

Entscheidungstheorie

Prof. Dr. Th. Augustin

13. Juli 2014

1 Einführung

1.1 Charakterisierung der Entscheidungstheorie als Theorie des rationalen Entscheidens unter Unsicherheit

1.1.1 Interdisziplinäre Bedeutung und Entwicklungsstränge

a) Rational Choice

b) Expertensysteme, entscheidungsunterstützende Systeme (Decision Support Systems)

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”.

c) Statistische Entscheidungstheorie

1.1.2 Grundlegende Typen von Entscheidungsmodellen

1.2 Die Grundform eines datenfreien Entscheidungsproblems (No-data-Problem)

Def. 1.1 (Datenfreies Entscheidungsproblem)

Ein *datenfreies Entscheidungsproblem* (no-data-problem) in Nutzenform (Verlustform) ist ein Tripel $(\mathbb{A}, \Theta, u(\cdot))$ bzw. $(\mathbb{A}, \Theta, l(\cdot))$, bestehend aus

- einer Menge \mathbb{A} („*Aktionenmenge*“),
- einer Menge Θ („*Zustandsmenge*“)

- und einer Abbildung („*Nutzenfunktion*“) ($u \hat{=} \text{utility}$)

$$\begin{aligned} u : \mathbb{A} \times \Theta &\rightarrow \mathbb{R} \\ (a, \vartheta) &\mapsto u(a, \vartheta) \end{aligned} \tag{1.1}$$

bzw. einer Abbildung („*Verlustfunktion*“) ($l \hat{=} \text{loss}$)

$$\begin{aligned} l : \mathbb{A} \times \Theta &\rightarrow \mathbb{R} \\ (a, \vartheta) &\mapsto l(a, \vartheta) \end{aligned} \tag{1.2}$$

Bsp. 1.2 (Das „Omelettenproblem“ von Savage)

Bem. 1.3 (Konsequenzenfunktion)

Für manche Anwendungen ist es sinnvoll, einen Schritt dazwischenzuschalten und – bei gegebenem \mathbb{A} und Θ – zunächst eine *Konsequenzenfunktion*

$$c : \mathbb{A} \times \Theta \rightarrow \mathcal{C}$$

$$(a, \vartheta) \mapsto c(a, \vartheta)$$

(mit \mathcal{C} ist die Menge potentieller Konsequenzen) zu betrachten und darauf eine Nutzenbewertung

$$u_{\mathcal{C}} : \mathcal{C} \rightarrow \mathbb{R}$$

$$c \mapsto u_{\mathcal{C}}(c)$$

bzw. eine Verlustbewertung

$$l_{\mathcal{C}} : \mathcal{C} \rightarrow \mathbb{R}$$

$$l \mapsto l_{\mathcal{C}}(c)$$

festzulegen.

$u(\cdot)$ und $l(\cdot)$ ergeben sich dann durch Superposition beider Funktionen als

$$u(a, \vartheta) = u_{\mathcal{C}}(c(a, \vartheta))$$

bzw.

$$l(a, \vartheta) = l_{\mathcal{C}}(c(a, \vartheta))$$

Bem. 1.4 (zu Nutzen- und Verlustfunktionen)

- In Definition 1.1 haben Verlust- und Nutzenfunktion (als Funktionen der Gestalt $\mathbb{A} \times \Theta \rightarrow \mathbb{R}$) formal genau dieselbe Struktur. Es muss also jeweils dazugesagt werden, was vorliegen soll. (Nutzen: Je mehr, umso besser; Verlust: Je weniger, umso besser)
- Diese Uneindeutigkeit liegt nicht zuletzt daran, dass man eigentlich mit *Präferenzordnungen* auf der Menge der Konsequenzen als grundlegende Entität arbeiten müsste. Die Nutzentheorie lehrt, wie man unter welchen Bedingungen aus Präferenzordnungen einen kardinalen (metrischen) Nutzen konstruiert. (Hier nur eventuell am Ende der Vorlesung betrachtet.)

- Die Kardinalität des Nutzens wird (zunächst) nicht bezweifelt; man kann also mit den Nutzeinheiten wie gewohnt rechnen. Insbesondere wird dem Folgenden eine Äquivalenz von Nutzen- und Verlustsicht unterstellt: durch Multiplizieren mit (-1) kann man dann jede Nutzenfunktion $u(\cdot)$ in eine, genau dieselbe Präferenzordnung widerspiegelnde Verlustfunktion umwandeln. Daher wird im Folgenden meist nur entweder von Nutzen- oder von Verlustfunktion gesprochen.

Bem. 1.5 (Notation im endlichen Fall; Einbettung in den \mathbb{R}^m)

Im Falle einer endlichen Aktionenmenge und einer endlichen Zustandsmenge wird folgende Notation verwendet:

$$\begin{aligned}\mathbb{A} &= \{a_1, \dots, a_i, \dots, a_n\} \\ \Theta &= \{\vartheta_1, \dots, \vartheta_j, \dots, \vartheta_m\}\end{aligned}\tag{1.3}$$

(also $|\mathbb{A}| = n$, $|\Theta| = m$; i Laufindex in \mathbb{A} , j Laufindex in Θ)

Nutzenfunktion, Verlustfunktion und Konsequenzenfunktion können dann als Matrizen $(u_{ij})_{\substack{i=1,\dots,n \\ j=1,\dots,m}}$, $(l_{ij})_{\substack{i=1,\dots,n \\ j=1,\dots,m}}$, $(c_{ij})_{\substack{i=1,\dots,n \\ j=1,\dots,m}}$ dargestellt werden.

Zum Beispiel:

$$\begin{array}{c|ccccc}
 & \mathcal{V}_1 & \mathcal{V}_2 & \dots & \dots & \mathcal{V}_m \\
 \hline
 a_1 & u_{11} & u_{12} & \dots & \dots & u_{1m} \\
 a_2 & u_{21} & u_{22} & \dots & \dots & u_{2m} \\
 \dots & \dots & \dots & \dots & \dots & \dots \\
 \dots & \dots & \dots & u_{ij} & \dots & \dots \\
 \dots & \dots & \dots & \dots & \dots & \dots \\
 a_n & u_{n1} & u_{n2} & \dots & \dots & u_{nm}
 \end{array} \tag{1.4}$$

Man spricht dann von *Nutzentafel*, *Verlusttafel* oder *Konsequenzentafel*.

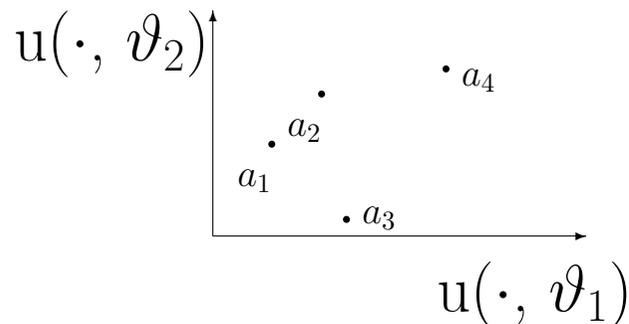
Einbettung in den \mathbb{R}^m :

Jede Aktion $a_i \in \mathbb{A}$ kann dann mit dem zugehörigen *Nutzenvektor*

$$\vec{u}(a_i) = (u_{i1}, u_{i2}, \dots, u_{ij}, \dots, u_{im})^T \in \mathbb{R}^m \quad (1.5)$$

identifiziert werden.

Ist $m = 2$ oder $m = 3$, so ist ferner eine graphische Darstellung möglich:



Wird diese Sichtweise verwendet, sprechen wir von direkter Nutzen- bzw. Verlustrepräsentation der Aktionenmenge.

Bem. 1.6 (Semantik des datenfreien Entscheidungsproblems)

Abstrahiert man bei einem konkreten Entscheidungsproblem und fasst es in die Form $(\mathbb{A}, \Theta, u(\cdot))$ bzw. $(\mathbb{A}, \Theta, l(\cdot))$ laut Definition 1.1, so sind damit implizit die folgende Reihe von grundsätzlichen Annahmen verbunden, die in der jeweiligen Anwendung kritisch zu hinterfragen sind:

a) \mathbb{A} sei bekannt.

b) Θ sei bekannt (Close-World-Assumption).

- c) Die Ergebnisse (Konsequenzen) seien eindeutig, d.h. aus dem Zusammenspiel jedes Elements a von \mathbb{A} und $\vartheta \in \Theta$ ergibt sich eine eindeutige bestimmte Konsequenz $c(a, \vartheta)$. Dabei können Konsequenzen durchaus Wahrscheinlichkeitsverteilungen sein.
- d) Es läßt sich eine eindeutig bestimmte reellwertige Nutzen-/Verlustfunktion angeben, die die individuellen Präferenzen des Entscheidungsträgers vollständig widerspiegelt.
- * Vgl. Bemerkung 1.4: Konstruktion von kardinalen (reellwertigen) Nutzenfunktionen aus bestimmte Bedingungen erfüllenden Präferenzordnungen \rightarrow „Nutzentheorie“

- * Durchgängig wird insbesondere angenommen, dass die Präferenzen eine vollständige Ordnung erlauben und in diesem Sinn eindimensional sind/gemacht werden können.

- * Es gibt also keine unterschiedlichen nichtabbildbaren „Nutzendimensionen“ (z.B. eine Diskrepanz zwischen kurz- und langfristigen Nutzen). Mehrdimensionale Nutzenbewertungen sind Gegenstand der sog. multikriteriellen Entscheidungstheorie. Es gibt auch eine Reihe von Ansätzen mit unscharfer Nutzenfunktionen.¹

¹Für eine erste Einführung vgl. z.B. Rommelfanger & Eickemeier (2001, Kap. 2.3 und 3.1.2)

- * Der Nutzen wird dabei als abstrakte Größe gesehen, in der z.B. auch Wertdispositionen mitintegriert sind.

Aspekte wie Wertorientierung, Fairness, Reziprozität, etc. können durchaus eine Rolle spielen, sie müssen durch die Nutzenfunktion abgebildet werden bzw. abbildbar sein.

- * Vorsicht: Bei monetären Konsequenzen ist im Allgemeinen der Nutzen nicht identisch mit (oder linear in) der Geldmenge. Nur wenn alle im Entscheidungsproblem betrachteten Geldmengen weit unterhalb des tatsächlichen Vermögensstandes sind, ist die Nutzenfunktion (praktisch) linear in den Geldbeträgen. Dies ist v.a. beim Berechnen von erwarteten Nutzen/Verlusten wichtig.

- e) Aktionen und Zustände seien wertfrei. (Eventuelle Bewertungen müssen in die Ergebnisse und damit in die Nutzenfunktion eingebaut werden.)
- f) Umwelt nicht beeinflussbar: „*Handlungsunabhängigkeit der Zustände*“
Gegebenenfalls „Implikationsschemata“ definieren (Festlegung der Umweltzustände ist keineswegs immer trivial!)
- g) Typ der Unsicherheit bekannt (Typ I, Typ I', oder Typ II, Typ II')
(bzw. später dann Verallgemeinerung)

- h) Keine zusätzliche Information (außer Wahrscheinlichkeit bei Typ I, Typ I'). Informationsgewinnung über Strategien formulieren → Statistische Entscheidungstheorie

- i) Einmalige Wahl der Entscheidung, keine Korrekturen

- j) Keine Wiederholung der Entscheidungssituation (wiederholte Entscheidung als eine Entscheidungsstrategie formulieren)

1.3 Typische Beispiele

1.3.1 Das Ausflugsproblem nach Chernoff & Moses (1959)

(in einer Adaption von Ferschl 1975)

- Mr. Nelson möchte morgen eine Bergwanderung unternehmen, und zwar in einer Jahreszeit, in der man mit dem plötzlichen Einfallen von Schlechtwetter zu rechnen hat

- **Handlungsmöglichkeiten:**

a_1 := leichte Bekleidung mitnehmen

a_2 := leichte Bekleidung plus Regenschirm mitnehmen

a_3 := wetterfeste, warme Bekleidung plus Regenschirm mitnehmen

- Die **relevanten Fakten** bzw. **Zustände** sind

ϑ_1 := schönes Wetter am Ausflugstag
 ϑ_2 := schlechtes Wetter am Ausflugstag

- Nutzentafel für Mr. Nelson:

	ϑ_1	ϑ_2
a_1	5	0
a_2	3	1
a_3	2	3

Die Bewertungen werden als Nutzen interpretiert

Potentielle datenbasierte Zusatzinfo (\rightarrow später): Barometerablesung

1.3.2 Teilnahme an einer Lotterie

Gegeben sei eine Urne mit

g grünen

b blauen

r restlichen

Kugeln. Man kann

a_1 nicht spielen

a_2 zum Preis von c_g auf grün setzen

a_3 zum Preis von c_b auf blau setzen.

Nun wird eine Kugel zufällig gezogen. Ist sie grün (bzw. blau), so erhält man, wenn man auf grün (bzw. blau) gesetzt hat, w_g bzw. w_b (w: win), wobei die Beträge so klein seien, dass die Linearität des Nutzens in der Geldmenge gewährleistet sei.

$\Theta = \{\vartheta_1, \vartheta_2, \vartheta_3\}$ mit

ϑ_1 gezogene Kugel ist grün

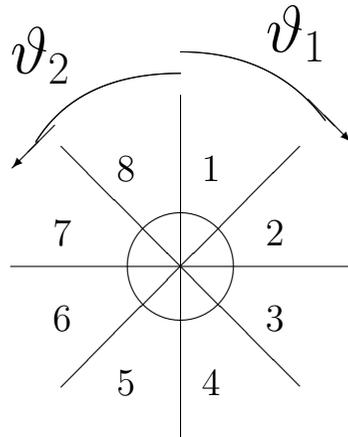
ϑ_2 gezogene Kugel ist blau

ϑ_3 gezogene Kugel hat sonstige Farbe

Für später wichtig: Sind g,b,r bekannt, so liegt eine Unsicherheitssituation vom Typ I (Risikosituation (i.e.S.)) vor!

1.3.3 „Kuchenteilen“

Gegeben sei ein Kuchen aus 8 durchnummerierten gleich großen Stücken.



- Der Entscheidungsträger wählt ein Stück z , $z \in \{1, \dots, 7\}$, bei dem der Kuchen geteilt wird.
- Es gebe zwei Umweltzustände: ϑ_1 und ϑ_2 . Tritt ϑ_1 ein, so erhält der Entscheidungsträger die Stücke 1 bis z , bei ϑ_2 hingegen die Stücke $z + 1$ bis 8.

	ϑ_1	ϑ_2
1	1	7
2	2	6
3	3	5
4	4	4
5	5	3
6	6	2
7	7	1

Nutzen = Anzahl Stücke! (Nichtsättigungsannahme)

- Es wird hier noch nichts vorausgesetzt, wie die Umweltzustände eintreten. Werden sie z.B. durch einen Gegenspieler erzeugt, so erhält man ein typisches Beispiel für die Situation des strategischen Spiels, also für Typ II - Unsicherheit

1.3.4 Investitionsproblem

Ein Unternehmen steht vor der Entscheidung, eine Marketing-Investition zu tätigen. Ihr Erfolg hängt von der Konjunkturlage im nächsten Halbjahr ab.

- **Aktionen:**

a_1 := Investition tätigen

a_2 := Investition nicht tätigen

- **Zustände:**

$\vartheta_1 :=$ Steigende Konjunktur

$\vartheta_2 :=$ Stagnation

$\vartheta_3 :=$ Konjunktur fällt

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

1.3.5 Aktienkauf (natürlich stark vereinfacht)

1.3.6 Produktionsplanung unter Szenarien

- q Güter
- $\mathbb{A} \subseteq (\mathbb{R}_0^+)^q$ beschrieben durch Kapazitätsbeschränkungen

-

$$a = \begin{pmatrix} a[1] \\ \vdots \\ a[q] \end{pmatrix}$$

mit $a[\ell]$: Produktion von $a[\ell]$ Einheiten von Gut ℓ , $\ell = 1, \dots, q$

- Umweltzustand ϑ : Szenario (mögliche Entwicklung des Marktes)

- Umsätze unter verschiedenen Szenarien ergeben Nutzenfunktion:

$$u : \mathbb{A} \times \Theta \rightarrow \mathbb{R}$$

$$(a, \vartheta) \mapsto u(a, \vartheta)$$

- Einfache Form: Stückpreisbasiert; unter Nichtsättigungsannahme; Vektor $c(\vartheta)$ mit $c[\ell](\vartheta)$: Stückpreis für Gut ℓ bei Szenario ϑ , $\ell = 1, \dots, q$, dann

$$u(a, \vartheta) = a'c(\vartheta) = \sum_{\ell=1}^q a[\ell] \cdot c[\ell](\vartheta)$$

bzw. bei Interpretation von $c(\vartheta)$ als Kosten:

$$l(a, \vartheta) = \sum_{\ell=1}^q a[\ell] \cdot c[\ell](\vartheta)$$

- 1.3.7 Einbettung stat. Tests in die Entscheidungstheorie I:
der datenfreie Kern eines Testproblems

- 1.3.8 Einbettung der Parameterschätzung in die
Entscheidungstheorie: der datenfreie Kern eines
Schätzproblems

1.4 Randomisierte Aktionen, Konvexität

1.4.1 Konvexe Menge

Def. 1.7 (konvexe Mengen)

Seien \mathbb{V} ein Vektorraum und z_1, \dots, z_n Elemente von \mathbb{V} .

a) Mit $\lambda_1, \lambda_2, \dots, \lambda_n \in \mathbb{R}$ mit $0 \leq \lambda_\ell \leq 1, \ell = 1, \dots, n$, und $\sum_{\ell=1}^n \lambda_\ell = 1$ sowie $z_1, z_2, \dots, z_n \in \mathbb{V}$ heißt

$$z := \lambda_1 \cdot z_1 + \lambda_2 \cdot z_2 + \dots + \lambda_n \cdot z_n \quad (1.6)$$

Konvexkombination von $z_1 \dots z_n$.

b) Für $x, y \in \mathbb{V}$ heißt die Menge

$$[x, y] := \{\lambda \cdot x + (1 - \lambda) \cdot y \mid 0 \leq \lambda \leq 1\} \quad (1.7)$$

aller Konvexkombinationen von x und y *Verbindungsstrecke* zwischen x und y .

c) Eine Teilmenge $\mathcal{M} \subseteq \mathbb{V}$ heißt *konvex*, wenn für alle $z_1, z_2 \in \mathcal{M}$ gilt:

$$z_1 \in \mathcal{M}, z_2 \in \mathcal{M} \Rightarrow [z_1, z_2] \subseteq \mathcal{M}.$$

Bsp. 1.9 (Beispiele für konvexe Mengen)

Lemma 1.10 (Äquivalentes Kriterium)

\mathcal{M} ist konvex im Sinne von Definition 1.7 genau dann, wenn gilt:

$$\forall q, \lambda_1, \dots, \lambda_q \geq 0, \sum_{\ell=1}^q \lambda_\ell = 1, z_1, \dots, z_q \in \mathcal{M} \text{ gilt : } \sum_{\ell=1}^q \lambda_\ell z_\ell \in \mathcal{M}$$

1.4.2 Mischungen von Verteilungen

Bem. 1.11 (Mischungen von Verteilungen)

Sei (Ω, \mathcal{A}) ein messbarer Raum und \mathcal{P} die Menge aller Wahrscheinlichkeitsmaße auf (Ω, \mathcal{A}) .

Gegeben seien q Elemente $p_1(\cdot), \dots, p_q(\cdot) \in \mathcal{P}$ (also Wahrscheinlichkeitsmaße) und reelle Zahlen $\lambda_1, \dots, \lambda_q$ mit $0 \leq \lambda_\ell \leq 1$, $\ell = 1, \dots, q$, und $\sum_{\ell=1}^q \lambda_\ell = 1$.

Definiert man die *Mischung* $\bar{p}(\cdot)$ von $p_1(\cdot), \dots, p_q(\cdot)$ vermöge

$$\bar{p}(A) = \sum_{\ell=1}^q \lambda_{\ell} p_{\ell}(A), \quad A \in \mathcal{A}, \quad (1.8)$$

so ist $\bar{p}(\cdot)$ wieder ein Element von \mathcal{P} . Die Menge \mathcal{P} ist folglich konvex.

Ferner gilt auch für $k \in \mathbb{N}$: Ist X k -fach integrierbar bzgl. $p_1(\cdot), p_2(\cdot), \dots, p_q(\cdot)$, so ist X auch k -fach integrierbar bezüglich $\bar{p}(\cdot)$. Für die Momente um 0 (!) gilt dann auch

$$\mathbb{E}_{\bar{p}(\cdot)} X^k = \sum_{\ell=1}^q \lambda_{\ell} \mathbb{E}_{p_{\ell}} X^k \quad (1.9)$$

Haben $p_1(\cdot), \dots, p_q(\cdot)$ die Dichten bzw. Wahrscheinlichkeitsfunktionen (bzw. allgemeiner die ν -Dichten) $f_1(\cdot), \dots, f_q(\cdot)$, so hat $\bar{p}(\cdot)$ die Dichte bzw.

Wahrscheinlichkeitsfunktion (bzw. allgemeiner die ν -Dichte)

$$\bar{f}(\cdot) = \sum_{\ell=1}^q \lambda_{\ell} \cdot f_{\ell}(\cdot). \quad (1.10)$$

Bem. 1.12

- Man beachte, dass im Allgemeinen für \mathcal{P} nicht eine Standardklasse von Verteilungen gewählt werden kann. Zum Beispiel ist die Menge aller Normalverteilungen nicht abgeschlossen gegenüber Mischungen. Dies mag zunächst enttäuschen, gewährt aber andererseits große Flexibilität in der Modellierung, da man durch Mischen eben sehr komplexe Formen erzeugen kann.
- Die Beschränkung auf Momente um 0 in (1.9) ist wesentlich; bspw. Varianzen kann man nicht so einfach zusammenzählen.

1.4.3 Konvexe Hülle, konvexe Polyeder

Def. 1.13 (Konvexe Hülle)

Sei \mathcal{M} eine beliebige Teilmenge von \mathbb{V} .

- a) Der Schnitt aller konvexen Obermengen von \mathcal{M} heißt *konvexe Hülle*.
(Die konvexe Hülle von \mathcal{M} ist also die „kleinste konvexe Menge, die \mathcal{M} umfasst“.)
Schreibweise: $\text{conv}(\mathcal{M})$

- b) Ist \mathcal{M} endlich, so heißt $\text{conv}(\mathcal{M})$ auch *konvexes Polyeder (Polytop)*.

Bem. 1.14

- Polyedrische Menge: Schnitt endlich vieler „Halbräume“
Konvexes Polyeder : beschränkte polyedrische Menge

- Der Begriff „konvexes Polyeder“ wird in der Literatur nicht ganz einheitlich gebraucht. Gelegentlich werden alle Mengen, die sich als Schnitt endlich vieler Halbräume darstellen lassen als konvexe Polyeder bezeichnet. Die Vorlesung folgt der Konvention, solche Mengen als polyedrische Mengen zu bezeichnen. Man kann zeigen, dass ein Polyeder im Sinne der Vorlesung dann genau eine polyedrische Menge ist, die zusätzlich beschränkt ist. Im \mathbb{R}^k werden die die Halbräume begrenzenden Hyperebenen als Begrenzungslinien bezeichnet.

Proposition 1.15 *Andere Charakterisierung der konvexen Hülle*

Die konvexe Hülle $\text{conv}(\mathcal{M})$ ist die Menge aller Konvex-Kombinationen von Punkten aus \mathcal{M} :

$$\begin{aligned} \text{conv}(\mathcal{M}) &= \\ &= \left\{ \sum_{\ell=1}^q \lambda_{\ell} z_{\ell} \mid \sum_{\ell=1}^q \lambda_{\ell} = 1, q \in \mathbb{N}, 0 \leq \lambda_{\ell} \leq 1, z_{\ell} \in \mathcal{M}, \ell = 1, \dots, q \right\} \end{aligned}$$

Bsp. 1.17 (nach Büning / Naeve / Trenkler / Waldmann)
(2000, p. 327f, 333f)

Dieses Beispiel ist typisch für die Produktionsplanung (vgl. Beispiel in Abschnitt 1.3.6); es wird in Kapitel 2.2 immer wieder verwendet.

Ein Unternehmer stelle die Produkte P_1 und P_2 her. Die dazu benötigten Mittel sind wie folgt beschränkt:

Maschine: maximal 1200h

Rohstoffe: maximal 3000 Mengeneinheiten (ME)

Arbeitskraft: maximal 125h

und verteilen sich wie folgt auf je eine Mengeneinheit des Produkts $P_\ell, \ell = 1, 2$

	P_1	P_2
Maschine	3h	2h
Rohstoff	5ME	10ME
Arbeitskraft	0h	0.5h

- a) Beschreiben Sie die Menge aller möglichen Produktionsmengen P_1 und P_2 , die mit den vorgegebenen Beschränkungen verträglich sind!
- b) Veranschaulichen Sie Ihr Ergebnis graphisch!
- c) Welche Produktionsmengen nützen die vorhandenen Produktionsfaktoren so weit wie möglich aus?

1.4.4 Extremalpunkte

Bem. 1.19 *Man kann zeigen: Die Eckpunkte eines konvexen Polyeders $\mathcal{M} \in \mathbb{R}^k$ sind alle Punkte, die die folgenden beiden Bedingungen erfüllen:*

i) x ist ein Schnittpunkt von (mindestens) k Begrenzungslinien.

ii) $x \in \mathcal{M}$.

Bsp. 1.21 (Eckpunkte)

Man betrachte Beispiel 1.17 und bestimme die Menge der Eckpunkte der Menge \mathbb{A} aller möglichen Produktionsmengen.

Satz 1.23 (Extremalpunkte)

a) Sei $\mathcal{M} \subseteq \mathbb{R}^n$ konvex, nichtleer, abgeschlossen und beschränkt.

Dann gilt:

a1) $\mathcal{E}(\mathcal{M}) \neq \emptyset$

a2) Jede lineare Funktion $f : \mathcal{M} \mapsto \mathbb{R}$ nimmt ihr Maximum und ihr Minimum auf $\mathcal{E}(\mathcal{M})$ an.

b) Ein konvexer Polyeder ist die konvexe Hülle seiner Extremalpunkte.

Bem. 1.24 (zu a2))

- Der Punkt a2) ist von extremer praktischer Bedeutung; insbesondere bildet er die Grundlage der linearen Optimierung (siehe später), also des Optimierens linearer Funktionen unter linearen Nebenbedingungen. Er führt das Auswerten der Funktion über unendlicher Menge auf das Auswerten von endlich vielen Punkten zurück.
- Für die Statistik ist a2) auch deshalb interessant, da der Erwartungswert eine lineare Funktion (bzw. ein lineares Funktional) in den zugrunde liegenden Wahrscheinlichkeitsmaßen ist.

1.4.5 Randomisierte Aktionen

Def. 1.25 (randomisierte (gemischte Aktionen))

Sei \mathbb{A} die Aktionenmenge eines datenfreien Entscheidungsproblems $(\mathbb{A}, \Theta, u(\cdot))$ (und $\sigma(\mathbb{A})$ eine σ -Algebra über \mathbb{A} , die alle Einpunktmengen $\{a\} \in \mathbb{A}$ enthält.)

Dann heißt jedes Wahrscheinlichkeitsmaß auf $(\mathbb{A}, \sigma(\mathbb{A}))$ *randomisierte (gemischte) Aktion*.

Die Menge aller randomisierten Aktionen auf $(\mathbb{A}, \sigma(\mathbb{A}))$ werde mit $\mathcal{M}^{\sigma(\mathbb{A})}(\mathbb{A})$ bezeichnet. Ist klar, welche σ -Algebra verwendet wird, so schreibt man kurz $\mathcal{M}(\mathbb{A})$.

Bem. 1.26 (Zur Semantik einer randomisierten Aktion)

Sei $\mathbb{A} = \{a_1, \dots, a_n\}$. Dann besteht die randomisierte Aktion $\tilde{a}(\cdot) \in \mathcal{M}(\mathbb{A})$ aus folgender Handlungsvorschrift:

Führe ein Zufallsexperiment auf $\{1, \dots, n\}$ mit $p(\{i\}) = \tilde{a}(\{a_i\})$, $i = 1, \dots, n$, durch und wähle Aktion a_i genau dann, wenn i eintritt.

D.h. es wird mit Wahrscheinlichkeit $\tilde{a}(\{a_i\})$ die Aktion a_i gewählt.

Man schreibt oft

$$\tilde{a} = \begin{bmatrix} a_1, & \dots, & a_n \\ \tilde{a}(\{a_1\}), & \dots, & \tilde{a}(\{a_n\}) \end{bmatrix}.$$

Wählt man $\mathcal{M}(\mathbb{A})$ als Aktionenmenge, so kann man darauf aufbauend ein eigenes Entscheidungsproblem formulieren. Dazu ist es noch nötig, den Nutzen/Verlust geeignet zu definieren (als Erwartungswert).

Def. 1.27 (gemischte Erweiterung)

Gegeben sei ein datenfreies Entscheidungsproblem $(\mathbb{A}, \Theta, u(\cdot))$ und eine geeignete σ -Algebra $\sigma(\mathbb{A})$ auf \mathbb{A} . Dann heißt das datenfreie Entscheidungsproblem $(\mathcal{M}^{\sigma(\mathbb{A})}(\mathbb{A}), \Theta, \tilde{u}(\cdot))$ mit

$$\begin{aligned} \tilde{u}(\cdot) : \mathcal{M}^{\sigma(\mathbb{A})} \times \Theta &\longrightarrow \mathbb{R} \\ (\tilde{a}, \vartheta) &\longmapsto \tilde{u}(\tilde{a}, \vartheta) \end{aligned}$$

und

$$\tilde{u}(\tilde{a}, \vartheta) := \mathbb{E}_{\tilde{a}}[u(a, \vartheta)] = \int u(a, \vartheta) \, d\tilde{a}(a) \quad (1.11)$$

die *gemischte Erweiterung* von $(\mathbb{A}, \Theta, u(\cdot))$ (bzgl. $\sigma(\mathbb{A})$).

Bem. 1.28 (zu Def.1.27)

- Die Verwendung des allgemeinen Maßintegrals erlaubt die simultane Betrachtung des stetigen und diskreten Falls sowie die Berücksichtigung gemischt stetig/diskreter Verteilungen. Insbesondere gilt:

Ist $\tilde{a}(\cdot)$ ein (Lebesgue-)stetiges Wahrscheinlichkeitsmaß mit Dichte $\tilde{f}(\cdot)$, so ist

$$\tilde{u}(\tilde{a}, \vartheta) = \int_{\mathbb{A}} u(a, \vartheta) \tilde{f}(a) da . \quad (1.12)$$

Ist $\tilde{a}(\cdot)$ diskret mit Wahrscheinlichkeitsfunktion $\tilde{p}(\cdot)$ auf $\mathbb{A} = \{a_1, a_2, \dots, \}$, so ist

$$\tilde{u}(\tilde{a}, \vartheta) = \sum_{i=1}^n u(a_i, \vartheta) \tilde{p}(a_i) . \quad (1.13)$$

- Beachte Formel (1.11) ist eine definatorische Festlegung, die zwar plausibel ist, aber keineswegs logisch zwingend.
- Die gemischte Erweiterung ist also wieder ein datenfreies Entscheidungsproblem. Man braucht folglich bei allgemeinen Definitionen und Sätzen nicht zu unterscheiden, ob ein „ursprüngliches Entscheidungsproblem“ vorliegt oder die zugehörige gemischte Erweiterung.

Bem. 1.29 (reine Aktionen)

Die randomisierten Aktionen der Form $\tilde{a}(\cdot) = \delta_{\{a\}}, \delta_{\{a\}} \in \mathbb{A}$ (Dirac-Maß = Einpunktmaß in der Menge $\{a\}$) werden als *reine Aktionen* bezeichnet und mit $a \in \mathbb{A}$ identifiziert.

Bem. 1.30 (Vom Sinn und Unsinn randomisierter Aktionen)

Korollar 1.31

$\mathcal{M}^{\sigma(\mathbb{A})}$ aus Definition 1.25 ist konvex.

Beweis:

Bem. 1.32 (Geometrische Deutung randomisierter Aktionen)

- $|\Theta| = m < \infty, \quad |\mathbb{A}| = n < \infty$
- a_i wurde mit Nutzenvektor $(\vec{u}_i = u(a_i, \vartheta_1), u(a_i, \vartheta_2), \dots, u(a_i, \vartheta_m))^T = (u_{i1}, u_{i2}, \dots, u_{im})^T$ identifiziert (vgl. Bem. 1.5) In Zeichen $a_i \hat{=} \vec{u}_i$.
- $\tilde{a} = \begin{bmatrix} a_1 & \dots & a_i & \dots & a_n \\ p_1 & \dots & p_i & \dots & p_n \end{bmatrix}$ mit $p_i = p(\{a_i\})$ hat definitionsgemäß (vgl. Def. 1.27) den Nutzenvektor

$$\vec{u} = \left(\sum_{i=1}^n p_i u_{i1}, \sum_{i=1}^n p_i u_{i2}, \dots, \sum_{i=1}^n p_i u_{im} \right)^T \quad (1.14)$$

Vektoriell sind in der direkten Nutzenrepräsentation (vgl. Bemerkung) durch die Identifizierung von randomisierten Aktionen mit ihrem Nutzenvektor dann randomisierte Aktionen und Aktionen, die ja mit ihrem Nutzenvektor identifiziert wurden, Objekte vom selben Typ, nämlich Punkte des \mathbb{R}^m .

Es ist (Rechenregel für Vektoren)

$$\tilde{a} \hat{=} \vec{u} = p_1 a_1 + p_2 a_2 + \dots + p_n a_n$$

Also ist jede randomisierte Aktion eine Konvexkombination von reinen Aktionen und umgekehrt.

Satz 1.33 (randomisierte Aktionen als konvexe Hülle)

Sind Θ und \mathbb{A} endlich, so gilt $\mathcal{M}(\mathbb{A}) \hat{=} \text{conv}(\mathbb{A})$. Damit ist also $\mathcal{M}(\mathbb{A})$ ein konvexes Polyeder.

1.4.6 Zur Randomisierung in der statistischen Entscheidungstheorie

Bem. 1.38 (Zwei Möglichkeiten zu randomisieren)

$$(\mathbb{A}, \Theta, l(\cdot)) \longrightarrow (\mathcal{D}, \Theta, \mathbf{R}(\cdot))$$

1. \mathbb{A} besteht ausschließlich aus reinen Aktionen, im zugeordneten Auswertungsproblem $(\mathcal{D}, \Theta, \mathbf{R}(\cdot))$ wird zur gemischten Erweiterung übergegangen, d.h. es wird zwischen Entscheidungsfunktionen randomisiert. („echte randomisierte Entscheidungsfunktion“).

2. Die Aktionenmenge \mathbb{A} besteht aus gemischten Aktionen, ist also die gemischte Erweiterung einer Menge \mathbb{A}_0
 Jede Entscheidungsfunktion ordnet jedem $x \in \mathcal{X}$ eine Aktion aus $\mathbb{A} = \mathcal{M}(\mathbb{A}_0)$, also eine randomisierte Aktion zu.

1.5 Datengestütztes Entscheiden – Grundbegriffe der statistischen Entscheidungstheorie

1.5.1 Ein (sehr) ausführliches Motivationsbeispiel

Stichprobeninformation, Entscheidungsfunktion und Risiko

Investitionsproblem (vgl. Bsp. in Abschnitt 1.3.4)

Aktionen:

a_1 investieren

a_2 nicht investieren

Marketing-Investition; Erfolg hängt von zukünftiger Konjunktur ab

Zustände:

ϑ_1 steigende Konjunktur

ϑ_2 Stagnation

ϑ_3 fallende Konjunktur

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

Natürlich wird man „nicht ins Blaue“ entscheiden, sondern man wird versuchen, zusätzliche Information über die Umweltzustände zu bekommen.

Beispielweise kann man einen Konjunkturtest heranziehen.

Prognoseaussagen X mit folgenden Werten

x_1 : Konjunktur wird steigen

x_2 : Stagnation wird erwartet

x_3 : Konjunktur wird fallen

Präziser: Es wird erwartet/ prognostiziert, dass ...

Allgemein: Informationsbeschaffungsexperimente

1.5.2 Grundlegendes zur statistischen Entscheidungstheorie

Grundbegriffe

- datenfreies Problem
- Informationsstruktur
- Entscheidungsfunktion
- Risikofunktion

Def. 1.39 (Datengestütztes Entscheidungsproblem)

Ein *datengestütztes Entscheidungsproblem* in Nutzen- bzw. Verlustform ist ein Tupel

$$\begin{aligned} & \left((\mathbb{A}, \Theta, u(\cdot)); (\mathcal{X}, \sigma(\mathcal{X}), (p_\vartheta)_{\vartheta \in \Theta}) \right) \\ \text{bzw.} & \left((\mathbb{A}, \Theta, l(\cdot)); (\mathcal{X}, \sigma(\mathcal{X}), (p_\vartheta)_{\vartheta \in \Theta}) \right), \end{aligned} \tag{1.15}$$

bestehend aus den Elementen

- eines datenfreien Entscheidungsproblems $(\mathbb{A}, \Theta, u(\cdot))$ in Nutzenform bzw. $(\mathbb{A}, \Theta, l(\cdot))$ in Verlustform und

- eines statistischen Modells $(\mathcal{X}, \sigma(\mathcal{X}), (p_\vartheta)_{\vartheta \in \Theta})$, dem sogenannten *Informationsbeschaffungsexperiment*, auch als *Informationsstruktur* bezeichnet.

Dabei werde implizit angenommen, dass $\sigma(\mathcal{X})$ alle Einpunktmengen $\{x\}$ mit $\{x\} \in \mathcal{X}$ enthält.

Man beachte, dass der Zustandsraum Θ des datenfreien Problems genau der Indexmenge der Wahrscheinlichkeitsmaße entspricht; das Informationsbeschaffungsexperiment soll ja (potentiell) informativ sein.

Def. und Bem. 1.40 Entscheidungsfunktionen, Auswertung datengestützter Entscheidungsprobleme

Gegeben sei ein datengestütztes Entscheidungsproblem (1.15) in Nutzenform bzw. in Verlustform.

Jede messbare Abbildung $d : \mathcal{X} \longrightarrow \mathbb{A}$ heißt *Entscheidungsfunktion* (Strategie).

Das (formal) datenfreie Entscheidungsproblem

$$\begin{array}{ll} (\mathcal{D}, \Theta, \mathbf{U}) & \text{in Nutzenform bzw.} \\ (\mathcal{D}, \Theta, \mathbf{R}) & \text{in Verlustformform,} \end{array}$$

mit

- $\emptyset \neq \mathcal{D} \subseteq \tilde{\mathbb{A}}^{\sigma(\mathcal{X})}$, der Menge aller messbaren Abbildungen von \mathcal{X} nach \mathbb{A} , und

-

$$\mathbf{U} : \mathcal{D} \times \Theta \longrightarrow \mathbb{R} \tag{1.16}$$

$$\begin{aligned} (d, \vartheta) &\longmapsto \mathbf{U}(d, \vartheta) = \int u(d(x), \vartheta) dp_{\vartheta}(x) \\ &= \mathbb{E}_{p_{\vartheta}} \underbrace{u(d(X), \vartheta)} \end{aligned}$$

bzw.

$$\mathbf{R} : \mathcal{D} \times \Theta \longrightarrow \mathbb{R} \tag{1.17}$$

$$\begin{aligned} (d, \vartheta) \longmapsto \mathbf{R}(d, \vartheta) &= \int l(d(x), \vartheta) dp_{\vartheta}(x) \\ &= \mathbb{E}_{p_{\vartheta}} l(d(X), \vartheta) \end{aligned}$$

heißt *zugeordnetes Auswertungsproblem* (eingeschränkt auf \mathcal{D}), und $(\mathbb{A}, \Theta, u(\cdot))$ bzw. $(\mathbb{A}, \Theta, l(\cdot))$ heißt dann zugehöriges *datenfreies Basisproblem*.

$\mathbf{R}(d, \vartheta)$ aus (1.17) heißt, als Funktion von ϑ für festes $d \in \mathcal{D}$ betrachtet, *Risikofunktion* der Entscheidungsfunktion d . (Die auf der Nutzenform basierende analoge Funktion $\mathbf{U}(d, \vartheta)$ wird öfter als *gain* bezeichnet. In der statistischen Entscheidungstheorie wird praktisch ausschließlich in der Verlustform gearbeitet.)

Bem. 1.41 (Auswertungsprinzip)

Entscheidungsfunktionen werden mit Hilfe des zugeordneten Auswertungsproblems beurteilt.

Bem. 1.42

- a) Da datengestützte Entscheidungsprobleme mit Hilfe des zugeordneten Auswertungsproblems ‚gelöst‘ werden und dieses formal die Struktur eines datenfreien Entscheidungsproblems besitzt, gelten die in Kapitel 2 später gemachten Aussagen über Eigenschaften optimaler Aktionen auch für Entscheidungsfunktionen. Damit werden wir einen „Berg“ an Korollaren erhalten. Wenn die Ergebnisse von Kapitel 2 allgemein meist im Kontext von [formal datenfreien] Entscheidungsproblemen formuliert werden, so gelten sie natürlich insbesondere für Entscheidungsfunktionen.

b) Man kann auch entscheiden, ob sich die Berücksichtigung von Zusatzinformation, die ja normalerweise auch mit Kosten verbunden ist, lohnt: Die Differenz aus dem Kriteriumswert (siehe Kapitel 2) der optimalen Entscheidungsfunktion im datengestützten Entscheidungsproblem und dem Kriteriumswert der optimalen Aktion in dem ursprünglich zugrunde gelegten datenfreien Basisproblem liefert den sogenannten *Informationswert*.

Bsp. 1.43 (Auswertungsproblem im Investitionsproblem)

Das Auswertungsproblem Betrachten Sie das Investitionsproblem mit der in der in Kapitel 1.5 beschriebenen Informationsstruktur:

<u>Datenfreies Basisproblem</u>	<u>Informationsstruktur</u>																												
<table style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="border-right: 1px solid black; padding: 5px;"></th> <th style="padding: 5px;">ϑ_1</th> <th style="padding: 5px;">ϑ_2</th> <th style="padding: 5px;">ϑ_3</th> </tr> </thead> <tbody> <tr> <td style="border-right: 1px solid black; padding: 5px;">a_1</td> <td style="padding: 5px;">10 000</td> <td style="padding: 5px;">2 000</td> <td style="padding: 5px;">-15 000</td> </tr> <tr> <td style="border-right: 1px solid black; padding: 5px;">a_2</td> <td style="padding: 5px;">1 000</td> <td style="padding: 5px;">1 000</td> <td style="padding: 5px;">0</td> </tr> </tbody> </table>		ϑ_1	ϑ_2	ϑ_3	a_1	10 000	2 000	-15 000	a_2	1 000	1 000	0	<table style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="border-right: 1px solid black; padding: 5px;"></th> <th style="padding: 5px;">x_1</th> <th style="padding: 5px;">x_2</th> <th style="padding: 5px;">x_3</th> </tr> </thead> <tbody> <tr> <td style="border-right: 1px solid black; padding: 5px;">ϑ_1</td> <td style="padding: 5px;">0.6</td> <td style="padding: 5px;">0.3</td> <td style="padding: 5px;">0.1</td> </tr> <tr> <td style="border-right: 1px solid black; padding: 5px;">ϑ_2</td> <td style="padding: 5px;">0.2</td> <td style="padding: 5px;">0.4</td> <td style="padding: 5px;">0.4</td> </tr> <tr> <td style="border-right: 1px solid black; padding: 5px;">ϑ_3</td> <td style="padding: 5px;">0.1</td> <td style="padding: 5px;">0.4</td> <td style="padding: 5px;">0.5</td> </tr> </tbody> </table>		x_1	x_2	x_3	ϑ_1	0.6	0.3	0.1	ϑ_2	0.2	0.4	0.4	ϑ_3	0.1	0.4	0.5
	ϑ_1	ϑ_2	ϑ_3																										
a_1	10 000	2 000	-15 000																										
a_2	1 000	1 000	0																										
	x_1	x_2	x_3																										
ϑ_1	0.6	0.3	0.1																										
ϑ_2	0.2	0.4	0.4																										
ϑ_3	0.1	0.4	0.5																										

Notation wiederum:

$$d(i_1, i_2, \dots, i_s) \quad \text{für} \quad d = \begin{pmatrix} x_1 & x_2 & \dots & x_s \\ a_{i_1} & a_{i_2} & \dots & a_{i_s} \end{pmatrix}$$

Datengestütztes Entscheidungsproblem: Auswertungsproblem

$$R(d, \vartheta) = \sum_x u(\underbrace{d(x)}_a, \vartheta) p_{\vartheta}(\{x\})$$

	ϑ_1	ϑ_2	ϑ_3
$d(1, 1, 1)$	10 000	2 000	-15 000
$d(1, 1, 2)$	9 100	1 600	- 7 500
$d(1, 2, 1)$	7 300	1 600	- 9 000
$d(1, 2, 2)$	6 400	1 200	- 1 500
$d(2, 1, 1)$	4 600	1 800	-13 500
$d(2, 1, 2)$	3 700	1 400	- 6 000
$d(2, 2, 1)$	1 900	1 400	- 7 500
$d(2, 2, 2)$	1 000	1 000	0

1.5.3 Einbettung der Test- und Schätztheorie in die statistische Entscheidungstheorie

Bem. und Bsp. 1.45 (Einbettung der Schätztheorie)

Die Aufgabe, einen Parameter aus einer i.i.d. Stichprobe zu schätzen, kann in die Entscheidungstheorie eingebettet werden.

$(X_1, \dots, X_n) \equiv \vartheta_0$ (Zweimal am Tag zeigt eine stehengebliebene Uhr die aktuelle Uhrzeit absolut korrekt an.).

Diese Einbettung umfasst auch Regressionsmodelle, betrachte dazu die folgende Bemerkung.

Bem. 1.46 (Zur Einbettung von Regressionsmodellen)

Bsp. 1.47 (Einbettung Testtheorie)

Auch die üblichen Testprobleme lassen sich in die Entscheidungstheorie einbetten.

Weiteres Beispiel für Test: the lady tasting tea → spätere Übungsaufgabe
Wir werden später generell auch randomisierte Aktionen zulassen.

1.5.4 Komplexität datengestützter Entscheidungsprobleme, konditionale versus frequentistische Sicht

2 **Entscheidungskriterien**

2.1 Vorbereitende Überlegungen

2.1.1 Entscheidungsregeln – Entscheidungsprinzipien

Bem. 2.1 (Entscheidungsregeln und Entscheidungsprinzipien)

- Eine Menge ist *linear geordnet* bzw. eine Ordnung \prec ist *vollständig*, wenn für zwei Elemente a, b stets gilt
$$a \prec b \text{ oder } a \succ b \text{ oder } a \sim b. \quad (\sim : \text{äquivalent})$$

Beispiele:

\mathbb{R} ist bezüglich der üblichen Ordnung vollständig geordnet, aber \mathbb{R}^n etwa bezüglich der komponentenweisen Ordnung nicht.

- Mit *Entscheidungsregeln* bringt man die Aktionen in eine vollständige, lineare Ordnung
 - + Ist der Aktionsraum endlich, dann gibt es „größte“/ „beste“ Aktionen.
 - Allerdings: Für die konkrete Gestalt von Entscheidungsregeln gibt es i.a. keinen universalen Gültigkeitsanspruch.
- *Entscheidungsprinzipien* sind Grundsätze für die Auswahl von Aktionen, die beanspruchen, allgemeine Rationalitätsprinzipien zu sein.

Dafür ist ihre „Ordnungskraft“ nicht so stark: Man wird im Allgemeinen nur gewisse ungünstige Aktionen als „unzulässig“ ausschließen können.

- Später werden wir sehen: Entscheidungs*prinzipien* sind unabhängig vom Unsicherheitsverständnis (ob Typ I oder Typ II), die Entscheidungs*regeln* hingegen nicht.

Zulässigkeit und Dominanzprinzip

Def. 2.2 (Dominanzrelationen)

Seien $a_1, a_2 \in \mathbb{A}$.

a) a_1 dominiert $a_2 \iff$

$$u(a_1, \vartheta) \geq u(a_2, \vartheta) \quad \text{für alle } \vartheta \in \Theta.$$

Man schreibt:

$$a_1 \succcurlyeq a_2$$

b) a_1 dominiert a_2 strikt \iff

$$u(a_1, \vartheta) \geq u(a_2, \vartheta) \quad \text{für alle } \vartheta \in \Theta.$$

und

$$u(a_1, \vartheta) > u(a_2, \vartheta) \quad \text{für mindestens ein } \vartheta \in \Theta.$$

Man schreibt:

$$a_1 \succ a_2$$

c) a_1 dominiert a_2 stark \iff

$$u(a_1, \vartheta) > u(a_2, \vartheta) \quad \text{für alle } \vartheta \in \Theta$$

Man schreibt:

$$a_1 \succ\succ a_2$$

d) a_1, a_2 sind *äquivalent* \iff
 $a_1 \succeq a_2$ und $a_2 \succeq a_1$

Man schreibt:

$$a_1 \sim a_2$$

- Für äquivalente Aktionen gilt:

$$a_1 \sim a_2 \iff u(a_1, \vartheta) = u(a_2, \vartheta) \quad \text{für alle } \vartheta \in \Theta$$

- Gemäß der in Bem. 1.6 postulierten Äquivalenz von Nutzen- und Verlustsicht kehren sich bei Verlusttafeln die „Vorzeichen“ in Def. 2.2 a) bis c) einfach um, z.B.:

$$a_1 \succeq a_2 \iff l(a_1, \vartheta) \leq l(a_2, \vartheta) \quad \text{für alle } \vartheta \in \Theta$$

- Die Dominanzrelationen gelten i.A. nicht für beliebige Paare von Aktionen: sie sind *nicht vollständig*. Sie liefern jedoch eine Klasseneinteilung der Aktionen.

Def. 2.3

Eine Aktion $a \in \mathbb{A}$ heißt *admissibel (zulässig)* \iff
 a wird von keiner Aktion strikt dominiert.

Die Aktion $a \in \mathbb{A}$ heißt *inadmissibel (unzulässig)* \iff
Es gibt eine Aktion a' , so dass $a' \succ a$.

Bem. 2.4 (Dominanzprinzip)

Im Folgenden werden zwei äquivalente Formulierungen gegeben:

Dominiert die Aktion a_1 die Aktion a_2 strikt, das heißt, gilt:

$$a_1 \succ a_2,$$

so ist es nicht vernünftig, die Aktion a_2 zu wählen.

oder:

Es ist nicht vernünftig, eine inadmissible Aktion zu wählen.

Bem. 2.5

Das Dominanzprinzip ist das Rationalitätsprinzip der hier zugrunde gelegten Form der Entscheidungstheorie schlechthin.

Die in Bemerkung 2.4 der getroffenen Grundannahmen der Gültigkeit des Dominanzprinzips kann man somit als Test für das Zutreffen des hier verwendeten Modellierungsrahmen ansehen.

Hat man bei der Betrachtung einer Entscheidungstafel das Gefühl, dass die Allgemeingültigkeit des Dominanzprinzips bezweifelt werden kann, muss man prüfen, ob nicht grundlegende Annahmen wie insbesondere die Handlungsunabhängigkeit der Zustände verletzt sind.

Bem. 2.6 (Potentielle Inadmissibilität beim Übergang zur gemischten Erweiterung)

Vorsicht: Man kann bereits an einfachen Beispielen sehen, dass beim Übergang einer „reinen“ Aktionenmenge \mathbb{A} zur gemischten Erweiterung $\mathcal{M}(\mathbb{A})$ Aktionen, die in \mathbb{A} zulässig sind, in $\mathcal{M}(\mathbb{A})$ inadmissibel werden können.

Def. 2.7 ((Wesentlich) vollständige Klassen)

- a) Eine Teilmenge $\mathbb{C} \subset \mathbb{A}$ heißt *vollständig (complete)* \iff
 Für jede Aktion $a \in \mathbb{A} \setminus \mathbb{C}$ gibt es ein $a^* \in \mathbb{C}$, so dass

$$a^* \succ a$$

- b) Eine Aktionenmenge $\mathbb{C} \subset \mathbb{A}$ heißt *wesentlich vollständig (essentially complete)* \iff

Zu jeder Aktion $a \in \mathbb{A}$ (oder $\mathbb{A} \setminus \mathbb{C}$) gibt es ein $a' \in \mathbb{C}$, so dass

$$a' \succeq a.$$

Bem. 2.8

- i) Vollständige Klassen müssen alle admissiblen Aktionen enthalten; sie können auch nicht-admissible Aktionen enthalten.

- ii) Interessant sind vor allem *minimale* Klassen, also (wesentlich) vollständige Klassen zu denen keine Untermengen existieren, die ebenfalls (wesentlich) vollständig sind. Eine minimal vollständige Klasse ist eindeutig, minimale wesentlich vollständige Klassen kann es mehrere geben.

iii) Minimal vollständige Klassen müssen nicht existieren. Wenn eine minimal vollständige Klasse \mathbb{C}_{\min} existiert, dann gilt

$$\mathbb{C}_{\min} = \bigcap_{\mathbb{C} \text{ vollständig}} \mathbb{C}$$

und

$$\mathbb{C}_{\min} = \mathbb{A}_{ad}.$$

iv) Eine minimale wesentlich vollständige Menge (keine Eindeutigkeit) bildet die weitestgehende Reduktion eines Entscheidungsproblems auf seinen „wesentlichen“ Kern. Sie enthält aus jeder Äquivalenzmenge von admissiblen Aktionen genau einen Repräsentanten.

Optimalitätskriterien

Def. 2.9 (Optimalitätskriterium / Entscheidungsregel)

Gegeben sei ein datenfreies Entscheidungsproblem $(\mathbb{A}, \Theta, u(\cdot))$.

Eine Abbildung

$$\begin{aligned}\Phi : \mathbb{A} &\longrightarrow \mathbb{R}^q \\ a &\longmapsto \Phi(a)\end{aligned}$$

heißt *q-stufiges Optimalitätskriterium* oder *Entscheidungskriterium*.

Die Entscheidungsregel lautet dann: Suche

$$a^* \in \mathbb{A} \quad \text{mit} \quad \Phi(a^*) \geq \Phi(a) \quad \forall a \in \mathbb{A}. \quad (2.1)$$

a^* heißt dann Φ -*optimale Aktion*, wobei im Fall $q \geq 2$ „ \geq “ die lexikographische Ordnung meint.

Bem. 2.10 (Nichtexistenz optimaler Aktionen)

Es muss weder eine optimale noch eine zulässige Aktion existieren.

Bem. 2.11

Es scheint mir nützlich, den Begriff der wesentlichen Vollständigkeit auch auf Optimalitätskriterien auszudehnen. Man würde dann eine Menge $\mathbb{C}_\Phi \subset \mathbb{A}$ als *wesentlich vollständig bezüglich dem Optimalitätskriterium Φ* bezeichnen, wenn es zu jedem $a \in \mathbb{A}$ ein $a' \in \mathbb{C}_\Phi$ gibt, so dass $\Phi(a') \geq \Phi(a)$.

2.2 Entscheiden unter Sicherheit

Betrachtet wird hier der Fall $|\Theta| = 1$ (bzw. die Situation $u(a, \vartheta)$ bzw. $l(a, \vartheta)$ konstant in $\vartheta \in \Theta$ für alle $a \in \mathbb{A}$). Da hier keine Unsicherheit über die Umweltzustände herrscht (bzw. der Nutzen konstant in den Umweltzuständen ist), muss man „nur“ noch diejenige(n) Aktion(en) suchen, die maximalen Nutzen liefert/liefern. Ein Problem dabei ist, dass bei den in der Praxis üblicherweise betrachteten Problemen \mathbb{A} hoch dimensional ist und nur durch eine Vielzahl von Restriktionen „indirekt“ beschrieben wird. Die Aufgabe, das Maximum (Minimum) einer Funktion (hier des Nutzens) über einer bestimmten Menge (hier der Aktionenmenge) zu finden, ist ein sehr gängiges Problem, das in vielen Kontexten auftritt und deshalb eine wichtige Rolle in der numerischen Mathematik spielt (Optimierung unter Nebenbedingungen).

Wir werden im folgenden v. a. lineare Optimierungsprobleme betrachten. Die entsprechenden Techniken und Resultate erlauben nicht nur

- eine effiziente Lösung von Entscheidungsproblemen unter Sicherheit

sondern bilden auch

- eine wesentliche Grundlage zur mathematischen Behandlung von Entscheidungsproblemen unter Unsicherheit, wie sie in den folgenden Abschnitten betrachtet werden.

2.2.1 Grundlegendes: Lineare Optimierung

- *Optimierung*: Suche Extremstellen (Minimum, Maximum) einer Funktion $f : D_1 \rightarrow \mathbb{R}$, über einer Menge $D_2 \subseteq D_1$. Dabei
 - * wird jetzt $f(\cdot)$ meist als *Zielfunktion* bezeichnet, und
 - * D_2 ergibt sich über „*Restriktionen, Nebenbedingungen*“

- *lineare Optimierung, lineare Programmierung*:
sowohl $f(\cdot)$ als auch die D_2 beschreibende Nebenbedingungen sind *linear* in den Komponenten („Variablen“) $\vec{x} \in D_1$
- *endliche lineare Optimierung* (hier im folgenden vorausgesetzt)
 - * $D_1 \subseteq \mathbb{R}^m$; $m < \infty$: die Anzahl der Variablen ist endlich.
 - * die Anzahl der linearen Nebenbedingungen ist endlich. Damit ist D_2 eine polyedrische Menge bzw. meist ein konvexes Polyeder.

2.2.2 Typische Beispiele und Problemstellung

Bsp. 2.12 (Produktionsplanung)

Maximiere Gesamtgewinn aus verschiedenen Produkten unter Restriktionen an

- Produktmenge (Lagermenge, Transportmenge)
- Produktionskapazität
- Ressourcenmenge

bei als fest angenommenem Preis und unbegrenztem Absatz.

Etwa im Beispiel 1.17. Ist beispielsweise bekannt, dass je Mengeneinheit von Produkt 1 ein Gewinn von 3€ und bei Produkt 2 von 4€ pro Mengeneinheit, erzielt wird, so lautet der Ansatz zur Bestimmung des gewinnoptimalen Plans:

Bsp. 2.13 (Mischungsprobleme)

Minimiere Kosten bei Einsatz verschiedener Ressourcen unter Nebenbedingungen an einzelne Komponenten (z.B. verschiedene Düngemittel mit verschiedenen Anteilen mehrerer Nährstoffe)

Bsp. 2.14 (Investitionsplanung)

Minimiere laufende Investitionskosten unter Bedingungen an Fertigungskapazität und Investitionsbudget.

Bsp. 2.15 (Transportoptimierung und Zuordnungsprobleme)

- Verteile Einheiten kostenoptimal unter Restriktionen an Einsatzbereiche (z.B. minimiere Fahrzeiten einer Menge von Notarztwägen an verschiedenen Orten unter der Bedingung, dass alle Einsatzorte bedient werden)
- verteile Servicekräfte auf verschiedene zu bedienende Bereiche optimal hinsichtlich der benötigten Einarbeitungszeit

Diese Fragestellungen führen meist explizit auf ganzzahlige Probleme, diese werden hier nur am Rande betrachtet

2.2.3 Standard-Probleme der linearen Optimierung und ihre Lösung

Da lineare Optimierung später auf verschiedene Art verwendet wird, werden hier die Grundkonzepte allgemein besprochen. Zu maximieren/minimieren ist eine lineare Funktion $c^T w$ bei gegebenem c (z.B. Preise/Kosten) in der Variablen w über einem konvexen Polyeder der Form $R \cdot w \geq b$ bzw. $R \cdot w \leq b$, also z.B. der Ationenmenge \mathbb{A} .

Def. 2.16 (Standard-Minimum-Problem und Standard-Maximum Problem)

Ein Optimierungsproblem der Form

$$\begin{array}{c} c^T \cdot w \\ (1 \times n)(n \times 1) \end{array} \rightarrow \min_w \quad (2.2)$$

unter den Nebenbedingungen

$$\begin{array}{c} R \cdot w \\ (m \times n)(n \times 1) \end{array} \geq \begin{array}{c} b \\ (m \times 1) \end{array}$$

$$\begin{array}{c} w \\ (n \times 1) \end{array} \geq \begin{array}{c} 0 \\ (n \times 1) \end{array} \quad (2.3)$$

heißt *Standard-Minimum-Problem*² (in der Variablen $w = (w[1], \dots, w[n])^T \in (\mathbb{R}_0^+)^n$ mit der Zielfunktion $c^T \cdot w$). Jedes die Restriktionen (2.3) erfüllendes \bar{w} mit $c^T \cdot \bar{w} > -\infty$ heißt *zulässig*³ oder *zulässige Lösung (des Standard-Minimum-Problems)*.

Eine zulässige Lösung w^* heißt *Optimallösung*⁴ des Standard-Minimum-Problems, wenn für jedes zulässige \bar{w} gilt: $c^T \cdot \bar{w} \geq c^T \cdot w^*$.

²Die in der Literatur verwendeten Bezeichnungen sind sehr heterogen. Gleiche Begriffe können bei unterschiedlichen Autoren durchaus Verschiedenes bedeuten; dies gilt insbesondere für die Termini „Standard-Form“ und „kanonische Form.“

³Vorsicht: Leider wird hier, wie in der Literatur üblich, der Begriff Zulässigkeit in einem doppelten Sinne gebraucht. Zulässigkeit von Aktionen als nicht strikte Dominiertheit und Zulässigkeit als Verträglichkeit mit den Nebenbedingungen. Im Entscheidungsproblem unter Unsicherheit ist dabei eine zulässige Aktion im ersten Sinne bereits eine Optimallösung (Dominanz und Optimalität fallen bei $|\Theta| = 1$ zusammen); ansonsten sind aber Verwechslungen wohl ausgeschlossen.

⁴Man beachte, daß hier nur von Zulässigkeit und Optimalität gesprochen wird, wenn der zugehörige Wert $c^T \cdot w$ endlich ist.

Analog heißt ein Optimierungsproblem *Standard-Maximum-Problem* (in der Variablen $w = (w[1], \dots, w[n])^T \in (\mathbb{R}_0^+)^n$ mit der Zielfunktion $c^T \cdot w$), wenn es folgende Gestalt besitzt:

$$\begin{array}{c} c^T \cdot w \\ (1 \times n)(n \times 1) \end{array} \rightarrow \max_w \quad (2.4)$$

unter den Nebenbedingungen

$$\begin{array}{c} R \cdot w \\ (m \times n)(n \times 1) \end{array} \leq \begin{array}{c} b \\ (m \times 1) \end{array} \\ \begin{array}{c} w \\ (n \times 1) \end{array} \geq \begin{array}{c} 0 \\ (n \times 1) \end{array} . \quad (2.5)$$

Jedes die Restriktionen (2.5) erfüllende \bar{w} mit $c^T \cdot \bar{w} < \infty$ heißt *zulässig* oder *zulässige Lösung* (des Standard-Maximum-Problems).

Eine zulässige Lösung w^* heißt *Optimallösung* des Standard-Maximum-Problems, wenn für jedes zulässige \bar{w} gilt: $c^T \cdot \bar{w} \leq c^T \cdot w^*$.

Dabei wurde von folgenden *Konventionen* Gebrauch gemacht:

1. Bei einem Vektor $y \in \mathbb{R}^q$ bezeichnet $y[\ell]$ für jedes $\ell \in \{1, \dots, q\}$ die ℓ -te Komponente von y .
2. Für zwei Vektoren $y \in \mathbb{R}^q$ und $z \in \mathbb{R}^q$ sind die Relationen „ \leq “, „ \geq “ und „ $=$ “ jeweils komponentenweise zu lesen. Beispielsweise gilt also:

$$y \leq z : \iff y[\ell] \leq z[\ell], \quad \forall \ell \in \{1, \dots, q\}.$$

Bem. 2.17 (Alle Probleme können auf die Standardformen gebracht werden.)

Jede lineare Optimierungsaufgabe läßt sich als ein Standard-Minimum-Problem und als ein Standard-Maximum-Problem formulieren:

- Durch Multiplizieren mit (-1) können Maximierungsprobleme in Minimierungsprobleme und Restriktionen der Form „ \leq “ in Nebenbedingungen der Gestalt „ \geq “ umgewandelt werden (und umgekehrt).
- Gleichheitsbedingungen in den Nebenbedingungen lassen sich in zwei Ungleichungen auflösen.

- Auch mit negativen Variablen kann umgegangen werden; man schreibt sie als Differenz zweier nichtnegativer Variablen.

Typischerweise bieten aus diesem Grund Programmpakete meist nur eine bestimmte Form an. □

Proposition 2.18 (Zur Lösungsmenge linearer Programme)

Gegeben sei ein Standard-Minimum-Problem (bzw. ein Standard-Maximum-Problem) gemäß Definition 2.16. Dann gilt:

1. Ist die Menge aller zulässigen Lösungen nach unten bzw. nach oben beschränkt, so existiert genau dann eine Optimallösung w^* , wenn es ein zulässiges w_0 gibt.
2. Die Menge W^* aller Optimallösungen ist konvex und abgeschlossen. \square

Bem. 2.19 (Zur Lösung von Ungleichungen)

Man kann mit Hilfe der linearen Optimierung auch überprüfen, ob ein Ungleichungssystem

$$R \cdot w \leq b$$

eine Lösung hat.

Dazu ersetzt man eine Komponente von b (z.B. $b[1]$) durch eine weitere Variable z und untersucht, ob das Problem

$$z \longrightarrow \min_{w,z}$$

unter den Nebenbedingungen

$$\begin{aligned} r_{11} \cdot w[1] + r_{12} \cdot w[2] + \dots - z &\leq 0 \\ r_{12} \cdot w[2] + r_{22} \cdot w[2] + \dots &\leq b[2] \\ &\dots \end{aligned}$$

eine Lösung kleiner gleich $b[1]$ hat.

Alternativ kann man eine künstliche, beschränkte Zielfunktion einführen und das Problem

$$v \longrightarrow \max_{v,w}$$

unter den Nebenbedingungen

$$R \cdot w \leq b$$

$$v \leq 1$$

betrachten. Man erhält hier theoretisch sogar alle Lösungen.

Bem. 2.20 (Berechnung von Optimallösungen)

- Zur Lösung linearer Optimierungsprobleme steht eine Vielzahl von Algorithmen zur Verfügung. Das älteste und bekannteste Verfahren ist das so genannte *Simplex-Verfahren*, das auch in allen gängigen Programmpaketen implementiert ist.

Es benutzt – wie seine vielen Abwandlungen (und Weiterentwicklungen) – die Tatsache, dass gemäß Satz 1.23 einer der Extrempunkte zu einer Optimallösung führen muss, und sucht daher in „geschickter Weise“ die Extrempunkte ab.

Daneben gibt es noch eine Reihe von sehr effizienten „innere-Punkte-Methoden“ → Literatur.

- Bei zwei (oder drei) Variablen ist auch eine graphische Lösung möglich.

Bsp. 2.21 (Graphische Lösung von Beispiel 1.17)

2.2.4 Dualität der linearen Optimierung I

Jedem linearen Programm in Standardform kann ein sogenanntes *duales Programm* zugeordnet werden. Es entsteht dadurch, daß man von einem Minimierungsproblem zu einem Maximierungsproblem (und umgekehrt) übergeht und dabei für jede Restriktion des ursprünglichen Problems eine neue Variable einführt. Als neue Koeffizientenmatrix der Restriktionen dient die Transponierte der ursprünglichen Koeffizientenmatrix. Ferner sind auch die Rollen von c und b zu vertauschen; c wird zum Spaltenvektor der Restriktionen, während nun b zu einem Faktor des die Zielfunktion bestimmenden Skalarprodukts wird.

Def. 2.22 Duale Standard-Minimum- und Standard-Maximum-Probleme

Gegeben sei das Standard-Minimum-Problem in der Variablen w aus Definition 2.16. Dann heißt die Optimierungsaufgabe

$$\begin{array}{c} b^T \cdot u \\ (1 \times m)(m \times 1) \end{array} \rightarrow \max_u \quad (2.6)$$

unter den Nebenbedingungen

$$\begin{array}{c} R^T \cdot u \\ (n \times m)(m \times 1) \end{array} \leq \begin{array}{c} c \\ (n \times 1) \end{array} \\ \begin{array}{c} u \\ (m \times 1) \end{array} \geq \begin{array}{c} 0 \\ (m \times 1) \end{array} \quad (2.7)$$

das *zugehörige duale Standard-Maximum-Problem*. In diesem Zusammenhang wird dann das ursprüngliche Problem als *primales Standard-Minimum-Problem* bezeichnet.

Analog wird bei gegebenem Standard-Maximum-Problem in der Variablen w die Optimierungsaufgabe

$$\underset{u}{b^T \cdot u} \rightarrow \min \quad (2.8)$$

$(1 \times m)(m \times 1)$

unter den Nebenbedingungen

$$\begin{aligned} \underset{(n \times m)(m \times 1)}{R^T \cdot u} &\underset{(n \times 1)}{\geq} \underset{(n \times 1)}{c} \\ \underset{(m \times 1)}{u} &\underset{(m \times 1)}{\geq} \underset{(m \times 1)}{0} \end{aligned} \quad (2.9)$$

als *zugehöriges duales Standard-Minimum-Problem* bezeichnet. Das ursprüngliche Problem heißt dann *primales Standard-Maximum-Problem*. \square

Der Zusammenhang zwischen Primalprogramm und Dualprogramm geht über die oben beschriebene, äußere Komplementarität weit hinaus:

Proposition 2.23 Zur Beziehung zwischen Primal-Problem und Dual-Problem

Gegeben sei das Standard-Minimum-Problem (2.2f.) in der Variablen w und das zugehörige duale Standard-Maximum-Problem (2.6f.) in der Variablen u . Dann gilt:

1. Das duale Standard-Minimum-Problem zu (2.6f.) ist wieder das ursprüngliche Standard-Minimum-Problem (2.2f.).
2. Ist \tilde{w} eine zulässige Lösung eines Standard-Minimum-Problems und \tilde{u} eine zulässige Lösung des zugehörigen dualen Standard-Maximum-Problems, so gilt

$$c^T \cdot \tilde{w} \geq b^T \cdot \tilde{u} .$$

\tilde{w} und \tilde{u} sind genau dann Optimallösungen, wenn in dieser Beziehung Gleichheit herrscht.

3. Das primale Standard-Minimum-Problem besitzt genau dann eine Optimallösung w^* , wenn für das duale Standard-Maximum-Problem eine Optimallösung u^* existiert. Gemäß oben sind dann beide Kriteriumswerte $c^T \cdot w^*$ und $b^T \cdot u^*$ gleich. \square

Bsp. 2.24 (Zur Dualität im Bsp. 1.17)

Bem. 2.25

Die Optimallösung $(u^*[1], \dots, u^*[m])^T$ des dualen Standard-Minimum-Problems besitzen auch eine inhaltliche Interpretation. Ihre Komponenten werden als *Schattenpreise* oder *Opportunitätskosten* bezeichnet und geben an, um wie viel sich der Zielfunktionswert ändert, wenn sich die zugehörige Restriktion im primalen Programm um eine Einheit erhöht, sofern diese Änderung als klein angesehen werden kann, so dass die Restriktionen nicht „grob“ verletzt werden. Man erhält also sozusagen den Betrag, den man maximal bereit wäre, für eine Erweiterung der Restriktion um eine Einheit zu zahlen.

Bsp. 2.26 (Schattenpreise im Bsp. 1.17)

(Fortsetzung von Bsp. 2.21)

2.2.5 Kanonische Form, Schlupfvariablen

In gewisser Weise ausgezeichnet sind spezielle Standard-Minimum-Probleme und Standard-Maximum-Probleme, bei denen alle Restriktionen in Gestalt von Gleichungen vorliegen. Dies lässt sich stets dadurch erreichen, dass man wie in Definition 2.27 in jeder Nebenbedingung eine nichtnegative Hilfsvariable („*Schlupfvariable*“) subtrahiert bzw. hinzuaddiert, die die Differenz zwischen beiden Seiten der Ungleichungen „auffängt“. In die Zielfunktion gehen Schlupfvariablen nicht mit ein. Wichtig ist es, auf ihre Nichtnegativität zu achten, da sonst die ursprünglichen Restriktionen nicht mehr erfüllt sein müssen.

Def. 2.27 (Kanonische Form)

Gegeben sei das Standard-Minimum-Problem (2.2f.) in der Variablen w .
Das Optimierungsproblem

$$c^T \cdot w \rightarrow \min_w \quad (2.10)$$

unter den Nebenbedingungen⁵

$$[R, -I_m] \cdot \begin{bmatrix} w \\ w_s \end{bmatrix} = b \quad (2.11)$$

$$\begin{matrix} w \\ (n \times 1) \end{matrix} \geq \begin{matrix} 0 \\ (n \times 1) \end{matrix} \quad (2.12)$$

$$\begin{matrix} w_s \\ (m \times 1) \end{matrix} \geq \begin{matrix} 0 \\ (m \times 1) \end{matrix} \quad (2.13)$$

⁵Dabei bezeichnet I_m für jedes $m \in \mathbb{N}$ die m -dimensionale Einheitsmatrix.

heißt *(zugehöriges) Standard-Minimum-Problem in kanonischer Form (in der (Haupt)Variablen w mit der Schlupfvariablen $w_s = (w_s[1], \dots, w_s[m])^T$).*

Analog heißt für das Standard-Maximum-Problem (2.4f.) in der Variablen w das Optimierungsproblem

$$c^T \cdot w \rightarrow \max_w \quad (2.14)$$

unter den Nebenbedingungen

$$[R, I_m] \cdot \begin{bmatrix} w \\ w_s \end{bmatrix} = b \quad (2.15)$$

$$\begin{matrix} w \\ (n \times 1) \end{matrix} \geq \begin{matrix} 0 \\ (n \times 1) \end{matrix} \quad (2.16)$$

$$\begin{matrix} w_s \\ (m \times 1) \end{matrix} \geq \begin{matrix} 0 \\ (m \times 1) \end{matrix} \quad (2.17)$$

(zugehöriges) Standard-Maximum-Problem in kanonischer Form (in der (Haupt-)Variablen w mit der Schlupfvariablen $w_s = (w_s[1], \dots, w_s[m])^T$). \square

Bsp. 2.28 Standard-Maximum Problem in kanonischer Form im Beispiel 1.17

Man bestimme das Standard-Maximum-Problem in der kanonischen Form.

Offensichtlich besteht folgender Zusammenhang zwischen den Lösungen eines Standard-Minimum- bzw. eines Standard-Maximum-Problems und den Lösungen des zugehörigen Problems in kanonischer Form:

Bem. 2.29 (Lösungen bei kanonischer Form und ursprünglicher Form)

Betrachtet man ein Standard-Minimum-Problem gemäß (2.2f.) bzw. ein Standard-Maximum-Problem laut (2.4f.) sowie die in (2.10ff.) und (2.14ff.) beschriebenen zugehörigen Standard-Minimum-Probleme bzw. Standard-Maximum-Probleme in kanonischer Form. Dann gilt:

1. Ist $(\bar{w}[1], \dots, \bar{w}[n], \bar{w}_s[1], \dots, \bar{w}_s[m])^T$ zulässig (resp. eine Optimallösung) für das Problem in kanonischer Form gemäß (2.10ff.) bzw. (2.14ff.), so ist $(\bar{w}[1], \dots, \bar{w}[n])^T$ zulässig (resp. eine Optimallösung) für das entsprechende Problem (2.2f.) bzw. (2.4f.).
2. Umgekehrt gibt es zu jeder zulässigen (resp. optimalen) Lösung $(\bar{w}[1], \dots, \bar{w}[n])^T$ des Problems (2.2f.) bzw. (2.4f.) eine zulässige (resp. optimale) Lösung $(\bar{w}[1], \dots, \bar{w}[n], \bar{w}_s[1], \dots, \bar{w}_s[m])^T$ des zugehörigen Problems in kanonischer Form laut (2.10ff.) bzw. (2.14ff.). \square

Der praktische Nutzen der Verwendung von Schlupfvariablen liegt darin, dass man durch sie erkennen kann, bei welchen Restriktionen noch „Spiel ist“, d.h. welche Kapazitätsbeschränkungen nicht zur Gänze ausgeschöpft werden.

2.2.6 Dualität II: Der Satz vom komplementären Schlupf

Vom theoretischen wie praktischen Blickwinkel ist der Satz vom komplementären Schlupf (s.u.) von großer Bedeutung. Wendet man zu seiner Vorbereitung die Dualitätsbetrachtungen aus Definition 2.22 auf Standard-Minimum- bzw. Standard-Maximum-Probleme in kanonischer Form an, so erhält man:

Bem. 2.30 Duale Programme bei Programmen in kanonischer Form

Ist ein Standard-Minimum-Problem in kanonischer Form der in (2.10ff.) beschriebenen Gestalt gegeben, so lautet das zugehörige duale Standard-Maximum-Problem in kanonischer Form:

$$b^T \cdot u \rightarrow \max_u \quad (2.18)$$

unter den Nebenbedingungen

$$\begin{bmatrix} R^T & I_n \end{bmatrix} \cdot \begin{bmatrix} u \\ u_r \end{bmatrix} = c \quad (2.19)$$

$$\begin{matrix} u \\ (m \times 1) \end{matrix} \geq \begin{matrix} 0 \\ (m \times 1) \end{matrix} \quad (2.20)$$

$$\begin{matrix} u_r \\ (n \times 1) \end{matrix} \geq \begin{matrix} 0 \\ (n \times 1) \end{matrix} . \quad (2.21)$$

Entsprechend lautet das zu einem Standard-Maximum-Problem in kanonischer Form (vgl. 2.14ff.) gehörende duale Standard-Minimum-Problem in kanonischer Form:

$$b^T \cdot u \rightarrow \min \quad (2.22)$$

unter den Nebenbedingungen

$$\begin{bmatrix} R^T & -I_n \end{bmatrix} \cdot \begin{bmatrix} u \\ u_r \end{bmatrix} = c \quad (2.23)$$

$$\begin{matrix} u \\ (m \times 1) \end{matrix} \geq \begin{matrix} 0 \\ (m \times 1) \end{matrix} \quad (2.24)$$

$$\begin{matrix} u_r \\ (n \times 1) \end{matrix} \geq \begin{matrix} 0 \\ (n \times 1) \end{matrix} . \quad (2.25)$$

□

Natürlich behalten die Dualitätsergebnisse aus Proposition 2.23 ihre Gültigkeit. Darüber hinaus lässt sich jedoch ein fundamentaler Zusammenhang zwischen den Optimallösungen eines Problems und den Schlupfvariablen des zugehörigen dualen Problems herleiten. Es kann nämlich aus der Existenz von echt von Null verschiedenen Komponenten einer Optimallösung auf das Verschwinden der entsprechenden Haupt- bzw. Schlupfvariablen in *allen* Optimallösungen des zugehörigen Dual-Problems geschlossen werden. Auch ist es möglich, zulässige Lösungen aufgrund ihres Schlupfes unter Umständen als optimal zu charakterisieren:

Proposition 2.31 (Der Satz vom komplementären Schlupf)

Betrachtet werde ein Standard-Minimum-Problem bzw. ein Standard-Problem in kanonischer Form gemäß (2.10ff.) bzw. (2.14ff.) und das zugehörige duale Problem in kanonischer Form, wie in (2.18ff.) bzw. (2.22ff.) beschrieben. Dann gilt:

1. Ist $(\tilde{w}^T, \tilde{w}_s^T)^T$ eine zulässige Lösung des primalen Problems und $(\tilde{u}^T, \tilde{u}_r^T)^T$ eine zulässige Lösung des zugehörigen dualen Problems, so sind beide genau dann optimal, wenn gilt

$$\tilde{w}^T \cdot \tilde{u}_r + \tilde{w}_s^T \cdot \tilde{u} = 0 . \quad (2.26)$$

Korollar 2.32

In der Situation von Proposition 2.31 gilt für alle $j = 1, \dots, m$:

a) Gibt es eine Optimallösung

$$(w^*[1], \dots, w^*[n], w_s^*[1], \dots, w_s^*[m])^T$$

für das primale Problem mit

$$w_s^*[j] > 0 ,$$

so ist

$$u^*[j] = 0$$

für alle Optimallösungen

$$(u^*[1], \dots, u^*[m], u_r^*[1], \dots, u_r^*[n])^T$$

des zugehörigen Dualproblems.

b) Gibt es eine Optimallösung

$$(u^*[1], \dots, u^*[m], u_r^*[1], \dots, u_r^*[n])^T$$

des Dualproblems mit

$$u^*[j] > 0 ,$$

so ist

$$w_s^*[j] = 0$$

für alle Optimallösungen

$$(w^*[1], \dots, w^*[n], w_s^*[1], \dots, w_s^*[m])^T$$

des primalen Problems.

□

Bsp. 2.33 (Der Satz vom komplementären Schlupf im Beispiel 2.28)

Illustrieren Sie anhand von Bsp 1.17 den Satz vom komplementären Schlupf!

2.2.7 Erweiterungen

Es gibt eine Reihe wichtiger Erweiterungen:

- nichtlineare Zielfunktionen

Hier gibt es eine Vielzahl von leistungsfähigen Algorithmen. Ist die Zielfunktion der Quotient zweier linearer Funktionen, so lässt sich eine geeignete Verallgemeinerung von Satz 1.23 und damit auch des Simplex-Algorithmus finden („*Quotientenoptimierung*“).

Bei allgemeinen Zielfunktionen wird das Extremum nicht notwendig in einem Extrempunkt angenommen.

Betrachtet man beispielsweise (Varianz der Bernoulli-Verteilung)

$$\begin{aligned} p(1 - p) &\rightarrow \max \\ p &\geq 0 \\ -p &\geq -1, \end{aligned}$$

so liegt die Optimallösung p^* bei $\frac{1}{2}$.

Dennoch lassen sich quadratische Optimierungsprobleme (wie z.B. das KQ-Kriterium unter Nebenbedingungen) mit einem Simplex-ähnlichen Verfahren lösen (etwa: Büning et. al., 2000, Kap. 8.9).

- *Parametrische Optimierung:*

Die Zielfunktion und/oder Restriktionen hängen von einem Parameter ab.

(Wichtig z.B. für Sensitivitätsanalysen) Auch hier sind Varianten z.B. des Simplex-Verfahren möglich.

- *Ganzzahlige Optimierung*

Sind die Lösungen inhaltlich zwingend ganzzahlig (z.B. die Aufteilung von wenigen unteilbaren Ressourcen), so ist eine eigenständige Betrachtung nötig.

Ein besonderer Spezialfall ist die *Boolesche Optimierung*, bei der nur 0, 1-Variablen zugelassen sind.

2.3 Minimax/Maximin Aktionen

2.3.1 Das Kriterium: Motivation und Definition

Bsp. 2.34 (Motivationsbeispiel Kuchen teilen) *vgl. das Bsp. in Abschnitt 1.3.3*

Def. 2.35 (Maximin/Minimax-Aktion)

a) Gegeben sei ein datenfreies Entscheidungsproblem $(\mathbb{A}, \Theta, u(\cdot))$ in Nutzenform.

Dann heißt $a^* \in \mathbb{A}$ *Maximin-Aktion*, wenn

$$\inf_{\vartheta \in \Theta} u(a^*, \vartheta) \geq \inf_{\vartheta \in \Theta} u(a, \vartheta) \quad \forall a \in \mathbb{A}. \quad (2.27)$$

b) Bei einem datenfreien Entscheidungsproblem in Verlustform $(\mathbb{A}, \Theta, l(\cdot))$ heißt $a^* \in \mathbb{A}$ *Minimax-Aktion*, wenn gilt

$$\sup_{\vartheta \in \Theta} l(a^*, \vartheta) \leq \sup_{\vartheta \in \Theta} l(a, \vartheta) \quad \forall a \in \mathbb{A}. \quad (2.28)$$

Bem. 2.36

Man verwendet als Kriteriumsfunktion

$$\begin{aligned}\Phi : \mathbb{A} &\rightarrow \mathbb{R} \\ a &\mapsto \inf_{\vartheta} u(a, \vartheta)\end{aligned}$$

bzw.

$$\begin{aligned}\Phi : \mathbb{A} &\rightarrow \mathbb{R} \\ a &\mapsto -\sup_{\vartheta} l(a, \vartheta)\end{aligned}$$

Abraham Wald (1902 – 1950)



Foto: Oberwolfach Photo Collection
(http://owpdb.mfo.de/person_detail?id=4398 [Stand: 01.07.2013])

Bsp. 2.37 (Ausflugs und Omelettenproblem)

Bsp. 2.38 (Maximin-Entscheidungsfunktion Investitionsproblem)

im

Betrachten Sie das Investitionsproblem mit der in der in Kapitel 1.5 beschriebenen Informationsstruktur:

Datenfreies Entscheidungsproblem Informationsstruktur

	ϑ_1	ϑ_2	ϑ_3
a_1	10 000	2 000	-15 000
a_2	1 000	1 000	0

	x_1	x_2	x_3
ϑ_1	0.6	0.3	0.1
ϑ_2	0.2	0.4	0.4
ϑ_3	0.1	0.4	0.5

Notation wiederum:

$$d(i_1, i_2, \dots, i_s) \quad \text{für} \quad d = \begin{pmatrix} x_1 & x_2 & \dots & x_s \\ a_{i_1} & a_{i_2} & \dots & a_{i_s} \end{pmatrix}$$

Bestimmen Sie die Maximin-Entscheidungsfunktion!

Datengestütztes Entscheidungsproblem: Auswertungsproblem

	ϑ_1	ϑ_2	ϑ_3	Minimum
$d(1, 1, 1)$	10 000	2 000	-15 000	- 15 000
$d(1, 1, 2)$	9 100	1 600	- 7 500	- 7 500
$d(1, 2, 1)$	7 300	1 600	- 9 000	- 9 000
$d(1, 2, 2)$	6 400	1 200	- 1 500	- 1 500
$d(2, 1, 1)$	4 600	1 800	-13 500	- 13 500
$d(2, 1, 2)$	3 700	1 400	- 6 000	- 6 000
$d(2, 2, 1)$	1 900	1 400	- 7 500	- 7 500
$d(2, 2, 2)$	1 000	1 000	0	- 0

Bem. 2.39 (Randomisieren und Maximin-Kriterium)

Der Übergang zur gemischten Erweiterung kann den Kriteriumswert erhöhen. Ist a^* Maximin-Aktion in $(\mathcal{M}(\mathbb{A}), \Theta, \tilde{u}(\cdot))$, so kann es sein, dass

$$\inf_{\vartheta \in \Theta} \tilde{u}(a^*, \vartheta) > \inf_{\vartheta \in \Theta} u(a, \vartheta) \quad \forall a \in \mathbb{A}$$

2.3.2 Berechnung/Auffinden von Minimax-/Maximin-Aktionen

**Bem. 2.40 (Die
Minimax/Maximin
Zustandsraum)**

**graphische
Aktionen**

**Bestimmung von
bei zweielementigem**

Bem. 2.42 (Vorüberlegung zur Bestimmung von Maximin-Aktionen über lineare Optimierung bei endlichen \mathbb{A} und Θ)

Proposition 2.43 (Bestimmung von Maximin-Aktionen über lineare Optimierung)

Gegeben sei die gemischte Erweiterung $(\mathcal{M}(\mathbb{A}), \Theta, u(\cdot))$ eines datenfreien Entscheidungsproblem in Nutzenform bei endlichem \mathbb{A} und endlichem Θ . Die randomisierte Aktion $\lambda^* = (\lambda(a_1), \dots, \lambda(a_n))$ ist Maximin-Lösung genau dann, wenn sie Optimallösung des folgenden Optimierungsproblems ist:

$$M \longrightarrow \max_{M, \lambda(a_1), \dots, \lambda(a_n)}$$

unter den Nebenbedingungen

$$\sum_{i=1}^n u(a_i, \vartheta_j) \lambda(a_i) \geq M, \quad j = 1, \dots, m,$$

$$\sum_{i=1}^n \lambda(a_i) = 1,$$

$$\lambda(a_i) \geq 0, \quad i = 1, \dots, n.$$

i) Es entsteht also folgendes Standardmaximum-Problem in den Variablen $M, \lambda(a_1), \dots, \lambda(a_n)$

$$M \longrightarrow \max_{M, \lambda(a_1), \dots, \lambda(a_n)}$$

unter den Nebenbedingungen

$$\begin{array}{rcl}
M - \sum_{i=1}^n u(a_i, \vartheta_1) \lambda(a_i) & \leq & 0 \\
\vdots & & \vdots \quad \vdots \\
M - \sum_{i=1}^n u(a_i, \vartheta_m) \lambda(a_i) & \leq & 0 \\
\sum_{i=1}^n \lambda(a_i) & \leq & 1 \\
-\sum_{i=1}^n \lambda(a_i) & \leq & -1 \\
\lambda(a_1) & \geq & 0 \\
\vdots & & \vdots \quad \vdots \\
\lambda(a_n) & \geq & 0 \\
M & \geq & 0,
\end{array}$$

wobei stillschweigend (wegen der Nichtnegativität von M) vorausgesetzt wurde, dass o. B. d. A. $u(a_i, \vartheta_j) \geq 0$ für alle $i = 1, \dots, n$ und $j = 1, \dots, m$ ist. (Sonst gehe man über zu $u(a_i, \vartheta_j) := u(a_i, \vartheta) - \min_{i,j} u(a_i, \vartheta_j)$, was dieselben optimalen Werte von $\lambda(a_1), \dots, \lambda(a_n)$ liefert.)

Also:

$$(1, 0, \dots, 0) \cdot \begin{pmatrix} M \\ \lambda(a_1) \\ \vdots \\ \lambda(a_n) \end{pmatrix} \longrightarrow \max_{M, \lambda(a_1), \dots, \lambda(a_n)}$$

unter den Nebenbedingungen

$$\begin{pmatrix} 1 & -u(a_1, \vartheta_1) & -u(a_2, \vartheta_1) & -\dots & -u(a_n, \vartheta_1) \\ 1 & -u(a_1, \vartheta_2) & -u(a_2, \vartheta_2) & -\dots & -u(a_n, \vartheta_2) \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & -u(a_1, \vartheta_m) & -u(a_2, \vartheta_m) & -\dots & -u(a_n, \vartheta_m) \\ 0 & 1 & 1 & \dots & 1 \\ 0 & -1 & -1 & \dots & -1 \end{pmatrix} \begin{pmatrix} M \\ \lambda(a_1) \\ \lambda(a_2) \\ \vdots \\ \lambda(a_n) \end{pmatrix} \leq \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ -1 \end{pmatrix}$$

$$\begin{pmatrix} M \\ \lambda(a_1) \\ \vdots \\ \lambda(a_n) \end{pmatrix} \geq \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Sucht man die optimale unrandomisierte Aktion, so kann man diese auch mithilfe Boolescher Optimierung aus diesem Optimierungsproblem gewinnen.

Def. 2.44 (Equalizer-Aktion, Äquilibrator-Aktion)

Eine Aktion $a \in \mathbb{A}$ mit in \mathcal{V} konstantem Nutzen heißt *Equalizer-Aktion* oder *Äquilibrator-Aktion*.

Bsp. 2.46 (\bar{X} als Minimax-Schätzer bei Normalverteilung)

Anwendung in der statistischen Entscheidungstheorie:

$$X_1, \dots, X_n \text{ i.i.d. } \sim N(\mu, \sigma_0^2)$$

Standardschätzer für μ ist \bar{X} , der auch unter allen erwartungstreuen Schätzern die gleichmäßig kleinste Varianz besitzt. (UMVU, siehe auch Kap. 1.5) Frage: Ist \bar{X} auch bezüglich des MSE als Risikofunktion Minimax-Schätzer?

2.3.3 Weitere Eigenschaften von Minimax-/Maximin-Aktionen

Proposition 2.47 (Invarianz gegenüber monotonen Transformationen des Nutzens)

Sei $g : \mathbb{R} \rightarrow \mathbb{R}$ eine monoton steigende Funktion.

Betrachtet man die Entscheidungsprobleme $(\mathbb{A}, \Theta, u(\cdot))$ und $(\mathbb{A}, \Theta, g \circ u(\cdot))$ mit

$$\begin{aligned} g \circ u : A \times \Theta &\rightarrow \mathbb{R} \\ (a, \vartheta) &\mapsto g(u(a, \vartheta)), \end{aligned}$$

so gilt:

- i) Ist a^* Maximin-Aktion in $(\mathbb{A}, \Theta, u(\cdot))$, so ist sie Maximin-Aktion in $(\mathbb{A}, \Theta, g \circ u(\cdot))$.
- ii) Ist g streng monoton und $|\Theta| < \infty$, so gilt sogar: Eine Aktion a^* ist Maximin-Aktion in $(\mathbb{A}, \Theta, u(\cdot))$ genau dann wenn sie Maximin-Aktion in $(\mathbb{A}, \Theta, g \circ u(\cdot))$ ist.

Bem. 2.48 (Tücken und formale Kritik des Maximin-Prinzips)

a) Bei unendlichen Mengen müssen Minimax/Maximin-Aktionen nicht existieren \longrightarrow Übergang zu ε -Minimax bzw. ε -Maximin-Aktionen

$$\inf_{\vartheta \in \Theta} u(a^*, \vartheta) \geq \inf_{\vartheta \in \Theta} u(a, \vartheta) - \varepsilon, \quad \forall a \in \mathbb{A}$$

(wird eher selten betrachtet, jedenfalls in der Vorlesung gar nicht.)

b) Minimax/Maximin-Aktionen müssen nicht eindeutig sein.

c) Minimax/Maximin-Aktionen müssen nicht zulässig sein.

d) Beim Übergang zur gemischten Erweiterung kann sich der Kriteriumswert erhöhen.

e) Kritik anhand der Metaregeln.

d1) Minimax/Maximin Kriterium ist nur bei endlichem Θ notwendig kompatibel mit der Dominanzrelation.

Kompatibilität mit der Dominanzrelation:

Ein Optimalitätskriterium Φ ist kompatibel mit der Dominanzrelation, wenn für alle $a, b \in \mathbb{A}$ gilt:

$$a \preceq b \implies \Phi(a) \leq \Phi(b) \quad (2.29)$$

$$a \prec b \implies \Phi(a) \leq \Phi(b) \quad (2.30)$$

$$a \prec\prec b \implies \Phi(a) < \Phi(b) \quad (2.31)$$

d2) Die Spaltenlinearität ist verletzt.

Spaltenlinearität:

Die durch Φ induzierte Präferenzordnung soll ungeändert bleiben, wenn für einen beliebigen Zustand ϑ_{j^*} die Spalte der Nutzenwerte um einen konstanten Betrag verändert wird

$$\begin{array}{ccc}
 u(a_1, \vartheta_{j^*}) & & u(a_1, \vartheta_{j^*}) + c \\
 u(a_2, \vartheta_{j^*}) & \longrightarrow & u(a_2, \vartheta_{j^*}) + c \\
 \vdots & & \vdots \\
 u(a_m, \vartheta_{j^*}) & & u(a_m, \vartheta_{j^*}) + c
 \end{array}$$

Satz 2.51 (Existenz Maximin-optimaler Lösungen bei endlichem Θ)

Gegeben sei ein datenfreies Entscheidungsproblem $(\mathbb{A}, \Theta, u(\cdot))$ mit endlicher Aktionsmenge \mathbb{A} und endlichem Zustandsraum Θ . Dann existiert eine Maximin-Aktion in \mathbb{A} und in $\mathcal{M}(\mathbb{A})$.

Beweis:

Korollar 2.53

Unter den gegebenen Voraussetzungen ist die Menge aller Maximin-Lösungen konvex und abgeschlossen.

**Bem. 2.54 (Maximin-Regel:
inhaltliche Kritik)**

Zusammenfassung

und

2.4 Bayes-Aktionen; konditionale Bayes-Inferenz

2.4.1 Entscheiden in der Risikosituation

Bsp. 2.55 (Beispiel: Lotterie vgl. Bsp. in Abschnitt 1.3.2)

Urne mit g

(hier: 50) grünen, b (30) blauen und r (20) restlichen Kugeln.

Man kann

a_1 nicht spielen

a_2 zum Preis von 20 auf Grün setzen

a_3 zum Preis von 10 auf Blau setzen

Nur einmaliger, zufälliger Zug aus der Urne, man erhält 100 Euro bei Treffer:

	grün	blau	rest	Min
a_1	0	0	0	0
a_2	80	-20	-20	-20
a_3	-10	90	-10	-10

Wie würden Sie sich entscheiden?

Def. 2.56

Gegeben seien

- ein Entscheidungsproblem $(\mathbb{A}, \Theta, u(\cdot))$ in Nutzenform
- eine σ -Algebra $\sigma(\Theta)$ über Θ , die alle Einpunktmengen $\{\vartheta\}$ enthält.
- Ein Wahrscheinlichkeitsmaß $\pi(\cdot)$ über $(\Theta, \sigma(\Theta))$.

Dann heißt für jedes $a \in \mathbb{A}$

$$\mathbb{E}_\pi(u(a)) := \int_{\Theta} u(a, \vartheta) \, d\pi(\vartheta) \quad (2.32)$$

der *Erwartungsnutzen* von a .

Bem. 2.57

- $u(a, \cdot)$ ist nun sozusagen eine Zufallsgröße $u(a)$!
Ein Zufallsexperiment determiniert ϑ , damit also auch den konkret sich einstellenden Nutzen

$$u(a) : \begin{array}{l} \Theta \rightarrow \mathbb{R} \quad \text{bzw.} \quad (\Theta, \sigma(\Theta)) \rightarrow (\mathbb{R}, \mathcal{B}) \\ \vartheta \mapsto u(a)(\vartheta) := u(a, \vartheta) \end{array}$$

- Zu (2.32) sei nochmals an die Definition des allgemeinen Maßintegral erinnert. Insbesondere gilt wiederum

* $\pi(\cdot)$ stetig mit Dichte $f(\cdot)$, dann

$$\mathbb{E}(u(a)) = \int_{\Theta} u(a, \vartheta) f(\vartheta) \, d\vartheta$$

* $\pi(\cdot)$ diskret mit Wahrscheinlichkeitsfunktion $\pi(\{\vartheta\})_{\vartheta \in \Theta}$

$$\mathbb{E}(u(a)) = \sum_{\vartheta \in \Theta} u(a, \vartheta) \pi(\{\vartheta\})$$

Def. 2.58 (Bernoullikriterium)

In der Situation von Def. 2.56 wähle man

$$\begin{aligned}\Phi(\cdot, \pi) &: \mathbb{A} \rightarrow \mathbb{R} \\ a &\longmapsto \Phi(a, \pi) = \mathbb{E}_\pi(u(a))\end{aligned}$$

als Kriteriumsfunktion!

Bem. 2.60

Gerade bei diesem Kriterium ist es absolut essentiell, sich daran zu erinnern, dass der Nutzen nicht notwendig linear in den Geldbeträgen ist.

Bem. 2.62 (Einbeziehen der Varianz)

V. a. bei wirtschaftlichen Anwendungen wird manchmal ergänzend die Varianz von $u(a, \cdot)$ (oder ein anderes Risikomaß) mit einbezogen; man spricht dann von (μ, σ) -Kriterien (μ Mittelwert, σ^2 Varianz).

Dies kann auf zwei Arten geschehen:

- 1) *Zweistufige* (μ, σ) -*Kriterien* (um zwischen Aktionen mit gleichen Erwartungswerten zu differenzieren), also etwa:

$$\tilde{\Phi} : \mathbb{A} \rightarrow \begin{pmatrix} \mathbb{E}_\pi u(a) \\ \mathbb{V}_\pi u(a) \end{pmatrix}$$

im Sinne der lexikographischen Ordnung (bzgl. $-v$), also

$$\begin{aligned} \tilde{\Phi}(a) > \tilde{\Phi}(a') &\iff \\ \mathbb{E}_\pi(u(a)) > \mathbb{E}_\pi(u(a')) &\text{ oder} \\ \mathbb{E}_\pi(u(a)) = \mathbb{E}_\pi(u(a')) &\text{ und } \mathbb{V}_\pi(u(a)) < \mathbb{V}_\pi(u(a')) \end{aligned}$$

2) *einstufige* (μ, σ) -*Kriterien* mit der Kriteriumsfunktion

$$\Phi(a) = \mathbb{E}_\pi(u(a)) - c \cdot \sqrt{\mathbb{V}_\pi(u(a))}$$

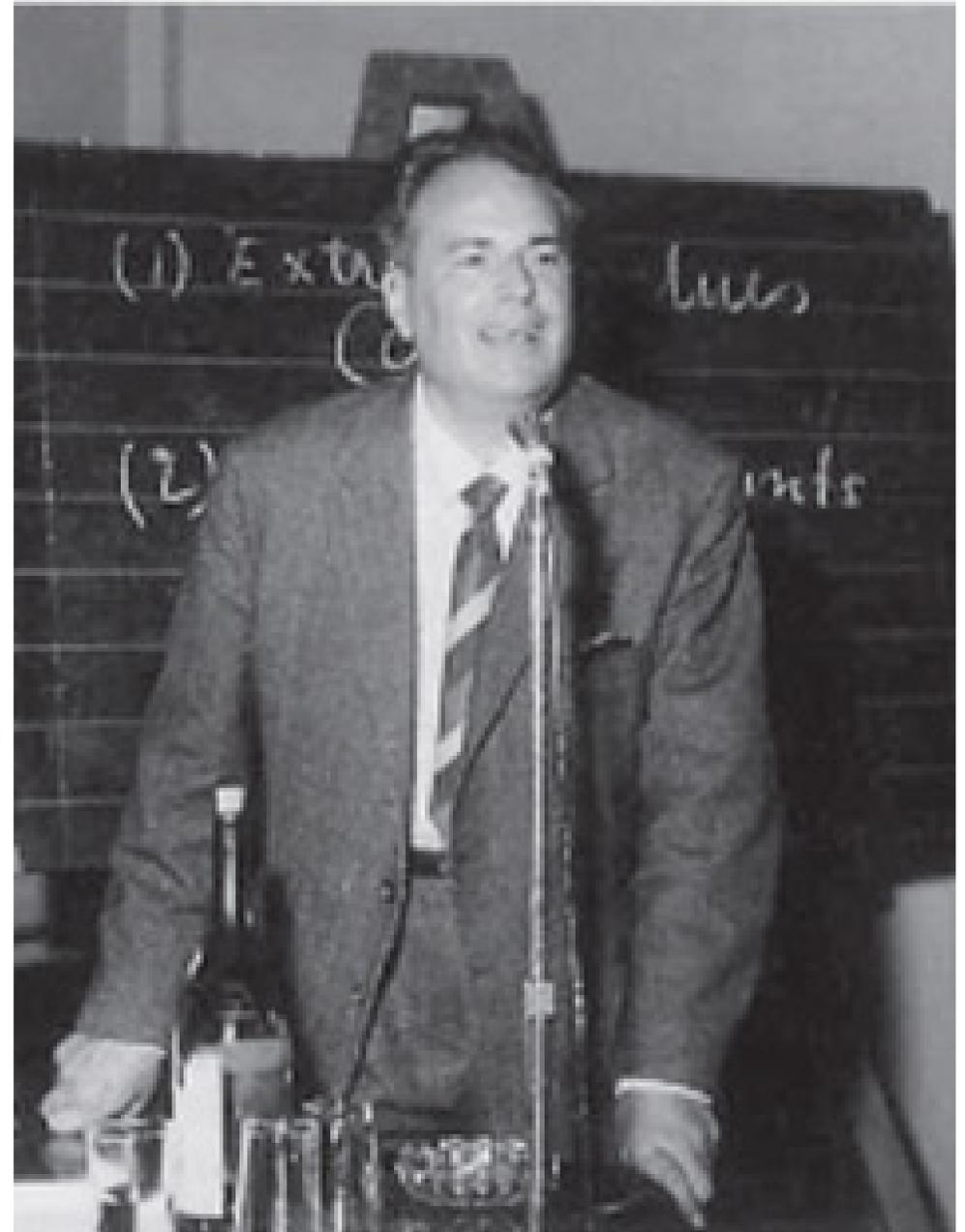
$c > 0$ (üblich): risikoavers, eventuell aber auch $c < 0$: risikofreundlich (z.B. Glücksspiel).

2.4.2 Grundlagen der Bayesianischen Ansätze

Bem. 2.63 (Erstes Bayesianisches / Subjektivistisches Paradigma)

Jede Situation unter Unsicherheit kann durch einen Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \pi(\cdot))$ beschrieben werden.

Im Kontext eines Entscheidungsproblems heißt die Wahrscheinlichkeitsverteilung $\pi(\cdot)$ über $(\Theta, \sigma(\Theta))$ (*a*) *Priori-Verteilung* (Verteilung „vor den Daten“) (zu $(\mathbb{A}, \Theta, u(\cdot))$).



Quelle: <http://www.sipta.org/isipta11/index.php?id=finetti>

2.4.3 Bayes-Aktionen: Definition, Bestimmung und wichtige Eigenschaften

Def. 2.64 (Bayes-Aktion)

Gegeben sei ein Entscheidungsproblem $(\mathbb{A}, \Theta, u(\cdot))$.

a) Ist $\pi(\cdot)$ eine Priori-Verteilung zu $(\mathbb{A}, \Theta, u(\cdot))$, so heißt

$$\Phi(\cdot, \pi) : \mathbb{A} \rightarrow \mathbb{R}$$

$$a \longmapsto \Phi(a, \pi) = \mathbb{E}_\pi(u(a)) = \int_{\Theta} u(a, \vartheta) d\pi(\vartheta)$$

Bayes-Kriterium zu $\pi(\cdot)$ und jede bezüglich dieses Kriteriums optimale Aktion a^* *Bayes-Aktion bezüglich $\pi(\cdot)$* .

$$B(\pi) := \Phi(a^*, \pi)$$

heißt *Bayes Nutzen bezüglich $\pi(\cdot)$*

b) Eine Aktion a^{**} heißt *Bayes-Aktion (schlechthin)*, wenn es eine Priori-Bewertung $\pi^{**}(\cdot)$ zu $(\mathbb{A}, \Theta, u(\cdot))$ gibt, so dass a^{**} Bayes Aktion bezüglich $\pi^{**}(\cdot)$ ist.

Bem. 2.65

Beachte: „Rein mathematisch“ besteht kein Unterschied zum Bernoullikriterium; semantisch handelt es sich aber bei der Wahrscheinlichkeitsbewertung der Zustände um ein anderes Objekt, nämlich um eine subjektive Wahrscheinlichkeitsbewertung.

Bem. 2.66

- Es ist einfach, Beispiele zu finden, bei denen alle Aktionen eines Entscheidungsproblems Bayes-Aktionen sein können. Vorwurf der Beliebigkeit.
- Hinweis auf das sog. Bayes-Umkehrproblem (Bem. 2.73): Zu welcher Aktion gibt es eine Priori-Bewertung, bezüglich derer die jeweilige Aktion Bayes-Aktion ist?

Bem. 2.68 (Der Bayes-Nutzen randomisierter Aktionen)

Für randomisierte Aktionen \tilde{a} ist unter schwachen Regularitätsbedingungen (Anwendbarkeit des Satzes von Fubini)

$$\begin{aligned} \Phi(\tilde{a}, \pi) = \mathbb{E}_\pi(u(\tilde{a})) &= \int_{\Theta} u(\tilde{a}, \vartheta) d\pi(\vartheta) \\ &= \int_{\Theta} \left(\int_{\mathbb{A}_0} u(a, \vartheta) d\tilde{a}(a) \right) d\pi(\vartheta) \end{aligned} \quad (2.33)$$

$$\begin{aligned} &= \int_{\mathbb{A}_0} \left(\int_{\Theta} u(a, \vartheta) d\pi(\vartheta) \right) d\tilde{a}(a) \\ &= \int_{\mathbb{A}_0} \Phi(a, \pi) d\tilde{a}(a) \end{aligned} \quad (2.34)$$

also gilt etwa bei endlichen \mathbb{A}_0 und Θ :

Hat $\tilde{a}(\cdot)$ die Wahrscheinlichkeitsfunktion $(\tilde{p}(a_i))_{i=1,\dots,n}$ auf \mathbb{A}_0 und $\pi(\cdot)$ die Wahrscheinlichkeitsfunktion $(\pi(\{\vartheta_j\}))_{j=1,\dots,m}$ auf Θ , so gilt:

$$\begin{aligned}
 \Phi(\tilde{a}, \pi) &= \sum_{j=1}^m u(\tilde{a}, \vartheta_j) \cdot \pi(\{\vartheta_j\}) \\
 &= \sum_{j=1}^m \left(\sum_{i=1}^n u(a_i, \vartheta_j) \cdot p(\{a_i\}) \right) \pi(\{\vartheta_j\}) \\
 &= \sum_{i=1}^n \left(\sum_{j=1}^m u(a_i, \vartheta_j) \cdot \pi(\{\vartheta_j\}) \right) p(\{a_i\}) \\
 &= \sum_{i=1}^n \Phi(a_i, \pi) \cdot p(\{a_i\})
 \end{aligned}$$

Proposition 2.70 (Berechnung von Bayes-Aktionen über lineare Optimierung bei endlichem Θ und endlichem \mathbb{A})

Bem. 2.73 Bayes-Umkehrproblem bei endlichem Θ und endlichem \mathbb{A}

Es soll die Menge Π_a aller Priori-Verteilungen angegeben werden, für die eine *gegebene Aktion* $a \in \mathbb{A}$ Bayes-Aktion in \mathbb{A} ist.

2.4.4 Wichtige Sätze über Bayes-Aktionen

Im folgenden $|\Theta| < \infty$, also endliche Zustandsmengen.

Viele Sätze gelten unter geeigneten Zusatzvoraussetzungen *mutatis mutandis* auch für unendliche Zustandsräume.

(\mathbb{A} nicht notwendig als endlich vorausgesetzt, da man ja auch $\mathcal{M}(\mathbb{A})$ betrachten will.)

a) Entbehrlichkeit randomisierter Aktionen

Satz 2.74 (Entbehrlichkeit randomisierter Aktionen)

Gegeben sei ein Entscheidungsproblem $(\mathcal{M}(\mathbb{A}_0), \Theta, \tilde{u}(\cdot))$, das die gemischte Erweiterung eines Entscheidungsproblems $(\mathbb{A}_0, \Theta, u(\cdot))$ mit endlicher Aktionenmenge \mathbb{A}_0 darstellt. Dann gilt für jede Priori-Verteilung $\pi(\cdot)$:

Ist $\tilde{a} = \begin{pmatrix} a_1 & \cdots & a_n \\ p_1 & \cdots & p_n \end{pmatrix}$ Bayes-Aktion zur Bewertung $\pi(\cdot)$, so gibt es auch eine reine Aktion $a \in \mathbb{A}_0$, die Bayes-Aktion zu $\pi(\cdot)$ in $\mathcal{M}(\mathbb{A}_0)$ ist.

Korollar 2.75

In der Situation von Satz 2.74 gilt für jede Priori-Verteilung $\pi(\cdot)$:
Bayes-Aktionen zu $\pi(\cdot)$ in \mathbb{A}_0 bleiben Bayes-Aktionen zu $\pi(\cdot)$ beim
Übergang zu $\mathcal{M}(\mathbb{A}_0)$.

Beweis des Korollars:

Angenommen a_0^* sei eine Bayes-Aktion zu $\pi(\cdot)$ zu $\pi(\cdot)$ in \mathbb{A}_0 , aber nicht
Bayes-Aktion zu $\pi(\cdot)$ in $\mathcal{M}(\mathbb{A}_0)$. Dann gibt es ein $\tilde{a}^* \in \mathcal{M}(\mathbb{A}_0)$ mit

$$\Phi(\tilde{a}^*, \pi) > \Phi(a_0^*, \pi). \quad (2.35)$$

Wegen Satz 2.74 gibt es dann eine reine Aktion $a^* \in \mathbb{A}_0$, so dass auch a^*
Bayes-Aktion in $\mathcal{M}(\mathbb{A}_0)$ ist.

D.h. $\Phi(\tilde{a}^*, \pi) = \Phi(a^*, \pi)$. Wegen (2.35) ist damit

$$\Phi(a^*, \pi) > \Phi(a_0^*, \pi).$$

Beachte, dass $a^* \in \mathbb{A}_0$; also ist a_0^* nicht Bayes-Aktion zu $\pi(\cdot)$ in \mathbb{A}_0 .
Widerspruch!

(Die Menge der reinen Bayes-Aktionen bildet so etwas wie eine wesentlich minimal vollständige Klasse bezüglich der durch $\Phi(\cdot, \pi)$ induzierten Ordnung)

Interpretation:

Bem. 2.76 (Zum Beweis von Satz 2.51)

(Beweis hier auf zwei Arten, inklusive einer **geometrischen Veranschaulichung**)

b) Bayes-Aktionen und Zulässigkeit

Satz 2.77 Zulässigkeit von Bayes-Aktionen zu nicht entarteten Priori-Verteilungen

Gegeben sei ein Entscheidungsproblem $(\mathbb{A}, \Theta, u(\cdot))$ mit Priori-Bewertung $\pi(\cdot)$ so, dass $\pi(\{\vartheta_j\}) > 0$ für alle j .

Dann gilt: Jede Bayes-Aktion a^* zu $\pi(\cdot)$ ist zulässig.

Interpretation:

Proposition 2.78

Gegeben sei ein Entscheidungsproblem $(\mathbb{A}, \Theta, u(\cdot))$. Ist $a \in \mathbb{A}$ eine zulässige Bayes-Aktion, so ist a auch zulässig im Problem $(\mathcal{M}(\mathbb{A}), \Theta, \tilde{u}(\cdot))$.

Neu ist hier die Aussage für $\pi(\{\vartheta_j\}) = 0$ für manche j . Für $\pi(\{\vartheta_j\}) > 0$ folgt die Aussage bereits aus Korollar 2.75 und Satz 2.77; nach dem Korollar ist a auch Bayes-Aktion in der gemischten Erweiterung, auf die dann Satz 2.77 angewendet werden kann.

Sind umgekehrt zulässige Aktionen auch Bayes-Aktionen?

Satz 2.79

Betrachtet werde ein Entscheidungsproblem $(\mathbb{A}, \Theta, u(\cdot))$. Ist \mathbb{A} konvex, z.B. $\mathbb{A} = \mathcal{M}(\mathbb{A}_0)$ für eine geeignete Aktionenmenge \mathbb{A}_0 , dann gilt: Zu jeder zulässigen Aktion a gibt es eine Priori Verteilung $\pi^a(\cdot)$ mit $\pi^a(\{\vartheta_j\}) > 0$, $j = 1, \dots, m$, so dass a Bayes-Aktion zur Bewertung $\pi^a(\cdot)$ ist.

Interpretation:

2.4.5 Bayes und Minimax

Def. 2.80 (Ungünstigste Priori-Verteilung)

Gegeben sei ein Entscheidungsproblem $(\mathbb{A}, \Theta, u(\cdot))$. Eine Verteilung $\pi^-(\cdot)$ über $(\Theta, \sigma(\Theta))$ heißt *ungünstigste Priori-Verteilung* für $(\mathbb{A}, \Theta, u(\cdot))$, falls für alle anderen Priori-Verteilungen $\pi(\cdot)$ aus der Menge Π aller Verteilungen über $(\Theta, \sigma(\Theta))$ und für die Bayes-Aktion a^- zu $\pi^-(\cdot)$ gilt:

$$\mathbb{E}_{\pi^-} u(a^-) \leq \mathbb{E}_{\pi} u(a^-)$$

Proposition 2.81

Unter sehr schwachen Regularitätsbedingungen gilt in obiger Situation:
 a^- ist Maximin-Aktion.

In Proposition ?? wurde ein Kriterium dafür angegeben, wann eine Äquilibrium-Aktion auch Maximin-(Minimax-)Aktion ist.

Bayes-Aktionen liefern ein *stärkeres* Kriterium, das auch *praktisch wichtiger* ist, weil die Feststellung, Bayes-Aktion zu sein, oft leichter zu treffen ist als die Feststellung, zulässige Aktion zu sein.

Proposition 2.82

Eine Äquilibrium-Aktion, die zugleich Bayes-Aktion ist, ist auch Maximin-Aktion.

Beweis:

Ist $\pi(\cdot) > 0$, so ist jede Bayes-Aktion zulässig, und der Satz ergibt sich aus Satz 2.77.

Sei nun allgemeiner $\pi(\cdot)$ beliebig und a_B :

- 1) Bayes-Aktion
- 2) equalizer-Aktion

Dann gibt es

1) eine Bewertung π , sodass

$$\Phi(a_B, \pi) \geq \Phi(a, \pi) \quad \text{für alle } a \in \mathbb{A} \quad (2.36)$$

2) eine Zahl c mit

$$u(a_B, \vartheta_j) = c \quad \text{für alle } j = 1, \dots, m \quad (2.37)$$

Angenommen, a_B wäre *nicht* eine Maximin-Aktion. Dann gäbe es eine Aktion a_M mit

$$\inf_{\ominus} u(a_M, \vartheta_j) > c. \quad (2.38)$$

Dann würde folgen

$$\begin{aligned} \Phi(a_M, \pi) &= \mathbb{E}_\pi u(a_M) = \int \underbrace{u(a_M, \vartheta)}_{>c \quad \forall \vartheta (\text{vgl. 2.38})} d\pi \\ &> \int c d\pi \stackrel{2.37}{=} \int u(a_B, \vartheta) d\pi = \mathbb{E}_\pi u(a_B) = \Phi(a_B, \pi) \end{aligned}$$

also

$$\Phi(a_M, \pi) > \Phi(a_B, \pi), \quad (2.39)$$

und a_B wäre keine Bayes-Aktion.
Widerspruch.

2.4.6 Konditionale Bayes-Inferenz: Begrifflicher Hintergrund und „Erinnerungen“

In Kapitel 1.5.4 wurde davon gesprochen, dass die Lösung/ Darstellung des datengestützten Entscheidungsproblems über das Auswertungsproblem und die damit verbunden Suche nach optimalen Entscheidungsfunktionen durchaus auch auf Kritik stösst. Diese stützt sich v.a. einerseits

- auf die immense computationale Komplexität

und andererseits ganz prinzipiell

- auf die Problematik kontrafaktischer Ereignisse bei der Bewertung von Entscheidungsfunktionen mittels der Risikofunktion.

Dies legt eine konditionale Sicht als mögliche Alternative nahe: Es werden optimale Lösungen für die konkret beobachtete Datenkonstellation $\{x\}$ gesucht („auf $\{x\}$ konditionierte Betrachtung“). Dies führt auf die „übliche Bayes-Inferenz“, die mit Hilfe der sogenannten Posteriori-Verteilung gegeben die Daten arbeitet, und in diesem Sinne hier zur Abgrenzung als „*konditionale Bayes-Inferenz*“ bezeichnet werde. Zur Vorbereitung werde an das Theorem von Bayes in seiner allgemeinen Form erinnert:

Proposition 2.83 (Allgemeines Theorem von Bayes)

Seien X und U zwei Zufallsvariablen mit gemeinsamer Wahrscheinlichkeitsfunktion $f_{X,U}(\cdot)$ bzw. Dichte $f_{X,U}(\cdot)$ (bezüglich eines dominierenden σ -finiten Maßes $\nu \otimes \lambda$) und den bedingten Wahrscheinlichkeitsfunktionen bzw. bedingten Dichten $f_{X|U}(\cdot|u)$ und $f_{U|X}(\cdot|x)$ (bezüglich ν bzw. λ).

Dann gilt:

$$f_{U|X}(u|x) = \frac{f_{X|U}(x|u) \cdot f_U(u)}{f_X(x)} \quad (2.41)$$

mit

$$f_X(x) = \int f_{X|U}(x|u) \cdot f_U(u) d\nu(u). \quad (2.42)$$

Bem. 2.84

Bei stetigem U mit Dichte $f_U(u)$ erhält man also Proposition 2.83 mit

$$f_X(x) = \int f_{X|U}(x|u) \cdot f_U(u) du. \quad (2.43)$$

Im Fall von diskreten Zufallsvariablen X und U – mit \mathcal{U} als Träger von U – ergibt sich

$$p(\{U = u\}|\{X = x\}) = \frac{p(\{X = x\}|\{U = u\}) \cdot p(\{U = u\})}{p(\{X = x\})} \quad (2.44)$$

mit

$$p(\{X = x\}) = \sum_{u \in \mathcal{U}} p(\{X = x\}|\{U = u\}) \cdot p(\{U = u\}) \quad (2.45)$$

Betrachtet werde im Folgenden immer einer dieser beiden Spezialfälle .
Die allgemeinere Formulierung über beliebige Dichten bezüglich geeigneter dominierender Maße ist unproblematisch.

Bem. 2.85 (Normierungskonstante)

$f_X(x)$ aus (2.42) spielt die Rolle einer reinen Normierungskonstante, die nicht von u abhängt. Häufig reicht es daher, $f_{X|U}(x|u) \cdot f_U(u)$ zu berechnen. Da man weiß, dass sich insgesamt eine Wahrscheinlichkeitsdichte ergeben muss, kennt man implizit auch die Normierungskonstante.

Man schreibt dann mit \propto als Symbol für „proportional zu“

$$f_{X|U}(x|u) \propto f_{X|U}(u|x) \cdot f_U(u).$$

Bem. 2.86 (Konditionale Bayes-Inferenz: Konzeptionelle Hintergr

Gegeben sei ein datengestütztes Entscheidungsproblem $((\mathbb{A}, \Theta, l(\cdot)); (\mathcal{X}, \sigma(\mathcal{X}), (p_{\vartheta}(\cdot))_{\vartheta \in \Theta})$, wobei $p_{\vartheta}(\cdot)$ die Wahrscheinlichkeitsfunktion bzw. Dichte $f_{\vartheta}(\cdot)$ besitze, und eine Priori-Verteilung auf einem geeigneten Maßraum $(\Theta, \sigma(\Theta))$ mit Dichte bzw. Wahrscheinlichkeitsfunktion $\pi(\cdot)$. Sei, als gedankliche Hilfskonstruktion, U („Umwelt“, „Natur“) eine „Zufallsgrösse“ (Zufallsvariable/-element), die das Eintreten des Umweltzustands ϑ beschreibt und X eine Zufallsgrösse, die den Ausgang des Informationbeschaffungsexperiments beschreibt. Dann gelte für die Dichte bzw. Wahrscheinlichkeitsfunktion $f_{X,U}(x, \vartheta)$ der gemeinsamen Verteilung von X und U für alle $x \in \mathcal{X}$ und $\vartheta \in \Theta$:

$$f_{X,U}(x, \vartheta) = \pi(\vartheta) \cdot f_{\vartheta}(x)$$

Das heißt für alle $x \in \mathcal{X}$ und $\vartheta \in \Theta$ gilt

$$f_{\vartheta}(x) = f_{X|U}(x|\vartheta) ;$$

die Verteilung der Zufallsgrösse aus der Informationsstruktur wird als bedingte Verteilung von X gegeben U interpretiert (!!).

Dann ergibt sich (!!!) für jedes x aus dem Satz von Bayes gemäß Proposition 2.83 für (eine Version der bzw.) die Dichte bzw. Wahrscheinlichkeitsfunktion $\pi(\vartheta|x)$ der bedingten Verteilung von U gegeben $X = x$:

$$\pi(\vartheta|x) = c(x) \cdot f_{\vartheta}(x) \cdot \pi(\vartheta) \quad (2.46)$$

mit

$$\frac{1}{c(x)} = f_X(x) = \int f_{X|U}(x|\vartheta)\pi(\vartheta)d\vartheta \quad (2.47)$$

im Falle von stetigem X und U , und bei diskretem X und U

$$\begin{aligned} \frac{1}{c(x)} = p(\{X = x\}) &= \sum_{j=1}^m p(\{X = x\}|\{U = \vartheta_j\}) \cdot \pi(\{U = \vartheta_j\}) \quad (2.48) \\ &= \sum_{j=1}^m p(\{X = x\}|\{U = \vartheta_j\}) \cdot \pi(\vartheta_j). \end{aligned}$$

Für jede Beobachtung $\in \mathcal{X}$ wird die bedingte Verteilung von U gegeben $X = x$ als *Posteriori-Verteilung des Parameters ϑ nach der Beobachtung x* bezeichnet. Die zugehörige Dichte bzw. Wahrscheinlichkeitsfunktion $\pi(\vartheta|x)$ heißt *Posteriori-Dichte nach der Beobachtung*, und $f_{\vartheta}(x)$ heißt *Likelihood*.

Die marginale Verteilung von X mit Dichte $f_X(x)$ aus (2.47) bzw. Wahrscheinlichkeitsfunktion $(p(\{X = x\}))_{x \in \mathcal{X}}$ aus (2.48) heißt *Priori-Prädiktive-Verteilung*.

Die Größen $f_X(x)$ und $p(\{X = x\})$ sind nicht zu verwechseln mit den als bedingte Verteilungen interpretierten $f_{\vartheta}(x)$ und $p_{\vartheta}(\{X = x\})$.

Analog gibt es auch eine *posteriori-prädiktive Verteilung*, wenn man in analoger Weise über die Posteriori-Verteilung herausintegriert bzw. -summiert. Dies ist dann die Wahrscheinlichkeitsverteilung der nächsten Beobachtung, basierend auf dem aktuellen Wissensstand.

Bem. 2.87 (Bayes-Postulat (nicht entscheidungstheoretisch))

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

Alle statistischen Analysen mögen sich ausschließlich auf die Posteriori zu stützen; darauf baut insbesondere auch die Konstruktion von

- Bayesschen-Punktschätzungen: *MPD-Schätzer (Maximum Posteriori Density-Schätzer)*
- Bayesschen-Intervallschätzungen: *HPD-Intervalle (Highest posterior density-Intervalle)*

- Bayes-Tests
 - Suffiziente Statistik: Enthält volle Information der Stichprobe über den Parameter (allgemeine statistische Theorie). Jetzt zwei verschiedene Arten von „enthält volle Information“; Wie passt Suffizienz in diesen Zusammenhang?
 - Posteriori-Verteilung enthält volle Information (vgl. Bem 2.87).
- ⇒ Posteriori hängt tatsächlich nur von suffizienter Statistik ab.

Proposition 2.88 (Suffizienz und Posteriori-Verteilung)

Ist in der Situation von Bemerkung 2.86 T eine für ϑ suffiziente Statistik mit Wahrscheinlichkeitsfunktion bzw. Dichte $g_{\vartheta}(\cdot)$, so hängt die Posteriori $\pi(\vartheta|x)$ nur mehr über $t = T(x)$ von x ab. Es gilt ⁶

$$\pi(\vartheta|x) \propto g_{\vartheta}(t) \cdot \pi(\vartheta)$$

Beweis:

Gemäß (2.46) ist

$$\pi(\vartheta|x) \propto f_{\vartheta}(x) \cdot \pi(\vartheta)$$

wobei wegen der Suffizienz von T sich $f_{\vartheta}(x)$ schreiben lässt als $f_{\vartheta}(x) = h_{X|T}(x) \cdot g_{\vartheta}(t)$. Einsetzen liefert die Behauptung.

⁶ \propto „proportional zu“, vgl. Bemerkung 2.85

Def. 2.89 (Vorbereitende Erinnerung: Exponentialfamilien)

Sei $(\mathcal{X}, \sigma(\mathcal{X}), (p_{\vartheta})_{\vartheta \in \Theta})$ ein statistisches Modell mit $\Theta \subseteq \mathbb{R}^q$.

- $(p_{\vartheta})_{\vartheta \in \Theta}$ bildet eine (oder p_{ϑ} ist für jedes $\vartheta \in \Theta$ ein Mitglied der) q -parametrische(n) *Exponentialfamilie* in (T_1, \dots, T_q) mit *natürlichem Parameter* $(c_1(\vartheta), \dots, c_q(\vartheta))$, wenn sich die Dichte $f_{\vartheta}(x)$ bezüglich eines dominierenden σ -finiten Maßes (also insbesondere Dichte/Wahrscheinlichkeitsfunktion) in die folgende Form bringen läßt: Mit $t_1 := T_1(\vec{x}), \dots, t_q := T_q(\vec{x})$ ist

$$f_{\vartheta}(x) = h(x) \cdot g(\vartheta) \cdot \exp\left(\sum_{\ell=1}^q c_{\ell}(\vartheta)t_{\ell}\right).$$

- Enthält Θ echt innere Punkte und sind $1, c_1(\vartheta), c_2(\vartheta), \dots, c_q(\vartheta)$ und $1, T_1(x), T_2(x), \dots, T_q(x)$ (f.-s.) jeweils linear unabhängig, so spricht man von einer *strikt* q -parametrischen Exponentialfamilie. (Der „natürliche Parameterraum“ hat wirklich die Dimension q .)

2.4.7 Konjugierte Verteilungen, Bayes-Lernen

Wir betrachten hier, unter Rückbezug auf das Bayes-Postulat (vgl. Bem. 2.87), das sukzessive Lernen aus Stichproben.

a) Ein Motivationsbeispiel

Bsp. 2.90 (Beta-Binomialmodell)

Absolutes Standardbeispiel

- Stichprobenmodell: Bernoulliverteilung (allgemein: Binomialverteilung)
zu Parameter ϑ

$$p_{\vartheta}(\{X_i = x_i\}) = \vartheta^{x_i}(1 - \vartheta)^{1-x_i}$$

jetzt im Bayes Kontext als bedingte Verteilung schreiben (wieder mit „Hilfsvariable“ U):

$$p(\{X_i = x_i\} \mid \{U = \vartheta\}) = \vartheta^{x_i}(1 - \vartheta)^{1-x_i}$$

- gebräuchliche Priori-Verteilung:

Betaverteilung, gilt als sehr flexibel, zwei Parameter $a > 0$, $b > 0$
hier als Priori verwendet, Bezeichnung $\pi(\cdot)$

$$\pi(\vartheta) = \frac{\vartheta^{a-1}(1-\vartheta)^{b-1}}{B(a,b)} \cdot I_{[0;1]}(\vartheta) \quad (2.49)$$

$B(a,b)$ ist eine reine Normierungskonstante.

Es gilt:

$$\text{Erwartungswert: } \frac{a}{a+b} \quad \text{Modus: } \frac{a-1}{a+b-2}, \quad a > 1, b > 1$$

Abbildung 1: Ruger, (1999) Test- und Schatztheorie I, Seite 193

2.4. BAYES-INFERENZ

193

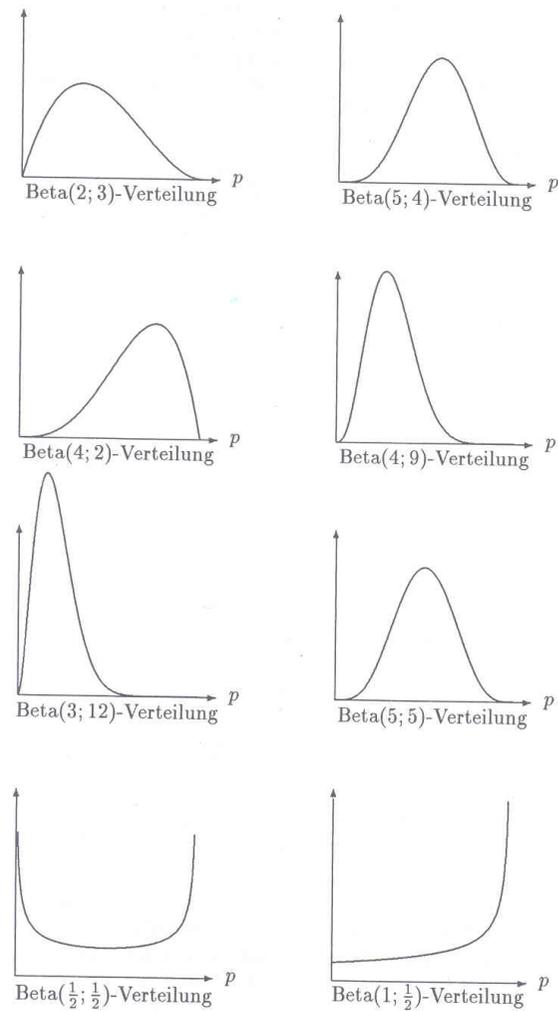


Abbildung 2.17: Einige Beta($a; b$)-Verteilungen.
 Die Beta($1; 1$)-Verteilung ist die Gleichverteilung. Die an der Senkrechten im Punkt 0.5 gespiegelte Dichte einer Beta($a; b$)-Verteilung ist die Beta($b; a$)-Verteilung.

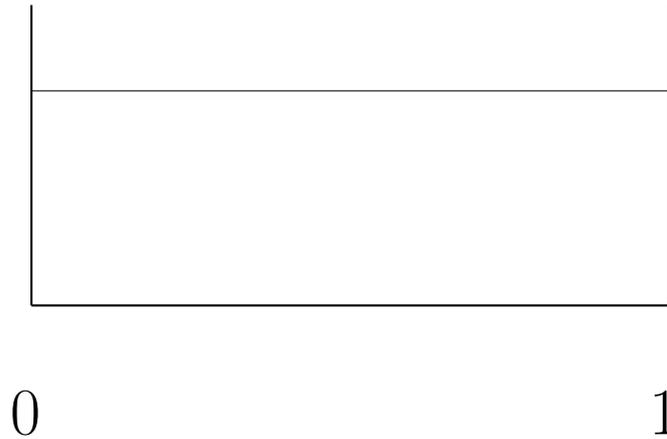
Jetzt Satz von Bayes anwenden: Posteriori nach einer Beobachtung berechnen.

$$\begin{aligned}
 \pi(\vartheta|x_i) &= \frac{\vartheta^{x_i}(1-\vartheta)^{1-x_i} \cdot \vartheta^{a-1}(1-\vartheta)^{b-1}}{\underbrace{\text{Norm.} \cdot B(a,b)}_{\text{Normierung}}} \cdot I_{[0;1]}(\vartheta) \\
 &\propto \vartheta^{x_i+a-1} \cdot (1-\vartheta)^{b-x_i} \cdot I_{[0;1]}(\vartheta) \\
 &= \vartheta^{a'-1} \cdot (1-\vartheta)^{b'-1} \cdot I_{[0;1]}(\vartheta)
 \end{aligned}$$

- Posteriori ist also wieder eine Betaverteilung, nun mit den Parametern

$$a' = a + x_i \quad \text{und} \quad b' = b - x_i + 1 = b + (1 - x_i).$$

Start z.B. mit $a^{(0)} = 1, b^{(0)} = 1$:



Gleichverteilung (als Nichtwissen verkaufbar?)

$x_1 = 1$ beobachtet $\Rightarrow a^{(1)} = a^{(0)} + 1 = 2, b^{(1)} = b^{(0)} + 0 = 1$

Beta(2, 1)-Verteilung

$$\pi(\vartheta \mid x_1) \propto \vartheta I_{[0;1]}(\vartheta)$$

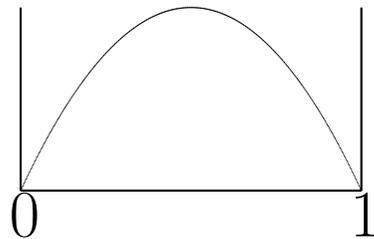
- Jetzt weiteres Experiment:

Updating-Prinzip: (neue) Priori = (alte Posteriori): Beta(2,1) Verteilung
neue Stichprobe x_2

neue Posteriori: Beta($a^{(2)}, b^{(2)}$) = Beta($a^{(1)} + x_i, b^{(1)} + (1 - x_i)$)

z.B. $x_2 = 0$ $\rightarrow a^{(2)} = 2 + 0 = 2, b^{(2)} = 1 + 1 - 0 = 2$

$$\pi_2(\vartheta | (x_1, x_2)') = \frac{\vartheta^{2-1} \cdot (1 - \vartheta)^1}{\text{Norm.}} = \frac{\vartheta^1 \cdot (1 - \vartheta)^1}{\text{Norm.}} = \frac{\vartheta - \vartheta^2}{\text{Norm.}} \quad \text{für } \vartheta \in [0; 1]$$



$$\pi_2(0|x) = \pi_2(1|x) = 0$$

- Weitere Beobachtung $x_3 = 1$:

neue Posteriori: $\text{Beta}(a^{(3)}, b^{(3)}) = \text{Beta}(a^{(2)} + x_i, b^{(2)} + (1 - x_i))$

$$a^{(3)} = 2 + 1, \quad b^{(3)} = 2$$

$$\pi_3(\vartheta | (x_1, x_2, x_3)') \propto \vartheta^2 (1 - \vartheta)^1 I_{[0;1]}(\vartheta) = \vartheta^2 - \vartheta^3 I_{[0;1]}(\vartheta)$$

- Weiter zusätzliche Beobachtung $x_4 = 1$:
neue Posteriori: $\text{Beta}(a^{(4)}, b^{(4)}) = \text{Beta}(a^{(3)} + x_i, b^{(3)} + (1 - x_i))$
 $a^{(4)} = 3 + 1, b^{(4)} = 2$

$$\pi_4(\vartheta | (x_1, x_2, x_3, x_4)') \propto \vartheta^3 (1 - \vartheta)^1$$

- Allgemein gilt bei n unabhängigen Wiederholungen:

Die Posteriori $\pi_n(\vartheta | (x_1, \dots, x_n)')$ ist eine

$$B \left(a^{(0)} + \sum_{i=1}^n x_i; b^{(0)} + n - \sum_{i=1}^n x_i \right) \text{ Verteilung.} \quad (2.50)$$

Man kann zeigen: Dasselbe Ergebnis erhält man, wenn man x_1, \dots, x_n auf einmal verarbeitet.

- In diesem Beispiel gilt für die posteriori-prädiktive Verteilung

$$p(X_{n+1} = 1 | X_1 = x_1, \dots, X_n = x_n)$$

$$= \mathbb{E}(\pi_n(\vartheta | x_1, \dots, x_n))$$

$$= \frac{a + \sum_{i=1}^n x_i}{a + b + n}.$$

Für die Gleichverteilung (vgl. oben) als Ausgangspriori ergibt sich wegen $a^{(0)} = b^{(0)} = 1$

$$\begin{aligned} & p(X_{n+1} = 1 | X_n = x_n, \dots, X_1 = x_1) \\ &= \frac{\left(\sum_{i=1}^n x_i \right) + 1}{n + 2}; \end{aligned}$$

b) Konjugiertheit: Definition und klassische Ergebnisse (vgl. „Schätzen und Testen I“)

Def. 2.91 (Konjugiertheit)

Eine Verteilungsfamilie Π von Priori-Verteilungen heißt zu einer Menge \mathcal{P} von Stichprobenverteilungen *konjugiert*, wenn für jede Priori-Verteilung $\pi(\cdot) \in \Pi$ und jedes $p(\cdot) \in \mathcal{P}$ die zugehörige Posteriori-Verteilung wieder ein Element von Π ist. Man sagt dann auch, dass jedes Element $\pi(\cdot) \in \Pi$ zu \mathcal{P} konjugiert ist.

Proposition 2.92 (Beispiele für Konjugiertheit: Beta-Binomial/Dirichlet-Multinomial-Modell/Gamma-Poisson-Modell, Selbstkonjugiertheit der Normalverteilung)

a) Die Menge der Betaverteilungen als Priori ist zur Menge der Bernoulliverteilungen konjugiert (vgl. Bsp. 2.90).

Allgemeiner gilt:

Ist $\vec{X} = (X_1, \dots, X_k)$ eine Stichprobe eines zum Parameter $\vec{\vartheta} = (\vartheta_1, \dots, \vartheta_k)$ multinomial-verteilten Untersuchungsmerkmals, besitzt \vec{X} also die Wahrscheinlichkeitsfunktion

$$f(\vec{x}|\vec{\vartheta}) \propto \prod_{j=1}^k \vartheta_j^{x_j}$$

und wählt man die sog. *Dirichlet-Verteilung* zum Parameter $\vec{\alpha} = (\alpha_1, \dots, \alpha_k)^T$

$$\pi(\vec{\vartheta}) = \prod_{j=1}^k \vartheta_j^{(\alpha_j-1)},$$

so ist die Posteriori-Verteilung eine Dirichlet-Verteilung mit dem Parameter $\alpha' = (\alpha'_1, \dots, \alpha'_k)^T$, wobei

$$\alpha'_j = \alpha_j + x_j - 1, \quad j = 1, \dots, k.$$

b) Ist $\vec{X} = (X_1, \dots, X_n)$ eine i.i.d. Stichprobe eines zum Parameter λ Poisson verteilten Untersuchungsmerkmals, besitzt \vec{X} also die Wahrscheinlichkeitsfunktion

$$f(\vec{x}|\lambda) = \frac{\lambda^{\sum_{i=1}^n x_i}}{x_1! x_2! \dots x_n!} e^{-n\lambda},$$

und wählt man als Priori-Verteilung eine Gamma-Verteilung mit Parametern a und b , d.h. eine Verteilung mit der Dichte

$$\pi(\lambda) = \frac{b^a}{\underbrace{\Gamma(a)}_{\text{Norm.konst.}}} \lambda^{a-1} e^{-b\lambda}, \quad (2.51)$$

so ist die Posteriori-Verteilung eine Gamma-Verteilung mit den

Parametern

$$a + \sum_{i=1}^n x_i \quad \text{und} \quad b + n.$$

Bsp. 2.93 (Normalverteilung)

Ist $\vec{X} = (X_1, \dots, X_n)$ eine i.i.d. Stichprobe eines mit den Parametern μ und σ^2 normalverteilten Untersuchungsmerkmals, so gilt:

- (i) Ist σ^2 bekannt und wählt man als Priori-Verteilung für μ eine Normalverteilung mit den Parametern ν und ρ^2 , so ist die a posteriori Verteilung $\pi(\mu|\vec{x})$ eine Normalverteilung mit den Parametern ν' und ρ'^2 mit

$$\nu' = \frac{\bar{x}\rho^2 + \nu\frac{\sigma^2}{n}}{\rho^2 + \frac{\sigma^2}{n}} \quad (2.52)$$

und

$$\rho^{2'} = \frac{\rho^2 \cdot \frac{\sigma^2}{n}}{\rho^2 + \frac{\sigma^2}{n}}. \quad (2.53)$$

- (ii) Ist μ bekannt, aber σ^2 unbekannt, so erhält man die konjugierte Verteilung, indem man $\frac{1}{\sigma^2}$ als gammaverteilt annimmt. Man sagt dann, σ^2 sei *invers gammaverteilt*.

Wie findet man solche konjugierten Paare?

Satz 2.94 (Zur Konjugiertheit in Exponentialfamilien)

Hat in der Situation von Def. 2.86 jedes Element der Menge \mathcal{P} der Stichprobenverteilungen eine Dichte bzw. Wahrscheinlichkeitsfunktion $f(x|\vartheta)$ der Form

$$f(x|\vartheta) \propto h(\vartheta) \exp(T(x) \cdot b(\vartheta)) \quad (2.54)$$

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

$$\pi(\vartheta) \propto [h(\vartheta)]^\alpha \exp(b(\vartheta) \cdot \beta), \quad (2.55)$$

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

$$\pi(\vartheta|x) \propto [h(\vartheta)]^{\alpha+1} \cdot \exp((T(x) + \beta) \cdot b(\vartheta)). \quad (2.56)$$

Beweis:

(2.56) ergibt sich unmittelbar durch Anwenden der Formel für die Posteriori-Verteilung auf (2.54) und (2.55). Dann ist (2.56) mit $\alpha' := \alpha + 1$ und $\beta' := \beta + T(x)$ von der Form (2.55), also sind tatsächlich Π und \mathcal{P} konjugiert.

Bem. 2.95 (zu Satz 2.94)

- Der Satz kann also direkt zur Konstruktion geeigneter, konjugierter Priori-Verteilungen verwendet werden, indem man die Stichprobenverteilung in die Form (2.54) bringt und dann eine Priori gemäß (2.55) wählt.
- $b(\vartheta)$ spielte in (2.54) und in (2.55) eine ganz unterschiedliche Rolle:
In (2.54) ist $b(\vartheta)$ der natürliche Parameter der Exponentialfamilie, aus der die Likelihood / Stichprobenverteilung stammt.
In (2.55) hingegen ist $b(\vartheta)$ die suffiziente Statistik für den natürlichen Parameter β der Exponentialfamilie, aus der die Priori stammt. (Bei der Priori ist ja der Wert von ϑ „zufällig“.)
- Ähnliches gilt für $h(\vartheta)$.
- De facto „datiert man einfach mittels der suffizienten Statistik auf“ (vgl. 2.88).

Bsp. 2.96 (Beispiele zu Satz 2.94)

Man bestimme in folgenden Situationen unter Verwendung von Satz 2.94 jeweils eine konjugierte Priori-Verteilung:

- a) X_1, \dots, X_n ist i.i.d. normalverteilt mit unbekanntem μ und bekannter Varianz σ^2

- b) X ist binomialverteilt zum unbekanntem Parameter p

- c) X_1, \dots, X_n ist i.i.d. Poisson-verteilt mit unbekanntem Parameter λ

2.4.8 (Reine) Bayes-Punktschätzung

Def. 2.97 (MPD-Schätzung)

Gegeben eine Beobachtung \vec{x} und die Posteriori-Verteilung mit Dichte bzw. Wahrscheinlichkeitsfunktion $\pi(\vartheta|\vec{x})$ heißt $\hat{\vartheta}$ mit

$$\pi(\hat{\vartheta}|\vec{x}) = \max_{\vartheta \in \Theta} \pi(\vartheta|\vec{x})$$

(reiner) Bayes-Schätzwert oder *Maximum (bzw Highest) Posteriori Density Schätzwert* (MPD- (bzw. HPD-) Schätzwert) oder *Posteriori-Modus-Schätzwert*. Die zugehörige Schätzfunktion $\hat{\vartheta}(\vec{X})$ heißt *reine Bayes-Schätzung* oder *MPD- (bzw. HPD-) Schätzung* bzw. *Posteriori-Modus-Schätzung*.

Bem. 2.98 (Zur MPD-Schätzung)

- a) Ist die Posteriori-Verteilung unimodal, so ist $\hat{\vartheta}$ der Modus der Posteriori.
- b) Ist der Zustandsraum Θ beschränkt und liegt dem Schätzproblem als Priori-Verteilung eine Gleichverteilung zugrunde, so gilt

$$\pi(\vartheta|\vec{x}) \propto f(\vec{x}|\vartheta) \cdot \pi(\vartheta) = f(\vec{x}|\vartheta) \cdot \text{Konstante}$$

D.h. der MPD-Schätzer ist dasjenige ϑ , das $f(x|\vartheta)$ maximiert, also der Maximum-Likelihood-Schätzwert.

c) Im Falle $\Theta = \mathbb{R}^+$ oder $\Theta = \mathbb{R}$ gibt es keine Gleichverteilung auf Θ , denn mit

$$f(x) = c \quad \text{ist} \quad \int_0^{\infty} f(x) dx = \int_0^{\infty} c dx = [x]_0^{\infty} = \infty$$

unabhängig von $c > 0$.

Man kann aber zeigen, dass viele der zentralen Ergebnisse der Bayes-Theorie erhalten bleiben, wenn man auch nicht normierbare σ -finite Maße als Prioris zulässt (z.B. Lebesgue Maß $\lambda(\cdot)$; $\lambda([a, b]) := b - a$: „*improper priors*“)

Bsp. 2.99 (Beta-Binomialmodell)

$\pi(\vartheta|x_1, \dots, x_n)$ ist $B(a + \sum_{i=1}^n x_i; b + n - \sum_{i=1}^n x_i) =: B(a', b')$ -verteilt.

Hat man bei der Priori $a=1=b$ gewählt, so ergibt sich mit $\frac{a' - 1}{a' + b' - 2}$ als Modus der $Beta(a', b')$ -Verteilung der MPD-Schätzwert

$$\hat{\vartheta} = \frac{1 + \sum_{i=1}^n x_i - 1}{1 + \sum_{i=1}^n x_i + 1 + n - \sum_{i=1}^n x_i - 2} = \frac{1}{n} \sum_{i=1}^n x_i,$$

also in der Tat der ML-Schätzwert.

2.4.9 Der Hauptsatz der Bayes-Entscheidungstheorie

Def. 2.100 (Posteriori-Verlust-optimale Aktionen, konditionale Bayes-Aktionen)

Gegeben sei ein datenbasiertes Entscheidungsproblem $((\mathbb{A}, \Theta, l(\cdot)); (\mathcal{X}, A, (p_\vartheta)_{\vartheta \in \Theta}))$ und eine Priori-Verteilung $\pi(\cdot)$ über $(\Theta, \sigma(\Theta))$.

Eine Aktion $a_x^* \in \mathbb{A}$ heißt *Posteriori-Verlust optimal* zur *Beobachtung* $x \in \mathcal{X}$ oder *konditionale Bayes-Aktion zu x und der Priori-Verteilung $\pi(\cdot)$* , wenn gilt

$$\mathbb{E}_{\pi(\cdot|x)} l(a_x^*, \vartheta) \leq \mathbb{E}_{\pi(\cdot|x)} l(a, \vartheta) \quad \forall a \in \mathbb{A}.$$

a_x^* ist also sozusagen Bayes-Aktion zur Posteriori-Verteilung $\pi(\cdot|x)$ als „aufdatierter Priori-Verteilung“ $\pi(\cdot|x)$.

Analog definiert man eine *Posteriori-Nutzen-Optimalität*.

2 Arten, Bayes-Entscheidungstheorie zu betreiben

datengestütztes Entscheidungsproblem
+
Informationsbeschaffungsexperiment
+
Priori-Verteilung;

↓
Auswertungsproblem + Priori-Verteilung
komplexer Aktionsraum: alle
Entscheidungsfunktionen

↓
Bayes-optimale
Entscheidungsfunktion $d^* : \mathcal{X} \rightarrow \mathbb{A}$
(→ Testfunktion, Schätzfunktion)

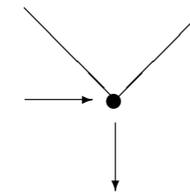
↓
konkrete Beobachtung x

↓
Bayes optimale Aktion
 $a^* = d^*(x)$

Priori-Verteilung

Stichprobenverteilung;
Informationsbeschaffungsexperiment

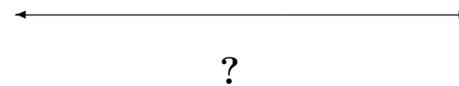
konkrete Beobachtung



Posteriori-Verteilung

↓ Bayes Postulat

Posteriori-Verlust optimale Aktion a_x^* ,
z.B. reine (optimale) Bayes Schätzung
 $\hat{\vartheta}_x$
reiner/optimaler Bayes Test φ_x



„Priori-Risiko“ optimale Aktion

Satz 2.101 (Hauptsatz der Bayes-Entscheidungstheorie)

Gegeben sei ein datengestütztes Entscheidungsproblem $((\mathbb{A}, \Theta, \ell(\cdot)); (\mathcal{X}, \mathcal{A}, (p_\vartheta)_{\vartheta \in \Theta}))$, bestehend aus einem *datenfreien Entscheidungsproblem* $(\mathbb{A}, \Theta, \ell(\cdot))$ und einer Informationsstruktur $(\mathcal{X}, \mathcal{A}, (p_\vartheta)_{\vartheta \in \Theta})$ sowie eine Priori-Verteilung $\pi(\cdot)$ über $(\Theta, \sigma(\Theta))$.

Eine Entscheidungsfunktion

$$\begin{aligned} d^* : \mathcal{X} &\longrightarrow \mathbb{A} \\ x &\longmapsto d^*(x) \end{aligned}$$

ist genau dann Bayes-optimal im zugehörigen Auswertungsproblem, wenn für jedes $x \in \mathcal{X}$ die zugehörige Aktion $d^*(x)$ Posteriori-Verlust optimal zur Beobachtung x ist.

Beweis: Für den diskreten Fall ⁷

- Vorneweg eine Hilfsüberlegung: Suche die Lage des Minimums \vec{z}_{min} einer Funktion $f(\vec{z})$ in $\vec{z} = (z_1, \dots, z_n)$, wobei mit $c_i \geq 0$, $i = 1, \dots, n$ gilt: $f(\vec{z}) = \sum_{i=1}^n c_i f_i(z_i)$, also die i -te Komponente von z nur im i -ten Summanden auftritt.

$$f(\vec{z}) = \sum_{i=1}^n c_i f_i(z_i) \rightarrow \min_{\vec{z}}$$

$$\iff f_i(z_i) \longrightarrow \min_{z_i} \quad \text{für jedes } i \text{ unabhängig von den anderen Summanden.}$$

- Der Deutlichkeit halber wird wieder eine Hilfsvariable U (vgl. Bem. 2.86) eingeführt und $p_{\vartheta}(\{X = x\})$ wird als $p(\{X = x\}|\{U = \vartheta\})$ geschrieben.

⁷für den allgemeinen Fall: siehe z.B. Rüger (1999, S. 283f.)

Angewendet auf Entscheidungsprobleme mit der Posteriori $\pi(\vartheta|x)$ ergibt sich mit dieser Notation

$$\pi(\vartheta|x) = \frac{p(\{X = x\}|\{U = \vartheta\}) \cdot \pi(\vartheta)}{p(\{X = x\})}. \quad (2.57)$$

Nun betrachte man Entscheidungsfunktionen $d(\cdot)$ im Auswertungsproblem:
Für die Risikofunktion

$$R(d, \vartheta) = \mathbb{E}_{p_\vartheta}(\ell(d(x), \vartheta))$$

gilt hier

$$R(d, \vartheta) = \sum_{x \in \mathcal{X}} \ell(d(x), \vartheta) \cdot p_\vartheta(\{X = x\}).$$

Die optimale Entscheidungsfunktion zur Priori $\pi(\cdot)$ minimiert unter allen d

$$\mathbb{E}_\pi(R(d, \vartheta)),$$

löst also

$$\sum_{\vartheta \in \Theta} \left(\sum_{x \in \mathcal{X}} \ell(d(x), \vartheta) \cdot p_\vartheta(\{X = x\}) \right) \cdot \pi(\vartheta) \rightarrow \min_d$$

$$\begin{aligned}
&\iff \sum_{\vartheta \in \Theta} \sum_{x \in \mathcal{X}} \left(\ell(d(x), \vartheta) \cdot \underbrace{p(\{X = x\} | U = \vartheta) \cdot \pi(\vartheta)}_{= \pi(\vartheta|x) \cdot p(\{X=x\})} \right) \rightarrow \min_d \\
&\stackrel{(2.57)}{\iff} \underbrace{\sum_{x \in \mathcal{X}}}_{\hat{=} \sum_{i=1}^n} \left(\underbrace{\sum_{\vartheta \in \Theta} \ell(d(x), \vartheta) \cdot \pi(\vartheta|x)}_{\hat{=} f_i(z_i)} \right) \cdot \underbrace{p(\{X = x\})}_{\hat{=} c_i; \text{ priori-prädiktiv, marginal}} \rightarrow \min_d
\end{aligned}$$

- Wegen der Hilfsüberlegung ist dies äquivalent dazu, für jedes feste x

$$\sum_{\vartheta \in \Theta} \ell(d(x), \vartheta) \cdot \pi(\vartheta|x)$$

separat zu minimieren nach $a_x := d(x)$ für festes x .

Dies liefert jeweils genau die Posteriori-Verlust optimale Aktion, also die Bayes-Aktion zur Posteriori als neuer Priori.

**Satz 2.102 (Bestimmung von Bayes-optimalen
Entscheidungsfunktionen, z.B. Rüger (1999, Satz 2.20))**

Gegeben sei das Schätzproblem als datengestütztes Entscheidungsproblem gemäß Kapitel 1.5 sowie eine Priori-Verteilung $\pi(\cdot)$.

Dann gilt:

- i) Wählt man die quadratische bzw. absolute Verlustfunktion, so gilt für die Bayes-optimale Entscheidungsfunktion $d_{quad}^*(\cdot)$ bzw. $d_{abs}^*(\cdot)$:
Für jedes x ist $d_{quad}^*(\cdot)$ genau der Erwartungswert und $d_{abs}^*(\cdot)$ der Median der Posteriori-Verteilung $\pi(\vartheta|x)$.

ii) Die HPD-Schätzung ergibt sich näherungsweise für kleine ϵ , wenn man die sogenannte Toleranzverlustfunktion zum Grade ϵ verwendet:

$$l_{\epsilon}(\hat{\vartheta}, \vartheta) = \begin{cases} 1 & |\hat{\vartheta} - \vartheta| > \epsilon \\ 0 & |\hat{\vartheta} - \vartheta| \leq \epsilon \end{cases}$$

2.4.10 „Asymptotische Objektivität“ der konditionalen Bayes-Inferenz

Motivationsbeispiel: Betrachte (2.52) und (2.53) für $n \rightarrow \infty$

$$\lim_{n \rightarrow \infty} \nu' = \frac{\bar{x}\rho^2 + 0}{\rho^2 + 0} = \bar{x} \quad (2.58)$$

$$\lim_{n \rightarrow \infty} \rho^{2'} = \frac{\rho^2 \cdot 0}{\rho^2 + 0} = 0 \quad (2.59)$$

- Die Grenzwerte (2.58) und (2.59) hängen nicht von ρ und ν , also von den Parametern der Priori-Verteilung ab. Hat man eine sehr große Stichprobe, so „verschwindet der Einfluss der Priori-Verteilung“. Dies

gilt ganz allgemein und wird oft als eine pragmatische Rechtfertigung dafür gesehen, bei großem Stichprobenumfängen „angenehme“, z.B. konjugierte, Prioris zu verwenden.

Satz 2.103 („Asymptotische Objektivität von Bayes-Verfahren“, „Konsistenzsatz“)

Sei $\Theta = \{\vartheta_1, \dots, \vartheta_m\}$ ein endlicher Parameterraum und $\vec{X} = (X_1, \dots, X_n)$ eine i.i.d. Stichprobe eines beliebig verteilten (reellwertigen) Untersuchungsmerkmals mit Dichten $f(x_i|\vartheta_{wahr})$, $\vartheta_{wahr} \in \Theta$.

Sei $\pi(\vartheta)$ die Wahrscheinlichkeitsfunktion der Priori-Verteilung auf Θ mit $\pi(\vartheta) > 0$ für alle ϑ . Dann gilt für die Wahrscheinlichkeitsfunktion der nach n Beobachtungen gebildeten Posteriori-Verteilung $\pi_n(\vartheta|x)$

$$\lim_{n \rightarrow \infty} \pi_n(\vartheta|x) = \begin{cases} 1 & \text{falls } \vartheta = \vartheta_{wahr} \\ 0 & \text{falls } \vartheta \neq \vartheta_{wahr} \end{cases}$$

Bem. 2.104 (Erneute kritische Diskussion des Bayes-Ansatzes)

2.5 Einige alternative Regeln (im Kontext der klassischen Entscheidungstheorie)

2.5.1 Die Laplace Regel



Foto:

<http://www.mathematik.de/ger/information/landkarte/gebiete/wahrscheinlichkeitstheorie/wahrscheinlichkeitstheorie.html>

[Stand: 25.06.13]

Def. 2.105 (Laplace-Regel)

Gegeben sei das datenfreie Entscheidungsproblem $(\mathbb{A}, \Theta, u(\cdot))$ mit endlichem $\Theta = \{\vartheta_1, \dots, \vartheta_m\}$.

Die Kriteriumsfunktionen

$$\begin{aligned} \tilde{\Phi} : \mathbb{A} &\longrightarrow \mathbb{R} \\ a &\longmapsto \sum_{j=1}^m u(a, \vartheta_j) \end{aligned} \tag{2.60}$$

und

$$\begin{aligned}\Phi : \mathbb{IA} &\longrightarrow \mathbb{R} \\ a &\longmapsto \frac{1}{m} \sum_{j=1}^m u(a, \vartheta_j)\end{aligned}\tag{2.61}$$

heißen *Laplace-Regel*.

Bem. 2.106

Die beiden Kriteriumsfunctonen (2.60) und (2.61) liefern dieselbe Ordnung auf der Aktionenmenge.

- a) Die Kriteriumsfuncton (2.61) entspricht einer Bayes-Regel mit Prioriverteilung $\pi(\cdot) = (\frac{1}{m}, \frac{1}{m}, \dots, \frac{1}{m})$ (mit $m = |\Theta|$), also einer Gleichverteilung auf Θ .

Damit geometrisch: Höhenlinien senkrecht auf Equiligator-Linie.

- b) Viele Eigenschaften der optimalen Aktion können deshalb aus den in Kapitel 2.4 formulierten Sätzen über Bayes-Regeln abgeleitet werden. (Zulässigkeit, Entbehrlichkeit randomisierter Aktionen,...)

c) Rechtfertigung durch „Prinzip vom unzureichenden Grund“ (Laplace):
Wenn nichts dafür spricht, dass eines der Elementarereignisse wahrscheinlicher ist als die anderen, dann sind sie gleichwahrscheinlich, also

$$\pi(\{\vartheta_1\}) = \pi(\{\vartheta_2\}) = \dots = \pi(\{\vartheta_m\}).$$

Da

$$\pi(\{\vartheta_1\}) + \pi(\{\vartheta_2\}) + \dots + \pi(\{\vartheta_m\}) = 1$$

ist zwangsläufig

$$\pi(\{\vartheta_j\}) = \frac{1}{m}, \quad j = 1, \dots, m.$$

- d) Verallgemeinerung auf unendliches Θ :
Theorie der nichtinformativen Prior-Verteilungen, siehe Bemerkung 2.108.

Bem. 2.107 (Beispiel und Kritik)

Abwandlung von Beispiel aus Kapitel 1.3.2, Lotterie
 Urnen bestehend aus einer unbekanntem Anzahl von grünen, blauen und
 restlichen (rote, schwarze, violette) Kugeln.

Man kann entweder

a_1 nicht spielen,

a_2 zum Preis von $c_g = 60\text{€}$ auf grün setzen oder

a_3 zum Preis von $c_b = 90\text{€}$ auf blau setzen.

Es wird eine Kugel zufällig gezogen. Man erhält 240€ , wenn die Kugel, auf
 die man gesetzt hat, gezogen wird.

	$\{g\}$	$\{b\}$	$\{\text{rest}\}$	
a_1	0	0	0	0
a_2	180	-60	-60	20 ← optimale Aktion
a_3	-90	150	-90	-10

Bem. 2.108 („Nichtinformative“ Priori-Verteilung und ihr Informa

In der konditionalen Inferenz gibt es verschiedene Versuche, ähnlich der Laplace-Regel, „nichtinformative“ Priori-Verteilungen zu definieren und diese dann als Standardbewertungen heranzuziehen.

- z.B. die Gleichverteilung, diese ist aber nicht invariant gegenüber Transformationen des Parameters. Man hat dann also „keine Information“ über ϑ , aber eine informative Priori z.B. über eine bijektive, nichtlineare Transformation von ϑ .
- z.B. Verteilungen, die invariant bezüglich bijektiver Transformationen des Parameters sind (Jeffrey-Regel).

- z.B. Verteilungen, die die Entropie maximieren (Jaynes-Regel)
- Ganz neue Möglichkeiten ergeben sich beim Übergang zu Credalmengen (siehe Kapitel 3).

2.5.2 Die Minimax-Regret-Regel von L.J. Savage, auch Niehans-Savage-Regel genannt

fnewpage

Def. 2.109 (Minimax-Regret-Aktion)

Gegeben sei ein datenfreies Entscheidungsproblem $(\mathbb{A}, \Theta, u(\cdot))$ in Nutzenform bzw. $(\mathbb{A}, \Theta, l(\cdot))$ in Verlustform.

Seien \mathbb{A} und Θ endlich, $\Theta = \{\vartheta_1, \dots, \vartheta_m\}$, $\mathbb{A} = \{a_1, \dots, a_n\}$.

Das datenfreie Entscheidungsproblem $(\mathbb{A}, \Theta, r(\cdot))$ in Verlustform mit

$$\begin{aligned} r : \mathbb{A} \times \Theta &\longrightarrow \mathbb{R} \\ (a_i, \vartheta_j) &\longmapsto r(a_i, \vartheta_j) \end{aligned}$$

und

$$r(a_i, \vartheta_j) = \max_{\ell=1, \dots, n} (u(a_\ell, \vartheta_j)) - u(a_i, \vartheta_j) \quad (2.62)$$

$$\text{bzw.} \quad r(a_i, \vartheta_j) = l(a_i, \vartheta_j) - \min_{\ell=1, \dots, n} (l(a_\ell, \vartheta_j)) \quad (2.63)$$

heißt **induziertes Regret Problem** bzw. **induzierte Regret-Tafel**.

Jedes $a^* \in \mathbb{A}$ mit

$$\max_{j=1, \dots, m} r(a^*, \vartheta_j) \leq \max_{j=1, \dots, m} r(a, \vartheta_j) \quad \text{für alle } a \in \mathbb{A}$$

heißt **Minimax-Regret-Aktion**.

Bsp. 2.110 (Beispiel und Kritik)

Proposition 2.111 (Bayes-Aktionen in Regrettafeln)

Gegeben sei ein datenfreies Entscheidungsproblem $(\mathbb{A}, \Theta, u(\cdot))$ bzw. $(\mathbb{A}, \Theta, l(\cdot))$ mit $\mathbb{A} < \infty$ und $|\Theta| < \infty$ und die Priori-Bewertung $\pi(\cdot)$. Eine Aktion a^* ist genau dann Bayes-Aktion zu $\pi(\cdot)$, wenn a^* Bayes-Aktion zu $\pi(\cdot)$ im induzierten Regret-Problem ist.

2.5.3 Das Hurwicz-Kriterium

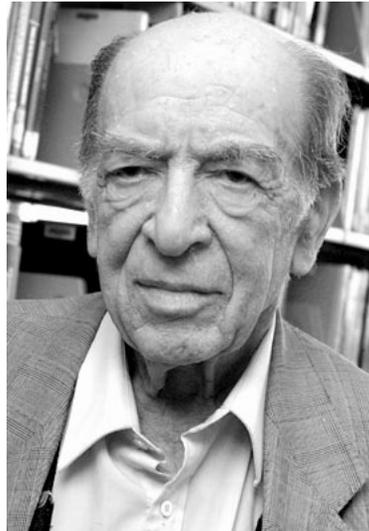


Foto: http://www.nobelprize.org/nobel_prizes/economic-sciences/laureates/2007/hurwicz-facts.html
[Stand: 25.06.13]

Def. 2.112 (Hurwicz-Kriterium)

Gegeben sei ein datenfreies Entscheidungsproblem $(\mathbb{A}, \Theta, u(\cdot))$ mit $|\Theta| < \infty$, sowie $\alpha \in [0, 1]$.

$$\begin{aligned} \Phi : \mathbb{A} &\longrightarrow \mathbb{R} \\ a &\longmapsto \Phi(a) \end{aligned}$$

mit

$$\Phi(a) = \alpha \max_j u(a, \vartheta_j) + (1 - \alpha) \min_j u(a, \vartheta_j) \quad (2.64)$$

heißt *Hurwicz-Kriterium* zum *Optimismusparameter* α .

2.5.4 Das Erfahrungskriterium von J.L. Hodges und E.L. Lehmann (1952)

Def. 2.113 (Erfahrungskriterium von Hodges & Lehmann)

Gegeben sei ein datenfreies Entscheidungsproblem $(\mathbb{A}, \Theta, u(\cdot))$, $|\Theta| < \infty$, $\mu \in [0, 1]$ und eine Priori-Bewertung $\pi(\cdot)$.

$$\begin{aligned} \Phi : \mathbb{A} &\longrightarrow \mathbb{R} \\ a &\longmapsto \Phi(a) \end{aligned}$$

mit

$$\begin{aligned} \Phi(a) = & \mu \cdot \left(\sum_{j=1}^m u(a, \vartheta_j) \pi(\{\vartheta_j\}) \right) \\ & + (1 - \mu) \cdot \left(\min_j (u(a, \vartheta_j)) \right) \end{aligned} \quad (2.65)$$

heißt *Erfahrungskriterium von Hodges und Lehmann* zum Vertrauensparameter μ .

2.6 Gleichmäßig beste Verfahren in der statistischen Entscheidungstheorie

Def. 2.114 (Gleichmäßig beste Verfahren)

Gegeben sei das auf eine Menge \mathcal{D}_0 eingeschränkte Auswertungsproblem $(\mathcal{D}_0, \Theta, R(\cdot))$ eines datengestütztes Entscheidungsproblem $((\mathbb{A}, \Theta, l(\cdot)); (\mathcal{X}, \sigma(\mathcal{X}), (p_\vartheta)_{\vartheta \in \Theta}))$

Eine Entscheidungsfunktion $d^* \in \mathcal{D}_0$ heißt *gleichmäßig bestes Verfahren* aus \mathcal{D}_0 , wenn für die Risikofunktion gilt:

$$R(d^*, \vartheta) \leq R(d, \vartheta) \quad \text{für alle } \vartheta \in \Theta \text{ und alle } d \in \mathcal{D}_0. \quad (2.66)$$

Bem. 2.115 (zu Def. 2.114)

- i) Man beachte, dass die Risikofunktion von d^* für *alle* $\vartheta \in \Theta$ nicht größer sein soll; d^* soll also im zugehörigen Auswertungsproblem $(\mathcal{D}_0, \Theta, R(\cdot))$ alle Elemente von \mathcal{D}_0 dominieren.
- ii) Dies ist für „großes \mathcal{D}_0 “ eine extrem starke Forderung – insbesondere im Lichte der elementaren Beispiele aus Kapitel 1.3 – aber v.a. bei Exponentialfamilie und geeigneter Einschränkung der Menge der betrachteten Entscheidungsfunktionen möglich (UMVU-Schätzer, gleichmäßig bester Test, siehe später).

Satz 2.116 (Lehmann und Scheffé (1950))

Gegeben sei ein Schätzproblem im Sinne Beispiel 1.45. Ferner sei

- $(p_{\vartheta})_{\vartheta \in \Theta}$ eine strikt q -parametrische Exponentialfamilie in $T(\vec{x}) = (T_1(\vec{x}), \dots, T_q(\vec{x}))$.
- $\ell(\hat{\vartheta}, \vartheta)$ konvex in $\hat{\vartheta}$ für alle ϑ .

Betrachtet man die Schätzung einer Transformation $\gamma(\vartheta)$ von ϑ und die Klasse \mathcal{D}_γ aller für $\gamma(\vartheta)$ erwartungstreuen Schätzer, so gilt:

Ist \mathcal{D}_γ nicht leer, so gibt es eine nichtrandomisierte Schätzfunktion der Form $\eta(T(\vec{X}))$, die gleichmäßig beste in der Klasse \mathcal{D}_γ ist.

Ist umgekehrt $\eta(T(\vec{X}))$ eine erwartungstreue Schätzfunktion für $\gamma(\vartheta)$, so ist $\eta(T(\vec{X}))$ gleichmäßig bestes Verfahren in \mathcal{D}_γ .

Bem. 2.117 (Zur Interpretation des Satzes)

- „Informationsdeutung“:

- Konstruktive Anwendung:

Korollar 2.118

Gegeben sei eine strikt q -parametrische Exponentialfamilie in $T(\vec{x}) = (T_1(\vec{x}), \dots, T_q(\vec{x}))$.

Dann ist für jede (messbare) Funktion $\eta(\cdot)$ der Schätzer $\eta(T_1, \dots, T_q)$ UMVU-Schätzer für $\mathbb{E}_\vartheta(\eta(T_1, \dots, T_q))$.

Bem. 2.119 (Zum Satz von Lehmann-Scheffé)

Die Beschränkung auf erwartungstreue Schätzer ist wesentlich.

Bem. 2.120

Korollar 2.118 wird typischerweise „andersherum“ angewendet. Will man eine Funktion $\gamma(\vartheta)$ schätzen, so sucht man eine Funktion $g(T)$, so dass $\mathbb{E}g(T) = \gamma(\vartheta)$. Gemäß Korollar 2.118 weiß man dann, dass $g(T)$ UMVU für $\gamma(\vartheta)$ ist.

Üblicherweise wird man versuchen einen Ansatz zu wählen, bei dem sich $g(\cdot)$ aus „einfachen Grundfunktionen“ zusammensetzt.

Bsp. 2.121

Gegeben sei eine i.i.d. Stichprobe eines normalverteilten Merkmals mit unbekanntem Mittelwert μ , aber bekannter Varianz σ^2 . Man bestimme einen UMVU Schätzer

- a) für μ und
- b) für $\exp(\mu)$.

Satz 2.123 (Optimale Tests)

Betrachtet werde das Testproblem als Entscheidungsproblem gemäß Beispiel 1.47 mit

a_0 für H_0 entscheiden

a_1 für H_1 entscheiden

a_0	0	1
a_1	1	0

$$l(a_0, \vartheta) = \begin{cases} 0 & \vartheta \in \Theta_0 \\ 1 & \vartheta \in \Theta_1 \end{cases}$$

$$l(a_1, \vartheta) = \begin{cases} 1 & \vartheta \in \Theta_0 \\ 0 & \vartheta \in \Theta_1 \end{cases}$$

Ferner sei $\Theta_0 = \{\vartheta | \vartheta \leq \vartheta_0\}$, $\Theta_1 = \{\vartheta | \vartheta \geq \vartheta_1\}$, $\vartheta_1 > \vartheta_0$, und ein Signifikanzniveau $\alpha \in (0, 1)$ vorgegeben.

Bildet $(P_{\vartheta}^{\oplus n})$ eine strikt einparametrische Exponentialfamilie in T (mit dem natürlichen Parameter $c(\vartheta)$, der in eindeutiger Beziehung zu ϑ stehe), so gibt es ein $\kappa \in \mathbb{R}$, so dass der Test

$$\varphi^*(\vec{x}) = \begin{cases} 1 & T(\vec{x}) > \kappa \\ \gamma & T(\vec{x}) = \kappa \\ 0 & T(\vec{x}) < \kappa \end{cases} \quad (2.67)$$

mit

$$\mathbb{E}_{\vartheta_0} \varphi^* = \alpha \quad (2.68)$$

gleichmäßig bester Test (UMP) ist, d.h es gilt

$$\mathbb{E}_{\vartheta} \varphi^* \geq \mathbb{E}_{\vartheta} \varphi, \quad \forall \vartheta \in \Theta_1,$$

für alle φ mit $\sup_{\vartheta \in \Theta_0} \mathbb{E}_{\vartheta} \varphi \leq \alpha$.

Bem. 2.124 (Zur Interpretation von Satz 2.123)

- Informationsdeutung:

- konstruktive Anwendung:

Bem. 2.125

- Die Aussage gilt allgemeiner für Verteilungen mit *monotonen Dichtequotienten*; bei denen also für alle $\vartheta_1 \in \Theta_1$, $\vartheta_0 \in \Theta_0$ und eine geeignete Funktion $T(\vec{X})$ der Quotient $\frac{f_{T||\vartheta_1}(t)}{f_{T||\vartheta_0}(t)}$ monoton in t ist.
- Auch bei der Fragestellung $H_0 : \vartheta = \vartheta_0$ gegen $H_1 : \vartheta \neq \vartheta_0$, gibt es bei Exponentialfamilien einen gleichmäßig besten Test, wenn man sich auf *unverfälschte* Tests (d.h. Gütefunktion $\geq \alpha$ für alle Elemente der Alternative) beschränkt.
Ähnliches gilt in der Situation $H_0 : \vartheta \in [\underline{\vartheta}_0, \bar{\vartheta}_0]$ gegen $H_1 : \vartheta \in [\underline{\vartheta}_1, \bar{\vartheta}_1]$, wenn man nur unverfälschte und ähnliche Tests (Gütefunktion am Rand der Nullhypothese $= \alpha$) betrachtet.

Bsp. 2.127 Anwendung von Satz 2.123 auf den Mittelwertstest bei der Normalverteilung und auf das Testen bei der Binomialverteilung

a) X_1, \dots, X_n *i.i.d.* $\sim N(\mu, \sigma^2)$

$H_0 : \mu \leq \mu_0$ gegen $H_1 : \mu \geq \mu_1 > \mu_0$ σ^2 bekannt

$$\begin{aligned}
 f_{X_1, \dots, X_n | \mu}(x_1, \dots, x_n) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} \cdot \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right) \\
 &= \underbrace{\left(\frac{1}{\sqrt{2\pi}}\right)^n}_{\text{konstant (bei bek. } \sigma^2)} \cdot \frac{1}{\sigma^n} \cdot \underbrace{\exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n x_i^2\right)}_{\text{nur von } X \text{ abhängig (bei bekanntem } \sigma^2)} \cdot \underbrace{\exp\left(-\frac{1}{2\sigma^2} n\mu^2\right)}_{\text{nur von } \mu \text{ abh. nicht von } x} \cdot \exp\left(\frac{\mu}{\sigma^2} \cdot \sum_{i=1}^n x_i\right)
 \end{aligned}$$

natürlicher Parameter $\frac{\mu}{\sigma^2}$

Suffiziente Statistik: $T = \sum_{i=1}^n X_i$

Die kritische Region $K(\kappa)$, d.h. der Bereich aller \vec{x} mit $\varphi(\vec{x}) = 1$, ist wegen (3.5) von der Form

$$K(\kappa) := \{\vec{x} | T(\vec{x}) > \kappa\}$$

mit κ so, dass

$$P_{\mu_0}(K(\kappa)) \stackrel{!}{=} \alpha.$$

(Man weiß ja wegen (2.68), dass die maximale Fehlerwahrscheinlichkeit an der Grenze der Nullhypothese angenommen wird.)

Das heißt, es soll gelten

$$P_{\mu_0}(K(\kappa)) = P_{\mu_0}(\{\vec{x} | T(\vec{x}) > \kappa\}) = P_{\mu_0}(\{\vec{x} | \sum_{i=1}^n X_i > \kappa\}) = \alpha.$$

Zur Berechnung dieser Wahrscheinlichkeit benutzt man

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

d.h.

$$\frac{\sum_{i=1}^n X_i - n\mu_0}{\sigma\sqrt{n}} \sim N(0, 1).$$

Also führt der Ansatz $P_{\mu_0}(K(\kappa)) \stackrel{!}{=} \alpha$, d.h.

$$P \left(\frac{\sum_{i=1}^n X_i - n\mu_0}{\sigma \cdot \sqrt{n}} > \underbrace{\frac{\kappa - n\mu}{\sigma \cdot \sqrt{n}}}_{\kappa' = \tau_\alpha} \right) \stackrel{!}{=} \alpha,$$

Fraktile der Normalverteilung

dazu, dass die kritische Region so zu wählen ist, dass

$$\begin{aligned} & \frac{\sum_{i=1}^n X_i - n \cdot \mu}{\sigma \cdot \sqrt{n}} > \tau_\alpha \\ \iff & \frac{\bar{X} - \mu}{\sigma} \cdot \sqrt{n} > \tau_\alpha \\ \iff & \bar{X} > \mu_0 + \frac{\tau_\alpha \cdot \sigma}{\sqrt{n}}. \end{aligned}$$

Da ferner $\mu_1 > \mu_0$ bei dieser Konstruktion beliebig gewählt werden konnte, gilt die Aussage für *alle* $\mu > \mu_0$.

Beachte: es ergibt sich die übliche kritische Region des Gauss Tests, dieser ist also UMP.

b) Bei der Bernoulliverteilung sei zu einem konkreten Beispiel übergegangen.

Man testet $H_0 : p \leq 0.5$ gegen $H_1 : p \geq 0.6$, wobei bei der i.i.d. Stichprobe X_1, \dots, X_n der Stichprobenumfang $n = 5$ sei und das Signifikanzniveau auf $\alpha = 0.1$ gesetzt sei.

Zur Anwendung von Satz 2.123 bringt man zunächst die Wahrscheinlichkeitsfunktion $f_p(\cdot)$ der Bernoulliverteilung auf

„Exponentialfamilien-Gestalt“.

$$\begin{aligned}
 f_p(x_1, \dots, x_n) &= \prod_{i=1}^n p^{x_i} \cdot (1-p)^{1-x_i} = \\
 &= (1-p)^n \cdot \left(\frac{p}{1-p}\right)^{\sum_{i=1}^n x_i} = \\
 &= (1-p)^n \cdot \exp\left(\ln\left(\left(\frac{p}{1-p}\right)^{\sum_{i=1}^n x_i}\right)\right) = \\
 &= (1-p)^n \cdot \exp\left(\sum_{i=1}^n x_i \cdot \ln\left(\frac{p}{1-p}\right)\right)
 \end{aligned}$$

Man erhält $T = \sum_{i=1}^n X_i$, und $\ln\left(\frac{p}{1-p}\right)$ ist der natürliche Parameter.

Der Ansatz

$$K(\kappa) := \{\vec{x} | T(\vec{x}) > \kappa\}$$

für die kritische Region führt auf ein Problem. Da $T = \sum_{i=1}^n X_i$ binomialverteilt ist, ergibt sich ⁸

$$P_{p_0}(K(\kappa)) = P_{p_0}(\{T > \kappa\}) = \sum_{\substack{j > \kappa \\ j \in \mathbb{N}}}^n \binom{5}{j} \underbrace{0.5^j \cdot 0.5^{5-j}}_{0.5^5}$$

⁸Wegen (2.68) setzt man hier wieder den oberen Randwert der Nullhypothese ein, als denjenigen Wert der Nullhypothese, der am schwersten von H_1 zu unterscheiden ist.

Allerdings gibt es kein κ , so dass diese Gleichung erfüllt ist:

$$\kappa \in (4, 5] \quad P(K(\kappa)) = 0$$

$$\kappa = 4 \Rightarrow P(K(\kappa)) = \binom{5}{5} \cdot \left(\frac{1}{2}\right)^5 = \frac{1}{32} < 0.1$$

$$\kappa \in (3, 4] \Rightarrow P(K(\kappa)) = P(K(4)) = \frac{1}{32}$$

$$\kappa = 3 \Rightarrow P(K(\kappa)) = \frac{1}{32} + \binom{5}{4} \cdot 0.5^5 = \frac{1}{32} + \frac{5}{32} = \frac{6}{32} > 0.1$$

Was tun? Klar ist

$\varphi(\vec{x}) = 1$, d.h. H_0 ablehnen, für alle \vec{x} mit $T(\vec{x}) = 5$ und

$\varphi(\vec{x}) = 0$, d.h. H_0 nicht ablehnen, für alle \vec{x} mit $T(\vec{x}) \leq 3$.

Ist $T(\vec{x}) = 4$, dann hat man so zu randomisieren, dass $\mathbb{E}_{0.5}\varphi = \alpha$. Man setzt hierzu $\varphi(\vec{x}) = \gamma$, falls $T(\vec{x}) = 4$, und erhält:

$$\begin{aligned}\mathbb{E}_{0.5}\varphi &= 1 \cdot P(\varphi(x) = 1) + \gamma \cdot P(\varphi(x) = \gamma) + \\ &\quad + 0 \cdot P(\varphi(x) = 0) \stackrel{!}{=} \alpha\end{aligned}$$

also

$$\begin{aligned}\gamma &= \frac{\alpha - P(\varphi(x) = 1)}{P(\varphi(x) = \gamma)} = \\ &= \frac{\alpha - P(T = 5)}{P(T = 4)} = \frac{0.1 - \frac{1}{32}}{\frac{5}{32}} = 0.44 .\end{aligned}$$

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 –)

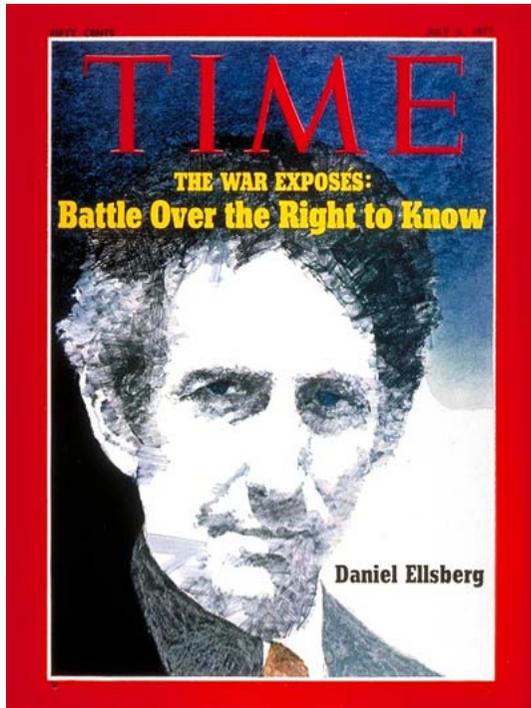


Foto: Time Magazine (July 1971)

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 2.63) 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

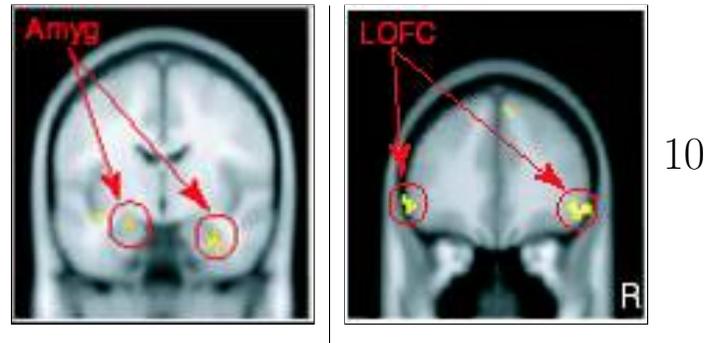
- **Wissen**

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

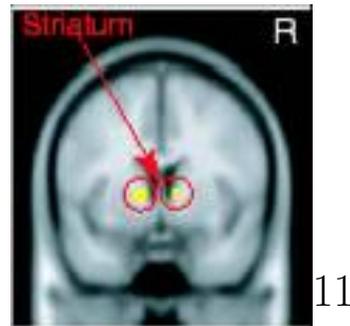
Bsp.: Wetter in New York – Wetter in Tirana

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

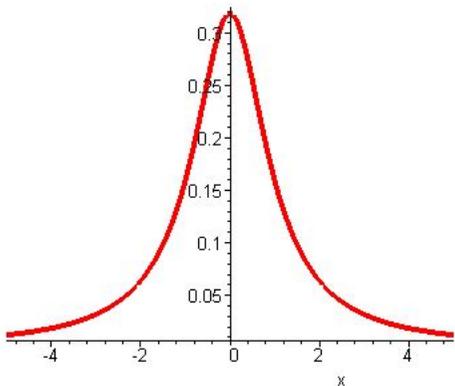
- Ordinale Wahrscheinlichkeiten, z.B. $\pi(\{\vartheta_{(1)}\}) \leq \pi(\{\vartheta_{(2)}\}) \leq \pi(\{\vartheta_{(3)}\})$, beschreibt eine Credal-Menge.
- Eine andere Interpretation/Anwendung von Credal-Mengen ist die Modellierung von Gruppenentscheidungen: verschiedene klassische Prior-Verteilungen.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

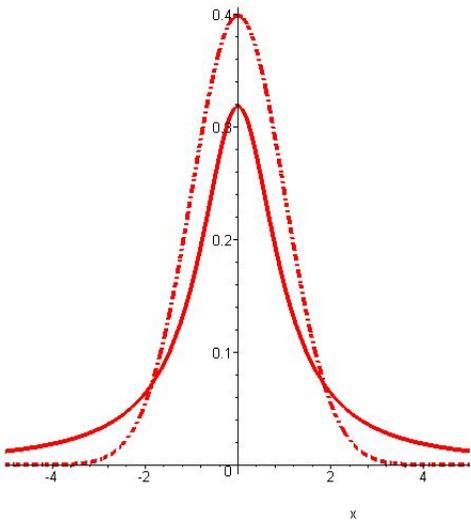
- Vorsichtige Analyse bei unvollständigen Daten (Manski (2003): *Partial Identification*). Betrachte die Menge aller mit den Daten kompatibler Modelle!
- Alternative Inferenzmethoden, insbesondere *logischer Wahrscheinlichkeitsbegriff* (Levi, Kyburg, Weichselberger/Wallner)
- Modellierung Unsicheren 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:

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

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

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

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

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

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

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

$$\begin{aligned} \Phi_{\eta}(a_i, \mathcal{M}) &:= \eta \cdot \underline{\mathbb{E}}_{\mathcal{M}}(u(a_i)) + (1 - \eta) \bar{\mathbb{E}}_{\mathcal{M}}(u(a_i)) \\ \Phi_{\eta}(a_1, \mathcal{M}) &= \eta \cdot \frac{1}{3} + (1 - \eta) \cdot \frac{1}{3} = \frac{1}{3} \\ \Phi_{\eta}(a_2, \mathcal{M}) &= \eta \cdot 0 + (1 - \eta) \cdot \frac{2}{3} = (1 - \eta) \cdot \frac{2}{3} \end{aligned}$$

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 ist äusserst problematisch (vgl. Kap. 2.5.1).

- 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. 2.86 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 2.94)

Eine zum Nachweis meist umständlichere, aber für die Interpretation oft anschaulichere und für die Verallgemeinerung besser geeignete, alternative Darstellung von Satz 2.94 lautet:

Hat in der Situation von Def. 2.86 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 2.93)

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

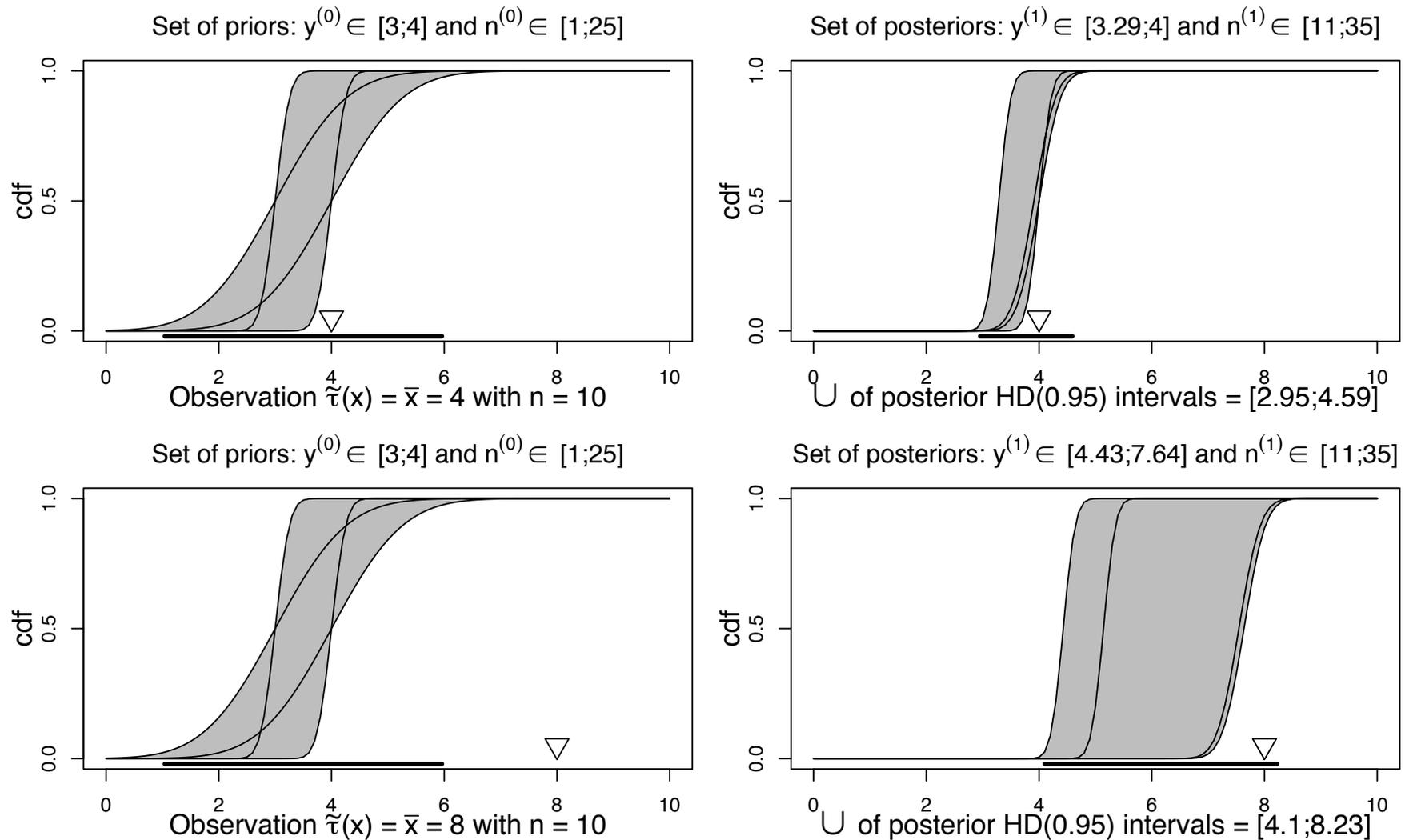
$$\text{pdc} \left(\frac{1}{n} \sum_{i=1}^n x_i; \underline{\nu}^{(0)}, \bar{\nu}^{(0)} \right) := \inf \left\{ \left| \frac{1}{n} \sum_{i=1}^n x_i - \nu^{(0)} \right| \mid \underline{\nu}^{(0)} \leq \nu^{(0)} \leq \bar{\nu}^{(0)} \right\}.$$

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.

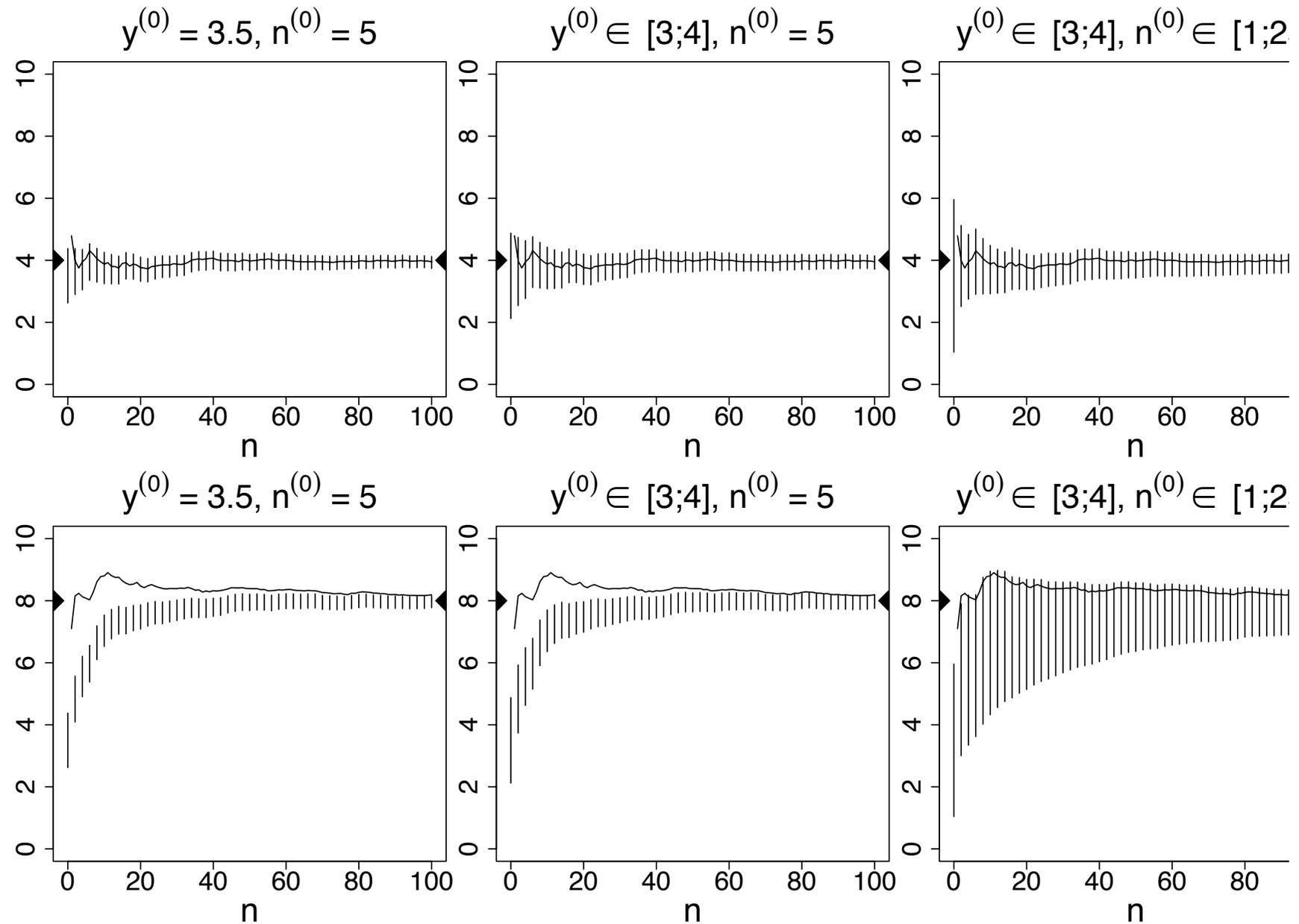
Quelle: Walter & Augustin (2009, S. 268)



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 2.103: Konsistenzsatz für Credal-)

In der Situation von Satz 2.103 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:

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