
Wahrscheinlichkeit

Dr. Dirk Hasenclever
IMISE, Leipzig

Hasenclever@IMISE.uni-Leipzig.de

Variabilität von Versuchsergebnissen

□ Grunderfahrung Zufall:

Mit dem gleichen Eingriff erhält man bei biologisch-medizinischen Versuchen in der Regel **unterschiedliche Ergebnisse** auch, wenn man alle bekannten **Störgrößen ausschaltet** und die **Versuchsbedingungen konstant** hält.

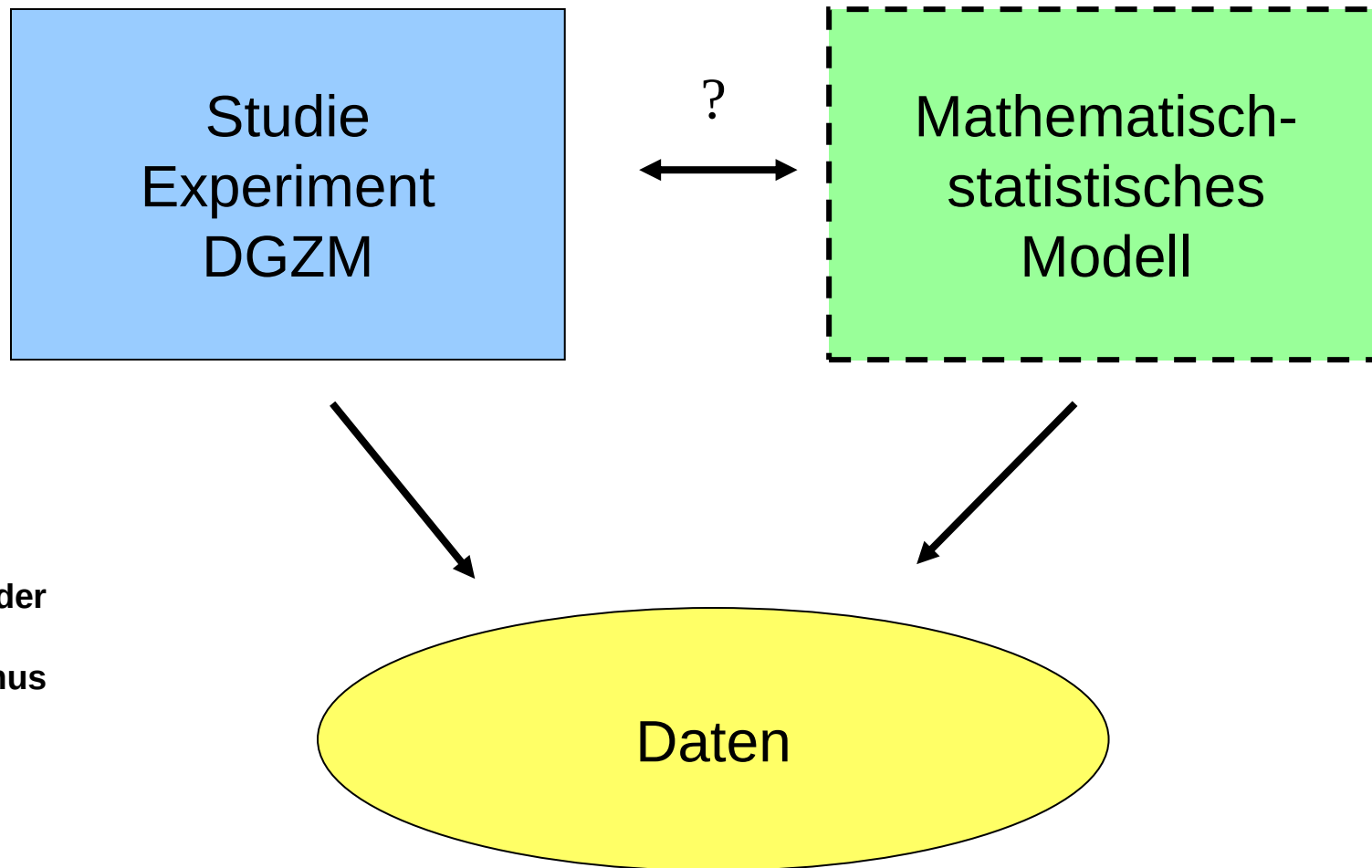
Induktion als Grundproblem der Statistik

- Versuche sind de facto nur an **endlichen Stichproben** durchführbar.
- Der **induktive Schluss**:
Endliche Stichprobe -> Allgemeine Aussage
über Versuchsprinzip
(Grundgesamtheit)
ist mit **evidentialer UNSICHERHEIT** behaftet
- **Statistik** kontrolliert diese Unsicherheit und
schützt vor **irreführenden Zufallsbefunden**.

You need to posit a statistical model in order to learn from data

- We cannot learn unless we make
a **learning enabling** “transcendental”
ansatz:
 - Data come from (in principle) identically repeatable random experiments:
 - E.g.: Take a patient with disease X, apply treatment Y and observe the outcome.
 - Independently identically repeatable random experiments or random variables (RV) are characterized by a
 - **probability function $\Pr$ that assigns a probability to every possible outcome of the experiment.**
-

Statistische Modellierung



You need to posit a statistical model in order to learn from data II

- This **probability function**
that nature has hidden in natural random experiments
is not directly observable, but is precisely
the only thing that we can learn.
 - **We can use statistics to learn about the unknown $\Pr$** or
about parameters (e.g. means) derived from it
 - when we are **given a large enough data sample** from the
experiment.
 - **Statisticians do not talk about data!**
but about unknown probability functions
or their derived parameters
in **natural random experiments**.
-

Mathematisches Jenseits

Statistisches Diesseits

X Zufallsvariable

x Realisierung der Zufallsvariablen X

$f(x) dx$ theoretische Dichte von X

$f^{\wedge}(x)$ empirische Dichteschätzung
(z.B. Histogramm)

$F(x) = \text{pr}(X \leq x) = \int_{-\infty}^x f(\xi) d\xi$
theoretische Verteilungsfunktion

$F^{\wedge}(x)$ empirische Verteilungsfunktion
 $= \#\{x_i \mid x_i \leq x\} / N$

Mathematische Stichprobe

Empirische Stichprobe

$X_1, X_2, \dots, X_N$
unabhängig und identisch verteilte
Zufallsvariable.

$x_1, x_2, \dots, x_N$
Realisierungen unabhängiger
und identisch verteilter Zufallsvariabler.

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

Empirischer Mittelwert: $\bar{x} = \frac{1}{N} \sum x_i$

theoretische Varianz

empirischer Varianzschätzer:

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

$$\hat{\sigma} = \frac{1}{N-1} \sum (x_i - \bar{x})^2$$

Was ist Wahrscheinlichkeit?

Definition 1.4: Sei $(\Omega, \mathfrak{S})$ ein meßbarer Raum und sei P eine Abbildung von $\mathfrak{S}$ in $\mathbb{R}$. P heißt ein Wahrscheinlichkeitsmaß oder kurz eine Wahrscheinlichkeit, wenn gilt:

$$(I) \quad A \in \mathfrak{S} \Rightarrow P(A) \geq 0.$$

$$(II) \quad (A_n \in \mathfrak{S} \text{ für alle } n \in \mathbb{N} \text{ und } A_i \cap A_k = \emptyset \text{ für } i \neq k) \Rightarrow$$

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n).$$

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

Erläuterung formaler W-Begriff I - diskret

Beispiel: Mingolfscores an Bahn 2 (geübte Freizeitspieler)

$\Omega = \{1,2,3,4,5,6,7\}$ mögliche Ergebnisse

$\Sigma = \text{Potenzmeng } e(\Omega)$ z.B. $\{1,2\} \in \Sigma$

Definiere pr auf Ω

$\text{pr}(i) := (0.1, 0.2, 0.2, 0.1, 0.1, 0.1, 0.2)[i]$

Wegen Additivität: $\text{pr}(\{1,2\}) = \text{pr}(1) + \text{pr}(2) = 0.3$

pr auch auf Σ fortgesetzt.

Erläuterung formaler W-Begriff I - stetig

Beispiel: Messwerte

Ω = positive reelle Zahlen (mögliche Ergebnisse)

Problem : $\text{pr}(\text{Einzelwert}) = 0$

Σ = Durchschnitte und Vereinigungen von Intervallen

Definiere pr auf Σ durch Angabe einer Dichte $f(x)dx$

$$\text{pr}([a, b]) := \int_a^b f(x)dx$$

Mathematische Wahrscheinlichkeit

Erläuterung

Ω bezeichnet die Menge der möglichen Ergebnisse des Zufallsexperiments.

S ist eine unter abzählbaren Durchschnitten und Vereinigungen abgeschlossene Menge von Teilmengen von Ω .

S beschreibt alle möglichen beobachtbaren Ereignisse.

Falls Ω endlich oder abzählbar, nimmt man die volle Potenzmenge von Ω .

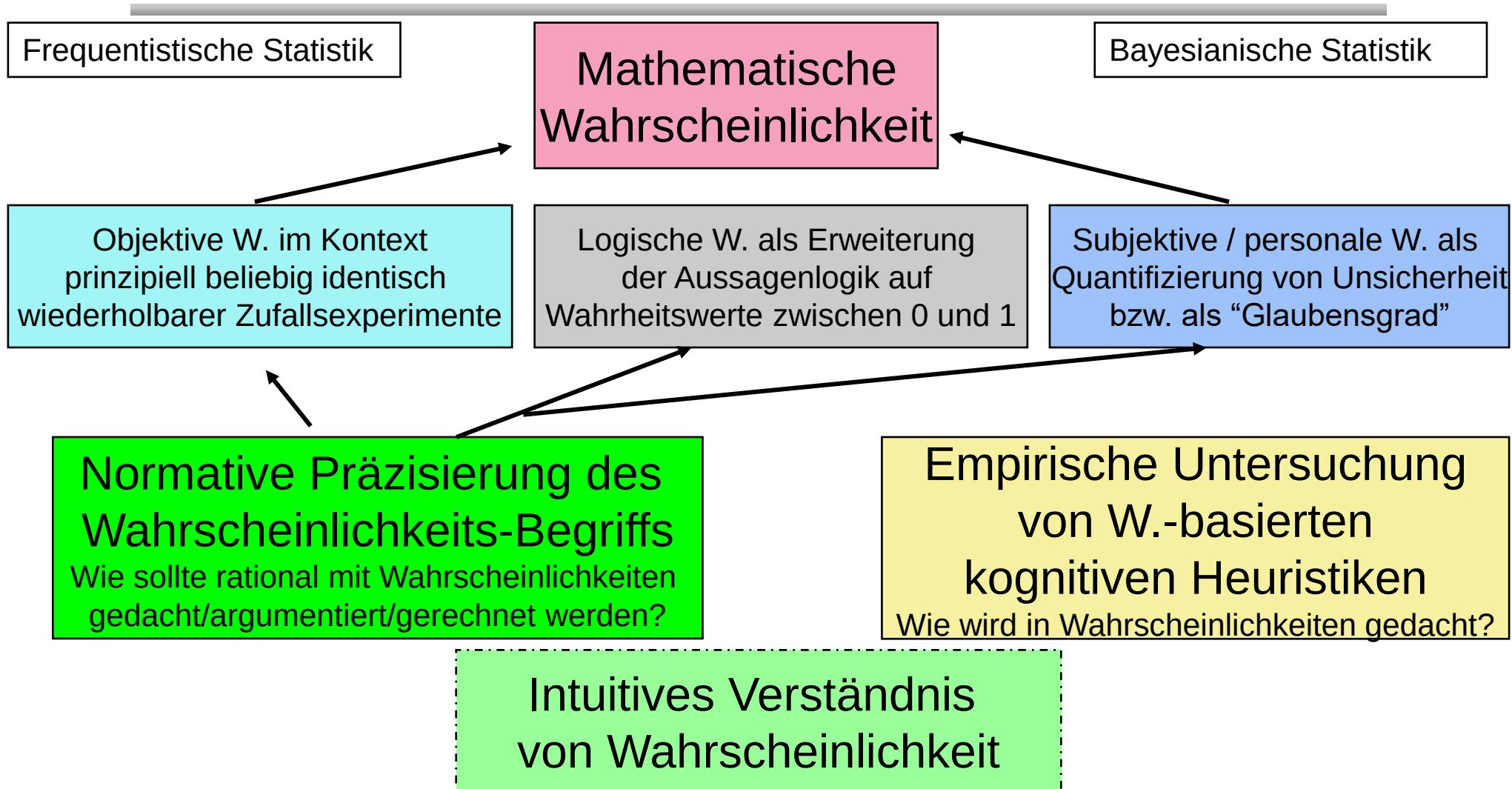
Ist Ω überabzählbar z.B. $\mathbb{R}$, so nimmt man meist für S die Menge der messbaren Borelmengen, die von Intervallen $[a,b]$ mit $a,b \in \mathbb{Q}$ erzeugt wird.

$\Pr$ ist eine Funktion, die jedem möglichen Ereignis $E \in S$ seine Wahrscheinlichkeit $\Pr(E)$ in $[0,1]$ zuordnet.

Mathematische Wahrscheinlichkeit und Erkenntnis?

**Der mathematische Begriff der
Wahrscheinlichkeit bedarf im Kontext der
Erkenntnisgewinnung einer
Realinterpretation!**

Wahrscheinlichkeiten



Zwei intuitive Quellen für Wahrscheinlichkeit

- Stabile Häufigkeiten bei wiederholbaren Versuchen
(besonders bei Symmetrie: Würfel, Urne etc.)
-> **frequentistischer W-Begriff**
- Rationale Einschätzungen von Eventualitäten beim Handeln unter Unsicherheit
Objektiviert sich als „fairer Wettquotient“
-> **personalistischer (Bayes) W-Begriff**
- Mathematische Axiome gelten für beide Realbegriffe.

Frequentistischer W-Begriff I

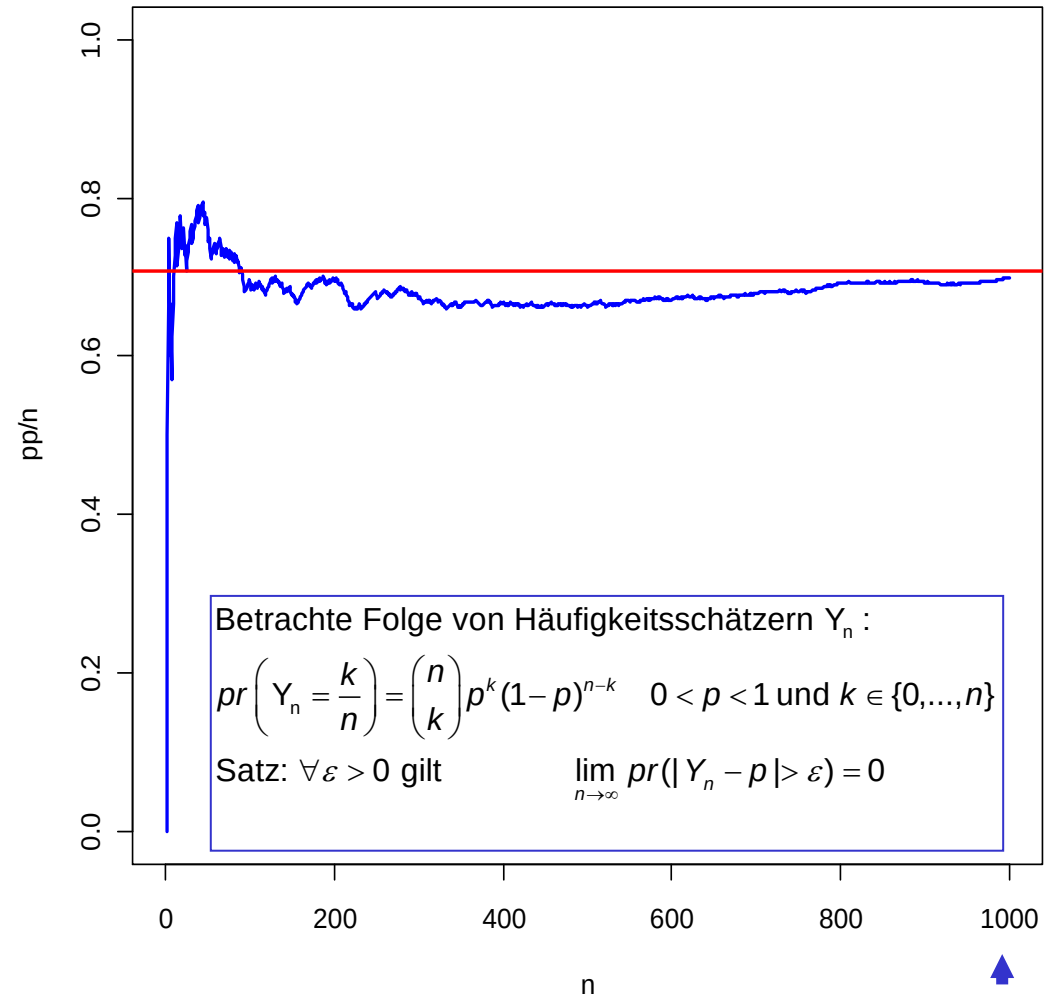
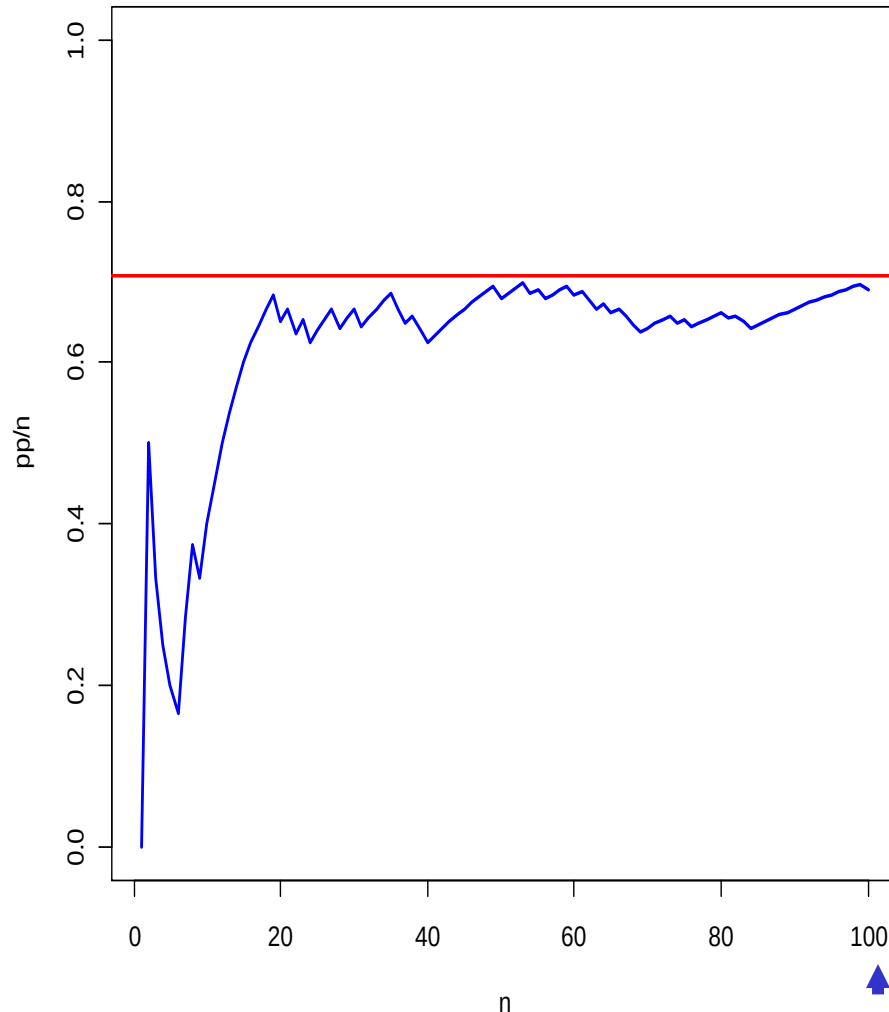
- Wahrscheinlichkeit ist **nur definiert** in Bezug auf **(identisch) wiederholbarer Zufallsexperimente**.
- Historische Prototypen:
Gleichverteilung durch Symmetrie:
Urnenmodell, Würfel, Kartenspiele etc.
- Umgangssprachlich geläufige Wahrscheinlichkeit von singulären Ereignissen (z.B. Ergebnis eines bestimmten Fußballspiels) ist **wissenschaftlich undefiniert**.

Frequentistischer W-Begriff II

- Die W. eines bestimmten Ergebnisses eines wiederholbaren Zufallsexperiments lässt sich durch die **empirische Häufigkeit** in einer hinreichend langen Serie von Experimenten approximativ schätzen.
- Allerdings ist die Konvergenz der empirischen Häufigkeiten nur stochastisch:

$$\forall \varepsilon > 0 : \lim_{n \rightarrow \infty} \Pr(|\hat{p}_n - p| > \varepsilon) \rightarrow 0$$

Law of large numbers: Proportion $\rightarrow$ Probability



Frequentistischer W-Begriff III

- Weil die Erklärung der Konvergenz der Häufigkeiten gegen die „wahre“ Wahrscheinlichkeit **selbst den Wahrscheinlichkeitsbegriff voraussetzt**, kann man die frequentistische Wahrscheinlichkeit **nicht als rein empirische Eigenschaft von Häufigkeitsfolgen** konzeptualisieren.
- Frequentistische Wahrscheinlichkeit wird daher verstanden als **objektive, aber latente Eigenschaft (identisch) wiederholbarer Zufallsexperimente**.

Bayes'scher Wahrscheinlichkeitsbegriff I

- Bayes'sche W. = **rationaler Grad von Glauben** oder **rationale Quantifizierung von Unsicherheit**.
- W. stellt subjektive Ungewissheit dar.
- Rede über **Wahrscheinlichkeit von Hypothesen** erlaubt, auch wenn diese „an sich“ wahr oder falsch.



Rev. Thomas Bayes (1702-1761)

Bayes'scher Wahrscheinlichkeitsbegriff II

- Bayes'sche Wahrscheinlichkeit kann **formalisiert** werden als:

- **Rationale Wettquotienten**

(de Finetti)

- **Erweiterung der Logik** auf

Wahrheitswerte zwischen 0 und 1

(Jeffreys, R. T. Cox, Jaynes)

Rationale Wettquotienten I

Subjektiver Glaube wird dargestellt im Wettverhalten.

Wette zwischen X und Y bzgl. Proposition H:

X wettet auf H und setzt u

Y wettet auf $\neg H$ und setzt v

Der Gewinner bekommt: $u+v$

Wettquotient von X: $u/(u+v)$

Faire Wette = maximaler akzeptabler Wettquotient
= Gewinnerwartung 0:

$$p*(+v)+(1-p)*(-u) = 0 \quad \Rightarrow$$

$$p = u/(u+v)$$

Fairer Wettquotient ist subjektive Wahrscheinlichkeit von X bzgl. H.

Rationale Wettquotienten II

Rationalitätskriterium für Systeme von Wetten

Wetten auf $H_1, \dots, H_n$: 2^n mögliche Kombinationen von Wahrheitswerten

System von Wetten irrational $\Leftrightarrow$ Verlust bei allen möglichen Ausgängen

Sei $Cr_X(H)$ der größte Wettquotient, den X bzgl. H akzeptiert.

Cr_X kohärent $\Leftrightarrow$ Es gibt kein endliches System $H_1, \dots, H_n$, so dass Wetten mit $Cr_X(H_i)$ zu notwendigem Verlust führt.

Rationale Wettquotienten III

Theorem (de Finetti):

Cr_x kohärent $\Leftrightarrow$

Cr_x erfüllt Axiome der Wahrscheinlichkeit
mit endlicher Additivität

Wahrscheinlichkeit als Erweiterung der Logik

Ziel: Induktive (erkenntnisleitende) Logik als Erweiterung der Logik der Propositionen

Intuitiver Grundbegriff:

q ist wahrscheinlicher als r gegeben Wissen/Daten p:

$$q | p > r | p$$

Frage: Kann man diesen Grundbegriff

- a) konsistent mit der Logik
der Propositionen verwenden
- b) durch Daten/neues Wissen kohärent
zum Lernen verwenden?

RT Cox theorem

□ Desiderata:

- Degrees of plausibility represented as real numbers
- Qualitative correspondence with common sense
 - Compatibility with Boolean Logic
- Consistency

RT Cox theorem II

□ Theorem:

- There exists a (~unique) strictly increasing function pr such that for all propositions A, B and information X :
 - $0 \leq Pr(A|X) \leq 1$
 - $Pr(A|X) = 0$ iff A known false given X
 - $Pr(A|X) = 1$ iff A known true given X
 - $pr(A \& B | X) = p(B | A, X) * pr(A|X)$
 - $pr(\neg A | X) = 1 - pr(A | X)$

Conditional probability

- **Fundamental Product Rule:**

- $$\begin{aligned}\Pr(A \& B) &= \Pr(A \mid B) * \Pr(B) \\ &= \Pr(B \mid A) * \Pr(A)\end{aligned}$$

- **Attention:**

- $\Pr(A \mid B) \neq \Pr(B \mid A)$

$$\begin{aligned}\Pr(\text{Student} \mid \text{Fan of RB Leipzig}) &\neq \\ \Pr(\text{Fan of RB Leipzig} \mid \text{Student})\end{aligned}$$

Independence

A and B are **independent**, if

$$\text{pr}(A \ \& \ B) = \text{pr}(A) \cdot \text{pr}(B)$$

or equivalent (via Product Rule)

$$\text{pr}(A \mid B) = \text{pr}(A)$$

$$\text{pr}(B \mid A) = \text{pr}(B)$$

Information on A contains no information about B and vice versa.

If an independence can be assumed, complexity is reduced massively and calculations are simplified.

Conditional probabilities – Law of Total Probability

- If there are N **alternative scenarios** $A_1, A_2, \dots, A_N$
 - That are pair-wise exclusive AND
 - Jointly cover all possibilities

Then for each B

$$\begin{aligned} \square \quad \Pr(B) = & \Pr(B \mid A_1) * \Pr(A_1) + \\ & \Pr(B \mid A_2) * \Pr(A_2) + \\ & \dots + \\ & \Pr(B \mid A_N) * \Pr(A_N) \end{aligned}$$

Bayes'scher Wahrscheinlichkeitsbegriff III

- Induktives Lernen bzgl. W. von Hypothesen
(Bayes'sche Lernformel)

**Unsicherheit bzgl. interessierenden
Parameters (Prior)**

+

**Daten eines informativen
Zufallsexperiments**



**Neuer Stand (reduzierter) Unsicherheit
bzgl. des Parameters (Posterior)**

Bayes'sche Lernformel

Seien $H_1 \dots H_k$ mögliche Hypothesen über ein Zufallsexperiment.
Seien D die beobachteten Daten.

Sei $pr(H_i)$, $i=1..k$, die Prior-Wahrscheinlichkeit von H_i
vor Beobachtung der Daten.

Dann gilt

Likelihood der Daten gegeben H_i

Posterior

Prior

$$pr(H_i | D) = \frac{pr(D | H_i) * pr(H_i)}{\sum_j pr(D | H_j) * pr(H_j)}$$

Normierungsfaktor

Sind Sie Frequentist oder Bayesianer?

- Kurzbeschreibungen von 100 Personen:
70 Rechtsanwälte und 30 Ingenieure.

Jack ist 45 Jahre alt, verheiratet, 4 Kinder. Er ist konservativ, sorgfältig und ehrgeizig.

Jack zeigt kein Interesse an politischen oder sozialen Themen und verbringt den größten Teil seiner Freizeit mit seinen vielfältigen Hobbies, darunter Heimwerkerei, Segeln und mathematische Rätsel

- Die Wahrscheinlichkeit, dass Jack einer der 30 Ingenieure unter den 100 Personen ist, beträgt:

(A) deutlich unter 50% (B) ca. 50% (C) deutlich über 50%

Nach Kahnemann & Tversky (1982)

Argument des Frequentisten

- Frequentistische Wahrscheinlichkeiten:
 $\text{pr}(R) = 0.7$
 $\text{pr}(I) = 0.3$
- Andere Wahrscheinlichkeits-Informationen stehen nicht zur Verfügung.
- Also ist Jack wahrscheinlich Rechtsanwalt.
- *Kahnemann & Tversky als Frequentisten wunderten sich, dass die meisten Versuchspersonen Antwort (B): ca. 50% angaben.*

Argument eines Bayesianers

- Bayesianer: Berücksichtige zusätzliches Vorwissen (subjektive Einschätzung):
- $pr(PS|I) = 0.60$ (PS=Interesse für Politik&Soziales)
 $pr(PS|R) = 0.95$ (PS gehört zum Beruf R.!!)
- $pr(\neg PS) = pr(\neg PS|R) * pr(R) + pr(\neg PS|I) * pr(I) = 0.155$
- $$pr(R | \neg PS) = \frac{pr(\neg PS | R) * pr(R)}{pr(\neg PS)} = \frac{0.05 * 0.7}{0.155} = 0.226$$
- **Satz von Bayes über bedingte Wahrscheinlichkeiten**
- Also ist Jack wahrscheinlich Ingenieur!

Paradigmenvergleich

Wahrscheinlichkeit

Frequentisten

Bayesianer

Daten

zufällig, austauschbar
=identisch verteilt

gegeben

Modelle

gegeben: wahr o. falsch
meist unbekannt

mehr oder weniger
wahrscheinlich

Scope

nur wiederholbare
Zufallsexperimente

Hypothesen und
singuläre Ereignisse
+ Zufallsexperimente

Philosoph. Flair

Realismus/Objektivismus
Empirismus

Rationale, kohärente
Bewertung von Unsicherheit

Hausaufgabe – das Ziegenproblem

- In einer Quizshow stehen Sie vor drei identischen Garagentüren.
- Hinter einer der dreien steht ein **roter Ferrari** für Sie, wenn Sie die richtige Tür öffnen.
- Sie stellen sich vor vorläufig vor die Tür1.
- Der Moderator öffnet daraufhin die Tür3:

Ätsch eine Ziege!

Und bietet Ihnen an, sich eventuell umzuentscheiden **!?!**

- **Was tun Sie und warum?**