



Wahrscheinlichkeitstheorie und Frequentistische Inferenz

BSc Psychologie WiSe 2021/22

Prof. Dr. Dirk Ostwald

(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



Methodenlehre I: Experimentelle und Neurowissenschaftliche Psychologie



Forschung



Lehre



Kontakt

Abteilungsleitung

> [Prof. Dr. Dirk Ostwald](#)

dirk.ostwald@ovgu.de

Tel.: +49 391 67 57370

Abteilungsassistentz

> [Birgit Müller](#)

birgit.mueller@ovgu.de

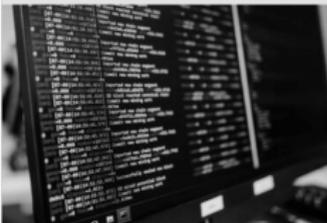
Tel.: +49 391 67 58464

Anschrift

Otto-von-Guericke-Universität
Magdeburg
Institut für Psychologie
Universitätsplatz 2
Gebäude 24
39106 Magdeburg

> [Anfahrt](#)

CBBS Imaging Plattform



Team



Wissenschaft und Psychologie

Datenwissenschaft und Statistik

Formalia

Studium und Diskussion

Selbstkontrollfragen

Wissenschaft und Psychologie

Datenwissenschaft und Statistik

Formalia

Studium und Diskussion

Selbstkontrollfragen

Wissenschaft

Wissenschaft bezeichnet den methodischen Prozess intersubjektiv nachvollziehbaren Forschens und Erkennens in einem bestimmten Bereich, der nach herkömmlichem Verständnis ein begründetes, geordnetes und gesichertes Wissen hervorbringt.

Wissenschaft kennzeichnet entsprechend das gesicherte und in einen rationalen Begründungszusammenhang gestellte Wissen, welches kommunizierbar und überprüfbar ist sowie bestimmten wissenschaftlichen Kriterien folgt.

Wissenschaft bezeichnet somit ein zusammenhängendes System von Aussagen, Theorien und Verfahrensweisen, das strengen Prüfungen der Geltung unterzogen wurde und mit dem Anspruch objektiver, überpersönlicher Gültigkeit verbunden ist.

Carrier, M. (2011) Lexikon der Philosophie, Reclam, Stuttgart, S. 312

Naturwissenschaften | Science

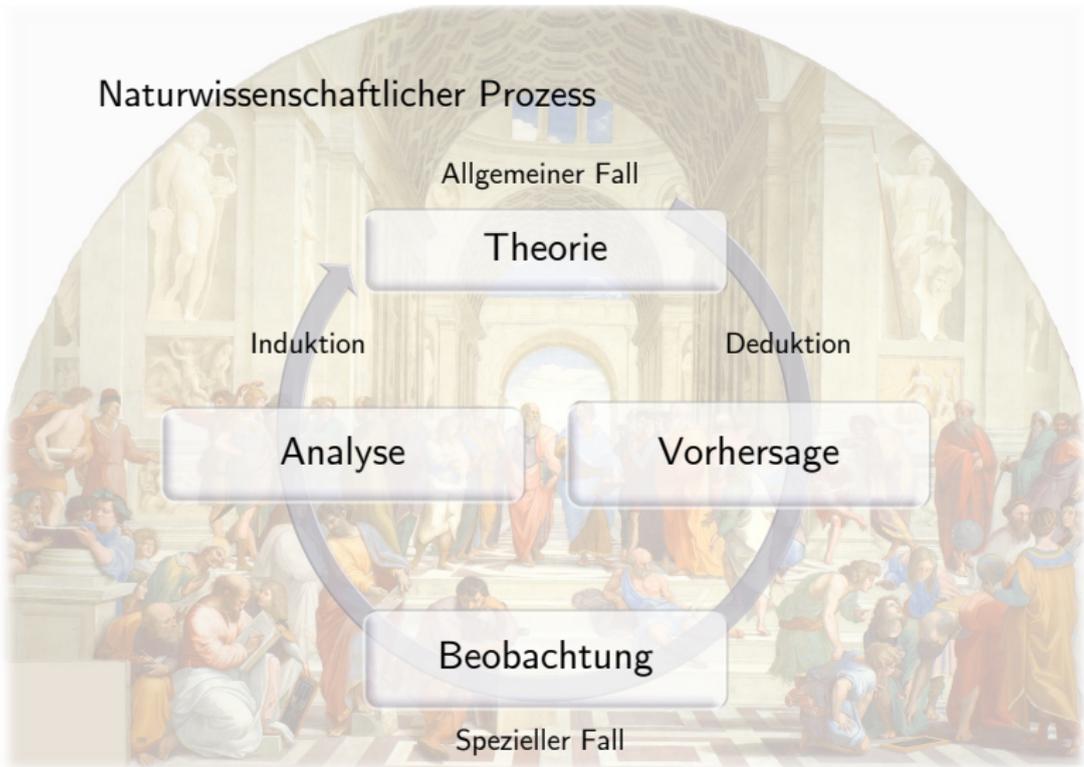
- Empirische Erforschung der Natur mit dem Ziel, Regelmäßigkeiten zu erkennen
- Quantitatives Beobachten, messen, analysieren des Verhaltens der Natur
- Grundlage zur Nutzbarmachung der Natur in den Ingenieursdisziplinen
- Physik, Chemie, Biologie, Medizin, **Psychologie**, Geologie, etc.

Geisteswissenschaften | Humanities

- Analytische Erforschung menschlicher Kultur
- Qualitative Sinnsuche, informelle Kritik, Spekulation
- Philosophie, Theologie, Geschichtswissenschaft, Literaturwissenschaft, etc.
- Naturwissenschaftliche Theoriebildung

Formalwissenschaften | Formal Sciences

- Analyse formaler System
- Sprachwerkzeuge
- Mathematik, Logik, theoretische Informatik, Rechtswissenschaft, etc.
- Naturwissenschaftliche Theoriebildung



Terminologie wissenschaftlicher Prozesse

Experiment

Kontrollierter Test einer wissenschaftlichen Theorie

Experimentelle Variablen

Unabhängige Variable \Leftrightarrow Experimentelle Manipulation

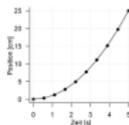
Abhängige Variable \Leftrightarrow Ergebnismaße der Manipulation



Isaac Newton
1689

$$\ddot{x}(t) = \frac{F}{m}, F = -mg$$

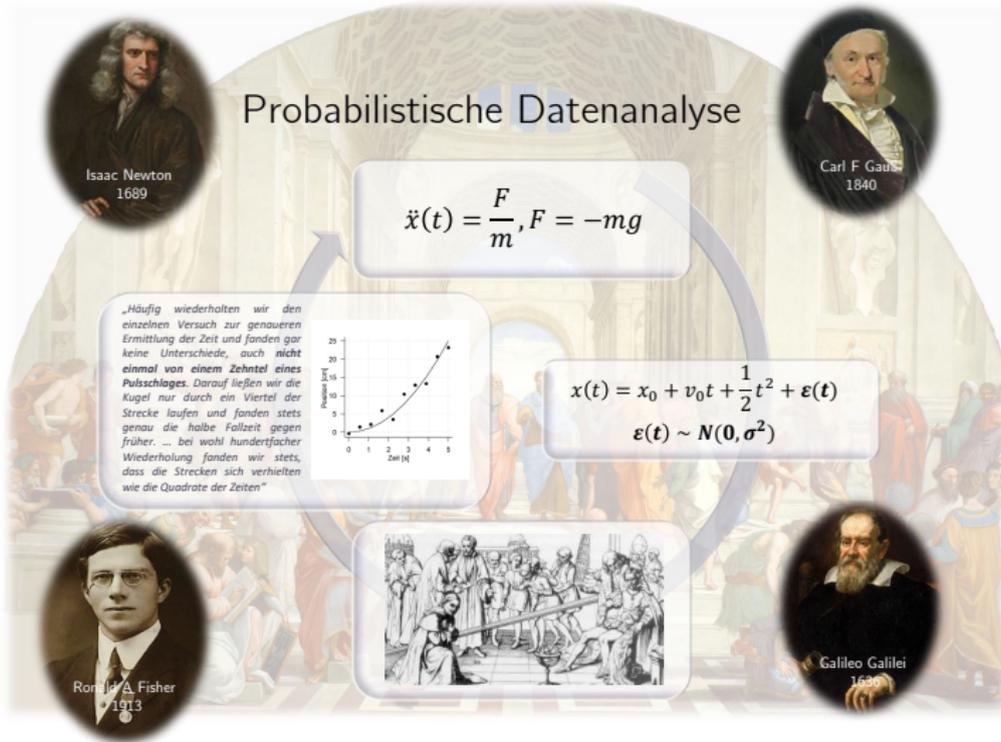
„Häufig wiederholten wir den einzelnen Versuch zur genaueren Ermittlung der Zeit und fanden gar keine Unterschiede, auch nicht einmal von einem Zehntel eines Pulsschlages. Darauf ließen wir die Kugel nur durch ein Viertel der Strecke laufen und fanden stets genau die halbe Fallzeit gegen früher. ... bei wohl hundertfacher Wiederholung fanden wir stets, dass die Strecken sich verhielten wie die Quadrate der Zeiten“



$$x(t) = x_0 + v_0 t + \frac{1}{2} t^2$$



Galileo Galilei
1639



Psychologie

Wissenschaft des menschlichen Erlebens, Verhaltens und Handelns

Beschreiben

- Benennen und Klassifizieren neuropsychologischer Phänomene

Erklären

- Entwicklung mechanistischer neuropsychologischer Modelle

Vorhersagen

- Prognose zukünftigen Erlebens, Verhaltens und Handelns

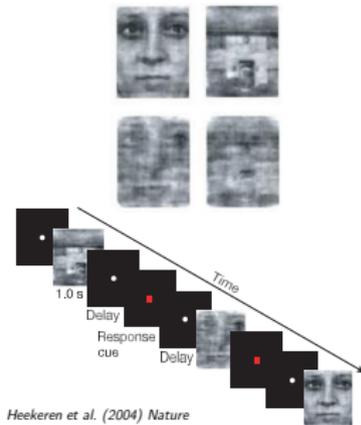
Verändern

- Prävention, Diagnose, Behandlung psychiatrischer Erkrankungen

Neurobiologische Verarbeitung von Sinnesreizen

Wie werden visuelle Stimuli im Gehirn verarbeitet?

Wie entscheiden Menschen, ob sie ein Haus oder ein Gesicht wahrnehmen?



→ Allgemeine Psychologie, Biologische Psychologie, Kognitive Neurowissenschaften

Neurobiologische Verarbeitung von Sinnesreizen - Verhaltensdaten

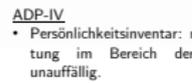
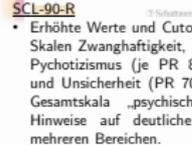
Trial	Kohärenz	Stimulus	Antwort	Reaktionszeit (ms)
1	Gesicht	60	Gesicht	479
2	Gesicht	20	Gesicht	483
3	Haus	40	Haus	321
4	Gesicht	40	Gesicht	369
5	Gesicht	60	Gesicht	478
6	Haus	60	Haus	836
7	Gesicht	40	Gesicht	794
8	Gesicht	60	Haus	797
9	Gesicht	20	Gesicht	464
10	Haus	20	Haus	563
11	Haus	20	Gesicht	590
12	Gesicht	40	Haus	429
13	Gesicht	40	Haus	751
14	Gesicht	60	Gesicht	876
15	Gesicht	60	Haus	851
16	Haus	40	Gesicht	492

Neurobiologische Verarbeitung von Sinnesreizen - Neurophysiologiedaten

ms	E1 (O1)	E2 (O2)	E3 (Cz)	E4 (Pz)	E5 (AF1)	E6 (AF2)
0	-0.682	0.404	-0.738	-1.82	0.04	0.423
2	17.814	18.271	17.445	17.90	14.71	15.381
4	18.101	18.387	17.453	17.37	17.86	18.888
6	3.695	2.148	4.356	4.39	3.29	2.917
8	-15.927	-13.785	-14.493	-14.38	-14.44	-17.026
10	-18.736	-16.447	-19.397	-18.02	-19.89	-20.089
12	-5.155	-6.549	-4.901	-3.42	-5.80	-4.349
14	14.131	11.346	11.808	13.87	13.49	13.435
16	20.931	19.074	18.354	20.50	21.53	17.981
18	9.441	9.169	6.949	7.12	6.60	6.892
20	-10.897	-11.638	-11.607	-10.51	-11.66	-11.150
22	-19.245	-20.782	-20.979	-19.59	-18.46	-21.143
24	-10.372	-11.178	-11.150	-8.66	-10.04	-11.485
26	8.590	7.720	9.665	9.57	8.52	7.609
28	21.212	21.199	18.541	19.04	19.53	20.762
30	12.311	13.931	13.714	12.70	13.51	12.338
32	-5.040	-5.536	-2.416	-5.49	-4.14	-7.970

Psychologische Diagnostik im psychotherapeutischen Erstgespräch

Ist eine Psychotherapie indiziert?



- Summe 47, Hinweis auf schwere depressive Episode.
- SCL-90-R**
 - Erhöhte Werte und Cutoff-Überschreitungen auf den Skalen Zwanghaftigkeit, Depressivität, Ängstlichkeit, Psychotizismus (je PR 80), Aggressivität (PR 73) und Unsicherheit (PR 70). Deutliche Belastung auf Gesamtskala „psychische Belastung“ (PR 80). Hinweise auf deutliche psychische Belastung in mehreren Bereichen.
- ADP-IV**
 - Persönlichkeitsinventar: minimale Cutoff-Überschreitung im Bereich der Paranoiden PS, sonst unauffällig.

→ Diagnostik, Persönlichkeitspsychologie, Verfahrenlehre, MSc Psychotherapie

Psychologische Diagnostik im psychotherapeutischen Erstgespräch - Verhaltensdaten

Item	Score
1	2
2	1
3	3
4	3
5	2
6	1
7	4
8	1
9	4
10	2
11	3
12	0
13	4
14	0
15	4
16	1

Evidenzbasierte Evaluation von Psychotherapieformen bei Depression

Welche Therapieform ist bei Depression wirksamer?

Online Psychotherapie



Klassische Psychotherapie



→ Klinische Psychologie, Klinische Diagnostik, MSc Psychotherapie

Evidenzbasierte Evaluation von Psychotherapieformen bei Depression

Becks Depressions-Inventar (BDI) zur Depressionsdiagnostik

BDI-II Fragebogen		Wahr	Schwach	Stark
		0-2		3-4
<p>Anleitung: Dieser Fragebogen enthält 21 Gruppen von Aussagen. Bitte lesen Sie jede dieser Gruppen von Aussagen sorgfältig durch und wählen Sie sich dann in jeder Gruppe eine Aussage heraus, die am besten beschreibt, wie Sie sich in der letzten zwei Wochen, einschließlich heute, gefühlt haben. Konzentrieren Sie die Aufmerksamkeit auf die Aussage an, die Sie sich herausgehoben haben (0, 1, 2 oder 3). Falls in einer Gruppe mehrere Aussagen gleichwertig auf Sie zutreffen, können Sie für Aussage mit der höchsten Zahl ein „Achten Sie bitte darauf, dass Sie in jeder Gruppe nicht mehr als eine Aussage ankreuzen, die die höchste Zahl (Gruppe II (Veränderungen der Schlafgewohnheiten) oder Gruppe 16 (Veränderungen des Appetits)).“</p>				
<p>1.) Traurigkeit</p> <p>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.</p>				
<p>2.) pessimismus</p> <p>0 Ich sehe nicht mutlos in die Zukunft. 1 Ich sehe mutlos in die Zukunft als Ganzes. 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.</p>				
<p>3.) Versagensgefühle</p> <p>0 Ich fühle mich nicht als Versager. 1 Ich habe häufiger Versagensgefühle. 2 Wenn ich zurückblicke, sehe ich eine Menge Fehlentscheidungen. 3 Ich habe das Gefühl, ich mache ein völliger Versager zu sein.</p>				
<p>4.) Verlust von Freude</p> <p>0 Ich kann die Dinge genauso gut genießen wie früher. 1 Ich kann die Dinge nicht mehr so genießen wie früher. 2 Dinge, die mir früher Freude gemacht haben, kann ich kaum mehr genießen. 3 Dinge, die mir früher Freude gemacht haben, kann ich überhaupt nicht mehr genießen.</p>				
<p>5.) Schuldgefühle</p> <p>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 meisten Zeit Schuldgefühle. 3 Ich habe ständig Schuldgefühle.</p>				
<p>6.) Bestürzungsgefühle</p> <p>0 Ich habe keine 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.</p>				
<p>7.) Selbstbehinderung</p> <p>0 Ich hätte von mir genauso viel wie immer. 1 Ich habe Vertrauen in mich verloren. 2 Ich vermute von mir enttäuscht. 3 Ich lehne mich völlig ab.</p>				
<p>8.) Selbstvorwürfe</p> <p>0 Ich kritisiere oder tadle mich nicht mehr als sonst. 1 Ich bin mir gegenüber kritischer als sonst. 2 Ich kritisiere mich für all meine Mängel. 3 Ich gebe mir die Schuld für alles Schlechte, was passiert.</p>				
<p>9.) Selbstmordgedanken</p> <p>0 Ich denke nicht daran, mir etwas anzutun. 1 Ich denke manchmal an Selbstmord, aber ich würde es nicht tun. 2 Ich möchte mich ein bisschen verletzen, um etwas mich umzubringen, wenn ich die Gelegenheit dazu hätte. 3 Ich würde mich umbringen, wenn ich die Gelegenheit dazu hätte.</p>				
<p>10.) Weinen</p> <p>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.</p>				

<p>11.) Unruhe</p> <p>0 Ich bin nicht unruhiger als sonst. 1 Ich bin unruhiger als sonst. 2 Ich bin so unruhig, dass es mir schwerfällt, still zu sitzen. 3 Ich bin so unruhig, dass ich mich ständig bewegen oder etwas tun muss.</p>	
<p>12.) Interessensverlust</p> <p>0 Ich habe das Interesse an anderen Menschen oder an Tätigkeiten nicht verloren. 1 Ich habe weniger Interesse an anderen Menschen oder an Dingen als sonst. 2 Ich habe das Interesse an anderen Menschen oder Dingen zum größten Teil verloren. 3 Es fällt mir schwer, mich überhaupt für irgend etwas zu interessieren.</p>	
<p>13.) Entscheidungsfähigkeit</p> <p>0 Ich bin so entscheidungsfähig wie immer. 1 Es fällt mir schwerer als sonst, Entscheidungen zu treffen. 2 Es fällt mir sehr viel schwerer als sonst, Entscheidungen zu treffen. 3 Ich habe Mühe, überhaupt Entscheidungen zu treffen.</p>	
<p>14.) Wertlosigkeit</p> <p>0 Ich fühle mich nicht wertlos. 1 Ich fühle mich für weniger wertvoll und nützlich als sonst. 2 Vergleichen mit anderen Menschen fühle ich mich viel weniger wert. 3 Ich fühle mich völlig wertlos.</p>	
<p>15.) Energieverlust</p> <p>0 Ich habe so viel Energie wie immer. 1 Ich habe weniger Energie als sonst. 2 Ich habe so wenig Energie, dass ich kaum noch etwas schaffe. 3 Ich habe keine Energie mehr, um überhaupt noch etwas zu tun.</p>	
<p>16.) Veränderungen der Schlafgewohnheiten</p> <p>0 Meine Schlafgewohnheiten haben sich nicht verändert. 1 Ich schlafe etwas mehr als sonst. 2 Ich schlafe etwas weniger als sonst. 3 Ich schlafe viel mehr als sonst. 4 Ich schlafe viel weniger als sonst. 5 Ich schlafe fast den ganzen Tag. 6 Ich wache 1-2 Stunden früher auf als gewöhnlich und kann dann nicht mehr einschlafen.</p>	
<p>17.) Reizbarkeit</p> <p>0 Ich bin nicht reizbarer als sonst. 1 Ich bin reizbarer als sonst. 2 Ich bin viel reizbarer als sonst. 3 Ich fühle mich dauernd gereizt.</p>	
<p>18.) Veränderungen des Appetits</p> <p>0 Mein Appetit hat sich nicht verändert. 1a Mein Appetit ist etwas schlechter als sonst. 1b Mein Appetit ist etwas größer als sonst. 2a Mein Appetit ist viel schlechter als sonst. 2b Mein Appetit ist viel größer als sonst. 3a Ich habe überhaupt keinen Appetit. 3b Ich habe ständig Heißhunger.</p>	
<p>19.) Konzentrationschwierigkeiten</p> <p>0 Ich kann mich so gut konzentrieren wie immer. 1 Ich kann mich nicht mehr so gut konzentrieren wie sonst. 2 Es fällt mir schwer, mich längere Zeit auf irgend etwas zu konzentrieren. 3 Ich kann mich überhaupt nicht mehr konzentrieren.</p>	
<p>20.) Ermüdung oder Erschöpfung</p> <p>0 Ich fühle mich nicht müde oder erschöpfter als sonst. 1 Ich werde schneller müde oder erschöpfter als sonst. 2 Für viele Dinge, die ich üblicherweise tue, bin ich zu müde oder erschöpft. 3 Ich bin so müde oder erschöpft, dass ich fast nichts mehr tun kann.</p>	
<p>21.) Verlust an sexuellem Interesse</p> <p>0 Mein Interesse an Sexualität hat sich in letzter Zeit nicht verändert. 1 Ich interessiere mich weniger für Sexualität als früher. 2 Ich interessiere mich jetzt viel weniger für Sexualität. 3 Ich habe das Interesse an Sexualität völlig verloren.</p>	
<p>Summe Werte 1-10: <input type="text"/></p>	
<p>Übersicht Werte 11-21: <input type="text"/></p>	

0 - 8 keine Depression

9 - 13 minimale Depression

14 - 19 leichte Depression

20 - 28 mittelschwere Depression

29 - 63 schwere Depression

Evidenzbasierte Evaluation von Psychotherapieformen bei Depression - Verhaltensdaten

Patient ID	Bedingung	Prae-PT-BDI	Post-PT-BDI
1	Online	32	29
2	Online	41	30
3	Online	38	15
4	Online	48	16
5	Klassisch	34	16
6	Klassisch	19	15
7	Online	26	14
8	Online	37	21
9	Online	34	17
10	Online	18	20
11	Online	33	25
12	Klassisch	35	20
13	Klassisch	36	11
14	Klassisch	35	19
15	Online	33	20
16	Online	28	17

Wissenschaft und Psychologie

Datenwissenschaft und Statistik

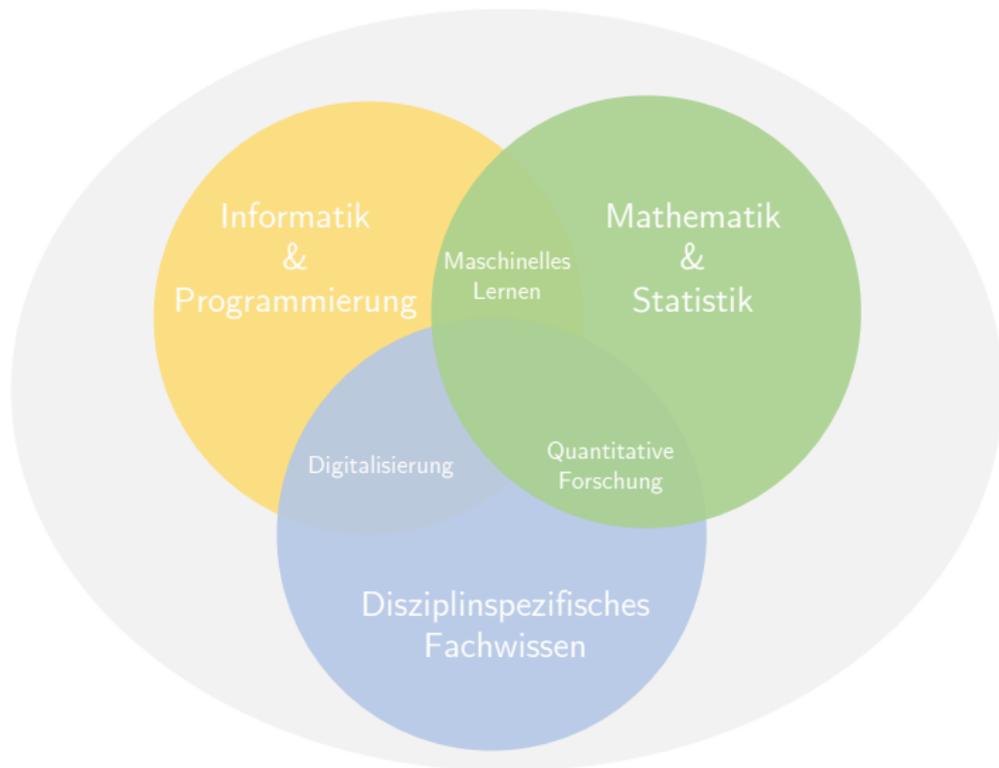
Formalia

Studium und Diskussion

Selbstkontrollfragen

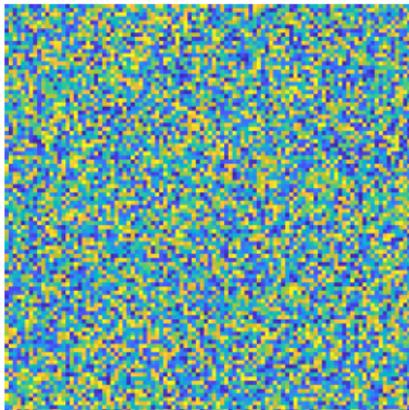
Datenwissenschaft

Die Kunst, aus Daten Sinn zu generieren

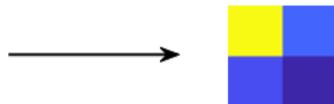


Datenwissenschaft ist Datenreduktion

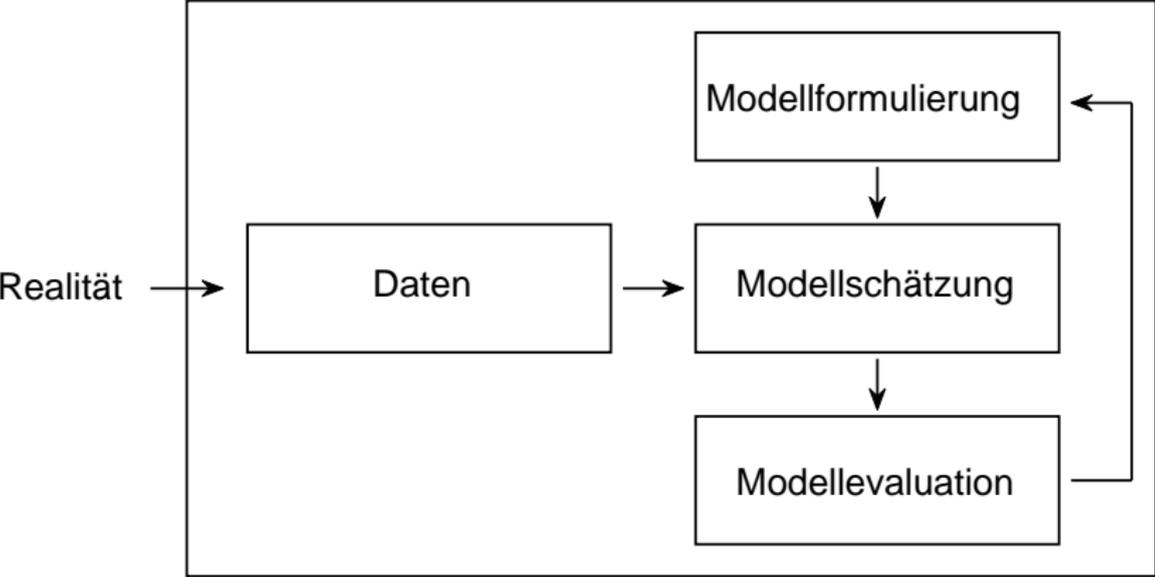
Rohdaten



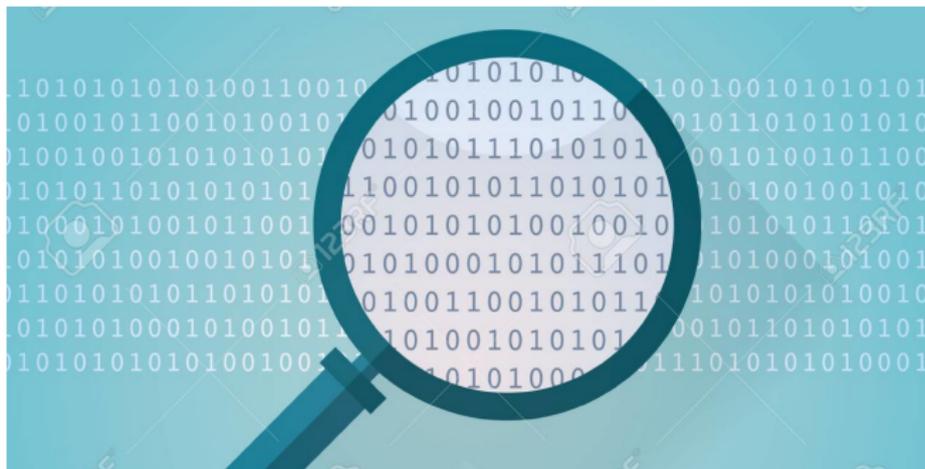
Reduzierte Daten



Datenwissenschaft ist Naturwissenschaft



Datenwissenschaft 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

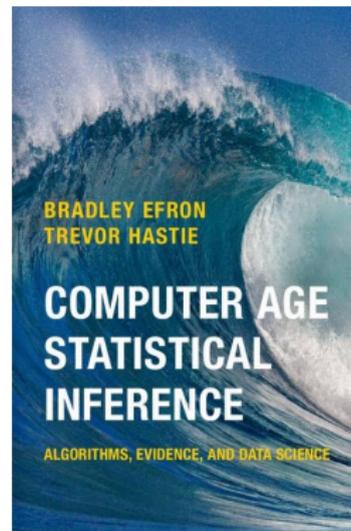
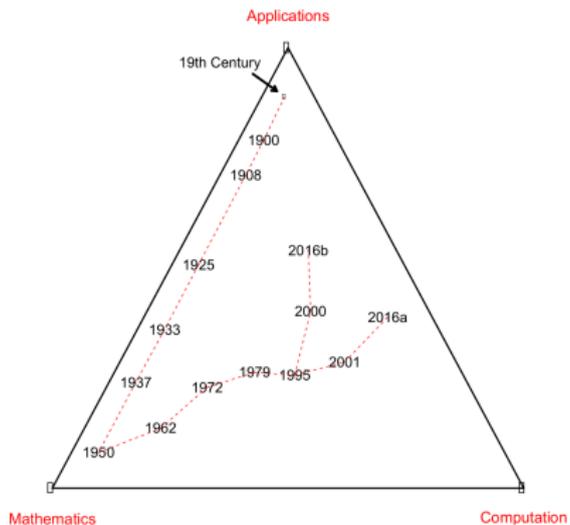
Datenwissenschaft in der Psychologie

Die Kunst, aus Verhaltens- und Neurophysiologiedaten
psychologischen Sinn zu generieren

Statistik

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





Statistik in der Psychologie

Die Kunst, aus Verhaltens- und Neurophysiologiedaten
psychologischen Sinn zu generieren
und seine assoziierte Unsicherheit zu quantifizieren

Klassische Partition der Statistik in der Psychologie

- Deskriptivstatistik
- Inferenzstatistik
- Multivariate Statistik

Sinnvolle Partition der psychologischen Datenwissenschaft

- Wahrscheinlichkeitstheoretische Grundlagen
- Frequentistische Inferenz und 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.

Fundamentale Annahmen der Frequentistischen Statistik

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

Fundamentale Annahmen der Bayesianischen Statistik

- Wahrscheinlichkeiten werden als Grade der Sicherheit, nicht als langfristige relative Häufigkeiten interpretiert. Aussagen der Form “Die Wahrscheinlichkeit, dass das Wintersemester 2021/22 vollständig in Präsenzlehre stattfindet, ist 0.8.” 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.

Datenwissenschaftliches Curriculum der OVGU Psychologie

Grundlagen der Mathematik und Informatik		
Semester	Inhalt	Thema
1	Mathematik	Algebra
2	Mathematik	Summen, Produkte, Rekursion
3	Mathematik	Folgen, Reihen, Grenzwerte
4	Mathematik	Funktionen
5	Mathematik	Differentialrechnung
6	Mathematik	Integralrechnung
7	Informatik	Grundbegriffe der Informatik

Blk 51		B1 Wahrscheinlichkeitstheorie und Frequentistische Inferenz 2 SWS	G2 Programmierung und Ökologische Statistik 2 SWS	
Woche	Inhalt	Thema	Inhalt	Thema
1	Einführung	Einführung	Einführung	Thema
2	Wahrscheinlichkeitstheorie	Wahrscheinlichkeitsräume	Grundlagen	8 und 8/10/10
3	Wahrscheinlichkeitstheorie	Zufallsvariablen	Grundlagen	Verteilungen
4	Wahrscheinlichkeitstheorie	Zufallsvariablen	Grundlagen	Matrizen und Arrays
5	Wahrscheinlichkeitstheorie	Erwartungswert und Kovarianz	Grundlagen	Listen und DataFrames
6	Wahrscheinlichkeitstheorie	Ungleichungen und Grenzwerte	Grundlagen	Kontrollstruktur und Schleifen
7	Wahrscheinlichkeitstheorie	Transformations der Normalverteilung	Grundlagen	Erweiteren und Scope
8	Frequentistische Inferenz	Statistische Modelle, Statistiken, Schätzer	Grundlagen	8 Base Graphics
9	Frequentistische Inferenz	Schätzergenschaften	Deskriptive Statistik	Strukturmanagement
10	Frequentistische Inferenz	Spezialstatistiken und Konfidenzintervalle	Deskriptive Statistik	Werkzeuge/Verfahrenen
11	Frequentistische Inferenz	Aggregiertest	Deskriptive Statistik	Verteilungsfunktionen und Quantile
12	Frequentistische Inferenz	T-Test	Deskriptive Statistik	Maße der zentralen Tendenz
13	Frequentistische Inferenz	Kovariationale Varianzanalyse	Deskriptive Statistik	Maße der Streuung/Varianz
14	Frequentistische Inferenz	Zusätzliche Varianzanalyse	Deskriptive Statistik	Lineare Zusammenhänge

Blk 52		B2 Allgemeines Lineares Modell und Bayesische Inferenz 2 SWS	A2 Forschungsmethoden 2 SWS	
Woche	Inhalt	Thema	Inhalt	Thema
1	Allgemeines Lineares Modell	Levenberg und Matrizen	Grundlagen	Quantitative Verfahren
2	Allgemeines Lineares Modell	Matrixale Normalverteilungen	Grundlagen	Quantitative Verfahren
3	Allgemeines Lineares Modell	Einfache lineare Regression und ANM	Studienfrage	Quantitative Verfahren
4	Allgemeines Lineares Modell	Parametererschätzung und Verteilungstheorie	Studienfrage	Qualitative Verfahren
5	Allgemeines Lineares Modell	Aggregiertest	Studienfrage	Qualitative Verfahren
6	Allgemeines Lineares Modell	Fundamentale Design I	Studienfrage	Qualitative Verfahren
7	Allgemeines Lineares Modell	Fundamentale Design II	Methodologie	Grundlagen
8	Allgemeines Lineares Modell	Multiple Regression und Partialle Korrelation	Methodologie	Grundlagen
9	Allgemeines Lineares Modell	Hierarchische Lineare Modelle	Methodologie	Grundlagen
10	Bayesische Inferenz	Grundlagen	Methodologie	Grundlagen
11	Bayesische Inferenz	Statistische Modelle	Methodologie	Grundlagen
12	Bayesische Inferenz	Allgemeines Lineares Modell	Methodologie	Grundlagen
13	Bayesische Inferenz	Bayesische Inferenz	Methodologie	Grundlagen
14	Bayesische Inferenz	Variationale Inferenz	Methodologie	Grundlagen

Blk 53		A1 und A3 Multivariate Datenanalyse 4 SWS
Woche	Inhalt	Thema
1	Grundlagen	Levenberg
2	Grundlagen	Matrizen
3	Grundlagen	Aggregiertest
4	Grundlagen	Matrixale Normalverteilungen
5	Achtensatzinformationen	Hauptkomponentenanalyse
6	Achtensatzinformationen	Faktoranalyse
7	Methoden der Frequentistischen Inferenz	Kovariante Korrelation
8	Methoden der Frequentistischen Inferenz	T-Test
9	Methoden der Frequentistischen Inferenz	Kovariante Varianzanalyse
10	Methoden der Frequentistischen Inferenz	Allgemeines Lineares Modell
11	Methoden des Multivariaten Lernens	Statistische Lerntheorie
12	Methoden des Multivariaten Lernens	ÜB und logistische Regression
13	Methoden des Multivariaten Lernens	Support Vector Machines
14	Methoden des Multivariaten Lernens	Neuronale Netze

Wissenschaft und Psychologie

Datenwissenschaft und Statistik

Formalia

Studium und Diskussion

Selbstkontrollfragen

Wintersemester 2021/2022

- Modul B1 Deskriptive Statistik
- Modul C2 Computergestützte Datenanalyse

Sommersemester 2022

- Modul B2 Inferenzstatistik
- Modul A2 Forschungsmethoden

Wintersemester 2021/2022

- **Modul B1 Deskriptive Statistik**
- Modul C2 Computergestützte Datenanalyse

Sommersemester 2022

- Modul B2 Inferenzstatistik
- Modul A2 Forschungsmethoden

Modul B1 Deskriptive Statistik | Wahrscheinlichkeitstheorie und Frequentistische Inferenz

- Donnerstags 13 - 15 Uhr, G40B-231
- Kursmaterialien (Folien, Videos) auf der **Kurswebseite**
- Ankündigungen über die **Moodleseite**
- Benotete Multiple Choice Klausur (30 Fragen) Ende Wintersemester 2021/22
- Klausurwiederholungstermin am Ende des Sommersemesters 2022
- Klausurtermin und Klausurort gemäß Prüfungsplan des **FNW Prüfungsamtes**

Modul B1 Deskriptive Statistik | Wahrscheinlichkeitstheorie und Frequentistische Inferenz

Datum	Einheit	Thema
14.10.2021	Einführung	(1) Einführung
21.10.2021	Wahrscheinlichkeitstheorie	(2) Wahrscheinlichkeitsräume
28.10.2021	Wahrscheinlichkeitstheorie	(3) Zufallsvariablen
04.11.2021	Wahrscheinlichkeitstheorie	(4) Zufallsvektoren
11.11.2021	Wahrscheinlichkeitstheorie	(5) Erwartungswert und Kovarianz
18.11.2021	Wahrscheinlichkeitstheorie	(6) Ungleichungen und Grenzwerte
25.11.2021	Wahrscheinlichkeitstheorie	(7) Normalverteilungstransformationen
02.12.2021	Frequentistische Inferenz	(8) Statistische Modelle, Statistiken, Schätzer
09.12.2021	Frequentistische Inferenz	(9) Schätzeigenschaften
16.12.2021	Frequentistische Inferenz	(10) Konfidenzintervalle
	Weihnachtspause	
06.01.2022	Frequentistische Inferenz	(11) Hypothesentests
13.01.2022	Frequentistische Inferenz	(12) T-Tests
20.01.2022	Frequentistische Inferenz	(13) Einfaktorielle Varianzanalyse
27.01.2022	Frequentistische Inferenz	(14) Zweifaktorielle Varianzanalyse
Feb 2022	Klausurtermin	
Jul 2022	Klausurwiederholungstermin	

Modul B1 Deskriptive Statistik | Wahrscheinlichkeitstheorie und Frequentistische Inferenz

- Vorlesungsfolien inklusive Selbstkontrollfragen sind klausurrelevant
- Altklausuren finden sich auf den Kurswebseiten früherer Jahre
- Empfohlen ist die Lektüre der entsprechenden Abschnitte in

Probabilistische Datenanalyse für die Wissenschaftliche Psychologie

- Als weiterführende Literatur bieten sich an
 - Georgii, H.O. (2015) Stochastik
 - Fahrmeir, L., Heumann, C., Künstler, R., Pigeot, I., Tutz, G. (2016) Statistik
 - DeGroot, M.H. & Shervish, M.J. (2012) Probability and Statistics
 - Casella, G. & Berger, R.L. (2012) Statistical Inference

Wintersemester 2021/2022

- Modul B1 Deskriptive Statistik
- **Modul C2 Computergestützte Datenanalyse**

Sommersemester 2022

- Modul B2 Inferenzstatistik
- **Modul A2 Forschungsmethoden**

Modul C2 Computergestützte Datenanalyse | Programmierung und Deskriptive Statistik

- Donnerstags 15 - 17 Uhr, G40B-231
- Kursmaterialien (Folien, Videos) auf der [Kurswebseite](#)
- Ankündigungen über die [Moodleseite](#)
- Benotete Multiple Choice Klausur (20 Fragen) Ende Wintersemester 2021/22
- Klausurwiederholungstermin am Ende des Sommersemesters 2022
- Klausurtermin und Klausurort gemäß Prüfungsplan des [FNW Prüfungsamtes](#)

Modul C2 Computergestützte Datenanalyse | Programmierung und Deskriptive Statistik

Datum	Einheit	Thema
14.10.2021	Einführung	(1) Einführung
21.10.2021	R Grundlagen	(2) R und RStudio
28.10.2021	R Grundlagen	(3) Vektoren
04.11.2021	R Grundlagen	(4) Matrizen und Arrays
11.11.2021	R Grundlagen	(5) Listen und Dataframes
18.11.2021	R Grundlagen	(6) Kontrollstruktur und Schleifen
25.11.2021	R Grundlagen	(7) Funktionen und Scope
02.12.2021	R Grundlagen	(8) R Base Graphics
09.12.2021	Deskriptive Statistik	(9) Datenmanagement
16.12.2021	Deskriptive Statistik	(10) Häufigkeitsverteilungen
	Weihnachtspause	
06.01.2022	Deskriptive Statistik	(11) Verteilungsfunktionen und Quantile
13.01.2022	Deskriptive Statistik	(12) Maße der zentralen Tendenz
20.01.2022	Deskriptive Statistik	(13) Maße der Datenvariabilität
27.01.2022	Deskriptive Statistik	(14) Bivariate Zusammenhangsmaße
Feb 2022	Klausurtermin	
Jul 2022	Klausurwiederholungstermin	

Modul C2 Computergestützte Datenanalyse | Programmierung und Deskriptive Statistik

- Vorlesungsfolien inklusive Selbstkontrollfragen sind klausurrelevant
- Altklausuren finden sich auf den Kurswebseiten früherer Jahre
- Empfohlen ist die Lektüre der entsprechenden Abschnitte in

Probabilistische Datenanalyse für die Wissenschaftliche Psychologie

- Als weiterführende Literatur bieten sich an
 - Sauer, S. (2019) Moderne Datenanalyse mit R
 - Wickham, H. (2019) Advanced R
 - Cotton, R. (2013) Learning R.
 - Murrell, P. (2021) R Graphics

Wissenschaft und Psychologie

Datenwissenschaft und Statistik

Formalia

Studium und Diskussion

Selbstkontrollfragen

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, (...)

⇒ Bachelorarbeit

Zur Vermittlung 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.

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, (...)

⇒ Masterarbeit

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

- a) multivariate Verfahren und Messtheorie

Q & A

Wissenschaft und Psychologie

Datenwissenschaft und Statistik

Formalia

Studium und Diskussion

Selbstkontrollfragen

Selbstkontrollfragen

1. Diskutieren Sie die Begriffe Naturwissenschaft, Geisteswissenschaft und Formalwissenschaft
2. Beschreiben Sie den naturwissenschaftlichen Prozess.
3. Definieren Sie die Begriffe Experiment, unabhängige Variable und abhängige Variable.
4. Nennen Sie vier Ziele der wissenschaftlichen Psychologie.
5. Beschreiben Sie Beispiele für die Anwendung wissenschaftlicher Methoden in der Psychologie.
6. Nennen Sie die drei Hauptkomponenten der Datenwissenschaft.
7. Nennen Sie drei Grundannahmen der Wahrscheinlichkeitstheorie.
8. Nennen Sie drei Grundannahmen der frequentistischen Statistik.
9. Nennen Sie drei Grundannahmen der Bayesianischen Statistik.



Wahrscheinlichkeitstheorie und Frequentistische Inferenz

BSc Psychologie WiSe 2021/22

Prof. Dr. Dirk Ostwald

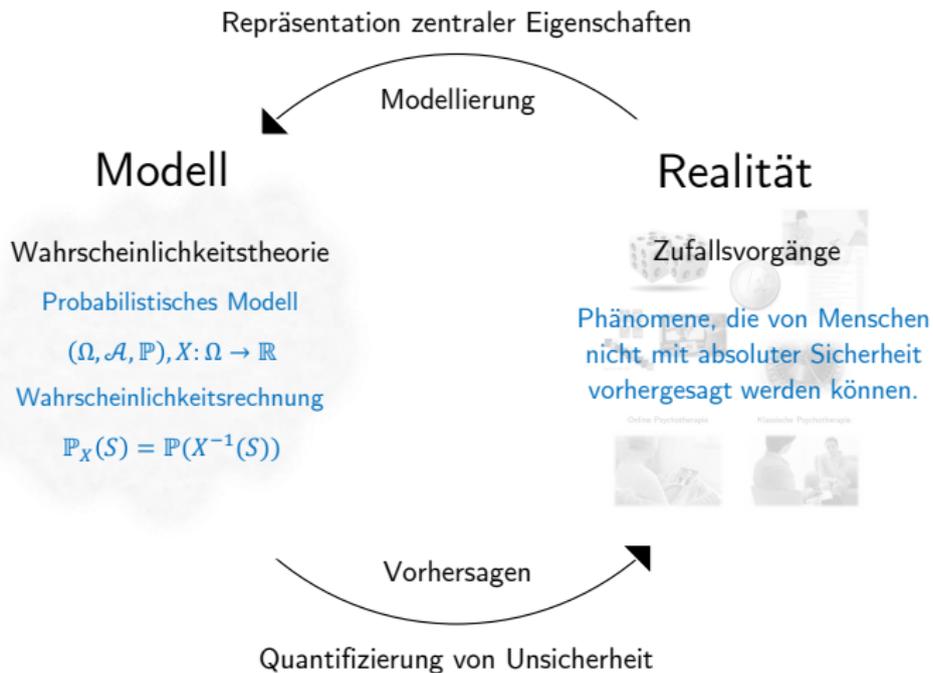
(2) Wahrscheinlichkeitsräume

Wahrscheinlichkeitstheorie und Frequentistische Inferenz

Datum	Einheit	Thema
14.10.2021	Einführung	(1) Einführung
21.10.2021	Wahrscheinlichkeitstheorie	(2) Wahrscheinlichkeitsräume
28.10.2021	Wahrscheinlichkeitstheorie	(3) Elementare Wahrscheinlichkeiten
04.11.2021	Wahrscheinlichkeitstheorie	(4) Zufallsvariablen
11.11.2021	Wahrscheinlichkeitstheorie	(5) Zufallsvektoren
18.11.2021	Wahrscheinlichkeitstheorie	(6) Erwartungswert und Kovarianz
25.11.2021	Wahrscheinlichkeitstheorie	(7) Ungleichungen und Grenzwerte
02.12.2021	Wahrscheinlichkeitstheorie	(8) Normalverteilungstransformationen
09.12.2021	Frequentistische Inferenz	(9) Modelle, Statistiken und Schätzer
16.12.2021	Frequentistische Inferenz	(10) Schätzereigenschaften
	Weihnachtspause	
06.01.2022	Frequentistische Inferenz	(11) Konfidenzintervalle
13.01.2022	Frequentistische Inferenz	(12) Hypothesentests
20.01.2022	Frequentistische Inferenz	(13) Einstichproben-T-Tests
27.01.2022	Frequentistische Inferenz	(14) Zweistichproben-T-Tests
Feb 2022	Klausurtermin	
Jul 2022	Klausurwiederholungstermin	

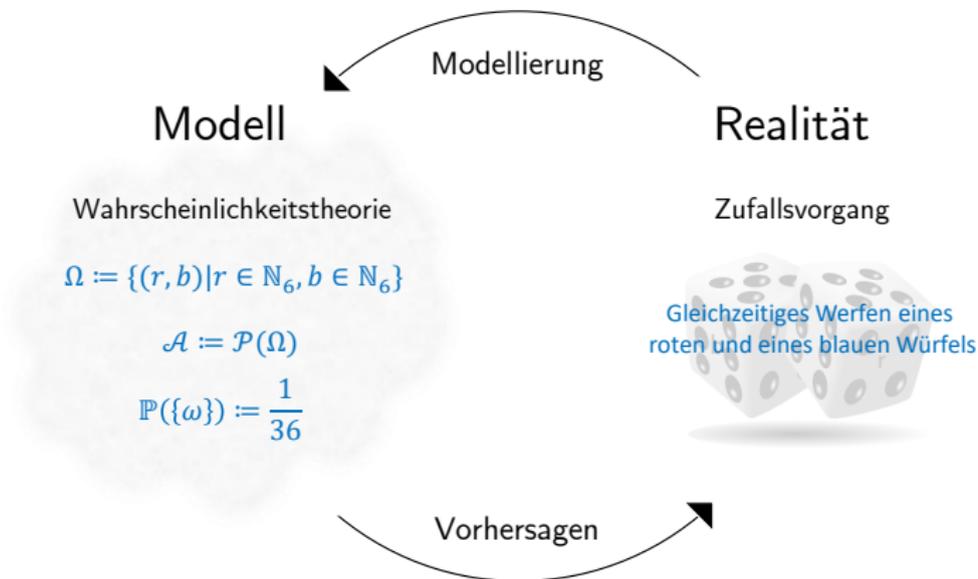
Statistik

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



Jedes Augenzahlpaar kommt im Mittel gleich häufig vor.

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



Die Wahrscheinlichkeit dafür, dass die Summe der gefallenen Augenzahlen 4 ist, ist $0.0\overline{83}$.

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 (Wahrscheinlichkeitsraum)

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

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

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

Bemerkung

- Die Definition benutzt den Begriff der σ -Algebra.

Definition (σ -Algebra)

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

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

ERGEBNISSE DER MATHEMATIK UND IHRER GRENZGEBIETE

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

GRUNDBEGRIFFE DER WAHRSCHEINLICHKEITS- RECHNUNG

VON
A. KOLMOGOROFF



BERLIN
VERLAG VON JULIUS SPRINGER
1933

„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]

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 $P(A)$ wenig unterscheiden wird.

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

1. A und B sind disjunkt, d. h. $AB = \emptyset$.
2. $AB \dots N = \emptyset$.
3. $AB \dots N = X$.
4. $A + B + \dots + N = X$.
5. Die Komplementmenge \bar{A} .
6. $A = \emptyset$.
7. $A = E$.

* Vgl. 1.1. Formel (3).

Im Falle der zufälligen Ereignisse.

1. Die Ereignisse A und B sind unvereinbar.
2. Die Ereignisse A, B, \dots, N sind unvereinbar.
3. Das Ereignis X besteht in der gleichzeitigen Realisation aller Ereignisse A, B, \dots, N .
4. Das Ereignis X besteht in der Entscheidung mindestens eines unter den Ereignissen A, B, \dots, N .
5. Das entgegengesetzte Ereignis \bar{A} besteht in der Nichtrealisation des Ereignisses A .
6. A ist unmöglich.
7. A muß notwendig vorkommen.

8. Ein System \mathcal{R} der Mengen A_1, A_2, \dots, A_n bildet eine Zerlegung der Menge E , wenn $A_1 + A_2 + \dots + A_n = E$ ist (das setzt bereits voraus, daß die Mengen A_i paarweise disjunkt sind).
9. B ist eine Untermenge von A : $B \subset A$.
8. Ein Versuch \mathcal{R} besteht darin, daß man feststellt, welches unter den Ereignissen A_1, A_2, \dots, A_n vorkommt. A_1, A_2, \dots, A_n sind die möglichen Ausgänge des Versuches \mathcal{R} .
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 $A + A = E$ und den Axiomen IV und V folgt

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

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

Da $E = 0$ ist, erhält man insbesondere

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

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

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

(der Additionssatz).

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

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

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

$$(6) \quad P(A|B) = P(A)P_A(B).$$

Ein Induktionschluß ergibt sodann die allgemeine Formel

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

(der Multiplikationssatz).

Man beweist auch leicht folgende Formeln:

$$(8) \quad P_A(B) \geq 0,$$

$$(9) \quad P_A(E) = 1,$$

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

Vergleicht man diese Formeln (8) bis (10) mit den Axiomen III bis V, so ergibt sich, daß das Mengensystem \mathcal{M} mit der Mengenfunktion $P_A(B)$

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.

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]

Ereignissystem \mathcal{A}

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

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

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

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

Theorem (Ereignissystem bei endlicher Ergebnismenge)

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

Beweis

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

Wahrscheinlichkeitsmaß \mathbb{P}

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

σ -Additivität des Wahrscheinlichkeitsmaßes \mathbb{P}

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

Probability Theory \Leftrightarrow Reasoning with Uncertainty

Definition

Erste Eigenschaften

Wahrscheinlichkeitsfunktionen

Beispiele

Selbstkontrollfragen

Theorem (Wahrscheinlichkeit des unmöglichen Ereignisses)

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

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

Beweis

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

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

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

□

Theorem (σ -Additivität bei endlichen Folgen disjunkter Ereignisse)

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

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

Beweis

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

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

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

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

□

Definition

Erste Eigenschaften

Wahrscheinlichkeitsfunktionen

Beispiele

Selbstkontrollfragen

Definition (Wahrscheinlichkeitsfunktion)

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

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

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

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

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

Bemerkungen

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

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

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

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

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

Bemerkung

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

Beweis

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

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

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

□.

Definition

Erste Eigenschaften

Wahrscheinlichkeitsfunktionen

Beispiele

Selbstkontrollfragen

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

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

Beispiele

Würfeln mit einem Würfel

Wir modellieren das 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 Werfens eines Würfels

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

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

Würfeln mit einem Würfel (fortgesetzt)

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

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

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

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

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

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

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

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

Beispiele

Gleichzeitiges Würfeln mit einem 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

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

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

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

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

Beschreibung	Mengenform
Auf dem blauen Würfel fällt eine Drei	$\omega \in A = \{(3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6)\}$
Auf dem roten 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	$\omega \in A = \{(1, 3), (3, 1), (2, 2)\}$

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

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

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

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

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

Werfen einer Münze

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

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

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 (Ω, \mathcal{A}) ist damit abgeschlossen.

Werfen einer Münze (fortgesetzt)

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

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

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

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

Beispiele

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 (Ω, \mathcal{A}) ist damit abgeschlossen. Ein Wahrscheinlichkeitsmaß \mathbb{P} kann wiederum durch Definition von $\mathbb{P}(\{\omega\})$ für alle $\omega \in \Omega$ festgelegt werden. Für das Modell zweier fairer Münzen könnte man etwa

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

wählen.

Definition

Erste Eigenschaften

Wahrscheinlichkeitsfunktionen

Beispiele

Selbstkontrollfragen

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 Zufallsvorgangs durch einen Wahrscheinlichkeitsraum 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>.

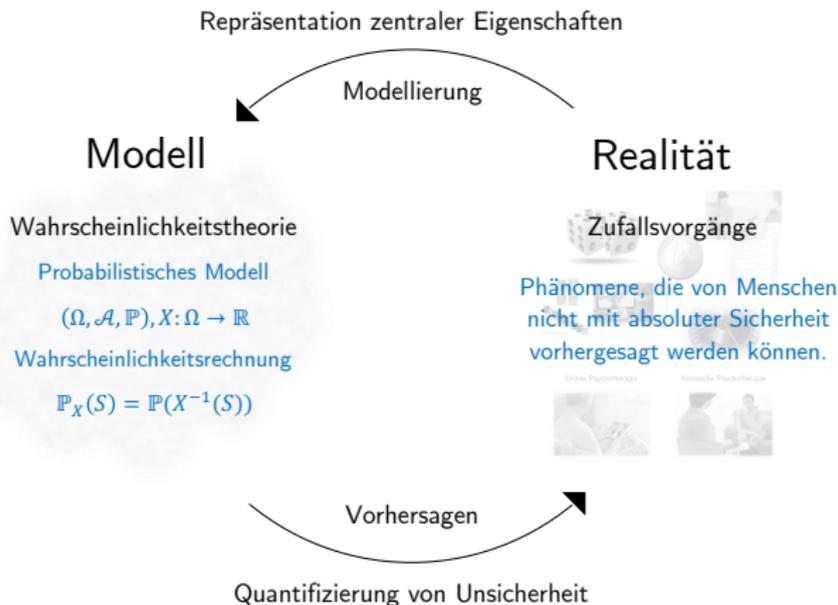


Wahrscheinlichkeitstheorie und Frequentistische Inferenz

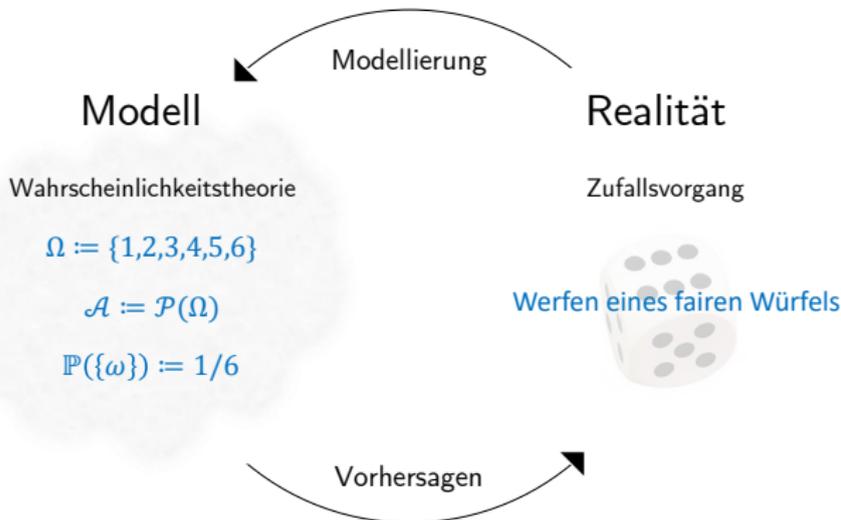
BSc Psychologie WiSe 2021/22

Prof. Dr. Dirk Ostwald

(3) Elementare Wahrscheinlichkeiten



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



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

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

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

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

mit den Eigenschaften

- $\mathbb{P}(A) \geq 0$ für alle $A \in \mathcal{A}$ (Nicht-Negativität),
- $\mathbb{P}(\Omega) = 1$ (Normiertheit),
- $\mathbb{P}(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$ 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* $A \in \mathcal{A}$.

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 Wahrscheinlichkeitsraummodell des Werfens eines fairen Würfels.
- Wir betrachten die Ereignisse

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

- 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 *inkluisiven oder*, also *und/oder*.
- Δ entspricht dem *exklusiven oder*, also *entweder..., oder ..., aber nicht beides*.
- $A \cup B$ entspricht also dem Ereignis, dass A und/oder B eintritt.
- $\omega \in A \cup B$ ist also schon erfüllt, wenn "nur" $\omega \in A$ oder "nur" $\omega \in B$ gilt.
- Für obiges Beispiel gilt

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

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

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

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

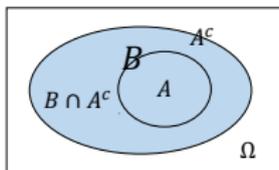
Beweis

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

A

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

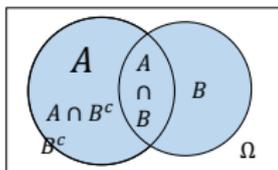
$$A \cap (B \cap A^c) = \emptyset$$



B

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

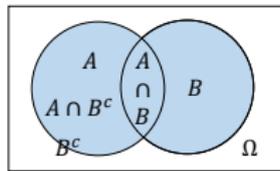
$$A = (A \cap B) \cup (A \cap B^c)$$



C

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

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



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

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

Weitere Eigenschaften

Beweis (fortgeführt)

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

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

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

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

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)

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

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

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

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

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

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

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

Mit 3. folgt dann

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

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

Interpretation

Gemeinsame Wahrscheinlichkeiten

Weitere Eigenschaften

Unabhängige Ereignisse

Bedingte Wahrscheinlichkeiten

Selbstkontrollfragen

Definition (Unabhängige Ereignisse)

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

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

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

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

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:

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

- Die Bedingung der beliebigen Untermengen von I sichert die paarweise unabhängig der $A_i, i \in I$.
- Der Sinn der Produkteigenschaft bei Unabhängigkeit erschließt sich im Kontext *bedingter Wahrscheinlichkeiten*.

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

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

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

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

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
 - $\mathbb{P}(\cup_{i=1}^{\infty} A_i|B) = \sum_{i=1}^{\infty} \mathbb{P}(A_i|B)$ für paarweise disjunkte $A_1, A_2, \dots \in \mathcal{A}$.
- \Rightarrow Die Rechenregeln der Wahrscheinlichkeitstheorie gelten für die Ereignisse links des Strichs.
- Üblicherweise gilt $\mathbb{P}(A|B) \neq \mathbb{P}(B|A)$, z.B.

$$\mathbb{P}(\text{Tod}|\text{Erhängen}) \neq \mathbb{P}(\text{Erhängen}|\text{Tod}).$$

- Eine Verallgemeinerung für $\mathbb{P}(B) = 0$ ist möglich, aber technisch aufwändig.

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

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

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

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

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

hat. Nach Definition der bedingten Wahrscheinlichkeit ergibt sich dann direkt

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

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

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

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 $\mathbb{P}(\cdot|B) \geq 0$. Dann gilt

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

Beweis

Mit der Definition der jeweiligen bedingten Wahrscheinlichkeit folgen direkt

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

und

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

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.

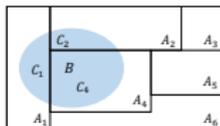
Theorem (Gesetz der totalen Wahrscheinlichkeit)

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

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

Beweis

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



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

□

Bemerkung

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

Theorem (Theorem von Bayes)

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

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

Beweis

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

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

□

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 Unabhängigkeit von A und B im Vordergrund.
- Im Rahmen der **Bayesianischen Inferenz** ist das Theorem von Bayes zentral; hier wird $\mathbb{P}(A_i)$ oft *Prior Wahrscheinlichkeit* und $\mathbb{P}(A_i|B)$ oft *Posterior Wahrscheinlichkeit des Ereignisses A_i* genannt. Wie oben erläutert entspricht $\mathbb{P}(A_i|B)$ dann der aktualisierten Wahrscheinlichkeit von A_i , wenn man um das Eintreten von B weiß.

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.

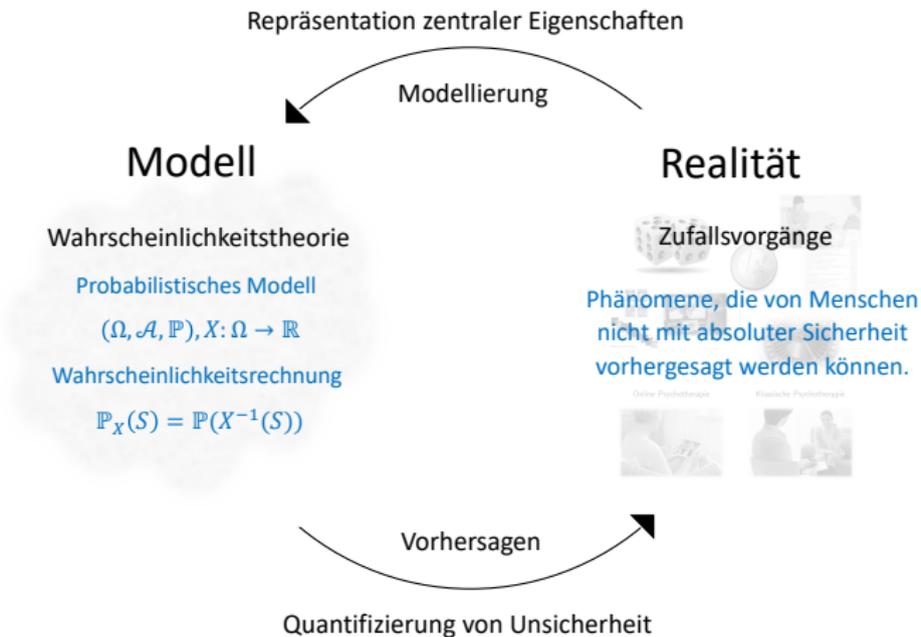


Wahrscheinlichkeitstheorie und Frequentistische Inferenz

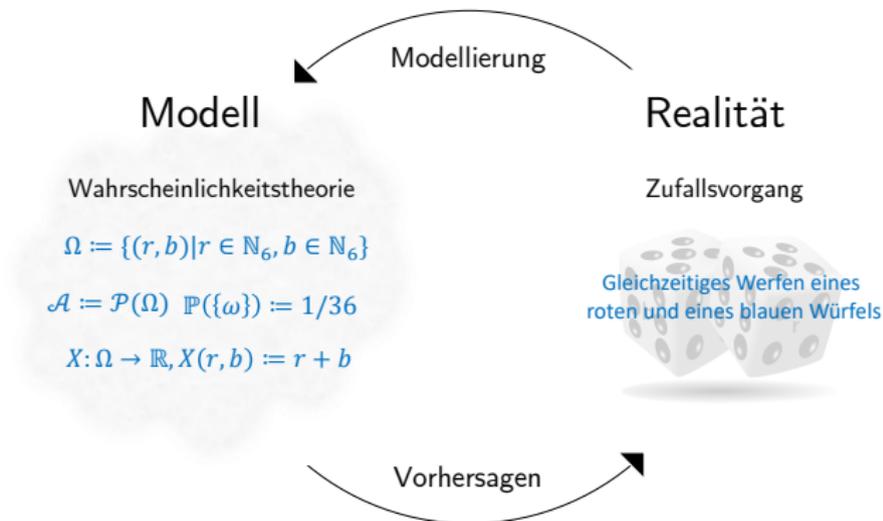
BSc Psychologie WiSe 2021/22

Prof. Dr. Dirk Ostwald

(4) Zufallsvariablen



Jedes Augenzahlpaar kommt im Mittel gleich häufig vor.
Basierend auf der Physik sollte jedes Augenzahlpaar die gleiche Wahrscheinlichkeit haben.



Die Summe der Augenzahlen ist eine Zufallsvariable mit Verteilung \mathbb{P}_X .

Konstruktion, Definition, Notation, Intuition

Wahrscheinlichkeitsmassfunktionen

Wahrscheinlichkeitsdichtefunktionen

Kumulative Verteilungsfunktionen

Selbstkontrollfragen

Konstruktion, Definition, Notation, Intuition

Wahrscheinlichkeitsmassfunktionen

Wahrscheinlichkeitsdichtefunktionen

Kumulative Verteilungsfunktionen

Selbstkontrollfragen

Konstruktion von Zufallsvariablen und Verteilungen

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

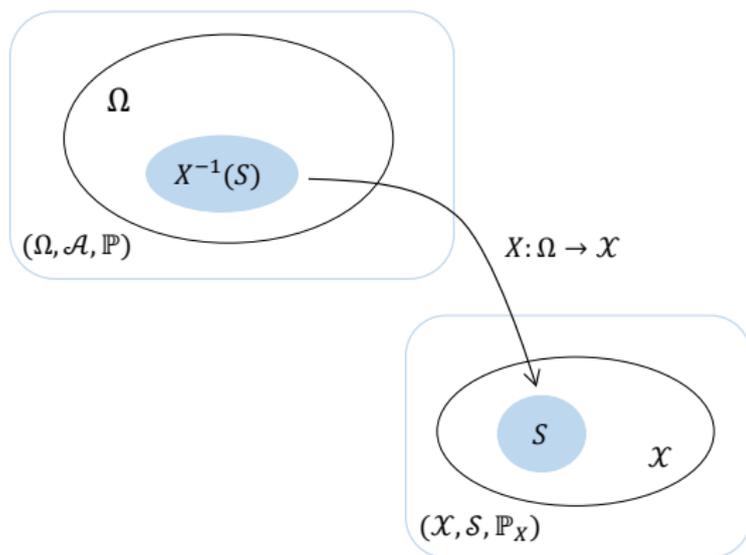
$$X^{-1}(S) := \{\omega \in \Omega \mid X(\omega) \in S\}. \quad (1)$$

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

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

zugeordnet werden.

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



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

Definition (Zufallsvariable)

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

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

Bemerkungen

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

Beispiel (Summe eines roten und eines blauen Würfels)

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

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

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

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

Beispiel (Summe zweier Würfel)

Bestimmung der Verteilung \mathbb{P}_X von X

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

Definition (Notation für Zufallsvariablen)

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

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

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

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

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

Aus diesen Konventionen folgen exemplarisch die folgenden Konventionen für Verteilungen:

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

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

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

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

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

Definition (Realisierungen von Zufallsvariablen und Datenpunkte)

Der einzelne Wert, den eine Zufallsvariable in einem Durchgang eines Zufallsvorgangs annimmt, heißt eine *Realisierung der Zufallsvariable*. In der probabilistischen Datenanalyse wird jeder beobachtete *Datenpunkt* als eine Realisierung einer Zufallsvariablen betrachtet.

Simulation von Zufallsvariablenrealisierungen

```
# Wahrscheinlichkeitsraummodell
Omega = 1:6                                # Ergebnisraum
pi     = rep(1,6)/6                         # Wahrscheinlichkeitsfunktion

# 1. Durchgang des Zufallsvorgangs
rb     = c(Omega[rmultinom(1,1,rep(1,6)/6) == 1], # Realisierung roter Würfel
           Omega[rmultinom(1,1,rep(1,6)/6) == 1]) # Realisierung blauer Würfel
X      = sum(rb)                             # Realisierung der Zufallsvariable X
print(rb)
```

```
> [1] 2 6
```

```
print(X)
```

```
> [1] 8
```

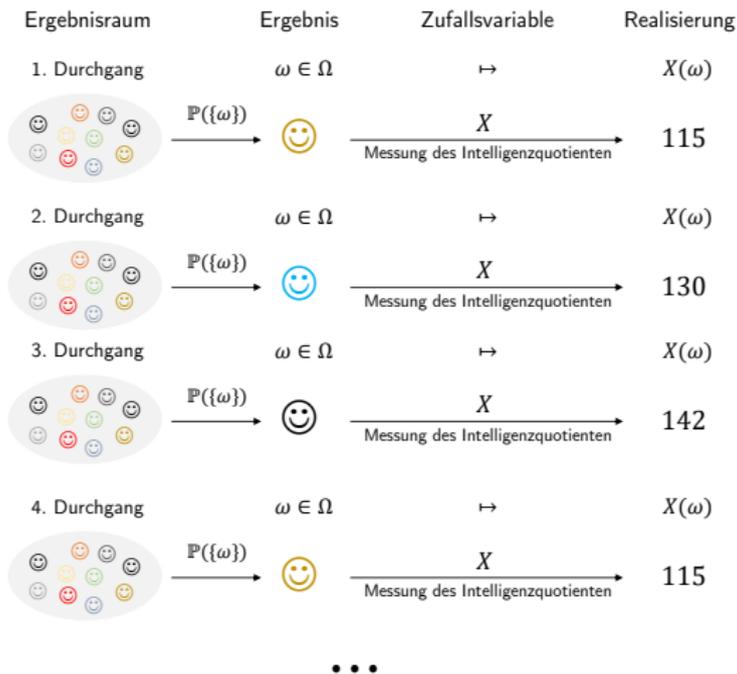
```
# 2. Durchgang des Zufallsvorgangs
rb     = c(Omega[rmultinom(1,1,rep(1,6)/6) == 1], # Realisierung roter Würfel
           Omega[rmultinom(1,1,rep(1,6)/6) == 1]) # Realisierung blauer Würfel
X      = sum(rb)                             # Realisierung der Zufallsvariable X
print(rb)
```

```
> [1] 3 1
```

```
print(X)
```

```
> [1] 4
```

Zufallsvariablen als Modelle von Messvorgängen



Konstruktion, Definition, Notation, Intuition

Wahrscheinlichkeitsmassfunktionen

Wahrscheinlichkeitsdichtefunktionen

Kumulative Verteilungsfunktionen

Selbstkontrollfragen

Definition (Diskrete ZV, Wahrscheinlichkeitsmassefunktion)

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

$$p : \mathcal{X} \rightarrow [0, 1], x \mapsto p(x) := \mathbb{P}_X(X = x). \quad (5)$$

Bemerkungen

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

Definition (Bernoulli-Zufallsvariable)

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

$$p : \mathcal{X} \rightarrow [0, 1], x \mapsto p(x) := \mu^x(1 - \mu)^{1-x} \text{ mit } \mu \in [0, 1]. \quad (6)$$

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

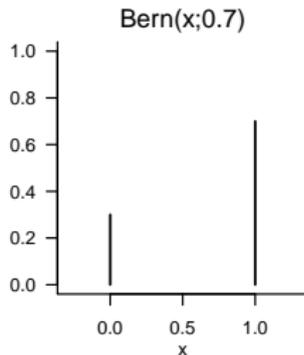
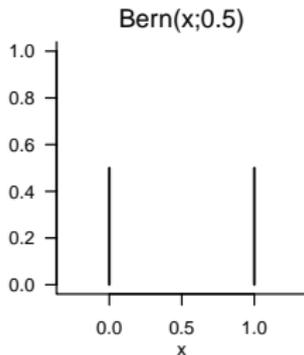
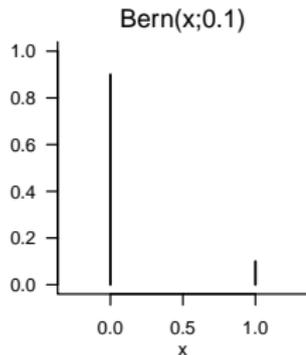
$$\text{Bern}(x; \mu) := \mu^x(1 - \mu)^{1-x}. \quad (7)$$

Bemerkungen

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

$$\mathbb{P}(X = 1) = \mu^1(1 - \mu)^{1-1} = \mu. \quad (8)$$

Bernoulli-Zufallsvariable



Definition (Diskret-gleichverteilte Zufallsvariable)

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

$$p : \mathcal{X} \rightarrow \mathbb{R}_{\geq 0}, x \mapsto p(x) := \frac{1}{|\mathcal{X}|}. \quad (9)$$

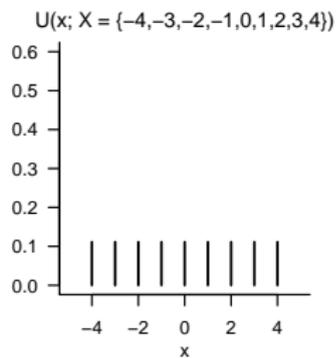
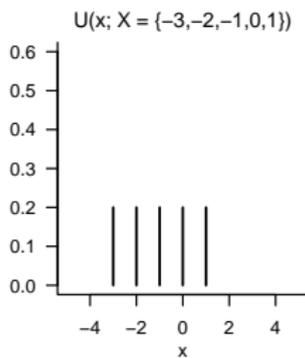
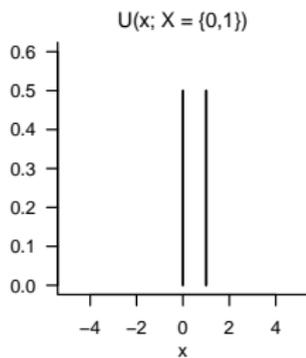
Dann sagen wir, dass X einer *diskreten Gleichverteilung* unterliegt und nennen X eine *diskret-gleichverteilte Zufallsvariable*. Wir kürzen dies mit $X \sim U(|\mathcal{X}|)$ ab. Die WMF einer diskret-gleichverteilten Zufallsvariable bezeichnen wir mit

$$U(x; |\mathcal{X}|) := \frac{1}{|\mathcal{X}|}. \quad (10)$$

Bemerkungen

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

Diskret-gleichverteilte Zufallsvariable



Konstruktion, Definition, Notation, Intuition

Wahrscheinlichkeitsmassfunktionen

Wahrscheinlichkeitsdichtefunktionen

Kumulative Verteilungsfunktionen

Selbstkontrollfragen

Definition (Kontinuierliche ZV, Wahrscheinlichkeitsdichtefunktion)

Eine Zufallsvariable X heißt *kontinuierlich*, wenn eine Funktion

$$p : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}, x \mapsto p(x) \quad (11)$$

existiert, so dass gilt

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

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

Bemerkungen

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

Definition (Normalverteilte und standardnormalverteilte Zufallsvariablen)

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

$$p : \mathbb{R} \rightarrow \mathbb{R}_{>0}, x \mapsto p(x) := \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x - \mu)^2\right). \quad (12)$$

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

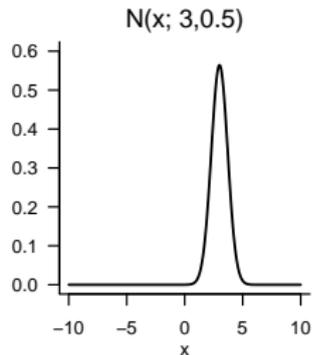
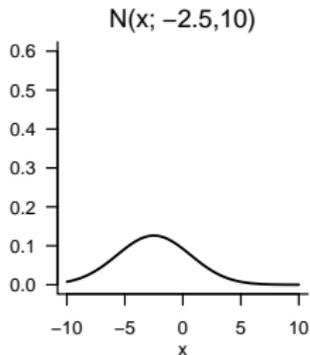
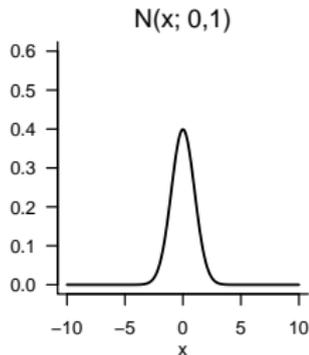
$$N(x; \mu, \sigma^2) := \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x - \mu)^2\right). \quad (13)$$

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

$$p : \mathbb{R}_{>0} \rightarrow \mathbb{R}_{>0}, x \mapsto p(x) := \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} \exp\left(-\frac{x}{\beta}\right), \quad (14)$$

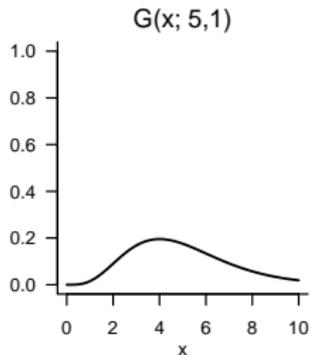
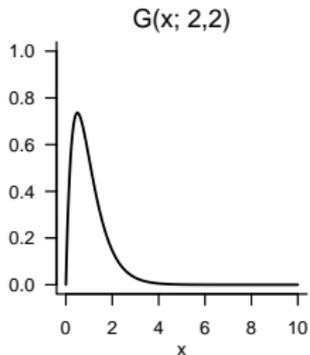
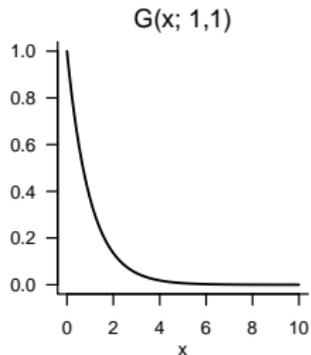
wobei Γ die Gammafunktion bezeichne. Dann sagen wir, dass X einer *Gammaverteilung mit Formparameter $\alpha > 0$ und Skalenparameter $\beta > 0$* unterliegt und nennen X eine *gammaverteilte Zufallsvariable*. Wir kürzen dies mit $X \sim G(\alpha, \beta)$ ab. Die WDF einer gammaverteilten Zufallsvariable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

$$G(x; \alpha, \beta) := \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} \exp\left(-\frac{x}{\beta}\right). \quad (15)$$

Bemerkung

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

Gamma-Zufallsvariablen



Definition (Beta-Zufallsvariable)

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

$$p : \mathcal{X} \rightarrow [0, 1], x \mapsto p(x) := \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1} \text{ mit } \alpha, \beta \in \mathbb{R}_{>0}, \quad (16)$$

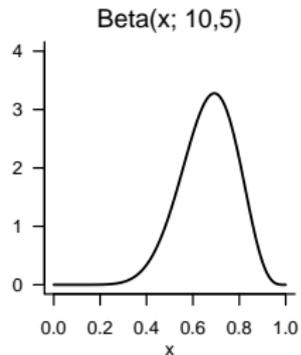
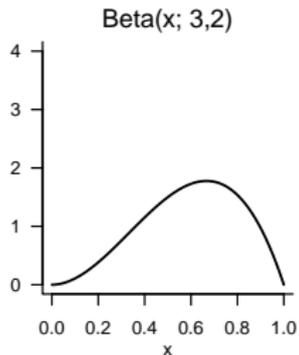
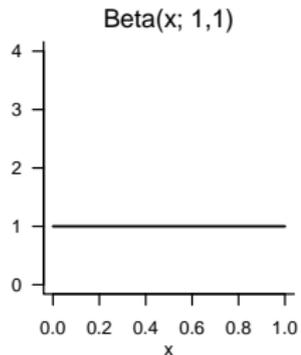
wobei Γ die Gammafunktion bezeichne. Dann sagen wir, dass X einer *Beta-Verteilung* mit Parametern $\alpha > 0$ und $\beta > 0$ unterliegt, und nennen X eine *beta-verteilte Zufallsvariable*. Wir kürzen dies mit $X \sim \text{Beta}(\alpha, \beta)$ ab. Die WDF einer beta-verteilten Zufallsvariable bezeichnen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

$$\text{Beta}(x; \alpha, \beta) := \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1}. \quad (17)$$

Bemerkung

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

Beta-Zufallsvariable



Definition (Gleichverteilte Zufallsvariable)

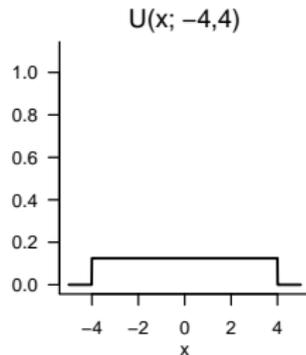
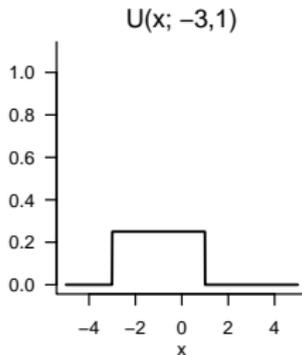
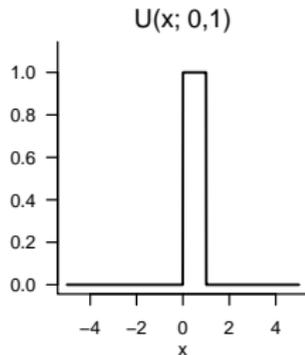
Es sei X eine kontinuierliche Zufallsvariable mit Ergebnisraum \mathbb{R} und WDF

$$p : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}, x \mapsto p(x) := \begin{cases} \frac{1}{b-a} & x \in [a, b] \\ 0 & x \notin [a, b] \end{cases}. \quad (18)$$

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

$$U(x; a, b) := \frac{1}{b-a}. \quad (19)$$

Gleichverteilte Zufallsvariablen



Konstruktion, Definition, Notation, Intuition

Wahrscheinlichkeitsmassfunktionen

Wahrscheinlichkeitsdichtefunktionen

Kumulative Verteilungsfunktionen

Selbstkontrollfragen

Definition (Kumulative Verteilungsfunktion)

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

$$P : \mathbb{R} \rightarrow [0, 1], x \mapsto P(x) := \mathbb{P}(X \leq x). \quad (20)$$

Bemerkungen

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

Theorem (Überschreitungswahrscheinlichkeit)

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

$$\mathbb{P}(X > x) = 1 - P(x) \text{ für alle } x \in \mathcal{X}. \quad (21)$$

Beweis

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

$$\Omega = \{\omega \in \Omega | X(\omega) > x\} \cup \{\omega \in \Omega | X(\omega) \leq x\} = \{X > x\} \cup \{X \leq x\}. \quad (22)$$

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

$$\begin{aligned} \mathbb{P}(\Omega) &= 1 \\ \Leftrightarrow \mathbb{P}(\{X > x\} \cup \{X \leq x\}) &= 1 \\ \Leftrightarrow \mathbb{P}(\{X > x\}) + \mathbb{P}(\{X \leq x\}) &= 1 \\ \Leftrightarrow \mathbb{P}(\{X > x\}) &= 1 - \mathbb{P}(\{X \leq x\}) \\ \Leftrightarrow \mathbb{P}(\{X > x\}) &= 1 - P(x). \end{aligned} \quad (23)$$

□

Theorem (Intervallwahrscheinlichkeiten)

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

$$\mathbb{P}(X \in]x_1, x_2]) = P(x_2) - P(x_1) \text{ für alle } x_1, x_2 \in \mathcal{X} \text{ mit } x_1 < x_2. \quad (24)$$

Beweis

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

$$\{X \leq x_1\} \cap \{x_1 < X \leq x_2\} = \emptyset \text{ und } \{X \leq x_1\} \cup \{x_1 < X \leq x_2\} = \{X \leq x_2\}. \quad (25)$$

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

$$\begin{aligned} \mathbb{P}(\{X \leq x_1\} \cup \{x_1 < X \leq x_2\}) &= \mathbb{P}(\{X \leq x_2\}) \\ \Leftrightarrow \mathbb{P}(\{X \leq x_1\}) + \mathbb{P}(\{x_1 < X \leq x_2\}) &= \mathbb{P}(\{X \leq x_2\}) \\ \Leftrightarrow \mathbb{P}(\{x_1 < X \leq x_2\}) &= \mathbb{P}(\{X \leq x_2\}) - \mathbb{P}(\{X \leq x_1\}) \\ \Leftrightarrow \mathbb{P}(\{x_1 < X \leq x_2\}) &= P(x_2) - P(x_1) \\ \Leftrightarrow \mathbb{P}(X \in]x_1, x_2]) &= P(x_2) - P(x_1). \end{aligned} \quad (26)$$

□

Theorem (Eigenschaften von kumulative Verteilungsfunktionen)

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

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

Bemerkungen

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

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 $x_1 < x_2$,

$$\{X \leq x_1\} = \{\omega \in \Omega | X(\omega) \leq x_1\} \subset \{\omega \in \Omega | X(\omega) \leq x_2\} = \{X \leq x_2\} \quad (27)$$

Also gilt

$$\mathbb{P}(\{X \leq x_1\}) \leq \mathbb{P}\{X \leq x_2\} \Rightarrow P(x_1) \leq P(x_2). \quad (28)$$

- (2) Wir verzichten auf einen Beweis

- (3) Wir definieren

$$P(x^+) = \lim_{y \rightarrow x, y > x} P(y). \quad (29)$$

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

$$\{X \leq x\} = \bigcap_{n=1}^{\infty} \{X \leq y_n\}. \quad (30)$$

Es gilt also

$$P(x) = \mathbb{P}(\{X \leq x\}) = \mathbb{P}(\bigcap_{n=1}^{\infty} \{X \leq y_n\}) = \lim_{n \rightarrow \infty} \mathbb{P}(\{X \leq y_n\}) = P(x^+), \quad (31)$$

wobei wir die dritte Gleichung unbegründet stehen lassen.

Kumulative Verteilungsfunktionen von diskreten Zufallsvariablen

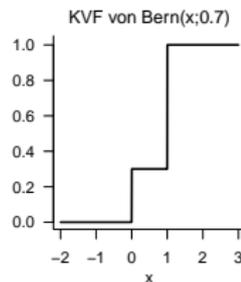
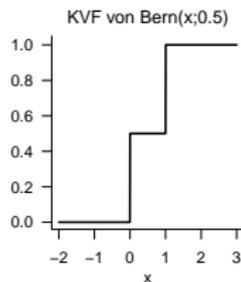
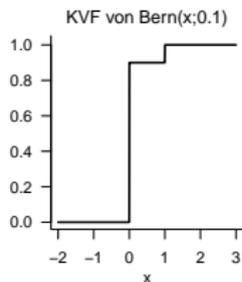
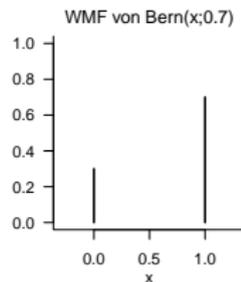
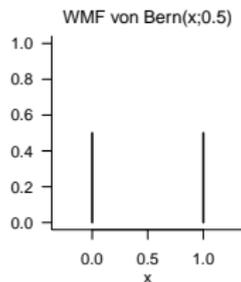
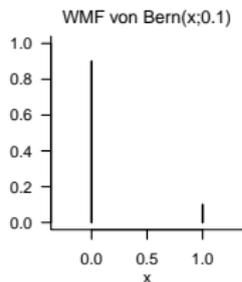
- Wenn $a < b$ und $\mathbb{P}(a < X < b) = 0$, dann ist P konstant horizontal auf $]a, b[$.
- An jedem Punkt x mit $\mathbb{P}(X = x) > 0$ springt die KVF um den Betrag $\mathbb{P}(X = x)$.
- \Leftrightarrow An jedem Punkt x mit $p(x) > 0$ springt die KVF um den Betrag $p(x) > 0$.
- Generell ist die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

$$P : \mathbb{R} \rightarrow [0, 1], x \mapsto P(x) := \sum_{k=0}^{\lfloor x \rfloor} \mathbb{P}(X = k) \quad (32)$$

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

Kumulative Verteilungsfunktionen

Bernoulli-Zufallsvariablen



Theorem (Kumulative Verteilungsfunktionen von kontinuierlichen ZVen)

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

$$P(x) = \int_{-\infty}^x p(t) dt \text{ und } p(x) = \frac{d}{dx} P(x). \quad (33)$$

Beweis

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

□

Bemerkungen

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

Beispiel (Normalverteilung)

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

- Die WDF von X ist

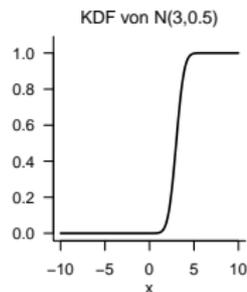
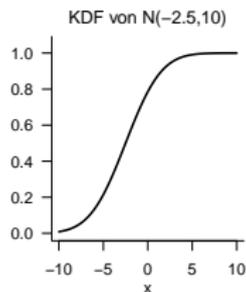
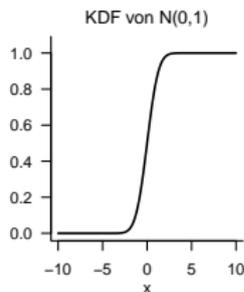
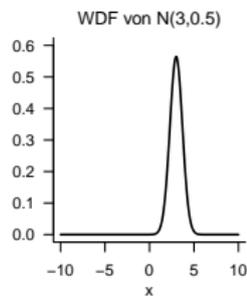
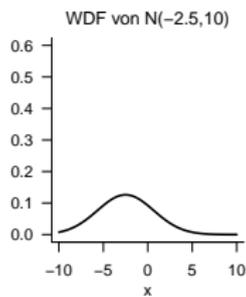
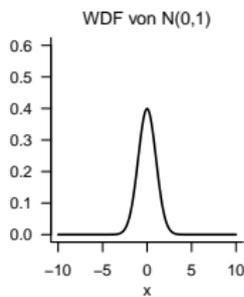
$$p : \mathbb{R} \rightarrow \mathbb{R}_{>0}, x \mapsto p(x) := \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x - \mu)^2\right).$$

- Die KVF von X ist

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

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

Normalverteilte Zufallsvariablen



Definition (Inverse Kumulative Verteilungsfunktion)

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

$$P^{-1} :]0, 1[\rightarrow \mathbb{R}, q \mapsto P^{-1}(q) := \{x \in \mathbb{R} | P(x) = q\} \quad (34)$$

die *inverse kumulative Verteilungsfunktion* von X .

Bemerkungen

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

Beispiel (Normalverteilung)

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

- Die KVF von X ist

$$P : \mathbb{R} \rightarrow]0, 1[, x \mapsto \mathbb{P}(X \leq x) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^x \exp\left(-\frac{1}{2\sigma^2}(\xi - \mu)^2\right) d\xi \quad (35)$$

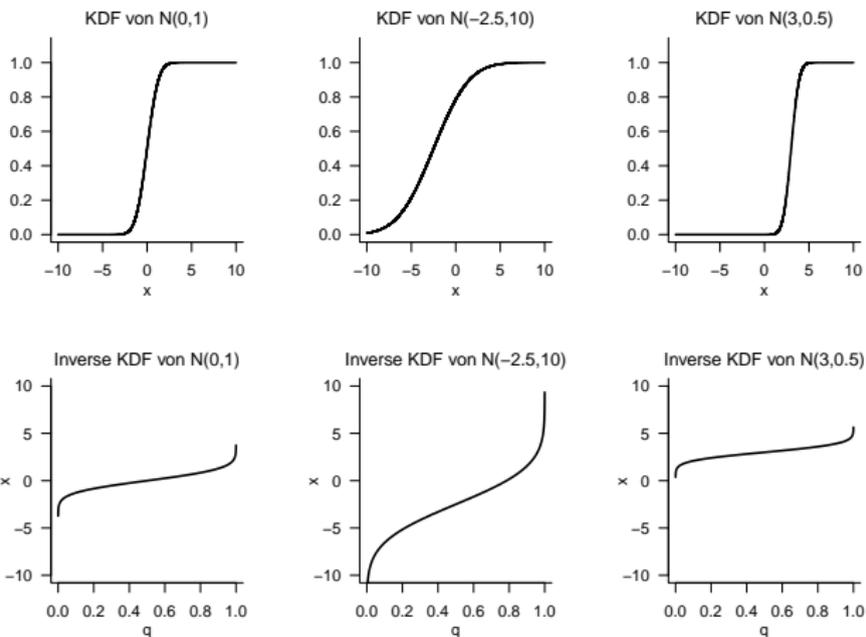
- Die inverse KVF von X ist

$$P^{-1} :]0, 1[\rightarrow \mathbb{R}, q \mapsto P^{-1}(q) = \{x \in \mathbb{R} | P(x) = q\}. \quad (36)$$

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

Kumulative Verteilungsfunktionen

Normalverteilte Zufallsvariablen



Konstruktion, Definition, Notation, Intuition

Wahrscheinlichkeitsmassfunktionen

Wahrscheinlichkeitsdichtefunktionen

Kumulative Verteilungsfunktionen

Selbstkontrollfragen

Selbstkontrollfragen

1. Definieren Sie den Begriff der Zufallsvariable.
2. Erläutern Sie die Gleichung $\mathbb{P}_X(X = x) = \mathbb{P}(\{X = x\})$.
3. Erläutern Sie die Bedeutung von $\mathbb{P}(X = x)$.
4. Definieren Sie den Begriff der Wahrscheinlichkeitsmassenfunktion.
5. Definieren Sie die Begriffe der Wahrscheinlichkeitsdichtefunktion.
6. Definieren Sie den Begriff der kumulativen Verteilungsfunktion.
7. Schreiben sie die Intervallwahrscheinlichkeit einer Zufallsvariable mithilfe 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.



Wahrscheinlichkeitstheorie und Frequentistische Inferenz

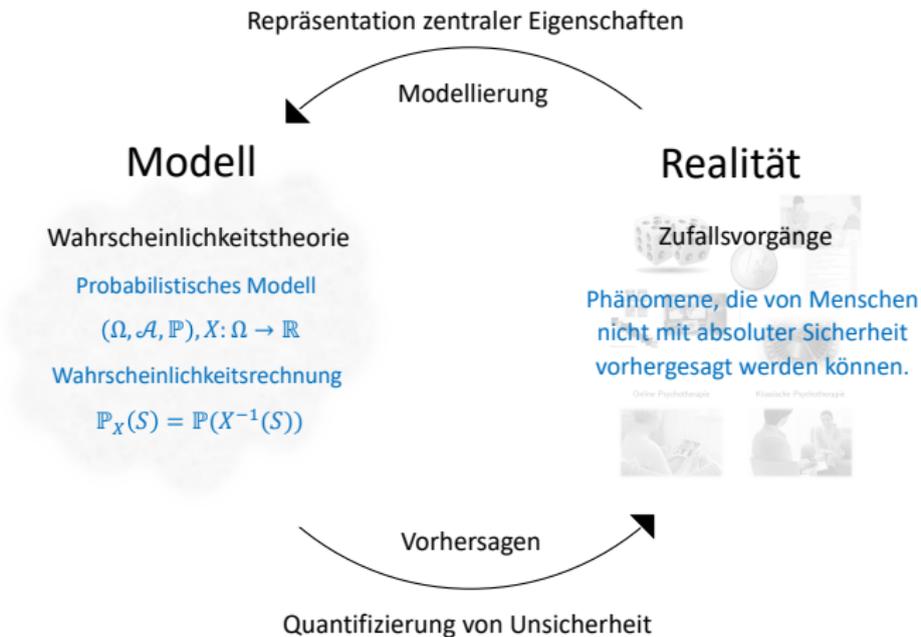
BSc Psychologie WiSe 2021/22

Prof. Dr. Dirk Ostwald

(5) Multivariate Verteilungen

Wahrscheinlichkeitstheorie und Frequentistische Inferenz

Datum	Einheit	Thema
14.10.2021	Einführung	(1) Einführung
21.10.2021	Wahrscheinlichkeitstheorie	(2) Wahrscheinlichkeitsräume
28.10.2021	Wahrscheinlichkeitstheorie	(3) Elementare Wahrscheinlichkeiten
04.11.2021	Wahrscheinlichkeitstheorie	(4) Zufallsvariablen
11.11.2021	Wahrscheinlichkeitstheorie	(5) Multivariate Verteilungen
18.11.2021	Wahrscheinlichkeitstheorie	(6) Erwartungswert und Kovarianz
25.11.2021	Wahrscheinlichkeitstheorie	(7) Ungleichungen und Grenzwerte
02.12.2021	Wahrscheinlichkeitstheorie	(8) Normalverteilungstransformationen
09.12.2021	Frequentistische Inferenz	(9) Modelle, Statistiken und Schätzer
16.12.2021	Frequentistische Inferenz	(10) Schätzeigenschaften
	Weihnachtspause	
06.01.2022	Frequentistische Inferenz	(11) Konfidenzintervalle
13.01.2022	Frequentistische Inferenz	(12) Hypothesentests
20.01.2022	Frequentistische Inferenz	(13) Einstichproben-T-Tests
27.01.2022	Frequentistische Inferenz	(14) Zweistichproben-T-Tests
31.01.2022	Klausur	16 - 17 Uhr, G44 - H6
Jul 2022	Klausurwiederholungstermin	



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

Modell

Modellierung

Realität

Wahrscheinlichkeitstheorie

$$X_{1j} \sim N(\mu_1, \sigma^2), j = 1, \dots, n_1$$

$$X_{2j} \sim N(\mu_2, \sigma^2), j = 1, \dots, n_2$$

Zufallsvorgang

Online Psychotherapie

Klassische Psychotherapie



Klinische Studie zum Vergleich der Effekte von
Klassischer und Online PT bei Depression

	Online	Classical	SD
1	35	35	10
2	35	35	10
3	35	35	10
4	35	35	10
5	35	35	10
6	35	35	10
7	35	35	10
8	35	35	10
9	35	35	10
10	35	35	10
11	35	35	10
12	35	35	10
13	35	35	10
14	35	35	10
15	35	35	10
16	35	35	10
17	35	35	10
18	35	35	10
19	35	35	10
20	35	35	10

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

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

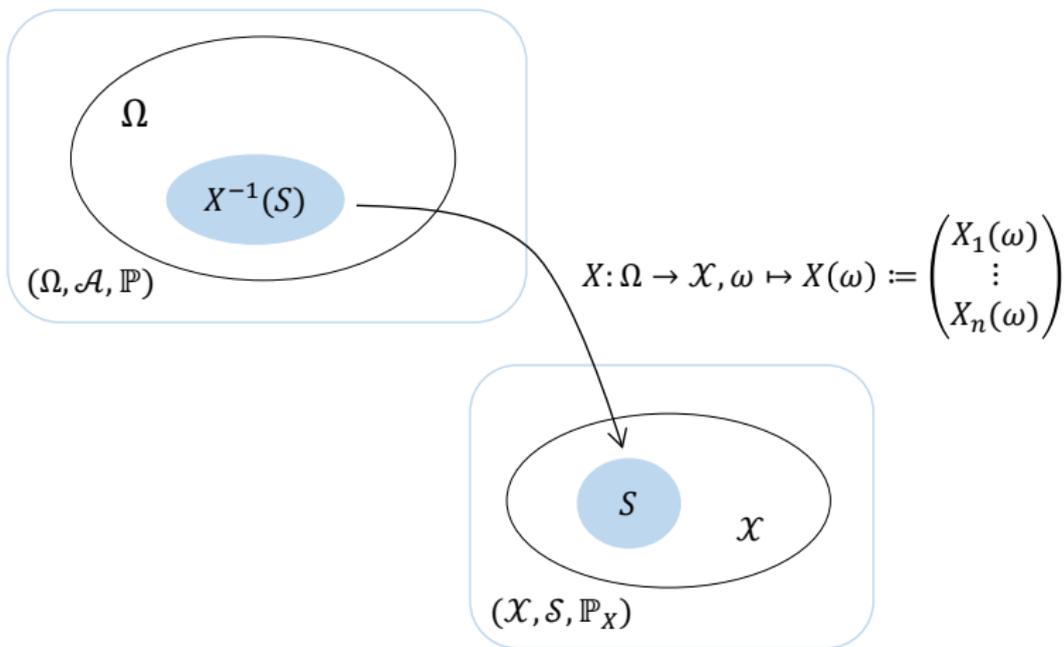
mit der *Messbarkeitseigenschaft*

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

Bemerkungen

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

Definition



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

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

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

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

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

die *multivariate Verteilung des Zufallsvektor* X .

Bemerkungen

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

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

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

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

Definition (Multivariate kumulative Verteilungsfunktionen)

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

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

multivariate kumulative Verteilungsfunktion von X .

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)

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

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

Bemerkungen

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

Beispiel (Multivariate Wahrscheinlichkeitsmassfunktion)

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

Eine exemplarische bivariate WMF der Form

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

ist dann durch nachfolgende Tabelle definiert.

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

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

Definition (Kontinuierlicher Zufallsvektor, multivariate WDF)

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

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

mit den Eigenschaften

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

$$(2) \mathbb{P}_X(x_1 \leq X \leq x_2) = \int_{x_{1_1}}^{x_{2_1}} \cdots \int_{x_{1_n}}^{x_{2_n}} p_X(s_1, \dots, s_n) ds_1 \cdots ds_n$$

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

$$\mathbb{P}_X(X = x) = \mathbb{P}_X(x \leq X \leq x) = \int_{x_1}^{x_1} \cdots \int_{x_n}^{x_n} p_X(s_1, \dots, s_n) ds_1 \cdots ds_n = 0 \quad (10)$$

Definition

Multivariate Verteilungen

Marginalverteilungen

Bedingte Verteilungen

Unabhängige Zufallsvariablen

Selbstkontrollfragen

Definition (Univariate Marginalverteilung)

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

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

definierte Verteilung die *ite univariate Marginalverteilung* von X .

Bemerkungen

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

Theorem (Marginale Wahrscheinlichkeitsmasse- und dichtefunktionen)

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

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

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

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

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.

Beispiel (Marginale Wahrscheinlichkeitsmassfunktionen)

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

Basierend auf der oben definierten WMF ergeben sich folgende marginalen WMFen p_{X_1} und p_{X_2}

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

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

Definition

Multivariate Verteilungen

Marginalverteilungen

Bedingte Verteilungen

Unabhängige Zufallsvariablen

Selbstkontrollfragen

Vorbemerkungen

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

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

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

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

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

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

Definition (Bedingte WMF, diskrete bedingte Verteilung)

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

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

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

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

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

Bemerkungen

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

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

- Bedingte Verteilungen sind (lediglich) normalisierte gemeinsame Verteilungen.

Beispiel (Bedingte Wahrscheinlichkeitsmassfunktionen)

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

Basierend auf der oben definierten WMF und den entsprechenden oben evaluierten marginalen WMFen ergeben sich folgende bedingte WMFen für $p_{X_2|X_1=x_1}$

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

Bemerkungen

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

Definition (Bedingte WDF, kontinuierliche bedingte Verteilungen)

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

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

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

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

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

Bemerkung

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

Definition

Multivariate Verteilungen

Marginalverteilungen

Bedingte Verteilungen

Unabhängige Zufallsvariablen

Selbstkontrollfragen

Definition (Unabhängige Zufallsvariablen)

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

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

Bemerkungen

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

Theorem (Unabhängigkeit und Faktorisierung der WMF/WDF)

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

X_1 und X_2 sind unabhängige Zufallsvariablen \Leftrightarrow

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

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

X_1 und X_2 sind unabhängige Zufallsvariablen \Leftrightarrow

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

Bemerkungen

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

Unabhängige Zufallsvariablen

Beispiel (Unabhängige diskrete Zufallsvariablen)

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

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

Da hier gilt, dass

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

sind die Zufallsvariablen X_1 und X_2 nicht unabhängig.

Unabhängige Zufallsvariablen

Beispiel (Unabhängige diskrete Zufallsvariablen)

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

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

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

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

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

Definition (n unabhängige Zufallsvariablen)

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

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

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

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

Bemerkung

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

Definition (Unabhängig und identisch verteilte Zufallsvariablen)

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

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

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

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

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

Bemerkungen

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

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 2021/22

Prof. Dr. Dirk Ostwald

(6) Erwartungswert, Varianz, Kovarianz

Wahrscheinlichkeitstheorie und Frequentistische Inferenz

Datum	Einheit	Thema
14.10.2021	Einführung	(1) Einführung
21.10.2021	Wahrscheinlichkeitstheorie	(2) Wahrscheinlichkeitsräume
28.10.2021	Wahrscheinlichkeitstheorie	(3) Elementare Wahrscheinlichkeiten
04.11.2021	Wahrscheinlichkeitstheorie	(4) Zufallsvariablen
11.11.2021	Wahrscheinlichkeitstheorie	(5) Multivariate Verteilungen
18.11.2021	Wahrscheinlichkeitstheorie	(6) Erwartungswert, Varianz, Kovarianz
25.11.2021	Wahrscheinlichkeitstheorie	(7) Ungleichungen und Grenzwerte
02.12.2021	Wahrscheinlichkeitstheorie	(8) Normalverteilungstransformationen
09.12.2021	Frequentistische Inferenz	(9) Modelle, Statistiken und Schätzer
16.12.2021	Frequentistische Inferenz	(10) Schätzereigenschaften
	Weihnachtspause	
06.01.2022	Frequentistische Inferenz	(11) Konfidenzintervalle
13.01.2022	Frequentistische Inferenz	(12) Hypothesentests
20.01.2022	Frequentistische Inferenz	(13) Einstichproben-T-Tests
27.01.2022	Frequentistische Inferenz	(14) Zweistichproben-T-Tests
31.01.2022	Klausur	16 - 17 Uhr, G44 - H6
Jul 2022	Klausurwiederholungstermin	

Repräsentation zentraler Eigenschaften

Modellierung

Modell

Realität

Wahrscheinlichkeitstheorie

Probabilistisches Modell

$$(\Omega, \mathcal{A}, \mathbb{P}), X: \Omega \rightarrow \mathbb{R}$$

Wahrscheinlichkeitsrechnung

$$\mathbb{P}_X(S) = \mathbb{P}(X^{-1}(S))$$

Zufallsvorgänge

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

Deine Psychologie

Klassische Psychologie



Vorhersagen

Quantifizierung von Unsicherheit

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

Modell

Modellierung

Realität

Wahrscheinlichkeitstheorie

$$X_{1j} \sim N(\mu_1, \sigma^2), j = 1, \dots, n_1$$

$$X_{2j} \sim N(\mu_2, \sigma^2), j = 1, \dots, n_2$$

$$\mathbb{E}(X_{ij}) = \mu_i \quad \mathbb{V}(X_{ij}) = \sigma^2 \quad \forall i, j$$

$$C(X_{ij}, X_{kl}) = 0 \quad \forall i \neq k, j \neq l$$

Zufallsvorgang

Online Psychotherapie

Klassische Psychotherapie



Klinische Studie zum Vergleich der Effekte von
Klassischer und Online PT bei Depression

	Online	KL	KL
1	Online	25	20
2	Online	14	20
3	Online	25	20
4	Online	14	20
5	Online	25	20
6	Online	25	20
7	Online	20	20
8	Online	15	17
9	Online	20	20
10	Online	20	20
11	Online	20	20
12	Online	20	20
13	Online	25	27
14	Online	20	14
15	Online	20	17
16	Online	20	20

Vorhersagen

Erwartungswert

Varianz und Standardabweichung

Stichprobenmittel, Stichprobenvarianz, Stichprobenstandardabweichung

Kovarianz und Korrelation

Stichprobenkovarianz und Stichprobenkorrelation

Selbstkontrollfragen

Erwartungswert

Varianz und Standardabweichung

Stichprobenmittel, Stichprobenvarianz, Stichprobenstandardabweichung

Kovarianz und Korrelation

Stichprobenkovarianz und Stichprobenkorrelation

Selbstkontrollfragen

Definition (Erwartungswert)

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

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

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

Bemerkungen

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

Beispiel (Erwartungswert einer Bernoulli Zufallsvariable)

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

Beweis

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

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

□

Erwartungswert

Beispiel (Erwartungswert einer normalverteilten Zufallsvariable)

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

Beweis

Wir halten zunächst ohne Beweis fest, dass

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

Mit der Definition des Erwartungswerts für kontinuierliche Zufallsvariablen gilt

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

Mit der Substitutionsregel

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

und der Definition von

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

gilt dann

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

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

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

verschwindet der Integralterm und wir erhalten

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

□

Theorem (Eigenschaften des Erwartungswerts)

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

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

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

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

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

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

Bemerkung

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

Beweis

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

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

Beweis (fortgeführt)

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

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

Beweis (fortgeführt)

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

(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_{X_1, \dots, X_n} .
Weil als X_1, \dots, X_n unabhängig vorausgesetzt sind, gilt

$$p_{X_1, \dots, X_n}(x_1, \dots, x_n) = \prod_{i=1}^n p_{X_i}(x_i). \quad (15)$$

Weiterhin gilt also

$$\begin{aligned} \mathbb{E} \left(\prod_{i=1}^n x_i \right) &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left(\prod_{i=1}^n x_i \right) p_{X_1, \dots, X_n}(x_1, \dots, x_n) dx_1 \dots dx_n \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \prod_{i=1}^n x_i \prod_{i=1}^n p_{X_i}(x_i) dx_1 \dots dx_n \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \prod_{i=1}^n x_i p_{X_i}(x_i) dx_1 \dots dx_n \\ &= \prod_{i=1}^n \int_{-\infty}^{\infty} x_i p_{X_i}(x_i) dx_i = \prod_{i=1}^n \mathbb{E}(X_i). \end{aligned} \quad (16)$$

□

Erwartungswert

Varianz und Standardabweichung

Stichprobenmittel, Stichprobenvarianz, Stichprobenstandardabweichung

Kovarianz und Korrelation

Stichprobenkovarianz und Stichprobenkorrelation

Selbstkontrollfragen

Definition (Varianz und Standardabweichung)

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

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

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

$$\mathbb{S}(X) := \sqrt{\mathbb{V}(X)}. \quad (18)$$

Bemerkungen

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

Varianz und Standardabweichung

Beispiel (Varianz einer Bernoulli Zufallsvariable)

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

$$\mathbb{V}(X) = \mu(1 - \mu). \quad (19)$$

Beweis

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

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

□

Theorem (Varianzverschiebungssatz)

Es sei X eine Zufallsvariable. Dann gilt

$$\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2. \quad (21)$$

Beweis

Mit der Definition der Varianz und der Linearität des Erwartungswerts gilt

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

□

Bemerkung

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

Varianz und Standardabweichung

Beispiel (Varianz einer normalverteilten Zufallsvariable)

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

Beweis

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

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

Mit der Substitutionsregel

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

und der Definition von

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

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

Varianz und Standardabweichung

Beweis (fortgeführt)

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

geschrieben werden. Also gilt

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

Varianz und Standardabweichung

Beweis (fortgeführt)

Wir halten weiterhin ohne Beweis fest, dass

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

Es ergibt sich dann

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

Mit der partiellen Integrationsregel

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

und der Definition von

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

Varianz und Standardabweichung

Beweis (fortgeführt)

und

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

so dass

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

gilt, ergibt sich dann

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

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

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

□

Theorem (Eigenschaften der Varianz)

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

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

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

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

Varianz und Standardabweichung

Beweis

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

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

Wurzelziehen ergibt dann das Resultat für die Standardabweichung.

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

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

Varianz und Standardabweichung

Beweis (fortgeführt)

Es ergibt sich also

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

Beweis (fortgeführt)

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

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

ist. Damit folgt also

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

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

□

Erwartungswert

Varianz und Standardabweichung

Stichprobenmittel, Stichprobenvarianz, Stichprobenstandardabweichung

Kovarianz und Korrelation

Stichprobenkovarianz und Stichprobenkorrelation

Selbstkontrollfragen

Definition (Stichprobenmittel, -varianz, -standardabweichung)

X_1, \dots, X_n seien Zufallsvariablen. Dann nennt man X_1, \dots, X_n auch eine *Stichprobe*.

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

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

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

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

- Die *Stichprobenstandardabweichung* ist definiert als

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

Bemerkungen

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

Beispiel (Stichprobenmittel, Stichprobenvarianz, Stichprobenstandardabweichung)

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

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

- Die Stichprobenmittelrealisation ist

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

- Die Stichprobenvarianzrealisation ist

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

- Die Stichprobenstandardabweichungrealisation ist

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

Erwartungswert

Varianz und Standardabweichung

Stichprobenmittel, Stichprobenvarianz, Stichprobenstandardabweichung

Kovarianz und Korrelation

Stichprobenkovarianz und Stichprobenkorrelation

Selbstkontrollfragen

Definition (Kovarianz und Korrelation)

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

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

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

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

Bemerkungen

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

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

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

Beispiel (Kovarianz und Korrelation zweier diskreter Zufallsvariablen)

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

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

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

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

und

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

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

Kovarianz und Korrelation

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

Theorem (Kovarianzverschiebungssatz)

X und Y seien Zufallsvariablen. Dann gilt

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

Beweis

Mit der Definition der Kovarianz gilt

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

□

Bemerkungen

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

Theorem (Korrelation und Unabhängigkeit)

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

Beweis

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

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

Mit dem Kovarianzverschiebungssatz folgt dann

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

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

Kovarianz und Korrelation

Beweis (fortgeführt)

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

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

Wir halten dann zunächst fest, dass

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

und

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

Mit dem Kovarianzverschiebungssatz ergibt sich dann

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

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

Kovarianz und Korrelation

Beweis (fortgeführt)

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

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

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

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

Es gilt also zum Beispiel

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

und damit sind X und Y nicht unabhängig.

□

Theorem (Varianzen von Summen und Differenzen von Zufallsvariablen)

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

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

Speziell gelten

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

und

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

Bemerkungen

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

Beweis

Wir halten zunächst fest, dass

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

Es ergibt sich also

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

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

□

Erwartungswert

Varianz und Standardabweichung

Stichprobenmittel, Stichprobenvarianz, Stichprobenstandardabweichung

Kovarianz und Korrelation

Stichprobenkovarianz und Stichprobenkorrelation

Selbstkontrollfragen

Definition (Stichprobenkovarianz und Stichprobenkorrelation)

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

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

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

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

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

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

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

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

Bemerkungen

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

Stichprobenkovarianz und Stichprobenkorrelation

Beispiel (Stichprobenkovarianz und Stichprobenkorrelation)

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

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

- Die Stichprobenmittelrealisation ergibt sich zu

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

- Die Stichprobenstandardabweichungrealisationen ergeben sich zu

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

- Die Stichprobenkovarianz- und Stichprobenkorrelationsrealisationen ergeben sich zu

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

Erwartungswert

Varianz und Standardabweichung

Stichprobenmittel, Stichprobenvarianz, Stichprobenstandardabweichung

Kovarianz und Korrelation

Stichprobenkovarianz und Stichprobenkorrelation

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 $\mathbb{E}(X^2)$ mithilfe der Varianz und des Erwartungswerts von X aus.
7. Was ist $\mathbb{V}(aX)$ für konstantes $a \in \mathbb{R}$?
8. Definieren Sie die Kovarianz und Korrelation zweier Zufallsvariablen X und Y .
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.

Selbstkontrollfragen

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?
19. Definieren Sie das Stichprobenmittel für eine Stichprobe zweidimensionaler Zufallsvektoren.
20. Definieren Sie die Stichprobenkovarianz einer Stichprobe von zweidimensionaler Zufallsvektoren.
21. Wann ergeben sich für die Stichprobenkovarianz hohe positive oder hohe negative Werte?
22. Wann ergeben sich für die Stichprobenkovarianz Werte nahe Null?
23. Definieren Sie den Stichprobenkorrelationskoeffizient.

Lange, Tatjana, and Karl Mosler. 2017. *Statistik kompakt*. Springer-Lehrbuch. Berlin, Heidelberg: Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-662-53467-0>.



Wahrscheinlichkeitstheorie und Frequentistische Inferenz

BSc Psychologie WiSe 2021/22

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)

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

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

Bemerkungen

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

Wahrscheinlichkeitsungleichungen

Beweis

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

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

weil X nicht-negativ ist. Es folgt dann

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

Dabei gilt die erste Ungleichung weil

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

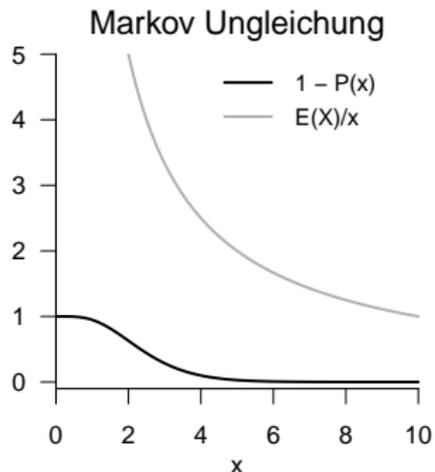
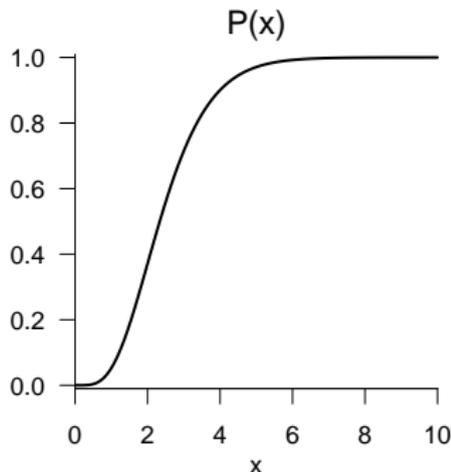
und die zweite Ungleichung gilt, weil $x \leq \xi$ für $\xi \in [x, \infty[$. Es folgt also, dass

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

□

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

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



Theorem (Chebyshev Ungleichung)

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

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

Bemerkungen

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

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

Wahrscheinlichkeitsungleichungen

Beweis

Wir halten zunächst fest, dass für $a, b \in \mathbb{R}$ gilt

$$a^2 \geq b^2 \Leftrightarrow |a| \geq |b|. \quad (7)$$

Dazu betrachten wir die folgenden vier möglichen Fälle

$$\begin{array}{llllll} |a| \geq |b| \text{ für } a > 0, b > 0 & \Leftrightarrow & a \geq b & \Leftrightarrow & a^2 \geq b^2 \\ |a| \geq |b| \text{ für } a > 0, b < 0 & \Leftrightarrow & a \geq b & \Leftrightarrow & a^2 \geq b^2 \\ |a| \geq |b| \text{ für } a < 0, b > 0 & \Leftrightarrow & -a \geq b & \Leftrightarrow & (-a)^2 \geq b^2 = a^2 \geq b^2 \\ |a| \geq |b| \text{ für } a < 0, b < 0 & \Leftrightarrow & -a \geq b & \Leftrightarrow & (-a)^2 \geq b^2 = a^2 \geq b^2 \end{array}$$

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

$$\begin{aligned} \mathbb{P}(Y \geq x^2) &\leq \frac{\mathbb{E}(Y)}{x^2} \\ \Leftrightarrow \mathbb{P}\left((X - \mathbb{E}(X))^2 \geq x^2\right) &\leq \frac{\mathbb{E}\left((X - \mathbb{E}(X))^2\right)}{x^2} \\ \Leftrightarrow \mathbb{P}(|X - \mathbb{E}(X)| \geq x) &\leq \frac{\mathbb{V}(X)}{x^2}. \end{aligned} \quad (8)$$

□

Wahrscheinlichkeitsungleichungen

Erwartungswertungleichungen

Gesetze der Großen Zahl

Zentrale Grenzwertsätze

Selbstkontrollfragen

Theorem (Cauchy-Schwarz-Ungleichung)

Es seien X, Y zwei Zufallsvariablen und $\mathbb{E}(XY)$ endlich. Dann gilt

$$\mathbb{E}(XY)^2 \leq \mathbb{E}(X^2) \mathbb{E}(Y^2) \quad (9)$$

Bemerkungen

- Analog gilt für Vektoren $x, y \in \mathbb{R}^n$, dass $\langle x, y \rangle^2 \leq \|x\| \cdot \|y\|$.
- Die Korrelationsungleichung ist eine direkte Konsequenz der Cauchy-Schwarz-Ungleichung.

Erwartungswertungleichungen

Beweis

Wir betrachten den Fall, dass $0 < \mathbb{E}(X^2) < \infty$ und $0 < \mathbb{E}(Y^2) < \infty$. Für alle $a, b \in \mathbb{R}$ gilt dann

$$0 \leq \mathbb{E} \left((aX + bY)^2 \right) \text{ und } 0 \leq \mathbb{E} \left((aX - bY)^2 \right). \quad (10)$$

Also folgt

$$0 \leq a^2 \mathbb{E}(X^2) + b^2 \mathbb{E}(Y^2) + 2ab \mathbb{E}(XY) \text{ und } 0 \leq a^2 \mathbb{E}(X^2) + b^2 \mathbb{E}(Y^2) - 2ab \mathbb{E}(XY). \quad (11)$$

Definition von $a := \sqrt{\mathbb{E}(Y^2)}$ und $b := \sqrt{\mathbb{E}(X^2)}$ ergibt dann

$$\begin{aligned} 0 &\leq \mathbb{E}(Y^2)\mathbb{E}(X^2) + \mathbb{E}(X^2)\mathbb{E}(Y^2) + 2\sqrt{\mathbb{E}(Y^2)}\sqrt{\mathbb{E}(X^2)}\mathbb{E}(XY) \\ \Leftrightarrow -2\sqrt{\mathbb{E}(Y^2)}\sqrt{\mathbb{E}(X^2)}\mathbb{E}(XY) &\leq 2\mathbb{E}(X^2)\mathbb{E}(Y^2) \\ \Leftrightarrow -\sqrt{\mathbb{E}(Y^2)\mathbb{E}(X^2)}\mathbb{E}(XY) &\leq \sqrt{\mathbb{E}(Y^2)\mathbb{E}(X^2)}\sqrt{\mathbb{E}(Y^2)\mathbb{E}(X^2)} \\ \Leftrightarrow -\mathbb{E}(XY) &\leq \sqrt{\mathbb{E}(Y^2)\mathbb{E}(X^2)}, \end{aligned} \quad (12)$$

und analog

$$\begin{aligned} 0 &\leq \mathbb{E}(Y^2)\mathbb{E}(X^2) + \mathbb{E}(X^2)\mathbb{E}(Y^2) - 2\sqrt{\mathbb{E}(Y^2)}\sqrt{\mathbb{E}(X^2)}\mathbb{E}(XY) \\ \Leftrightarrow 2\sqrt{\mathbb{E}(Y^2)}\sqrt{\mathbb{E}(X^2)}\mathbb{E}(XY) &\leq 2\mathbb{E}(X^2)\mathbb{E}(Y^2) \\ \Leftrightarrow \sqrt{\mathbb{E}(Y^2)\mathbb{E}(X^2)}\mathbb{E}(XY) &\leq \sqrt{\mathbb{E}(Y^2)\mathbb{E}(X^2)}\sqrt{\mathbb{E}(Y^2)\mathbb{E}(X^2)} \\ \Leftrightarrow \mathbb{E}(XY) &\leq \sqrt{\mathbb{E}(Y^2)\mathbb{E}(X^2)}. \end{aligned} \quad (13)$$

Zusammengenommen ergibt sich für den betrachteten Fall also die Cauchy-Schwarz-Ungleichung. Ein vollständiger Beweis findet sich bei DeGroot and Schervish (2012), Theorem 4.6.2.

Theorem (Korrelationsungleichung)

X und Y seien Zufallsvariablen mit $\mathbb{V}(X), \mathbb{V}(Y) > 0$. Dann gilt

$$\rho(X, Y)^2 = \frac{\mathbb{C}(X, Y)^2}{\mathbb{V}(X)\mathbb{V}(Y)} \leq 1. \quad (14)$$

Bemerkung

- Es gilt also $\rho(X, Y)^2 \leq 1 \Leftrightarrow |\rho(X, Y)| \leq 1 \Leftrightarrow \rho(X, Y) \in [-1, 1]$.

Beweis

Mit der Cauchy-Schwarz-Ungleichung für zwei Zufallsvariablen U und V gilt, dass

$$(\mathbb{E}(UV))^2 \leq \mathbb{E}(U^2)\mathbb{E}(V^2). \quad (15)$$

Wir definieren nun $U := X - \mathbb{E}(X)$ und $V := Y - \mathbb{E}(Y)$.

Dann besagt die Cauchy-Schwarz Ungleichung, dass

$$\mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y)))^2 \leq \mathbb{E}((X - \mathbb{E}(X))^2) \mathbb{E}((Y - \mathbb{E}(Y))^2). \quad (16)$$

Also gilt

$$\mathbb{C}(X, Y)^2 \leq \mathbb{V}(X)\mathbb{V}(Y) \Leftrightarrow \frac{\mathbb{C}(X, Y)^2}{\mathbb{V}(X)\mathbb{V}(Y)} \leq 1. \quad (17)$$

□

Theorem (Jensensche Ungleichung)

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

$$g(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda g(x_1) + (1 - \lambda)g(x_2) \quad (18)$$

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

$$\mathbb{E}(g(X)) \geq g(\mathbb{E}(X)). \quad (19)$$

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

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

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

$$\mathbb{E}(g(X)) \leq g(\mathbb{E}(X)). \quad (21)$$

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 $\mathbb{E}(\ln X) \leq \ln \mathbb{E}(X)$.

Erwartungswertungleichungen

Beweis

Wir zeigen die Ungleichung für den Fall einer konkaven Funktion g .

- Es sei f eine Tangente am Punkt $g(\mathbb{E}(X))$.
- f sei also eine linear-affine Funktion der Form

$$f(X) := aX + b \text{ für } a, b \in \mathbb{R} \text{ mit } f(\mathbb{E}(X)) = g(\mathbb{E}(X)). \quad (22)$$

- Weil g konkav ist, gilt $g(x) \leq ax + b$ für alle $x \in \mathbb{R}$, also $g(X) \leq aX + b$.
- Also gilt

$$\mathbb{E}(g(X)) \leq \mathbb{E}(aX + b) = a\mathbb{E}(X) + b = f(\mathbb{E}(X)) = g(\mathbb{E}(X)). \quad (23)$$

□

Wahrscheinlichkeitsungleichungen

Erwartungswertungleichungen

Gesetze der Großen Zahl

Zentrale Grenzwertsätze

Selbstkontrollfragen

Überblick

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

Definition (Konvergenz in Wahrscheinlichkeit)

Eine Folge von Zufallsvariable X_1, X_2, \dots *konvergiert gegen eine Zufallsvariable X in Wahrscheinlichkeit*, wenn für jedes noch so kleine $\epsilon > 0$ gilt, dass

$$\lim_{n \rightarrow \infty} \mathbb{P}(|X_n - X| < \epsilon) = 1 \Leftrightarrow \lim_{n \rightarrow \infty} \mathbb{P}(|X_n - X| \geq \epsilon) = 0 \quad (24)$$

Die Konvergenz von X_1, X_2, \dots gegen X in Wahrscheinlichkeit wird geschrieben als

$$X_n \xrightarrow[n \rightarrow \infty]{P} X \quad (25)$$

Bemerkungen

- $X_n \xrightarrow[n \rightarrow \infty]{P} X$ heißt, dass sich die Wahrscheinlichkeit, dass X_n in dem zufälligen Intervall $]X - \epsilon, X + \epsilon[$ liegt, unabhängig davon, wie klein dieses Intervall sein mag, 1 nähert, wenn n gegen Unendlich strebt.
- Intuitiv heißt das, dass sich für eine konstante Zufallsvariable $X := a$ die Verteilung von X_n mehr und mehr um a konzentriert, wenn n gegen Unendlich strebt.

Theorem (Schwaches Gesetz der Großen Zahl)

X_1, \dots, X_n seien unabhängig und gleichverteilte Zufallsvariablen mit $\mathbb{E}(X_i) = \mu$ für alle $i = 1, \dots, n$.

Weiterhin bezeichne

$$\bar{X}_n := \frac{1}{n} \sum_{i=1}^n X_i \quad (26)$$

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

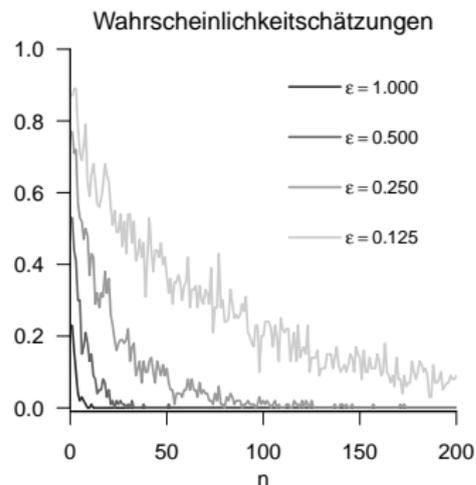
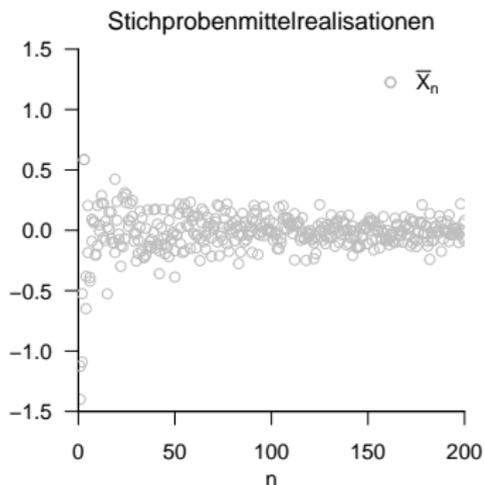
$$\bar{X}_n \xrightarrow[n \rightarrow \infty]{P} \mu. \quad (27)$$

Bemerkungen

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

Beispiel ($X_1, \dots, X_n \sim N(0, 1)$)

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



Wahrscheinlichkeitsungleichungen

Erwartungswertungleichungen

Gesetze der Großen Zahl

Zentrale Grenzwertsätze

Selbstkontrollfragen

Überblick

- Die Zentralen Grenzwertsätze besagen grob, dass die Summe von unabhängigen Zufallsvariablen asymptotisch, d.h. für unendlich viele Zufallsvariablen, normalverteilt ist.
- Modelliert man eine Messgröße Y also als Summe eines deterministischen Einflusses μ und der Summe $\varepsilon := \sum_{i=1}^n X_i$ einer Vielzahl von unabhängigen Zufallsvariablen $X_i, i = 1, \dots, n$, welche unbekannte Störeinflüsse beschreiben, so ist für großes n die Annahme

$$Y = \mu + \varepsilon \text{ mit } \varepsilon \sim N(m, s^2) \quad (28)$$

also mathematisch gerechtfertigt. Wie wir später sehen werden, liegt die Annahme in Gleichung (28) vielen statischen Modellen zugrunde.

- In der "Lindenberg und Lévy" Form des Zentralen Grenzwertsatzes werden unabhängig und identische Zufallsvariablen vorausgesetzt. In der "Liapunov" Form werden nur unabhängige Zufallsvariablen vorausgesetzt. Der Beweis der "Lindenberg und Lévy" Form ist einfacher als der Beweis der "Liapunov" Form. Wir verzichten hier aber auf die Angabe von Beweisen.
- In beiden Formulierungen des Zentralen Grenzwertsatzes die betrachtete Konvergenz von Zufallsvariablen die *Konvergenz in Verteilung*, welche wir zunächst einführen.

Definition (Konvergenz in Verteilung)

Eine Folge X_1, X_2, \dots von Zufallsvariablen *konvergiert in Verteilung gegen eine Zufallsvariable* X , wenn

$$\lim_{n \rightarrow \infty} P_{X_n}(x) = P_X(x). \quad (29)$$

für alle x an denen P_X stetig ist.

Die Konvergenz in Verteilung von X_1, X_2, \dots gegen X wird geschrieben als

$$X_n \xrightarrow[n \rightarrow \infty]{D} X, \quad (30)$$

Gilt $X_n \xrightarrow[n \rightarrow \infty]{D} X$, dann heißt die Verteilung von X die *asymptotische Verteilung der Folge* X_1, X_2, \dots .

Bemerkungen

- $X \xrightarrow[n \rightarrow \infty]{D} X$ ist eine Aussage über die Konvergenz von KVF's.
- Konvergenz in Wahrscheinlichkeit impliziert Konvergenz in Verteilung.

Theorem (Zentraler Grenzwertsatz nach Lindenberg und Lévy)

X_1, \dots, X_n seien unabhängig und identisch verteilte Zufallsvariablen mit Erwartungswert $\mu := \mathbb{E}(X_i)$ und Varianz $\sigma^2 := \mathbb{V}(X_i)$, $0 < \sigma^2 < \infty$ für alle $i = 1, \dots, n$.

Dann gilt für jedes feste x , dass

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \leq x \right) = \Phi(x), \quad (31)$$

wobei Φ KVF der Standardnormalverteilung bezeichnet.

Bemerkung

- Wir beweisen später noch folgende Implikationen des Zentralen Grenzwertsatzes:
 - Für $n \rightarrow \infty$ gilt asymptotisch $\sum_{i=1}^n X_i \sim N(n\mu, n\sigma^2)$.
 - Für $n \rightarrow \infty$ gilt asymptotisch $\bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right)$.

Zentrale Grenzwertsätze

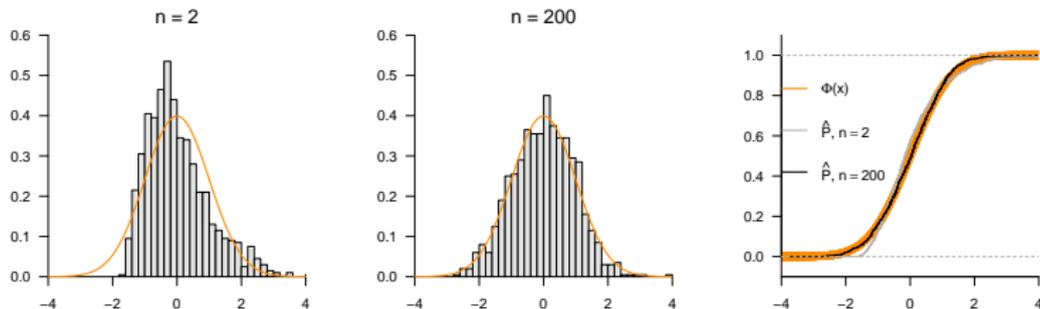
Beispiel ($X_1, \dots, X_n \sim \chi^2(k)$)

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

$$Y_n := \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \quad (32)$$

basierend auf 1000 Realisationen von Y_n für $n = 2$ und $n = 200$, sowie die WDF von $N(0, 1)$.

- Die rechte Abbildung zeigt die entsprechenden (empirischen) kumulativen Verteilungsfunktionen.



Theorem (Zentraler Grenzwertsatz nach Liapounov)

X_1, \dots, X_n seien unabhängige aber nicht notwendigerweise identisch verteilten Zufallsvariablen mit den Eigenschaften

$$\mathbb{E}(|X_i - \mu_i|^3) < \infty \text{ und } \lim_{n \rightarrow \infty} \frac{\sum_{i=1}^n \mathbb{E}(|X_i - \mu_i|^3)}{(\sum_{i=1}^n \sigma_i^2)^{3/2}} = 0. \quad (33)$$

Es sei weiterhin $\mu_i := \mathbb{E}(X_i)$ und $\sigma_i^2 = \mathbb{V}(X_i)$ für $i = 1, \dots, n$. Dann gilt für jedes feste x , dass

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\frac{\sum_{i=1}^n X_i - \sum_{i=1}^n \mu_i}{\sqrt{\sum_{i=1}^n \sigma_i^2}} \leq x \right) = \Phi(x), \quad (34)$$

wobei Φ KVF der Standardnormalverteilung bezeichnet.

Bemerkungen

- Für $n \rightarrow \infty$ gilt also asymptotisch $\sum_{i=1}^n X_i \sim N \left(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2 \right)$.
- Die Summe vieler unabhängiger Zufallsvariablen ist asymptotisch normalverteilt.

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.

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 2021/22

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 $X_i \sim N(\mu, \sigma^2)$, $i = 1, \dots, n$ durch `rnorm(n,mu,sigma)` gegeben ist.

```
rnorm(1,0,1)           #  $X_1 \sim N(0,1)$ 
```

```
[1] -1.4
```

```
rnorm(1,10,1)         #  $X_1 \sim N(10,1)$ 
```

```
[1] 10.3
```

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

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

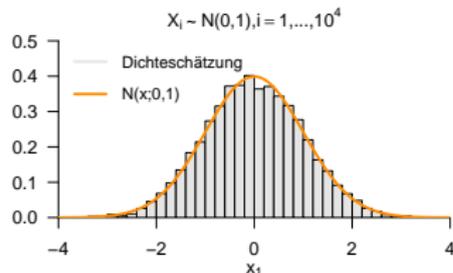
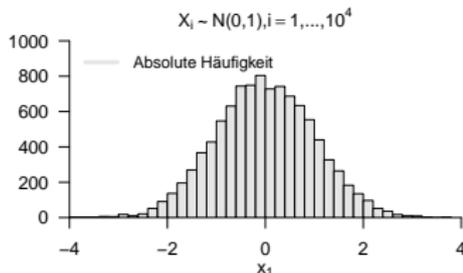
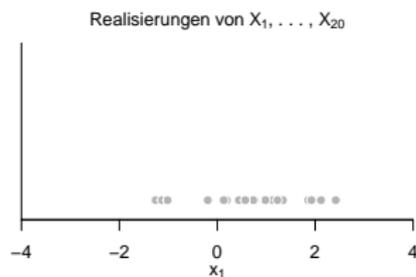
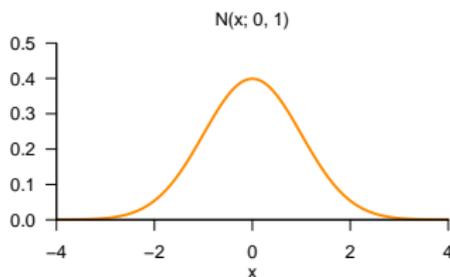
```
rnorm(1e1,5,sqrt(2))  #  $X_i \sim N(5,2)$ ,  $i = 1, \dots, 10$  (u.i.v.)
```

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

Vorbemerkungen

Realisierungen von Zufallsvariablen

Die empirische Verteilung unabhängig und identisch simulierter Zufallsvariablenrealisationen entspricht der Verteilung der Zufallsvariable. Die empirische Verteilung stellt man mit Histogrammen (Häufigkeitsverteilungen) oder histogrammbasierten 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 X normalverteilt ist, wie ist dann eine Zufallsvariable Y , die sich durch Transformation von X ergibt, verteilt?"

Für die in dieser Vorlesungseinheit behandelten Fälle gilt, dass man explizit Wahrscheinlichkeitsdichtefunktionen für die Verteilung der transformierten Zufallsvariable angeben kann. Diese gehören zu den klassischen Resultaten der frequentistischen 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. $X \sim N(0, 1)$ und ihre Transformation $Y := X^2$ und sind $x_1 = 0.10$, $x_2 = -0.20$, $x_3 = 0.80$ drei u.i.v. Realisierungen von X , so entspricht dies den u.i.v. Realisierungen $y_1 = x_1^2 = 0.01$, $y_2 = x_2^2 = 0.04$, $y_3 = x_3^2 = 0.64$ von Y . In diesem Beispiel fällt auf, dass Y keine negativen Werte annimmt, die Verteilung von Y ordnet negativen Werten daher Wahrscheinlichkeitsdichten von 0 zu.

Simulation der Transformation normalverteilter Zufallsvariablen in R

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

# Quadrieren einer Zufallsvariable
X      = rnorm(n, mu, sqrt(sigsqr)) # Realisierungen x_i, i = 1, ..., n von X
Y      = X^2                # Realisierungen y_i = x_i^2 von Y

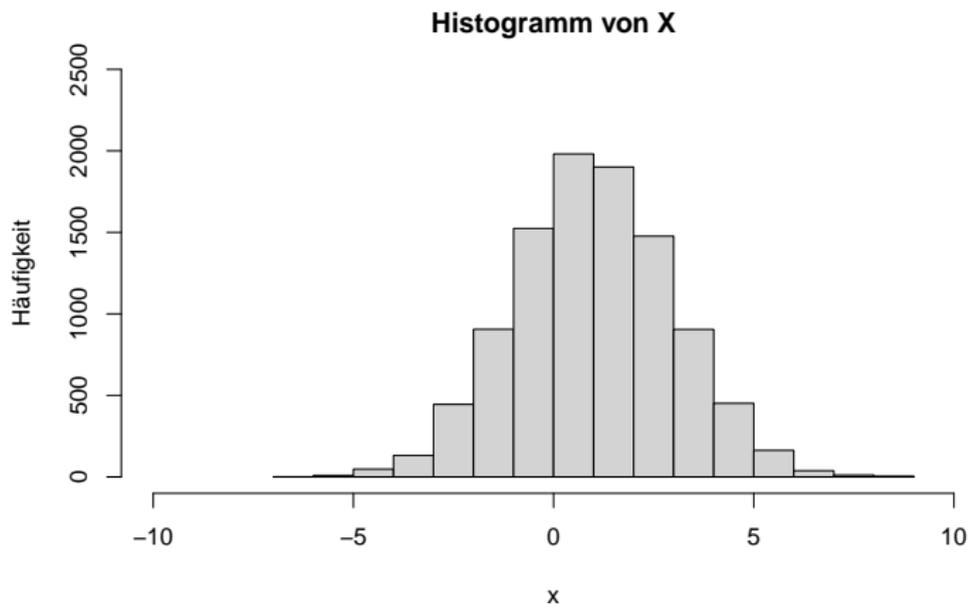
# Ausgabe der ersten acht Werte
print(X[1:8], digits = 2)
```

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

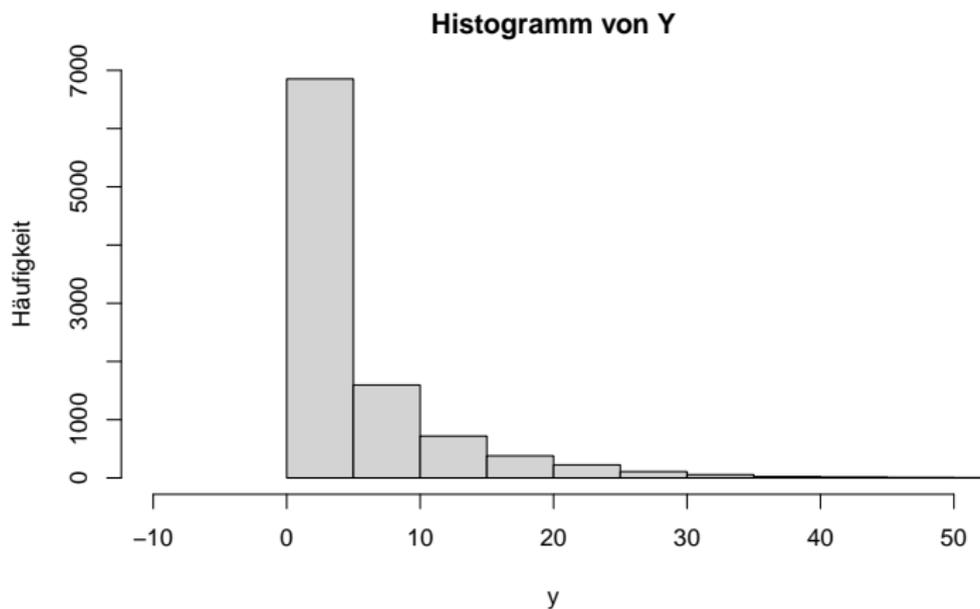
```
print(Y[1:8], digits = 2)
```

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

Simulation der Transformation normalverteilter Zufallsvariablen in R



Simulation der Transformation normalverteilter Zufallsvariablen in R



Überblick

Im Abschnitt **Transformationstheoreme** stellen wir zunächst einige generelle Werkzeuge zum Berechnen der WDFen von transformierten Zufallsvariablen bereit. Diese Werkzeuge sind von der allgemeinen Form "Wenn X eine Zufallsvariable mit WDF p_X und $Y := f(X)$ die durch f transformierte Zufallsvariable ist, dann gilt für die WDF von Y die folgende Formel: $p_Y := \{\text{Formel}\}$."

Im Abschnitt **Standardtransformationen** diskutieren wir sechs Standardtransformationen normalverteilter Zufallsvariablen, die in der frequentistischen Inferenz und damit im weiteren Verlauf des Kurses zentrale Rollen spielen. Diese Aussagen sind von der allgemeinen Form "Wenn $X_i, i = 1, \dots, n$ unabhängig und identisch normalverteilte Zufallsvariablen sind und $Y := f(X_1, \dots, X_n)$ eine Transformation dieser Zufallsvariablen ist, dann ist die WDF von Y durch die Formel $p_Y := \{\text{Formel}\}$ gegeben und man nennt die Verteilung von Y *Verteilungsname*."

Die Aussagen im Abschnitt **Standardtransformationen** sind für die 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.

Ausblick

Das probabilistische Standardmodell von n Datenpunkten hat die Form

$$X_i := \mu_i + \varepsilon_i, i = 1, \dots, n \quad (1)$$

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

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

Statistiken und Schätzer, also Funktionen von $X_i, i = 1, \dots, n$, 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_Y von $Y := f(X)$, wenn X eine Zufallsvariable mit WDF p_X ist und f eine bijektive Funktion ist.

Das **univariate WDF Transformationstheorem bei linear affinen Abbildungen** gibt eine Formel zur Berechnung der WDF p_Y von $Y := f(X)$ an, wenn X eine Zufallsvariable mit WDF p_X ist und f eine linear-affine Funktion ist.

Das **univariate WDF Transformationstheorem bei stückweisen bijektiven Abbildungen** gibt eine Formel zur Berechnung der WDF p_Y von $Y := f(X)$ an, wenn X eine Zufallsvariable mit WDF p_X ist und f zumindest in Teilen bijektiv ist.

Das **multivariate WDF Transformationstheorem bei bijektiven Abbildungen** liefert eine Formel zur Berechnung der WDF p_Y von $Y := f(X)$, wenn X ein Zufallsvektor mit WDF p_X ist und f eine bijektive multivariate vektorwertige Funktion ist.

Das **Faltungstheorem** liefert eine Formel zur Berechnung der WDF p_Y von $Y := X_1 + X_2$, wenn X_1 und X_2 zwei Zufallsvariablen mit WDFen p_{X_1} und p_{X_2} sind.

Theorem (Univariate WDF Transformation bei bijektiven Abbildungen)

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

$$Y := f(X) \quad (2)$$

wobei die univariate reellwertige Funktion $f :]a, b[\rightarrow \mathbb{R}$ differenzierbar und bijektiv auf $]a, b[$ sei. $f(]a, b[)$ sei das Bild von $]a, b[$ unter f . Schließlich sei $f^{-1}(y)$ der Wert der Umkehrfunktion von $f(x)$ für $y \in f(]a, b[)$ und $f'(x)$ sei die Ableitung von f an der Stelle x . Dann ist die WDF von Y gegeben durch

$$p_Y : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}, y \mapsto p_Y(y) := \begin{cases} \frac{1}{|f'(f^{-1}(y))|} p_X(f^{-1}(y)) & \text{für } y \in f(]a, b[) \\ 0 & \text{für } y \in \mathbb{R} \setminus f(]a, b[). \end{cases} \quad (3)$$

Bemerkungen

- Linear-affine Abbildungen sind ein wichtiger Anwendungsfall.
- Die Z -Transformation ist ein wichtiger Anwendungsfall.

Transformationstheoreme

Beweis

Wir halten zunächst fest, dass weil f eine differenzierbare bijektive Funktion auf $]a, b[$ ist, f entweder strikt wachsend oder strikt fallend ist. Nehmen wir zunächst an, dass f auf $]a, b[$ strikt wachsend ist. Dann ist auch f^{-1} für alle $y \in f(]a, b[)$ wachsend, und es gilt

$$P_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(f(X) \leq y) = \mathbb{P}\left(f^{-1}(f(X)) \leq f^{-1}(y)\right) = \mathbb{P}\left(X \leq f^{-1}(y)\right) = P_X\left(f^{-1}(y)\right).$$

P_Y ist also differenzierbar an allen Stellen y , an denen sowohl f^{-1} als auch P_X differenzierbar sind. Mit der Kettenregel und dem Satz von der Umkehrabbildung $(f^{-1}(x))' = 1/f'(f^{-1}(x))$, folgt dann, dass die WDF p_Y sich ergibt wie folgt:

$$p_Y(y) = \frac{d}{dy} P_Y(y) = \frac{d}{dy} P_X\left(f^{-1}(y)\right) = p_X\left(f^{-1}(y)\right) \frac{d}{dy} f^{-1}(y) = \frac{1}{f'\left(f^{-1}(y)\right)} p_X\left(f^{-1}(y)\right),$$

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

$$P_Y(y) = \mathbb{P}(f(X) \leq y) = \mathbb{P}\left(f^{-1}(f(X)) \geq f^{-1}(y)\right) = \mathbb{P}\left(X \geq f^{-1}(y)\right) = 1 - P_X\left(f^{-1}(y)\right),$$

Mit der Kettenregel und dem Satz von der Umkehrabbildung folgt dann

$$p_Y(y) = \frac{d}{dy} (1 - P_Y(y)) = -\frac{d}{dy} P_X\left(f^{-1}(y)\right) = -p_X\left(f^{-1}(y)\right) \frac{d}{dy} f^{-1}(y) = -\frac{1}{f'\left(f^{-1}(y)\right)} p_X\left(f^{-1}(y)\right).$$

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)

X sei eine Zufallsvariable mit WDF p_X und es sei

$$Y = f(X) \text{ mit } f(X) := aX + b \text{ f\"ur } a \neq 0. \quad (4)$$

Dann ist die WDF von Y gegeben durch

$$p_Y : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}, y \mapsto p_Y(y) := \frac{1}{|a|} p_X \left(\frac{y - b}{a} \right). \quad (5)$$

Bemerkung

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

Transformationstheoreme

Beweis

Wir halten zunächst fest, dass

$$f^{-1} : \mathbb{R} \rightarrow \mathbb{R}, y \mapsto f^{-1}(y) = \frac{y - b}{a} \quad (6)$$

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

$$f(f^{-1}(x)) = a \left(\frac{x - b}{a} \right) + b = x - b + b = x \text{ für alle } x \in \mathbb{R} \quad (7)$$

einsieht. Wir halten weiterhin fest, dass

$$f' : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto f'(x) = \frac{d}{dx}(ax + b) = a. \quad (8)$$

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

$$\begin{aligned} p_Y : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}, y \mapsto p_Y(y) &= \frac{1}{|f'(f^{-1}(y))|} p_X(f^{-1}(y)) \\ &= \frac{1}{|a|} p_X\left(\frac{y - b}{a}\right). \end{aligned} \quad (9)$$

□

Theorem (WDF Transformation bei stückweise bijektiven Abbildungen)

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

$$Y = f(X), \quad (10)$$

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

$$p_Y : \mathcal{Y} \rightarrow \mathbb{R}_{\geq 0}, y \mapsto p_Y(y) := \sum_{i=1}^k 1_{\mathcal{Y}_i}(y) \frac{1}{|f_i'(f_i^{-1}(y))|} p_X(f_i^{-1}(y)). \quad (11)$$

gegeben.

Bemerkungen

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

Theorem (Multivariate WDF Transformation bei bijektiven Abbildungen)

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

$$Y := f(X), \quad (12)$$

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

$$J^f(x) = \left(\frac{\partial}{\partial x_j} f_i(x) \right)_{1 \leq i \leq n, 1 \leq j \leq n} \in \mathbb{R}^{n \times n} \quad (13)$$

die Jacobi-Matrix von f an der Stelle $x \in \mathbb{R}^n$, $|J^f(x)|$ die Determinante von $J^f(x)$, und es sei $|J^f(x)| \neq 0$ für alle $x \in \mathbb{R}^n$. Dann ist eine WDF von Y durch

$$p_Y : \mathbb{R}^n \rightarrow \mathbb{R}_{\geq 0}, y \mapsto p_Y(y) := \begin{cases} \frac{1}{|J^f(f^{-1}(y))|} p_X(f^{-1}(y)) & \text{for } y \in f(\mathbb{R}^n) \\ 0 & \text{for } y \in \mathbb{R}^n \setminus f(\mathbb{R}^n) \end{cases} \quad (14)$$

gegeben.

Bemerkungen

- Wir verzichten auf einen Beweis.
- Es handelt sich um eine direkte Generalisierung des univariaten Falls.
- Die T - und F -Transformationen sind wichtige Anwendungsfälle.

Theorem (Summe unabhängiger Zufallsvariable, Faltung)

X_1 und X_2 seien zwei kontinuierliche unabhängige Zufallsvariablen mit WDF p_{X_1} und p_{X_2} , respektive. $Y := X_1 + X_2$ sei die Summe von X_1 und X_2 . Dann ergibt sich eine WDF der Verteilung von Y als

$$p_Y(y) = \int_{-\infty}^{\infty} p_{X_1}(y - x_2)p_{X_2}(x_2) dx_2 = \int_{-\infty}^{\infty} p_{X_1}(x_1)p_{X_2}(y - x_1) dx_1 \quad (15)$$

Die Formel für die WDF p_Y heißt *Faltung* oder *Konvolution* von p_{X_1} und p_{X_2} .

Bemerkung

- Die Summen- und Mittelwerttransformation sind wichtige Anwendungsfälle.

Transformationstheoreme

Beweis

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

$$f : \mathbb{R}^2 \rightarrow \mathbb{R}^2, x \mapsto f(x) := \begin{pmatrix} x_1 + x_2 \\ x_2 \end{pmatrix} := \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \quad (16)$$

Die inverse Funktion von f ist dann gegeben durch

$$f : \mathbb{R}^2 \rightarrow \mathbb{R}^2, z \mapsto f(z) := \begin{pmatrix} z_1 - x_2 \\ z_2 \end{pmatrix} \quad (17)$$

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

$$f^{-1}(f(x)) = f^{-1} \begin{pmatrix} x_1 + x_2 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 + x_2 - x_2 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \quad (18)$$

einsieht. Die Jacobimatrix von f ergibt sich zu

$$J^f(x) = \begin{pmatrix} \frac{\partial}{\partial x_1} f_1(x) & \frac{\partial}{\partial x_2} f_1(x) \\ \frac{\partial}{\partial x_1} f_2(x) & \frac{\partial}{\partial x_2} f_2(x) \end{pmatrix} = \begin{pmatrix} \frac{\partial}{\partial x_1} (x_1 + x_2) & \frac{\partial}{\partial x_2} (x_1 + x_2) \\ \frac{\partial}{\partial x_1} x_2 & \frac{\partial}{\partial x_2} x_2 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \quad (19)$$

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

Transformationstheoreme

Beweis (fortgeführt)

Wir halten weiterhin fest, dass die Unabhängigkeit von X_1 und X_2 impliziert, dass

$$p_{X_1, X_2}(x_1, x_2) = p_{X_1}(x_1)p_{X_2}(x_2) \quad (20)$$

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

$$\begin{aligned} p_Z(z) &= \frac{1}{|Jf(f^{-1}(z))|} p_X(f^{-1}(z)) \\ &= \frac{1}{1} p_{X_1, X_2}(z_1 - x_2, x_2) \\ &= p_{X_1}(z_1 - x_2)p_{X_2}(x_2) \end{aligned} \quad (21)$$

Integration über x_2 ergibt dann eine WDF für die marginale Verteilung von Z_1

$$p_{Z_1}(z_1) = \int_{-\infty}^{\infty} p_{X_1}(z_1 - x_2)p_{X_2}(x_2) dx_2 \quad (22)$$

Mit $Z_1 = X_1 + X_2 = Y$ ergibt sich dann die erste Form des Faltungstheorems zu

$$p_Y(y) = \int_{-\infty}^{\infty} p_{X_1}(y - x_2)p_{X_2}(x_2) dx_2. \quad (23)$$

□

Vorbemerkungen

Transformationstheoreme

Standardtransformationen

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

Selbstkontrollfragen

Überblick

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

Das **Mittelwerttransformationstheorem** besagt, dass das Stichprobenmittel unabhängig normalverteilter Zufallsvariablen wiederum normalverteilt ist und gibt die Parameter dieser Verteilung an.

Das **Z-Transformationstheorem** besagt, dass Subtraktion des Erwartungswertparameters und gleichzeitige Division mit der Wurzel des Varianzparameters die Verteilung einer normalverteilten Zufallsvariable in eine Standardnormalverteilung 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 $i = 1, \dots, n$ seien $X_i \sim N(\mu_i, \sigma_i^2)$ unabhängige normalverteilte Zufallsvariablen. Dann gilt für die Summe $Y := \sum_{i=1}^n X_i$, dass

$$Y \sim N\left(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2\right) \quad (24)$$

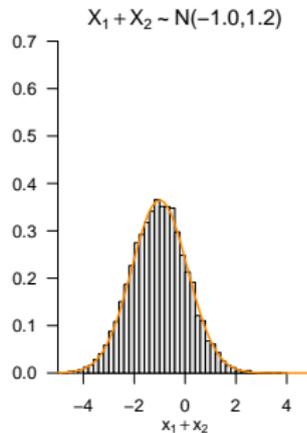
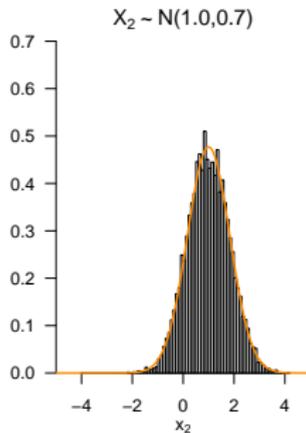
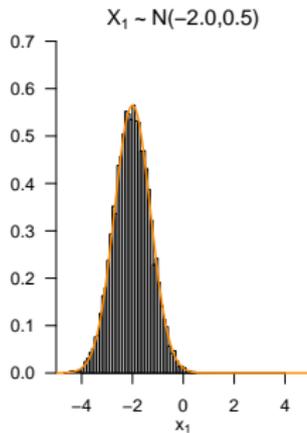
Für unabhängige und identisch normalverteilte Zufallsvariablen $X_i \sim N(\mu, \sigma^2)$ gilt folglich

$$Y \sim N(n\mu, n\sigma^2). \quad (25)$$

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 $X_1 \sim N(\mu_1, \sigma_1^2)$, $X_2 \sim N(\mu_2, \sigma_2^2)$, und $Y := X_1 + X_2$ gilt, dass $Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$. Für $n > 2$ folgt das Theorem dann durch Iteration. Mit der Definition der WDF der Normalverteilung erhalten wir zunächst

$$\begin{aligned} p_Y(y) &= \int_{-\infty}^{\infty} p_{X_1}(x_1) p_{X_2}(y - x_1) dx_1 \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{1}{2} \left(\frac{x_1 - \mu_1}{\sigma_1}\right)^2\right) \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left(-\frac{1}{2} \left(\frac{y - x_1 - \mu_2}{\sigma_2}\right)^2\right) dx_1 \\ &= \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma_1\sigma_2} \exp\left(-\frac{1}{2} \left(\frac{x_1 - \mu_1}{\sigma_1}\right)^2 - \frac{1}{2} \left(\frac{y - x_1 - \mu_2}{\sigma_2}\right)^2\right) dx_1. \end{aligned} \tag{26}$$

Mit einigem algebraischen Aufwand erhält man die Identität

$$\begin{aligned} -\frac{1}{2} \left(\frac{x_1 - \mu_1}{\sigma_1}\right)^2 - \frac{1}{2} \left(\frac{y - x_1 - \mu_2}{\sigma_2}\right)^2 \\ = -\frac{(y - \mu_1 - \mu_2)^2}{2(\sigma_1^2 + \sigma_2^2)} - \frac{((\sigma_1^2 + \sigma_2^2)x_1 - \sigma_1^2 y + \mu_2 \sigma_1^2 - \mu_1 \sigma_2^2)^2}{2\sigma_1^2 \sigma_2^2 (\sigma_1^2 + \sigma_2^2)}, \end{aligned} \tag{27}$$

so dass weiterhin gilt, dass

Beweis (fortgeführt)

$$\begin{aligned} p_Y(y) &= \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma_1\sigma_2} \exp\left(-\frac{(y - \mu_1 - \mu_2)^2}{2(\sigma_1^2 + \sigma_2^2)} - \frac{((\sigma_1^2 + \sigma_2^2)x_1 - \sigma_1^2 y + \mu_2\sigma_1^2 - \mu_1\sigma_2^2)^2}{2\sigma_1^2\sigma_2^2(\sigma_1^2 + \sigma_2^2)}\right) dx_1 \\ &= \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma_1\sigma_2} \exp\left(-\frac{(y - \mu_1 - \mu_2)^2}{2(\sigma_1^2 + \sigma_2^2)}\right) \exp\left(-\frac{((\sigma_1^2 + \sigma_2^2)x_1 - \sigma_1^2 y + \mu_2\sigma_1^2 - \mu_1\sigma_2^2)^2}{2\sigma_1^2\sigma_2^2(\sigma_1^2 + \sigma_2^2)}\right) dx_1 \\ &= \frac{1}{2\pi\sigma_1\sigma_2} \exp\left(-\frac{(y - \mu_1 - \mu_2)^2}{2(\sigma_1^2 + \sigma_2^2)}\right) \int_{-\infty}^{\infty} \exp\left(-\frac{((\sigma_1^2 + \sigma_2^2)x_1 - \sigma_1^2 y + \mu_2\sigma_1^2 - \mu_1\sigma_2^2)^2}{2\sigma_1^2\sigma_2^2(\sigma_1^2 + \sigma_2^2)}\right) dx_1. \end{aligned} \tag{28}$$

Beweis (fortgeführt)

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

$$\int_{-\infty}^{\infty} \exp\left(-\frac{((\sigma_1^2 + \sigma_2^2)x_1 - \sigma_1^2 y + \mu_2 \sigma_1^2 - \mu_1 \sigma_2^2)^2}{2\sigma_1^2 \sigma_2^2 (\sigma_1^2 + \sigma_2^2)}\right) dx_1 = \frac{\sqrt{2\pi} \sigma_1 \sigma_2}{\sqrt{\sigma_1^2 + \sigma_2^2}}. \quad (29)$$

Es ergibt sich also

$$\begin{aligned} p_Y(y) &= \frac{1}{2\pi \sigma_1 \sigma_2} \frac{\sqrt{2\pi} \sigma_1 \sigma_2}{\sqrt{\sigma_1^2 + \sigma_2^2}} \exp\left(-\frac{(y - \mu_1 - \mu_2)^2}{2(\sigma_1^2 + \sigma_2^2)}\right) \\ &= \frac{(2\pi)^{-1} (2\pi)^2}{\sqrt{\sigma_1^2 + \sigma_2^2}} \exp\left(-\frac{(y - \mu_1 - \mu_2)^2}{2(\sigma_1^2 + \sigma_2^2)}\right) \\ &= \frac{1}{\sqrt{2\pi} \sqrt{\sigma_1^2 + \sigma_2^2}} \exp\left(-\frac{(y - \mu_1 - \mu_2)^2}{2(\sigma_1^2 + \sigma_2^2)}\right). \end{aligned} \quad (30)$$

Beweis (fortgeführt)

Schließlich folgt, dass

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

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.

□

Vorbemerkungen

Transformationstheoreme

Standardtransformationen

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

Selbstkontrollfragen

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

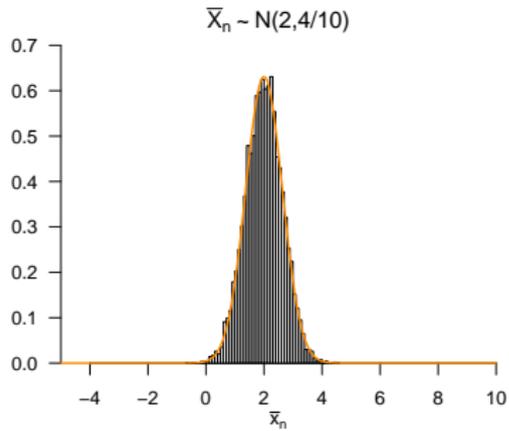
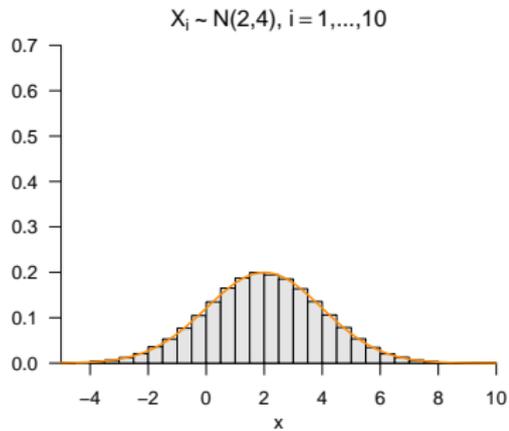
Für $i = 1, \dots, n$ seien $X_i \sim N(\mu, \sigma^2)$ unabhängig und identisch normalverteilte Zufallsvariablen. Dann gilt für das Stichprobenmittel $\bar{X}_n := \frac{1}{n} \sum_{i=1}^n X_i$, dass

$$\bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right). \quad (32)$$

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 $\bar{X}_n = \frac{1}{n} Y$ mit $Y := \sum_{i=1}^n X_i \sim N(n\mu, n\sigma^2)$. Einsetzen in das univariate WDF Transformationstheorem für lineare Funktionen ergibt dann

$$\begin{aligned} p_{\bar{X}_n}(\bar{x}_n) &= \frac{1}{|1/n|} N\left(n\bar{x}_n; n\mu, n\sigma^2\right) \\ &= \frac{n}{\sqrt{2\pi n\sigma^2}} \exp\left(-\frac{1}{2n\sigma^2} (n\bar{x}_n - n\mu)^2\right) \\ &= \frac{n}{\sqrt{2\pi n\sigma^2}} \exp\left(-\frac{1}{2n\sigma^2} (n\bar{x}_n - n\mu)^2\right) \\ &= nn^{-\frac{1}{2}} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(n\bar{x}_n)^2}{2n\sigma^2} + \frac{2(n\bar{x}_n)(n\mu)}{2n\sigma^2} - \frac{(n\mu)^2}{2n\sigma^2}\right) \\ &= \sqrt{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{n\bar{x}_n^2}{2\sigma^2} + \frac{2n\bar{x}_n\mu}{2\sigma^2} - \frac{n\mu^2}{2\sigma^2}\right) \\ &= \frac{1}{1/\sqrt{n}} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\bar{x}_n^2}{2(\sigma^2/n)} + \frac{2\bar{x}_n\mu}{2(\sigma^2/n)} - \frac{\mu^2}{2(\sigma^2/n)}\right) \\ &= \frac{1}{\sqrt{2\pi(\sigma^2/n)}} \exp\left(-\frac{1}{2(\sigma^2/n)} (\bar{x}_n - \mu)^2\right) \\ &= N\left(\bar{x}_n; \mu, \sigma^2/n\right) \end{aligned} \tag{33}$$

Vorbemerkungen

Transformationstheoreme

Standardtransformationen

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

Selbstkontrollfragen

Definition (Z -Zufallsvariable)

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

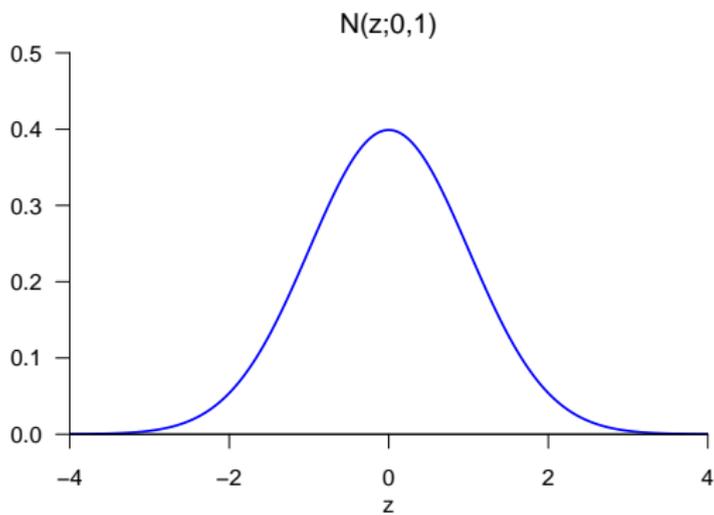
$$p : \mathbb{R} \rightarrow \mathbb{R}_{>0}, z \mapsto p(z) := \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right). \quad (34)$$

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

$$N(z; 0, 1) := \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right). \quad (35)$$

Bemerkung

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



Theorem (Z-Transformation)

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

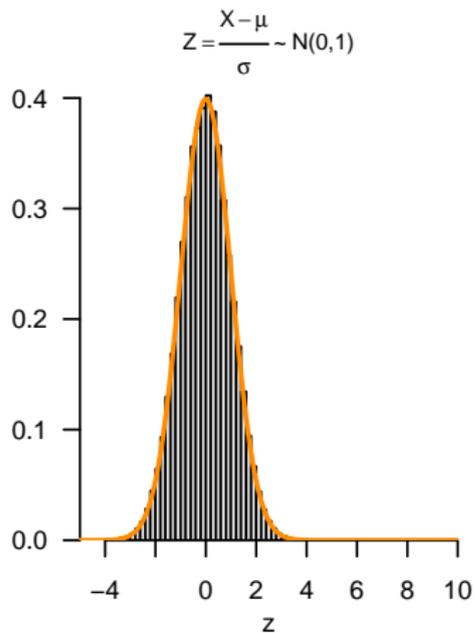
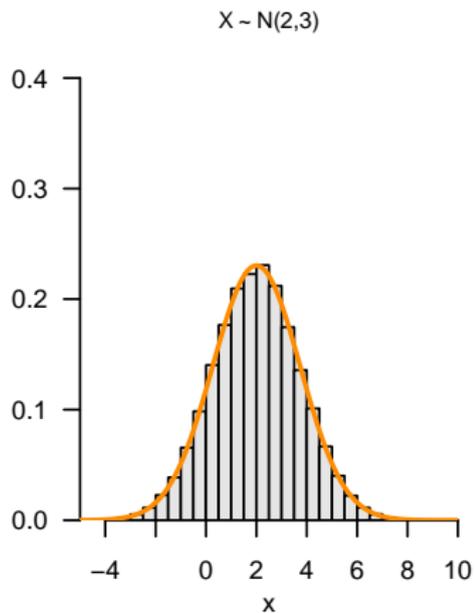
$$Z := \frac{X - \mu}{\sigma} \quad (36)$$

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

Bemerkungen

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

$$\zeta : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto \zeta(x) := \frac{x - \mu}{\sigma} =: z \quad (37)$$

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

$$\zeta^{-1} : \mathbb{R} \rightarrow \mathbb{R}, z \mapsto \zeta^{-1}(z) := \sigma z + \mu \quad (38)$$

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

$$\zeta^{-1}(z) = \zeta^{-1} \left(\frac{x - \mu}{\sigma} \right) = \frac{\sigma(x - \mu)}{\sigma} + \mu = x - \mu + \mu = x. \quad (39)$$

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

$$\zeta'(x) = \frac{d}{dx} \left(\frac{x - \mu}{\sigma} \right) = \frac{d}{dx} \left(\frac{x}{\sigma} - \frac{\mu}{\sigma} \right) = \frac{1}{\sigma}. \quad (40)$$

Beweis (fortgeführt)

Einsetzen in das univariate WDF Transformationstheorem für lineare Funktionen ergibt dann

$$\begin{aligned} p_Z(z) &= \frac{1}{|1/\sigma|} N(\sigma z + \mu; \mu, \sigma^2) \\ &= \frac{1}{1/\sqrt{\sigma^2}} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2} (\sigma z + \mu - \mu)^2\right) \\ &= \frac{\sqrt{\sigma^2}}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2} \sigma^2 z^2\right) \\ &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} z^2\right) \\ &= N(z; 0, 1) \end{aligned} \tag{41}$$

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

Definition (χ^2 -Zufallsvariable)

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

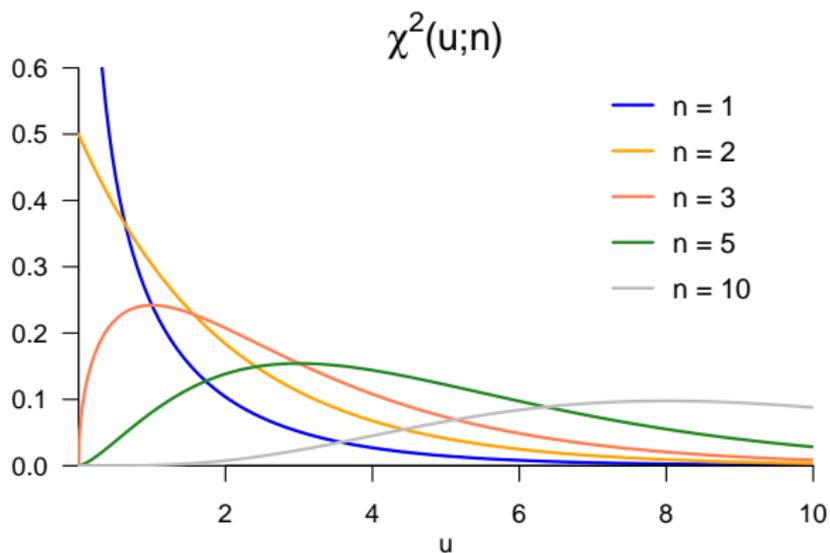
$$p : \mathbb{R}_{>0} \rightarrow \mathbb{R}_{>0}, u \mapsto p(u) := \frac{1}{\Gamma\left(\frac{n}{2}\right) 2^{\frac{n}{2}}} u^{\frac{n}{2}-1} \exp\left(-\frac{1}{2}u\right), \quad (42)$$

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

$$\chi^2(u; n) := \frac{1}{\Gamma\left(\frac{n}{2}\right) 2^{\frac{n}{2}}} u^{\frac{n}{2}-1} \exp\left(-\frac{1}{2}u\right). \quad (43)$$

Bemerkung

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



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

Theorem (χ^2 -Transformation)

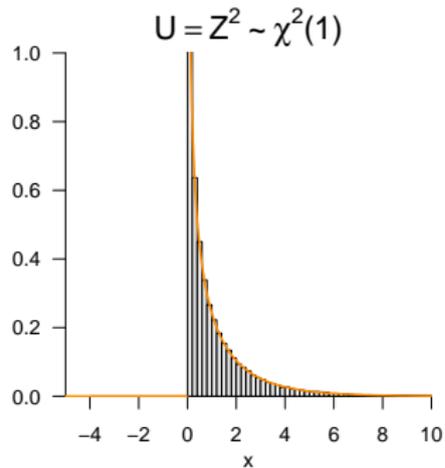
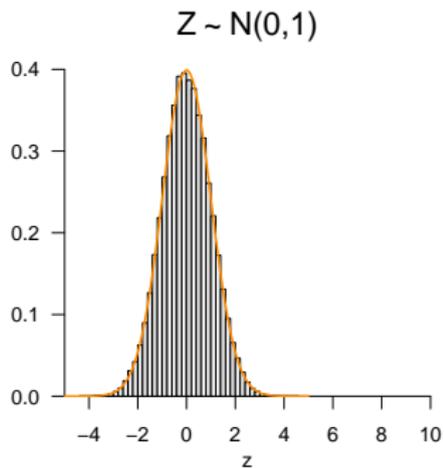
$Z_1, \dots, Z_n \sim N(0, 1)$ seien unabhängig und identisch verteilte Z -Zufallsvariablen. Dann ist die Zufallsvariable

$$U := \sum_{i=1}^n Z_i^2 \quad (44)$$

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

Bemerkungen

- Die T-Statistik und T-Tests sind wichtige Anwendungsfälle.
- Die F-Statistik und Varianzanalysen sind wichtige Anwendungsfälle.



χ^2 -Transformation

Beweis

Wir zeigen das Theorem nur für den Fall $n := 1$ mithilfe des WDF Transformationstheorems für stückweise bijektive Abbildungen. Danach ist die WDF einer Zufallsvariable $U := f(Z)$, welche aus der Transformation einer Zufallsvariable Z mit WDF p_Z durch eine stückweise bijektive Abbildung hervorgeht, gegeben durch

$$p_U(u) = \sum_{i=1}^k 1_{\mathcal{U}_i} \frac{1}{|f'_i(f_i^{-1}(u))|} p_Z(f_i^{-1}(u)). \quad (45)$$

Wir definieren

$$\mathcal{U}_1 :=] - \infty, 0[, \mathcal{U}_2 :=]0, \infty[, \text{ und } \mathcal{U}_i := \mathbb{R}_{>0} \text{ für } i = 1, 2, \quad (46)$$

sowie

$$f_i : \mathcal{Z}_i \rightarrow \mathcal{U}_i, x \mapsto f_i(z) := z^2 =: u \text{ für } i = 1, 2. \quad (47)$$

Die Ableitung und die Umkehrfunktion der f_i ergeben sich zu

$$f'_i : \mathcal{Z}_i \rightarrow \mathcal{Z}_i, x \mapsto f'_i(z) = 2z \text{ für } i = 1, 2, \quad (48)$$

und

$$f_1^{-1} : \mathcal{U}_1 \rightarrow \mathcal{U}_1, u \mapsto f_1^{-1}(u) = -\sqrt{u} \text{ und } f_2^{-1} : \mathcal{U}_2 \rightarrow \mathcal{U}_2, u \mapsto f_2^{-1}(u) = \sqrt{u}, \quad (49)$$

respektive.

Beweis (fortgeführt)

Einsetzen in Gleichung (45) ergibt dann

$$\begin{aligned} p_U(u) &= 1_{\mathcal{U}_1}(u) \frac{1}{|f_1'(f_1^{-1}(u))|} p_Z\left(f_1^{-1}(u)\right) + 1_{\mathcal{U}_2}(u) \frac{1}{|f_2'(f_2^{-1}(u))|} p_Z\left(f_2^{-1}(u)\right) \\ &= \frac{1}{|2(-\sqrt{u})|} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(-\sqrt{u})^2\right) + \frac{1}{|2(\sqrt{u})|} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(\sqrt{u})^2\right) \\ &= \frac{1}{2\sqrt{u}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}u\right) + \frac{1}{2\sqrt{u}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}u\right) \\ &= \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{u}} \exp\left(-\frac{1}{2}u\right). \end{aligned} \tag{50}$$

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

$$\frac{1}{\Gamma\left(\frac{1}{2}\right) 2^{\frac{1}{2}}} u^{\frac{1}{2}-1} \exp\left(-\frac{1}{2}u\right) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{u}} \exp\left(-\frac{1}{2}u\right) \tag{51}$$

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

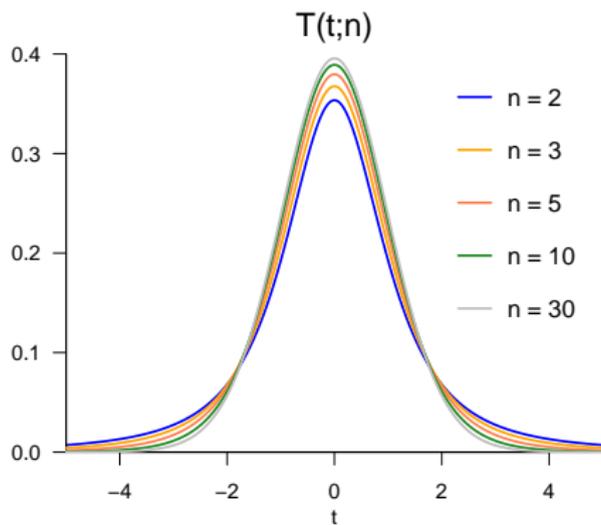
Definition (t -Zufallsvariable)

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

$$p : \mathbb{R} \rightarrow \mathbb{R}_{>0}, t \mapsto p(t) := \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}, \quad (52)$$

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

$$T(t; n) := \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}. \quad (53)$$



- Die Verteilung ist um 0 symmetrisch
- Steigendes n verschiebt Wahrscheinlichkeitsmasse aus den Ausläufen zum Zentrum.
- Ab $n = 30$ gilt $T(t; n) \approx N(0, 1)$.

Theorem (T-Transformation)

$Z \sim N(0, 1)$ sei eine Z -Zufallsvariable, $U \sim \chi^2(n)$ sei eine χ^2 -Zufallsvariable mit n Freiheitsgraden, und Z und U seien unabhängige Zufallsvariablen. Dann ist die Zufallsvariable

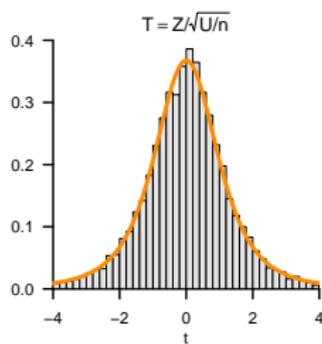
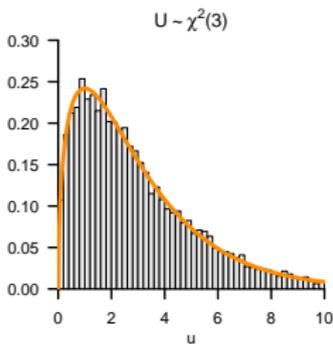
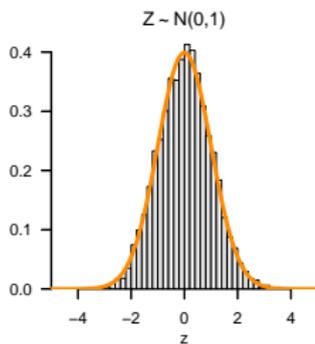
$$T := \frac{Z}{\sqrt{U/n}} \quad (54)$$

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-Statistik und T-Tests sind wichtige Anwendungsfälle.

T-Transformation



T-Transformation

Beweis

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

$$p_{Z,U}(z, u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right) \frac{1}{\Gamma(\frac{n}{2})2^{\frac{n}{2}}} u^{\frac{n}{2}-1} \exp\left(-\frac{1}{2}u\right). \quad (55)$$

gegeben ist. Wir betrachten dann die multivariate vektorwertige Abbildung

$$f : \mathbb{R}^2 \rightarrow \mathbb{R}^2, (z, u) \mapsto f(z, u) := \left(\frac{z}{\sqrt{u/n}}, u \right) =: (t, w) \quad (56)$$

und benutzen das multivariate WDF Transformationstheorem für bijektive Abbildungen um die WDF von (t, w) herzuleiten. Dazu erinnern wir uns, dass wenn X ein n -dimensionaler Zufallsvektor mit WDF p_X und $Y := f(X)$ für eine differenzierbare und bijektive Abbildung $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ ist, die WDF des Zufallsvektors Y durch

$$p_Y : \mathbb{R}^n \rightarrow \mathbb{R}_{\geq 0}, y \mapsto p_Y(y) := \frac{1}{|Jf(f^{-1}(y))|} p_X(f^{-1}(y)) \quad (57)$$

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

$$f^{-1} : \mathbb{R}^2 \rightarrow \mathbb{R}^2, (t, w) \mapsto f^{-1}(t, w) := \left(\sqrt{w/nt}, w \right). \quad (58)$$

T-Transformation

Beweis (fortgeführt)

Dies ergibt sich direkt aus

$$f^{-1}(f(z, u)) = f^{-1}\left(\frac{z}{\sqrt{u/n}}, u\right) = \left(\frac{\sqrt{u/n}z}{\sqrt{u/n}}, u\right) = (z, u) \text{ für alle } (z, u) \in \mathbb{R}^2. \quad (59)$$

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

$$|J^f(z, u)| = \left| \begin{array}{cc} \frac{\partial}{\partial z} \left(\frac{z}{\sqrt{u/n}} \right) & \frac{\partial}{\partial u} \left(\frac{z}{\sqrt{u/n}} \right) \\ \frac{\partial}{\partial z} u & \frac{\partial}{\partial u} u \end{array} \right| = \left(\frac{v}{n} \right)^{-1/2}, \quad (60)$$

gegeben ist, sodass folgt, dass

$$\frac{1}{|J^f(f^{-1}(z, u))|} = \left(\frac{w}{n} \right)^{1/2}. \quad (61)$$

Einsetzen in Gleichung (57) ergibt dann

$$p_{T,W}(t, w) = \left(\frac{w}{n} \right)^{1/2} p_{Z,V} \left(\sqrt{w/nt}, w \right), \quad (62)$$

T-Transformation

Beweis (fortgeführt)

Es folgt also

$$\begin{aligned} p_T(t) &= \int_0^\infty p_{T,W}(t, w) dw \\ &= \int_0^\infty \left(\frac{w}{n}\right)^{1/2} p_{Z,V}\left(\sqrt{w/nt}, w\right) dw \\ &= \int_0^\infty \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\sqrt{w/nt}\right)^2\right) \frac{1}{\Gamma\left(\frac{n}{2}\right)2^{\frac{n}{2}}} w^{\frac{n}{2}-1} \exp\left(-\frac{1}{2}w\right) \left(\frac{w}{n}\right)^{1/2} dw \\ &= \frac{1}{\sqrt{2\pi}} \frac{1}{\Gamma\left(\frac{n}{2}\right)2^{\frac{n}{2}} n^{\frac{1}{2}}} \int_0^\infty \exp\left(-\frac{1}{2}\frac{w}{n}t^2\right) w^{\frac{n}{2}-1} \exp\left(-\frac{1}{2}w\right) w^{1/2} dw \\ &= \frac{1}{\sqrt{2\pi}} \frac{1}{\Gamma\left(\frac{n}{2}\right)2^{\frac{n}{2}} n^{\frac{1}{2}}} \int_0^\infty \exp\left(-\frac{1}{2}\frac{w}{n}t^2 - \frac{1}{2}w\right) w^{\frac{n}{2}-1} w^{\frac{1}{2}} dw \\ &= \frac{1}{\sqrt{2\pi}} \frac{1}{\Gamma\left(\frac{n}{2}\right)2^{\frac{n}{2}} n^{\frac{1}{2}}} \int_0^\infty \exp\left(-\frac{1}{2}\left(\frac{w}{n}t^2 + w\right)\right) w^{\frac{n+1}{2}-1} dw \\ &= \frac{1}{\sqrt{2\pi}} \frac{1}{\Gamma\left(\frac{n}{2}\right)2^{\frac{n}{2}} n^{\frac{1}{2}}} \int_0^\infty \exp\left(-\frac{1}{2}\left(1 + \frac{t^2}{n}\right)w\right) w^{\frac{n+1}{2}-1} dw \end{aligned} \tag{63}$$

Beweis (fortgeführt)

Wir stellen dann fest, dass der Integrand auf der linken Seite der obigen Gleichung dem Kern einer Gamma WDF mit Parametern $\alpha = \frac{n+1}{2}$ und $\beta = \frac{2}{1 + \frac{t^2}{n}}$ entspricht, wie man leicht einsieht:

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

Beweis (fortgeführt)

Es ergibt sich also

$$p_T(t) = \frac{1}{\sqrt{2\pi}} \frac{1}{\Gamma(\frac{n}{2}) 2^{\frac{n}{2}} n^{\frac{1}{2}}} \int_0^\infty \Gamma\left(w; \frac{n+1}{2}, \frac{2}{1 + \frac{t^2}{n}}\right) dw. \quad (64)$$

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

$$p_T(t) = \frac{1}{\sqrt{2\pi}} \frac{1}{\Gamma(\frac{n}{2}) 2^{\frac{n}{2}} n^{\frac{1}{2}}} \Gamma\left(\frac{n+1}{2}\right) \left(\frac{2}{1 + \frac{t^2}{n}}\right)^{\frac{n+1}{2}}. \quad (65)$$

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

□

Vorbemerkungen

Transformationstheoreme

Standardtransformationen

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

Selbstkontrollfragen

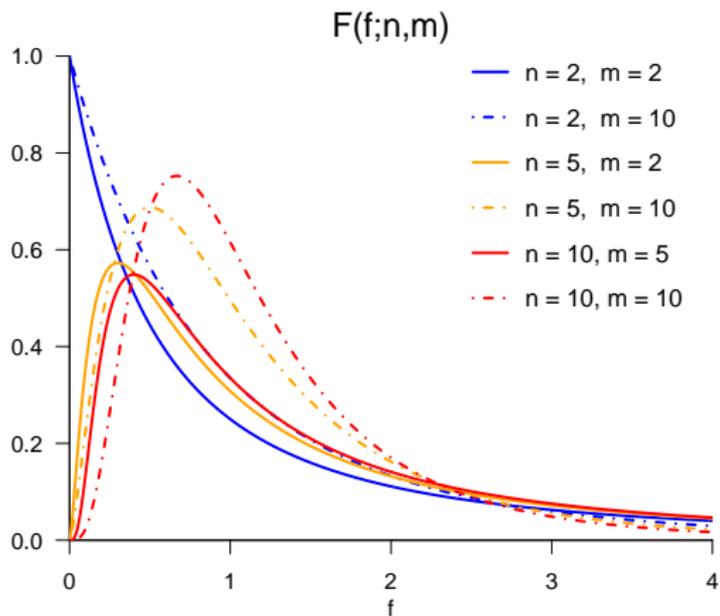
Definition (f -Zufallsvariable)

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

$$p_F : \mathbb{R} \rightarrow \mathbb{R}_{>0}, f \mapsto p_F(f) := m^{\frac{m}{2}} n^{\frac{n}{2}} \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right) \Gamma\left(\frac{n}{2}\right)} \frac{f^{\frac{m}{2}-1}}{\left(1 + \frac{m}{n} f\right)^{\frac{m+n}{2}}}, \quad (66)$$

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

$$F(f; n, m) := m^{\frac{m}{2}} n^{\frac{n}{2}} \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right) \Gamma\left(\frac{n}{2}\right)} \frac{f^{\frac{m}{2}-1}}{\left(1 + \frac{m}{n} f\right)^{\frac{m+n}{2}}}. \quad (67)$$



Theorem (F -Transformation)

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

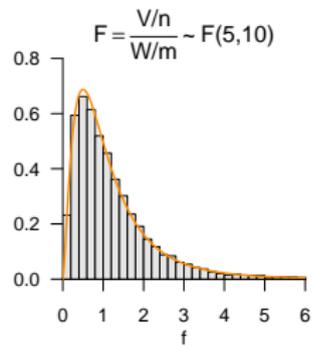
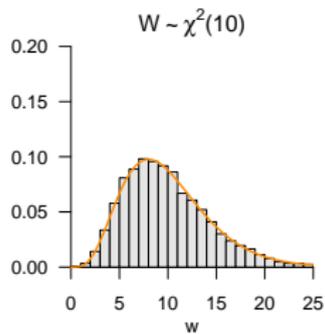
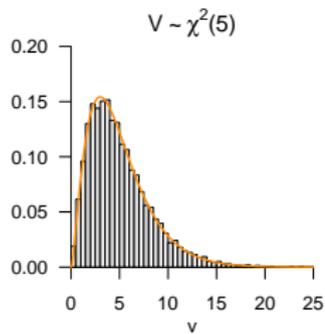
$$F := \frac{V/n}{W/m} \quad (68)$$

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 2021/22

Prof. Dr. Dirk Ostwald

(9) Grundbegriffe Frequentistischer Inferenz

Wahrscheinlichkeitstheorie und Frequentistische Inferenz

Datum	Einheit	Thema
14.10.2021	Einführung	(1) Einführung
21.10.2021	Wahrscheinlichkeitstheorie	(2) Wahrscheinlichkeitsräume
28.10.2021	Wahrscheinlichkeitstheorie	(3) Elementare Wahrscheinlichkeiten
04.11.2021	Wahrscheinlichkeitstheorie	(4) Zufallsvariablen
11.11.2021	Wahrscheinlichkeitstheorie	(5) Multivariate Verteilungen
18.11.2021	Wahrscheinlichkeitstheorie	(6) Erwartungswert, Varianz, Kovarianz
25.11.2021	Wahrscheinlichkeitstheorie	(7) Ungleichungen und Grenzwerte
02.12.2021	Wahrscheinlichkeitstheorie	(8) Transformationen der Normalverteilung
09.12.2021	Frequentistische Inferenz	(9) Grundbegriffe Frequentistischer Inferenz
16.12.2021	Frequentistische Inferenz	(10) Parameterschätzung
	Weihnachtspause	
	Frequentistische Inferenz	(11) Konfidenzintervalle
13.01.2022	Frequentistische Inferenz	(12) Hypothesentests
20.01.2022	Frequentistische Inferenz	(13) Einstichproben-T-Tests
27.01.2022	Frequentistische Inferenz	(14) Zweistichproben-T-Tests
31.01.2022	Klausur	16 - 17 Uhr, G44 - H6
Jul 2022	Klausurwiederholungstermin	

Evidenzbasierte Evaluation von Psychotherapieformen bei Depression

Online Psychotherapie



Klassische Psychotherapie



Becks Depression Inventar (BDI) zur Depressionsdiagnostik

Evidenzbasierte Evaluation von Psychotherapieformen bei Depression

Experimentelle Bedingung
(Gruppen von $n = 50$)

Psychotherapie

Klassisch

Pre-BDI



Post-BDI

Online

Pre-BDI



Post-BDI

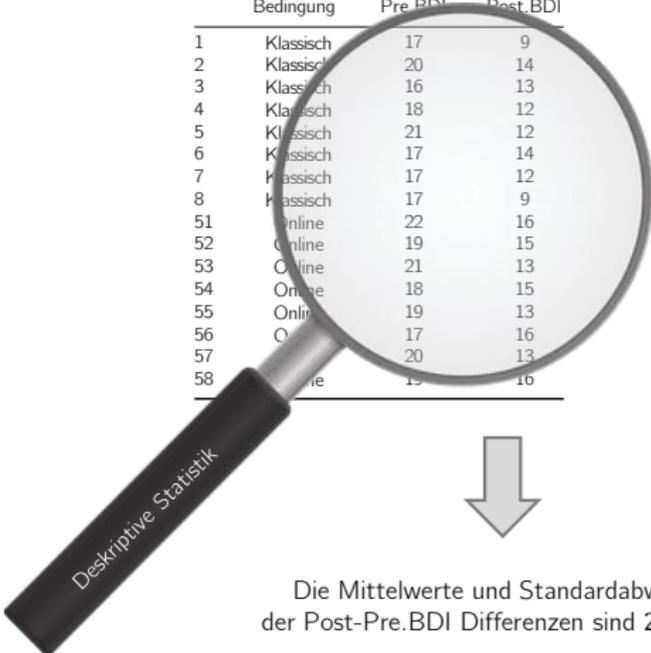
Einlesen des Datensatzes mit read.table()

```
fname = file.path(getwd(), "9_Daten", "psychotherapie_datensatz.csv")  
D = read.table(fname, sep = ",")
```

Daten der ersten acht Proband:innen jeder Gruppe

	Bedingung	Pre.BDI	Post.BDI
1	Klassisch	17	9
2	Klassisch	20	14
3	Klassisch	16	13
4	Klassisch	18	12
5	Klassisch	21	12
6	Klassisch	17	14
7	Klassisch	17	12
8	Klassisch	17	9
51	Online	22	16
52	Online	19	15
53	Online	21	13
54	Online	18	15
55	Online	19	13
56	Online	17	16
57	Online	20	13
58	Online	19	16

Deskriptive Statistik



	Bedingung	Pre.BDI	Post.BDI
1	Klassisch	17	9
2	Klassisch	20	14
3	Klassisch	16	13
4	Klassisch	18	12
5	Klassisch	21	12
6	Klassisch	17	14
7	Klassisch	17	12
8	Klassisch	17	9
51	Online	22	16
52	Online	19	15
53	Online	21	13
54	Online	18	15
55	Online	19	13
56	Online	17	16
57	Online	20	13
58	Online	19	16



Die Mittelwerte und Standardabweichung
der Post-Pre.BDI Differenzen sind 2.89 ± 1.67 .

Frequentistische Statistik

Wahrscheinlichkeitstheorie

Probabilistisches Modell

$$(\Omega, \mathcal{A}, \mathbb{P}), X: \Omega \rightarrow \mathbb{R}$$

Wahrscheinlichkeitsrechnung



Frequentistische Inferenzstatistik

	Bedingung	Pre BDI	Post BDI
1	Klassisch	17	9
2	Klassisch	20	14
3	Klassisch	16	13
4	Klassisch	18	12
5	Klassisch	21	12
6	Klassisch	17	14
7	Klassisch	17	12
8	Klassisch	17	9
51	Online	22	16
52	Online	19	15
53	Online	21	13
54	Online	18	15
55	Online	19	13
56	Online	17	16
57	Online	20	13
58	Online	19	16



Wir lehnen die Nullhypothese keines Unterschiedes zwischen den Therapiebedingungen ab ($T = 2.7, p < 0.05$).

Bayesianische Statistik

Wahrscheinlichkeitstheorie

Probabilistisches Modell

$$(\Omega, \mathcal{A}, \mathbb{P}), X: \Omega \rightarrow \mathbb{R}$$

Wahrscheinlichkeitsrechnung



	Bedingung	Pre_BDI	Post_BDI
1	Klassisch	17	9
2	Klassisch	20	14
3	Klassisch	16	13
4	Klassisch	18	12
5	Klassisch	21	12
6	Klassisch	17	14
7	Klassisch	17	12
8	Klassisch	17	9
51	Online	22	16
52	Online	19	15
53	Online	21	13
54	Online	18	15
55	Online	19	13
56	Online	17	16
57	Online	20	13
58	Online	19	16



Die Wahrscheinlichkeit für einen Unterschied der Erwartungswertparameter ist größer als 0.63.

Maschinelles Lernen

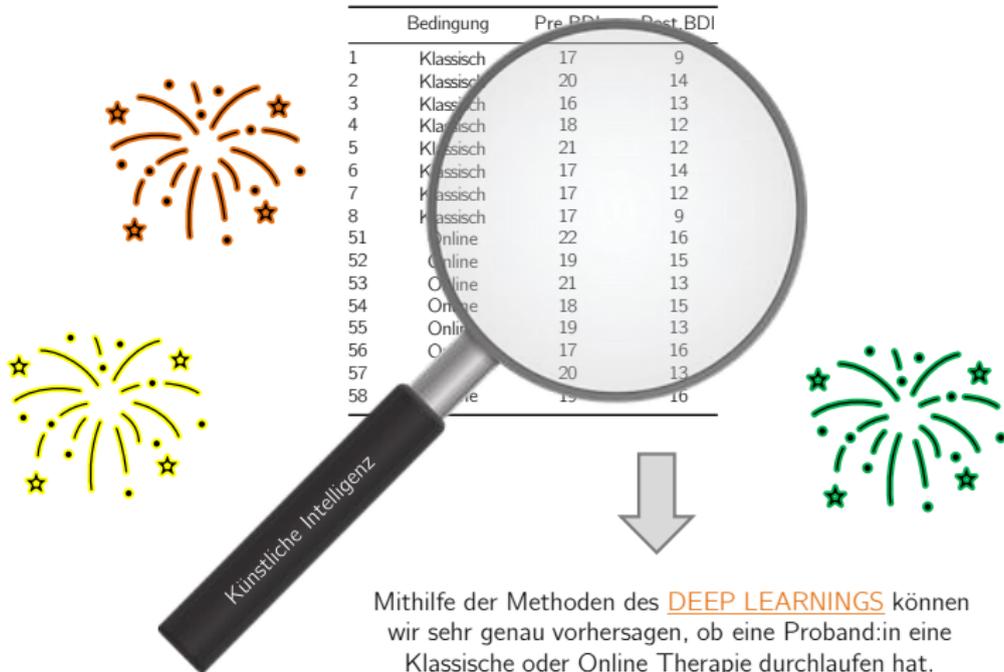


	Bedingung	Pre_BDI	Post_BDI
1	Klassisch	17	9
2	Klassisch	20	14
3	Klassisch	16	13
4	Klassisch	18	12
5	Klassisch	21	12
6	Klassisch	17	14
7	Klassisch	17	12
8	Klassisch	17	9
51	Online	22	16
52	Online	19	15
53	Online	21	13
54	Online	18	15
55	Online	19	13
56	Online	17	16
57	Online	20	13
58	Online	19	16



Wir können mit 87% Genauigkeit vorhersagen, ob eine Proband:in eine Klassische oder Online Therapie durchlaufen hat.

Künstliche Intelligenz



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

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

bestehend aus einem *Stichprobenraum* \mathcal{X} , einer σ -Algebra \mathcal{A} auf \mathcal{X} und einer mindestens zweielementigen Menge $\{\mathbb{P}_\theta | \theta \in \Theta\}$ von Wahrscheinlichkeitsmaßen auf $(\mathcal{X}, \mathcal{A})$, die mit einer Menge Θ indiziert sind.

Bemerkungen

- Im Gegensatz zum W -Raum betrachten wir zwei oder mehr W -Maße.
- Das jeweils zugrundeliegende Wahrscheinlichkeitsmaß ist mit $\theta \in \Theta$ indiziert.
- Für Erwartungswerte und (Ko)Varianzen bezüglich \mathbb{P}_θ schreiben wir $\mathbb{E}_\theta, \mathbb{V}_\theta, \mathbb{C}_\theta$.

Definition (Parametrische, Standard- und Produktmodelle)

- Ein statistisches Modell \mathcal{M} heißt ein *parametrisches Modell*, wenn $\Theta \subset \mathbb{R}^d$ für ein $d \in \mathbb{N}$ ist. Für $d = 1$ heißt \mathcal{M} ein *einparametrisches Modell*.
- Ein statistisches Modell \mathcal{M} heißt ein *diskretes Modell*, wenn \mathcal{X} diskret ist. Jedes \mathbb{P}_θ besitzt dann eine WMF p_θ definiert durch $p_\theta := \mathbb{P}_\theta(x)$. Ein statistisches Modell \mathcal{M} heißt ein *stetiges Modell*, wenn $\mathcal{X} \subset \mathbb{R}^n$ ist und jedes \mathbb{P}_θ eine WDF p_θ besitzt.
- Für ein statistisches Modell $\mathcal{M}_0 := (\mathcal{X}_0, \mathcal{A}_0, \{\mathbb{P}_\theta^0 | \theta \in \Theta\})$ und $n > 1$ heißt das statistische Modell $\mathcal{M} := (\mathcal{X}, \mathcal{A}, \{\mathbb{P}_\theta | \theta \in \Theta\})$, für welches $\mathcal{X} := \times_{i=1}^n \mathcal{X}_0$ das n -fache kartesische Produkt von \mathcal{X}_0 mit sich selbst ist, \mathcal{A} die entsprechende Produkt- σ -Algebra ist, und $\{\mathbb{P}_\theta | \theta \in \Theta\}$ die entsprechende Menge an Produktmaßen $\prod_{i=1}^n \mathbb{P}_\theta^0$ ist, das zu \mathcal{M}_0 gehörige *Produktmodell*.

Bemerkungen

- Der Vorgang des Beobachtens, der ein zufälliges Ergebnis liefert, wird mit Zufallsvektoren X , die Werte in \mathcal{X} annehmen, beschrieben. Im Kontext statistischer Modelle nennt man diese Zufallsvektoren *Beobachtung*, *Messung*, oder *Stichprobe*. Ihre Realisierungen $x \in \mathcal{X}$, also konkret vorliegende Datenwerte, werden *Beobachtungswert*, *Messwert*, *Stichprobenwert* oder *Datensatz* genannt.
- Produktmodelle modellieren die unabhängige Wiederholung eines Einzelexperimentes. Der entsprechende Zufallsvektor $X := (X_1, \dots, X_n)$ entspricht dann einer Menge von unabhängigen Zufallsvariablen/vektoren $X_1, \dots, X_n \sim \mathbb{P}_\theta^0$.
- Wenn für ein Produktmodell die Menge \mathcal{X}_0 eindimensional ist, also z.B. $\mathcal{X}_0 = \mathbb{R}$ gilt, spricht man von einem *univariaten statistischen Modell*. Wenn für ein Produktmodell die Menge \mathcal{X}_0 mehrdimensional ist, also z.B. $\mathcal{X}_0 = \mathbb{R}^m$, $m > 1$ ist, spricht man auch von *multivariaten statistischen Modellen*. Wir betrachten in diesem Kurs nur univariate statistische Modelle.
- Die am häufigsten angewendeten statistischen Modelle sind parametrische stetige Produktmodelle.
- In einem konkreten Datenanalyseproblem nimmt man an, dass die beobachteten Stichprobenwerte $x = (x_1, \dots, x_n)$ der Stichprobe $X = (X_1, \dots, X_n)$ durch $\mathbb{P}_{\hat{\theta}}$ generiert wurde, wobei $\hat{\theta}$ hier den *wahren, aber unbekanntem, Parameterwert* bezeichnet. Der wahre, aber unbekanntem, Parameterwert $\tilde{\theta}$ bleibt auch nach der statistischen Analyse unbekannt.
- In der mathematischen Analyse von Inferenzmethoden betrachtet man alle möglichen wahren, aber unbekanntem, Parameterwerte, schreibt also einfach $\{\mathbb{P}_\theta \mid \theta \in \Theta\}$.

Beispiel (1)

Definition (Bernoullimodell)

Das univariate parametrische Produktmodell

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

mit

$$\mathcal{X} := \{0, 1\}^n, \mathcal{A} := \mathcal{P}(\{0, 1\}^n), \theta := \mu, \Theta :=]0, 1[, \quad (3)$$

also

$$\{\mathbb{P}_\theta | \theta \in \Theta\} := \left\{ \prod_{i=1}^n \text{Bern}(\mu) | \mu \in]0, 1[\right\}, \quad (4)$$

und damit

$$X_1, \dots, X_n \sim \text{Bern}(\mu) \text{ mit } \mu \in]0, 1[, \quad (5)$$

heißt *Bernoullimodell*.

Bemerkung

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

Beispiel (2)

Definition (Normalverteilungsmodell)

Das univariate parametrische Produktmodell

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

mit

$$\mathcal{X} := \mathbb{R}^n, \mathcal{A} := \mathcal{B}(\mathbb{R}^n), \theta := (\mu, \sigma^2), \Theta := \mathbb{R} \times \mathbb{R}_{>0}, \quad (7)$$

also

$$\{\mathbb{P}_\theta | \theta \in \Theta\} := \left\{ \prod_{i=1}^n N(\mu, \sigma^2) | (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_{>0} \right\}, \quad (8)$$

und damit

$$X_1, \dots, X_n \sim N(\mu, \sigma^2) \text{ mit } (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_{>0} \quad (9)$$

heißt Normalverteilungsmodell

Bemerkungen

- Das Normalverteilungsmodell ist Grundlage der allermeisten populären statistischen Verfahren.
- Diese Verfahren werden im Allgemeinen Linearen Modell integrativ betrachtet.
- Das Modell ist grundlegend durch normalverteiltes Rauschen, nicht "normalverteilte Populationen," motiviert.

Statistische Modelle

Statistiken und Schätzer

Standardprobleme Frequentistischer Inferenz

Selbstkontrollfragen

Definition (Statistik)

$\mathcal{M} := (\mathcal{X}, \mathcal{A}, \{\mathbb{P}_\theta | \theta \in \Theta\})$ sei ein statistisches Modell und (Σ, \mathcal{S}) sei ein Messraum. Dann wird eine Zufallsvariable

$$S : \mathcal{X} \rightarrow \Sigma \quad (10)$$

Statistik genannt.

Bemerkungen

- Beobachtungen und Statistiken werden durch Zufallsvariablen modelliert.
- Beobachtungen modellieren die stochastische Generation von Daten
- Statistiken modellieren von Datenanalyt:innen konstruierte Funktionen, die aufgrund von Beobachtungen essentielle Größen extrahieren, aus denen sich sinnvolle Schlüsse ziehen lassen.

Beispiele (Statistiken)

\mathcal{M} sei das Normalverteilungsmodell. Dann sind zum Beispiel folgende Zufallsvariablen Statistiken:

- Das *Stichprobenmittel*

$$\bar{X}_n : \mathbb{R}^n \rightarrow \mathbb{R}, X \mapsto \bar{X}_n(X) := \frac{1}{n} \sum_{i=1}^n X_i, \quad (11)$$

- Die *Stichprobenvarianz*

$$S_n^2 : \mathbb{R}^n \rightarrow \mathbb{R}_{\geq 0}, X \mapsto S_n^2(X) := \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2, \quad (12)$$

- Die *Stichprobenstandardabweichung*

$$S_n : \mathbb{R}^n \rightarrow \mathbb{R}_{\geq 0}, X \mapsto S_n(X) := \sqrt{S_n^2}, \quad (13)$$

- Die *T-Statistik*

$$T : \mathbb{R}^n \rightarrow \mathbb{R}, X \mapsto T(X) := \sqrt{n} \left(\frac{\bar{X}_n}{S_n} \right). \quad (14)$$

Definition (Schätzer)

$\mathcal{M} := (\mathcal{X}, \mathcal{A}, \{\mathbb{P}_\theta | \theta \in \Theta\})$, (Σ, \mathcal{S}) sei ein Messraum, und $\tau : \Theta \rightarrow \Sigma$ sei eine Abbildung, die jedem $\theta \in \Theta$ eine Kenngröße $\tau(\theta) \in \Sigma$ zuordnet. Eine Statistik

$$\hat{\tau} : \mathcal{X} \rightarrow \Sigma \tag{15}$$

heißt dann ein *Schätzer* für τ .

Bemerkungen

- Typische Beispiele für τ sind
 - $\tau(\theta) := \theta$ für die Schätzung von θ ,
 - $\tau(\theta) := \theta_i$ mit $\theta \in \mathbb{R}^d$, $d > 1$ für die Schätzung einer Komponente von θ ,
 - $\tau(\theta) := \mathbb{E}_\theta(X_0)$ für die Schätzung des Erwartungswerts einer Beobachtung,
 - $\tau(\theta) := \mathbb{V}_\theta(X_0)$ für die Schätzung der Varianz einer Beobachtung.
- Schätzer nehmen Zahlwerte in Σ an und heißen deshalb auch *Punktschätzer*.
- Nicht jeder Schätzer ist ein guter Schätzer, man definiert deshalb *Schätzgütekriterien*.
- Für $\hat{\tau}$ bei $\tau(\theta) := \theta$ schreibt man auch $\hat{\theta}$

Beispiele (Schätzer)

\mathcal{M} sei das Normalverteilungsmodell.

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

$$\tau : \mathbb{R} \times \mathbb{R}_{>0} \rightarrow \mathbb{R}, (\mu, \sigma^2) \mapsto \tau(\mu, \sigma^2) := \mu. \quad (16)$$

Ebenso ist \bar{X}_n ein Schätzer für

$$\tau : \mathbb{R} \times \mathbb{R}_{>0} \rightarrow \mathbb{R}, (\mu, \sigma^2) \mapsto \tau(\mu, \sigma^2) := \mathbb{E}_{\mu, \sigma^2}(X_0) = \mu. \quad (17)$$

- Weiterhin ist die konstante Funktion

$$\hat{\tau} : \mathbb{R}^n \rightarrow \mathbb{R}, x \mapsto \hat{\tau}(X) := 42 \quad (18)$$

ein Schätzer für

$$\tau : \mathbb{R} \times \mathbb{R}_{>0} \rightarrow \mathbb{R}_{>0}, (\mu, \sigma^2) \mapsto \tau(\mu, \sigma^2) := \sigma^2. \quad (19)$$

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

Gütekriterien für Schätzer sind der Inhalt von Vorlesungseinheit (10) Schätzereigenschaften.

Statistische Modelle

Statistiken und Schätzer

Standardprobleme Frequentistischer Inferenz

Selbstkontrollfragen

(1) Parameterschätzung

Ziel der Parameterschätzung ist es, einen möglichst guten Tipp für den wahren, aber unbekanntem, Parameterwert (oder eine Funktion dessen) abzugeben, typischerweise basierend auf der Beobachtung einer Realisierung von $X_1, \dots, X_n \sim p_\theta$.

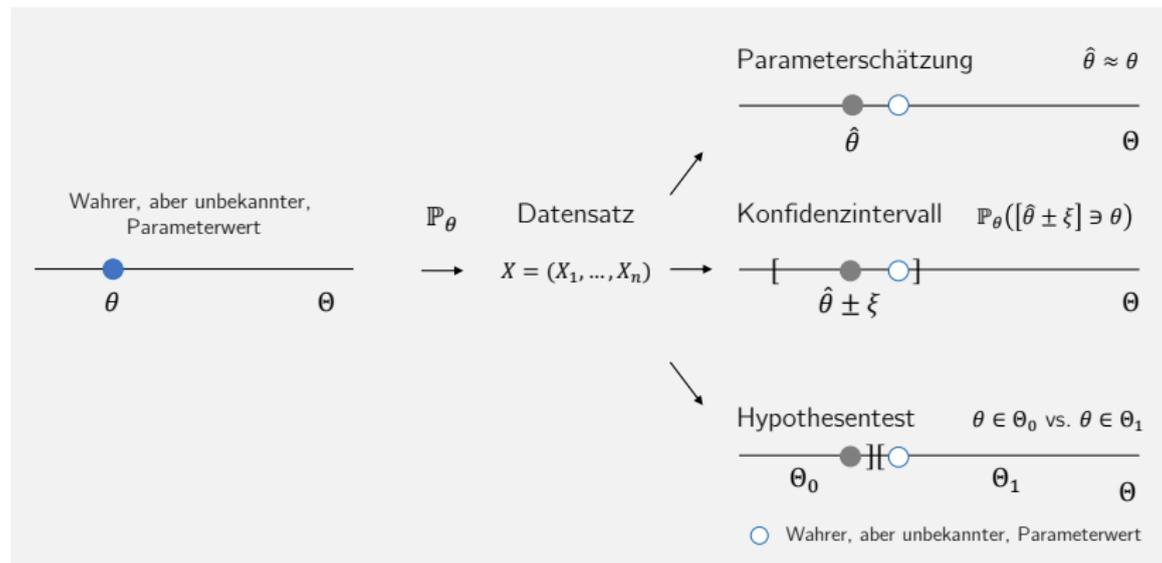
(2) Konfidenzintervalle

Das Ziel der Bestimmung von Konfidenzintervallen ist es, basierend auf der Verteilung möglicher Parameterschätzwerte eine quantitative Aussage über die mit dem Schätzwert assoziierte Unsicherheit zu treffen.

(3) Hypothesentests

Das Ziel der Auswertung von Hypothesentests ist es, basierend auf der angenommenen Verteilung der Beobachtungen X_1, \dots, X_n in einer möglichst sinnvollen Form zu entscheiden, ob der wahre, aber unbekanntem Parameterwert, in einer von zwei sich gegenseitig ausschließenden Untermengen des Parameterraumes, welche man als Hypothesen bezeichnet, liegt.

Standardprobleme Frequentistischer Inferenz



Standardprobleme Frequentistischer Inferenz

Standardannahmen frequentistischer Inferenz

\mathcal{M} sei ein statistisches Modell mit $X_1, \dots, X_n \sim p_\theta$. Es wird angenommen, dass ein vorliegender Datensatz eine der möglichen Realisierungen von $X_1, \dots, X_n \sim p_\theta$ ist.

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

$$\text{Datensatz (1)} : x^{(1)} = \left(x_1^{(1)}, x_2^{(1)}, \dots, x_n^{(1)} \right) \text{ mit } \bar{x}_n^{(1)} = \frac{1}{n} \sum_{i=1}^n x_i^{(1)}$$

$$\text{Datensatz (2)} : x^{(2)} = \left(x_1^{(2)}, x_2^{(2)}, \dots, x_n^{(2)} \right) \text{ mit } \bar{x}_n^{(2)} = \frac{1}{n} \sum_{i=1}^n x_i^{(2)}$$

$$\text{Datensatz (3)} : x^{(3)} = \left(x_1^{(3)}, x_2^{(3)}, \dots, x_n^{(3)} \right) \text{ mit } \bar{x}_n^{(3)} = \frac{1}{n} \sum_{i=1}^n x_i^{(3)}$$

$$\text{Datensatz (4)} : x^{(4)} = \left(x_1^{(4)}, x_2^{(4)}, \dots, x_n^{(4)} \right) \text{ mit } \bar{x}_n^{(4)} = \frac{1}{n} \sum_{i=1}^n x_i^{(4)}$$

$$\text{Datensatz (5)} : x^{(5)} = \dots$$

Um die Qualität statistischer Methoden zu beurteilen betrachtet die frequentistische Statistik deshalb die Wahrscheinlichkeitsverteilungen von Schätzern und Statistiken unter Annahme von $X_1, \dots, X_n \sim p_\theta$. Was zum Beispiel ist die Verteilung von $\bar{x}_n^{(1)}, \bar{x}_n^{(2)}, \bar{x}_n^{(3)}, \bar{x}_n^{(4)}, \dots$ also die Verteilung der Zufallsvariable $\bar{X}_n := \frac{1}{n} \sum_{i=1}^n X_i$?

Wenn eine statistische Methode im Sinne der frequentistischen 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.

Evidenzbasierte Evaluation von Psychotherapieformen bei Depression

Online Psychotherapie



Klassische Psychotherapie



Becks Depression Inventar (BDI) zur Depressionsdiagnostik

Evidenzbasierte Evaluation von Psychotherapieformen bei Depression

Experimentelle Bedingung
(Gruppen von $n = 50$)

Psychotherapie

Klassisch

Pre-BDI



Post-BDI

Online

Pre-BDI



Post-BDI

Standardprobleme Frequentistischer Inferenz

Einlesen des Datensatzes mit `read.table()`

```
fname = file.path(getwd(), "9_Daten", "psychotherapie_datensatz.csv")  
D = read.table(fname, sep = ",")
```

Daten der ersten acht Proband:innen jeder Gruppe

	Bedingung	Pre.BDI	Post.BDI
1	Klassisch	17	9
2	Klassisch	20	14
3	Klassisch	16	13
4	Klassisch	18	12
5	Klassisch	21	12
6	Klassisch	17	14
7	Klassisch	17	12
8	Klassisch	17	9
51	Online	22	16
52	Online	19	15
53	Online	21	13
54	Online	18	15
55	Online	19	13
56	Online	17	16
57	Online	20	13
58	Online	19	16

Standardannahmen der Frequentistischen Inferenz

Wir legen das Normalverteilungsmodell zugrunde, d.h. wir nehmen an, dass die BDI Werte Realisierungen von u.i.v. Zufallsvariablen

$$X_{ijk} \sim N\left(\mu_{ij}, \sigma^2\right), i = 1, 2, j = 1, 2, k = 1, \dots, n_i \quad (20)$$

wobei $i \in \{1, 2\}$ die experimentelle Bedingung (1 = Klassisch, 2 = Online), $j \in \{1, 2\}$ den Zeitpunkt der Messung (1 = Pre, 1 = Post) und $k \in \mathbb{N}_i$ den Proband:innen Index in der i ten experimentellen Bedingung bezeichnen sollend Dies entspricht der Annahme, dass sich der BDI Wert einer Proband:in durch Addition einer normalverteilten Fehlervariable mit Erwartungswertparameter 0 und Varianzparameter σ^2 zu den innerhalb einer Versuchsbedingung und einer Messung identischen Wert μ_{ij} ergibt.

Standardprobleme der Frequentistischen Inferenz

- (1) Was sind sinnvolle Tipps für die wahren, aber unbekanntenen, Parameterwerte μ_{ij} 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

$$(\mu_{12} - \mu_{11}) - (\mu_{21} - \mu_{22}) \neq 0? \quad (21)$$

Statistische Modelle

Statistiken und Schätzer

Standardprobleme Frequentistischer Inferenz

Selbstkontrollfragen

Selbstkontrollfragen

1. Definieren und erläutern Sie den Begriff des statistischen Modells.
2. Definieren und erläutern Sie den Begriff eines parametrischen statistischen Produktmodells.
3. Erläutern Sie den Unterschied zwischen univariaten und multivariaten statistischen Modellen.
4. Formulieren Sie das Bernoulli-Modell.
5. Formulieren Sie das Normalverteilungsmodell.
6. Definieren und erläutern Sie den Begriff der Statistik.
7. Definieren und erläutern Sie den Begriff des Schätzers.
8. Nennen und erläutern Sie die Standardprobleme der frequentistischen Inferenz.
9. Erläutern Sie die Standardannahmen der frequentistischen Inferenz.



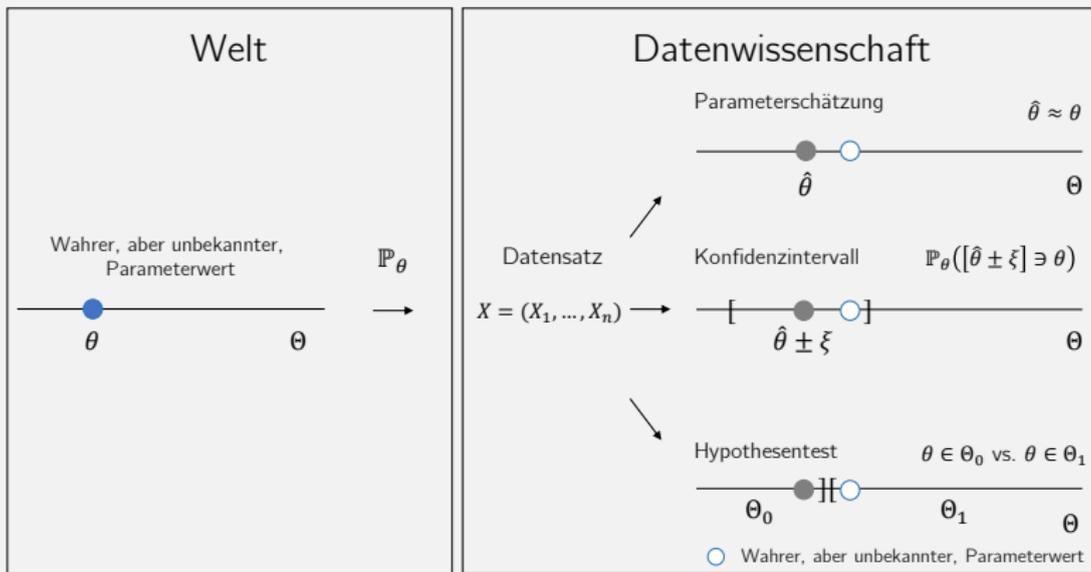
Wahrscheinlichkeitstheorie und Frequentistische Inferenz

BSc Psychologie WiSe 2021/22

Prof. Dr. Dirk Ostwald

(10) Parameterschätzung

Modell und Standardprobleme der Frequentistischen Inferenz



(1) Parameterschätzung

Ziel der Parameterschätzung ist es, einen möglichst guten Tipp für den wahren, aber unbekanntem, Parameterwert (oder eine Funktion dessen) abzugeben, typischerweise basierend auf der Beobachtung einer Realisierung von $X_1, \dots, X_n \sim p_\theta$.

(2) Konfidenzintervalle

Das Ziel der Bestimmung von Konfidenzintervallen ist es, basierend auf der Verteilung möglicher Parameterschätzwerte eine quantitative Aussage über die mit dem Schätzwert assoziierte Unsicherheit zu treffen.

(3) Hypothesentests

Das Ziel der Auswertung von Hypothesentests ist es, basierend auf der angenommenen Verteilung der Beobachtungen X_1, \dots, X_n in einer möglichst sinnvollen Form zu entscheiden, ob der wahre, aber unbekanntem Parameterwert, in einer von zwei sich gegenseitig ausschließenden Untermengen des Parameterraumes, welche man als Hypothesen bezeichnet, liegt.

Standardannahmen Frequentistischer Inferenz

\mathcal{M} sei ein statistisches Modell mit $X_1, \dots, X_n \sim \mathbb{P}_\theta$. Es wird angenommen, dass ein vorliegender Datensatz eine der möglichen Realisierungen von $X_1, \dots, X_n \sim \mathbb{P}_\theta$ ist. Aus frequentistischer Sicht kann man die Erhebung von Datensätzen unendlich oft wiederholen und zu jedem Datensatz Statistiken auswerten.

$$\text{Datensatz (1)} : x^{(1)} = \left(x_1^{(1)}, x_2^{(1)}, \dots, x_n^{(1)} \right), \text{ Statistik (1): } S : \mathbb{R}^n \rightarrow \Sigma, x^{(1)} \mapsto S \left(x^{(1)} \right)$$

$$\text{Datensatz (2)} : x^{(2)} = \left(x_1^{(2)}, x_2^{(2)}, \dots, x_n^{(2)} \right), \text{ Statistik (2): } S : \mathbb{R}^n \rightarrow \Sigma, x^{(2)} \mapsto S \left(x^{(2)} \right)$$

$$\text{Datensatz (3)} : x^{(3)} = \left(x_1^{(3)}, x_2^{(3)}, \dots, x_n^{(3)} \right), \text{ Statistik (3): } S : \mathbb{R}^n \rightarrow \Sigma, x^{(3)} \mapsto S \left(x^{(3)} \right)$$

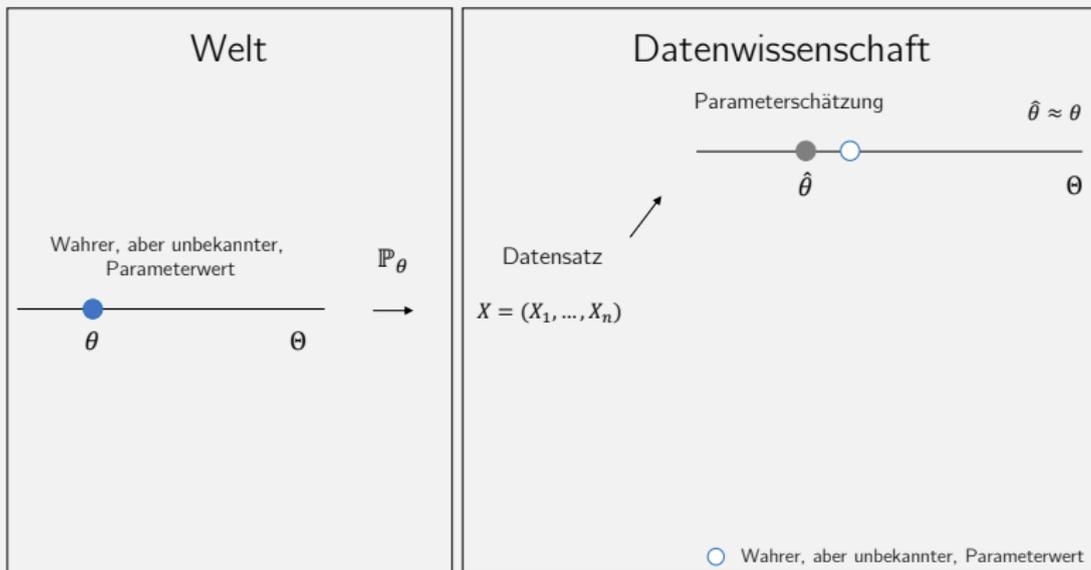
$$\text{Datensatz (4)} : x^{(4)} = \left(x_1^{(4)}, x_2^{(4)}, \dots, x_n^{(4)} \right), \text{ Statistik (4): } S : \mathbb{R}^n \rightarrow \Sigma, x^{(4)} \mapsto S \left(x^{(4)} \right)$$

...

Um die Qualität statistischer Methoden zu beurteilen betrachtet die frequentistische Statistik deshalb die Wahrscheinlichkeitsverteilungen von Statistiken und Schätzern unter der Annahme von $X_1, \dots, X_n \sim p_\theta$.

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 realen Normalfall nur eines vorliegenden Datensatzes, kann sie auch "schlecht" sein.

Modell und Standardprobleme der Frequentistischen Inferenz



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)

$\mathcal{M} := (\mathcal{X}, \mathcal{A}, \{\mathbb{P}_\theta | \theta \in \Theta\})$ sei ein statistisches Modell, (Θ, \mathcal{S}) sei ein Messraum, und $\hat{\theta} : \mathcal{X} \rightarrow \Theta$ sei ein 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 $\hat{\theta}(x)$ unterschieden.

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)

Per se garantiert keine der obengenannten Methoden, dass die mit ihrer Hilfe generierten Parameterschätzer in einem wohldefinierten Sinn gute Schätzer sind.

Die Eigenschaften von durch die Maximum-Likelihood Methode generierten Schätzern sind generell wünschenswert. Wir betrachten also in der Folge nur die Maximum-Likelihood Methode genauer. Mithilfe der Maximum-Likelihood Methode generierte Parameterpunktschätzer nennen wir *Maximum-Likelihood (ML) Schätzer*.

Grundbegriffe

Maximum-Likelihood Schätzer

Schätzereigenschaften bei endlichen Stichproben

Asymptotische Schätzereigenschaften

Eigenschaften von Maximum-Likelihood Schätzern

Selbstkontrollfragen

Definition (Likelihood-Funktion und Log-Likelihood-Funktion)

\mathcal{M} sei ein parametrisches statistisches Produktmodell mit WMF oder WDF p_θ , so dass $X_1, \dots, X_n \sim p_\theta$. Dann ist die *Likelihood Funktion* definiert als

$$L_n : \Theta \rightarrow [0, \infty], \theta \mapsto L_n(\theta) := \prod_{i=1}^n p_\theta(x_i). \quad (1)$$

und die *Log-Likelihood-Funktion* ist definiert als

$$\ell_n : \Theta \rightarrow \mathbb{R}, \theta \mapsto \ell_n(\theta) := \ln L_n(\theta). \quad (2)$$

Bemerkungen

- L_n ist eine Funktion des Parameters eines statistischen Modells.
- Werte von L_n sind die gemeinsamen W-Massen bzw. W-Dichten von Daten x_1, \dots, x_n .
- Generell gibt es keinen Grund anzunehmen, dass L_n über Θ zu 1 integriert.
- Die Likelihood-Funktion ist also keine WMF oder WDF.
- Die Log-Likelihood-Funktion ist die logarithmierte Likelihood-Funktion.

Definition (Maximum-Likelihood Schätzer)

\mathcal{M} sei ein parametrisches statistisches Produktmodell mit Parameter $\theta \in \Theta$. Ein *Maximum-Likelihood (ML) Schätzer* von θ ist definiert als

$$\hat{\theta}_n^{\text{ML}} : \mathcal{X} \rightarrow \Theta, x \mapsto \hat{\theta}_n^{\text{ML}}(x) := \arg \max_{\theta \in \Theta} L_n(\theta) = \arg \max_{\theta \in \Theta} \ell_n(\theta). \quad (3)$$

Bemerkungen

- $L_n(\theta) := \prod_{i=1}^n p_\theta(x_i)$ hängt von x_1, \dots, x_n ab, also hängt auch $\hat{\theta}_n^{\text{ML}}(x)$ von x_1, \dots, x_n ab.
- Weil \ln monoton steigend ist, entspricht eine Maximumstelle von ℓ_n einer Maximumstelle von L_n .
- Das Arbeiten mit der Log-Likelihood-Funktion ist oft einfacher als mit der Likelihood Funktion.
- Multiplikation von L_n mit einer positive Konstante, die nicht von θ abhängt, verändert einen ML Schätzer nicht, konstante additive Terme in der Log-Likelihood können also vernachlässigt werden.
- Maximum-Likelihood Schätzung ist ein Optimierungsproblem

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 analytische Optimierung in klassischen Beispielen und
- Methoden der numerische Optimierung im Anwendungskontext.

Beispiel (Bernoullimodell)

\mathcal{M} sei das Bernoullimodell, also $X_1, \dots, X_n \sim \text{Bern}(\mu)$.

(1) Formulierung der Log-Likelihood-Funktion

Es gilt

$$L_n :]0, 1[\rightarrow]0, 1[, \mu \mapsto L_n(\mu) := \prod_{i=1}^n \mu^{x_i} (1 - \mu)^{1-x_i} = \mu^{\sum_{i=1}^n x_i} (1 - \mu)^{n - \sum_{i=1}^n x_i}. \quad (4)$$

Logarithmieren ergibt

$$\ell_n :]0, 1[\rightarrow \mathbb{R}, \mu \mapsto \ell_n(\mu) = \ln \mu \sum_{i=1}^n x_i + \ln(1 - \mu) \left(n - \sum_{i=1}^n x_i \right). \quad (5)$$

Beispiel (Bernoullimodell)

(2) Auswertung der Ableitung der Log-Likelihood-Funktion und Nullsetzen

Es gilt

$$\begin{aligned}\frac{d}{d\mu} \ell_n(\mu) &= \frac{d}{d\mu} \left(\ln \mu \sum_{i=1}^n x_i + \ln(1 - \mu) \binom{n}{\sum_{i=1}^n x_i} \right) \\ &= \frac{d}{d\mu} \ln \mu \sum_{i=1}^n x_i + \frac{d}{d\mu} \ln(1 - \mu) \binom{n}{\sum_{i=1}^n x_i} \\ &= \frac{1}{\mu} \sum_{i=1}^n x_i - \frac{1}{1 - \mu} \binom{n}{\sum_{i=1}^n x_i}.\end{aligned}\tag{6}$$

Die sogenannte *Maximum-Likelihood Gleichung* ergibt sich in diesem Beispiel also zu

$$\frac{1}{\hat{\mu}_n^{\text{ML}}} \sum_{i=1}^n x_i - \frac{1}{1 - \hat{\mu}_n^{\text{ML}}} \binom{n}{\sum_{i=1}^n x_i} = 0.\tag{7}$$

Maximum-Likelihood Schätzer

Beispiel (Bernoullimodell)

(3) Auflösen nach potentiellen Maximumstellen

Es gilt

$$\begin{aligned} & \frac{1}{\hat{\mu}_n^{\text{ML}}} \sum_{i=1}^n x_i - \frac{1}{1 - \hat{\mu}_n^{\text{ML}}} \left(n - \sum_{i=1}^n x_i \right) = 0 \\ \Leftrightarrow & \hat{\mu}_n^{\text{ML}} (1 - \hat{\mu}_n^{\text{ML}}) \left(\frac{1}{\hat{\mu}_n^{\text{ML}}} \sum_{i=1}^n x_i - \frac{1}{1 - \hat{\mu}_n^{\text{ML}}} \left(n - \sum_{i=1}^n x_i \right) \right) = 0 \\ \Leftrightarrow & \sum_{i=1}^n x_i - \hat{\mu}_n^{\text{ML}} \sum_{i=1}^n x_i - n \hat{\mu}_n^{\text{ML}} + \hat{\mu}_n^{\text{ML}} \sum_{i=1}^n x_i = 0 \quad (8) \\ \Leftrightarrow & n \hat{\mu}_n^{\text{ML}} = \sum_{i=1}^n x_i \\ \Leftrightarrow & \hat{\mu}_n^{\text{ML}} = \frac{1}{n} \sum_{i=1}^n x_i. \end{aligned}$$

Beispiel (Bernoullimodell)

$\hat{\mu}_n^{\text{ML}} = \frac{1}{n} \sum_{i=1}^n x_i$ ist also ein potentieller Maximum-Likelihood Schätzer von μ . Dies kann durch Betrachten der zweiten Ableitung von ℓ_n verifiziert werden, worauf hier verzichtet werden soll.

$$\hat{\mu}_n^{\text{ML}} : \{0, 1\}^n \rightarrow [0, 1], x \mapsto \hat{\mu}_n^{\text{ML}}(x) := \frac{1}{n} \sum_{i=1}^n x_i \quad (9)$$

ist also ein Maximum-Likelihood Schätzer von μ im Bernoullimodell.

Beispiel (Normalverteilungsmodell)

\mathcal{M} sei das Normalverteilungsmodell, also $X_1, \dots, X_n \sim N(\mu, \sigma^2)$

(1) Formulierung der Log-Likelihood-Funktion

Es gilt

$$\begin{aligned} L_n : \mathbb{R} \times \mathbb{R}_{>0} &\rightarrow \mathbb{R}_{>0}, (\mu, \sigma^2) \mapsto L_n(\mu, \sigma^2) := \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x_i - \mu)^2\right) \\ &= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right). \end{aligned} \quad (10)$$

Logarithmieren ergibt

$$\ell_n : \mathbb{R} \times \mathbb{R}_{>0} \rightarrow \mathbb{R}, (\mu, \sigma^2) \mapsto \ell_n(\mu, \sigma^2) = -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2. \quad (11)$$

Beispiel (Normalverteilungsmodell)

(2) Auswertung der Ableitung der Log-Likelihood-Funktion und Nullsetzen

Es ergibt sich

$$\frac{d}{d\mu} \ell_n(\mu) = -\frac{d}{d\mu} \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 = -\frac{1}{2\sigma^2} \sum_{i=1}^n \frac{d}{d\mu} (x_i - \mu)^2 = \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu). \quad (12)$$

Weiterhin ergibt sich

$$\frac{d}{d\sigma^2} \ell_n(\sigma^2) = -\frac{n}{2} \frac{d}{d\sigma^2} \ln \sigma^2 - \frac{d}{d\sigma^2} \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - \mu)^2. \quad (13)$$

Die Maximum-Likelihood Gleichungen haben also die Form

$$\begin{aligned} \sum_{i=1}^n (x_i - \hat{\mu}_n^{\text{ML}}) &= 0 \\ -\frac{n}{2\hat{\sigma}_n^{\text{ML}2}} + \frac{1}{2\hat{\sigma}_n^{\text{ML}4}} \sum_{i=1}^n (x_i - \mu)^2 &= 0 \end{aligned} \quad (14)$$

Beispiel (Normalverteilungsmodell)

(3) Auflösen nach potentiellen Maximumstellen

Es ergibt sich

$$\sum_{i=1}^n (x_i - \hat{\mu}_n^{\text{ML}}) = 0 \Leftrightarrow \sum_{i=1}^n x_i = n \hat{\mu}_n^{\text{ML}} \Leftrightarrow \hat{\mu}_n^{\text{ML}} = \frac{1}{n} \sum_{i=1}^n x_i. \quad (15)$$

Also ist $\hat{\mu}_n^{\text{ML}} = n^{-1} \sum_{i=1}^n x_i$ ein potentieller Maximum-Likelihood Schätzer von μ . Einsetzen ergibt dann weiterhin

$$\begin{aligned} -\frac{n}{2\hat{\sigma}_n^{2\text{ML}}} + \frac{1}{2\hat{\sigma}_n^{4\text{ML}}} \sum_{i=1}^n (x_i - \hat{\mu}_n^{\text{ML}})^2 &= 0 \\ \Leftrightarrow -n\hat{\sigma}_n^{2\text{ML}} + \sum_{i=1}^n (x_i - \hat{\mu}_n^{\text{ML}})^2 &= 0 \end{aligned} \quad (16)$$

$$\Leftrightarrow \hat{\sigma}_n^{2\text{ML}} = \frac{1}{n} \sum_{i=1}^n (x_i - \hat{\mu}_n^{\text{ML}})^2.$$

$\hat{\sigma}_n^{2\text{ML}} = n^{-1} \sum_{i=1}^n (x_i - \hat{\mu}_n^{\text{ML}})^2$ ist also potentieller Maximum-Likelihood Schätzer von σ^2 .

Beispiel (Normalverteilungsmodell)

Beide potentiellen Maximum-Likelihood Schätzer können durch Betrachten der zweiten Ableitung von ℓ_n verifiziert werden, worauf hier verzichtet werden soll.

Also sind

$$\hat{\mu}_n^{\text{ML}} : \mathbb{R}^n \rightarrow \mathbb{R}, x \mapsto \hat{\mu}_n^{\text{ML}}(x) := \frac{1}{n} \sum_{i=1}^n x_i \quad (17)$$

und

$$\hat{\sigma}_n^{2\text{ML}} : \mathbb{R}^n \rightarrow \mathbb{R}_{\geq 0}, x \mapsto \hat{\sigma}_n^{2\text{ML}}(x) := \frac{1}{n} \sum_{i=1}^n (x_i - \hat{\mu}_n^{\text{ML}})^2. \quad (18)$$

die Maximum-Likelihood Schätzer von μ und σ^2 im Normalverteilungsmodell. $\hat{\mu}_n^{\text{ML}}$ ist identisch mit dem Stichprobenmittel \bar{X}_n , $\hat{\sigma}_n^{2\text{ML}}$ ist dagegen nicht identisch mit der Stichprobenvarianz S_n^2 .

Anwendungsbeispiel

Experimentelle Bedingung
(Gruppen von $n = 50$)

Psychotherapie

Klassisch

Pre-BDI



Post-BDI

Online

Pre-BDI



Post-BDI

Anwendungsbeispiel

Standardannahmen der Frequentistischen Inferenz

Wir legen das Normalverteilungsmodell zugrunde, d.h. wir nehmen an, dass die BDI Werte Realisierungen von u.i.v. Zufallsvariablen

$$X_{ijk} \sim N\left(\mu_{ij}, \sigma^2\right), i = 1, 2, j = 1, 2, k = 1, \dots, n_i \quad (19)$$

wobei $i \in \{1, 2\}$ die experimentelle Bedingung (1 = Klassisch, 2 = Online), $j \in \{1, 2\}$ den Zeitpunkt der Messung (1 = Pre, 1 = Post) und $k \in \mathbb{N}_i$ den Proband:innen Index in der i ten experimentellen Bedingung bezeichnen sollend Dies entspricht der Annahme, dass sich der BDI Wert einer Proband:in durch Addition einer normalverteilten Fehlervariable mit Erwartungswertparameter 0 und Varianzparameter σ^2 zu den innerhalb einer Versuchsbedingung und einer Messung identischen Wert μ_{ij} ergibt.

Standardprobleme der Frequentistischen Inferenz

- (1) Was sind sinnvolle Tipps für die wahren, aber unbekannt, Parameterwerte μ_{ij} 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

$$(\mu_{12} - \mu_{11}) - (\mu_{21} - \mu_{22}) \neq 0? \quad (20)$$

Anwendungsbeispiel

Wir beschränken uns hier zunächst auf die Pre.BDI Daten der Klassischen Bedingung, formal

$$X_k \sim N(\mu, \sigma) \text{ mit } X_k := X_{11k} \text{ für } k = 1, \dots, n \text{ und } \mu := \mu_{11}. \quad (21)$$

```
# Einlesen und Auswahl der Daten
fname = file.path(getwd(), "10_Daten", "psychotherapie_datensatz.csv")
D = read.table(fname, sep = ",")
X = D$Pre.BDI[D$Bedingung == "Klassisch"] # i = 1 (Klassisch), j = 1 (Pre) Daten

# Maximum-Likelihood Schätzung der Gruppen-spezifischen Erwartungswertparameter \mu_{11}
mu_hat_11 = mean(X) # mean(x) berechnet das Stichprobenmittel des Datensatzes x
print(mu_hat_11) # Ausgabe
```

> [1] 18.2

```
# Maximum-Likelihood Schätzung der Gruppen-spezifischen Varianzparameter \sigma^2_{ij}
n = length(X) # Anzahl der Datenpunkte
sigsqr_hat = ((n-1)/n)*var(X) # var(x) berechnet die Stichprobenvarianz des Datensatzes x
print(sigsqr_hat)
```

> [1] 3.43

Basierend auf dem Prinzip der Maximum-Likelihood Schätzung sind also

$$\hat{\mu}_{50}^{\text{ML}} = 18.19, \text{ und } \hat{\sigma}_{50}^{2\text{ML}} = 3.43 \quad (22)$$

sinnvolle Tipps für μ und σ^2 basierend auf den vorliegenden 50 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 $\mathcal{M} := \{\mathcal{X}, \mathcal{A}, \{\mathbb{P}_\theta | \theta \in \Theta\}\}$ mit n -dimensionalen Stichprobenraum (z.B. $\mathcal{X} := \mathbb{R}^n$), d -dimensionalen Parameterraum $\Theta \subset \mathbb{R}^d$ und gegebener WMF oder WDF p_θ für alle $\theta \in \Theta$ aus. $X_1, \dots, X_n \sim p_\theta$ bezeichnet die zu \mathcal{M} gehörende Stichprobe unabhängig und identisch verteilter Zufallsvariablen, es gilt also $X_0 \sim p_\theta$ und $X_i \sim p_\theta$ für alle $i = 1, \dots, n$.

Für einen Messraum (Σ, \mathcal{S}) sei $\hat{\tau} : \mathcal{X} \rightarrow \Sigma$ ein Schätzer von $\tau : \Theta \rightarrow \Sigma$. Wir betrachten Erwartungswerts-, und Standardabweichungsschätzer, also Schätzer für

$$\tau : \Theta \rightarrow \Sigma, \theta \mapsto \tau(\theta) \text{ mit } \tau(\theta) := \mathbb{E}_\theta(X_0), \tau(\theta) := \mathbb{V}_\theta(X_0), \text{ und } \tau(\theta) := \mathbb{S}_\theta(X_0) \quad (23)$$

respektive, sowie Parameterschätzer, also Schätzer für

$$\tau : \Theta \rightarrow \Sigma, \tau(\theta) := \theta. \quad (24)$$

In der Folge führen wir *Frequentistische Schätzereigenschaften* ein. Frequentistische Schätzereigenschaften betrachten die Verteilung der Schätzwerte $\hat{\tau}(x_1, \dots, x_n)$ in Abhängigkeit von der Verteilung der Stichprobenwerte x_1, \dots, x_n . Weil die Stichprobenwerte zufällig sind, sind auch die Schätzwerte zufällig; ein Schätzer $\hat{\tau}$ ist also 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 $n \rightarrow \infty$.

Ein Schätzer $\hat{\tau}_n$ heißt **erwartungstreu**, wenn sein Erwartungswert dem wahren, aber unbekanntem, Wert $\tau(\theta)$ für alle $\theta \in \Theta$ gleicht.

Die **Varianz** eines Schätzers $\hat{\tau}_n$ ist die Varianz der Zufallsvariable $\hat{\tau}_n(X)$; der **Standardfehler** eines Schätzers $\hat{\tau}_n$ ist die Standardabweichung der Zufallsvariable $\hat{\tau}_n(X)$.

Die **Cramér-Rao-Ungleichung** gibt eine untere Schranke für die Varianz erwartungstreuer Schätzer an. Ein erwartungstreuer Schätzer mit Varianz gleich der in der Cramér-Rao-Ungleichung gegebenen unteren Schranke hat die kleinstmögliche Varianz aller erwartungstreuen Schätzer und ist in diesem Sinne ein "optimaler" Schätzer.

Der **mittlere quadratische Fehler** von $\hat{\tau}_n$ ist der Erwartungswert der quadrierten Abweichung von $\hat{\tau}_n(X)$ von $\tau(\theta)$ über Stichproben vom Umfang n .

Definition (Fehler, Systematischer Fehler, und Erwartungstreue)

$X_1, \dots, X_n \sim p_\theta$ sei eine Stichprobe und $\hat{\tau}_n$ sei ein Schätzer für τ .

- Der *Fehler* von $\hat{\tau}_n$ ist definiert als

$$\hat{\tau}_n(X) - \tau(\theta). \quad (25)$$

- Der *systematische Fehler (Bias)* von $\hat{\tau}_n$ ist definiert als

$$B(\hat{\tau}_n) := \mathbb{E}_\theta(\hat{\tau}_n(X)) - \tau(\theta). \quad (26)$$

- $\hat{\tau}_n$ heißt *erwartungstreu (unbiased)*, wenn

$$B(\hat{\tau}_n) = 0 \Leftrightarrow \mathbb{E}_\theta(\hat{\tau}_n(X)) = \tau(\theta) \text{ für alle } \theta \in \Theta, n \in \mathbb{N}. \quad (27)$$

Andernfalls heißt $\hat{\tau}_n$ *verzerrt (biased)*.

Bemerkungen

- Der Fehler hängt von einer Realisation der Stichprobe ab.
- Der systematische Fehler ist der erwartete Fehler über viele Stichprobenrealisationen.
- Ein Parameterschätzer ist erwartungstreu, wenn $\mathbb{E}_\theta(\hat{\theta}_n(X)) = \theta$.

Theorem (Erwartungstreue von Stichprobenmittel und Stichprobenvarianz)

$X_1, \dots, X_n \sim p_\theta$ sei die Stichprobe eines statistischen parametrischen Produktmodells \mathcal{M} .

- Das Stichprobenmittel

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

ist ein erwartungstreuer Schätzer des Erwartungswerts $\mathbb{E}_\theta(X_0)$.

- Die Stichprobenvarianz

$$S_n^2 := \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \quad (29)$$

ist ein erwartungstreuer Schätzer der Varianz $\mathbb{V}_\theta(X_0)$.

Beweis

Um die Notation zu vereinfachen, definieren wir $\mathbb{E} := \mathbb{E}_\theta$ und $\mathbb{V} := \mathbb{V}_\theta$. Mit der Linearität von Erwartungswerten ergibt sich dann

$$\mathbb{E}(\bar{X}_n) = \mathbb{E}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} \sum_{i=1}^n \mathbb{E}(X_i) = \frac{1}{n} \sum_{i=1}^n \mathbb{E}(X_0) = \frac{1}{n} n \mathbb{E}(X_0) = \mathbb{E}(X_0).$$

Dies zeigt die Erwartungstreue des Stichprobenmittels als Schätzer des Erwartungswertes.

Um die Erwartungstreue der Stichprobenvarianz zu zeigen, halten wir zunächst fest, dass

$$\mathbb{V}(\bar{X}_n) = \mathbb{V}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} \sum_{i=1}^n \mathbb{V}(X_i) = \frac{1}{n^2} \sum_{i=1}^n \mathbb{V}(X_0) = \frac{1}{n^2} n \mathbb{V}(X_0) = \frac{\mathbb{V}(X_0)}{n}.$$

gilt. Weiterhin halten wir ohne Beweis fest, dass

$$\sum_{i=1}^n (X_i - \bar{X}_n)^2 = \sum_{i=1}^n (X_i - \mu - \bar{X}_n + \mu)^2 = \sum_{i=1}^n (X_i - \mu)^2 - n(\bar{X}_n - \mu)^2.$$

Es ergibt sich somit

$$\begin{aligned}\mathbb{E} \left((n-1)S_n^2 \right) &= \mathbb{E} \left(\sum_{i=1}^n (X_i - \bar{X}_n)^2 \right) \\ &= \mathbb{E} \left(\sum_{i=1}^n (X_i - \mu)^2 - n(\bar{X}_n - \mu)^2 \right) \\ &= \sum_{i=1}^n \mathbb{E} \left((X_i - \mu)^2 \right) - n\mathbb{E} \left((\bar{X}_n - \mu)^2 \right) \\ &= n\mathbb{V}(X_0) - n\mathbb{V}(\bar{X}_n) \\ &= n\mathbb{V}(X_0) - n \frac{\mathbb{V}(X_0)}{n} \\ &= n\mathbb{V}(X_0) - \mathbb{V}(X_0) \\ &= (n-1)\mathbb{V}(X_0)\end{aligned}$$

Schließlich ergibt sich

$$\mathbb{E}(S_n^2) = \mathbb{E} \left(\frac{1}{n-1} (n-1)S_n^2 \right) = \frac{1}{n-1} \mathbb{E} \left((n-1)S_n^2 \right) = \frac{1}{n-1} (n-1)\mathbb{V}(X_0) = \mathbb{V}(X_0)$$

und damit die Erwartungstreue der Stichprobenvarianz als Schätzer der Varianz.

Theorem (Verzerrtheit der Stichprobenstandardabweichung)

$X_1, \dots, X_n \sim p_\theta$ sei die Stichprobe eines statistischen parametrischen Produktmodells \mathcal{M} . Dann ist die Stichprobenstandardabweichung

$$S_n := \sqrt{S_n^2} \quad (30)$$

ein verzerrter Schätzer der Standardabweichung $\mathbb{S}_\theta(X_0)$.

Beweis

Wir halten zunächst fest, dass $\sqrt{\cdot}$ eine strikt konkave Funktion und $\sigma^2 > 0$ ist. Dann aber gilt mit der Jensenschen Ungleichung $\mathbb{E}(f(X)) < f(\mathbb{E}(X))$ für strikt konkave Funktionen, dass

$$\mathbb{E}(S) = \mathbb{E}\left(\sqrt{S_n^2}\right) < \sqrt{\mathbb{E}(S_n^2)} = \sqrt{\mathbb{V}_\theta(X_0)} = \mathbb{S}_\theta(X_0). \quad (31)$$

□

Bemerkung

- Nichtlineare Transformationen von erwartungstreuen Schätzern liefern oft verzerrte Schätzer.

Simulation ($X_1, \dots, X_{12} \sim N(\mu, \sigma^2)$) mit $n = 12$, $\mu = 1.7$, $\sigma^2 = 2$, $\sigma \approx 1.41$)

```
# Modellformulierung
mu      = 1.7          # wahrer, aber unbekannter, Erwartungswertparameter
sigsqr  = 2            # wahrer, aber unbekannter, Varianzparameter
n       = 12          # Stichprobengroesse n
ns      = 1e4         # Anzahl der Simulationen
X_bar   = rep(NaN,ns) # Stichprobenmittelarray
S_sqr   = rep(NaN,ns) # Stichprobenvarianzarray
S       = rep(NaN,ns) # Stichprobenstandardabweichungarray

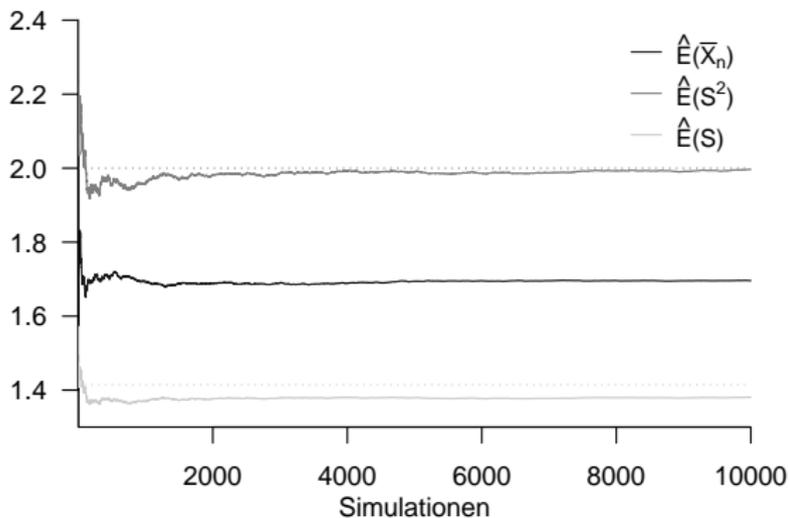
# Simulationsiterationen
for(s in 1:ns){

  # Stichprobenrealisation von X_1, ..., X_{12}
  x      = rnorm(n,mu,sqrt(sigsqr))

  # Erwartungswert-, Varianz-, Standardabweichungsschaetzer
  X_bar[s] = mean(x)      # Stichprobenmittel
  S_sqr[s] = var(x)      # Stichprobenvarianz
  S[s]     = sd(x)       # Stichprobenstandardabweichung
}

# Erwartungswertschaetzung
E_hat_X_bar = cumsum(X_bar)/(1:ns) # \mathbb{E}(\bar{X}_n) Schaetzungen
E_hat_S_sqr = cumsum(S_sqr)/(1:ns) # \mathbb{E}(S^2) Schaetzungen
E_hat_S     = cumsum(S) / (1:ns)  # \mathbb{E}(S) Schaetzungen
```

Simulation ($X_1, \dots, X_{12} \sim N(\mu, \sigma^2)$ mit $n = 12$, $\mu = 1.7$, $\sigma^2 = 2$, $\sigma \approx 1.41$)



Definition (Varianz und Standardfehler)

$X_1, \dots, X_n \sim p_\theta$ sei eine Stichprobe und $\hat{\tau}_n$ sei ein Schätzer von τ .

- Die *Varianz* von $\hat{\tau}_n$ ist definiert als

$$\mathbb{V}_\theta(\hat{\tau}_n) := \mathbb{E}_\theta \left((\hat{\tau}_n(X) - \mathbb{E}_\theta(\hat{\tau}_n(X)))^2 \right). \quad (32)$$

- Der *Standardfehler* von $\hat{\tau}_n$ ist definiert als

$$\text{SE}(\hat{\tau}_n) := \sqrt{\mathbb{V}_\theta(\hat{\tau}_n)} \quad (33)$$

Bemerkungen

- Die Varianz eines Schätzers $\hat{\tau}_n$ ist die Varianz der Zufallsvariable $\hat{\tau}_n(X)$.
- Der Standardfehler eines Schätzers $\hat{\tau}_n$ ist die Standardabweichung von $\hat{\tau}_n(X)$.

Theorem (Standardfehler des Stichprobenmittels)

$X_1, \dots, X_n \sim p_\theta$ sei die Stichprobe eines parametrischen statistischen Produktmodells \mathcal{N} . Dann ist der *Standardfehler des Stichprobenmittels* gegeben durch

$$\text{SE}(\bar{X}_n) = \frac{\mathbb{S}_\theta(X_0)}{\sqrt{n}}. \quad (34)$$

Der Standardfehler des Stichprobenmittels heißt auch *Standardfehler des Mittelwertes*.

Beweis

Per definitionem und mit $\mathbb{V}_\theta(\bar{X}_n) = \mathbb{V}_\theta(X_0)/n$, ergibt sich

$$\text{SE}(\bar{X}_n) = \sqrt{\mathbb{V}_\theta(\bar{X}_n)} = \sqrt{\frac{\mathbb{V}_\theta(X_0)}{n}} = \frac{\mathbb{S}_\theta(X_0)}{\sqrt{n}}. \quad (35)$$

Bemerkungen

- Der Standardfehler des Mittelwerts beschreibt die Variabilität des Stichprobenmittels.
- Da $\mathbb{S}_\theta(X_0)$ unbekannt ist, ist auch $\text{SE}(\bar{X}_n)$ unbekannt.
- Ein verzerrter Schätzer für den Standardfehler des Stichprobenmittels ist gegeben durch $\hat{\text{SE}}(\bar{X}_n) = \frac{S_n}{\sqrt{n}}$.

Beispiel (Standardfehler des Bernoulli Parameter ML Schätzers)

Es sei $X_1, \dots, X_n \sim \text{Bern}(\mu)$ und $\hat{\mu}_n^{\text{ML}}$ der ML Schätzer für μ . Dann ist

$$\text{SE}(\hat{\mu}_n^{\text{ML}}) = \sqrt{\frac{\mu(1-\mu)}{n}}. \quad (36)$$

Beweis

Es gilt

$$\begin{aligned} \text{SE}(\hat{\mu}_n^{\text{ML}}) &= \sqrt{\mathbb{V}_\mu(\hat{\mu}_n^{\text{ML}})} = \sqrt{\mathbb{V}_\mu\left(\frac{1}{n} \sum_{i=1}^n X_i\right)} \\ &= \sqrt{\frac{1}{n^2} \sum_{i=1}^n \mathbb{V}_\mu(X_i)} = \sqrt{\frac{n\mu(1-\mu)}{n^2}} = \sqrt{\frac{\mu(1-\mu)}{n}}, \end{aligned} \quad (37)$$

wobei die dritte Gleichung mit der Unabhängigkeit der X_i , $i = 1, \dots, n$ und die vierte Gleichung mit der Varianz $\mathbb{V}_\mu(X_0) = \mathbb{V}_\mu(X_i) = \mu(1-\mu)$, $i = 1, \dots, n$ der Bernoulli Stichprobenvariablen folgt. \square

Bemerkung

- Ein Schätzer für den Standardfehler $\text{SE}(\hat{\mu}_n^{\text{ML}})$ ist $\hat{\text{SE}}(\hat{\mu}_n^{\text{ML}}) = \sqrt{\frac{\hat{\mu}_n^{\text{ML}}(1-\hat{\mu}_n^{\text{ML}})}{n}}$

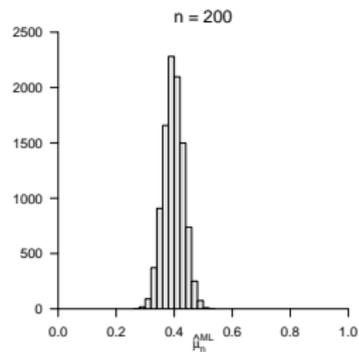
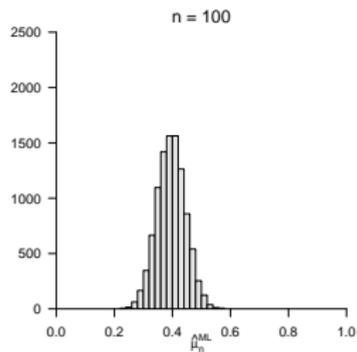
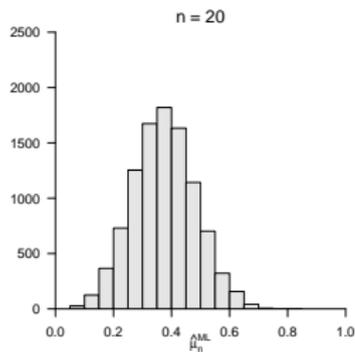
Simulation ($X_1, \dots, X_n \sim \text{Bern}(\mu)$ mit $\mu = 0.4$)

```
# 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(NA, length(n_all)*ns), # ML Schätzerarray
                    nrow = length(n_all))

# Stichprobengroesseniterationen
for(i in seq_along(n_all)){

  # Simulationsiterationen
  for(s in 1:ns){
    x           = rbinom(n_all[i],1,mu) # Stichprobenrealisation von X_1,...,X_n
    mu_hat_ML[i,s] = mean(x)           # Stichprobenmittel
  }
}
```

Simulation ($X_1, \dots, X_n \sim \text{Bern}(\mu)$ mit $\mu = 0.4$)



- Die Varianz bzw. der Standardfehler von $\hat{\mu}_n^{\text{ML}}$ hängen von n ab.

Vorbemerkungen zur Cramér-Rao-Ungleichung

Je kleiner die Varianz eines Schätzers, desto besser. Weil aber Stichproben streuen, kann die Varianz von erwartungstreuen Schätzern nicht beliebig klein sein.

Die **Cramér-Rao-Ungleichung** gibt eine untere Schranke für die Varianz erwartungstreuer Schätzer an. Ein erwartungstreuer Schätzer mit Varianz gleich der in der Cramér-Rao-Ungleichung gegebenen unteren Schranke hat die kleinstmögliche Varianz aller erwartungstreuer Schätzer und ist in diesem Sinne "optimal."

Die Cramér-Rao-Ungleichung basiert auf dem Begriff der **Fisher-Information**. Wir diskutieren deshalb zunächst die Begriffe der **Scorefunktion** und der darauf basierenden **Fisher-Information**.

Die vorgestellten Resultate gelten im Allgemeinen nur unter eine Reihe von Annahmen, den sogenannten **Fisher-Regularitätsbedingungen**.

Fisher-Regularitätsbedingungen

1. Θ ist ein offenes Intervall, d.h. θ liegt nicht an einer Parameterraumgrenze.
2. Der Träger von p_θ hängt nicht von θ ab.
3. WMFs oder WDF mit unterschiedlichem $\theta \in \Theta$ sind unterschiedlich.
4. Die Likelihood-Funktion ist zweimal stetig differenzierbar.
5. Integration und Differentiation dürfen vertauscht werden.

Definition (Scorefunktion und Fisher-Information)

$X_1, \dots, X_n \sim p_\theta$ sei die Stichprobe eines parametrischen statistischen Produktmodells \mathcal{M} mit eindimensionalem Parameter θ und ℓ_n sei die zugehörige Log-Likelihood-Funktion.

- Die erste Ableitung der Log-Likelihood-Funktion ℓ_n wird *Scorefunktion der Stichprobe* X_1, \dots, X_n genannt und mit

$$S_n(\theta) := \frac{d}{d\theta} \ell_n(\theta). \quad (38)$$

bezeichnet. Für $n = 1$ schreiben wir $S(\theta) := S_1(\theta)$ und nennen $S(\theta)$ *Scorefunktion einer Zufallsvariable*.

- Die negative zweite Ableitung der Log-Likelihood-Funktion ℓ_n wird *Fisher-Information der Stichprobe* X_1, \dots, X_n genannt und mit

$$I_n(\theta) := -\frac{d^2}{d\theta^2} \ell_n(\theta). \quad (39)$$

bezeichnet. Für $n = 1$ schreiben wir $I(\theta) := I_1(\theta)$ und nennen $I(\theta)$ die *Fisher-Information einer Zufallsvariable*.

Definition (Erwartete und beobachtete Fisher-Information)

$X_1, \dots, X_n \sim p_\theta$ sei die Stichprobe eines parametrischen statistischen Produktmodells \mathcal{M} mit eindimensionalem Parameter θ , ℓ_n sei die zugehörige Log-Likelihood-Funktion und $\hat{\theta}_n^{\text{ML}}$ sei ein ML-Schätzer von θ .

- Die *beobachtete Fisher-Information der Stichprobe* X_1, \dots, X_n ist definiert als

$$I_n(\hat{\theta}_n^{\text{ML}}) := -\frac{d^2}{d\theta^2} \ell_n(\hat{\theta}_n^{\text{ML}}), \quad (40)$$

d.h. die beobachtete Fisher-Information der Stichprobe X_1, \dots, X_n ist die Fisher-Information an der Stelle des ML-Schätzers $\hat{\theta}_n^{\text{ML}}$.

- Die *erwartete Fisher-Information der Stichprobe* X_1, \dots, X_n ist definiert als

$$J_n(\theta) := \mathbb{E}_\theta(I_n(\theta)). \quad (41)$$

Für $n = 1$ schreiben wir $J(\theta) := J_1(\theta)$ und nennen $J(\theta)$ die *erwartete Fisher-Information einer Zufallsvariable*.

Theorem (Additivität der Fisher-Information)

$X_1, \dots, X_n \sim p_\theta$ sei die Stichprobe eines parametrischen statistischen Produktmodells \mathcal{M} mit eindimensionalem Parameter θ , ℓ_n sei die zugehörige Log-Likelihood-Funktion, und $I_n(\theta)$ und $J_n(\theta)$ seien die Fisher-Information und die erwartete Fisher-Information der Stichprobe X_1, \dots, X_n , respektive. Dann gilt

$$I_n(\theta) = nI_1(\theta) \text{ und } J_n(\theta) = nJ_1(\theta). \quad (42)$$

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.
- Für einen Beweis verweisen wir auf den Appendix.

Theorem (Erwartungswert und Varianz der Scorefunktion)

Der Erwartungswert der Scorefunktion einer Zufallsvariable ist

$$\mathbb{E}_\theta(S(\theta)) = 0 \quad (43)$$

und die Varianz der Scorefunktion einer Zufallsvariable ist

$$\mathbb{V}_\theta(S(\theta)) = J(\theta). \quad (44)$$

Bemerkungen * Der Erwartungswert der Ableitung der Log-Likelihood-Funktion ist Null. * Die erwartete Fisher-Information ist gleich der Varianz der Scorefunktion. * Für einen Beweis verweisen wir auf den Appendix.

Beispiel (Scorefunktion und Fisher-Information für den Bernoulli Parameter)

Es sei $X_1, \dots, X_n \sim \text{Bern}(\mu)$ mit $\mu \in]0, 1[$. Dann gilt:

- Die Scorefunktion der Stichprobe ist

$$S_n :]0, 1[\rightarrow \mathbb{R}, \mu \mapsto S_n(\mu) := \frac{1}{\mu} \sum_{i=1}^n x_i - \frac{1}{1-\mu} \left(n - \sum_{i=1}^n x_i \right). \quad (45)$$

- Die Fisher-Information der Stichprobe ist

$$I_n :]0, 1[\rightarrow \mathbb{R}, \mu \mapsto I_n(\mu) := \frac{x}{\mu^2} + \frac{(1-x)^2}{1-\mu}. \quad (46)$$

- Die beobachtete Fisher-Information der Stichprobe ist

$$I_n :]0, 1[\rightarrow \mathbb{R}, \hat{\mu}_n^{\text{ML}} \mapsto I_n \left(\hat{\mu}_n^{\text{ML}} \right) := \frac{x}{\hat{\mu}_n^{\text{ML}2}} + \frac{(1-x)}{1-\hat{\mu}_n^{\text{ML}}}. \quad (47)$$

- Die erwartete Fisher-Information der Stichprobe ist

$$J_n :]0, 1[\rightarrow \mathbb{R}, \mu \mapsto J_n(\mu) := \frac{n}{\mu(1-\mu)}. \quad (48)$$

Für einen Beweis verweisen wir auf den Appendix.

Beispiel (Erwartungswert-Parameter der Normalverteilung)

Es sei $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ und σ^2 als bekannt vorausgesetzt. Dann gilt:

- Die Scorefunktion der Stichprobe ist

$$S_n : \mathbb{R} \rightarrow \mathbb{R}, \mu \mapsto S_n(\mu) := \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu). \quad (49)$$

- Die Fisher-Information der Stichprobe ist

$$I_n : \mathbb{R} \rightarrow \mathbb{R}, \mu \mapsto I_n(\mu) := \frac{n}{\sigma^2}. \quad (50)$$

- Die beobachtete Fisher-Information der Stichprobe ist

$$I_n(\hat{\mu}_n^{\text{ML}}) = \frac{n}{\sigma^2}. \quad (51)$$

- Die erwartete Fisher-Information der Stichprobe ist

$$J_n : \mathbb{R} \rightarrow \mathbb{R}, \mu \mapsto J_n(\mu) := \frac{n}{\sigma^2}. \quad (52)$$

Für einen Beweis verweisen wir auf den Appendix.

Beispiel (Varianz-Parameter der Normalverteilung)

Es sei $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ und μ als bekannt vorausgesetzt. Dann gilt:

- die Scorefunktion ist

$$S_n : \mathbb{R}_{>0} \rightarrow \mathbb{R}, \sigma^2 \mapsto S_n(\sigma^2) := -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - \mu)^2 \quad (53)$$

- die Fisher-Information der Stichprobe ist

$$I_n : \mathbb{R}_{>0} \rightarrow \mathbb{R}, \sigma^2 \mapsto I_n(\sigma^2) := \frac{1}{\sigma^6} \sum_{i=1}^n (x_i - \mu)^2 - \frac{n}{2\sigma^4} \quad (54)$$

- die beobachtete Fisher-Information der Stichprobe ist

$$I_n(\hat{\sigma}_n^{2\text{ML}}) = \frac{n}{2\hat{\sigma}_{\text{ML}}^4} \quad (55)$$

- die erwartete Fisher-Information der Stichprobe ist

$$J_n : \mathbb{R}_{>0} \rightarrow \mathbb{R}, \sigma^2 \mapsto J_n(\sigma^2) := \frac{n}{2\sigma^4}. \quad (56)$$

Für einen Beweis verweisen wir auf den Appendix.

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

$$\mathbb{V}_\theta(\hat{\tau}_n) \geq \frac{\left(\frac{d}{d\theta}\tau(\theta)\right)^2}{J(\theta)}. \quad (57)$$

Im Speziellen gilt für $\tau(\theta) := \theta$ und somit $\hat{\tau}_n = \hat{\theta}_n$ und $\left(\frac{d}{d\theta}\tau(\theta)\right)^2 = 1$, dass

$$\mathbb{V}_\theta(\hat{\theta}_n) \geq \frac{1}{J(\theta)}. \quad (58)$$

Die rechte Seite obiger Ungleichungen heißt *Cramér-Rao-Schranke*.

Bemerkungen

- Die Varianz eines erwartungstreuen Schätzers $\hat{\theta}$ von θ ist größer oder gleich der reziproken erwarteten Fisher-Information $J(\theta)$.
- Wenn $\mathbb{V}_\theta(\hat{\theta}_n) = \frac{1}{J(\theta)}$ ist, ist die Varianz des Schätzers minimal.

Beweis

Wir halten zunächst fest, dass für die Zufallsvariablen $S(\theta)$ und $\hat{\tau}_n$ mit der Korrelationsungleichung und $\mathbb{V}_\theta(S(\theta)) = J(\theta)$ gilt, dass

$$\begin{aligned}\frac{\mathbb{C}_\theta(S_n(\theta), \hat{\tau}_n)^2}{\mathbb{V}_\theta(S(\theta))\mathbb{V}_\theta(\hat{\tau}_n)} &\leq 1 \\ \Leftrightarrow \mathbb{V}_\theta(\hat{\tau}_n) &\geq \frac{\mathbb{C}_\theta(S(\theta), \hat{\tau}_n)^2}{J(\theta)}.\end{aligned}\tag{59}$$

Mit dem Translationstheorem für Kovarianzen, $\mathbb{E}_\theta(S(\theta)) = 0$ und der Erwartungstreue von $\hat{\tau}_n$ ergibt sich dann

$$\mathbb{C}_\theta(S(\theta), \hat{\tau}_n) = \frac{d}{d\theta} \tau(\theta)\tag{60}$$

wie unten gezeigt wird. Also gilt

$$\mathbb{V}_\theta(\hat{\tau}_n) \geq \frac{\left(\frac{d}{d\theta} \tau(\theta)\right)^2}{J(\theta)}.\tag{61}$$

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

$$\begin{aligned}C_{\theta}(S(\theta), \hat{\tau}_n) &= \mathbb{E}_{\theta}(S(\theta)\hat{\tau}_n) - \mathbb{E}_{\theta}(S(\theta))\mathbb{E}_{\theta}(\hat{\tau}_n) = \mathbb{E}_{\theta}(S(\theta)\hat{\tau}_n) \\&= \int S(\theta) \hat{\tau}_n p_{\theta}(x) dx \\&= \int \frac{d}{d\theta} \ln L(\theta) \hat{\tau}_n p_{\theta}(x) dx \\&= \int \frac{\frac{d}{d\theta} L(\theta)}{L(\theta)} \hat{\tau}_n p_{\theta}(x) dx \\&= \int \frac{\frac{d}{d\theta} L(\theta)}{p_{\theta}(x)} \hat{\tau}_n p_{\theta}(x) dx && (62) \\&= \int \frac{d}{d\theta} L(\theta) \hat{\tau}_n dx \\&= \frac{d}{d\theta} \int L(\theta) \hat{\tau}_n dx \\&= \frac{d}{d\theta} \int \hat{\tau}_n p_{\theta}(x) dx = \frac{d}{d\theta} \mathbb{E}_{\theta}(\hat{\tau}_n) = \frac{d}{d\theta} \tau(\theta).\end{aligned}$$

Definition (Mittlerer quadratischer Fehler)

$X_1, \dots, X_n \sim p_\theta$ sei die Stichprobe eines parametrischen statistischen Produktmodells \mathcal{M} und $\hat{\tau}_n$ ein Schätzer für τ . Dann ist der *mittlere quadratische Fehler* (engl. *mean squared error*) von $\hat{\tau}_n$ definiert als

$$\text{MQF}(\hat{\tau}_n) := \mathbb{E}_\theta \left((\hat{\tau}_n(X) - \tau(\theta))^2 \right). \quad (63)$$

Bemerkungen

- Der MQF von $\hat{\tau}_n$ ist die erwartete quadrierte Abweichung von $\hat{\tau}_n(X)$ von $\tau(\theta)$.
- Die Varianz von $\hat{\tau}_n$ ist die erwartete quadrierte Abweichung von $\hat{\tau}_n$ von $\mathbb{E}_\theta(\hat{\tau}_n(X))$.
- $\mathbb{E}_\theta(\hat{\tau}_n(X))$ kann mit $\tau(\theta)$ übereinstimmen, muss es aber nicht.

Theorem (Zerlegung des mittleren quadratischen Fehlers)

$X_1, \dots, X_n \sim p_\theta$ sei die Stichprobe eines parametrischen statistischen Produktmodells \mathcal{M} , $\hat{\tau}_n$ sei ein Schätzer für τ , und $\text{MQF}(\hat{\tau}_n)$ sei der mittlere quadratische Fehler von $\hat{\tau}_n$. Dann gilt

$$\text{MQF}(\hat{\tau}_n) = \text{B}(\hat{\tau}_n)^2 + \text{V}_\theta(\hat{\tau}_n). \quad (64)$$

Bemerkungen

- $\text{MQF} = \text{Bias}^2 + \text{Varianz}$.
- Der MQF kann als Bias-Variance Abwägungskriterium benutzt werden.
- Kleine Schätzerverzerrungen können gegenüber einer großen Schätzervarianz präferiert werden.

Beweis

Zur Vereinfachung der Notation seien $\tau := \tau(\theta)$, $\hat{\tau}_n := \hat{\tau}_n(X)$ und $\bar{\tau}_n := \mathbb{E}_\theta(\hat{\tau}_n(X))$ zur Vereinfachung der Notation. Dann gilt:

$$\begin{aligned}
 \mathbb{E}_\theta \left((\hat{\tau}_n - \tau)^2 \right) &= \mathbb{E}_\theta \left((\hat{\tau}_n - \bar{\tau}_n + \bar{\tau}_n - \tau)^2 \right) \\
 &= \mathbb{E}_\theta \left((\hat{\tau}_n - \bar{\tau}_n)^2 + 2(\hat{\tau}_n - \bar{\tau}_n)(\bar{\tau}_n - \tau) + (\bar{\tau}_n - \tau)^2 \right) \\
 &= \mathbb{E}_\theta \left((\hat{\tau}_n - \bar{\tau}_n)^2 \right) + 2\mathbb{E}_\theta \left((\hat{\tau}_n - \bar{\tau}_n)(\bar{\tau}_n - \tau) \right) + \mathbb{E}_\theta \left((\bar{\tau}_n - \tau)^2 \right) \\
 &= \mathbb{E}_\theta \left((\hat{\tau}_n - \bar{\tau}_n)^2 \right) + 2\mathbb{E}_\theta \left(\hat{\tau}_n \bar{\tau}_n - \hat{\tau}_n \tau - \bar{\tau}_n \bar{\tau}_n + \bar{\tau}_n \tau \right) + \mathbb{E}_\theta \left((\bar{\tau}_n - \tau)^2 \right) \\
 &= \mathbb{E}_\theta \left((\hat{\tau}_n - \bar{\tau}_n)^2 \right) + 2 \left(\bar{\tau}_n \bar{\tau}_n - \bar{\tau}_n \tau \right) + \mathbb{E}_\theta \left((\bar{\tau}_n - \tau)^2 \right) \\
 &= \mathbb{E}_\theta \left((\hat{\tau}_n - \bar{\tau}_n)^2 \right) + 0 + \mathbb{E}_\theta \left((\bar{\tau}_n - \tau)^2 \right) \\
 &= \mathbb{E}_\theta \left((\bar{\tau}_n - \tau)^2 \right) + \mathbb{E}_\theta \left((\hat{\tau}_n - \bar{\tau}_n)^2 \right) \\
 &= \mathbb{E}_\theta \left((\mathbb{E}_\theta(\hat{\tau}_n) - \tau)^2 \right) + \mathbb{E}_\theta \left((\hat{\tau}_n - \mathbb{E}_\theta(\hat{\tau}_n))^2 \right) \\
 &= (\mathbb{E}_\theta(\hat{\tau}_n) - \tau)^2 + \mathbb{V}_\theta(\hat{\tau}_n) \\
 &= \mathbf{B}(\hat{\tau}_n)^2 + \mathbb{V}_\theta(\hat{\tau}_n).
 \end{aligned}$$

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.

Ein Schätzer $\hat{\tau}_n$ für τ heißt **asymptotisch erwartungstreu**, wenn der Erwartungswert von $\hat{\tau}_n$ für große Stichprobengrößen $n \rightarrow \infty$ gleich dem wahren, aber unbekanntem, Wert $\tau(\theta)$ ist.

Ein Schätzer $\hat{\tau}_n$ für τ heißt **konsistent**, wenn für große Stichprobengrößen $n \rightarrow \infty$ die Wahrscheinlichkeit dafür, dass $\hat{\tau}_n(X)$ vom wahren, aber unbekanntem, Wert $\tau(\theta)$ abweicht beliebig klein wird.

Ein Schätzer $\hat{\tau}_n$ für τ heißt **asymptotisch normalverteilt**, wenn für große Stichprobengrößen $n \rightarrow \infty$, die Verteilung von $\hat{\tau}_n$ durch eine Normalverteilung gegeben ist.

Ein Schätzer $\hat{\tau}_n$ für τ heißt **asymptotisch effizient**, wenn für große Stichprobengrößen $n \rightarrow \infty$ die Verteilung von $\hat{\tau}_n$ durch eine Normalverteilung mit Erwartungswertparameter $\tau(\theta)$ und Varianzparameter gleich der Cramér-Rao-Schranke gegeben ist.

Definition (Asymptotische Erwartungstreue)

$X_1, \dots, X_n \sim p_\theta$ sei die Stichprobe eines parametrischen statistischen Produktmodells \mathcal{M} und $\hat{\tau}_n$ sei ein Schätzer für τ . $\hat{\tau}_n$ heißt *asymptotisch erwartungstreu*, wenn

$$\lim_{n \rightarrow \infty} \mathbb{E}_\theta(\hat{\tau}_n(X)) = \tau(\theta) \text{ für alle } \theta \in \Theta. \quad (65)$$

Bemerkungen

- Asymptotisch erwartungstreue Schätzer sind für “unendlich große” Stichproben erwartungstreu.
- Erwartungstreue Schätzer sind immer auch asymptotisch erwartungstreu.

Beispiel (Ein asymptotisch erwartungstreuer Schätzer)

Es sei \mathcal{M} das Normalverteilungsmodell, also $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ und

$$\hat{\sigma}_n^{2\text{ML}} := \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \quad (66)$$

sei der ML Schätzer von σ^2 . Mit der Erwartungstreue der Stichprobenvarianz ergibt sich dann

$$\mathbb{E}_{\mu, \sigma^2} \left(\hat{\sigma}_n^{2\text{ML}} \right) = \mathbb{E}_{\mu, \sigma^2} \left(\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \right) = \frac{1}{n} \mathbb{E}_{\mu, \sigma^2} \left(\sum_{i=1}^n (X_i - \bar{X}_n)^2 \right) = \frac{n-1}{n} \sigma^2$$

Also gilt $\mathbb{E}_{\mu, \sigma^2} \left(\hat{\sigma}_n^{2\text{ML}} \right) \neq \sigma^2$. $\hat{\sigma}_n^{2\text{ML}}$ ist also ein verzerrter Schätzer von σ^2 . Allerdings gilt $(n-1)/n \rightarrow 1$ für $n \rightarrow \infty$, so dass

$$\lim_{n \rightarrow \infty} \mathbb{E}_{\mu, \sigma^2} \left(\hat{\sigma}_n^{2\text{ML}} \right) = \lim_{n \rightarrow \infty} \frac{n-1}{n} \sigma^2 = \sigma^2 \lim_{n \rightarrow \infty} \frac{n-1}{n} = \sigma^2. \quad (67)$$

Im Limit großer Stichprobenumfänge ist der ML Schätzer des Varianzparameters einer Normalverteilung also erwartungstreu.

Simulation $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ mit $\mu = 1, \sigma^2 = 2$

```

# Modellformulierung
mu      = 1                # wahrer, aber unbekannter, Erwartungswertparameter
sigsqr  = 2                # wahrer, aber unbekannter, Varianzparameter
n       = seq(1,100, by = 2) # Stichprobengroessen
ns      = 1e3              # Anzahl Simulation pro Stichprobengroesse
sigsqr_ml = matrix(        # \hat{\sigma^2}^{ML} Array
  rep(NaN, length(n)*ns),
  ncol = length(n))

# Stichprobengroesseniterationen
for(i in seq_along(n)){

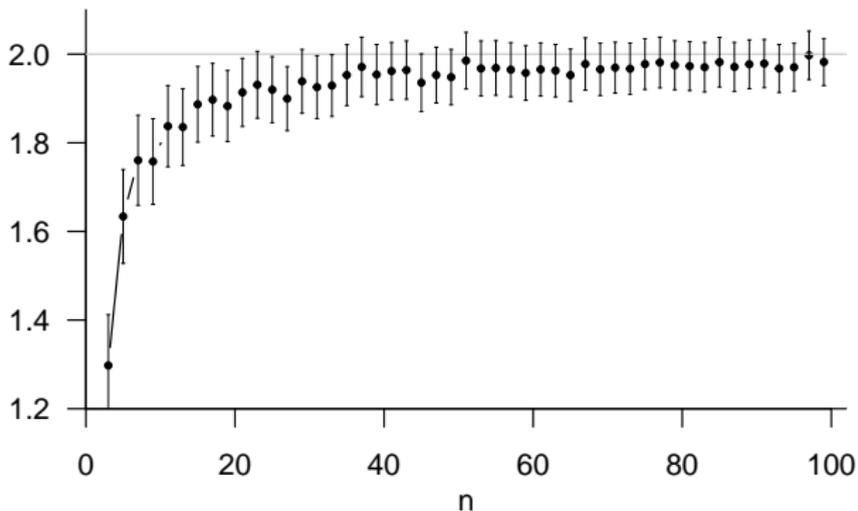
  # Simulationsiterationen
  for(s in 1:ns){

    # Stichprobenrealisation
    x = rnorm(n[i], mu, sqrt(sigsqr))

    # \hat{\sigma^2}^{ML}
    sigsqr_ml[s,i] = ((n[i]-1)/n[i])*var(x)
  }
}
E_sigsqr_ml = colMeans(sigsqr_ml) # Erwartungswertschaetzung

```

Simulation $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ mit $\mu = 1, \sigma^2 = 2$



Definition (Konsistenz)

$X_1, \dots, X_n \sim p_\theta$ sei die Stichprobe eines parametrischen statistischen Produktmodells \mathcal{M} und $\hat{\tau}_n$ sei ein Schätzer von τ . Eine Folge von Schätzern $\hat{\tau}_1, \hat{\tau}_2, \dots$ wird dann eine *konsistente Folge von Schätzern* genannt, wenn für jedes $\epsilon > 0$ und jedes $\theta \in \Theta$ gilt, dass

$$\lim_{n \rightarrow \infty} \mathbb{P}_\theta (|\hat{\tau}_n(X) - \tau(\theta)| \geq \epsilon) = 0.$$

Wenn $\hat{\tau}_1, \hat{\tau}_2, \dots$ eine konsistente Folge von Schätzern ist, dann heißt $\hat{\tau}_n$ *konsistenter Schätzer*.

Bemerkungen

- Für $n \rightarrow \infty$ wird die Wahrscheinlichkeit, dass $\hat{\tau}_n(X)$ beliebig nah bei $\tau(\theta)$ liegt, groß.
- Für $n \rightarrow \infty$ wird die Wahrscheinlichkeit, dass $\hat{\tau}_n(X)$ von $\tau(\theta)$ abweicht, klein.
- Diese Eigenschaften gelten für alle möglichen wahren, aber unbekanntenen, Parameterwerte.
- Die Konvergenz ist *Konvergenz in Wahrscheinlichkeit*.
- Konsistenz von Schätzern kann direkt oder mit Kriterien nachgewiesen werden.

Theorem (Mittlerer-Quadratischer-Fehler-Kriterium für Konsistenz)

$X_1, \dots, X_n \sim p_\theta$ sei die Stichprobe eines parametrischen statistischen Produktmodells \mathcal{M} und $\hat{\tau}_n$ sei ein Schätzer von τ . Wenn $\lim_{n \rightarrow \infty} \text{MQF}(\hat{\tau}_n) = 0$ gilt, dann ist $\hat{\tau}_n$ ein konsistenter Schätzer.

Beweis

Mit der Chebychev-Ungleichung folgt, dass

$$\mathbb{P}_\theta (|\hat{\tau}_n(X) - \tau(\theta)| \geq \epsilon) \leq \frac{\mathbb{E}_\theta \left((\hat{\tau}_n(X) - \tau(\theta))^2 \right)}{\epsilon^2} \quad (68)$$

Grenzwertbildung ergibt dann

$$\lim_{n \rightarrow \infty} \mathbb{P}_\theta (|\hat{\tau}_n(X) - \tau(\theta)| \geq \epsilon) \leq \frac{1}{\epsilon^2} \lim_{n \rightarrow \infty} \mathbb{E}_\theta \left((\hat{\tau}_n(X) - \tau(\theta))^2 \right). \quad (69)$$

Wenn also $\lim_{n \rightarrow \infty} \mathbb{E}_\theta \left((\hat{\tau}_n(X) - \tau(\theta))^2 \right) = 0$ gilt, dann gilt mit $\mathbb{P}_\theta (|\hat{\tau}_n(X) - \tau(\theta)| \geq \epsilon) \geq 0$, dass

$$\lim_{n \rightarrow \infty} \mathbb{P}_\theta (|\hat{\tau}_n(X) - \tau(\theta)| \geq \epsilon) = 0. \quad (70)$$

Also ist $\hat{\tau}_n$ ein konsistenter Schätzer.

Theorem (Bias-Varianz-Kriterium für Konsistenz)

$X_1, \dots, X_n \sim p_\theta$ sei die Stichprobe eines parametrischen statistischen Produktmodells \mathcal{M} und $\hat{\tau}_n$ sei ein Schätzer von τ . Wenn

$$\lim_{n \rightarrow \infty} B(\hat{\tau}_n) = 0 \text{ und } \lim_{n \rightarrow \infty} \mathbb{V}_\theta(\hat{\tau}_n) = 0 \quad (71)$$

gilt, dann ist $\hat{\tau}_n$ ein konsistenter Schätzer

Beweis

Wenn $n \rightarrow \infty$, dann gilt $B(\hat{\tau}_n) \rightarrow 0$, also auch $B(\hat{\tau}_n)^2 \rightarrow 0$. Wenn für $n \rightarrow \infty$ sowohl $B(\hat{\tau}_n)^2 \rightarrow 0$ als auch $\mathbb{V}_\theta(\hat{\tau}_n) \rightarrow 0$, dann gilt auch $\lim_{n \rightarrow \infty} \text{MQF}(\hat{\theta}_n) = 0$. Also gilt mit dem MQF-Kriterium, dass $\hat{\tau}_n$ konsistent ist.

Beispiel (Konsistenz des Stichprobenmittels)

Für $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ folgt die Konsistenz des Stichprobenmittels als Schätzer für $\mathbb{E}(X_0)$ aus dem Bias-Varianz-Kriterium durch

$$B(\bar{X}_n) = 0 \text{ und } \lim_{n \rightarrow \infty} \mathbb{V}_\theta(\bar{X}_n) = \lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{V}(X_0) = 0. \quad (72)$$

Simulation \bar{X}_n bei $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ mit $\mu = 1, \sigma^2 = 2$

```
# Modellformulierung
mu      = 1
sigsqr  = 2
n       = seq(1,1e3,by = 10)
eps     = c(0.15, 0.10, 0.05)
ne      = length(eps)
nn      = length(n)
ns      = 1e3
E       = array(rep(NA,n,ne,ns),
               dim = c(nn,ne,ns))

# Simulation
for(e in seq_along(eps)){
  for(i in seq_along(n)){
    for(s in 1:ns){

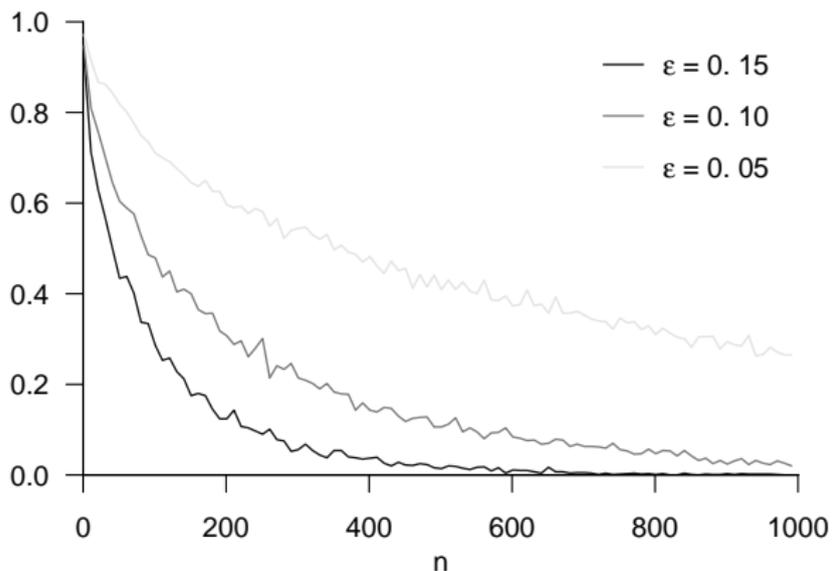
      # Stichprobenrealisationen
      x = rnorm(n[i], mu, sqrt(sigsqr))
      if(abs(mean(x) - mu) >= eps[e]){
        E[i,e,s] = 1
      } else {
        E[i,e,s] = 0
      }
    }
  }
}

# Schaetzung von \mathbb{P}(|\hat{\tau}_n(X) - \tau(\theta)| \ge \epsilon)
P_hat  = apply(E, c(1,2), mean)

# w.a.u \mu Wert
# w.a.u. \sigma^2 Wert
# Stichprobengroesse n
# \epsilon Werte
# Anzahl \epsilon Werte
# Anzahl Stichprobengroessen
# Anzahl Simulationen
# Ereignisindikatorarray

# \epsilon Iterationen
# n Iterationen
# Simulationsiterationen
```

Simulation \bar{X}_n bei $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ mit $\mu = 1, \sigma^2 = 2$



Definition (Asymptotische Normalität)

$X_1, \dots, X_n \sim p_\theta$ sei die Stichprobe eines parametrischen statistischen Produktmodells \mathcal{M} und $\hat{\theta}_n$ sei ein Parameterschätzer für θ . Weiterhin sei $\tilde{\theta} \sim N(\mu, \sigma^2)$ eine normalverteilte Zufallsvariable mit Erwartungswertparameter μ und Varianzparameter σ^2 . Wenn $\hat{\theta}_n$ in Verteilung gegen $\tilde{\theta}$ konvergiert, dann heißt $\hat{\theta}_n$ *asymptotisch normalverteilt* und wir schreiben

$$\hat{\theta}_n \stackrel{a}{\sim} N(\mu, \sigma^2). \quad (73)$$

Bemerkung

- Konvergenz in Verteilung heißt $\lim_{n \rightarrow \infty} P_n(\hat{\theta}_n) = P(\tilde{\theta})$.

Definition (Asymptotische Effizienz)

$X_1, \dots, X_n \sim p_\theta$ sei die Stichprobe eines parametrischen statistischen Produktmodells \mathcal{M} und $\hat{\theta}_n$ sei ein Parameterschätzer für θ . Weiterhin sei $J_n(\theta)$ die erwartete Fisher-Information der Stichprobe X_1, \dots, X_n . Wenn gilt, dass

$$\hat{\theta}_n \stackrel{a}{\sim} N\left(\theta, J_n(\theta)^{-1}\right), \quad (74)$$

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.

Grundbegriffe

Maximum-Likelihood Schätzer

Schätzereigenschaften bei endlichen Stichproben

Asymptotische Schätzereigenschaften

Eigenschaften von Maximum-Likelihood Schätzern

Selbstkontrollfragen

Theorem (Eigenschaften von ML Schätzern)

$X_1, \dots, X_n \sim p_\theta$ sei die Stichprobe eines parametrischen statistischen Produktmodells \mathcal{M} und $\hat{\theta}_n^{\text{ML}}$ sei ein ML Schätzer für θ . Dann gilt, dass $\hat{\theta}_n^{\text{ML}}$

- (1) nicht notwendigerweise erwartungstreu, aber
- (2) konsistent,
- (3) asymptotisch normalverteilt,
- (4) asymptotisch erwartungstreu, und
- (5) asymptotisch effizient ist.

Für einen Beweis verweisen wir auf Held and Sabanés Bové (2014), Abschnitt 3.4

Simulation der asymptotische Effizienz des Bernoulli ML Parametschätzers

- Es sei $X_1, \dots, X_n \sim \text{Bern}(\mu)$.
- Der ML Schätzer von μ ist $\hat{\mu}_n^{\text{ML}} = \frac{1}{n} \sum_{i=1}^n X_i$.
- Die erwartete Fisher-Information der Stichprobe X_1, \dots, X_n ist

$$J_n(\mu) = \frac{n}{\mu(1-\mu)}. \quad (75)$$

- Also gilt

$$\hat{\mu}_n^{\text{ML}} \stackrel{a}{\sim} N\left(\mu, \frac{\mu(1-\mu)}{n}\right). \quad (76)$$

Eigenschaften von Maximum-Likelihood Schätzern

Simulation der asymptotische Effizienz des Bernoulli ML Parameterschätzers

```
# Modellformulierung
mu           = 0.4                                # w.a.u. Parameterwert
n_all       = c(1e1,5e1,1e2)                     # Stichprobengroesse n
ns          = 1e4                                 # Anzahl der Simulationen
mu_hat_ML   = matrix(                             # ML Schaetzerarray
  rep(NA,
      length(n_all)*ns),
  nrow = length(n_all))

mu_hat_ML_r = 1e3                                 # ML Schaetzerraumaufloesung
mu_hat_ML_x = seq(0,1,len = mu_hat_ML_r)         # ML Schaetzerraum
mu_hat_ML_p = matrix(rep(NA, length(n_all)*mu_hat_ML_r), # ML WDF Array
  nrow = length(n_all))

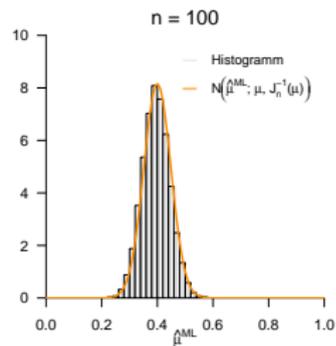
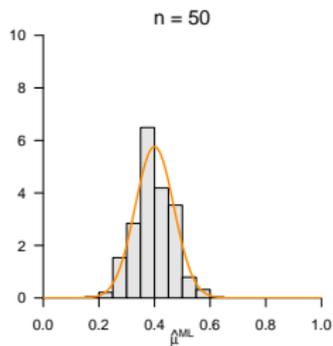
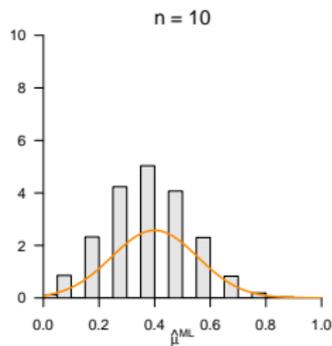
# Stichprobengroesseniterationen
for(i in seq_along(n_all)){

  # Simulationsiterationen
  for(s in 1:ns){
    x           = rbinom(n_all[i],1,mu)           # Stichprobenrealisation
    mu_hat_ML[i,s] = mean(x)                       # ML Schaetzer
  }

  # WDF der asymptotischen Verteilung
  mu_hat_ML_p[i,] = dnorm(mu_hat_ML_x, mu, sqrt(mu*(1-mu)/n_all[i]))
}
```

Eigenschaften von Maximum-Likelihood Schätzern

Simulation der asymptotische Effizienz des Bernoulli ML Parametschätzers



Selbstkontrollfragen

1. Definieren und erläutern Sie den Begriff des Parameterpunktschätzers.
2. Definieren Sie die Begriffe der Likelihood-Funktion und der Log-Likelihood Funktion.
3. Definieren Sie den Begriff des Maximum-Likelihood Schätzers.
4. Erläutern Sie das Vorgehen zur ML Schätzung für ein parametrisches statistisches Produktmodell.
5. Geben Sie den ML Schätzer für den Parameter μ des Bernoullimodells an.
6. Geben Sie den ML Schätzer für den Parameter μ des Normalverteilungmodells an.
7. Geben Sie den ML Schätzer für den Parameter σ^2 des Normalverteilungmodells an.
8. Definieren und erläutern Sie den Begriff der Erwartungstreue eines Schätzers.
9. Definieren Sie die Begriffe der Varianz und des Standardfehlers eines Schätzers.
10. Definieren Sie den Begriff der Scorefunktion einer Stichprobe.
11. Definieren Sie den Begriff der Fisher-Information einer Stichprobe.
12. Definieren Sie den Begriff der erwarteten Fisher-Information einer Stichprobe.
13. Geben Sie das Theorem zur Cramér-Rao-Ungleichung wieder.
14. Erläutern Sie den Begriff der Cramér-Rao-Schranke.
15. Definieren Sie den Begriff des MQFs eines Schätzers.
16. Geben Sie das Theorem zur Zerlegung des MQFs wieder.
17. Erläutern Sie den Begriff des asymptotischen Erwartungstreue eines Schätzers
18. Erläutern Sie den Begriff der Konsistenz eines Schätzers.
19. Erläutern Sie den Begriff der asymptotischen Normalität eines Schätzers.
20. Erläutern Sie den Begriff der asymptotischen Effizienz eines Schätzers.
21. Nennen Sie fünf Eigenschaften eines ML Schätzers.

Appendix

Appendix

Beweis der Additivität der Fisher-Information

Wir zeigen das Resultat für die erwartete Fisher-Information, das Resultat für die Fisher-Information ist dann implizit. Per definitionem und mit der Linearität von Ableitungen und Erwartungswerte gilt

$$\begin{aligned} J_n(\theta) &= \mathbb{E}_\theta \left(-\frac{d^2}{d\theta^2} \ell_n(\theta) \right) \\ &= \mathbb{E}_\theta \left(-\frac{d^2}{d\theta^2} \ln \left(\prod_{i=1}^n p_\theta(x_i) \right) \right) \\ &= \mathbb{E}_\theta \left(-\frac{d^2}{d\theta^2} \sum_{i=1}^n \ln p_\theta(x_i) \right) \\ &= \mathbb{E}_\theta \left(-\frac{d^2}{d\theta^2} \sum_{i=1}^n \ln p_\theta(x_1) \right) \\ &= \mathbb{E}_\theta \left(-\frac{d^2}{d\theta^2} \ell_1(\theta) n \right) \\ &= n \mathbb{E}_\theta \left(-\frac{d^2}{d\theta^2} \ell_1(\theta) \right) \\ &= n J_n(\theta). \end{aligned} \tag{77}$$

Appendix

Beweis von Erwartungswert und Varianz der Scorefunktion

Wir betrachten nur den Fall, dass p_θ eine WDF ist und zeigen zunächst, dass $\mathbb{E}_\theta(S(\theta)) = 0$ ist:

$$\begin{aligned}\mathbb{E}_\theta(S(\theta)) &= \int S(\theta) p_\theta(x) dx \\ &= \int \frac{d}{d\theta} \ell(\theta) p_\theta(x) dx \\ &= \int \frac{d}{d\theta} \ln L(\theta) p_\theta(x) dx \\ &= \int \frac{1}{L(\theta)} \frac{d}{d\theta} L(\theta) p_\theta(x) dx && (78) \\ &= \int \frac{1}{p_\theta(x)} \frac{d}{d\theta} L(\theta) p_\theta(x) dx \\ &= \int \frac{d}{d\theta} L(\theta) dx \\ &= \frac{d}{d\theta} \int p_\theta(x) dx = \frac{d}{d\theta} 1 = 0.\end{aligned}$$

Appendix

Mit der Definition der Varianz folgt dann sofort, dass $\mathbb{V}_\theta(S(\theta)) = \mathbb{E}_\theta \left(S(\theta)^2 \right)$ ist. Als nächstes zeigen wir, dass $J(\theta) = \mathbb{E}_\theta \left(S(\theta)^2 \right)$ und deshalb $\mathbb{V}_\theta(S(\theta)) = J(\theta)$ ist:

$$\begin{aligned} J(\theta) &= \mathbb{E}_\theta \left(-\frac{d^2}{d\theta^2} \ln L(\theta) \right) \\ &= \mathbb{E}_\theta \left(-\frac{d}{d\theta} \frac{\frac{d}{d\theta} L(\theta)}{L(\theta)} \right) \\ &= \mathbb{E}_\theta \left(-\frac{\frac{d^2}{d\theta^2} L(\theta) L(\theta) - \frac{d}{d\theta} L(\theta) \frac{d}{d\theta} L(\theta)}{L(\theta) L(\theta)} \right) \\ &= -\mathbb{E}_\theta \left(\frac{\frac{d^2}{d\theta^2} L(\theta)}{L(\theta)} \right) + \mathbb{E}_\theta \left(\frac{\left(\frac{d}{d\theta} L(\theta) \right)^2}{(L(\theta))^2} \right) \\ &= -\int \frac{\frac{d^2}{d\theta^2} L(\theta)}{L(\theta)} p_\theta(x) dx + \int \frac{\left(\frac{d}{d\theta} L(\theta) \right)^2}{(L(\theta))^2} p_\theta(x) dx \\ &= -\frac{d^2}{d\theta^2} \int p_\theta(x) dx + \int \left(\frac{1}{L(\theta)} \frac{d}{d\theta} L(\theta) \right)^2 p_\theta(x) dx \\ &= -\frac{d^2}{d\theta^2} 1 + \int \left(\frac{d}{d\theta} \ln L(\theta) \right)^2 p_\theta(x) dx = \mathbb{E}_\theta \left(S(\theta)^2 \right). \end{aligned} \tag{79}$$

Beweis der Scorefunktion und Fisher-Information für den Bernoulli Parameter

Die Scorefunktion wurde bereits im Kontext der Maximum-Likelihood-Schätzung von μ hergeleitet. Wir betrachten die Fisher-Information einer einzelnen Bernoulli Zufallsvariable X :

$$\begin{aligned} I(\mu) &:= -\frac{d^2}{d\mu^2} \ell_1(\mu) \\ &= -\frac{d^2}{d\mu^2} \ln p_\mu(x) \\ &= -\frac{d^2}{d\mu^2} (x \ln \mu + (1-x) \ln(1-\mu)) \\ &= -\frac{d}{d\mu} \left(\frac{d}{d\mu} (x \ln \mu + (1-x) \ln(1-\mu)) \right) \\ &= -\frac{d}{d\mu} \left(\frac{x}{\mu} + \frac{(1-x)}{1-\mu} \right) \\ &= -\left(-\frac{x}{\mu^2} - \frac{(1-x)}{1-\mu} \right) \\ &= \frac{x}{\mu^2} + \frac{(1-x)^2}{1-\mu} . \end{aligned} \tag{80}$$

Damit ergibt sich die erwartete Fisher-Information der Zufallsvariable X als

$$\begin{aligned} J(\mu) &= \mathbb{E}_\mu(I(\mu)) \\ &= \mathbb{E}_\mu \left(\frac{X}{\mu^2} + \frac{(1-X)^2}{1-\mu} \right) \\ &= \frac{\mathbb{E}_\mu(X)}{\mu^2} + \frac{(1-\mathbb{E}_\mu(X))^2}{1-\mu} \\ &= \frac{\mu}{\mu^2} + \frac{(1-\mu)^2}{1-\mu} \\ &= \frac{1}{\mu(1-\mu)}. \end{aligned} \tag{81}$$

Mit der Additivitätseigenschaft der Fisher-Information und der Definition der beobachteten Fisher-Information ergibt sich dann sofort

$$I_n(\mu) = \frac{nx}{\mu^2} + \frac{n(1-x)^2}{1-\mu}, \text{ und } J_n(\mu) = \frac{n}{\mu(1-\mu)}. \tag{82}$$

Appendix

Beweis der Scorefunktion und Fisher-Information für den Erwartungswertparameter der Normalverteilung

Wir erinnern uns, dass die Log-Likelihood-Funktion der Stichprobe $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ bei bekanntem Varianzparameter σ^2 durch

$$\ell_n : \mathbb{R} \rightarrow \mathbb{R}, \mu \mapsto \ell_n(\mu) := -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2. \quad (83)$$

gegeben ist. Damit ergibt sich die Scorefunktion als

$$S_n(\mu) = \frac{d}{d\mu} \ell_n(\mu) = \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu) \quad (84)$$

Die Fisher-Information der Stichprobe X_1, \dots, X_n ergibt sich als

$$I_n(\mu) = -\frac{d^2}{d\mu^2} \ell_n(\mu) = -\frac{d}{d\mu} S_n(\mu) = -\frac{1}{\sigma^2} \frac{d}{d\mu} \left(\sum_{i=1}^n x_i - n\mu \right) = \frac{n}{\sigma^2}. \quad (85)$$

Die beobachtete Fisher-Information ist die Fisher-Information an der Stelle des Maximum-Likelihood-Schätzers $\hat{\mu}_n^{\text{ML}}$. Die erwartete Fisher-Information schließlich ergibt sich als

$$J_n(\mu) = \mathbb{E}_\mu(I_n(\mu)) = \mathbb{E}_\mu \left(\frac{n}{\sigma^2} \right) = \frac{n}{\sigma^2}. \quad (86)$$

Appendix

Beweis der Scorefunktion und Fisher-Information für den Varianzparameter der Normalverteilung

Wir erinnern uns, dass die Log-Likelihood-Funktion der Stichprobe $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ bei bekanntem Erwartungswert-Parameter μ durch

$$\ell_n : \mathbb{R}_{>0} \rightarrow \mathbb{R}, \sigma^2 \mapsto \ell_n(\sigma^2) := -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2. \quad (87)$$

gegeben ist. Die Scorefunktion ergibt sich also als

$$S_n(\sigma^2) = \frac{d}{d\sigma^2} \ell_n(\sigma^2) = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - \mu)^2. \quad (88)$$

Die Fisher-Information der Stichprobe X_1, \dots, X_n ergibt sich als

$$\begin{aligned} I_n(\sigma^2) &= -\frac{d}{d\sigma^2} S_n(\sigma^2) = -\left(\frac{n}{2\sigma^4} - \frac{1}{\sigma^6} \sum_{i=1}^n (x_i - \mu)^2 \right) \\ &= \frac{1}{\sigma^6} \sum_{i=1}^n (x_i - \mu)^2 - \frac{n}{2\sigma^4}. \end{aligned} \quad (89)$$

Die beobachtete Fisher-Information ist die Fisher-Information an der Stelle des Maximum-Likelihood-Schätzers $\hat{\sigma}_n^{2\text{ML}}$.

$$\begin{aligned}
 I_n(\hat{\sigma}_n^{2\text{ML}}) &= \frac{\sum_{i=1}^n (x_i - \mu)^2}{\left(\hat{\sigma}_n^{2\text{ML}}\right)^3} - \frac{n}{2 \left(\hat{\sigma}_n^{2\text{ML}}\right)^2} \\
 &= \frac{\sum_{i=1}^n (x_i - \mu)^2}{\frac{1}{n^3} \left(\sum_{i=1}^n (x_i - \mu)^2\right)^3} - \frac{n}{2 \left(\hat{\sigma}_n^{2\text{ML}}\right)^2} \\
 &= \frac{1}{\frac{1}{n^3} \left(\sum_{i=1}^n (x_i - \mu)^2\right)^2} - \frac{n}{2 \left(\hat{\sigma}_n^{2\text{ML}}\right)^2} \\
 &= \frac{n}{\left(\hat{\sigma}_n^{2\text{ML}}\right)^2} - \frac{n}{2 \left(\hat{\sigma}_n^{2\text{ML}}\right)^2} \\
 &= \frac{n}{2 \left(\hat{\sigma}_n^{2\text{ML}}\right)^2} \\
 &= \frac{n}{2 \hat{\sigma}_n^4 \text{ML}}.
 \end{aligned} \tag{90}$$

Die erwartete Fisher-Information ergibt sich als

$$\begin{aligned} J_n(\sigma^2) &= \mathbb{E}_{\sigma^2}(I_n(\sigma^2)) \\ &= \mathbb{E}_{\sigma^2} \left(\frac{1}{\sigma^6} \sum_{i=1}^n (x_i - \mu)^2 - \frac{n}{2\sigma^4} \right) \\ &= \frac{1}{\sigma^6} \sum_{i=1}^n \mathbb{E}_{\sigma^2} \left((x_i - \mu)^2 \right) - \frac{n}{2\sigma^4} \\ &= \frac{1}{\sigma^6} \sum_{i=1}^n \sigma^2 - \frac{n}{2\sigma^4} \\ &= \frac{n\sigma^2}{\sigma^6} - \frac{n}{2\sigma^4} \\ &= \frac{n}{\sigma^4} - \frac{n}{2\sigma^4} \\ &= \frac{n}{2\sigma^4}. \end{aligned} \tag{91}$$

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 2021/22

Prof. Dr. Dirk Ostwald

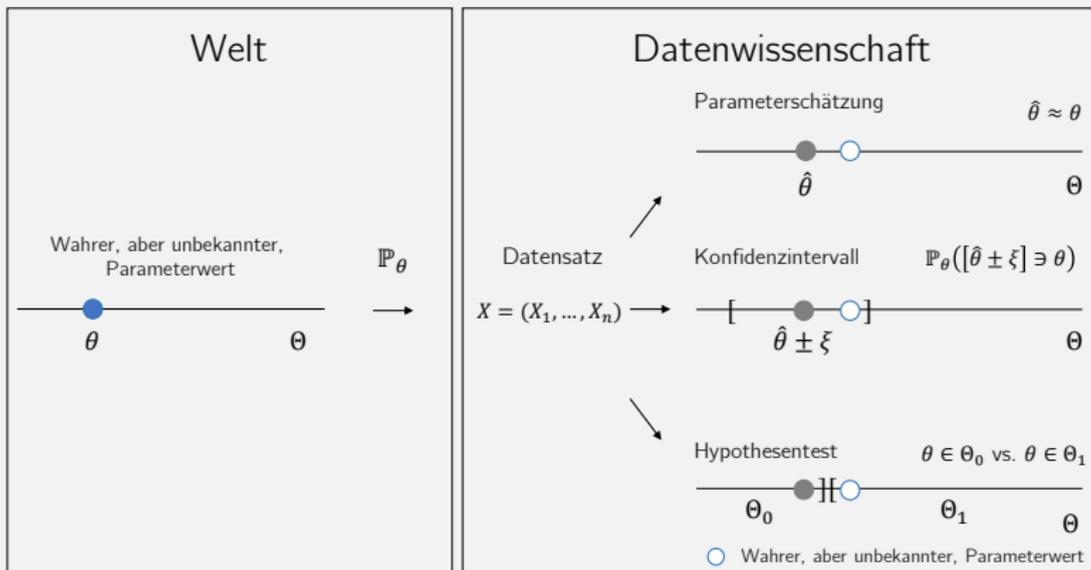
(11) Konfidenzintervalle

Wahrscheinlichkeitstheorie und Frequentistische Inferenz

Datum	Einheit	Thema
14.10.2021	Einführung	(1) Einführung
21.10.2021	Wahrscheinlichkeitstheorie	(2) Wahrscheinlichkeitsräume
28.10.2021	Wahrscheinlichkeitstheorie	(3) Elementare Wahrscheinlichkeiten
04.11.2021	Wahrscheinlichkeitstheorie	(4) Zufallsvariablen
11.11.2021	Wahrscheinlichkeitstheorie	(5) Multivariate Verteilungen
18.11.2021	Wahrscheinlichkeitstheorie	(6) Erwartungswert, Varianz, Kovarianz
25.11.2021	Wahrscheinlichkeitstheorie	(7) Ungleichungen und Grenzwerte
02.12.2021	Wahrscheinlichkeitstheorie	(8) Transformationen der Normalverteilung
09.12.2021	Frequentistische Inferenz	(9) Grundbegriffe Frequentistischer Inferenz
16.12.2021	Frequentistische Inferenz	(10) Parameterschätzung
	Weihnachtspause	
13.01.2022	Frequentistische Inferenz	(11) Konfidenzintervalle
20.01.2022	Frequentistische Inferenz	(12) Hypothesentests Frage-Antwort-Session
27.01.2022	Frequentistische Inferenz	Frage-Antwort-Session
31.01.2022	Klausur	16 - 17 Uhr, G44 - H6
Jul 2022	Klausurwiederholungstermin	

- (12) Hypothesentests ist als Online Vorlesung verfügbar, der Termin dient als Frage-Antwort-Session.
- (13) Einstichproben-T-Tests ist als **nicht klausurrelevante** Online Vorlesung verfügbar.
- (14) Zweistichproben-T-Tests ist als **nicht klausurrelevante** Online Vorlesung verfügbar.

Modell und Standardprobleme der Frequentistischen Inferenz



(1) Parameterschätzung

Ziel der Parameterschätzung ist es, einen möglichst guten Tipp für den wahren, aber unbekanntem, Parameterwert (oder eine Funktion dessen) abzugeben, typischerweise basierend auf der Beobachtung einer Realisierung von $X_1, \dots, X_n \sim p_\theta$.

(2) Konfidenzintervalle

Das Ziel der Bestimmung von Konfidenzintervallen ist es, basierend auf der Verteilung möglicher Parameterschätzwerte eine quantitative Aussage über die mit dem Schätzwert assoziierte Unsicherheit zu treffen.

(3) Hypothesentests

Das Ziel der Auswertung von Hypothesentests ist es, basierend auf der angenommenen Verteilung der Beobachtungen X_1, \dots, X_n in einer möglichst sinnvollen Form zu entscheiden, ob der wahre, aber unbekanntem Parameterwert, in einer von zwei sich gegenseitig ausschließenden Untermengen des Parameterraumes, welche man als Hypothesen bezeichnet, liegt.

Standardannahmen Frequentistischer Inferenz

\mathcal{M} sei ein statistisches Modell mit $X_1, \dots, X_n \sim \mathbb{P}_\theta$. Es wird angenommen, dass ein vorliegender Datensatz eine der möglichen Realisierungen von $X_1, \dots, X_n \sim \mathbb{P}_\theta$ ist. Aus frequentistischer Sicht kann man die Erhebung von Datensätzen unendlich oft wiederholen und zu jedem Datensatz Statistiken auswerten.

$$\text{Datensatz (1)} : x^{(1)} = \left(x_1^{(1)}, x_2^{(1)}, \dots, x_n^{(1)} \right), \text{ Statistik (1): } S : \mathbb{R}^n \rightarrow \Sigma, x^{(1)} \mapsto S \left(x^{(1)} \right)$$

$$\text{Datensatz (2)} : x^{(2)} = \left(x_1^{(2)}, x_2^{(2)}, \dots, x_n^{(2)} \right), \text{ Statistik (2): } S : \mathbb{R}^n \rightarrow \Sigma, x^{(2)} \mapsto S \left(x^{(2)} \right)$$

$$\text{Datensatz (3)} : x^{(3)} = \left(x_1^{(3)}, x_2^{(3)}, \dots, x_n^{(3)} \right), \text{ Statistik (3): } S : \mathbb{R}^n \rightarrow \Sigma, x^{(3)} \mapsto S \left(x^{(3)} \right)$$

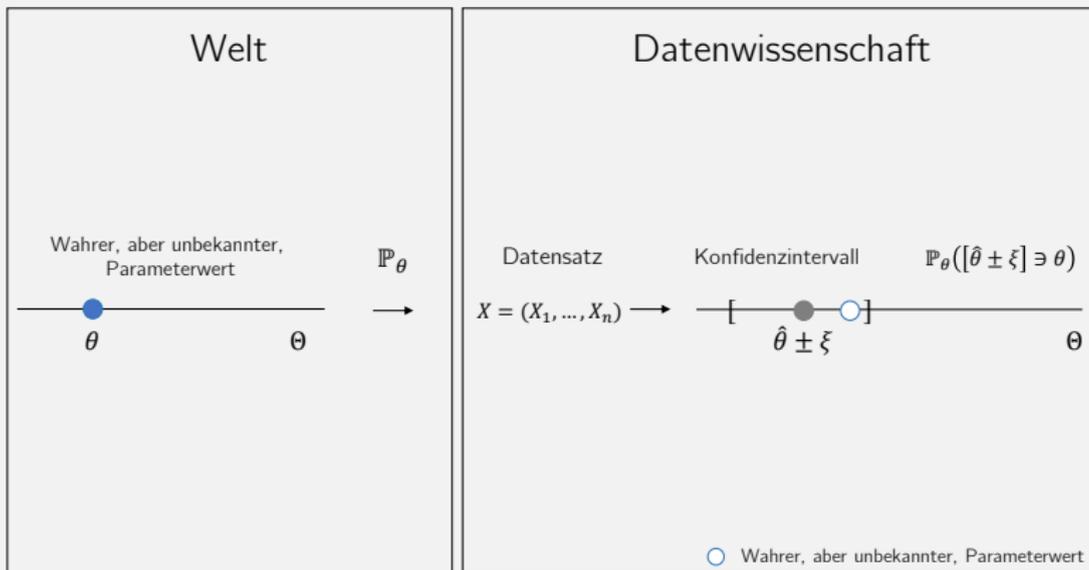
$$\text{Datensatz (4)} : x^{(4)} = \left(x_1^{(4)}, x_2^{(4)}, \dots, x_n^{(4)} \right), \text{ Statistik (4): } S : \mathbb{R}^n \rightarrow \Sigma, x^{(4)} \mapsto S \left(x^{(4)} \right)$$

...

Um die Qualität statistischer Methoden zu beurteilen betrachtet die frequentistische Statistik deshalb die Wahrscheinlichkeitsverteilungen von Statistiken und Schätzern unter der Annahme von $X_1, \dots, X_n \sim p_\theta$.

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 realen Normalfall nur eines vorliegenden Datensatzes, kann sie auch "schlecht" sein.

Modell und Standardprobleme der Frequentistischen Inferenz



Definition

Beispiele

Selbstkontrollfragen

Appendix

Definition

Beispiele

Selbstkontrollfragen

Appendix

Definition (δ -Konfidenzintervall)

Es sei $X = X_1, \dots, X_n \sim p_\theta$ eine Stichprobe, $\delta \in]0, 1[$, und $G_u(X)$ und $G_o(X)$ seien zwei Statistiken. Dann ist ein δ -Konfidenzintervall ein Intervall der Form $K_n := [G_u, G_o]$, so dass

$$\mathbb{P}_\theta (K_n \ni \theta) = \mathbb{P}_\theta (G_u(X) \leq \theta \leq G_o(X)) = \delta \text{ für alle } \theta \in \Theta \text{ gilt.} \quad (1)$$

δ heißt das *Konfidenzniveau* oder die *Überdeckungswahrscheinlichkeit* des Konfidenzintervalls. Die Statistiken $G_u(X)$ und $G_o(X)$ sind die unteren und oberen Grenzen des Konfidenzintervalls.

Bemerkungen

- θ ist fest, nicht zufällig, und unbekannt.
- K_n ist ein zufälliges Intervall, weil $G_u(X)$ und $G_o(X)$ Zufallsvariablen sind.
- $K_n \ni \theta$ bedeutet $\theta \in K_n$, aber K_n ist zufällig und steht deshalb vorn (cf. $\mathbb{P}(X = x)$).
- Ein δ -Konfidenzintervall überdeckt den wahren Wert θ mit Wahrscheinlichkeit δ .
- Oft wird $\delta = 0.95$ gewählt, also *95%-Konfidenzintervalle* betrachtet.

Zwei Interpretationen von δ -Konfidenzintervallen

- (1) Wird ein Zufallsvorgang unter gleichen Umständen häufig wiederholt, so überdeckt das zugehörige δ -Konfidenzintervall den wahren, aber unbekanntem Parameterwert in $\delta \cdot 100\%$ der Fälle. Technischer ausgedrückt, für unabhängig gezogene Stichproben einer Verteilung mit wahren, aber unbekanntem, Parameter θ überdeckt ein entsprechendes δ -Konfidenzintervall θ in $\delta \cdot 100\%$ aller Fälle.
- (2) Man betrachte eine Folge von Zufallsvorgängen mit wahren, aber unbekanntem, Parametern $\theta_1, \theta_2, \dots$ und stelle sich vor δ -Konfidenzintervalle für eben jene von Folge von wahren, aber unbekanntem Parametern $\theta_1, \theta_2, \dots$ zu konstruieren. Dann überdecken $\delta \cdot 100\%$ der Konfidenzintervalle den wahren, aber unbekanntem, Wert $\theta_i, i = 1, 2, \dots$

Definition (Pivot)

Ein *Pivot* ist eine Funktion einer Stichprobe X_1, \dots, X_n (also eine Statistik) und eines wahren, aber unbekanntem, Parameters θ , deren Verteilung nicht von θ abhängt.

Bemerkung

- Ein Pivot ist eine Statistik, die vom wahren, aber unbekanntem Parameter θ abhängt, deren Verteilung aber nicht von dem wahren, aber unbekanntem Parameter θ abhängt. Mithilfe von Pivots kann man Konfidenzintervalle für alle Werte von θ konstruieren.

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) Erwartungswertparameter bei Normalverteilung mit bekannter Varianz
- (2) Varianzparameter bei Normalverteilung
- (3) Erwartungswertparameter bei Normalverteilung mit unbekannter Varianz

Definition

Beispiele

Selbstkontrollfragen

Appendix

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) **Erwartungswertparameter bei Normalverteilung mit bekannter Varianz**
- (2) Varianzparameter bei Normalverteilung
- (3) Erwartungswertparameter bei Normalverteilung mit unbekannter Varianz

(1) Definition des statistischen Modells

Es sei $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ eine Stichprobe, wobei μ unbekannt sei und $\sigma^2 > 0$ bekannt sei. Wir entwickeln ein δ -Konfidenzintervall für den Erwartungswertparameter μ .

(2) Definition der Statistik

Wir betrachten die *Z-Konfidenzintervallstatistik*

$$Z := \frac{\sqrt{n}}{\sigma} (\bar{X}_n - \mu) \quad \text{mit} \quad \bar{X}_n := \frac{1}{n} \sum_{i=1}^n X_i. \quad (2)$$

(3) Analyse der Verteilung der Statistik

Für die *Z-Konfidenzintervallstatistik* gilt $Z \sim N(0, 1)$. Für einen Beweis verweisen wir auf den Appendix. Die *Z-Konfidenzintervallstatistik* ist eine Funktion der Stichprobe X_1, \dots, X_n (via \bar{X}_n) und μ , während ihre Verteilung also nicht von μ abhängt. Die *Z-Konfidenzintervallstatistik* ist damit ein Pivot. Wir erinnern daran, dass wir die WDF einer *Z-Zufallsvariable* mit ϕ , die KVF einer *Z-Zufallsvariable* mit Φ und die inverse KVF einer *Z-Zufallsvariable* mit Φ^{-1} bezeichnen.

(3) Analyse der Verteilung der Statistik (fortgeführt)

```
# Modellformulierung
mu      = 10
sigsqr  = 4
n       = 12
ns      = 1e4
res     = 1e3

# w.a.u. Erwartungswertparameter
# wahrer bekannter Varianzparameter
# Stichprobengroesse
# Anzahl Stichprobenrealisierungen
# Ausgangsraumaufloesung

# analytische Definitionen und Resultate
x_i     = seq(3,17,len = res)
x_bar   = seq(3,17,len = res)
z       = seq(-4,4,len = res)
p_x_i   = dnorm(x_i,mu,sqrt(sigsqr))
p_x_bar = dnorm(x_bar,mu,sqrt(sigsqr/n))
p_z     = dnorm(z,0,1)

# x_i Raum
# x_bar Raum
# z Raum
# x_i WDF
# x_bar WDF
# z WDF

# Simulation
X_i     = rep(NA,n,ns)
X_bar   = rep(NA,n,ns)
Z       = rep(NA,n,ns)

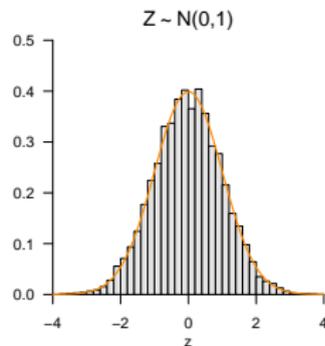
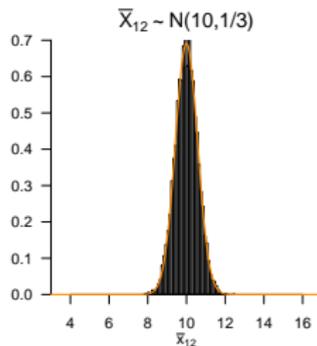
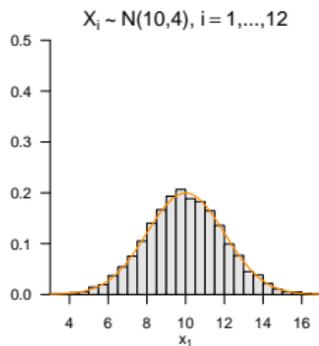
# X_i Array
# \bar{X}_{12} Array
# Z Array

for(s in 1:ns){
  X       = rnorm(n,mu,sqrt(sigsqr))
  X_i[s]  = X[1]
  X_bar[s] = mean(X)

  # Simulationsiterationen
  # Stichprobenrealisierung
  # X_i
  # Stichprobenmittelrealisierung

  # Z-Statistik Realisation
  Z[s]    = sqrt(n)*(X_bar[s] - mu)/sqrt(sigsqr)
}
}
```

(3) Analyse der Verteilung der Statistik (fortgeführt)

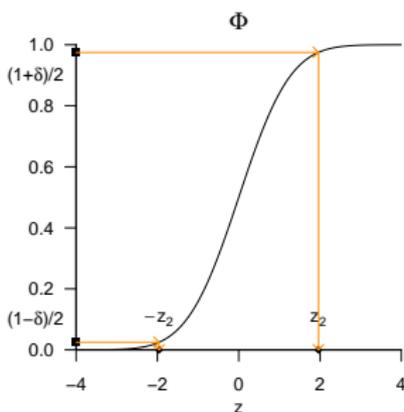
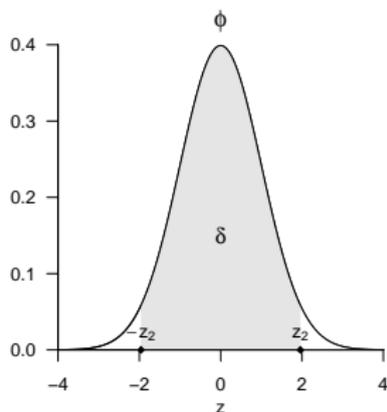


(4) Etablierung der Konfidenzbedingung

Für $\delta \in]0, 1[$, seien

$$z_1 := \Phi^{-1}\left(\frac{1-\delta}{2}\right) \text{ und } z_2 := \Phi^{-1}\left(\frac{1+\delta}{2}\right) \quad (3)$$

Es gilt dann $(1+\delta)/2 - (1-\delta)/2 = \delta$ und zum Beispiel gilt für $\delta = 0.95$, dass $z_1 = \Phi^{-1}(0.025) = -1.96$ und $z_2 = \Phi^{-1}(0.975) = 1.96$. Weiterhin gilt mit der Symmetrie von $N(0, 1)$, dass $z_1 = -z_2$. Es gilt hier also per Definition $\mathbb{P}(-z_2 \leq Z \leq z_2) = \delta$.



(4) Etablierung der Konfidenzbedingung (fortgeführt)

Mit der Definition von z_2 wie oben folgt dann aber

$$\begin{aligned}
 \delta &= \mathbb{P}(-z_2 \leq Z \leq z_2) \\
 &= \mathbb{P}\left(-z_2 \leq \frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mu) \leq z_2\right) \\
 &= \mathbb{P}\left(-\frac{\sigma}{\sqrt{n}}z_2 \leq \bar{X}_n - \mu \leq \frac{\sigma}{\sqrt{n}}z_2\right) \\
 &= \mathbb{P}\left(-\bar{X}_n - \frac{\sigma}{\sqrt{n}}z_2 \leq -\mu \leq -\bar{X}_n + \frac{\sigma}{\sqrt{n}}z_2\right) \\
 &= \mathbb{P}\left(\bar{X}_n + \frac{\sigma}{\sqrt{n}}z_2 \geq \mu \geq \bar{X}_n - \frac{\sigma}{\sqrt{n}}z_2\right) \\
 &= \mathbb{P}\left(\bar{X}_n - \frac{\sigma}{\sqrt{n}}z_2 \leq \mu \leq \bar{X}_n + \frac{\sigma}{\sqrt{n}}z_2\right) \\
 &= \mathbb{P}\left(\left[\bar{X}_n - \frac{\sigma}{\sqrt{n}}z_2, \bar{X}_n + \frac{\sigma}{\sqrt{n}}z_2\right] \ni \mu\right).
 \end{aligned} \tag{4}$$

Wir können nun das gesuchte Konfidenzintervall definieren.

(5) Definition des Konfidenzintervalls

Es sei $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ eine Stichprobe bei bekanntem Varianzparameter σ^2 und unbekanntem Erwartungswertparameter μ , es sei $\delta \in]0, 1[$ und es sei $z_\delta := \Phi^{-1}\left(\frac{1+\delta}{2}\right)$. Definiere

$$K_n := \left[\bar{X}_n - \frac{\sigma}{\sqrt{n}} z_\delta, \bar{X}_n + \frac{\sigma}{\sqrt{n}} z_\delta \right]. \quad (5)$$

Dann gilt wie oben gezeigt, dass

$$\mathbb{P}(K_n \ni \mu) = \delta. \quad (6)$$

Damit ist K_n ein δ -Konfidenzintervall für μ . Man beachte, dass K_n ein zufälliges Intervall ist, weil \bar{X}_n eine Zufallsvariable ist.

(5) Definition des Konfidenzintervalls (fortgeführt)

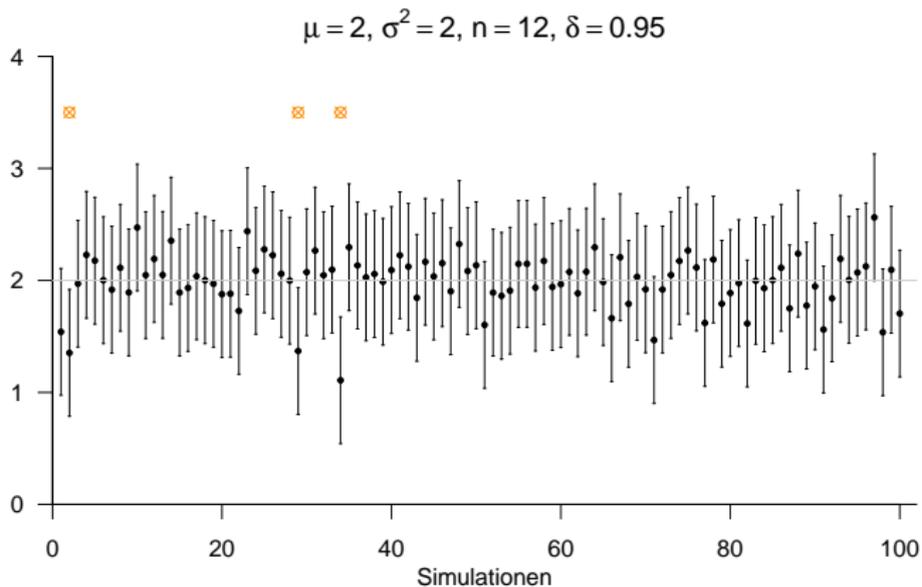
Simulation

```
# Modellformulierung
mu      = 2                                # w.a.u. Erwartungswertparameter
sigsqr  = 1                                # w.a.u. Varianzparameter
sigma   = sqrt(sigsqr)                     # w.a.u. St.Abweichungsparameter
n       = 12                                # Stichprobengroesse
delta   = 0.95                              # Konfidenzbedingung
phi_inv = qnorm((1+delta)/2)                # \Phi^{-1}((\delta + 1)/2)

# Simulation
ns      = 1e2                                # Anzahl Simulationen
X_bar   = rep(NaN,ns)                         # Stichprobenmittelarray
C       = matrix(rep(NaN,2*ns), ncol = 2)     # Konfidenzintervallarray
for(i in 1:ns){
  X      = rnorm(n,mu,sigma)                  # Stichprobenrealisierung
  X_bar[i] = mean(X)                          # Stichprobenmittel
  C[i,1]  = X_bar[i] - (sigma/sqrt(n))*phi_inv # untere KI Grenze
  C[i,2]  = X_bar[i] + (sigma/sqrt(n))*phi_inv # obere KI Grenze
}
```

(5) Definition des Konfidenzintervalls (fortgeführt)

Simulation



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) Erwartungswertparameter bei Normalverteilung mit bekannter Varianz
- (2) **Varianzparameter bei Normalverteilung**
- (3) Erwartungswertparameter bei Normalverteilung mit unbekannter Varianz

(1) Definition des statistischen Modells

Es sei $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ eine Stichprobe mit unbekanntem Varianzparameter $\sigma^2 > 0$ und bekanntem oder unbekanntem Erwartungswertparameter μ . Wir entwickeln ein δ -Konfidenzintervall für den Varianzparameter σ^2 .

(2) Definition der Statistik

Wir betrachten die U -Konfidenzintervallstatistik

$$U := \frac{n-1}{\sigma^2} S_n^2 \text{ mit } S_n^2 := \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2. \quad (7)$$

(3) Analyse der Verteilung der Statistik

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 X_1, \dots, X_n (via S_n^2) und σ^2 , während ihre Verteilung nicht von σ^2 abhängt. Die U -Konfidenzintervallstatistik ist also ein Pivot. Wir erinnern daran, dass wir WDF einer χ^2 -verteilten Zufallsvariable mit χ^2 , die KVF einer χ^2 -verteilten Zufallsvariable mit Ξ und die inverse KVF einer χ^2 -verteilten Zufallsvariable mit Ξ^{-1} bezeichnen.

(3) Analyse der Verteilung der Statistik (fortgeführt)

```
# Modellformulierung
mu      = 10
sigsqr  = 4
n       = 12
ns      = 1e4
res     = 1e3

# wahre Erwartungswertparameter
# wahre bekannter Varianzparameter
# Stichprobengroesse
# Anzahl Stichprobenrealisierungen
# Ausgangsraumaufloesung

# analytische Definitionen und Resultate
x_i     = seq(3,17,len = res)
x_bar   = seq(3,17,len = res)
u       = seq(0,30,len = res)
p_x_i   = dnorm(x_i,mu,sqrt(sigsqr))
p_x_bar = dnorm(x_bar,mu,sqrt(sigsqr/n))
p_u     = dchisq(u,n-1)

# x_1 Raum
# x_bar Raum
# normalisierte S_n^2 Raum
# x_1 WDF
# x_bar WDF
# u WDF

# Simulation
X_i     = rep(NA,n,ns)
X_bar   = rep(NA,n,ns)
S_sqr   = rep(NA,n,ns)
U       = rep(NA,n,ns)

# X_1 Array
# \bar{X} Array
# S^2 Array
# U Array

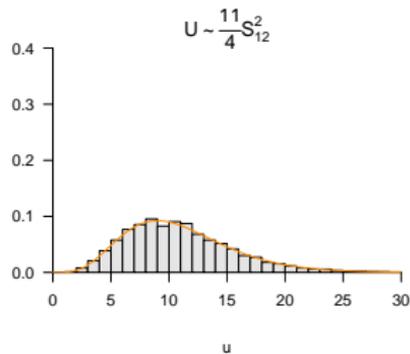
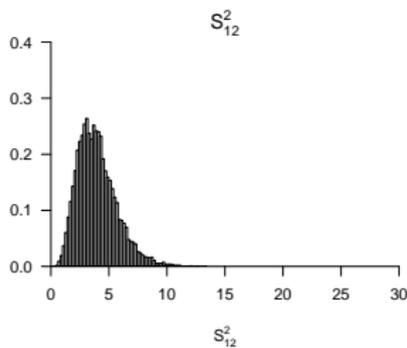
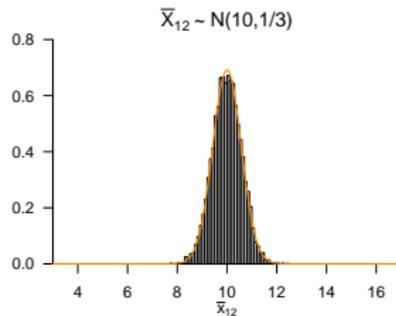
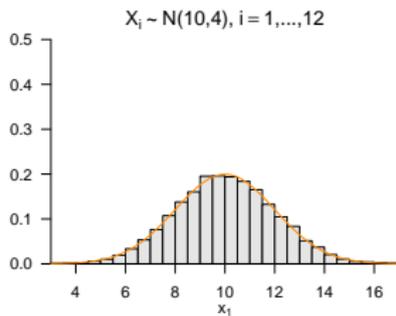
for(s in 1:ns){
  X      = rnorm(n,mu,sqrt(sigsqr))
  X_i[s] = X[1]
  X_bar[s] = mean(X)
  S_sqr[s] = var(X)

  # Simulationsiterationen
  # Stichprobenrealisierung
  # X_i
  # Stichprobenmittelrealisierung
  # Stichprobenvarianzrealisierung

  # U-Statistik Realisation
  U[s]   = ((n-1)/sigsqr)*S_sqr[s]
}

```

(3) Analyse der Verteilung der Statistik (fortgeführt)

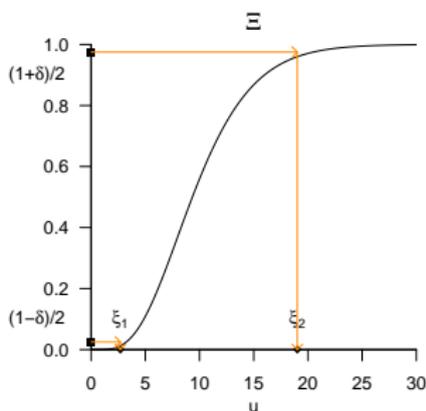
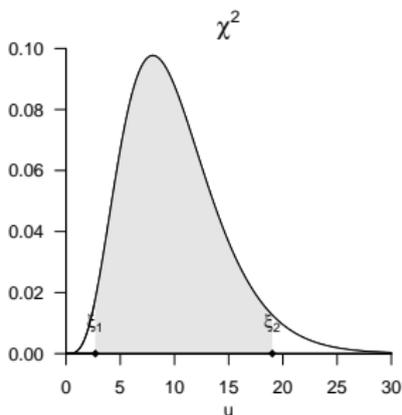


(4) Etablierung der Konfidenzbedingung

Für $\delta \in]0, 1[$ seien

$$\xi_1 := \Xi^{2^{-1}} \left(\frac{1 - \delta}{2}; n - 1 \right) \text{ und } \xi_2 := \Xi^{2^{-1}} \left(\frac{1 + \delta}{2}; n - 1 \right) \quad (8)$$

Es gilt dann $(1 + \delta)/2 - (1 - \delta)/2 = \delta$ gilt und zum Beispiel gilt für $n = 10$ und $\delta = 0.95$, $\xi_1 := \Xi^{2^{-1}}(0.025; 9) = 2.70$ und $\xi_2 := \Xi^{2^{-1}}(0.975; 9) = 19.0$. Es gilt hier also per Definition $\mathbb{P}(\xi_1 \leq U \leq \xi_2) = \delta$.



(4) Etablierung der Konfidenzbedingung (fortgeführt)

Mit der Definition von ξ_1 und ξ_2 wie oben folgt dann aber

$$\begin{aligned}
 \delta &= \mathbb{P}(\xi_1 \leq U \leq \xi_2) \\
 &= \mathbb{P}\left(\xi_1 \leq \frac{n-1}{\sigma^2} S_n^2 \leq \xi_2\right) \\
 &= \mathbb{P}\left(\xi_1^{-1} \geq \frac{\sigma^2}{(n-1)S_n^2} \geq \xi_2^{-1}\right) \\
 &= \mathbb{P}\left(\frac{(n-1)S_n^2}{\xi_1} \geq \sigma^2 \geq \frac{(n-1)S_n^2}{\xi_2}\right) \\
 &= \mathbb{P}\left(\frac{(n-1)S_n^2}{\xi_2} \leq \sigma^2 \leq \frac{(n-1)S_n^2}{\xi_1}\right) \\
 &= \mathbb{P}\left(\left[\frac{(n-1)S_n^2}{\xi_2}, \frac{(n-1)S_n^2}{\xi_1}\right] \ni \sigma^2\right).
 \end{aligned} \tag{9}$$

Wir können nun das gesuchte Konfidenzintervall definieren.

(5) Definition des Konfidenzintervalls

Es sei $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ eine Stichprobe mit unbekanntem Varianzparameter σ^2 und bekanntem oder unbekanntem Erwartungswertparameter μ , und es seien weiterhin $\delta \in]0, 1[$, $\xi_1 := \Xi^{2^{-1}}\left(\frac{1-\delta}{2}; n-1\right)$ und $\xi_2 := \Xi^{2^{-1}}\left(\frac{1+\delta}{2}; n-1\right)$. Definiere

$$K_n := \left[\frac{(n-1)S_n^2}{\xi_2}, \frac{(n-1)S_n^2}{\xi_1} \right]. \quad (10)$$

Dann gilt wie oben gezeigt

$$\mathbb{P}(K_n \ni \sigma^2) = \delta. \quad (11)$$

Damit ist K_n ein δ -Konfidenzintervall für σ^2 . Man beachte, dass K_n ein zufälliges Intervall ist, weil S_n^2 eine Zufallsvariable ist.

(5) Definition des Konfidenzintervalls (fortgeführt)

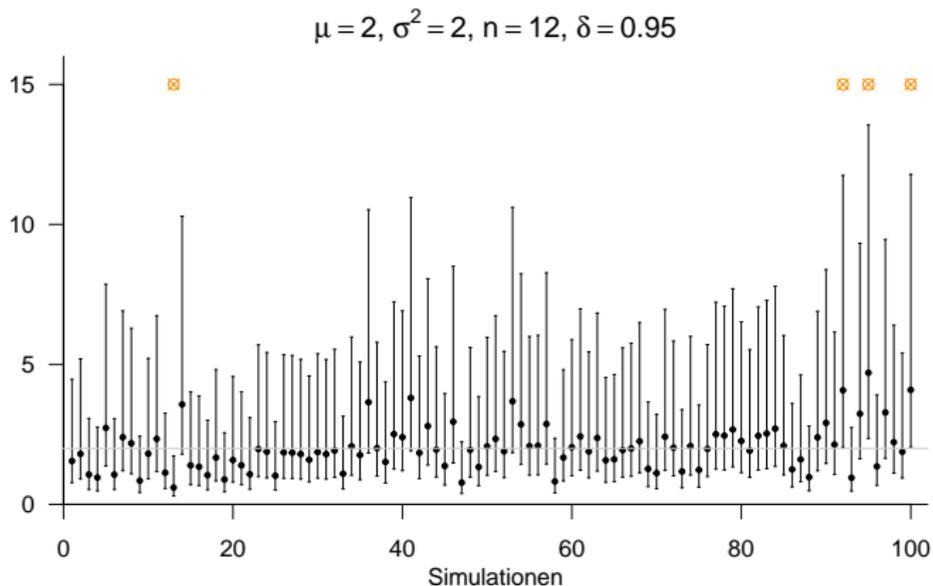
Simulation

```
# Modellformulierung
mu      = 2                # w.a.u. Erwartungswertparameter
sigsqr  = 2                # w.a.u. Varianzparameter
n       = 12              # Stichprobengroesse
delta   = 0.95            # Konfidenzbedingung
xi_1    = qchisq((1-delta)/2, n - 1) # \chi^2((1-\delta)/2; n - 1)
xi_2    = qchisq((1+delta)/2, n - 1) # \chi^2((1+\delta)/2; n - 1)

# Simulation
ns      = 1e2              # Anzahl Simulationen
X_bar   = rep(NaN,ns)      # Stichprobenmittelarray
S2      = rep(NaN,ns)      # Stichprobenvarianzarray
C       = matrix(rep(NaN,2*ns), ncol = 2) # Konfidenzintervallarray
for(i in 1:ns){           # Simulationsiterationen
  X      = rnorm(n,mu,sqrt(sigsqr)) # Stichprobenrealisierung
  S2[i]  = var(X)           # Stichprobenvarianz
  C[i,1] = (n-1)*S2[i]/xi_2 # untere KI Grenze
  C[i,2] = (n-1)*S2[i]/xi_1 # obere KI Grenze
}
```

(5) Definition des Konfidenzintervalls (fortgeführt)

Simulation



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) Erwartungswertparameter bei Normalverteilung mit bekannter Varianz
- (2) Varianzparameter bei Normalverteilung
- (3) **Erwartungswertparameter bei Normalverteilung mit unbekannter Varianz**

(1) Definition des statistischen Modells

Es sei $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ eine Stichprobe, wobei μ und $\sigma^2 > 0$ unbekannt seien. Wir entwickeln ein δ -Konfidenzintervall für den Erwartungswertparameter μ .

(2) Definition der Statistik

Wir betrachten die T -Konfidenzintervallstatistik

$$T := \frac{\sqrt{n}}{S_n}(\bar{X}_n - \mu) \text{ mit } \bar{X}_n := \frac{1}{n} \sum_{i=1}^n X_i \text{ und } S_n := \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2}. \quad (12)$$

(3) Analyse der Verteilung der Statistik

Für die T -Konfidenzintervallstatistik gilt $T \sim t(n-1)$, die T -Konfidenzintervallstatistik ist also eine t -verteilte Zufallsvariable mit Freiheitsgradparameter $n-1$. Für einen Beweis verweisen wir auf den Appendix. Die T -Konfidenzintervallstatistik ist eine Funktion der Stichprobe X_1, \dots, X_n (via \bar{X}_n und S_n), während ihre Verteilung also weder von μ noch von σ^2 abhängt. Die T -Konfidenzintervallstatistik ist also ein Pivot. Wir erinnern daran, dass wir die WDF einer t -verteilten Zufallsvariable mit t , die KVF einer t -verteilten Zufallsvariable mit ψ und die inverse KVF einer t -verteilten Zufallsvariable mit ψ^{-1} bezeichnen.

(3) Analyse der Verteilung der Statistik (fortgeführt)

```

# Modellformulierung
mu      = 10                                # w. a. u. Erwartungswertparameter
sigsqr  = 4                                # wahrer bekannter Varianzparameter
n       = 12                                # Stichprobengroesse
ns      = 1e4                                # Anzahl Stichprobenrealisierungen
res     = 1e3                                # Ausgangsraumaufloesung

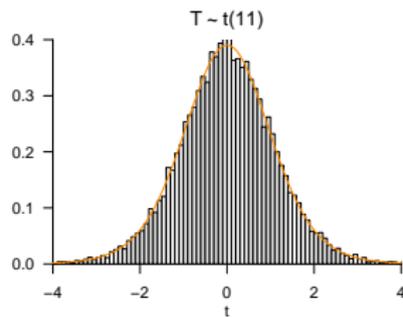
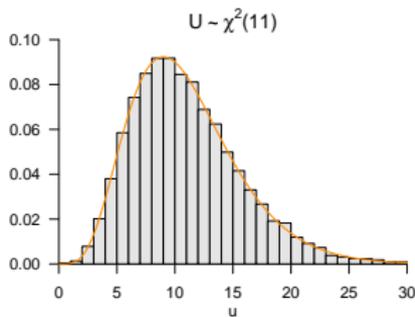
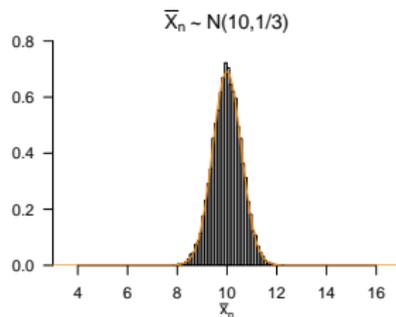
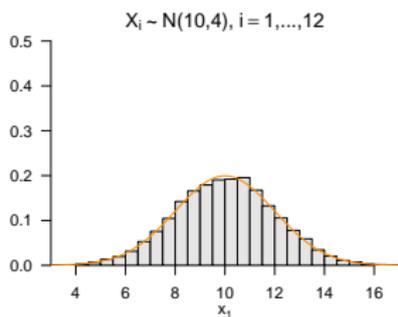
# analytische Definitionen und Resultate
x_i     = seq(3,17,len = res)               # x_1 Raum
x_bar   = seq(3,17,len = res)               # x_bar Raum
u       = seq(0,30,len = res)               # normalisierte s_{12}^2 Raum
t       = seq(-4,4,len = res)               # t Raum
p_x_i   = dnorm(x_i,mu,sqrt(sigsqr))        # x_1 WDF
p_x_bar = dnorm(x_bar,mu,sqrt(sigsqr/n))    # x_bar WDF
p_u     = dchisq(u,n-1)                     # u WDF
p_t     = dt(t,n-1)                         # t WDF

# Simulation
X_i     = rep(NaN,ns)                       # X_1 Array
X_bar   = rep(NaN,ns)                       # \bar{X} Array
U       = rep(NaN,ns)                       # U Array
Tee     = rep(NaN,ns)                       # T Array
for(s in 1:ns){
  X      = rnorm(n,mu,sqrt(sigsqr))         # Simulationsiterationen
  X_i[s] = X[1]                             # Stichprobenrealisierung
  X_bar[s] = mean(X)                         # X_i
  U[s]    = (n-1)/(sigsqr)*var(X)           # Stichprobenmittelrealisierung
  Tee[s]  = sqrt(n)*((X_bar[s] - mu)/sqrt(var(X))) # U Statistik Realisierung

# T-Statistik Realisierung
Tee[s]   = sqrt(n)*((X_bar[s] - mu)/sqrt(var(X)))
}

```

(3) Analyse der Verteilung der Statistik (fortgeführt)

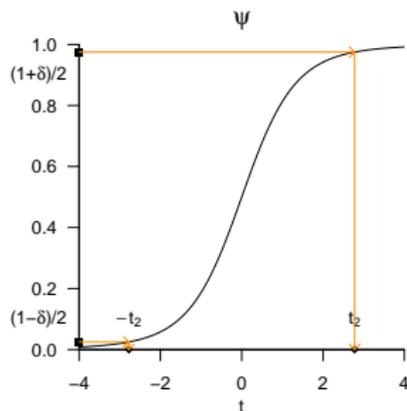
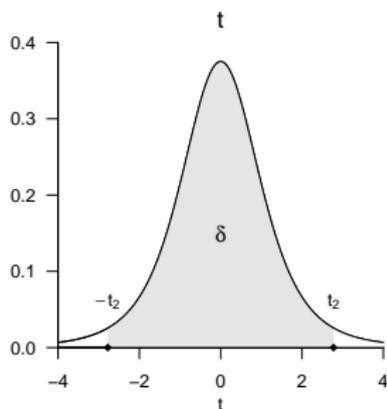


(4) Etablierung der Konfidenzbedingung

Für $\delta \in]0, 1[$ seien

$$t_1 := \psi^{-1}\left(\frac{1-\delta}{2}; n-1\right) \text{ und } t_2 := \psi^{-1}\left(\frac{1+\delta}{2}; n-1\right) \quad (13)$$

Es gilt dann $(1+\delta)/2 - (1-\delta)/2 = \delta$ und zum Beispiel gilt für $n = 5$ und $\delta = 0.95$, $t_1 = \psi^{-1}(0.025; 4) = -2.57$ und $t_2 = \psi^{-1}(0.975; 4) = 2.57$. Weiterhin gilt mit der Symmetrie von $t(n-1)$, $t_1 = -t_2$. Es gilt hier also per Definition $\mathbb{P}(-t_2 \leq T \leq t_2) = \delta$.



(4) Etablierung der Konfidenzbedingung (fortgeführt)

Mit der definition von t_2 wie oben folgt dann aber

$$\begin{aligned}
 \delta &= \mathbb{P}(-t_2 \leq T \leq t_2) \\
 &= \mathbb{P}\left(-t_2 \leq \frac{\sqrt{n}}{S_n}(\bar{X}_n - \mu) \leq t_2\right) \\
 &= \mathbb{P}\left(-\frac{S_n}{\sqrt{n}}t_2 \leq \bar{X}_n - \mu \leq \frac{S_n}{\sqrt{n}}t_2\right) \\
 &= \mathbb{P}\left(-\bar{X}_n - \frac{S_n}{\sqrt{n}}t_2 \leq -\mu \leq -\bar{X}_n + \frac{S_n}{\sqrt{n}}t_2\right) \\
 &= \mathbb{P}\left(\bar{X}_n + \frac{S_n}{\sqrt{n}}t_2 \geq \mu \geq \bar{X}_n - \frac{S_n}{\sqrt{n}}t_2\right) \\
 &= \mathbb{P}\left(\bar{X}_n - \frac{S_n}{\sqrt{n}}t_2 \leq \mu \leq \bar{X}_n + \frac{S_n}{\sqrt{n}}t_2\right) \\
 &= \mathbb{P}\left[\left[\bar{X}_n - \frac{S_n}{\sqrt{n}}t_2, \bar{X}_n + \frac{S_n}{\sqrt{n}}t_2\right] \ni \mu\right].
 \end{aligned} \tag{14}$$

Wir können nun das gesuchte Konfidenzintervall definieren.

(5) Definition des Konfidenzintervalls

Es sei $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ mit unbekanntem Parameter μ und σ^2 , es sei $\delta \in]0, 1[$, und es sei $t_\delta := \psi^{-1}\left(\frac{1+\delta}{2}; n-1\right)$. Definiere

$$K_n := \left[\bar{X}_n - \frac{S_n}{\sqrt{n}} t_\delta, \bar{X}_n + \frac{S_n}{\sqrt{n}} t_\delta \right]. \quad (15)$$

Dann gilt wie oben gezeigt

$$\mathbb{P}(K_n \ni \mu) = \delta. \quad (16)$$

Damit ist K_n ein δ -Konfidenzintervall für μ . Man beachte, dass K_n ein zufälliges Intervall ist, weil \bar{X}_n und S_n Zufallsvariablen sind.

(5) Definition des Konfidenzintervalls (fortgeführt)

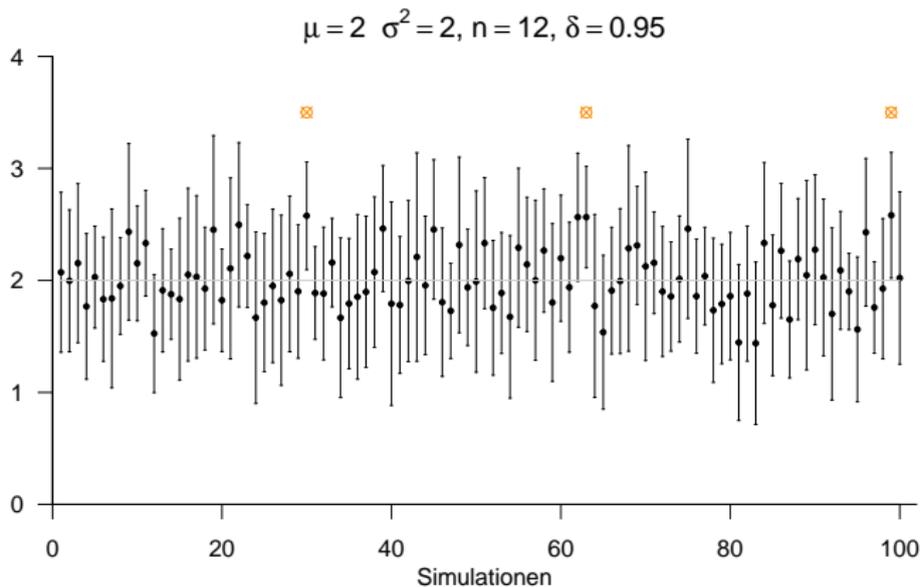
Simulation

```
# Modellformulierung
mu      = 2                                # w.a.u. Erwartungswertparameter
sigsqr  = 1                                # w.a.u. Varianzparameter
sigma   = sqrt(sigsqr)                     # w.a.u. St.Abweichungsparameter
n       = 12                               # Stichprobengroesse
delta   = 0.95                             # Konfidenzbedingung
psi_inv = qt((1+delta)/2,n-1)              # \psi^{-1}((\delta + 1)/2, n-1)

# Simulation
ns      = 1e2                              # Anzahl Simulationen
X_bar   = rep(NaN,ns)                       # Stichprobenmittelarray
S       = rep(NaN,ns)                       # St.Abweichungsarray
C       = matrix(rep(NaN,2*ns), ncol = 2)    # Konfidenzintervallarray
for(i in 1:ns){
  X      = rnorm(n,mu,sigma)                 # Stichprobenrealisierung
  X_bar[i] = mean(X)                         # Stichprobenmittel
  S[i]   = sd(X)                             # Stichprobenstandardabweichung
  C[i,1] = X_bar[i] - (S[i]/sqrt(n))*psi_inv # untere KI Grenze
  C[i,2] = X_bar[i] + (S[i]/sqrt(n))*psi_inv # obere KI Grenze
}
```

(5) Definition des Konfidenzintervalls (fortgeführt)

Simulation



(5) Definition des Konfidenzintervalls (fortgeführt)

Simulation der zweiten Interpretation eines Konfidenzintervalls

```
# Anzahl Simulationen mit \theta_1, \theta_2, ...
ns      = 1e2                                # Anzahl Simulationen

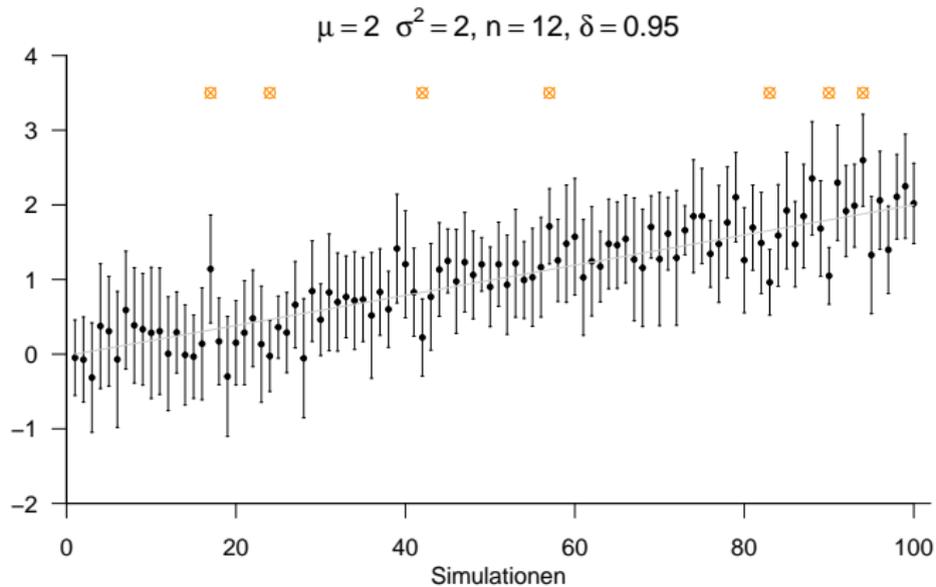
# Modellformulierung
mu      = 2 * seq(0,1,len = ns)              # w.a.u. Erwartungswertparameter
sigsqr  = 1                                  # w.a.u. Varianzparameter
sigma   = sqrt(sigsqr)                       # w.a.u. St. Abweichungsparameter
n       = 12                                 # Stichprobengroesse
delta   = 0.95                               # Konfidenzbedingung
psi_inv = qt((1+delta)/2,n-1)                # \psi^{-1}((\delta + 1)/2, n-1)

# Simulation

X_bar   = rep(NaN,ns)                         # Stichprobenmittelarray
S       = rep(NaN,ns)                         # St. Abweichungsarray
C       = matrix(rep(NaN,2*ns), ncol = 2)     # Konfidenzintervallarray
for(i in 1:ns){
  X      = rnorm(n,mu[i],sigma)                # Stichprobenrealisierung
  X_bar[i] = mean(X)                          # Stichprobenmittel
  S[i]   = sd(X)                              # Stichprobenstandardabweichung
  C[i,1] = X_bar[i] - (S[i]/sqrt(n))*psi_inv  # untere KI Grenze
  C[i,2] = X_bar[i] + (S[i]/sqrt(n))*psi_inv  # obere KI Grenze
}
```

(5) Definition des Konfidenzintervalls (fortgeführt)

Simulation der zweiten Interpretation eines Konfidenzintervalls



Definition

Beispiele

Selbstkontrollfragen

Appendix

Definition

Beispiele

Selbstkontrollfragen

Appendix

Selbstkontrollfragen

1. Definieren Sie den Begriff des δ -Konfidenzintervalls (δ -KIs).
2. Geben Sie zwei Interpretationen eines δ -KIs.
3. Erläutern Sie die typischen Schritte zur Konstruktion eines δ -KIs.
4. Definieren Sie die Z -Konfidenzintervallstatistik und geben Sie ihre Verteilung an.
5. Geben Sie das δ -KI für den Erwartungswert einer Normalverteilung bei bekannter Varianz an.
6. Definieren Sie die U -Statistik und geben Sie ihre Verteilung an.
7. Geben Sie das δ -KI für den Varianzparameter einer Normalverteilung an.
8. Definieren Sie die T -Konfidenzintervallstatistik und geben Sie ihre Verteilung an.
9. Geben Sie das δ -KI für den Erwartungswert einer Normalverteilung bei unbekannter Varianz an.

Definition

Beispiele

Selbstkontrollfragen

Appendix

Appendix

Beweis der Verteilung der Z -Konfidenzintervallstatistik

Wir halten zunächst fest, dass mit der Mittelwerttransformation für unabhängig und identisch normalverteilte Zufallsvariablen gilt, dass $\bar{X}_n \sim N(\mu, \frac{\sigma^2}{n})$ (cf. (7) Transformationen der Normalverteilungen). Die Verteilung der Z -Konfidenzintervallsstatistik folgt dann mit dem WDF Transformationstheorem bei linear-affinen Abbildungen (cf. *ibid.*) durch Transformation von \bar{X}_n mit der Funktion

$$f : \mathbb{R} \rightarrow \mathbb{R}, \bar{x}_n \mapsto f(\bar{x}_n) := \frac{\sqrt{n}}{\sigma} \bar{x}_n - \frac{\sqrt{n}}{\sigma} \mu \quad (17)$$

Für die WDF von Z gilt also

$$\begin{aligned} p_Z(z) &= \frac{\sigma}{\sqrt{n}} N\left(\frac{\sigma}{\sqrt{n}}\left(z + \frac{\sqrt{n}}{\sigma}\mu\right); \mu, \frac{\sigma^2}{n}\right) \\ &= \frac{\sigma}{\sqrt{n}} \frac{1}{\sqrt{2\pi \frac{\sigma^2}{n}}} \exp\left(-\frac{1}{2 \frac{\sigma^2}{n}} \left(\frac{\sigma}{\sqrt{n}}\left(z + \frac{\sqrt{n}}{\sigma}\mu\right) - \mu\right)^2\right) \\ &= \frac{\sigma}{\sqrt{n}} \frac{\sqrt{n}}{\sigma} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} \frac{n}{\sigma^2} \left(\frac{\sigma}{\sqrt{n}}z + \mu - \mu\right)^2\right) \\ &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} z^2\right) \\ &= N(z; 0, 1) \end{aligned} \quad (18)$$

□

Appendix

Beweis der Verteilung der T -Konfidenzintervallstatistik

Wir halten zunächst fest, dass

$$T = \sqrt{n} \left(\frac{\bar{X}_n - \mu}{S_n} \right) = \frac{\bar{X}_n - \mu}{S_n / \sqrt{n}} = \frac{\sigma \sqrt{n} (\bar{X}_n - \mu)}{S_n} = \frac{\frac{\sqrt{n}}{\sigma} (\bar{X}_n - \mu)}{S_n / \sigma} = \frac{\frac{\sqrt{n}}{\sigma} (\bar{X}_n - \mu)}{\sqrt{\frac{1}{\sigma^2} S_n^2}} \quad (19)$$

An der rechten Seite obiger Gleichung sehen wir also, dass die T -Konfidenzintervallstatistik als Quotient der Zufallsvariablen

$$\frac{\sqrt{n}}{\sigma} (\bar{X}_n - \mu) \text{ und } \sqrt{\frac{1}{\sigma^2} S_n^2} \quad (20)$$

geschrieben werden kann. Dabei hat die Zufallsvariable im Zähler die Form der Z -Konfidenzintervallstatistik, kann also als

$$\frac{\sqrt{n}}{\sigma} (\bar{X}_n - \mu) =: Z \quad (21)$$

geschrieben werden. Die Zufallsvariable im Nenner dagegen hat die Form der Quadratwurzel einer mit $n - 1$ multiplizierten U -Konfidenzintervallstatistik, kann also als

$$\sqrt{\frac{1}{\sigma^2} S_n^2} = \sqrt{\frac{1}{n-1} \frac{n-1}{\sigma^2} S_n^2} := \sqrt{\frac{1}{n-1} U} \quad (22)$$

geschrieben werden.

Beweis der Verteilung der T -Konfidenzintervallstatistik (fortgeführt)

Insgesamt kann die T Konfidenzintervallstatistik also als

$$T = \frac{Z}{\sqrt{\frac{U}{n-1}}} \text{ mit } Z \sim N(0, 1) \text{ und } U \sim \chi^2(n-1) \quad (23)$$

geschrieben werden. Casella and Berger (2012) zeigen in Abschnitt 5.3, dass die Zufallsvariablen im Zähler und Nenner dieser Formulierung der T -Konfidenzintervallstatistik unabhängig sind.

Es ergibt sich also, dass die T -Konfidenzintervallstatistik als Quotient zweier unabhängiger Zufallsvariablen geschrieben werden kann, wobei die Zählerzufallsvariable eine standardnormalverteilte Zufallsvariable ist und die Nennerzufallsvariable die Quadratwurzel einer durch die Anzahl ihrer Freiheitsgrade geteilten χ^2 -verteilten Zufallsvariable ist.

Dann aber folgt mit dem T -Transformationstheorem (cf. (8) Transformationen der Normalverteilung), dass die T -Konfidenzintervallstatistik eine t -verteilte Zufallsvariable mit Freiheitsgradparameter entsprechend dem Freiheitsgradparameter der χ^2 -verteilten Zufallsvariable, hier also $n - 1$, ist. Es folgt demnach $T \sim t(n - 1)$.

□

References

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



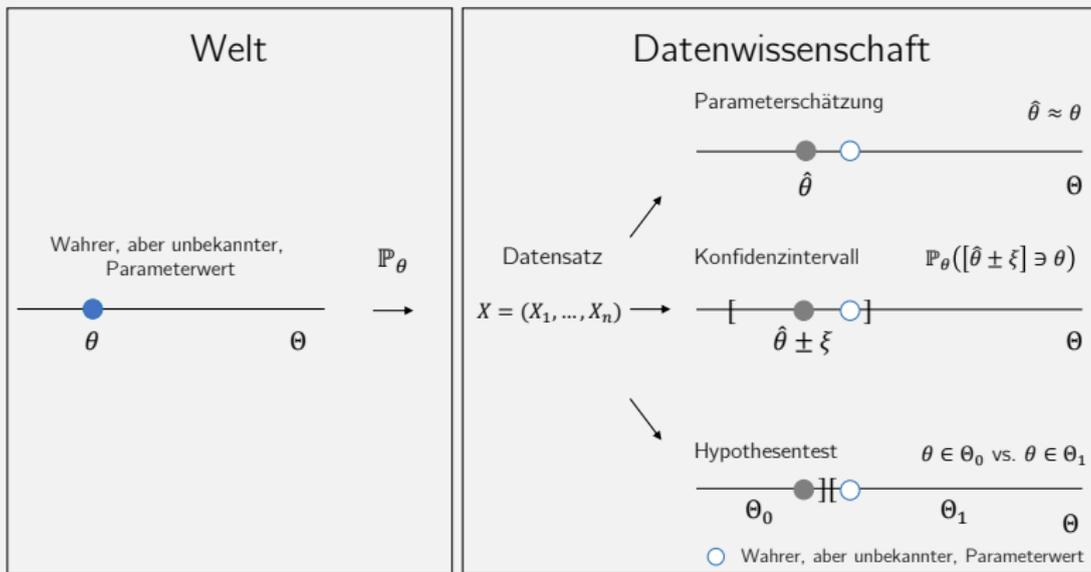
Wahrscheinlichkeitstheorie und Frequentistische Inferenz

BSc Psychologie WiSe 2021/22

Prof. Dr. Dirk Ostwald

(12) Hypothesentests

Modell und Standardprobleme der Frequentistischen Inferenz



(1) Parameterschätzung

Ziel der Parameterschätzung ist es, einen möglichst guten Tipp für den wahren, aber unbekanntem, Parameterwert (oder eine Funktion dessen) abzugeben, typischerweise basierend auf der Beobachtung einer Realisierung von $X_1, \dots, X_n \sim p_\theta$.

(2) Konfidenzintervalle

Das Ziel der Bestimmung von Konfidenzintervallen ist es, basierend auf der Verteilung möglicher Parameterschätzwerte eine quantitative Aussage über die mit dem Schätzwert assoziierte Unsicherheit zu treffen.

(3) Hypothesentests

Das Ziel der Auswertung von Hypothesentests ist es, basierend auf der angenommenen Verteilung der Beobachtungen X_1, \dots, X_n in einer möglichst sinnvollen Form zu entscheiden, ob der wahre, aber unbekanntem Parameterwert, in einer von zwei sich gegenseitig ausschließenden Untermengen des Parameterraumes, welche man als Hypothesen bezeichnet, liegt.

Standardannahmen Frequentistischer Inferenz

\mathcal{M} sei ein statistisches Modell mit $X_1, \dots, X_n \sim \mathbb{P}_\theta$. Es wird angenommen, dass ein vorliegender Datensatz eine der möglichen Realisierungen von $X_1, \dots, X_n \sim \mathbb{P}_\theta$ ist. Aus frequentistischer Sicht kann man die Erhebung von Datensätzen unendlich oft wiederholen und zu jedem Datensatz Statistiken auswerten.

$$\text{Datensatz (1)} : x^{(1)} = \left(x_1^{(1)}, x_2^{(1)}, \dots, x_n^{(1)} \right), \text{ Statistik (1): } S : \mathbb{R}^n \rightarrow \Sigma, x^{(1)} \mapsto S \left(x^{(1)} \right)$$

$$\text{Datensatz (2)} : x^{(2)} = \left(x_1^{(2)}, x_2^{(2)}, \dots, x_n^{(2)} \right), \text{ Statistik (2): } S : \mathbb{R}^n \rightarrow \Sigma, x^{(2)} \mapsto S \left(x^{(2)} \right)$$

$$\text{Datensatz (3)} : x^{(3)} = \left(x_1^{(3)}, x_2^{(3)}, \dots, x_n^{(3)} \right), \text{ Statistik (3): } S : \mathbb{R}^n \rightarrow \Sigma, x^{(3)} \mapsto S \left(x^{(3)} \right)$$

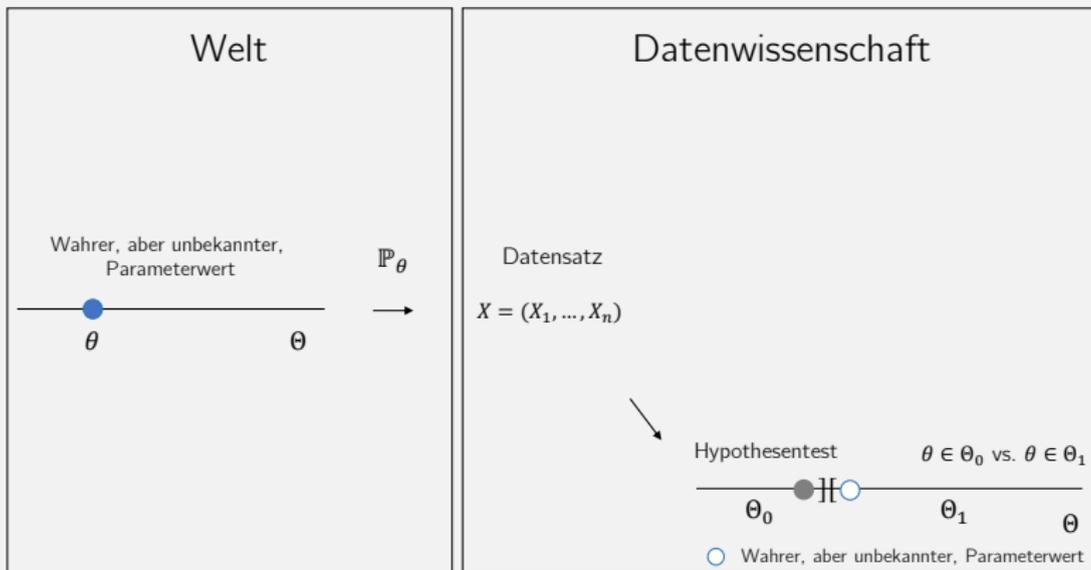
$$\text{Datensatz (4)} : x^{(4)} = \left(x_1^{(4)}, x_2^{(4)}, \dots, x_n^{(4)} \right), \text{ Statistik (4): } S : \mathbb{R}^n \rightarrow \Sigma, x^{(4)} \mapsto S \left(x^{(4)} \right)$$

...

Um die Qualität statistischer Methoden zu beurteilen betrachtet die frequentistische Statistik deshalb die Wahrscheinlichkeitsverteilungen von Statistiken und Schätzern unter der Annahme von $X_1, \dots, X_n \sim p_\theta$.

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 realen Normalfall nur eines vorliegenden Datensatzes, kann sie auch "schlecht" sein.

Modell und Standardprobleme der Frequentistischen Inferenz



Vorbemerkungen

Grundlegende Definitionen

Z-Test

p-Werte

Konfidenzintervalle und Hypothesentests

Selbstkontrollfragen

Vorbemerkungen

Grundlegende Definitionen

Z-Test

p-Werte

Konfidenzintervalle und Hypothesentests

Selbstkontrollfragen

Vorbemerkungen

Grundlegende Logik statistischer Hypothesentests

Man hat einen Datensatz x_1, \dots, x_n vorliegen und nimmt an, dass es sich dabei um die Realisation einer Stichprobe handelt, zum Beispiel von $X_1, \dots, X_n \sim N(\mu, \sigma^2)$.

Man berechnet basierend auf dem Datensatz eine *Teststatistik*, zum Beispiel das anhand der Stichprobenvarianz und der Stichprobengröße normalisierte Stichprobenmittel $\sqrt{n}\bar{x}_n/s_n$.

Man fragt sich, wie wahrscheinlich es wäre, den beobachteten oder einen extremeren Wert der Teststatistik unter der Annahme eines *Nullmodells* zu observieren. Dabei meint man mit *Nullmodell* intuitiv ein Wahrscheinlichkeitsverteilungsmodell bei dem kein "interessanter Effekt" vorliegt, also zum Beispiel $\mu = 0$ gilt. Die Wahrscheinlichkeit ist wie immer frequentistisch zu verstehen, d.h. als idealisierte relative Häufigkeit, wenn man viele Stichprobenrealisationen des Nullmodells generieren würde.

Ist die betrachtete Wahrscheinlichkeit dafür, den beobachteten oder einen extremeren Wert der Teststatistik unter Annahme des Nullmodells zu observieren groß, so sagt man sich "Nunja, dann ist es wohl ganz plausibel, dass das Nullmodell die Daten generiert hat." Im Wissenschaftsjargon spricht man von einem "nicht-signifikanten Ergebnis."

Ist die betrachtete Wahrscheinlichkeit dafür, den beobachteten oder einen extremeren Wert der Teststatistik unter Annahme des Nullmodells zu observieren dagegen klein, so sagt man sich "Aha, dann ist es wohl nicht so plausibel, dass das Nullmodell die Daten generiert hat." Im Wissenschaftsjargon spricht man von einem "signifikanten Ergebnis."

Wie immer in der frequentistischen Statistik weiß man nach Durchführung dieser Prozedur nicht, ob im vorliegenden Fall nun wirklich das Nullmodell oder ein anderes Modell die Daten generiert hat, sondern man weiß nur, wie oft man bei dieser Prozedur im Mittel richtig oder falsch liegen würde, wenn alle Annahmen zuträfen und man diese Prozedur sehr oft wiederholen würde.

Vorbemerkungen

Grundlegende Definitionen

Z-Test

p-Werte

Konfidenzintervalle und Hypothesentests

Selbstkontrollfragen

Definition (Statistische Hypothesen und Testszenario)

$X_1, \dots, X_n \sim p_\theta$ sei eine Stichprobe mit WMF oder WDF p_θ , \mathcal{X} sei der Ergebnisraum des Zufallsvektors $X := (X_1, \dots, X_n)$, und Θ sei der Parameterraum des zugrundeliegenden statistischen Modells. Weiterhin sei $\{\Theta_0, \Theta_1\}$ eine Partition des Parameterraumes, so dass $\Theta = \Theta_0 \cup \Theta_1$ und $\Theta_0 \cap \Theta_1 = \emptyset$ gelten. Eine *statistische Hypothese* ist dann eine Aussage über den wahren, aber unbekanntem, Parameterwert θ in Hinblick auf die Untermengen Θ_0 und Θ_1 des Parameterraums. Speziell werden die Aussagen

- $\theta \in \Theta_0$ als *Nullhypothese* 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 X genau festgelegt.
- Bei einer zusammengesetzten Hypothese ist nur die Verteilungsklasse von X festgelegt.

Definition (Einseitige und zweiseitige Hypothesen)

$\Theta := \mathbb{R}$ sei ein eindimensionaler Parameterraum und θ_0 sei ein Element von Θ . Dann werden zusammengesetzte Nullhypothesen der Form

$$\Theta_0 :=] - \infty, \theta_0] \text{ oder } \Theta_0 := [\theta_0, \infty[\quad (1)$$

einseitige Nullhypothesen genannt und auch in der Form

$$H_0 : \theta \leq \theta_0 \text{ oder } H_0 : \theta \geq \theta_0 \quad (2)$$

geschrieben. Die entsprechenden Alternativhypothesen haben dabei die Form

$$\Theta_1 :=]\theta_0, \infty[\text{ oder } \Theta_1 :=] - \infty, \theta_0[\text{ bzw. } H_1 : \theta > \theta_0 \text{ oder } H_1 : \theta < \theta_0. \quad (3)$$

Bei einer einfachen Nullhypothese der Form

$$\Theta_0 := \{\theta_0\} \text{ bzw. } H_0 : \theta = \theta_0 \quad (4)$$

wird die Alternativhypothese

$$\Theta_1 := \Theta \setminus \{\theta_0\} \text{ bzw. } H_1 : \theta \neq \theta_0 \Leftrightarrow \Theta_1 :=] - \infty, \theta_0[\cup]\theta_0, \infty[\quad (5)$$

zweiseitige Alternativhypothese genannt.

Definition (Test)

In einem Testszenario ist ein *Test* ϕ eine Abbildung aus dem Ergebnisraum \mathcal{X} nach $\{0, 1\}$,

$$\phi : \mathcal{X} \rightarrow \{0, 1\}, x \mapsto \phi(x). \quad (6)$$

Dabei repräsentiert

- $\phi(x) = 0$ den Vorgang des Nichtablehnens der Nullhypothese.
- $\phi(x) = 1$ den Vorgang des Ablehnens der Nullhypothese.

Bemerkung

- Weil die Stichprobe X ein Zufallsvektor ist, ist ein Test ϕ eine Zufallsvariable.

Definition (Standardtest)

Ein *Standardtest* ist definiert durch die Verkettung einer *Teststatistik*

$$\gamma : \mathcal{X} \rightarrow \mathbb{R} \quad (7)$$

und einer *Entscheidungsregel*

$$\delta : \mathbb{R} \rightarrow \{0, 1\}. \quad (8)$$

Ein Standardtest kann also geschrieben werden als

$$\phi := \delta \circ \gamma : \mathcal{X} \rightarrow \{0, 1\}. \quad (9)$$

Bemerkungen

- Weil die Stichprobe X ein Zufallsvektor ist, sind sowohl γ als auch δ Zufallsvariablen.
- Wir betrachten in der Folge nur Standardtests.

Definition (Kritischer Bereich)

Die Untermenge K des Ergebnisraums des Zufallsvektors $X := (X_1, \dots, X_n)$, für die ein Test den Wert 1 annimmt, heißt *kritischer Bereich* des Tests,

$$K := \{x \in \mathcal{X} \mid \phi(x) = 1\} \subset \mathcal{X}. \quad (10)$$

Bemerkungen

- Die Ereignisse $\phi(X) = 1$ und $X \in K$ sind äquivalent.
- Die Ereignisse $\phi(X) = 1$ und $X \in K$ haben die gleiche Wahrscheinlichkeit.

Definition (Ablehnungsbereich)

Die Untermenge A des Ergebnisraums einer Teststatistik, für die der Test den Wert 1 annimmt, heißt *Ablehnungsbereich* des Tests,

$$A := \{\gamma(x) \in \mathbb{R} \mid \phi(x) = 1\} \subset \mathbb{R}. \quad (11)$$

Bemerkungen

- Die Ereignisse $\phi(X) = 1$ und $\gamma(X) \in A$ sind äquivalent.
- Die Ereignisse $\phi(X) = 1$ und $\gamma(X) \in A$ haben die gleiche Wahrscheinlichkeit.

Definition (Kritischer Wert-basierte Tests)

Ein *kritischer Wert-basierter Test* ist ein Standardtest, bei dem die Entscheidungsregel δ von einem kritischen Wert $k \in \mathbb{R}$ abhängt. Speziell ist

- ein *einseitiger kritischer Wert-basierter Test* von der Form

$$\phi : \mathcal{X} \rightarrow \{0, 1\}, x \mapsto \phi(x) := 1_{\{\gamma(x) \geq k\}} = \begin{cases} 1 & \gamma(x) \geq k \\ 0 & \gamma(x) < k \end{cases} \quad (12)$$

- ein *zweiseitiger kritischer Wert-basierter Test* von der Form

$$\phi : \mathcal{X} \rightarrow \{0, 1\}, x \mapsto \phi(x) := 1_{\{|\gamma(x)| \geq k\}} = \begin{cases} 1 & |\gamma(x)| \geq k \\ 0 & |\gamma(x)| < k \end{cases} \quad (13)$$

Bemerkung

- Wir betrachten in der Folge nur kritischer Wert-basierte Tests.

Definition (Richtige Testentscheidungen und Testfehler)

Das Nichtablehnen der Nullhypothese, wenn die Nullhypothese zutrifft, sowie das Ablehnen der Nullhypothese, wenn die Nullhypothese nicht zutrifft, werden *richtige Testentscheidungen* genannt. Es können weiterhin zwei Arten von Testfehlern auftreten: das Ablehnen der Nullhypothese, wenn die Nullhypothese zutrifft, heißt *Typ I Fehler*, das Nichtablehnen der Nullhypothese, wenn die Alternativhypothese zutrifft, heißt *Typ II Fehler*.

Die untenstehende Graphik gibt eine Übersicht.

		Testentscheidung	
		$\phi(X) = 0$	$\phi(X) = 1$
Wahrer Parameterwert	$\theta \in \theta_0$	Richtige Entscheidung	Typ I Fehler
	$\theta \in \theta_1$	Typ II Fehler	Richtige Entscheidung

Definition (Testgütefunktion)

Für einen Test ϕ ist die *Testgütefunktion* definiert als

$$q_\phi : \Theta \rightarrow [0, 1], \theta \mapsto q_\phi(\theta) := \mathbb{P}_\theta(\phi = 1). \quad (14)$$

Für $\theta \in \Theta_1$ heißt q_ϕ auch *Powerfunktion* oder *Trennschärfefunktion*.

Bemerkungen

- \mathbb{P}_θ bezeichnet die Verteilung von ϕ unter der Annahme $X_1, \dots, X_n \sim p_\theta$.
- Es gilt $\mathbb{P}_\theta(\phi = 1) = \mathbb{P}_\theta(X \in K) = \mathbb{P}_\theta(\gamma \in A)$
- Für jedes $\theta \in \Theta$ liefert q_ϕ die Wahrscheinlichkeit, dass H_0 durch ϕ abgelehnt wird.
- Bei Poweranalysen betrachtet man q_ϕ als Funktion aller Testszenario und Testparameter.
- Ändert sich ϕ , z.B. weil sich der kritische Wert von ϕ ändert, dann ändert sich $q_\phi(\theta)$.
- Im Idealfall hätte man einen Test ϕ mit

$$q_\phi(\theta) = \mathbb{P}_\theta(\phi = 1) = 0 \text{ für } \theta \in \Theta_0 \text{ und } q_\phi(\theta) = \mathbb{P}_\theta(\phi = 1) = 1 \text{ für } \theta \in \Theta_1. \quad (15)$$

- Die Testentscheidung eines solchen ϕ wäre mit Wahrscheinlichkeit 1 richtig.

Intuition zur Testkonstruktion

Im Idealfall hätte man einen Test ϕ mit

$$q_\phi(\theta) = \mathbb{P}_\theta(\phi = 1) = 0 \text{ für } \theta \in \Theta_0 \text{ und } q_\phi(\theta) = \mathbb{P}_\theta(\phi = 1) = 1 \text{ für } \theta \in \Theta_1. \quad (16)$$

⇒ Gut sind kleine Werte von q_ϕ für $\theta \in \Theta_0$ und große Werte von q_ϕ für $\theta \in \Theta_1$.

Generell gibt es Abhängigkeiten zwischen den Werten von q_ϕ für $\theta \in \Theta_0$ und $\theta \in \Theta_1$:

Sei zum Beispiel ϕ_α der Test definiert durch $\phi_\alpha(X) := 0$ für alle $X \in \mathcal{X}$, also der Test, der die Nullhypothese, unabhängig von den beobachteten Daten, *niemals ablehnt*. Für diesen Test gilt $q_{\phi_\alpha}(\theta) = 0$ für $\theta \in \Theta_0$. Allerdings gilt für diesen Test auch $q_{\phi_\alpha}(\theta) = 0$ für $\theta \in \Theta_1$.

Andersherum sei ϕ_b der Test definiert durch $\phi_b(X) := 1$ für alle $X \in \mathcal{X}$, also ein Test, der die Nullhypothese, unabhängig von den beobachteten Daten, *immer ablehnt*. Für diesen Test gilt $q_{\phi_b}(\theta) = 1$ für $\theta \in \Theta_1$. Allerdings gilt für diesen Test auch $q_{\phi_b}(\theta) = 1$ für $\theta \in \Theta_0$.

In der Konstruktion eines Tests muss also eine angemessene Balance zwischen kleinen Werten von q_ϕ für $\theta \in \Theta_0$ und großen Werten von q_ϕ für $\theta \in \Theta_1$ gefunden werden.

Intuition zur Testkonstruktion

Die populärste Methode, eine Balance zwischen zwischen kleinen Werten von q für $\theta \in \Theta_0$ und großen Werten von q für $\theta \in \Theta_1$ zu finden, ist in einem ersten Schritt ein $\alpha_0 \in [0, 1]$ zu wählen und sicher zu stellen, dass

$$q_\phi(\theta) \leq \alpha_0 \text{ für alle } \theta \in \Theta_0. \quad (17)$$

Eine konventionelle Wahl für 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 *Signifikanzniveau* bezeichnet) und des Testumfangs α (auch als *effektives Niveau* bezeichnet).

Definition (Level- α_0 -Test, Signifikanzlevel α_0 , Testumfang α)

q_ϕ sei die Testgütefunktion eines Tests ϕ und es sei $\alpha_0 \in [0, 1]$. Dann heißt ein Test ϕ , für den gilt, dass

$$q_\phi(\theta) \leq \alpha_0 \text{ für alle } \theta \in \Theta_0 \quad (18)$$

ein *Level- α_0 -Test* und man sagt, dass der Test das *Signifikanzlevel* α_0 hat. Die Zahl

$$\alpha := \max_{\theta \in \Theta_0} q_\phi(\theta) \in [0, 1] \quad (19)$$

heißt der *Testumfang* von ϕ .

Bemerkungen

- α ist die größtmögliche Wahrscheinlichkeit für einen Typ I Fehler.
- Ein Test ist dann, und nur dann, ein Level- α_0 -Test, wenn $\alpha \leq \alpha_0$ gilt.
- Bei einer einfachen Nullhypothese gilt für den Testumfang, dass $\alpha = q_\phi(\theta_0) = \mathbb{P}_{\theta_0}(\phi = 1)$.

Typ I Fehlerwahrscheinlichkeit vs. Testumfang vs. Signifikanzlevel

Bei einfacher Θ_0 ist der Testumfang gleich der Wahrscheinlichkeit eines Typ I Fehlers

$$\alpha := \max_{\theta \in \Theta_0} q_\phi(\theta) = \max_{\theta \in \{\theta_0\}} q_\phi(\theta) = q_\phi(\theta_0) = \mathbb{P}_{\theta_0}(\phi = 1). \quad (20)$$

Bei zusammengesetzter Θ_0 gibt es je nach Wert von $\theta \in \Theta_0$ verschiedene Wahrscheinlichkeiten für einen Typ I Fehler. Die größte dieser Wahrscheinlichkeiten ist der Testumfang

$$\alpha := \max_{\theta \in \Theta_0} q_\phi(\theta) = \max_{\theta \in \Theta_0} \mathbb{P}_\theta(\phi = 1). \quad (21)$$

Ein Test hat Signifikanzlevel α_0 , wenn der Testumfang kleiner oder gleich α_0 ist.

$$\alpha = \max_{\theta \in \Theta_0} q_\phi(\theta) \leq \alpha_0 \quad (22)$$

Ein Test, bei dem das Signifikanzlevel größer als der Testumfang ist, heißt *konservativ*.

Ein Test, bei dem das Signifikanzlevel gleich dem Testumfang ist, heißt *exakt*.

Zur Wahl von Nullhypothese und Alternativhypothese

Das Vorgehen in der Testkonstruktion zunächst durch die Wahl eines Signifikanzniveaus den Testumfang zu begrenzen und erst in einem zweiten Schritt dafür zu sorgen, dass die Wahrscheinlichkeit von $\phi = 1$ bei $\theta \in \Theta_1$ bei diesem Signifikanzniveau möglichst groß ist, induziert eine Asymmetrie in der Behandlung von Null- und Alternativhypothese. Implizit wichtet man mit diesem Vorgehen Typ I Fehler als schwerwiegender als Typ II Fehler.

Dies wiederum impliziert eine mögliche Strategie zur Festlegung von Null- und Alternativhypothese: Die Nullhypothese ist die Hypothese, hinsichtlich deren assoziierter Testentscheidung man eher keinen Fehler machen möchte bzw. deren Fehlerwahrscheinlichkeit man primär kontrollieren möchte.

In der wissenschaftlichen Anwendung ist es Standard, die falsche Konfirmation der eigenen Theorie als einen schwerwiegenderen Fehler als die falsche Ablehnung der eigenen Theorie zu werten.

Die falsche Konfirmation der eigenen Theorie sollte also ein Typ I Fehler, das falsche Ablehnen der eigenen Theorie ein Typ II Fehler sein.

Damit die falsche Konfirmation der eigenen Theorie einen Typ I Fehler, also das Ablehnen von H_0 bei Zutreffen von H_0 , darstellt, muss die eigene Theorie als Alternativhypothese aufgestellt werden. Die Alternativhypothese fälschlicherweise abzulehnen wird damit ein Typ II Fehler.

Dirk Ostwalds Meinung zum Hypothesentesten in der Wissenschaft

Frequentistisches Hypothesentesten ist als Entscheidungsproblem ohne klar und explizit definierte Entscheidungsnutzenfunktion formuliert und deshalb recht mühselig zu analysieren und zu studieren. Es gibt sehr viel zugänglichere Theorien zu Entscheidungen unter Unsicherheit (vgl. 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.

Vorbemerkungen

Grundlegende Definitionen

Z-Test

p-Werte

Konfidenzintervalle und Hypothesentests

Selbstkontrollfragen

- (1) Statistisches Modell, Hypothesen, Teststatistik, Test
- (2) Analyse der Testgütefunktion
- (3) Testumfangkontrolle
- (4) Analyse der Powerfunktion

Z-Test (1) Statistisches Modell, Hypothesen, Teststatistik, Test

Statistisches Modell

$X_1, \dots, X_n \sim N(\mu, \sigma^2)$ sei die Stichprobe eines **Normalverteilungsmodells** mit **unbekanntem Erwartungswertparameter** μ und **bekanntem Varianzparameter** $\sigma^2 > 0$. Der Parameter dieses Modells ist $\theta = \mu$ und der Parameterraum dieses Modells ist $\Theta = \mathbb{R}$.

Testhypothesen, Teststatistik, Test

Für ein $\mu_0 \in \mathbb{R}$ betrachten wir die einfache Nullhypothese und die zusammengesetzte Alternativhypothese

$$H_0 : \mu = \mu_0 \Leftrightarrow \Theta_0 := \{\mu_0\} \text{ und } H_1 : \mu \neq \mu_0 \Leftrightarrow \Theta_1 := \mathbb{R} \setminus \{\mu_0\}, \quad (23)$$

respektive. Weiterhin betrachten wir die *Z-Teststatistik*

$$Z := \sqrt{n} \left(\frac{\bar{X}_n - \mu_0}{\sigma} \right) \quad (24)$$

und definieren den zweiseitigen kritischen Wert-basierten Test

$$\phi(X) := 1_{\{|Z| \geq k\}} = \begin{cases} 1 & |Z| \geq k \\ 0 & |Z| < k \end{cases}. \quad (25)$$

Theorem (Testgütefunktion bei zweiseitigem Z-Test, H_0 einfach)

Es sei $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ die Stichprobe eines Normalverteilungsmodells mit unbekanntem Erwartungswertparameter μ und bekanntem Varianzparameter $\sigma^2 > 0$. Für $\Theta := \mathbb{R}$ und ein $\mu_0 \in \mathbb{R}$ sei ein TestszENARIO durch

$$H_0 : \mu = \mu_0 \Leftrightarrow \Theta_0 := \{\mu_0\} \text{ und } H_1 : \mu \neq \mu_0 \Leftrightarrow \Theta_1 := \mathbb{R} \setminus \{\mu_0\}, \quad (26)$$

gegeben. Weiterhin sei mit der Z-Teststatistik

$$Z := \sqrt{n} \left(\frac{\bar{X}_n - \mu_0}{\sigma} \right) \quad (27)$$

in diesem Szenario der zweiseitige kritischer Wert-basierte Test

$$\phi(X) := 1_{\{|Z| \geq k\}} = \begin{cases} 1 & |Z| \geq k \\ 0 & |Z| < k \end{cases}. \quad (28)$$

definiert. Dann ist die Testgütefunktion von ϕ gegeben durch

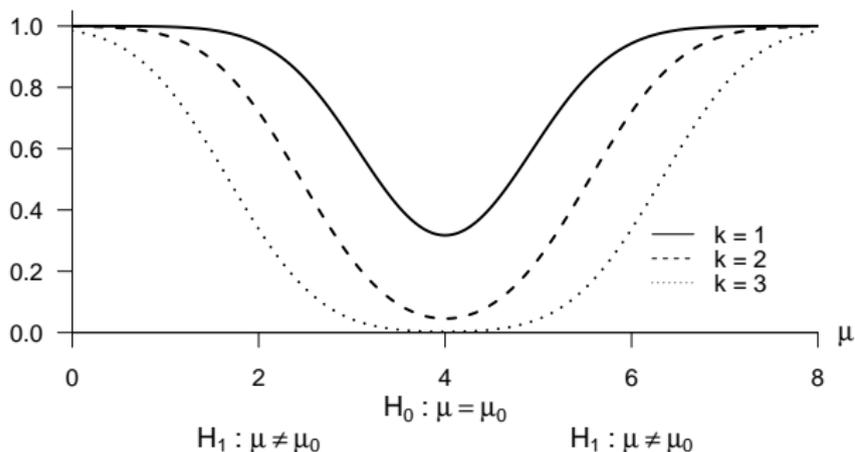
$$q_\phi : \mathbb{R} \rightarrow [0, 1], \mu \mapsto q_\phi(\mu) := 1 - \Phi \left(k - \frac{\sqrt{n}}{\sigma} (\mu - \mu_0) \right) + \Phi \left(-k - \frac{\sqrt{n}}{\sigma} (\mu - \mu_0) \right), \quad (29)$$

wobei Φ die KVF der Standardnormalverteilung bezeichnet.

Z-Test (2) Analyse der Testgütefunktion

Testgütefunktion q_ϕ für $\mu_0 = 4$, $n = 15$, $\sigma^2 = 9$ und drei zweiseitigen kritischen Wert-basierten Tests ϕ mit kritischen Werten $k = 1, 2, 3$ (vgl. DeGroot and Schervish (2012), Abbildung 9.1).

$$q_\phi(\mu) = \mathbb{P}_\mu(\phi = 1)$$



Z-Test (2) Analyse der Testgütefunktion

Beweis

Die Testgütefunktion des betrachteten Test im vorliegenden Testscenario ist definiert als

$$q_\phi : \mathbb{R} \rightarrow [0, 1], \mu \mapsto q_\phi(\mu) := \mathbb{P}_\mu(\phi = 1). \quad (30)$$

Wir sind also an der Wahrscheinlichkeit $\mathbb{P}_\mu(\phi = 1)$ mit $\mu \in \mathbb{R}$ interessiert, d.h. an der Wahrscheinlichkeit dafür, dass der betrachtete Test den Wert 1 annimmt, wenn der wahre, aber unbekannte, Parameterwert μ ist. Wir hatten oben festgehalten, dass die Wahrscheinlichkeiten dafür, dass ein Test den Wert 1 annimmt und dafür, dass die zugehörige Teststatistik im Ablehnungsbereich des Tests liegt, gleich sind.

Um $\mathbb{P}_\mu(\phi = 1)$ auswerten zu können, benötigen wir also zunächst einmal die Verteilung der Teststatistik für $\mu \in \mathbb{R}$. Sei also $\mu \in \mathbb{R}$. Dann betrachten wir im vorliegenden Testscenario die Stichprobe $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ mit unbekanntem Erwartungswertparameter μ und bekanntem Varianzparameter $\sigma^2 > 0$ und die Z -Teststatistik $Z := \sqrt{n}((\bar{X}_n - \mu_0)/\sigma)$. Man beachte, dass hier sowohl $\mu = \mu_0$ als auch $\mu \neq \mu_0$ zugelassen sind. Wir befinden uns also nicht nur im in (11) Konfidenzintervalle betrachteten Szenario $\mu = \mu_0$. Es gilt also zunächst, die Verteilung der Z -Teststatistik für ein generelles $\mu \in \mathbb{R}$ zu bestimmen.

Dazu folgen wir der in (11) Konfidenzintervalle etablierten Herangehensweise: Wir halten zunächst fest, dass mit der Mittelwerttransformation für unabhängig und identisch normalverteilte Zufallsvariablen gilt, dass $\bar{X}_n \sim N(\mu, \sigma^2/n)$ (cf. (8) Transformationen der Normalverteilung). Die Verteilung der Z -Teststatistik folgt dann mit dem WDF Transformationstheorem bei linear-affinen Abbildungen (cf. *ibid.*) durch Transformation von \bar{X}_n mit der Funktion

$$f : \mathbb{R} \rightarrow \mathbb{R}, \bar{x}_n \mapsto f(\bar{x}_n) := \frac{\sqrt{n}}{\sigma} \bar{x}_n - \frac{\sqrt{n}}{\sigma} \mu_0 \quad (31)$$

Z-Test (2) Analyse der Testgütefunktion

Beweis (fortgeführt)

Wir erhalten

$$\begin{aligned} p_Z(z) &= \frac{\sigma}{\sqrt{n}} N\left(\frac{\sigma}{\sqrt{n}}\left(z + \frac{\sqrt{n}}{\sigma}\mu_0\right); \mu, \frac{\sigma^2}{n}\right) \\ &= \frac{\sigma}{\sqrt{n}} \frac{1}{\sqrt{2\pi \frac{\sigma^2}{n}}} \exp\left(-\frac{1}{2 \frac{\sigma^2}{n}} \left(\frac{\sigma}{\sqrt{n}}\left(z + \frac{\sqrt{n}}{\sigma}\mu_0\right) - \mu\right)^2\right) \\ &= \frac{\sigma}{\sqrt{n}} \frac{\sqrt{n}}{\sigma} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} \frac{n}{\sigma^2} \left(\frac{\sigma}{\sqrt{n}}z - (\mu - \mu_0)\right)^2\right) \\ &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} \frac{n}{\sigma^2} \left(\frac{\sigma^2}{n}z^2 - 2\frac{\sigma}{\sqrt{n}}z(\mu - \mu_0) + (\mu - \mu_0)^2\right)\right) \\ &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{n}{\sigma^2} \frac{\sigma^2}{n} z^2 - 2\frac{n}{\sigma^2} \frac{\sigma}{\sqrt{n}} z(\mu - \mu_0) + \frac{n}{\sigma^2} (\mu - \mu_0)^2\right)\right) \\ &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(z^2 - 2z \frac{\sigma}{\sqrt{n}} \frac{\sqrt{n}}{\sigma} (\mu - \mu_0) + \left(\frac{\sqrt{n}}{\sigma}\right)^2 (\mu - \mu_0)^2\right)\right) \\ &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(z^2 - 2z \frac{\sqrt{n}}{\sigma} (\mu - \mu_0) + \left(\frac{\sqrt{n}}{\sigma} (\mu - \mu_0)\right)^2\right)\right) \\ &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(z - \frac{\sqrt{n}}{\sigma} (\mu - \mu_0)\right)^2\right) \end{aligned} \tag{32}$$

Z-Test (2) Analyse der Testgütefunktion

Beweis (fortgeführt)

Wir haben also gefunden, dass für die Z -Teststatistik für beliebiges $\mu \in \mathbb{R}$ gilt, dass

$$Z \sim N\left(\frac{\sqrt{n}}{\sigma}(\mu - \mu_0), 1\right). \quad (33)$$

Für den Fall, dass $\mu_0 = \mu$ ist, die Nullhypothese also zutrifft, folgt $Z \sim N(0, 1)$. Als nächstes bestimmen wir den Ablehnungsbereich des betrachteten Tests. Per Definition gilt

$$A := \{Z \in \mathbb{R} \mid \phi(X) = 1\} \quad (34)$$

Es ergibt sich also

$$\begin{aligned} A &= \{Z \in \mathbb{R} \mid \phi(X) = 1\} \\ &= \{Z \in \mathbb{R} \mid |Z| \geq k\} \\ &= \left\{ Z \in \mathbb{R} \mid \begin{cases} Z \geq k, & Z \geq 0 \\ -Z \geq k, & Z < 0 \end{cases} \right\} \\ &= \left\{ Z \in \mathbb{R} \mid \begin{cases} Z \geq k, & Z \geq 0 \\ Z \leq -k, & Z < 0 \end{cases} \right\} \\ &= \{Z \in \mathbb{R} \mid Z < 0, Z \leq -k\} \cup \{Z \in \mathbb{R} \mid Z > 0, Z \geq k\} \\ &=] - \infty, -k] \cup]k, \infty[. \end{aligned} \quad (35)$$

Z-Test (2) Analyse der Testgütefunktion

Beweis (fortgeführt)

Mit diesem Ablehnungsbereich des betrachteten Tests ergibt sich dann

$$\begin{aligned}q_{\phi}(\mu) &= \mathbb{P}_{\mu}(\phi = 1) \\&= \mathbb{P}_{\mu}(Z \in]-\infty, -k] \cup]k, \infty[) \\&= \mathbb{P}_{\mu}(Z \in]-\infty, -k]) + \mathbb{P}_{\mu}(Z \in [k, \infty[) \\&= \mathbb{P}_{\mu}(Z \leq -k) + \mathbb{P}_{\mu}(Z \geq k) \\&= \mathbb{P}_{\mu}(Z \leq -k) + (1 - \mathbb{P}_{\mu}(Z \leq k)) \\&= 1 - \mathbb{P}_{\mu}(Z \leq k) + \mathbb{P}_{\mu}(Z \leq -k)\end{aligned}\tag{36}$$

Wir sind also an der Wahrscheinlichkeiten $\mathbb{P}_{\mu}(Z \leq -k)$ und $\mathbb{P}_{\mu}(Z \leq k)$ interessiert. Dazu halten wir zunächst fest, dass für jede normalverteilte Zufallsvariable $X \sim N(m, s^2)$ gilt, dass

$$\mathbb{P}(X \leq x) = \mathbb{P}\left(\frac{X - m}{s} \leq \frac{x - m}{s}\right) = \mathbb{P}\left(\xi \leq \frac{x - m}{s}\right) = \Phi\left(\frac{x - m}{s}\right)\tag{37}$$

gilt, wobei ξ hier die durch $(X - m)/s$ definierte Z -verteilte Zufallsvariable bezeichnet. Im vorliegenden Fall folgt mit $m = \frac{\sqrt{n}}{\sigma}(\mu - \mu_0)$, $s = 1$ und $x = k$ bzw. $x = -k$ also

$$q_{\phi}(\mu) = 1 - \mathbb{P}_{\mu}(Z \leq k) + \mathbb{P}_{\mu}(Z \leq -k) = 1 - \Phi\left(k - \frac{\sqrt{n}}{\sigma}(\mu - \mu_0)\right) + \Phi\left(-k - \frac{\sqrt{n}}{\sigma}(\mu - \mu_0)\right)$$

und wir haben alles gezeigt. □

Theorem (Testumfangkontrolle bei zweiseitigem Z-Test, H_0 einfach)

Es sei $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ die Stichprobe eines Normalverteilungsmodells mit unbekanntem Erwartungswertparameter μ und bekanntem Varianzparameter $\sigma^2 > 0$. Für $\Theta := \mathbb{R}$ und ein $\mu_0 \in \mathbb{R}$ sei ein Testscenario durch

$$H_0 : \mu = \mu_0 \Leftrightarrow \Theta_0 := \{\mu_0\} \text{ und } H_1 : \mu \neq \mu_0 \Leftrightarrow \Theta_1 := \mathbb{R} \setminus \{\mu_0\}, \quad (38)$$

gegeben. Weiterhin sei mit der Z-Teststatistik

$$Z := \sqrt{n} \left(\frac{\bar{X}_n - \mu_0}{\sigma} \right) \quad (39)$$

in diesem Szenario der zweiseitige kritische Wert-basierte Test

$$\phi(X) := 1_{\{|Z| \geq k\}} = \begin{cases} 1 & |Z| \geq k \\ 0 & |Z| < k \end{cases}. \quad (40)$$

definiert. Dann ist ϕ ein Level- α_0 -Test mit Testumfang α_0 , wenn der kritische Wert k definiert ist durch

$$k := \Phi^{-1} \left(1 - \frac{\alpha_0}{2} \right) \quad (41)$$

Wir bezeichnen diesen kritischen Wert mit k_{α_0} .

Z-Test (3) Testumfangkontrolle

Beweis

Damit der betrachtete Test ein Level- α_0 -Test ist, muss bekanntlich $q_\phi(\mu) \leq \alpha_0$ für alle $\mu \in \{\mu_0\}$, also hier $q_\phi(\mu_0) \leq \alpha_0$, gelten. Weiterhin ist der Testumfang des betrachteten Tests durch $\alpha = \max_{\mu \in \{\mu_0\}} q_\phi(\mu)$, also hier durch $\alpha = q_\phi(\mu_0)$ gegeben. Wir müssen also zeigen, dass die Wahl von

$$k_{\alpha_0} := \Phi^{-1}(1 - \alpha_0/2) \quad (42)$$

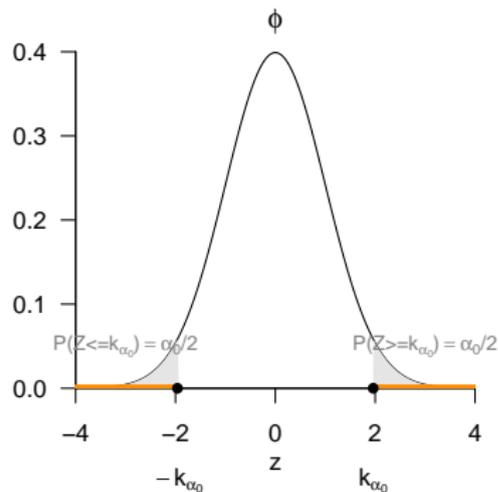
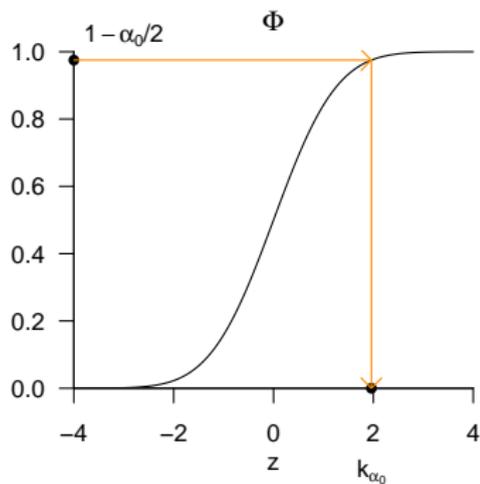
garantiert, dass ϕ ein Level- α_0 -Test mit Testumfang α_0 ist. Sei also $k := k_{\alpha_0}$. Dann gilt

$$\begin{aligned} q_\phi(\mu_0) &= 1 - \Phi(k_{\alpha_0}) + \Phi(-k_{\alpha_0}) \\ &= 1 - \Phi(k_{\alpha_0}) + (1 - \Phi(k_{\alpha_0})) \\ &= 2 - 2\Phi(k_{\alpha_0}) \\ &= 2(1 - \Phi(k_{\alpha_0})) \\ &= 2 \left(1 - \Phi \left(\Phi^{-1}(1 - \alpha_0/2) \right) \right) \\ &= 2(1 - 1 + \alpha_0/2) \\ &= 2\alpha_0/2 \\ &= \alpha_0, \end{aligned} \quad (43)$$

wobei die zweite Gleichung mit der Symmetrie der Standardnormalverteilung folgt. Es folgt also direkt, dass bei der Wahl von $k = k_{\alpha_0}$, $q_\phi(\mu_0) \leq \alpha_0$ ist und der betrachtete Test somit ein Level- α_0 -Test ist. Weiterhin folgt direkt, dass der Testumfang des betrachteten Tests bei der Wahl von $k = k_{\alpha_0}$ gleich α_0 ist.

Z-Test (3) Testumfangkontrolle

Wahl von $k_{\alpha_0} := \Phi^{-1}(1 - \alpha_0/2)$ mit $\alpha_0 := 0.05$ und resultierender Ablehnungsbereich



Praktisches Vorgehen

- Man nimmt an, dass ein vorliegender Datensatz x_1, \dots, x_n eine Realisation von $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ mit unbekanntem μ und bekanntem Varianzparameter $\sigma^2 > 0$ ist.
- Man möchte entscheiden ob für ein $\mu_0 \in \mathbb{R}$ eher $H_0 : \mu = \mu_0$ oder $H_1 : \mu \neq \mu_0$ zutrifft.
- Man wählt ein Signifikanzniveau α_0 und bestimmt den zugehörigen kritischen Wert k_{α_0} . Zum Beispiel gilt bei Wahl von $\alpha_0 := 0.05$, dass $k_{0.05} = \Phi^{-1}(1 - 0.05/2) \approx 1.96$ ist.
- Anhand von n, μ_0, σ^2 und \bar{x}_n berechnet man die Realisierung der Z -Teststatistik

$$z := \sqrt{n} \left(\frac{\bar{x}_n - \mu_0}{\sigma} \right) \quad (44)$$

- Wenn z größer-gleich k_{α_0} ist oder wenn z kleiner-gleich $-k_{\alpha_0}$ ist, lehnt man die Nullhypothese ab, andernfalls lehnt man sie nicht ab.
- Die oben entwickelte Theorie des Z -Tests garantiert dann, dass man in höchstens $\alpha_0 \cdot 100$ von 100 Fällen die Nullhypothese fälschlicherweise ablehnt.

Z-Test (3) Testumfangkontrolle

Simulation

```
# Modellspezifikation
mu           = 2           # w.a.u. Erwartungswertparameter
sigsqr      = 1           # bekannter Varianzparameter
n           = 12          # Stichprobengröße
mu_0        = mu          # Nullhypothese H_0
alpha_0     = 0.05        # Signifikanzniveau
k_alpha_0   = qnorm(1 - (alpha_0/2)) # kritischer Wert

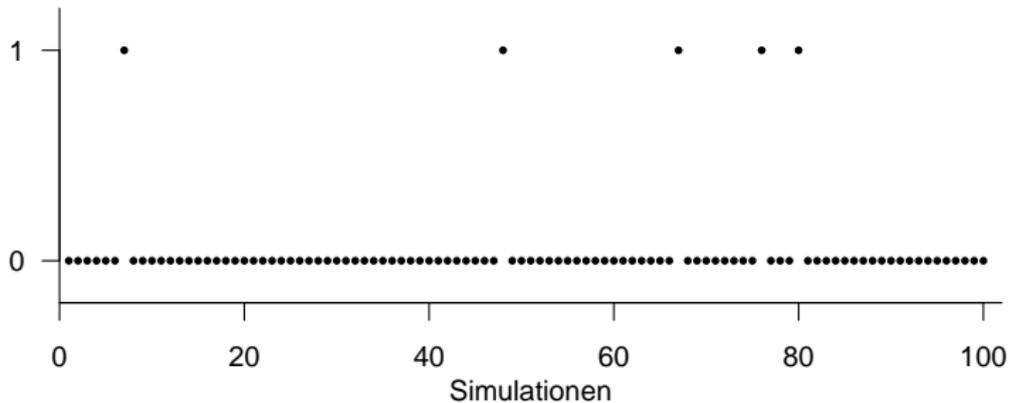
# Simulation
n_sim       = 1e4         # Anzahl von Stichprobenrealisierung
phi         = rep(NA, n_sim) # Testergebnisarray
for(i in 1:n_sim){
  X         = rnorm(n, mu, sqrt(sigsqr)) # Stichprobenrealisation
  Z         = sqrt(n)*(mean(X) - mu_0)/sqrt(sigsqr) # Teststatistik
  if(abs(Z) >= k_alpha_0){ # Test 1_{|Z(X)| >= k_alpha_0}
    phi[i] = 1 # Ablehnen von H_0
  } else {
    phi[i] = 0 # Nicht Ablehnen von H_0
  }
}
cat("Geschätzter Testumfang alpha =", mean(phi)) # Ausgabe
```

> Geschätzter Testumfang alpha = 0.0517

Z-Test (3) Testumfangkontrolle

Simulation

$$\phi(x) \text{ für } \mu = 2, \sigma^2 = 2, n = 12, \mu_0 = 2, \alpha_0 = 0.05$$



Z-Test (4) Analyse der Powerfunktion

Wir betrachten die Testgütefunktion

$$q_\phi : \mathbb{R} \rightarrow [0, 1], \mu \mapsto q_\phi(\mu) := 1 - \Phi\left(k - \frac{\sqrt{n}}{\sigma}(\mu - \mu_0)\right) + \Phi\left(-k - \frac{\sqrt{n}}{\sigma}(\mu - \mu_0)\right) \quad (45)$$

bei bekanntem und festem Wert von $\sigma^2 > 0$, bei festgelegter Nullhypothese, also festem Wert von μ_0 und bei kontrolliertem Testumfang, also für $k := k_{\alpha_0} = \Phi^{-1}(1 - \alpha_0/2)$ mit festem α_0 , als Funktion der übrigen Testszenarioparameter.

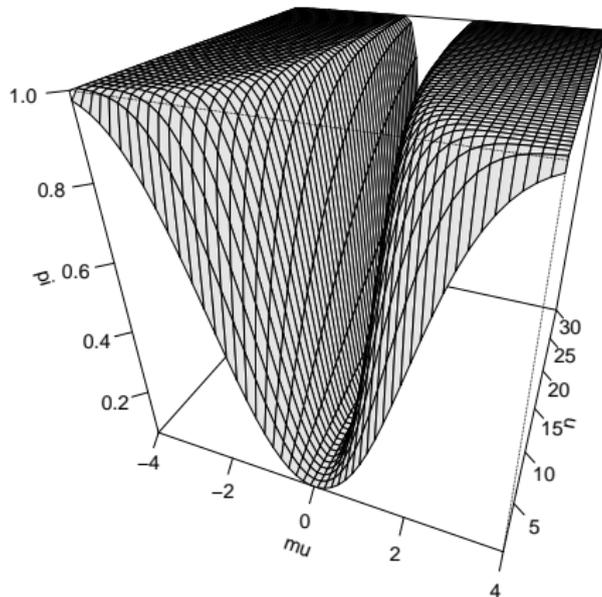
Es ergibt sich die bivariate reellwertige Funktion

$$\begin{aligned} \pi : \mathbb{R} \times \mathbb{N} &\rightarrow [0, 1], (\mu, n) \mapsto \\ \pi(\mu, n) &:= 1 - \Phi\left(k_{\alpha_0} - \frac{\sqrt{n}}{\sigma}(\mu - \mu_0)\right) + \Phi\left(-k_{\alpha_0} - \frac{\sqrt{n}}{\sigma}(\mu - \mu_0)\right) \end{aligned} \quad (46)$$

Bei festgelegten $\sigma^2 > 0, \mu_0, k_{\alpha_0}$ hängt für $\mu \notin \{\mu_0\}$ die Powerfunktion des zweiseitigen Z-Tests mit einfacher Nullhypothese also vom unbekanntem Wert μ und von der Stichprobengröße n ab. Wir visualisieren diese Abhängigkeiten untenstehend.

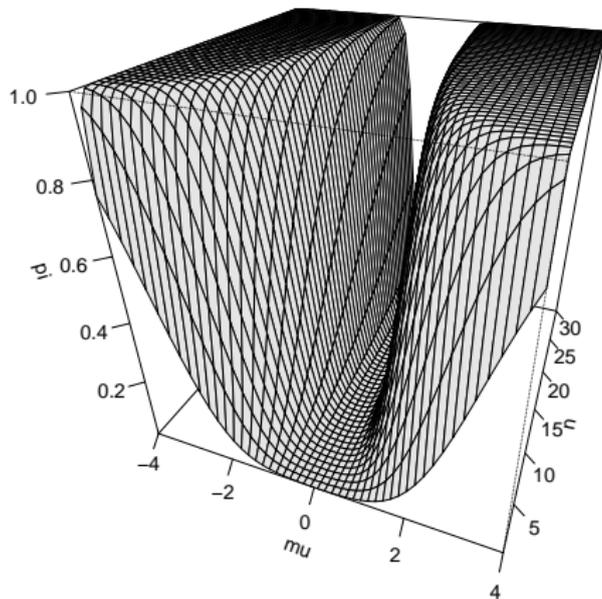
Z-Test (4) Analyse der Powerfunktion

Powerfunktion für $\mu_0 = 0$ und $\sigma^2 = 1$ bei $\alpha_0 = 0.05$



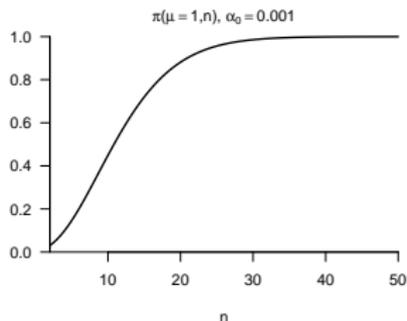
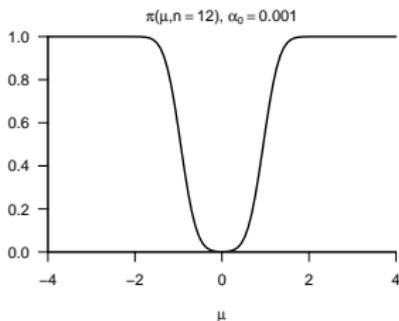
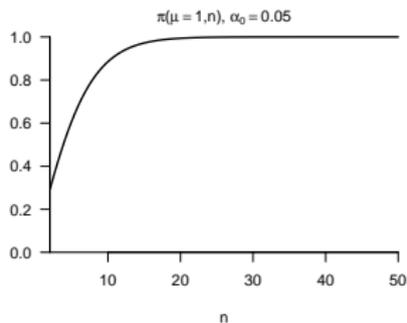
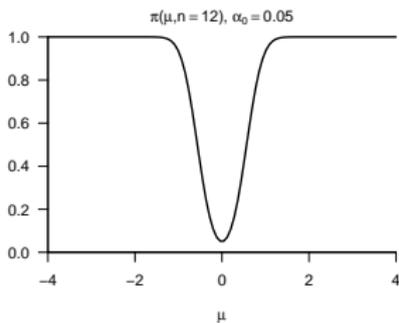
Z-Test (4) Analyse der Powerfunktion

Powerfunktion für $\mu_0 = 0$ und $\sigma^2 = 1$ bei $\alpha_0 = 0.001$



Z-Test (4) Analyse der Powerfunktion

Powerfunktionen für $\mu_0 = 0$ und $\sigma^2 = 1$



Z-Test (4) Analyse der Powerfunktion

Praktisches Vorgehen

Mit größerem n steigt die Powerfunktion des Tests an

- Ein großer Stichprobenumfang ist besser als ein kleiner Stichprobenumfang.
- Kosten für die Erhöhung des Stichprobenumfangs werden aber nicht berücksichtigt.

⇒ Die Theorie statistischer Hypothesentests ist nicht besonders lebensnah.

Die Powerfunktion hängt vom wahren, aber unbekanntem, Parameterwert μ ab.

⇒ Wenn man μ schon kennen würde, würde man den Test nicht durchführen.

Generell wird folgendes Vorgehen favorisiert

- Man legt das Signifikanzniveau α_0 fest und evaluiert die Powerfunktion.
- Man wählt einen Mindestparameterwert μ^* , den man mit $\pi(n, \mu) = \beta$ detektieren möchte.
- Ein konventioneller Wert ist $\beta = 0.8$.
- Man liest die für $\pi(n, \mu = \mu^*) = \beta$ nötige Stichprobengröße n ab.

Z-Test (4) Analyse der Powerfunktion

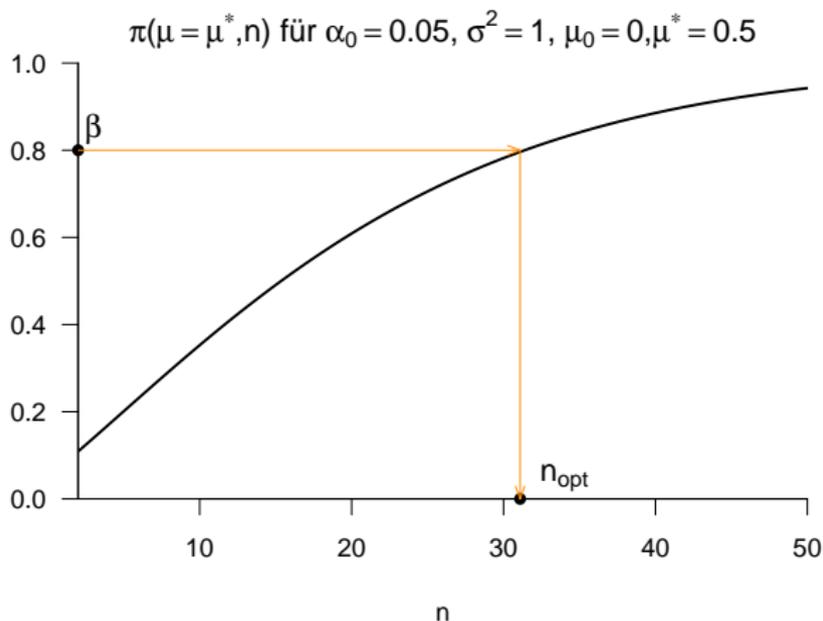
Praktisches Vorgehen

```
# Szenariospezifikation
sigma      = 1                # bekanntes \sigma
mu_0       = 0                # einfache Nullhypothese
n_min      = 2                # n Minimum
n_max      = 50               # n Maximum
n_res      = 1e2              # n Auflösung
n          = seq(n_min,n_max, len = n_res) # n Raum
alpha_0    = 0.05             # Signifikanzniveau
k_alpha_0  = qnorm(1 - alpha_0/2) # kritischer Wert

# Poweranalyse
mu_star    = .5                # Mindestparameterwert
pi_n       = (1-pnorm( k_alpha_0-sqrt(n)/sigma*(mu_star-mu_0))
              +pnorm(-k_alpha_0-sqrt(n)/sigma*(mu_star-mu_0))) # Powerfunktion
beta       = 0.8                # gewünschter Powerfunktionswert
i          = 1                # Indexinitialisierung
n_min      = NaN               # minimales n Initialisierung
while(pi_n[i] < beta){
  n_min    = n[i]              # Solange |\pi(n, \mu^*)| < |\beta|
  i        = i + 1            # Aufnahme des minimal nötigen ns
}                                # und Erhöhung des Indexes
cat("Minimal nötiges n =", ceiling(n_min)) # Ausgabe
```

```
> Minimal nötiges n = 32
```

Z-Test (4) Analyse der Powerfunktion



Vorbemerkungen

Grundlegende Definitionen

Z-Test

p-Werte

Konfidenzintervalle und Hypothesentests

Selbstkontrollfragen

Motivation

- Es werde ein zweiseitiger Z-Test mit Signifikanzlevel $\alpha_0 = 0.05$ durchgeführt.
 - H_0 wird abgelehnt, wenn $|Z| \geq \Phi^{-1}(1 - 0.05/2) \approx 1.96$.
 - Nehmen wir an, es werde $z = 2.01$ beobachtet.
 - Das Testergebnis lautet " H_0 Ablehnen".
 - Nehmen wir an, es werde $z = 3.81$ beobachtet.
 - Das Testergebnis lautet " H_0 Ablehnen".
 - Der alleinige Bericht des Testergebnis supprimiert interessante Information.
- ⇒ Neben der Testumfangkontrolle durch z.B. $\alpha_0 = 0.05$ ist es daher üblich, alle Werte von α_0 anzugeben, für die ein Level- α_0 -Test zum Ablehnen von H_0 führen würde.
- Bei $z = 2.01$ würde H_0 für jedes α_0 mit $2.01 \geq \Phi^{-1}(1 - \alpha_0/2)$ abgelehnt werden.
 - Bei $z = 3.81$ würde H_0 für jedes α_0 mit $3.81 \geq \Phi^{-1}(1 - \alpha_0/2)$ abgelehnt werden.
- Das kleinste Signifikanzlevel α_0 , bei dem man H_0 basierend auf einem Wert der Teststatistik ablehnen würde, wird *p-Wert* genannt.

Definition (p-Wert)

ϕ sei ein kritischer Wert-basierter Test. Der *p-Wert* ist das kleinste Signifikanzlevel α_0 , bei welchem man die Nullhypothese basierend auf einem vorliegendem Wert der Teststatistik ablehnen würde.

Beispiel (Zweiseitiger Z-Test mit einfacher Nullhypothese)

- Bei $Z = z$ würde H_0 für jedes α_0 mit $|z| \geq \Phi^{-1}(1 - \alpha_0/2)$ abgelehnt werden. Für diese α_0 gilt, wie unten gezeigt,

$$\alpha_0 \geq 2\mathbb{P}(Z \geq |z|). \quad (47)$$

- Das kleinste $\alpha_0 \in [0, 1]$ mit $\alpha_0 \geq 2\mathbb{P}(Z \geq |z|)$ ist dann $\alpha_0 = 2\mathbb{P}(Z \geq |z|)$, also folgt

$$\text{p-Wert} = 2\mathbb{P}(Z \geq |z|) = 2(1 - \Phi(|z|)). \quad (48)$$

- Zum Beispiel ist für $Z = 2.01$ der p-Wert 0.04, für $Z = -2.01$ ist der p-Wert auch 0.04, für $Z = 3.81$ ist der p-Wert 0.0001, und für $Z = -3.81$ ist der p-Wert auch 0.0001

Beispiel (Zweiseitiger Z-Test mit einfacher Nullhypothese, fortgeführt)

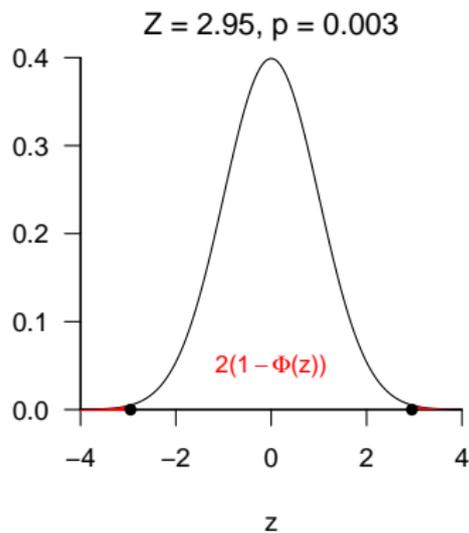
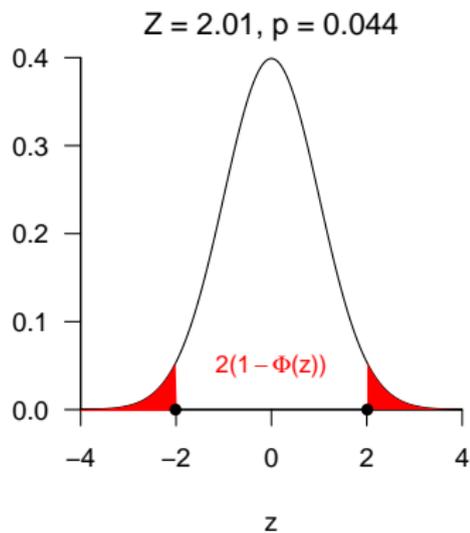
Es bleibt zu zeigen, dass gilt

$$|z| \geq \Phi^{-1}(1 - \alpha_0/2) \Leftrightarrow \alpha_0 \geq 2\mathbb{P}(Z \geq |z|) \quad (49)$$

Wir haben

$$\begin{aligned} |z| &\geq \Phi^{-1}\left(1 - \frac{\alpha_0}{2}\right) \\ \Leftrightarrow |z| &\geq \Phi^{-1}\left(1 - \frac{\alpha_0}{2}\right) \\ \Leftrightarrow \Phi(|z|) &\geq \Phi\left(\Phi^{-1}\left(1 - \frac{\alpha_0}{2}\right)\right) \\ \Leftrightarrow \Phi(|z|) &\geq 1 - \frac{\alpha_0}{2} \\ \Leftrightarrow \mathbb{P}(Z \leq |z|) &\geq 1 - \frac{\alpha_0}{2} \\ \Leftrightarrow \frac{\alpha_0}{2} &\geq 1 - \mathbb{P}(Z \leq |z|) \\ \Leftrightarrow \frac{\alpha_0}{2} &\geq \mathbb{P}(Z \geq |z|) \\ \Leftrightarrow \alpha_0 &\geq 2\mathbb{P}(Z \geq |z|). \end{aligned} \quad (50)$$

Beispiel (Zweiseitiger Z-Test mit einfacher Nullhypothese)



Bemerkungen

- p-Werte spiegeln die Antwort auf die intuitive Frage wie wahrscheinlich es im frequentistischen Sinne wäre, den beobachteten oder einen extremeren Wert der Teststatistik unter der Annahme eines Nullmodells zu observieren.
- p-Werte sind ist extrem populär, ihre uninformierte Benutzung ist aber auch sehr umstritten.
- [The American Statistician \(2019\) Statistical Inference in the 21st Century: A World Beyond \$p < 0.05\$](#)
- p-Werte werden, wie Hypothesentestergebnisse generell, leider oft überinterpretiert.
- Es gibt basierend auf dem Gesagten keinen Grund dies anzunehmen, trotzdem vorsorglich:
 - p-Werte quantifizieren nicht die Wahrscheinlichkeit, dass die Nullhypothese wahr ist.
 - Aufgrund von $p < 0.05$ sollte man nicht glauben, dass ein Effekt existiert.
 - Aufgrund von $p > 0.05$ sollte man nicht glauben, dass ein Effekt nicht existiert.
- p-Werte sind eine Möglichkeit ein Signal-zu-Rauschen Verhältnis zu quantifizieren.
- p-Werte sind eine Möglichkeit Unsicherheit zu quantifizieren.

Vorbemerkungen

Grundlegende Definitionen

Z-Test

p-Werte

Konfidenzintervalle und Hypothesentests

Selbstkontrollfragen

Theorem (Dualität von Konfidenzintervallen und Hypothesentests I)

$X = (X_1, \dots, X_n) \sim p_\theta$ sei eine Stichprobe mit Ergebnisraum \mathcal{X} und Parameterraum Θ . Weiterhin sei $[G_u(X), G_o(X)]$ ein δ -Konfidenzintervall für θ .

Dann ist der Hypothesentest

$$\phi_\theta : \mathcal{X} \rightarrow \{0, 1\}, x \mapsto \phi(x) := \begin{cases} 0, & [G_u(x), G_o(x)] \ni \theta \\ 1, & [G_u(x), G_o(x)] \not\ni \theta \end{cases} \quad (51)$$

ein Test vom Signifikanzniveau $\alpha_0 = 1 - \delta$ für die Hypothesen

$$\Theta_0 := \{\theta\} \text{ und } \Theta_1 := \Theta \setminus \{\theta\}. \quad (52)$$

Beweis

Aufgrund der einfachen Nullhypothese und somit $\alpha_0 = \alpha$ folgt

$$\alpha_0 = \alpha = \mathbb{P}_\theta(\phi(X) = 1) = \mathbb{P}_\theta([G_u(x), G_o(x)] \not\ni \theta) = 1 - \mathbb{P}_\theta([G_u(x), G_o(x)] \ni \theta) = 1 - \delta. \quad (53)$$

Bemerkungen

- θ bezeichnet hier das Element der einfachen Nullhypothese
- Mit δ -Konfidenzintervallen kann man also Tests vom Signifikanzniveau $\alpha_0 = 1 - \delta$ konstruieren.

Konfidenzintervalle und Hypothesentests

Beispiel (Konstruktion eines Hypothesentests aus einem Konfidenzintervall)

Wir haben bereits gesehen, dass für eine Stichprobe $X_1, \dots, X_n \sim N(\mu_0, \sigma^2)$ mit unbekanntem Erwartungswertparameter μ_0 und bekanntem Varianzparameter $\sigma^2 > 0$, $\delta \in]0, 1[$ und $z_\delta := \Phi^{-1}\left(\frac{1+\delta}{2}\right)$,

$$C_n := \left[\bar{X}_n - \frac{\sigma}{\sqrt{n}} z_\delta, \bar{X}_n + \frac{\sigma}{\sqrt{n}} z_\delta \right]. \quad (54)$$

ein δ -Konfidenzintervall definiert.

Mit der Dualität von Konfidenzintervallen und Hypothesentests können wir also folgenden Test für die Hypothesen $\Theta_0 = \{\mu_0\}$ und $\Theta_1 = \mathbb{R} \setminus \mu_0$ definieren:

$$\phi : \mathcal{X} \rightarrow \{0, 1\}, x \mapsto \phi(x) := \begin{cases} 0, & \left[\bar{X}_n - \frac{\sigma}{\sqrt{n}} z_\delta, \bar{X}_n + \frac{\sigma}{\sqrt{n}} z_\delta \right] \ni \mu_0 \\ 1, & \left[\bar{X}_n - \frac{\sigma}{\sqrt{n}} z_\delta, \bar{X}_n + \frac{\sigma}{\sqrt{n}} z_\delta \right] \not\ni \mu_0 \end{cases} \quad (55)$$

Dann gilt

$$\begin{aligned} \mathbb{P}_{\mu_0}(\phi(X) = 1) &= 1 - \mathbb{P}_{\mu_0}(\phi(X) = 0) \\ &= 1 - \mathbb{P}_{\mu_0} \left(\left[\bar{X}_n - \frac{\sigma}{\sqrt{n}} z_\delta, \bar{X}_n + \frac{\sigma}{\sqrt{n}} z_\delta \right] \ni \mu_0 \right) \\ &= 1 - \delta. \end{aligned} \quad (56)$$

und wir haben gezeigt, dass ϕ ein Test vom Signifikanzniveau $\alpha_0 = 1 - \delta$ ist.

Konfidenzintervalle und Hypothesentests

Simulation

```
# Modellformulierung
mu_0 = 2 # w. a. u. Erwartungswertparameter/Nullhypothese
sigma = 1 # Standardabweichungsparameter
n = 12 # Stichprobengröße
delta = 0.95 # Konfidenzbedingung
phi_inv = qnorm((1+delta)/2) #  $\Phi^{-1}((\delta + 1)/2)$ 

# Simulationssetup
ns = 1e2 # Anzahl Simulationen
X_bar = rep(NA,n,ns) # Stichprobenmittelarray
C = matrix(rep(NA,n,2*ns), ncol = 2) # Konfidenzintervallarray
kfn = rep(NA,n,ns) # Überdeckungsindikatorarray
phi = rep(NA,n,ns) # Testarray

# Simulationsiterationen
for(i in 1:ns){

  # Datengeneration
  X = rnorm(n,mu_0,sigma) # Stichprobenrealisierung

  # Konfidenzintervallevaluation
  X_bar[i] = mean(X) # Stichprobenmittel
  C[i,1] = X_bar[i] - (sigma/sqrt(n))*phi_inv # untere KI Grenze
  C[i,2] = X_bar[i] + (sigma/sqrt(n))*phi_inv # obere KI Grenze

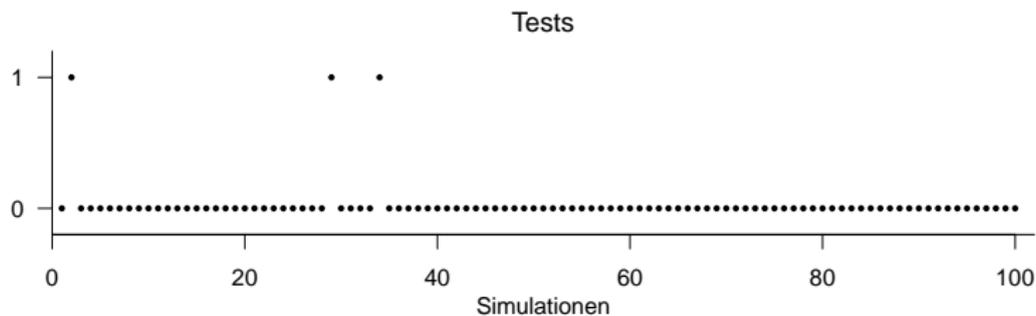
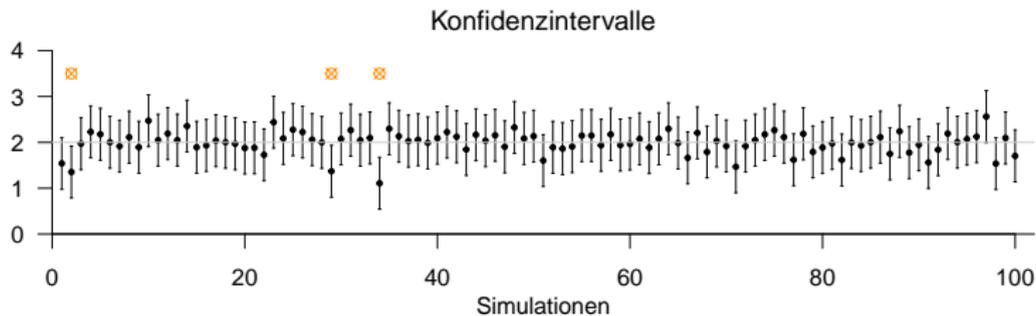
  # Überdeckungsindikatorevaluation
  if(C[i,1] <= mu_0 & mu_0 <= C[i,2]){
    kfn[i] = 1} else{
    kfn[i] = 0}

  # Testevaluation
  if(C[i,1] <= mu_0 & mu_0 <= C[i,2]){
    phi[i] = 0} else{
    phi[i] = 1}
}
cat("Geschätztes Konfidenzniveau = ", mean(kfn),
    "\nGeschätzter Testumfang = ", mean(phi))
```

```
> Geschätztes Konfidenzniveau = 0.97
> Geschätzter Testumfang = 0.03
```

Konfidenzintervalle und Hypothesentests

Simulation



Theorem (Dualität von Konfidenzintervallen und Hypothesentests II)

Unter den gleichen Annahmen wie im vorherigen Theorem sei

$$\Phi := \{\phi_\theta(X) | \theta \in \Theta\} \quad (57)$$

eine Familie von Tests, so dass $\phi_\theta(X)$ ein Test mit Signifikanzniveau α_0 für die Hypothesen

$$\Theta_0 := \{\theta\} \text{ and } \Theta_1 := \Theta \setminus \{\theta\}. \quad (58)$$

sei.

Man nehme weiter an, dass die Menge

$$C_n := \{\theta \in \Theta | \phi_\theta(X) = 0\} \quad (59)$$

als $C_n = [G_u(X), G_o(X)]$ für entsprechend bestimmte $G_u(X)$ und $G_o(X)$ geschrieben werden kann. Dann ist C_n ein $\delta := 1 - \alpha_0$ Konfidenzintervall für θ .

Beweis

Aufgrund der einfachen Nullhypothese und somit $\alpha_0 = \alpha$ folgt das Level $\delta = 1 - \alpha_0$ des Konfidenzintervalls aus

$$\delta = \mathbb{P}_\theta(C_n \ni \theta) = \mathbb{P}_\theta(\phi_\theta(X) = 0) = 1 - \mathbb{P}_\theta(\phi_\theta(X) = 1) = 1 - \alpha_0. \quad (60)$$

Bemerkung

- Mit Tests von Signifikanzniveau α_0 kann man also $1 - \alpha_0$ -Konfidenzintervalle konstruieren.

Vorbemerkungen

Grundlegende Definitionen

Z-Test

p-Werte

Konfidenzintervalle und Hypothesentests

Selbstkontrollfragen

Selbstkontrollfragen

1. Erläutern Sie die grundlegende Logik statistischer Hypothesentests.
2. Geben Sie die Definition statistischer Hypothesen und eines Testszenarios wieder.
3. Definieren Sie die Begriffe der einfachen und zusammengesetzten Hypothesen.
4. Definieren Sie die Begriffe der einseitigen und zweiseitigen Hypothesen.
5. Definieren Sie den Begriff des Tests.
6. Definieren Sie den Begriff des Standardtests.
7. Definieren Sie den Begriff des kritischen Bereichs eines Tests.
8. Definieren Sie den Begriff des 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 Signifikanzniveaus und des Level- α_0 -Tests.
14. Definieren Sie den Begriff des Testumfangs.
15. Erläutern Sie die prinzipielle Strategie zur Wahl von Null- und Alternativhypothesen in der Wissenschaft.
16. Nennen Sie vier Schritte in der Konstruktion eines Tests.

Selbstkontrollfragen

17. Definieren Sie das statistische Modell eines Z-Tests.
18. Definieren Sie die Hypothesen eines Z-Tests mit einfacher Nullhypothese und zweiseitiger Alternativhypothese.
19. Definieren Sie den zweiseitigen Z-Test.
20. Skizzieren Sie qualitativ Testgütefunktionen eines zweiseitigen Z-Tests für verschiedene kritische Werte.
21. Wie muss der kritische Wert eines zweiseitigen Z-Tests definiert sein, damit der Test ein Level- α_0 -Test ist?
22. Skizzieren Sie qualitativ die Bestimmung des kritischen Wertes k_{α_0} bei einem zweiseitigen Z-Test.
23. Erläutern Sie das praktische Vorgehen zur Durchführung eines zweiseitigen Z-Tests.
24. Von welchen Werten hängt die Powerfunktion eines zweiseitigen Z-Tests ab?
25. Skizzieren Sie qualitativ die Powerfunktion des zweiseitigen Z-Tests bei fester Stichprobengröße.
26. Skizzieren Sie qualitativ die Powerfunktion des zweiseitigen Z-Tests bei festem Erwartungswertparameter.
27. Erläutern Sie das favorisierte praktische Vorgehen zur Durchführung einer Poweranalyse.
28. Erläutern Sie die Motivation zur Auswertung von p-Werten.
29. Definieren Sie den Begriff des p-Werts.
30. Geben Sie das erste Theorem zur Dualität von Konfidenzintervallen und Hypothesentests wieder.
31. Geben Sie das zweite Theorem zur Dualität von Konfidenzintervallen und Hypothesentests wieder.
32. Erläutern Sie die Dualität von Konfidenzintervallen und Hypothesentests.

References

- DeGroot, Morris H., and Mark J. Schervish. 2012. *Probability and Statistics*. 4th ed. Boston: Addison-Wesley.
- Horvath, Lilla, Stanley Colcombe, Michael Milham, Shruti Ray, Philipp Schwartenbeck, and Dirk Ostwald. 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>.
- 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.



Wahrscheinlichkeitstheorie und Frequentistische Inferenz

BSc Psychologie WiSe 2021/22

Prof. Dr. Dirk Ostwald

(13) Einstichproben-T-Tests

T-Teststatistik

Einstichproben-T-Tests

Zweiseitige Einstichproben-T-Tests

Einseitige Einstichproben-T-Tests

Anwendungsbeispiel

Selbstkontrollfragen

T-Teststatistik

Einstichproben-T-Tests

Zweiseitige Einstichproben-T-Tests

Einseitige Einstichproben-T-Tests

Anwendungsbeispiel

Selbstkontrollfragen

Definition (T-Teststatistik)

$X_1, \dots, X_n \sim N(\mu, \sigma^2)$ sei die Stichprobe eines Normalverteilungmodells, \bar{X}_n bezeichne das Stichprobenmittel, S_n bezeichne die Stichprobenstandardabweichung, und es sei $\mu_0 \in \mathbb{R}$. Die *T-Teststatistik* ist definiert als

$$T := \sqrt{n} \left(\frac{\bar{X}_n - \mu_0}{S_n} \right). \quad (1)$$

Bemerkungen

- Im Gegensatz zur T-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 (systematischer Variabilität) zu Rauschen (unsystematischer Variabilität) verstanden werden:

$$\sqrt{\text{Stichprobengröße}} \left(\frac{\text{Signal}}{\text{Rauschen}} \right) = \sqrt{n} \left(\frac{\bar{X}_n - \mu_0}{S_n} \right) \quad (2)$$

- Die T-Teststatistik ist eine skalare Deskription des Effekt vs. Variabilität Verhältnisses eines Datensatzes.
- In der T-Teststatistik wird die Effektgröße in Einheiten der Stichprobenstandardabweichung gemessen:
 - $T = 1 \Leftrightarrow \sqrt{n}(\bar{X}_n - \mu_0) = 1S_n$
 - $T = 2 \Leftrightarrow \sqrt{n}(\bar{X}_n - \mu_0) = 2S_n$

Definition (Nichtzentrale t -Zufallsvariable)

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

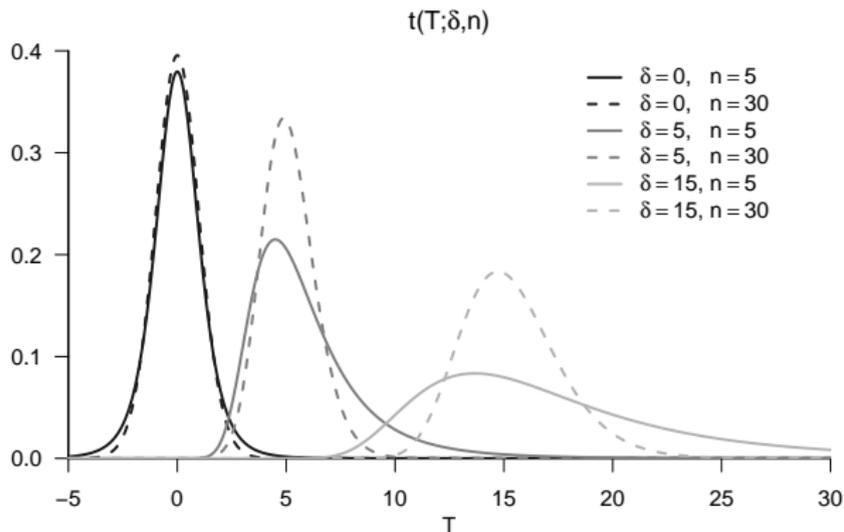
$$p : \mathbb{R} \rightarrow \mathbb{R}_{>0}, t \mapsto p(t) := \frac{1}{2^{\frac{n-1}{2}} \Gamma\left(\frac{n}{2}\right) (n\pi)^{\frac{1}{2}}} \times \int_0^\infty \tau^{\frac{n-1}{2}} \exp\left(-\frac{\tau}{2}\right) \exp\left(-\frac{1}{2} \left(t \left(\frac{\tau}{n}\right)^{\frac{1}{2}} - \delta\right)^2\right) d\tau. \quad (3)$$

Dann sagen wir, dass T einer nichtzentralen t -Verteilung mit Nichtzentralitätsparameter δ und Freiheitsgradparameter n unterliegt und nennen T eine *nichtzentrale t -Zufallsvariable mit Nichtzentralitätsparameter δ und Freiheitsgradparameter n* . Wir kürzen dies mit $t(\delta, n)$ ab. Die WDF einer nichtzentralen t -Zufallsvariable bezeichnen wir mit $t(T; \delta, n)$. Die KVF und inverse KVF einer nichtzentralen t -Zufallsvariable bezeichnen wir mit $\psi(\cdot; \delta, n)$ und $\psi^{-1}(\cdot; \delta, n)$, respektive.

Bemerkungen

- Eine nichtzentrale t -Zufallsvariable mit $\delta = 0$ ist eine t -Zufallsvariable.
- Es gilt also $t(T; 0, n) = t(T; n)$.
- Die funktionale Form der WDF findet sich zum Beispiel in Lehmann (1986), Seite 254, Gl. (80).

Wahrscheinlichkeitsdichtefunktionen nichtzentraler t -Verteilungen



Theorem (Nichtzentrale T-Transformation)

$X \sim N(\mu, 1)$ sei eine normalverteilte Zufallsvariable, $U \sim \chi^2(n)$ sei eine χ^2 Zufallsvariable mit Freiheitsgradparameter n , und X und U seien unabhängige Zufallsvariablen. Dann ist die Zufallsvariable

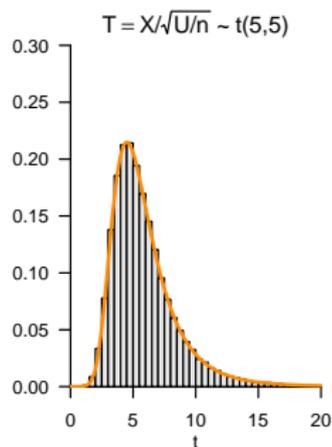
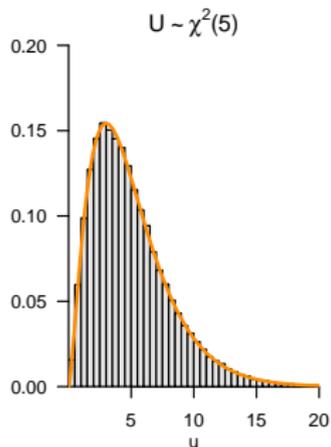
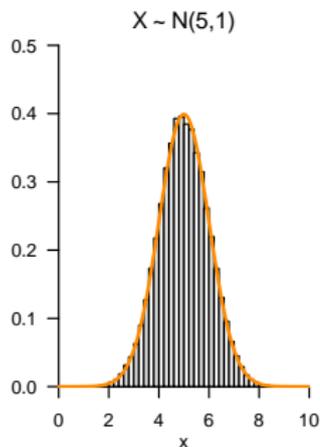
$$T := \frac{X}{\sqrt{U/n}} \quad (4)$$

eine nichtzentrale t -Zufallsvariable mit Nichtzentralitätsparameter μ und Freiheitsgradparameter n , es gilt also $T \sim t(\mu, n)$

Bemerkung

- Wir verzichten auf einen Beweis.

Nichtzentrale T-Transformation



Theorem (Verteilung der T-Teststatistik)

$X_1, \dots, X_n \sim N(\mu, \sigma^2)$ sei die Stichprobe eines Normalverteilungmodells, \bar{X}_n sei das Stichprobenmittel, S_n sei die Stichprobenstandardabweichung, und es sei $\mu_0 \in \mathbb{R}$. Dann ist die T-Teststatistik

$$T := \sqrt{n} \left(\frac{\bar{X}_n - \mu_0}{S_n} \right) \quad (5)$$

eine nichtzentrale t -Zufallsvariable mit Nichtzentralitätsparameter

$$d = \sqrt{n} \left(\frac{\mu - \mu_0}{\sigma} \right) \quad (6)$$

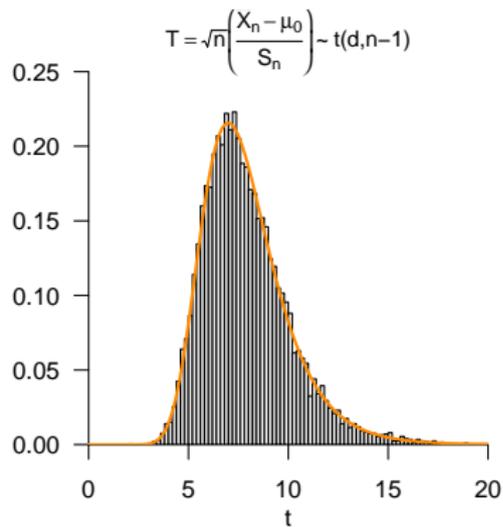
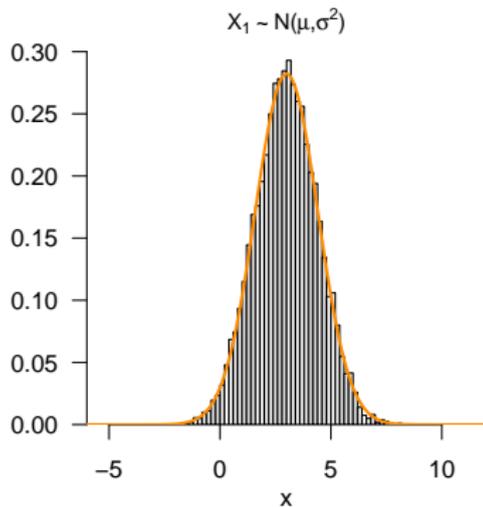
und Freiheitsgradparameter $n - 1$, es gilt also $T \sim t(d, n - 1)$

Bemerkung

- Wir verzichten auf einen Beweis.

T-Teststatistik

T-Teststatistik bei $n = 12, \mu = 3, \sigma^2 = 2, \mu_0 = 0$



T-Teststatistik

Einstichproben-T-Tests

Zweiseitige Einstichproben-T-Tests

Einseitige Einstichproben-T-Tests

Anwendungsbeispiel

Selbstkontrollfragen

Anwendungsszenario

- **Eine Stichprobe** experimenteller Einheiten.
- Annahme der unabhängigen und identischen Normalverteilung $N(\mu, \sigma^2)$ der Daten.
- μ und σ^2 unbekannt.
- Quantifizieren der Unsicherheit beim inferentiellen Vergleich von μ mit μ_0 beabsichtigt.

Anwendungsbeispiele

- Gruppenanalysen mit Wechsler Adult Intelligence Scale
 - $\mu \neq \mu_0 := 100 \Rightarrow$ Evidenz für über- oder unterdurchschnittliche WAIS Performanz
- BDI Score Datenanalyse einer Gruppe psychiatrischer Patient:innen
 - $\mu > \mu_0 := 18 \Rightarrow$ Klinisch auffällige Gruppe
- Gruppenanalysen in der funktionellen Kernspintomographie
 - $\mu > \mu_0 := 0 \Rightarrow$ Evidenz für regionale Gehirnaktivierung

Hypothesenszenarien

Einfache Nullhypothese, einfache Alternativhypothese $H_0 : \mu = \mu_0, H_1 : \mu = \mu_1$

- Theoretisch wichtiges Szenario (Neymann-Pearson Lemma)
- Praktische Relevanz eher gering

Einfache Nullhypothese, zusammengesetzte Alternativhypothese $H_0 : \mu = \mu_0, H_1 : \mu \neq \mu_0$

- Zweiseitiger Einstichproben-T-Test mit ungerichteter Hypothese
- Ungerichtete Fragestellung nach einem Unterschied

Zusammengesetzte Nullhypothese/Alternativhypothese $H_0 : \mu \leq \mu_0, H_1 : \mu > \mu_0$

- Einseitiger Einstichproben-T-Test mit gerichteter Hypothese
- Gerichtete Fragestellung nach einem positiven Unterschied

Zusammengesetzte Nullhypothese/Alternativhypothese $H_0 : \mu \geq \mu_0, H_1 : \mu < \mu_0$

- Gerichtete Fragestellung nach einem negativen Unterschied
- Qualitativ äquivalente Theorie zum umgekehrten Fall

Im Folgenden näher betrachtete Hypothesenszenarien

Einfache Nullhypothese, einfache Alternativhypothese $H_0 : \mu = \mu_0, H_1 : \mu = \mu_1$

- Theoretisch wichtiges Szenario (Neymann-Pearson Lemma)
- Praktische Relevanz eher gering

Einfache Nullhypothese, zusammengesetzte Alternativhypothese $H_0 : \mu = \mu_0, H_1 : \mu \neq \mu_0$

- Zweiseitiger Einstichproben-T-Test mit ungerichteter Hypothese
- Ungerichtete Fragestellung nach einem Unterschied

Zusammengesetzte Nullhypothese/Alternativhypothese $H_0 : \mu \leq \mu_0, H_1 : \mu > \mu_0$

- Einseitiger Einstichproben-T-Test mit gerichteter Hypothese
- Gerichtete Fragestellung nach einem positiven Unterschied

Zusammengesetzte Nullhypothese/Alternativhypothese $H_0 : \mu \geq \mu_0, H_1 : \mu < \mu_0$

- Gerichtete Fragestellung nach einem negativen Unterschied
- Qualitativ äquivalente Theorie zum umgekehrten Fall

Gliederung

- (1) Statistisches Modell in klassischer Form
- (2) Statistisches Modell in generativer Form
- (3) Testhypothesen, Teststatistik, Test
- (4) Analyse der Testgütefunktion
- (5) Testumfangkontrolle
- (6) p-Werte
- (7) Analyse der Powerfunktion

T-Teststatistik

Einstichproben-T-Tests

Zweiseitige Einstichproben-T-Tests

Einseitige Einstichproben-T-Tests

Anwendungsbeispiel

Selbstkontrollfragen

(1) Statistisches Modell in klassischer Form

$X_1, \dots, X_n \sim N(\mu, \sigma^2)$ sei die Stichprobe eines Normalverteilungsmodells mit unbekanntem Erwartungswertparameter μ und unbekanntem Varianzparameter $\sigma^2 > 0$. Als Parameter von Interesse betrachten wir $\theta = \mu$, so dass sich der Parameterraum von Interesse zu $\Theta = \mathbb{R}$ ergibt.

(2) Statistisches Modell in generativer Form

Es sei

$$X_i = \mu + \varepsilon_i \text{ mit } \varepsilon_i \sim N(0, \sigma^2) \text{ f\"ur } i = 1, \dots, n \quad (7)$$

wobei

- $X_i, i = 1, \dots, n$ beobachtbare Zufallsvariablen,
- $\mu \in \mathbb{R}$ den festen identischen Erwartungswertparameter über Stichprobenvariablen, und
- $\varepsilon_i, i = 1, \dots, n$ unabhängige normalverteilte nicht-beobachtbare Zufallsvariablen

bezeichnen.

Zweiseitige Einstichproben-T-Tests

(2) Statistisches Modell in generativer Form (fortgeführt)

Die generative Form betont, dass im vorliegenden Modell beobachtete Daten durch einen systematischen deterministischen Prozess (hier μ) unter dem additiven Einfluss einer Vielzahl unabhängiger und deshalb in der Summe normalverteilter Störprozesse (hier $\varepsilon_i, i = 1, \dots, n$) erzeugt konzipiert werden.

Beweis

Die Äquivalenz beider Modellformen folgt direkt aus der Transformation normalverteilter Zufallsvariablen durch linear-affine Funktionen (cf. (6) Transformationen der Normalverteilung). Speziell gilt im vorliegenden Fall für $\varepsilon_i \sim N(0, \sigma^2)$, dass

$$X_i = f(\varepsilon_i) \text{ mit } f: \mathbb{R} \rightarrow \mathbb{R}, \varepsilon_i \mapsto f(\varepsilon_i) := \varepsilon_i + \mu. \quad (8)$$

Mit den WDF Transformationstheorem bei linear-affinen Abbildungen folgt dann

$$\begin{aligned} p_{X_i}(x_i) &= \frac{1}{|1|} p_{\varepsilon_i} \left(\frac{x_i - \mu}{1} \right) \\ &= N(x_i - \mu; 0, \sigma^2) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left(-\frac{1}{2\sigma^2} (x_i - \mu - 0)^2 \right) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left(-\frac{1}{2\sigma^2} (x_i - \mu)^2 \right) \\ &= N(x_i; \mu, \sigma^2), \end{aligned} \quad (9)$$

also $X_i \sim N(\mu, \sigma^2)$.

(3) Testhypothesen, Teststatistik, Test

Für ein $\mu_0 \in \mathbb{R}$ betrachten wir die einfache Nullhypothese und die zusammengesetzte Alternativhypothese

$$H_0 : \mu = \mu_0 \Leftrightarrow \Theta_0 := \{\mu_0\} \text{ und } H_1 : \mu \neq \mu_0 \Leftrightarrow \Theta_1 := \mathbb{R} \setminus \{\mu_0\}, \quad (10)$$

respektive. Weiterhin betrachten wir die T-Teststatistik

$$T := \sqrt{n} \left(\frac{\bar{X}_n - \mu_0}{S_n} \right) \quad (11)$$

und definieren den zweiseitigen kritischen Wert-basierten Test

$$\phi(X) := 1_{\{|T| \geq k\}} = \begin{cases} 1 & |T| \geq k \\ 0 & |T| < k \end{cases}. \quad (12)$$

(4) Analyse der Testgütefunktion

Theorem (Testgütefunktion)

ϕ sei der im obigen Testscenario definierte Test. Dann ist die Testgütefunktion von ϕ gegeben durch

$$q_\phi : \mathbb{R} \rightarrow [0, 1], \mu \mapsto q_\phi(\mu) := 1 - \psi(k; d_\mu, n - 1) + \psi(-k; d_\mu, n - 1) \quad (13)$$

wobei $\psi(\cdot; d_\mu, n - 1)$ die KVF der nichtzentralen t -Verteilung mit Nichtzentralitätsparameter

$$d_\mu := \sqrt{n} \left(\frac{\mu - \mu_0}{\sigma} \right) \quad (14)$$

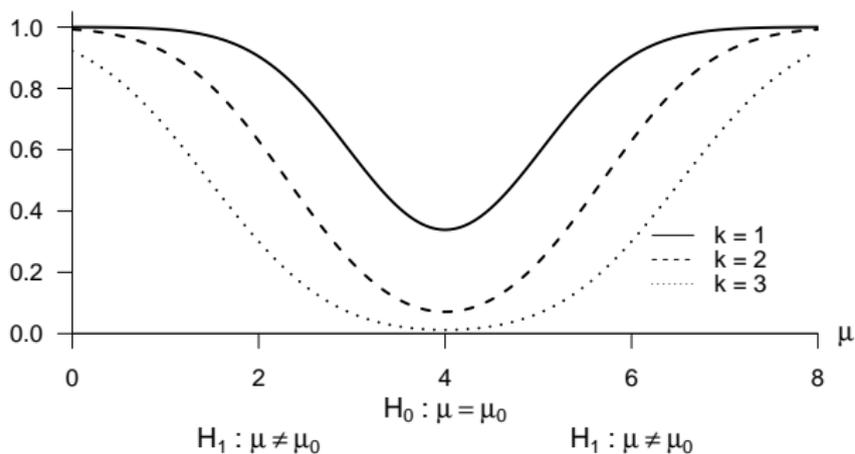
und Freiheitsgradparameter $n - 1$ bezeichnet.

Zweiseitige Einstichproben-T-Tests

(4) Analyse der Testgütefunktion

Testgütefunktion q_ϕ für $\sigma^2 = 9$, $\mu_0 = 4$, $n = 12$ und $k = 1, 2, 3$.

$$q_\phi(\mu) = \mathbb{P}_\mu(\phi = 1)$$



(4) Analyse der Testgütefunktion

Beweis

Die Testgütefunktion des betrachteten Test im vorliegenden Testszenario ist definiert als

$$q_\phi : \mathbb{R} \rightarrow [0, 1], \mu \mapsto q_\phi(\mu) := \mathbb{P}_\mu(\phi = 1). \quad (15)$$

Da die Wahrscheinlichkeiten für $\phi = 1$ und dafür, dass die zugehörige Teststatistik im Ablehnungsbereich des Tests liegt gleich sind, benötigen wir die also zunächst die Verteilung der Teststatistik. Wir haben oben bereits gesehen, dass die T-Teststatistik

$$T := \sqrt{n} \left(\frac{\bar{X}_n - \mu_0}{S_n} \right) \quad (16)$$

unter der Annahme $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ nach einer nichtzentralen t -Verteilung $t(d, n - 1)$ mit Nichtzentralitätsparameter

$$d = \sqrt{n} \left(\frac{\mu - \mu_0}{\sigma} \right) \quad (17)$$

verteilt ist. Der Ablehnungsbereich des zweiseitigen T-Tests ergibt sich, wie in ähnlicher Form bei der Betrachtung des zweiseitigen Z-Tests gesehen, zu

$$A =] - \infty, -k] \cup]k, \infty[. \quad (18)$$

(4) Analyse der Testgütefunktion

Beweis (fortgeführt)

Mit diesem Ablehnungsbereich ergibt sich dann

$$\begin{aligned}q_{\phi}(\mu) &= \mathbb{P}_{\mu}(\phi = 1) \\&= \mathbb{P}_{\mu}(T \in]-\infty, -k] \cup]k, \infty[) \\&= \mathbb{P}_{\mu}(T \in]-\infty, -k]) + \mathbb{P}_{\mu}(T \in [k, \infty[) \\&= \mathbb{P}_{\mu}(T \leq -k) + \mathbb{P}_{\mu}(T \geq k) \\&= \mathbb{P}_{\mu}(T \leq -k) + (1 - \mathbb{P}_{\mu}(T \leq k)) \\&= 1 - \mathbb{P}_{\mu}(T \leq k) + \mathbb{P}_{\mu}(T \leq -k) \\&= 1 - \psi(k; d_{\mu}, n - 1) + \psi(-k; d_{\mu}, n - 1),\end{aligned}\tag{19}$$

wobei $\psi(\cdot; d_{\mu}, n - 1)$ die KVF der nichtzentralen T-Verteilung mit Nichtzentralitätsparameter d_{μ} und Freiheitsgradparameter $n - 1$ bezeichnet.

□

(5) Testumfangkontrolle

Theorem (Testumfangkontrolle)

ϕ sei der im obigen Testszenario definierte Test. Dann ist ϕ ein Level- α_0 -Test mit Testumfang α_0 , wenn der kritische Wert definiert ist durch

$$k_{\alpha_0} := \psi^{-1} \left(1 - \frac{\alpha_0}{2}; n - 1 \right), \quad (20)$$

wobei $\psi^{-1}(\cdot; n - 1)$ die inverse KVF der t -Verteilung mit $n - 1$ Freiheitsgraden ist.

Zweiseitige Einstichproben-T-Tests

(5) Testumfangkontrolle

Beweis

Damit der betrachtete Test ein Level- α_0 -Test ist, muss bekanntlich $q_\phi(\mu) \leq \alpha_0$ für alle $\mu \in \{\mu_0\}$, also hier $q_\phi(\mu_0) \leq \alpha_0$, gelten. Weiterhin ist der Testumfang des betrachteten Tests durch $\alpha = \max_{\mu \in \{\mu_0\}} q_\phi(\mu)$, also hier durch $\alpha = q_\phi(\mu_0)$ gegeben. Wir müssen also zeigen, dass die Wahl von k_{α_0} garantiert, dass ϕ ein Level- α_0 -Test mit Testumfang α_0 ist. Dazu merken wir zunächst an, dass für $\mu = \mu_0$ gilt, dass

$$\begin{aligned}q_\phi(\mu_0) &= 1 - \psi(k; d_{\mu_0}, n - 1) + \psi(-k; d_{\mu_0}, n - 1) \\ &= 1 - \psi(k; 0, n - 1) + \psi(-k; 0, n - 1) \\ &= 1 - \psi(k; n - 1) + \psi(-k; n - 1),\end{aligned}\tag{21}$$

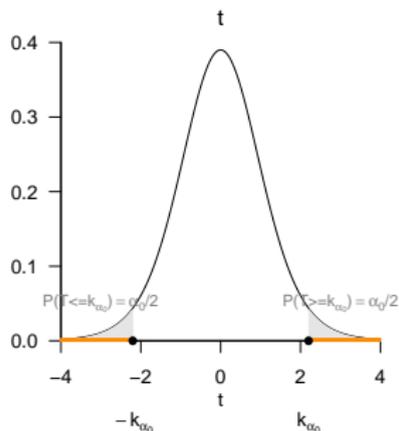
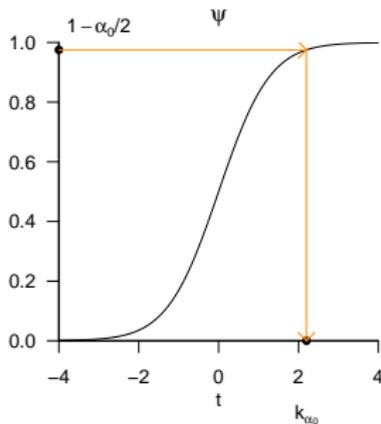
wobei $\psi(\cdot; d, n - 1)$ und $\psi(\cdot; n - 1)$ die KVF der nichtzentralen t -Verteilung mit Nichtzentralitätsparameter d und Freiheitsgradparameter $n - 1$ sowie der t -Verteilung mit Freiheitsgradparameter $n - 1$, respektive, bezeichnen. Sei nun also $k := k_{\alpha_0}$. Dann gilt

$$\begin{aligned}q_\phi(\mu_0) &= 1 - \psi(k_{\alpha_0}, n - 1) + \psi(-k_{\alpha_0}, n - 1) \\ &= 1 - \psi(k_{\alpha_0}, n - 1) + (1 - \psi(k_{\alpha_0}, n - 1)) \\ &= 2(1 - \psi(k_{\alpha_0}, n - 1)) \\ &= 2 \left(1 - \psi \left(\psi^{-1}(1 - \alpha_0/2, n - 1), n - 1 \right) \right) \\ &= 2(1 - 1 + \alpha_0/2) \\ &= \alpha_0,\end{aligned}\tag{22}$$

wobei die zweite Gleichung mit der Symmetrie der t -Verteilung folgt. Es folgt also direkt, dass bei der Wahl von $k = k_{\alpha_0}$, $q_\phi(\mu_0) \leq \alpha_0$ ist und der betrachtete Test somit ein Level- α_0 -Test ist. Weiterhin folgt direkt, dass der Testumfang des betrachteten Tests bei der Wahl von $k = k_{\alpha_0}$ gleich α_0 ist.

(5) Testumfangkontrolle

Wahl von $k_{\alpha_0} := \psi^{-1}(1 - \alpha_0/2; n - 1)$ mit $n = 12$, $\alpha_0 := 0.05$ und Ablehnungsbereich



(5) Testumfangkontrolle

Praktisches Vorgehen

- Man nimmt an, dass ein vorliegender Datensatz x_1, \dots, x_n eine Realisation von $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ mit unbekanntem Parametern μ und $\sigma^2 > 0$ ist.
- Man möchte entscheiden ob für ein $\mu_0 \in \mathbb{R}$ eher $H_0 : \mu = \mu_0$ oder $H_1 : \mu \neq \mu_0$ zutrifft.
- Man wählt ein Signifikanzniveau α_0 und bestimmt den zugehörigen Freiheitsgradparameter-abhängigen kritischen Wert k_{α_0} . Zum Beispiel gilt bei Wahl von $\alpha_0 := 0.05$ und $n = 12$, also Freiheitsgradparameter 11, dass $k_{0.05} = \psi^{-1}(1 - 0.05/2; 11) \approx 2.20$ ist.
- Anhand von n, μ_0, \bar{x}_n und s_n berechnet man die Realisierung der T-Teststatistik

$$t := \sqrt{n} \left(\frac{\bar{x}_n - \mu_0}{s_n} \right) \quad (23)$$

- Wenn t größer-gleich k_{α_0} ist oder wenn t kleiner-gleich $-k_{\alpha_0}$ ist, lehnt man die Nullhypothese ab, andernfalls lehnt man sie nicht ab.
- Die oben entwickelte Theorie garantiert dann, dass man in höchstens $\alpha_0 \cdot 100$ von 100 Fällen die Nullhypothese fälschlicherweise ablehnt.

(6) p-Werte

Bestimmung des p-Wertes

- Per Definition ist der p-Wert das kleinste Signifikanzlevel α_0 , bei welchem man die Nullhypothese basierend auf einem vorliegendem Wert der Teststatistik ablehnen würde.
- Bei $T = t$ würde H_0 für jedes α_0 mit $|t| \geq \psi^{-1}(1 - \alpha_0/2; n - 1)$ abgelehnt werden. Für diese α_0 gilt, wie unten gezeigt,

$$\alpha_0 \geq 2\mathbb{P}(T \geq |t|). \quad (24)$$

- Das kleinste $\alpha_0 \in [0, 1]$ mit $\alpha_0 \geq 2\mathbb{P}(T \geq |t|)$ ist dann $\alpha_0 = 2\mathbb{P}(T \geq |t|)$, also folgt

$$\text{p-Wert} = 2\mathbb{P}(T \geq |t|) = 2(1 - \psi(|t|; n - 1)). \quad (25)$$

- Im Gegensatz zum Z-Test hängt bei T-Tests der p-Wert auch von der Stichprobengröße ab.
- Zum Beispiel ist für $T = 2.00$ und $n = 10$ der p-Wert 0.076, für $T = 2.00$ und $n = 100$ ist der p-Wert dagegen 0.048.

Zweiseitige Einstichproben-T-Tests

Bestimmung des p-Wertes

- Es bleibt zu zeigen, dass gilt

$$|t| \geq \psi^{-1}(1 - \alpha_0/2; n - 1) \Leftrightarrow \alpha_0 \geq 2\mathbb{P}(T \geq |t|) \quad (26)$$

- Dies aber folgt aus

$$\begin{aligned} & |t| \geq \psi^{-1}\left(1 - \frac{\alpha_0}{2}; n - 1\right) \\ \Leftrightarrow & \psi(|t|; n - 1) \geq \psi\left(\psi^{-1}\left(1 - \frac{\alpha_0}{2}; n - 1\right); n - 1\right) \\ \Leftrightarrow & \psi(|t|; n - 1) \geq 1 - \frac{\alpha_0}{2} \\ \Leftrightarrow & \mathbb{P}(T \leq |t|) \geq 1 - \frac{\alpha_0}{2} \\ \Leftrightarrow & \frac{\alpha_0}{2} \geq 1 - \mathbb{P}(T \leq |t|) \\ \Leftrightarrow & \frac{\alpha_0}{2} \geq \mathbb{P}(T \geq |t|) \\ \Leftrightarrow & \alpha_0 \geq 2\mathbb{P}(T \geq |t|). \end{aligned} \quad (27)$$

(7) Analyse der Powerfunktion

Wir betrachten die Testgütfunktion

$$q_\phi : \mathbb{R} \rightarrow [0, 1], \mu \mapsto q_\phi(\mu) := 1 - \psi(k_{\alpha_0}; d_\mu, n - 1) + \psi(-k_{\alpha_0}; d_\mu, n - 1) \quad (28)$$

bei kontrolliertem Testumfang, also für $k_{\alpha_0} := \psi^{-1}(1 - \alpha_0/2; n - 1)$ mit festem α_0 als Funktion des Nichtzentralitätsparameters und des Stichprobenumfangs. Namentlich hängt hier k_{α_0} auch von n ab.

Es ergibt sich die bivariate reellwertige Funktion

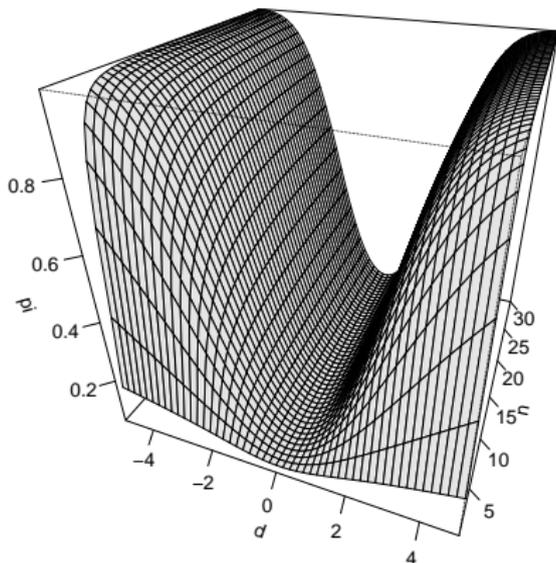
$$\pi : \mathbb{R} \times \mathbb{N} \rightarrow [0, 1], (d, n) \mapsto \pi(d, n) := 1 - \psi(k_{\alpha_0}; d, n - 1) + \psi(-k_{\alpha_0}; d, n - 1) \quad (29)$$

Bei festgelegten α_0 hängt die Powerfunktion des zweiseitigen T-Tests mit einfacher Nullhypothese also vom unbekanntem Wert d und von der Stichprobengröße n ab. Wir visualisieren diese Abhängigkeiten untenstehend.

Zweiseitige Einstichproben-T-Tests

(7) Analyse der Powerfunktion

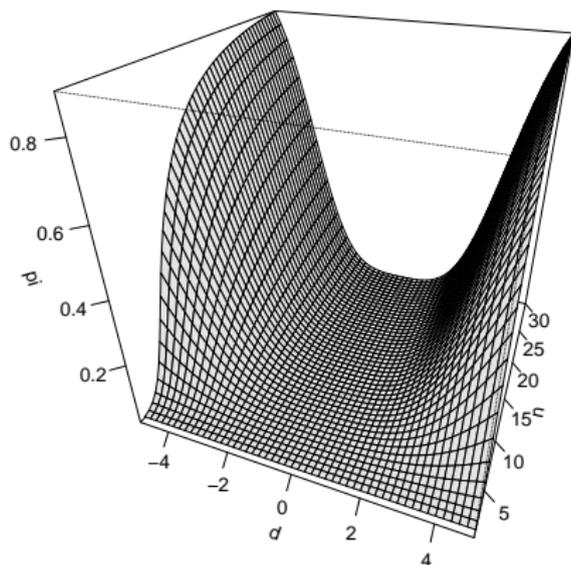
Powerfunktion für $\alpha_0 = 0.05$



Zweiseitige Einstichproben-T-Tests

(7) Analyse der Powerfunktion

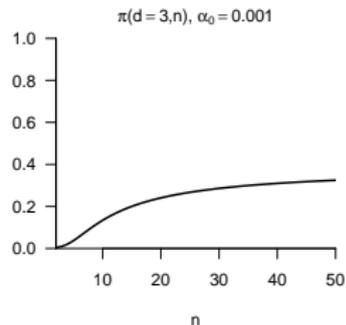
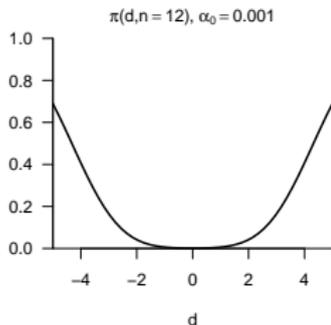
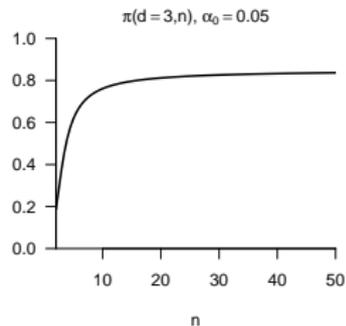
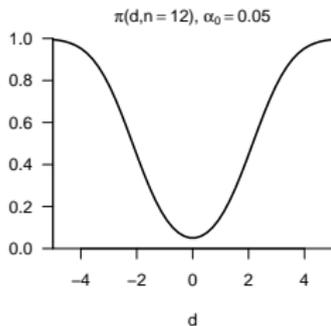
Powerfunktion für $\alpha_0 = 0.001$



Zweiseitige Einstichproben-T-Tests

(7) Analyse der Powerfunktion

Powerfunktionen für $\mu_0 = 0$



Zweiseitige Einstichproben-T-Tests

(7) Analyse der Powerfunktion

Praktisches Vorgehen

Mit größerem n steigt die Powerfunktion des Tests an

- Ein großer Stichprobenumfang ist besser als ein kleiner Stichprobenumfang.
- Kosten für die Erhöhung des Stichprobenumfangs werden aber nicht berücksichtigt.

⇒ Die Theorie statistischer Hypothesentests ist nicht besonders lebensnah.

Die Powerfunktion hängt vom wahren, aber unbekanntem, Parameterwert $d = \sqrt{n}(\mu - \mu_0)/\sigma$ ab.

⇒ Wenn man d schon kennen würde, würde man den Test nicht durchführen.

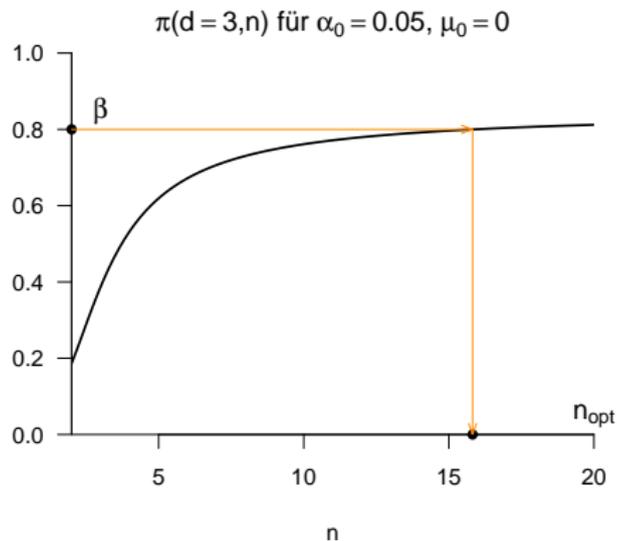
Generell wird folgendes Vorgehen favorisiert

- Man legt das Signifikanzniveau α_0 fest und evaluiert die Powerfunktion.
- Man wählt einen Mindestparameterwert d^* , den man mit $\pi(d, n) = \beta$ detektieren möchte.
- Ein konventioneller Wert ist $\beta = 0.8$.
- Man liest die für $\pi(d = d^*, n) = \beta$ nötige Stichprobengröße n ab.

Zweiseitige Einstichproben-T-Tests

(7) Analyse der Powerfunktion

Praktisches Vorgehen



T-Teststatistik

Einstichproben-T-Tests

Zweiseitige Einstichproben-T-Tests

Einseitige Einstichproben-T-Tests

Anwendungsbeispiel

Selbstkontrollfragen

(1) Statistisches Modell in klassischer Form

$X_1, \dots, X_n \sim N(\mu, \sigma^2)$ sei die Stichprobe eines Normalverteilungsmodells mit unbekanntem Erwartungswertparameter μ und unbekanntem Varianzparameter $\sigma^2 > 0$. Als Parameter von Interesse betrachten wir $\theta = \mu$, so dass sich der Parameterraum von Interesse zu $\Theta = \mathbb{R}$ ergibt.

(2) Statistisches Modell in generativer Form

Es sei

$$X_i = \mu + \varepsilon_i \text{ mit } \varepsilon_i \sim N(0, \sigma^2) \text{ f\"ur } i = 1, \dots, n \quad (30)$$

wobei

- $X_i, i = 1, \dots, n$ beobachtbare Zufallsvariablen,
- $\mu \in \mathbb{R}$ den festen identischen Erwartungswertparameter über Stichprobenvariablen, und
- $\varepsilon_i, i = 1, \dots, n$ unabhängige normalverteilte nicht-beobachtbare Zufallsvariablen

bezeichnen.

(3) Testhypothesen, Teststatistik, Test

Für ein $\mu_0 \in \mathbb{R}$ betrachten wir die zusammengesetzte Nullhypothese

$$H_0 : \mu \leq \mu_0 \Leftrightarrow \Theta_0 :=] - \infty, \mu_0] \quad (31)$$

und die zusammengesetzte Alternativhypothese

$$H_1 : \mu > \mu_0 \Leftrightarrow \Theta_1 :=]\mu_0, \infty[. \quad (32)$$

Weiterhin betrachten wir die T-Teststatistik

$$T := \sqrt{n} \left(\frac{\bar{X}_n - \mu_0}{S_n} \right) \quad (33)$$

und definieren den einseitigen kritischen Wert-basierten Test

$$\phi(X) := 1_{\{T \geq k\}} = \begin{cases} 1 & T \geq k \\ 0 & T < k \end{cases}. \quad (34)$$

(4) Analyse der Testgütefunktion

Theorem (Testgütefunktion)

ϕ sei der im obigen Testszenario definierte Test. Dann ist die Testgütefunktion von ϕ gegeben durch

$$q_\phi : \mathbb{R} \rightarrow [0, 1], \mu \mapsto q_\phi(\mu) := 1 - \psi(k; d_\mu, n - 1) \quad (35)$$

wobei $\psi(\cdot; d_\mu, n - 1)$ die KVF der nichtzentralen t -Verteilung mit Nichtzentralitätsparameter

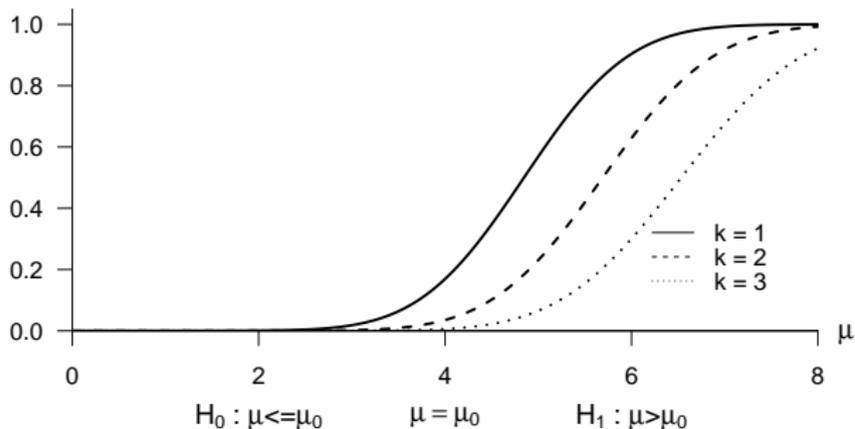
$$d_\mu := \sqrt{n} \left(\frac{\mu - \mu_0}{\sigma} \right) \quad (36)$$

und Freiheitsgradparameter $n - 1$ bezeichnet.

(4) Analyse der Testgütefunktion

Testgütefunktion q_ϕ für $\sigma^2 = 9$, $\mu_0 = 4$, $n = 12$ und $k = 1, 2, 3$.

$$q_\phi(\mu) = \mathbb{P}_\mu(\phi = 1)$$



(4) Analyse der Testgütefunktion

Beweis

Die Testgütefunktion des betrachteten Test im vorliegenden Testszenario ist definiert als

$$q_\phi : \mathbb{R} \rightarrow [0, 1], \mu \mapsto q_\phi(\mu) := \mathbb{P}_\mu(\phi = 1). \quad (37)$$

Da die Wahrscheinlichkeiten für $\phi = 1$ und dafür, dass die zugehörige Teststatistik im Ablehnungsbereich des Tests liegt gleich sind, benötigen wir die also zunächst die Verteilung der Teststatistik. Wir haben oben bereits gesehen, dass die T-Teststatistik

$$T := \sqrt{n} \left(\frac{\bar{X}_n - \mu_0}{S_n} \right) \quad (38)$$

unter der Annahme $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ nach einer nichtzentralen t -Verteilung $t(d, n - 1)$ mit Nichtzentralitätsparameter

$$d = \sqrt{n} \left(\frac{\mu - \mu_0}{\sigma} \right) \quad (39)$$

verteilt ist. Mit dem Ablehnungsbereich des einseitigen T-Tests $A :=]k, \infty[$ ergibt sich dann

$$q_\phi(\mu) = \mathbb{P}_\mu(\phi = 1) = \mathbb{P}_\mu(T \in]k, \infty[) = \mathbb{P}_\mu(T \geq k) = 1 - \mathbb{P}_\mu(T \leq k) = 1 - \psi(k; d, n - 1), \quad (40)$$

wobei $\psi(\cdot; d, n - 1)$ weiterhin die KVF der nichtzentralen T-Verteilung mit Nichtzentralitätsparameter d und Freiheitsgradparameter $n - 1$ bezeichnet.

(5) Testumfangkontrolle

Theorem (Testumfangkontrolle)

ϕ sei der im obigen Szenario definierte Test. Dann ist ϕ ein Level- α_0 -Test mit Testumfang α_0 , wenn der kritische Wert k definiert ist durch

$$k_{\alpha_0} := \psi^{-1}(1 - \alpha_0; n - 1), \quad (41)$$

wobei $\psi^{-1}(\cdot; n - 1)$ die inverse KVF der t -Verteilung mit $n - 1$ Freiheitsgraden ist.

(5) Testumfangkontrolle

Beweis

Damit der betrachtete Test ein Level- α_0 -Test ist, muss bekanntlich $q_\phi(\mu) \leq \alpha_0$ für alle $\mu \in]-\infty, \mu_0]$ gelten. Weiterhin ist der Testumfang des betrachteten Tests durch $\alpha = \max_{]-\infty, \mu_0]} q_\phi(\mu)$ gegeben. Mit der Monotonie von q_ϕ , die wir hier als selbstevident voraussetzen, gilt $\alpha = q_\phi(\mu_0)$. Wir müssen also zeigen, dass die Wahl von k_{α_0} garantiert, dass ϕ ein Level- α_0 -Test mit Testumfang α_0 ist. Dazu merken wir zunächst an, dass für $\mu = \mu_0$ gilt, dass

$$\begin{aligned}q_\phi(\mu_0) &= 1 - \psi(k; d_{\mu_0}, n - 1) \\ &= 1 - \psi(k; 0, n - 1) \\ &= 1 - \psi(k; n - 1)\end{aligned}\tag{42}$$

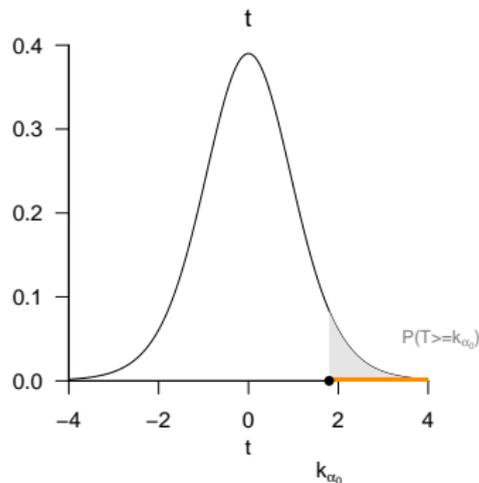
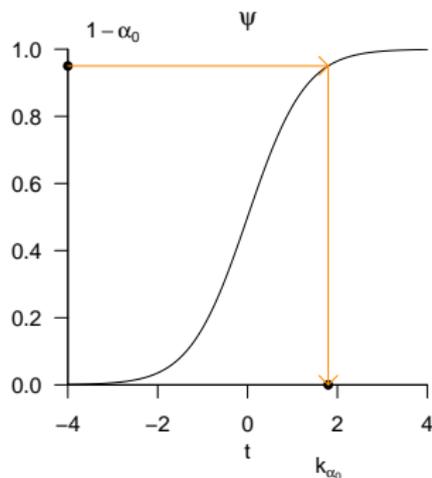
wobei $\psi(\cdot; d, n - 1)$ und $\psi(\cdot; n - 1)$ die KVF der nichtzentralen t -Verteilung mit Nichtzentralitätsparameter d und Freiheitsgradparameter $n - 1$ sowie der t -Verteilung mit Freiheitsgradparameter $n - 1$, respektive, bezeichnen. Sei nun also $k := k_{\alpha_0}$. Dann gilt

$$\begin{aligned}q_\phi(\mu_0) &= 1 - \psi(k_{\alpha_0}) \\ &= 1 - \psi\left(\psi^{-1}(1 - \alpha_0)\right) \\ &= 1 - 1 + \alpha_0 \\ &= \alpha_0,\end{aligned}\tag{43}$$

Es folgt also direkt, dass bei der Wahl von $k = k_{\alpha_0}$, $q_\phi(\mu_0) \leq \alpha_0$ ist und der betrachtete Test somit ein Level- α_0 -Test ist. Weiterhin folgt direkt, dass der Testumfang des betrachteten Tests bei der Wahl von $k = k_{\alpha_0}$ gleich α_0 ist.

(5) Testumfangkontrolle

Wahl von $k_{\alpha_0} := \psi^{-1}(1 - \alpha_0; n - 1)$ mit $n = 12$ $\alpha_0 := 0.05$ und resultierender Ablehnungsbereich



(5) Testumfangkontrolle

Praktisches Vorgehen

- Man nimmt an, dass ein vorliegender Datensatz x_1, \dots, x_n eine Realisation von $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ mit unbekanntem Parametern μ und $\sigma^2 > 0$ ist.
- Man möchte entscheiden ob für ein $\mu_0 \in \mathbb{R}$ eher $H_0 : \mu \leq \mu_0$ oder $H_1 : \mu > \mu_0$ zutrifft.
- Man wählt ein Signifikanzniveau α_0 und bestimmt den zugehörigen Freiheitsgradparameter-abhängigen kritischen Wert k_{α_0} . Zum Beispiel gilt bei Wahl von $\alpha_0 := 0.05$ und $n = 12$, also Freiheitsgradparameter 11, dass $k_{0.05} = \psi^{-1}(1 - 0.05; 11) \approx 1.80$ ist.
- Anhand von n, μ_0, \bar{x}_n und s_n berechnet man die Realisierung der T-Teststatistik

$$t := \sqrt{n} \left(\frac{\bar{x}_n - \mu_0}{s_n} \right) \quad (44)$$

- Wenn t größer-gleich k_{α_0} ist, lehnt man die Nullhypothese ab, andernfalls nicht.
- Die oben entwickelte Theorie garantiert dann, dass man in höchstens $\alpha_0 \cdot 100$ von 100 Fällen die Nullhypothese fälschlicherweise ablehnt.

(6) p-Werte

Bestimmung des p-Wertes

- Per Definition ist der p-Wert das kleinste Signifikanzlevel α_0 , bei welchem man die Nullhypothese basierend auf einem vorliegendem Wert der Teststatistik ablehnen würde.
- Bei $T = t$ würde H_0 für jedes α_0 mit $t \geq \psi^{-1}(1 - \alpha_0; n - 1)$ abgelehnt werden. Für diese α_0 gilt, wie unten gezeigt,

$$\alpha_0 \geq \mathbb{P}(T \geq t). \quad (45)$$

- Das kleinste $\alpha_0 \in [0, 1]$ mit $\alpha_0 \geq \mathbb{P}(T \geq t)$ ist dann $\alpha_0 = \mathbb{P}(T \geq t)$, also folgt

$$\text{p-Wert} = \mathbb{P}(T \geq t) = 1 - \psi(t; n - 1). \quad (46)$$

- Im Gegensatz zum zweiseitigen T-Tests ergeben sich für negative Werte von t sehr große p-Werte.
- Zum Beispiel ist für $T = 2.00$ und $n = 10$ der p-Wert 0.038, für $T = -2.00$ und $n = 10$ ist der p-Wert dagegen 0.961.

(6) p-Werte

Bestimmung des p-Wertes

- Es bleibt zu zeigen, dass gilt

$$t \geq \psi^{-1}(1 - \alpha_0; n - 1) \Leftrightarrow \alpha_0 \geq 2\mathbb{P}(T \geq t) \quad (47)$$

- Dies aber folgt aus

$$\begin{aligned} & t \geq \psi^{-1}(1 - \alpha_0; n - 1) \\ \Leftrightarrow & \psi(t; n - 1) \geq \psi(\psi^{-1}(1 - \alpha_0; n - 1); n - 1) \\ \Leftrightarrow & \psi(t; n - 1) \geq 1 - \alpha_0 \\ \Leftrightarrow & \mathbb{P}(T \leq t) \geq 1 - \alpha_0 \\ \Leftrightarrow & \alpha_0 \geq 1 - \mathbb{P}(T \leq t) \\ \Leftrightarrow & \alpha_0 \geq \mathbb{P}(T \geq t) \end{aligned} \quad (48)$$

(7) Analyse der Powerfunktion

Wir betrachten die Testgütefunktion

$$q_\phi : \mathbb{R} \rightarrow [0, 1], \mu \mapsto q_\phi(\mu) := 1 - \psi(k_{\alpha_0}; d_\mu, n - 1) \quad (49)$$

bei kontrolliertem Testumfang, also für $k_{\alpha_0} := \psi^{-1}(1 - \alpha_0; n - 1)$ mit festem α_0 als Funktion des Nichtzentralitätsparameters und des Stichprobenumfangs. Namentlich hängt hier k_{α_0} auch von n ab.

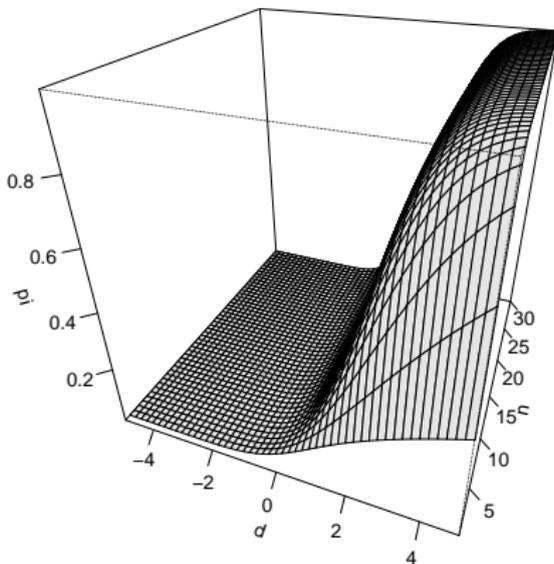
Es ergibt sich die bivariate reellwertige Funktion

$$\pi : \mathbb{R} \times \mathbb{N} \rightarrow [0, 1], (d, n) \mapsto \pi(\mu, n) := 1 - \psi(k_{\alpha_0}; d, n - 1) \quad (50)$$

Bei festgelegten α_0 hängt die Powerfunktion des einseitige T-Tests mit zusammengesetzter Nullhypothese also vom unbekanntem Wert d und von der Stichprobengröße n ab. Wir visualisieren diese Abhängigkeiten untenstehend.

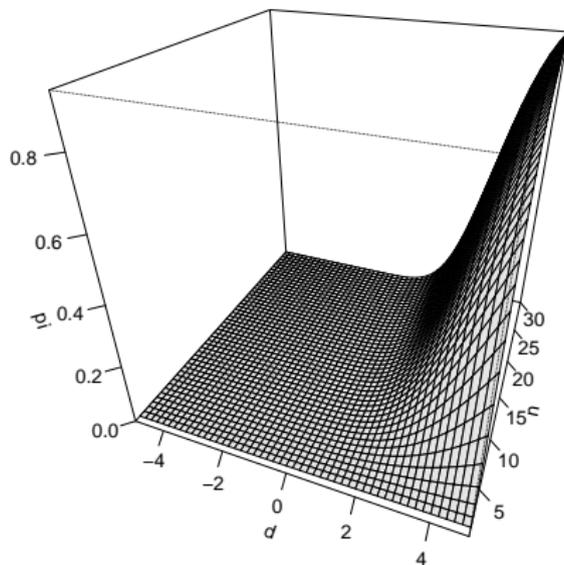
Einseitige Einstichproben-T-Tests

Powerfunktion für $\alpha_0 = 0.05$



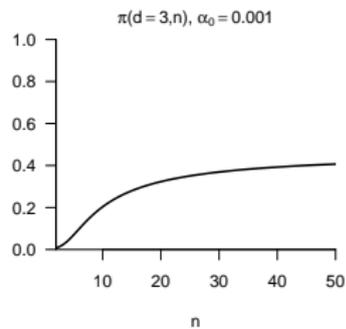
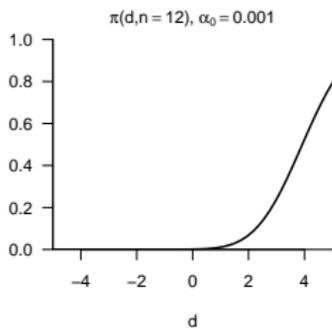
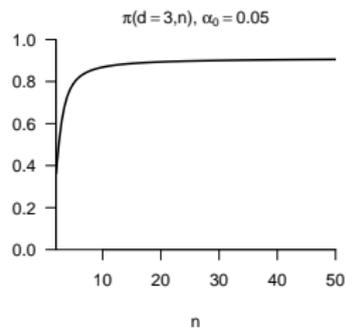
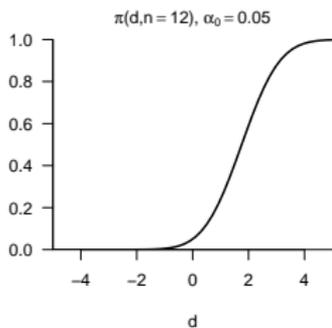
Einseitige Einstichproben-T-Tests

Powerfunktion für $\alpha_0 = 0.001$



Einseitige Einstichproben-T-Tests

Powerfunktionen für $\mu_0 = 0$



(7) Analyse der Powerfunktion

Praktisches Vorgehen

Mit größerem n steigt die Powerfunktion des Tests an

- Ein großer Stichprobenumfang ist besser als ein kleiner Stichprobenumfang.
- Kosten für die Erhöhung des Stichprobenumfangs werden aber nicht berücksichtigt.

⇒ Die Theorie statistischer Hypothesentests ist nicht besonders lebensnah.

Die Powerfunktion hängt vom wahren, aber unbekanntem, Parameterwert $d = \sqrt{n}(\mu - \mu_0)/\sigma$ ab.

⇒ Wenn man d schon kennen würde, würde man den Test nicht durchführen.

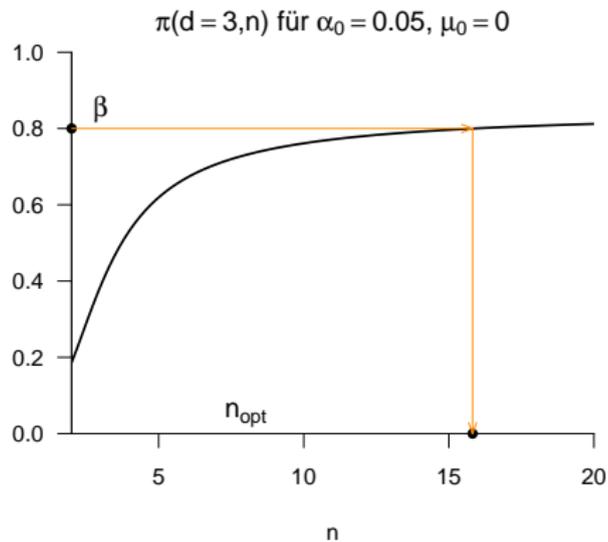
Generell wird folgendes Vorgehen favorisiert

- Man legt das Signifikanzniveau α_0 fest und evaluiert die Powerfunktion.
- Man wählt einen Mindestparameterwert d^* , den man mit $\pi(d, n) = \beta$ detektieren möchte.
- Ein konventioneller Wert ist $\beta = 0.8$.
- Man liest die für $\pi(d = d^*, n) = \beta$ nötige Stichprobengröße n ab.

Einseitige Einstichproben-T-Tests

(7) Analyse der Powerfunktion

Praktisches Vorgehen



T-Teststatistik

Einstichproben-T-Tests

Zweiseitige Einstichproben-T-Tests

Einseitige Einstichproben-T-Tests

Anwendungsbeispiel

Selbstkontrollfragen

T-Teststatistik

Einstichproben-T-Tests

Zweiseitige Einstichproben-T-Tests

Einseitige Einstichproben-T-Tests

Anwendungsbeispiel

Selbstkontrollfragen

Selbstkontrollfragen

1. Definieren Sie die T-Teststatistik.
2. Erläutern Sie die T-Teststatistik.
3. Geben Sie das Theorem zur nichtzentralen T-Transformation wieder.
4. Geben Sie das Theorem zur Verteilung der T-Teststatistik wieder.
5. Erläutern Sie das Anwendungsszenario eines Einstichproben-T-Tests.
6. Erläutern Sie mögliche Hypothesenszenarien eines Einstichproben-T-Tests.
7. Geben Sie das statistische Modell eines Einstichproben-T-Tests in klassischer Form an.
8. Geben Sie das statistische Modell eines Einstichproben-T-Tests in generativer Form an.
9. Erläutern Sie die Bedeutung des eines statistischen Modells in generativer Form.
10. Definieren Sie den zweiseitigen Einstichproben-T-Test.
11. Skizzieren Sie die Testgütefunktion eines zweiseitigen Einstichproben-T-Test
12. Geben Sie die Definition des kritischen Werts für einen zweiseitigen Level- α_0 -Einstichproben-T-Test wieder.
13. Erläutern Sie das praktische Vorgehen bei Durchführung eines zweiseitigen Level- α_0 -Einstichproben-T-Tests.
14. Geben Sie die Definition des p-Wertes Werts für einen zweiseitigen Einstichproben-T-Test wieder.
15. Von welchen Werten hängt die Powerfunktion eines Einstichproben-zweiseitigen T-Tests ab?
16. Skizzieren Sie die Powerfunktion des zweiseitigen Einstichproben-T-Tests bei fester Stichprobengröße.
17. Skizzieren Sie die Powerfunktion des zweiseitigen Einstichproben-T-Tests bei festem Nichtzentralitätsparameter.

References

Lehmann, E. L. 1986. *Testing Statistical Hypotheses*. Wiley Series in Probability and Statistics.



Inferenzstatistik

BSc Psychologie WiSe 2021/22

Prof. Dr. Dirk Ostwald

(14) Zweistichproben-T-Tests

Zweistichproben-T-Tests bei unabhängigen Stichproben

Zweistichproben-T-Tests bei abhängigen Stichproben

Anwendungsbeispiele

Selbstkontrollfragen

Zweistichproben-T-Tests bei unabhängigen Stichproben

Zweistichproben-T-Tests bei abhängigen Stichproben

Anwendungsbeispiele

Selbstkontrollfragen

Zweistichproben-T-Test bei unabhängigen Stichproben

Anwendungsszenario

- **Zwei Stichproben** experimenteller Einheiten
- Annahme unabhängiger identischer Normalverteilungen $N(\mu_1, \sigma_1^2)$ und $N(\mu_2, \sigma_2^2)$
- $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2$ unbekannt.
- Quantifizieren der Unsicherheit beim inferentiellen Vergleich von μ_1 mit μ_2 beabsichtigt.

Anwendungsbeispiele

- BDI Score Datenanalyse bei zwei Gruppen psychiatrischer Patient:innen
 - Gruppe 1 Psychoanalyse, Gruppe 2 Kognitive Verhaltenstherapie
 - $\mu_1 \neq \mu_2$? \Leftrightarrow Welche Therapie ist besser?
- **Forcierte Schwimmtestdatenanalyse bei zwei Gruppen genmanipulierter Mäuse**
 - Gruppe 1 Wildtyp, Gruppe 2 Serotoninrezeptormutation
 - $\mu_1 \neq \mu_2$? \Leftrightarrow Trägt Serotoninrezeptor zum Schwimmtestverhalten bei?

Modellszenarien

- Annahme identischer Varianzen
- Annahme eines bekannten Varianzverhältnisses
- Keine Annahmen zu Varianzen

Hypothesenszenarien

- $H_0 : \mu_1 = \mu_2$ und $H_1 : \mu_1 \neq \mu_2$
- $H_0 : \mu_1 \leq \mu_2$ und $H_1 : \mu_1 > \mu_2$
- $H_0 : \mu_1 \geq \mu_2$ und $H_1 : \mu_1 < \mu_2$

Wir betrachten hier exemplarisch $H_0 : \mu_1 = \mu_2$ und $H_1 : \mu_1 \neq \mu_2$

Modellszenarien

- **Annahme identischer Varianzen**
- Annahme eines bekannten Varianzverhältnisses
- Keine Annahmen zu Varianzen

Hypothesenszenarien

- $H_0 : \mu_1 = \mu_2$ und $H_1 : \mu_1 \neq \mu_2$
- $H_0 : \mu_1 \leq \mu_2$ und $H_1 : \mu_1 > \mu_2$
- $H_0 : \mu_1 \geq \mu_2$ und $H_1 : \mu_1 < \mu_2$

(1) Statistisches Modell in klassischer Form

$X_{11}, \dots, X_{1n_1} \sim N(\mu_1, \sigma^2)$ sei eine Stichprobe eines Normalverteilungsmodells mit unbekanntem Erwartungswertparameter μ_1 und unbekanntem Varianzparameter $\sigma^2 > 0$. $X_{21}, \dots, X_{2n_2} \sim N(\mu_2, \sigma^2)$ sei eine weitere Stichprobe eines Normalverteilungsmodells mit unbekanntem Erwartungswertparameter μ_2 und unbekanntem Varianzparameter $\sigma^2 > 0$. Die Varianzparameter beider Stichproben werden also als identisch vorausgesetzt. Der Parameter von Interesse ist (μ_1, μ_2) , der Parameterraum des Modells ist $\Theta := \mathbb{R}^2$.

(2) Statistisches Modell in generativer Form

Es sei

$$X_{ij} = \mu_i + \varepsilon_{ij} \text{ mit } \varepsilon_{ij} \sim N(0, \sigma^2) \text{ für } i = 1, 2 \text{ und } j = 1, \dots, n_i \quad (1)$$

wobei

- i die Stichproben indiziert,
- j die experimentellen Einheiten indiziert,
- n_i die Stichprobengrößen sind,
- X_{ij} beobachtbare Zufallsvariablen sind,
- μ_i feste Erwartungswertparameter der Stichprobenvariablen sind, und
- ε_{ij} unabhängige normalverteilte nicht-beobachtbare Zufallsvariablen sind.

(3) Testhypothesen, Teststatistik, Test

Wir betrachten die einfache Nullhypothese

$$H_0 : \mu_1 = \mu_2 \Leftrightarrow \Theta_0 := \{(\mu_1, \mu_2) \in \mathbb{R}^2 \mid \mu_1 = \mu_2\} \quad (2)$$

und die zusammengesetzte Alternativhypothese

$$H_1 : \mu_1 \neq \mu_2 \Leftrightarrow \Theta_1 := \{(\mu_1, \mu_2) \in \mathbb{R}^2 \mid \mu_1 \neq \mu_2\}. \quad (3)$$

Weiterhin betrachten wir die *Zweistichproben-T-Teststatistik*

$$T := \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \left(\frac{\bar{X}_1 - \bar{X}_2}{S_{12}} \right) \quad (4)$$

wobei für $i = 1, 2$

$$\bar{X}_i := \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij} \text{ und } S_{12} := \sqrt{\frac{\sum_{j=1}^{n_1} (X_{1j} - \bar{X}_1)^2 + \sum_{j=1}^{n_2} (X_{2j} - \bar{X}_2)^2}{(n_1 - 1) + (n_2 - 1)}} \quad (5)$$

die Stichprobenmittelwerte und die *gepoolte Stichprobenstandardabweichung*, respektive, bezeichnen. Schließlich definieren wir den zweiseitigen kritischen Wert-basierten Test

$$\phi(X) := 1_{\{|T| \geq k\}}. \quad (6)$$

(4) Analyse der Testgütefunktion

Theorem (Testgütefunktion)

Es sei ϕ der im obigen Modell formulierte Zweistichproben-T-Test. Dann ist die Testgütefunktion von ϕ gegeben durch

$$q_\phi : \mathbb{R}^2 \rightarrow [0, 1], (\mu_1, \mu_2) \mapsto q_\phi(\mu_1, \mu_2) \\ := 1 - \psi(k; d_{\mu_1, \mu_2}, n_1 + n_2 - 2) + \psi(-k; d_{\mu_1, \mu_2}, n_1 + n_2 - 2) \quad (7)$$

wobei $\psi(\cdot; d_{\mu_1, \mu_2}, n_1 + n_2 - 2)$ die KVF der nichtzentralen t -Verteilung mit Nichtzentralitätsparameter

$$d_{\mu_1, \mu_2} := \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \frac{\mu_1 - \mu_2}{\sigma} \quad (8)$$

und Freiheitsgradparameter $n_1 + n_2 - 2$ bezeichnet.

Bemerkungen

- q_ϕ ist eine bivariate reellwertige Funktion.
- q_ϕ kann alternativ als univariate reellwertige Funktion von $\Delta := \mu_1 - \mu_2$ konzipiert werden.
- Im Vergleich zum Einstichprobenszenario gelten

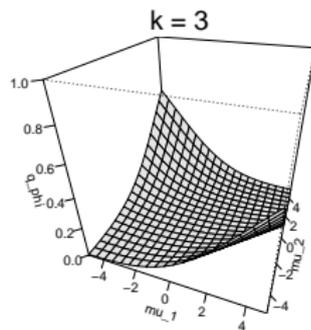
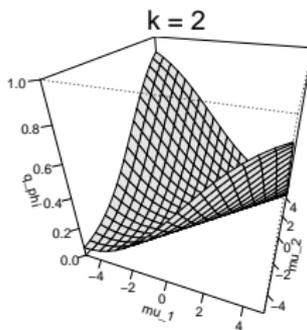
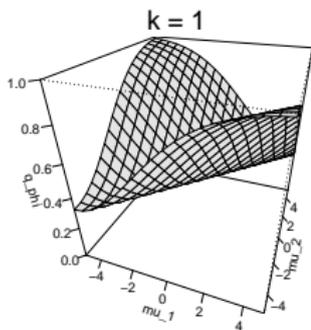
$$n \hookrightarrow n_1 + n_2 - 2, \quad \sqrt{n} \hookrightarrow \sqrt{\frac{n_1 n_2}{n_1 + n_2}}, \quad \mu - \mu_0 \hookrightarrow \mu_1 - \mu_2 \quad (9)$$

- Für einen Beweisansatz, siehe DeGroot and Schervish (2012) Seite 591.

(4) Analyse der Testgütefunktion

Testgütefunktion q_ϕ für $\sigma^2 = 9, n_1 = 12, n_2 = 12$.

$$q_\phi(\mu) = \mathbb{P}_\mu(\phi = 1)$$



(5) Testumfangkontrolle

Theorem (Testumfangkontrolle)

ϕ sei der im obigen Testscenario definierte Test. Dann ist ϕ ein Level- α_0 -Test mit Testumfang α_0 , wenn der kritische Wert definiert ist durch

$$k_{\alpha_0} := \psi^{-1} \left(1 - \frac{\alpha_0}{2}; n_1 + n_2 - 2 \right), \quad (10)$$

wobei $\psi^{-1}(\cdot; n_1 + n_2 - 2)$ die inverse KVF der t -Verteilung mit $n_1 + n_2 - 2$ Freiheitsgraden ist.

Bemerkungen

- Das Resultat folgt in Analogie zum Einstichproben-T-Test.
- Im Vergleich zum Einstichproben-T-Testfall gilt lediglich

$$n - 1 \hookrightarrow n_1 + n_2 - 2. \quad (11)$$

(5) Testumfangkontrolle

Praktisches Vorgehen

- Man nimmt an, dass zwei vorliegende Datensätze x_{11}, \dots, x_{1n_1} und x_{21}, \dots, x_{2n_2} Realisationen von $X_{11}, \dots, X_{1n_1} \sim N(\mu_1, \sigma^2)$ und $X_{21}, \dots, X_{2n_2} \sim N(\mu_2, \sigma^2)$ mit unbekanntem Parametern μ_1, μ_2, σ^2 sind.
- Man möchte entscheiden, ob eher $H_0 : \mu_1 = \mu_2$ oder $H_1 : \mu_1 \neq \mu_2$ zutrifft.
- Man wählt ein Signifikanzniveau α_0 und bestimmt den zugehörigen Freiheitsgradparameter-abhängigen kritischen Wert k_{α_0} . Zum Beispiel gilt bei Wahl von $\alpha_0 := 0.05$ und $n_1 = 12, n_2 = 12$, also Freiheitsgradparameter $12+12-2 = 22$, dass $k_{0.05} = \psi^{-1}(1 - 0.05/2; 22) \approx 2.07$ ist.
- Anhand von $n_1, n_2, \bar{x}_1, \bar{x}_2$ und der gepoolten Stichprobenstandardabweichung s_{12} berechnet man die Realisierung der Zweistichproben-T-Teststatistik

$$t := \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \left(\frac{\bar{x}_1 - \bar{x}_2}{s_{12}} \right) \quad (12)$$

- Wenn t größer-gleich k_{α_0} ist oder wenn t kleiner-gleich $-k_{\alpha_0}$ ist, lehnt man die Nullhypothese ab, andernfalls lehnt man sie nicht ab.
- Die oben entwickelte Theorie des Zweistichproben-T-Tests garantiert dann, dass man in höchstens $\alpha_0 \cdot 100$ von 100 Fällen die Nullhypothese fälschlicherweise ablehnt.

(6) p-Werte

Bestimmung des p-Wertes

- Per Definition ist der p-Wert das kleinste Signifikanzlevel α_0 , bei welchem man die Nullhypothese basierend auf einem vorliegendem Wert der Teststatistik ablehnen würde.
- Bei $T = t$ würde H_0 für jedes α_0 mit $|t| \geq \psi^{-1}(1 - \alpha_0/2; n_1 + n_2 - 2)$ abgelehnt werden. Für diese α_0 gilt, wie bereits mehrfach gezeigt,

$$\alpha_0 \geq 2\mathbb{P}(T \geq |t|). \quad (13)$$

- Das kleinste $\alpha_0 \in [0, 1]$ mit $\alpha_0 \geq 2\mathbb{P}(T \geq |t|)$ ist dann $\alpha_0 = 2\mathbb{P}(T \geq |t|)$, also folgt

$$\text{p-Wert} = 2\mathbb{P}(T \geq |t|) = 2(1 - \psi(|t|; n_1 + n_2 - 2)). \quad (14)$$

- Im Vergleich zum Einstichprobenfall gilt lediglich $n \mapsto n_1 + n_2 - 2$.

(7) Analyse der Powerfunktion

Wir betrachten die Testgütefunktion

$$\begin{aligned} q_\phi : \mathbb{R}^2 &\rightarrow [0, 1], (\mu_1, \mu_2) \mapsto q_\phi(\mu_1, \mu_2) \\ &:= 1 - \psi(k; d_{\mu_1, \mu_2}, n_1 + n_2 - 2) + \psi(-k; d_{\mu_1, \mu_2}, n_1 + n_2 - 2) \end{aligned} \quad (15)$$

als Funktion des Nichtzentralitätsparameters und der Summe der Stichprobenumfänge $n := n_1 + n_2$ bei kontrolliertem Testumfang, also für $k_{\alpha_0} := \psi^{-1}(1 - \alpha_0/2; n - 2)$ mit festem α_0 .

Es ergibt sich die multivariate reellwertige Funktion

$$\begin{aligned} \pi : \mathbb{R} \times \mathbb{N}^2 &\rightarrow [0, 1], (d, n) \mapsto \\ \pi(d, n) &:= 1 - \psi(k_{\alpha_0}; d, n - 2) + \psi(-k_{\alpha_0}; d, n - 2) \end{aligned} \quad (16)$$

Bei festgelegten α_0 hängt die Powerfunktion des zweiseitigen T-Tests mit einfacher Nullhypothese also vom unbekanntem Wert d und von der Summe der Stichprobengrößen n ab. De-facto handelt es sich also um die gleiche Powerfunktion wie beim zweiseitigen Einstichproben-T-Test mit dem einzigen Unterschied, dass für den Freiheitsgradparameter $n - 2$ anstelle von $n - 1$ gilt. Wir verzichten auf eine erneute Visualisierung.

(7) Analyse der Powerfunktion

Praktisches Vorgehen

Mit größerem $n = n_1 + n_2$ steigt die Powerfunktion des Tests an

- Ein großer Stichprobenumfang ist besser als ein kleiner Stichprobenumfang.
- Kosten für die Erhöhung des Stichprobenumfangs werden aber nicht berücksichtigt.
- Ungleichgewichte zwischen n_1 und n_2 werden durch die Tatsache ausglichen, dass Datenpunkte einer Stichproben auch zur Varianzschätzung in der anderen Stichprobe beitragen, da eine identische Varianz vorausgesetzt wurde.

Die Powerfunktion hängt vom wahren, aber unbekanntem, Parameterwert $d = \sqrt{n}(\mu_1 - \mu_0)/\sigma$ ab.

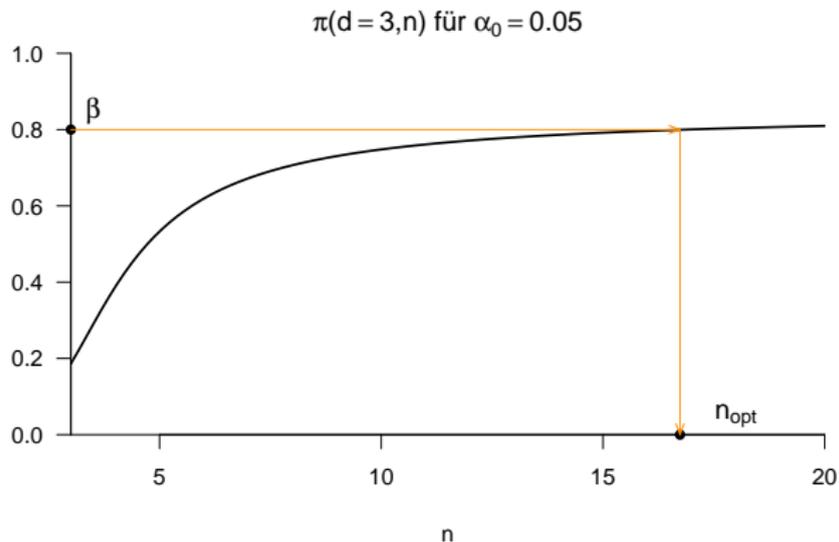
⇒ Wenn man d schon kennen würde, würde man den Test nicht durchführen.

Generell wird folgendes Vorgehen favorisiert

- Man legt das Signifikanzniveau α_0 fest und evaluiert die Powerfunktion.
- Man wählt einen Mindestparameterwert d^* , den man mit $\pi(d, n) = \beta$ detektieren möchte.
- Ein konventioneller Wert ist $\beta = 0.8$.
- Man liest die für $\pi(d = d^*, n) = \beta$ nötige Stichprobengröße n ab.

(7) Analyse der Powerfunktion

Praktisches Vorgehen



Modellszenarien

- Annahme identischer Varianzen
- **Annahme eines bekannten Varianzverhältnisses**
- Keine Annahmen zu Varianzen

Hypothesenszenarien

- $H_0 : \mu_1 = \mu_2$ und $H_1 : \mu_1 \neq \mu_2$
- $H_0 : \mu_1 \leq \mu_2$ und $H_1 : \mu_1 > \mu_2$
- $H_0 : \mu_1 \geq \mu_2$ und $H_1 : \mu_1 < \mu_2$

Überblick

- Die obige Theorie kann für eine bekannte Konstante $k > 0$ auf den Fall des statistischen Modells

$$X_{11}, \dots, X_{1n_1} \sim N(\mu_1, \sigma_1^2) \text{ und } X_{21}, \dots, X_{2n_2} \sim N(\mu_2, k\sigma_1^2) \quad (17)$$

erweitert werden.

- In diesem Fall hat die Zweistichproben-T-Teststatistik definiert durch (cf. DeGroot and Schervish (2012), S. 593)

$$T := \frac{(n_1 + n_2 - 2)^{\frac{1}{2}} (\bar{X}_1 - \bar{X}_2)}{\left(\frac{1}{n_1} + \frac{k}{n_2}\right)^{\frac{1}{2}} \left(S_1^2 + \frac{S_2^2}{k}\right)^{\frac{1}{2}}} \quad (18)$$

mit den jeweiligen Stichprobenmittelwerten \bar{X}_1 und \bar{X}_2 , sowie den Stichprobenvarianzen S_1^2 und S_2^2 eine nichtzentrale t -Verteilung mit Freiheitsgradparameter $n_1 + n_2 - 2$.

- Wir wollen diesen Fall hier nicht weiter vertiefen.

Modellszenarien

- Annahme identischer Varianzen
- Annahme eines bekannten Varianzverhältnisses
- **Keine Annahmen zu Varianzen**

Hypothesenszenarien

- $H_0 : \mu_1 = \mu_2$ und $H_1 : \mu_1 \neq \mu_2$
- $H_0 : \mu_1 \leq \mu_2$ und $H_1 : \mu_1 > \mu_2$
- $H_0 : \mu_1 \geq \mu_2$ und $H_1 : \mu_1 < \mu_2$

Überblick

- Wenn im Rahmen des statistischen Modells

$$X_{11}, \dots, X_{1n_1} \sim N(\mu_1, \sigma_1^2) \text{ und } X_{21}, \dots, X_{2n_2} \sim N(\mu_2, \sigma_2^2) \quad (19)$$

alle vier Parameter $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2$ sowie das Varianzverhältnis $k = \sigma_2^2 / \sigma_1^2$ als unbekannt vorausgesetzt werden, **kann obige Theorie nicht auf diesen Fall angewendet werden**. Das Problem der statistischen Inferenz wird in diesem Fall als **Behrens-Fisher Problem** bezeichnet.

- Eine populäre Methode zur statistischen Inferenz im Behrens-Fisher Problem geht auf Welch (1938) zurück. Diese Methode bedient sich der Teststatistik

$$\tilde{T} := (\bar{X}_1 - \bar{X}_2) / \left(\left(\frac{S_1^2}{n_1(n_1 - 1)} + \frac{S_2^2}{n_2(n_2 - 1)} \right)^{1/2} \right), \quad (20)$$

wobei S_1^2 und S_2^2 die Stichprobenvarianzen bezeichnen. Die Verteilung von \tilde{T} ist unbekannt, kann aber mit einer (nichtzentralen) t -Verteilung approximiert werden.

- Der Freiheitsgradparameter dieser approximierenden (nichtzentralen) t -Verteilung ergibt sich dabei mithilfe der beobachteten Werte s_1^2 und s_2^2 von S_1^2 und S_2^2 , respektive, zu

$$\nu = \left(\frac{s_1^2}{n_1(n_1 - 1)} + \frac{s_2^2}{n_2(n_2 - 1)} \right) / \left(\frac{1}{(n_1 - 1)^3} \left(\frac{s_1^2}{n_1} \right)^2 + \frac{1}{(n_2 - 1)^3} \left(\frac{s_2^2}{n_2} \right)^2 \right). \quad (21)$$

- Wir wollen die theoretische Begründung dieses Vorgehens hier nicht weiter vertiefen.

Zweistichproben-T-Tests bei unabhängigen Stichproben

Zweistichproben-T-Tests bei abhängigen Stichproben

Anwendungsbeispiele

Selbstkontrollfragen

Anwendungsszenario

- **Zwei Messungen an einer Gruppe experimenteller Einheiten.**
- Annahme "abhängiger" Normalverteilungen $N(\mu_1, \sigma_1^2)$ und $N(\mu_2, \sigma_2^2)$
- $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2$ unbekannt.
- Quantifizieren der Unsicherheit beim inferentiellen Vergleich von μ_1 mit μ_2 beabsichtigt.

Anwendungsbeispiele

- BDI Datenanalyse bei Psychotherapie
 - Erste Messung vor Psychotherapie, zweite Messung nach Psychotherapie
 - $\mu_1 \neq \mu_2?$ \Leftrightarrow Hat Psychotherapie einen Einfluss auf BDI?
- Reaktionszeitdatenanalyse zur visuellen Wahrnehmung
 - Bedingung 1: Hoher Kontrast, Bedingung 2: Niedriger Kontrast
 - $\mu_1 \neq \mu_2?$ \Leftrightarrow Hat Stimuluskontrast Einfluss auf perzeptuelle Entscheidungen?

Zweistichproben-T-Test bei abhängigen Stichproben

(1) Statistisches Modell in generativer Form

Es sei

$$X_{ij} = \mu_i + \beta_j + \varepsilon_{ij} \text{ mit } \beta_j \sim N(0, \sigma_\beta^2) \text{ und } \varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2) \text{ für } i = 1, 2, j = 1, \dots, n, \quad (22)$$

wobei

- i die Messungen (abhängige Stichproben) und j die experimentellen Einheiten indizieren,
- X_{ij} beobachtbare Zufallsvariablen sind,
- μ_i feste Erwartungswertparameter der Messungen sind,
- β_j unabhängige normalverteilte nicht-beobachtbare Zufallsvariablen sind, und
- ε_{ij} unabhängige normalverteilte nicht-beobachtbare Zufallsvariablen sind.

Alle β_i und ε_{ij} werden als voneinander unabhängig vorausgesetzt.

- Die X_{ij} modellieren Datenpunkte,
- die μ_i modellieren feste Messungs-abhängige Effekte,
- die β_j modellieren zufällige experimentelle Einheits-abhängige Effekte,
- die ε_{ij} modellieren (wie immer) zufällige Störeffekte.

Lineare Modelle mit festen Effekten (fixed effects) und zufälligen Effekten (random effects) heißen gemischte Modelle (linear mixed models).

Zweistichproben-T-Test bei abhängigen Stichproben

(1) Statistisches Modell in generativer Form

Datengeneration

```
# Modellspezifikation
n           = 12
mu          = c(10,12)
sigsqr_beta = 5
sigsqr_veps = 1

# Stichprobengröße
# feste Effektparameter \mu_1, mu_2
# zufälliger Effektparameter \sigsqr_beta
# Störvariablenparameter \sigsqr_varepsilon

# Datengeneration
X = data.frame(
  mu_1 = rep(NaN,n),
  mu_2 = rep(NaN,n),
  beta_j = rep(NaN,n),
  veps_1_j = rep(NaN,n),
  veps_2_j = rep(NaN,n),
  X_1j = rep(NaN,n),
  X_2j = rep(NaN,n)
)

# Datensatzinitialisierung
# \mu_1
# \mu_2
# \beta_j,
# \varepsilon_{1j},
# \varepsilon_{2j},
# X_{1j}, j = 1, \dots, n
# X_{2j}, j = 1, \dots, n

# Iterationen über experimentelle Einheiten
for(j in 1:n){
  beta_j = rnorm(1,0,sqrt(sigsqr_beta)) # zufälliger experimenteller Einheits-abhängiger Effekt
  X[[j,3]] = beta_j # \beta_j

  # Iterationen über Messungen
  for(i in 1:2){
    eps_ij = rnorm(1,0,sqrt(sigsqr_veps)) # zufälliger Störvariablenwert
    X[[j,i]] = mu[i] # \mu_i
    X[[j,3+i]] = eps_ij # \varepsilon_{ij}
    X[[j,5+i]] = mu[i] + beta_j + eps_ij # X_{ij} = \mu_i + \beta_j + \varepsilon_{ij}
  }
}
```

Zweistichproben-T-Test bei abhängigen Stichproben

(1) Statistisches Modell in generativer Form

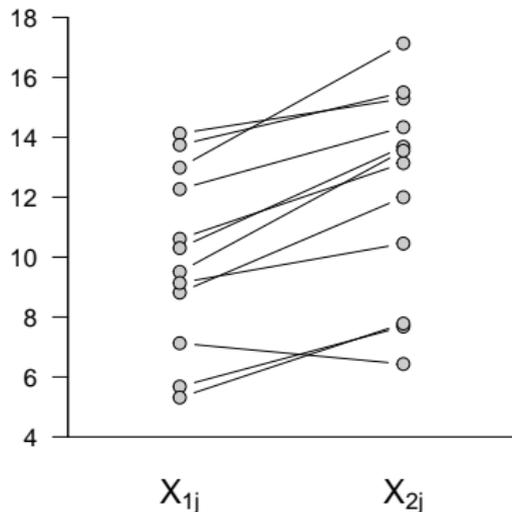
Datenbeispiel

mu_1	mu_2	beta_j	veps_1_j	veps_2_j	X_1j	X_2j
10	12	-3.131	0.255	-2.437	7.12	6.43
10	12	-0.012	0.622	1.148	10.61	13.14
10	12	-4.074	-0.247	-0.244	5.68	7.68
10	12	-0.632	-0.554	0.629	8.81	12.00
10	12	4.618	-1.631	0.512	12.99	17.13
10	12	-4.166	-0.522	-0.053	5.31	7.78
10	12	1.214	-0.914	0.468	10.30	13.68
10	12	0.812	-1.305	0.738	9.51	13.55
10	12	4.223	-0.097	-0.936	14.12	15.29
10	12	-0.036	-0.827	-1.512	9.14	10.45
10	12	2.092	0.176	0.244	12.27	14.34
10	12	3.630	0.112	-0.134	13.74	15.50

Zweistichproben-T-Test bei abhängigen Stichproben

(1) Statistisches Modell in generativer Form

Datenbeispiel



Zweistichproben-T-Test bei abhängigen Stichproben

(1) Statistisches Modell in generativer Form

Theorem (Verteilung der Differenzvariablen)

Das oben formulierte statistische Modell eines Zweistichproben-T-Test bei abhängigen Stichproben sei gegeben und für $j = 1, \dots, n$ sei $X_j := X_{j2} - X_{j1}$ die Differenzvariable der Zufallsvariablen X_{j1} und X_{j2} . Dann gilt

$$X_1, \dots, X_n \sim N(\mu, \sigma^2) \text{ mit } \mu := \mu_2 - \mu_1 \text{ und } \sigma^2 := 2\sigma_\epsilon^2. \quad (23)$$

Bemerkungen

- Hinsichtlich der Differenzvariablen X_j handelt es sich um das Einstichproben-T-TestszENARIO.
- Die Hypothesen des Zweistichproben-T-Test bei abhängigen Stichproben hinsichtlich der Erwartungswertparameter μ_1 und μ_2 übertragen sich wie folgt auf die Einstichproben-T-Test Hypothesen in bezug auf den Erwartungswertparameter μ :

Zweistichproben-T-Test			Einstichproben-T-Test		
$H_0 : \mu_1 = \mu_2$	$H_1 : \mu_1 \neq \mu_2$	\Rightarrow	$H_0 : \mu = 0$	$H_1 : \mu \neq 0$	
$H_0 : \mu_1 \leq \mu_2$	$H_1 : \mu_1 > \mu_2$	\Rightarrow	$H_0 : \mu \geq 0$	$H_1 : \mu < 0$	
$H_0 : \mu_1 \geq \mu_2$	$H_1 : \mu_1 < \mu_2$	\Rightarrow	$H_0 : \mu \leq 0$	$H_1 : \mu > 0$	

Zweistichproben-T-Test bei abhängigen Stichproben

(1) Statistisches Modell in generativer Form

Beweis

Wir halten zunächst fest, dass

$$\begin{aligned}X_j &= X_{2j} - X_{1j} \\&= \mu_2 + \beta_j + \varepsilon_{2j} - \mu_1 - \beta_j - \varepsilon_{1j} \\&= \mu_2 - \mu_1 + \varepsilon_{2j} - \varepsilon_{1j} \\&= \mu + \varepsilon_j,\end{aligned}\tag{24}$$

wobei wir $\mu := \mu_2 - \mu_1$ und $\varepsilon_j := \varepsilon_{2j} - \varepsilon_{1j}$ definiert haben.

In Generalisierung des Summentransformationstheorems für normalverteilte Zufallsvariablen (cf. (6) Transformationen der Normalverteilung) halten wir weiterhin fest, dass für unabhängige Zufallsvariablen $X_i \sim N(\mu_i, \sigma_i^2)$, $i = 1, \dots, n$ und Koeffizienten a_i , $i = 1, \dots, n$ gilt, dass

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

Mit $a_1 := 1$ und $a_2 := -1$ folgt dann aus $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon)$, $i = 1, 2$ direkt, dass

$$\varepsilon_j \sim N\left(0, 2\sigma_\varepsilon^2\right)\tag{26}$$

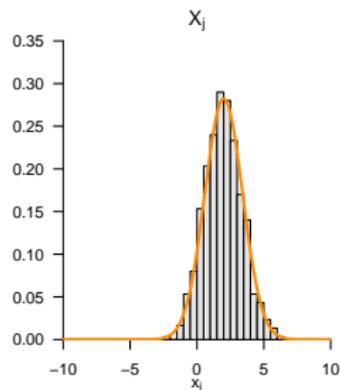
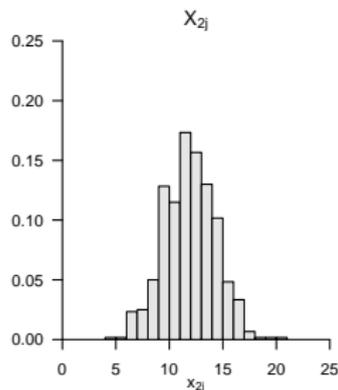
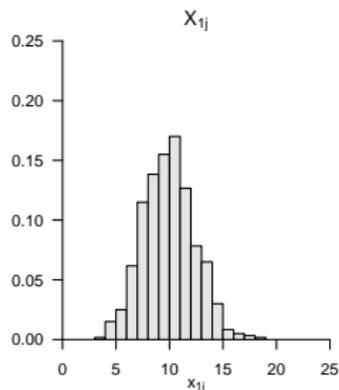
Mit der Definition $\sigma^2 := 2\sigma_\varepsilon^2$ und dem Transformationstheorem für normalverteilte Zufallsvariablen bei linear-affiner Transformation folgt das Resultat dann direkt.

Zweistichproben-T-Test bei abhängigen Stichproben

(1) Statistisches Modell in generativer Form

Visualisierung des Zweistichproben-T-Test Theorems

$$n = 12, \mu_1 = 10, \mu_2 = 12, \sigma_\beta^2 = 5, \sigma_\varepsilon^2 = 1$$



(3) Praktisches Vorgehen

- Man nimmt an, dass ein vorliegender Datensatz von Messpaaren x_{11} und x_{21} , x_{12} und x_{22} , ..., x_{1n} und x_{2n} eine Realisation von X_{ij} , $i = 1, 2, j = 1, \dots, n$ wie in der Beschreibung des generativen Modells definiert ist.
 - Man bildet die Differenzwerte $x_1 = x_{21} - x_{11}$, $x_2 = x_{22} - x_{12}$, ..., $x_n = x_{2n} - x_{1n}$.
- ⇒ Man nutzt die Theorie des Einstichproben-T-Tests zum Testen der Hypothesen des Zweistichproben-T-Tests mit abhängigen Stichproben.

Zweistichproben-T-Test bei abhängigen Stichproben

(3) Praktisches Vorgehen

Datenbeispiel

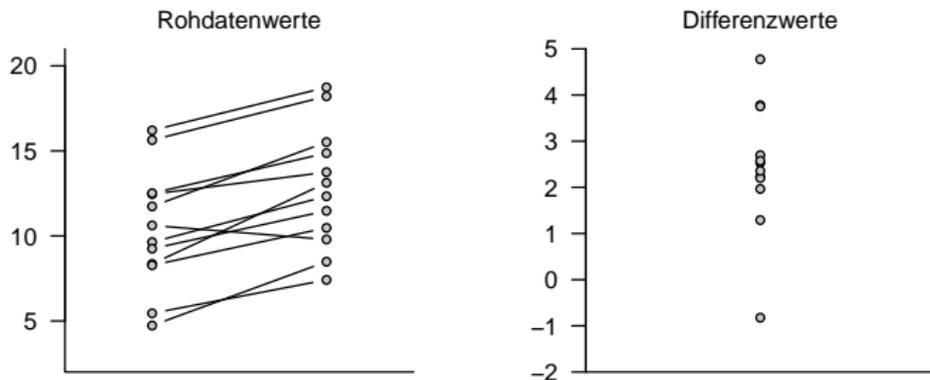
μ_1	μ_2	β_j	veps_{1_j}	veps_{2_j}	X_{1j}	X_{2j}	X_j
10	12	-4.271	-0.279	-0.313	5.45	7.42	1.966
10	12	2.387	0.070	-0.639	12.46	13.75	1.291
10	12	-0.112	-0.251	0.445	9.64	12.33	2.696
10	12	6.161	0.047	0.578	16.21	18.74	2.531
10	12	0.264	-1.912	0.862	8.35	13.13	4.774
10	12	-0.544	-0.206	0.019	9.25	11.47	2.225
10	12	0.066	0.550	-2.274	10.62	9.79	-0.824
10	12	5.998	-0.361	0.213	15.64	18.21	2.575
10	12	2.402	-0.665	1.114	11.74	15.52	3.779
10	12	-0.550	-1.178	-0.976	8.27	10.47	2.202
10	12	2.382	0.132	0.489	12.51	14.87	2.357
10	12	-3.800	-1.471	0.284	4.73	8.48	3.755

⇒ Einstichproben-T-Test basierend auf $X_j, j = 1, \dots, n$

Zweistichproben-T-Test bei abhängigen Stichproben

(3) Praktisches Vorgehen

Datenbeispiel



⇒ Einstichproben-T-Test basierend auf $X_j, j = 1, \dots, n$

Zweistichproben-T-Tests bei unabhängigen Stichproben

Zweistichproben-T-Tests bei abhängigen Stichproben

Anwendungsbeispiele

Selbstkontrollfragen

Zweistichproben-T-Tests bei unabhängigen Stichproben

Zweistichproben-T-Tests bei abhängigen Stichproben

Anwendungsbeispiele

Selbstkontrollfragen

Selbstkontrollfragen

1. Erläutern Sie das Anwendungsszenario eines Zweistichproben-T-Tests bei unabhängigen Stichproben.
2. Nennen Sie drei Modellszenarien für Zweistichproben-T-Tests bei unabhängigen Stichproben.
3. Geben Sie das statistische Modell eines Zweistichproben-T-Tests bei unabhängigen Stichproben in klassischer Form an.
4. Geben Sie das statistische Modell eines Zweistichproben-T-Tests bei unabhängigen Stichproben in generativer Form an.
5. Definieren Sie die gepoolte Stichprobenstandardabweichung.
6. Definieren Sie die Zweistichproben-T-Teststatistik.
7. Definieren Sie den kritischen Wert für einen zweiseitigen Level- α_0 -Zweistichproben-T-Tests bei unabhängigen Stichproben.
8. Erläutern Sie das praktische Vorgehen bei Durchführung eines zweiseitigen Level- α_0 -Zweistichproben-T-Tests bei unabhängigen Stichproben.
9. Erläutern Sie das Szenario eines Zweistichproben-T-Tests bei unabhängigen Stichproben mit bekanntem Varianzverhältnis.
10. Definieren Sie das Behrens-Fisher Problem.
11. Erläutern Sie das Anwendungsszenario eines Zweistichproben-T-Tests bei abhängigen Stichproben.
12. Geben Sie das statistische Modell eines Zweistichproben-T-Tests bei abhängigen Stichproben in generativer Form an.
13. Geben Sie das Theorem zur Verteilung der Differenzvariablen im Zweistichproben-T-Tests bei abhängigen Stichproben wieder.
14. Erläutern Sie das praktische Vorgehen bei Durchführung eines Zweistichproben-T-Tests bei abhängigen Stichproben.

References

DeGroot, Morris H., and Mark J. Schervish. 2012. *Probability and Statistics*. 4th ed. Boston: Addison-Wesley.

Welch, B. L. 1938. "The Significance of the Difference Between Two Means When the Population Variances Are Unequal." *Biometrika* 29 (3-4): 350–62. <https://doi.org/10.1093/biomet/29.3-4.350>.