

Unsicherheit und Uneindeutigkeit aus der Sicht der statistischen Modellbildung

Thomas Augustin

Institut für Statistik, LMU München

1. Wahrscheinlichkeit und Statistik
2. Wahrscheinlichkeit: historische Entwicklung und aktuelle Hauptrichtungen
3. Ambiguität als Herausforderung:
 - 3.1 Weichere Modelle mit klassischen Wahrscheinlichkeiten
 - 3.2 Imprecise Probabilities – Intervallwahrscheinlichkeit

Vereinfachtes Argumentationsmuster

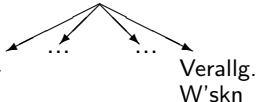
Klassische Statistik

- Präzise Modelle (und Wsk)
- Klarer Optimalitätsbegriff

Komplexität →

Explizite Modellierung
von Uneindeutigkeit

Weichere Modelle



Modellpluralität

Weichere Optimalität



Die Rolle der Wahrscheinlichkeitsrechnung in der Statistik I

- ▶ Stark vereinfacht:
- ▶ Wahrscheinlichkeit erlaubt Quantifizierung des Stichproben- und Inferenzfehlers bei Zufallsstichproben
- ▶ Bei Modellen zudem Modellfehler, unerklärter Rest
 - ▶ Beobachtete Daten y_i = wahrer Zusammenhang $f(x_i)$ + Fehler ϵ_i
 - ▶ ϵ_i Fehler: Zufällig

Die Rolle der Wahrscheinlichkeitsrechnung in der Statistik III

- ▶ Der Wahrscheinlichkeitsbegriff prädeterminiert über das Inferenzkonzept hinaus auch den Optimalitätsbegriff in Entscheidungsproblemen

	Zustand 1	...	Zustand m
Aktion 1	Nutzenwerte		
Aktion 2			
...			
Aktion n			

z.B. Ausflugsproblem, Investitionsproblem

2. Wahrscheinlichkeit: Historische Entwicklung und aktuelle Hauptrichtungen

Weichselberger (2001, Physika, Kap. 1)

Der Urzustand

- ▶ Wahrscheinlichkeit als „philosophischer“ Begriff, Probabilismus in der Theologie
- ▶ Wahr – schein – lichkeit
- ▶ prove – ability

Drei historische Wurzeln

- ▶ Philosophische Auseinandersetzung, Probabilismus in der Theologie
- ▶ Mathematische Theorie der Glücksspiele
- ▶ Politische Arithmetik: Massenerscheinungen im öffentlichen Leben

- ▶ Zunehmende Quantifizierung: „ Das Messbare messen, das Nichtmessbare messbar machen.“(Galilei)

- ▶ John Grant: ... Bills of Mortality (probabilistische Gesetzmäßigkeiten statt beschreibende Häufigkeiten)
- ▶ Port-Royal-Logik (Arnauld, Nicole):
 - Zusammenwirken der Philosophie des Unsicheren und der Mathematik der Glücksspiele
 - „Von der richtigen Beurteilung zukünftiger Begebenheiten“: Universelle Anwendbarkeit des Wahrscheinlichkeitsbegriffes auf zukünftige Ereignisse aller Art

Aktuelle Hauptrichtungen des Wahrscheinlichkeitsbegriffs

- ▶ axiomatisch - formal (Kolmogoroff, 1933)
- ▶ objektiv(istisch), objektbezogen:
 - ▶ Wahrscheinlichkeit als Eigenschaft des Objektes
 - ▶ oft: zufällige Folgen (aleatorisch, frequentistisch)
- ▶ subjektiv(istisch), subjektbezogen:
 - ▶ Wahrscheinlichkeit als Eigenschaft des Betrachters
 - ▶ Unsicherheit
- ▶ logisch, schlussbezogen:
 - ▶ Passt nicht in Kolmogorowsches System
 - ▶ Wahrscheinlichkeit als Eigenschaft des Schlusses von der Prämisse auf die Konklusion
 - ▶ $P(A||B)$: Prämisse $\xrightarrow{P(A||B)}$ Konklusion

Subjektive Wahrscheinlichkeit I

- ▶ Umfassend: **Jede** Form der Unsicherheit (statt Zufälligkeit) durch Wahrscheinlichkeiten quantifizierbar; (man denke an Zufallsszahlen)
- ▶ Behaviouristischer Zugang: Interpretation bzw. Messung über Verhalten (insbesondere in Wettsituationen)
 - ▶ Fiktiver Markt mit Wetten
 - ▶ Man kann Wetten kaufen oder verkaufen
Kauf: Gewinn 1 Euro, falls A eintritt
Marktpreis e variabel
 - ▶ Suche Indifferenzpunkt $p(A)$ zwischen Kauf (Wette auf A) und Verkauf (Wette auf „Nicht- A “)



- ▶ $p(A)$ Wahrscheinlichkeit von A
- ▶ Kohärenzbedingungen für Kombinationen von Wetten auf verschiedene Ereignisse
- ▶ (fast) äquivalent zu Kolmogorowscher Axiomatik

Subjektive Wahrscheinlichkeit II

- ▶ unmittelbare Inferenztheorie (Bayesianer)
 - ▶ A priori Verteilung $\pi(\theta)$ über Parameter
 - ▶ Verteilung der Daten $f(x|\theta)$
 - ▶ Posteriori Verteilung $\pi(\theta|x)$ über den Satz von Bayes

Wahrscheinlichkeit(Parameter)

+

Daten

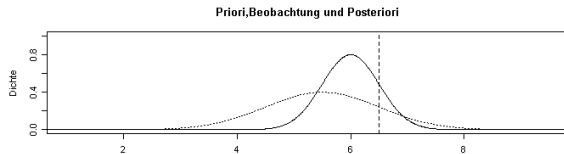
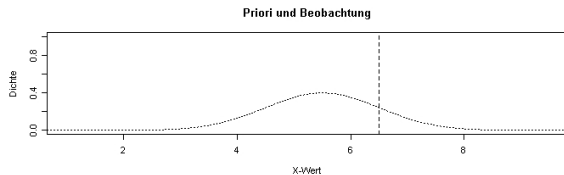
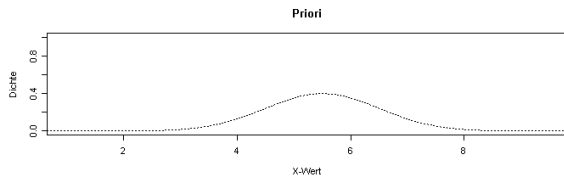
⇒ Wahrscheinlichkeit(Parameter|Daten)

Bayes-Inferenz über Mittelwert bei Normalverteilung

- ▶ Stichprobe(nmittel) $X \sim N(\theta, 1): f(x|\theta)$
- ▶ Priori $\theta \sim N(\nu, 1): \pi(\theta)$
- ▶ Posteriori $\theta|x \sim N(\frac{1}{2} \cdot (\nu + x), 0.5): \pi(\theta|x)$

Priori-Daten-Konflikt, Beispiel 1

Beispiel 1: $\nu = 5.5$, $x = 6.5$, $\frac{\nu+x}{2} = 6$



Logische Wahrscheinlichkeit

- ▶ Wahrscheinlichkeit als Eigenschaft des Schlusses von der Prämisse auf die Konklusion
- ▶ $P(A||B)$: Prämisse $\xrightarrow{P(A||B)}$ Konklusion
- ▶ Auflösung des Subjektivismus – Objektivismus Gegensatzes
- ▶ Symmetrischer Ansatz

Wahrscheinlichkeit(Daten||Parameter)



Wahrscheinlichkeit(Parameter||Daten)

- ▶ Dempster, Levi, Kyburg, Hampel, Weichselberger/Wallner

Auswirkungen auf die Entscheidungstheorie

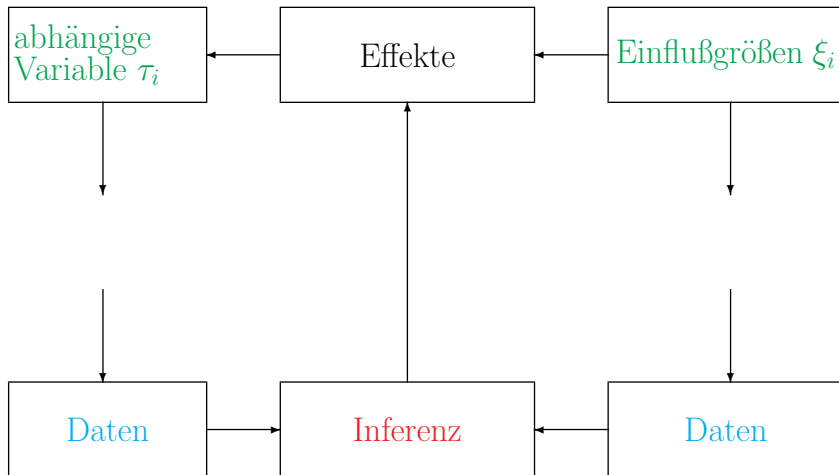
	Zustand 1	...	Zustand m
Aktion 1	Nutzenwerte		
Aktion 2			
...			
Aktion n			

- ▶ objektivistisch: keine Zustandswahrscheinlichkeiten (Nichtwissen), Entscheidungsproblem als Spiel gegen die feindliche Natur; Maximin-Lösung (s.a. Rawls)
- ▶ subjektiv: „virtuelle Lotteriesituation“, präzise Zustandswahrscheinlichkeiten: erwarteter Nutzen
- ▶ Was tun mit partiellem Wissen?

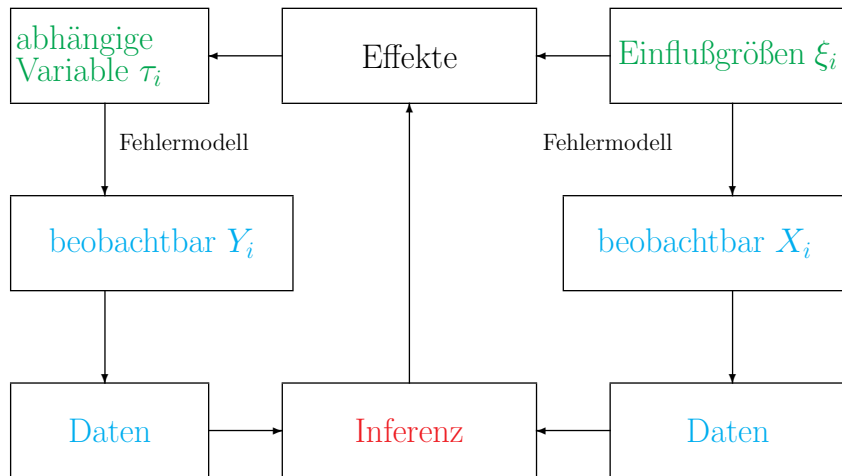
3. Ambiguität als Herausforderung

- ▶ Sehr erfolgreiche Verbreitung der Statistik in allen empirisch arbeitenden Wissenschaften
- ▶ Weichere Modelle mit klassischen Wahrscheinlichkeiten
 - ▶ Fehler in-den-Variablen-Modellen
 - ▶ Robuste Verfahren
 - ▶ Nichtparametrische Verfahren
 - ▶ Sensitivitätsanalysen
- ▶ Darüberhinaus Zweifel an der prinzipiellen Eignung von Wahrscheinlichkeiten zur Modellierung Unsicheren Wissens
 - ▶ künstliche Intelligenz
 - ▶ Ökonomie
- ▶ Imprecise Probabilities – Intervallwahrscheinlichkeit

3.1. Weichere Modelle mit klassischen Wahrscheinlichkeiten



Fehler-in-den-Variablen-Modelle



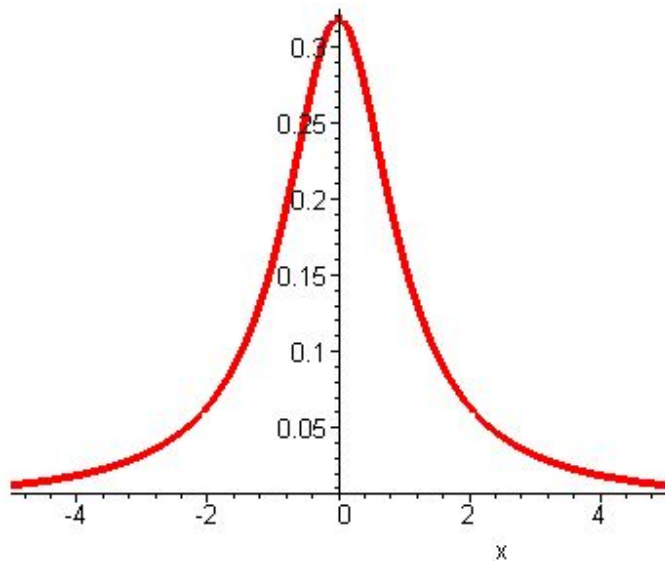
Messfehler und Fehlklassifikation

- ▶ Unterscheide zwischen
 - ▶ idealer und
 - ▶ beobachtbarer Ebene
- ▶ auch wichtig in amtlicher Statistik bei bewusst zur Anonymisierung kontaminierten Daten
- ▶ Nichtberücksichtigung von Messfehlern / Fehlklassifikation kann zu gravierenden Verzerrungen der Schätzung führen
- ▶ Grobe Erklärung: Variabilität im Messprozess wird versehentlich dem Modell zugeschrieben
- ▶ stetige Variable: *Messfehler* (z.B. Carroll et al., 2006, Chapman & Hall; Schneeweiss & Augustin, 2007, ASTA)
- ▶ diskrete Variable *Fehlklassifikation* (z.B. Küchenhoff et al., 2006, Biometrics)

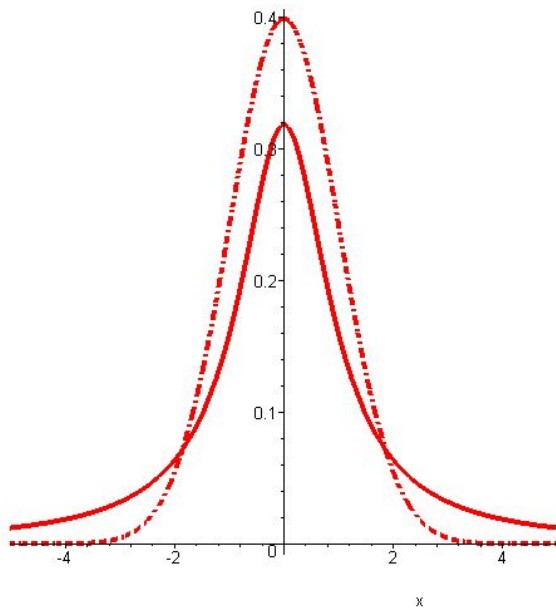
Nichtparametrische („Verteilungsfreie“) Verfahren

- ▶ Qualitative Verteilungsannahmen (z.B. Symmetrische Verteilung) mit entsprechenden Hypothesen (z.B. Median=0) anstatt z.B. Normalverteilung
- ▶ „Vorsichtige Datenverarbeitung“ statt Werte, Ränge oder grosser–kleiner Zählungen

Robuste Verfahren



Robuste Verfahren



Betrachte Stichprobenmittel \bar{X} .

- ▶ Sind $X_1, \dots, X_n \sim N(\mu, 1)$ (normalverteilt), dann

$$\bar{X} \sim N\left(\mu, \frac{1}{n}\right)$$

Man kann aus der Stichprobe lernen: je größer der Stichprobenumfang, desto genauer der Schätzer \bar{X} .

- ▶ Sind $X_1, \dots, X_n \sim C(\mu, 1)$ (Cauchy-verteilt), dann ist

$$\bar{X} \sim C(\mu, 1)$$

Man kann nicht dazulernen, egal wie groß der Stichprobenumfang ist.

- ▶ Viele optimale Verfahren verhalten sich desaströs unter minimalen Abweichungen vom zugrundegelegtem idealen Modell
- ▶ Idee: Schließe Versicherung ab (Versicherungsprämie \triangleq leichter Effizienzverlust im Idealfall)
- ▶ Einfachstes Beispiel: Median statt ausreißerempfindlichem arithmetischen Mittel
- ▶ Betrachte statt Modell $f(x|\vartheta)$: Modell „ungefähr $f(x|\vartheta)$ “

- ▶ Idee (Ingenieurwissenschaften): Bleiben substanzwissenschaftliche Ergebnisse erhalten unter systematischer Variation der Modellannahmen?
- ▶ Analyse von Counterfactuals (Rubin)
- ▶ Fehlende Daten bei systematischem Fehlermuster (MNAR)
- ▶ Bayesianische Sensitivitätsanalyse (Good)

3.2. Imprecise Probabilities – Intervallwahrscheinlichkeit

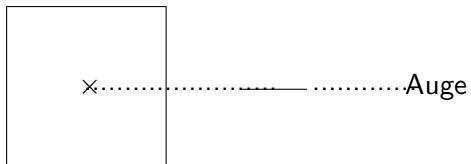
Klir and Wierman (Uncertainty-based Information, Physika, 1998, S.1)

„For three hundred years [...] uncertainty was conceived solely in terms of probability theory. This seemingly unique connection between uncertainty and probability is now challenged [... by several other] theories, which are demonstrably capable of characterizing situations under uncertainty. [...] [...] it became clear that there are several distinct types of uncertainty. That is, it was realized that **uncertainty** is a **multidimensional concept**. [...] That] multidimensional nature of uncertainty was obscured when uncertainty was conceived solely in terms of probability theory, in which it is manifested by only one of its dimensions“.

Unsicherheit

- ▶ Ideale Stochastizität (Risiko): perfekter Zufallsmechanismus, oder äquivalent
- ▶ Ambiguität (nichtstochastische Unsicherheit, Unbestimmtheit)
- ▶ Vagheit

Projektion



„Are all uncertainties risks?“

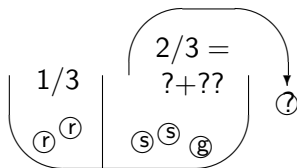
- ▶ Keynes (1921, Treatise on Probability):
 - ▶ Logische Wahrscheinlichkeiten
 - ▶ Nicht notwendigerweise alle Ergebnisse hinsichtlich ihrer Wahrscheinlichkeit vergleichbar → „non-numerical probabilities“
- ▶ Knight (1921: Risk, Uncertainty and Profit)
 - ▶ Risiko (Wahrscheinlichkeitsbewertung exakt möglich) versus
 - ▶ Unsicherheit
 - ▶ Unterscheide Ziehen aus Urnen mit bekannten und unbekanntem Anteilen

- ▶ Rand Corporation
- ▶ Pentagon Papers
- ▶ Alternativer Nobelpreis

Ellsberg-Paradox

Gedankenexperiment unter Ökonomen und Statistikern

- ▶ Gegeben sei eine Urne mit
 roten
 gelben Kugeln
 schwarzen
- ▶ Anteil der *roten* Kugeln ist genau $\frac{1}{3}$
- ▶ Der Anteil der *gelben* und *schwarzen* Kugeln ist hingegen unbekannt
- ▶ Eine Kugel wird zufällig gezogen



Ellsberg-Paradox — Die Priori im Hintergrund

- ▶ $\{r\} > \{s\}$ und zugleich $\{r \vee g\} < \{s \vee g\}$
- ▶ Bayesianisches Paradigma: es gibt zu jeder Unsicherheitssituation eine (klassische) Wahrscheinlichkeitsbewertung $\pi(\cdot)$, die die subjektive Einschätzung beschreibt.
- ▶ Für die priori Verteilung $\pi(\cdot)$ muss also gelten:

$$[\pi(\{r\}) > \pi(\{s\})] \quad \wedge \quad [\pi(\{r\}) + \pi(\{g\}) < \pi(\{s\}) + \pi(\{g\})]$$

⇒ Es kann *keine klassische* Wahrscheinlichkeitsverteilung über die Umweltzustände geben, die die am häufigsten beobachteten Präferenzen widerspiegelt.

Konsequenzen aus dem Ellsberg-Experiment

- ▶ Gedankenexperiment: Nicht nur empirische Verletzung von Rationalitätsprämissen, sondern bewusste und als rational empfundene Ablehnung durch führende Forscher der Entscheidungstheorie
- ▶ Es gibt also Präferenzordnungen in Entscheidungssituationen unter Unsicherheit, die als *rational* empfunden werden, die aber *nicht* mittels *klassischer* Wahrscheinlichkeit beschreibbar sind.
- ▶ Ambiguität (nichtstochastische Unsicherheit) als konstitutives Element
- ▶ Soll die Entscheidungstheorie ihrem Anspruch als Theorie des *rationalen Entscheidens* unter Unsicherheit gerecht werden, so muss sie solche Situationen modellieren können.“
- ▶ Beachte: Die Einfachheit des Beispiels stützt das Argument.

Quelle:



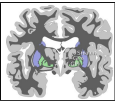
**Neural Systems Responding to Degrees of
Uncertainty in Human Decision-Making**

Ming Hsu et al.

Science 310, 1680 (2005)

hier: nach Jelizarow (2007)

Gehirnregionen und ihre Funktionen

	Bild	Funktion	Geschwindigkeit
Amygdala		Überwachung	schnell
Orbitofrontaler Kortex			
Striatum		Belohnungserwartungssystem	langsam nachgeordnet

▶ **Kartenstapel**

▶ Risikobedingung

Entscheidung zwischen purem Risiko (Zusammensetzung des Kartenstapels bekannt) und sicherem Geldbetrag

▶ Ambiguitätsbedingung

Entscheidung zwischen purer Ambiguität (Zusammensetzung des Kartenstapels unbekannt) und sicherem Geldbetrag

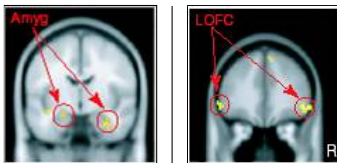
▶ **Wissen**

Entscheidungen zwischen einem sicheren Geldbetrag und der Wette auf Bejahung oder Verneinung von Aussagen bzw. Ereignissen, deren Inhalt entweder eine Risiko- oder Ambiguitätsbedingung darstellt.

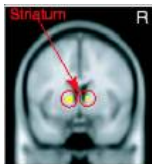
Bsp.: Wetter in New York — Wetter in Tirana

Aktivität von Hirnregionen unter Risiko/Ambiguität

- ▶ Während der ambigen Bedingung – im Vergleich zur riskanten Bedingung – waren besonders aktiv *OFC* (*Orbitofrontaler Kortex*) und *Amygdala*



- ▶ Während der riskanten Bedingung – im Vergleich zur ambigen Bedingung – war besonders aktiv: *Striatum*



Gehirnregionen und ihre Funktionen

	Bild	Funktion	Geschwindigkeit	besonders aktiv bei
Amygdala		Überwachung	schnell	Ambiguität
OFC				
Striatum		Belohnungs- erwartungs- system	langsam nachge- ordnet	Risiko

Deskriptive Modelle verallgemeinerter Entscheidungstheorie

- ▶ Behavioral Finance; Auflösung des Homo oeconomicus, Framing Effekte
- ▶ Prospect Theorie: Kahneman & Tversky (1979, Econometrica)
 - * Vorverarbeitung des Entscheidungsproblems in Editing Phase
 - * Wertfunktion (\neq Erwartungsnutzen) bezogen auf Referenzpunkt
 - * Transformation der Wahrscheinlichkeiten durch Gewichte
- ▶ Rank-dependent Expected Utility Theory: Quiggin (1982, JEcon Beh Org)
 - * Gewichtsfunktion auf Verteilungsfunktion bezogen
 - * Wahrscheinlichkeitsgewichtung abhängig von Rangplatz(z.B. von Pessimisten wird die Wahrscheinlichkeit des ungünstigsten Ereignisses zu hoch eingeschätzt)

Erste Schritte zur mathematischen Modellierung des Ellsberg-Experiments

Situation beschreibbar:

- ▶ einerseits durch *Mengen* klassischer Wahrscheinlichkeitsmasse im Sinne Kolmogorovs. Die priori Information besteht aus der Menge aller klassischen Wahrscheinlichkeitsmasse $\pi(\cdot)$ auf $(\{r, g, s\}, P(\{r, g, s\}))$, womit $\pi(\{r\}) = \frac{1}{3}$ und $\pi(\{g, s\}) = \frac{2}{3}$.
- ▶ Modellierung durch *intervallwertige Wahrscheinlichkeiten*

$$\pi(\{r\}) = \left[\frac{1}{3}; \frac{1}{3} \right], \quad \pi(\{r, g\}) = \left[\frac{1}{3}; 1 \right]$$

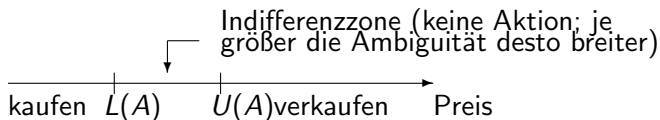
$$\pi(\{g\}) = \left[0; \frac{2}{3} \right], \quad \pi(\{r, s\}) = \left[\frac{1}{3}; 1 \right]$$

$$\pi(\{s\}) = \left[0; \frac{2}{3} \right], \quad \pi(\{g, s\}) = \left[\frac{2}{3}; \frac{2}{3} \right]$$

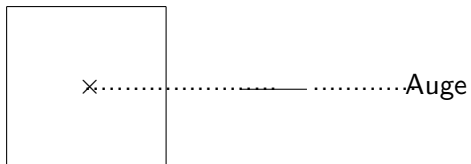
Die Idee ist verallgemeinerbar!

Imprecise Probabilities – Intervallwahrscheinlichkeit

- ▶ P : Ereignisse \rightarrow Intervalle in $[0; 1]$
- ▶ $P(A) = [L(A), U(A)]$
- ▶ Breite drückt Ambiguität aus
- ▶ Idee sehr naheliegend (spätestens seit Boole, 1854)
- ▶ Axiomatisierung
 - ▶ Spezielle Teilklassen: Dempster (1966), Huber (1973), Shafer (1976)
 - ▶ Allgemein:
Walley (1991): behavioristisch
Weichselsberger (2000, 2001): interpretationsunabhängig



Projektion



Positive und negative Symmetrie

- ▶ Keine der beiden Farben (gelb oder schwarz) ist wahrscheinlicher:

- * $\pi(\{s\}) = \pi(\{g\}) = \frac{1}{3} !?$

- * $\pi(\{s\}) = \pi(\{g\}) = [0; \frac{2}{3}]$

- ▶ Unterscheide zwischen

- ▶ Wissen um Symmetrie
("positive Symmetrie")

und

- ▶ Nichtwissen von Asymmetrie
("negative Symmetrie")
- ▶ $\frac{2}{3}$ gelbe und schwarze Kugeln in unbekannter Zusammensetzung

Society for Imprecise Probabilities, Theories and Applications:
www.sipta.org

- ▶ Modellierung partiellen Wissens in der Entscheidungstheorie
- ▶ Weiterrechnen mit Konfidenzintervallen
- ▶ Ordinale Wahrscheinlichkeiten z.B.
 $\pi(\{\vartheta_{(1)}\}) \leq \pi(\{\vartheta_{(2)}\}) \leq \pi(\{\vartheta_{(3)}\})$ beschreibt Credal-Menge
- ▶ Modellierung Unsicheren Wissens in Expertensystemen
(Medizin, auch Wirtschaftswissenschaften)
- ▶ Gruppenentscheidungen: verschiedene klassische Prioris.

Typische Modellklassen und Anwendungen II

- ▶ Formaler Überbau über Robuste Statistik: Umgebungsmodelle
- ▶ Robuste Bayes Analyse: Mengen von Prioris
- ▶ Vorsichtige Analyse bei unvollständigen Daten (Manski (2003): Partial Identification)
- ▶ Credal Classification
- ▶ Formaler Überbau über Sensitivitätsanalyse
- ▶ Finanz- und Finanzierungsmathematik (auch enger Zusammenhang zu Risikomaßen, Knightian Uncertainty)
- ▶ Alternative Inferenzmethoden, insbesondere logischer Wahrscheinlichkeitsbegriff (Dempster, Levi, Kyburg, Hampel, Weichselberger/Wallner)
- ▶ Modellierung von Priori-Daten-Konflikten

Manski's Law of Decreasing Credibility

- ▶ The credibility of inference decreases with the strength of the assumptions maintained.

C. Manski. (2003, Partial Identification, Springer)

- ▶ „Unschärfere“, aber dafür zuverlässigere Aussagen

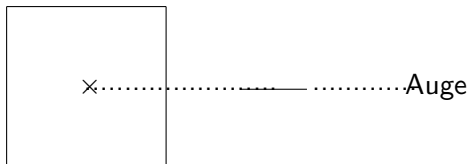
Modellierung partiellen Wissens in der Entscheidungstheorie

- ▶ Ambiguität in Intervallbreite ausdrücken
- ▶ Adäquate Modellierung von partiellem Wissen
- ▶ Intervallwertiger Erwartungsnutzen (Choquet Erwartungsnutzen, Gamma-Maximin)
- ▶ Extremfälle:
 - ▶ perfekte probabilistische Information: einpunktiges Intervall → klass. Erwartungsnutzen
 - ▶ vollständiges Nichtwissen: Wahrscheinlichkeit $[0; 1]$, Maximin Lösung (Spieltheorie), Schleier des Nichtwissens
- ▶ Effiziente Berechnung durch lineare Optimierung (Utkin & Augustin (2005), Kikuti et al. (2005))

Bayes-Inferenz über Mittelwert bei Normalverteilung

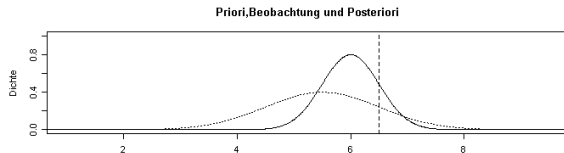
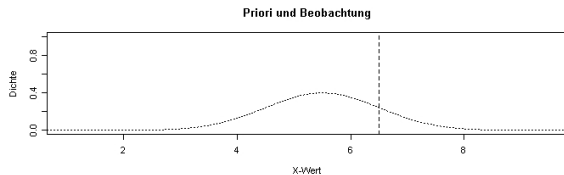
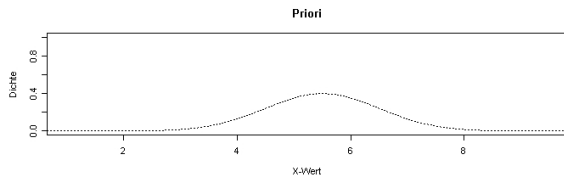
- ▶ Stichprobe(nmittel) $X \sim N(\theta, 1): f(x|\theta)$
- ▶ Priori $\theta \sim N(\nu, 1): \pi(\theta)$
- ▶ Posteriori $\theta|x \sim N(\frac{1}{2} \cdot (\nu + x), 0.5); \pi(\theta|x)$

Projektion



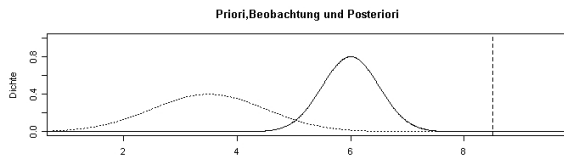
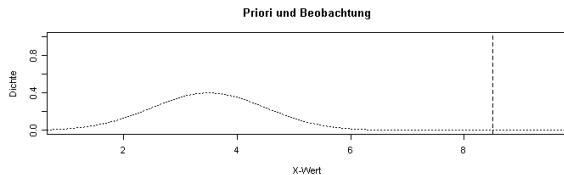
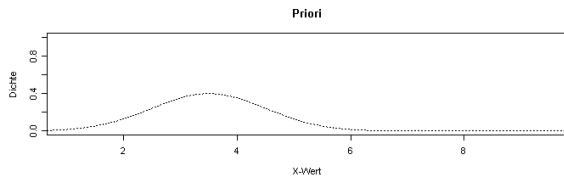
Priori-Daten-Konflikt, Beispiel 1

Beispiel 1: $\nu = 5.5$, $x = 6.5$, $\frac{\nu+x}{2} = 6$

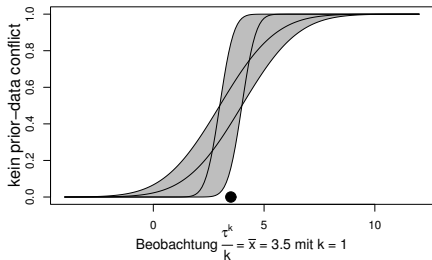


Priori-Daten-Konflikt, Beispiel 2

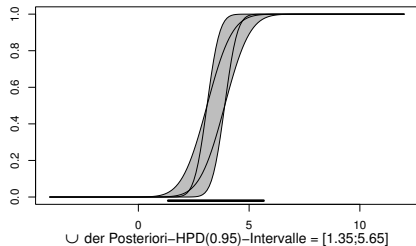
Beispiel 2: $\nu = 3.5$, $x = 8.5$, $\frac{\nu+x}{2} = 6$



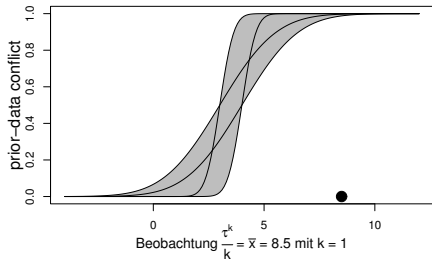
Menge der Priorverteilungen mit $y^{(0)}$ in [3;4] und $n^{(0)}$ in [0.25;4]



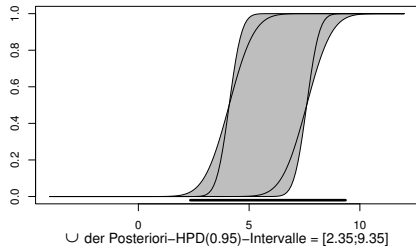
Menge der Posteriorverteilungen mit $y^{(1)}$ in [3.1;3.9] und $n^{(1)}$ in [1.25;5]



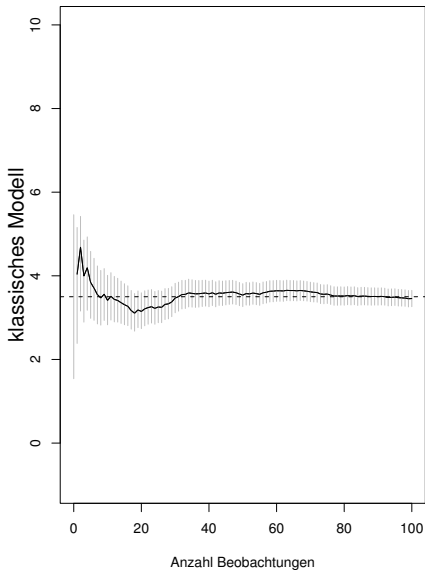
Menge der Priorverteilungen mit $y^{(0)}$ in [3;4] und $n^{(0)}$ in [0.25;4]



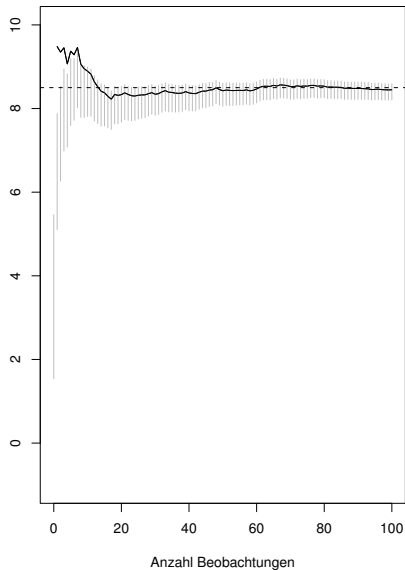
Menge der Posteriorverteilungen mit $y^{(1)}$ in [4.1;7.6] und $n^{(1)}$ in [1.25;5]



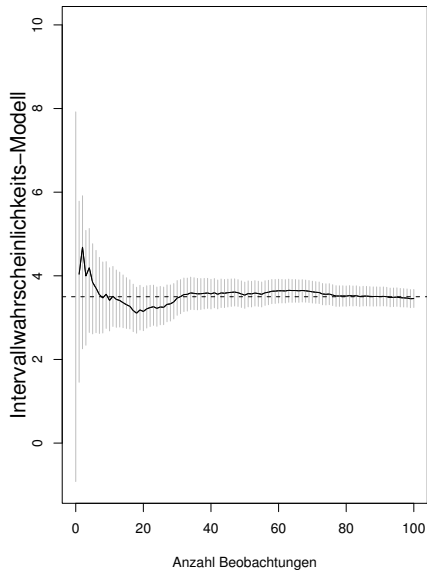
\cup der HPD(0.95) für $y^{(0)} \in [3.5;3.5]$, $n^{(0)} \in [1;1]$



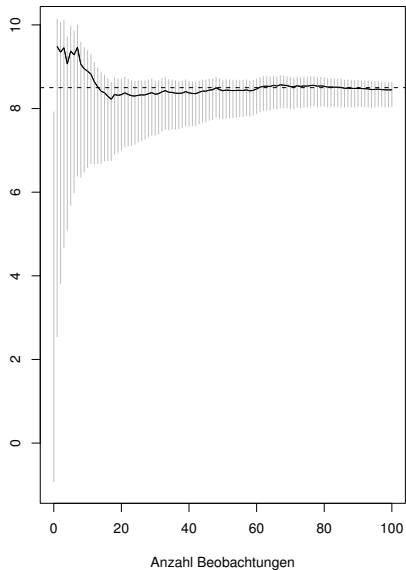
\cup der HPD(0.95) für $y^{(0)} \in [3.5;3.5]$, $n^{(0)} \in [1;1]$



\cup der HPD(0.95) für $y^{(0)} \in [3;4]$, $n^{(0)} \in [0.25;4]$



\cup der HPD(0.95) für $y^{(0)} \in [3;4]$, $n^{(0)} \in [0.25;4]$



Vereinfachtes Argumentationsmuster

