

3 Entscheidungstheorie unter einem allgemeineren Wahrscheinlichkeitsbegriff

3.1 Vorbemerkung: Unsicherheitstheorien

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”.

„Echtes Entscheidungsproblem unter Unsicherheit“

PARTIELLES WISSEN über Eintretensneigung von Umweltzuständen

solche
situationen

nein

engeren Sinn
wird gleichgesetzt
vielleicht gegen die
klar

ja

(virtuelle) Risikosituation mit
subjektiven Wahrscheinlichkeiten
||
wird gleichgesetzt
Risikosituation mit objektiven
Wahrscheinlichkeiten

klar

Erwartungsnutzen

Bayes

ja, wenn man allgemeineren
Begriff von Wahrscheinlichkeit
verwendet

3.2 Das Ellsberg-„Paradox“ – Vom Ungenügen klassischer Wahrscheinlichkeit

Zur Person: Daniel Ellsberg (geb. 1931 –)

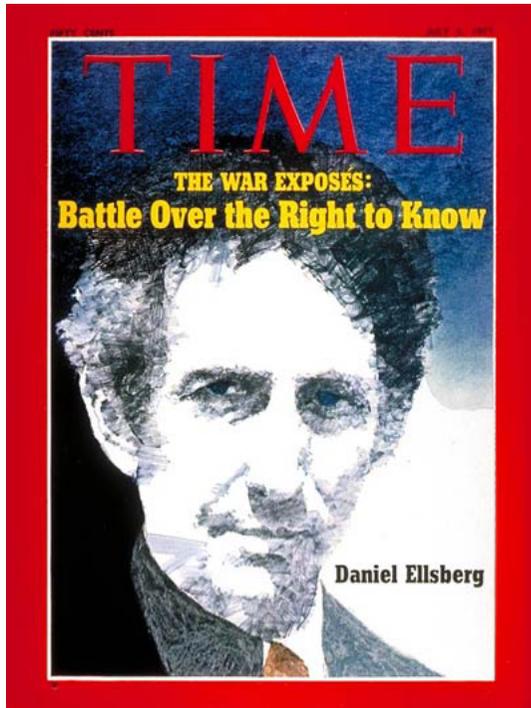


Foto: Time Magazine (July 1971)

Bsp. 3.1 (Das Ellsberg-Experiment)

a) **Das Gedankenexperiment:** Gedankenexperiment in Harvard unter Statistikern und Ökonomen, also Forschern die mit Wahrscheinlichkeitsrechnung vertraut sind, über ihre Präferenzen in folgendem Experiment:

- 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
- Situation I:
 - * Man darf wählen zwischen
 - a_1 : 100\$, falls „rot“ gezogen wird
und
 - a_2 : 100\$, falls „schwarz“ gezogen wird
 - * was würden Sie bevorzugen?

* typische Antwort

$$a_1 \succ a_2$$

● Situation II:

* Nun wird gelb zum Joker und man kann wählen zwischen

a_3 : 100\$; falls „rot“ oder „gelb“ gezogen wird
und

a_4 : 100\$; falls „schwarz“ oder „gelb“ gezogen wird

* typische Antwort

$$a_3 \prec a_4$$

b) **Die Priori im Hintergrund**

Laut dem Bayesianischen Paradigma (vgl. Kapitel 2.66) gibt es zu jeder Unsicherheitssituation eine (klassische) Wahrscheinlichkeitsbewertung $\pi(\cdot)$, die die subjektive Einschätzung der Unsicherheit und die daraus abgeleiteten Präferenzen beschreibt. Geht man davon aus, dass jeder Entscheidungsträger seinen subjektiven Erwartungsnutzen maximiert, so kann man $\pi(\cdot)$ durch das beobachtetes Verhalten näher charakterisieren:

Aus Daniel Ellsbergs "Risk, Ambiguity, and the Savage Axioms", p. 655f:

[...] Responses do vary. There are those who do *not* violate the axioms, or say they won't, even in these situations (e.g., G. Debreu, R. Schlaiffer,

P. Samuelson); such subjects tend to apply the axioms rather than their intuition, and when in doubt, to apply some form of the Principle of Insufficient Reason. Some violate the axioms cheerfully, even with gusto (J. Marschak, N. Dalkey); others sadly but persistently, having looked into their hearts, found conflicts with the axioms and decided, in Samuelson's phrase, to satisfy their preferences and let the axioms satisfy themselves. Still others (H. Raiffa) tend, intuitively, to violate the axioms but feel guilty about it and go back into further analysis.

The important finding is that, after rethinking all their „offending“ decisions in the light of the axioms, a number of people who are not only sophisticated but reasonable decide that they wish to persist in their choices. This includes people who previously felt a „first-order commitment“ to the axioms, many of them surprised and some dismayed to find that they wished, in these situations, to violate the Sure-thing Principle. Since this group included L. J. Savage, when last tested by me

(I have been reluctant to try him again), it seems to deserve respectful consideration.[...]

Bem. 3.2 (Konsequenzen aus dem Ellsberg-Experiment I)

- i) Das Ellsberg-Experiment wendet sich eigentlich ursprünglich gegen das sog. sure-thing-principle, wie es von Savage (1954) zur Fundierung seines Ansatzes verwendet wurde. Dieses Konzept wird in der Vorlesung nicht detailliert betrachtet, entspricht aber letztendlich dem hier verwendeten Bayes-Ansatz.
- ii) Bei der zugrundegelegenen Mehrheitsentscheidung in dem Gedankenexperiment muss für die zugrunde gelegte Priori-Verteilung also gelten:

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

\implies Es kann *keine klassische* Wahrscheinlichkeitsverteilung über $(\Theta, \mathcal{P}(\Theta))$ geben, die die am häufigsten beobachteten Präferenzen widerspiegelt.

iii) Man beachte, dass es sich hier um ein *Gedanken*-Experiment handelt. Es geht also nicht darum, dass sich Entscheidungsträger in konkreten Situationen – etwa unter dem Einfluss von Werbung oder Suggestivformulierungen – irrational verhalten, sondern darum, dass das oben beschriebene Verhalten von erfahrenen Ökonomen und Statistikern als *rational* empfunden wird. Viele betonen, dass sie bei der Präferenz bleiben, auch wenn sie die Axiome der Wahrscheinlichkeitsrechnung verletzt.

Es gibt also Präferenzen in Entscheidungssituationen unter Unsicherheit, die als *rational* empfunden werden, die aber *nicht direkt* mittels *klassischer* Wahrscheinlichkeit beschreibbar sind.

iii) Ein ähnlicher Widerspruch ergibt sich für die am zweithäufigsten gewählte Präferenz $a_1 \prec a_2$ und $a_3 \succ a_4$.

iv) Die Einfachheit des Beispiels stützt das Argument.

Bem. 3.3 (Konsequenzen aus dem E.-Experiment II:)

Nichtadditivität; Ambiguität

- Das zentrale Problem läuft technisch darauf hinaus, dass

$$\pi(\{g, s\}) \neq \pi(\{g\}) + \pi(\{s\}), \quad (3.1)$$

was natürlich das Additionsaxiom für Wahrscheinlichkeiten verletzt.

- „verwerfe solche Wahrscheinlichkeitsbewertungen als irrational“
- versus: „Diese Präferenzen sind aber nicht irrational: Nichtadditivität von Sicherheiten.“

- These: „Es gibt Präferenzen, die – auch nach Reflexion – als rational empfunden werden, aber nicht mittels klassischer Wahrscheinlichkeit beschreibbar sind. Soll die Entscheidungstheorie ihrem Anspruch als Theorie des *rationalen Entscheidens* unter Unsicherheit gerecht werden, so muss sie solche Situationen modellieren können.“
- Das Problem rührt daher, dass sozusagen die Wahrscheinlichkeitsbewertungen „unterschiedlich sicher“ sind. Bei $\{g, s\}$ ist die gesamte Unsicherheit auf *Zufälligkeit* reduziert (Anteil $\frac{2}{3}$ bekannt), während bei $\{g\}$ zusätzlich der Anteil unbekannt ist. Dieser nichtstochastische Aspekt (*Unbestimmtheit, Ambiguity*) der Unsicherheit stellt ein konstitutives Element der Entscheidungssituation dar und muss geeignet modelliert werden.

- So gilt die Ausgangssituation auch als typisch bei der Konstruktion subjektiver Wahrscheinlichkeiten aus Expertenstatements der Form „ziemlich sicher, dass Krankheit r oder g “, aber „Evidenz nicht aufteilbar zwischen $\{r\}$ und $\{g\}$ “.

Bem. 3.4 Exkurs: Neuronale Unterscheidung von Risiko und Ambiguität⁹

	Funktion	Geschwindigkeit
Amygdala	Überwachung	schnell
OFC		
Striatum	Belohnungs- erwartungs- system	nachge- langsam ordnet

⁹Quelle: Hsu, M., et al. (2005, p. 1681)

OFC: Orbitofrontaler Kortex

Experimentelle Behandlungen

- **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

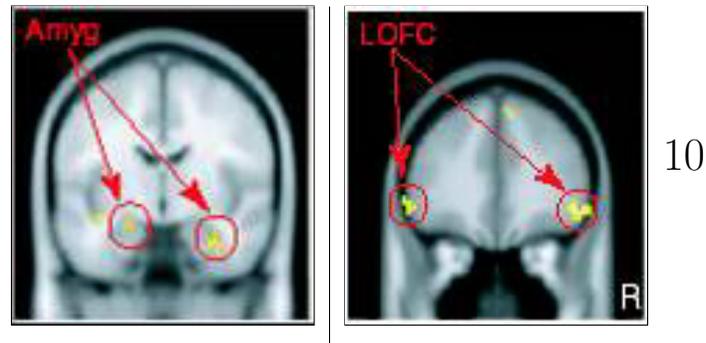
- **Wissen**

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

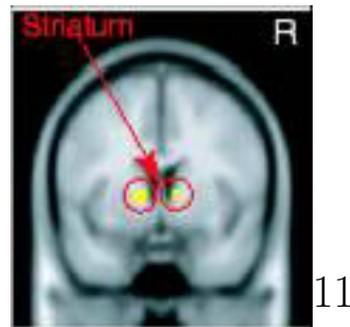
Bsp.: Wetter in New York – Wetter in Tirana

Aktivität von Hirnregionen unter Risiko und 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*



¹⁰Quelle: Hsu, M., Bhatt, M., Adolphs, R., Tranel, D., Camerer C.: Neural systems responding to degrees of uncertainty in human decision-making, *Science* 310, 1681, 2005.

¹¹Quelle: Hsu, M., et al. (2005, p. 1682)

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

Bsp. 3.5 (Erste Modellierung des Ellsberg-Experiments)

Erste Schritte zur mathematischen Modellierung:

Die Situation ist beschreibbar:

- einerseits durch *Mengen* klassischer Wahrscheinlichkeitsmaße im Sinne Kolmogorovs. Die Priori-Information besteht aus der Menge aller klassischen¹² Wahrscheinlichkeitsmaße $\pi(\cdot)$ auf $(\{r, g, s\}, \mathcal{P}(\{r, g, s\}))$, mit $\pi(\{r\}) = \frac{1}{3}$ und $\pi(\{g, s\}) = \frac{2}{3}$.

¹²Zur Unterscheidung von intervallwertigen Wahrscheinlichkeiten werden Wahrscheinlichkeitsbewertungen im üblichen Sinn als klassische Wahrscheinlichkeit bezeichnet. Für klassische Wahrscheinlichkeiten werden kleine Buchstaben verwendet, für intervallwertige große Buchstaben.

- andererseits durch *Intervalle*, intervallwertige Wahrscheinlichkeiten:

$$\begin{aligned} \Pi(\{r\}) &= \left[\frac{1}{3}; \frac{1}{3} \right], & \Pi(\{r, g\}) &= \left[\frac{1}{3}; 1 \right] \\ \Pi(\{g\}) &= \left[0; \frac{2}{3} \right], & \Pi(\{r, s\}) &= \left[\frac{1}{3}; 1 \right] \\ \Pi(\{s\}) &= \left[0; \frac{2}{3} \right], & \Pi(\{g, s\}) &= \left[\frac{2}{3}; \frac{2}{3} \right] \end{aligned}$$

Die Idee ist verallgemeinerbar!

3.3 Grundbegriffe verallgemeinerter Wahrscheinlichkeiten

3.3.1 Credalmengen, Intervallwahrscheinlichkeiten und ihre Strukturen

Ziel: Verallgemeinerte Wahrscheinlichkeitsrechnung:

- konsequente Verallgemeinerung der klassischen traditionellen Wahrscheinlichkeitsrechnung, d.h. des auf die Axiomatik von Kolmogorov aufbauenden Kalküls, das als Spezialfall mitenthalten sein soll und als Richtschnur dient. Jede Mengenfunktion $p(\cdot)$, die Axiome von Kolmogorov erfüllt, wird im folgenden zur Unterscheidung als *klassische Wahrscheinlichkeit* bezeichnet.
- angemessene Berücksichtigung des Ausmaßes an Ambiguität

- Zuverlässigkeit statt (Über)präzision.

Manski's *Law of Decreasing Credibility*: „The credibility of inferences decreases with the strength of the assumptions maintained.“ (Manski, 2003, p. 1)

Bem. 3.6 Grundlegende Ansätze, Credalmenge, Intervallwahrscheinlichkeit und Struktur

Ähnlich wie bei der Modellierung des Ellsberg-Experiments gibt es auch allgemein zwei (verwandte (s.u.)) Ansätze zu Verallgemeinerten Wahrscheinlichkeiten über einem Messraum (Ω, \mathcal{A}) :

a) *Credalmengen*

- *Mengen* \mathcal{M} klassischer Wahrscheinlichkeiten als Grundentität: Die Menge in ihrer Gesamtheit beschreibt die Wahrscheinlichkeit; kein Element ist wahrscheinlicher als ein anderes. (Legen einer Verteilung über diese Menge würde nach Mischungsprozess auf eine klassische Wahrscheinlichkeit führen, also das Grundproblem nicht lösen.)
- \mathcal{M} heißt dann *Credal-Menge* (Isaac Levi: Credal Set)
- „Größe“ der Menge reflektiert Ausmaß der Ambiguität:

- * perfekte probabilistische Information, ideale Stochastizität, Risikosituation
→ Menge, die aus einem Punkt besteht; klassische Wahrscheinlichkeit als Spezialfall

- * Unsicherheit i.e.S., volle Ambiguität
→ Menge *aller* Wahrscheinlichkeitsverteilungen auf dem gegebenem Messraum

- Sehr allgemeine Modellbildung möglich: Man trifft eine Aussage nur über die Wahrscheinlichkeitskomponenten derjenigen Ereignisse, über die man etwas weiß; die restlichen Werte ergeben sich implizit.

b) Intervallwahrscheinlichkeit

- Bei einem Meßraum (Ω, \mathcal{A}) betrachtet man statt der Abbildung

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

die Abbildung

$$\begin{aligned} P : \mathcal{A} &\rightarrow \mathcal{Z}_0([0; 1]) \\ A &\mapsto P(A) = [\underbrace{L}_{\text{lower}}(A), \underbrace{U}_{\text{upper}}(A)], \end{aligned}$$

wobei $\mathcal{Z}_0([0; 1])$ die Menge aller abgeschlossenen Intervalle in $[0; 1]$ ist.¹³

¹³Es sei nochmals an die Konvention erinnert: klassische Wahrscheinlichkeit wird mit Kleinbuchstaben symbolisiert, Intervallwahrscheinlichkeit mit Großbuchstaben

c) Formal sind beide Ansätze eng verwandt, denn es gilt:

- Aus jeder nichtleeren Menge \mathcal{M} von klassischen Wahrscheinlichkeiten kann man eine Intervallbewertung erzeugen, indem man festsetzt:

$$L(A) = \inf_{p \in \mathcal{M}} p(A) \quad U(A) = \sup_{p \in \mathcal{M}} p(A), \quad \forall A \in \mathcal{A} \quad (3.2)$$

- Umgekehrt kann man zu jeder intervallwertigen Bewertung $P(\cdot) = [L(\cdot), U(\cdot)]$ die Menge aller damit kompatiblen klassischen Wahrscheinlichkeiten betrachten:

$$\mathcal{M} := \{p(\cdot) \mid p(\cdot) \text{ ist klassische Wahrscheinlichkeit} \wedge \quad (3.3) \\ \mid L(A) \leq p(A) \leq U(A), \forall A \in \mathcal{A}\}.$$

* und dann heißt \mathcal{M} *Struktur* (oder Core (Kern)) der Intervallbewertung $P(\cdot)$.

* Ist $\mathcal{M} \neq \emptyset$, so heißt $P(\cdot)$ *R-Wahrscheinlichkeit*

* Gilt zudem für alle $A \in \mathcal{A}$:

$$L(A) = \inf_{p \in \mathcal{M}} p(A) \quad \text{und} \quad U(A) = \sup_{p \in \mathcal{M}} p(A), \quad (3.4)$$

so spricht man von *F-Wahrscheinlichkeit* oder Intervallwahrscheinlichkeit (im eigentlichen Sinn).

Bem. 3.7 Einige weitere Bemerkungen zur Intervallwahrscheinlichk

- $L(\cdot)$ und $U(\cdot)$ werden häufig als *Kapazität* (oder *fuzzy-Maß*) bezeichnet.
- Innerhalb des Intervalls herrscht völlige Unsicherheit; kein Wert ist „wahrscheinlicher“.

- perfekte probabilistische Information, ideale Stochastizität: $L(\cdot) = U(\cdot)$, das Intervall besteht aus einem Punkt \rightarrow klassische Wahrscheinlichkeit als Spezialfall.
- Unsicherheit i.e.S., volle Ambiguität führt auf das Intervall $[0;1]$. Dies ist das korrekte Modell für absolutes Nichtwissen bezüglich Wahrscheinlichkeiten; man weiß nur, dass die Wahrscheinlichkeitskomponente jedes Ereignissen zwischen 0 und 1 liegt.
- Obige Definition der Intervallwahrscheinlichkeit ist konsequent interpretationsunabhängig (Weichselberger (2000, 2001)): direkte Verallgemeinerung der Kolmogorovschen Axiomatik.

- Es gibt auch zwei operationale Definitionen von Intervallwahrscheinlichkeiten, die – ähnlich wie in der klassischen Theorie – umgekehrt auch als Interpretationen einer axiomatischen Vorgehensweise gesehen werden können:
 - * subjektivisch: untere und obere Wettquotienten (behavioristisch, Walley, wohl am stärksten verbreitet), Verallgemeinerung des de Finetti-Ansatzes. Wetten mit unterschiedlichem Kauf- und Verkaufspreis.

- * frequentistisch: Einhüllen der Häufigkeitspunkte nicht notwendig konvergierender Folgen relativer Häufigkeiten; verallgemeinerte von Mises-Kollektive.
- Unabhängig davon, ob man eine frequentistische oder subjektivistische Interpretation bevorzugt, gibt es zwei verschiedene Auffassungen wie man die Struktur und die Intervalle (bzw. auch eine Credalmenge) versteht:
 - * Die sog. *epistemische* Sicht beruht auf der Vorstellung, es gebe eine wahre klassische Wahrscheinlichkeit, die man aber nur (noch?) nicht kennt. Die Credalmenge bzw. die Intervalle zeichnen dann einfach Bereiche aus, in denen die klassische Wahrscheinlichkeit liegt. (Situation im Ellsberg-Experiment; es gibt einen wahren Anteil der gelben Kugeln. Das ist auch die Sicht des Fundamentaltheorems von de Finetti).

* Demgegenüber versteht der *ontologische* Ansatz Wahrscheinlichkeit grundsätzlich als intervallwertige/mengenwertige Entität.

Für die Modellierung spielt die Unterscheidung erst bei komplexeren Modellen (z.B. unabhängige Koppelung) eine wichtige Rolle, weswegen hier nicht genauer darauf eingegangen wird.

- Die Idee mit intervallwertigen Wahrscheinlichkeiten zu arbeiten, ist natürlich naheliegend und beileibe nicht neu (spätestens seit Boole vor ca. 150 Jahren)
Relativ neu ist aber die Existenz eines leistungsfähigen Kalküls basierend auf einer „sauberen“ Axiomatik.

- Bemerkung: Die Begriffe R- und F-Wahrscheinlichkeiten stammen von Weichselberger (2000, 2001), der zeigt, dass man damit bereits eine interpretationsunabhängige tragfähige Axiomatik bilden kann. (Zusatzforderungen an $L(\cdot)$ und $U(\cdot)$, die die Flexibilität einschränken würden, sind nicht notwendig.)
- Bei R-Wahrscheinlichkeit ist per definitionem die Menge \mathcal{M} aus (3.4) nicht leer. Die Intervallbewertung $P(\cdot) = [L(\cdot), U(\cdot)]$ ist im „wahrscheinlichkeitsbezogenen Sinn“ nicht widersprüchlich; es gibt klassische Wahrscheinlichkeiten, die mit den Intervallgrenzen $L(\cdot)$ und $U(\cdot)$ verträglich sind. Dies ist eine Minimalforderung an Bewertungen im Wahrscheinlichkeitskontext, allerdings können $L(\cdot)$ und/oder $U(\cdot)$ „zu weit“ sein, d.h. es kann Ereignisse $A \in \mathcal{A}$ geben mit.

$$L(A) < p(A) \quad \forall p \in \mathcal{M}$$

und/oder

$$U(A) > p(A) \quad \forall p \in \mathcal{M}.$$

- * Allgemeine Überprüfung, ob R-Wsk und F-Wahrscheinlichkeit vorliegt: lineare Optimierung (Weichselberger 2001, Kap. 4.1)

- * Subjektivistisch interpretiert bedeutet R-Wahrscheinlichkeit:
Es können keine Wettsysteme aufgestellt werden, die zu sicherem Verlust führen (Walley (1991): *avoiding sure loss*), eventuell ist aber, wegen „zu weiter Grenzen“, die Indifferenz zu stark ausgeprägt, und es werden dann vorteilhafte Wetten nicht akzeptiert.
Es gibt also mehrere R-Wahrscheinlichkeiten mit derselben Struktur.

- Bei F-Wahrscheinlichkeit passen hingegen Intervallgrenzen und Struktur in eineindeutiger Weise zusammen. Keine der Intervallgrenzen ist zu weit. Die in der Struktur enthaltenen probabilistischen Information wird in den Intervallgrenzen voll wiedergespiegelt. Das Weltverhalten ist „kohärent“ (Walley (1991), *coherent*).

* Ermittelt man aus den Grenzen die Struktur, verwendet diese als Credalmenge und produziert über (3.2) eine Intervallbewertung $\tilde{P}(\cdot)$, so ist $P(\cdot) = \tilde{P}(\cdot)$; man erhält also in diesem Fall die Ausgangswahrscheinlichkeiten.

* Eine sehr schwache notwendige Bedingung für F-Wahrscheinlichkeiten ist

$$L(A) = 1 - U(A^C), \quad \forall A \in \mathcal{A}.$$

Bei $|\Omega| > 2$ ist dies aber keineswegs hinreichend.

- * Die Überprüfung, ob eine F-Wahrscheinlichkeit vorliegt, kann wieder mittels linearer Optimierung erfolgen (Weichselberger (2001, Kap. 4.2))

Bem. 3.8 Zentrale Grundlage zur numerischen Behandlung, Strukturen als konvexe Polyeder

Ist $\Omega = \{\omega_1, \dots, \omega_q\}$ endlich mit $|\Omega| = q$, so ist die Struktur jeder F-Wahrscheinlichkeit $\Pi(\cdot)$ über $(\Omega, \mathcal{P}(\Omega))$ ein konvexes Polyeder im \mathbb{R}^q (, wobei wieder klassische Wahrscheinlichkeit $p(\cdot)$ mit der Wahrscheinlichkeitsfunktion $f(\cdot)$ vermöge $f(\omega_\ell) = p(\{\omega_\ell\})$, $\ell = 1, \dots, q$, mit einem Punkt des \mathbb{R}^q identifiziert wird.).

Im Allgemeinen ist jede Credal-Menge, die durch *endlich* viele *lineare* Bedingungen an ihre Elemente, also an die klassische Wahrscheinlichkeiten $\pi(\cdot)$ beschrieben wird, ein konvexes Polyeder. Jede dieser Bedingungen hat somit die Form

$$\underline{\alpha} \leq \sum_{\ell=1}^q \alpha_\ell p(\{\omega_\ell\}) \leq \bar{\alpha}$$

für jeweils geeignete $\underline{\alpha}, \bar{\alpha}, \alpha_1, \dots, \alpha_m$.

Insbesondere ist dann jeweils die Extremalpunktmenge $\mathcal{E}(\mathcal{M})$ nicht leer und endlich.

3.3.2 Typische Modellklassen und Anwendungen

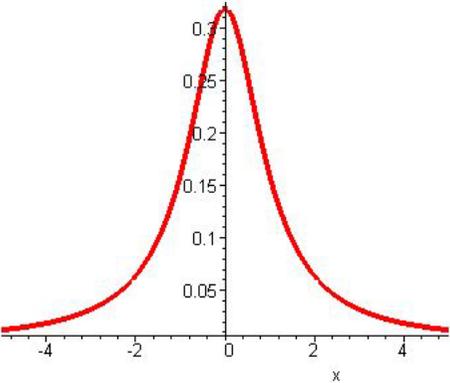
- Ordinale Wahrscheinlichkeiten, z.B. $\pi(\{\vartheta_{(1)}\}) \leq \pi(\{\vartheta_{(2)}\}) \leq \pi(\{\vartheta_{(3)}\})$, beschreibt eine Credal-Menge.
- Eine andere Interpretation/Anwendung von Credal-Mengen ist die Modellierung von Gruppenentscheidungen: verschiedene klassische Prior-Verteilungen.
- Kontaminationsmodelle: Robuste Statistik (s.u.)
- *Robuste Bayes Analyse*: Mengen von Prioris

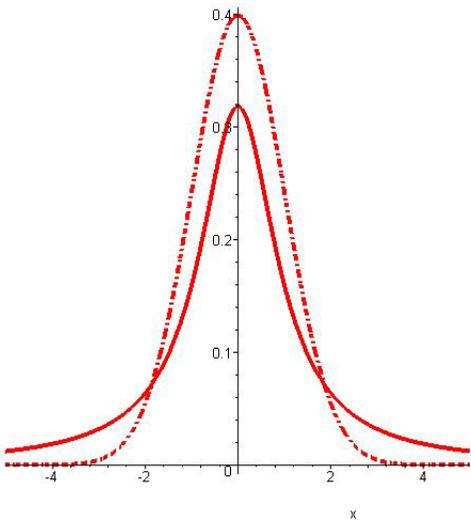
- Vorsichtige Analyse bei unvollständigen Daten (Manski (2003): *Partial Identification*). Betrachte die Menge aller mit den Daten kompatibler Modelle!
- Alternative Inferenzmethoden, insbesondere *logischer Wahrscheinlichkeitsbegriff* (Levi, Kyburg, Weichselberger/Wallner)
- Modellierung Unsicheren Wissens in Expertensystemen (Medizin, auch Wirtschaftswissenschaften); künstliche Intelligenz: Dempster-Shafer-Theorie
- Finanz- und Finanzierungsmathematik (auch enger Zusammenhang zur sog. Kohärenz von Risikomaßen)

- In manchen Anwendungen wird auch vorgeschlagen, bei der Modellbildung statt mit Plug-in-Punktschätzern mit Plug-in-Intervallschätzern zu arbeiten, um die Tatsache zu nutzen, dass eine Intervallschätzung für eine Wahrscheinlichkeit wesentlich mehr Information enthält als ein entsprechender Punktschätzer.
- Entscheidungstheorie (hier sogleich näher betrachtet)

Bem. 3.9 (Zum Hintergrund: Robustheit)

Grundlegend für die Entwicklung der robusten Statistik war die Erkenntnis, dass bei parametrischen Modellen optimale statistische Verfahren sich potentiell desaströs bei minimalen Abweichungen von der Verteilungsannahme verhalten.





Betrachtet man beispielsweise die Schätzung des Lageparameters aus einer i.i.d. Stichprobe X_1, \dots, X_n , so gilt für das arithmetische Mittel \bar{X} :

$$\text{i) } X_i \sim N(\mu, \sigma^2), i = 1, \dots, n, \quad \longrightarrow \bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

$$\text{ii) } X_i \sim \text{Cauchy}(a, b), i = 1, \dots, n \quad \longrightarrow \bar{X} \sim \text{Cauchy}(a, b).$$

Interpretation:

Man versucht deshalb, sich gegen solche Effekte zu „versichern“, indem man Verfahren entwickelt, die im „idealen Modell“ nicht ganz so leistungsfähig sind („Versicherungsprämie“), dafür aber bei kleiner Abweichung vom idealen Modell nicht zusammenbrechen.

Bem. 3.10 (Typische Modellklassen der robusten Statistik)

Man ersetzt ein „ideales statistisches Modell“ $(\Omega, \mathcal{A}, (p_\vartheta)_{\vartheta \in \Theta})$ mit präzisen Wahrscheinlichkeiten p_ϑ durch entsprechende Credalmengen \mathcal{M}_ϑ , die aus allen Verteilungen bestehen, die in einem gewissen Sinn nahe an $p_\vartheta(\cdot)$ sind. Die Credalmenge wird zum Beispiel über Wahrscheinlichkeitsmetriken (siehe z.B. Rüger (2002, S. 41ff)) beschrieben oder durch Schranken an die Dichten oder die Verteilungsfunktionen. Ein besonders gängiges Modell ist auch das ε -*Kontaminations-Modell*, bei dem man für „kleines“ $\varepsilon > 0$ betrachtet:

$$\begin{aligned} \mathcal{M}_{\vartheta, \varepsilon} = \{ & q(\cdot) \in \mathcal{P}(\Omega, \mathcal{A}) \mid \\ & q(\cdot) = (1 - \varepsilon)p_\vartheta(\cdot) + \varepsilon \cdot r(\cdot), \\ & r \in \mathcal{P}(\Omega, \mathcal{A}) \} \end{aligned} \quad (3.5)$$

mit $\mathcal{P}(\Omega, \mathcal{A})$ wieder als der Menge aller Wahrscheinlichkeitsmaße über (Ω, \mathcal{A}) .

3.4 Erste Anwendungen in der Entscheidungstheorie

3.4.1 Verallgemeinerte Erwartungswerte

intervallwertige Wahrscheinlichkeiten \Rightarrow intervallwertige Erwartungswerte
Zwei Definitionen in der Literatur üblich.

Def. 3.11 (Verallgemeinerte Erwartungswerte)

Betrachtet werden eine Zufallsvariable X auf einem Meßraum (Ω, \mathcal{A}) und eine F-Wahrscheinlichkeit $P(\cdot) = [L(\cdot); U(\cdot)]$ mit Struktur \mathcal{M} .

a) X heißt \mathcal{M} -integrierbar, wenn X für jedes $p \in \mathcal{M}$ p-integrierbar ist.

b) Dann heißt für \mathcal{M} -integrierbares X

$$\mathbb{E}_{\mathcal{M}}X := [\underline{\mathbb{E}}_{\mathcal{M}}X; \bar{\mathbb{E}}_{\mathcal{M}}X] := \left[\inf_{p \in \mathcal{M}} \mathbb{E}_p X; \sup_{p \in \mathcal{M}} \mathbb{E}_p X \right] \quad (3.3)$$

intervallwertiger Erwartungswert i.e.S.

b) Ist X zudem nicht negativ, so heißen die Ausdrücke

$$\int X \, dL := \int L(\{X > x\}) \, dx \quad (3.4)$$

$$\int X \, dU := \int U(\{X > x\}) \, dx \quad (3.5)$$

Choquet-Integrale zu $L(\cdot)$ und $U(\cdot)$.

Bem. 3.12 (Zu Def. 3.11)

- i) Die Definitionen in Teil a) und b) lassen sich unmittelbar auf Credal-Mengen ausdehnen.
- ii) Das Choquet-Integral ist v.a. in den Wirtschaftswissenschaften populär („*Choquet Expected Utility*“).
- iii) Ist Ω endlich und \mathcal{M} die Struktur einer F-Wahrscheinlichkeit oder allgemein ein konvexes Polyeder (vgl. Bemerkung 3.8), so können $\underline{\mathbb{E}}_{\mathcal{M}}X$ und $\bar{\mathbb{E}}_{\mathcal{M}}X$ einfach durch lineare Optimierung bestimmt werden:

$$\sum_{\omega \in \Omega} X(\omega)p(\{\omega\}) \longrightarrow \max_{p(\cdot)}$$

unter den durch \mathcal{M} beschriebenen Nebenbedingungen.

iv) Allgemein gilt

$$\int X \, dL \leq \underline{\mathbb{E}}_{\mathcal{M}} X \leq \overline{\mathbb{E}}_{\mathcal{M}} X \leq \int X \, dU \quad (3.6)$$

v) Unter gewissen Regularitätsbedingungen, die zum Beispiel bei endlichem Ω erfüllt sind, gilt:

$$\int X \, dL = \underline{\mathbb{E}}_{\mathcal{M}} X \quad \text{für alle } \mathcal{M}\text{-integrierbaren } X$$

genau dann, wenn $L(\cdot)$ *zwei-monoton*¹⁴ ist, d.h. wenn für alle $A, B \in \mathcal{A}$

$$L(A \cup B) + L(A \cap B) \geq L(A) + L(B) \quad (3.7)$$

gilt.

¹⁴Der Begriff der Zweimonotonie ist rein technischer Natur; lässt sich inhaltlich nicht direkt interpretieren.

Bem. 3.13 (Exkurs: Woher kommen die Formeln (3.4) und (3.5)?)

Es gilt allgemein für nichtnegatives X

$$\mathbb{E}_P X = \int_0^\infty p(\{X > x\}) \, dx, \quad (3.8)$$

sofern die rechte Seite endlich ist.

In dieser Beziehung $p(\cdot)$ durch $L(\cdot)$ und $U(\cdot)$ ersetzen.

Beweis von (3.8) durch partielle Integration:

$$\begin{aligned} \int_0^\infty p(\{X > x\}) \, dt &= \int_0^\infty \underbrace{(1 - F(x))}_u \cdot \underbrace{1}_{v'} \, dx = \\ &= \left[\underbrace{(1 - F(x))}_u \cdot \underbrace{x}_v \right]_0^\infty - \int_0^\infty \underbrace{-f(x)}_{u'} \underbrace{x}_v \, dx. \end{aligned}$$

Damit gilt, falls

$$\lim_{x \rightarrow \infty} (1 - F(x)) \cdot x = \lim_{x \rightarrow \infty} x - x \cdot F(x) = 0 \quad (3.9)$$

in der Tat

$$\int_0^\infty p(\{X > x\}) \, dt = 0 + \int_0^\infty x \cdot f(x) \, dx = \mathbb{E}X.$$

(3.9) stellt nur sicher, dass die rechte Seite von (3.8) endlich ist. Angenommen, es gibt ein x_0 so, dass $x - x \cdot F(x) \geq c > 0$ für alle $x \geq x_0$, dann wäre

$$\begin{aligned} & \int_0^{\infty} (1 - F(x)) \, dx = \\ &= \int_0^{x_0} (1 - F(x)) \, dx + \int_{x_0}^{\infty} (1 - F(x)) \, dx = \\ &\geq \int_0^{x_0} (1 - F(x)) \, dx + \int_{x_0}^{\infty} c \, dx = \infty \end{aligned}$$

Diskreter Spezialfall: kann X nur natürliche Zahlen annehmen, so ist

$$\mathbb{E}X = \sum_{n \in \mathbb{N}} P(\{X \geq n\})$$

Beweis:

$$\begin{aligned}
 \mathbb{E}X &= \sum_{n \in \mathbb{N}} n \cdot P(\{X = n\}) = \\
 &= 1 \cdot P(X = 1) + 2 \cdot P(X = 2) + 3 \cdot P(X = 3) + \dots \\
 &= P(X = 1) + \\
 &\quad P(X = 2) + P(X = 2) + \\
 &\quad \underbrace{P(X = 3)} + \underbrace{P(X = 3)} + \underbrace{P(X = 3)} + \\
 &\quad \dots \\
 &= P(X \geq 1) + P(X \geq 2) + P(X \geq 3) + \dots
 \end{aligned}$$

3.4.2 Verallgemeinerter Erwartungsnutzen und Optimalitätskriterien

Einen guten Überblick bietet: Troffaes (2007, International Journal Approximate Reasoning).

Def. 3.14 (Verallgemeinerter Erwartungsnutzen)

Gegeben sei ein datenfreies Entscheidungsproblem $(\mathbb{A}, \Theta, u(\cdot))$, und eine *verallgemeinerte Priori-Bewertung*, also eine Intervallwahrscheinlichkeit $\Pi(\cdot)$ mit Struktur $\mathcal{M} \neq \emptyset$ oder eine konvexe Credal-Menge $\mathcal{M} \neq \emptyset$ auf $(\Theta, \sigma(\Theta))$. Dann heißt

$$\mathbb{E}_{\mathcal{M}}(u(a)) = \left[\inf_{\pi \in \mathcal{M}} \mathbb{E}_{\pi}(u(a)); \sup_{\pi \in \mathcal{M}} \mathbb{E}_{\pi}(u(a)) \right]$$

der *verallgemeinerte Erwartungsnutzen* der Aktion a zur verallgemeinerten Priori-Bewertung $\Pi(\cdot)$ bzw. \mathcal{M} .

Bem. 3.15 (Intervallordnungen)

Da Intervalle in der Regel unvollständige Ordnungen produzieren, bedarf die Forderung „maximiere den verallgemeinerten Erwartungsnutzen $\mathbb{E}_{\mathcal{M}}(u(a))$ unter allen $a \in \mathbb{A}$ “ noch der genaueren Präzisierung, die sich in verschiedenen Optimalitätsbegriffen niederschlägt:

- Entweder: Aussagen analog zur Zulässigkeit: *E-Zulässigkeit*, *Intervallordnungen im eigentlichen Sinn*, die i.A. keine vollständige Ordnungen produzieren.

- oder alternativ: Intervalle durch *eine* reelle Zahl beschreiben (Mittelpunkt, untere Intervallgrenze, etc. ...): *Repräsentationen* z.B. durch minimalen Erwartungsnutzen oder gewichtete Summe des maximalen und minimalen Erwartungsnutzens.

Bem. 3.11 (Optimale Aktionen)

a1) Betrachtet werde die Situation von Def. 3.14. Dann heißt

$$\begin{aligned}\Phi(\cdot, \mathcal{M}) : \mathbb{A} &\longrightarrow \mathbb{R} \\ a &\longmapsto \underline{\mathbb{E}}_{\mathcal{M}}(u(a))\end{aligned}$$

Max E Min-Kriterium und jede Aktion $a^* \in \mathbb{A}$ mit

$$\Phi(a^*, \mathcal{M}) \geq \Phi(a, \mathcal{M}), \quad \forall a \in \mathbb{A},$$

also mit

$$\underline{\mathbb{E}}_{\mathcal{M}}u(a^*) \geq \underline{\mathbb{E}}_{\mathcal{M}}(u(a)), \quad \forall a \in \mathbb{A}$$

Max E Min Aktion zur verallgemeinerten Priori-Bewertung $\Pi(\cdot)$ bzw. \mathcal{M} .

Man würde dann genau erwarten, dass sich im Extremfall perfekte probabilistische Information des Bayeskriterium und bei völliger Ambiguität das Maximin-Kriterium ergibt. Dies ist in der Tat so, vgl. 3.4.4.

a2) Für $0 \leq \eta \leq 1$ heißt

$$\Phi_{\eta}(\mathcal{M}) : \mathbb{A} \longrightarrow \mathbb{R}$$

$$\mathbb{A} \longmapsto \eta \cdot \underline{\mathbb{E}}_{\mathcal{M}}(u(a)) + (1 - \eta)\bar{\mathbb{E}}_{\mathcal{M}}(u(a))$$

lineare Repräsentation mit der Vorsicht η und jede Aktion a^{η} mit*

$$\Phi_{\eta}(a^{*\eta}, \mathcal{M}) \geq \Phi_{\eta}(a, \mathcal{M}), \quad \forall a \in \mathbb{A},$$

also mit

$$\eta \cdot \underline{\mathbb{E}}_{\mathcal{M}} u(a^{*\eta}) + (1 - \eta) \overline{\mathbb{E}}_{\mathcal{M}} u(a^{*\eta}) \geq \\ \eta \cdot \underline{\mathbb{E}}_{\mathcal{M}}(u(a)) + (1 - \eta) \overline{\mathbb{E}}_{\mathcal{M}}(u(a)), \forall a \in \mathbb{A},$$

optimale Aktion zur verallgemeinerten Priori-Bewertung $\Pi(\cdot)$ bzw. \mathcal{M} bei der Vorsicht η .

- b) Eine Aktion $a_\pi \in \mathbb{A}$ heißt *E-zulässig (E-admissible)* bezüglich der verallgemeinerten Priori-Bewertung $\Pi(\cdot)$ bzw. \mathcal{M} , falls es ein $\pi(\cdot) \in \mathcal{M}$ gibt, so dass a_π Bayes-Aktion zu $\pi(\cdot)$ ist.

Bem. 3.12

- Der Begriff Max E Min (Maximiere den minimalen Erwartungswerts) geht auf die sogenannte Theorie der *linearen partiellen Information* von Kofler und Menges zurück, die in den Wirtschaftswissenschaften in den 80iger Jahren schon einmal populär war.

Andere gängige Bezeichnungen für dieses Kriterium sind insbesondere Γ -*Maximinkriterium* und in Situationen, in denen in (3.6) Gleichheit gilt, *Choquet-Erwartungsnutzen-Optimalität*.
- Dem Max E Min-Kriterium liegt eine skeptische Perspektive zugrunde, man führt sozusagen einen „lokalen Maximin-Ansatz“ durch, indem man sich auf das ungünstigste Element von \mathcal{M} konzentriert.

- Analog entspricht die lineare Repräsentation mit der Vorsicht η einem „lokalen Hurwicz-Kriterium“. Die Vorsicht η modelliert die Einstellung zur Ambiguität; $\eta > \frac{1}{2}$ bedeutet Ambiguitätsaversion, $\eta < \frac{1}{2}$ Ambiguitätsfreudigkeit. Der Fall $\eta = 1$ führt auf das Max E Min-Prinzip. (Man vergleiche hierzu auch Satz 3.14.)

3.4.3 Zurück zum Ellsberg-Experiment

Bsp. 3.13 (Modellierung des Ellsberg-Experiments)

Situation 1

a_1 auf $\{r\}$ setzen

a_2 auf $\{s\}$ setzen

	ϑ_1	ϑ_2	ϑ_3
	r	g	s
a_1	1	0	0
a_2	0	0	1

vgl. vorne: nochmals wieder in Situation 1 für jedes π

$$\mathbb{E}_\pi(u(a_1)) = \sum_{j=1}^3 u(a_1, \vartheta_j) \pi(\{\vartheta_j\}) = \pi(\{r\})$$

$$\mathbb{E}_\pi(u(a_2)) = \sum_{j=1}^3 u(a_2, \vartheta_j) \pi(\{\vartheta_j\}) = \pi(\{s\})$$

also mit

$$\mathcal{M} = \left\{ \pi(\cdot) \text{ Wsk auf } \mathcal{P}(\{r, s, g\}) \right. \\ \left. \left| \pi(\{r\}) = \frac{1}{3}; \pi(\{s, g\}) = \frac{2}{3} \right. \right\}.$$

$$\underline{\mathbb{E}}_{\mathcal{M}}(u(a_1)) = \inf_{\pi \in \mathcal{M}} \mathbb{E}_\pi(u(a_1)) = \inf_{\pi \in \mathcal{M}} \pi(\{r\}) = \frac{1}{3}$$

und

$$\underline{\mathbb{E}}_{\mathcal{M}}(u(a_2)) = \inf_{\pi \in \mathcal{M}} \mathbb{E}_\pi(u(a_2)) = \inf_{\pi \in \mathcal{M}} \pi(\{s\}) = 0$$

sowie

$$\bar{\mathbb{E}}_{\mathcal{M}}(u(a_1)) = \sup_{\pi \in \mathcal{M}} \mathbb{E}_{\pi}(u(a_1)) = \sup_{\pi \in \mathcal{M}} \pi(\{r\}) = \frac{1}{3}$$

und

$$\bar{\mathbb{E}}_{\mathcal{M}}(u(a_2)) = \sup_{\pi \in \mathcal{M}} \mathbb{E}_{\pi}(u(a_2)) = \sup_{\pi \in \mathcal{M}} \pi(\{s\}) = \frac{2}{3}.$$

Folglich ist mit

$$\Phi_{\eta}(a_i, \mathcal{M}) := \eta \cdot \underline{\mathbb{E}}_{\mathcal{M}}(u(a_i)) + (1 - \eta) \bar{\mathbb{E}}_{\mathcal{M}}(u(a_i))$$

$$\Phi_{\eta}(a_1, \mathcal{M}) = \eta \cdot \frac{1}{3} + (1 - \eta) \cdot \frac{1}{3} = \frac{1}{3}$$

$$\Phi_{\eta}(a_2, \mathcal{M}) = \eta \cdot 0 + (1 - \eta) \cdot \frac{2}{3} = (1 - \eta) \cdot \frac{2}{3}$$

und damit

$$\Phi_{\eta}(a_1, \mathcal{M}) > \Phi_{\eta}(a_2, \mathcal{M}) \iff \eta > \frac{1}{2}.$$

in Situation 2

a_3 auf $\{r\}$ oder $\{g\}$ setzen

a_4 auf $\{s\}$ oder $\{g\}$ setzen

	r	g	s
a_3	1	1	0
a_4	0	1	1

$$\mathbb{E}_\pi(u(a_3)) = \sum_{j=1}^3 u(a_3, \vartheta_j) \pi(\{\vartheta_j\}) = \pi(\{r\}) + \pi(\{g\})$$

$$\mathbb{E}_\pi(u(a_4)) = \sum_{j=1}^3 u(a_4, \vartheta_j) \pi(\{\vartheta_j\}) = \pi(\{g\}) + \pi(\{s\})$$

also

$$\underline{\mathbb{E}}_{\mathcal{M}}(u(a_3)) = \inf_{\pi \in \mathcal{M}} \mathbb{E}_{\pi}(u(a_3)) = \inf_{\pi \in \mathcal{M}} (\pi(\{r\}) + \pi(\{g\})) = \frac{1}{3}$$

und¹⁵

$$\underline{\mathbb{E}}_{\mathcal{M}}(u(a_4)) = \inf_{\pi \in \mathcal{M}} \mathbb{E}_{\pi}(u(a_4)) = \inf_{\pi \in \mathcal{M}} (\pi(\{g\}) + \pi(\{s\})) = \frac{2}{3}$$

sowie

$$\overline{\mathbb{E}}_{\mathcal{M}}(u(a_3)) = \sup_{\pi \in \mathcal{M}} \mathbb{E}_{\pi}(u(a_3)) = \sup_{\pi \in \mathcal{M}} (\pi(\{r\}) + \pi(\{g\})) = 1$$

¹⁵Hier ist Vorsicht geboten: Man muss die internen Restriktionen berücksichtigen; es muss „dasselbe $\pi(\cdot)$ sein, bezüglich dem das Infimum gebildet wird“:

$$\inf_{\pi \in \mathcal{M}} (\pi(\{g\}) + \pi(\{s\})) \neq \inf_{\pi \in \mathcal{M}} \pi(\{g\}) + \inf_{\pi \in \mathcal{M}} \pi(\{s\}).$$

Dies ist der wesentliche Unterschied zwischen Intervallwahrscheinlichkeiten und einer naiven Intervallarithmetik.

und

$$\bar{\mathbb{E}}_{\mathcal{M}}(u(a_4)) = \sup_{\pi \in \mathcal{M}} \mathbb{E}_{\pi}(u(a_4)) = \sup_{\pi \in \mathcal{M}} (\pi(\{g\}) + \pi(\{s\})) = \frac{2}{3}.$$

Damit ist hier mit dem Kriterium $\Phi_{\eta}(\cdot, \mathcal{M})$ von oben

$$\Phi(a_3, \mathcal{M}) = \eta \cdot \frac{1}{3} + (1 - \eta) \cdot 1 = 1 - \frac{2}{3}\eta$$

und

$$\Phi(a_4, \mathcal{M}) = \eta \cdot \frac{2}{3} + (1 - \eta) \cdot \frac{2}{3} = \frac{2}{3}$$

$$\Phi(a_4, \mathcal{M}) > \Phi(a_3, \mathcal{M}) \iff \frac{2}{3} > 1 - \frac{2}{3}\eta \iff \eta > \frac{1}{2}.$$

Die ambiguitätsaversen Personen ($\eta > \frac{1}{2}$) bevorzugen a_1 vor a_2 und a_4 vor a_3 , die ambiguitätsfreudigen a_2 vor a_1 und a_3 vor a_4 . Damit sind genau die am häufigsten geäußerten Präferenzen, die ja der klassischen Wahrscheinlichkeitsrechnung widersprochen haben, geeignet modelliert.

3.4.4 Verallgemeinerter Erwartungsnutzen als Überbau über die klassischen Kriterien der Entscheidungstheorie

Die Intervallwahrscheinlichkeit und der darauf aufbauende verallgemeinerter Erwartungsnutzen wurde als Mittelweg zwischen Bayes- und Maximin-Sicht motiviert. In der Tat gilt der folgende Satz:

Satz 3.14 (Klassische Entscheidungskriterien als Extrempole I)

In der Situation von Def. 3.16 gilt:

- a) Herrscht ideale Stochastizität, ist also $\mathcal{M} = \{\pi\}$ (einelementig), so gilt:
Eine Aktion $a^* \in \mathbb{A}$ ist genau dann optimal im Sinne der in Definition 3.11 genannten Kriterien, wenn a^* Bayes-Aktion (im klassischen Sinn von Kapitel 2.4) zu $\pi(\cdot)$ ist.

- b) Herrscht Unsicherheit i.e.S., besteht \mathcal{M} also aus allen klassischen Wahrscheinlichkeiten auf $(\Theta, \sigma(\Theta))$, so gilt:
Eine Aktion $a^* \in \mathbb{A}$ ist genau dann Max E Min Aktion zu \mathcal{M} , wenn a^* Maximin-Aktion im Sinne von Kapitel 2.3 ist, und optimale Aktion zu \mathcal{M} bei der Vorsicht η , wenn a^* Hurwicz-Aktion zum Optimismusparameter $1 - \eta$ im Sinne von Kapitel 2.5 ist.

Es lassen sich ähnliche Aussagen für die E-Zulässigkeit treffen:

Proposition 3.15 Klassische Entscheidungskriterien als Extrempole II

- i) Herrscht in der Situation von 3.16 ideale Stochastizität, ist also $\mathcal{M} = \{\pi\}$ (einelementig), so gilt: Die Menge der E-zulässigen Aktionen zu \mathcal{M} ist identisch zu der Menge der Bayes-Aktionen zu $\pi(\cdot)$.
- ii) Herrscht Unsicherheit i.e.S., besteht \mathcal{M} also aus allen klassischen Wahrscheinlichkeiten auf $(\Theta, \sigma(\Theta))$ und ist \mathbb{A} konvex, so gilt mit $\mathcal{M}^+ := \{\pi(\cdot) \in \mathcal{M} \mid \pi(\cdot) > 0\}$: $a^* \in \mathbb{A}$ ist genau dann zulässig (im Sinne von Kapitel 2.0), wenn a^* E-zulässig bezüglich \mathcal{M}^+ ist.

3.4.5 Berechnung optimaler Aktionen über die Extremalpunktmenge

Die Tatsache, dass gemäß Bem. 3.8 auf endlichen Räumen Intervall-Wahrscheinlichkeiten und weitere typische Credalmengen konvexe Polyeder sind, ist für die konkrete Berechnung optimaler Aktionen sehr hilfreich, denn sie erlaubt den Rückzug auf die endlichen, nichtleeren Extremalpunktmenge.

Satz 3.16 (Berechnung verallgemeinerter Erwartungsnutzen)

Gegeben sei ein datenfreies Entscheidungsproblem $(\mathbb{A}, \Theta, u(\cdot))$, und eine *verallgemeinerte Priori-Bewertung*, also eine Intervallwahrscheinlichkeit $\Pi(\cdot)$ mit Struktur $\mathcal{M} \neq \emptyset$ oder eine Credal-Menge $\mathcal{M} \neq \emptyset$ auf $(\Theta, \sigma(\Theta))$, die zusätzlich die Bedingungen von Bem. 3.8 erfüllt. Dann gilt mit $\mathcal{E}(\mathcal{M})$ als Extremalpunktmenge:

$$\underline{\mathbb{E}}_{\mathcal{M}}(u(a)) = \min_{\pi \in \mathcal{E}(\mathcal{M})} \sum_{j=1}^m u(a, \vartheta_j) \pi(\{\vartheta_j\})$$

und

$$\bar{\mathbb{E}}(u(a)) = \max_{\pi \in \mathcal{E}(\mathcal{M})} \sum_{j=1}^m u(a, \vartheta_j) \pi(\{\vartheta_j\}).$$

Bem. 3.17

- Dieser Satz ist eigentlich ein Korollar zu Satz 1.23 aus Kapitel 1, wenn man sich an Bemerkung 3.8 erinnert, demzufolge Strukturen und entsprechende Credalmengen konvexe Polyeder bilden.
- Zur Berechnung des verallgemeinerten Erwartungsnutzen brauchen also nur endlich viele klassische Erwartungswerte bestimmt werden.

Bsp. 3.18 (Investitionsproblem)

Gegeben sei das Investitionsproblem (vgl. das Beispiel aus Abschnitt 1.3.4)

	ϑ_1	ϑ_2	ϑ_3
a_1	10000	2000	-15000
a_2	1000	1000	0

und eine Experteneinschätzung in Form einer ordinalen Wahrscheinlichkeit, d.h. der Experte ordnet nur die Umweltzustände nach der Wahrscheinlichkeit ihres Eintretens. Gegeben sei die Credalmenge

$$\mathcal{M} = \{\pi(\cdot) \mid \pi(\{\vartheta_1\}) \geq \pi(\{\vartheta_2\}) \geq \pi(\{\vartheta_3\})\}.$$

Berechne die Max E Min-Aktion!

\mathcal{M} besteht aus linearen Restriktionen, die sich aus der ordinalen Wahrscheinlichkeitsbewertung ergeben, z.B.

$$\pi(\{\vartheta_{j_1}\}) \geq \pi(\{\vartheta_{j_2}\}) \iff 0 \leq \pi(\{\vartheta_{j_1}\}) - \pi(\{\vartheta_{j_2}\}),$$

und den trivialen Bedingungen:

$$0 \leq \pi(\{\vartheta_j\}) \leq 1,$$

$$\pi(\{\vartheta_1\}) + \pi(\{\vartheta_2\}) + \pi(\{\vartheta_3\}) = 1,$$

also ist \mathcal{M} in der Tat ein konvexes Polyeder.

Berechnet man die Extrempunktmenge $\mathcal{E}(\mathcal{M})$, so ergibt sich ¹⁶

¹⁶(siehe z. B. E. Kofler: Prognosen und Stabilität bei unvollständiger Information. Campus (Frankfurt; New York), 1989.)

$$\mathcal{E}(\mathcal{M}) = \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \\ 0 \end{pmatrix} \begin{pmatrix} \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{3} \end{pmatrix} \right\}.$$

Damit ist bei a_1

$$\begin{aligned} \underline{\mathbb{E}}_{\mathcal{M}}(u(a_1)) &= \inf_{\pi \in \mathcal{M}} \mathbb{E}_{\pi}(u(a_1)) = \min_{\pi \in \mathcal{E}(\mathcal{M})} \left(\sum_{j=1}^m u(a_1, \vartheta_j) \pi(\{\vartheta_j\}) \right) \\ &= \min \left(10000 \cdot 1 + 0; 10000 \cdot \frac{1}{2} + 2000 \cdot \frac{1}{2} + 0; \frac{1}{3}(10000 + 20000 - 15000) \right) \\ &= \min(10000; 6000; -1000) = -1000. \end{aligned}$$

und für a_2

$$\begin{aligned} \underline{\mathbb{E}}_{\mathcal{M}}(u(a_2)) &= \inf_{\pi \in \mathcal{M}} \mathbb{E}_{\pi}(u(a_2)) = \min_{\pi \in \mathcal{E}(\mathcal{M})} \left(\sum_{j=1}^m u(a_2, \vartheta_j) \pi(\{\vartheta_j\}) \right) \\ &= \min \left(1000 \cdot 1; 1000 \cdot \frac{1}{2} + 1000 \cdot \frac{1}{2}; \frac{1}{3}(1000 + 1000 + 0) \right) \\ &= \frac{2000}{3} \end{aligned}$$

$$\underline{\mathbb{E}}_{\mathcal{M}}(u(a_1)) = \min_{\pi \in \mathcal{E}(\mathcal{M})} \mathbb{E}_{\pi}(u(a_1)) < \min_{\pi \in \mathcal{E}(\mathcal{M})} \mathbb{E}_{\pi}(u(a_2)) = \underline{\mathbb{E}}_{\mathcal{M}}(u(a_2)).$$

Also ist a_2 Max E Min-Aktion.

Bem. 3.19 (Praktische Berechnung)

- Bei größeren Problemen ist die Berechnung der Extrempunktmenge durchaus aufwendig, aber es gibt spezielle Algorithmen (aus der „computational geometry“, die mittlerweile auch in R verwendbar sind).¹⁷
- Die Berechnung von $\underline{\mathbb{E}}_{\mathcal{M}}(u(a))$ und $\overline{\mathbb{E}}_{\mathcal{M}}(u(a))$ für festes $a \in \mathbb{A}$ kann effizient mit linearer Optimierung erfolgen (vgl. Bem. 3.8 und Bem. 3.12). In vielen Modellen kann auch auf die Darstellung über das Choquet-Integral zurückgegriffen werden, das eine explizite Berechnungsformel liefert. Man kann ferner sogar zeigen, dass sich die Berechnung der Max E Min optimalen randomisierten Aktionen auf ein *einziges* lineares Optimierungsproblem zurückführen lässt.¹⁸

¹⁷vgl. das RCDD-Package (Geyer and Meeden, 2013, CRAN) als Interface für Fukudas „cdd-Bibliothek“.

¹⁸Vgl. Utkin & Augustin (2005, Proceedings of the Fourth International Symposium on Imprecise Probabilities and Their Applications), wo auch Algorithmen für die anderen Kriterien entwickelt werden.

3.5 Konditionale Bayes-Inferenz mit Credalmengen

3.5.1 Generalized Bayes Rule

Bem. 3.20 (Zur Kritik des Ansatzes, Priori-Credal-Mengen)

Natürlich hängt (2.50) neben den Stichprobenergebnis auch noch von $a^{(0)}$ und $b^{(0)}$ entscheidend ab.

+ „Vorwissen kommt mit herein“

- Aber: Wann hat man schon so präzises Vorwissen und braucht dann noch eine Stichprobe?

- Was tut man bei Nichtwissen?

Die Gleichverteilung ist als Modell für Nichtwissen nicht unproblematisch. (vgl. auch Kritik Laplace-Regel) Hat man in diesem Sinne also „keine Information“ über \mathcal{V} , so ergibt sich eine informative Priori z.B. über eine bijektive, nichtlineare Transformation von \mathcal{V} .

- Modellierung von partiellem Vorwissen und Nichtwissen durch Credalmengen:
 - * Lasse a und/oder b in Bereich variieren und betrachte die (konvexe Hülle aller) entstehenden Verteilungen als Priori-Credalmenge. („robuste Bayes-Analyse“)
 - * Idee: Fast „völliges Nichtwissen“, alle möglichen Werte von a und b → „*near ignorance prior*“; in anderer Parametrisierung besser darstellbar, siehe Bem. 3.23

Bem. 3.21 (Robuste Bayes-Analyse, Generalized Bayes Rule)

In der Situation von Bem. 2.89 kann man auch mit *Priori-Credalmengen* \mathcal{M} arbeiten. Dann heißt

$$\mathcal{M}_{\cdot|x} = \left\{ \pi(\cdot|x) \mid \exists \pi(\cdot) \in \mathcal{M} : \pi(\cdot|x) \text{ ist Posteriori-Verteilung (3.10)} \right. \\ \left. \text{von } \mathcal{V} \text{ gegeben } x \text{ bezüglich } \pi(\cdot) \right\}$$

Posteriori-Credalmenge gegeben x bezüglich \mathcal{M} . Man spricht dann von *einer robusten Bayes-Analyse*; (3.11) wird dann oft als *Generalized Bayes Rule* (GBR) bezeichnet.

Bem. 3.22 (Bayes Postulat (nicht entscheidungstheoretisch))

Nach der Beobachtung der Stichprobe enthält die (klassische) Posteriori-Verteilung bzw. die Posteriori-Credalmenge die volle Information, d.h. sie beschreibt das Wissen über den unbekannt Parameter vollständig.

Alle statistischen Analysen haben sich ausschließlich auf die Posteriori zu stützen; darauf aufbauend insbesondere Konstruktion von

- Bayesschen-Punktschätzungen: *MPD-Schätzer (Maximum Posteriori Density-Schätzer)*
- Bayessche-Intervallschätzung: *HPD-Intervalle (Highest posterior density-Intervalle)*
- Bayes-Tests

3.5.2 Konjugiertheit und verallgemeinerte Bayes-Inferenz

Bem. 3.23 (Eine alternative Darstellung von Satz 2.97)

Eine zum Nachweis meist umständlichere, aber für die Interpretation oft anschaulichere und für die Verallgemeinerung besser geeignete, alternative Darstellung von Satz 2.97 lautet:

Hat in der Situation von Def. 2.89 jedes Element der Menge \mathcal{P} der Stichprobenverteilungen eine Dichte bzw. Wahrscheinlichkeitsfunktion $f(x|\vartheta)$ der Form

$$f(x|\vartheta) \propto \exp\left(\psi(\vartheta)\tau(\vec{x}) - n \cdot d(\vartheta)\right) \quad (3.11)$$

und jedes Element der Menge Π , aus der die Priori-Verteilung stammt, eine Dichte bzw. Wahrscheinlichkeitsfunktion der Form

$$\pi(\vartheta) \propto \exp\left(n^{(0)}\left(\psi(\vartheta)y^{(0)} - d(\vartheta)\right)\right) \quad (3.12)$$

so sind Π und \mathcal{P} konjugiert. Es gilt dann

$$\pi(x|\vartheta) \propto \exp\left(n^{(1)}\left(\psi(\vartheta)y^{(1)} - d(\vartheta)\right)\right) \quad (3.13)$$

mit

$$n^{(1)} = n^{(0)} + n \quad (3.14)$$

und

$$y^{(1)} = \frac{n^{(0)}y^{(0)} + \tau(x)}{n^{(0)} + n}. \quad (3.15)$$

$y^{(0)}$ ist typischerweise ein Lageparameter, $n^{(0)}$ kann man als virtuelle Beobachtungen interpretieren, auf denen das Priori-Wissen beruht.

$n^{(0)}$ tritt immer gemeinsam mit n auf, nämlich im Ausdruck $n^{(0)} + n$

Mit $\bar{\tau}(\vec{x}) = \frac{1}{n}\tau(\vec{x})$ läßt sich die Aufdatierung des Parameters schreiben als

$$y^{(1)} = \frac{n^{(0)}}{n^{(0)} + n}y^{(0)} + \frac{n}{n^{(0)} + n}\bar{\tau}(\vec{x}) \quad (3.16)$$

also als gewichtetes Mittel der Priori-Vermutung und des Stichprobenmittels.

Beispielsweise ist dann die Beta-priori aus (2.49)

$$\begin{aligned}\pi(\vartheta) &\propto \vartheta^{a-1}(1-\vartheta)^{b-1}I_{[0;1]}(\vartheta) \\ &= \vartheta^{n^{(0)}y^{(0)}-1}(1-\vartheta)^{n^{(0)}(1-y^{(0)})-1}I_{[0;1]}(\vartheta)\end{aligned}$$

mit (festem) $n^{(0)} > 0$ und (festem) $y^{(0)} \in (0; 1)$. $y^{(0)}$ ist dann genau der Erwartungswert der Priori-Verteilung; ferner gilt für die priori-prädiktive Verteilung der nächsten unabhängigen Beobachtung X_{neu} :

$$p(X_{neu} = 1) = y^{(0)}$$

.

Bem. 3.24 (Robuste Bayes-Analyse in konjugierten Modellen)

- Die Darstellung in Bem. 3.23 ermöglicht eine elegante robuste Bayes-Analyse. Priori-Credalmengen erzeugt man durch intervallwertige Priori-Parameter:¹⁹

$[\underline{y}^{(0)}, \bar{y}^{(0)}]$ typischerweise intervallwertiger Priori-Mittelwert bzw. bzw. Lageparameter

und/oder

$[\underline{n}^{(0)}, \bar{n}^{(0)}]$ intervallwertige virtuelle Beobachtungen, auf denen das Priori-Wissen beruht

¹⁹Walley (1991/1996): Binomial-/Multinomialmodell. Quaeghebeur & deCooman (2005): Exponentialfamilien mit festem $n^{(0)}$; Walter & Augustin (2009): $n^{(0)}$ zusätzlich variabel.

Lässt man $\underline{y}^{(0)}$ und $\bar{y}^{(0)}$ gegen die Grenzen des zulässigen Priori-Parameterbereichs gehen, so erhält man sogenannte „*near ignorance*“-Modelle.

Beispielsweise gilt dann für die priori-prädiktive Verteilung

$$p(X = 1) = \left[\lim_{y^{(0)} \downarrow 0} y^{(0)}, \lim_{y^{(0)} \uparrow 1} y^{(0)} \right] = [0; 1]$$

und analog für $p(X = 0)$, was Nichtwissen über ϑ deutlich ausdrückt.

- Für die posteriori-prädiktive Verteilung ergibt sich mit festem „Priori-Gewichts-Parameter“ $n^{(0)}$ nach n Beobachtungen x_1, \dots, x_n :

$$\begin{aligned}
 & p(X_{neu} = 1 | X_1 = x_1, \dots, X_n = x_n) \\
 &= \left[\lim_{y^{(0)} \downarrow 0} \frac{n^{(0)}y^{(0)} + \sum_{i=1}^n x_i}{n^{(0)} + n}; \lim_{y^{(0)} \uparrow 1} \frac{n^{(0)}y^{(0)} + \sum_{i=1}^n x_i}{n^{(0)} + 1} \right] \\
 &= \left[\frac{\sum_{i=1}^n x_i}{n^{(0)} + n}; \frac{n^{(0)} + \sum_{i=1}^n x_i}{n^{(0)} + n} \right].
 \end{aligned}$$

Die Breite

$$\frac{n^{(0)}}{n^{(0)} + n}$$

des Intervalls, nimmt in n monoton ab: Für kleines n sind aus Nichtwissen nur schwache Folgerungen ziehbar; für größeres n wird man präziser.

Beachte, dies ist kein Konfidenzintervall, sondern eine „intervallwertige Punktschätzung“!

Die entsprechende Verallgemeinerung auf mehrkategoriale Beobachtungen ist das sog. *Imprecise-Dirichlet-Model* (IDM, Walley (1996, J. Royal Statistical Society B)). Es gilt als Grundlage vieler weiterführender Anwendungen. Man kann zeigen: Die Priori- und Posteriori-Verteilungen sind unabhängig von der Kategorisierung; Zusammenfassen/Präzisieren der Kategorien ändert die Inferenz nicht (vgl. im Gegensatz dazu die Auseinandersetzung mit der Laplace-Regel in Kap. 2.3)

Bem. 3.25 (Fortsetzung von Beispiel 2.96)

Man betrachte die Bayes-Inferenz für den Mittelwert einer Normalverteilung aus einer i.i.d. Stichprobe. Geht man im Sinne von Bem. 3.24 zu Priori-Credalmengen über, wobei ν in einem Intervall $[\underline{\nu}^{(0)}, \bar{\nu}^{(0)}]$ und $n^{(0)}$ in $[\underline{n}^{(0)}, \bar{n}^{(0)}]$ variiert, so gilt²⁰

²⁰Walter&Augustin(2009, Remark 4.1)

$$\underline{\nu}^{(1)} = \begin{cases} \frac{\bar{n}^{(0)} \underline{\nu}^{(0)} + \sum_{i=1}^n x_i}{\bar{n}^{(0)} + n}, & \text{falls } \frac{1}{n} \sum_{i=1}^n x_i \geq \underline{\nu}^{(0)} \\ \frac{\underline{n}^{(0)} \bar{\nu}^{(0)} + \sum_{i=1}^n x_i}{\underline{n}^{(0)} + n}, & \text{falls } \frac{1}{n} \sum_{i=1}^n x_i < \underline{\nu}^{(0)} \end{cases}$$

$$\bar{\nu}^{(1)} = \begin{cases} \frac{\bar{n}^{(0)} \bar{\nu}^{(0)} + \sum_{i=1}^n x_i}{\bar{n}^{(0)} + n}, & \text{falls } \frac{1}{n} \sum_{i=1}^n x_i \leq \bar{\nu}^{(0)} \\ \frac{\underline{n}^{(0)} \bar{\nu}^{(0)} + \sum_{i=1}^n x_i}{\underline{n}^{(0)} + n}, & \text{falls } \frac{1}{n} \sum_{i=1}^n x_i > \bar{\nu}^{(0)} \end{cases}$$

Daraus ergibt sich insbesondere, dass

$$\begin{aligned} \bar{\nu}^{(1)} - \underline{\nu}^{(1)} &= \frac{\bar{n}^{(0)}(\bar{\nu}^{(0)} - \underline{\nu}^{(0)})}{\bar{n}^{(0)} + n} \\ &+ \text{pdc} \left(\frac{1}{n} \sum_{i=1}^n x_i; \underline{\nu}^{(0)}, \bar{\nu}^{(0)} \right) \frac{n(\bar{n}^{(0)} - \underline{n}^{(0)})}{(\bar{n}^{(0)} + n)(\underline{n}^{(0)} + n)} \end{aligned}$$

Dabei ist

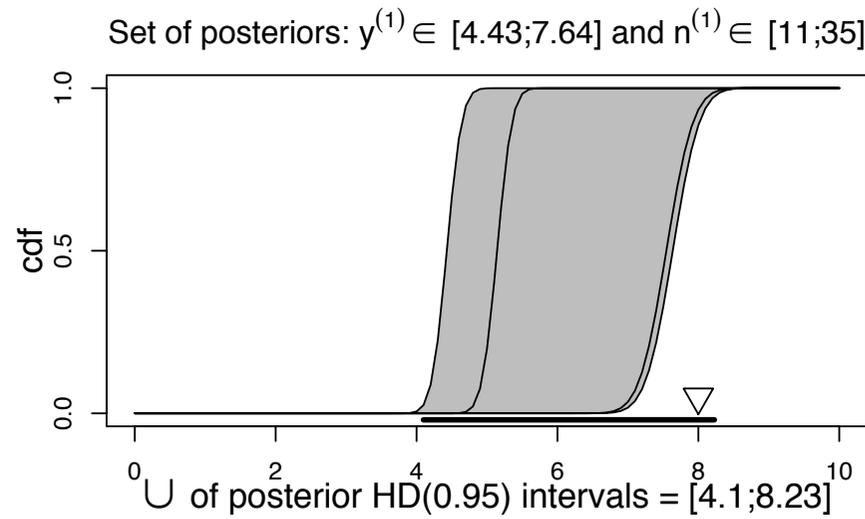
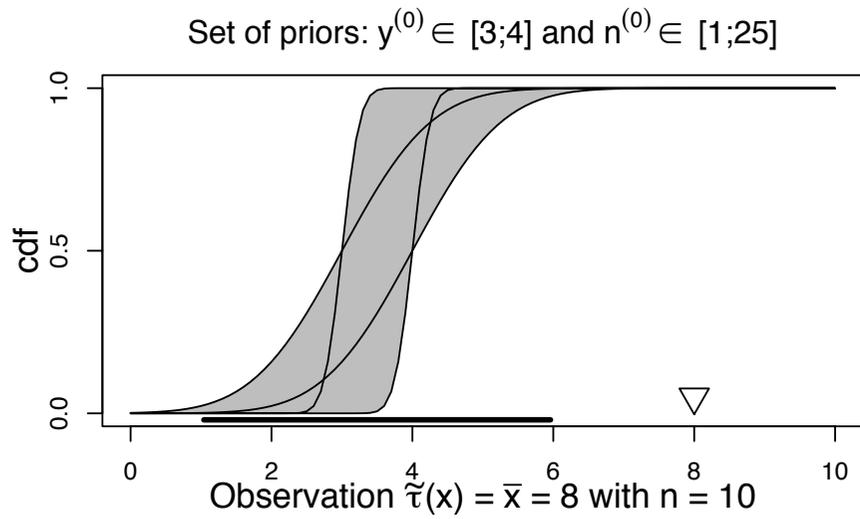
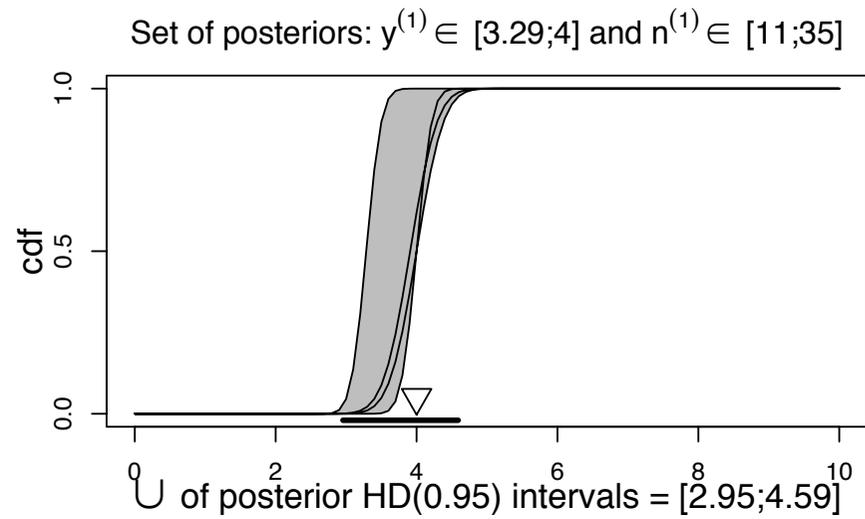
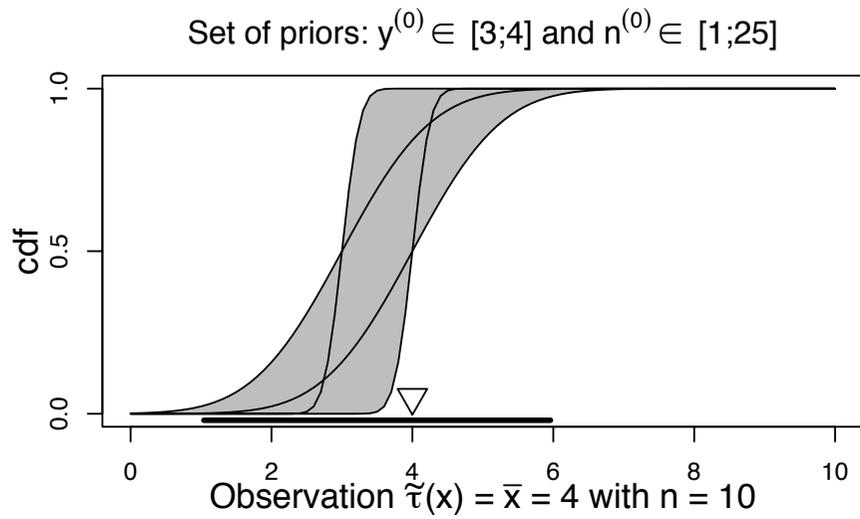
$$\text{pdc} \left(\frac{1}{n} \sum_{i=1}^n x_i; \underline{\nu}^{(0)}, \bar{\nu}^{(0)} \right) := \inf \left\{ \left| \frac{1}{n} \sum_{i=1}^n x_i - \nu^{(0)} \right| \mid \underline{\nu}^{(0)} \leq \nu^{(0)} \leq \bar{\nu}^{(0)} \right\}.$$

pdc misst das Ausmaß des Priori-Daten-Konflikts, also wie weit die Stichprobenbeobachtung $\frac{1}{n} \sum_{i=1}^n x_i$ von dem Priori-Mittelwert $[\underline{\nu}^{(0)}, \bar{\nu}^{(0)}]$ entfernt ist. (0, wenn innerhalb des Intervalls, sonst Differenz zur nächsten Grenze.)

Damit gilt also:

- Bei „nicht überraschenden Beobachtungen“ ist die Posteriori-Unschärfe klein.
- Bei „überraschenden Beobachtungen“ hingegen gilt: Die Posteriori-Unschärfe ist groß; bei abgeleiteten Folgerungen ist man sehr vorsichtig.

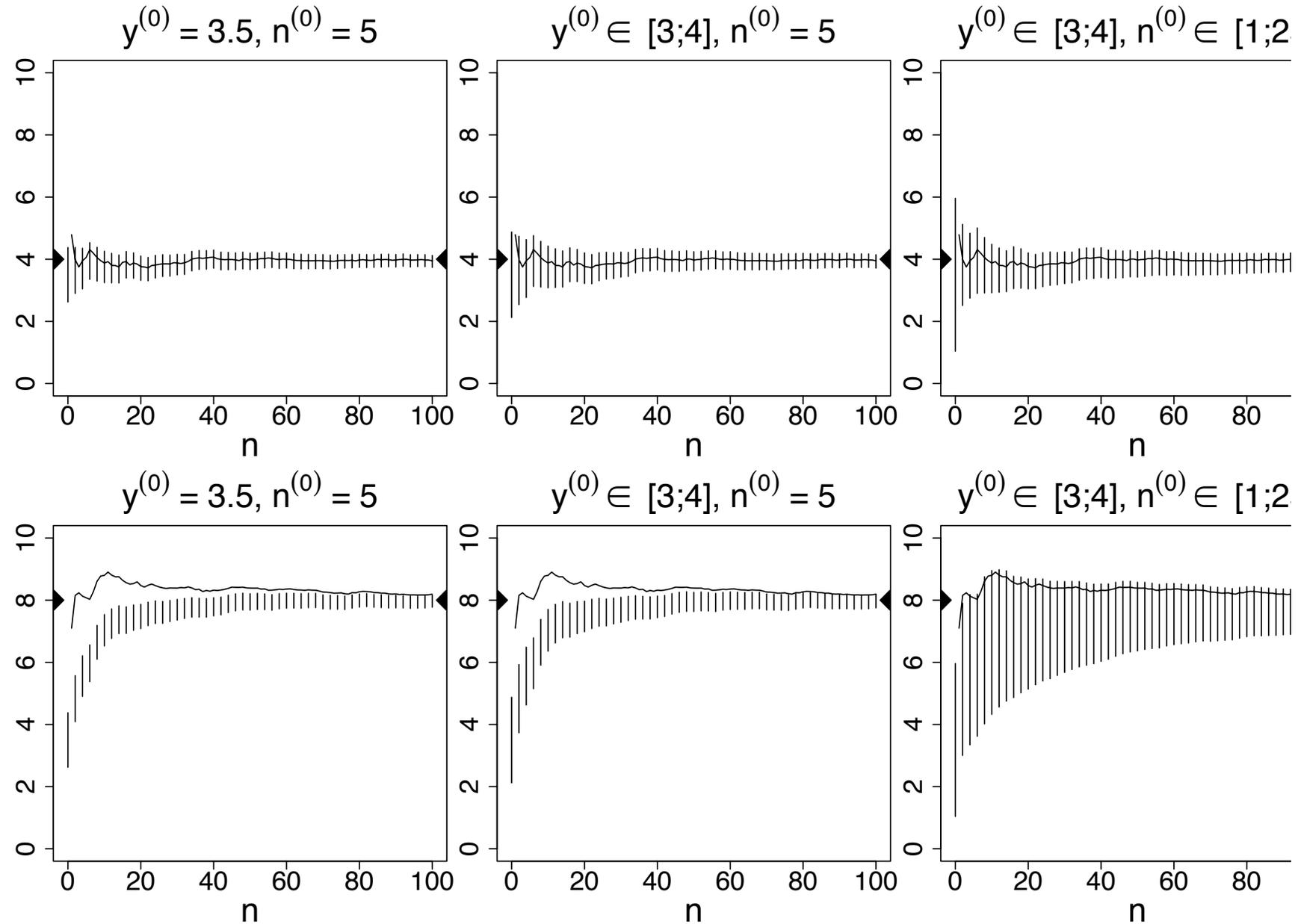
Quelle: Walter & Augustin (2009, S. 268)



Die Bilder zeigen die Inferenz mit einer Menge von Normalverteilungen (angedeutet durch die Verteilungsfunktionen) gebildeten Priori-Credalmenge (jeweils links) und die zugehörigen Posteriori Credalmengen, einmal im Fall ohne Priori-Daten-Konflikt (oben) und mit Priori-Daten-Konflikt (unten). In der Tat ist im zweiten Fall die Credalmenge sehr „gross“; man kann in dieser widersprüchlichen Situation kaum weitergehende Schlüsse tätigen, während die obere Aussage relativ präzise Aussagen ermöglicht.

Die Abbildungen zeigen das arithmetische Mittel (schwarze Linie) und HPD-Intervalle zum Sicherheitsgrad 95 Prozent bei der konjugierten Inferenz den Mittelwert einer Normalverteilung für verschiedene Stichprobenumfänge in einer typischen Situation ohne Priori-Daten-Konflikt (oben) und mit Priori-Daten-Konflikt (unten). Links ist die klassische Modellierung basierend auf einer Priori-Normalverteilung, rechts eine geeignet konstruierte Credalmenge. Man sieht, dass die HPD Schätzungen hier deutlich breiter sind als im Fall ohne Priori-Daten Konflikt, während im klassischen Fall die Intervalle gleich lang bleiben und nur verschoben werden.

Quelle: Walter & Augustin (2009, S. 268)



Korollar 3.26 (Korollar zu Satz 2.106: Konsistenzsatz für Credal-)

In der Situation von Satz 2.106 gilt:

Ist \mathcal{M} eine Priori-Credalmenge mit $\pi(\cdot) > 0, \forall \pi \in \mathcal{M}$, so zieht sich die nach n Beobachtungen gebildete Posteriori-Credalmenge $\mathcal{M}_{|x}^{(n)}$ im Punkt ϑ_{wahr} zusammen:

$$\begin{aligned} & \lim_{n \rightarrow \infty} \left(\inf_{\pi(\cdot|x) \in \mathcal{M}_{|x}^{(n)}} \pi(\vartheta|x) \right) \\ &= \lim_{n \rightarrow \infty} \left(\sup_{\pi(\cdot|x) \in \mathcal{M}_{|x}^{(n)}} \pi(\vartheta|x) \right) = \begin{cases} 1 & \text{falls } \vartheta = \vartheta_{wahr} \\ 0 & \text{falls } \vartheta. \end{cases} \end{aligned}$$