Generalized Bayesian Inference and Prior-data Conflict

Gero Walter

Department of Statistics
Ludwig-Maximilians-University Munich

May 16th, 2008

Institut für Statistik münchen

Introduction — What is Prior-data Conflict?

A name for situations in which *informative prior beliefs* and *trusted data* are in conflict with each other.

Example: (Walley 1991)

$$
\begin{array}{r|ccl}
\text{Data:} & X & \sim & \mathrm{N}(\vartheta, 1) \\
\text{Conjugate prior:} & \vartheta & \sim & \mathrm{N}(\mu, 1) \\
\hline
\text{Posterior:} & \vartheta \mid x & \sim & \mathrm{N}\left(\dfrac{\mu + x}{2}, \dfrac{1}{2}\right)
\end{array}
$$

Introduction — What is Prior-data Conflict?

A name for situations in which *informative prior beliefs* and *trusted data* are in conflict with each other.

Example: (Walley 1991)

Data:	X	$\sim$	$\mathrm{N}(\vartheta, 1)$
Conjugate prior:	ϑ	$\sim$	$\mathrm{N}(\mu, 1)$
Posterior:	$\vartheta \mid x$	$\sim$	$\mathrm{N}\left(\frac{\mu + x}{2}, \frac{1}{2}\right)$

Consider

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

Introduction — Imprecise Probabilities

IP models promise to solve these problems:
they can reflect the amount of knowledge they stand for.

➡ Multidimensional nature of uncertainty:
stochastic uncertainty $\longleftrightarrow$ ambiguity in probability assignments

Introduction — Imprecise Probabilities

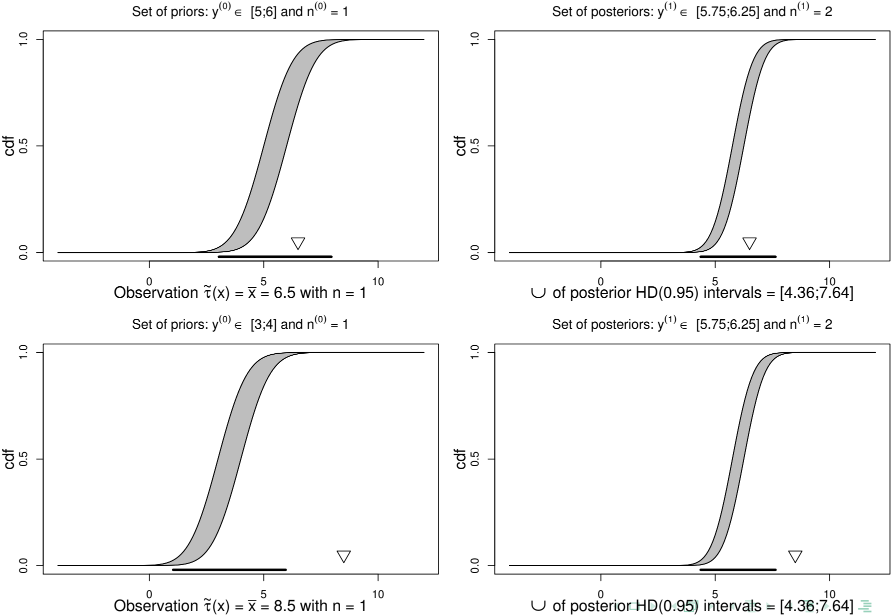
IP models promise to solve these problems:
they can reflect the amount of knowledge they stand for.

→ Multidimensional nature of uncertainty:
stochastic uncertainty $\longleftrightarrow$ ambiguity in probability assignments

Bad news:

- IDM under prior information or
- models in the framework of Quaeghebeur & de Cooman or
- iLUCK-models

ignore prior-data conflict just like precise models!


Set of priors: y^(0) ∈ [5;6] and n^(0) = 1
cdf
1.0
0.5
0.0
0
5
10
Observation τ̃(x) = x̄ = 6.5 with n = 1
Set of posteriors: y^(1) ∈ [5.75;6.25] and n^(1) = 2
cdf
∪ of posterior HD(0.95) intervals = [4.36;7.64]
Set of priors: y^(0) ∈ [3;4] and n^(0) = 1
cdf
Observation τ̃(x) = x̄ = 8.5 with n = 1
Set of posteriors: y^(1) ∈ [5.75;6.25] and n^(1) = 2
cdf
∪ of posterior HD(0.95) intervals = [4.36;7.64]

Recalling (i)LUCK-models — Construction of Conjugate Priors

➜ Conjugate priors can be constructed as follows:

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

➜ Updating yields then Posterior:

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

$$y^{(1)} = \frac{n^{(0)} y^{(0)} + \tau(x)}{n^{(0)} + n} \qquad \text{and} \qquad n^{(1)} = n^{(0)} + n\,.$$

(i)LUCK-models — Interpretation of Parameters

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

- For samples from a $\mathrm{N}(\mu, 1)$, $p(\mu)$ is a $\mathrm{N}(y^{(0)}, \frac{1}{n^{(0)}})$
- For samples from a $\mathrm{M}(\boldsymbol{\theta})$, $p(\boldsymbol{\theta})$ is a $\mathrm{Dir}(n^{(0)}, y^{(0)})$ ($y_j^{(0)} = t_j \hat{=}$ prior probability for class j, $n^{(0)} = s$)

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

With $\tau(x) = \sum_{i=1}^{n} \tau(x_i)$ and $\tilde{\tau}(x) =: \frac{1}{n} \sum_{i=1}^{n} \tau(x_i)$:

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

Method of Quaeghebeur and de Cooman

Construction of *imprecise* prior: **Vary** $y^{(0)}$ **in a convex set** $\mathcal{Y}^{(0)}$

→ Prior credal set contains *all convex mixtures* of distributions with $y^{(0)} \in \mathcal{Y}^{(0)}$

→ Set $\mathcal{Y}^{(1)}$ defining the imprecise posterior is easily derived by

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

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

Linearity: Vertices of $\mathcal{Y}^{(0)} \longrightarrow$ Vertices of $\mathcal{Y}^{(1)}$

LUCK-models & iLUCK-models

(First) Generalization:

➡ LUCK-models (Linearly Updated Conjugate prior Knowledge):

Prior $p(\vartheta)$ and posterior $p(\vartheta \mid x)$ such that

$$p(\vartheta) \propto \exp\left\{ n^{(0)} \left[\langle \psi, y^{(0)} \rangle - \mathbf{b}(\psi) \right] \right\} \quad \text{and}$$
$$p(\vartheta \mid x) \propto \exp\left\{ n^{(1)} \left[\langle \psi, y^{(1)} \rangle - \mathbf{b}(\psi) \right] \right\}, \quad \text{where}$$

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

➡ iLUCK-models (imprecise LUCK-models):

Vary $y^{(0)}$ in a set $\mathcal{Y}^{(0)}$ (method of Quaeghebeur and de Cooman)

iLUCK-model — Main Parameter Posterior Imprecision

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

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

. . . does not depend on the sample statistic!

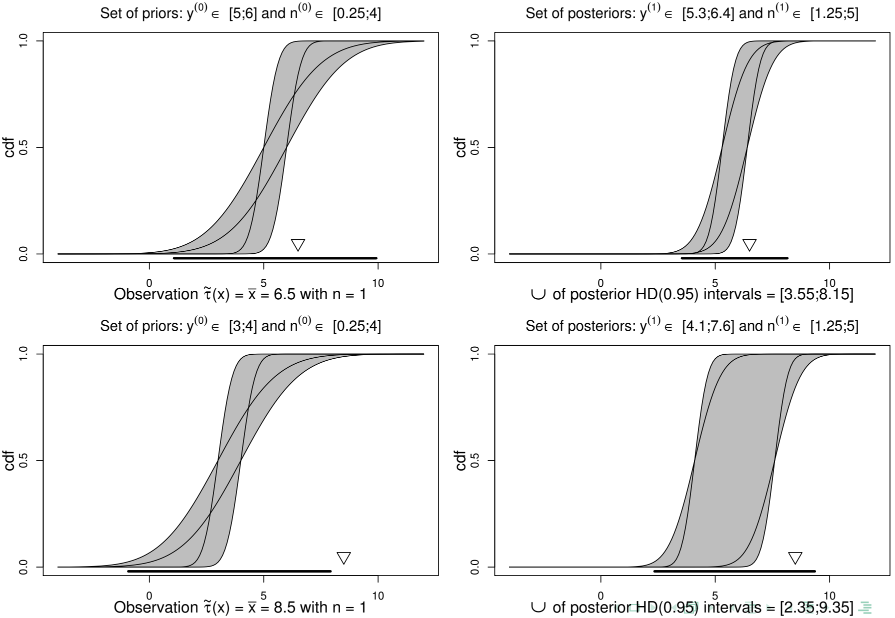
iLUCK-model — Main Parameter Posterior Imprecision

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

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

. . . does not depend on the sample statistic!

➥ ⚠ For any sample of size n, posterior imprecision is reduced by the same amount! ⚠



Set of priors: y(0) ∈ [5;6] and n(0) ∈ [0.25;4]
Observation τ̃(x) = x̄ = 6.5 with n = 1
Set of posteriors: y(1) ∈ [5.3;6.4] and n(1) ∈ [1.25;5]
∪ of posterior HD(0.95) intervals = [3.55;8.15]
Set of priors: y(0) ∈ [3;4] and n(0) ∈ [0.25;4]
Observation τ̃(x) = x̄ = 8.5 with n = 1
Set of posteriors: y(1) ∈ [4.1;7.6] and n(1) ∈ [1.25;5]
∪ of posterior HD(0.95) intervals = [2.35;9.35]
cdf
0.0
0.5
1.0
0
5
10

Generalized iLUCK-models — Formal Definition

Definition (Generalized iLUCK-models)

Consider a set of LUCK-models $\big(p(\vartheta), p(\vartheta \,|\, x)\big)$ that is produced by $y^{(0)}$ varying in some set $\mathcal{Y}^{(0)} \subset \mathcal{Y}$ and, in addition, $n^{(0)}$ varying in a set $\mathcal{N}^{(0)} \subset \mathbb{R}^+$. Let furthermore again the credal sets $\mathcal{P}$ and $\mathcal{P}_{|x}$ consist of all convex mixtures obtained from this variation of $p(\vartheta)$ and $p(\vartheta \,|\, x)$. Then $\big(\mathcal{P}, \mathcal{P}_{|x}\big)$ is called the corresponding *generalized iLUCK-model* based on $\mathcal{Y}^{(0)}$ and $\mathcal{N}^{(0)}$.

$$\mathcal{Y}^{(0)} \times \mathcal{N}^{(0)} \quad \longrightarrow \quad \mathcal{Y}^{(1)} \times \mathcal{N}^{(1)}$$

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

Generalized iLUCK-models — Update Step

$$\underline{y}^{(1)} = \begin{cases} \dfrac{\overline{n}^{(0)}\underline{y}^{(0)} + \tau(x)}{\overline{n}^{(0)} + n} & \text{if} \quad \tilde{\tau}(x) \geq \underline{y}^{(0)} \\[2ex] \dfrac{\underline{n}^{(0)}\underline{y}^{(0)} + \tau(x)}{\underline{n}^{(0)} + n} & \text{if} \quad \tilde{\tau}(x) < \underline{y}^{(0)} \end{cases}$$

$\iff$ prior-data conflict (second case)

$$\overline{y}^{(1)} = \begin{cases} \dfrac{\overline{n}^{(0)}\overline{y}^{(0)} + \tau(x)}{\overline{n}^{(0)} + n} & \text{if} \quad \tilde{\tau}(x) \leq \overline{y}^{(0)} \\[2ex] \dfrac{\underline{n}^{(0)}\overline{y}^{(0)} + \tau(x)}{\underline{n}^{(0)} + n} & \text{if} \quad \tilde{\tau}(x) > \overline{y}^{(0)} \end{cases}$$

$\iff$ prior-data conflict (second case)

Generalized iLUCK-models — Update Step

	$\tilde{\tau}(x) < \underline{y}^{(0)}$	$\underline{y}^{(0)} \leq \tilde{\tau}(x) \leq \overline{y}^{(0)}$	$\tilde{\tau}(x) > \overline{y}^{(0)}$
Calculation of $\underline{y}^{(1)}$ via	$\underline{n}^{(0)}$	$\overline{n}^{(0)}$	$\overline{n}^{(0)}$
Calculation of $\overline{y}^{(1)}$ via	$\overline{n}^{(0)}$	$\overline{n}^{(0)}$	$\underline{n}^{(0)}$

Generalized iLUCK-models — Main Parameter Posterior Imprecision

$$\overline{y}^{(1)} - \underline{y}^{(1)} = \frac{\overline{n}^{(0)}(\overline{y}^{(0)} - \underline{y}^{(0)})}{\overline{n}^{(0)} + n} + \Delta\left(\tilde{\tau}(x); \underline{y}^{(0)}, \overline{y}^{(0)}\right) \frac{n(\overline{n}^{(0)} - \underline{n}^{(0)})}{(\overline{n}^{(0)} + n)(\underline{n}^{(0)} + n)},$$

where

$$\Delta\left(\tilde{\tau}(x); \underline{y}^{(0)}, \overline{y}^{(0)}\right) = \inf\left\{\left|\tilde{\tau}(x) - y^{(0)}\right| : \underline{y}^{(0)} \le y^{(0)} \le \overline{y}^{(0)}\right\}$$

is the distance of observation $\tilde{\tau}(x)$ to prior interval $\left[\underline{y}^{(0)}; \overline{y}^{(0)}\right]$.

- $\Delta(\ ) = 0 \iff$ same amount of imprecision as in iLUCK-models if $\overline{n}^{(0)} = \underline{n}^{(0)}$
- $\Delta(\ ) > 0 \iff$ prior-data conflict: wider interval!

Why then not choose $\overline{n}^{(0)} = n^{(0)}$?

The factor to $\Delta(\ )$ gets maximal if $n = \sqrt{\underline{n}^{(0)}\overline{n}^{(0)}}$.

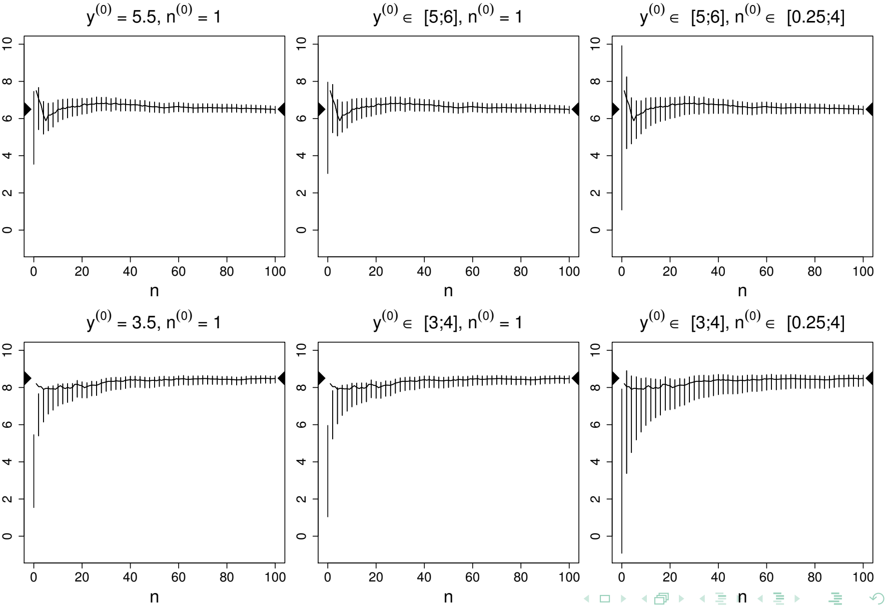
$\iff \overline{y}^{(1)} - \underline{y}^{(1)}$ maximal for fixed $\Delta(\ )$.

$\iff$ Same weight on the prior and on the sample.

$\iff \sqrt{\underline{n}^{(0)}\overline{n}^{(0)}}$ is the 'global' prior strength

to be compared to $n^{(0)}$.

For $n \longrightarrow \infty$, it doesn't matter, as $\dfrac{\overline{n}^{(1)} - \underline{n}^{(1)}}{\overline{n}^{(1)}} \longrightarrow 0$.


$y^{(0)} = 5.5$, $n^{(0)} = 1$
$y^{(0)} \in [5;6]$, $n^{(0)} = 1$
$y^{(0)} \in [5;6]$, $n^{(0)} \in [0.25;4]$
$y^{(0)} = 3.5$, $n^{(0)} = 1$
$y^{(0)} \in [3;4]$, $n^{(0)} = 1$
$y^{(0)} \in [3;4]$, $n^{(0)} \in [0.25;4]$
0
2
4
6
8
10
0
20
40
60
80
100
n

Concluding Remarks

informative prior beliefs and *trusted data*
& updating by generalized Bayes' rule:
→ use generalized iLUCK-models

See manuscript "Imprecision and Prior-data Conflict in Generalized Bayesian Inference" (tba for JSTP Special Issue on Imprecision).

⚠ generalized iLUCK-models might not do well when the degree of conflict is very strong ⚠

→ alternative learning rules?