Prior-Data Conflict and Generalized Bayesian Inference

Gero Walter

Institut für Statistik
Ludwig-Maximilians-Universität München

March 31st, 2010

Generalized Bayesian Inference – General Idea

Bayesian Inference on some parameter θ:

prior knowledge on θ + data x → updated knowledge on θ

Generalized Bayesian Inference – General Idea

Bayesian Inference on some parameter θ:

prior knowledge on θ	+	data x	→	updated knowledge on θ
prior distribution $p(\theta)$	+	likelihood $f(x \mid \theta)$	→	posterior distribution $p(\theta \mid x)$

Generalized Bayesian Inference – General Idea

Bayesian Inference on some parameter θ:

prior knowledge on θ	+	data x	→	updated knowledge on θ
prior distribution $p(\theta)$	+	likelihood $f(x \mid \theta)$	→	posterior distribution $p(\theta \mid x)$
set of priors	+	likelihood	→	**set of** posteriors

Generalized Bayesian Inference – General Idea

Bayesian Inference on some parameter θ:

prior knowledge on θ	+	data x	→	updated knowledge on θ
prior distribution $p(\theta)$	+	likelihood $f(x \mid \theta)$	→	posterior distribution $p(\theta \mid x)$
set of priors	+	likelihood	→	**set of** posteriors

Tractability: use **conjugate** priors, i.e.
choose $p(\theta)$ such that $p(\theta \mid x)$ is from the same parametric class
→ update only parameters!

Prior-Data Conflict

Prior-Data Conflict $\hat{=}$ situation in which...

- ...*informative prior beliefs* and *trusted data* (sampling model correct, no outliers, etc.) are in conflict
- "... the prior [places] its mass primarily on distributions in the sampling model for which the observed data is surprising." (Evans & Moshonov, 2006)

Dirichlet-Multinomial-Model

Data :	$\mathbf{k}$	$\sim$	$\mathrm{M}(\boldsymbol{\theta})$	$(\sum k_j = n)$
conjugate prior:	$\boldsymbol{\theta}$	$\sim$	$\mathrm{Dir}(\boldsymbol{\alpha})$	$(\sum \theta_j = 1)$
posterior:	$\boldsymbol{\theta} \mid \mathbf{k}$	$\sim$	$\mathrm{Dir}(\boldsymbol{\alpha} + \mathbf{k})$	

$$\mathbb{E}[\theta_j] = \frac{\alpha_j}{\sum \alpha_i}$$

$$\mathbb{V}(\theta_j) = \frac{\alpha_j(\sum \alpha_i - \alpha_j)}{(\sum \alpha_i)^2(\sum \alpha_i + 1)} = \frac{\mathbb{E}[\theta_j](1 - \mathbb{E}[\theta_j])}{\sum \alpha_i + 1}$$

Dirichlet-Multinomial-Model — Alternative Parameterisation

$$\frac{\alpha_j}{\sum \alpha_i} = \mathbb{E}[\theta_j] =: y_j^{(0)} \qquad \sum \alpha_i =: n^{(0)}$$

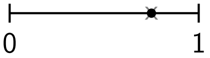
Data :	$\mathbf{k}$	$\sim$	$\mathrm{M}(\boldsymbol{\theta})$
conjugate prior:	$\boldsymbol{\theta}$	$\sim$	$\mathrm{Dir}(n^{(0)}, \mathbf{y}^{(0)})$
posterior:	$\boldsymbol{\theta} \mid \mathbf{k}$	$\sim$	$\mathrm{Dir}(n^{(1)}, \mathbf{y}^{(1)})$

$$y_j^{(1)} = \frac{n^{(0)}}{n^{(0)} + n} \cdot y_j^{(0)} + \frac{n}{n^{(0)} + n} \cdot \frac{k_j}{n}, \qquad n^{(1)} = n^{(0)} + n$$

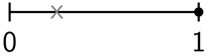
$$\mathbb{V}(\theta_j) = \frac{y_j^{(0)}(1 - y_j^{(0)})}{n^{(0)} + 1}$$

Prior-Data Conflict — Simple Example

Case (i): $y_j^{(0)} = 0.75\,,$ $(n^{(0)} = 8)$ $\quad k_j/n = 0.75$ $(n = 16)$

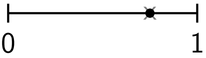


Case (ii): $y_j^{(0)} = 0.25\,,$ $(n^{(0)} = 8)$ $\quad k_j/n = 1$ $(n = 16)$

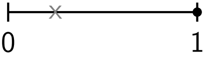


Prior-Data Conflict — Simple Example

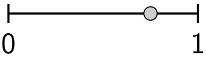
Case (i): $y_j^{(0)} = 0.75\,,$ $(n^{(0)} = 8)$ $\quad k_j/n = 0.75$ $(n = 16)$



Case (ii): $y_j^{(0)} = 0.25\,,$ $(n^{(0)} = 8)$ $\quad k_j/n = 1$ $(n = 16)$



→ $\mathbb{E}[\theta_j \mid \mathbf{k}] = y_j^{(1)} = 0.75, \quad \mathbb{V}(\theta_j \mid \mathbf{k}) = 3/400$



$(\mathbb{V}(\theta_j) = 1/48)$

⚠ Posterior inferences do not reflect uncertainty due to unexpected observations! ⚠

Conjugate Priors

Weighted average structure is underneath *all common* conjugate priors for exponential family sampling distributions!

$X \overset{iid}{\sim}$ linear, canonical exponential family , i.e.

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

Conjugate Priors

Weighted average structure is underneath *all common* conjugate priors for exponential family sampling distributions!

$X \overset{iid}{\sim}$ linear, canonical exponential family , i.e.

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

→ conjugate prior:

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

Conjugate Priors

Weighted average structure is underneath *all common* conjugate priors for exponential family sampling distributions!

$X \overset{iid}{\sim}$ linear, canonical exponential family , i.e.

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

➡ conjugate prior:

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

➡ (conjugate) posterior:

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

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

Conjugate Priors — Interpretation of $y^{(0)}$ and $n^{(0)}$

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

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

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

- for samples from a $\mathrm{N}(\mu, 1)$, $p(\mu)$ is a $\mathrm{N}(y^{(0)}, \frac{1}{n^{(0)}})$
- for samples from a $\mathrm{Po}(\lambda)$, $p(\lambda)$ is a $\mathrm{Ga}(n^{(0)}y^{(0)}, n^{(0)})$

  ➡ $\mathbb{E}[\lambda] = y^{(0)}$, $\mathbb{V}(\lambda) = \frac{y^{(0)}}{n^{(0)}}$

Why Generalize Bayesian Inference?

Assigning a certain prior distribution on θ

$\Longleftrightarrow$ Defining a conglomerate of probability statements (on θ).

Why Generalize Bayesian Inference?

Assigning a certain prior distribution on θ

$\Longleftrightarrow$ Defining a conglomerate of probability statements (on θ).

Standard Bayesian theory lacks the ability to specifiy the degree of uncertainty in these probability statements.

Variance or stretch of a distribution for describing uncertainty?

Why Generalize Bayesian Inference?

Assigning a certain prior distribution on θ

⟷ Defining a conglomerate of probability statements (on θ).

Standard Bayesian theory lacks the ability to specifiy the degree of uncertainty in these probability statements.

Variance or stretch of a distribution for describing uncertainty?

→ Does not work in the case of prior-data conflict:
In conjugate updating, the posterior variance does not depend on the degree of prior-data conflict in most cases.

→ How to express the precision of a probability statement?

Generalized Bayesian Inference — Basic Idea

Use **set of** priors ➡ base inferences on **set of** posteriors obtained by element-wise updating

➥ numbers become intervals, e.g.

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

$$P(\theta \in A) \quad \rightarrow \quad \left[\underline{P}(\theta \in A),\, \overline{P}(\theta \in A)\right] = \left[\min P_p(\theta \in A),\, \max P_p(\theta \in A)\right]$$

Generalized Bayesian Inference — Basic Idea

Use **set of** priors → base inferences on **set of** posteriors obtained by element-wise updating

↳ numbers become intervals, e.g.

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

$$P(\theta \in A) \quad \rightarrow \quad \left[\underline{P}(\theta \in A),\, \overline{P}(\theta \in A)\right] = \left[\min P_p(\theta \in A),\, \max P_p(\theta \in A)\right]$$

Shorter intervals ↔ more precise probability statements

↳ differentiate between

- stochastic uncertainty ("risk") vs.
- non-stochastic uncertainty ("ambiguity")

Generalized Bayesian Inference — Basic Idea

Sets of distributions ⟷ Probability / Expectation Intervals
(„credal sets“)

Probability ↓ Weichselberger (2001)

Expectation Intervals ↓ Walley (1991)

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

Generalized Bayesian Inference — Basic Idea

Sets of distributions ⟷ Probability / Expectation Intervals
(„credal sets“) ↓ ↓
Weichselberger (2001) Walley (1991)

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

First Approach: so-called *iLUCK-models* (Walter & Augustin, 2009)
Dir-Mult-Model: ⟷ **I**mprecise **D**irichlet **M**odel (Walley 1996)

1. use conjugate priors as constructed by general method (prior parameters $y^{(0)}$, $n^{(0)}$) [IDM: t, s]

Generalized Bayesian Inference — Basic Idea

Sets of distributions ⟷ Probability / Expectation Intervals
(„credal sets") ↓ ↓
Weichselberger (2001) Walley (1991)

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

First Approach: so-called *iLUCK-models* (Walter & Augustin, 2009)
Dir-Mult-Model: ⟷ **I**mprecise **D**irichlet **M**odel (Walley 1996)

1. use conjugate priors as constructed by general method (prior parameters $y^{(0)}$, $n^{(0)}$) [IDM: t, s]
2. construct sets of priors via sets of parameters $y^{(0)} \in \mathcal{Y}^{(0)}$ ($n^{(0)}$ fixed) [IDM: often $t \in [0,1]$, $s = 1$ or 2]

Generalized Bayesian Inference — Basic Idea

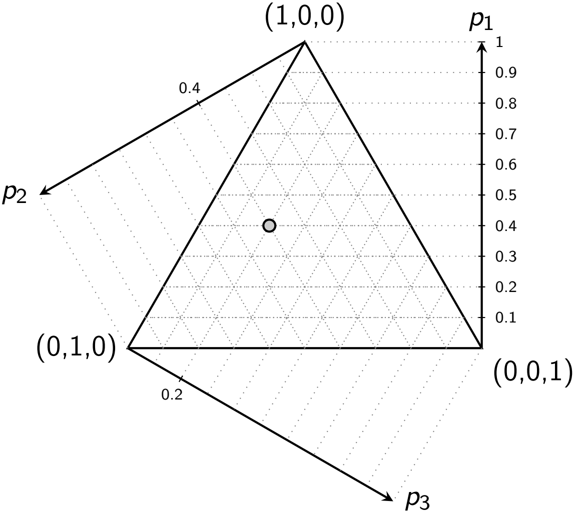
Sets of distributions ⟷ Probability / Expectation Intervals
(„credal sets“)
↓ Weichselberger (2001) ↓ Walley (1991)

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

First Approach: so-called *iLUCK-models* (Walter & Augustin, 2009)
Dir-Mult-Model: ⟷ **I**mprecise **D**irichlet **M**odel (Walley 1996)

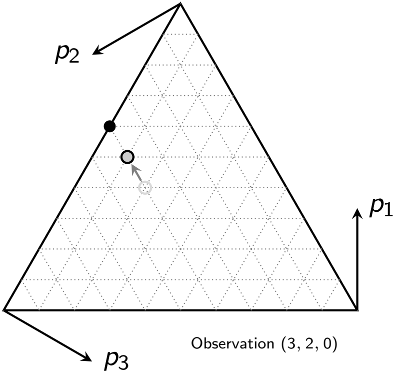
1. use conjugate priors as constructed by general method (prior parameters $y^{(0)}$, $n^{(0)}$) [IDM: t, s]
2. construct sets of priors via sets of parameters $y^{(0)} \in \mathcal{Y}^{(0)}$ ($n^{(0)}$ fixed) [IDM: often $t \in [0,1]$, $s = 1$ or 2]
3. set of posteriors $\hat{=}$ set of (element-wise) updated priors
   ↳ very easy to handle: $\mathcal{Y}^{(0)}$ updated linearly to $\mathcal{Y}^{(1)}$

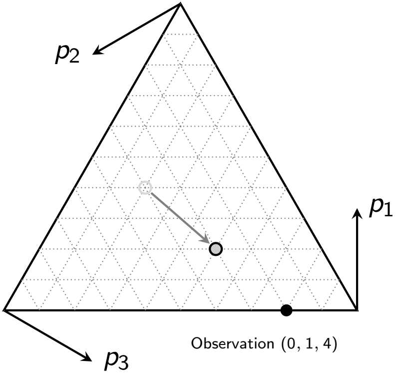
3-dimensional Dirichlet Distribution: Barycentric Graph



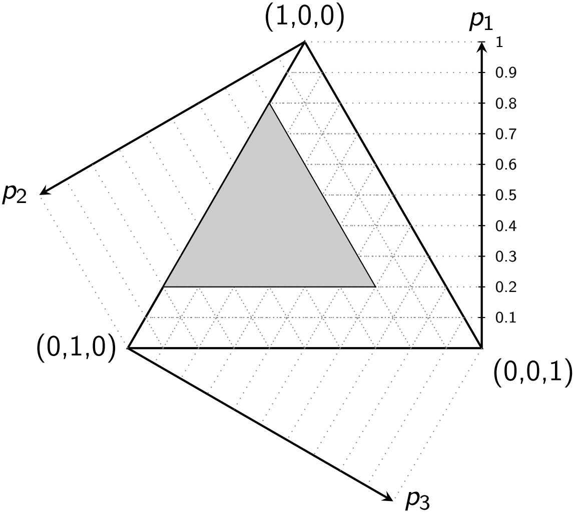
$y_j^{(0)} = (0.4, 0.4, 0.2)$
$(n^{(0)} = 5)$

Update Step for 3-dimensional Dirichlet Distribution



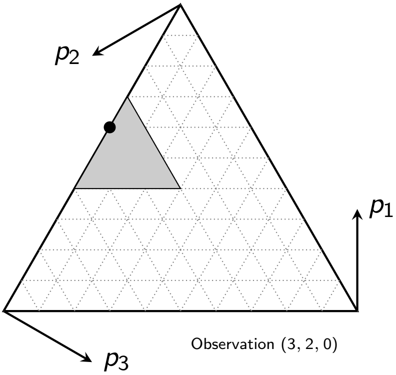


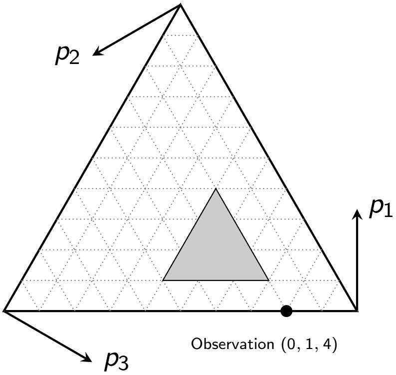
Barycentric Graph for IDM / iLUCK-model



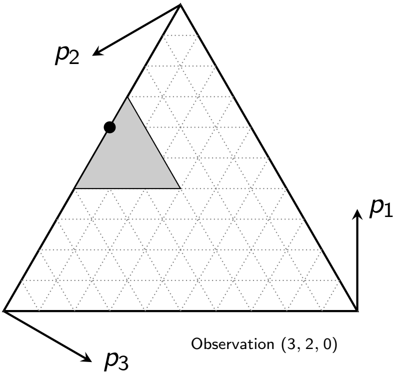
$y_1^{(0)} \in [0.2,\ 0.8]$
$y_2^{(0)} \in [0.2,\ 0.8]$
$y_3^{(0)} \in [0.0,\ 0.6]$
$(n^{(0)} = 5)$

Update Step in the IDM / iLUCK-model





Update Step in the IDM / iLUCK-model



→ same imprecision in both cases ?!?

Prior-Data Conflict in iLUCK-models

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

Prior-Data Conflict in iLUCK-models

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

➡ Posterior imprecision does not depend on $\tau(x)$!

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

Generalized iLUCK-models

Second approach: so-called *generalized iLUCK-models* (Walter & Augustin, 2009)

1. use conjugate priors as constructed by general method (prior parameters $y^{(0)}$, $n^{(0)}$)

Generalized iLUCK-models

Second approach: so-called *generalized iLUCK-models* (Walter & Augustin, 2009)

1. use conjugate priors as constructed by general method (prior parameters $y^{(0)}$, $n^{(0)}$)
2. construct sets of priors via sets of parameters $y^{(0)} \in \mathcal{Y}^{(0)} \times \boxed{n^{(0)} \in \mathcal{N}^{(0)}}$

➥ weigh prior information $\mathcal{Y}^{(0)}$ and sample information $\tilde{\tau}(x)$ more flexible in

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

Generalized iLUCK-models

Second approach: so-called *generalized iLUCK-models* (Walter & Augustin, 2009)

1. use conjugate priors as constructed by general method (prior parameters $y^{(0)}$, $n^{(0)}$)
2. construct sets of priors via sets of parameters $y^{(0)} \in \mathcal{Y}^{(0)} \times \boxed{n^{(0)} \in \mathcal{N}^{(0)}}$
   ➥ weigh prior information $\mathcal{Y}^{(0)}$ and sample information $\tilde{\tau}(x)$ more flexible in

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

3. set of posteriors $\hat{=}$ set of (element-wise) updated priors
   ➥ still easy to handle: described as set of $(y^{(1)}, n^{(1)})$'s

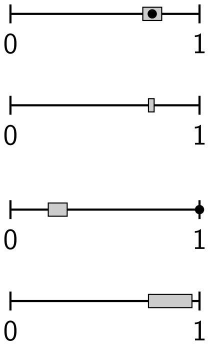
Generalized iLUCK-models — 1-dim Example

Case (i): $y_j^{(0)} \in [0.7,\ 0.8]\,,$ $(n^{(0)} \in [1, 8])$ $\quad k_j/n = 0.75$ $(n = 16)$

$\rightarrow$ $y_j^{(1)} \in [0.73,\ 0.76]\,,$ $(n^{(0)} \in [17, 24])$

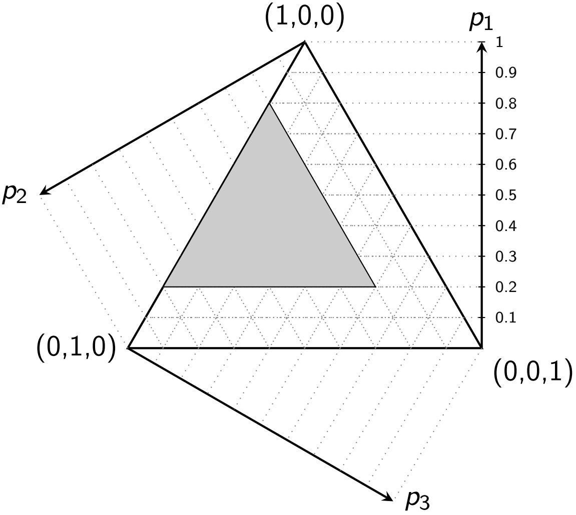
Case (ii): $y_j^{(0)} \in [0.2,\ 0.3]\,,$ $(n^{(0)} \in [1, 8])$ $\quad k_j/n = 1$ $(n = 16)$

$\rightarrow$ $y_j^{(1)} \in [0.73,\ 0.96]\,,$ $(n^{(0)} \in [17, 24])$



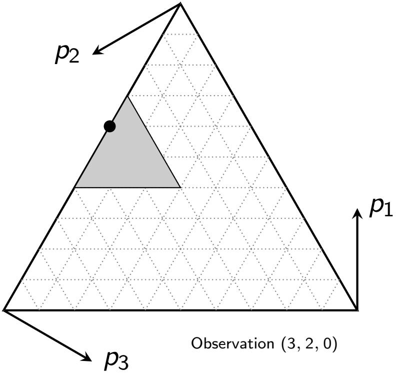
Generalized iLUCK-models lead to cautious inferences if, and only if, caution is needed.

Generalized iLUCK-model— 3-dim Example



$y_1^{(0)} \in [0.2,\ 0.8]$
$y_2^{(0)} \in [0.2,\ 0.8]$
$y_3^{(0)} \in [0.0,\ 0.6]$
$n^{(0)} \in [1,\ 5]$

Update step in the generalized iLUCK-model



Summary & References

- If observed data is unexpected under the prior model, this surprise is often not reflected in posterior inferences when conjugate priors are used.
- Fundamentally, prior-data conflict points to the issue of specifying the precision of probability statements in general.
- iLUCK-models like the IDM ignore on prior-data conflict just like standard conjugate models.
- Generalized iLUCK-models offer a general, manageable, and powerful calculus for Bayesian inference with sets of priors, allowing for a sensible reaction to prior-data conflict by increased imprecision of inferences.

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