Generalized iLUCK-models and prior-data conflict

A contribution to Generalized Bayesian Inference

Gero Walter

Institut für Statistik
Ludwig-Maximilians-Universität München

4. Juni 2009

Generalisierte Bayes-Inferenz – Idee

Bayes-Inferenz über einen Parameter θ:

Priori-Wissen über θ + Daten x → aufdatiertes Wissen über θ

Generalisierte Bayes-Inferenz – Idee

Bayes-Inferenz über einen Parameter θ:

Priori-Wissen über θ	+	Daten x	→	aufdatiertes Wissen über θ
Priori-Verteilung $p(\theta)$	+	Likelihood $f(x \mid \theta)$	→	Posteriori-Verteilung $p(\theta \mid x)$

Generalisierte Bayes-Inferenz – Idee

Bayes-Inferenz über einen Parameter θ:

Priori-Wissen über θ	+	Daten x	→	aufdatiertes Wissen über θ
Priori-Verteilung $p(\theta)$	+	Likelihood $f(x \mid \theta)$	→	Posteriori-Verteilung $p(\theta \mid x)$
Menge von Prioris $\mathcal{M}_\theta$	+	Likelihood $f(x \mid \theta)$	→	**Menge von** Posterioris $\mathcal{M}_{\theta\mid x}$

Generalisierte Bayes-Inferenz – Idee

➡ eindimensionale Größen werden intervallwertig, z.B.

$$\mathbb{E}[\theta] \longrightarrow \left[\underline{\mathbb{E}}[\theta], \overline{\mathbb{E}}[\theta]\right] = \left[\min_{p \in \mathcal{M}_\theta} \mathbb{E}_p[\theta], \max_{p \in \mathcal{M}_\theta} \mathbb{E}_p[\theta]\right]$$

$$P(\theta \in A) \longrightarrow \left[\underline{P}(\theta \in A), \overline{P}(\theta \in A)\right]$$

$$= \left[\min_{p \in \mathcal{M}_\theta} P_p(X \in A), \max_{p \in \mathcal{M}_\theta} P_p(X \in A)\right]$$

Generalisierte Bayes-Inferenz – Idee

➡ eindimensionale Größen werden intervallwertig, z.B.

$$\mathbb{E}[\theta] \longrightarrow \left[\underline{\mathbb{E}}[\theta], \overline{\mathbb{E}}[\theta]\right] = \left[\min_{p \in \mathcal{M}_\theta} \mathbb{E}_p[\theta], \max_{p \in \mathcal{M}_\theta} \mathbb{E}_p[\theta]\right]$$

$$P(\theta \in A) \longrightarrow \left[\underline{P}(\theta \in A), \overline{P}(\theta \in A)\right]$$

$$= \left[\min_{p \in \mathcal{M}_\theta} P_p(X \in A), \max_{p \in \mathcal{M}_\theta} P_p(X \in A)\right]$$

Mengen von Verteilungen („credal sets“) ↔ Intervalle für Wahrscheinlichkeiten / Erwartungswerte

Wahrscheinlichkeiten ↕ Weichselberger (2001)

Erwartungswerte ↕ Walley (1991)

➡ The Society for Imprecise Probability: Theories and Applications (www.sipta.org)

Unsicherheit ist mehrdimensional

„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 has become 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.“

Klir & Wierman: Uncertainty-based Information, Physika, 1998, S.1

Unsicherheit ist mehrdimensional

- *risk* / Risiko / ideale stochastische Unsicherheit: „Zufalls-Mechanismus“ bekannt/klar
- *ambiguity* / Ambiguität / Nicht-stochastische Unsicherheit: „Zufalls-Mechanismus“ unbekannt/unklar
- ...

(siehe auch Knight (1921): *risk* vs. *uncertainty*)

Unsicherheit ist mehrdimensional

- *risk* / Risiko / ideale stochastische Unsicherheit: „Zufalls-Mechanismus“ bekannt/klar
- *ambiguity* / Ambiguität / Nicht-stochastische Unsicherheit: „Zufalls-Mechanismus“ unbekannt/unklar
- ...

(siehe auch Knight (1921): *risk* vs. *uncertainty*)

→ mit Mengen von Verteilungen kann *ambiguity* explizit ins Modell einbezogen werden:

einzelne Verteilung	⟷	*risk*
Größe der Menge	⟷	*ambiguity*

Additivität $P(A \cup B) = P(A) + P(B)$

(subjektive) Bayesianer: jede Art von Unsicherheit als (subjektive) Wahrscheinlichkeitsverteilung darstellbar

Probleme bei

- Entwicklung von Expertensystemen

Additivität $P(A \cup B) = P(A) + P(B)$

(subjektive) Bayesianer: jede Art von Unsicherheit als (subjektive) Wahrscheinlichkeitsverteilung darstellbar

Probleme bei

- Entwicklung von Expertensystemen
- bei Experimenten in der Tradition von Ellsberg

Expertenangaben („elicitation“) führen ständig zu nicht-additiven Wahrscheinlichkeiten! (Vermischung von *risk* und *ambiguity*.)

$\Longrightarrow$ untere/obere Wahrscheinlichkeiten sind nicht-additive Maße

Modellierung von Nichtwissen

„objektive“ Bayesianer, die keine subjektive Priori verwenden möchten, benutzen eine „nichtinformative“ Priori

Probleme:

- nicht eindeutig

Modellierung von Nichtwissen

„objektive" Bayesianer, die keine subjektive Priori verwenden möchten, benutzen eine „nichtinformative" Priori

Probleme:

- nicht eindeutig
- Nichtwissen über Asymmetrie $\neq$ Wissen um Symmetrie

Modellierung von Nichtwissen

„objektive“ Bayesianer, die keine subjektive Priori verwenden möchten, benutzen eine „nichtinformative“ Priori

Probleme:

- nicht eindeutig
- Nichtwissen über Asymmetrie $\neq$ Wissen um Symmetrie
- *eine* Priori kann kein Nichtwissen über einen Parameter ausdrücken

Modellierung von Nichtwissen

„objektive" Bayesianer, die keine subjektive Priori verwenden möchten, benutzen eine „nichtinformative" Priori

Probleme:

- nicht eindeutig
- Nichtwissen über Asymmetrie $\neq$ Wissen um Symmetrie
- *eine* Priori kann kein Nichtwissen über einen Parameter ausdrücken

→ eine Menge von Verteilungen prinzipiell schon:
$P(\theta \in A) = (0, 1)$

Prior-data conflict

= Situationen, in denen *informatives Priori-Wissen* und *gesicherte Daten* (keine Ausreißer, etc.) in Konflikt stehen

Beispiel: (Walley 1991)

$$\begin{array}{r|ccl} \text{Daten:} & X & \sim & \mathrm{N}(\vartheta, 1) \\ \text{konjugierte priori:} & \vartheta & \sim & \mathrm{N}(\mu, 1) \\ \hline \text{Posteriori:} & \vartheta \mid x & \sim & \mathrm{N}\left(\frac{\mu + x}{2}, \frac{1}{2}\right) \end{array}$$

Prior-data conflict

= Situationen, in denen *informatives Priori-Wissen* und *gesicherte Daten* (keine Ausreißer, etc.) in Konflikt stehen

Beispiel: (Walley 1991)

$$
\begin{array}{r|ccl}
\text{Daten:} & X & \sim & \mathrm{N}(\vartheta, 1) \\
\text{konjugierte priori:} & \vartheta & \sim & \mathrm{N}(\mu, 1) \\
\hline
\text{Posteriori:} & \vartheta \mid x & \sim & \mathrm{N}\left(\dfrac{\mu + x}{2}, \dfrac{1}{2}\right)
\end{array}
$$

- Fall (i): $\mu = 5.5$, $x = 6.5$ $\implies \vartheta \sim \mathrm{N}(6, \frac{1}{2})$
- Fall (ii): $\mu = 3.5$, $x = 8.5$ $\implies \vartheta \sim \mathrm{N}(6, \frac{1}{2})$

Prior-data conflict

= Situationen, in denen *informatives Priori-Wissen* und *gesicherte Daten* (keine Ausreißer, etc.) in Konflikt stehen

Beispiel: (Walley 1991)

$$\begin{array}{r|ccl} \text{Daten:} & X & \sim & \mathrm{N}(\vartheta, 1) \\ \text{konjugierte priori:} & \vartheta & \sim & \mathrm{N}(\mu, 1) \\ \hline \text{Posteriori:} & \vartheta \mid x & \sim & \mathrm{N}\left(\frac{\mu + x}{2}, \frac{1}{2}\right) \end{array}$$

- Fall (i): $\mu = 5.5$, $x = 6.5$ $\implies \vartheta \sim \mathrm{N}(6, \frac{1}{2})$
- Fall (ii): $\mu = 3.5$, $x = 8.5$ $\implies \vartheta \sim \mathrm{N}(6, \frac{1}{2})$

➡ mit Mengen von Prioris Fall (i) und (ii) unterscheidbar (Fall (i): kleine Menge, Fall (ii): große Menge)

- Generalisierte Bayes-Inferenz: Idee
- Warum Intervallwahrscheinlichkeit?
  - Unsicherheit ist mehrdimensional
  - Additivität
  - Modellierung von Nichtwissen
  - Prior-data conflict
- **Generalisierte Bayes-Inferenz mit LUCK-models**
  - **LUCK-models**
  - **iLUCK-models**
  - **generalized iLUCK-models**
- Zusammenfassung

Generalisierte Bayes-Inferenz mit LUCK-models: Konzept

Wie Mengen von Prioris sinnvoll beschreiben?

- benutze **konjugierte** Prioris
  - ↳ nur Parameter der Priori aufdatieren

Generalisierte Bayes-Inferenz mit LUCK-models: Konzept

Wie Mengen von Prioris sinnvoll beschreiben?

- benutze **konjugierte** Prioris
  - ↦ nur Parameter der Priori aufdatieren
- statt einem Parameter eine (konvexe) Menge von Priori-Parametern
  - ↦ wenn Aufdatierung linear ist, dann ist die Menge der Posteriori-Parameter einfach ermittelbar!

Generalisierte Bayes-Inferenz mit LUCK-models: Konzept

Wie Mengen von Prioris sinnvoll beschreiben?

- benutze **konjugierte** Prioris
  - ↦ nur Parameter der Priori aufdatieren
- statt einem Parameter eine (konvexe) Menge von Priori-Parametern
  - ↦ wenn Aufdatierung linear ist, dann ist die Menge der Posteriori-Parameter einfach ermittelbar!
- Idee von Quaeghebeur & de Cooman (2005): allgemeine Konstruktion von konjugierten Prioris führt automatisch zu linear aufdatierten Parametern!

Konstruktion von konjugierten Prioris:

$X \overset{iid}{\sim}$ *linear, canonical exponential familiy*, d.h.

$$p(x \mid \theta) \propto \exp\left\{\langle \psi, \tau(x) \rangle - n\mathbf{b}(\psi)\right\} \qquad \left[\psi \text{ Transformation von } \theta\right]$$

Konstruktion von konjugierten Prioris:

$X \overset{iid}{\sim}$ *linear, canonical exponential familiy*, d.h.

$$p(x \mid \theta) \propto \exp\left\{\langle\psi, \tau(x)\rangle - n\mathbf{b}(\psi)\right\} \qquad \Big[\psi \text{ Transformation von } \theta\Big]$$

→ konjugierte Priori:

$$p(\theta) \propto \exp\left\{n^{(0)}\left[\langle\psi, y^{(0)}\rangle - \mathbf{b}(\psi)\right]\right\}$$

Konstruktion von konjugierten Prioris:

$X \overset{iid}{\sim}$ *linear, canonical exponential familiy*, d.h.

$$p(x \mid \theta) \propto \exp\left\{\langle\psi, \tau(x)\rangle - n\mathbf{b}(\psi)\right\} \qquad \left[\psi \text{ Transformation von } \theta\right]$$

➡ konjugierte Priori:

$$p(\theta) \propto \exp\left\{n^{(0)}\left[\langle\psi, y^{(0)}\rangle - \mathbf{b}(\psi)\right]\right\}$$

➡ (konjugierte) Posteriori:

$$p(\theta \mid x) \propto \exp\left\{n^{(1)}\left[\langle\psi, y^{(1)}\rangle - \mathbf{b}(\psi)\right]\right\}$$

wobei $y^{(1)} = \dfrac{n^{(0)}y^{(0)} + \tau(x)}{n^{(0)} + n}$ und $n^{(1)} = n^{(0)} + n\,.$

➥ jedes $\big(p(\theta), p(\theta \mid x)\big)$, das so aufdatiert wird = LUCK-model

LUCK-models: Interpretation von $y^{(0)}$ und $n^{(0)}$

$\boxed{y^{(0)}\text{: “main prior parameter”}}$

- für Stichproben aus $\mathrm{N}(\mu, 1)$ ist $p(\mu)$ eine $\mathrm{N}(y^{(0)}, \frac{1}{n^{(0)}})$
- für Stichproben aus $\mathrm{M}(\boldsymbol{\theta})$ ist $p(\boldsymbol{\theta})$ eine $\mathrm{Dir}(n^{(0)}, y^{(0)})$
  ($y_j^{(0)} = t_j \hat{=}$ Priori-W'keit für Kategorie j, $n^{(0)} = s$; $s \cdot t_j = \alpha_j$)

LUCK-models: Interpretation von $y^{(0)}$ und $n^{(0)}$

$y^{(0)}$: "main prior parameter"

- für Stichproben aus $\mathrm{N}(\mu, 1)$ ist $p(\mu)$ eine $\mathrm{N}(y^{(0)}, \frac{1}{n^{(0)}})$
- für Stichproben aus $\mathrm{M}(\boldsymbol{\theta})$ ist $p(\boldsymbol{\theta})$ eine $\mathrm{Dir}(n^{(0)}, y^{(0)})$
  ($y_j^{(0)} = t_j \mathrel{\hat{=}}$ Priori-W'keit für Kategorie j, $n^{(0)} = s$; $s \cdot t_j = \alpha_j$)

$n^{(0)}$: "prior strength" oder "pseudocounts"

mit $\tilde{\tau}(x) =: \frac{1}{n}\tau(x)$: $\quad [\ \tau(x) = \sum_{i=1}^{n} \tau(x_i)\ ]$

$$y^{(1)} = \frac{n^{(0)}}{n^{(0)} + n} \cdot y^{(0)} + \frac{n}{n^{(0)} + n} \cdot \tilde{\tau}(x)\,.$$

Mengen von LUCK-models – iLUCK-model

iLUCK-model: variiere $y^{(0)}$ in $\mathcal{Y}^{(0)}$ [$\mathcal{Y}^{(0)}$ konvex] $\Longleftrightarrow$ erlaubt unscharfes Wissen über den zentralen Parameter $y^{(0)}$

→ Priori-„credal set" enthält *alle finite konvexen Mischungen* von $p(\theta)$s mit $y^{(0)} \in \mathcal{Y}^{(0)}$

Mengen von LUCK-models – iLUCK-model

iLUCK-model: variiere $y^{(0)}$ in $\mathcal{Y}^{(0)}$ [$\mathcal{Y}^{(0)}$ konvex] $\Longleftrightarrow$ erlaubt unscharfes Wissen über den zentralen Parameter $y^{(0)}$

- Priori-„credal set" enthält *alle finite konvexen Mischungen* von $p(\theta)$s mit $y^{(0)} \in \mathcal{Y}^{(0)}$
- Posteriori-„credal set" einfach berechenbar: *alle finite konvexen Mischungen* von $p(\theta \mid x)$s mit

$$y^{(1)} \in \mathcal{Y}^{(1)} = \frac{n^{(0)}}{n^{(0)} + n} \cdot \mathcal{Y}^{(0)} + \frac{n}{n^{(0)} + n} \cdot \tilde{\tau}(x)$$

Mengen von LUCK-models – iLUCK-model

iLUCK-model: variiere $y^{(0)}$ in $\mathcal{Y}^{(0)}$ [$\mathcal{Y}^{(0)}$ konvex] $\Longleftrightarrow$ erlaubt unscharfes Wissen über den zentralen Parameter $y^{(0)}$

- Priori-„credal set“ enthält *alle finite konvexen Mischungen* von $p(\theta)$s mit $y^{(0)} \in \mathcal{Y}^{(0)}$
- Posteriori-„credal set“ einfach berechenbar: *alle finite konvexen Mischungen* von $p(\theta \mid x)$s mit

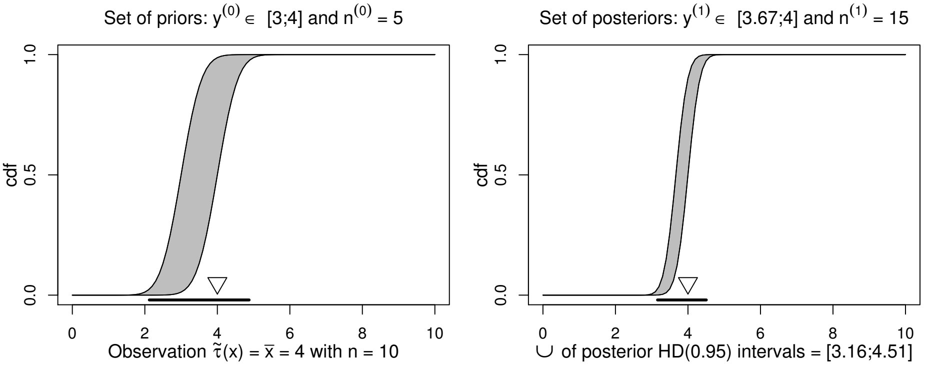
$$y^{(1)} \in \mathcal{Y}^{(1)} = \frac{n^{(0)}}{n^{(0)} + n} \cdot \mathcal{Y}^{(0)} + \frac{n}{n^{(0)} + n} \cdot \tilde{\tau}(x)$$

Aber: $\mathcal{Y}^{(0)}$ muss endliche Grenzen haben

- „perfektes“ Nichtwissen hier nicht darstellbar

iLUCK-model: Beispiel $X \sim \mathrm{N}(\mu, 1)$

$\rightarrow \mu \sim \mathrm{N}(y^{(0)}, \frac{1}{n^{(0)}})$



iLUCK-model: Beispiel $X \sim \mathrm{M}(\theta)$

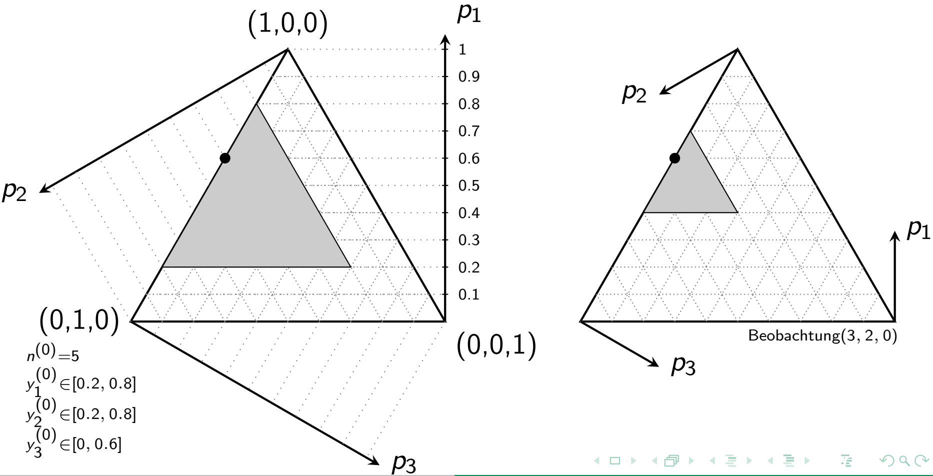
→ $\theta \sim \mathrm{Dir}(n^{(0)}, y^{(0)})$ ↔ **Imprecise Dirichlet Model (IDM)**

→ Walley (*JRSS*, 1991), Bernard (*IJAR Special Issue*, 2009)

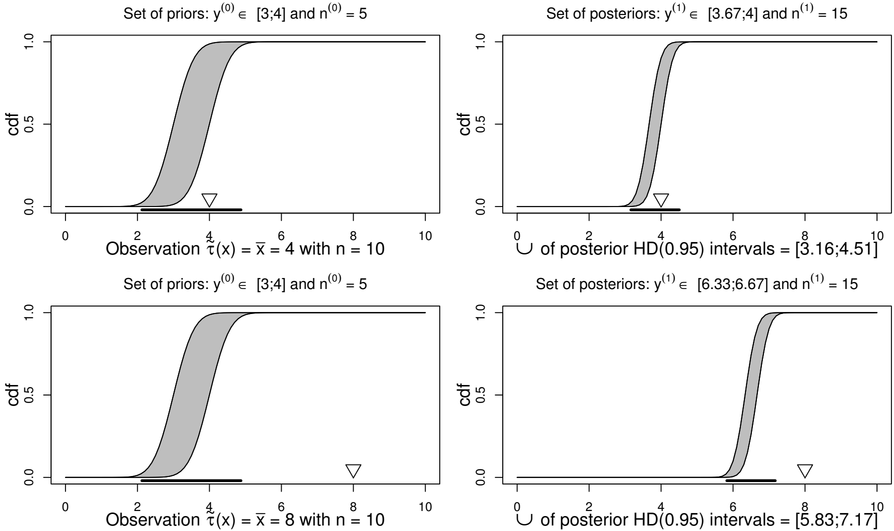
iLUCK-model: Beispiel $X \sim \mathrm{M}(\theta)$

→ $\theta \sim \mathrm{Dir}(n^{(0)}, y^{(0)})$ ↔ **Imprecise Dirichlet Model (IDM)**

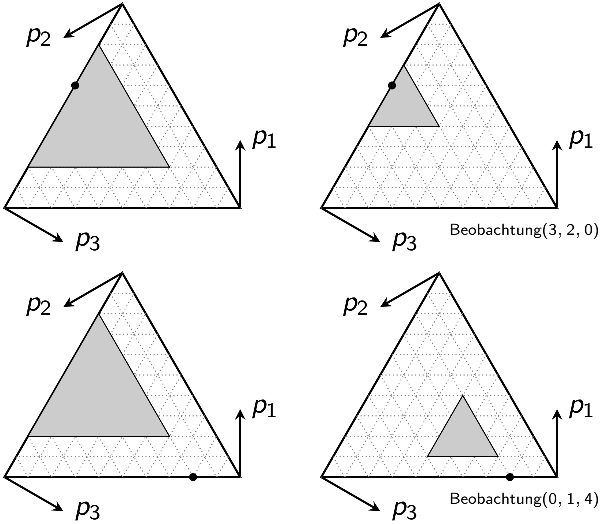
→ Walley (*JRSS*, 1991), Bernard (*IJAR Special Issue*, 2009)



Probleme bei *prior-data conflict*: $X \sim \mathrm{N}(\mu, 1)$



Probleme bei *prior-data conflict*: $X \sim \mathrm{M}(\theta)$ **(IDM)**



iLUCK-model — Posteriori-Unschärfe in y

$$\overline{y}^{(1)} - \underline{y}^{(1)} = \frac{n^{(0)}\overline{y}^{(0)} + \tau(x)}{n^{(0)} + n} - \frac{n^{(0)}\underline{y}^{(0)} + \tau(x)}{n^{(0)} + n}$$

$$= \frac{n^{(0)}(\overline{y}^{(0)} - \underline{y}^{(0)})}{n^{(0)} + n}$$

...hängt nicht von der suffizienten Statistik $\tau(x)$ ab!

iLUCK-model — Posteriori-Unschärfe in y

$$\overline{y}^{(1)} - \underline{y}^{(1)} = \frac{n^{(0)}\overline{y}^{(0)} + \tau(x)}{n^{(0)} + n} - \frac{n^{(0)}\underline{y}^{(0)} + \tau(x)}{n^{(0)} + n}$$

$$= \frac{n^{(0)}(\overline{y}^{(0)} - \underline{y}^{(0)})}{n^{(0)} + n}$$

…hängt nicht von der suffizienten Statistik $\tau(x)$ ab!

Für jede Stichprobe mit Umfang n reduziert sich die Posteriori-Unschärfe im gleichen Umfang!

generalized iLUCK-models (Walter & Augustin, 2009):

variiere $y^{(0)}$ in $\mathcal{Y}^{(0)}$ **und** $n^{(0)}$ in $\mathcal{N}^{(0)}$ $\Longleftrightarrow$
flexiblere Gewichtung der Priori-Information $\mathcal{Y}^{(0)}$ und Stichproben-Information $\tilde{\tau}(x)$ in

$$y^{(1)} \in \mathcal{Y}^{(1)} = \frac{n^{(0)}}{n^{(0)} + n} \cdot \mathcal{Y}^{(0)} + \frac{n}{n^{(0)} + n} \cdot \tilde{\tau}(x)$$

➜ Priori-„credal set“ enthält *alle finiten konvexen Mischungen* von $p(\theta)$s mit $y^{(0)} \in \mathcal{Y}^{(0)}$ **und** $n^{(0)} \in \mathcal{N}^{(0)}$

generalized iLUCK-models (Walter & Augustin, 2009):

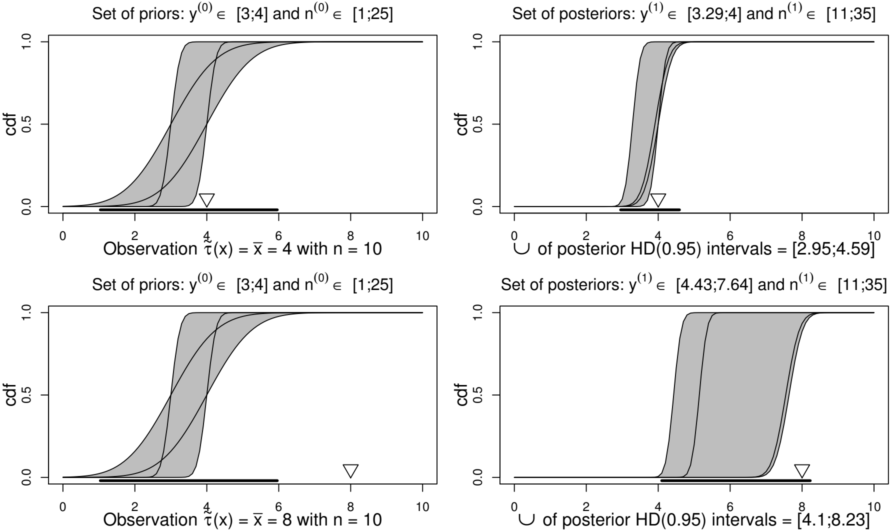
variiere $y^{(0)}$ in $\mathcal{Y}^{(0)}$ **und** $n^{(0)}$ in $\mathcal{N}^{(0)}$ $\iff$
flexiblere Gewichtung der Priori-Information $\mathcal{Y}^{(0)}$ und Stichproben-Information $\tilde{\tau}(x)$ in

$$y^{(1)} \in \mathcal{Y}^{(1)} = \frac{n^{(0)}}{n^{(0)} + n} \cdot \mathcal{Y}^{(0)} + \frac{n}{n^{(0)} + n} \cdot \tilde{\tau}(x)$$

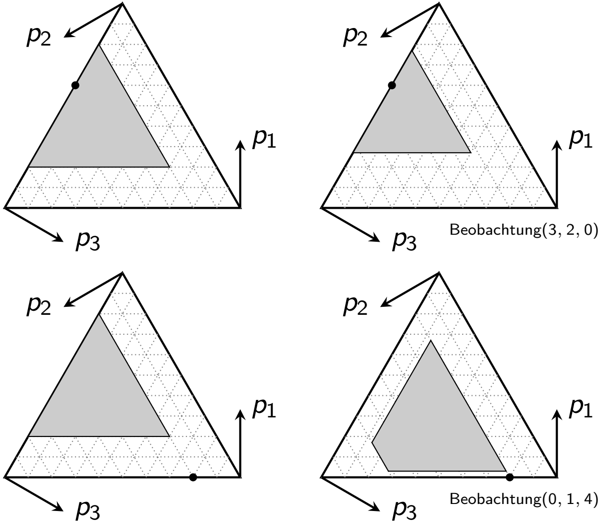
- Priori-„credal set“ enthält *alle finiten konvexen Mischungen* von $p(\theta)$s mit $y^{(0)} \in \mathcal{Y}^{(0)}$ **und** $n^{(0)} \in \mathcal{N}^{(0)}$
- Posteriori-„credal set“ immer noch relativ einfach berechenbar: *alle finiten konvexen Mischungen* von $p(\theta \mid x)$s mit

$$\left\{ \left(n^{(1)}, y^{(1)} \right) \middle| \, n^{(1)} = n^{(0)} + n, \, y^{(1)} = \frac{n^{(0)} y^{(0)} + \tau(x)}{n^{(0)} + n}, \, n^{(0)} \in \mathcal{N}^{(0)}, \, y^{(0)} \in \mathcal{Y}^{(0)} \right\}$$

generalized iLUCK-model: Beispiel $X \sim N(\mu, 1)$



generalized iLUCK-model: Beispiel $X \sim \mathrm{M}(\theta)$



Zusammenfassung

- iLUCK-models bieten ein allgemeines (Exponentialfamilie!), handhabbares und leistungsfähiges Kalkül für Bayes-Inferenz mit Mengen von Prioris.
  → *ambiguity* im Priori-Wissen wird explizit ins Modell einbezogen!

Zusammenfassung

- iLUCK-models bieten ein allgemeines (Exponentialfamilie!), handhabbares und leistungsfähiges Kalkül für Bayes-Inferenz mit Mengen von Prioris.
  → *ambiguity* im Priori-Wissen wird explizit ins Modell einbezogen!
- mit generalized iLUCK-models werden iLUCK-models so erweitert, dass auch *prior-data conflict* berücksichtigt wird.

Zusammenfassung

- iLUCK-models bieten ein allgemeines (Exponentialfamilie!), handhabbares und leistungsfähiges Kalkül für Bayes-Inferenz mit Mengen von Prioris.
  → *ambiguity* im Priori-Wissen wird explizit ins Modell einbezogen!
- mit generalized iLUCK-models werden iLUCK-models so erweitert, dass auch *prior-data conflict* berücksichtigt wird.

Walter, G. , Augustin, T.: Imprecision and prior-data conflict in generalized Bayesian inference. *Journal of Statistical Theory and Practice*, 2009.