Vorlesung: Statistik II für Wirtschaftswissenschaft

Prof. Dr. Helmut Küchenhoff

Institut für Statistik, LMU München

Sommersemester 2017

Termine und Informationen

Homepage:
`http://www.stablab.stat.uni-muenchen.de/lehre/veranstaltungen/statistik2wiwi/index.html`

Vorlesung:

Prof. Helmut Küchenhoff

Di 16:00 - 18:00	Audi max

Übung (wöchentlich):

Ansprechperson: Andre Klima, Matthias Aßenmacher

Übung 1:	Mi. 12.15 - 13.45 Uhr	Geschwister-Scholl-Platz 1 D209
Übung 2:	Mi. 12.15 - 13.45 Uhr	Geschwister-Scholl-Platz 1 E004
Übung 3:	Mi. 14.15 - 15.45 Uhr	Geschwister-Scholl-Platz 1 D209
Übung 4:	Mi. 14.15 - 15.45 Uhr	Geschwister-Scholl-Platz 1 E004
Übung 5:	Do. 18.00 - 19.30 Uhr	Geschwister-Scholl-Platz 1 D209
Übung 6:	Fr. 10.15 - 11.45 Uhr	Geschwister-Scholl-Platz 1 M114

Literatur

L.Fahrmeir, Ch. Heumann, R.Künstler, I.Pigeot, G.Tutz:
Statistik - Der Weg zur Datenanalyse Springer-Verlag, 8. Auflage, 2016

Dank

an Christian Heumann für Materialien und Folien

Was ist Statistik?

March of science 22.4.2017

- Let's make facts great again
- Grab them by data
- We need evidence based policy

Definition Statistik

Statistik als Wissenschaft bezeichnet eine Methodenlehre, die sich mit der Erhebung, der Darstellung, der Analyse und der Bewertung von Daten auseinander setzt. Ein zentraler Aspekt ist dabei die Modellbildung mit zufälligen Komponenten.

Teilgebiete:

- Deskriptive Statistik: beschreibend
- Explorative Datenanalyse: Suche nach Strukturen
- Induktive Statistik: Schlüsse von Daten auf Grundgesamtheit oder allgemeine Phänomene

Beispiel 1: Präsidentschaftswahl in Frankreich

Prognose 20:00 Franz. TV

Macron	Le Pen	Fillon	Melenchon	Hamon
23%	22%	19%	19 %	6.8%

Ergebnis:

Macron	Le Pen	Fillon	Melenchon	Hamon
24%	21.3%	20%	19.6 %	6.4%

Schluss von Stichprobe auf Grundgesamtheit

Induktive Statistik

- Schluss von Daten auf allgemeine Phänomene
- Zentrales Mittel für Erkenntnisse
- Umgang mit Unsicherheit
- Rationale Grundlage von Entscheidungen
- Unterschiedliche Ansätze

Programm

1. Wahrscheinlichkeitsbegriff
2. Rechnen mit Wahrscheinlichkeiten
3. Zufallsgrößen
4. Spezielle Verteilungsmodelle
5. Grenzwertsätze
6. Schätzen
7. Statistische Tests
8. Inferenz bei Regression
9. Bayes–Inferenz

Wahrscheinlichkeitsrechnung für Wirtschaftswissenschaft?

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 Wirtschaftswissenschaften gilt typischerweise nur: wenn A dann folgt eher B als C.
2. Wahrscheinlichkeiten und Umgang mit Unsicherheit spielen in der Wirtschaft eine wichtige Rolle. Bei naiver Herangehensweise (ohne Wahrscheinlichkeitsrechnung) kann man sich leicht täuschen. Riskoberwertung ist ein zentraler Aspekt bei unternehmerischem Handeln
3. Stichprobenverfahren und statistische Modelle spielen in den (empirisch orientierten) Wirtschaftswissenschaften eine zentrale Rolle. Für das Verständnis sind Grundlagenkenntnisse in Wahrscheinlichkeitsrechnung zentral

Klassische Aspekte und Meilensteine

Wahrscheinlichkeit

- 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)

Historische Wurzeln

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.

Laplace'scher Wahrscheinlichkeitsbegriff

Laplace (1749 - 1827)

- Aufbauend auf Symmetrieüberlegungen
- Wahrscheinlichkeit eines Ereignisses A:

$$P(A) := \frac{\text{Anzahl der für } A \text{ günstigen Fälle}}{\text{Anzahl der (gleich) möglichen Fälle}}$$

Beispiel

Wurf eines fairen Würfels

- Wahrscheinlichkeit des Ereignisses A: Es wird eine gerade Zahl gewürfelt

  möglich: $\{1, 2, 3, 4, 5, 6\}$

  günstig: $\{2, 4, 6\}$

$$\Longrightarrow P(A) = \frac{3}{6} = \frac{1}{2}$$

- Erfolgreiche Anwendung v.a. auf Glücksspiele, in der Physik (stochastische Mechanik) und in der Stichprobentheorie bei einer **einfachen Zufallsauswahl**
- Intuitiv einleuchtend, aber beschränkte Anwendbarkeit

Probleme

Warum reichen Laplace-Wahrscheinlichkeiten nicht?

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

Beispiel: Wie wird das Wetter morgen? 3 Möglichkeiten:

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

Wahrscheinlichkeitsbegriffe

Objektivistisch / frequentistische Richtungen / aleatorische Wahrscheinlichkeiten

- 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“

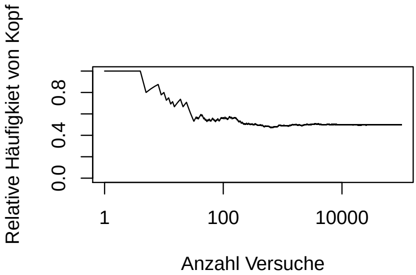
Für ein Ereignis A:

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

n_A : Anzahl der Erfolge
n : Anzahl der Versuche

Experimente

- Buffon (1707-1788) : 4040 Würfe , 2048 "Kopf"
- Karl Pearson (1857-1936) 24000 Würfe, 12012 "Kopf"
- Computersimulation 100.000 Würfe 49972 "Kopf"



Frequentistischer Wahrscheinlichkeitsbegriff II

Probleme bei der Definition

- Einmalige Ereignisse
- Grenzwertdefinition
- Experimentdurchführung

Subjektivistische Richtungen I

- 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

  $\Longrightarrow$ verschiedene Subjekte können durchaus zu unterschiedlichen Bewertungen kommen.

Subjektivistische Richtungen II

- 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:

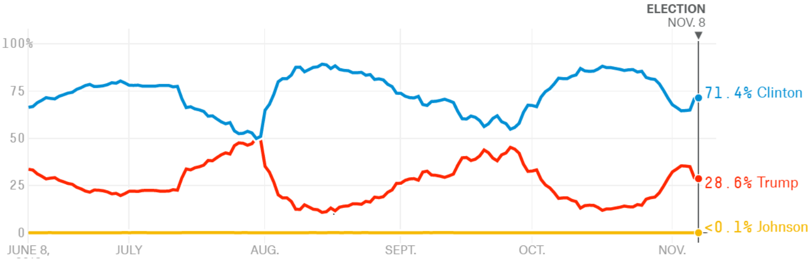
Münzwurf:	Einsatz auf „Zahl“ bis zu 0.5 € sinnvoll
Würfel:	Einsatz auf „5 oder 6“ bis zu 1/3 € sinnvoll

Subjektiver Wahrscheinlichkeitsbegriff II

Probleme

- subjektiv = unwissenschaftlich ?
- Wettdefinition
- Informationsstand

Beispiel: US Wahl



https://projects.fivethirtyeight.com/2016-election-forecast/

Beispiele

- Wahl in Frankreich http://www.economist.com/blogs/graphicdetail/2017/04/france-s-presidential-election
- Wettmärkte http://www.paddypower.com/bet/politics
- Wahlistik
- Fussballwetten https://www.oddset.de/de

Überprüfung von Aussagen über Wahrscheinlichkeiten

- Nicht durch Einzelfälle
- Relative Häufigkeiten
- (Imagiäre) Wetten

Zur Kommunikation von Wahrscheinlichkeiten

Literatur:
D. Kahnemann, P. Slovic, A. Tversky: Judgement under uncertainty: Heuristics and biases Cambridge press 1982.

Darstellung durch natürliche Häufigkeiten (nach Gigerenzer)

- Superrepräsentative Stichprobe vorstellen
- 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.

Beispiel: Beipackzettel

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).

Beispiel: Lotto

6 aus 49

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

$$\frac{1}{\binom{49}{6}} = \frac{1}{13983816} = 0.000000072$$

- „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

Risiko

- Häufig als Wahrscheinlichkeit verwendet
- Manchmal auch als Paar von Wahrscheinlichkeit und Höhe eines Verlustes
- Produkt aus Wahrscheinlichkeit und Schaden
- Entscheidungstheorie unterschiedet verschiedenes Risikoverhalten

Value at risk

- Risikomaß für Wertpapiere
- Der Verlust, der mit einer Wahrscheinlichkeit von $1 - \alpha$ innerhalb eines bestimmten Zeitraums nicht überschritten wird.
- Für verschiedene Portfolios einsetzbar
- Anwendungen auch für Firmen
- Aufsichtsbehörden

Risikodarstellung in der Medizin

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)

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"

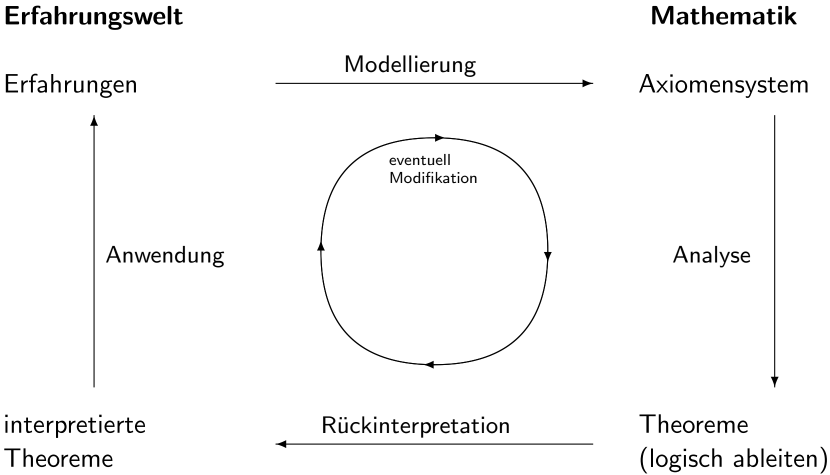
- Reduktion der Todesfälle von 41 auf 32 pro 1000 Patienten mit hohem Chorlesterin ($32 = 41 \cdot (1\text{-}0.22) = 41 \cdot 0.78$)
  Wahrscheinlichkeit für Todesfall: Reduktion von 4.1% auf 3.2%
  Absolute Risikodifferenz: 0.9%
- Reduktion um 22% (relatives Risiko 0.78) „22% werden gerettet"
- Es müssen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

Mathematisch-formaler Wahrscheinlichkeitsbegriff

Axiome

- Axiomatik nach Kolmogoroff
- typische Anwendung der axiomatischen Methode:
  Axiom: Nicht bezweifelte Grundannahme für Kalkül
- Die 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.

Die axiomatische Methode



Ausblick

- In der Tat gibt es auch Kritik an dieser Axiomatik: zu streng und überpräzise $\longrightarrow$ 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

Zufallsvorgänge

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 Zufallsvorgangs sind möglich? (Was kann alles passieren?)

b) Mit welcher Wahrscheinlichkeit treten die einzelnen Ergebnisse ein?

Formale Beschreibung I

Ergebnisraum

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

Beispiele:

- $\Omega = \{1, \dots, 6\}$ beschreibt die möglichen Ergebnisse eines Würfelexperiments

  Ein mögliches Ergebnis: $\omega = 4$; $\omega = 17$ ist kein mögliches Ergebnis.

- $\Omega = \mathbb{R}_0^+$ beschreibt die möglichen Erwerbseinkommen

  Ein mögliches Ergebnis: $\omega = 17513$ €

- Ziehung einer Person: $\Omega = \{1, \dots, N\}$

  Ein mögliches Ergebnis: $\omega = 17$

Formale Beschreibung II

Ereignisse

Ereignisse sind **Teilmengen** von Ω

Beispiele:

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

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

Ereignisoperationen

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

Beispiele:

Ω	$= \{1,2,3,4,5,6\}$	
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“

Wahrscheinlichkeit (formale Definition)

Wahrscheinlichkeit

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

$$\begin{aligned} P : \mathcal{P}(\Omega) &\rightarrow \mathbb{R} \\ A &\mapsto P(A) \end{aligned}$$

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 $P(\{3\})$.

Axiome von Kolmogoroff

Die drei Axiome

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 Kolmogoroff (1933)

- Die Axiome von Kolmogoroff 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.
- Die Axiome passen zu den beiden bisher diskutierten Wahrscheinlichkeitsbegriffen

Zusammenfassung

- Wahrscheinlichkeitsbegriffe wichtig für Evidenz bei Unsicherheit und Entscheidungen
- Kommunikation schwierig
- Subjektive Wahrscheinlichkeiten
- Frequentistischer Begriff
- Berechnung von Wahrscheinlichkeiten wichtige Aufgabe

Axiome nach Kolmogoroff

Gegeben sei ein Zufallsexperiment mit Ergebnisraum Ω (Menge der möglichen Ergebnisse)

Axiom 1

Jedem Ereignis A , $A \subset \Omega$ ist eine Wahrscheinlichkeit $P(A)$ zugeordnet, die Werte zwischen 0 und 1 annehmen kann:

$$0 \leq P(A) \leq 1.$$

Axiom 2

Das sichere Ereignis hat die Wahrscheinlichkeit 1:

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

Axiom 3

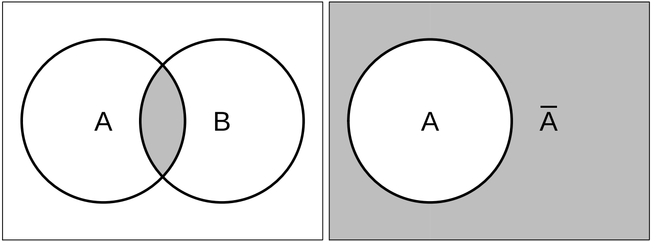
Sind A_1 und A_2 disjunkte Ereignisse, so ist

$$P(A_1 \cup A_2) = P(A_1) + P(A_2).$$

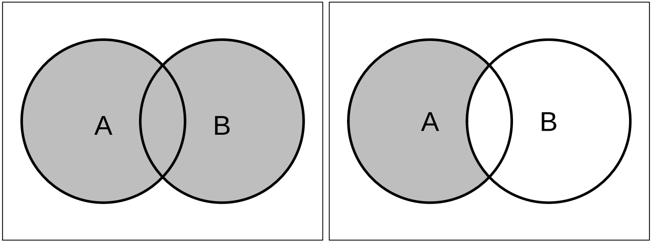
Venn Diagramme

Veranschaulichung von Wahrscheinlichkeiten durch Flächen :

Schnittmenge und Komplement:



Vereinigung und Differenz



Folgerungen

Folgerung 1

Die Wahrscheinlichkeit für das zu A komplementäre Ereignis $\bar{A}$ ist

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

Beweis

$$\begin{array}{rr} \text{Axiom 2:} & P(\Omega) = 1 \\ \Leftrightarrow & P(A \cup \bar{A}) = 1 \\ \overset{\text{Axiom 3}}{\Leftrightarrow} & P(A) + P(\bar{A}) = 1 \\ \Leftrightarrow & P(\bar{A}) = 1 - P(A) \end{array}$$

Folgerungen

Folgerung 2

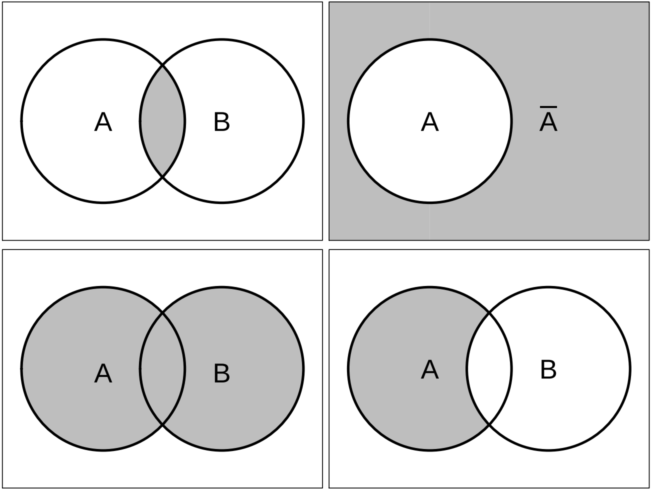
Die Wahrscheinlichkeit des unmöglichen Ereignisses $\emptyset$ ist

$$P(\emptyset) = 0$$

Beweis

$$P(\emptyset) = P(\bar{\Omega}) \overset{\text{Folgerung 1}}{=} 1 - P(\Omega) \overset{\text{Axiom 2}}{=} 0$$

Venn Diagramme



Folgerungen

Folgerung 3

Die Wahrscheinlichkeit, dass von zwei Ereignissen A_1 und A_2, die sich nicht notwendig gegenseitig ausschließen, mindestens eins eintritt, ist

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$$

Beweis

$$\begin{aligned}
P(A_1 \cup A_2) &\overset{\substack{\text{disjunkte}\\\text{Zerlegung}}}{=} P\left(A_1 \backslash A_2 \cup A_2 \backslash A_1 \cup (A_1 \cap A_2)\right) \\
&\overset{\text{Axiom 3}}{=} P(A_1 \backslash A_2) + P(A_2 \backslash A_1) + P(A_1 \cap A_2) \\
&\overset{\text{kreative 0}}{=} \underbrace{P(A_1 \backslash A_2) + P(A_1 \cap A_2)}_{P(A_1)} \\
&\qquad + \underbrace{P(A_2 \backslash A_1) + P(A_1 \cap A_2)}_{P(A_2)} \\
&\qquad - P(A_1 \cap A_2) \\
&= P(A_1) + P(A_2) - P(A_1 \cap A_2)
\end{aligned}$$

Folgerungen

Folgerung 4

Für $A \subseteq B$ gilt stets

$$P(A) \leq P(B)$$

Beweis

$$\begin{array}{cccc}
 & B & \overset{\text{disjunkte Zerlegung}}{=} & A \cup (\bar{A} \cap B) \\
\Rightarrow & P(B) & = & P(A \cup (\bar{A} \cap B)) \\
\overset{\text{Axiom 3}}{\Leftrightarrow} & P(B) & = & P(A) + \underbrace{P(\bar{A} \cap B)}_{\geq 0 \text{ (Axiom 1)}} \\
\Rightarrow & P(B) & \geq & P(A)
\end{array}$$

Folgerungen

Folgerung 5

Sei $A_1, ..., A_n$ eine vollständige Zerlegung des Ereignisraums Ω in paarweise disjunkte Ereignisse. Für ein beliebiges Ereignis B gilt dann

$$P(B) = \sum_{i=1}^{n} P(B \cap A_i)$$

Zusammenfassung

Rechenregeln für Wahrscheinlichkeiten

- $0 \leq P(A) \leq 1$
- $P(\Omega) = 1$
- $P(\emptyset) = 0$
- $P(\bar{A}) = 1 - P(A)$
- $P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$
- $P(A_1 \cup A_2) = P(A_1) + P(A_2)$, falls A_1 und A_2 disjunkt sind
- $P(B) = \sum_{i=1}^{n} P(B \cap A_i)$, falls A_i eine vollständige Zerlegung von Ω bilden

Laplacesche Wahrscheinlichkeit

Definition Laplacesche Wahrscheinlichkeit

Liegt ein Zufallsexperiment zugrunde, bei dem

- die Ergebnismenge *endlich* ist und
- alle Ergebnisse *gleichwahrscheinlich* sind,

dann bildet der Quotient aus

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

die Laplace-Wahrscheinlichkeit.

Die Mächtigkeiten $|A|$ und $|\Omega|$ können z.B. mit Hilfe von kombinatorischen Regeln bestimmt werden.

Ziehen aus einer Grundgesamtheit

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 $n = 1, ..., N$

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

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?

Lösung nach Laplace

- Wahrscheinlichkeit für „SPD-Wähler“

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

  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.

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

- 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.

Zufallsstichproben

- Ziehung von mehreren n Einheiten aus der Grundgesamtheit
- Ziehung mit und ohne Zurücklegen
- 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:

- Mit Hilfe einer nummerierten Liste der Grundgesamtheit Hilfe von Computerprogrammen
- Ersatzmechanismen : Random dialing (Telefon), Random Walks etc.
- **Nicht aufs gerate Wohl**. (Ich spreche Leute an)

Ziehen mit Zurücklegen

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

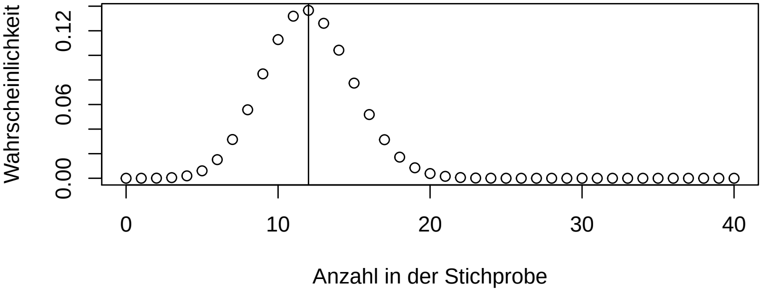
Beispiel: Stichprobentheorie

Ziehe Stichprobe vom Umfang n aus Grundgesamtheit von N=1000 mit Zurücklegen. Annahme: In Grundgesamtheit sind 300 SPD Wähler

$$
\begin{aligned}
n = 1 \qquad & P(1SPD) = 0.3 \\
n = 2 \qquad & P(0SPD) = \tfrac{700 \cdot 700}{1000 \cdot 1000} = 0.49 \\
& P(1SPD) = \tfrac{300 \cdot 700}{1000 \cdot 1000} \cdot 2 = 0.42 \\
& P(2SPD) = \tfrac{300 \cdot 300}{1000 \cdot 1000} = 0.09
\end{aligned}
$$

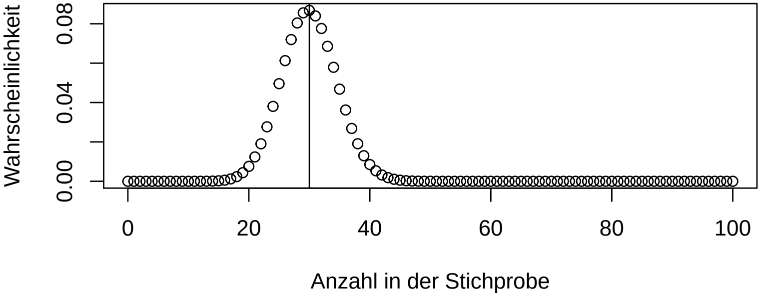
Beispiel: n=40

Ziehe Stichprobe vom Umfang n aus Grundgesamtheit von N=1000 mit Zurücklegen. Annahme: In Grundgesamtheit sind 300 SPD Wähler Berechnung für große n mit Hilfe der Binomialverteilung



Beispiel: n=100

Ziehe Stichprobe vom Umfang n aus Grundgesamtheit von N=1000 mit Zurücklegen. Annahme: In Grundgesamtheit sind 300 SPD Wähler
Berechnung für große n mit Hilfe der Binomialverteilung



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“.
- $\Omega = \{\{\omega_1, \ldots, \omega_n\} : \omega_j \in \{1, \ldots, N\}, \omega_j \neq \omega_i \text{ für } j \neq i\}$
- Anzahl der Stichproben:

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

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

Definition der bedingten Wahrscheinlichkeit

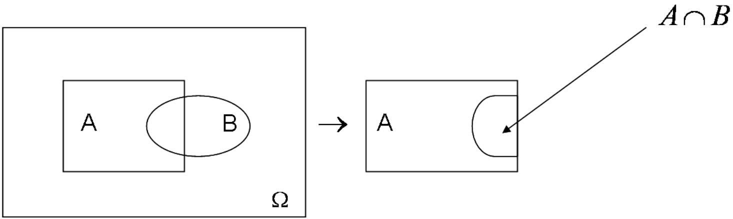
In dem Beispiel betrachten wir das Risiko gegeben „schwerer Fall“. Das Risiko wird berechnet durch

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

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

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

Einschränkung des Ergebnisraumes und bedingte Wahrscheinlichkeit



Bedingte Wahrscheinlichkeit: Beispiel

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“
Krankenh U	400	500	0.56
Krankenh K	30	70	0.7

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

Beispiel: Würfeln

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

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

Multiplikationssatz

Satz

Für zwei beliebige Ereignisse A und B gilt:

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

Beweis

Nach Definition gilt:

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

$$\text{und} \quad P(A|B) = \frac{P(A \cap B)}{P(B)} \Leftrightarrow P(A|B) \cdot P(B) = P(A \cap B)$$

zusammen ergibt sich

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

Fußball Beispiel

Wie groß ist die Wahrscheinlichkeit, das Halbfinale zu gewinnen ?
Gesucht: $P(B)$ mit $B =$ „Sieg im Halbfinale"
Siegchancen sind abhängig vom jeweiligen Gegner!
$\Longrightarrow$ bedingte Wahrscheinlichkeiten.

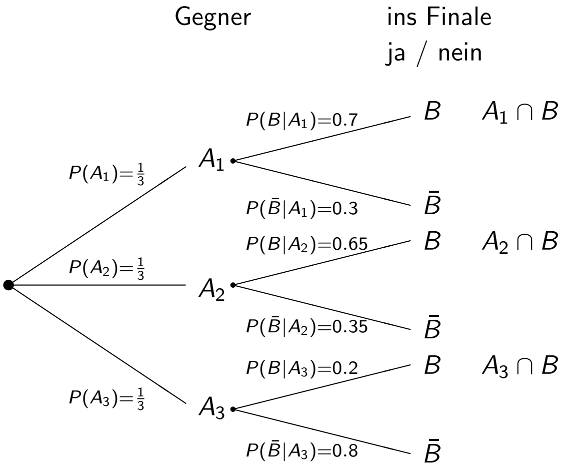
A_1	Gegner ist Mannschaft	1
A_2	"	2
A_3	"	3

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 $\Longrightarrow$ Annahme: $P(A_1) = P(A_2) = P(A_3) = \frac{1}{3}$

Wahrscheinlichkeitsbaum (Fußball Beispiel)



Fußball Beispiel(2)

Welche „Wege“ im Wahrscheinlichkeitsbaum führen zu B?
Nutze Multiplikationssatz

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

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

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

Entlang der Äste multiplizieren, dann summieren

Satz von der totalen Wahrscheinlichkeit

Satz

Bilden die Ereignisse $A_1, ..., A_n$ eine *vollständige* Zerlegung von $\Omega = \bigcup_{i=1}^{n} A_i$ in paarweise disjunkte Ereignisse, so gilt für ein beliebiges Ereignis B:

$$P(B) = \sum_{i=1}^{n} P(B|A_i) \cdot P(A_i).$$

Satz von Bayes

Satz

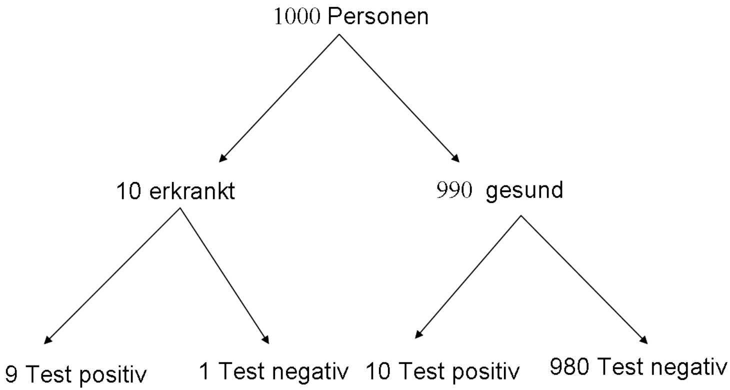
Für beliebige Ereignisse A und B mit $P(A), P(B) > 0$ gilt:

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}.$$

Bilden die A_i eine vollständige Zerlegung von Ω und ist B irgendein Ereignis, so gilt unter Zuhilfenahme des *Satzes von der totalen Wahrscheinlichkeit*:

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

Medizinische Tests



Medizinische Tests 2

K:	Krank
G:	Gesund
TP:	test positiv
TN:	Test negativ

Gegeben:

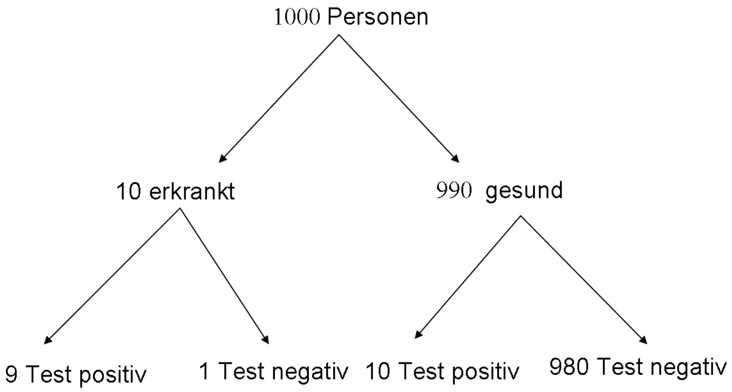
$$\begin{aligned} P(K) &= 10/1000 = 0.01 \\ P(TP|K) &= 9/10 = 0.9 \\ P(TP|G) &= 10/990 = 0.0101 \end{aligned}$$

$$P(K|TP) = ???$$

Lösung mit Satz von Bayes

$$
\begin{aligned}
P(K|TP) &= \frac{P(K \cap TP)}{P(TP)} \\
&= \frac{P(TP|K) \cdot P(K)}{P(TP|K) \cdot P(K) + P(TP|G) \cdot P(G)} \\
&= \frac{0.9 \cdot 0.01}{0.9 \cdot 0.01 + 0.0101 \cdot 0.99} = 0.474
\end{aligned}
$$

Lösung mit Population



Beachte: Die Bedingung entspricht der Bezugspopulation 9 von 19 Patienten mit positivem Test sind tatsächlich krank:

$$P(K|TP) = 9/19 = 0.474$$

Unabhängigkeit

Definition stochastisch unabhängig

Zwei zufällige Ereignisse A und B heißen genau dann voneinander stochastisch unabhängig, wenn

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

gilt, d.h., wenn die Wahrscheinlichkeit für das gleichzeitige Eintreten von A und B gleich dem Produkt der beiden Einzelwahrscheinlichkeiten ist.

Folgerungen

- Sind zwei Ereignisse A und B unabhängig so folgt, dass das Eintreten des Ereignisses B keinen Einfluss auf das Eintreten von A hat, d.h. es gilt:

$$P(A|B) = P(A|\bar{B}) = P(A)$$

- Man kann unter der der Annahme der Unabhängigkeit Wahrscheinlichkeiten berechnen:
  A: Beim ersten Wurf 6
  B: Beim zweiten Wurf 6

$$P(A \cap B) = P(A)P(B) = 1/36$$

Zufallsgrößen

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:

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

(Abbildung des Ergebnisraums auf die reellen Zahlen)

Im Beispiel 1:

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

Würfelwurf mit fairem Würfel

Betrachte ein Spiel mit den Gewinnen:

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

Die Wahrscheinlichkeiten P_X ergeben sich wie folgt:

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

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

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

Wahrscheinlichkeitsverteilung einer diskreten Zufallsgröße

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* (Auszahlungs*regel* 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 Auszahlungs*betrag*).
- 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.

Wahrscheinlichkeitsfunktion

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

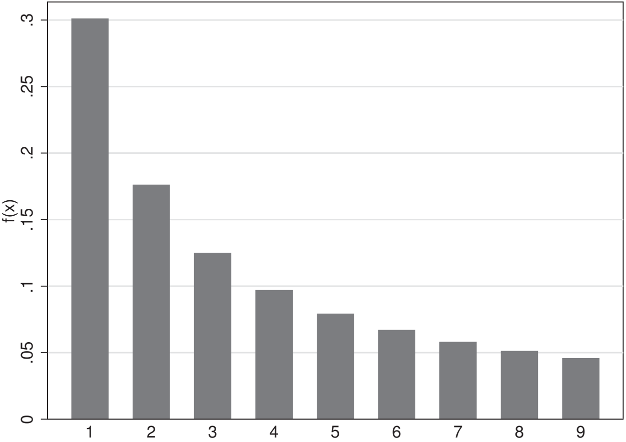
$$f(x) = \begin{cases} P(X = x_i) = p_i, & x = x_i \in \{x_1, x_2, \ldots, x_k, \ldots\} \\ 0, & \text{sonst.} \end{cases}$$

Beispiel: Benfords Gesetz

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

$$X = \text{„Anfangsziffer von Zahlen“}$$

die Wahrscheinlichkeitsfunktion

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

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

Zum Rechnen mit Zufallsvariablen

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

Zufallsvariablen können durch die Wahrscheinlichkeiten $P(X \leq x)$ eindeutig beschrieben werden.

Definition Verteilungsfunktion

Die Verteilungsfunktion einer Zufallsvariablen X ist definiert durch

$$F(x) = P(X \leq x) = P(-\infty < X \leq x).$$

Sie hat folgende Eigenschaften:

- $0 \leq F(x) \leq 1$
- $\lim_{x \to -\infty} F(x) = 0$
- $\lim_{x \to \infty} F(x) = 1$
- schwach monoton wachsend
- rechtsseitig stetig

Verteilungsfunktion

Berechnung der Verteilungsfunktion von diskreten Zufallsvariablen

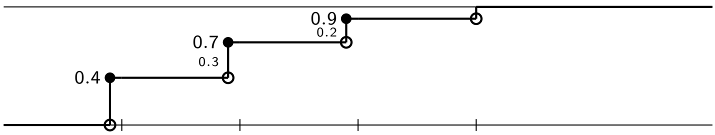
Die Verteilungsfunktion einer diskreten Zufallsvariablen ermittelt sich über die Summe der Wahrscheinlichkeiten p_i, deren zugehörige Träger x_i kleiner-gleich dem abgefragten Wert sind:

$$F(x) = \sum_{x_i \leq x} p_i$$

Die Verteilungsfunktion von diskreten Zufallsvariablen ist damit

- eine *Treppenfunktion*
- mit *Sprungstellen* an den möglichen Werten x_i der jeweiligen ZV,
- die *Sprunghöhen* gleichen den zugehörigen W'keiten p_i.

Beispiel: Haushaltsgröße



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.

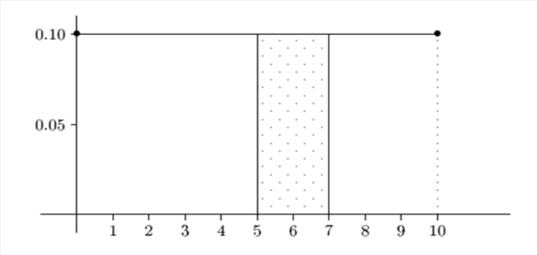
gesucht: $P(T=2) = ?$

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

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

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

Berechnung von Wahrscheinlichkeiten



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

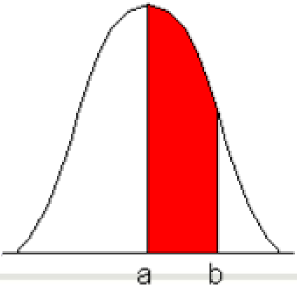
Dichtefunktion

Definition Dichtefunktion

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. f heißt dann Dichtefunktion der Zufallsgröße



Eigenschaften der Dichte

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

Beispiel: Warten auf den Bus

- Verteilungsfunktion

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

- Dichtefunktion

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

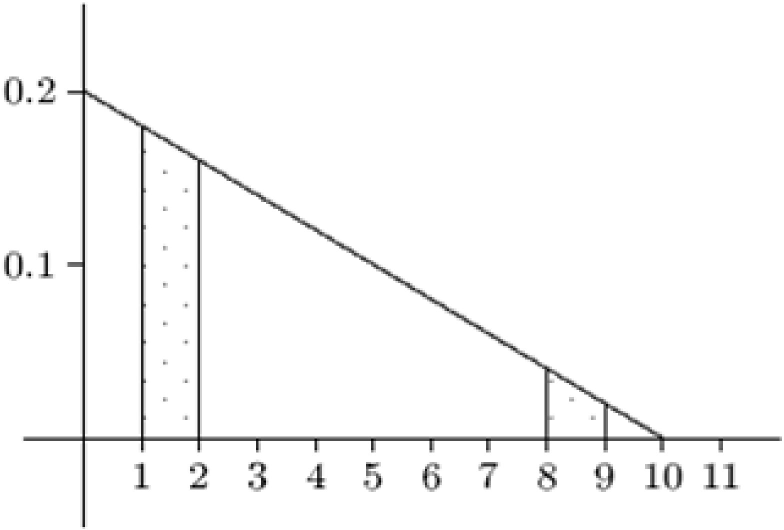
Eigenschaften von stetigen Zufallsgrößen

Für eine stetige Zufallsgröße X mit Verteilungsfunktion F gilt für alle a und b

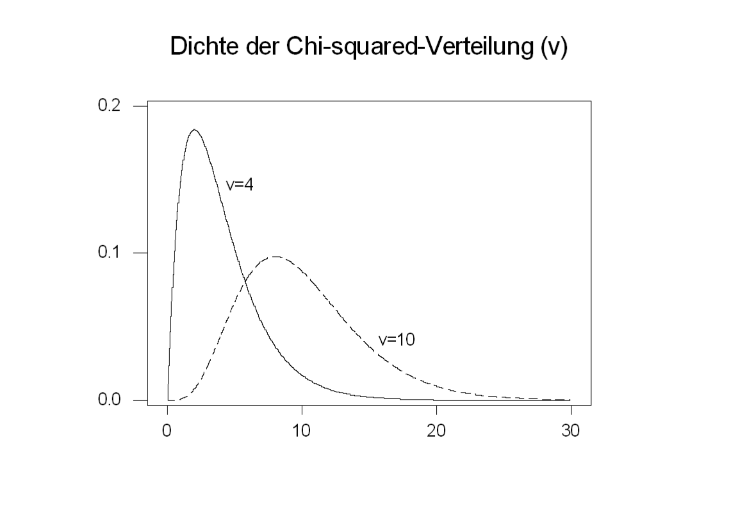
$$
\begin{aligned}
P(X = a) = P(X = b) &= 0 \\
P(a < X \leq b) &= F(b) - F(a) \\
P(a \leq X \leq b) &= F(b) - F(a) \\
P(a < X < b) &= F(b) - F(a)
\end{aligned}
$$

.

Warten auf den Bus (2): Interpretation ?



Warten auf den Bus (3): Interpretation ?



Bemerkungen

- Stetige Zufallsvariablen sind für die Modellbildung sehr wichtig
- Insbesondere ergeben sich Approximationsmöglichkeiten für diskrete Zufallsvariablen durch stetige Zufallsvariablen bei größeren Stichprobenumfängen

Erwartungswert und Varianz

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

Insbesondere:

a) „durchschnittlicher Wert" $\longrightarrow$ Erwartungswert, z.B.

- „mittleres" Einkommen,
- „durchschnittliche" Körpergröße,
- fairer Preis eines Spiels.

b) Streuung (Dispersion), z.B. wie stark schwankt das Einkommen, die Körpergröße etc.

Erwartungswert diskreter Zufallsgrößen

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

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

- Würfelwurf:
  $\mathbb{E}(X) = \frac{1}{6} \cdot 1 + \frac{1}{6} \cdot 2 + \frac{1}{6} \cdot 3 + \frac{1}{6} \cdot 4 + \frac{1}{6} \cdot 5 + \frac{1}{6} \cdot 6 = 3.5$
- Summe zweier Würfel:
  $\mathbb{E}(S) = \frac{1}{36} \cdot 2 + \frac{2}{36} \cdot 3 + \ldots + \frac{2}{36} \cdot 11 + \frac{1}{36} \cdot 12 = 7$
- Antwort ja oder nein:
  $\mathbb{E}(X) = P(X = 0) \cdot 0 + P(X = 1) \cdot 1 = P(X = 1)$
- Wette mit Einsatz E und Gewinn 1 bei Gewinnwahrscheinlichkeit p
  $\mathbb{E}(X) = p \cdot (1 - E) + (1 - p) \cdot (-E) = p - pE - E + pE = p - E$
  Erwarteter Gewinn positiv, falls $E < p$.

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

$$\begin{aligned}\mathbb{E}(T) &= \int_{-\infty}^{\infty} xf(x)dx \\ &= \int_{0}^{10} \frac{1}{10} xdx = 5\end{aligned}$$

Varianz und Standardabweichung von Zufallsgrößen

- Lageparameter: Erwartungswert
- Streuungsparameter: Varianz und Standardabweichung

Wie stark weichen die Ausprägungen im Durchschnitt vom Erwartungswert ab?

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

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

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

Beispiel I zur Varianz

Y: Einmal Würfeln und Multiplikation mit 2

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

Beispiel II zur Varianz II

S: Würfeln mit 2 Würfeln

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

Varianz bei der Wartezeit auf den Bus

$$
\begin{aligned}
Var(T) &= \int_{-\infty}^{\infty} (x-5)^2 f(x) dx \\
&= \int_{0}^{10} (x-5)^2 \frac{1}{10} dx \\
&= \frac{25}{3} \\
\sigma_T &= \sqrt{\frac{25}{3}} = 2.9
\end{aligned}
$$

Bemerkungen

- 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*.

Verschiebungssatz

Es gilt:

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

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.“

Varianz von linearen Transformationen

$Y = a + b \cdot X$

$Var(Y) = b^2 \cdot Var(X)$
$\sigma_Y = |b| \cdot \sigma_Y$

Verschiebungen ändern nichts an Streuung

Beispiel zur Varianz

X: Einmal Würfeln
Y: Einmal Würfeln und Multiplikation mit 2

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

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

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

Varianz von Summen von Zufallsgrößen

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

$$Var(X_1 + X_2) = Var(X_1) + Var(X_2)$$

Beispiele:

- zweimaliges Würfeln
- Ziehen von 2 Personen

Beachte: Unabhängigkeit ist wichtige Voraussetzung

Bemerkungen I

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

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

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

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

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

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

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

Standardisierte Zufallsvariable

Standardisierung

Die Zufallsvariable

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

heißt *standardisierte Zufallsvariable*. Es gilt

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

$$\begin{aligned}
\mathbb{E}(Z) &= \mathbb{E}\left(\frac{X - \mathbb{E}(X)}{\sqrt{\mathrm{Var}(X)}}\right) = \frac{1}{\sqrt{\mathrm{Var}(X)}} \cdot \mathbb{E}(X - \mathbb{E}(X)) \\
&= \frac{1}{\sqrt{\mathrm{Var}(X)}} \cdot (\mathbb{E}(X) - \mathbb{E}(\mathbb{E}(X))) = \frac{1}{\sqrt{\mathrm{Var}(X)}} \cdot (\mathbb{E}(X) - \mathbb{E}(X)) = 0 \\
\mathrm{Var}(Z) &= \mathrm{Var}\left(\frac{X - \mathbb{E}(X)}{\sqrt{\mathrm{Var}(X)}}\right) = \mathrm{Var}\left(\frac{X}{\sqrt{\mathrm{Var}(X)}} - \frac{\mathbb{E}(X)}{\sqrt{\mathrm{Var}(X)}}\right) \\
&= \mathrm{Var}\left(\frac{X}{\sqrt{\mathrm{Var}(X)}}\right) = \left(\frac{1}{\sqrt{\mathrm{Var}(X)}}\right)^2 \cdot \mathrm{Var}(X) = 1
\end{aligned}$$

Beispiel: Chuck-a-Luck

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

Chuck-a-Luck: Erwartungswert

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:

$X_1, X_2, \ldots, X_6$	Gewinn, wenn beim ersten Wurf ein Einsatz auf $1, 2, \ldots, 6$ gesetzt wird.
$Y_1, Y_2, \ldots, Y_6$	Gewinn, wenn beim zweiten Wurf ein Einsatz auf $1, 2, \ldots, 6$ 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 $\mathbb{E}(X_i) = \mathbb{E}(Y_i) = -\frac{17}{216}$ folgt:

$$\begin{aligned}
\mathbb{E}(2X_6) &= 2\,\mathbb{E}(X_6) = -\frac{34}{216} \\
\mathbb{E}(X_1 + X_6) &= \mathbb{E}(X_1) + \mathbb{E}(X_6) = -\frac{34}{216} \\
\mathbb{E}(X_6 + Y_6) &= \mathbb{E}(X_6) + \mathbb{E}(Y_6) = -\frac{34}{216}
\end{aligned}$$

  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

$$\begin{aligned}
\mathrm{Var}(G) &= \left(3+\frac{17}{216}\right)^2\cdot\frac{1}{216}+\left(2+\frac{17}{216}\right)^2\cdot\frac{15}{216}+\left(1+\frac{17}{216}\right)^2\cdot\frac{75}{216} \\
&\quad+\left(-1+\frac{17}{216}\right)^2\cdot\frac{125}{216} \\
&= 0.04388156+0.30007008+0.40402836+0.4911961 = \\
&= 1.2391761 \\
\sqrt{\mathrm{Var}(G)} &= 1.113183
\end{aligned}$$

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

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

und

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

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
$P(X_1 + X_6 = x)$	0.29630	0.44444	0.11111	0.12037	0.02778

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

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.

Bernoulliverteilung

Definition Bernoulliverteilung

Ein Experiment mit nur zwei Ergebnissen ($1 =$ Erfolg, $0 =$ Misserfolg) gehorcht einer Bernoulliverteilung.

Kurzschreibweise: $X \sim B(1, p)$

$$P(X = x) = \begin{cases} p & \text{falls } x = 1 \\ 1 - p & \text{falls } x = 0 \end{cases}$$

Bernoulliverteilung

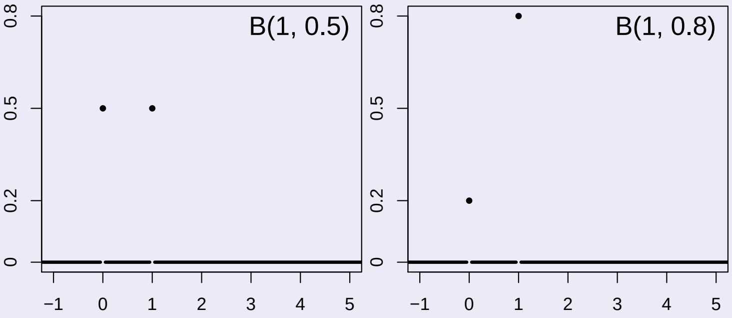
Erwartungswert und Varianz

$$E(X) = p$$

$$Var(X) = p\,(1 - p)$$

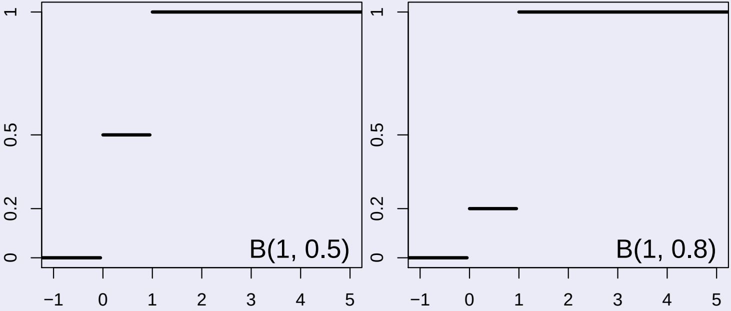
Bernoulliverteilung

graphische Beispiele der Wahrscheinlichkeitsfunktion



Bernoulliverteilung

graphische Beispiele der Verteilungsfunktion



Bernoulliverteilung

Beispiel

Betrachtet wird das Ergebnis eines einmaligen Münzwurfs mit einer unfairen Münze:

$$\text{Ausprägungen: } 1 \text{ (Kopf)}, 0 \text{ (Zahl)}$$

$$P(X = 1) = \frac{2}{3}$$

$$E(X) = \frac{2}{3}$$

$$Var(X) = \frac{2}{3} \cdot (1 - \frac{2}{3}) = \frac{2}{9}$$

Binomialverteilung

Definition Binomialverteilung

Werden n *unabhängige* und *identische* Bernoulliexperimente durchgeführt, so folgt die *Anzahl der Erfolge* einer Binomialverteilung.

Kurzschreibweise: $X \sim B(n, p)$

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, ..., n$$

Binomialverteilung

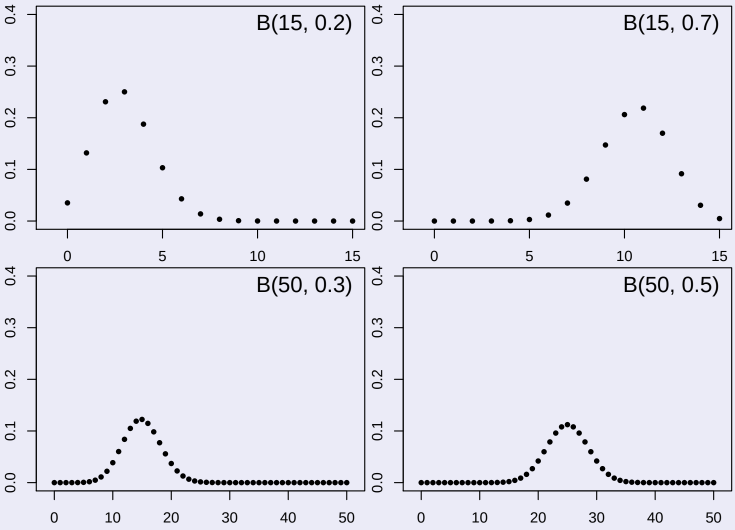
Erwartungswert und Varianz

$$E(X) = np$$

$$Var(X) = np(1-p)$$

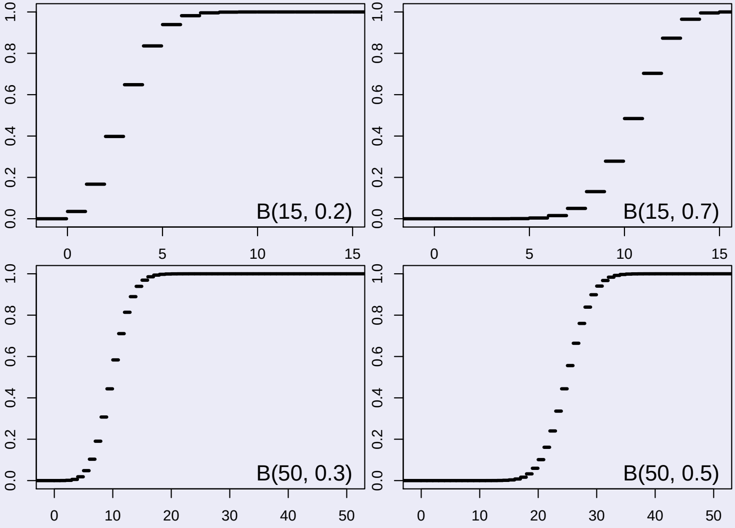
Binomialverteilung

graphische Beispiele der Wahrscheinlichkeitsfunktion



Binomialverteilung

graphische Beispiele der Verteilungsfunktion



Binomialverteilung

Beispiel

Betrachtet wird die Anzahl des Ereignisses „Kopf oben“ beim zehnmaligen Münzwurf mit einer unfairen Münze:

$$
\begin{aligned}
n &= 10 \\
p &= \frac{2}{3} \\
E(X) &= 10 \cdot \frac{2}{3} = 6,67 \\
Var(X) &= 10 \cdot \frac{2}{3} \cdot \left(1 - \frac{2}{3}\right) = 2,22 \\
P(X = 7) &= \binom{10}{7} \cdot \left(\frac{2}{3}\right)^{7} \cdot \left(1 - \frac{2}{3}\right)^{10-7} = 0,26
\end{aligned}
$$

Beispiel: Wahlprognose

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

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

$X \sim B(100, 0.3)$

$$\begin{aligned} P(X \geq 50) &= P(X = 50) + P(X = 51) + \ldots + P(X = 100) \\ &= \binom{100}{50} \cdot 0.3^{50} \cdot 0.7^{50} + \ldots \\ &= 0.00002206 \end{aligned}$$

geometrische Verteilung

Definition geometrische Verteilung

Interessiert man sich für die *Anzahl* der Versuche, bis bei einem Bernoulliexperiment ein Erfolg beobachtet wird, so folgt dieser Versuchsaufbau einer geometrischen Verteilung.

Kurzschreibweise: $X \sim G(p)$

$$P(X = x) = p\,(1-p)^{x-1}, \quad x \in \mathbb{N}$$

geometrische Verteilung

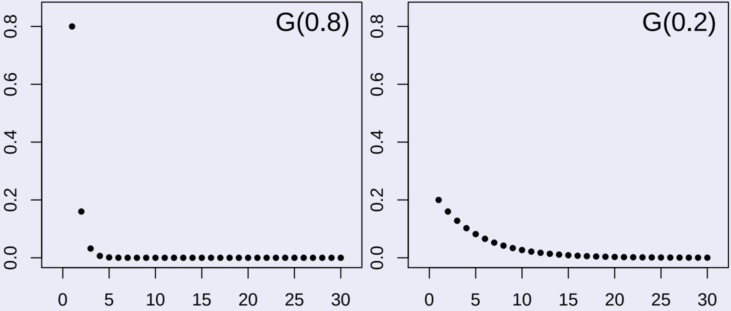
Erwartungswert und Varianz

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

$$Var(X) = \frac{1}{p}\left(\frac{1}{p} - 1\right)$$

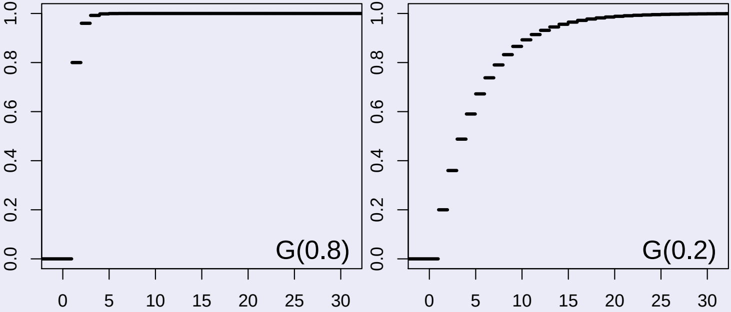
geometrische Verteilung

graphische Beispiele der Wahrscheinlichkeitsfunktion



geometrische Verteilung

graphische Beispiele der Verteilungsfunktion



geometrische Verteilung

Beispiel

Betrachtet wird die Anzahl der Würfe, bis eine 1 gewürfelt wird.
Dies ist geometrisch verteilt mit $p = \frac{1}{6}$, also $X \sim G(\frac{1}{6})$.

$$E(X) = \frac{1}{1/6} = 6$$

$$Var(X) = \frac{1}{1/6}(\frac{1}{1/6} - 1) = 30$$

Im Mittel fällt beim sechsten Wurf eine 1.

Poissonverteilung

Definition Poissonverteilung

Soll die Wahrscheinlichkeit für die *Häufigkeit* bzw. *Anzahl* des Eintretens eines bestimmten Ereignisses innerhalb eines *fest vorgegebenen Intervalls* der Länge t (hier nur $t = 1$) beschrieben werden, so lässt sich dies mit einer Poissonverteilung modellieren.

Kurzschreibweise: $X \sim Po(\lambda), \quad \lambda > 0$

$$P(X = x) = \frac{\lambda^x}{x!} \cdot \exp(-\lambda), \quad x \in \mathbb{N}_0$$

Poissonverteilung

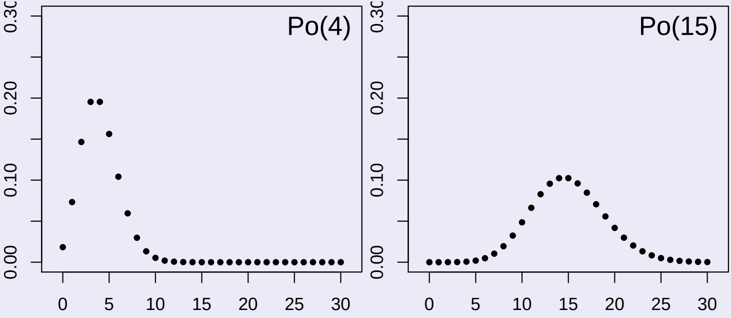
Erwartungswert und Varianz

$$E(X) = \lambda$$

$$Var(X) = \lambda$$

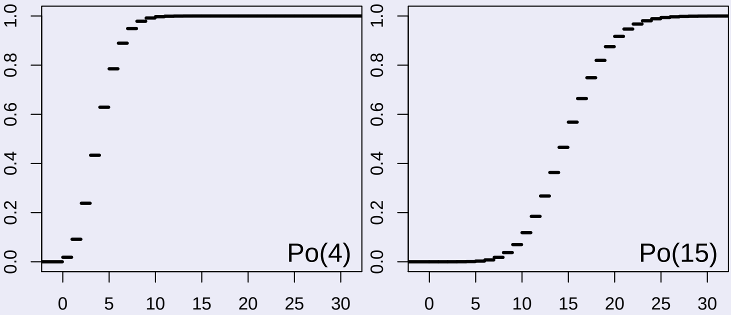
Poissonverteilung

graphische Beispiele der Wahrscheinlichkeitsfunktion



Poissonverteilung

graphische Beispiele der Verteilungsfunktion



Poissonverteilung

Additionssatz

Sind $X \sim Po(a)$ und $Y \sim Po(b)$ unabhängige Zufallsvariablen, dann gilt:

$$X + Y \sim Po(a + b).$$

Poissonverteilung

Beispiel

Bei einer Hotline weiß man aus Erfahrung, dass dort am Freitag zwischen 15 und 16 Uhr 7 ($= \lambda$) Kunden den Dienst in Anspruch nehmen. Die Wahrscheinlichkeit dafür, dass es mal 9 Kunden sind, beträgt:

$$P(X = 9) = \frac{7^9}{9!} \cdot \exp(-7) = 0,1014.$$

Exponentialverteilung

Definition Exponentialverteilung

Wird die stetige Wartezeit bis zum Eintreten eines Ereignisses betrachtet und wird gefordert, dass die weitere Wartezeit unabhängig von der bereits verstrichenen Wartezeit ist, so bietet sich die Exponentialverteilung zur Modellierung dieses Problems an.

Kurzschreibweise: $X \sim Expo(\lambda)$

$$f(x) = \begin{cases} \lambda \cdot \exp(-\lambda x) & \text{für } x \geq 0 \\ 0 & \text{sonst} \end{cases}$$

$$F(x) = \begin{cases} 1 - \exp(-\lambda x) & \text{für } x \geq 0 \\ 0 & \text{sonst} \end{cases}$$

Die Exponentialverteilung ist damit das stetige Analogon zur geometrischen Verteilung.

Exponentialverteilung

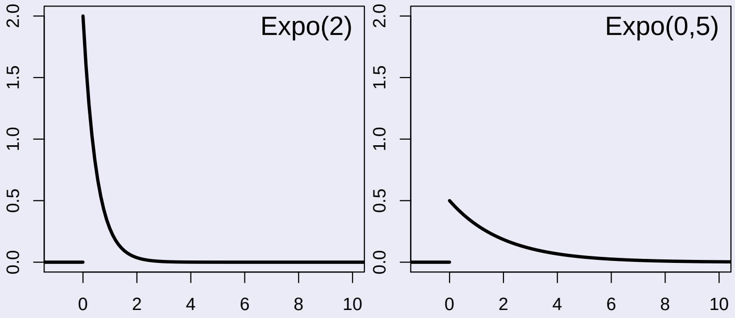
Erwartungswert und Varianz

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

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

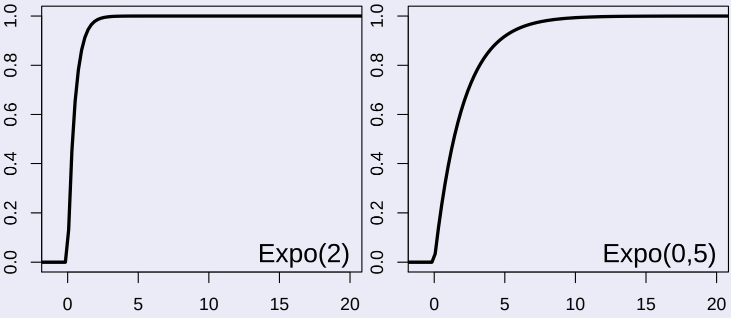
Exponentialverteilung

graphische Beispiele der Dichtefunktion



Exponentialverteilung

graphische Beispiele der Verteilungsfunktion



Exponentialverteilung

Zusammenhang zwischen Exponential- und Poissonverteilung

Die Anzahl der Ereignisse Y innerhalb eines Kontinuums ist poissonverteilt mit Parameter λ genau dann, wenn die Wartezeit zwischen zwei Ereignissen exponentialverteilt mit Parameter λ ist.

Exponentialverteilung

Beispiel

Die Zufallsvariable X: „Lebensdauer einer Glühbirne einer Schaufensterbeleuchtung“ sei exponentialverteilt mit Parameter $\lambda = 10$. Damit gilt:

$$E(X) = \frac{1}{10}; \quad Var(X) = \frac{1}{10^2} = \frac{1}{100}$$

Die Zufallsvariable Y: „Anzahl der ausgefallenen Glühbirnen“ ist damit poissonverteilt mit Parameter $\lambda = 10$ und damit $E(Y) = 10$ sowie $Var(Y) = 10$.

Betrachten wir als Kontinuum ein Jahr, so erhalten wir für die erwartete Anzahl der ausgefallenen Glühbirnen pro Jahr

$$E(Y) = 10 \text{ Glühbirnen pro Jahr}$$

und für die zu erwartende Wartezeit zwischen zwei Ausfällen

$$E(X) = \frac{1}{10} \text{ Jahr} = 36,5 \text{ Tage.}$$

Pareto-Verteilung

Verteilung zur Modellierung von Einkommen. Kurzschreibweise:
$X \sim Pareto(k, \alpha)$ Verteilungsfunktion

$$F(x) = \begin{cases} 1 - \left(\frac{k}{x}\right)^{\alpha} & \text{für } x \geq k \\ 0 & \text{sonst} \end{cases}$$

Dichte :

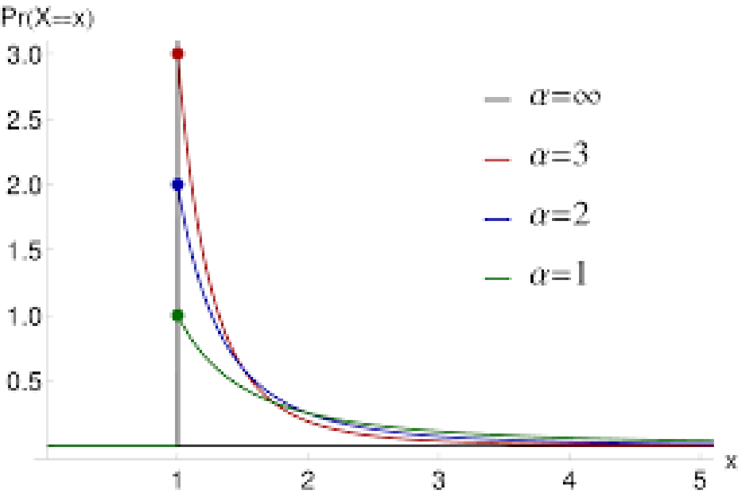
$$f(x) = \begin{cases} \frac{\alpha k^{\alpha}}{x^{\alpha+1}} & \text{für } x \geq k \\ 0 & \text{sonst} \end{cases}$$

Erwartungswert:

$$E(X) = \frac{\alpha}{\alpha - 1} k$$

Pareto-Verteilung

graphische Beispiele der Dichtefunktion



Normalverteilung

Definition Normalverteilung

Die Normalverteilung ist die in der Statistik am häufigsten verwendete stetige Verteilung. Ihre Verteilung liegt (recht) eng und symmetrisch um ihren Erwartungswert.

Kurzschreibweise: $X \sim N(\mu, \sigma^2)$

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

Anwendungen

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

Normalverteilung

Erwartungswert und Varianz

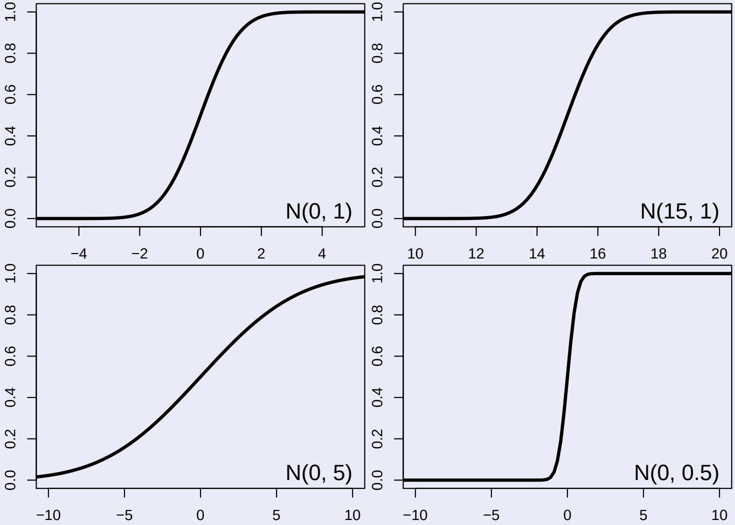
$$E(X) = \mu$$

$$Var(X) = \sigma^2$$

Dies sind zugleich die Parameter der Verteilung.

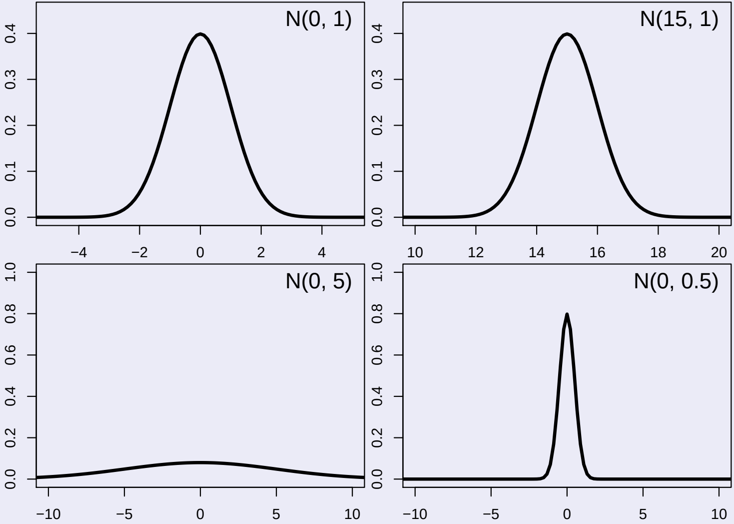
Normalverteilung

graphische Beispiele der Verteilungsfunktion

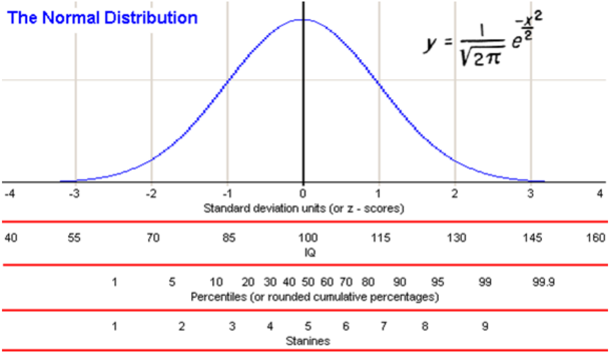


Normalverteilung

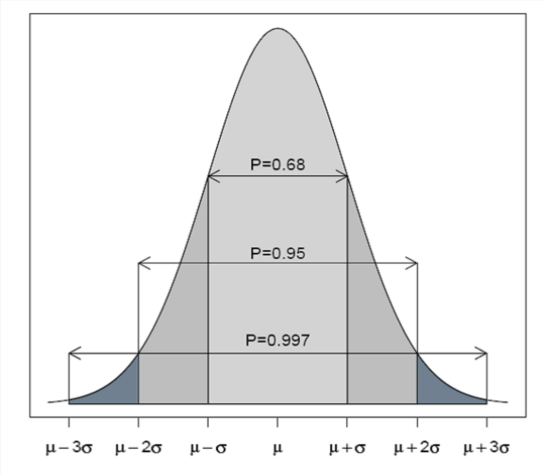
graphische Beispiele der Dichtefunktion



Normalverteilung II



Normalverteilung III



Normalverteilung

Standardisierung

Sei $X \sim N(\mu, \sigma^2)$. Dann ist

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

Normalverteilung

Häufig kommen unabhängig und identisch verteilte Zufallsgrößen vor. Man spricht dann von **iid** (**i**ndependently **i**dentically **d**istributed) Zufallsgrößen.

Additionssatz

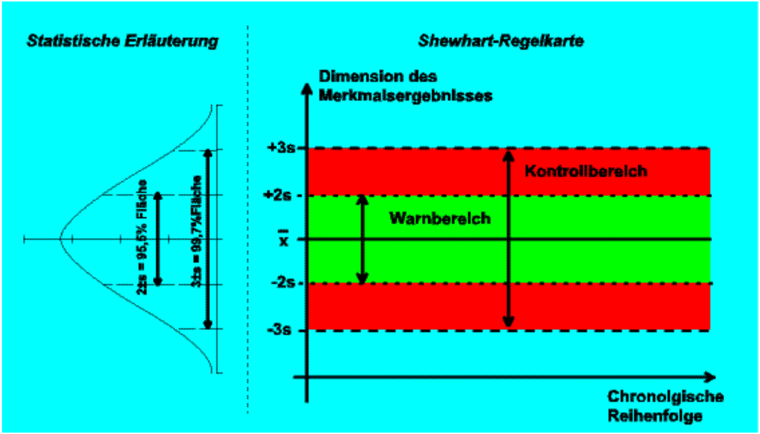
Seien $X_1, \ldots, X_n \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$, dann ist deren Summe normalverteilt:

$$\sum_{i=1}^{n} X_i \sim N\left(n\mu, n\sigma^2\right).$$

Das arithmetische Mittel der Zufallsvariablen $X_1, \ldots, X_n$ ist ebenfalls normalverteilt:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \sim N\left(\mu, \frac{\sigma^2}{n}\right).$$

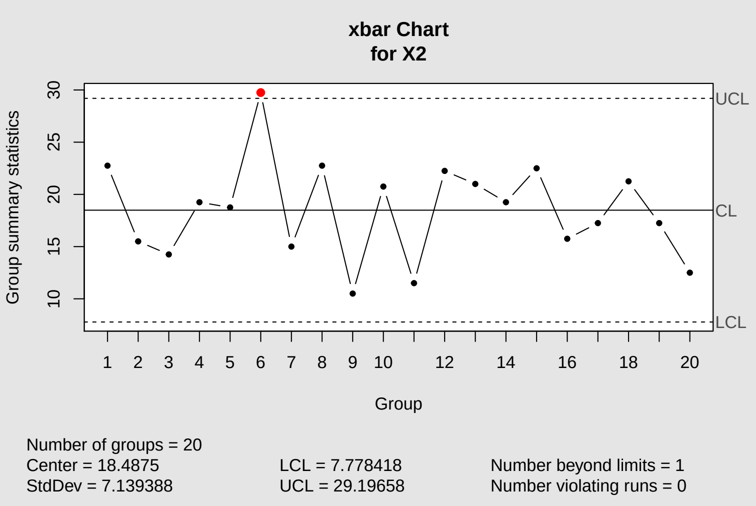
Anwendung aus der Qualitätskontrolle



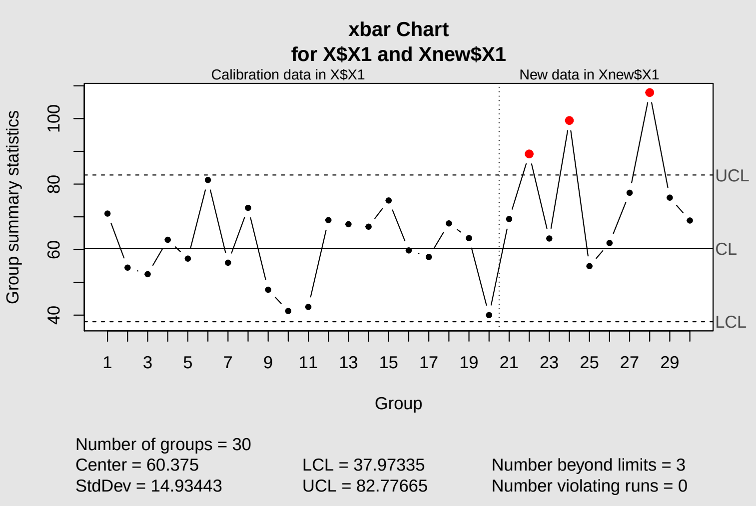
Beispiel mit Proben von 4 Einheiten

Eingriffsgrenzen:

$$\bar{X} + 3 \cdot \frac{\sigma}{\sqrt{4}}$$



Beispiel mit Proben von 4 Einheiten



Normalverteilung

Rechenregeln

- Sei $\Phi(\frac{x-\mu}{\sigma})$ die Verteilungsfunktion der Standardnormalverteilung zu einer beliebigen Normalverteilung $F(x)$.
- Seien a und b beliebige reelle Zahlen, $z_a = \frac{a-\mu}{\sigma}$ und $z_b = \frac{b-\mu}{\sigma}$ deren Standardisierungen und
- sei z ein beliebiges Quantil der Standardnormalverteilung.

Dann gilt:

$$
\begin{aligned}
P(X \leq b) &= F(b) = \Phi(z_b) \\
P(X > b) &= 1 - \Phi(z_b) \\
P(a \leq X \leq b) &= \Phi(z_b) - \Phi(z_a) \\
\Phi(-z) &= 1 - \Phi(z) \\
\Phi(0) &= 0,5 \\
P(-a \leq X \leq a) &= 2\Phi(z_a) - 1
\end{aligned}
$$

Normalverteilung

wichtige Quantile der Standardnormalverteilung

Quantile, die oft beim Testen von Hypothesen verwendet werden:

$\alpha = 0,05$: $z_{1-\alpha} = z_{0,95} = 1,64$

$\alpha = 0,05$: $z_{1-\frac{\alpha}{2}} = z_{0,975} = 1,96$

$\alpha = 0,01$: $z_{1-\alpha} = z_{0,99} = 2,33$

$\alpha = 0,01$: $z_{1-\frac{\alpha}{2}} = z_{0,995} = 2,58$

Quantilbestimmung

Ein beliebiges Quantil x_p einer nichtstandardisierten Normalverteilung kann durch folgende Rechnung bestimmt werden:

$$x_p = \mu + \sigma \cdot z_p$$

t-Verteilung

Definition t-Verteilung

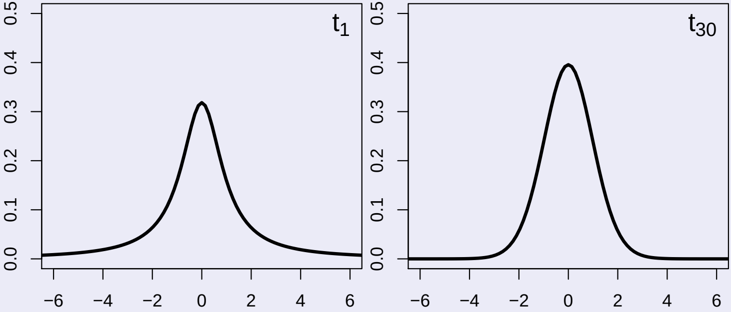
Seien X und $Y_1, \ldots Y_n$ unabhängige Zufallsvariablen mit $X \sim N(0,1)$ und $Y_i \sim N(0,1)$. Dann ist der Quotient

$$\frac{X}{\sqrt{\sum_{i=1}^{n} Y_i^2/n}} \sim t_n$$

t-verteilt mit n Freiheitsgraden.

t-Verteilung

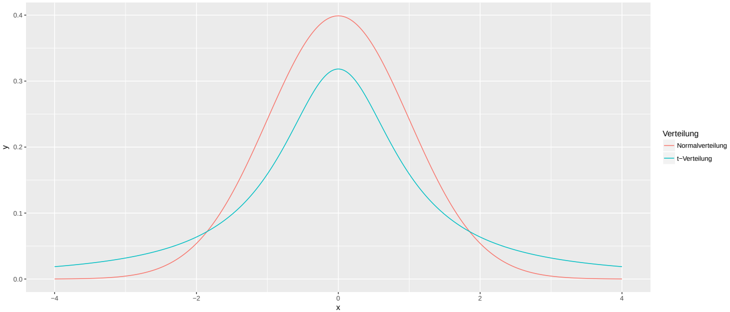
Beispiele der Dichtefunktion



Anwendungen bei Finanzdaten

- Häufig wird die Normalverteilung für die Verteilung von Renditen genutzt
- Problematisch, da die Wahrscheinlichkeit von extremen Ausreißern (Crash) unterschätzt wird
- Abhilfe: Verwende Verteilungen mit **heavy tails** z.B. die t-Verteilung

Vergleich t-Verteilung und Normalverteilung



Mehrdimensionale Zufallsvariablen

Analog zu den Maßzahlen und Überlegungen aus der deskriptiven Statistik:

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

also z.B. $\omega \in \Omega$, zufällig gezogene Person und damit $X(\omega)$ und $Y(\omega)$ Auswertung der Merkmale jeweils an *derselben* Person.
$\Rightarrow$ zweidimensionale Zufallsvariable $\binom{X}{Y}$ (wie bei Zusammenhangsanalyse in Statistik I)
Das Hauptinteresse gilt (entsprechend der Kontingenztafel in Statistik I) der gemeinsamen Verteilung

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

Zweidimensionale Verteilungen

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 $\binom{X}{Y}$ bzw. der Variablen X und Y.
Randwahrscheinlichkeiten:

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

Bedingte Verteilungen

$$
\begin{aligned}
P(X = x_i | Y = y_j) &= \frac{P(X = x_i, Y = y_j)}{P(Y = y_j)} \\
P(Y = y_j | X = x_i) &= \frac{P(X = x_i, Y = y_j)}{P(X = x_i)}
\end{aligned}
$$

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

Seien X und Y zwei Zufallsvariablen. Dann heißt

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

Kovarianz von X und Y.

Rechenregeln

- $\text{Cov}(X, X) = \text{Var}(X)$
- $\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X) \cdot \mathbb{E}(Y)$
- $\text{Cov}(X, Y) = \text{Cov}(Y, X)$
- Mit $\tilde{X} = a_X X + b_X$ und $\tilde{Y} = a_Y Y + b_Y$ ist
$$\text{Cov}(\tilde{X}, \tilde{Y}) = a_X \cdot a_Y \cdot \text{Cov}(X, Y)$$
- $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2 \cdot \text{Cov}(X, Y)$

Korrelation

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 $\mathrm{Var}(X) > 0$ und $\mathrm{Var}(Y) > 0$, so gilt $\rho(X, Y) = 0$ genau dann, wenn $\mathrm{Cov}(X, Y) = 0$.

Grenzwertsätze: Einführung

Big Data: Beobachtung von *großen* Datensätzen

- 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 $X_1, \ldots, X_n$, 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 $X_1, \ldots, X_n$ eine Stichprobe eines Merkmals $\tilde{X}$ bei einer einfachen Zufallsauswahl sind.

Beispiel:
$\tilde{X}$ 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 $X_1, \ldots, X_n$ ist wieder eine Zufallsvariable, z.B. das arithmetische Mittel oder die Stichprobenvarianz

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

Wahrscheinlichkeitsaussagen möglich $\Longrightarrow$ Wahrscheinlichkeitsrechnung anwenden

- Gerade bei diesen Zufallsgrößen ist die Abhängigkeit von n oft wichtig, man schreibt dann $\bar{X}_n$, $\tilde{S}_n^2$
- Sind $X_1, \ldots, X_n$ 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

Erwartungswert und Varianz von $\bar{X}_n$

$X_1, X_2, \ldots, X_n$ seien unabhängig und identisch verteilt.

$$X_1, X_2, \ldots, X_n \qquad i.i.d.$$

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

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

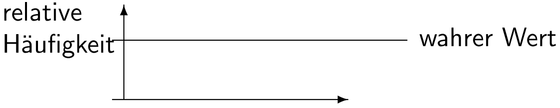
Diese Eigenschaften bilden die Grundlage für die folgenden Sätze.

Das schwache Gesetz der großen Zahlen

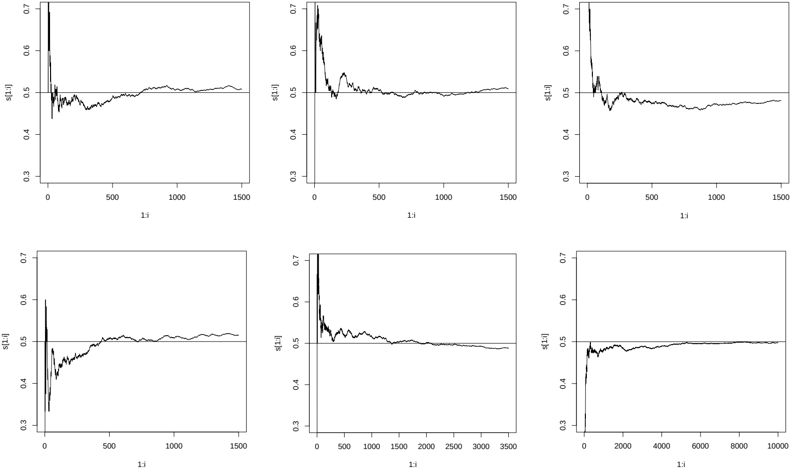
Betrachte für wachsenden Stichprobenumfang n:

- $X_1, \ldots, X_n$ 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 Korridore/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:

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

Theorem von Bernoulli

Seien $X_1, \ldots, X_n$, 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 $\epsilon > 0$

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

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

Zwei wichtige Konsequenzen

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 ($\pi \mathrel{\hat{=}}$ 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 $X_1, \ldots, X_n$ 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 $\epsilon > 0$:

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

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)

Die Verteilungsfunktion

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

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

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

Hauptsatz der Statistik

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

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

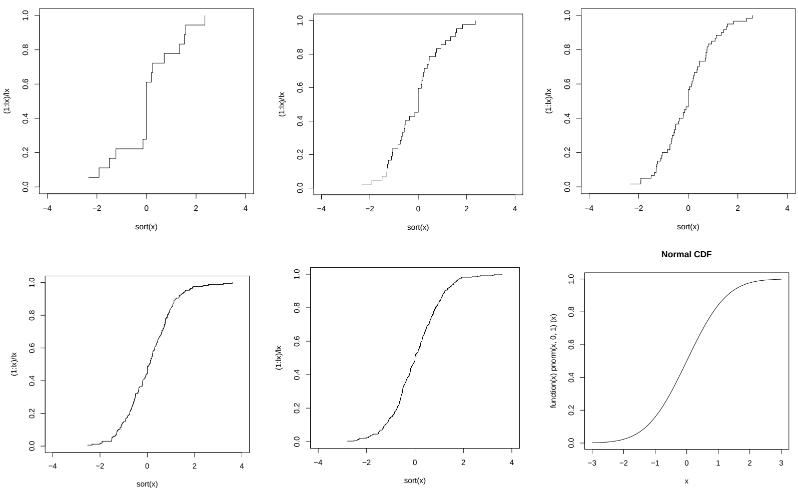
gilt für jedes $c > 0$

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

Interpretation

- „Erträglichkeitsschranke“ c vorgegeben. Wsk, dass maximaler Abstand größer c ist geht für hinreichend großes n gegen 0 $\Longrightarrow$ 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 representativ für die Grundgesamtheit, d.h. man kann Verteilungsgesetzmäßigkeiten durch Beobachtungen erlernen (grundlegend für die Statistik) $\rightarrow$ „Hauptsatz “.

Beispiele



Der zentrale Grenzwertsatz I

- Gibt es für große Stichprobenumfänge Regelmäßigkeiten im Verteilungstyp?

- Gibt es eine Standardverteilung, mit der man oft bei großen empirischen Untersuchungen rechnen kann?

Der zentrale Grenzwertsatz II

Seien $X_1, \ldots, X_n$ i.i.d. mit $\mathbb{E}(X_i) = \mu$ und $\mathrm{Var}(X_i) = \sigma^2 > 0$ 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: $Z_n \overset{a}{\sim} \mathrm{N}(0;1)$, d.h. es gilt für jedes z

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

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

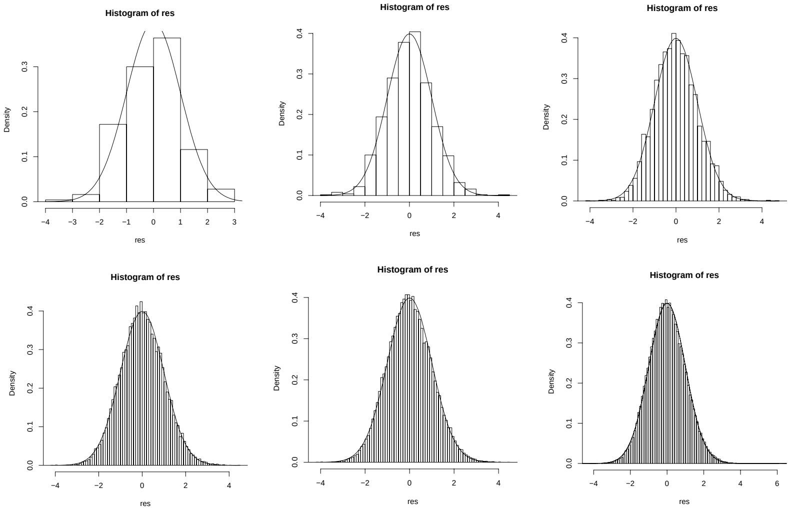
Die Funktion kommt durch *Standardisieren* und durch *geeignetes Mitteln* zustande.

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

$$\sum X_i \quad \longrightarrow \text{verliert sich; } \mathrm{Var}(\sum X_i) \to \infty$$

$$\frac{1}{n}\sum x_i \quad \longrightarrow \mathrm{Var}\left(\frac{1}{n}\sum X_i\right) \to 0$$

Beispiele



Histogram of res
Density
res
Histogram of res
Density
res
Histogram of res
Density
res
Histogram of res
Density
res
Histogram of res
Density
res
Histogram of res
Density
res

Anwendung des zentralen Grenzwertsatz auf $\bar{X}$ I

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

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)*?

Anwendung des zentralen Grenzwertsatz auf $\bar{X}$ II

Aus dem zentralen Grenzwertsatz folgt:

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

$$= \frac{n\bar{X}_n - n\mu}{\sqrt{n}\cdot\sigma} = \frac{\bar{X}_n-\mu}{\sigma/\sqrt{n}} \overset{a}{\sim} \mathrm{N}(0,1)$$

oder auch

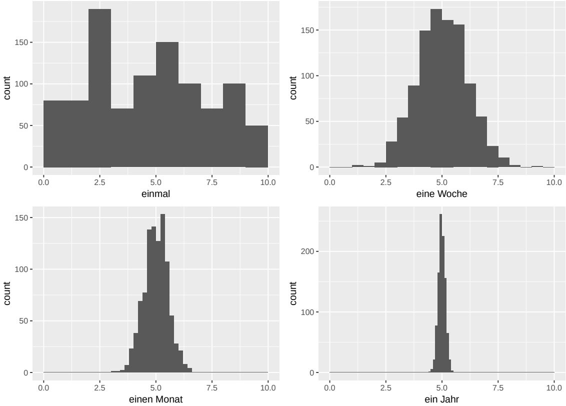
$$\bar{X}_n \overset{a}{\sim} \mathrm{N}\left(\mu, \frac{\sigma^2}{n}\right).$$

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

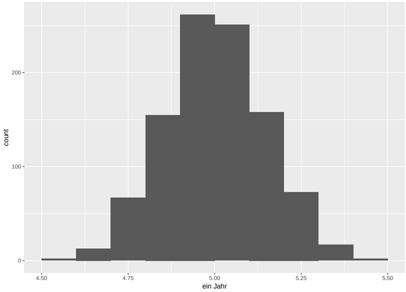
* Schwankung im richtigen Wert (μ)
* Ausschläge werden kleiner

Warten auf den Bus

Bestimme Wartezeit, Durchschnittliche Wartezeit in 1 Woche, 1 Monat, 1 Jahr



1 Jahr



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 $Y_1, \ldots, Y_n$ 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 $Y_i \sim Bin(1, \pi)$ und es gilt

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

Approximation der Binomialverteilung II

Damit lässt sich der zentrale Grenzwertsatz anwenden:

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

und damit

$$\frac{X-\mathbb{E}(X)}{\sqrt{\mathrm{Var}(X)}} \overset{a}{\sim} \mathrm{N}(0,1)$$

so dass

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

falls n groß genug.

Faustregeln

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

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

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

Statistische Inferenz

Ziel: Etwas über die reale Welt lernen, indem man Daten auswertet

1. Schlüsse von Stichprobendaten auf die Grundgesamtheit
2. Schlüsse von Experimentaldaten auf ein allgemeines Phänomen
3. Schlüsse von Beobachtungsdaten auf allgemeine Zusammenhänge
4. Prognosen für die Zukunft mit Hilfe von Daten aus der Vergangenheit

Beispiele:

1. Analysen aus dem SOEP (Sozioökonomisches Panel), Wahlumfragen
2. Klinische Studie zur Wirkung eines Medikaments
3. Verkäufe und Fernsehwerbung
4. Wirtschaftsprognosen (Wachstum, Inflation etc.)

Wahrscheinlichkeit und Statistische Inferenz

- Stichprobe zufällig gezogen
- Ergebnis von Experiment enthält stochastische Komponenten
- Modelle für Beobachtungen enthalten stochastische Terme und Annahmen

Inhalte

1. Berücksichtigung der Zufälligkeit
2. Folgen für die Aussagekraft
3. Fehlerabschätzung

Voraussetzungen für das Anwenden statistischer Inferenz

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

Punktschätzung

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?

Beispiel 1: Schätzer $\bar{X}$

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	$\bar{X}$	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 1: Schätzer $\bar{X}$

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	$\bar{X}$	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

“Glück“

Punktschätzung

Beachte: Auswahl zufällig $\Rightarrow$ Schätzung zufällig

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

Schätzfunktionen

Definition

Sei $X_1, \dots, X_n$ eine 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 (z.B. Bildung des Mittelwerts) auf sie an.“

Beispiele für Schätzfunktionen

- Arithmetisches Mittel der Stichprobe:

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

Für binäre (0-1)-Größen X_i ist $\bar{X}$ auch die relative Häufigkeit des Auftretens von „$X_i = 1$“ in der Stichprobe.

- Stichprobenvarianz:

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

Beispiele für unübliche Schätzfunktionen

- Größter Stichprobenwert:

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

- Kleinster Stichprobenwert:

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

Qualitätsmerkmal eines Schätzers

Erwartungstreue, Bias:
Gegeben sei eine Stichprobe $X_1, \ldots, X_n$ und eine Schätzfunktion $T = g(X_1, \ldots, X_n)$ (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.

Bias und Erwartungstreue für $\bar{X}$

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

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

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

Bias und Erwartungstreue für S^2

Es gilt (Beachte hier: Division durch n):

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

Man erhält also einen (leicht) verzerrten Schätzer.
Für die Stichprobenvarianz gilt daher:

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

Also ist S^2 erwartungstreu für σ^2. Diese Eigenschaft ist auch die Motivation für die Division durch n-1.

Nichtlineare Funktionen

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 $\sqrt{\sigma^2} = \sigma$.

Wahlumfrage

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

Grundgesamtheit: Dichotomes Merkmal

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

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

Stichprobe $X_1, \ldots, X_n$ vom Umfang n:

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

Anteil als erwartungstreuer Schätzer

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 alleine ist ein schwaches Kriterium!

Betrachte die offensichtlich unsinnige Schätzfunktion:

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

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 erwartungtreu:

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

Deshalb betrachtet man zusätzlich die Effizienz eines Schätzers.

Effizienz I

Beispiel Wahlumfrage:

Gegeben sind zwei erwartungstreue Schätzer (n sei gerade):

$$
\begin{aligned}
T_1 &= \frac{1}{n}\sum_{i=1}^{n} X_i \\
T_2 &= \frac{1}{n/2}\sum_{i=1}^{n/2} X_i
\end{aligned}
$$

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!

Effizienz II

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

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

und damit:

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

Effizienz III

Analog kann man zeigen:

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

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".

Effizienz IV

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

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

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

Beste Schätzer

- 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 (*u*niformly *m*inimum *v*ariance *u*nbiased), falls

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

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

UMVU-Schätzer

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

Verzerrte Schätzer

- Ist $X_1, \ldots, X_n$ 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 $T = g(X_1, \ldots, X_n) = 42$, der die Stichprobe nicht beachtet, Varianz 0.

MSE

Man zieht dann den sogenannten *Mean Squared Error*

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

zur Beurteilung heran. Es gilt

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

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

$$\lim_{n \to \infty} E(\hat{\vartheta}) = \vartheta$$

  bzw.

$$\lim_{n \to \infty} \mathrm{Bias}(\hat{\vartheta}) = 0$$

  gelten.

- Abschwächung des Begriffs der Erwartungstreue: Gilt nur noch bei einer unendlich großen Stichprobe.
- Erwartungstreue Schätzer sind auch asymptotisch erwartungstreu.

Konsistenz

- 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

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

Konsistenz von $\bar{X}$

Der MSE von $\bar{X}$ ist gegeben durch

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

$\bar{X}$ ist also ein MSE-konsistenter Schäter 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 jeweiliger Partei organisiert.

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

Ein Bus fährt vor, sie wollen anhand einer Stichprobe ermitteln, welcher Bus vorgefahren ist.

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

- Welche Partei? Wohl A, aber B ist 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 $X_1, \ldots, X_n$ mit der konkreten Realisation $x_1, \ldots, x_n$.

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

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

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

- 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).

Maximum–Likelihood–Prinzip: Beispiel

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

$$6 \quad 5 \quad 3 \quad 4 \quad 4$$

Wahrscheinlichkeit der Stichprobe für gegebenes π:

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

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

Wahrscheinlichkeit für einige Werte von π

π	$P(X_1 = 6, \ldots, X_5 = 4 \vert \pi)$
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 $L(\vartheta) = P_\vartheta(X_1 = x_1, \ldots, X_n = x_n)$, nun als Funktion von ϑ gesehen, die *Likelihood* (deutsch: Plausibilität, Mutmaßlichkeit) von ϑ gegeben die Realisation $x_1, \ldots, x_n$.

Deduktion und Induktion

- Deduktiv (Wahrscheinlichkeitsrechnung): ϑ bekannt, $x_1, \ldots, x_n$ zufällig („unbekannt").
- Induktiv (Statistik): ϑ unbekannt, $x_1, \ldots, x_n$ bekannt.

Deduktiv	Induktiv
geg: Parameter bekannt	**ges:** Plausibilität des Parameters
$P_\vartheta(X_1 = x_1, \ldots, X_n = x_n)$	$P_\vartheta(X_1 = x_1, \ldots, X_n = x_n)$
Funktion von $x_1, \ldots, x_n$ bei festem ϑ	Funktion von ϑ bei festen $x_1, \ldots, x_n$
ges: Wskt. von Beobachtungen	**geg:** Beobachtungen bekannt

Definition Maximum Likelihood

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

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

heißt *Likelihood* des Parameters ϑ bei den Beobachtungen $x_1, \dots, x_n$.

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*.

Likelihood bei stetige Verteilungen

- 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.

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

ML-Schätzung für π einer Bernoulliverteilung I

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

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

$$\begin{aligned} P(X_i = 1) &= \pi \\ P(X_i = 0) &= 1 - \pi \\ P(X_i = x_i) &= \pi^{x_i} \cdot (1-\pi)^{1-x_i}\,, \quad x_i \in \{0; 1\}. \end{aligned}$$

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

ML-Schätzung für π einer Bernoulliverteilung II

- Bestimme die Likelihoodfunktion:

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

ML-Schätzung für π einer Bernoulliverteilung III

- Berechne die logarithmierte Likelihoodfunktion:

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

- Ableiten der Log-Likelihood (nach π):

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

ML-Schätzung für π einer Bernoulliverteilung IV

- Bemerkung zur Log-Likelihood:

  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 $g(L(x))$.

ML-Schätzung für π einer Bernoulliverteilung V

- Berechnung des ML-Schätzers durch Nullsetzen der abgeleiteten Loglikelihoodfunktion:

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

also:

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

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

ML-Schätzung bei Normalverteilung I

- Bestimme die Likelihoodfunktion:

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

- Bestimme die Log-Likelihoodfunktion:

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

ML-Schätzung bei Normalverteilung II

- Ableiten und Nullsetzen der Log-Likelihoodfunktion:

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

ML-Schätzung bei Normalverteilung

- 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

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

Fazit

- Der ML-Schätzer $\hat{\mu} = \bar{X}$ für μ stimmt mit dem üblichen Schätzer für den Erwartungswert überein.
- Der ML-Schätzer $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$ für σ^2 ist verzerrt, d.h. nicht erwartungstreu.

Einige allgemeine Eigenschaften von ML-Schätzern

- ML-Schätzer $\hat{\vartheta}$ sind im Allgemeinen nicht erwartungstreu.
- ML-Schätzer $\hat{\vartheta}$ sind asymptotisch erwartungstreu.
- ML-Schätzer $\hat{\vartheta}$ sind konsistent.

Zusammenfassung und Ausblick

- Schätztheorie ist ein zentrales Werkzeug statistischer Inferenz.
- Stochastische Aussagen über Schätzfehler möglich.
- Bias und MSE sind zentrale Kenngrößen.
- Likelihood wichtige allgemeine Basis.
- Intervallschätzungen werde im nächsten Kapitel besprochen.

Intervallschätzung: Motivation

Annahme:
Der wahre Anteil der CDU/CSU - Wähler 2017 liegt bei genau 40.0%.
Wie groß ist die Wahrscheinlichkeit, in einer Zufallsstichprobe von 1000 Personen genau einen relativen Anteil von 40.0% von CDU/CSU Wählern zu erhalten?

$$
\begin{aligned}
X_i &= \begin{cases} 1, \text{CDU/CSU} \\ 0, \text{sonst} \end{cases} \\
P(X_i = 1) &= \pi = 0.4 \\
X &= \sum_{i=1}^{n} X_i \sim B(n, \pi) \text{ mit } n = 1000 \\
\hat{\pi} &= \frac{X}{n}
\end{aligned}
$$

Punktschätzer

Berechnung der Wahrscheinlichkeit:

$$\begin{aligned} P(X = 400) &= \binom{n}{x} \cdot \pi^x \cdot (1-\pi)^{n-x} \\ &= \binom{1000}{400} \cdot 0.4^{400} \cdot (1-0.4)^{600} \\ &= 0.026 \end{aligned}$$

Mit Wahrscheinlichkeit von etwa 97.4% verfehlt der Schätzer den wahren Wert.
Beim Runden auf ganze Prozente muss der Anteil der CDU/CSU - Wähler in der Stichprobe zwischen 395 und 404 liegen, um 40% zu erhalten:

$$P(395 \leq X \leq 404) = 0.25$$

Auch beim Runden auf ganze Prozente ergibt sich mit Wahrscheinlichkeit 75% ein falscher Wert.

Konsequenzen

- Vorsicht bei der Interpretation, insbesondere bei „knappen Ergebnissen“
- Angabe der Genauigkeit
- Geeignete Wahl des Stichprobenumfangs
- Es ist häufig nicht sinnvoll, sich genau auf einen Wert festzulegen. Oft ist die Angabe eines Intervalls, von dem man hofft, dass es den wahren Wert überdeckt, vorzuziehen:
  $\Rightarrow$ *Intervallschätzung*

Schätzgenauigkeit

Anteilschätzer:
Schätzung des Anteils in der Grundgesamtheit (bzw. der Erfolgswahrscheinlichkeit) π durch relative Häufigkeit in der Stichprobe.
Gegeben: i.i.d. Stichprobe $X_1, \ldots X_n$ mit $X_i \in \{0, 1\}$

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

Dann kann die Schätzgenauigkeit durch die Standardabweichung von $\hat{\pi}$ charakterisiert werden:

$$SE(\hat{\pi}) = \sqrt{\frac{\hat{\pi} \cdot (1 - \hat{\pi})}{n}}$$

Die Standardabweichung eines Schätzers wird auch häufig als **Standardfehler** (**englisch: standard error**) bezeichnet.

Berechnung des Standardfehlers

Standardfehler für verschieden Stichprobenumfänge n und (wahre) Erfolgswahrscheinlichkeiten n: Angaben in Prozentpunkten.

n	$\pi = 10\%$	$\pi = 40\%$	$\pi = 50\%$
20	6.71	10.95	11.18
100	3.00	4.90	5.00
1000	0.95	1.55	1.58
2000	0.67	1.10	1.12
5000	0.42	0.69	0.71

Beachte: π unbekannt.
Höchste Werte für $\pi = 0.5$. Daher können diese Werte als obere Grenze verwendet werden. Bei einem Stichprobenumfang von n =1000 liegt der Standardfehler (SE) also unter 1.58%.

Mittelwertsschätzung

Schätzung des Mittelwertes in der Grundgesamtheit (bzw. des Erwartungswertes μ bei einem Experiment) durch den Mittwelwert $\bar{X}$ in der Stichprobe.
Gegeben: i.i.d. Stichprobe $X_1, \dots X_n$ mit $E(X_i) = \mu$

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

Dann kann die Schätzgenauigkeit durch die Standardabweichung von $\hat{\mu}$ charakterisiert werden:

$$SE(\hat{\mu}) = \sqrt{\frac{\sigma^2}{n}} = SEM$$

Die Standardabweichung wird auch häufig als **Standardfehler** (**englisch: standard error** oder **standard error of the mean (SEM)**) bezeichnet.

Beispiel: Schätzgenauigkeit bei Umsatz von Kunden

Big Data Anwendung:
Eine Firma möchte die durchschnittliche Dauer der Internetnutzung ihrer 1 Million Kunden schätzen.
Konservative, d.h. eher zu hohe, Schätzung der Standardabweichung:
$\sigma = 120$ Minuten
Berechnung des Standardfehlers bei verschiedenen Stichprobengrößen:

n	SE (Minuten)
20	26.83
100	12.00
1000	3.79
2000	2.68
5000	1.70

Es ist also nicht immer nötig, die Daten von allen Kunden auszuwerten.
Man kann sich oft auf eine Zufallsstichprobe beschränken.

Standardfehler und Angabe von Schwankungsbreiten

- Standardfehler wichtiges Kriterium, aber eher schwer zu kommunizieren
- Alternative: Schwankungsbreiten und Unsicherheit
- Benutze asymptotische Normalverteilung

Die Schätzer $\hat{\pi}$ und $\hat{\mu}$ sind asymptotisch normalverteilt.
Ist der Standardfehler des Schätzer gegeben, so gilt:

$$\begin{aligned} P(\hat{\pi} \in [\pi - 2 \cdot SE(\hat{\pi}); \pi + 2 \cdot SE(\hat{\pi})]) &= 0.95 \\ P(\hat{\mu} \in [\mu - 2 \cdot SE(\hat{\mu}); \mu + 2 \cdot SE(\hat{\mu})]) &= 0.95 \end{aligned}$$

Illustration mit R.

Symmetrische Intervallschätzung

Allgemeiner Ansatz:
Basierend auf einer Schätzfunktion $T = g(X_1, \ldots, X_n)$ 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:
$I(T) = [0, 100\%]$ 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 von Konfidenzintervallen

Definition

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

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

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.

Konstruktion von Konfidenzintervallen

Praktische Vorgehensweise: Suche Zufallsvariable Z_ϑ, die

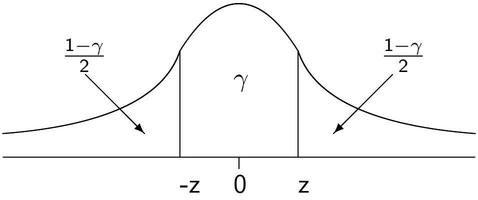
- den gesuchten Parameter ϑ enthält und
- deren Verteilung aber nicht mehr von dem Parameter abhängt („*Pivotgröße*“, dt. Angelpunkt).
- Dann wähle den Bereich C_Z so, dass $P_\vartheta(Z_\vartheta \in C_Z) = \gamma$ und
- löse nach ϑ auf.

Konfidenzintervall für den Mittelwert (normalverteiltes Merkmal, Varianz bekannt)

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

1. Starte mit der Verteilung von $\bar{X}$:
$$\bar{X} \sim \mathrm{N}(\mu, \sigma^2/n).$$
2. Dann erfüllt
$$Z = \frac{\bar{X} - \mu}{\sigma} \cdot \sqrt{n} \sim \mathrm{N}(0; 1)$$
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:

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

beziehungsweise

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

Wichtige Quantile der Standardnormalverteilung

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

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

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

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

Herleitung KI

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

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

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

KI für Mittelwert (NV mit bekannter Varianz)

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, desto genauer!)
  Aufpassen, die Genauigkeit nimmt nur mit $\sqrt{n}$ zu. Halbierung des Intervalls, Vervierfachung 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

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

(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.

$\longrightarrow$ Wir benötigen die t-Verteilung.

Eigenschaften der t-Verteilung

- Je größer ν ist, umso ähnlicher sind sich die $t(\nu)$-Verteilung und die Standardnormalverteilung.
  - Für $\nu \to \infty$ 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 $\frac{\bar{X}-\mu}{\sigma}\sqrt{n}$.

Konfidenzintervall zum Konfidenzniveau γ

Ausgehend von

$$P\left(-t^{(n-1)}_{\frac{1+\gamma}{2}} \leq \frac{\bar{X}-\mu}{S}\cdot\sqrt{n} \leq t^{(n-1)}_{\frac{1+\gamma}{2}}\right) = \gamma$$

wie im Beispiel mit bekannter Varianz nach μ auflösen (mit S statt σ):

$$P\left(\bar{X} - \frac{t^{(n-1)}_{\frac{1+\gamma}{2}}\cdot S}{\sqrt{n}} \leq \mu \leq \bar{X} + \frac{t^{(n-1)}_{\frac{1+\gamma}{2}}\cdot S}{\sqrt{n}}\right) = \gamma$$

Damit ergibt sich:

Konfidenzintervall für μ bei unbekannter Varianz

$$\left[\bar{X} \pm \frac{t^{(n-1)}_{\frac{1+\gamma}{2}}\cdot S}{\sqrt{n}}\right]$$

Eigenschaften

- Es gelten analoge Aussagen zum Stichprobenumfang und Konfidenzniveau wie bei bekannter Varianz.
- Für jedes γ (und jedes ν) gilt:

  $$t^{\nu}_{\frac{1+\gamma}{2}} > z_{\frac{1+\gamma}{2}} .$$

  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 Füllgewichtzum Sicherheitsniveau von 95%.

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: $s = 10g$.

Beispiel: Konfidenzintervall zum Konfidenzniveau γ

- Konstruktion des Konfidenzintervalls:
  - Da die Varianz σ^2 unbekannt ist, muss das Konfidenzintervall basierend auf der t-Verteilung konstruiert werden:

$$[\bar{X} \pm \frac{t^{(n-1)}_{\frac{1+\gamma}{2}} \cdot S}{\sqrt{n}}]$$

    Aus dem Sicherheitsniveau $\gamma = 0.95$ errechnet sich $\frac{1+\gamma}{2} = 0.975$.
    Quantil der t-Verteilung bei 0.975 und 15 Freiheitsgraden ($T = \frac{\bar{X}-\mu}{S}\sqrt{n}$ ist t-verteilt mit n-1 Freiheitsgeraden) liefert $t^{15}_{0.975} = 2.13$.
  - Einsetzen liefert damit:

$$[245 \pm 2.13 \cdot \frac{10}{4}] = [239.675; 250.325]$$

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 $\pi = P(X = 1)$ einer Bernoulli-Zufallsgröße X

- $X_1, \ldots, X_n$ i.i.d. Stichprobe
- n hinreichend groß (Faustregel $n > 30$)
- vorgegebenes Sicherheitsniveau γ

Approximatives Konfidenzintervall für π

$$\boxed{\hat{\pi} \pm z_{\frac{1+\gamma}{2}} \cdot \sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}}$$

$\hat{\pi}$ = Anteil aus der Stichprobe

$z_{\frac{1+\gamma}{2}}$ ist das $\frac{1+\gamma}{2}$-Quantil der Standardnormalverteilung.

Beispiel: Wahlumfrage

- Gegeben:
  - $n = 500$
  - $\hat{\pi} = 46.5\%$
  - $\gamma = 95\%$ und damit $z_{\frac{1+\gamma}{2}} = 1.96$
- Konfidenzintervall:

$$\left[\hat{\pi} \pm z_{\frac{1+\gamma}{2}} \cdot \sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}\right] = \left[0.465 \pm 1.96 \cdot \sqrt{\frac{0.465(1-0.465)}{500}}\right]$$

$$= [0.421; 0.508]$$

Inhaltliche Bemerkung (Beispiel: Wahlumfrage)

- 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 werden bei Wahlumfragen Zusatzinformation verwendet (insbesondere auch frühere Wahlergebnisse). „Gebundene Hochrechnung“
- Zu der Unsicherheit durch die Stichprobenziehung kommen weitere Probleme wie falsche Antworten, Antwortverweigerung, Nicht–Erreichbarkeit von Personen. Dies kann zu Verzerrungen und deutlicher Unterschätzug des Fehlers führen.

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

$$g \geq z_{\frac{1+\gamma}{2}} \cdot \sqrt{\frac{\pi(1-\pi)}{n}}$$

Auflösen nach n:

$$n \geq \frac{1}{g^2} z^2_{\frac{1+\gamma}{2}} \cdot \pi(1-\pi)$$

Beachte: $\pi(1-\pi) \leq 0.25$

Beispiel: Stichprobenplanung bei Anteilsschätzung

Gegeben:

- Konfidenzniveau: 0.95
- Genauigkeit: 10%

Bestimmung von n:

$$n \geq \frac{1}{g^2} z^2_{\frac{1+\gamma}{2}} \cdot \pi(1-\pi) = \frac{1}{0.1^2} 1.96^2 \cdot 0.25 = 96.04$$

Beachte: $\pi(1-\pi) \leq 0.25$

Also sollten ca. 100 Personen befragt werden.

Bei $g = 5\%$ ergibt sich $n = 385$
Bei $g = 1\%$ ergibt sich $n = 9604$

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: $X_1, \ldots, X_{n_X}$, X_i i.i.d.
  - Daten aus Gruppe 2: $Y_1, \ldots, Y_{n_Y}$, Y_j i.i.d.

- Stichprobenumfänge hinreichend groß ($n_X \geq 30$, $n_Y \geq 30$)
- Schätzung: $\bar{X} - \bar{Y} = \frac{1}{n_X} \sum_{i=1}^{n_X} X_i - \frac{1}{n_Y} \sum_{i=j}^{n_Y} Y_j$

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

- $S_d = \sqrt{\frac{S_X^2}{n_X} + \frac{S_Y^2}{n_Y}}$
- $z_{\frac{1+\gamma}{2}}$ ist das $\frac{1+\gamma}{2}$-Quantil der Standardnormalverteilung

Beispiel: Radiohördauer Ost-West

Westen:	$\bar{x} = 11.4$ Stunden und	$s_X = 8.4$	m = 259
Osten:	$\bar{y} = 9.5$ Stunden und	$s_Y = 8.4$	n= 941

$$\sqrt{\frac{s_X^2}{n_X} + \frac{s_Y^2}{n_Y}} \approx 0.6$$

Wir berechnen ein 99% - Konfidenzintervall:

$$k_u = \bar{x} - \bar{y} - z_{\frac{1+\gamma}{2}} \cdot \sqrt{\frac{s_X^2}{n_X} + \frac{s_Y^2}{n_Y}} = 0.38$$

$$k_o = \bar{x} - \bar{y} + z_{\frac{1+\gamma}{2}} \cdot \sqrt{\frac{s_X^2}{n_X} + \frac{s_Y^2}{n_Y}} = 3.42$$

Die Differenz liegt also zwischen 0.38 und 3.42 h/Woche
Werte für 95% - KI: $[0.74h; 3.1h]$

Zusammenfassung

- Konfidenzintervalle sind zentrales Instrument statistischer Inferenz.
- Unsicherheit der Aussagen direkt interpretierbar.
- Interpretation des Sicherheitsniveaus problematisch.
- (Fehl–)Interpretation als Wahrscheinlichkeit für den unbekannten Parameter in manchen Fällen vertretbar (Bayes-Inferenz).

Hypothese

„Behauptung einer Tatsache, deren Überprüfung noch aussteht" (Leutner in: Endruweit, Trommsdorff: Wörterbuch der Soziologie, 1989).

Statistischer Test: Überprüfung von Hypothesen anhand einer Stichprobe

Idealtypische Vorgehensweise
Wissenschaftlicher Fortschritt durch Falsifikation von Hypothesen

Statistische Testtheorie

Schließe von Stichprobe oder 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 einer Stichprobe möglicherweise ein **reines Zufallsprodukt** oder **mit großer Sicherheit** auf einen **realen Effekt** zurückzuführen?
$\longrightarrow$ Dazu notwendig:
Formale Entscheidungsregel = Statistischer Test

Beispiel: Münzdrehen (2€)

Zeitungsberichte: 2€ Münzen nicht „fair“

Münzhypothese

- Vermutung:
  2€Münze nicht fair
- Überprüfung: 10-Mal die Münze werfen, Anzahl ,,Zahl'' notieren

Mögliche Ergebnisse des Experiments

- 5-Mal „Zahl“
  $\longrightarrow$ deutet nicht auf eine unfaire Münze hin
- 10-Mal „Zahl“
  $\longrightarrow$ verdächtig, die Münze ist vermutlich nicht fair
- 0-Mal „Zahl“
  $\longrightarrow$ verdächtig, die Münze ist vermutlich nicht fair
- 8-Mal „Zahl“
  $\longrightarrow$?? 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

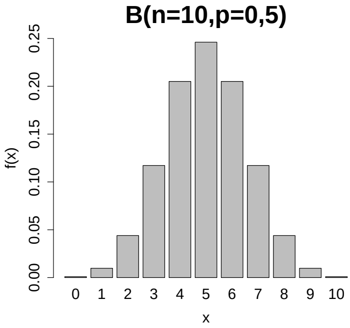
$$\sum_{i=1}^{10} X_i \quad \sim \quad B(n = 10, \pi)$$

- Welche Ergebnisse sind wahrscheinlich, falls die Münze fair ist?
- **Falls die Münze fair ist**, so ist die Anzahl „Zahl“ binomialverteilt mit $p = 0.5$.

$$\sum_{i=1}^{10} X_i \quad \sim \quad B(n = 10, \pi = 0.5)$$

- **Falls die Münze fair ist**, so sollte $\sum_{i=1}^{10} X_i$ mit einer Wahrscheinlichkeit von 95 % nicht weit entfernt vom Erwartungswert 5 liegen.

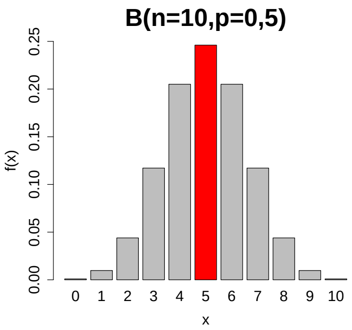
Binomialverteilung



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

$$\sum =$$

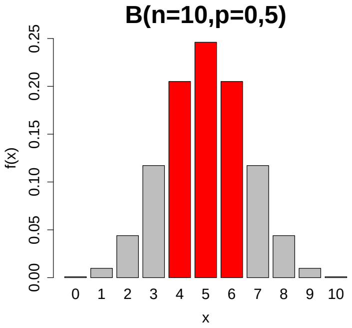
Binomialverteilung



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 0.246	0.205	0.117	0.044	0.010	0.001

$$\sum = 0.246$$

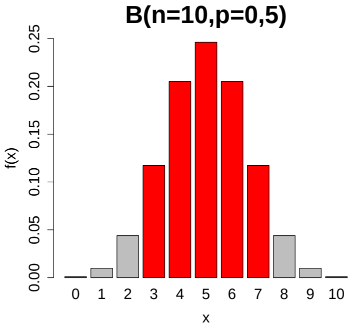
Binomialverteilung



x	0	1	2	3	4	5	6	7	8	9	10
f(x)	0.001	0.010	0.044	0.117	0.205 0.205	0.246 0.246	0.205 0.205	0.117	0.044	0.010	0.001

$$\sum = 0.656$$

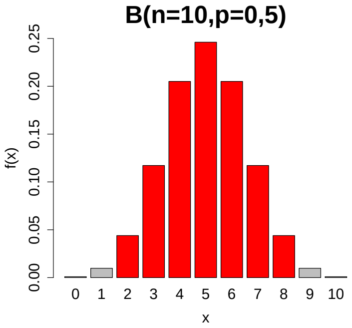
Binomialverteilung



x	0	1	2	3	4	5	6	7	8	9	10
f(x)	0.001	0.010	0.044	0.117 0.117	0.205 0.205	0.246 0.246	0.205 0.205	0.117 0.117	0.044	0.010	0.001

$$\sum = 0.890$$

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
			0.044	0.117	0.205	0.246	0.205	0.117	0.044		

$$\sum = 0.978$$

Münzhypothese

- Falls die Münze fair ist, so liegt die Anzahl von „Zahl“ bei $n = 10$ Würfen mit einer Wahrscheinlichkeit von mindestens 95% im Bereich

$$\{2, 3, 4, 5, 6, 7, 8\}$$

- Falls die Anzahl von „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 von „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.

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

$$H_0 \quad : \quad \pi = 0.5$$

- Gegenhypothese H_1= Alternative zur Nullhypothese.
  Beispiel: Die Münze ist nicht fair

$$H_1 \quad : \quad \pi \neq 0.5$$

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 $n = 10$ Würfen. Unter H_0 gilt:

$$T = \sum_{i=1}^{10} X_i \quad \sim B(n = 10, \pi = \mathbf{0.5})$$

Statistischer Test: Annahme- und Ablehnbereich

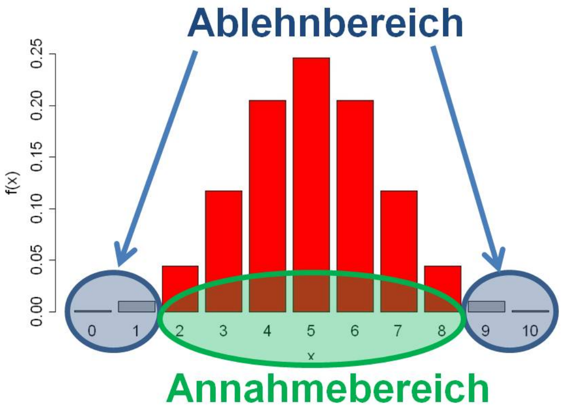
- Der Annahmebereich des Tests ist der Bereich, in dem die Prüfgröße T mit einer hohen Wahrscheinlichkeit (mindestens $1 - \alpha$) liegt.
  Beispiel: $\alpha = 0.05$ und

$$\text{Annahmebereich} \quad = \quad \{2, 3, 4, 5, 6, 7, 8\}$$

- 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: $\alpha = 0.05$ und

$$\text{Ablehnbereich} \quad = \quad \{0, 1, 9, 10\}$$

Beispiel Annahme- und Ablehnbereich



Statistischer Test: Experiment und Entscheidung

Wir ziehen eine Stichprobe und berechnen den Wert der Teststatistik T.

1. Fall: Der Wert der Teststatistik liegt im Annahmebereich.
$\longrightarrow$ Wir behalten die Nullhypothese H_0 bei.

2. Fall: Der Wert der Teststatistik liegt im Ablehnbereich.
$\longrightarrow$ 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	$\surd$	Fehler 1.Art
H_0 falsch	Fehler 2. Art	$\surd$

Man kann nicht beide Fehlerwahrscheinlichkeiten gleichzeitig kontrollieren! (Tradeoff!)
$\Rightarrow$ asymmetrische Vorgehensweise:

Der Fehler 1. Art wird kontrolliert durch die Angabe einer Oberschranke α („Signifikanzniveau“)

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.
  $\Rightarrow$ Nur wenn die Wirkung des Medikaments überzeugend ist, soll es zugelassen werden.

Fehler 1. Art (α-Fehler):

- Die Nullhypothese wird abgelehnt, obwohl sie in Wirklichkeit wahr 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.

  $\Rightarrow$ 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 vergegebenen 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.

$$\downarrow$$

Ungleichbehandlung beider Fehlerarten
$\rightarrow$ Grund für Formulierung der eigentlichen Forschungsfrage
als statistische Alternative:
Entscheidung für H_1 durch α statistisch abgesichert!

Veranschaulichung

- 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.

$$\downarrow$$

- Fehler 1. Art: Unschuldiger wird verurteilt
- Fehler 2. Art: Schuldiger wird nicht verurteilt

p-Wert

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: Zweiseitiger Test

- P_{H_0}(10 „Zahl“)+ P_{H_0}(0 „Zahl“)= 0.002
  10 „Zahl“ $\Rightarrow$ p–Wert 0.002
- P_{H_0}(9 „Zahl“)= 0.01
  $\longrightarrow$ P_{H_0}(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 „Zahl“)= 0.044
  $\longrightarrow$$P_{H_0}$(mindestens 8 „Zahl“ oder höchstens 2 „Zahl“)
  $= 2 \cdot (0.001 + 0.01 + 0.044) = 0.110$
  8 „Zahl“ $\Rightarrow$ p–Wert $= 0.110$
- P_{H_0}(7 „Zahl“)= 0.117
  $\longrightarrow$ P_{H_0}(mehr als 7 „Zahl“ oder höchstens 3 „Zahl“)
  $= 2 \cdot (0.001 + 0.01 + 0.044 + 0.117) = 0.344$
  7 „Zahl“ $\Rightarrow$ p–Wert =0.344

Testentscheidung durch p-Wert

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.
Illustration mit R
Je kleiner der p–Wert desto weniger passen die Daten zur Nullhypothese

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.

Testtheorie

Motivation

Die Prüfung einer statistischen Hypothese H_0 erfolgt mit statistischen Tests.

Ausgangspunkt ist die *Beobachtung* einer Zufallsvariablen in einer *zufälligen Stichprobe* oder einem *Experiment* .

Mittels der daraus gewonnenen *Schätzungen* der unbekannten Parameter will man zu einer *Aussage* über die Glaubwürdigkeit der Hypothese H_0 gelangen.

Testtheorie

Definition Hypothesenraum

Der statistische Test stellt eine Methode dar, Verteilungsannahmen über eine Zufallsvariable X anhand einer konkreten Stichprobe zu überprüfen.

Die Menge aller für die Zufallsvariable X in Frage kommenden Verteilungen wird als *Hypothesenraum* Ω bezeichnet. Diese Menge ist *vor* der Durchführung eines Test festzulegen.

Definition parametrisches Testproblem

Betrachtet man einen Hypothesenraum Ω, der nur Verteilungen *einer* Familie (z.B. Normalverteilung) enthält, so ist die Festlegung von Ω äquivalent zur Festlegung des Parameterraums Θ, der alle möglichen Werte eines Verteilungsparameters θ enthält. In diesem Fall spricht man von einem *parametrischen Testproblem*.

Testtheorie

Definition Nullhypothese und Alternative

Bei einem parametrischen Testproblem wird der Hypothesenraum (Parameterraum) in zwei Teilmengen aufgeteilt:

Nullhypothese die zu testende Hypothese, die durch den Test widerlegt werden soll: $H_0 = \{\theta | \theta \in \Theta_0\}$

Alternative diejenige Hypothese, die durch den Test gezeigt werden soll: $H_1 = \{\theta | \theta \in \Theta_1\}$

Dabei gilt stets: $\Theta_0 \cap \Theta_1 = \emptyset$

Definition Signifikanztest

Ein Test heißt *Signifikanztest*, wenn die Nullhypothese direkt an die Alternative „grenzt“, d.h., wenn die minimale Distanz zwischen beiden Hypothesen gleich Null ist (z.B. $H_0 : \mu = \mu_0$ gegen $H_1 : \mu \neq \mu_0$).

Testtheorie

Definition Testgröße

Die Funktion $T(\mathbf{X}) = T(X_1, ..., X_n)$ der Stichprobenvariblen $\mathbf{X} = (X_1, ..., X_n)$ heißt *Testgröße* oder *Prüfgröße*.

Für die konkrete Stichprobe $(x_1, ..., x_n)$ ergibt sich $t = T(x_1, ..., x_n)$ als Realisation der Zufallsgröße $T(\mathbf{X})$.

Definition kritischer Bereich und Annahmebereich

Der Wertebereich der Zufallsgröße $T(X_1, ..., X_n)$ wird in zwei Teilbereiche zerlegt:

kritischer Bereich K H_0 wird abgelehnt, falls

$$t = T(x_1, ..., x_n) \in K$$

Annahmebereich $\bar{K}$ H_0 wird beibehalten, falls

$$t = T(x_1, ..., x_n) \notin K$$

Testtheorie

Definition Fehler 1. und 2. Art

Bei der Durchführung eines statistischen Tests können folgende vier Situationen auftreten:

	H_0 wird beibehalten	H_0 wird abgelehnt
H_0 wahr	richtige Entscheidung	*Fehler 1. Art*
H_1 wahr	*Fehler 2. Art*	richtige Entscheidung

Definition Signifikanzniveau und Niveau-α-Test

Bei der Konstruktion eines Tests gibt man sich für die Wahrscheinlichkeit des Fehlers 1. Art eine Schranke α vor (z.B. $\alpha = 0,05$), die nicht überschritten werden darf.
Diese Schranke bezeichnet man als *Signifikanzniveau* des Tests.
Der zugehörige Test heißt dann *Signifikanztest zum Niveau* α oder kurz *Niveau-α-Test*.

Testtheorie

ein- und zweiseitige Tests

Fall	Null-hypothese	Alternativ-hypothese	Testproblem
(a)	$\theta = \theta_0$	$\theta \neq \theta_0$	zweiseitig
(b)	$\theta \geq \theta_0$	$\theta < \theta_0$	einseitig
(c)	$\theta \leq \theta_0$	$\theta > \theta_0$	einseitig

Testtheorie

allgemeines Vorgehen bei Tests

1. Verteilungsannahme über die Zufallsvariable X
2. Formulierung von H_0 und H_1
3. Vorgabe der Irrtumswahrscheinlichkeit α
4. Konstruktion bzw. Wahl einer geeigneten Testgröße $T(X) = T(X_1, ..., X_n)$ als Funktion der Stichprobenvariablen X
5. Wahl eines kritischen Bereichs K mit $P_\theta(T(X) \in K) \leq \alpha$ für alle $\theta \in \Theta_0$
6. Berechnung der Realisierung $t = T(X_1, ..., X_n)$ der Testgröße anhand der konkreten Stichprobe $(x_1, ..., x_n)$
7. Entscheidungsregel:
   - $t \in K$: H_0 ablehnen, damit H_1 nachgewiesen

Testtheorie

Testentscheidung mit p-Werten

Beim Einsatz von Statistiksoftware zum Prüfen von Hypothesen werden unsere üblichen Schritte – insbesondere der kritische Wert – nicht angezeigt. Statt dessen wird der konkrete Wert der Teststatistik und der zugehörige **p-Wert** (engl. p–value) oder die sog. **Signifikanz** ausgegeben.

Die Testentscheidung lautet dann: H_0 ablehnen, falls der p-value kleiner oder gleich dem vorgegebenem Signifikanzniveau α ist, ansonsten H_0 nicht ablehnen.

Zweiseitiger approximativer Test auf den Anteilswert

- X Bernoulli-Variable mit $\pi = P(X = 1)$.
- Zweiseitige Hypothese über den Anteilswert p

$$\begin{aligned} H_0 &: \quad \pi = \pi_0 \\ H_1 &: \quad \pi \neq \pi_0 \end{aligned}$$

- Testgröße: Anteil in der Stichprobe $X_1, \ldots, X_n$

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

- Stichprobenumfang n ist genügend groß

Zweiseitiger approximativer Test auf den Anteilswert

Hypothesen: $H_0 : \pi = \pi_0$ versus $H_1 : \pi \neq \pi_0$

Testentscheidung zum Signifikanzniveau α

Annahmebereich

$$\pi_0 \pm z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\pi_0(1-\pi_0)}{n}}$$

H_0 wird abgelehnt, falls

$$\hat{\pi} < \pi_0 - z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\pi_0(1-\pi_0)}{n}}$$

oder

$$\hat{\pi} > \pi_0 + z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\pi_0(1-\pi_0)}{n}}$$

Beispiel: Münzwurf bei Stern TV 2002

- Nullhypothese: $\pi = \pi_0 = 0.5$ („Münze ist fair.")
- Signifikanzniveau: $\alpha = 0.01$
- $n = 800$ Münzwürfe
  $\rightarrow$ Normalverteilung
- Annahmebereich

$$\pi_0 \pm z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\pi_0(1-\pi_0)}{n}} = 0.5 \pm z_{1-\frac{0.01}{2}} \cdot \sqrt{\frac{0.5(1-0.5)}{800}}$$
$$= 0.5 \pm 0.046$$

- H_0 wird beibehalten, falls: $\hat{\pi} \in [0.454; 0.546]$

Wert bei Stern TV (2002) : 501/800

Dualität Test und Konfidenzintervall

- Annahmebereich: Wir behalten H_0 bei, falls die Testgröße $\hat{\pi}$ in der Nähe von π_0 liegt:
- Äquivalente Formulierung über ein Konfidenzintervall: Wir behalten H_0 bei, falls π_0 in der Nähe der Testgröße liegt
- Wir behalten H_0 bei, falls π_0 im Konfidenzintervall für den Anteil liegt
- Dabei hängen das Konfindenzniveau 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 Test auf den Anteilswert

- X Bernoulli-Variable mit $\pi = P(X = 1)$.
- Einseitige Hypothese über den Anteilswert π

$$\begin{aligned} H_0 &: \quad \pi \leq \pi_0 \\ H_1 &: \quad \pi > \pi_0 \end{aligned}$$

- Testgröße: Anteil in der Stichprobe $X_1, \ldots, X_n$

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

- 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

Annahmebereich

$$\hat{\pi} \quad \leq \quad \pi_0 + z_{1-\alpha} \cdot \sqrt{\frac{\pi_0(1-\pi_0)}{n}}$$

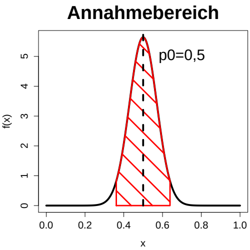
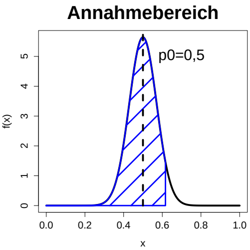
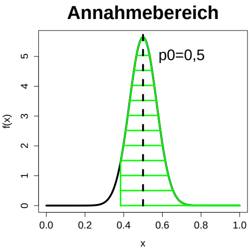
H_0 wird abgelehnt, falls

$$\hat{\pi} \quad > \quad \pi_0 + z_{1-\alpha} \cdot \sqrt{\frac{\pi_0(1-\pi_0)}{n}}$$

$z_{1-\alpha}$ 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 $\alpha = 0.05$

$H_0 : \pi = 0.5$ $H_1 : \pi \neq 0.5$	$H_0 : \pi \leq 0.5$ $H_1 : \pi > 0.5$	$H_0 : \pi \geq 0.5$ $H_1 : \pi < 0.5$
		
$[0.36; 0.64]$	$[0; 0.62]$	$[0.38; 1]$

Zusammenfassung

- Signifikanztest weiteres zentrales Instrument der statistischen Inferenz
- Konstruktion über Nullhypothese (i.d.R. Gegenteil der Forschungshypothese)
- Statistische Signifikanz entspricht Falsifizierung der Nullhypothese
- Enger Zusammenhang mit Konfidenzintervallen

Konstruktion von statistischen Tests

1. Forschungshypothese
2. Operationalisierung über die zu beobachtende Zufallsvariable X und deren Parameter
3. Formulierung von H_0 typischerweise als Gegenteil der Forschungshypothese und H_1
4. Konstruktion bzw. Wahl einer geeigneten Testgröße $T(X) = T(X_1, ..., X_n)$ als Funktion der erhobenen Daten. Die Testgröße beinhaltet die Information der Daten bezüglich H_0.
5. Aus der Verteilung von $T(X)$ unter der Nullhypothese erhält man Ablehnbereich bzw. p-Wert
6. Entscheidungsregel: H_0 ablehnen, falls Testgröße im Ablehnbereich bzw. p–Wert $\leq \alpha$

Typen von Tests

- Ein–Stichproben–Fall vs. Zwei– oder Mehr–Stichproben–Fall
- Parametrisch vs. Non-Parametrisch
- Lageparameter, Verteilungen, andere Parameter

Test auf den Erwartungswert

- Wir interessieren uns für den Erwartungswert μ einer metrischen Zufallsgröße.
  Beispiele: Alter, Einkommen, Körpergröße, Scorewert, ...
- Wir können einseitige oder zweiseitige Hypothesen formulieren.
- Beispiele
  - Der Mittelwert der Länge eines Teils in der Produktion liegt bei 12.50 cm.
  - Der Blutdruck einer Person wird durch eine Intervention niedriger.

Zweiseitiger Gauss-Test auf den Erwartungswert μ

Voraussetzung: Stichprobenumfang n genügend groß (Faustregel $n > 30$)

2. X Zufallsgröße mit Erwartungwert μ.
3. Hypothese über μ:

$$\begin{aligned} H_0 &: \quad \mu = \mu_0 \\ H_1 &: \quad \mu \neq \mu_0 \end{aligned}$$

4. Testgröße: Normierter Mittelwert in der Stichprobe $X_1, \ldots, X_n$.

$$\begin{aligned} T &:= \frac{\bar{X} - \mu_0}{S}\sqrt{n} \\ S^2 &:= \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 \end{aligned}$$

Bezeichnung für T: t-Wert oder z-Wert

Zweiseitiger Gauss-Test auf den Erwartungswert μ

5. (Approximative) Verteilung von T unter H_0

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

6. Testentscheidung:

$$p - Wert = 2 \cdot [1 - \Phi(|T|)] = 2 \cdot \left[1 - \Phi\left(|\overline{X} - \mu_0| / \sqrt{\frac{S^2}{n}}\right)\right]$$

Φ ist die Verteilungsfunktion der Standardnormalverteilung
Ablehnung für

$$|T| > z_{1-\alpha/2}$$

$z_{1-\alpha/2}$ ist das (1-$\alpha/2$) - Quantil der Standardnormalverteilung

Einseitiger Gauss-Test auf den Erwartungswert μ

Voraussetzung: Stichprobenumfang n genügend groß (Faustregel $n > 30$)

2. X Zufallsgröße mit Erwartungwert μ.
3. Hypothese über μ:

$$\begin{aligned} H_0 &: \quad \mu \leq \mu_0 \\ H_1 &: \quad \mu > \mu_0 \end{aligned}$$

4. Testgröße: Normierter Mittelwert in der Stichprobe $X_1, \ldots, X_n$.

$$\begin{aligned} T &:= \frac{\bar{X} - \mu_0}{S} \sqrt{n} \\ S^2 &:= \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2 \end{aligned}$$

Bezeichnung für T: t-Wert oder z-Wert

Einseitiger Gauss-Test auf den Erwartungswert μ

5. (Approximative) Verteilung von T unter H_0

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

6. Testentscheidung :

$$p - Wert = [1 - \Phi(T)] = \left[1 - \Phi\left((\overline{X} - \mu_0)/\sqrt{\frac{S^2}{n}}\right)\right]$$

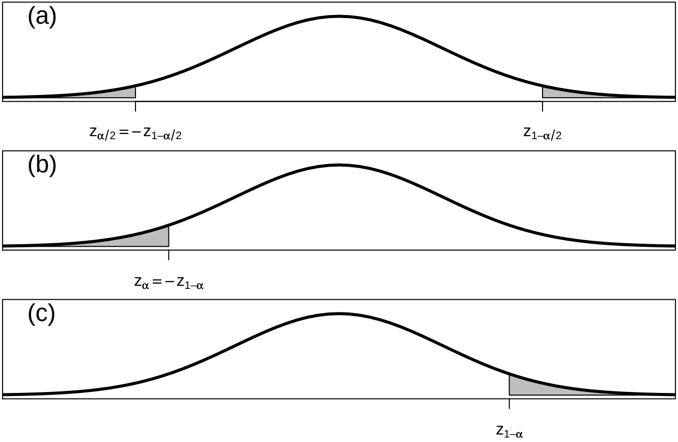
Φ ist die Verteilungsfunktion der Standardnormalverteilung
Ablehnung für

$$T > z_{1-\alpha}$$

$z_{1-\alpha}$ ist das (1-α) - Quantil der Standardnormalverteilung

Ablehnbereich einfacher Gauss-Test

Graphisch dargestellt liegt der kritische Bereich für die unterschiedlichen Fälle an den markierten Enden:



t–Test

Wird bei kleineren Stichproben verwendet.
Voraussetzung: X annähernd normalverteilt

1. X Zufallsgröße mit Erwartungwert μ.
2. Hypothese über μ:

$$\begin{aligned} H_0 &: \quad \mu = \mu_0 \\ H_1 &: \quad \mu \neq \mu_0 \end{aligned}$$

3. Testgröße: Normierter Mittelwert in der Stichprobe $X_1, \ldots, X_n$.

$$\begin{aligned} T &:= \frac{\bar{X} - \mu_0}{S}\sqrt{n} \\ S^2 &:= \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 \end{aligned}$$

Bezeichnung für T: t-Wert

t–Test

4. Verteilung von T unter H_0

$$T \sim t^{n-1}$$

   t-Verteilung mit n-1 Freiheitsgeraden

5. Testentscheidung :

$$p - Wert = 2 \cdot [1 - F_{t;n-1}(|T|)]$$

   $F_{t;n-1}$ ist die Verteilungsfunktion der t-Verteilung mit n-1 Freiheitsgeraden Ablehnung für

$$|T| > t^{n-1}_{1-\alpha/2}$$

   $t^{n-1}_{1-\alpha/2}$ ist das (1-α/2) - Quantil der t^{n-1} - Verteilung

Veränderung des Blutdrucks nach einer Intervention

- Nullhypothese: Die Blutdruckdifferenz ist 0.

$$H_0 \quad : \quad \mu = 0$$
$$H_1 \quad : \quad \mu \neq 0$$

- Testgröße: Durchschnittliche Bluddruckdifferenz
- n= 22 $\longrightarrow$ zweiseitiger t -Test

Ergebnisse mit R

data: bdd
t = -1.8237, df = 21, p-value = 0.08246
Alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
-5.9034075 0.3870439
sample estimates:
mean of x
-2.758182

Vorzeichentest

Non-Parametrischer Test zur Lage einer Verteilung

2. Betrachtet wird der Median einer Verteilung von beliebiger Struktur
3. $H_0 : x_{med} = \delta_0$
   $H_1 : x_{med} \neq \delta_0$
4. T $=$ Anzahl der Werte $< \delta_0$
5. $T \sim B(n; 0.5)$
6. Testentscheidung

$$p - Wert = min(2 \cdot (1 - F_{B(n;0.5)}(T-1); 2 \cdot (F_{B(n;0.5)}(T))$$

$F_{B(n;0.5)}$: Verteilungsfunktion der Binomialverteilung

χ^2-Anpassungstest

Motivation

Wir wollen prüfen, ob eine Zufallsgröße einer bestimmten Verteilung genügt.
Beispiel: Der Würfel ist fair (alle Zahlen habe die Wahrscheinlichkeit 1/6)
Die Testgröße wird so konstruiert, dass sie die Abweichungen der unter H_0 erwarteten von den tatsächlich beobachteten absoluten Häufigkeiten misst.

Der Test wird zunächst für kategoriale Größen definiert. Bei stetigem Größen kann der Test angewendet werden, wenn die Stichprobe $\mathbf{X}$ in k (oft willkürlich gewählten) Klassen eingeteilt wird..

χ^2-Anpassungstest

1. Die diskrete Zufallsgröße X mit möglichen Werten 1,...,k hat eine bestimmte Verteilung $F_0(x)$
2. $H_0 : P(X = i) = \pi_i$
   $H_1 : P(X = i) \neq \pi_i$ für mindestens ein i
3. Konstruktion der Testgröße

$$T(\mathbf{X}) = \sum_{i=1}^{k} \frac{(N_i - n\pi_i)^2}{n\pi_i}$$

wobei

- N_i die absolute Häufigkeit der Stichprobe $\mathbf{X}$ für die i-te Klasse angibt
- π_i die Wahrscheinlichkeit, dass X in die Klasse i fällt
- n die Größe der Stichprobe beinhaltet.

χ^2-Anpassungstest

4. Verteilung der Testgröße

$$T_{H_0} \sim \chi^2_{k-1}$$

Die χ^2-Verteilung gilt nur asymptotisch und ist zumeist hinreichend genau, wenn höchstens 1/5 der erwarteten Klassenbesetzungen $n\pi_i$ kleiner als 5 und alle $n\pi_i$ größer als 1 sind.

5. Testentscheidung
Kritischer Bereich: $K = (c_{k-1;1-\alpha}; \infty)$
$c_{k-1;1-\alpha}$ ist das (1-α) - Quantil der χ^2_{k-1} - Verteilung

Approximativer Test auf Erwartungswert–Differenz bei unabhängigen Stichproben

1. X und Y sind zwei Größen mit Erwartungswerten μ_X und μ_Y
2. $X_1, \ldots, X_{n_X}$ und $Y_1, \ldots, Y_{n_Y}$ **unabhängige** Stichproben
3. $H_0 : \mu_X = \mu_Y$
   $H_1 : \mu_X \neq \mu_Y$
4. Testgröße: standardisierte Differenz der Mittelwerte

$$T = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{s_X^2}{n_X} + \frac{s_Y^2}{n_Y}}}$$

5. $T \sim N(0, 1)$ bei großen Stichprobenumfängen
   (Faustregel: Stichprobenumfänge $n_X, n_Y > 30$)

Approximativer Test auf Erwartungswert–Differenz bei unabhängigen Stichproben

6. Testentscheidung:

$$p - Wert = 2 \cdot [1 - \Phi(|T|)]$$

   Φ ist die Verteilungsfunktion der Standardnormalverteilung
   Ablehnung für

$$|T| > z_{1-\alpha/2}$$

   $z_{1-\alpha/2}$ ist das (1-α/2) - Quantil der Standardnormalverteilung

Die entsprechenden einseitigen Tests sind analog zum approximativen Gauss-Test (verwende $1 - \alpha$ Quantile)

Beispiel: Radio-Hördauer Ost-West

- Hören Personen in den alten Bundesländern im Schnitt mehr Radio?
  X : Hördauer im den alten Bundesländern,
  Y : Hördauer in den neuen Bundesländern

$$\begin{aligned} H_0 &: \quad \mu_X - \mu_Y \leq 0 \\ H_1 &: \quad \mu_X - \mu_Y > 0 \end{aligned}$$

- Befragung unter 253 Personen aus den alten Bundesländern und 932 Personen aus den neuen Bundesländern
  - unverbundene Stichproben $X_1, \ldots, X_{253}$ und $Y_1, \ldots, Y_{932}$
  - Stichprobengrößen $n_X = 253, n_Y = 932 > 30$
- Durchschnittliche Hördauer:
  11.4 h (Standardabweichung 8.4 h) in den alten Bundesländern
  9.5 h (Standardabweichung 8.4 h) in den neuen Bundesländern

Beispiel: Radio-Hördauer Ost-West

- Signifikanzniveau: $\alpha = 0.1$
- Differenz der Radio-Hördauer

$$\overline{X} - \overline{Y} = 11.4 - 9.5 = 1.9$$

- Testgröße

$$T = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{s_X^2}{n_X} + \frac{s_Y^2}{n_Y}}} = 1.9/0.65 = 2.9$$

- p–Wert : 0.001865813
- H_0 wird abgelehnt, Personen aus den alten Bundesländern hören signifikant länger Radio.

Doppelter t-Test auf die Erwartungswertdifferenz bei unabhängigen Stichproben

1. Vergleich zweier Mittelwerte
2. X und Y sind zwei Größen mit Erwartungswerten μ_X und μ_Y
   X und Y sind **normalverteilt**.
3. $H_0 : \mu_X = \mu_Y$
   $H_1 : \mu_X \neq \mu_Y$
4. Testgröße: Standardisierte Differenz der Mittelwerte

$$T = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{s_X^2}{n_X} + \frac{s_Y^2}{n_Y}}}$$

Doppelter t-Test auf die Erwartungswertdifferenz bei unabhängigen Stichproben

5

$$p - Wert = 2 \cdot [1 - F_{t;k}(|T|)]$$

$F_{t;k}$ ist die Verteilungsfunktion der t-Verteilung mit k Freiheitsgeraden

$$k = \frac{\left(\frac{s_X^2}{n_X} + \frac{s_Y^2}{n_Y}\right)^2}{\frac{1}{n_X - 1} \cdot (\frac{s_X^2}{n_X})^2 + \frac{1}{n_Y - 1} \cdot (\frac{s_Y^2}{n_Y})^2}$$

Ablehnung für

$$|T| > t^k_{1-\alpha/2}$$

$t^k_{1-\alpha/2}$ ist das (1-$\alpha/2$) - Quantil der t^k - Verteilung

Tests auf Erwartungswertdifferenz bei abhängigen Stichproben

1. Gegeben ist eine verbundene Stichprobe $X_1, \ldots, X_n$ und $Y_1, \ldots, Y_n$
2. Bilde die Differenz

$$D_i = X_i - Y_i \quad i = 1, \ldots, n$$

3. Berechne Standardabweichung der Differenz

$$s_d = \sqrt{\frac{1}{n-1} \cdot \sum_{i=1}^{n} (d_i - \bar{d})^2}$$

4. Führe einen Test auf den Erwartungswert von D durch
   - $n > 30 \longrightarrow$ Gauß-Test
   - D normalverteilt $\longrightarrow$ t-Test

Der Wilcoxon Test für unabhängige Stichproben

Test ist identisch mit dem Mann-Whitney-U-Test

1. Unterschied in der Lage zweier Verteilungen
2. X und Y sind zwei Größen mit Medianen med_X und med_Y
3. $H_0 : \text{med}_X = \ med_Y$ vs. $H_1 : \text{med}_X \neq \text{med}_Y$
4. Testgröße Gegeben zwei unabhängige Stichproben $X_1, \ldots, X_n$ und $Y_1, \ldots, Y_m$
   Grundidee: Betrachte die Ränge aus allen Beobachtungen X_i und Y_j und bezeichne diese mit $\text{rg}(X_i)$ und $\text{rg}(Y_j)$, z.B.
   $X_1 = 3,\ X_2 = 5,\ Y_1 = 6,\ Y_2 = 1,\ Y_3 = 4\ \Rightarrow$
   $\text{rg}(X_1) = 2$, $\text{rg}(X_2) = 4$, $\text{rg}(Y_1) = 5$, $\text{rg}(Y_2) = 1$, $\text{rg}(Y_3) = 3$

$$T = \sum_{i=1}^{n} \text{rg}(X_i)$$

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 für große und kleine T.

χ^2-Unabhängigkeitstest

1. Sind zwei kategoriale Zufallsgrößen unabhängig? Unterscheiden sich zwei Anteile?
2. Zwei Zufallsgrößen X und Y mit k bzw. l Ausprägungen

$$p_{ij} = P(X = i, Y = j)$$
$$p_{i\bullet} = P(X = i) \quad p_{\bullet j} = P(Y = j)$$

3. Hypothesen:

$$\begin{aligned} H_0 &: \quad \text{X und Y sind stochastisch unabhängig} \\ & \quad p_{ij} = p_{i\bullet} \cdot p_{\bullet j} \text{ für alle } i = 1, \ldots, k,\ j = 1, \ldots, l \\ H_1 &: \quad \text{X und Y sind stochastisch abhängig} \\ & \quad p_{ij} \neq p_{i\bullet} \cdot p_{\bullet j} \text{ für mindestens eine ij-Kombination} \end{aligned}$$

χ^2-Unabhängigkeitstest

4. Prüfgröße:

$$\chi^2 = \sum_{i=1}^{k} \sum_{j=1}^{l} \frac{(n_{ij} - e_{ij})^2}{e_{ij}}$$

5. Verteilung:

$$\chi^2 \sim \chi^2_{(k-1)(l-1)}$$

Annahmebereich

$$\chi^2 \leq c_{1-\alpha,\,(k-1)(l-1)}$$

Dabei ist $c_{1-\alpha,\,(k-1)(l-1)}$ das

- $(1-\alpha)$-Quantil der χ^2-Verteilung
- mit $(k-1)\cdot(l-1)$ Freiheitsgraden.

Beispiel: χ^2-Unabhängigkeitstest

$$e_{ij} = \frac{n_{i\bullet} n_{\bullet j}}{n}$$

Geschlecht \ Fußball	ja	nein	Summe
m	87	10	97
w	23	31	54
Summe	110	41	151

Erwartete Besetzungszahlen bei Unabhängigkeit

	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)	$\frac{54 \cdot 110}{151} \approx 39$	$\frac{54 \cdot 41}{151} \approx 15$

$$\begin{aligned}
\chi^2 &= \sum_{i=1}^{k} \sum_{j=1}^{l} \frac{(n_{ij} - e_{ij})^2}{e_{ij}} \\
&\approx \frac{(87-71)^2}{71} + \frac{(10-26)^2}{26} + \frac{(23-39)^2}{39} + \frac{(31-15)^2}{15} \\
&\approx 37.09
\end{aligned}$$

Beispiel: χ^2-Unabhängigkeitstest

- Signifikanzniveau: $\alpha = 0.01$
- Überprüfung mit Faustregel:
  Erwartete Besetzungszahlen $e_{ij} \geq 5 \quad \checkmark$
- 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

Unabhängigkeit und Differenz von Anteilen

Die beide Fragen:

- Gibt es Unterschiede in den Anteilen von $Y = 1$ zwischen zwei Gruppen?
- Gibt es einen Zusammenhang zwischen Gruppen–Zugehörigkeit und einem binären Merkmal Y?

sind äquivalent.

Differenz von Anteilen bei abhängigen Stichproben

Voraussetzungen:

- X und Y sind zwei Bernoulli-Größen mit

$$\begin{aligned} p_X &= P(X = 1) \\ p_Y &= P(Y = 1) \end{aligned}$$

- $(X_1, Y_1), \ldots, (X_n, Y_n)$ **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}

Hier kann der χ^2 -Unahängigkeitstest angewendet werden
Für kleine Stichproben: Fisher-Test

Zusammenfassung

- Konstruktion von statistischen Tests verläuft nach einfachen Prinzipien
- Hervorragende Übersicht und Darstellung in Fahrmeier et al. (2016)
- Viele weitere Tests vorhanden
- Immer Angabe von Schätzern und Konfidenzintervallen (nicht nur p–Werte!)

Lineare Regressionsmodelle

Deskriptive Statistik:

Gegeben Datenpunkte (Y_i, X_i) schätze die beste Gerade

$$Y_i = \beta_0 + \beta_1 X_i,\ i = 1, \ldots, n.$$

(mit der Methode der kleinsten Quadrate)

Statistisches Modell

- Linearer Zusammenhang
- Im Folgenden:
  Probabilistische Modelle in Analogie zu den deskriptiven Modellen aus Statistik I

Lineare Einfachregression

Zunächst Modelle mit nur *einer* unabhängigen Variable.

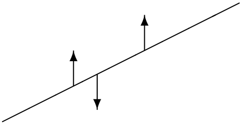
Statistische Sichtweise:

- Modell

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

β_1 Wirkung der Änderung von X_i um eine Einheit auf Y

- gestört durch zufällige Fehler ε_i



Modellannahmen

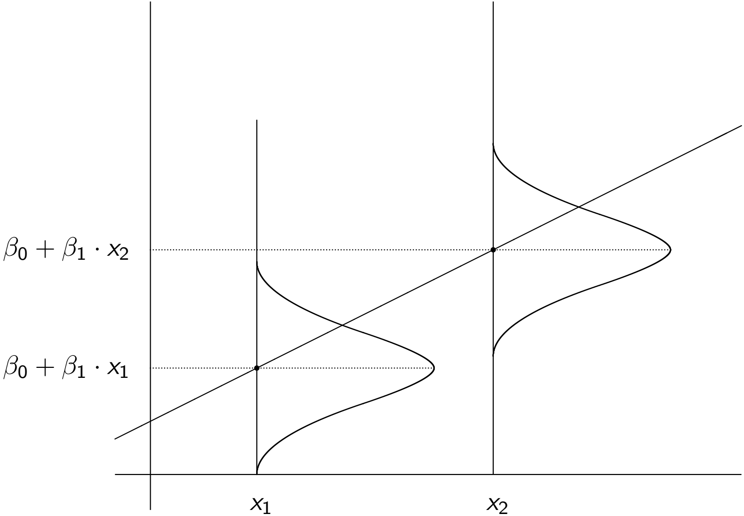
Beobachtung von Datenpaaren (X_i, Y_i), $i = 1, \ldots, n$ mit

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i,$$

wobei sich die Annahmen auf den zufälligen Störterm beziehen:

- $E(\varepsilon_i) = 0$
- $Var(\varepsilon_i) = \sigma^2$ 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)

Lineare Einfachregression



Schätzung der Parameter

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 entsprechen den KQ-Schätzer aus Statistik 1

$$
\begin{aligned}
\hat{\beta}_1 &= \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n} (X_i - \bar{X})^2}, \\
\hat{\beta}_0 &= \bar{Y} - \hat{\beta}_1 \bar{X}, \\
\hat{\sigma}^2 &= \frac{1}{n-2} \sum_{i=1}^{n} \hat{\varepsilon}_i^2
\end{aligned}
$$

mit den Residuen

$$\hat{\varepsilon}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i.$$

Konstruktion von Testgrößen

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

$$\frac{\hat{\beta}_1 - \beta_1}{\hat{\sigma}_{\hat{\beta}_1}} \sim t^{(n-2)}.$$

Konfidenzintervalle

- $\hat{\beta}_0$ 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, das bei den t- Tests verwendet wurde

  Konfidenzintervalle zum Sicherheitsgrad γ:

$$\begin{array}{ll} \text{für } \beta_0 \ : & [\hat{\beta}_0 \pm \hat{\sigma}_{\hat{\beta}_0} \cdot t^{(n-2)}_{\frac{1+\gamma}{2}}] \\ \text{für } \beta_1 \ : & [\hat{\beta}_1 \pm \hat{\sigma}_{\hat{\beta}_1} \cdot t^{(n-2)}_{\frac{1+\gamma}{2}}] \end{array}$$

Tests für die Parameter des Modells

Mit der Teststatistik

$$T_{\beta_1^*} = \frac{\hat{\beta}_1 - \beta_1^*}{\hat{\sigma}_{\hat{\beta}_1}}$$

ergibt sich

	Hypothesen			kritische Region
I.	$H_0 : \beta_1 \leq \beta_1^*$	gegen	$\beta_1 > \beta_1^*$	$T \geq t_{1-\alpha}^{(n-2)}$
II.	$H_0 : \beta_1 \geq \beta_1^*$	gegen	$\beta_1 < \beta_1^*$	$T \leq t_{1-\alpha}^{(n-2)}$
III.	$H_0 : \beta_1 = \beta_1^*$	gegen	$\beta_1 \neq \beta_1^*$	$\lvert T \rvert \geq t_{1-\frac{\alpha}{2}}^{(n-2)}$

(analog für $\hat{\beta}_0$).

Von besonderem Interesse ist der Fall $\beta_1^* = 0$ (Steigung gleich 0): Hiermit kann man überprüfen, ob die $X_1, \ldots, X_n$ einen signifikanten Einfluss hat oder nicht.

Beispiel : Mietspiegel

Call:
lm(formula = nmqm   wfl, data = mietsp2015)
Coefficients:

	Estimate	Std. Error	t value	$Pr(>\|t\|)$
(Intercept)	11.72	0.46	26.286	$< 2e-16$
wfl	-0.0226	0.005787	-3.905	< 0.00012

Multiples Regressionsmodell

Beispiel: Mietspiegel

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i$$

mit

$$\begin{aligned} X_1 &= \begin{cases} 1 & \text{Gute Lage} \\ 0 & \text{schlechte Lage} \end{cases} \\ X_2 &= \text{Wohnfläche} \\ Y &= \text{Miete} \end{aligned}$$

Multiples Regressionsmodell: Interpretation

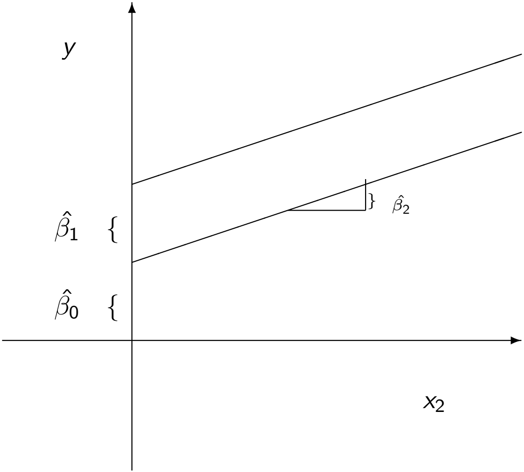
- Geschätzte Regressionsgerade für gute Lage

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 \cdot 1 + \hat{\beta}_2 \cdot x_{2i}$$

- Geschätzte Regressionsgerade für die schlechte Lage :

$$\begin{aligned} \hat{y}_i &= \hat{\beta}_0 + \hat{\beta}_1 \cdot 0 + \hat{\beta}_2 \cdot x_{2i} \\ &= \hat{\beta}_0 + \hat{\beta}_2 \cdot x_{2i} \end{aligned}$$

Grundidee (ANCOVA)



Mehr als 2 Gruppen

Lösungsansatz

Hier ist eine direkte Lösung nicht sinnvoll.
Grundidee:

- aus einem nominalen Regressor mit k Merkmalsausprägungen
- $k-1$ neue Regressoren (Dummys) gebildet werden.
- Eine Merkmalsausprägung des ursprünglichen Regressors wird zur *Referenzkategorie*.

Nominale Regressoren

Dummykodierung

Nach Wahl der Referenzkategorie $j \in \{1, ..., k\}$ ergeben sich die Dummys $X_i, i = 1, ..., k$ und $i \neq j$ mit folgenden Werten:

$$x_i = \begin{cases} 1 & \text{falls Kategorie } i \text{ vorliegt,} \\ 0 & \text{sonst.} \end{cases}$$

Nominale Regressoren

Beispiel

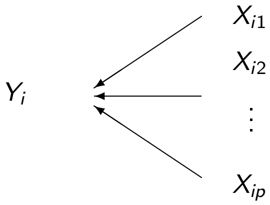
Gegeben seien folgende Daten:

lfd Nr.	Alter	Studienfach
1	19	BWL
2	22	Sonstige
3	20	VWL
$\vdots$	$\vdots$	$\vdots$

Mit der Kodierung BWL = 1, VWL = 2, Sonstige = 3 erhalten wir bei Wahl der Referenzkategorie = 3 (Sonstige) zwei Dummys X_1 (für BWL) und X_2 (für VWL) gemäß folgendem Schema:

Ausprägung von X	Wert von X_1	Wert von X_2
1 BWL	1	0
2 VWL	0	1
3 Sonstige	0	0

Multiples Regressionsmodell



abhängige Variable	unabhängige Variablen
metrisch/quasistetig	metrische/quasistetige oder dichotome (0/1) Variablen (kategoriale Variablen mit mehr Kategorien $\rightarrow$ Dummy-Kodierung)

Multiple lineare Regression

- Analoger Modellierungsansatz, aber mit mehreren erklärenden Variablen:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_p X_{ip} + \varepsilon_i$$

- Schätzung von $\beta_0, \beta_1, \ldots, \beta_p$ und σ^2 sinnvollerweise über Matrixrechnung bzw. Software.

  Aus dem R-Output sind $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_p$ sowie $\hat{\sigma}_{\hat{\beta}_0}, \hat{\sigma}_{\hat{\beta}_1}, \ldots, \hat{\sigma}_{\hat{\beta}_p}$ ablesbar.

Schätzung im multiplen Modell

- Darstellung in Matrix-Form
- KQ- Methode und Maximum-Likelihood - Methode stimmen überein
- Berechnung effizient mit Matrix-Kalkül
- Zu den Parametern können jeweils die Standardfehler geschätzt werden.

Multiple lineare Regression

Es gilt für jedes $j = 0, \ldots, p$

$$\frac{\hat{\beta}_j - \beta_j}{\hat{\sigma}_{\hat{\beta}_j}} \sim t^{(n-p-1)}$$

und man erhält wieder Konfidenzintervalle für β_j:

$$[\hat{\beta}_j \pm \hat{\sigma}_{\hat{\beta}_j} \cdot t^{(n-p-1)}_{\frac{1+\gamma}{2}}]$$

sowie entsprechende Tests.

Multiple lineare Regression: Tests

Von besonderem Interesse ist wieder der Test

$$H_0 : \beta_j = 0, \;\; H_1 : \beta_j \neq 0.$$

Der zugehörige p-Wert findet sich im Ausdruck (Vorsicht mit Problematik des multiplen Testens!).
Man kann auch simultan testen, z.B.

$$\beta_1 = \beta_2 = \ldots = \beta_p = 0.$$

Dies führt zu einem sogenannten F-Test ($\longrightarrow$ Software).

Sind alle X_{ij} 0/1-wertig, so erhält man eine sogenannte *Varianzanalyse*, was dem Vergleich von mehreren Mittelwerten entspricht.

Varianzanalyse (Analysis of Variance, ANOVA)

- 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).

Varianzanalyse: Beispiel

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}

$$\begin{array}{ll} j = 1, \ldots, J & \text{Faktorstufen} \\ i = 1, \ldots, n_j & \text{Personenindex in der } j\text{-ten Faktorstufe} \end{array}$$

Modellformulierung

Modell (Referenzcodierung):

$$Y_{ij} = \mu_J + \beta_j + \varepsilon_{ij} \qquad j = 1, \ldots, J, \; i = 1, \ldots, n_j,$$

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$(0, \sigma^2)$, $\quad \varepsilon_{11}, \varepsilon_{12}, \ldots, \varepsilon_{Jn_J}$ unabhängig.

Testproblem:

$$H_0 \; : \; \beta_1 = \beta_2 = \ldots \beta_{j-1} = 0$$

gegen

$$H_1 \; : \; \beta_j \neq 0 \quad \text{für mindestens ein j}$$

Streuungszerlegung

Mittelwerte:

$\bar{Y}_{\bullet\bullet}$	Gesamtmittelwert in der Stichprobe
$\bar{Y}_{\bullet j}$	Mittelwert in der j-ten Faktorstufe

Es gilt (vgl. Statistik I) die Streuungszerlegung:

$$\sum_{j=1}^{J}\sum_{j=1}^{n_j}(Y_{ij}-\bar{Y}_{\bullet\bullet})^2 = \sum_{j=1}^{J}\underbrace{n_j(\bar{Y}_{\bullet j}-\bar{Y}_{\bullet\bullet})^2}_{\substack{=SQE \\ \text{Variabilität \textbf{der} Gruppen}}} + \underbrace{\sum_{j=1}^{J}\sum_{i=1}^{n_j}(Y_{ij}-\bar{Y}_{\bullet j})^2}_{\substack{=SQR \\ \text{Variabilität \textbf{in den} Gruppen}}}$$

F-Test

Die **Testgröße**

$$F = \frac{SQE/(J-1)}{SQR/(n-J)}$$

ist geeignet zum Testen der Hypothesen

$$H_0 \quad : \quad \beta_1 = \beta_2 = \ldots \beta_{j-1} = 0$$

gegen

$$H_1 \quad : \quad \beta_j \neq 0 \quad \text{für mindestens ein j}$$

Testprozedur

- **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.

  $\Rightarrow$ Testen spezifischer Hypothesen über die Effekte β_j. Dabei tritt allerdings die Problematik des multiplen Testens auf.

Zusammenfassung

- Testen von Regressionsmodellen wesentliches Werkzeug
- Gleichzeitige Berücksichtigung vieler Einflüsse möglich
- Viel Möglichkeiten zum Testen (F-Tests)
- Regressionsmodell Ausgangspunkt für viele neue Verfahren (Big Data, Algorithmen, KI)

Bayes- Schätzung und Bayes-Lernen

Subjektiver Wahrscheinlichkeitsbegriff

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.

Beispiel: Trifft Müller beim Elfmeter?

3 Theorien :

UH	immer p =1.0
AR	p= 0.8
JL	p=0.7

Ansatz: Theorien habe gleiche Wahrscheinlichkeit

$$P(UH) = P(AR) = P(IL) = 1/3$$

Daten : Ein Treffer $X_1 = 1$

Berechnung der Wahrscheinlichkeit der Beobachtung unter den verschiedenen Theorien (Likelihood)

$$\begin{aligned} P(X_1 = 1|UH) &= 1 \\ P(X_1 = 1|AR) &= 0.8 \\ P(X_1 = 1|JL) &= 0.7 \end{aligned}$$

ML-Prinzip: Theorie UH stimmt

Satz von Bayes

Bilden die A_i eine vollständige Zerlegung von Ω und ist B irgendein Ereignis, so gilt:

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

In unserem Fall entsprechen die A_j den Theorien $A_1 = UH$ etc. und das Ereignis B entspricht den Daten $X_1 = 1$

$$P(UH) = P(AR) = P(JL) = 1/3$$

$$\begin{aligned} P(X_1 = 1|UH) &= 1 \\ P(X_1 = 1|AR) &= 0.8 \\ P(X_1 = 1|JL) &= 0.7 \end{aligned}$$

Posteriori- Wahrscheinlichkeiten

Wir berechnen jetzt die Posteriori-Wahrscheinlichkeit der Theorien gegeben die Daten:

$$P(UH|X_1 = 1) = \frac{P(UH) \cdot P(X_1 = 1|UH)}{P(UH) \cdot P(X_1 = 1|UH) + P(AR) \cdot P(X_1 = 1|AR) + P(JL) \cdot P(X_1 = 1|JL)}$$

$$\begin{aligned}
P(UH|X = 1) &= \frac{\frac{1}{3} \cdot 1}{\frac{1}{3} \cdot 1 + \frac{1}{3} \cdot 0.8 + \frac{1}{3} \cdot 0.7} \\
&= \frac{1}{1 + 0.8 + 0.7} = 0.4 \\
P(AR|X = 1) &= \frac{0.8}{1 + 0.8 + 0.7} = 0.32 \\
P(JL|X = 1) &= \frac{0.7}{1 + 0.8 + 0.7} = 0.28
\end{aligned}$$

Leichte Verschiebung

Verwendung weiterer Daten

13 von 15 (TM)
Berechnung der Likelihood mit Hilfe der Binomialverteilung(i.i.d)

$$
\begin{aligned}
P(X_2 = 13|UH) &= 0 \\
P(X_2 = 13|AR) &= 0.23 \\
P(X_2 = 13|JL) &= 0.09
\end{aligned}
$$

Wir berechnen jetzt die Posteriori-Wahrscheinlichkeit der Theorien gegeben die Daten:

$$
\begin{aligned}
P(UH|X_2 = 13) &= \frac{0}{0 + 0.23 + 0.09} = 0 \\
P(AR|X_2 = 13) &= \frac{0.23}{0 + 0.23 + 0.09} = 0.72 \\
P(JL|X_2 = 13) &= \frac{0.09}{0 + 0.23 + 0.09} = 0.28
\end{aligned}
$$

Es spricht Einiges für AR

Gerd Müller

50 von 62 (GM)
Berechnung der Likelihood mit Hilfe der Binomialverteilungsannahme (i.i.d)

$$\begin{aligned} P(X_3 = 50|UH) &= 0 \\ P(X_3 = 50|AR) &= 0.126 \\ P(X_3 = 50|JL) &= 0.021 \end{aligned}$$

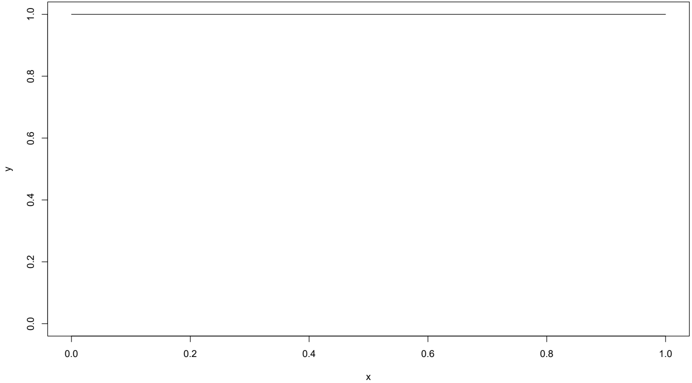
Wir berechnen jetzt die Posteriori-Wahrscheinlichkeit der Theorien gegeben die Daten:

$$\begin{aligned} P(UH|X_3 = 50) &= \frac{0}{0 + 0.126 + 0.021} = 0 \\ P(AR|X_3 = 50) &= \frac{0.126}{0 + 0.126 + 0.021} = 0.86 \\ P(JL|X_3 = 50) &= \frac{0.021}{0 + 0.126 + 0.021} = 0.14 \end{aligned}$$

Es spricht Einiges für AR

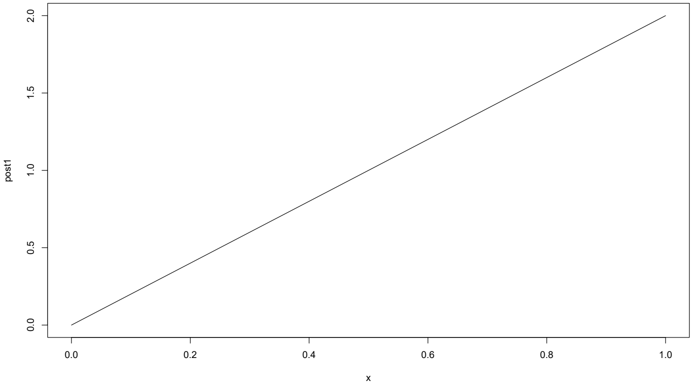
Inferenz über unbekannten Parameter

Parameter θ unbekannt
Apriori Gleichverteilung



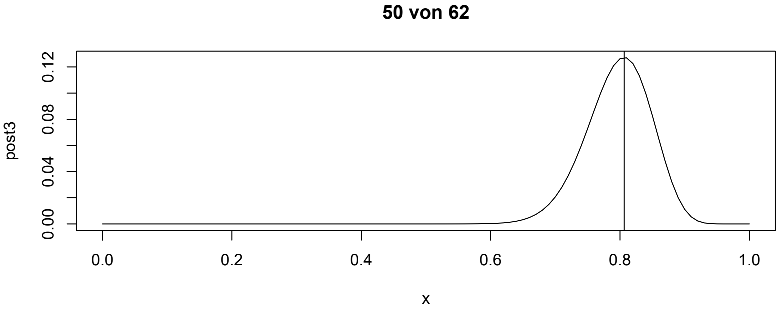
Berechnung der Posteriori–Verteilung für einen Treffer

Inferenz über unbekannten Parameter



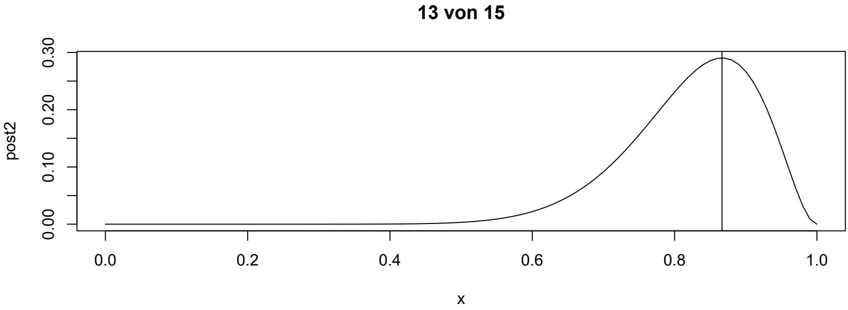
Gerd Müller

50 von 62 (GM)
Berechnung der Likelihood und der Posteriori–Verteilung mit Hilfe der Binomialverteilungsannahme (i.i.d)



Thomas Müller

13 von 15 (TM)
Berechnung der Likelihood und der Posteriori–Verteilung mit Hilfe der Binomialverteilungsannahme (i.i.d)



Berechnung der Posteriori–Verteilung

Inferenz für Parameter θ bei Beobachtung x

- $f(\theta)$ Priori - Verteilung von θ
- $f(x|\theta)$ Wahrscheinlichkeitsfunktion/Dichte von x bei Parameter θ
- $f(x)$ a priori Randverteilung von x
- $f(\theta|x)$ Posteriori - Verteilung von θ gegeben die Beobachtung x

Berechnung der Posteriori-Dichte

$$f(\theta|x) = \frac{f(x|\theta)f(\theta)}{f(x)} = \frac{f(x|\theta)f(\theta)}{\int f(x|\theta)f(\theta)d\theta}$$

Allgemeine Form

Bei Beobachtungen $x_1, \ldots x_n$ wird die gemeinsam Dichte betrachtet. Man erhält für unabhängige Beobachtungen:

$$f(x_1, \ldots, x_n|\theta) = f(x_1|\theta) \cdot f(x_2|\theta) \cdots f(x_n|\theta) = L(\theta)$$

$L(\theta)$ ist die Likelihoodfunktion

Bayes-Inferenz

Bayes–Inferenz

Die Wahrscheinlichkeits- oder Dichtefunktion von X gegeben der Parameter θ ist gegeben durch

$$f(x|\theta)$$

Die Likelihood ist

$$L(\theta) = f(x_1, \ldots, x_n|\theta)$$

Für den unbekannten Parameter θ ist die Priori-Dichte gegeben

$$f(\theta)$$

Dann gilt für die Posteriori- Dichte von θ

$$f(\theta|x_1, \ldots, x_n) = \frac{f(x_1, \ldots, x_n)|\theta)f(\theta)}{f(x)} = \frac{L(\theta)f(\theta)}{\int L(\theta)f(\theta)d\theta}$$

Bemerkungen

- Die Posteriori-Verteilung $f(\theta|x_1,\ldots,x_n)$ enthält die gesamte Information der Daten über den Parameter θ
- Die Posteriori-Verteilung hat folgende Darstellung :

$$f(\theta|x_1,\ldots,x_n) = cL(\theta)f(\theta)$$

  Dabei kann c als von θ unabhängiger Normierungsfaktor angesehen werden
- Ein zentrales Problem bei der Bayes–Inferenz ist die Wahl der Priori-Verteilung. Man wählt häufig sog. nicht-informativen Priori-Verteilungen

Binomialverteilung

$X \sim Bin(n, \theta)$

$$f(x|\theta) = \binom{n}{x} \theta^x \cdot (1-\theta)^{n-x}$$

Priori - Dichte

$$f(\theta) = 1 \text{ für } 0 \leq \theta \leq 1$$

Posteriori

$$f(\theta|x) = c \binom{n}{x} \theta^x \cdot (1-\theta)^{n-x}$$

Schätzung

Die Inferenz erfolgt mit der Posteriori-Verteilung f_{POST} Punktschätzung von θ durch

- Posteriori - Modus, d.h. Maximum von f_{POST}
- Posteriori - Erwartungswert, d.h. Erwartungswert von θ unter $f_{POST}(\theta)$

Intervallschätzung

Strategie: Finde Intervall, in dem der Parameter mit Wahrscheinlichkeit γ liegt

$$P(\theta \in [\theta_u, \theta_o]) = \gamma$$

Bezeichnung: Kredibilitätsintervalle

Beispiel: Normalverteilung (Fahrmeir et al.)

Beispiel übernommen aus:
L. Fahrmeir, Ch. Heumann, R. Künstler, I. Pigeot und G. Tutz: Statistik - Der Weg zur Datenanalyse, (8. Auflage), Springer-Verlag, 2016.

Seien $X_1, \ldots, X_n$ unabhängige Wiederholungen von $X \sim N(\mu, \sigma^2)$, wobei μ zu schätzen ist, aber nun σ^2 als bekannt angenommen wird. Als a priori Dichte für μ wählen wir eine $N(\mu_0, \sigma_0^2)$-Verteilung, also

$$f(\mu) = \frac{1}{(2\pi\sigma_0^2)^{1/2}} \exp\left\{-\frac{(\mu - \mu_0)^2}{2\sigma_0^2}\right\}.$$

Die a posteriori Dichte ergibt sich also aus:

$$f(\mu \mid x_1, \ldots, x_n) = \frac{L(\mu, \sigma) f(\mu)}{\int L(\mu, \sigma) f(\mu)\, d\mu}$$

Beispiel: Normalverteilung (Fortsetzung)

Es ergibt sich eine Normalverteilung

$$\mu \mid x_1, \ldots, x_n \sim N(\tilde{\mu}, \tilde{\sigma}^2)$$

mit a posteriori Erwartungswert

$$\tilde{\mu} = \frac{n\sigma_0^2}{n\sigma_0^2 + \sigma^2}\bar{x} + \frac{\sigma^2}{n\sigma_0^2 + \sigma^2}\mu_0$$

und a posteriori Varianz

$$\tilde{\sigma}^2 = \frac{\sigma^2}{n + \sigma^2/\sigma_0^2} \, .$$

Beispiel: Normalverteilung (Zusammenfassung)

- Für $\sigma_0^2 \to 0$ ("exaktes Vorwissen") gilt $\tilde{\mu} \to \mu_0$
- Für $\sigma_0^2 \to \infty$ ("kein Vorwissen") ergibt sich $\tilde{\mu} \to \bar{x}$, also die Maximum Likelihood-Schätzung $\hat{\mu} = \bar{x}$ aus der Stichprobe
- Analog für die Varianz

Also:

$$\mu \mid x_1, \ldots, x_n \to N\left(\bar{x}, \sigma^2/n\right) \quad \text{für } \sigma_0^2 \to \infty$$

bei nichtvorhandenem Vorwissen über μ, und

$$\mu \mid x_1, \ldots, x_n \to N(\mu_0, 0)$$

bei sicherem Vorwissen $\mu = \mu_0$.

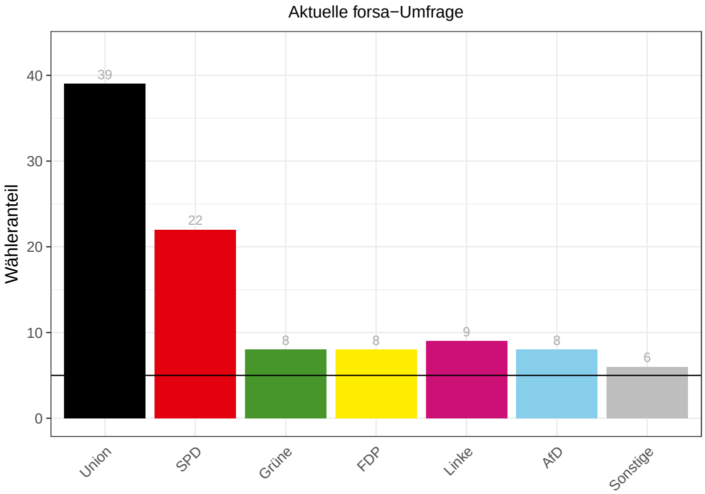
Der "Hyperparameter" σ_0^2 steuert also den Kompromiss zwischen Stichprobeninformation und subjektiver a priori Information.

Vergleich: Frequentistisch vs. Bayes–Inferenz

- Da die Likelihood wesentlich in die Berechnung eingeht gibt es teilweise ähnliche Ergebnisse, z.B Kredibilitätsintervalle mit nicht informativer Prioi sind in bestimmten Fällen identisch mit Konfidenzintervallen
- Unterschiedliche Interpretation
- Laufende wissenschaftliche Diskussion

Auswertung von Umfragen: Wahlistik

Verwende Umfrageergebnisse, z.B.



Bayes Schätzung von Wahrscheinlichkeiten