# The Effect of Prior-Data Conflict in Bayesian Linear Regression

Gero Walter

Institut für Statistik
Ludwig-Maximilians-Universität München

April 14th, 2010

- Prior-Data Conflict
  - Simple Example: Dirichlet-Multinomial Model
  - Conjugate Priors
- Bayesian Linear Regression
  - Standard Conjugate Prior (SCP)
  - Canonically Constructed Conjugate Prior (CCCP)
- Generalized Bayesian Inference
  - Basic Idea
  - Generalized iLUCK-models
  - Outlook & Summary

# 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)

# Simple Example: Dirichlet-Multinomial-Model

$$
\begin{array}{r|ccl}
\text{Data:} & \mathbf{k} & \sim & \mathrm{M}(\boldsymbol{\theta}) \\
\text{conjugate prior:} & \boldsymbol{\theta} & \sim & \mathrm{Dir}(\boldsymbol{\alpha}) \\
\hline
\text{posterior:} & \boldsymbol{\theta} \mid \mathbf{k} & \sim & \mathrm{Dir}(\boldsymbol{\alpha} + \mathbf{k})
\end{array}
\qquad
\begin{array}{l}
(\sum k_j = n) \\
(\sum \theta_j = 1)
\end{array}
$$

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

# Simple Example: Dirichlet-Multinomial-Model

$$\begin{array}{r|ccl}
\text{Data:} & \mathbf{k} & \sim & \mathrm{M}(\boldsymbol{\theta}) \\
\text{conjugate prior:} & \boldsymbol{\theta} & \sim & \mathrm{Dir}(\boldsymbol{\alpha}) \\
\hline
\text{posterior:} & \boldsymbol{\theta} \mid \mathbf{k} & \sim & \mathrm{Dir}(\boldsymbol{\alpha} + \mathbf{k})
\end{array}
\qquad
\begin{array}{l}
(\sum k_j = n) \\
(\sum \theta_j = 1)
\end{array}$$

$$\mathbb{E}[\theta_j] = \frac{\alpha_j}{\sum \alpha_i} =: y_j^{(0)} \qquad \mathbb{V}(\theta_j) = \frac{\mathbb{E}[\theta_j](1 - \mathbb{E}[\theta_j])}{\sum \alpha_i + 1} = \frac{y_j^{(0)}(1 - y_j^{(0)})}{n^{(0)} + 1}$$

$$\begin{array}{r|ccl}
\text{Data :} & \mathbf{k} & \sim & \mathrm{M}(\boldsymbol{\theta}) \\
\text{conjugate prior:} & \boldsymbol{\theta} & \sim & \mathrm{Dir}(n^{(0)}, \mathbf{y}^{(0)}) \\
\hline
\text{posterior:} & \boldsymbol{\theta} \mid \mathbf{k} & \sim & \mathrm{Dir}(n^{(1)}, \mathbf{y}^{(1)})
\end{array}
\qquad
n^{(0)} = \sum \alpha_i$$

# Simple Example: Dirichlet-Multinomial-Model

$$\begin{array}{r|ccl}
\text{Data:} & \mathbf{k} & \sim & \mathrm{M}(\boldsymbol{\theta}) \\
\text{conjugate prior:} & \boldsymbol{\theta} & \sim & \mathrm{Dir}(\boldsymbol{\alpha}) \\
\hline
\text{posterior:} & \boldsymbol{\theta} \mid \mathbf{k} & \sim & \mathrm{Dir}(\boldsymbol{\alpha} + \mathbf{k})
\end{array}
\qquad
\begin{array}{l}
(\sum k_j = n) \\
(\sum \theta_j = 1)
\end{array}$$

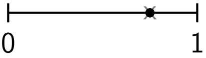
$$\mathbb{E}[\theta_j] = \frac{\alpha_j}{\sum \alpha_i} =: y_j^{(0)} \quad \mathbb{V}(\theta_j) = \frac{\mathbb{E}[\theta_j](1 - \mathbb{E}[\theta_j])}{\sum \alpha_i + 1} = \frac{y_j^{(0)}(1 - y_j^{(0)})}{n^{(0)} + 1}$$

$$\begin{array}{r|ccl}
\text{Data :} & \mathbf{k} & \sim & \mathrm{M}(\boldsymbol{\theta}) \\
\text{conjugate prior:} & \boldsymbol{\theta} & \sim & \mathrm{Dir}(n^{(0)}, \mathbf{y}^{(0)}) \\
\hline
\text{posterior:} & \boldsymbol{\theta} \mid \mathbf{k} & \sim & \mathrm{Dir}(n^{(1)}, \mathbf{y}^{(1)})
\end{array}
\qquad
n^{(0)} = \sum \alpha_i$$

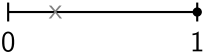
$$y_j^{(1)} = \frac{n^{(0)}}{n^{(0)} + n} \cdot y_j^{(0)} + \frac{n}{n^{(0)} + n} \cdot \frac{k_j}{n} \qquad n^{(1)} = n^{(0)} + n$$

# Simple Example: Dirichlet-Multinomial-Model

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

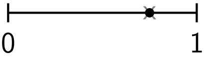


Case (ii): $y_j^{(0)} = 0.25\,,$ $(n^{(0)} = 8)$ $\quad k_j/n = 1$ $(n = 16)$

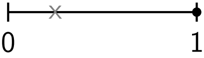


# Simple Example: Dirichlet-Multinomial-Model

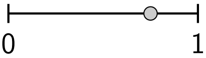
Case (i): $y_j^{(0)} = 0.75\,, \quad (n^{(0)} = 8) \qquad k_j/n = 0.75 \quad (n = 16)$



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



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



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

⚠ 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.

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

➡ conjugate prior:

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

# Conjugate 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 $n^{(1)} = n^{(0)} + n\,.$

# Bayesian Linear Regression

Are posterior inferences in Bayesian linear regression influenced by prior-data conflict?

$$z_i = x_i^{\mathsf{T}} \beta + \varepsilon_i \quad \Big[ x_i \in \mathbb{R}^p \,,\ \beta \in \mathbb{R}^p \Big]$$

$$z = \mathbf{X}\beta + \varepsilon$$

# Bayesian Linear Regression

Are posterior inferences in Bayesian linear regression influenced by prior-data conflict?

$$z_i = x_i^{\mathsf{T}} \beta + \varepsilon_i \quad \left[x_i \in \mathbb{R}^p\,,\; \beta \in \mathbb{R}^p\right] \qquad \varepsilon_i \overset{iid}{\sim} \mathrm{N}(0, \sigma^2)$$

$$z = \mathbf{X}\beta + \varepsilon \qquad \rightarrow \qquad z \mid \beta, \sigma^2 \sim \mathrm{N}(\mathbf{X}\beta, \sigma^2 \mathbf{I})$$

# Bayesian Linear Regression

Are posterior inferences in Bayesian linear regression influenced by prior-data conflict?

$$z_i = x_i^{\mathsf{T}} \beta + \varepsilon_i \quad \left[ x_i \in \mathbb{R}^p \,, \; \beta \in \mathbb{R}^p \right] \qquad \varepsilon_i \overset{iid}{\sim} \mathrm{N}(0, \sigma^2)$$

$$z = \mathbf{X}\beta + \varepsilon \qquad \longrightarrow \quad z \mid \beta, \sigma^2 \sim \mathrm{N}(\mathbf{X}\beta, \sigma^2 \mathbf{I})$$

Prior on $(\beta, \sigma^2)$: generally taken as $p(\beta, \sigma^2) = p(\beta \mid \sigma^2) p(\sigma^2)$.

- SCP: standard conjugate prior model (e.g., O'Hagan 1994)
- CCCP: "canonically constructed conjugate prior"

# Standard Conjugate Prior (SCP)

$$\beta \mid \sigma^2 \sim \mathrm{N}_p(m^{(0)}, \sigma^2 \mathbf{M}^{(0)}) \quad \text{(multivariate Normal)}$$

$$\sigma^2 \sim \mathrm{IG}\,(a^{(0)}, b^{(0)}) \qquad \text{(Inverse Gamma, e.g. } p(\sigma^2) \propto \frac{e^{-\frac{b^{(0)}}{\sigma^2}}}{(\sigma^2)^{a^{(0)}+1}}\text{)}$$

# Standard Conjugate Prior (SCP)

$$\beta \mid \sigma^2 \sim \mathrm{N}_p(m^{(0)}, \sigma^2 \mathbf{M}^{(0)}) \quad \text{(multivariate Normal)}$$

$$\sigma^2 \sim \mathrm{IG}\,(a^{(0)}, b^{(0)}) \qquad \text{(Inverse Gamma, e.g. } p(\sigma^2) \propto \frac{e^{-\frac{b^{(0)}}{\sigma^2}}}{(\sigma^2)^{a^{(0)}+1}}\text{)}$$

$$\begin{aligned}
m^{(1)} &= \left(\mathbf{M}^{(0)^{-1}} + \mathbf{X}^\mathsf{T}\mathbf{X}\right)^{-1} \left(\mathbf{M}^{(0)^{-1}} m^{(0)} + \mathbf{X}^\mathsf{T} z\right) \\
\mathbf{M}^{(1)} &= \left(\mathbf{M}^{(0)^{-1}} + \mathbf{X}^\mathsf{T}\mathbf{X}\right)^{-1} \\
a^{(1)} &= a^{(0)} + \frac{n}{2} \\
b^{(1)} &= b^{(0)} + \frac{1}{2}\left(z^\mathsf{T} z + m^{(0)^\mathsf{T}} \mathbf{M}^{(0)^{-1}} m^{(0)} - m^{(1)^\mathsf{T}} \mathbf{M}^{(1)^{-1}} m^{(1)}\right)
\end{aligned}$$

# SCP: Update step for $\beta \mid \sigma^2$

$$\begin{aligned} \mathbb{E}[\beta \mid \sigma^2] &= m^{(0)} \\ \mathbb{E}[\beta \mid \sigma^2, z] &= m^{(1)} = (\mathbf{I} - \mathbf{A})\, m^{(0)} + \mathbf{A}\, \hat{\beta}^{\mathsf{LS}} \\ &\qquad \text{where } \mathbf{A} = \left(\mathbf{M}^{(0)^{-1}} + \mathbf{X}^{\mathsf{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathsf{T}}\mathbf{X} \end{aligned}$$

# SCP: Update step for $\beta \mid \sigma^2$

$$
\begin{aligned}
\mathbb{E}[\beta \mid \sigma^2] &= m^{(0)} \\
\mathbb{E}[\beta \mid \sigma^2, z] &= m^{(1)} = (\mathbf{I} - \mathbf{A})\, m^{(0)} + \mathbf{A}\, \hat{\beta}^{\text{LS}} \\
&\qquad\qquad\qquad \text{where } \mathbf{A} = \left(\mathbf{M}^{(0)^{-1}} + \mathbf{X}^{\mathsf{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathsf{T}}\mathbf{X} \\
\mathbb{V}(\beta \mid \sigma^2) &= \sigma^2 \mathbf{M}^{(0)} \\
\mathbb{V}(\beta \mid \sigma^2, z) &= \sigma^2 \mathbf{M}^{(1)} = \sigma^2 \left(\mathbf{M}^{(0)^{-1}} + \mathbf{X}^{\mathsf{T}}\mathbf{X}\right)^{-1}
\end{aligned}
$$

➡ $\mathbb{V}(\beta_j \mid \sigma^2, z) < \mathbb{V}(\beta_j \mid \sigma^2)$

# SCP: Update step for $\beta \mid \sigma^2$

$$\begin{aligned}
\mathbb{E}[\beta \mid \sigma^2] &= m^{(0)} \\
\mathbb{E}[\beta \mid \sigma^2, z] &= m^{(1)} = (\mathbf{I} - \mathbf{A})\, m^{(0)} + \mathbf{A}\, \hat{\beta}^{\mathrm{LS}} \\
&\qquad \text{where } \mathbf{A} = \left(\mathbf{M}^{(0)^{-1}} + \mathbf{X}^{\mathsf{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathsf{T}}\mathbf{X} \\
\mathbb{V}(\beta \mid \sigma^2) &= \sigma^2 \mathbf{M}^{(0)} \\
\mathbb{V}(\beta \mid \sigma^2, z) &= \sigma^2 \mathbf{M}^{(1)} = \sigma^2 \left(\mathbf{M}^{(0)^{-1}} + \mathbf{X}^{\mathsf{T}}\mathbf{X}\right)^{-1}
\end{aligned}$$

➡ $\mathbb{V}(\beta_j \mid \sigma^2, z) < \mathbb{V}(\beta_j \mid \sigma^2)$

Actually of interest:

$$\begin{aligned}
\mathbb{E}[\beta] &= m^{(0)}, & \mathbb{V}(\beta) &= \frac{b^{(0)}}{a^{(0)} - 1}\mathbf{M}^{(0)} = \mathbb{E}[\sigma^2]\,\mathbf{M}^{(0)} \\
\mathbb{E}[\beta \mid z] &= m^{(1)}, & \mathbb{V}(\beta \mid z) &= \frac{b^{(1)}}{a^{(1)} - 1}\mathbf{M}^{(1)} = \mathbb{E}[\sigma^2 \mid z]\,\mathbf{M}^{(1)}
\end{aligned}$$

# SCP: Update step for $\sigma^2$

$$\mathbb{E}[\sigma^2 \mid z] = \frac{2a^{(0)} - 2}{2a^{(0)} + n - 2}\,\mathbb{E}[\sigma^2] + \frac{n-p}{2a^{(0)} + n - 2}\,\hat\sigma^2_{\mathrm{LS}} + \frac{p}{2a^{(0)} + n - 2}\,\hat\sigma^2_{\mathrm{PDC}}$$

$$\begin{aligned}\hat\sigma^2_{\mathrm{LS}} &= \frac{1}{n-p}(z - \mathbf{X}\hat\beta^{\mathrm{LS}})^{\mathsf{T}}(z - \mathbf{X}\hat\beta^{\mathrm{LS}}) \\ \hat\sigma^2_{\mathrm{PDC}} &= \frac{1}{p}(m^{(0)} - \hat\beta^{\mathrm{LS}})^{\mathsf{T}}\Big(\mathbf{M}^{(0)} + (\mathbf{X}^{\mathsf{T}}\mathbf{X})^{-1}\Big)^{-1}(m^{(0)} - \hat\beta^{\mathrm{LS}})\end{aligned}$$

➥ $\mathbb{E}[\hat\sigma^2_{\mathrm{PDC}} \mid \sigma^2] = \sigma^2$

# SCP: Update step for $\sigma^2$

$$\mathbb{E}[\sigma^2 \mid z] = \frac{2a^{(0)} - 2}{2a^{(0)} + n - 2}\,\mathbb{E}[\sigma^2] + \frac{n-p}{2a^{(0)} + n - 2}\,\hat{\sigma}^2_{\mathrm{LS}} + \frac{p}{2a^{(0)} + n - 2}\,\hat{\sigma}^2_{\mathrm{PDC}}$$

$$\begin{aligned} \hat{\sigma}^2_{\mathrm{LS}} &= \frac{1}{n-p}(z - \mathbf{X}\hat{\beta}^{\mathrm{LS}})^{\mathrm{T}}(z - \mathbf{X}\hat{\beta}^{\mathrm{LS}}) \\ \hat{\sigma}^2_{\mathrm{PDC}} &= \frac{1}{p}(m^{(0)} - \hat{\beta}^{\mathrm{LS}})^{\mathrm{T}}\left(\mathbf{M}^{(0)} + (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\right)^{-1}(m^{(0)} - \hat{\beta}^{\mathrm{LS}}) \end{aligned}$$

➥ $\mathbb{E}[\hat{\sigma}^2_{\mathrm{PDC}} \mid \sigma^2] = \sigma^2$

Weights:

- $2a^{(0)} - 2$ for $\mathbb{E}[\sigma^2]$: think of $\mathbb{V}(\sigma^2) = \frac{(b^{(0)})^2}{(a^{(0)}-1)^2(a^{(0)}-2)}$
- $n - p$ for $\hat{\sigma}^2_{\mathrm{LS}}$: usual *df*s in least-squares estimate
- $p$ for $\hat{\sigma}^2_{\mathrm{PDC}}$: $\dim(\beta) =$ number of dimensions in which prior-data conflict is possible

# SCP: Summary

$$\mathbb{E}[\beta \mid z] = m^{(1)}$$

$$\mathbb{V}(\beta \mid z) = \mathbb{E}[\sigma^2 \mid z] \cdot \mathbf{M}^{(1)}$$

# SCP: Summary

weighted average of prior and LS estimate

$$\mathbb{E}[\beta \mid z] = m^{(1)}$$

$$\mathbb{V}(\beta \mid z) = \mathbb{E}[\sigma^2 \mid z] \cdot \mathbf{M}^{(1)}$$

# SCP: Summary

weighted average of prior and LS estimate

$$\mathbb{E}[\beta \mid z] = m^{(1)}$$

$$\mathbb{V}(\beta \mid z) = \mathbb{E}[\sigma^2 \mid z] \cdot \mathbf{M}^{(1)}$$

may increase due to prior-data conflict (weight $p$)

# SCP: Summary

weighted average of prior and LS estimate

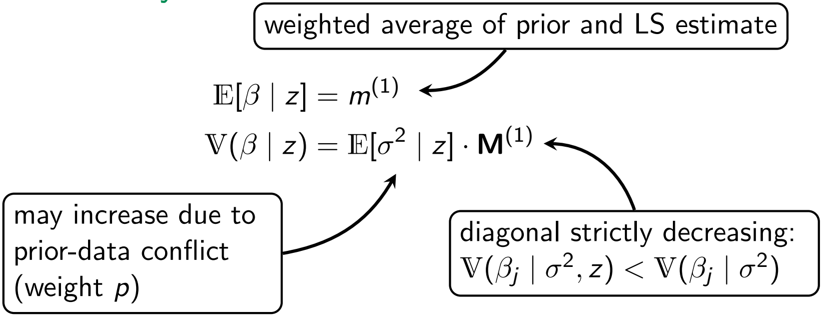
$$\mathbb{E}[\beta \mid z] = m^{(1)}$$

$$\mathbb{V}(\beta \mid z) = \mathbb{E}[\sigma^2 \mid z] \cdot \mathbf{M}^{(1)}$$

may increase due to prior-data conflict (weight $p$)

diagonal strictly decreasing: $\mathbb{V}(\beta_j \mid \sigma^2, z) < \mathbb{V}(\beta_j \mid \sigma^2)$

# SCP: Summary



weighted average of prior and LS estimate

$$\mathbb{E}[\beta \mid z] = m^{(1)}$$

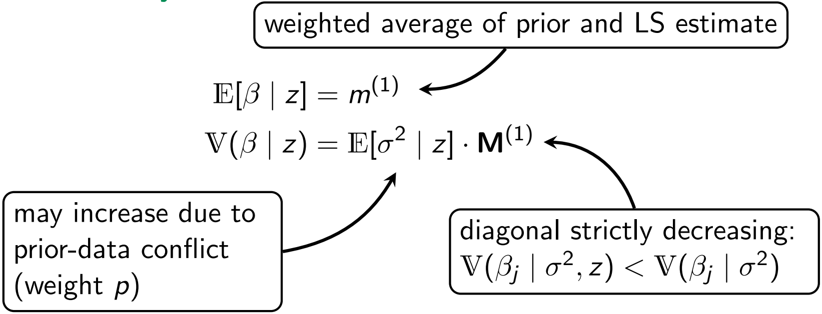
$$\mathbb{V}(\beta \mid z) = \mathbb{E}[\sigma^2 \mid z] \cdot \mathbf{M}^{(1)}$$

may increase due to prior-data conflict (weight $p$)

diagonal strictly decreasing: $\mathbb{V}(\beta_j \mid \sigma^2, z) < \mathbb{V}(\beta_j \mid \sigma^2)$

→ Posterior variance may increase due to prior-data conflict. But:

# SCP: Summary



weighted average of prior and LS estimate

$$\mathbb{E}[\beta \mid z] = m^{(1)}$$

$$\mathbb{V}(\beta \mid z) = \mathbb{E}[\sigma^2 \mid z] \cdot \mathbf{M}^{(1)}$$

may increase due to prior-data conflict (weight $p$)

diagonal strictly decreasing: $\mathbb{V}(\beta_j \mid \sigma^2, z) < \mathbb{V}(\beta_j \mid \sigma^2)$

→ Posterior variance may increase due to prior-data conflict. But:

- Effect of possible increase of $\mathbb{E}[\sigma^2 \mid z]$ contrasted by automatic decrease of $\mathbf{M}^{(1)}$!
- Variance increase *by the same factor* for all $\beta_j$s!

# Canonically Constructed Conjugate Prior (CCCP)

CCCP turns out as a special case of SCP:

$$\beta \mid \sigma^2 \sim \mathrm{N}_p\Big(m^{(0)}\,,\ \sigma^2 \underbrace{\frac{n}{n^{(0)}}\,(\mathbf{X}^\mathsf{T}\mathbf{X})^{-1}}_{=:\mathbf{M}^{(0)}}\Big)$$

$$\sigma^2 \sim \mathrm{IG}\left(a^{(0)},\ b^{(0)}\right)$$

# Canonically Constructed Conjugate Prior (CCCP)

CCCP turns out as a special case of SCP:

some linear trafo. of $y_1^{(0)} \in \mathbb{R}^p$

$$\beta \mid \sigma^2 \sim \mathrm{N}_p\left(m^{(0)},\ \sigma^2 \underbrace{\frac{n}{n^{(0)}} (\mathbf{X}^\mathsf{T}\mathbf{X})^{-1}}_{=:\mathbf{M}^{(0)}}\right)$$

restricted covariance structure

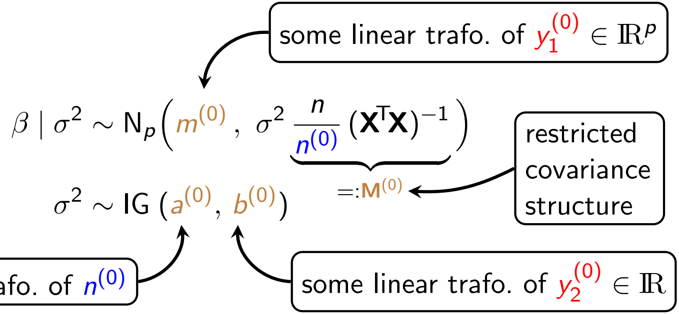
$$\sigma^2 \sim \mathrm{IG}\left(a^{(0)},\ b^{(0)}\right)$$

some linear trafo. of $n^{(0)}$

some linear trafo. of $y_2^{(0)} \in \mathbb{R}$

# Canonically Constructed Conjugate Prior (CCCP)

CCCP turns out as a special case of SCP:



some linear trafo. of $y_1^{(0)} \in \mathbb{R}^p$

$$\beta \mid \sigma^2 \sim \mathrm{N}_p\Big(m^{(0)}\,,\ \underbrace{\sigma^2\,\frac{n}{n^{(0)}}\,(\mathbf{X}^\mathsf{T}\mathbf{X})^{-1}}_{=:\mathbf{M}^{(0)}}\Big)$$

restricted covariance structure

$$\sigma^2 \sim \mathrm{IG}\left(a^{(0)},\ b^{(0)}\right)$$

some linear trafo. of $n^{(0)}$

some linear trafo. of $y_2^{(0)} \in \mathbb{R}$

$$\mathbb{E}[\beta \mid \sigma^2, z] = m^{(1)} = \frac{n^{(0)}}{n^{(0)} + n}\mathbb{E}[\beta \mid \sigma^2] + \frac{n}{n^{(0)} + n}\hat{\beta}^{\mathrm{LS}}$$

$$\mathbb{V}(\beta \mid \sigma^2, z) = \sigma^2\mathbf{M}^{(1)} = \sigma^2\frac{n}{n^{(0)} + n}\,(\mathbf{X}^\mathsf{T}\mathbf{X})^{-1}$$

# CCCP: Update step for $\sigma^2$

Again, for the posterior on $\beta$ it holds that

$$\mathbb{E}[\beta \mid z] = m^{(1)}$$
$$\mathbb{V}(\beta \mid z) = \mathbb{E}[\sigma^2 \mid z] \cdot \mathbf{M}^{(1)}$$

such that the update step for $\mathbb{E}[\sigma^2]$ is of most interest:

# CCCP: Update step for $\sigma^2$

Again, for the posterior on $\beta$ it holds that

$$\mathbb{E}[\beta \mid z] = m^{(1)}$$
$$\mathbb{V}(\beta \mid z) = \mathbb{E}[\sigma^2 \mid z] \cdot \mathbf{M}^{(1)}$$

such that the update step for $\mathbb{E}[\sigma^2]$ is of most interest:

$$\mathbb{E}[\sigma^2 \mid z] = \frac{n^{(0)} + p}{n^{(0)} + n + p} \mathbb{E}[\sigma^2] + \frac{n - p}{n^{(0)} + n + p} \hat{\sigma}^2_{\mathrm{LS}} + \frac{p}{n^{(0)} + n + p} \hat{\sigma}^2_{\mathrm{PDC}}$$

$$\hat{\sigma}^2_{\mathrm{PDC}} = \frac{1}{p}(m^{(0)} - \hat{\beta}^{\mathrm{LS}})^{\mathrm{T}} \frac{n^{(0)}}{n^{(0)} + n} \mathbf{X}^{\mathrm{T}}\mathbf{X}(m^{(0)} - \hat{\beta}^{\mathrm{LS}}) \qquad \Big[\mathbb{E}[\,\cdot \mid \sigma^2] = \sigma^2\Big]$$

For CCCP, two other interesting decompositions of $\mathbb{E}[\sigma^2 \mid z]$ exist.

# CCCP: Summary

$$\mathbb{E}[\beta \mid z] = m^{(1)} = \frac{n^{(0)}}{n^{(0)} + n}\mathbb{E}[\beta \mid \sigma^2] + \frac{n}{n^{(0)} + n}\hat{\beta}^{\mathsf{LS}}$$

# CCCP: Summary

$$\mathbb{E}[\beta \mid z] = m^{(1)} = \frac{n^{(0)}}{n^{(0)} + n}\mathbb{E}[\beta \mid \sigma^2] + \frac{n}{n^{(0)} + n}\hat{\beta}^{\mathsf{LS}}$$

$$\mathbb{V}(\beta \mid z) = \mathbb{E}[\sigma^2 \mid z] \cdot \mathbf{M}^{(1)}$$

elements strictly decreasing

may increase due to prior-data conflict (weight $p$)

# CCCP: Summary

$$\mathbb{E}[\beta \mid z] = m^{(1)} = \frac{n^{(0)}}{n^{(0)} + n}\mathbb{E}[\beta \mid \sigma^2] + \frac{n}{n^{(0)} + n}\hat{\beta}^{\mathrm{LS}}$$

$$\begin{aligned}\mathbb{V}(\beta \mid z) &= \mathbb{E}[\sigma^2 \mid z] \cdot \mathbf{M}^{(1)} \\ &= \frac{n^{(0)} + p}{n^{(0)} + n + p}\,\frac{n^{(0)}}{n^{(1)}}\,\underbrace{\mathbb{E}[\sigma^2]\frac{n}{n^{(0)}}(\mathbf{X}^{\mathsf{T}}\mathbf{X})^{-1}}_{\mathbb{V}(\beta)} \\ &\quad + \frac{n - p}{n^{(0)} + n + p}\,\frac{n}{n^{(1)}}\,\underbrace{\hat{\sigma}^2_{\mathrm{LS}}(\mathbf{X}^{\mathsf{T}}\mathbf{X})^{-1}}_{\mathbb{V}(\hat{\beta}^{\mathrm{LS}})} \\ &\quad + \frac{p}{n^{(0)} + n + p}\,\frac{n}{n^{(1)}}\,\hat{\sigma}^2_{\mathrm{PDC}}(\mathbf{X}^{\mathsf{T}}\mathbf{X})^{-1}\end{aligned}$$

elements strictly decreasing

may increase due to prior-data conflict (weight $p$)

# Generalized Bayesian Inference: Basic Idea

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

# 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

**Lottery A**
Number of winning tickets:
exactly known as 5 out of 100
→ $P(\text{win}) = 5/100$

**Lottery B**
Number of winning tickets:
not exactly known, supposedly
between 1 and 7 out of 100
→ $P(\text{win}) = [1/100, 7/100]$

# Generalized Bayesian Inference: Basic Idea

Sets of distributions ↔ Probability Intervals / Expectation Intervals
(„credal sets“) (lower/upper prob.) (lower/upper prevision)

(lower/upper prob.) ↓ Weichselberger (2001)

(lower/upper prevision) ↓ Walley (1991)

# Generalized Bayesian Inference: Basic Idea

Sets of distributions ↔ Probability Intervals / Expectation Intervals

(„credal sets“) (lower/upper prob.) (lower/upper prevision)

↓ ↓

Weichselberger (2001) Walley (1991)

$$\overline{P}(A) = 1 - \underline{P}(A^c) \qquad \overline{\mathbb{E}}[X] = -\underline{\mathbb{E}}(-X)$$

# Generalized Bayesian Inference: Basic Idea

Sets of distributions („credal sets“) ⟷ Probability Intervals (lower/upper prob.) / Expectation Intervals (lower/upper prevision)

(lower/upper prob.) ↓ Weichselberger (2001)

(lower/upper prevision) ↓ Walley (1991)

$$\overline{P}(A) = 1 - \underline{P}(A^c) \qquad \overline{\mathbb{E}}[X] = -\underline{\mathbb{E}}(-X)$$

→ The Society for Imprecise Probability: Theories and Applications
(ISIPTA conferences, summer schools,… www.sipta.org)

# Generalized Bayesian Inference: Basic Idea

Sets of distributions ⟷ Probability Intervals / Expectation Intervals
(„credal sets“) (lower/upper prob.) (lower/upper prevision)

↓ Weichselberger (2001)

↓ Walley (1991)

$$\overline{P}(A) = 1 - \underline{P}(A^c) \qquad \overline{\mathbb{E}}[X] = -\underline{\mathbb{E}}(-X)$$

→ The Society for Imprecise Probability: Theories and Applications
(ISIPTA conferences, summer schools,… www.sipta.org)

→ Frank Coolen (my host)

# Generalized Bayesian Inference: Basic Idea

Sets of distributions („credal sets“) ⟷ Probability Intervals (lower/upper prob.) / Expectation Intervals (lower/upper prevision)

(lower/upper prob.) ↓ Weichselberger (2001)

(lower/upper prevision) ↓ Walley (1991)

$$\overline{P}(A) = 1 - \underline{P}(A^c) \qquad \overline{\mathbb{E}}[X] = -\underline{\mathbb{E}}(-X)$$

- ➡ The Society for Imprecise Probability: Theories and Applications (ISIPTA conferences, summer schools,... www.sipta.org)
- ➡ Frank Coolen (my host)
- ➡ Thomas Augustin (my advisor)

# Generalized Bayesian Inference: Basic Idea

Sets of distributions ↔ Probability Intervals / Expectation Intervals
(„credal sets“)
(lower/upper prob.) (lower/upper prevision)
↓ ↓
Weichselberger (2001) Walley (1991)

$$\overline{P}(A) = 1 - \underline{P}(A^c) \qquad \overline{\mathbb{E}}[X] = -\underline{\mathbb{E}}(-X)$$

- The Society for Imprecise Probability: Theories and Applications
  (ISIPTA conferences, summer schools,... www.sipta.org)

- Frank Coolen (my host)
- Thomas Augustin (my advisor)
- ...

# Generalized iLUCK-models

Model for Bayesian inference with sets of priors
(Walter & Augustin, 2009)

1. use conjugate priors from general construction method (prior parameters $y^{(0)}$, $n^{(0)}$)

# Generalized iLUCK-models

Model for Bayesian inference with sets of priors
(Walter & Augustin, 2009)

1. use conjugate priors from general construction 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 iLUCK-models

Model for Bayesian inference with sets of priors
(Walter & Augustin, 2009)

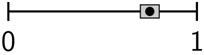
1. use conjugate priors from general construction method (prior parameters $y^{(0)}$, $n^{(0)}$)
2. construct sets of priors via sets of parameters $y^{(0)} \in \mathcal{Y}^{(0)} \times n^{(0)} \in \mathcal{N}^{(0)}$
3. set of posteriors $\hat{=}$ set of (element-wise) updated priors
   - ➡ still easy to handle: described as set of $(y^{(1)}, n^{(1)})$'s

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

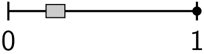
$$n^{(1)} = n^{(0)} + n$$

# Generalized iLUCK-models: Simple Example

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

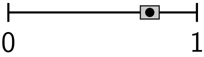


Case (ii): $y_j^{(0)} \in [0.2, 0.3]\,,$ $(n^{(0)} \in [1, 8])$ $k_j/n = 1$ $(n = 16)$

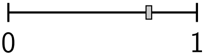


# Generalized iLUCK-models: Simple Example

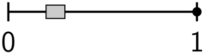
Case (i): $y_j^{(0)} \in [0.7,\, 0.8]\,,$ $(n^{(0)} \in [1, 8])$ $k_j/n = 0.75$ $(n = 16)$



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

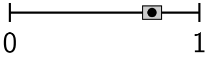


Case (ii): $y_j^{(0)} \in [0.2,\, 0.3]\,,$ $(n^{(0)} \in [1, 8])$ $k_j/n = 1$ $(n = 16)$

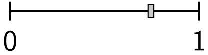


# Generalized iLUCK-models: Simple Example

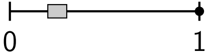
Case (i): $y_j^{(0)} \in [0.7,\, 0.8]\,,$ $(n^{(0)} \in [1, 8])$ $\quad k_j/n = 0.75$ $(n = 16)$



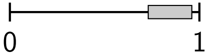
→ $y_j^{(1)} \in [0.73,\, 0.76]$ $(n^{(0)} \in [17, 24])$



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



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



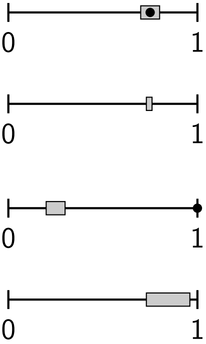
# Generalized iLUCK-models: Simple Example

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

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

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

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



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

# Generalized iLUCK-models in Bayesian Linear Regression

➡ Apply generalized iLUCK-models to Bayesian linear regression in order to improve behaviour!

# Generalized iLUCK-models in Bayesian Linear Regression

➞ Apply generalized iLUCK-models to Bayesian linear regression in order to improve behaviour!

CCCP model allows to employ generalized iLUCK-model inference:

- Set intervals for $\mathbb{E}[\beta_j] = m_j^{(0)}$s
  ➞ update step has dimension-individual sensitivity to prior-data conflict, i.e. interval lengths in different dimensions depend on degree of prior-data conflict in that dimension.

# Generalized iLUCK-models in Bayesian Linear Regression

➡ Apply generalized iLUCK-models to Bayesian linear regression in order to improve behaviour!

CCCP model allows to employ generalized iLUCK-model inference:

- Set intervals for $\mathbb{E}[\beta_j] = m_j^{(0)}$s
  ➡ update step has dimension-individual sensitivity to prior-data conflict, i.e. interval lengths in different dimensions depend on degree of prior-data conflict in that dimension.
- Fixed covariance structure $\mathbf{M}^{(0)} = \frac{n}{n^{(0)}}(\mathbf{X}^\mathsf{T}\mathbf{X})^{-1}$ in $\mathbb{V}(\beta) = \mathbb{E}[\sigma^2] \cdot \mathbf{M}^{(0)}$ seems restrictive, but interval for $n^{(0)}$ gives variety. (Defining a more general set of symmetric positive definite matrices could be very difficult!)

# Generalized iLUCK-models in Bayesian Linear Regression

➡ Apply generalized iLUCK-models to Bayesian linear regression in order to improve behaviour!

CCCP model allows to employ generalized iLUCK-model inference:

- Set intervals for $\mathbb{E}[\beta_j] = m_j^{(0)}$s
  ➡ update step has dimension-individual sensitivity to prior-data conflict, i.e. interval lengths in different dimensions depend on degree of prior-data conflict in that dimension.
- Fixed covariance structure $\mathbf{M}^{(0)} = \frac{n}{n^{(0)}}(\mathbf{X}^\mathsf{T}\mathbf{X})^{-1}$ in $\mathbb{V}(\beta) = \mathbb{E}[\sigma^2] \cdot \mathbf{M}^{(0)}$ seems restrictive, but interval for $n^{(0)}$ gives variety. (Defining a more general set of symmetric positive definite matrices could be very difficult!)
- Set interval for $\mathbb{E}[\sigma^2]$. (Update step is a bit tricky, though.)

# Summary & References

- If observed data is unexpected under the prior model, this surprise is often not adequately reflected in posterior inferences when conjugate priors are used (Lin. Reg.: same inflation factor for all $\beta$s).

# Summary & References

- If observed data is unexpected under the prior model, this surprise is often not adequately reflected in posterior inferences when conjugate priors are used (Lin. Reg.: same inflation factor for all $\beta$s).
- Interval probability / sets of probabilities offer the possibility to specify the precision of probability statements explicitly (Lottery B).

# Summary & References

- If observed data is unexpected under the prior model, this surprise is often not adequately reflected in posterior inferences when conjugate priors are used (Lin. Reg.: same inflation factor for all $\beta$s).
- Interval probability / sets of probabilities offer the possibility to specify the precision of probability statements explicitly (Lottery B).
- Generalized iLUCK-models offer a general, manageable, and powerful calculus for Bayesian inference with sets of priors, allowing for adequate reaction to prior-data conflict by increased imprecision.

# Summary & References

- If observed data is unexpected under the prior model, this surprise is often not adequately reflected in posterior inferences when conjugate priors are used (Lin. Reg.: same inflation factor for all $\beta$s).
- Interval probability / sets of probabilities offer the possibility to specify the precision of probability statements explicitly (Lottery B).
- Generalized iLUCK-models offer a general, manageable, and powerful calculus for Bayesian inference with sets of priors, allowing for adequate reaction to prior-data conflict by increased imprecision.

Walter, G. , Augustin, T.: Bayesian linear regression — different conjugate models and their (in)sensitivity to prior-data conflict. In: Kneib, T. , Tutz, G. (eds.), *Statistical Modelling and Regression Structures – Festschrift in the Honour of Ludwig Fahrmeir*, 2010.

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