Prior-Data Conflict and Generalized Bayesian Inference

Gero Walter

Institut für Statistik
Ludwig-Maximilians-Universität München

24.03.2010

- Prior-Data Conflict
- Generalized Bayesian Inference
- R-package `luck`

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

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

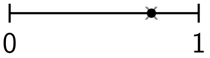
Data :	$\mathbf{k}$	$\sim$	$\mathrm{M}(\boldsymbol{\theta})$
conjugate prior:	$\boldsymbol{\theta}$	$\sim$	$\mathrm{Dir}(n^{(0)}, \mathbf{y}^{(0)})$
posterior:	$\boldsymbol{\theta} \mid \mathbf{k}$	$\sim$	$\mathrm{Dir}(n^{(1)}, \mathbf{y}^{(1)})$

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

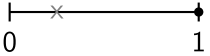
$$\mathbb{V}(\theta_j) = \frac{y_j^{(0)}(1 - y_j^{(0)})}{n^{(0)} + 1}$$

Prior-Data Conflict — Example

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

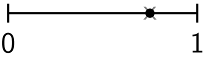


Case (ii): $y_j^{(0)} = 0.25\,, \quad (n^{(0)} = 8)$ $\qquad k_j/n = 1 \quad (n = 16)$

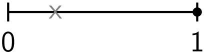


Prior-Data Conflict — Example

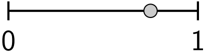
Case (i): $y_j^{(0)} = 0.75\,, \quad (n^{(0)} = 8)$ $\qquad k_j/n = 0.75 \quad (n = 16)$



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



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



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

⚠ 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.

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

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

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).

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

Why Generalize Bayesian Inference?

Assigning a certain prior distribution on θ
$\Longleftrightarrow$ Defining a conglomerate of probability statements (on θ).

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).

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“)

↓ Weichselberger (2001) ↓ 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)

Approach here: so-called *generalized i*LUCK*-models* (Walter & Augustin, 2009)

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

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)

Approach here: so-called *generalized i*LUCK*-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 $y^{(0)} \in \mathcal{Y}^{(0)} \times n^{(0)} \in \mathcal{N}^{(0)}$

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)

Approach here: so-called *generalized i*LUCK*-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 $y^{(0)} \in \mathcal{Y}^{(0)} \times n^{(0)} \in \mathcal{N}^{(0)}$
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 Bayesian Inference — Example

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

$(n^{(0)} \in [1,8]) \qquad (n = 16)$

→ $y_j^{(1)} \in [0.73,\, 0.76]\,,$

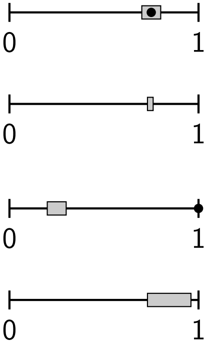
$(n^{(0)} \in [17,24])$

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

$(n^{(0)} \in [1,8]) \qquad (n = 16)$

→ $y_j^{(1)} \in [0.73,\, 0.96]\,,$

$(n^{(0)} \in [17,24])$

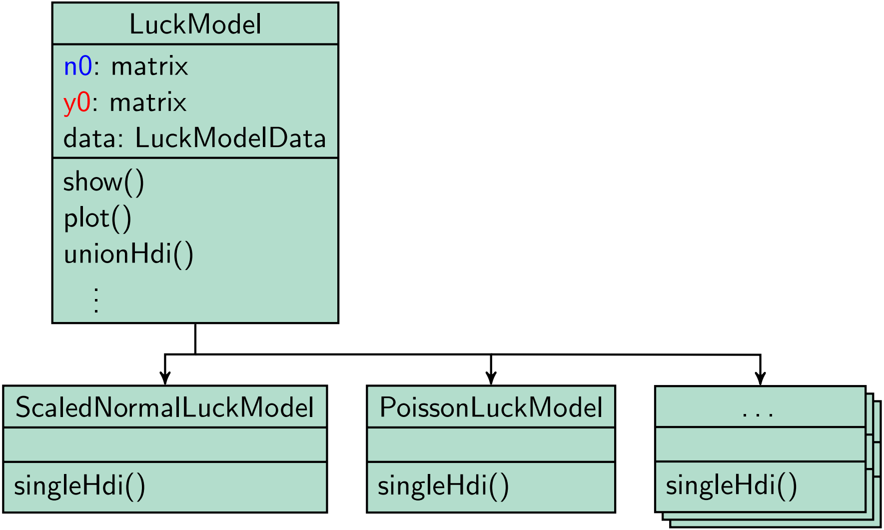


Generalized iLUCK-models lead to cautious inferences if, and only if, caution is needed.

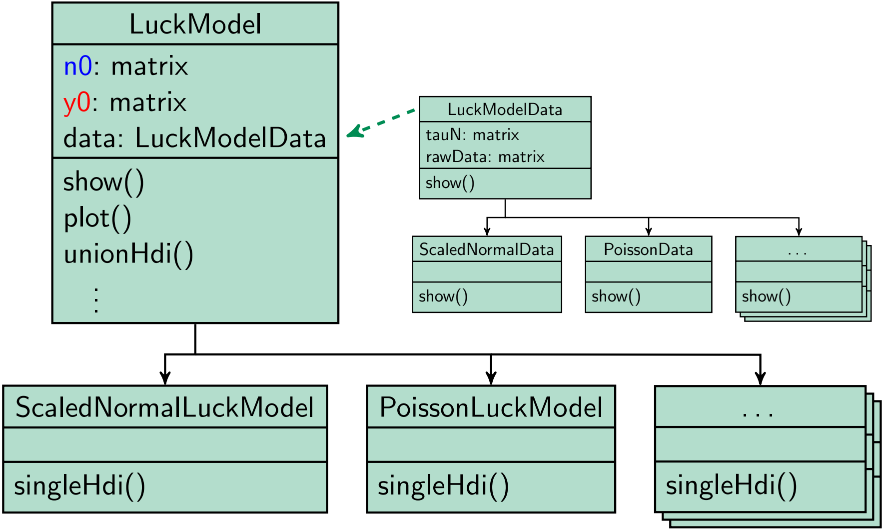
R-package `luck` (under development)

- S4 implementation of general prior structure
  (parameter sets $y^{(0)} \in \mathcal{Y}^{(0)} \times n^{(0)} \in \mathcal{N}^{(0)}$)
  and basic utilities
- Lean subclasses for inferences in various data situations
  (data from ScaledNormal, Poisson,...)
- Project page:
  `http://r-forge.r-project.org/projects/luck/`

R-package luck — Class Structure



R-package `luck` — Class Structure



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.
- 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.