Statistische Methoden der Datenanalyse

Wintersemester 2011/2012

Albert-Ludwigs-Universität Freiburg

Prof. Markus Schumacher

Physikalisches Institut Westbau 2 OG Raum 008

Telefonnummer 07621 203 7612

E-Mail: Markus.Schumacher@physik.uni-freiburg.de

Vorlesung basiert in weiten Teilen auf der von Glen Cowan.

http://terascale.physik.uni-freiburg.de/lehre/ws_1112/statmethoden_ws1112

Kapitel 1

Grundlegende Konzepte der Wahrscheinlichkeitstheorie

Der Begriff der Wahrscheinlichkeit

In den Naturwissenschaften gibt es verschiedene Elemente der Unsicherheit:

Unsicherheit: Ausgang eines Experimentes/Messung unvorhersagbar bei Wiederholung

Gründe: die Theorie ist nicht deterministisch z. B. Quantenmechanik,

zufällige Messfehler (Messapparatur, Experimentator/Beobachter) anwesend auch ohne Quanteneffekte

Randbedingungen, die man im Prinzip kennen könnte, aber es nicht tut, z.B. durch Begrenzungen an Zeit, Geld, ...

Wir können diese Unsicherheit durch das Konzept der WAHRSCHEINLICHKEIT quantifizieren

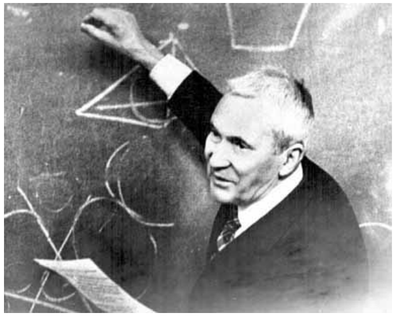
Axiomatische Definition der Wahrscheinlichkeit

Betrachte Menge *S* (Grundgesamtheit) mit Untermengen *A*, *B*, ...
Ordne jeder Teilmenge eine Zahl P zwischen 0 und 1 zu, so dass gilt:

$$\text{For all } A \subset S, P(A) \geq 0$$

$$P(S) = 1$$

$$\text{If } A \cap B = \emptyset, P(A \cup B) = P(A) + P(B)$$



Kolmogorov
Axiome (1933)

Aus diesen Axiomen können wir weitere Eigenschaften ableiten

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

$$P(A \cup \overline{A}) = 1$$

$$P(\emptyset) = 0$$

$$\text{if } A \subset B, \text{ then } P(A) \leq P(B)$$

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

Bedingte Wahrscheinlichkeit

Wir definieren weiterhin die bedingte Wahrscheinlichkeit:
A gegeben *B* (mit *P(B) ≠ 0*)

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

z.B. Würfeln: $P(n < 3\,|\,n\,\text{even}) = \frac{P((n{<}3) \cap n\,\text{even})}{P(\text{even})} = \frac{1/6}{3/6} = \frac{1}{3}$

Wenn Teilmengen *A, B* unbabhängig dann gilt $P(A \cap B) = P(A)P(B)$

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

Nicht zu verwechseln mit disjunkten Teilmengen: i.e. $A \cap B = \emptyset$

Die bedingten Wahrscheinlichkeiten erfüllen Kolomogorov-Axiome

Das Theorem von Bayes

Aus der Definition der bedingten Wahrscheinlichkeit erhalten wir

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

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

Bayes-Theorem

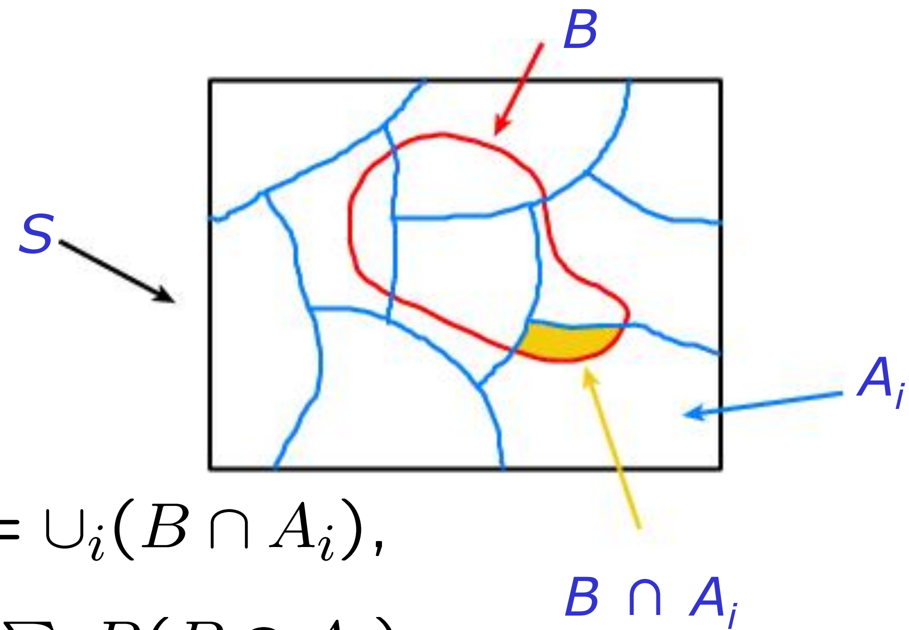
$$\boxed{P(A|B) = \frac{P(B|A)P(A)}{P(B)}}$$

Erstmals publiziert (posthum) durch den Reverend Thomas Bayes (1702−1761)

An essay towards solving a problem in the doctrine of chances, Philos. Trans. R. Soc. **53** (1763) 370; Reprint in Biometrika, **45** (1958) 293

Betrachte eine Untermenge B der Grundgesamtheit S,

Die in disjunkte Untermengen A_i zerlegt ist, so dass gilt $\cup_i A_i = S$,



→ $B = B \cap S = B \cap (\cup_i A_i) = \cup_i (B \cap A_i)$,

→ $P(B) = P(\cup_i (B \cap A_i)) = \sum_i P(B \cap A_i)$

→ $P(B) = \sum_i P(B|A_i) P(A_i)$ Gesetz der Gesamtwkt.

Bayes-Theorem wird zu

$$\boxed{P(A|B) = \frac{P(B|A)P(A)}{\sum_i P(B|A_i)P(A_i)}}$$

Ein Beispiel zur Anwendung von Bayes-Theorem

Annahme: die Wahrscheinlichkeit (für jeden) AIDS zu haben ist:

$$P(\text{AIDS}) = 0.001$$
$$P(\text{no AIDS}) = 0.999$$

$\leftarrow$ A-Priori-Wahrscheinlichkeiten, i.e. bevor ein Test durchgeführt wurde

Betrachte einen AIDS-Test: Ergebnis ist + or -

$$P(+|\text{AIDS}) = 0.98$$
$$P(-|\text{AIDS}) = 0.02$$

$\leftarrow$ Wahrscheinlichkeiten dafür, eine infizierte Person (in-)korrekt zu identifizieren

$$P(+|\text{no AIDS}) = 0.03$$
$$P(-|\text{no AIDS}) = 0.97$$

$\leftarrow$ Wahrscheinlichkeiten dafür, eine nicht infizierte Person (in-)korrekt zu identifizieren

Nehmen Sie an, dass Ihr Ergebnis "+" ist.
Wie beunruhigt sollten Sie sein?

Ein Beispiel zum Bayes-Theorem (Fortsetzung)

Die Wahrscheinlichkeit AIDS zu haben, wenn der Test "+"

$$P(\text{AIDS}|+) = \frac{P(+|\text{AIDS})P(\text{AIDS})}{P(+|\text{AIDS})P(\text{AIDS}) + P(+|\text{no AIDS})P(\text{no AIDS})}$$

$$= \frac{0.98 \times 0.001}{0.98 \times 0.001 + 0.03 \times 0.999}$$

$$= 0.032 \qquad \leftarrow \text{Posterior-Wahrscheinlichkeit}$$

i.e. "wahrscheinlich" sind Sie nicht erkrankt !

Ihre Sichtweise: mein Grad an Glaube, dass ich AIDS habe ist 3.2%

Die Sichtweise ihres Artzes: 3.2% von Leuten unter diesen Bedingungen haben AIDS

Erstaunliches Ergebnis? → Deshalb oft B-Probe bei solchen Tests

Interpretation der Wahrscheinlichkeit

Bisher nur axiomatische Definition. Nicht hilfreich in der Anwendung.

Wir brauchen:

- was sind die Elemente der Menge?
- wie ordnen wir diesen die Wahrscheinlichkeitswerte zu?

2 Schulen:

a) Frequentisten / Klassische Wahrscheinlichkeit

b) Bayesianer / Subjektive Wahrscheinlichkeit

Bemerkung: Bayes-Theorem gilt in beiden und wird von beiden benutzt.
Aber die Interpretation ist unterschiedlich.

Frequentistische Definition/Interpretation

Grundgesamtheit S = alle mögliche Ergebnisse eines Experimentes, das hypothetisch wiederholbar ist

Untermengen/Ereignis *A*, *B*, = Messung liefert Ergebnis, dass im Wertebereich von A bzw. B liegt...

Elementarereignis: Menge mit einem Element

Wahrscheinlichkeit = Grenzfall der relativen Häufigkeit

$$P(A) = \lim_{n \to \infty} \frac{\text{Anzahl der Ergebnisse } A \text{ in } n \text{ Messungen}}{n}$$

Frequentistische Statistik: generelle Philosophie

In frequentistischer Statistik, werden Wahrscheinlichkeiten nur Daten zugeordnet, i.e. Ergebnissen von wiederholbaren Beobachtungen

Wahrscheinlickeit = Grenzfall der relativen Häufigkeit

Wahrscheinlichkeit für

P (Higgs boson exists),
P (9.81 m/s^2 < g < 9.82m/s^2)

P(morgen 26.10. 2012 regnet es in Freiburg)

sind entweder 0 or 1, aber wir wissen nicht welche.

Die Werkzeuge der frequentistischen Statistik sagen uns, was wir zu erwarten haben, unter der Annahme über gewisse Wahrscheinlichkeiten, über hypothetisch wiederholbare Beobachtungen.

Die bevorzugten Theorien (Modelle, Hypothesen, …) sind die jenigen, für die unsere Beobachtungen als “normal” betrachtet werden.

Bayessianische Interpretation/Definition

Grundgesamtheit S = Menge der Hypothesen

Untermengen A,B = eine oder mehrere Hypothesen (exklusive, d.h.
sich ausschließende Aussagen,
die wahr oder falsch sind)

Wahrscheinlichkeit = Grad des Glaubens (Zuversicht),
dass Hypothese A wahr ist

$$P(A) = \text{degree of belief that } A \text{ is true}$$

Zuordnung der Wahrscheinlichkeiten im Prinzip beliebig

Kann Wahrscheinlichkeiten beliebigen Aussagen zuordnen
z.N. P(morgen 26.10.2011 regnet es in Freiburg) = 0.95

Umfasst frequentistische Def. “Ereignis tritt mit XY% Häufigkeit auf”.

Bayesianische Statistik: Generelle Philospohie

In Bayesianischer Statistik: weise "subjektive Wahrscheinlichkeit" = "Grad des Glaubens" den verschiedenen Hypothesen zu:

Wkt. solche Daten zu beobachten unter der Annahme der Hypothese *H* (the likelihood)

A-Priori-Wkt., i.e. vor der Datennahme

$$P(H|\vec{x}) = \frac{P(\vec{x}|H)\pi(H)}{\int P(\vec{x}|H)\pi(H)\,dH}$$

Posterior-Wkt., i.e. nach Auswertung der Daten

Normierung beinhaltet Summe über alle möglichen Hypothesen

Bayes-Theorem hat einen "wenn-dann"-Charakter:
Wenn Ihre A-Priori-Wkt π (*H*) wären, dann sagt es Ihnen wie diese Wkt. sich im Licht der Beobachtung/Daten ändern.

Keine generelle Regel für Wahl der A-Priori-Wkt! →Subjektiv!

Bemerkung zu zwei Schulen

Subjektivität in der Annahme der A-Priori-Wahrscheinlichkeit in der Bayesianischen Schule erscheint vielen als nicht “wissenschaftlich”. Das ist falsch! Aber Abhängigkeit von A-Priori-Wahrscheinlichkeiten enthält gewisse Beliebigkeit.

Häufige Annahme: flach=gleichverteilt
Aber in welcher Größe?
z.B. Erkrankung unabhängig vom Alter oder proportional zum Alter?

In Physik (Naturwissenschaften) frequentische Interpretation oft nützlicher, aber subjektive Wahrscheinlichkeit kann eine “natürlichere” Behandlung nicht wiederholbarer Messungen/ sich nicht wiederholender Phänomene liefern:
z.B. - systematische Fehler
- Wahrscheinlichkeit, dass ein neues Phänomen existiert, …