



Vorlesung: Statistik II für Studierende der Soziologie und Nebenfachstudierende

Prof. Dr. Helmut Küchenhoff

Institut für Statistik, LMU München

SoSe 2013

Besonderer Dank gilt Prof. Augustin, der mir das Material zur Verfügung gestellt hat.



Termine und Informationen

Homepage:

http://www.statistik.lmu.de/institut/ag/statsoz_neu/lehre/2013_SoSe/Stat2Soz_13/index.html

Vorlesung:

Prof. Helmut Küchenhoff

Mi. 12:00 bis 14:00 Uhr Hauptgebäude M 018

Do. 12:00 bis 14:00 Uhr Hauptgebäude M 018

Übung 1 Di. 12:15 bis 13:45 Uhr HGB-E 004 Hauptgebäude

Übung 2 Di. 14:15 bis 15:45 Uhr HGB-E 004 Hauptgebäude

Tutorium Mo. 12:15 bis 13:45 Uhr HGB-M 010 Hauptgebäude



Literatur

L. Fahrmeir, R. Künstler, I. Pigeot, G. Tutz:

Statistik - Der Weg zur Datenanalyse

Springer-Verlag, 7. Auflage

Helmut Küchenhoff et al. (2006):

Statistik für Kommunikationswissenschaftler

2., überarbeitete Auflage

UVK Verlagsgesellschaft mbH, Konstanz



- ① Einführung
- ① Wahrscheinlichkeitsrechnung
- ② Zufallsvariablen und ihre Verteilung
- ③ Statistische Inferenz
- ④ Hypothesentests
- ⑤ Regression



**0 Einführung**

1 Wahrscheinlichkeitsrechnung

2 Zufallsvariablen und ihre Verteilung

3 Statistische Inferenz

4 Hypothesentests

5 Regression

Einleitung

Deskriptive Statistik (Statistik I):

- Beschreibung von Daten (Grundgesamtheit oder Stichprobe).
- Keine Verallgemeinerung von einer Stichprobe auf die zugehörige Grundgesamtheit angestrebt.

Induktive Statistik (Statistik II):

- Induktion: Schluss vom Teil auf das Ganze, von vielen Einzelbeobachtungen auf allgemeine Gesetze
- Schluss von einer Stichprobe auf Eigenschaften der Grundgesamtheit (*Inferenz*)

Beispiel 1: Bundestagswahl 2009

Prognose 18:00 Infratest Dimap (ARD)

CDU/CSU	SPD	FDP	Linke	Grüne	Sonstige
33,5	22,5	15	12,5	10,5	6



Basis: Nachwahlbefragung 100 000 Wahlberechtigte nach Verlassen der Wahllokale
 Infos unter
<http://www.tagesschau.de/wahl/umfragen/infratest102.html>

Beispiel 1: Bundestagswahl 2009

- Grundgesamtheit: Alle Wähler der Bundestagswahl
- Stichprobe: 100 000 Wähler
- Gesucht: Information über *alle* Wähler, also die *Grundgesamtheit*

Stichprobe dient zum Lernen über die Grundgesamtheit

Studie mit Medizinischer Klinik in Großhadern

Prospektive Analyse kardialer Notfälle

01.05. - 31.07.2006

Retrospektive Analyse

01.05. - 3.07.2003

01.05. - 3.07.2005

Analyse der Protokolle von 24 Notarztstandorten aus München sowie näherer und weiterer Umgebung

Einschlusskriterien:

ST-Hebungsinfarkt, Nicht-ST-Hebungsinfarkt/instab. Angina
Herzrythmusstörungen mit ausgeprägter/lebensbedrohlicher Symptomatik
Herzrythmusstörungen mit geringer Symptomatik

Ziele und Methoden:

- Beschreibung des Zusammenhangs zwischen emotionalem Stress und Auftreten von Herzinfarkten
- Berücksichtigung von Störgrößen
- Schluss von Beobachtung auf allgemeines Gesetz
- Bewertung des Risikos

Beispiel 3: Lebenszufriedenheit und Alter

- Gibt es eine Midlife Crisis?
- Analysen von Panel-Daten zur subjektiven Lebenszufriedenheit mit
- semiparametrischen Regressionsmodellen
- In Zusammenarbeit mit Sonja Greven, Andrea Wiencierz, Christoph Wunder

Beispiel 3: Lebenszufriedenheit und Alter

Datengrundlage

- Daten stammen aus den Haushaltsstichproben A (Westdeutsche) und C (Ostdeutsche) des Sozio-Ökonomischen Panels (SOEP)
- für die ausgewählten Modellvariablen liegen Beobachtungen aus den Jahren 1992, 1994 bis 2006 vor
- durchschnittliche Anzahl von Beobachtungen pro Person: 7.77
- in die Modellberechnungen gingen 102 708 vollständige Beobachtungen von 13 224 Individuen ein
- Anzahl Beobachtungen pro Jahr:

1992	1994	1995	1996	1997	1998	1999
8 145	7 720	7 943	7 606	8 052	7 550	7 403
2000	2001	2002	2003	2004	2005	2006
7 628	7 092	7 068	7 000	6 876	6 543	6 082

Beispiel 3: Lebenszufriedenheit und Alter

Methode: Multiples Lineares Regressionsmodell

- Zielgröße: Subjektive Lebenszufriedenheit
- Einflussgrößen: Alter Gesundheit, Gehalt usw.
- Hauptfrage: Wie hängen Lebenszufriedenheit und Alter zusammen
- Alterseffekt wird nicht parametrisch modelliert

Beispiel 3: Lebenszufriedenheit und Alter

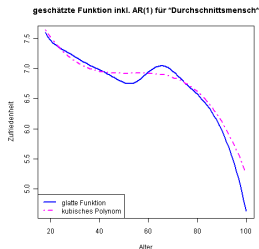
Ergebnisse für das Regressionsmodell

Variable	Coefficient	Standard error
Sex: female	0.074	(0.015)
Disability status: disabled	-0.452	(0.014)
Nights stayed in hospital	-0.012	(0.000)
Years of education	0.034	(0.002)
Log of net household income	0.492	(0.010)
Log of household size	-0.194	(0.012)
German	0.053	(0.020)
Full time employed	0.079	(0.011)
Part time employed	0.019	(0.012)
Unemployed	-0.597	(0.014)
Single	-0.174	(0.017)
Divorced	-0.137	(0.018)
Widowed	-0.196	(0.023)
West-Germany	0.511	(0.017)

Aus: WUNDER, C., WIENCIERZ, A., SCHWARZE, J. and KÜCHENHOFF, H.(2011). Well-Being over the Life Span: Semiparametric Evidence from British and German Longitudinal Data. *Review of Economics and Statistics*. Accepted for publication.

Beispiel 3: Lebenszufriedenheit und Alter

Ergebnis für Alterseffekt



Midlife-Crisis nur bei glatter Funktion erkennbar.

Beispiel 3: Lebenszufriedenheit und Alter

Ziele und Methoden

- Zusammenhänge analysieren
- Komplexe Einflüsse
- flexibles Modell

Studie in Zusammenarbeit mit Prof. Adam (LMU)

Fragestellung: Schmeckt mit Sauerstoff angereichertes Mineralwasser besser als gewöhnliches Mineralwasser ?

- Doppelblindstudie
- KontrollGruppe: zweimal das gleiche Wasser ohne O_2
- VerumGruppe: Beim zweiten Mal mit O_2 angereichertes Mineralwasser

Ergebnis (Clausnitzer et al., 2004):

Placebo: 76% gaben an, dass das zweite Wasser anders schmeckt

Verum : 89 % gaben an, dass das zweite Wasser anders schmeckt

Signifikanter Effekt → Zulassung von



• Punktschätzung:

- Zum Beispiel: Wie groß ist der Anteil der schwarz-gelb-Wähler unter allen Wahlberechtigten?
- Wie erhält man aus der Stichprobe gute Schätzwerte für Charakteristika („Parameter“) der Grundgesamtheit?
- Wann ist ein Schätzverfahren gut/besser als ein anderes?

• Bereichsschätzung:

- Typischerweise stimmt der Punktschätzer nicht mit dem wahren Wert überein.
- Realistischer: Gib einen Bereich an, „der den wahren Anteil der schwarz-gelb-Wähler mit hoher Wahrscheinlichkeit enthält“.
- Ungenauere Aussagen, dafür aber zuverlässiger.

Hypothesentests

- Überprüfe aus substanzwissenschaftlicher Theorie abgeleitete Hypothesen über die Grundgesamtheit anhand der Daten.
- Zum Beispiel: Verdienen Männer wirklich mehr als Frauen?
- Ist der Unterschied „signifikant“?

Regressionsmodelle incl. Varianzanalyse

- Modelle zur Beschreibung des Einflusses von Variablen
- Zum Beispiel: Wie hängt die Lebenszufriedenheit vom Alter ab ?

- Zentrales Problem der induktiven Statistik: *Jeder* Induktionsschluss ist potentiell fehlerbehaftet
- Beispielsweise ist der wahre Anteil der Wähler einer Partei in der Grundgesamtheit nicht exakt mithilfe der Stichprobe vorhersagbar.
- Entscheidende Idee: Kontrolle des Fehlers mit Hilfe der Wahrscheinlichkeitsrechnung.
 - Stichprobenziehung *zufällig*
 - Verwende Methoden der Wahrscheinlichkeitsrechnung zur Quantifizierung/Kontrolle des Fehlers

- Verwende statistische Modelle mit zufälligen Komponenten
- „Gesetze“ bzw. Aussagen enthalten stochastischen Aspekt

Kap. 1: Wahrscheinlichkeitsrechnung

Kap. 2: Wie nutzt man Wahrscheinlichkeitsüberlegungen für die Statistik?
Induktive Statistik



1 Einführung

2 Wahrscheinlichkeitsrechnung

3 Zufallsvariablen und ihre Verteilung

4 Statistische Inferenz

5 Hypothesentests

6 Regression

Wahrscheinlichkeitsrechnung für Sozialwissenschaftler ?

- 1 Probabilistisches Denken (d.h. das Denken in Wahrscheinlichkeiten) unerlässlich! Strenge Kausalitäten (wenn A dann folgt immer B) findet man bestenfalls vereinzelt in Naturwissenschaften, in den Sozialwissenschaften gilt typischerweise nur: wenn A dann folgt eher B als C .
- 2 Wahrscheinlichkeiten und Umgang mit Unsicherheit spielen in der Gesellschaft eine wichtige Rolle. Bei naiver Herangehensweise (ohne Wahrscheinlichkeitsrechnung) kann man sich leicht täuschen. Z.B. bewerten sowohl medizinische Experten als auch Laien Risiken oft falsch.
- 3 Stichprobenverfahren und statistische Modelle spielen in den (empirisch orientierten) Sozialwissenschaften eine zentrale Rolle. Für das Verständnis sind Grundlagenkenntnisse in Wahrscheinlichkeitsrechnung zentral

- Götz Rohwer, Ulrich Pötter: Wahrscheinlichkeit: Begriff und Rhetorik in der Sozialforschung, Weinheim (u.a.): Juventa-Verlag 2002
- Gigerenzer, G.: Das Einmaleins der Skepsis. BTB, 2.Auf. 2005.
- Nate Silver: The Signal and the Noise: Why So Many Predictions Fail but Some Don't. Penguin Press 2012.

Ein Zufallsvorgang (Zufallsexperiment) führt zu einem von mehreren, sich gegenseitig ausschließenden Ergebnissen. Es ist vor der Durchführung ungewiss, welches Ergebnis eintreten wird.

Was benötigen wir zur Beschreibung eines Zufallsvorganges?

Zwei wesentliche Aspekte:

- a) Welche Ergebnisse eines Zufallsvorganges sind möglich? (Was kann alles passieren?)
- b) Mit welcher Wahrscheinlichkeit treten die einzelnen Ergebnisse ein?

Formale Beschreibung

Festlegen eines *Ergebnisraums* (Grundraum, Stichprobenraum) Ω , der alle möglichen *Ergebnisse* ω enthält.

Beispiele:

- $\Omega = \{1, \dots, 6\}$ beschreibt die möglichen Ergebnisse eines Würfelexperimentes
Ein mögliches Ergebnis: $\omega = 4$; $\omega = 17$ ist kein mögliches Ergebnis.
- $\Omega = \mathbb{R}_0^+$ beschreibt die möglichen Erwerbseinkommen
Ein mögliches Ergebnis: $\omega = 17513 \text{ €}$
- Ziehung einer Person: $\Omega = \{1, \dots, N\}$
Ein mögliches Ergebnis: $\omega = 17$

Ereignisse

Ereignisse sind **Teilmengen** von Ω

Beispiele:

1. „gerade Zahl“ = $\{2, 4, 6\}$
2. „1 oder 2“ = $\{1, 2\}$
3. „Einkommen zwischen 1000 und 2000 h €“ = $\{\omega | 1000 \leq \omega \leq 2000\}$
4. „Person ist weiblich“ = $\{\text{alle Nummern, die zu Frauen gehören}\}$

Ereignissen sollen Wahrscheinlichkeiten zugeordnet werden.
Wir bezeichnen Ereignisse mit A, B, C, ...

Definition

Eine Menge ist eine Zusammenfassung verschiedener Objekte zu einem Ganzen. Die einzelnen Objekte einer Menge werden Elemente genannt.

Mengen werden üblicherweise mit Großbuchstaben bezeichnet, z.B. A, B, C, Ω, \dots

Mengen werden benutzt, um den Ausgang von Zufallsexperimenten zu beschreiben.

Beispiele: Standardmengen

\mathbb{N}	$= \{1, 2, 3, \dots\}$: Menge der natürlichen Zahlen,
\mathbb{N}_0	$= \{0, 1, 2, 3, \dots\}$: Menge der natürlichen Zahlen inklusive 0,
\mathbb{Z}	$= \{0, \pm 1, \pm 2, \dots\}$: Menge der ganzen Zahlen,
\mathbb{R}	$= (-\infty, \infty)$: Menge der reellen Zahlen,
\emptyset		: leere Menge.

- $\Omega = \{\text{Hörer einer Vorlesung}\}$.
- $\Omega = \{1, 2, 3, 4, 5, 6\}$ (Menge der Ergebnisse eines Würfelwurfs).
Die Reihenfolge der Aufzählung spielt (im Gegensatz zu Tupeln) keine Rolle:

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

- $B = \{K, Z\}$ (Menge der Ergebnisse eines Münzwurfs, $K = \text{Kopf}$, $Z = \text{Zahl}$).
- Charakterisierung von Mengen mit einer gewissen Eigenschaft:

$$\{1, 2, 3, \dots, 10\} = \{x \mid x \text{ ist eine natürliche Zahl } \leq 10\}$$

Die Menge aller x mit der Eigenschaft „ x ist eine natürliche Zahl ≤ 10 “.

Grundlegende Begriffe der Mengenlehre

Illustration anhand folgender Mengen

$$\begin{aligned} \Omega &= \{\text{CDU/CSU, SPD, FDP, Grüne, Linke, Sonstige}\} \\ A &= \{\text{CDU/CSU, SPD, FDP, Grüne}\} \\ B &= \{\text{CDU/CSU, SPD, FDP}\} \\ C &= \{\text{SPD, FDP, Grüne}\} \end{aligned}$$

Elementeigenschaft:

x ist Element der Menge A : $x \in A$

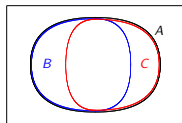
x ist nicht Element der Menge A : $x \notin A$

Es gilt $\text{SPD} \in C$, aber $\text{CDU/CSU} \notin C$.

- Teilmengen:** A ist Teilmenge von B , in Zeichen $A \subset B$, wenn jedes Element von A auch in B ist.

$B \subset A$

C ist keine Teilmenge von B



$\Omega \leftarrow$ größte Menge als Kasten

Im Beispiel:

$A \subset \Omega, B \subset \Omega, C \subset \Omega, B \subset A, C \subset A,$

aber beispielsweise: weder $B \subset C$ noch $C \subset B$ noch $C = B$

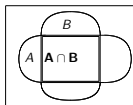
Für jede Menge A gilt:

- $\emptyset \subset A$ Die leere Menge ist Teilmenge jeder Menge, denn jedes Element in \emptyset ist in A
- $A \subset A$ d.h. „ \subset “ enthält implizit „ $=$ “, deshalb in Literatur manchmal auch \subseteq statt \subset

Schnittmenge

Die Schnittmenge $A \cap B$ ist die Menge aller Elemente, die sowohl in A als auch in B enthalten sind:

$$A \cap B = \{x | x \in A \text{ und } x \in B\}$$



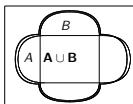
Eigenschaften

- * Gilt $A \subset B$, so ist $A \cap B = A$.
- * Für jede Menge A gilt: $A \cap A = A$ und $A \cap \emptyset = \emptyset$.
- * Zwei Mengen A und B mit $A \cap B = \emptyset$, d.h. zwei Mengen, die kein gemeinsames Element haben, heißen *disjunkt*.
- * Die Schnittmenge aus n Mengen A_1, \dots, A_n enthält alle Elemente, die in jeder der Mengen A_1, \dots, A_n enthalten sind und wird bezeichnet mit

$$\bigcap_{i=1}^n A_i := A_1 \cap A_2 \cap \dots \cap A_n.$$

Vereinigungsmenge: Die Vereinigungsmenge $A \cup B$ ist die Menge aller Elemente, die in A oder B enthalten sind:

$$A \cup B = \{x | x \in A \text{ oder } x \in B\}$$



- * Vorsicht: Das „oder“ ist *nicht* exklusiv gemeint, also nicht „entweder oder“, sondern als „in A oder in B oder in beiden“.
- * Die Vereinigungsmenge aus n Mengen A_1, \dots, A_n enthält alle Elemente, die in mindestens einer der Mengen A_1, \dots, A_n enthalten sind und wird bezeichnet mit

$$\bigcup_{i=1}^n A_i := A_1 \cup A_2 \cup \dots \cup A_n$$

$A \cup B$: Vereinigung = „A oder B“
 $A \cap B$: Durchschnitt = „A und B“
 A^C : Komplement = „Nicht A“

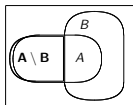
Beispiele:

$\Omega = \{1, 2, 3, 4, 5, 6\}$
 $A = \{2, 4, 6\}$ „gerade“
 $B = \{4, 5, 6\}$ „groß“
 $A \cup B = \{2, 4, 5, 6\}$ „gerade oder groß“
 $A \cap B = \{4, 6\}$ „gerade und groß“
 $A^C = \{1, 3, 5\}$ „ungerade“
 $B^C = \{1, 2, 3\}$ „klein“

Differenzmenge

Die Differenzmenge $A \setminus B$ ist die Menge aller Elemente, die in A, aber nicht in B enthalten sind:

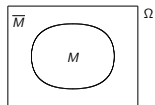
$$A \setminus B = \{x \mid x \in A \text{ aber } x \notin B\}$$



Komplementärmenge

Die Komplementärmenge \bar{A} bezüglich einer Grundmenge Ω ist die Menge aller Elemente von Ω , die nicht in A sind:

$$\bar{A} = \{x \in \Omega \mid x \notin A\} = \{x : x \notin A\}$$



- * Die Komplementärmenge ist nur unter Bezugnahme auf eine Grundmenge Ω definierbar.
- * Es gilt $\bar{A} = \Omega \setminus A$.
- * Es existieren noch weitere Schreibweisen für die Komplementärmenge, z.B. $A^c, \mathcal{C}A$.
- * „Tertium non datur“ (Grundlegendes Prinzip der Mengenlehre (und der Logik): Für jedes Element $x \in \Omega$ gilt entweder $x \in A$ oder $x \in \bar{A}$

Potenzmenge: Die Potenzmenge $\mathcal{P}(A)$ ist die Menge aller Teilmengen von A :

$$\mathcal{P}(A) = \{M \mid M \subset A\}.$$

Im Beispiel:

$$\mathcal{P}(B) = \{\emptyset, \{\text{CDU/CSU}\}, \{\text{SPD}\}, \{\text{FDP}\}, \{\text{CDU/CSU, SPD}\}, \{\text{CDU/CSU, FDP}\}, \{\text{SPD, FDP}\}, B\}$$

Mächtigkeit: Die Mächtigkeit $|A|$ einer Menge A ist die Anzahl der Elemente von A

Im Beispiel:

$$|B| = 3$$

Für jede Menge B gilt $|\mathcal{P}(B)| = 2^{|B|}$; im Beispiel 2^3

Rechenregeln für Mengen

- 1 Kommutativgesetze (Vertauschung):

$$A \cap B = B \cap A, \quad A \cup B = B \cup A.$$

- 2 Assoziativgesetze (Zusammenfassen):

$$(A \cap B) \cap C = A \cap (B \cap C). \\ (A \cup B) \cup C = A \cup (B \cup C).$$

Rechenregeln für Mengen

- 3 Distributivgesetze (Ausklammern/Ausmultiplizieren):

$$(A \cup B) \cap C = (A \cap C) \cup (B \cap C). \\ (A \cap B) \cup C = (A \cup C) \cap (B \cup C).$$

- 4 De Morgansche Regeln:

$$\overline{(A \cup B)} = \bar{A} \cap \bar{B} \\ \overline{(A \cap B)} = \bar{A} \cup \bar{B}$$

- 5 Aus $A \subset B$ folgt $\bar{B} \subset \bar{A}$.
 6 Für die Differenzmenge gilt $A \setminus B = A \cap \bar{B}$.
 7 Für die Potenzmenge gilt $|\mathcal{P}(A)| = 2^{|A|}$.
 vgl. Bsp: $|B| = 3 \quad |\mathcal{P}(B)| = 2^3 = 8$

Das kartesische Produkt zweier Mengen

$$A = \{a_1, a_2, a_3, \dots, a_k\}$$

$$B = \{b_1, b_2, b_3, \dots, b_m\}$$

ist die Menge

$$A \times B := \{(a_i, b_j) \mid i = 1, \dots, k, j = 1, \dots, m\}$$

Sie besteht also aus allen möglichen Kombinationen, so dass

$$A \times B = \{(a_1, b_1), (a_1, b_2), (a_1, b_3), \dots, (a_1, b_m),$$

$$(a_2, b_1), (a_2, b_2), (a_2, b_3), \dots, (a_2, b_m),$$

$$\vdots$$

$$(a_k, b_1), (a_k, b_2), (a_k, b_3), \dots, (a_k, b_m)\}$$

Verallgemeinerungen:

- Das kartesische Produkt der Mengen $\Omega_1, \Omega_2, \dots, \Omega_n$ wird mit

$$\times_{j=1}^n \Omega_j = \Omega_1 \times \Omega_2 \times \dots \times \Omega_n$$

bezeichnet und besteht aus allen möglichen n -Tupeln, die sich (unter Beachtung der Reihenfolge) aus Elementen aus $\Omega_1, \Omega_2, \dots, \Omega_n$ bilden lassen.

- Die Mengen $\Omega_1, \Omega_2, \dots, \Omega_n$ müssen nicht endlich sein; für endliche Mengen gilt

$$|\times_{j=1}^n \Omega_j| = |\Omega_1| \cdot |\Omega_2| \cdot \dots \cdot |\Omega_n|$$

- Kartesische Produkte werden verwendet, um Ergebnisse komplexer Experimente aus Einzelexperimenten zusammensetzen.

$$A = \{1, 2, 3\}$$

$$B = \{1, 2, 3, 4\}$$

$$A \times B = \{(1, 1), (1, 2), (1, 3), (2, 1), \dots\}$$

Achtung: Bei den Elementen von $A \times B$ handelt es sich um Tupel, das heißt die Reihenfolge ist wichtig! (z.B. $(1, 2)$ ist etwas anderes als $(2, 1)$, vgl. Statistik I Kapitel 5)

Wahrscheinlichkeit (formale Definition)

Definition

Eine Wahrscheinlichkeitsfunktion ordnet jedem Ereignis seine Wahrscheinlichkeit zu. Eine Wahrscheinlichkeit ist also eine Abbildung von Ereignissen (Elementen der Potenzmenge von Ω) auf reelle Zahlen:

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

$$A \mapsto P(A)$$

Dabei sollen gewisse fundamentale Rechenregeln gelten, z.B.

- 108 kann keine Wahrscheinlichkeit sein, nur Zahlen zwischen 0 und 1.
- $P(\{2, 3\})$ muss mindestens so groß sein wie $P(\{3\})$.

Die einelementigen Teilmengen (also die Ereignisse, die ein Ergebnis in ω enthalten) werden als *Elementarereignisse* bezeichnet.

Beispiel: Bei einem fairen Würfel gilt für die Elementarereignisse

$$P(\text{„Augenzahl 1“}) = P(\{1\}) = \frac{1}{6}$$

$$P(\text{„Augenzahl 2“}) = P(\{2\}) = \frac{1}{6}$$

$$\vdots$$

$$P(\text{„Augenzahl 6“}) = P(\{6\}) = \frac{1}{6}$$

Beispiel: Dreimaliger Münzwurf

Wir werfen dreimal unabhängig voneinander eine faire Münze und notieren jeweils, ob die Münze Wappen oder Zahl anzeigt. Man beschreibt den Ergebnisraum und berechne die Wahrscheinlichkeit, mindestens einmal Wappen zu erhalten und genau zweimal Wappen zu erhalten.

Laplace-Wahrscheinlichkeiten oder kombinatorische Wahrscheinlichkeiten

Häufig: Alle möglichen Elementarereignisse sind *gleich* wahrscheinlich, d.h. $P(\{\omega_j\}) = \frac{1}{|\Omega|}$. In diesem Fall sprechen wir von einem *Laplace-Experiment*.

Abzählregel:

$$P(A) = \frac{\text{Anzahl der für } A \text{ günstigen Ergebnisse}}{\text{Anzahl aller möglichen Ergebnisse}}$$

Laplace-Wahrscheinlichkeit: In einem Laplace-Experiment gilt für $P(A)$ mit $|A| = M$ und $|\Omega| = N$:

$$P(A) = \frac{|A|}{|\Omega|} = \frac{M}{N}.$$

Beispiel: Dreimaliger Münzwurf

Der zugehörige Ergebnisraum lautet:

$$\Omega = \{(W, W, W), (W, W, Z), (W, Z, W), (Z, W, W), (W, Z, Z), (Z, W, Z), (Z, Z, W), (Z, Z, Z)\}$$

Ω enthält die acht Ergebnisse dieses Zufallsexperiments, d.h. $|\Omega| = 8$. Da wir vorausgesetzt haben, dass die Münze fair ist, kann dieser Zufallsvorgang als Laplace-Experiment angesehen werden, und es gilt:

$$p(\{\omega\}) = \frac{1}{|\Omega|} = \frac{1}{8}, \quad \omega \in \Omega.$$

Berechnung der Wahrscheinlichkeit

Damit besitzt jedes mögliche Ergebnis die gleiche Wahrscheinlichkeit $1/8$, d.h. dreimal hintereinander Wappen zu erhalten, ist genauso wahrscheinlich wie etwa die Kombination (W, Z, W) .

Die Wahrscheinlichkeit für das Ereignis $A =$ „mindestens einmal Wappen“ berechnet sich nun direkt als Quotient der Anzahl der für A günstigen Ergebnisse und der Anzahl aller möglichen Ergebnisse. Da bei sieben Elementarereignissen von Ω mindestens einmal Wappen auftritt, ist $|A| = 7$ und somit

$$P(A) = \frac{7}{8}.$$

Für das Ereignis $B = \{\text{genau zweimal Wappen}\}$ gilt:

$$B = \{(W, W, Z), (W, Z, W), (Z, W, W)\}$$

also $|B| = 3$ und damit $P(B) = \frac{3}{8}$.

	Royal Flush
	Straight Flush
	Four of a Kind
	Full House
	Flush
	Straight
	Tripling
	Two Pair
	One Pair
	High Card

Poker: Texas Holdem



Deine Hand



Board

In diesem Beispiel hat man ein Flush, denn man kann die drei Kreuzkarten in der Mitte zusammen mit den zwei Kreuzkarten auf der Hand verwenden.

Beispiel 2



Deine Hand



Board

In diesem Beispiel hat man ein Paar Könige mit einer Dame als Kicker.

Beispiel: Poker

Berechnung von Wahrscheinlichkeiten

Meine Hand: 2 Asse

Ein Gegner: Alle Möglichkeiten für 2 Karten des Gegners:

Alle Möglichkeiten für die 5 gemeinsamen Karten.

Berechnung mit PokerStove (frei im Internet)

2,097,572,400 Möglichkeiten

Günstige Fälle 1781508418

$$P(\text{Gewinn}) = \frac{1781508418}{2097572400} = 85.204\%$$

Beispiel: Es wird ein Studierender der Vorlesung gezogen und nach seiner Wahlabsicht gefragt.

Dazu nehmen wir an, dass es N Studierende in der Vorlesung gibt und dass sie durchnummeriert sind $n = 1, \dots, N$

$$P(\text{Student Nr } n \text{ wird gezogen}) = 1/N$$

Alle haben die gleiche Ziehungswahrscheinlichkeit.

- Wie groß ist die Wahrscheinlichkeit, dass er/sie ein SPD Wähler ist?
- Wie groß ist die Wahrscheinlichkeit, dass eine Frau gezogen wird?

- Wahrscheinlichkeit für „SPD-Wähler“

$$\begin{aligned} P(\text{SPD}) &= \frac{\text{Anzahl der für SPD günstigen Ergebnisse}}{\text{Anzahl aller möglichen Ergebnisse}} \\ &= \frac{\text{Anzahl der SPD Wähler}}{\text{Anzahl aller Studierenden der Vorlesung}} \end{aligned}$$

Die Wahrscheinlichkeit ist also die relative Häufigkeit f_{SPD} der SPD Wähler in der Grundgesamtheit.

- Wahrscheinlichkeit für „Frau“?

Relative Häufigkeiten und Wahrscheinlichkeiten

Die Argumentation des Beispiels gilt ganz allgemein.

$$P(\text{Eine Person mit der Eigenschaft } E \text{ wird gezogen}) = f_E$$

- Die relativen Häufigkeiten/Anteile aus der Grundgesamtheit pflanzen sich also in der entsprechenden Wahrscheinlichkeitsverteilung in der Stichprobe fort.
- Dies ist ganz entscheidend, denn dadurch kann man also durch eine Stichprobe etwas über die Häufigkeitsverhältnisse in der Grundgesamtheit lernen.
- Wichtig ist dabei, dass die Stichprobe zufällig gezogen wird!

Urnenmodelle

Es hat sich eingebürgert, gewisse Grundsituationen, die in der praktischen Stichprobenziehung immer wieder vorkommen, als „Urnenmodelle“ zu beschreiben. Man stellt sich eine Urne mit Kugeln vor und zieht daraus dann in einer bestimmten Art und Weise eine bestimmte Anzahl von Kugeln (*Stichprobe*). In der Sozialwissenschaft entspricht jede „Kugel“ einer interessierenden Einheit (Person, Haushalt) und die „Urne“ einer gedachten Gesamtliste dieser Einheiten. Eine typische Unterscheidung der verschiedenen Ziehungsvorgänge besteht darin, ob eine Einheit mehrfach in eine Stichprobe gelangen kann („Ziehen mit Zurücklegen“, die gezogene Kugel wird also wieder in die Urne zurückgelegt) oder „Ziehen ohne Zurücklegen“ (die gezogenen Kugeln bleiben also außerhalb der Urne).

- das Ziehen von n Kugeln ohne Zurücklegen entspricht einem n -fachen einmaligen Zug aus der Urne.
- Typischerweise sind Stichproben ohne Zurücklegen praktisch einfacher zu realisieren und zu rechtfertigen.
- Für sehr große Grundgesamtheiten sind die Unterschiede zwischen mit und ohne Zurücklegen verschwindend gering.

Die praktische Umsetzung erfolgt mit Hilfe von Computerprogrammen (Pseudozufallszahlen)

- Grundgesamtheit mit N Zahlen $G = \{1, \dots, N\}$.
- Ziehe Stichprobe vom Umfang n **mit** Zurücklegen.
- Zur Beschreibung des Zufallsvorgangs müssen wir die Anzahl der potentiell möglichen Stichprobenergebnisse bestimmen (jede Stichprobe ist gleichwahrscheinlich).
- $\Omega = \{(\omega_1, \dots, \omega_n) \mid \omega_j \in \{1, \dots, N\}\}$, das selbe Element kann mehrfach vorkommen.
- $|\Omega| = \underbrace{N \cdot N \cdot \dots \cdot N}_n = N^n$, d.h. N^n potentiell mögliche Stichproben vom Umfang n .

Ziehen ohne Zurücklegen

- Grundgesamtheit mit N Einheiten (mit Nummern identifiziert) $G = \{1, \dots, N\}$.
- Ziehe Stichprobe vom Umfang n **ohne** Zurücklegen.
- $\Omega = \{(\omega_1, \dots, \omega_n) : \omega_j \in \{1, \dots, N\}, \omega_j \neq \omega_i \text{ für } i \neq j\}$, jedes Element kann nur einmal vorkommen.
- Anzahl möglicher Stichproben:

$$\begin{aligned}
 |\Omega| &= N \cdot (N-1) \cdot \dots \cdot N - n + 1 = \\
 &\quad \uparrow \quad \quad \uparrow \quad \quad \quad \uparrow \\
 &\quad 1. \text{ Ziehung} \quad 2. \text{ Ziehung} \quad n\text{-te Ziehung} \\
 &= \frac{N \cdot (N-1) \cdot \dots \cdot (N-n+1)(N-n) \cdot \dots \cdot 2 \cdot 1}{(N-n)(N-n-1) \cdot \dots \cdot 2 \cdot 1} = \\
 &= \frac{N!}{(N-n)!}
 \end{aligned}$$

Einfache Zufallsstichprobe

Ziehen **ohne** Zurücklegen **ohne** Berücksichtigung der Reihenfolge

- Ziehe n Kugeln aus einer Urne mit N nummerierten Kugeln. Die Reihenfolge der Ziehungen spielt keine Rolle, d.h. die Stichprobe „4,1,7“ wird nicht unterschieden von „7,1,4“.
- $\Omega = \{(\omega_1, \dots, \omega_n) : \omega_j \in \{1, \dots, N\}, \omega_j \neq \omega_i \text{ für } j \neq i\}$
- Anzahl der Stichproben:

$$|\Omega| = \frac{N!}{(N-n)!n!} = \binom{N}{n}$$

Man berücksichtigt zunächst die Reihenfolge. Dann gibt es $N!$ Möglichkeiten, die Kugeln anzuordnen. Da nur die ersten n Kugeln zählen, sind jeweils $(N - n)!$ Anordnungen, die sich nur ab der $(n + 1)$ -ten Kugel unterscheiden, äquivalent, also hat man

$$\frac{N!}{(N - n)!}$$

Möglichkeiten, geordnete Stichproben zu erzeugen.

Da die Reihenfolge unter den ersten n Kugeln aber keine Rolle spielen soll, sind jeweils $n!$ geordnete Stichproben wiederum äquivalent, es gibt also insgesamt

$$\frac{N!}{(N - n)!n!}$$

verschiedene Stichproben.

Der *Binomialkoeffizient* $\binom{N}{n}$ ist definiert als

$$\binom{N}{n} = \frac{N!}{(N - n)! \cdot n!}.$$

Es gilt:

$$\binom{N}{0} = 1, \binom{N}{1} = N, \binom{N}{N} = 1, \binom{N}{n} = 0, \text{ falls } N < n.$$

Beispiel: Lottozahlen

- Frage: Wie viele verschiedene mögliche Ziehungen gibt es?
- Antwort: Entspricht der Ziehung ohne Zurücklegen von 6 Zahlen aus 49, d.h.

$$|\Omega| = \binom{49}{6} = \frac{49!}{43! \cdot 6!} = 13983816.$$

Dann gilt

$$P(\text{„6 Richtige“}) = \frac{1}{13983816} = 0.000000072$$

Beispiel: Lottozahlen, 5 richtige

- Frage: Wie groß ist die Wahrscheinlichkeit, 5 Richtige zu bekommen?
- Anzahl der günstigen Fälle:

$$\binom{6}{5} \cdot \binom{43}{1}$$

↑

↑

Anzahl der
Möglichkeiten:
5 Richtige
aus 6 Zahlen

Anzahl der
Möglichkeiten:
1 Falsche
aus 43 Zahlen

$$P(\text{„5 Richtige“}) = \frac{\binom{6}{5} \binom{43}{1}}{\binom{49}{6}} = \frac{6 \cdot 43}{13983816} = 0.0000184.$$

Warum reichen Laplace-Wahrscheinlichkeiten nicht?

Essentielle Voraussetzung: alle Fälle müssen „gleich möglich“ (also gleich wahrscheinlich) sein

Beispiel: Wie wird das Wetter morgen? 3 Möglichkeiten:

$$\{\text{Sonne, Regen, Gemischt}\} \implies P(\text{„Sonne“}) = \frac{1}{3}$$

Eine Funktion P (P steht für Probability), die Ereignissen aus Ω reelle Zahlen zuordnet, heißt *Wahrscheinlichkeit*, wenn gilt

(K1) $P(A) \geq 0$ für alle Ereignisse $A \subset \Omega$.

(K2) $P(\Omega) = 1$.

(K3) Falls $A \cap B = \emptyset$, dann gilt $P(A \cup B) = P(A) + P(B)$

Axiome von Kolmogorov (1933)

- Die Axiome von Kolmogorov stellen zunächst eine reine Definition dar, die festlegt, was eine Wahrscheinlichkeit sein soll.
- Es gibt verschiedene Versuche Wahrscheinlichkeiten operational zu definieren (also durch eine Messvorschrift) und verschiedene Interpretationen, die die Axiomatik mit Leben füllen sollen.
- Aus hier nicht zu erörternden mathematischen Gründen
 - darf man bei überabzählbar unendlichen Ergebnisräumen, z.B. also im Fall $\Omega = \mathbb{R}$, nicht alle Teilmengen von Ω als Ereignisse zulassen. Alle Mengen, „an die man normalerweise denkt“ sind aber zugelassen.
 - muss man bei unendlichen Ergebnisräumen in (K3) eigentlich unendliche Summen zulassen.

Rechenregeln für Wahrscheinlichkeiten

- $P(\bar{A}) = 1 - P(A)$
- Für nicht notwendigerweise disjunkte Mengen A, B gilt

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

- Falls A_1, A_2, \dots, A_n paarweise disjunkt sind, also $A_i \cap A_j = \emptyset$ für $i \neq j$, dann gilt:

$$P\left(\bigcup_{i=1}^n A_i\right) = P(A_1 \cup A_2 \cup \dots \cup A_n) = P(A_1) + P(A_2) + \dots + P(A_n)$$

Es folgt, dass, sofern Ω endlich ist, die Wahrscheinlichkeit durch die Wahrscheinlichkeit der Elementarereignisse vollständig bestimmt ist:

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

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

Alle Elementarereignisse sind gleich wahrscheinlich, d.h.

$$\begin{aligned} p(\{1\}) &= p(\{2\}) = \dots p(\{6\}) := a; \\ \text{wegen } p(\{1\}) + p(\{2\}) + \dots + p(\{6\}) &= 1 \\ \Leftrightarrow 6 \cdot a &= 1 \\ \Leftrightarrow a &= \frac{1}{6} \end{aligned}$$

Sei $A = \{\text{gerade Augenzahl}\} = \{2, 4, 6\}$:

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

Die Axiomatik ist insbesondere nötig, um mit Situationen mit nicht gleich wahrscheinlichen Elementarereignissen rechnen zu können.

Betrachtet werde ein verfälschter Wurfel mit

$$\begin{aligned} P(\{1\}) = P(\{3\}) = P(\{5\}) &= \frac{1}{12} \\ P(\{2\}) = P(\{4\}) = P(\{6\}) &= \frac{1}{4} \end{aligned}$$

Übung:

Gegeben seien ferner die Ereignisse

A	die Person wählt Partei	1 oder 3
B	"	4 oder 6
C	"	3 oder 4

und berechnen Sie $P(A)$, $P(B)$, $P(B \cap C)$ und $P(A \cup C)$ mit Hilfe der Axiome.

Grundlegendes zum Begriff „Wahrscheinlichkeit“

- Was ist eigentlich Wahrscheinlichkeit?
- Was bedeutet: „Mit Wahrscheinlichkeit $\frac{2}{3}$ wird es morgen regnen?“
- Stark gebräuchlich in der Umgangssprache als graduelle Abschwächung von Sicherheit („wahrscheinlich kommt Max“). Weg vom simplen Ja/Nein.
- Teilweise sogar quantifiziert:
„Die Niederschlagswahrscheinlichkeit für morgen beträgt 30%“
- Medizinische Beipackzettel: „seltene Nebenwirkungen“

Literatur zum Begriff „Wahrscheinlichkeit“

Ausführlichere und weiterführende Literatur:

- Rohrer, G., Pötter, U. (2002): *Wahrscheinlichkeit, Begriff und Rhetorik in der Sozialforschung*. Juventa, Weinheim und München.
- Schneider, I. (Hg.) (1988): *Die Entwicklung der Wahrscheinlichkeit von den Anfängen bis 1933. Einführungen und Texte*. Wissenschaftliche Buchgesellschaft, Darmstadt.

- Wahrscheinlichkeit im Glücksspiel, v.a. Würfelspiel: Profanisierung erst im Mittelalter, dort erst als Zufall gedeutet, vorher oft als Gottesurteil etc.
 - * Cardano (1501-1576)
 - * Gallilei (1546-1642)
 - * Briefwechsel zwischen Pascal (1623-1662) und Fermat (1601-1665), erste systematische Wahrscheinlichkeitsrechnung: Lösung für Frage, wie Einsätze gerecht aufzuteilen sind, wenn Spiel unterbrochen wurde
 - * Huygens (1629-1695)
- Wahrscheinlichkeit (Prove-ability → probability)

- Mathematisierung von Glücksspiel
- als philosophischer/theologischer Begriff
- der Philosophie des Unsicheren und
- der Mathematik der Glücksspiele
- Jacob Bernoulli (1654 - 1705)
Binomialverteilung.
Theorem von Bernoulli: durch genügend große Versuchsreihen kann der Unterschied zwischen der relativen Häufigkeit eines Ereignisses und seiner Wahrscheinlichkeit beliebig gering gemacht werden.

Wahrscheinlichkeitsbegriffe

Objektivistisch // frequentistische Richtungen // aleatorische Wahrscheinlichkeiten

- Anschluss an die göttliche Ordnung
- Wahrscheinlichkeiten beschreiben tatsächlich vorhandene, zufällige Gesetzmäßigkeiten
- Objektbezogen: Wahrscheinlichkeit ist eine Eigenschaft des untersuchten Objekts (z.B. Würfel), objektiv \longleftrightarrow objektbezogen (wie z.B. spezifisches Gewicht, Länge)
- Häufigkeitsinterpretation bzw. sogar -Definition Wahrscheinlichkeit als relative Häufigkeiten in „unendlich langen“ reproduzierbaren Experimenten

Frequentistischer Wahrscheinlichkeitsbegriff I

R. von Mises (1883 - 1953):

„Die Wahrscheinlichkeit eines Ereignisses ist die langfristige relative Häufigkeit seines Auftretens“

Für ein Ereignis A:

$$P(A) = \lim_{n \rightarrow \infty} \frac{n_A}{n}$$

n_A : Anzahl der Erfolge

n : Anzahl der Versuche

Probleme bei der Definition:

- Einmalige Ereignisse
- Grenzwertdefinition
- Experimentdurchführung

- Wahrscheinlichkeit hat ausschließlich mit Unsicherheit, nicht mit Zufälligkeit zu tun (Man kann auch über völlig deterministische Aspekte unsicher sein!)
- Wahrscheinlichkeit ist Eigenschaft des untersuchenden Subjekts \implies verschiedene Subjekte können durchaus zu unterschiedlichen Bewertungen kommen.
- Anwendung auch auf Aussagen. Bsp: Die Wahrscheinlichkeit, dass die Regierungskoalition die gesamte Legislaturperiode hält, ist...
- behaviouristischer Standpunkt: Wahrscheinlichkeiten äußern sich im Verhalten und können so gemessen werden z.B. bei Wetten
- Wichtig: subjektiv sind die Wahrscheinlichkeiten aber nicht die Rechenregeln.

Subjektiver Wahrscheinlichkeitsbegriff I

Laplace, Ramsey, de Finetti:

„Die Wahrscheinlichkeit eines Ereignisses ist der Grad der Überzeugung, mit der ein Beobachter aufgrund eines bestimmten Informationsstandes an das Eintreten eines Ereignisses glaubt“

$P(A)$ ist der Wetteinsatz in Euro, den eine Person höchstens einzugehen bereit ist, falls diese bei Eintreten von A einen Euro gewinnt.

Beispiele:

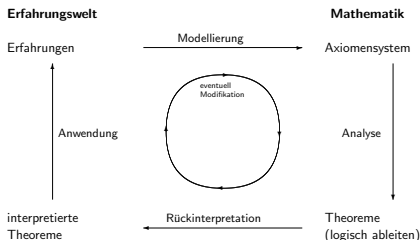
- Münzwurf: Einsatz auf „Zahl“ bis zu 0.5 € sinnvoll
Würfel: Einsatz auf „5 oder 6“ bis zu 1/3 € sinnvoll

Subjektiver Wahrscheinlichkeitsbegriff II

Probleme:

- subjektiv = unwissenschaftlich ?
- Wettdefinition
- Informationsstand

- Axiomatik nach Kolmogoroff
- typische Anwendung der axiomatischen Methode:
Axiom: Nicht bezweifelte Grundannahme für Kalkül
- Die Kolmogoroffsche Axiomatik ist eine reine Definition, die sich zunächst im „luftleeren“ Raum bewegt. Es wird rein formal festgelegt, was eine Wahrscheinlichkeit sein soll.
- Die Axiomatik ist *verträglich* sowohl mit der *Häufigkeits-* als auch mit der *Wettinterpretation*.
- Die Axiome von Kolmogoroff geben an, wie man mit Wahrscheinlichkeiten rechnet.
- Welche Phänomene man durch Wahrscheinlichkeiten beschreiben darf und wie die Ergebnisse zu interpretieren sind, ist aber damit nicht geklärt.



Ausblick

- In der Tat gibt es auch Kritik an dieser Axiomatik: „zu streng und überpräzise“ → aktueller Forschungsgegenstand (*Imprecise Probabilities*, *Intervallwahrscheinlichkeit*); hier nicht näher thematisiert: Kolmogoroff als absolute „Wahrheit“. Kritik:
 - * Modellierung unsicheren (partiell widersprüchlichen, unvollständigen) Expertenwissens
 - * Ökonomie: Entscheidungen unter komplexer Unsicherheit widersprechen Prognosen aus der üblichen Wahrscheinlichkeitsrechnung
- logischer Wahrscheinlichkeitsbegriff
Wahrscheinlichkeit kommt Schlüssen zu: Wahrscheinlichkeit als logischer Grad mit dem aus einer Prämisse auf die Konklusion geschlossen werden darf (frühere Formalisierungsversuche gelten heute als gescheitert; aber Renaissance der Thematik)

Zur Kommunikation von Wahrscheinlichkeiten

Darstellung durch natürliche Häufigkeiten (nach Gigerenzer)

- * „Superrepräsentative Stichprobe vorstellen“, in der sich genau die Häufigkeitsverhältnisse in der Grundgesamtheit wiederfinden, z.B. 10 000 Personen
- * Dann $P(A) = 0.1756$ vorstellen als: 1756 Personen haben die Eigenschaft A.
- + einfachere Kommunikation von Wahrscheinlichkeiten und Risiken, reduziert Fehler beim Rechnen mit Wahrscheinlichkeiten
Experimente mit Ärzten zeigen, dass die Darstellungsform Wahrscheinlichkeiten vs. natürliche Häufigkeiten) einen starken Einfluss auf die Korrektheit von Berechnungen hat.
- Gefahr der Verschleierung von Unsicherheit: die „natürlichen Häufigkeiten“ sind zu *erwartende Durchschnittswerte*, wenn man sehr viele Stichproben hätte.

Angabe des Risikos von Nebenwirkungen auf Beipackzetteln:

- sehr häufig: mehr als 1 von 10 Behandelten
- häufig: weniger als 1 von 10, aber mehr als 1 von 100 Behandelten
- gelegentlich: weniger als 1 von 100, aber mehr als 1 von 1000 Behandelten
- selten: weniger als 1 von 1000, aber mehr als 1 von 10000 Behandelten
- sehr selten: 1 Fall oder weniger von 10000 Behandelten, einschließlich Einzelfälle

Welche Nebenwirkungen können bei der Anwendung von * Auftreten?**

Gelegentlich wurde über das Auftreten von Mundschleimhautentzündungen, Kopfschmerzen, Ohrengeräuschen berichtet.

Selten können auftreten: Beschwerden im Magen-Darm-Bereich (z.B. Sodbrennen, Übelkeit, Erbrechen oder Durchfall).

- Beim Lotto ist die Wahrscheinlichkeit bei einem Spiel einen 6er zu bekommen:

$$\frac{1}{\binom{49}{6}} = \frac{1}{13983816} = 0.00000072$$

- „Einmal in 14 Millionen Spielen“
- „Einmal in 20.000 Jahren bei wöchentlichem Spielen“
- „Es ist wahrscheinlicher, den Tag der Ziehung nicht mehr zu erleben, als zu gewinnen“
- Simulationsexperiment

Es gibt drei Arten der Beschreibung von Risiken für die menschliche Gesundheit:

- Absolutes Risiko:**
Angabe von Krankheitswahrscheinlichkeiten, jeweils getrennt für die Gruppe mit und ohne Risikofaktor
- Relatives Risiko:**
Verhältnis der Krankheitswahrscheinlichkeiten mit und ohne Risikofaktor
- Anzahl der zusätzlich geschädigten Personen**
(erwarteter Effekt)

- Absolutes Risiko:**
Ohne WM-Spiel: 14.6/1.34 Millionen = 1.09:100.000
Mit WM-Spiel: 2.9:100000
(Zum Vergleich: Tägliches Sterberisiko laut Sterbetafel für M, 50: 0.00460442/365 = 1.26/100.000)
- Relatives Risiko:**
Bei Spielen der deutschen Mannschaft um den Faktor 2.9/1.09 = 2.66 erhöht
Konfidenzintervall (2.33 - 3.04)
- Anzahl der zusätzlichen Fälle:**
Gesamtfälle im Studiengebiet bei D-Spielen: 302
Zu erwartende Fälle: 302/2.66 = 114
Zusätzliche Fälle: 302 - 114 = 188
Konfidenzintervall (172 - 203)

Hochrechnung auf Deutschland: 8000 - 11000

Beispiel: Wirkung von Pravastatin

„Menschen mit hohem Cholesterinspiegel können das Risiko eines erstmaligen Herzinfarkts sehr schnell um 22 Prozent vermindern, wenn sie einen häufig angewandten Wirkstoff namens Pravastatin einnehmen“

- Reduktion der Todesfälle von 41 auf 32 pro 1000 Patienten mit hohem Cholesterin ($32 = 41 \cdot (1 - 0.22) = 41 \cdot 0.78$)
Wahrscheinlichkeit für Todesfall: Reduktion von 4.1% auf 3.2%
Absolute Risikodifferenz: 0.9%
- Reduktion um 22% (relatives Risiko 0.78) „22% werden gerettet“
- Es müssen 111 Patienten behandelt werden, um ein Menschenleben zu retten.
Number needed to treat = $1/\text{Absolute Risikodifferenz} = 1/0.009 = 111.11$

Bedingte Wahrscheinlichkeit I

„Herzoperation in Krankenhaus“

Überleben der Operation

Alle Fälle	Operation überlebt	Operation nicht überlebt	P(nicht ü) „Risiko“
Krankenhaus U	500	500	0.5
Krankenhaus K	900	100	0.1

Frage: „In welchem Krankenhaus würden Sie sich behandeln lassen?“

Bedingte Wahrscheinlichkeit II

Schwere der behandelten Fälle

	schwere Fälle	leichte Fälle
Krankenhaus U	900	100
Krankenhaus K	100	900

Frage: „Bleiben Sie bei Ihrer Entscheidung?“

Bedingte Wahrscheinlichkeit III

Überleben der Operation aufgeteilt nach der Schwere der behandelten Fälle

Schwere Fälle	Operation überlebt	Operation nicht überlebt	P(nicht ü) „Risiko“
Krankenhaus U	400	500	0.56
Krankenhaus K	30	70	0.7

Leichte Fälle	Operation überlebt	Operation nicht überlebt	P(nicht ü) „Risiko“
Krankenhaus U	100	0	0
Krankenhaus K	870	30	0.033

In dem Beispiel betrachten wir das Risiko gegeben „schwerer Fall“.
Das Risiko wird berechnet durch

$$\frac{\text{Anzahl (schwere Fälle und nicht überlebt)}}{\text{Anzahl(schwere Fälle)}}$$

Allgemein definieren wir die Wahrscheinlichkeit von
„Ereignis B gegeben A“

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

B: Nicht überleben
A: Schwerer Fall

Krankenhaus U

$$\begin{aligned} P(B) &= 500/1000 = 0.5 \\ P(A) &= 900/1000 = 0.9 \\ P(A \cap B) &= 500/1000 = 0.5 \\ P(B|A) &= 0.5/0.9 = 0.56 \end{aligned}$$

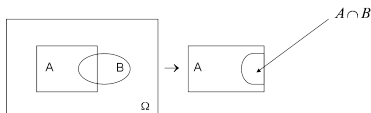
Krankenhaus K

$$\begin{aligned} P(B) &= 100/1000 = 0.1 \\ P(A) &= 100/1000 = 0.1 \\ P(A \cap B) &= 70/1000 = 0.07 \\ P(B|A) &= 0.07/0.1 = 0.7 = 70\% \end{aligned}$$

Schwere Fälle	OP überlebt	OP nicht überl.	P(nicht ü) „Risiko“
Krankenhaus U	400	500	0.56
Krankenhaus K	30	70	0.7

Leichte Fälle	OP überlebt	OP nicht überl.	P(nicht ü) „Risiko“
Krankenhaus U	100	0	0
Krankenhaus K	870	30	0.033

Einschränkung des Ergebnisraumes und bedingte Wahrscheinlichkeit



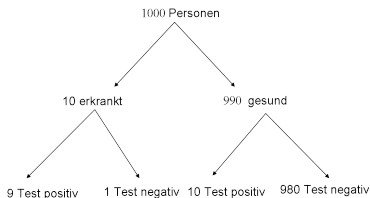
Beispiel: Würfeln

$$\begin{aligned} \Omega &= \{1,2,3,4,5,6\} \\ A &= \{2,4,6\} \text{ „gerade“} \\ B &= \{4,5,6\} \text{ „groß“} \\ A \cap B &= \{4,6\} \end{aligned}$$

$$\begin{aligned} P(A) &= 3/6 \\ P(A \cap B) &= 2/6 \\ P(B|A) &= P(A \cap B)/P(A) = (2/6)/(3/6) = 2/3 \end{aligned}$$

Interpretation:

Wenn bekannt ist, dass die gewürfelte Zahl gerade ist, steigt die Wahrscheinlichkeit für „groß“ auf 2/3.



Beachte: Die Bedingung entspricht der Bezugspopulation:

- 9 von 10 Kranken werden als solche erkannt:
 $P(\text{Test OK (positiv)} \mid \text{Patient krank}) = 9/10$
- 980 von 990 Gesunden werden als solche erkannt:
 $P(\text{Test OK (negativ)} \mid \text{Patient gesund}) = 98/99$
- 9 von 19 Patienten mit positivem Test sind tatsächlich krank:
 $P(\text{Diagnose OK (richtig)} \mid \text{Test positiv}) = 9/19$

Ohne Test: $P(\text{Patient krank}) = 1/100$

Bezugspopulation von zentraler Bedeutung

Beispiel: Wahlverhalten und Schicht

B : zufällig gezogene Person wählt Partei X
 A : zufällig gezogene Person gehört der Oberschicht an

$P(B)$: W'keit, dass zufällig gezogene Person Partei X wählt

$P(B \cap A)$: W'keit, dass zufällig gewählte Person Partei X wählt und der Oberschicht angehört.

$P(B|A)$: W'keit, dass zufällig gewählte Person Partei X wählt, wenn sie der Oberschicht angehört.

Unterschied:

$P(B|A)$: Man betrachtet nur Angehörige der Oberschicht (A ist sicher eingetreten)

Interpretation analog zu bedingten Häufigkeiten in Statistik I: $P(B|A)$ ist die Wahrscheinlichkeit von B wenn man bereits weiß, dass A gilt.

Bezug zu Statistik I

- Die Beziehung zu Statistik I und den bedingten relativen Häufigkeiten ergibt sich, wenn man wieder die durch A und B erzeugte (2×2) -Tafel betrachtet.
- An Stichprobenmodell denken: Grundgesamtheit Ω

$$P(B) \hat{=} f_{\bullet 1}, \quad P(A \cap B) \hat{=} f_{11}$$

$\hat{=}$ „entspricht“

	1	2	
1	f_{11}	f_{12}	$f_{1\bullet}$
2	f_{21}	f_{22}	$f_{2\bullet}$
	$f_{\bullet 1}$	$f_{\bullet 2}$	

- 1000 HörerInnen
- 600 männlich (M), davon 450 positiv (Pos)
- 400 weiblich (W), davon 300 positiv (Pos)

Wir ziehen zufällig 1 Hörer.

$$P(M) = 0.6$$

$$P(W) = 0.4$$

$$P(Pos) = 0.75$$

$$P(M \cap Pos) = 0.45$$

$$P(Pos|M) = P(M \cap Pos)/P(M) = 0.45/0.6 = 0.75 = P(Pos)$$

$$P(M|Pos) = P(M \cap Pos)/P(Pos) = 0.45/0.75 = 0.6 = P(M)$$

Interpretation: Die Ereignisse „Männlich“ und „Positiv“ sind unabhängig.

Äquivalent sind unter $P(A) > 0$ und $P(B) > 0$

1. $P(B|A) = P(B)$
2. $P(A|B) = P(A)$
3. $P(A \cap B) = P(A) \cdot P(B)$

Definition

Zwei Ereignisse A und B heißen *stochastisch unabhängig* wenn gilt:

$$P(A \cap B) = P(A) \cdot P(B)$$

Diese Definition der Unabhängigkeit besitzt den Vorteil, dass man nicht $P(A) = 0$, $P(B) = 0$ ausschließen muss; für die Interpretation ist aber die Bezugnahme auf bedingte Wahrscheinlichkeiten viel anschaulicher.

Koppelung von unabhängigen Experimenten

Mit dem Begriff der Unabhängigkeit (und bedingten Wahrscheinlichkeiten) kann man komplexere Situationen aus „Einzelbausteinen“ zusammensetzen:

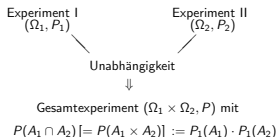
Bisher: Unabhängigkeit als zu überprüfende Eigenschaft

$$P(A_1 \cap A_2) = P(A_1) \cdot P(A_2) \implies \text{unabhängig.}$$

Jetzt: Unabhängigkeit inhaltlich postulieren. Gegeben $P(A_1)$, $P(A_2)$ und A_1 und A_2 unabhängig. Dann gilt:

$$P(A_1 \cap A_2) = P(A_1) \cdot P(A_2).$$

Formale Darstellung



Werfen eines Würfels ($\Omega_1 = \{1, \dots, 6\}$) und eines Oktaeders ($\Omega_2 = \{1, \dots, 8\}$) unabhängig voneinander.

$$A_1 \subset \Omega_1: A_1 = \{5, 6\}, \quad A_2 \subset \Omega_2: A_2 = \{7, 8\}$$

$$A_1 \cap A_2: \text{„5 oder 6 mit Würfel und 7 oder 8 mit Oktaeder“}$$

Dann definiert man

$$P(A_1 \cap A_2) [= P(A_1 \times A_2) =] = P_1(A_1) \cdot P_2(A_2);$$

also erhält man bei einem fairem Würfel und einem fairem Oktaeder mit

$$P_1(\{j\}) = \frac{1}{6}, \quad j = 1, \dots, 6, \quad \text{und} \quad P_2(\{j\}) = \frac{1}{8}, \quad j = 1, \dots, 8,$$

$$P(A_1 \cap A_2) = \frac{1}{3} \cdot \frac{1}{4} = \frac{1}{12}.$$

Diese Konstruktion führt man für alle möglichen $A_1 \subset \Omega_1, A_2 \subset \Omega_2$ durch.

Koppelung abhängiger Experimente

Als nächster Schritt werden komplexere Experimente aus viel einfacheren, voneinander abhängigen Einzelerperimenten gebaut. Gerade bei komplexeren Anwendungen ist es meist bedeutend einfacher, (und auch sicherer, da sich die Chance erhöht, korrektes Expertenwissen zu erhalten) bedingte statt unbedingte Wahrscheinlichkeiten anzugeben. Beispielsweise kann man versuchen, die Wahrscheinlichkeit eines Ereignisses dadurch zu bestimmen, dass man als Zwischenschritt „auf alle Eventualitäten bedingt“ und zunächst die entsprechenden bedingten Wahrscheinlichkeiten bestimmt. (→ Baumstruktur)

Von besonderer Bedeutung ist der Fall *unabhängiger und identischer Wiederholungen*, bei dem dasselbe Experiment wiederholt durchgeführt wird.

Zufallsstichprobe vom Umfang n mit Zurücklegen

Das Experiment „Ziehen einer Person und Ermittlung ihrer Parteipräferenz“ wird n -mal unabhängig (Befragte dürfen sich nicht gegenseitig beeinflussen!) durchgeführt.

Berechnungen und Beispiele folgen später mit Hilfe der Binomialverteilung

Fußball Beispiel

Wie groß ist die Wahrscheinlichkeit, das Halbfinale zu gewinnen ?

Gesucht: $P(B)$ mit $B = \text{„Sieg im Halbfinale“}$

Siegchancen sind abhängig vom jeweiligen Gegner!

⇒ bedingte Wahrscheinlichkeiten.

A_1	Gegner ist Mannschaft	1
A_2	„	2
A_3	„	3

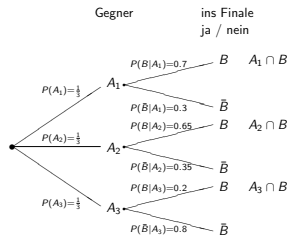
Bedingte Wahrscheinlichkeiten leicht(er) anzugeben:

$$P(B|A_1) = 0.7$$

$$P(B|A_2) = 0.65$$

$$P(B|A_3) = 0.2$$

Gegner wird ausgelost ⇒ Annahme: $P(A_1) = P(A_2) = P(A_3) = \frac{1}{3}$



Welche „Wege“ im Wahrscheinlichkeitsbaum führen zu B ?

$$\left. \begin{aligned} P(A_1 \cap B) &= P(A_1) \cdot P(B|A_1) = \frac{1}{3} \cdot 0.7 \\ P(A_2 \cap B) &= P(A_2) \cdot P(B|A_2) = \frac{1}{3} \cdot 0.65 \\ P(A_3 \cap B) &= P(A_3) \cdot P(B|A_3) = \frac{1}{3} \cdot 0.2 \end{aligned} \right\} \text{insgesamt: } 0.52$$

Verallgemeinerung: Vollständige Zerlegung

- A_1, A_2, A_3 bilden eine vollständige Zerlegung.
- $(A_1 \cap B)$, $(A_2 \cap B)$ und $(A_3 \cap B)$ sind disjunkt und ergeben in der Vereinigung B

Damit ergibt sich

$$\begin{aligned} P(B) &= P((A_1 \cap B) \cup (A_2 \cap B) \cup (A_3 \cap B)) \\ &= P(A_1 \cap B) + P(A_2 \cap B) + P(A_3 \cap B) \\ &= P(B|A_1) \cdot P(A_1) + P(B|A_2) \cdot P(A_2) + P(B|A_3) \cdot P(A_3) = 0.52 \end{aligned}$$

Entlang der Äste multiplizieren, dann summieren
Das Ergebnis lässt sich verallgemeinern auf:

- Beliebige Ereignisse B
- und vollständige Zerlegungen $(A_i)_{i=1, \dots, k}$.

Satz von der totalen Wahrscheinlichkeit

Gegeben sei eine vollständige Zerlegung A_1, A_2, \dots, A_k . Dann gilt für jedes Ereignis B

$$P(B) = \sum_{j=1}^k P(B|A_j) \cdot P(A_j) = \sum_{j=1}^k P(B \cap A_j).$$

Allgemeiner erlauben bedingte Wahrscheinlichkeiten die Modellierung komplexer „Experimente“, welche aus sukzessiven „Einzelexperimenten“ bestehen, bei denen die Ergebnisse jeweils von den vorherigen Experimenten abhängen dürfen (insb. dynamische stochastische Modelle).

- Komplexere Urnenmodelle ohne Zurücklegen, Wahrscheinlichkeit im n -ten Zug ist davon abhängig, welche Kugeln vorher gezogen wurden.
- Sicherheitsstudie zu Kernkraftwerken: Wahrscheinlichkeit für komplexe Pfade praktisch nicht angebar, aber eben bedingte Einzelwahrscheinlichkeiten.
- Markovmodelle (dynamische Modelle mit „einfacher Bedingung“)

- Koppelung abhängiger Experimente
 $\Omega_1 = \Omega_2 = \dots = \Omega_T = \{a_1, \dots, a_k\}$.
- Der Index $t = 1, \dots, T$ bezeichnet die Zeit.
- $A_{t,j}$ ist das Ereignis: Zustand j zum Zeitpunkt t

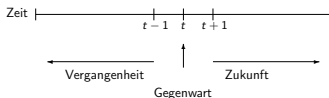
Markov-Eigenschaft:

$$P(A_{t+1,j_{t+1}} | A_{t,j_t} \cap A_{t-1,j_{t-1}} \cap \dots) = P(A_{t+1,j_{t+1}} | A_{t,j_t}), \quad (4.1)$$

Man spricht von einem Markovmodell mit den Zuständen a_1, \dots, a_k .

Sind die so genannten Übergangswahrscheinlichkeiten in (4.1) unabhängig von der Zeit, gilt also $P(A_{t+1,j} | A_{t,l}) \equiv p_{jl}$ für alle t, j, l , so heißt das Markovmodell *homogen*.

Markovmodelle II



Markov-Eigenschaft: „Gegeben den Zustand in der Gegenwart sind Vergangenheit und Zukunft unabhängig, d.h. die Zukunft hängt nur von der Gegenwart ab, aber nicht von der Vergangenheit“.

Für die Prognose der weiteren Entwicklung zählt also nur der aktuelle Stand, nicht aber, wie man dorthin gelangt ist.

Bei sozialen Prozessen ist die Markoveigenschaft und die Homogenität immer kritisch zu hinterfragen!

Markovmodelle: Typische Anwendungen:

- Glücksspiel: Die Wahrscheinlichkeit $P(A_{t+1,j})$ mit $A_{t+1,j}$ = „Spieler hat zum Zeitpunkt $t+1$ Kapitalbestand a_j “ hängt nur vom Kapitalbestand zum Zeitpunkt t ab.
- BWL: Konsumententscheidungen / Produktwahl
- Demographie: Geburts- und Todesprozesse
- Bildet das Wetter mit $\Omega =$
 $\{\text{Sonniger Tag, bewölkter Tag, regnerischer Tag, verschneiter Tag}\}$
 eine Markovkette?
- Soziologie: z.B. Modelle sozialer Mobilität, Mobilität in Betrieben
 - Rapoport (1980): Mathematische Methoden in der Sozialwissenschaft, Physika
 - Bartholomew (1982): Stochastic Models for Social Processes, Wiley

(nach Bartholomew (1982), S. 18f.)

Wie entwickelt sich der soziale Status durch die Generationen?

- Markoveigenschaft bedeutet hier: Status der Kinder nur abhängig vom Status der Eltern, aber nicht mehr zusätzlich vom Status der Großeltern oder vorheriger Generationen
- Homogenität bedeutet hier: Wahrscheinlichkeit für alle denkbaren Statuswechsel zeitlich konstant: z.B. Wechselhäufigkeit von Landwirtschaftssektor in Dienstleistungssektor?

Datengrundlage: männliche Generationenfolge in Marion County, Indiana (1905 – 1912)

		Söhne	a_1	a_2	a_3
Väter	nicht handwerkliche Tätigkeit	a_1	0.594	0.396	0.009
	handwerkliche Tätigkeit	a_2	0.211	0.782	0.007
	landwirtschaftliche Tätigkeit	a_3	0.252	0.641	0.108

Die obige Matrix enthält die (geschätzten) Übergangswahrscheinlichkeiten

$$i\text{-te Zeile, } j\text{-te Spalte: } P(A_{2j}|A_{1i})$$

Wahrscheinlichkeit, dass die **zweite** Generation in Zustand **j** ist unter der Bedingung, dass die **erste** Generation im Zustand **i** ist.

Interpretation Übergangsmatrix

Beispiel: Sohn „nicht handwerklich“ unter der Bedingung Vater „landwirtschaftlich“

$$P(A_{21}|A_{13}) = 0.252$$

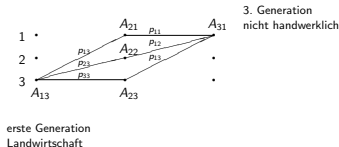
Für feste A_{1i} ist $P(A_{2j}|A_{1i})$ als Funktion in A_{2j} eine Wahrscheinlichkeitsverteilung, d.h. die Zeileneinträge summieren sich zu 1.

Inhaltliche Interpretation:

- Man sieht bei der handwerklichen Tätigkeit eine starke Tendenz zur Statuskonstanz ($P(A_{22}|A_{12}) = 0.782$)
- ähnliches gilt abgeschwächt für die nicht handwerkliche Tätigkeit ($P(A_{21}|A_{11}) = 0.594$),
- während sich der landwirtschaftliche Sektor deutlich auflöst; hier bleibt nur etwa jeder Zehnte ($P(A_{23}|A_{13}) = 0.108$), und ein „Zugewinn“ aus anderen Sektoren findet praktisch nicht statt ($P(A_{23}|A_{11}) = 0.009$, $P(A_{23}|A_{12}) = 0.007$ liegen jeweils unter einem Prozent).

Weitere Berechnungen

Wie groß ist die Wahrscheinlichkeit, dass der Enkel eines in der Landwirtschaft Tätigen eine Tätigkeit im nicht handwerklichen Sektor ausüben wird? Wahrscheinlichkeit, dass die 3. Generation nicht handwerklich unter der Bedingung, dass 1. Generation in Landwirtschaft



$$\begin{aligned}
 P(A_{31}|A_{13}) &= \sum_{l=1}^3 P(A_{31} \cap A_{2l}|A_{13}) \\
 &= \sum_{l=1}^3 P(A_{31}|(A_{2l} \cap A_{13})) \cdot P(A_{2l}|A_{13}) \\
 &= \sum_{l=1}^3 P(A_{31}|A_{2l}) \cdot P(A_{2l}|A_{13}) \\
 &= \sum_{l=1}^3 p_{l3} \cdot p_{1l} = p_{13} \cdot p_{11} + p_{23} \cdot p_{12} + p_{33} \cdot p_{13} \\
 &= 0.252 \cdot 0.594 + 0.641 \cdot 0.211 + 0.108 \cdot 0.252 = 0.312
 \end{aligned}$$

Kritische Aspekte

- + interessantes und sehr leistungsfähiges Modellierungsinstrument aber nicht in Ehrfrucht vor Methode erstarren, sondern Annahme kritisch hinterfragen
- Markoveigenschaft nicht unproblematisch: zusätzliche Rolle der Großväter!
- Zeitliche Homogenität problematisch (in der Tat gute 30 Jahre später p_{33} nochmals mehr als halbiert)

Prognosen

Kennt man die Startverteilung $P(A_{11}), P(A_{12}), P(A_{13})$, so kann man die weitere Verteilung auf die Sektoren berechnen.

$$\begin{aligned}
 P(A_{2j}) &= \sum_{m=1}^3 P(A_{2j}|A_{1m}) \cdot P(A_{1m}) \\
 P(A_{3j}) &= \sum_{l=1}^3 P(A_{3j}|A_{2l}) \cdot P(A_{2l}) = \\
 &= \sum_{l=1}^3 P(A_{3j}|A_{2l}) \cdot \sum_{m=1}^3 P(A_{2l}|A_{1m}) \cdot P(A_{1m}) = \\
 &= \sum_{l=1}^3 p_{jl} \cdot \sum_{m=1}^3 p_{lm} \cdot P(A_{1m}) = \sum_{l=1}^3 \sum_{m=1}^3 p_{jl} p_{lm} \cdot P(A_{1m})
 \end{aligned}$$

usw.

Man kann auch (mit weiterführenden Methoden) eine Gleichgewichtsverteilung bestimmen.

Das Theorem von Bayes

Ziel: Vertauschen von „Bedingung und Ereignis“

Also: gegeben $P(B|A)$, gesucht $P(A|B)$

Satz von Bayes

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A) \cdot P(B|A)}{P(B|A) \cdot P(A) + P(B|A^c) \cdot P(A^c)}$$

Allgemeiner nicht nur bei Dichotomie A und \bar{A} , sondern bei beliebiger vollständiger Zerlegung A_1, \dots, A_k anwendbar:

Satz von Bayes (allgemeine Formulierung)

Sei A_1, \dots, A_k eine vollständige Zerlegung von Ω (wobei $P(A_i) > 0$, $P(B|A_i) > 0$, $i = 1, \dots, k$ und $P(B) > 0$ erfüllt seien). Dann gilt

$$P(A_j|B) = \frac{P(B|A_j) \cdot P(A_j)}{\sum_{i=1}^k P(B|A_i) \cdot P(A_i)}$$

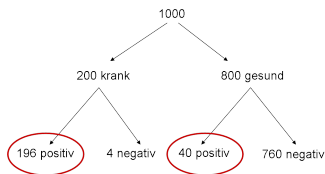
T+: Test positiv
T-: Test negativ

K+: krank
K-: gesund

- Sensitivität $P(T+ | K+) = 0.98$
- Spezifität $P(T- | K-) = 0.95$
- Prävalenz $P(K+) = 0.2 \Rightarrow P(K-) = 0.8$

Wie hoch ist die Wahrscheinlichkeit, dass eine positiv getestete Person wirklich krank ist?

Prädiktiver Wert = $P(K+ | T+)$



Positiver prädiktiver Wert: $\frac{196}{196+40} = 0.83$

Lösung durch Satz von Bayes

- Sensitivität $P(T+ | K+) = 0.98$
- Spezifität $P(T- | K-) = 0.95$
- Prävalenz $P(K+) = 0.2$

Dann gilt für den positiven prädiktiven Wert

$$\begin{aligned}
 P(K+ | T+) &= \frac{P(K+ \cap T+)}{P(T+)} \\
 &= \frac{P(K+) \cdot P(T+ | K+)}{P(K+) \cdot P(T+ | K+) + P(K-) \cdot P(T+ | K-)} \\
 &= \frac{0.2 \cdot 0.98}{0.2 \cdot 0.98 + 0.8 \cdot 0.05} \\
 &= \frac{0.196}{0.196 + 0.040}
 \end{aligned}$$

Inhaltliche Bemerkungen

- Problematik: Flächendeckendes Screening nicht unumstritten, da viele falsch-positive Ergebnisse. Gegenposition: Anwendung nur auf Risikopatienten.
- Bei Mammographie oder PSA-Test auf Prostatakrebs teilweise sogar noch viel geringere Spezifität.
- Wert der mathematischen Theorie: Wenn es etwas komplexer wird, verlässt einen sofort der „gesunde Menschenverstand“. Untersuchungen (u.a. von Gigerenzer) haben gezeigt, dass viele Ärzte sich dieser Problematik nicht bewusst sind.

Das Theorem von Bayes liefert ein Schema für das probabilistische Lernen aus Beobachtungen („Aufdatieren von Wahrscheinlichkeiten“).

$$\left. \begin{array}{l} \text{priori} \\ + \text{ Daten} \end{array} \right\} \rightarrow \text{posteriori}$$

Es dient als Grundlage der sog. *Bayesianischen Inferenz*, einer bestimmten Schule der statistischen Methodologie. Dabei geht es darum, aus Daten zu lernen, indem man die subjektiven Wahrscheinlichkeiten $P(A_i)$ für bestimmte Modellparameter mit Hilfe der Daten (B) aufdatiert, und somit zur besseren Wahrscheinlichkeitsaussagen für die Modellparameter kommt.

Frage: Mit welcher Wahrscheinlichkeit trifft Schweinsteiger beim Elfmeter. Gesucht ist also ein der Parameter P_S . Es werden dazu 3 Experten gefragt:

$$\begin{array}{ll} \text{FB: } M_1 & P_S = 1 \\ \text{JH: } M_2 & P_S = 0.6 \\ \text{JL: } M_3 & P_S = 0.9 \end{array}$$

Priori - Wahrscheinlichkeiten: $P(T_1) = P(T_2) = P(T_3) = \frac{1}{3}$

Daten D: 1 Elfmeter 1 Treffer

Gesucht: $P(M_i|D)$

Lösung mit Satz von Bayes:

$$P(M_1|D) = \frac{P(M_1) \cdot P(D|M_1)}{P(D|M_1) \cdot P(M_1) + P(D|M_2) \cdot P(M_2) + P(D|M_3) \cdot P(M_3)}$$

Es ergibt sich: $P(M_1|D) = 0.4$; $P(M_2|D) = 0.24$; $P(M_3|D) = 0.36$

Beispiel zur Bayes Inferenz II: 8 von 10

Weitere Berechnung für 10 Elfmeter 8 Treffer (D2)

Es ergibt sich: $P(M_1|D2) = 0$; $P(M_2|D2) = 0.13$; $P(M_3|D2) = 0.87$

- Daten widerlegen Modell 1
- Daten sprechen für Modell 3
- Inferenz durch priori und Daten

Bayes Inferenz

- Modelle M_i mit Priori Wahrscheinlichkeiten $P(M_i)$
- Daten D, die mit den Modellen auf verschiedene Weisen beschrieben werden
- Modellspezifikation liefert für jedes Modell $P(D|M_i)$ Diese wird auch als likelihood bezeichnet
- Satz von Bayes liefert:

$$P(M_i|D) = P(D|M_i) * P(M_i) * \frac{1}{P(D)} = \text{Likelihood} * \text{Priori} * \text{Konstante}$$

- Verallgemeinerung auf komplexe Situationen (z.B. Regressionsmodelle) möglich
- Expertenwissen und frühere Analysen können durch priori Verteilung eingebracht werden



- 1 Einführung
- 1 Wahrscheinlichkeitsrechnung
- 2 Zufallsvariablen und ihre Verteilung**
- 3 Statistische Inferenz
- 3 Hypothesentests
- 3 Regression

Ergebnisse von Zufallsexperimenten werden als Zahlen dargestellt

Beispiele:

1. Augenzahl beim Werfen zweier Würfel
2. Zeit beim Warten auf den Bus
3. Antwort ja = 1, nein = 0

Formal: Eine Zufallsgröße oder Zufallsvariable ist eine Abbildung:

$$X : \Omega \rightarrow \mathbb{R}$$

(Abbildung des Ergebnisraums auf die reellen Zahlen)

Im Beispiel 1:

$$\begin{aligned} (1,1) &\rightarrow 2 \\ (1,2) &\rightarrow 3 \\ (2,1) &\rightarrow 3 \\ (2,2) &\rightarrow 4 \end{aligned}$$

Wahrscheinlichkeitsverteilung einer diskreten Zufallsgröße

Betrachte ein Spiel mit den Gewinnen:

ω	$X(\omega)$
≤ 3	10 €
= 4, 5	20 €
= 6	100 €

Die Wahrscheinlichkeiten P_X ergeben sich wie folgt:

$$\begin{aligned} P_X(\{10\}) &= P_X(\text{man erhält 10€}) \\ &= P(\text{man hat etwas gewürfelt, das zu 10€ führt}) \\ &= P(\{1, 2, 3\}) = \frac{1}{2} \end{aligned}$$

$$\begin{aligned} P_X(\{20\}) &= P_X(\text{von allem, das zu 20€ führt}) \\ &= P(\{4, 5\}) = \frac{2}{6} \end{aligned}$$

$$P_X(\{100\}) = P_X(\{6\}) = \frac{1}{6}$$

Eine Zufallsgröße heißt diskret, falls sie nur endlich viele oder abzählbar viele Werte annehmen kann (typischerweise ganze Zahlen)

- P_X heißt *Wahrscheinlichkeitsverteilung* von X .
- X (als Variable) beschreibt den Ausgang eines Zufallsexperiments vor der *Durchführung* (Auszahlungsregel beim Würfelspiel: wenn 3 dann 10 Euro, wenn ..., dann ...).
- x (als Realisation) gibt den Wert der Variablen nach Durchführung des Zufallsexperiments an (daher „Realisation“, konkreter Auszahlungsbetrag).
- In der Verwendung analog zur Unterscheidung Merkmal / Merkmalsausprägung in Statistik I.
- Es ist häufig üblich, bei P_X den Index wegzulassen, also $P(\{x\})$ statt $P_X(\{x\})$ zu schreiben.

Definition

Die Menge

$$\mathcal{X} := \{x \in \mathbb{R} \mid P(\{x\}) > 0\}$$

heißt Träger von X .

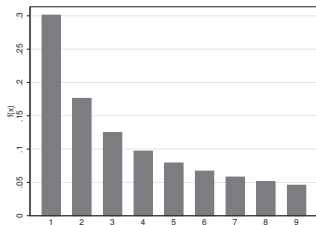
Die Wahrscheinlichkeitsfunktion $f(x)$ einer diskreten Zufallsvariable X ist für $x \in \mathbb{R}$ definiert durch

$$f(x) = \begin{cases} P(X = x_i) = p_i, & x = x_i \in \mathcal{X} = \{x_1, x_2, \dots, x_k, \dots\} \\ 0, & \text{sonst.} \end{cases}$$

Newcomb (1835–1909) und später Frank Benford (1883–1948) machten die verblüffende Entdeckung, dass die Anfangsziffern 1–9 von ganzen Zahlen in vielen Fällen nicht gleich häufig vorkommen. Am häufigsten ist die Anfangsziffer 1, am zweithäufigsten die Anfangsziffer 2 usw. Beispiele sind

- die Häufigkeit der Anfangsziffern von Zahlen in Zeitungsartikeln
- die Häufigkeit der Anfangsziffern von Steuerdokumenten
- die Häufigkeit der ersten Ziffer der Dateigröße von gespeicherten Dateien.

Wahrscheinlichkeitsfunktion I (Benfords Gesetz)



Wahrscheinlichkeitsfunktion II (Benfords Gesetz)

Benford publizierte für die Zufallsvariable

$X =$ „Anfangsziffer von Zahlen“

die Wahrscheinlichkeitsfunktion

$$f(x) = P(X = x) = \begin{cases} \log_{10} \left(\frac{x+1}{x} \right), & x = 1, \dots, 9 \\ 0, & \text{sonst} \end{cases}$$

Benfords Gesetz findet zum Beispiel Anwendung bei der Fahndung nach Steuerbetrügern, bei der Überprüfung von Wahlergebnissen

Sei X die Zufallsvariable *Anzahl der Haushaltsmitglieder* mit der Verteilung

$$P(\{X=1\})=0.4$$

$$P(\{X=2\})=0.3$$

$$P(\{X=3\})=0.2$$

$$P(\{X=4\})=0.1$$

(Annahme: Nur bis zu 4-Personen-Haushalte).

Man berechne die Wahrscheinlichkeit, bei einfacher Zufallsauswahl vom Umfang 1 einen Mehrpersonenhaushalt zu erhalten und die Wahrscheinlichkeit des Ereignisses „Die Zahl der Haushaltsmitglieder ist gerade“.

$$\begin{aligned}P(\{X > 1\}) &= P(X = 2) + P(X = 3) + P(X = 4) \\ &= 0.3 + 0.2 + 0.1 \\ &= 0.6\end{aligned}$$

$$P(\{X \text{ gerade}\}) = 0.3 + 0.1 = 0.4$$

Verteilungsfunktion II

Die Wahrscheinlichkeitsverteilung einer (diskreten) Zufallsvariablen X ist durch die Verteilungsfunktion eindeutig gegeben.

Es gilt zum Beispiel

$$P(a < X \leq b) = F(b) - F(a).$$

Die Ereignisse $\{X \leq a\} = \{\omega | X(\omega) \leq a\}$, $\{a < X \leq b\}$ und $\{X > b\}$ sind disjunkt und ergeben in ihrer Vereinigung Ω . Also gilt

$$\begin{aligned}1 &= P(\Omega) = P(X \leq a) + P(a < X \leq b) + P(X > b) \\ \Leftrightarrow 1 - P(X \leq a) - P(X > b) &= P(a < X \leq b) \\ \Leftrightarrow P(X \leq b) - P(X \leq a) &= P(a < X \leq b)\end{aligned}$$

Viele interessierende Ereignisse besitzen folgende Form:

$$\{X \leq a\} \quad \text{oder} \quad \{X \in [a, b]\} = \{a \leq X \leq b\},$$

wobei a und b feste reelle Zahlen sind.

$P(\{X \leq a\})$ für variables a entspricht der empirischen Verteilungsfunktion.

Definition

Sei X eine Zufallsvariable. Die Funktion

$$F: \mathbb{R} \rightarrow [0; 1]$$

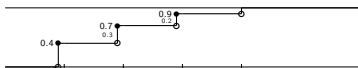
$$x \mapsto F(x)$$

$$F(x) := P(X \leq x)$$

heißt *Verteilungsfunktion*.

Beispiel

- $x < 1$:
 $F(x) = P(X \leq x) = P(X < 1) = 0$
- $x = 1$:
 $F(x) = P(X \leq x) = P(X \leq 1) = 0.4$
- $1 < x < 2$:
 $F(x) = P(X \leq x) = P(X \leq 1) + P(1 < X \leq x) = P(X \leq 1) = 0.4$
- $x = 2$:
 $F(x) = P(X \leq x) = P(X \leq 1) + P(X = 2) = 0.4 + 0.3 = 0.7$
- $x \leq 2$:
 $F(x) = P(X \leq x) = P(X \leq 2) = 0.7$
- $x = 3$:
 $F(x) = P(X \leq 3) = P(X \leq 2) + P(X = 3) = 0.7 + 0.2 = 0.9$
- $3 < x < 4$:
 $F(x) = P(X \leq x) = P(X \leq 3) = 0.9$
- $x = 4$:
 $F(x) = P(X \leq 4) = P(X \leq 3) + P(X = 4) = 1$
- $x > 4$:
 $F(x) = P(X \leq x) = 1$



Allgemein gilt:

$F(x)$ ist eine stückweise konstante Treppenfunktion und $P(X = x)$ ist genau die Sprunghöhe der Verteilungsfunktion im Punkt x .

Konzept der Dichtefunktion

Beispiel:

Wir betrachten eine Zufallsvariable T mit Wertebereich im Intervall $[0; 10]$

Warten auf den Bus, der alle 10 Minuten fährt. T kann also jeden Wert zwischen 0 und 10 annehmen.

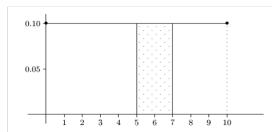
gesucht: $P(T=2) = ?$

$$P(T=2) = P(1.5 < T < 2.5) = 1/10$$

$$P(T=2) = P(1.99 < T < 2.01) = 2/1000$$

$$P(T=2) = 0 ???$$

Berechnung von Wahrscheinlichkeiten

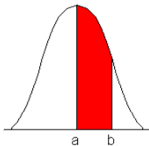


$$P(5 \leq T \leq 7) = \text{Fläche unter der Kurve}$$

Definition

Zu einer stetigen Zufallsgröße ist f Dichtefunktion, falls gilt:

$$P(a \leq X \leq b) = \int_a^b f(x) dx$$



- $f(x) \geq 0$
- $\int_{-\infty}^{\infty} f(x) dx = 1$
- $F(t) = \int_{-\infty}^t f(x) dx$
- $F'(x) = f(x)$ (Dichte ist Ableitung der Verteilungsfunktion)

Beispiel: Warten auf den Bus

- Verteilungsfunktion

$$F(x) = P(X \leq x) = \begin{cases} 0 & x < 0 \\ 0.1x & 0 \leq x \leq 10 \\ 1 & x > 10 \end{cases}$$

- Dichtefunktion

$$f(x) = \begin{cases} 0.1 & 0 \leq x \leq 10 \\ 0 & \text{sonst} \end{cases}$$

Eigenschaften Verteilungsfunktion

Allgemeiner als zuvor gilt hier

$$\begin{aligned} P(a < X \leq b) &= P(a \leq X \leq b) \\ &= P(a < X < b) = F(b) - F(a) \end{aligned}$$

da $P(X = a) = P(X = b) = 0$.

- Stetige Zufallsvariablen sind in der Wahrscheinlichkeitsrechnung sehr wichtig.
- Insbesondere ergeben sich Approximationsmöglichkeiten für diskrete durch stetige Zufallsvariablen bei größeren Stichprobenumfängen. Damit lassen sich zahlreiche Berechnungen vereinfachen (auch wenn die stetige Formulierung zunächst komplizierter wirkt).

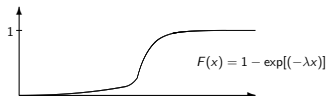
Eine Zufallsvariable X heißt stetig, wenn es eine Funktion $f(x) \geq 0$ gibt, so dass für jedes Intervall $[a, b]$

$$P(a \leq X \leq b) = \int_a^b f(x) dx = \text{Fläche zwischen } a \text{ und } b \text{ unter der Funktion}$$

gilt.

Typische Verteilungsfunktion I

(z.B. zur Beschreibung der Dauer von Arbeitslosigkeit)



Typische Verteilungsfunktion II

Die Kurve ist unterschiedlich steil. Sie hat zwar in keinem Punkt eine Sprungstelle ($P(X = x) = 0$), aber in jedem kleinen Intervall um x ist:

$$P(x - h < X < x + h) = F(x + h) - F(x - h)$$

durchaus unterschiedlich.

Die Steigung

$$\lim_{h \rightarrow 0} \frac{F(x + h) - F(x - h)}{2h}$$

enthält also wesentliche Information über P . Diese entspricht der Dichtefunktion.

Ziel: Charakterisiere Verteilungen von Zufallsvariablen durch Kenngrößen (in Analogie zu Lage- und Streuungsmaßen der deskriptiven Statistik).

Insbesondere:

- a) „durchschnittlicher Wert“ \rightarrow Erwartungswert, z.B.
- „mittleres“ Einkommen,
 - „durchschnittliche“ Körpergröße,
 - fairer Preis eines Spiels.
- b) Streuung (Dispersion), z.B. wie stark schwankt das Einkommen, die Körpergröße etc.

X sei eine diskrete Zufallsgröße mit den möglichen Werten x_1, \dots, x_n .

Dann ist der Erwartungswert $\mathbb{E}(X)$:

$$\mathbb{E}(X) = \sum_{i=1}^n x_i P(X = x_i)$$

„Der Wert, der sich bei häufiger Wiederholung als Mittelwert ergibt.“

Beispiele Erwartungswert

- Würfelfwurf:
 $\mathbb{E}(X) = \frac{1}{6} \cdot 1 + \frac{1}{6} \cdot 2 + \frac{1}{6} \cdot 3 + \frac{1}{6} \cdot 4 + \frac{1}{6} \cdot 5 + \frac{1}{6} \cdot 6 = 3.5$
- Summe zweier Würfel:
 $\mathbb{E}(S) = \frac{1}{36} \cdot 2 + \frac{2}{36} \cdot 3 + \dots + \frac{2}{36} \cdot 11 + \frac{1}{36} \cdot 12 = 7$
- Antwort ja oder nein:
 $\mathbb{E}(X) = P(X = 0) \cdot 0 + P(X = 1) \cdot 1 = P(X = 1)$

Erwartungswert stetiger Zufallsgrößen

Erwartungswert stetiger ZG:

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x \cdot f(x) dx$$

Integral statt Summe, Dichte statt Wahrscheinlichkeit

Beispiel: Warten auf den Bus

$$\begin{aligned} \mathbb{E}(T) &= \int_{-\infty}^{\infty} x f(x) dx \\ &= \int_0^{10} \frac{1}{10} x dx = 5 \end{aligned}$$

- Lageparameter: Erwartungswert
- Streuungsparameter: Varianz und Standardabweichung

Wie stark weichen die Ausprägungen im Durchschnitt vom Erwartungswert ab?

$$\text{diskret: } \text{Var}(X) = \mathbb{E}((X - \mathbb{E}(X))^2) = \sum_{i=1}^n (x_i - \mathbb{E}(X))^2 P(X = x_i)$$

$$\text{stetig: } \text{Var}(X) = \mathbb{E}((X - \mathbb{E}(X))^2) = \int_{-\infty}^{\infty} (x - \mathbb{E}(X))^2 f(x) dx$$

$$\sigma_X = \sqrt{\text{Var}(X)}$$

Y: Einmal Würfeln und Multiplikation mit 2

$$\begin{aligned} \mathbb{E}(Y) &= 7 \\ \text{Var}(Y) &= \frac{1}{6} \cdot (2-7)^2 + \frac{1}{6} \cdot (4-7)^2 + \frac{1}{6} \cdot (6-7)^2 \\ &\quad + \frac{1}{6} \cdot (8-7)^2 + \frac{1}{6} \cdot (10-7)^2 + \frac{1}{6} \cdot (12-7)^2 \\ &= 11.67 \\ \sigma &= 3.41 \end{aligned}$$

Beispiel II zur Varianz II

S: Würfeln mit 2 Würfeln

$$\begin{aligned} \mathbb{E}(S) &= 7 \\ \text{Var}(S) &= \frac{1}{36} \cdot (2-7)^2 + \frac{2}{36} \cdot (3-7)^2 + \frac{3}{36} \cdot (4-7)^2 + \dots \\ &\quad + \frac{2}{36} \cdot (11-7)^2 + \frac{1}{36} \cdot (12-7)^2 \\ &= 5.833 \\ \sigma &= 2.41 \end{aligned}$$

Varianz bei der Wartezeit auf den Bus

$$\begin{aligned} \text{Var}(T) &= \int_{-\infty}^{\infty} (x-5)^2 f(x) dx \\ &= \int_0^{10} (x-5)^2 \frac{1}{10} dx \\ &= \frac{25}{3} \\ \sigma_T &= \sqrt{\frac{25}{3}} = 2.9 \end{aligned}$$

- Die Varianz gibt die mittlere quadratische Abweichung vom Erwartungswert an. Durch das Quadrieren werden Abweichungen nach unten (negative Werte) auch positiv gezählt.
- Damit Erwartungswert und Varianz sinnvoll interpretiert werden können, muss eine metrische Skala zugrundeliegen.
- Allgemein bezeichnet man $\mathbb{E}(X^k)$ als *k-tes Moment*.

Es gilt:

$$\text{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$$

Quadrat in der Klammer Quadrat außerhalb der Klammer

- Verschiebungssatz für theoretische Überlegungen und Übungsaufgaben gutes Hilfsmittel
- Für Berechnungen mit dem Computer sollte er nicht benutzt werden (numerische Probleme)

Erwartungswert von linearen Transformationen

Der Erwartungswert lässt sich bei linearen Transformationen berechnen durch:

$$Y = a + b \cdot X$$

Dann folgt:

$$\mathbb{E}(Y) = a + b \cdot \mathbb{E}(X)$$

„Erwartungswert ist linear“

Beispiel

Einfacher Würfelwurf: X

Lineare Transformation: $Y = 10 \cdot X - 20$

„Ich zahle 20 € und erhalte das 10fache meiner Zahl.“

$$\mathbb{E}(Y) = 10 \cdot \mathbb{E}(X) - 20 = 10 \cdot 3.5 - 20 = 15$$

„Ich gewinne im Mittel 15 € pro Spiel.“

$$Y = a + b \cdot X$$

$$\text{Var}(Y) = b^2 \cdot \text{Var}(X)$$

$$\sigma_Y = |b| \cdot \sigma_X$$

Verschiebungen ändern nichts an Streuung

X: Einmal Würfeln

Y: Einmal Würfeln und Multiplikation mit 2

$$\begin{aligned}\text{Var}(X) &= \frac{1}{6} \cdot (1 - 3.5)^2 + \frac{1}{6} \cdot (2 - 3.5)^2 + \frac{1}{6} \cdot (3 - 3.5)^2 \\ &\quad + \frac{1}{6} \cdot (4 - 3.5)^2 + \frac{1}{6} \cdot (5 - 3.5)^2 + \frac{1}{6} \cdot (6 - 3.5)^2 \\ &= 2.917 \\ \sigma_X &= 1.705\end{aligned}$$

$$\begin{aligned}\text{Var}(Y) &= 4 \cdot 2.917 = 11.67 \\ \sigma_Y &= 2 \cdot 1.705 = 3.41\end{aligned}$$

Unabhängigkeit von Zufallsgrößen

Zwei Zufallsgrößen X und Y heißen unabhängig, falls alle zu X gehörigen Ereignisse von allen zu Y gehörigen Ereignissen unabhängig sind.

Beispiele:

X: Antwort der 1. Person

Y: Antwort der 2. Person

X: 1. Würfelwurf

Y: 2. Würfelwurf

Erwartungswert von Summen von Zufallsgrößen

Für beliebige Zufallsgrößen X_1 und X_2 gilt:

$$\mathbb{E}(X_1 + X_2) = \mathbb{E}(X_1) + \mathbb{E}(X_2)$$

Beispiele:

- zweimaliges Würfeln
- Ziehen von 2 Personen

Beachte: Unabhängigkeit wird nicht vorausgesetzt

Für **unabhängige** Zufallsgrößen X_1 und X_2 gilt:

$$\text{Var}(X_1 + X_2) = \text{Var}(X_1) + \text{Var}(X_2)$$

Beispiele:

- zweimaliges Würfeln
- Ziehen von 2 Personen

Beachte: Unabhängigkeit ist wichtige Voraussetzung

Bemerkungen II

Im Allgemeinen gilt:

$$\mathbb{E}(g(X)) \neq g(\mathbb{E}(X))$$

also z.B.

$$\mathbb{E}\left(\frac{1}{X}\right) \neq \frac{1}{\mathbb{E}(X)}$$

und

$$\mathbb{E}(X^2) \neq (\mathbb{E}(X))^2.$$

- Der Erwartungswert ist immer additiv aufspaltbar, die Varianz dagegen nur bei Unabhängigkeit!
- Die Additivität der Varianz unter Unabhängigkeit gilt nicht für die Standardabweichung σ :

$$\sqrt{\text{Var}(X + Y)} \neq \sqrt{\text{Var}(X)} + \sqrt{\text{Var}(Y)}$$

- Man beachte explizit, dass gilt $\text{Var}(-X) = \text{Var}(X)$ und damit unter Unabhängigkeit

$$\text{Var}(X - Y) = \text{Var}(X) + \text{Var}(Y).$$

$$\text{Var}(X - Y) = \text{Var}(X) + \text{Var}(-Y) = \text{Var}(X) + (-1)^2 \cdot \text{Var}(Y)$$

Standardisierte Zufallsvariable

Definition

Die Zufallsvariable

$$Z := \frac{X - \mathbb{E}(X)}{\sqrt{\text{Var}(X)}}$$

heißt *standardisierte Zufallsvariable*. Es gilt

$$\mathbb{E}(Z) = 0 \text{ und } \text{Var}(Z) = 1.$$

$$\begin{aligned} \mathbb{E}(Z) &= \mathbb{E}\left(\frac{X - \mathbb{E}(X)}{\sqrt{\text{Var}(X)}}\right) = \frac{1}{\sqrt{\text{Var}(X)}} \cdot \mathbb{E}(X - \mathbb{E}(X)) \\ &= \frac{1}{\sqrt{\text{Var}(X)}} \cdot (\mathbb{E}(X) - \mathbb{E}(\mathbb{E}(X))) = \frac{1}{\sqrt{\text{Var}(X)}} \cdot (\mathbb{E}(X) - \mathbb{E}(X)) = 0 \end{aligned}$$

$$\begin{aligned} \text{Var}(Z) &= \text{Var}\left(\frac{X - \mathbb{E}(X)}{\sqrt{\text{Var}(X)}}\right) = \text{Var}\left(\frac{X}{\sqrt{\text{Var}(X)}} - \frac{\mathbb{E}(X)}{\sqrt{\text{Var}(X)}}\right) \\ &= \text{Var}\left(\frac{X}{\sqrt{\text{Var}(X)}}\right) = \left(\frac{1}{\sqrt{\text{Var}(X)}}\right)^2 \cdot \text{Var}(X) = 1 \end{aligned}$$

Beim Spiel Chuck-a-Luck werden drei Würfel geworfen. Der Spieler setzt auf eine der Zahlen 1, 2, 3, 4, 5, 6. Zeigt keiner der Würfel die gesetzte Zahl, so ist der Einsatz verloren. Andernfalls erhält der Spieler (zusätzlich zu seinem Einsatz) für jeden Würfel, der die gesetzte Zahl zeigt, einen Betrag in Höhe des Einsatzes. Wahrscheinlichkeitsfunktion des Gewinns nach einem Spiel:

G = Gewinn	Würfelkombinationen	Anzahl	Wahrscheinlichkeit
3	666	1	1/216
2	66a, 6a6, a66 mit a=1,2,3,4,5	15	15/216
1	6ab, a6b, ab6, mit a,b=1,2,3,4,5	75	75/216
-1	abc mit a,b,c=1,2,3,4,5	125	125/216
Summe		216	1

Für den Erwartungswert erhält man

$$E(G) = 3 \cdot \frac{1}{216} + 2 \cdot \frac{15}{216} + 1 \cdot \frac{75}{216} - 1 \cdot \frac{125}{216} = -\frac{17}{216} = -0.078$$

also einen erwarteten Verlust von 7.8% des Einsatzes.

Chuck-a-Luck: Spielstrategie

Betrachte die Zufallsvariablen:

X_1, X_2, \dots, X_6 Gewinn, wenn beim ersten Wurf ein Einsatz auf 1, 2, ..., 6 gesetzt wird.

Y_1, Y_2, \dots, Y_6 Gewinn, wenn beim zweiten Wurf ein Einsatz auf 1, 2, ..., 6 gesetzt wird.

Mögliche Spielstrategien und zugehörige Gewinne:

$2X_6$ Gewinn, wenn beim ersten Wurf ein zweifacher Einsatz auf 6 gesetzt wird (Strategie 1).

$X_1 + X_6$ Gewinn, wenn beim ersten Wurf jeweils ein Einsatz auf 1 und 6 gesetzt wird (Strategie 2).

$X_6 + Y_6$ Gewinn, wenn beim ersten und zweiten Wurf ein Einsatz auf 6 gesetzt wird (Strategie 3).

Chuck-a-Luck: Erwartungswerte

- Erwartungswerte:

Aus $E(X_i) = E(Y_j) = -\frac{17}{216}$ folgt:

$$E(2X_6) = 2E(X_6) = -\frac{34}{216}$$

$$E(X_1 + X_6) = E(X_1) + E(X_6) = -\frac{34}{216}$$

$$E(X_6 + Y_6) = E(X_6) + E(Y_6) = -\frac{34}{216}$$

d.h. bei den drei Strategien sind die Erwartungswerte alle gleich!

- Trotzdem gibt es deutliche Unterschiede in den drei Strategien:

Strategie	Wertebereich	$P(\{-2\})$
$2X_6$	-2,2,4,6	0.579
$X_1 + X_6$	-2,0,1,2,3	0.296
$X_6 + Y_6$	-2,0,1,2,3,4,5,6	0.335

Chuck-a-Luck: Varianz

- Varianz des Gewinns nach einem Spiel

$$\begin{aligned}\text{Var}(G) &= \left(3 + \frac{17}{216}\right)^2 \cdot \frac{1}{216} + \left(2 + \frac{17}{216}\right)^2 \cdot \frac{15}{216} + \left(1 + \frac{17}{216}\right)^2 \cdot \frac{75}{216} \\ &\quad + \left(-1 + \frac{17}{216}\right)^2 \cdot \frac{125}{216} \\ &= 0.04388156 + 0.30007008 + 0.40402836 + 0.4911961 = \\ &= 1.2391761\end{aligned}$$

$$\sqrt{\text{Var}(G)} = 1.113183$$

- Nach den Rechenregeln für Varianzen erhält man für die Strategien 1 und 3:

$$\text{Var}(2X_6) = 4 \text{Var}(X_6) = 4 \cdot 1.2391761 = 4.956704$$

und

$$\text{Var}(X_6 + Y_6) = \text{Var}(X_6) + \text{Var}(Y_6) = 1.2391761 + 1.2391761 = 2.4783522.$$



Chuck-a-Luck: Fazit

- * Strategie 1, also $2X_6$, ist am riskantesten.
- * Die Gewinnchancen sind bei Strategie 1 aber größer als bei Strategie 2.
- * Am wenigsten riskant ist Strategie 2.



Chuck-a-Luck: Varianz

- Da X_1 und X_6 nicht unabhängig sind, muss hier die Varianz explizit berechnet werden.
- Wahrscheinlichkeitsverteilung von $X_1 + X_6$:

x	-2	0	1	2	3
$P(X_1 + X_2 = x)$	0.29630	0.44444	0.11111	0.12037	0.02778

$$\begin{aligned}\text{Var}(X_1 + X_6) &= \left(-2 + \frac{34}{216}\right)^2 \cdot 0.29630 + \left(0 + \frac{34}{216}\right)^2 \cdot 0.44444 + \\ &\quad + \left(1 + \frac{34}{216}\right)^2 \cdot 0.11111 + \left(2 + \frac{34}{216}\right)^2 \cdot 0.12037 + \\ &\quad + \left(3 + \frac{34}{216}\right)^2 \cdot 0.02778 = \\ &= 2.003001\end{aligned}$$



Binomialverteilung

Konstruktionsprinzip:

- Ein Zufallsexperiment wird n mal unabhängig durchgeführt.
- Wir interessieren uns jeweils nur, ob ein bestimmtes Ereignis A eintritt oder nicht.
- X = „Häufigkeit, mit der Ereignis A bei n unabhängigen Versuchen eintritt“.
- Träger von X : $\mathcal{X} = \{0, 1, 2, \dots, n\}$.



- Bezeichne $\pi = P(A)$ die Wahrscheinlichkeit für A in *einem* Experiment.
- Das Ereignis $X = x$ tritt z.B. auf, wenn in den ersten x Versuchen A eintritt und anschließend nicht mehr. Die Wahrscheinlichkeit dafür ist

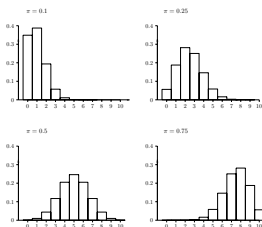
$$P(A_1 \cap \dots \cap A_x \cap \bar{A}_{x+1} \cap \dots \cap \bar{A}_n) = \underbrace{\pi \cdot \dots \cdot \pi}_{x \text{ mal}} \cdot \underbrace{(1 - \pi) \cdot \dots \cdot (1 - \pi)}_{n-x \text{ mal}} = \pi^x (1 - \pi)^{n-x}.$$

- Insgesamt gibt es $\binom{n}{x}$ Möglichkeiten für die Verteilung der x Erfolge (Auftreten von A) auf n Plätze. Damit gilt:

$$P(X = x) = \binom{n}{x} \pi^x (1 - \pi)^{n-x}.$$

Wahrscheinlichkeitshistogramme

Binomialverteilungen mit $n = 10$



Definition

Eine Zufallsvariable heißt *binomialverteilt* mit den Parametern n und π , kurz $X \sim B(n, \pi)$, wenn sie die Wahrscheinlichkeitsfunktion

$$f(x) = \begin{cases} \binom{n}{x} \pi^x (1 - \pi)^{n-x}, & x = 0, 1, \dots, n \\ 0, & \text{sonst} \end{cases}$$

besitzt.

Die $B(1, \pi)$ -Verteilung heißt auch *Bernoulliverteilung*.

Erwartungswert und Varianz I

- Zur Berechnung von Erwartungswert und Varianz der Binomialverteilung ist folgende Darstellung hilfreich:

$$X = X_1 + \dots + X_n$$

mit den binären Variablen

$$X_i = \begin{cases} 1 & \text{falls } A \text{ beim } i\text{-ten Versuch eintritt,} \\ 0 & \text{sonst.} \end{cases}$$

- Die X_i sind stochastisch unabhängig mit

$$\begin{aligned} \mathbb{E}(X_i) &= 0 \cdot P(X_i = 0) + 1 \cdot P(X_i = 1) = \pi \\ \text{Var}(X_i) &= \mathbb{E}(X_i^2) - (\mathbb{E}(X_i))^2 = 1 \cdot P(X_i = 1) - \pi^2 = \pi(1 - \pi). \end{aligned}$$

- Erwartungswert der Binomialverteilung:

$$\mathbb{E}(X) = \mathbb{E}(X_1 + \dots + X_n) = \mathbb{E}(X_1) + \dots + \mathbb{E}(X_n) = n\pi$$

Die direkte Berechnung über

$$\mathbb{E}(X) = \sum_{i=1}^n i \binom{n}{i} \pi^i (1-\pi)^{n-i} = \dots = n\pi$$

ist deutlich komplizierter!

- Varianz der Binomialverteilung:

$$\text{Var}(X) = \text{Var}(X_1 + \dots + X_n) = \text{Var}(X_1) + \dots + \text{Var}(X_n) = n\pi(1-\pi)$$



- 100 Wahlberechtigte werden befragt.
- 30% aller Wahlberechtigten wählen Partei S.

→ Wie groß ist die Wahrscheinlichkeit, dass von den 100 Befragten mehr als 50 die Partei S wählen?

$$X \sim B(100, 0.3)$$

$$\begin{aligned} P(X \geq 50) &= P(X = 50) + P(X = 51) + \dots + P(X = 100) \\ &= \binom{100}{50} \cdot 0.3^{50} \cdot 0.7^{50} + \dots \\ &= 0.0000206 \end{aligned}$$



Eigenschaften der Binomialverteilung

- Symmetrieeigenschaft (vertausche Rolle von A und \bar{A}):
Sei $X \sim B(n, \pi)$ und $Y = n - X$. Dann gilt $Y \sim B(n, 1 - \pi)$.
- Summeneigenschaft: Seien $X \sim B(n, \pi)$ und $Y \sim B(m, \pi)$. Sind X und Y unabhängig, so gilt

$$X + Y \sim B(n + m, \pi)$$

Wichtige Voraussetzung: Gleiches π



Poisson Verteilung

(vgl. z.B. Fahrmeir et. al))

Eine weitere wichtige diskrete Verteilung ist die Poisson-Verteilung. Sie modelliert die Anzahl (eher seltener) Ereignisse in einem Zeitintervall (Unfälle, Todesfälle; Sozialkontakte, deviante Verhaltensmuster, etc.).

Definition

Eine Zufallsvariable X mit der Wahrscheinlichkeitsfunktion

$$f(x) = P(X = x) = \begin{cases} \frac{\lambda^x}{x!} e^{-\lambda}, & x \in \{0, 1, \dots\} \\ 0, & \text{sonst} \end{cases}$$

heißt Poisson-verteilt mit Parameter (oder Rate) $\lambda > 0$, kurz $X \sim \text{Po}(\lambda)$.
Es gilt

$$\mathbb{E}(X) = \lambda, \quad \text{Var}(X) = \lambda$$



Die Poisson-Verteilung kann auch als Näherungsmodell für eine Binomialverteilung gesehen werden, wenn die Anzahl der Versuchswiederholungen n groß und die „Trefferwahrscheinlichkeit“ π sehr klein ist (seltene Ereignisse!).

Der Erwartungswert λ ist dann gleich $n \cdot \pi$.

Es gilt also abgekürzt geschrieben

$$X \sim B(n, \pi) \xrightarrow[\pi \text{ klein}]{n \text{ groß}} X \cong Po(n \cdot \pi)$$

Hat man mehrere unabhängige „Poisson-Prozesse“, also dynamische Simulationen, bei denen die Ereignisanzahl Poisson-verteilt ist, also z.B. verschiedene deviante Verhaltensmuster, so ist die Gesamtanzahl der einzelnen Ereignisanzahlen wieder Poisson-verteilt: genauer gilt

Sind $X \sim Po(\lambda_X)$, $Y \sim Po(\lambda_Y)$ voneinander unabhängig, so gilt

$$X + Y \sim Po(\lambda_X + \lambda_Y).$$

Beachte, die Unabhängigkeit (genauer die Unkorreliertheit, siehe später) ist wesentlich. Hat man als Extremfall, z.B. zwei Ereignisse bei denen das eine das andere voraussetzt (Scheidungen, Scheidungen mit Streit um das Sorgerecht für Kinder), so ist die Gesamtzahl nicht mehr Poisson-verteilt. Es muss gelten, wenn $X + Y$ Poisson-verteilt wäre:

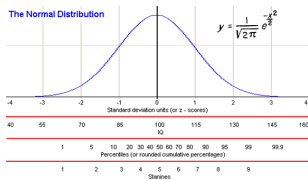
$$\text{Var}(X + Y) = \mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y) = \text{Var}(X) + \text{Var}(Y),$$

was aber bei abhängigen (korrelierten) X und Y verletzt ist.

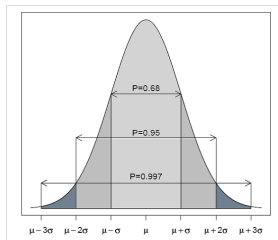
Normalverteilung I

Die Normalverteilung ist wohl das wichtigste Verteilungsmodell der Statistik, denn

- viele Zufallsvariablen sind (nach Transformation) (ungefähr) normalverteilt.
- beim Zusammenwirken vieler zufälliger Einflüsse ist der geeignet aggregierte Gesamteffekt oft approximativ normalverteilt (Zentraler Grenzwertsatz, Kap. 1.7).
- die asymptotische Grenzverteilung, also die Verteilung bei unendlich großem Stichprobenumfang, typischer statistischer Größen ist die Normalverteilung.



Normalverteilung II



Definition

Eine stetige Zufallsvariable X heißt normalverteilt mit den Parametern μ und σ^2 , in Zeichen $X \sim N(\mu, \sigma^2)$, wenn für ihre Dichte gilt:

$$f(x) = \frac{1}{\sqrt{2\pi} \cdot \sigma} \exp\left(-\frac{1}{2\sigma^2}(x - \mu)^2\right), \quad x \in \mathbb{R}$$

und standardnormalverteilt, in Zeichen $X \sim N(0; 1)$, falls $\mu = 0$ und $\sigma^2 = 1$ gilt (π ist hier die Kreiszahl $\pi = 3.14\dots$).

Grundlegende Eigenschaften

- a) Die Dichte der Standardnormalverteilung wird oft mit $\varphi(x)$ bezeichnet, also

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right)$$

und die zugehörige Verteilungsfunktion mit

$$\Phi(x) = \int_{-\infty}^x \varphi(u) du.$$

- b) $\Phi(x)$ lässt sich nicht in geschlossener Form durch bekannte Funktionen beschreiben \implies numerische Berechnung, Tabellierung.
 c) μ und σ^2 sind genau der Erwartungswert und die Varianz, also, wenn $X \sim N(\mu, \sigma^2)$, dann

$$\mathbb{E}(X) = \mu \text{ und } \text{Var}(X) = \sigma^2.$$

- d) Die Dichte ist symmetrisch um μ , d.h. $f(\mu - x) = f(\mu + x)$.

Rechnen mit der Normalverteilung I

- Es gilt:

$$\Phi(-x) = 1 - \Phi(x)$$

(folgt aus der Symmetrie der Dichte).

- Tabelliert sind die Werte der Verteilungsfunktion $\Phi(z) = P(Z \leq z)$ für $z \geq 0$.
Ablesebeispiel: $\Phi(1.75) = 0.9599$
- Funktionswerte für negative Argumente: $\Phi(-z) = 1 - \Phi(z)$
- Die z-Quantile ergeben sich über die Umkehrfunktion.
Beispielsweise ist $z_{0,9599} = 1.75$ und $z_{0,9750} = 1.96$.

Rechnen mit der Normalverteilung II

Zentrale Idee: X zu standardnormalverteilter Zufallsvariable umformen, d.h. standardisieren. Dabei muss die rechte Seite analog mit transformiert werden:

$$\begin{aligned}X \leq a &\Leftrightarrow X - \mu \leq a - \mu \\ &\Leftrightarrow \frac{X - \mu}{\sigma} \leq \frac{a - \mu}{\sigma}\end{aligned}$$

das heißt

$$P(X \leq a) = P\left(\frac{X - \mu}{\sigma} \leq \frac{a - \mu}{\sigma}\right).$$

Wegen

$$\frac{X - \mu}{\sigma} \sim N(0; 1)$$

gilt dann

$$P\left(\frac{X - \mu}{\sigma} \leq \frac{a - \mu}{\sigma}\right) = \Phi\left(\frac{a - \mu}{\sigma}\right),$$

so dass sich der folgende Zusammenhang ergibt:

$$P(X \leq a) = \Phi\left(\frac{a - \mu}{\sigma}\right).$$



Beispiel: IQ

Der IQ ist so konstruiert, dass er in der Bevölkerung normalverteilt ist mit Erwartungswert 100 und Standardabweichung 15.

- $P(IQ \leq 110) = ?$
- $P(IQ \leq 70) = ?$
- Wie groß ist q , damit gilt $P(IQ \geq q) = 0.01$?



Abgeschlossenheit gegenüber Linearkombinationen

Seien X_1 und X_2 unabhängig und $X_i \sim N(\mu_i, \sigma_i^2)$, $i = 1, 2$. Ferner seien b, a_1, a_2 feste reelle Zahlen. Dann gilt

$$\begin{aligned}Y_1 &:= a_1 X_1 + b \sim N(a_1 \mu_1 + b; a_1^2 \sigma_1^2) \\ Y_2 &:= a_1 X_1 + a_2 X_2 \sim N(a_1 \mu_1 + a_2 \mu_2; a_1^2 \sigma_1^2 + a_2^2 \sigma_2^2).\end{aligned}$$

Das Ergebnis lässt sich auf mehrere Summanden verallgemeinern.



Lebensdauerverteilungen I

Zur Modellierung von Lebensdauern (im weiteren Sinne) ist die Normalverteilung selten geeignet, da

- Nur Werte > 0 möglich
- Symmetrie selten sinnvoll

Typische Verteilungen sind die Exponentialverteilung, die Gammaverteilung, und die Weibullverteilung



Moderner Zweig vieler empirischer Untersuchungen: Lebensdaueranalyse bzw. allgemeiner Ereignisanalyse. Im Folgenden nur eine kurze Einführung, weiterführende Texte sind

- Rohwer und Pötter (2001): *Grundzüge der sozialwissenschaftlichen Statistik*, Teil III. Juventa, Soziologische Grundlagentexte.
- Blossfeld, Hamerle, Mayer (1986): *Ereignisanalyse: Statistische Theorie und Anwendungen in den Wirtschafts- und Sozialwissenschaften*. Campus.
- Diekmann und Mitter (1984): *Methoden zur Analyse von Zeitverläufen*. Teubner.
- Blossfeld und Rohwer (1995): *Techniques of Event History Modelling*. Erlbaum.

Betrachtet wird die Zufallsgröße „Zeit bis zu einem Ereignis“, z.B. Tod, Rückkehr aus Arbeitslosigkeit, Konkurs. Um den zeitlichen Aspekt (time) zu betonen, wird die interessierende Zufallsvariable häufig mit T statt mit X bezeichnet.

Die Exponentialverteilung

In Zeichen $X \sim \text{Ex}(\lambda)$ Der Parameter λ charakterisiert die Verteilung. Der Erwartungswert (Lebenserwartung) ist $\frac{1}{\lambda}$.

- 1 **Modell:** X ist die Lebensdauer eines Objekts, das nicht altert.
- 2 **Dichte, Verteilungsfunktion und Momente**

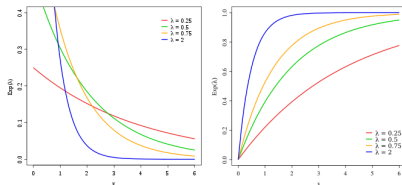
$$f_X(x) = \lambda e^{-\lambda x}$$

$$F_X(x) = 1 - e^{-\lambda x}$$

$$E(X) = \frac{1}{\lambda}$$

$$V(X) = \frac{1}{\lambda^2}$$

Dichte und Verteilungsfunktion



Die Survivorfunktion einer Verteilung ist definiert durch:

$$S(x) := P(X \geq x) = 1 - F(x)$$

Die Hazardrate ist definiert durch:

$$\lambda(x) := \lim_{h \rightarrow 0} \frac{P(x \leq X \leq x+h | X \geq x)}{h}$$

- Teil 1: bedingte Wahrscheinlichkeit mit Argument $\{x \leq X \leq x+h\}$ (Tod zwischen den Zeitpunkten x und $x+h$)
- Teil 2: bedingendes Ereignis $\{X \geq x\}$: Überleben bis mindestens zum Zeitpunkt x
- Teil 3: Intensität relativ zur Größe des betrachteten Intervalls $[x, x+h]$ mit Breite h .
- Teil 4: Grenzwert h gegen 0 betrachten, d.h. h sehr klein machen.
- Insgesamt: grobe, anschauliche Deutung:
Risiko, im nächsten Moment zu „sterben“, wenn man bis zum Zeitpunkt x „überlebt“ hat.
- Beachte: $\lambda(\cdot)$ ist keine Wahrscheinlichkeit, kann Werte zwischen 0 und unendlich annehmen.
- Sehr anschauliches Instrument zur Beschreibung von Lebensdauerverteilungen.

Zusammenhänge zwischen S, F und Hazard

Es gelten folgende Zusammenhänge:

$$\lambda(x) = \frac{f(x)}{S(x)}$$

$$S(x) = \exp\left(-\int_0^x \lambda(u) du\right)$$

$$F(x) = 1 - \exp\left(-\int_0^x \lambda(u) du\right)$$

$$f(x) = \lambda(x) \cdot S(x)$$

Beispiel: Warten auf den Bus

X ist die Zeit bis der Bus kommt (10 Minuten Takt).

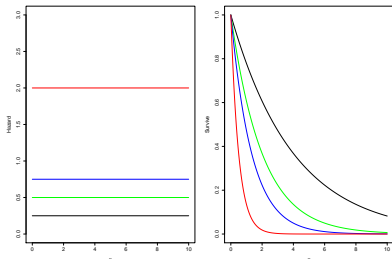
Für $x \in (0; 10)$ gilt:

$$f(x) = 0.1$$

$$F(x) = 0.1x$$

$$S(x) = 1 - 0.1x = 0.1 * (10 - x)$$

$$\lambda(x) = \frac{f(x)}{S(x)} = \frac{1}{10 - x}$$



$X \sim \text{Wb}(c, \alpha)$

- 1 **Modell:** Verteilung für Bruchfestigkeit von Materialien. Die Verteilung ist auch durch ihre Hazardrate charakterisiert und wird daher auch als Lebensdauerverteilung benutzt.

- 2 **Dichte, Verteilungsfunktion und Momente**

$$f(x) = cx^{c-1}/\alpha^c \cdot \exp\left(-\left(\frac{x}{\alpha}\right)^c\right)$$

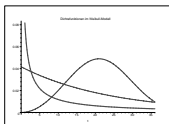
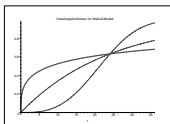
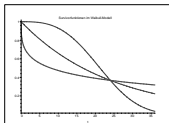
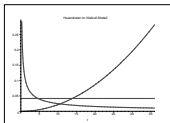
$$F(x) = 1 - \exp\left(-\left(\frac{x}{\alpha}\right)^c\right)$$

- 3 **Hazardrate**

$$\lambda(x) = \frac{f(x)}{1 - F(x)} = \frac{c}{\alpha} \left(\frac{x}{\alpha}\right)^{c-1}$$

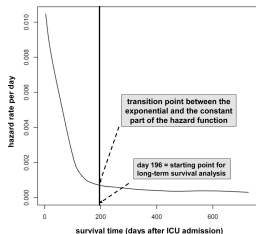
- 4 Für $c=1$ erhält man die Exponentialverteilung

Beispiele Weibullverteilungen



Beispiel: Überleben von Intensivpatienten

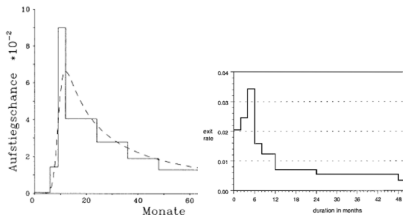
- Studie in Kooperation mit W. Hartl (Klinikum Großhadern)
- 1462 Patienten, die mehr als 4 Tage auf der Intensivstation waren
- Fragestellung: Wie ist der Risikoverlauf (Hazard) für Intensivpatienten
- Wie lange dauert es bis die Hazardrate konstant wird?
- Modell mit Weibullverteilung in zwei Phasen



- J. Brüderl (1990): Mobilitätsprozesse in Betrieben
- Personaldaten 1976-1984 der Arbeiter eines großen süddeutschen Maschinenbauunternehmens
- Analyse von Zeitdauern bis zur Beförderung bzw. Verlassen des Betriebs

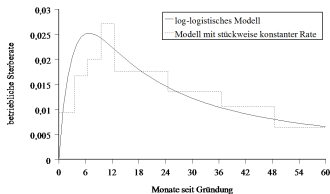
Das Honeymoon-Modell nach J. Brüderl

Hazardrate Aufstieg und Verlassen des Betriebs



Die Münchener Gründerstudie

- Brüderl/Preisendörfer/Ziegler (1996) Der Erfolg neugegründeter Betriebe. Duncker & Humblot.
- Gewerbedaten der IHK München/Oberbayern 1985/86
- Mündliche Befragung von 1.849 Unternehmensgründern im Jahr 1990



Modellierung mit der Log-logistischen Verteilung

Gerade in der Soziologie beobachtet man häufig *große* Stichprobenumfänge.

- Was ist das Besondere daran?
- Vereinfacht sich etwas und wenn ja was?
- Kann man „Wahrscheinlichkeitsgesetzmäßigkeiten“ durch Betrachten vielfacher Wiederholungen erkennen?

Das i.i.d.-Modell

Betrachtet werden diskrete oder stetige Zufallsvariablen X_1, \dots, X_n , die *i.i.d.* (independently, identically distributed) sind, d.h. die

- 1) unabhängig sind und
- 2) die gleiche Verteilung besitzen.

Ferner sollen der Erwartungswert μ und die Varianz σ^2 existieren. Die Verteilungsfunktion werde mit F bezeichnet.

Dies bildet insbesondere die Situation ab in der X_1, \dots, X_n eine Stichprobe eines Merkmals \bar{X} bei einer einfachen Zufallsauswahl sind.

Beispiel:

\bar{X} Einkommen, n Personen zufällig ausgewählt

X_1	Einkommen der	ersten	zufällig ausgewählten Person
X_2	Einkommen der	zweiten	zufällig ausgewählten Person
\vdots		\vdots	
X_n	Einkommen der	n -ten	zufällig ausgewählten Person

Stichprobenvariable

Jede Funktion von X_1, \dots, X_n ist wieder eine Zufallsvariable, z.B. das arithmetische Mittel oder die Stichprobenvarianz

$$\frac{1}{n} \sum_{i=1}^n X_i \quad \bar{S}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

Wahrscheinlichkeitsaussagen möglich \implies Wahrscheinlichkeitsrechnung anwenden

- Gerade bei diesen Zufallsgrößen ist die Abhängigkeit von n oft wichtig, man schreibt dann \bar{X}_n, \bar{S}_n^2
- Sind X_1, \dots, X_n jeweils $\{0, 1\}$ -Variablen, so ist \bar{X}_n gerade die empirische *relative Häufigkeit* von Einsen in der Stichprobe vom Umfang n . Notation: H_n

X_1, X_2, \dots, X_n seien unabhängig und identisch verteilt.

$$X_1, X_2, \dots, X_n \quad i.i.d.$$

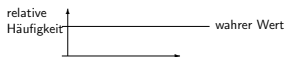
Ist $\mathbb{E}(X_i) = \mu$ und $\text{Var}(X_i) = \sigma^2$, so gilt:

$$\begin{aligned} \mathbb{E}(X_1 + X_2 + \dots + X_n) &= n\mu \\ \text{Var}(X_1 + X_2 + \dots + X_n) &= n\sigma^2 \\ \mathbb{E}\left(\frac{1}{n}(X_1 + X_2 + \dots + X_n)\right) &= \mu \\ \text{Var}\left(\frac{1}{n}(X_1 + X_2 + \dots + X_n)\right) &= \frac{\sigma^2}{n} \end{aligned}$$

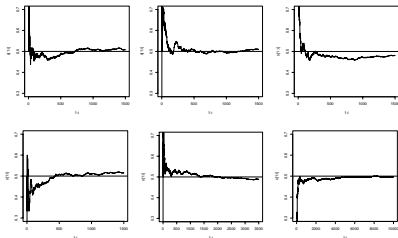
Diese Eigenschaften bilden die Grundlage für die folgenden Sätze.

Betrachte für wachsenden Stichprobenumfang n :

- X_1, \dots, X_n i.i.d.
- $X_i \in \{0, 1\}$ binäre Variablen mit $\pi = P(X_i = 1)$
Beispiele: Pro/Contra, Kopf/Zahl, A tritt ein/A tritt nicht ein
- $H_n =$ die relative Häufigkeit der Einsen in den ersten n Versuchen.



Simulationen



Beobachtungen

- 1 Am Anfang sehr unterschiedlicher, unregelmäßiger Verlauf der Pfade.
- 2 Mit wachsendem n pendeln sich die Pfade immer stärker um π herum ein, d.h. mit wachsendem Stichprobenumfang konvergiert die relative Häufigkeiten eines Ereignisses gegen seine Wahrscheinlichkeit.
- 3 Formalisierung von 2.: Legt man sehr kleine Korridor/Intervalle um π , so ist bei sehr großem n der Wert von H_n fast sicher in diesem Korridor.

Das Ereignis „Die relative Häufigkeit H_n liegt im Intervall der Breite 2ε um π “, lässt sich schreiben als:

$$\begin{aligned} \pi - \varepsilon &\leq H_n \leq \pi + \varepsilon \\ -\varepsilon &\leq H_n - \pi \leq \varepsilon \\ |H_n - \pi| &\leq \varepsilon \end{aligned}$$

Seien X_1, \dots, X_n , i.i.d. mit $X_i \in \{0, 1\}$ und $P(X_i = 1) = \pi$. Dann gilt für

$$H_n = \frac{1}{n} \sum_{i=1}^n X_i$$

(relative Häufigkeit der „Einsen“) und beliebig kleines $\epsilon > 0$

$$\lim_{n \rightarrow \infty} P(|H_n - \pi| \leq \epsilon) = 1$$

Anschauliche Interpretation: Die relative Häufigkeit eines Ereignisses nähert sich praktisch sicher mit wachsender Versuchszahl an die Wahrscheinlichkeit des Ereignisses an.

1) Häufigkeitsinterpretation von Wahrscheinlichkeiten:

$P(A)$, die Wahrscheinlichkeit eines Ereignisses A , kann man sich vorstellen als Grenzwert der relativen Häufigkeit des Eintretens von A in einer unendlichen Versuchsreihe identischer Wiederholungen eines Zufallsexperiments.

2) Induktion: Man kann dieses Ergebnis nutzen, um Information über eine unbekannte Wahrscheinlichkeit ($\pi \hat{=}$ Anteil in einer Grundgesamtheit) zu erhalten.

Sei z.B. π der (unbekannte) Anteil der SPD Wähler, so ist die relative Häufigkeit in der Stichprobe eine „gute Schätzung für π “. Je größer die Stichprobe ist, umso größer ist die Wahrscheinlichkeit, dass die relative Häufigkeit sehr nahe beim wahren Anteil π ist.



Gesetz der großen Zahl (allgemein)

Das Ergebnis lässt sich verallgemeinern auf Mittelwerte beliebiger Zufallsvariablen:

Gegeben seien X_1, \dots, X_n i.i.d. Zufallsvariablen mit (existierendem) Erwartungswert μ und (existierender) Varianz σ^2 . Dann gilt für

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

und beliebiges $\epsilon > 0$:

$$\lim_{n \rightarrow \infty} P(|\bar{X}_n - \mu| \leq \epsilon) = 1$$

Schreibweise:

$$\bar{X}_n \xrightarrow{P} \mu$$

(„Stochastische Konvergenz“, „ X_n konvergiert in Wahrscheinlichkeit gegen μ “.)



Konsequenz

- **Interpretation des Erwartungswerts:** μ kann in der Tat interpretiert werden als Durchschnittswert in einer unendlichen Folge von Wiederholungen des Zufallsexperiments.
- **Spiele.** Wenn ein Spiel mit negativem Erwartungswert häufig gespielt wird, verliert man mit sehr hoher Wahrscheinlichkeit (Grund für Rentabilität von Spielbanken und Wettbüros)



Jetzt betrachten wir die empirische Verteilungsfunktion: In jedem Punkt x ist $F_n(x)$ vor der Stichprobe eine Zufallsvariable, also ist F_n eine zufällige Funktion

Wie vergleicht man die zufällige Funktion $F_n(x)$ mit der Funktion $F(x)$? Der Abstand hängt ja von dem Punkt x ab, in dem gemessen wird!

Idee: Maximaler Abstand

$$\max_{x \in \mathbb{R}} |F_n^{X_1, \dots, X_n}(x) - F(x)|$$

Existiert nicht immer; formal muss man das sogenannte Supremum betrachten.

Seien X_1, \dots, X_n i.i.d. mit Verteilungsfunktion F und sei $F_n(x)$ die empirische Verteilungsfunktion der ersten n Beobachtungen. Mit

$$D_n := \sup_x |F_n(x) - F(x)|,$$

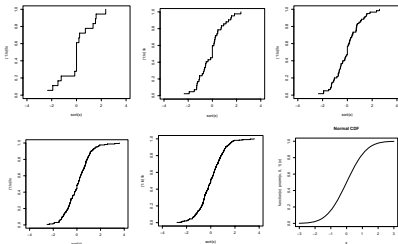
gilt für jedes $c > 0$

$$\lim_{n \rightarrow \infty} P(D_n > c) = 0.$$

Interpretation

- „Erträglichkeitsschranke“ c vorgegeben. Wsk, dass maximaler Abstand größer c ist geht für hinreichend großes n gegen 0 \Rightarrow überall kleiner Abstand. Man kann $\{D_n > c\}$ interpretieren als „Die Stichprobe führt den Betrachter hinter das Licht.“. Dann ist also die Wahrscheinlichkeit mit hinreichend großem n praktisch null.
- Anschaulich: Praktisch sicher spiegelt die empirische Verteilungsfunktion einer unendlichen Stichprobe die wahre Verteilungsfunktion wider.
- Falls die Stichprobe groß genug ist, so wird letztendlich immer repräsentativ für die Grundgesamtheit, d.h. man kann Verteilungsgesetzmäßigkeiten durch Beobachtungen erlernen (grundlegend für die Statistik) \rightarrow „Hauptsatz“.

Beispiele



- Gibt es für große Stichprobenumfänge Regelmäßigkeiten im Verteilungstyp?
- Gibt es eine Standardverteilung, mit der man oft bei großen empirischen Untersuchungen rechnen kann?

Seien X_1, \dots, X_n i.i.d. mit $\mathbb{E}(X_i) = \mu$ und $\text{Var}(X_i) = \sigma^2 > 0$ sowie

$$Z_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma} \right).$$

Dann gilt: Z_n ist *asymptotisch standardnormalverteilt*, in Zeichen: $Z_n \stackrel{\mathcal{L}}{\sim} N(0; 1)$, d.h. es gilt für jedes z

$$\lim_{n \rightarrow \infty} P(Z_n \leq z) = \Phi(z).$$

Für die Eingangsfragen gilt also:

Ja, wenn man die Variablen geeignet mittelt und standardisiert, dann kann man bei großem n näherungsweise mit der Normalverteilung rechnen. Dabei ist für festes n die Approximation umso besser, je „symmetrischer“ die ursprüngliche Verteilung ist.

Standardisieren

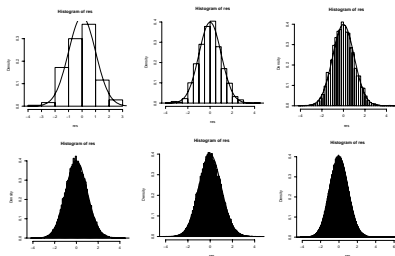
Die Funktion kommt durch *Standardisieren* und durch *geeignetes mitteln* zustande.

Dabei ist es wichtig, durch \sqrt{n} (und nicht durch n) zu teilen.

$$\sum X_i \quad \rightarrow \text{verliert sich; } \text{Var}(\sum X_i) \rightarrow \infty$$

$$\frac{1}{n} \sum x_i \quad \rightarrow \text{Var} \left(\frac{1}{n} \sum X_i \right) \rightarrow 0$$

Beispiele



Gemäß dem Gesetz der großen Zahlen weiß man: $\bar{X}_n \rightarrow \mu$

Für die Praxis ist es aber zudem wichtig, die konkreten Abweichungen bei großem aber endlichem n zu quantifizieren, etwa zur Beantwortung folgender Fragen:

- Gegeben eine Fehlermarge ε und Stichprobenumfang n : Wie groß ist die Wahrscheinlichkeit, dass \bar{X} höchstens um ε von μ abweicht?
- Gegeben eine Fehlermarge ε und eine „Sicherheitswahrscheinlichkeit“ γ : Wie groß muss man n mindestens wählen, damit mit mindestens Wahrscheinlichkeit γ das Stichprobenmittel höchstens um ε von μ abweicht (Stichprobenplanung)?

Aus dem zentralen Grenzwertsatz folgt:

$$\begin{aligned} \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma} \right) &= \frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n} \cdot \sigma} \\ &= \frac{n\bar{X}_n - n\mu}{\sqrt{n} \cdot \sigma} = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \approx N(0, 1) \end{aligned}$$

oder auch

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

$\frac{\sigma^2}{n}$ wird mit wachsendem n immer kleiner

- * Schwankung im richtigen Wert (μ)
- * Ausschläge werden kleiner

Approximation der Binomialverteilung I

Sei $X \sim B(n, \pi)$. Kann man die Verteilung von X approximieren?

Hier hat man zunächst nur ein X . Der zentrale Grenzwertsatz gilt aber für eine Summe vieler Glieder.

Idee: Schreibe X als Summe von binären Zufallsvariablen.

X ist die Anzahl der Treffer in einer i.i.d. Folge Y_1, \dots, Y_n von Einzelversuchen, wobei

$$Y_i = \begin{cases} 1 & \text{Treffer} \\ 0 & \text{kein Treffer} \end{cases}$$

Derselbe Trick wurde bei der Berechnung von Erwartungswerten angewendet.

Die Y_i sind i.i.d. Zufallsvariablen mit $Y_i \sim \text{Bin}(1, \pi)$ und es gilt

$$X = \sum_{i=1}^n Y_i, \quad \mathbb{E}(Y_i) = \pi, \quad \text{Var}(Y_i) = \pi \cdot (1 - \pi).$$

Approximation der Binomialverteilung II

Damit lässt sich der zentrale Grenzwertsatz anwenden:

$$\begin{aligned} \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\frac{Y_i - \pi}{\sqrt{\pi(1-\pi)}} \right) &= \frac{1}{\sqrt{n}} \frac{\sum Y_i - n \cdot \pi}{\sqrt{\pi(1-\pi)}} \\ &= \frac{\sum Y_i - n \cdot \pi}{\sqrt{n \cdot \pi(1-\pi)}} \approx N(0, 1) \end{aligned}$$

und damit

$$\frac{X - \mathbb{E}(X)}{\sqrt{\text{Var}(X)}} \approx N(0, 1)$$

so dass

$$P(X \leq x) \approx \Phi\left(\frac{x - n \cdot \pi}{\sqrt{n \cdot \pi(1-\pi)}}\right)$$

falls n groß genug.

Es gibt verschiedene Faustregeln, ab wann diese Approximation gut ist, z.B.

$$n \cdot \pi \geq 5 \quad \text{und} \quad n \cdot (1 - \pi) \geq 5 \\ n \cdot \pi(1 - \pi) \geq 9$$

Wichtig: Ob die Approximation hinreichend genau ist, hängt insbesondere vom substanzwissenschaftlichen Kontext ab.

Durch die Approximation der *diskreten* Binomialverteilung durch die *stetige* Normalverteilung geht der diskrete Charakter verloren. Man erhält als Approximation $P(X = x) \approx 0$ für jedes $x \in N$, was gerade für mittleres n unerwünscht ist.

Benutze deshalb

$$P(X \leq x) = P(X \leq x + 0.5)$$

bei ganzzahligem $x \in N$.

Man erhält als bessere Approximation

$$P(X \leq x) \approx \Phi\left(\frac{x + 0.5 - n\pi}{\sqrt{n\pi(1 - \pi)}}\right)$$

$$P(X = x) \approx \Phi\left(\frac{x + 0.5 - n\pi}{\sqrt{n\pi(1 - \pi)}}\right) - \Phi\left(\frac{x - 0.5 - n\pi}{\sqrt{n\pi(1 - \pi)}}\right)$$

Beispiel

Ein Politiker ist von einer gewissen umstrittenen Maßnahme überzeugt und überlegt, ob es taktisch geschickt ist, zur Unterstützung der Argumentation eine Mitgliederbefragung zu dem Thema durchzuführen. Er wählt dazu 200 Mitglieder zufällig aus und beschließt, eine Mitgliederbefragung zu „riskieren“, falls er in der Stichprobe mindestens 52% Zustimmung erhält.

Wie groß ist die Wahrscheinlichkeit, in der Stichprobe mindestens 52% Zustimmung zu erhalten, obwohl der wahre Anteil nur 48% beträgt?

Lösung

- X Anzahl der Ja-Stimmen
- X ja/nein \Rightarrow Binomialmodell
- $X \sim B(n, \pi)$ mit $n = 200$ und $\pi = 48\%$
- $n \cdot \pi = 96$ und $n \cdot (1 - \pi) = 104$: Faustregel erfüllt, die Normalapproximation darf also angewendet werden.

Gesucht: W'keit dass mind. 52%, also 104 Mitglieder, zustimmen, d.h.

$$\begin{aligned} P(X \geq 104) &= 1 - P(X \leq 103) \\ &= 1 - \Phi\left(\frac{x + 0.5 - n\pi}{\sqrt{n \cdot \pi(1 - \pi)}}\right) \\ &= 1 - \Phi\left(\frac{103.5 - 200 \cdot 0.48}{\sqrt{200 \cdot 0.48(1 - 0.48)}}\right) \\ &= 1 - \Phi(1.06) \\ &= 1 - 0.8554 = 14.5\% \end{aligned}$$

Im Folgenden Beschränkung auf den diskreten Fall und zweidimensionale Zufallsvariablen.

Vorstellung: Auswerten eines mehrdimensionalen Merkmals

$$\begin{pmatrix} \tilde{X} \\ \tilde{Y} \end{pmatrix}$$

also z.B. $\omega \in \tilde{\Omega}$, zufällig gezogene Person und damit $\tilde{X}(\omega)$ und $\tilde{Y}(\omega)$
Auswertung der Merkmale jeweils an *derselben* Person.

⇒ zweidimensionale Zufallsvariable $\begin{pmatrix} X \\ Y \end{pmatrix}$ (wie bei Zusammenhangsanalyse in Statistik I)

Das Hauptinteresse gilt (entsprechend der Kontingenztafel in Statistik I) der gemeinsamen Verteilung

$$P(\{X = x_i\} \cap \{Y = y_j\})$$



Betrachtet werden zwei eindimensionale diskrete Zufallselemente X und Y (zu demselben Zufallsexperiment). Die Wahrscheinlichkeit

$$P(X = x_i, Y = y_j) := P(\{X = x_i\} \cap \{Y = y_j\})$$

in Abhängigkeit von x_i und y_j heißt *gemeinsame Verteilung* der mehrdimensionalen Zufallsvariable $\begin{pmatrix} X \\ Y \end{pmatrix}$ bzw. der Variablen X und Y .
Randwahrscheinlichkeiten:

$$p_{i\bullet} = P(X = x_i) = \sum_{j=1}^m P(X = x_i, Y = y_j)$$

$$p_{\bullet j} = P(Y = y_j) = \sum_{i=1}^k P(X = x_i, Y = y_j)$$



Bedingte Verteilungen

$$P(X = x_i | Y = y_j) = \frac{P(X = x_i, Y = y_j)}{P(Y = y_j)}$$

$$P(Y = y_j | X = x_i) = \frac{P(X = x_i, Y = y_j)}{P(X = x_i)}$$

Stetiger Fall: Zufallsvariable mit zweidimensionaler Dichtefunktion $f(x, y)$:

$$P(a \leq X \leq b, c \leq Y \leq d) = \int_a^b \left(\int_c^d f(x, y) dy \right) dx$$



Kovarianz

Definition

Seien X und Y zwei Zufallsvariablen. Dann heißt

$$\sigma_{X,Y} := \text{Cov}(X, Y) = \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y)))$$

Kovarianz von X und Y .



- $\text{Cov}(X, X) = \text{Var}(X)$
- $\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X) \cdot \mathbb{E}(Y)$
- $\text{Cov}(X, Y) = \text{Cov}(Y, X)$
- Mit $\tilde{X} = a_X X + b_X$ und $\tilde{Y} = a_Y Y + b_Y$ ist

$$\text{Cov}(\tilde{X}, \tilde{Y}) = a_X \cdot a_Y \cdot \text{Cov}(X, Y)$$
- $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2 \cdot \text{Cov}(X, Y)$

Definition

Zwei Zufallsvariablen X und Y mit $\text{Cov}(X, Y) = 0$ heißen unkorreliert.

- Stochastisch unabhängige Zufallsvariablen sind unkorreliert. Die Umkehrung gilt jedoch im allgemeinen nicht.
- Vergleiche Statistik I: Kovarianz misst nur lineare Zusammenhänge.

Korrelationskoeffizient

Definition

Gegeben seien zwei Zufallsvariablen X und Y . Dann heißt

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)}\sqrt{\text{Var}(Y)}}$$

Korrelationskoeffizient von X und Y .

Eigenschaften des Korrelationskoeffizienten

- Mit $\tilde{X} = a_X X + b_X$ und $\tilde{Y} = a_Y Y + b_Y$ ist

$$|\rho(\tilde{X}, \tilde{Y})| = |\rho(X, Y)|.$$
- $-1 \leq \rho(X, Y) \leq 1$.
- $|\rho(X, Y)| = 1 \iff Y = aX + b$
- Sind $\text{Var}(X) > 0$ und $\text{Var}(Y) > 0$, so gilt $\rho(X, Y) = 0$ genau dann, wenn $\text{Cov}(X, Y) = 0$.

X_1 Gewinn, wenn beim ersten Wurf ein Einsatz auf 1 gesetzt wird.

X_6 Gewinn, wenn beim ersten Wurf ein Einsatz auf 6 gesetzt wird.

Kovarianz zwischen X_1 und X_6 :

(x_1, x_6)	$P(X_1 = x_1, X_6 = x_6)$	(x_1, x_6)	$P(X_1 = x_1, X_6 = x_6)$
$(-1, -1)$	$\frac{64}{216}$	$(-1, 3)$	$\frac{1}{216}$
$(-1, 1)$	$\frac{48}{216}$	$(3, -1)$	$\frac{1}{216}$
$(1, -1)$	$\frac{48}{216}$	$(1, 1)$	$\frac{24}{216}$
$(-1, 2)$	$\frac{12}{216}$	$(1, 2)$	$\frac{3}{216}$
$(2, -1)$	$\frac{12}{216}$	$(1, 2)$	$\frac{3}{216}$

$$\Rightarrow E(X_1 \cdot X_6) = -50/216 = -0.23148$$

$$\text{Cov}(X_1, X_6) = -0.23148 - (-0.0787) \cdot (-0.0787) = -0.23768$$

X_1 und X_6 sind negativ korreliert.

Grundprinzipien der induktiven Statistik

1 Einführung

2 Wahrscheinlichkeitsrechnung

3 Zufallsvariablen und ihre Verteilung

4 Statistische Inferenz

5 Hypothesentest

6 Regression

Ziel: Inferenzschluss, Repräsentationsschluss: Schluss von einer Stichprobe auf Eigenschaften der Grundgesamtheit, aus der sie stammt.

- Von Interesse sei ein Merkmal \tilde{X} in der Grundgesamtheit $\tilde{\Omega}$.
- Ziehe eine Stichprobe $(\omega_1, \dots, \omega_n)$ von Elementen aus $\tilde{\Omega}$ und werte \tilde{X} jeweils aus.
- Man erhält Werte x_1, \dots, x_n . Diese sind Realisationen der i.i.d. Zufallsvariablen oder Zufallselemente X_1, \dots, X_n , wobei die Wahrscheinlichkeitsverteilung der X_1, \dots, X_n genau die Häufigkeitsverhältnisse in der Grundgesamtheit widerspiegelt.

Ziel: Schlüsse von Stichprobe auf Grundgesamtheit
Schlüsse von Experiment auf allgemeines Phänomen

Zentrale Fragen:

- Wie kann die Zufälligkeit in korrekter Weise berücksichtigt werden?
- Wann sind Ergebnisse in der Stichprobe zufallsbedingt?
- Wie sind korrekte Schlüsse möglich?

Voraussetzungen für das Anwenden statistischer Inferenz

- Stichprobe sollte zufällig sein
- Experimentelle Situation
- Nicht nötig (geeignet) bei Vollerhebungen
- Nicht geeignet bei Vollerhebungen mit geringem Rücklauf

1 Schätzen:

Von Interesse ist der Wert eines Parameters in der Grundgesamtheit, z.B. Mittelwert oder Anteil

- Punktschätzung: Angabe eines Wertes
- Intervallschätzung (Konfidenzintervall): Angabe eines Bereiches, in dem der Wert mit hoher Sicherheit liegt

2 Testen (Signifikanztest):

Untersuchung, ob eine bestimmte Hypothese mit Hilfe der Daten widerlegt werden kann

z.B. Gewisse Satzkonstruktionen führen zu schnellerer Reaktion

Beispiele:

- Punktschätzung: z.B. wahrer Anteil 0.4751
- Intervallschätzung: z.B. wahrer Anteil liegt zwischen 0.46 und 0.48
- Hypothesentest: Die Annahme, der Anteil liegt höchstens bei 50% kann nicht aufrecht erhalten werden

Zentrale Fragestellung

Wie kommt man von Realisationen x_1, \dots, x_n von i.i.d. Zufallsvariablen X_1, \dots, X_n auf die Verteilung der X_i ?

- Dazu nimmt man häufig an, man kenne den Grundtyp der Verteilung der X_1, \dots, X_n . Unbekannt seien nur einzelne Parameter davon.
Beispiel: X_i sei normalverteilt, unbekannt seien nur μ, σ^2 .
⇒ *parametrische Verteilungsannahme* (meist im Folgenden)
- Alternativ: Verteilungstyp nicht oder nur schwach festgelegt (z.B. symmetrische Verteilung)
⇒ *nichtparametrische Modelle*
- Klarenweise gilt im Allgemeinen (generelles Problem bei der Modellierung): Parametrische Modelle liefern schärfere Aussagen – wenn ihre Annahmen zutreffen. Wenn ihre Annahmen nicht zutreffen, dann existiert die große Gefahr von Fehlschlüssen.

Beispiel:

Parameter: Mittelwert der täglichen Fernsehdauer von Jugendlichen in Deutschland

Schätzung: Mittelwert der Fernsehdauer in der Stichprobe
 oder: Median aus der Stichprobe?
 oder: Mittelwert ohne größten und kleinsten Wert?

Grundgesamtheit

1	2	3	4	5
1.30	1.31	1.32	1.40	1.42

Wahrer Wert: 1.35

Ziehe Stichprobe vom Umfang $n=2$ und berechne \bar{X}

S_1	S_2	X	P
1	2	1.305	0.1
1	3	1.310	0.1
1	4	1.350	0.1
1	5	1.360	0.1
2	3	1.315	0.1
2	4	1.355	0.1
2	5	1.365	0.1
3	4	1.360	0.1
3	5	1.370	0.1
4	5	1.410	0.1

„Pech“

Beispiel 2: Würfeln mit potentiell gefälschtem Würfel

Wie groß ist der Erwartungswert beim Würfeln mit potentiell gefälschtem Würfel?

Ziehe Stichprobe und berechne Mittelwert \bar{X}

\bar{X} liefert plausible Schätzung für den wahren (theoretischen) Mittelwert.

Simulation mit R

Punktschätzung

Beachte: Auswahl zufällig \Rightarrow Schätzung zufällig

- Die Merkmale der gezogenen n Einheiten sind also Zufallsgrößen.
- Bezeichnung: X_1, \dots, X_n .
- Wird der Parameter einer Merkmalsverteilung durch eine Funktion der Zufallsgrößen X_1, \dots, X_n der Stichprobe geschätzt, so spricht man bei diesem Vorgang von **Punktschätzung**.
- Die dabei benutzte Funktion wird auch **Schätzfunktion**, **Schätzstatistik** oder kurz **Schätzer** genannt.

Definition

Sei X_1, \dots, X_n i.i.d. Stichprobe. Eine Funktion

$$T = g(X_1, \dots, X_n)$$

heißt *Schätzer* oder *Schätzfunktion*.

Inhaltlich ist $g(\cdot)$ eine Auswertungsregel der Stichprobe:
„Welche Werte sich auch in der Stichprobe ergeben, ich wende das durch $g(\cdot)$ beschriebene Verfahren auf sie an. (z.B. Mittelwert)“

Beispiele für Schätzfunktionen II

- Größter Stichprobenwert:

$$X_{(n)} = g(X_1, \dots, X_n) = \max_{i=1, \dots, n} X_i$$

- Kleinster Stichprobenwert:

$$X_{(1)} = g(X_1, \dots, X_n) = \min_{i=1, \dots, n} X_i$$

- Arithmetisches Mittel der Stichprobe:

$$\bar{X} = g(X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^n X_i$$

Für binäre, dummy-kodierte X_i ist \bar{X} auch die relative Häufigkeit des Auftretens von „ $X_i = 1$ “ in der Stichprobe

- Stichprobenvarianz:

$$\tilde{S}^2 = g(X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - (\bar{X})^2$$

- Korrigierte Stichprobenvarianz:

$$S^2 = g(X_1, \dots, X_n) = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n-1} \left(\sum_{i=1}^n X_i^2 - n \cdot \bar{X}^2 \right)$$

Qualitätsmerkmal eines Schätzers

Erwartungstreue, Bias:

Gegeben sei eine Stichprobe X_1, \dots, X_n und eine Schätzfunktion $T = g(X_1, \dots, X_n)$ (mit existierendem Erwartungswert).

- T heißt *erwartungstreu* für den Parameter ϑ , falls gilt

$$\mathbb{E}_\vartheta(T) = \vartheta$$

für alle ϑ .

- Die Größe

$$\text{Bias}_\vartheta(T) = \mathbb{E}_\vartheta(T) - \vartheta$$

heißt *Bias* (oder *Verzerrung*) der Schätzfunktion. Erwartungstreue Schätzfunktionen haben per Definition einen Bias von 0.

Man schreibt $\mathbb{E}_\vartheta(T)$ und $\text{Bias}_\vartheta(T)$, um deutlich zu machen, dass die Größen von dem wahren ϑ abhängen.

Das arithmetische Mittel $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ ist erwartungstreu für den Mittelwert μ einer Grundgesamtheit

Aus X_1, \dots, X_n i.i.d. und $\mathbb{E}_\mu(X_1) = \mathbb{E}_\mu(X_2) = \dots = \mu$ folgt:

$$\begin{aligned}\mathbb{E}(\bar{X}) &= \mathbb{E}_\mu \left(\frac{1}{n} \sum_{i=1}^n X_i \right) = \frac{1}{n} \mathbb{E}_\mu \left(\sum_{i=1}^n X_i \right) \\ &= \frac{1}{n} \sum_{i=1}^n \mathbb{E}(X_i) \\ &= \frac{1}{n} \sum_{i=1}^n \mu = \frac{1}{n} \cdot n \cdot \mu = \mu\end{aligned}$$

Sei σ^2 die Varianz in der Grundgesamtheit. Es gilt

$$\mathbb{E}_{\sigma^2}(\tilde{S}^2) = \frac{n-1}{n} \sigma^2,$$

also ist \tilde{S}^2 *nicht* erwartungstreu für σ^2 .

$$\text{Bias}_{\sigma^2}(\tilde{S}^2) = \frac{n-1}{n} \sigma^2 - \sigma^2 = -\frac{1}{n} \sigma^2$$

(Für $n \rightarrow \infty$ geht $\text{Bias}_{\sigma^2}(\tilde{S}^2)$ gegen 0, \tilde{S}^2 ist „asymptotisch erwartungstreu“.)

Für die korrigierte Stichprobenvarianz hingegen gilt:

$$\begin{aligned}\mathbb{E}_{\sigma^2}(S^2) &= \mathbb{E}_{\sigma^2} \left(\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \right) \\ &= \mathbb{E}_{\sigma^2} \left(\frac{1}{n-1} \cdot \frac{n}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \right) \\ &= \mathbb{E}_{\sigma^2} \left(\frac{n}{n-1} S^2 \right) = \frac{n}{n-1} \cdot \frac{n-1}{n} \sigma^2 = \sigma^2\end{aligned}$$

Also ist S^2 erwartungstreu für σ^2 . Diese Eigenschaft ist auch die Motivation für die Korrektur der Stichprobenvarianz.

Vorsicht: Im Allgemeinen gilt für beliebige, nichtlineare Funktionen g

$$\mathbb{E} g(X) \neq g(\mathbb{E}(X)).$$

Man kann also nicht einfach z.B. $\sqrt{\cdot}$ und \mathbb{E} vertauschen. In der Tat gilt: S^2 ist zwar erwartungstreu für σ^2 , aber $\sqrt{S^2}$ ist nicht erwartungstreu für $\sqrt{\sigma^2} = \sigma$.

Gegeben sei eine Stichprobe der wahlberechtigten Bundesbürger. Geben Sie einen erwartungstreuen Schätzer des Anteils der rot-grün Wähler an.

Grundgesamtheit: Dichotomes Merkmal

$$\bar{X} = \begin{cases} 1 & \text{rot/grün: ja} \\ 0 & \text{rot/grün: nein} \end{cases}$$

Der Mittelwert π von \bar{X} ist der Anteil der rot/grün-Wähler in der Grundgesamtheit.

Stichprobe X_1, \dots, X_n vom Umfang n :

$$X_i = \begin{cases} 1 & i\text{-te Person wählt rot/grün} \\ 0 & \text{sonst} \end{cases}$$

Aus den Überlegungen zum arithmetischen Mittel folgt, dass

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

ein erwartungstreuer Schätzer für den hier betrachteten Parameter π ist. Also verwendet man die relative Häufigkeit in der Stichprobe, um den wahren Anteil π in der Grundgesamtheit zu schätzen.

Bedeutung der Erwartungstreue

Erwartungstreue ist ein schwaches Kriterium!

Betrachte die offensichtlich unsinnige Schätzfunktion

$$T_2 = g_2(X_1, \dots, X_n) = X_1,$$

d.h. $T_2 = 100\%$, falls der erste Befragte rot-grün wählt und $T_2 = 0\%$ sonst.

Die Schätzfunktion ignoriert fast alle Daten, ist aber erwartungstreu:

$$\mathbb{E}(T_2) = \mathbb{E}(X_1) = \mu$$

Deshalb betrachtet man zusätzlich die Effizienz eines Schätzers

Effizienz I

Beispiel Wahlumfrage

Gegeben sind zwei erwartungstreue Schätzer (n sei gerade):

$$T_1 = \frac{1}{n} \sum_{i=1}^n X_i$$

$$T_2 = \frac{1}{n/2} \sum_{i=1}^{n/2} X_i$$

Was unterscheidet formal T_1 von dem unsinnigen Schätzer T_2 , der die in der Stichprobe enthaltene Information nicht vollständig ausnutzt? Vergleiche die Schätzer über ihre Varianz, nicht nur über den Erwartungswert!

Wenn n so groß ist, dass der zentrale Grenzwertsatz angewendet werden kann, dann gilt approximativ

$$\frac{1}{\sqrt{n}} \frac{\sum_{i=1}^n (X_i - \pi)}{\sqrt{\pi(1-\pi)}} = \frac{\sum_{i=1}^n X_i - n \cdot \pi}{\sqrt{n} \sqrt{\pi(1-\pi)}} = \frac{\frac{1}{n} \sum_{i=1}^n X_i - \pi}{\sqrt{\frac{\pi(1-\pi)}{n}}} \sim \mathbb{N}(0; 1)$$

und damit

$$T_1 = \frac{1}{n} \sum_{i=1}^n X_i \sim \mathbb{N}\left(\pi; \frac{\pi(1-\pi)}{n}\right).$$

Analog kann man zeigen:

$$T_2 = \frac{1}{n/2} \sum_{i=1}^{n/2} X_i \sim \mathbb{N}\left(\pi; \frac{\pi(1-\pi)}{n/2}\right).$$

T_1 und T_2 sind approximativ normalverteilt, wobei T_1 eine deutlich kleinere Varianz als T_2 hat.

T_1 und T_2 treffen beide im Durchschnitt den richtigen Wert π . T_1 schwankt aber weniger um das wahre π , ist also „im Durchschnitt genauer“.



Ein erwartungstreuer Schätzer ist umso besser, je kleiner seine Varianz ist.

$$\text{Var}(T) = \text{Erwartete quadratische Abweichung von } T \text{ von } \underbrace{\mathbb{E}(T)}_{=\vartheta!}$$

Je kleiner die Varianz, umso mehr konzentriert sich die Verteilung eines erwartungstreuen Schätzers um den wahren Wert.



- Gegeben seien zwei erwartungstreue Schätzfunktionen T_1 und T_2 für einen Parameter ϑ . Gilt

$$\text{Var}_{\vartheta}(T_1) \leq \text{Var}_{\vartheta}(T_2) \text{ für alle } \vartheta$$

und

$$\text{Var}_{\vartheta^*}(T_1) < \text{Var}_{\vartheta^*}(T_2) \text{ für mindestens ein } \vartheta^*$$

so heißt T_1 *effizienter als* T_2 .

- Eine für ϑ erwartungstreue Schätzfunktion T heißt *UMVU-Schätzfunktion* für ϑ (uniformly minimum variance unbiased), falls

$$\text{Var}_{\vartheta}(T) \leq \text{Var}_{\vartheta}(T^*)$$

für alle ϑ und für alle erwartungstreuen Schätzfunktionen T^* .



- *Inhaltliche Bemerkung:* Der (tiefere) Sinn von Optimalitätskriterien wird klassischerweise insbesondere auch in der *Gewährleistung von Objektivität* gesehen.
- Ist X_1, \dots, X_n eine i.i.d. Stichprobe mit $X_i \sim N(\mu, \sigma^2)$, dann ist
 - \bar{X} UMVU-Schätzfunktion für μ und
 - S^2 UMVU-Schätzfunktion für σ^2 .

- Ist X_1, \dots, X_n mit $X_i \in \{0, 1\}$ eine i.i.d. Stichprobe mit $\pi = P(X_i = 1)$, dann ist die relative Häufigkeit \bar{X} UMVU-Schätzfunktion für π .
- Bei nicht erwartungstreuen Schätzern macht es keinen Sinn, sich ausschließlich auf die Varianz zu konzentrieren.
- Z.B. hat der unsinnige Schätzer $T = g(X_1, \dots, X_n) = 42$, der die Stichprobe nicht beachtet, Varianz 0.

MSE

Man zieht dann den sogenannten *Mean Squared Error*

$$\text{MSE}_\vartheta(T) = \mathbb{E}_\vartheta(T - \vartheta)^2$$

zur Beurteilung heran. Es gilt

$$\text{MSE}_\vartheta(T) = \text{Var}_\vartheta(T) + (\text{Bias}_\vartheta(T))^2.$$

Der MSE kann als Kompromiss zwischen zwei Auffassungen von Präzision gesehen werden: möglichst geringe systematische Verzerrung (Bias) und möglichst geringe Schwankung (Varianz).

Asymptotische Erwartungstreue

- * Eine Schätzfunktion heißt asymptotisch erwartungstreu, falls

$$\lim_{n \rightarrow \infty} E(\hat{\theta}) = \theta$$

bzw.

$$\lim_{n \rightarrow \infty} \text{Bias}(\hat{\theta}) = 0$$

gelten.

- * Abschwächung des Begriffs der Erwartungstreue: Gilt nur noch bei einer unendlich großen Stichprobe.
- * Erwartungstreue Schätzer sind auch asymptotisch erwartungstreu.
- * Sowohl S^2 als auch \tilde{S}^2 sind asymptotisch erwartungstreu.

- Für komplexere Modelle ist oft die Erwartungstreue der Verfahren ein zu restriktives Kriterium. Man fordert deshalb oft nur, dass sich der Schätzer wenigstens für große Stichproben gut verhält. Hierzu gibt es v.a. zwei verwandte aber „etwas“ unterschiedliche Kriterien.
- Ein Schätzer heißt (MSE-)konsistent oder konsistent im quadratischen Mittel, wenn gilt

$$\lim_{n \rightarrow \infty} (\text{MSE}(T)) = 0.$$

Der MSE von \bar{X} ist gegeben durch

$$\text{MSE}(\bar{X}) = \text{Var}(\bar{X}) + \text{Bias}^2(\bar{X}) = \frac{\sigma^2}{n} + 0 = \frac{\sigma^2}{n} \rightarrow 0.$$

\bar{X} ist also ein MSE-konsistenter Schätzer für den Erwartungswert. Anschaulich bedeutet die Konsistenz, dass sich die Verteilung des Schätzers für wachsenden Stichprobenumfang n immer stärker beim richtigen Wert „zusammenzieht“. Er trifft also für unendlich große Stichproben praktisch sicher den wahren Wert. (Dies gilt als eine Minimalanforderung an statistische Verfahren.)

Maximum-Likelihood-Prinzip I

Sie wissen als Wirt, dass heute die Lokalparteien ihre Busausflüge unternehmen: Es werden Busse mit je 100 Personen von der jeweiligen Partei organisiert.

- Bus I: 85% Partei A, 15% Partei B
- Bus II: 15% Partei A, 85% Partei B

Bus fährt vor, anhand Stichprobe ermitteln, ob Bild von ... von der Wand genommen werden soll oder nicht.

Stichprobe von 10 Personen ergibt 80% Anhänger der Partei A.

- Welche Partei: wohl A, aber B nicht ausgeschlossen bei unglücklicher Auswahl.
- Warum: A ist plausibler, da die Wahrscheinlichkeit, ungefähr den in der Stichprobe beobachteten Wert zu erhalten (bzw. erhalten zu haben) bei Bus I wesentlich größer ist als bei Bus II.

Maximum-Likelihood-Prinzip II

Aufgabe: Schätze den Parameter ϑ eines parametrischen Modells anhand einer i.i.d. Stichprobe X_1, \dots, X_n mit der konkreten Realisation x_1, \dots, x_n .

Idee der Maximum-Likelihood (ML) Schätzung für diskrete Verteilungen:

- Man kann für jedes ϑ die Wahrscheinlichkeit ausrechnen, genau die Stichprobe x_1, \dots, x_n zu erhalten:

$$P_{\vartheta}(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = \prod_{i=1}^n P_{\vartheta}(X_i = x_i)$$

- Je größer für ein gegebenes ϑ_0 die Wahrscheinlichkeit ist, die konkrete Stichprobe erhalten zu haben, umso plausibler ist es, dass tatsächlich ϑ_0 der wahre Wert ist (gute Übereinstimmung zwischen Modell und Daten).

I.i.d. Stichprobe vom Umfang $n = 5$ aus einer $B(10, \pi)$ -Verteilung:

6 5 3 4 4

Wahrscheinlichkeit der Stichprobe für gegebenes π :

$$P(X_1 = 6, \dots, X_5 = 4 | \pi) = P(X_1 = 6 | \pi) \cdot \dots \cdot P(X_5 = 4 | \pi) \\ = \binom{10}{6} \pi^6 (1 - \pi)^4 \cdot \dots \cdot \binom{10}{4} \pi^4 (1 - \pi)^6.$$

„ $P(\dots | \pi)$ Wahrscheinlichkeit, wenn π der wahre Parameter ist“

π	$P(X_1 = 6, \dots, X_5 = 4 \pi)$
0.1	0.0000000000001
0.2	0.0000000227200
0.3	0.0000040425220
0.4	0.0003025481000
0.5	0.0002487367000
0.6	0.0000026561150
0.7	0.0000000250490
0.8	0.0000000000055
0.9	0.0000000000000

Man nennt daher $L(\vartheta) = P_{\vartheta}(X_1 = x_1, \dots, X_n = x_n)$, nun als Funktion von ϑ gesehen, die *Likelihood* (deutsch: Plausibilität, Mutmaßlichkeit) von ϑ gegeben die Realisation x_1, \dots, x_n . Derjenige Wert $\hat{\vartheta} = \hat{\vartheta}(x_1, \dots, x_n)$, der $L(\vartheta)$ maximiert, heißt *Maximum-Likelihood-Schätzwert*; die zugehörige Schätzfunktion $T(X_1, \dots, X_n)$ *Maximum-Likelihood-Schätzer*

$$P_{\vartheta}(X_1 = x_1, \dots, X_n = x_n)$$

- Deduktiv (Wahrscheinlichkeitsrechnung): ϑ bekannt, x_1, \dots, x_n zufällig („unbekannt“).
- Induktiv (Statistik): ϑ unbekannt, x_1, \dots, x_n bekannt.

Deduktiv

Induktiv

geg: Parameter bekannt

ges: Plausibilität des Parameters

$P_{\vartheta}(X_1 = x_1, \dots, X_n = x_n)$
Funktion von x_1, \dots, x_n
bei festem ϑ

$P_{\vartheta}(X_1 = x_1, \dots, X_n = x_n)$
Funktion von ϑ
bei festem x_1, \dots, x_n

ges: Wskt von Beobachtungen

geg: Beobachtung bekannt

Definition Maximum Likelihood

Gegeben sei die Realisation x_1, \dots, x_n einer i.i.d. Stichprobe. Die Funktion in ϑ

$$L(\vartheta) = \begin{cases} \prod_{i=1}^n P_{\vartheta}(X_i = x_i) & \text{falls } X_i \text{ diskret} \\ \prod_{i=1}^n f_{\vartheta}(x_i) & \text{falls } X_i \text{ stetig.} \end{cases}$$

heißt *Likelihood* des Parameters ϑ bei der Beobachtung x_1, \dots, x_n .

Derjenige Wert $\hat{\vartheta} = \hat{\vartheta}(x_1, \dots, x_n)$, der $L(\vartheta)$ maximiert, heißt *Maximum-Likelihood-Schätzwert*; die zugehörige Schätzfunktion $T(X_1, \dots, X_n)$ *Maximum-Likelihood-Schätzer*.

- In diesem Fall verwendet man die Dichte

$$f_{\vartheta}(x_1, \dots, x_n) = \prod_{i=1}^n f_{\vartheta}(x_i)$$

als Maß für die Plausibilität von ϑ .

- Für die praktische Berechnung maximiert man statt der Likelihood typischerweise die Log-Likelihood

$$l(\vartheta) = \ln(L(\vartheta)) = \ln \prod_{i=1}^n P_{\vartheta}(X_i = x_i) = \sum_{i=1}^n \ln P_{\vartheta}(X_i = x_i)$$

bzw.

$$l(\vartheta) = \ln \prod_{i=1}^n f_{\vartheta}(x_i) = \sum_{i=1}^n \ln f_{\vartheta}(x_i).$$

$$X_j = \begin{cases} 1 & \text{falls Rot/Grün} \\ 0 & \text{sonst} \end{cases}$$

Verteilung der X_j : Binomialverteilung $B(1, \pi)$ (Bernoulliverteilung)

$$P(X_j = 1) = \pi$$

$$P(X_j = 0) = 1 - \pi$$

$$P(X_j = x_j) = \pi^{x_j} \cdot (1 - \pi)^{1-x_j}, \quad x_j \in \{0; 1\}.$$

Hier ist π der unbekannte Parameter, der allgemeine mit ϑ bezeichnet wird.

ML Schätzung für π einer Bernoulliverteilung II

- Bestimme die Likelihoodfunktion

$$\begin{aligned} L(\pi) &= P(X_1 = x_1, \dots, X_n = x_n) \\ &= \prod_{i=1}^n \pi^{x_i} (1 - \pi)^{1-x_i} \\ &= \pi^{\sum_{i=1}^n x_i} \cdot (1 - \pi)^{n - \sum_{i=1}^n x_i} \end{aligned}$$

ML Schätzung für π einer Bernoulliverteilung III

- Berechne die logarithmierte Likelihoodfunktion

$$l(\pi) = \ln(P(X_1 = x_1, \dots, X_n = x_n)) = \sum_{i=1}^n x_i \cdot \ln(\pi) + (n - \sum_{i=1}^n x_i) \cdot \ln(1 - \pi)$$

- Ableiten (nach π):

$$\frac{\partial}{\partial \pi} l(\pi) = \frac{\sum_{i=1}^n x_i}{\pi} + \frac{n - \sum_{i=1}^n x_i}{1 - \pi} \cdot (-1)$$

- Bemerkung zur Loglikelihood

Der Logarithmus ist streng monoton wachsend. Allgemein gilt für streng monoton wachsende Funktionen g : x_0 Stelle des Maximums von $L(x) \iff x_0$ auch Stelle des Maximums von $g(L(x))$.

- Berechnung des ML-Schätzers durch Nullsetzen der abgeleiteten Loglikelihoodfunktion

$$\begin{aligned} \frac{\partial}{\partial \pi} l(\pi) = 0 &\iff \frac{\sum_{i=1}^n x_i}{\pi} = \frac{n - \sum_{i=1}^n x_i}{1 - \pi} \\ &\iff (1 - \pi) \sum_{i=1}^n x_i = n \cdot \pi - \pi \sum_{i=1}^n x_i \\ &\iff \sum_{i=1}^n x_i = n \cdot \pi \end{aligned}$$

also

$$\hat{\pi} = \frac{\sum_{i=1}^n x_i}{n}$$

Also ist \bar{X} der Maximum-Likelihood-Schätzer für π .

- Bestimme die Likelihoodfunktion

$$\begin{aligned} L(\mu, \sigma^2) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi(\sigma^2)}} \exp\left(-\frac{1}{2\sigma^2}(x_i - \mu)^2\right) \\ &= \frac{1}{2\pi^{\frac{n}{2}}(\sigma^2)^{\frac{n}{2}}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right) \end{aligned}$$

- Bestimme die Log-Likelihoodfunktion

$$\begin{aligned} l(\mu, \sigma^2) &= \ln(L(\mu, \sigma^2)) \\ &= \ln(1) - \frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 \end{aligned}$$

- Ableiten und Nullsetzen der Loglikelihoodfunktion

$$\begin{aligned} \frac{\partial l(\mu, \sigma^2)}{\partial \mu} &= \frac{1}{2\sigma^2} 2 \cdot \sum_{i=1}^n (x_i - \mu) = 0 \\ \frac{\partial l(\mu, \sigma^2)}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} \sum_{i=1}^n (x_i - \mu)^2 = 0 \end{aligned}$$

- Auflösen der beiden Gleichungen nach μ und σ^2

Aus der ersten Gleichung erhalten wir

$$\sum_{i=1}^n x_i - n\mu = 0 \quad \text{also} \quad \hat{\mu} = \bar{x}.$$

Aus der zweiten Gleichung erhalten wir durch Einsetzen von $\hat{\mu} = \bar{x}$

$$\sum_{i=1}^n (x_i - \bar{x})^2 = n\sigma^2$$

also

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

- Der ML-Schätzer $\hat{\mu} = \bar{X}$ für μ stimmt mit dem üblichen Schätzer für den Erwartungswert überein.
- Der ML-Schätzer $\hat{\sigma}^2 = \hat{S}^2$ für σ^2 ist verzerrt, d.h. nicht erwartungstreu.

Einige allgemeine Eigenschaften von ML-Schätzern

- ML-Schätzer $\hat{\theta}$ sind im Allgemeinen nicht erwartungstreu.
- ML-Schätzer $\hat{\theta}$ sind asymptotisch erwartungstreu.
- ML-Schätzer $\hat{\theta}$ sind konsistent (und meist in einem asymptotischen Sinne effizient).

Intervallschätzung: Motivation und Hinführung

Der wahre Anteil der rot-grün Wähler 2009 war genau 33.7%. Wie groß ist die Wahrscheinlichkeit, in einer Zufallsstichprobe von 1000 Personen genau einen relativen Anteil von 33.7% von rot-grün Anhängern erhalten zu haben?

$$\begin{aligned} X_i &= \begin{cases} 1, \text{ rot/grün} \\ 0, \text{ sonst} \end{cases} \\ P(X_i = 1) &= \pi = 0.337 \\ X &= \sum_{i=1}^n X_i \sim B(n, \pi) \text{ mit } n = 1000 \end{aligned}$$

$$\begin{aligned}
 P(X = 337) &= \binom{n}{x} \cdot \pi^x \cdot (1 - \pi)^{n-x} \\
 &= \binom{1000}{337} \cdot 0.337^{337} \cdot (1 - 0.337)^{663} \\
 &= 0.02668164
 \end{aligned}$$

D.h., mit Wahrscheinlichkeit von etwa 97.3%, verfehlt der Schätzer den wahren Wert.

- Insbesondere Vorsicht bei der Interpretation „knapper Ergebnisse“ (z.B. Anteil 50.2%)
- Suche Schätzer mit möglichst kleiner Varianz, um „im Durchschnitt möglichst nahe dran zu sein“
- Es ist häufig auch gar nicht nötig, sich genau auf einen Wert festzulegen. Oft reicht die Angabe eines Intervalls, von dem man hofft, dass es den wahren Wert überdeckt: *Intervallschätzung*

Symmetrische Intervallschätzung

Basierend auf einer Schätzfunktion $T = g(X_1, \dots, X_n)$ sucht man:

$$I(T) = [T - a, T + a]$$

„**Trade off**“ bei der Wahl von a :

- Je größer man a wählt, also je breiter man das Intervall $I(T)$ macht,
- umso größer ist die Wahrscheinlichkeit, dass $I(T)$ den wahren Wert überdeckt,
- *aber* umso weniger aussagekräftig ist dann die Schätzung.

Extremfall im Wahlbeispiel: $I(T) = [0, 1]$ überdeckt sicher π , macht aber eine wertlose Aussage

Typisches Vorgehen

- Man gebe sich durch inhaltliche Überlegungen einen Sicherheitsgrad (*Konfidenzniveau*) γ vor.
- Dann konstruiert man das Intervall so, dass es mindestens mit der Wahrscheinlichkeit γ den wahren Parameter überdeckt.

Definition

Gegeben sei eine i.i.d. Stichprobe X_1, \dots, X_n zur Schätzung eines Parameters ϑ und eine Zahl $\gamma \in (0; 1)$. Ein zufälliges Intervall $C(X_1, \dots, X_n)$ heißt *Konfidenzintervall* zum *Sicherheitsgrad* γ (Konfidenzniveau γ), falls für jedes ϑ gilt:

$$P_{\vartheta}(\vartheta \in \underbrace{C(X_1, \dots, X_n)}_{\text{zufälliges Intervall}}) \geq \gamma.$$

Die Wahrscheinlichkeitsaussage bezieht sich auf das Ereignis, dass das zufällige Intervall den festen, wahren Parameter überdeckt. Streng genommen darf man im objektivistischen Verständnis von Wahrscheinlichkeit nicht von der *Wahrscheinlichkeit* sprechen, „dass ϑ in dem Intervall liegt“, da ϑ nicht zufällig ist und somit keine Wahrscheinlichkeitsverteilung besitzt.

Praktische Vorgehensweise: Suche Zufallsvariable Z_{ϑ} , die

- den gesuchten Parameter ϑ enthält und
- deren Verteilung aber nicht mehr von dem Parameter abhängt, („Pivotgröße“, dt. Angelpunkt).
- Dann wähle den Bereich C_Z so, dass $P_{\vartheta}(Z_{\vartheta} \in C_Z) = \gamma$ und
- löse nach ϑ auf.

Konfidenzintervall für den Mittelwert (normalverteiltes Merkmal, Varianz bekannt)

X_1, \dots, X_n i.i.d. Stichprobe gemäß $X_i \sim N(\mu, \sigma^2)$, wobei σ^2 bekannt sei.

1. Starte mit der Verteilung von \bar{X} :

$$\bar{X} \sim N(\mu, \sigma^2/n).$$

2. Dann erfüllt

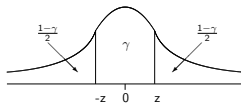
$$Z = \frac{\bar{X} - \mu}{\sigma} \cdot \sqrt{n} \sim N(0; 1)$$

die obigen Bedingungen an eine Pivotgröße.

3. Bestimme jetzt einen Bereich $[-z, z]$, wobei z so gewählt sei, dass

$$P(Z \in [-z; z]) = \gamma$$

KI-Bestimmung Strategie



Bestimmung von z :

$$P(Z \in [-z; z]) = \gamma \iff P(Z \geq z) = \frac{1-\gamma}{2}$$

beziehungsweise

$$P(Z \leq z) = 1 - \frac{1-\gamma}{2} = \frac{2-1+\gamma}{2} = \frac{1+\gamma}{2}.$$

Die Größe z heißt das $\frac{1+\gamma}{2}$ -Quantil und wird mit $z_{\frac{1+\gamma}{2}}$ bezeichnet.

$$\gamma = 90\% \quad \frac{1+\gamma}{2} = 95\% \quad z_{0.95} = 1.65$$

$$\gamma = 95\% \quad \frac{1+\gamma}{2} = 97.5\% \quad z_{0.975} = 1.96$$

$$\gamma = 99\% \quad \frac{1+\gamma}{2} = 99.5\% \quad z_{0.995} = 2.58$$

$$P\left(-z_{\frac{1+\gamma}{2}} \leq Z_{\mu} \leq z_{\frac{1+\gamma}{2}}\right) = P\left(-z_{\frac{1+\gamma}{2}} \leq \frac{\bar{X} - \mu}{\sigma} \cdot \sqrt{n} \leq z_{\frac{1+\gamma}{2}}\right) = \gamma$$

Jetzt nach μ auflösen (Ziel: $P(\dots \leq \mu \leq \dots)$):

$$\begin{aligned} \gamma &= P\left(-\frac{z_{\frac{1+\gamma}{2}} \cdot \sigma}{\sqrt{n}} \leq \bar{X} - \mu \leq \frac{z_{\frac{1+\gamma}{2}} \cdot \sigma}{\sqrt{n}}\right) \\ &= P\left(-\bar{X} - \frac{z_{\frac{1+\gamma}{2}} \cdot \sigma}{\sqrt{n}} \leq -\mu \leq -\bar{X} + \frac{z_{\frac{1+\gamma}{2}} \cdot \sigma}{\sqrt{n}}\right) \\ &= P\left(\bar{X} - \frac{z_{\frac{1+\gamma}{2}} \cdot \sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + \frac{z_{\frac{1+\gamma}{2}} \cdot \sigma}{\sqrt{n}}\right) \end{aligned}$$

KI für Mittelwert (NV mit bekanntem σ)

Damit ergibt sich:

Konfidenzintervall für μ bei bekannter Varianz

$$\left[\bar{X} - \frac{z_{\frac{1+\gamma}{2}} \cdot \sigma}{\sqrt{n}}, \bar{X} + \frac{z_{\frac{1+\gamma}{2}} \cdot \sigma}{\sqrt{n}}\right] = \left[\bar{X} \pm \frac{z_{\frac{1+\gamma}{2}} \cdot \sigma}{\sqrt{n}}\right]$$

Eigenschaften

- Je größer σ , desto größer das Intervall!
(Größeres $\sigma \Rightarrow$ Grundgesamtheit bezüglich des betrachteten Merkmals heterogener, also größere Streuung von $\bar{X} \Rightarrow$ ungenauere Aussagen.)
- Je größer γ , desto größer $z_{\frac{1+\gamma}{2}}$
(Je mehr Sicherheit/Vorsicht desto breiter das Intervall)
- Je größer n und damit \sqrt{n} , desto schmaler ist das Intervall
(Je größer der Stichprobenumfang ist, desto genauer!)
Aufpassen, die Genauigkeit nimmt nur mit \sqrt{n} zu. Halbierung des Intervalls, Vervielfachung des Stichprobenumfangs.
Kann man zur *Stichprobenplanung* verwenden!

Konfidenzintervall für den Mittelwert (normalverteiltes Merkmal, Varianz unbekannt)

Neben dem Erwartungswert ist auch σ^2 unbekannt und muss entsprechend durch den UMVU-Schätzer

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2,$$

(mit $S = \sqrt{S^2}$) geschätzt werden. Allerdings ist

$$Z = \frac{\bar{X} - \mu}{S} \cdot \sqrt{n}$$

jetzt nicht mehr normalverteilt, denn S ist zufällig.

→ Wir führen deshalb ein neues Verteilungsmodell ein.

t-Verteilung

Definition

Gegeben sei eine i.i.d. Stichprobe X_1, \dots, X_n mit $X_i \sim \mathcal{N}(\mu, \sigma^2)$. Dann heißt die Verteilung von

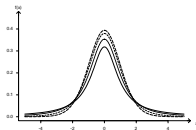
$$Z = \frac{\bar{X} - \mu}{S} \cdot \sqrt{n}$$

t-Verteilung (oder Student-Verteilung) mit $\nu = n - 1$ Freiheitsgraden.

In Zeichen: $Z \sim t(\nu)$.

Dichte einer t-Verteilung

Die Dichte einer t-Verteilung ist der Dichte der Standardnormalverteilung sehr ähnlich: Sie ist auch symmetrisch um 0, besitzt aber etwas höhere Dichte für extreme Werte („schwerere Enden“).



Dichten von t-Verteilungen für

$\nu = 1$ (—), $\nu = 2$ (···), $\nu = 5$ (- - -) und $\nu = 20$ (- · -) Freiheitsgrade.

Unsicherheit durch zusätzliche Schätzung von σ lässt Daten stärker schwanken.

Eigenschaften t-Verteilung

- Je größer ν ist, umso ähnlicher sind sich die $t(\nu)$ -Verteilung und die Standardnormalverteilung.
 - Für $\nu \rightarrow \infty$ sind sie gleich.
 - Ab $\nu = 30$ gilt der Unterschied als vernachlässigbar.
- Je größer n , desto geringer ist der Unterschied zwischen S^2 und σ^2 und damit zwischen $\frac{\bar{X} - \mu}{S} \sqrt{n}$ und $\frac{\bar{X} - \mu}{\sigma} \sqrt{n}$.

Ausgehend von

$$P\left(-t_{\frac{1-\gamma}{2}}^{(n-1)} \leq \frac{\bar{X} - \mu}{S} \cdot \sqrt{n} \leq t_{\frac{1-\gamma}{2}}^{(n-1)}\right) = \gamma$$

wie im Beispiel mit bekannter Varianz nach μ auflösen (mit S statt σ)

$$P\left(\bar{X} - \frac{t_{\frac{1-\gamma}{2}}^{(n-1)} \cdot S}{\sqrt{n}} \leq \mu \leq \bar{X} + \frac{t_{\frac{1-\gamma}{2}}^{(n-1)} \cdot S}{\sqrt{n}}\right) = \gamma$$

Damit ergibt sich:

Konfidenzintervall für μ bei unbekannter Varianz

$$\left[\bar{X} \pm \frac{t_{\frac{1-\gamma}{2}}^{(n-1)} \cdot S}{\sqrt{n}} \right]$$

- Es gelten analoge Aussagen zum Stichprobenumfang und Konfidenzniveau wie bei bekannter Varianz.
- Für jedes γ (und jedes ν) gilt

$$t_{\frac{1-\gamma}{2}}^{\nu} > z_{\frac{1-\gamma}{2}}$$

also ist das t-Verteilungs-Konfidenzintervall (etwas) breiter. Hintergrund: Da σ^2 unbekannt ist, muss es geschätzt werden. Dies führt zu etwas größerer Ungenauigkeit.

- Je größer ν , umso kleiner ist der Unterschied. Für $n \geq 30$ rechnet man einfach auch bei der t-Verteilung mit $z_{\frac{1-\gamma}{2}}$.

Beispiel Konfidenzintervall zum Konfidenzniveau γ

Eine Maschine füllt Gummibärchen in Tüten ab, die laut Aufdruck 250g Füllgewicht versprechen. Wir nehmen im folgenden an, dass das Füllgewicht normalverteilt ist. Bei 16 zufällig aus der Produktion herausgegriffenen Tüten wird ein mittleres Füllgewicht von 245g und eine Stichprobenstreuung (Standardabweichung) von 10g festgestellt.

- Berechnen Sie ein Konfidenzintervall für das mittlere Füllgewicht zum Sicherheitsniveau von 95%.
- Wenn Ihnen zusätzlich bekannt würde, dass die Stichprobenstreuung gleich der tatsächlichen Streuung ist, wäre dann das unter a) zu berechnende Konfidenzintervall für das mittlere Füllgewicht breiter oder schmaler? Begründen Sie ihre Antwort ohne Rechnung.

Beispiel Konfidenzintervall zum Konfidenzniveau γ

- Füllgewicht normalverteilt. ($\mu = 250g$ nicht benötigt)
- 16 Tüten gezogen $\Rightarrow n = 16$.
- Mittleres Füllgewicht in der Stichprobe: $\bar{x} = 245g$.
- Stichprobenstreuung: $s = 10g$.

a) Konstruktion des Konfidenzintervalls:

- Da die Varianz σ^2 unbekannt ist, muss das Konfidenzintervall basierend auf der t-Verteilung konstruiert werden:

$$\left[\bar{X} \pm t_{\frac{1-\gamma}{2}}(n-1) \cdot \frac{S}{\sqrt{n}} \right]$$

Aus dem Sicherheitsniveau $\gamma = 0.95$ errechnet sich $\frac{1+\gamma}{2} = 0.975$.

Nachschauen in t-Tabelle bei 0.975 und 15 Freiheitsgraden ($T = \frac{\bar{X}-\mu}{S/\sqrt{n}}$ ist t-verteilt mit $n-1$ Freiheitsgraden) liefert $t_{0.975} = 2.13$.

- Einsetzen liefert damit

$$\left[245 \pm 2.13 \cdot \frac{10}{4} \right] = [239.675; 250.325]$$

b) Jetzt sei σ^2 bekannt.

- Dann kann man mit dem Normalverteilungs-Intervall rechnen:

$$\left[\bar{X} \pm z_{\frac{1-\gamma}{2}} \cdot \frac{\sigma}{\sqrt{n}} \right]$$

Da jetzt σ bekannt, ist die Unsicherheit geringer und damit das Konfidenzintervall schmäler.

- In der Tat ist $z_{\frac{1-\gamma}{2}} < t_{\frac{1-\gamma}{2}}$.
- Rechnerisch ergibt sich mit $z_{\frac{1-\gamma}{2}} = 1.96$ das Konfidenzintervall

$$[240.100; 249.900]$$

Approximative Konfidenzintervalle

Ist der Stichprobenumfang groß genug, so kann wegen des zentralen Grenzwertsatzes das Normalverteilungs-Konfidenzintervall auf den Erwartungswert beliebiger Merkmale (mit existierender Varianz) angewendet werden. Man erhält approximative Konfidenzintervalle, die meist auch der Berechnung mit Software zugrundeliegen.

Approximatives Konfidenzintervall für den Mittelwert (n groß)

$$\left[\bar{X} \pm z_{\frac{1-\gamma}{2}} \cdot \frac{S}{\sqrt{n}} \right]$$

$\frac{S}{\sqrt{n}}$ wird als Standardfehler (Standard error) bezeichnet.

Approximatives Konfidenzintervall für einen Anteil

Gesucht: Konfidenzintervall für den Anteilswert $p = P(X = 1)$ einer Bernoulli-Zufallsgröße X

- X_1, \dots, X_n i.i.d. Stichprobe
- n hinreichend groß (Faustregel $n > 30$)
- vorgegebenes Sicherheitsniveau γ („gamma“)

Approximatives Konfidenzintervall für π

$$R \pm z_{\frac{1-\gamma}{2}} \cdot \sqrt{\frac{R(1-R)}{n}}$$

R = Anteil aus der Stichprobe

$z_{\frac{1-\gamma}{2}}$ ist das $\frac{1+\gamma}{2}$ -Quantil der Standardnormalverteilung.

- Gegeben:
 - $n = 500$
 - $\bar{X} = 46.5\%$
 - $\gamma = 95\%$ und damit $z_{\frac{1-\gamma}{2}} = 1.96$
- Konfidenzintervall:

$$\begin{aligned} \left[\bar{X} \pm z_{\frac{1-\gamma}{2}} \cdot \sqrt{\frac{\bar{X}(1-\bar{X})}{n}} \right] &= \left[0.465 \pm 1.96 \cdot \sqrt{\frac{0.465(1-0.465)}{500}} \right] \\ &= [0.421; 0.508] \end{aligned}$$

- Man beachte die relativ große Breite, trotz immerhin mittelgroßer Stichprobe
- Zum Sicherheitsniveau 95% ist keine eindeutige Aussage über die Mehrheitsverhältnisse möglich. Berücksichtigen, wenn man über Wahlumfrage urteilt
- In der Praxis sind aber Wahlumfragen etwas genauer, da man Zusatzinformation verwendet (insbesondere auch frühere Wahlergebnisse). „Gebundene Hochrechnung“

Bestimmung des Stichprobenumfangs für die Anteilsschätzung

- Genauigkeit ist inhaltlich vorzugeben
- Je genauer und sicherer, desto größer muss der Stichprobenumfang sein
- Genauigkeit: Halbe Länge g des Konfidenzintervalls
- Gib Konfidenzniveau (oft 95%) vor und bestimme n so, dass g kleiner ist als bestimmter Wert

Konkrete Umsetzung

- γ : Konfidenzniveau
- g : Genauigkeit

$$g \geq z_{\frac{1-\gamma}{2}} \cdot \sqrt{\frac{R(1-R)}{n}}$$

Auflösen nach n :

$$n \geq \frac{1}{g^2} z_{\frac{1-\gamma}{2}}^2 \cdot R(1-R)$$

Beachte: $R(1-R) \leq 0.25$

Gegeben:

- Konfidenzniveau: 0.05
- Genauigkeit: 10%

Bestimmung von n :

$$n \geq \frac{1}{g^2} z_{\frac{1-\gamma}{2}}^2 \cdot R(1-R) = \frac{1}{0.1^2} 1.96^2 \cdot 0.25 = 96.04$$

Beachte: $R(1-R) \leq 0.25$

Also sollten ca. 100 Personen befragt werden.

Bei $g = 5\%$ ergibt sich $n = 385$

Bei $g = 1\%$ ergibt sich $n = 9604$

- Differenz von Mittelwerten bei unabhängigen Stichproben
- Differenz von Anteilen bei unabhängigen Stichproben
- Differenz von Mittelwerten bei verbundenen Stichproben

Konfidenzintervall für die Differenz von Mittelwerten (unabhängige Stichproben)

Unterschied der Mittelwerte zwischen zwei Gruppen $\mu_X - \mu_Y$

- Zwei voneinander stochastisch unabhängige Stichproben
 - Daten aus Gruppe 1: X_1, \dots, X_m, X_i i.i.d.
 - Daten aus Gruppe 2: Y_1, \dots, Y_n, Y_j i.i.d.
- Stichprobenumfänge hinreichend groß ($n \geq 30, m \geq 30$)
- Schätzung: $\bar{X} - \bar{Y} = \frac{1}{n} \sum_{i=1}^n X_i - \frac{1}{m} \sum_{j=1}^m Y_j$

Approximatives KI für Differenz von Mittelwerten (unabhängigen Stichproben, n groß)

$$\left[(\bar{X} - \bar{Y}) - z_{\frac{1-\gamma}{2}} \cdot S_d; (\bar{X} - \bar{Y}) + z_{\frac{1-\gamma}{2}} \cdot S_d \right]$$

mit

- $S_d = \sqrt{\frac{s_X^2}{m} + \frac{s_Y^2}{n}}$
- $z_{\frac{1-\gamma}{2}}$ ist das $\frac{1+\gamma}{2}$ -Quantil der Standardnormalverteilung

Beispiel: Radiohördauer Ost-West

Westen: $\bar{x} = 11.4$ Stunden und $s_X = 8.4$

Osten: $\bar{y} = 9.5$ Stunden und $s_Y = 8.4$

$$\sqrt{\frac{s_X^2}{m} + \frac{s_Y^2}{n}} \approx 0.6$$

$$k_u = \bar{x} - \bar{y} - z_{\frac{1-\gamma}{2}} \cdot \sqrt{\frac{s_X^2}{m} + \frac{s_Y^2}{n}} = 0.38$$

$$k_o = \bar{x} - \bar{y} + z_{\frac{1-\gamma}{2}} \cdot \sqrt{\frac{s_X^2}{m} + \frac{s_Y^2}{n}} = 3.42$$

Die Differenz liegt also zwischen 0.38 und 3.42 h/Woche

Konfidenzintervall für die Differenz von Anteilen (unabhängige Stichproben)

Vergleich von Anteilen über die Differenz $p_X - p_Y$

- Zwei voneinander stochastisch unabhängige Stichproben
 - Daten aus Gruppe 1: X_1, \dots, X_m, X_i i.i.d.
 - Daten aus Gruppe 2: Y_1, \dots, Y_n, Y_j i.i.d.
- Stichprobenumfänge hinreichend groß ($n \geq 30, m \geq 30$)
- Schätzung: $R_X - R_Y$ (R_X und R_Y : Anteil in den Gruppen 1 und 2)

Approximatives KI für Differenz von Anteilen (unabhängigen Stichproben, n groß)

$$\left[(R_X - R_Y) - z_{\frac{1+\gamma}{2}} \cdot S_d; (R_X - R_Y) + z_{\frac{1+\gamma}{2}} \cdot S_d \right]$$

mit

- $S_d = \sqrt{\frac{R_X \cdot (1-R_X)}{m} + \frac{R_Y \cdot (1-R_Y)}{n}}$
- $z_{\frac{1+\gamma}{2}}$ ist das $\frac{1+\gamma}{2}$ -Quantil der Standardnormalverteilung

Beispiel: Ist Fernsehen informativ?

	nein	ja	
alte BL: X	47	206	253
neue BL: Y	185	747	932
	232	953	1185

$$R_X - R_Y = \frac{206}{253} - \frac{747}{932} = 0.81 - 0.80$$

$$S_d = \sqrt{\frac{R_X \cdot (1-R_X)}{m} + \frac{R_Y \cdot (1-R_Y)}{n}} = 0.03$$

Konfidenzintervall: $[-0.04; 0.07]$

Verbundene Stichproben

- Gleiche Größe zweimal (davor - danach)
- Zwei Größen bei derselben Person
- „Matched Pair“

Hauptidee:

Verwende Differenzen $W_i = X_i - Y_i$

Konfidenzintervall für die Differenz von Mittelwerten (verbundene Stichproben)

Unterschied zwischen zwei verbundenen Gruppen $\mu_X - \mu_Y$

- Zwei verbundene Stichproben
- Daten: $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$,
- Stichprobenumfang hinreichend groß ($n \geq 30$)
- Schätzung: $\bar{W} = \frac{1}{n} \sum_{i=1}^n W_i = \frac{1}{n} \sum_{i=1}^n (X_i - Y_i)$

Approximatives KI für Differenz von Mittelwerten (verbundene Stichproben, n groß)

$$\left[\bar{W} - z_{\frac{1+\gamma}{2}} \cdot \frac{S_w}{\sqrt{n}}; \bar{W} + z_{\frac{1+\gamma}{2}} \cdot \frac{S_w}{\sqrt{n}} \right]$$

mit:

- $S_w^2 = \frac{1}{n-1} \sum_{i=1}^n (W_i - \bar{W})^2$
- $z_{\frac{1+\gamma}{2}}$ ist das $\frac{1+\gamma}{2}$ -Quantil der Standardnormalverteilung



- 1 Einführung
- 2 Wahrscheinlichkeitsrechnung
- 3 Zufallsvariablen und ihre Verteilung
- 4 Statistische Inferenz
- 5 Hypothesentests**
- 6 Regression

Hypothese

„Behauptung einer Tatsache, deren Überprüfung noch aussteht“
(Leutner in: Endrueit, Trommsdorff: Wörterbuch der Soziologie, 1989).

Statistischer Test: Überprüfung von Hypothesen über die Grundgesamtheit anhand einer Stichprobe

Idealtypische Vorgehensweise

Wissenschaftlicher Fortschritt durch Falsifikation von Hypothesen

Statistische Testtheorie

Schließe von Stichprobe Experiment auf Grundgesamtheit bzw. Allg. Gesetz

Vorgehen:

- Inhaltliche Hypothese aufstellen
- Operationalisierung
- Inhaltliche Hypothese in statistische Hypothese „Übersetzen“
- Statistischer Test

Bemerkungen

- **Statistische Tests:**
Die am häufigsten verwendete Art statistischer Inferenz
- **Statistische Signifikanz:**
Zentrales Argument bei vielen empirischen Arbeiten
- **Voraussetzung für Testverfahren:**
Zufallsstichprobe oder Experiment

Ist ein beobachtetes Phänomen in Stichproben ein **reines Zufallsprodukt** oder **mit großer Sicherheit** auf einen **realen Effekt** zurückzuführen?

→ Dazu notwendig:

Formale Entscheidungsregel = Statistischer Test

Zeitungsberichte: 2€Münzen nicht „fair“



- Vermutung:
2€- Münze nicht fair
- Überprüfung: 10-Mal die Münze werfen, Anzahl „Zahl“ notieren

Mögliche Ergebnisse des Experiments

- 5-Mal "Zahl"
→ deutet nicht auf eine unfaire Münze hin
- 10-Mal "Zahl"
→ verdächtig, die Münze ist vermutlich nicht fair
- 0-Mal "Zahl"
→ verdächtig, die Münze ist vermutlich nicht fair
- 8-Mal "Zahl"
→ ?? mehr Zahlwürfe als erwartet. **Zufall? Oder Münze nicht fair?**

Münzhypothese

- Vermutung:
2€- Münze nicht fair
- Statistische Formulierung:
 X Bernoulli-Variable

$$X = \begin{cases} 1 & \text{"Zahl"} \\ 0 & \text{"Adler"} \end{cases}$$

- Wahrscheinlichkeit für Zahl

$$\pi = P(X = 1)$$

- „Die Münze ist nicht fair“ heißt

$$\pi \neq 0.5$$

Überprüfung der Münzhypothese

- Experiment: Wir werfen $n = 10$ -Mal die Münze

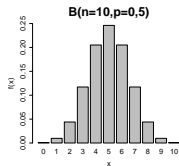
$$\sum_{i=1}^n X_i \sim B(n = 10, \pi)$$

- Welche Ergebnisse sind wahrscheinlich, falls die Münze fair ist?
- **Falls die Münze fair ist**, so ist die Anzahl „Zahl“ binomialverteilt mit $p = 0.5$.

$$\sum_{i=1}^{10} X_i \sim B(n = 10, \pi = 0.5)$$

- **Falls die Münze fair ist**, so sollte $\sum_{i=1}^{10} X_i$ mit einer Wahrscheinlichkeit von **90 %** nicht weit entfernt vom Erwartungswert 5 liegen.

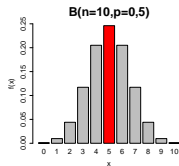
Binomialverteilung



x	0	1	2	3	4	5	6	7	8	9	10
f(x)	0.001	0.010	0.044	0.117	0.205	0.246	0.205	0.117	0.044	0.010	0.001

$$\Sigma =$$

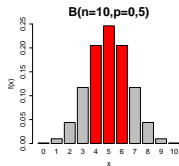
Binomialverteilung



x	0	1	2	3	4	5	6	7	8	9	10
f(x)	0.001	0.010	0.044	0.117	0.205	0.246	0.205	0.117	0.044	0.010	0.001

$$\Sigma = 0.246$$

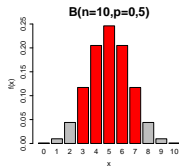
Binomialverteilung



x	0	1	2	3	4	5	6	7	8	9	10
f(x)	0.001	0.010	0.044	0.117	0.205	0.246	0.205	0.117	0.044	0.010	0.001

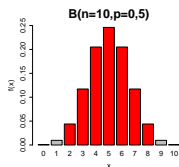
$$\Sigma = 0.656$$

Binomialverteilung



x	0	1	2	3	4	5	6	7	8	9	10
f(x)	0.001	0.010	0.044	0.117	0.205	0.246	0.205	0.117	0.044	0.010	0.001

$$\Sigma = 0.890$$



x	0	1	2	3	4	5	6	7	8	9	10
f(x)	0.001	0.010	0.044	0.117	0.205	0.246	0.205	0.117	0.044	0.010	0.001

$$\sum = 0.978$$

Statistischer Test: Hypothese

Statistischer Test

Untersuchung, ob man eine Hypothese über die Grundgesamtheit mit Hilfe einer Stichprobe widerlegen kann.

- **Nullhypothese** H_0 = Hypothese, die widerlegt werden soll.
Beispiel: Die Münze ist fair

$$H_0 : \pi = 0.5$$

- **Gegenhypothese** H_1 = Alternative zur Nullhypothese.
Beispiel: Die Münze ist nicht fair

$$H_1 : \pi \neq 0.5$$

- Falls die Münze fair ist, so liegt die Anzahl "Zahl" bei $n = 10$ Würfeln mit einer Wahrscheinlichkeit von mindestens 90% im Bereich

$$\{2, 3, 4, 5, 6, 7, 8\}$$

- Falls die Anzahl "Zahl" im Bereich $\{0, 1, 9, 10\}$ liegt, kann dies zwei Ursachen haben.

- 1 Ein sehr unwahrscheinliches Ereignis ist eingetreten.
- 2 Unsere Annahme, dass die Münze fair ist, stimmt nicht.

Entscheidungsregel, statistischer Test

Falls die Anzahl "Zahl" im Bereich $\{0, 1, 9, 10\}$ liegt, verwerfen wir die Vermutung, dass die Münze fair ist und gehen davon aus, dass die Münze nicht fair ist.

(Wir können uns natürlich irren.)

Statistischer Test: Prüfgröße, Teststatistik

- Eine **Prüfgröße (Teststatistik)** T ist eine zufällige Größe,
 - 1 anhand der wir entscheiden, ob die Nullhypothese H_0 plausibel ist.
 - 2 deren Verteilung wir kennen, falls die Nullhypothese H_0 zutrifft.
- Beispiel: Anzahl "Zahl" bei $n = 10$ Würfeln. Unter H_0 gilt:

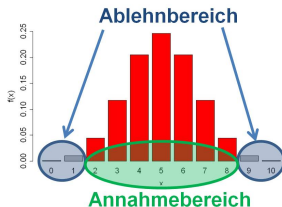
$$T = \sum_{i=1}^{10} X_i \sim B(n = 10, \pi = 0.5)$$

- Der **Annahmebereich** des Tests ist der Bereich, in dem die Prüfgröße T mit einer hohen Wahrscheinlichkeit (mindestens $1 - \alpha$) liegt.
Beispiel: $\alpha = 0.1$ und

$$\text{Annahmebereich} = \{2, 3, 4, 5, 6, 7, 8\}$$

- α heißt das **Signifikanzniveau** des Tests.
- Der **Ablehnbereich (kritische Bereich)** ist der Bereich, in dem die Prüfgröße T mit einer kleinen Wahrscheinlichkeit (höchstens α) liegt.
Beispiel: $\alpha = 0.1$ und

$$\text{Ablehnbereich} = \{0, 1, 9, 10\}$$



Statistischer Test: Experiment und Entscheidung

Wir ziehen eine Stichprobe und berechnen den Wert der Teststatistik T .

- Fall:** Der Wert der Teststatistik liegt im Annahmebereich.
→ Wir behalten die Nullhypothese H_0 bei.
- Fall:** Der Wert der Teststatistik liegt im Ablehnbereich.
→ Wir lehnen die Nullhypothese H_0 zugunsten der Gegenhypothese H_1 ab.

Festlegung des Signifikanzniveaus α

Beim Testen sind folgende Entscheidungen möglich:

H_0 : ablehnen oder H_0 : beibehalten

Damit sind zwei verschiedene Arten von Fehlern möglich:

Wahrheit \ Aktion	H_0 beibehalten	H_0 ablehnen
H_0 wahr	✓	Fehler 1. Art
H_0 falsch	Fehler 2. Art	✓

Man kann nicht beide Fehlerwahrscheinlichkeiten gleichzeitig kontrollieren! (Tradeoff!)

⇒ asymmetrische Vorgehensweise:

Der Fehler 1. Art wird kontrolliert durch die Angabe einer Obergrenze α („Signifikanzniveau“)

Übliche Werte für den Fehler erster Art sind:

$$\alpha = 0.1, \quad \alpha = 0.05, \quad \alpha = 0.01 \quad \alpha = 0.001$$

- Implizit wird also der Fehler 1. Art als schwerwiegender betrachtet.
- „konservative Perspektive“: Nullhypothese erst ablehnen, wenn wirklich nicht mehr mit den Daten verträglich.
- z.B. in der Medizin: H_0 : keine Wirkung.
⇒ Nur wenn die Wirkung des Medikaments überzeugend ist, soll es zugelassen werden.

- Die Nullhypothese wird abgelehnt, obwohl sie in Wirklichkeit richtig ist. Z.B.: Man behauptet, es bestünde ein Zusammenhang, obwohl in Wirklichkeit kein Zusammenhang besteht.
- Der Fehler 1. Art soll klein sein (üblich sind 5% oder 10%). Allerdings kann man nicht fordern, dass der Fehler 1. Art bei 0% liegen soll, sonst würde man die Nullhypothese nie ablehnen können.
⇒ Fehler 2. Art

Fehler 2. Art (β -Fehler):

- Die Nullhypothese wird beibehalten, obwohl sie in Wirklichkeit falsch ist.
- Ein guter statistischer Test garantiert bei einem vergebenen niedrigen Signifikanzniveau (als Schranke für den Fehler 1. Art) auch einen möglichst geringen Fehler 2. Art.

Folgerungen

- Die Nullhypothese wird höchstens mit Wahrscheinlichkeit α fälschlicherweise verworfen.
- Die Wahrscheinlichkeit für den Fehler 2. Art können wir nicht kontrollieren.



Ungleichbehandlung beider Fehlerarten
→ Grund für Formulierung eigentlicher Forschungsfrage
als statistische Alternative:
Entscheidung für H_1 durch α statistisch abgesichert!

- Ein Angeklagter steht vor Gericht.
- Hypothesen
 H_0 : „Angeklagter ist unschuldig“
und
 H_1 : „Angeklagter ist schuldig“
- Urteil: schuldig/nicht schuldig
- H_0 und H_1 sind so formuliert, da das Gericht die Schuld des Angeklagten beweisen muss, und nicht der Angeklagte seine Unschuld.



- Fehler 1. Art: Unschuldiger wird verurteilt
- Fehler 2. Art: Schuldiger wird nicht verurteilt

p-Wert

Der **p-Wert** ist die Wahrscheinlichkeit, dass die Testgröße

- den beobachteten Wert oder einen noch extremeren Wert („weiter weg von H_0 “) annimmt
- unter der Bedingung, dass H_0 wahr ist.

Bemerkungen

- 1 Für die Berechnung der p -Werte benötigt man eine Statistik-Software oder Tabellen.
- 2 Viele Statistik-Programme geben als Ergebnis eines statistischen Tests nur den p -Wert aus.

p-Wert Bestimmung: Einseitiger Test

- $P_{H_0}(10 \text{ Zahl}) = 0.001$
10 Zahl \Rightarrow p-Wert 0.001
- $P_{H_0}(9 \text{ Zahl}) = 0.01$
 $\rightarrow P_{H_0}(\text{mehr als 9 Zahl})$
 $= 0.001 + 0.01 = 0.011$
9 Zahl \Rightarrow p-Wert = 0.011
- $P_{H_0}(8 \text{ Zahl}) = 0.044$
 $\rightarrow P_{H_0}(\text{mehr als 8 Zahl})$
 $= 0.001 + 0.01 + 0.044 = 0.055$
8 Zahl \Rightarrow p-Wert = 0.055
- $P_{H_0}(7 \text{ Zahl}) = 0.117$
 $\rightarrow P_{H_0}(\text{mehr als 7 Zahl})$
 $= 0.001 + 0.01 + 0.044 + 0.117 = 0.172$
7 Zahl \Rightarrow p-Wert = 0.172

p-Wert Bestimmung: Zweiseitiger Test

- $P_{H_0}(10 \text{ Zahl}) + P_{H_0}(0 \text{ Zahl}) = 0.002$
10 Zahl \Rightarrow p-Wert 0.002
- $P_{H_0}(9 \text{ Zahl}) = 0.01$
 $\rightarrow P_{H_0}(\text{mindestens 9 Zahl oder höchstens 1 Zahl})$
 $= 0.001 + 0.01 + 0.01 + 0.001 = 0.022$
9 Zahl \Rightarrow p-Wert = 0.022
- $P_{H_0}(8 \text{ Zahl}) = 0.044$
 $\rightarrow P_{H_0}(\text{mindestens 8 Zahl oder höchstens 2 Zahl})$
 $= 2 * (0.001 + 0.01 + 0.044) = 0.055$
8 Zahl \Rightarrow p-Wert = 0.110
- $P_{H_0}(7 \text{ Zahl}) = 0.117$
 $\rightarrow P_{H_0}(\text{mehr als 7 Zahl oder höchstens 2 Zahl})$
 $= 2 * (0.001 + 0.01 + 0.044 + 0.117) = 0.344$
7 Zahl \Rightarrow p-Wert = 0.344

- Die Nullhypothese wird zum Niveau α genau dann abgelehnt, wenn der p-Wert kleiner als α ist
- Je kleiner der p-Wert desto weniger passen die Daten zur Nullhypothese

p-Wert und Signifikanzniveau

Die Nullhypothese wird genau dann abgelehnt, wenn der p-Wert kleiner oder gleich α ist.

Das ermöglicht ein direktes Ablesen der Testentscheidung aus entsprechenden Computerprogrammen. Daher wird der p-Wert meist zu den Test angegeben.

p-Wert: Interpretation

- Wahrscheinlichkeit betrifft das Auftreten der Daten und nicht die Wahrscheinlichkeit von H_0
- p-Wert ist **kein Maß** für die Stärke des Effekts. Daher sollten Begriffe wie "hochsignifikant" eher vermieden werden.
- Angabe des p-Wertes immer mit Schätzung des Effekts und Konfidenzintervall
- Bei kleinen p-Werten sollte nur $p < 0.001$ o.ä. angegeben werden.

Zweiseitiger approximativer Test auf den Anteilswert

- X Bernoulli-Variablen mit $p = P(X = 1)$.
- Zweiseitige Hypothese über den Anteilswert p

$$H_0 : p = p_0$$

$$H_1 : p \neq p_0$$

- Testgröße: Anteil in der Stichprobe X_1, \dots, X_n

$$R = \frac{1}{n} \sum_{i=1}^n X_i$$

- Stichprobenumfang n ist genügend groß

Hypothesen: $H_0: p = p_0$ versus $H_1: p \neq p_0$

Testentscheidung zum Signifikanzniveau α

Annahmehereich

$$p_0 \pm z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{p_0(1-p_0)}{n}}$$

H_0 wird abgelehnt, falls

$$R < p_0 - z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{p_0(1-p_0)}{n}}$$

oder

$$R > p_0 + z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{p_0(1-p_0)}{n}}$$

- Nullhypothese: $p = p_0 = 0.5$ („Münze ist fair.“)
- Signifikanzniveau: $\alpha = 0.05$
- $n = 50$ Münzwürfe
- Faustregel gültig?
 $50 \cdot 0.5 \cdot (1 - 0.5) = 50 \cdot 0.25 = 12.5 > 9 \rightarrow$ Normalverteilung
- Annahmehereich

$$\begin{aligned} p_0 \pm z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{p_0(1-p_0)}{n}} &= 0.5 \pm z_{1-\frac{0.05}{2}} \cdot \sqrt{\frac{0.5(1-0.5)}{50}} \\ &= 0.5 \pm z_{0.975} \cdot \sqrt{0.005} \\ &\approx 0.5 \pm 1.96 \cdot 0.07 \\ &= 0.5 \pm 0.14 \end{aligned}$$

- H_0 wird beibehalten, falls: $R \in [0.36; 0.64]$

Einseitiger Test auf den Anteilswert

- X Bernoulli-Variablen mit $\pi = P(X = 1)$.
- Einseitige Hypothese über den Anteilswert π

$$H_0: \pi \leq \pi_0$$

$$H_1: \pi > \pi_0$$

- Testgröße: Anteil in der Stichprobe X_1, \dots, X_n

$$R = \frac{1}{n} \sum_{i=1}^n X_i$$

- Stichprobenumfang n ist genügend groß
(Faustregel: $n\pi_0(1-\pi_0) > 9$)

Einseitiger Test auf den Anteilswert

Hypothesen: $H_0: \pi \leq \pi_0$ vs. $H_1: \pi > \pi_0$

Testentscheidung zum Signifikanzniveau α

Annahmehereich

$$R \leq \pi_0 + z_{1-\alpha} \cdot \sqrt{\frac{\pi_0(1-\pi_0)}{n}}$$

H_0 wird abgelehnt, falls

$$R > \pi_0 + z_{1-\alpha} \cdot \sqrt{\frac{\pi_0(1-\pi_0)}{n}}$$

$z_{1-\alpha}$ ist das $(1-\alpha)$ -Quantil der Standardnormalverteilung.

- X Bernoulli-Variablen mit $\pi = P(X = 1)$.
- Einseitige Hypothese über den Anteilswert p

$$H_0 : \pi \geq \pi_0$$

$$H_1 : \pi < \pi_0$$

- Testgröße: Anteil in der Stichprobe X_1, \dots, X_n

$$R = \frac{1}{n} \sum_{i=1}^n X_i$$

- Stichprobenumfang n ist genügend groß
(Faustregel: $n\pi_0(1 - \pi_0) > 9$)

Hypothesen: $H_0 : \pi \geq \pi_0$ vs. $H_1 : \pi < \pi_0$

Testentscheidung zum Signifikanzniveau α

Annahmebereich

$$R \geq \pi_0 - z_{1-\alpha} \cdot \sqrt{\frac{\pi_0(1 - \pi_0)}{n}}$$

H_0 wird abgelehnt, falls

$$R < \pi_0 - z_{1-\alpha} \cdot \sqrt{\frac{\pi_0(1 - \pi_0)}{n}}$$

$z_{1-\alpha}$ ist das $(1 - \alpha)$ -Quantil der Standardnormalverteilung.

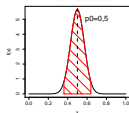
Vergleich einseitige Tests und zweiseitiger Test

Test auf Anteil mit einer Stichprobe der Größe $n = 50$ und Signifikanzniveau $\alpha = 0.05$

$$H_0 : \pi = 0.5$$

$$H_1 : \pi \neq 0.5$$

Annahmebereich

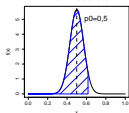


[0.36; 0.64]

$$H_0 : \pi \leq 0.5$$

$$H_1 : \pi > 0.5$$

Annahmebereich

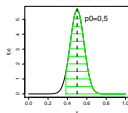


[0; 0.62]

$$H_0 : \pi \geq 0.5$$

$$H_1 : \pi < 0.5$$

Annahmebereich



[0.38; 1]

p-Wert: Ein- und zweiseitiger Test

In vielen Fällen halbiert sich der p-Wert bei einseitiger Fragestellung.

Beispiel: Binomialtest 10 Experimente T Anzahl Treffer.

Ergebnis: $T = 9$

- zweiseitiger Test: $H_0 : \pi = 0.5$ gegen $H_1 : \pi \neq 0.5$.

Extrem oder gleich extreme Werte von T : 9, 10, 1, 0

p-Wert =

$$P_{H_0}(T \in \{9, 10, 0, 1\}) = 0.01 + 0.001 + 0.01 + 0.001 = 0.022$$

- einseitiger Test: $H_0 : \pi \leq 0.5$ gegen $H_1 : \pi > 0.5$

Extrem oder gleich extreme Werte von T : 9, 10

$$\text{p-Wert} = P_{H_0}(T \in \{9, 10\}) = 0.11$$

- Wir interessieren uns für den Erwartungswert μ einer metrischen Zufallsgröße.
Beispiele: Alter, Einkommen, Körpergröße, Scorewert ...
- Wir können einseitige oder zweiseitige Hypothesen formulieren.
- Beispiele
 - "Vor 10 Jahren betrug die Durchschnittsgröße von Studienanfängern und -anfängerinnen 167 cm. Heute sind sie im Schnitt größer als 167 cm."
 - Die Differenz zwischen Gewicht vor und nach einer Diät ist 0.

- X Zufallsgröße mit Erwartungswert μ .

- Zweiseitige Hypothese über μ :

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu \neq \mu_0$$

- Testgröße: Mittelwert in der Stichprobe X_1, \dots, X_n :

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

- Stichprobenumfang n ist genügend groß (Faustregel $n > 30$)

Zweiseitiger Gauss-Test auf den Erwartungswert μ

Hypothesen: $H_0 : \mu = \mu_0$ vs. $H_1 : \mu \neq \mu_0$

Testentscheidung zum Signifikanzniveau α

Annahmehereich

$$\mu_0 \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{S^2}{n}}$$

mit

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

H_0 wird abgelehnt, falls

$$\bar{X} < \mu_0 - z_{1-\frac{\alpha}{2}} \sqrt{\frac{S^2}{n}} \quad \text{oder} \quad \bar{X} > \mu_0 + z_{1-\frac{\alpha}{2}} \sqrt{\frac{S^2}{n}}$$

$z_{1-\frac{\alpha}{2}}$ ist das $(1 - \alpha/2)$ -Quantil der Standardnormalverteilung.

Zweiseitiger Gauss-Test: p-Wert

Der p-Wert ergibt sich zu :

$$2 \cdot \left[1 - \Phi \left(\frac{|\bar{X} - \mu_0|}{\sqrt{\frac{S^2}{n}}} \right) \right]$$

Φ ist die Verteilungsfunktion der Standard Normalverteilung

$$(\bar{X} - \mu_0) / \sqrt{\frac{S^2}{n}}$$

wird als t-Wert (oder z-Wert) bezeichnet.

Beispiel: Verändert sich der Blutdruck nach einer Intervention

Beispiel zweiseitiger Test

- Nullhypothese: Die Blutdruckdifferenz ist 0.

$$H_0 : \mu = 0$$

$$H_1 : \mu \neq 0$$

- Testgröße: Durchschnittlicher Blutdruck
- Faustregel $n = 62 > 30$ ist erfüllt
→ zweiseitiger **Gauß-Test**
- Falls wir annehmen, dass die Blutdruckdifferenz normalverteilt ist, können wir auch den zweiseitigen **t-Test** anwenden.
(Standardeinstellung in SPSS)

Dualität Test und Konfidenzintervall

- Annahmereich:** Wir behalten H_0 bei, falls die Testgröße T in der Nähe von μ_0 liegt:
- Äquivalente Formulierung über ein **Konfidenzintervall:** Wir behalten H_0 bei, falls μ_0 in der Nähe der Testgröße liegt
- Wir behalten H_0 bei, falls μ_0 im Konfidenzintervall für die Differenz liegt
- Dabei hängen das Konfidenzniveau γ und das Signifikanzniveau α wie folgt zusammen:
 $1 - \alpha = \gamma$
- Dies gilt sehr allgemein für zweiseitige Test und Konfidenzintervalle
- Dies Prinzip kann zur Konstruktion von Konfidenzintervallen verwendet werden

Einseitiger Gauss-Test auf den Erwartungswert μ

- X Zufallsgröße mit Erwartungswert μ .
- Einseitige Hypothese über μ :

$$H_0 : \mu \leq \mu_0$$

$$H_1 : \mu > \mu_0$$

- Testgröße: Mittelwert in der Stichprobe X_1, \dots, X_n :

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

- Stichprobenumfang n ist genügend groß
(Faustregel $n > 30$)

Einseitiger Gauss-Test auf den Erwartungswert μ

Hypothesen: $H_0 : \mu \leq \mu_0$ vs. $H_1 : \mu > \mu_0$

Testentscheidung zum Signifikanzniveau α

Annahmereich

$$\bar{X} \leq \mu_0 + z_{1-\alpha} \sqrt{\frac{S^2}{n}}$$

H_0 wird abgelehnt, falls

$$\bar{X} > \mu_0 + z_{1-\alpha} \sqrt{\frac{S^2}{n}}$$

$z_{1-\alpha}$ ist das $(1 - \alpha)$ -Quantil der Standardnormalverteilung.

Was tun wir, falls die Faustregel $n > 30$ nicht erfüllt ist?

Zusätzliche Voraussetzung

- Zufallsgröße X ist normalverteilt.

Wir stellen keine Bedingung an die Stichprobengröße n .

t-Test

- Der einseitige bzw. der zweiseitige t-Test auf den Erwartungswert μ hat die gleiche Form wie der einseitige bzw. zweiseitige Gauss-Test.
- Der t-Test unterscheidet sich vom Gauss-Test dadurch, dass wir das **Quantil z der Standardnormalverteilung** durch das **Quantil t der t -Verteilung mit $n - 1$ Freiheitsgraden** ersetzen.

Approximativer Test auf Erwartungswert-Differenz bei unabhängigen Stichproben

Voraussetzungen:

- X und Y sind zwei Größen mit Erwartungswerten μ_X und μ_Y
- X_1, \dots, X_m und Y_1, \dots, Y_n **unabhängige** Stichproben
- Testgröße: standardisierte Differenz der Mittelwerte

$$T = \frac{\bar{X} - \bar{Y}}{\frac{1}{m} \sum_{i=1}^m X_i - \frac{1}{n} \sum_{i=1}^n Y_i}$$

- Faustregel: Stichprobenumfänge $m, n > 30$

Verbundene, abhängige Stichprobe

Zwei Stichproben heißen **verbunden**, falls an einem Merkmalsträger (z.B. einer Person) zwei vergleichbare Merkmale erhoben werden. Man nennt verbundene Stichproben oft auch **abhängige Stichproben**.

Beispiel: Das Ziel einer medizinischen Studie ist es, die Wirkung eines cholesterin-senkenden Medikaments zu überprüfen.

- Unterteilung der Probanden und Probandinnen in 2 Gruppen: 1 Gruppe erhält das Medikament, 1 Gruppe erhält ein Placebo.
unverbundene Stichprobe
- Alle Probanden und Probandinnen erhalten das Medikament. Von allen Personen wird der Cholesterinspiegel am Anfang und am Ende der Studie erhoben.
verbundene Stichprobe

Differenz von Erwartungswerten bei unabhängigen Stichproben

Annahmereich

Für die beiden einseitigen Tests und den zweiseitigen Test auf die Differenz

$$\begin{array}{lll} H_0 : \mu_X - \mu_Y = d_0 & H_0 : \mu_X - \mu_Y \leq d_0 & H_0 : \mu_X - \mu_Y \geq d_0 \\ H_1 : \mu_X - \mu_Y \neq d_0 & H_1 : \mu_X - \mu_Y > d_0 & H_1 : \mu_X - \mu_Y < d_0 \end{array}$$

ist der Annahmereich

$$\text{mit } d_0 \pm z_{1-\frac{\alpha}{2}} \cdot s \quad] -\infty, d_0 + z_{1-\alpha} \cdot s] \quad [d_0 - z_{1-\alpha} \cdot s, \infty [$$

$$s = \sqrt{\frac{s_X^2}{m} + \frac{s_Y^2}{n}}$$

- Hören Personen in den alten Bundesländern im Schnitt mehr Radio?
 X : Hördauer im den alten Bundesländern,
 Y : Radiodauer in den neuen Bundesländern

$$H_0 : \mu_X - \mu_Y \leq 0$$

$$H_1 : \mu_X - \mu_Y > 0$$

- Befragung unter 253 Personen aus den alten Bundesländern und 932 Personen aus den neuen Bundesländern
 - unverbundene Stichproben X_1, \dots, X_{253} und Y_1, \dots, Y_{932}
 - Stichprobengrößen $m = 253, n = 932 > 30$
- Durchschnittliche Hördauer:
 - 11.4 h (Standardabweichung 8.4 h) in den alten Bundesländern
 - 9.5 h (Standardabweichung 8.4 h) in den neuen Bundesländern



Doppelter t -Test auf die Erwartungswertdifferenz bei unabhängigen Stichproben

Voraussetzungen:

- X und Y sind zwei Größen mit Erwartungswerten μ_X und μ_Y
- X_1, \dots, X_m und Y_1, \dots, Y_n **unabhängige** Stichproben
- Testgröße: Differenz der Mittelwerte

$$T = \bar{X} - \bar{Y} = \frac{1}{m} \sum_{i=1}^m X_i - \frac{1}{n} \sum_{i=1}^n Y_i$$

- X und Y sind **normalverteilt**.
- Die Varianzen sind gleich $\sigma_X^2 = \sigma_Y^2$



- Signifikanzniveau: $\alpha = 0, 1$
- Differenz der Radio-Hördauer

$$\bar{X} - \bar{Y} = 11.4 - 9.5 = 1.9$$

- Annahmebereich

$$\begin{aligned} \bar{X} - \bar{Y} &\leq z_{1-\alpha} \cdot \sqrt{\frac{s_X^2}{m} + \frac{s_Y^2}{n}} \\ &= z_{0.9} \cdot \sqrt{\frac{8.4^2}{932} + \frac{8.4^2}{253}} \\ &\approx 1.28 \cdot 0.65 \\ &\approx 0.83 \end{aligned}$$

- H_0 wird abgelehnt, Personen aus den alten Bundesländern hören signifikant länger Radio.



Doppelter t -Test auf die Erwartungswertdifferenz bei unabhängigen Stichproben

Annahmebereich

Für die beiden einseitigen t -Tests und den zweiseitigen t -Test auf die Differenz

$$\begin{array}{lll} H_0 : \mu_X - \mu_Y = d_0 & H_0 : \mu_X - \mu_Y \leq d_0 & H_0 : \mu_X - \mu_Y \geq d_0 \\ H_1 : \mu_X - \mu_Y \neq d_0 & H_1 : \mu_X - \mu_Y > d_0 & H_1 : \mu_X - \mu_Y < d_0 \end{array}$$

ist der Annahmebereich

$$d_0 \pm t_{(m+n-2; 1-\frac{\alpha}{2})} s_d \quad] -\infty, d_0 + t_{(m+n-2; 1-\alpha)} s_d] \quad [d_0 - t_{(m+n-2; 1-\alpha)} s_d, \infty[$$

mit

$$\begin{aligned} s_d &= s_{X,Y} \cdot \sqrt{\frac{1}{m} + \frac{1}{n}} \\ s_{X,Y}^2 &= \frac{1}{m+n-2} ((m-1)S_X^2 + (n-1)S_Y^2) \end{aligned}$$

$t_{(m+n-2; 1-\frac{\alpha}{2})}$ und $t_{(m+n-2; 1-\alpha)}$ sind die Quantile der t -Verteilung mit $m+n-2$ Freiheitsgraden.



- ④ Gegeben ist eine verbundene Stichprobe
 X_1, \dots, X_n und Y_1, \dots, Y_n

- ④ Bilde die Differenz

$$W_i = X_i - Y_i \quad i = 1, \dots, n$$

- ④ Führe einen Test auf den Erwartungswert von W durch

- $n > 30 \rightarrow$ Gauß-Test
- W normalverteilt \rightarrow t -Test

- X und Y sind zwei Bernoulli-Größen mit

$$p_X = P(X = 1)$$

$$p_Y = P(Y = 1)$$

- X_1, \dots, X_m und Y_1, \dots, Y_n **unabhängige** Stichproben
- Testgröße: Differenz der Anteile

$$\begin{aligned} T &= R_X - R_Y \\ &= \frac{1}{m} \sum_{i=1}^m X_i - \frac{1}{n} \sum_{i=1}^n Y_i \end{aligned}$$

- Faustregel: Stichprobenumfänge $m, n > 30$

Differenz von Anteilen bei unabhängigen Stichproben

Annahmebereich

Für die beiden einseitigen Tests und den zweiseitigen Test auf die Differenz

$$H_0: p_X - p_Y = 0 \quad H_0: p_X - p_Y \leq 0 \quad H_0: p_X - p_Y \geq 0$$

$$H_1: p_X - p_Y \neq 0 \quad H_1: p_X - p_Y > 0 \quad H_1: p_X - p_Y < 0$$

ist der Annahmebereich

$$\left[-z_{1-\frac{\alpha}{2}} \cdot s_R; z_{1-\frac{\alpha}{2}} \cdot s_R \right] \quad] -\infty, z_{1-\alpha} \cdot s_R] \quad [-z_{1-\alpha} \cdot s_R, \infty[$$

mit

$$s_R = \sqrt{R(1-R) \left(\frac{1}{m} + \frac{1}{n} \right)} \quad \text{und} \quad R = \frac{m \cdot R_X + n \cdot R_Y}{m + n}$$

Beispiel: Ist Fernsehen informativ?

Weiterführung des Beispiels aus dem Thema „Schätzen“

- Beurteilen Personen aus den alten Bundesländern den Informationsgehalt im Fernsehen anders als Personen aus den neuen Bundesländern?

zweiseitiger Test

X : Person aus den alten Bundesländern hält Fernsehen für informativ

Y : Person aus den neuen Bundesländern hält Fernsehen für informativ

- Signifikanzniveau: $\alpha = 0.05$
- Umfrage: 253 Personen aus den alten Bundesländern, 932 Personen aus den neuen Bundesländern:
 „Halten Sie Fernsehen für informativ? Ja/Nein“
 - **unverbundene Stichproben** X_1, \dots, X_{253} und Y_1, \dots, Y_{932}
 - **Stichprobengrößen** $m = 253, n = 932 > 30$

Beispiel: Ist Fernsehen informativ?

- alte Bundesländer: 206 Personen halten Fernsehen für informativ
neue Bundesländer: 747 Personen halten Fernsehen für informativ
- Anteile

$$R_X = \frac{206}{253} \approx 0.81 \quad \text{und} \quad R_Y = \frac{747}{932} \approx 0.80$$

- Standardfehler

$$R = \frac{m \cdot R_X + n \cdot R_Y}{m + n} = \frac{253 \cdot 0.81 + 932 \cdot 0.8}{932 + 253}$$

$$= \frac{950.53}{1185} \approx 0.802$$

$$s_R = \sqrt{R(1-R) \left(\frac{1}{m} + \frac{1}{n} \right)} = \sqrt{0.802 \cdot (1 - 0.802) \left(\frac{1}{932} + \frac{1}{253} \right)}$$

$$\approx 0.03$$

Zusammenhang zwischen 2 kategorialen Merkmalen

Sind zwei kategoriale Merkmale unabhängig?

Beispiele

- Gibt es einen Zusammenhang zwischen besuchter Schule (Hauptschule, Realschule, Gymnasium) und Fernsehkonsum (hoch/niedrig)?
- Gibt es einen Zusammenhang zwischen Geschlecht (m/w) und der Affinität zu Fußball (Fan/kein Fan)?
- ...

→ Empirische Untersuchung mittels Kontingenztafeln und Kennzahlen

Beispiel: Ist Fernsehen informativ?

- Annahmehbereich

$$\pm z_{1-\frac{\alpha}{2}} \cdot s_R \approx \pm z_{0.975} \cdot 0.03$$

$$\approx \pm 1.96 \cdot 0.03$$

$$\approx \pm 0.059$$

- Differenz der Anteile in der Stichprobe

$$R_X - R_Y \approx 0.81 - 0.8 = 0.01$$

- H_0 wird beibehalten, der Unterschied zwischen alten und neuen Bundesländern ist nicht signifikant

Randverteilungen

Die zu den einzelnen Merkmalen gehörigen empirischen Häufigkeitsverteilungen heißen **Randverteilungen** oder **marginale** Verteilungen.

Fußball \ Geschlecht	ja	nein	Summe
m	87	10	97
w	23	31	54
Summe	110	41	151

Randverteilung Geschlecht

Randverteilung Fußballfan

m	w		ja	nein	
97	54	← Absolute Häufigkeiten	110	41	← Absolute Häufigkeiten
$\frac{97}{151}$	$\frac{54}{151}$	← Relative Häufigkeiten	$\frac{110}{151}$	$\frac{41}{151}$	← Relative Häufigkeiten

Unter der **bedingten Verteilung** von X gegeben $Y = B_j$ versteht man die Verteilung von X in der Teilgesamtheit der Untersuchungseinheiten, die die Ausprägung B_j des Merkmals Y aufweisen.

	Fußball		
Geschlecht	ja	nein	Summe
m	87	10	97
w	23	31	54
Summe	110	41	151

Verteilung „Fußballfan“ bei Männern

ja	nein	Ges.
87	10	97
90%	10%	100%

Verteilung „Geschlecht“ bei Fußballfans

m	w	Ges.
87	23	110
79%	21%	100%

- Falls die beiden Merkmale unabhängig voneinander sind, so sollte

„relative Häufigkeit Fußballfan“
 \approx „relative Häufigkeit Fußballfans unter Frauen“
 \approx „relative Häufigkeit Fußballfans unter Männern“

- Formel:**

$$\text{bedingte relative Häufigkeit } \frac{n_{ij}}{n_{j\bullet}} = \frac{n_{i\bullet}}{n} \text{ relative Randhäufigkeit}$$

- Daraus folgt

$$e_{ij} := \frac{n_{i\bullet} n_{\bullet j}}{n}$$

ist die **unter Unabhängigkeit erwartete Anzahl**.

- Falls die Merkmale **unabhängig** sind, sollte gelten:

$$e_{ij} \approx n_{ij}$$

χ^2 -Unabhängigkeitstest

- Zwei Zufallsgrößen X und Y mit k bzw. l Ausprägungen

$$p_{ij} = P(X = i, Y = j)$$

$$p_{i\bullet} = P(X = i) \quad p_{\bullet j} = P(Y = j)$$

- Hypothesen:

H_0 : X und Y sind stochastisch unabhängig

$$p_{ij} = p_{i\bullet} \cdot p_{\bullet j} \text{ für alle } i = 1, \dots, k, j = 1, \dots, l$$

H_1 : X und Y sind stochastisch abhängig

$$p_{ij} \neq p_{i\bullet} \cdot p_{\bullet j} \text{ für mindestens eine } ij\text{-Kombination}$$

- Prüfgröße:

$$\chi^2 = \sum_{i=1}^k \sum_{j=1}^l \frac{(n_{ij} - e_{ij})^2}{e_{ij}}$$

- Faustregel $n_{ij} \geq 5$ für alle i, j

χ^2 -Unabhängigkeitstest

Annahmebereich

Für den χ^2 -Unabhängigkeitstest ist der Annahmebereich

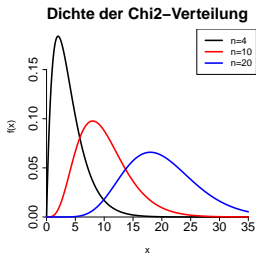
$$\chi^2 \leq q_{1-\alpha, (k-1)(l-1)}$$

Die Nullhypothese wird abgelehnt, falls

$$\chi^2 > q_{1-\alpha, (k-1)(l-1)}$$

Dabei ist $q_{1-\alpha, (k-1)(l-1)}$ das

- $(1 - \alpha)$ -Quantil
- der χ^2 -Verteilung
- mit $(k - 1) \cdot (l - 1)$ Freiheitsgraden.



Beispiel

- Signifikanzniveau: $\alpha = 0.01$
- Faustregel gültig?
Besetzungszahlen $n_{ij} \geq 5$
- Bestimmung der Freiheitsgrade: $k = l = 2$

$$\text{Freiheitsgrade} = (k - 1) \cdot (l - 1) = (2 - 1) \cdot (2 - 1) = 1$$

$$q_{1-0.01; (2-1)(2-1)} = q_{0.99; 1} \approx 6,63$$

- H_0 wird abgelehnt

Beispiel

$$e_{ij} = \frac{n_{i \cdot} n_{\cdot j}}{n}$$

Erwartete Besetzungszahlen bei Unabhängigkeit

Subjekt	ja	nein	Summe
m	97	10	97
w	23	31	54
Summe	110	41	151

	ja ($j=1$)	nein ($j=2$)
m ($i=1$)	$\frac{97 \cdot 110}{151} \approx 71$	$\frac{97 \cdot 41}{151} \approx 26$
w ($i=2$)	$\frac{54 \cdot 110}{151} \approx 39$	$\frac{54 \cdot 41}{151} \approx 15$

$$\begin{aligned} \chi^2 &= \sum_{i=1}^k \sum_{j=1}^l \frac{(n_{ij} - e_{ij})^2}{e_{ij}} \\ &\approx \frac{(87 - 71)^2}{71} + \frac{(10 - 26)^2}{26} + \frac{(23 - 39)^2}{39} + \frac{(31 - 15)^2}{15} \\ &\approx 37.09 \end{aligned}$$

Unabhängigkeit und Differenz von Anteilen

Die beide Fragen:

- Gibt es Unterschiede in den Anteilen von $Y = 1$ zweier Gruppen ?
- Gibt es einen Zusammenhang zwischen Gruppen-Zugehörigkeit und einem binären Merkmal Y ?

sind äquivalent.

Die beiden Testprozeduren Gauss Test für Differenzen von Anteilen und χ^2 -Unabhängigkeitstest führen zum gleichen Ergebnis.

Voraussetzungen:

- X und Y sind zwei Bernoulli-Größen mit

$$p_X = P(X = 1)$$

$$p_Y = P(Y = 1)$$

- $(X_1, Y_1), \dots, (X_n, Y_n)$ **abhängige, verbundene** Stichproben
- Absolute Häufigkeiten werden in einer Kontingenztafel festgehalten

	Y=0	Y=1
X=0	n_{11}	n_{12}
X=1	n_{21}	n_{22}

Nichtparametrische Tests

- Bis auf den χ^2 -Unabhängigkeits-Test bauen alle Tests auf der (zumindestens approximativen Gültigkeit der) Normalverteilungsannahme auf.
- Problematisch, z.B.
 - bei kleinen Stichprobenumfängen (z.B. in der Vorbereitung von strukturierten Beobachtungen, bei nicht reaktiven Verfahren oder in der Psychologie und Medizin)
 - oder bei ordinalen Daten mit wenigen unterschiedlichen Ausprägungen.
- Hier kann die unreflektierte Anwendung der Standardtests zu Fehlergebnissen führen.
- Ein wichtiger Ausweg: nichtparametrische Tests = Verteilungsfreie Verfahren
- Hier wird die Information in den Beobachtungen auf Ränge, bzw. größer/kleiner Vergleiche reduziert.
- Bekannteste Beispiele: Wilcoxon-Test, Vorzeichenstest.

Voraussetzungen:

- $(X_1, Y_1), \dots, (X_n, Y_n)$ **abhängige, verbundene** binäre Zufallsvariable
- Absolute Häufigkeiten werden in einer Kontingenztafel festgehalten

	Y=0	Y=1
X=0	n_{11}	n_{12}
X=1	n_{21}	n_{22}

Nullhypothese $H_0: P(X = 1) = P(Y = 1)$

Die beiden Zellen n_{11} und n_{22} liefern keine Information zum Unterschied der beiden Wahrscheinlichkeiten $P(X=1)$ und $P(Y=1)$. Betrachte die beiden Zellen der Wechsler: n_{12} und n_{21}

Unter der Nullhypothese sollte $n_{12} = n_{21}$ gelten.

Daher ist unter H_0 :

$$n_{12} \sim B(n_{12} + n_{21}, 0.5)$$

Man führt also einen Binomialtest mit $\pi = 0.5$ und $n = n_{12} + n_{21}$ durch.

Der Wilcoxon Test für unabhängige Stichproben

Voraussetzungen:

- X und Y sind zwei Größen mit Medianen med_X und med_Y
- X_1, \dots, X_m und Y_1, \dots, Y_n **unabhängige** Stichproben
- $H_0: med_X = med_Y$ vs. $H_1: med_X \neq med_Y$

Grundidee: Betrachte die Ränge aus allen Beobachtungen X_i und Y_j und bezeichne diese mit $rg(X_i)$ und $rg(Y_j)$, z.B.

$X_1 = 3, X_2 = 5, Y_1 = 6, Y_2 = 1, Y_3 = 4 \Rightarrow$

$rg(X_1) = 2, rg(X_2) = 4, rg(Y_1) = 5, rg(Y_2) = 1, rg(Y_3) = 3$

Testgröße:

$$T = \sum_{i=1}^m rg(X_i)$$

Die exakte Verteilung von T kann berechnet werden. Für hinreichend große n und m kann sie durch eine NV approximiert werden. Ablehnung von H_0 falls große und kleine T .

- Gegeben sei ein rein zufälliger Datensatz mit 50 Variablen ohne irgendeinen Zusammenhang.
 - Testen aller Variablenpaare auf einen Zusammenhang (1225 Tests)
Bei vorgegebener Irrtumswahrscheinlichkeit von 5% gilt für die Anzahl fälschlich verworfener Nullhypothesen $X \sim B(1225, 0.05)$ und somit $E(X) = 61.25$.
Im Durchschnitt wird also mehr als 61 mal die Nullhypothese, dass kein Zusammenhang besteht, verworfen.
- ⇒ wenige, sinnvolle Hypothesen *vorher inhaltlich* überlegen. In den Daten entdeckte „Zusammenhänge“ als statistisch signifikant nachzuweisen, ist (fast) zirkulär.
- Es gibt Ansätze, wie man bei großen Hypothesensystemen diesem Problem entkommt.
- ⇒ Theorie des multiplen Testens
- Z.B. Adjustierung des Irrtumswahrscheinlichkeit: Statt α betrachte man $\alpha/($ Anzahl der Tests). Diese spezielle Korrektur ist aber meist überkonservativ und kann durch bessere Korrekturen ersetzt werden.

Bestimmung mit Obergrenze für Fehler 2. Art (β -Fehler)
Interessante Größen:

-
- $\beta = P(\text{Nullhypothese wird nicht abgelehnt} \mid H_1 \text{ wahr})$
- $1 - \beta$ wird auch als Power bezeichnet
- Üblicher Wert: Power = 0.8 bzw. $\beta = 0.2$

Fehler 2. Art im Binomialtest

Interessante Größen:

- $H_0: \pi = 0.5$ versus
 $H_1: \pi > 0.5$
- $\beta: P(H_0 \text{ wird beibehalten} \mid H_1)$
- π_w : wahre Erfolgswahrscheinlichkeit
- c : kritischer Wert
(Überschreiten führt zu Ablehnung von H_0)

Numerische Werte:

$$n = 10 \text{ und } c = 7$$

$$\pi_w = 0.6 \text{ und } \beta = 0.95$$

$$\pi_w = 0.7 \text{ und } \beta = 0.85$$

$$\pi_w = 0.8 \text{ und } \beta = 0.62$$

$$n = 20 \text{ und } c = 13$$

$$\pi_w = 0.6 \text{ und } \beta = 0.87$$

$$\pi_w = 0.7 \text{ und } \beta = 0.58$$

$$\pi_w = 0.8 \text{ und } \beta = 0.02$$

$$n = 100 \text{ und } c = 57$$

$$\pi_w = 0.6 \text{ und } \beta = 0.37$$

$$\pi_w = 0.7 \text{ und } \beta = 0.007$$

$$\pi_w = 0.8 \text{ und } \beta < 0.0001$$

Fehler 2. Art im Beispiel des Binomialtests

Fehler 2. Art sinkt hier mit:

- Wachsender wahrer Erfolgswahrscheinlichkeit ($\pi > 0.5$)
- Wachsendem Stichprobenumfang

- **Festlegen von:**

α, β , Effektgröße π_w

- **Beispiel:**

$\pi_w = 0.7$

$\Rightarrow n = 20 \Rightarrow \beta = 0.58$ zu klein

$\Rightarrow n = 100 \Rightarrow \beta = 0.07$ zu groß

- **Wähle kleinstes n mit:**

$\beta \leq 0.2$

- **Hier:**

$n = 37, \beta = 0.19$

$n = 36, \beta = 0.26$

Vergleich zweier Gruppen

- Zielgröße: Y
- Relevanter Unterschied: d
- Annahme zur Streuung: σ
- Signifikanzniveau: $\alpha = 0.05$
- Fehler 2.Art: $\beta = 0.2$

Bestimmung des Stichprobenumfangs beim t-Test

Einige Werte für den Stichprobenumfang im Zwei-Stichproben t-Test

- unter $\alpha = 0.05, \beta = 0.2$
- zweiseitige Fragestellung

Beachte: Es kommt nur auf das Verhältnis d/σ (Effektstärke) an.

d/σ	$n_1 = n_2$
2	5
1.33	10
1	17
0.66	37
0.58	50
0.29	100

Zusammenfassung Tests

Statistischer Test

Ein statistischer Test ist eine Untersuchung, ob man eine Hypothese über die Grundgesamtheit mit Hilfe einer Stichprobe widerlegen kann

Wichtige Tests

- einseitiger Test \leftrightarrow zweiseitiger Test
- Tests auf Anteile, Erwartungswerte und deren Differenzen
 - eine Stichprobe
 - zwei Stichproben \rightarrow verbunden oder unverbunden
- Test auf Unabhängigkeit: χ^2 -Unabhängigkeitstest

- Statistische Signifikanz bedeutet: Das Ergebnis ist nicht durch Zufall erklärbar.
- Statistische Signifikanz bedeutet nicht unbedingt, dass der Unterschied relevant ist.
→ Daher immer die Größe des Effekts (Schätzung) angeben!
- Ein statistischer Test liefert aber allein noch keinen kausalen Zusammenhang.
- Bei nicht signifikantem Ergebnis immer Konfidenzintervalle der entsprechenden Parameter angeben. Sie geben eine Idee, wie groß der nicht nachgewiesene Effekt sein könnte.

- 1 Einführung
- 1 Wahrscheinlichkeitsrechnung
- 2 Zufallsvariablen und ihre Verteilung
- 3 Statistische Inferenz
- 4 Hypothesentests
- 5 Regression

Lineare Regressionsmodelle

Statistisches Modell

Deskriptive Statistik:

Gegeben Datenpunkte (Y_i, X_i) schätze die beste Gerade

$$Y_i = \beta_0 + \beta_1 X_i, \quad i = 1, \dots, n.$$

(mit der Methode der kleinsten Quadrate)

- Linearer Zusammenhang
- Im Folgenden:
Probabilistische Modelle in Analogie zu den deskriptiven Modellen aus Statistik I

Zunächst Modelle mit nur *einer* unabhängigen Variable.

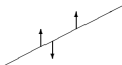
Statistische Sichtweise:

- Modell

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

β_1 „Elastizität“: Wirkung der Änderung von X_i um eine Einheit

- gestört durch zufällige Fehler ε_i

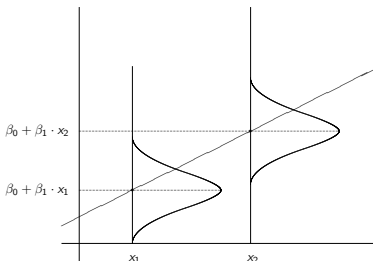


Beobachtung von Datenpaaren (X_i, Y_i) , $i = 1, \dots, n$ mit

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i,$$

wobei sich die Annahmen auf den zufälligen Störterm beziehen:

- $E(\varepsilon_i) = 0$
- $\text{Var}(\varepsilon_i) = \sigma^2$ für alle i gleich
- $\varepsilon_{i1}, \varepsilon_{i2}$ stochastisch unabhängig für $i_1 \neq i_2$
- $\varepsilon_i \sim N(0, \sigma^2)$ (zusätzlich, bei großen Stichproben nicht erforderlich)



Die Schätzwerte werden üblicherweise mit $\hat{\beta}_0, \hat{\beta}_1$ und $\hat{\sigma}^2$ bezeichnet. In der eben beschriebenen Situation gilt:

- Die (Maximum Likelihood) Schätzer lauten:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2},$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X},$$

$$\hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^n \hat{\varepsilon}_i^2$$

mit den geschätzten Residuen

$$\hat{\varepsilon}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i.$$

Mit

$$\hat{\sigma}_{\hat{\beta}_0} := \frac{\hat{\sigma} \sqrt{\sum_{i=1}^n X_i^2}}{\sqrt{n \sum_{i=1}^n (X_i - \bar{X})^2}}$$

gilt

$$\frac{\hat{\beta}_0 - \beta_0}{\hat{\sigma}_{\hat{\beta}_0}} \sim t^{(n-2)}$$

und analog mit

$$\hat{\sigma}_{\hat{\beta}_1} := \frac{\hat{\sigma}}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2}}$$

gilt

$$\frac{\hat{\beta}_1 - \beta_1}{\hat{\sigma}_{\hat{\beta}_1}} \sim t^{(n-2)}.$$

- $\hat{\beta}_0$ und $\hat{\beta}_1$ sind die *KQ*-Schätzer aus Statistik I. Unter Normalverteilung fällt hier das *ML*- mit dem *KQ*-Prinzip zusammen.
- Man kann unmittelbar Tests und Konfidenzintervalle ermitteln (völlig analog zum Vorgehen in Kapitel 3 und 4).

Konfidenzintervalle zum Sicherheitsgrad γ :

$$\text{für } \beta_0 : \left[\hat{\beta}_0 \pm \hat{\sigma}_{\hat{\beta}_0} \cdot t_{\frac{1-\gamma}{2}}^{(n-2)} \right]$$

$$\text{für } \beta_1 : \left[\hat{\beta}_1 \pm \hat{\sigma}_{\hat{\beta}_1} \cdot t_{\frac{1-\gamma}{2}}^{(n-2)} \right]$$

Tests für die Parameter des Modells

Mit der Teststatistik

$$T_{\hat{\beta}_1} = \frac{\hat{\beta}_1 - \beta_1^*}{\hat{\sigma}_{\hat{\beta}_1}}$$

ergibt sich

	Hypothesen	gegen	kritische Region
I.	$H_0 : \beta_1 \leq \beta_1^*$	$\beta_1 > \beta_1^*$	$T \geq t_{1-\alpha}^{(n-2)}$
II.	$H_0 : \beta_1 \geq \beta_1^*$	$\beta_1 < \beta_1^*$	$T \leq t_{1-\alpha}^{(n-2)}$
III.	$H_0 : \beta_1 = \beta_1^*$	$\beta_1 \neq \beta_1^*$	$ T \geq t_{1-\frac{\alpha}{2}}^{(n-2)}$

(analog für $\hat{\beta}_0$).

Von besonderem Interesse ist der Fall $\beta_1^* = 0$ (Steigung gleich 0): Hiermit kann man überprüfen, ob die X_1, \dots, X_n einen signifikanten Einfluss hat oder nicht.

Typischer Output

Koeffizienten*

			Standardisierte Koeffizienten	T	Signifikanz
	B	Standardfehler	Beta		
Konstante	$\hat{\beta}_0$	$\hat{\sigma}_{\hat{\beta}_0}$	5)	1)	3)
Unabhängige Variable	$\hat{\beta}_1$	$\hat{\sigma}_{\hat{\beta}_1}$	6)	2)	4)

- 1) Wert der Teststatistik

$$T_{\beta_0} = \frac{\hat{\beta}_0}{\hat{\sigma}_{\beta_0}}$$

zum Testen von $H_0: \beta_0 = 0$ gegen $H_1: \beta_0 \neq 0$.

- 2) Analog: Wert von

$$T_{\beta_1} = \frac{\hat{\beta}_1}{\hat{\sigma}_{\beta_1}}$$

zum Testen von $H_0: \beta_1 = 0$ gegen $H_1: \beta_1 \neq 0$.

- 3) p-Wert zu 1)

- 4) p-Wert zu 2)

- 5), 6) hier nicht von Interesse.

Die Testentscheidung „ $\hat{\beta}_1$ signifikant von 0 verschieden“ entspricht dem statistischen Nachweis eines Einflusses von X .

Beispiel: Arbeitszeit und Einkommen

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i$$

mit

$$X_1 = \begin{cases} 1 & \text{männlich} \\ 0 & \text{weiblich} \end{cases}$$

$$X_2 = \text{(vertragliche) Arbeitszeit}$$

$$Y = \text{Einkommen}$$

Multiplres Regressionsmodell: Interpretation

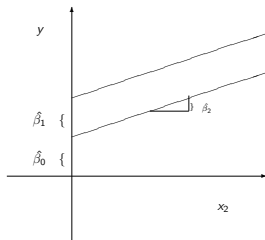
- Geschätzte Regressionsgerade für die Männer:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 \cdot 1 + \hat{\beta}_2 \cdot x_{2i}$$

- Geschätzte Regressionsgerade für die Frauen:

$$\begin{aligned} \hat{y}_i &= \hat{\beta}_0 + \hat{\beta}_1 \cdot 0 + \hat{\beta}_2 \cdot x_{2i} \\ &= \hat{\beta}_0 + \hat{\beta}_2 \cdot x_{2i} \end{aligned}$$

Grundidee (ANCOVA)



- β_0 Grundlevel
- β_2 durchschnittlicher Stundenlohn
- β_1 Zusatzeffekt des Geschlechts zum Grundlevel

Die 0-1 Variable dient als Schalter, mit dem man den Männereffekt an/abschaltet.

Nominales Merkmal mit q Kategorien,
z.B. X = Parteipräferenz mit

$$X = \begin{cases} 1 & \text{CDU/CSU oder FDP} \\ 2 & \text{SPD oder Grüne} \\ 3 & \text{Sonstige} \end{cases}$$

Man darf X nicht einfach mit Werten 1 bis 3 besetzen, da es sich um ein nominales Merkmal handelt.

Dummycodierung II

Idee: Mache aus der einen Variable mit q (hier 3) Ausprägungen $q - 1$ (hier 2) Variablen mit den Ausprägungen ja/nein ($\cong 0/1$). Diese *Dummyvariablen* dürfen dann in der Regression verwendet werden.

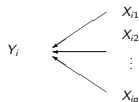
$$X_1 = \begin{cases} 1 & \text{CDU/CSU oder FDP} \\ 0 & \text{andere} \end{cases}$$

$$X_2 = \begin{cases} 1 & \text{SPD, Grüne} \\ 0 & \text{andere} \end{cases}$$

Durch die Ausprägungen von X_1 und X_2 sind alle möglichen Ausprägungen von X vollständig beschrieben:

X	Text	X_1	X_2
1	CDU/CSU, FDP	1	0
2	SPD, Grüne	0	1
3	Sonstige	0	0

Multipl. Regressionsmodell



abhängige Variable

unabhängige Variablen

metrisch/quasistetig

metrische/quasistetige oder
dichotome (0/1) Variablen
(kategoriale Variablen mit mehr Kategorien \rightarrow
Dummy-Kodierung)

- Analoger Modellierungsansatz, aber mit mehreren erklärenden Variablen:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip} + \varepsilon_i$$

- Schätzung von $\beta_0, \beta_1, \dots, \beta_p$ und σ^2 sinnvollerweise über Matrixrechnung bzw. Software.
Aus dem SPSS-Output sind $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p$ sowie $\hat{\sigma}_{\hat{\beta}_0}, \hat{\sigma}_{\hat{\beta}_1}, \dots, \hat{\sigma}_{\hat{\beta}_p}$ ablesbar.

Es gilt für jedes $j = 0, \dots, p$

$$\frac{\hat{\beta}_j - \beta_j}{\hat{\sigma}_{\hat{\beta}_j}} \sim t^{(n-p-1)}$$

und man erhält wieder Konfidenzintervalle für β_j :

$$[\hat{\beta}_j \pm \hat{\sigma}_{\hat{\beta}_j} \cdot t_{\frac{1-\alpha}{2}}^{(n-p-1)}]$$

sowie entsprechende Tests.

Multiple linear Regression: Tests

Von besonderem Interesse ist wieder der Test

$$H_0 : \beta_j = 0, \quad H_1 : \beta_j \neq 0.$$

Der zugehörige p-Wert findet sich im Ausdruck (Vorsicht mit Problematik des multiplen Testens!).

Man kann auch simultan testen, z.B.

$$\beta_1 = \beta_2 = \dots = \beta_p = 0.$$

Dies führt zu einem sogenannten F-Test (\rightarrow Software).

Sind alle X_{ij} 0/1-wertig, so erhält man eine sogenannte *Varianzanalyse*, was dem Vergleich von mehreren Mittelwerten entspricht.

Multiple linear Regression: Interpretation

- Für Befragte mit $X_{ij} = 0$ für alle j gilt:

$$\mathbb{E}(Y) = \beta_0$$

- Ist $X_{i1} = 1$ und $X_{ij} = 0$ für $j \geq 2$, so gilt

$$\mathbb{E}(Y) = \beta_0 + \beta_1$$

- Ist $X_{i1} = 1$ und $X_{i2} = 1$, sowie $X_{ij} = 0$ für $j \geq 3$, so gilt

$$\mathbb{E}(Y) = \beta_0 + \beta_1 + \beta_2$$

- etc.

- Vor allem in der angewandten Literatur, etwa in der Psychologie, wird die Varianzanalyse unabhängig vom Regressionsmodell entwickelt.
- Ziel: Mittelwertvergleiche in mehreren Gruppen, häufig in (quasi-) experimentellen Situationen.
- Verallgemeinerung des t-Tests. Dort nur zwei Gruppen.
- Hier nur *einfaktorielle Varianzanalyse* (Eine Gruppierungsvariable).

Modellformulierung

Modell (Referenzcodierung):

$$Y_{ij} = \mu_j + \beta_j + \varepsilon_{ij} \quad j = 1, \dots, J, \quad i = 1, \dots, n_j,$$

mit

μ_j	Mittelwert der Referenz
β_j	Effekt der Kategorie j im Vergleich zur Referenz J
ε_{ij}	zufällige Störgröße
ε_{ij}	$\sim N(0, \sigma^2)$, $\varepsilon_{11}, \varepsilon_{12}, \dots, \varepsilon_{Jn_j}$ unabhängig.

Testproblem:

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_{j-1} = 0$$

gegen

$$H_1 : \beta_j \neq 0 \quad \text{für mindestens ein } j$$

Einstellung zu Atomkraft anhand eines Scores, nachdem ein Film gezeigt wurde.

3 Gruppen („Faktorstufen“):

- Pro-Atomkraft-Film
- Contra-Atomkraft-Film
- ausgewogener Film

Varianzanalyse: Vergleich der Variabilität in und zwischen den Gruppen

Beobachtungen: Y_{ij}

$$\begin{aligned} j &= 1, \dots, J && \text{Faktorstufen} \\ i &= 1, \dots, n_j && \text{Personenindex in der } j\text{-ten Faktorstufe} \end{aligned}$$

Streuungszerlegung

Mittelwerte:

$$\begin{aligned} \bar{Y}_{\bullet\bullet} & \text{ Gesamtmittelwert in der Stichprobe} \\ \bar{Y}_{\bullet j} & \text{ Mittelwert in der } j\text{-ten Faktorstufe} \end{aligned}$$

Es gilt (vgl. Statistik I) die Streuungszerlegung:

$$\sum_{j=1}^J \sum_{i=1}^{n_j} (Y_{ij} - \bar{Y}_{\bullet\bullet})^2 = \sum_{j=1}^J n_j (\bar{Y}_{\bullet j} - \bar{Y}_{\bullet\bullet})^2 + \underbrace{\sum_{j=1}^J \sum_{i=1}^{n_j} (Y_{ij} - \bar{Y}_{\bullet j})^2}_{= SQE}$$

$$\begin{aligned} \text{Variabilität der Gruppen} &= SQR \\ \text{Variabilität in den Gruppen} & \end{aligned}$$

Die **Testgröße**

$$F = \frac{SQE/(J-1)}{SQR/(n-J)}$$

ist geeignet zum Testen der Hypothesen

$$H_0 : \beta_1 = \beta_2 = \dots \beta_{j-1} = 0$$

gegen

$$H_1 : \beta_j \neq 0 \text{ für mindestens ein } j$$

Beispiel: Wer gewinnt die Bundestagswahl

Literatur: Helmut Norpht und Thomas Gschwend: The chancellor model: Forecasting German elections. International Journal of Forecasting 26 (2010) 42-53.

Prognosemodell zur Vorhersage des Wahlausgangs

- **Kritische Region:** große Werten von F

Also H_0 ablehnen, falls

$$T > F_{1-\alpha}(J-1, n-J),$$

mit dem entsprechenden $(1-\alpha)$ -Quantil der F -Verteilung mit $(J-1)$ und $(n-J)$ Freiheitsgraden.

- (Je größer die Variabilität zwischen den Gruppen im Vergleich zu der Variabilität in den Gruppen, desto unplausibler ist die Nullhypothese, dass alle Gruppenmittelwerte gleich sind.)
- Bei Ablehnung des globalen Tests ist dann oft von Interesse, welche Gruppen sich unterscheiden.
 ⇒ Testen spezifischer Hypothesen über die Effekte β_j . Dabei tritt allerdings die Problematik des multiplen Testens auf.

Quelle: T. Gschwend und H. Nordpoh

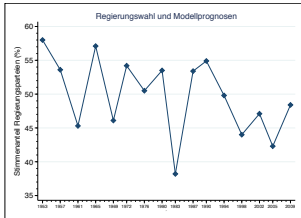
How the Chancellor Model Picks Merkel in 2013

Thomas Gschwend (Mannheim) & Helmut Norpoh (Stony Brook)

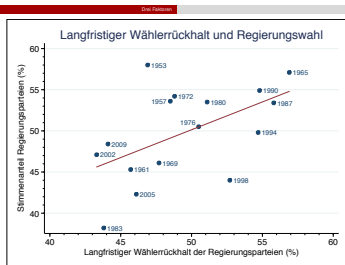
Workshop "Election Forecasting"

Center for Advanced Studies (CAS)
LMU München

Was wollen wir erklären?



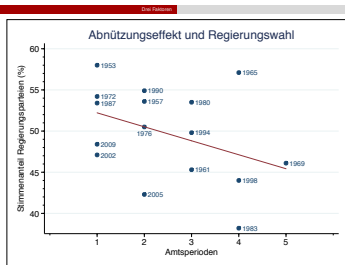
Gschwend & Nordpoth (Stammann & Storz Books) CAG Workshop 'Coalition Forecasting', LMU 15. Juli 2013 2/14



Gschwend & Nordpoth (Stammann & Storz Books) CAG Workshop 'Coalition Forecasting', LMU 15. Juli 2013 3/14



Gschwend & Nordpoth (Stammann & Storz Books) CAG Workshop 'Coalition Forecasting', LMU 15. Juli 2013 4/14



Gschwend & Nordpoth (Stammann & Storz Books) CAG Workshop 'Coalition Forecasting', LMU 15. Juli 2013 5/14

$$y_i = \beta_0 + \beta_1 PAR_i + \beta_2 KAN_i + \beta_3 \log(AMT_i) + \varepsilon_i$$

Vorgehen:

- 1 Schätzung der Parameter mit der KQ Methode
- 2 Bestimmung der aktuellen Einflussgrößen PAR, KAN AMT
- 3 Prognose durch Einsetzen in der Regressionsgleichung
- 4 Unsicherheit der Prognose aus der Schätzung der Varianz von ε_i und der Unsicherheit der Schätzung ergibt Prognoseintervall

Das Modell

Table 1

Statistical Estimates of Vote Predictors

VOTE PREDICTORS	COEFFICIENT (S.E.)
Chancellor Support	0.40*** (.03)
Long-term Partisanship	0.74*** (.07)
Term (logged)	-2.8*** (.54)
Constant	-7.9* (3.6)
R ²	0.96
Root Mean Squared Error	1.1
(N)	(16)
Durbin-Watson d	1.73
Ljung-Box Q (4 lags)	2.90

Note: Model estimation based on elections 1953–2009.

* $p < .05$; ** $p < .01$; *** $p < .001$

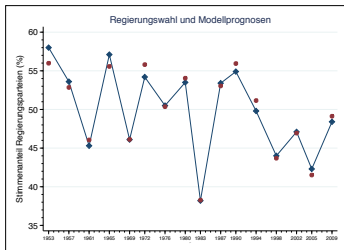
Gschwend & Nordpoh (Stammann & Storz Book)

CAG Workshop "Election Forecasting", LMU

15. Juli 2013 6/14

Quelle: T. Gschwend und H. Nordpoh

Robustheit und Performance



Gschwend & Nordpoh (Stammann & Storz Book)

CAG Workshop "Election Forecasting", LMU

15. Juli 2013 7/14

Quelle: T. Gschwend und H. Nordpoh

Robustheit und Performance

	25. August 2002	23. August 2005	24. August 2009
Modellvorhersage	47,1	42,0	52,9
Institute			
Allensbach	45,0	36,6	51,0
Emnid	46,0	38,0	50,0
Forsa	45,0 - 47,0	36,0	50,0
FGW	47,0	38,0	51,0
Dimap	46,5	38,0	51,0
18 Uhr-Prognose	22. September 2002	18. September 2005	27. September 2009
ARD	46,5	42,5	48,5
ZDF	47,0	41,0	48,0
RTL/n-tv	47,5	42,0	48,0
Endergebnis	47,1	42,3	45,4

Gschwend & Nordpoh (Stammann & Storz Book)

CAG Workshop "Election Forecasting", LMU

15. Juli 2013 8/14

Vielmann 02.2012

Wie kommen wir zu einer Prognose?

Stimmenanteil für Schwarz/Gelb 2013 =

$$-7,9 + 0,74 \cdot (PAR) + 0,40 \cdot (KAN) - 2,8 \cdot (AMT)$$

- 1 Langfristiger Wählerückhalt von CDU/FDP ist 46,4% (= PAR).
- 2 Abnutzungseffekt bei 2. Amtsperiode von CDU/FDP ist $\log(2)$ (= AMT).
- 3 Kanzlerunterstützung (= KAN) heute noch nicht bekannt!

Aber Juli I Politbarometer: Merkel (62%):
Steinbrück (29%). Daher KAN = 68%



Vielmann 02.2012

Wie kommen wir zu einer Prognose?

Stimmenanteil für Schwarz/Gelb 2013 =

$$-7,9 + 0,74 \cdot (PAR) + 0,40 \cdot (KAN) - 2,8 \cdot (AMT) = \underline{51,7\%}$$

- 1 Langfristiger Wählerückhalt von CDU/FDP ist 46,4% (= PAR).
- 2 Abnutzungseffekt bei 2. Amtsperiode von CDU/FDP ist $\log(2)$ (= AMT).
- 3 Kanzlerunterstützung (= KAN) heute noch nicht bekannt!

Aber Juli I Politbarometer: Merkel (62%):
Steinbrück (29%). Daher KAN = 68%



Gschwend & Nordpoh (Mannheim & Stong Brook)

CAG Workshop "Election Forecasting", LMU

16. Juli 2013 12:14

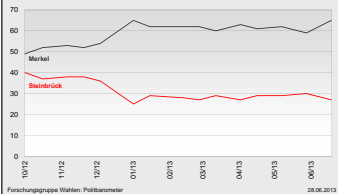
Gschwend & Nordpoh (Mannheim & Stong Brook)

CAG Workshop "Election Forecasting", LMU

16. Juli 2013 12:14

Vielmann 02.2012

Gewünschte/r Bundeskanzler/in

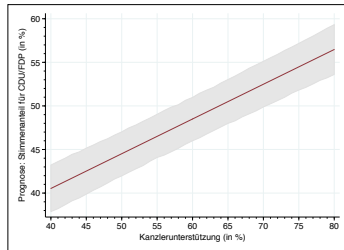
FORSCHUNGSGRUPPE
WAHLEN e.V.

Gschwend & Nordpoh (Mannheim & Stong Brook)

CAG Workshop "Election Forecasting", LMU

16. Juli 2013 12:14

Bedingte Vielmann 02.2012



Gschwend & Nordpoh (Mannheim & Stong Brook)

CAG Workshop "Election Forecasting", LMU

16. Juli 2013 12:14

- Regressionsmodelle wichtiges Instrument zur Prognose
- Sorgfältige Wahl des Modells und der Einflussgrößen Voraussetzung
- Zentrale Annahme: Die Wirkungsstruktur gilt zu dem Zeitpunkt der Prognose weiterhin.
- Kausalität muss inhaltlich diskutiert werden
- Vorsicht vor Überanpassung des Modells an den Lerndaten
- Aber: Es gibt viele erfolgreiche Prognosestrategien ohne inhaltliche Interpretation