Inhalt

Wiederholung: Bayes-Inferenz

- Parameter $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ nicht deterministisch, sondern als zufällig angenommen.
- Volles Wahrscheinlichkeitsmodell für alle beobachteten und unbeobachteten Größen bestehend aus:
  1. ***Beobachtungsmodell:*** bedingte Verteilung der Daten gegeben unbekannte Parameter $\boldsymbol{\theta}$, $p(\boldsymbol{y}|\boldsymbol{\theta})$.
  2. ***Priori-Verteilung*** $p(\boldsymbol{\theta})$: drückt Vorwissen/Annahmen über $\boldsymbol{\theta}$ aus.
- Mathematisch günstig: zum Beobachtungsmodell ***konjugierte Prioris***. (Prioris und Posterioris in der gleichen Verteilungsfamilie.)
- Informationsgehalt der Priori: ***nicht*** oder ***schwach informative Prioris***.

Wiederholung: Bayes-Inferenz

- **Statistische Schlüsse** basieren auf der **Posteriori-Verteilung** $p(\boldsymbol{\theta}|\boldsymbol{y})$, $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, der bedingten Verteilung der unbeobachteten Größen gegeben die beobachteten Daten.
- **Berechnung** mit dem Satz von Bayes:

$$p(\boldsymbol{\theta}|\boldsymbol{y}) = \frac{p(\boldsymbol{y}|\boldsymbol{\theta})\, p(\boldsymbol{\theta})}{\int_{\boldsymbol{\Theta}} p(\boldsymbol{y}|\boldsymbol{\theta})\, p(\boldsymbol{\theta})\, d\boldsymbol{\theta}} \propto p(\boldsymbol{y}|\boldsymbol{\theta})\, p(\boldsymbol{\theta})$$

  mit der **Normierungskonstanten** der Posteriori-Dichte im Nenner.

Wiederholung: Bayes-Inferenz

- Übliche **Punktschätzer** der Bayesianischen Inferenz:
  - Posteriori-Erwartungswert $\hat{\boldsymbol{\theta}} = E(\boldsymbol{\theta}|\boldsymbol{y})$,
  - Posteriori-Median $\hat{\boldsymbol{\theta}} = \inf\{\boldsymbol{\theta} : F(\boldsymbol{\theta}|\boldsymbol{y}) \geq 0.5\}$,
  - Posteriori-Modus $\hat{\boldsymbol{\theta}} = \arg\max\{p(\boldsymbol{\theta}|\boldsymbol{y})\}$.

- **Problem:** Posteriori-Verteilung meist analytisch unzugänglich

- **Lösung**: Verwendung von ***MCMC-Verfahren***, mit denen (abhängige) Zufallszahlen aus der Posteriori Verteilung gezogen werden können.

Grundidee MCMC

- Konstruiere Markov-Kette (MK), deren stationäre Verteilung mit der Posteriori Verteilung übereinstimmt.
- Zustände der MK entsprechen gezogenen Zufallszahlen, die (nach entsprechender Konvergenzzeit (Burn-In-Phase) der MK) abhängige Stichprobe aus Posteriori darstellen.
- Abhängigkeit kann durch geeignetes Ausdünnen der Stichprobe reduziert werden.
- Interessierende Größen (z.B. P.-Erwartungswert), werden dann aus dieser (ausgedünnten) Stichprobe durch die **empirischen Analoga** geschätzt.
- Bekanntester Algorithmus: ***Metropolis-Hastings-Algorithmus***, Spezialfall: ***Gibbs-Sampler***

Beobachtungsmodell

Beobachtungsmodell:

$$\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{b}, \boldsymbol{\vartheta} \sim N(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{b}, \boldsymbol{R}(\boldsymbol{\vartheta}))$$

entspricht $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{b} + \boldsymbol{\varepsilon}$ mit $\boldsymbol{\varepsilon} \sim N(\boldsymbol{0}, \boldsymbol{R}(\boldsymbol{\vartheta}))$.

Priori für β

- Jetzt auch „**feste**“ **Effekte** β als **Zufallsgrößen**
- kein Vorwissen über $\beta \Rightarrow$ ***nichtinformative Priori***, d.h.

$$p(\boldsymbol{\beta}) \propto const,$$

  sonst

$$\boldsymbol{\beta} \sim N(\boldsymbol{m}, \boldsymbol{M}) \text{ mit bekanntem EWert } \boldsymbol{m}, \text{ Kovarianz } \boldsymbol{M}.$$

- nichtinformative Priori $p(\boldsymbol{\beta}) \propto const$ ergibt sich als Grenzfall der Priori-Normalverteilung (NV) für ***Präzisionsmatrix*** $\boldsymbol{M}^{-1} \to \mathbf{0}$.
- zum Beobachtungsmodell **konjugierte Normalverteilung** für $\boldsymbol{\beta}$ $\Rightarrow$ Posteriori-Inferenz vergleichsweise einfach.

Prioris für b, ε

Üblicherweise:

$$\boldsymbol{b} \sim N(\mathbf{0}, \boldsymbol{G}(\vartheta)); \quad \varepsilon \sim N(\mathbf{0}, \boldsymbol{R}(\vartheta))$$

- Kovarianzmatrizen $\boldsymbol{G} = \text{Cov}(\boldsymbol{b})$ und $\boldsymbol{R} = \text{Cov}(\varepsilon)$ hängen i.A. von unbekannten ***Hyperparametern*** im Vektor ϑ ab.
- ***Voller Bayes-Ansatz:***
  ϑ ebenfalls **Zufallsvariable**, mit (Hyper-)Priori $p(\vartheta)$, die in Ermittlung der Posteriori mit einfließt.
- ***empirischer Bayes-Ansatz:***
  ϑ als unbekannter, aber **fester Parameter**.

Weitere Annahme: Zufallsgrößen β, $\boldsymbol{b}$ und ε a priori **unabhängig**.

Gemeinsame Posteriori bei NV-Priori für β und b

$$
\begin{aligned}
p(\boldsymbol{\beta}, \boldsymbol{b}|\boldsymbol{y}) &\propto p(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{b}) p(\boldsymbol{\beta}) p(\boldsymbol{b}) \\
&\propto \exp\left(-\frac{1}{2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{Z}\boldsymbol{b})'\boldsymbol{R}^{-1}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{Z}\boldsymbol{b}) \right. \\
&\qquad \left. - \frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{m})'\boldsymbol{M}^{-1}(\boldsymbol{\beta} - \boldsymbol{m}) - \frac{1}{2}\boldsymbol{b}'\boldsymbol{G}^{-1}\boldsymbol{b}\right).
\end{aligned}
$$

$\Rightarrow \begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{b} \end{pmatrix} \Big| \boldsymbol{y} \sim N(\boldsymbol{\mu}_{\beta,b}, \boldsymbol{\Sigma}_{\beta,b})$ **(Übung)** mit

$$
\boldsymbol{\Sigma}_{\beta,b} = \left(\boldsymbol{C}'\boldsymbol{R}^{-1}\boldsymbol{C} + \boldsymbol{A}\right)^{-1}; \quad \boldsymbol{C} = [\boldsymbol{X}|\boldsymbol{Z}]; \quad \boldsymbol{A} = \begin{bmatrix} \boldsymbol{M}^{-1} & 0 \\ 0 & \boldsymbol{G}^{-1} \end{bmatrix}
$$

$$
\boldsymbol{\mu}_{\beta,b} = \boldsymbol{\Sigma}_{\beta,b}\left(\tilde{\boldsymbol{m}} + \boldsymbol{C}'\boldsymbol{R}^{-1}\boldsymbol{y}\right); \quad \tilde{\boldsymbol{m}} = \begin{pmatrix} \boldsymbol{M}^{-1}\boldsymbol{m} \\ \boldsymbol{0} \end{pmatrix}
$$

Posteriori bei nichtinformativer Priori für β

Nichtinformative Priori $p(\boldsymbol{\beta}) \propto \textit{const}$ entspricht Präzisionsmatrix $\boldsymbol{M}^{-1} = \boldsymbol{0}$:

$$
\begin{aligned}
p(\boldsymbol{\beta}, \boldsymbol{b} | \boldsymbol{y}) &\propto p(\boldsymbol{y} | \boldsymbol{\beta}, \boldsymbol{b}) p(\boldsymbol{b}) \\
&\propto \exp\left(-\frac{1}{2} (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{Z}\boldsymbol{b})' \boldsymbol{R}^{-1} (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{Z}\boldsymbol{b}) - \frac{1}{2} \boldsymbol{b}' \boldsymbol{G}^{-1} \boldsymbol{b} \right) \\
&\Rightarrow \text{Posteriori-EW für } \begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{b} \end{pmatrix}: \ \left(\boldsymbol{C}' \boldsymbol{R}^{-1} \boldsymbol{C} + \boldsymbol{A} \right)^{-1} \boldsymbol{C}' \boldsymbol{R}^{-1} \boldsymbol{y} \\
&\Rightarrow \text{Posteriori-Kov. für } \begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{b} \end{pmatrix}: \ \left(\boldsymbol{C}' \boldsymbol{R}^{-1} \boldsymbol{C} + \boldsymbol{A} \right)^{-1}
\end{aligned}
$$

- **Posteriori** äquivalent zu **penalisiertem KQ-Kriterium** $KQ_{pen}(\boldsymbol{\beta}, \boldsymbol{b})$ in (16)
- **Posteriori-Modus** als Maximierer identisch mit **BLUP-Schätzern**, die $KQ_{pen}(\boldsymbol{\beta}, \boldsymbol{b})$ minimieren. **Posteriori-EW** = Posteriori-Modus wegen NV.
- Posteriori-Kov. identisch mit Kovarianz in (21).

Empirische Bayes-Schätzung

- ***Empirische Bayes-Schätzer*** für $\boldsymbol{\beta}, \boldsymbol{b}$ durch Einsetzen der geschätzten Kovarianzmatrizen $\hat{\boldsymbol{G}} = \boldsymbol{G}(\hat{\boldsymbol{\vartheta}})$ und $\hat{\boldsymbol{R}} = \boldsymbol{R}(\hat{\boldsymbol{\vartheta}})$ in vorige Ausdrücke.
- Schätzung von $\hat{\boldsymbol{\vartheta}}$ durch Maximieren der **marginalen Likelihood** für $\boldsymbol{\vartheta}$:

$$\hat{\boldsymbol{\vartheta}} = \arg\max p(\boldsymbol{y}|\boldsymbol{\vartheta}) = \arg\max \int p(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{b}, \boldsymbol{\vartheta}) p(\boldsymbol{b}|\boldsymbol{\vartheta}) p(\boldsymbol{\beta}) d\boldsymbol{\beta} d\boldsymbol{b}.$$

- Für nicht-informative $p(\boldsymbol{\beta}) \propto const$ ist die **marginale Likelihood**

$$p(\boldsymbol{y}|\boldsymbol{\vartheta}) = \int p(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{b}, \boldsymbol{\vartheta}) p(\boldsymbol{b}|\boldsymbol{\vartheta}) d\boldsymbol{\beta} d\boldsymbol{b} = \int p(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{\vartheta}) d\boldsymbol{\beta}$$

  proportional zur **restringierten Likelihood** $\exp\{l_R(\boldsymbol{\vartheta})\}$, siehe (20).
  $\Rightarrow$ emp. Bayes-Schätzer in diesem Fall äquivalent zu **REML-Schätzer** $\hat{\boldsymbol{\vartheta}}_{REML}$ und den dazugehörigen EBLUPs für $\boldsymbol{\beta}$ und $\boldsymbol{b}$.

Volle Bayes-Inferenz

- **Voller Bayes-Ansatz**: Priori-Verteilung $p(\boldsymbol{\vartheta})$ auch für unbekannte Parameter $\boldsymbol{\vartheta}$; $\boldsymbol{\beta}$, $\boldsymbol{b}|\boldsymbol{\vartheta}$ und $\boldsymbol{\vartheta}$ als unabhängig angenommen.

- Inferenz basiert auf **Posteriori-Verteilung**

$$p(\boldsymbol{\beta}, \boldsymbol{b}, \boldsymbol{\vartheta}|\boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{b}, \boldsymbol{\vartheta})p(\boldsymbol{\beta})p(\boldsymbol{b}|\boldsymbol{\vartheta})p(\boldsymbol{\vartheta})$$

- $p(\boldsymbol{\beta}, \boldsymbol{b}, \boldsymbol{\vartheta}|y)$ echte Posteriori-Dichte, wenn zur **Normierung** gilt:

$$p(\boldsymbol{y}) = \int p(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{b}, \boldsymbol{\vartheta})p(\boldsymbol{\beta})p(\boldsymbol{b}|\boldsymbol{\vartheta})p(\boldsymbol{\vartheta})d\boldsymbol{\beta}d\boldsymbol{b}d\boldsymbol{\vartheta} < \infty.$$

- Bei echter, **informativer** Priori mit $\int p(\boldsymbol{\vartheta})d\boldsymbol{\vartheta} = 1$ existiert auch $p(\boldsymbol{\beta}, \boldsymbol{b}, \boldsymbol{\vartheta}|\boldsymbol{y})$.

- Für **nichtinformative** Priori mit $\int p(\boldsymbol{\vartheta})d\boldsymbol{\vartheta} = \infty$ Existenz der Posteriori nicht allgemein gesichert.

Beziehungen zur Likelihood-Inferenz

Bei nichtinformativer Priori $p(\boldsymbol{\vartheta}) \propto \mathit{const}$ und Existenz der Posteriori:

- REML-Schätzer $\hat{\boldsymbol{\vartheta}}_{REML}$ als Maximierer der marginalen Likelihood = Posteriori-Modus der marginalen Posteriori von $\boldsymbol{\vartheta}$ wegen

$$p(\boldsymbol{\vartheta}|\boldsymbol{y}) = p(\boldsymbol{y}|\boldsymbol{\vartheta})\frac{p(\boldsymbol{\vartheta})}{p(\boldsymbol{y})} \propto p(\boldsymbol{y}|\boldsymbol{\vartheta}).$$

- Bei zusätzlich nichtinformativer Priori $p(\boldsymbol{\beta}) \propto \mathit{const}$: ML-Schätzer $\hat{\boldsymbol{\vartheta}}_{ML}$ als Maximierer der Likelihood = $\boldsymbol{\vartheta}$-Komponente des Posteriori-Modus der gemeinsamen Posteriori von $\boldsymbol{\beta}$ und $\boldsymbol{\vartheta}$ wegen

$$p(\boldsymbol{\beta},\boldsymbol{\vartheta}|y) = p(\boldsymbol{y}|\boldsymbol{\vartheta},\boldsymbol{\beta})\frac{p(\boldsymbol{\vartheta})p(\boldsymbol{\beta})}{p(\boldsymbol{y})} \propto p(\boldsymbol{y}|\boldsymbol{\vartheta},\boldsymbol{\beta}).$$

Inferenz

- Normierungskonstante der Posteriori i.A. nicht analytisch zugänglich
  $\Rightarrow$ Posteriori-Dichte $p(\boldsymbol{\beta}, \boldsymbol{b}, \boldsymbol{\vartheta} | \boldsymbol{y})$ **nicht in geschlossener Form** darstellbar
- volle Bayes-Inferenz daher üblicherweise mittels ***MCMC-Simulation***
  (Details siehe z.B. Fahrmeir et al. (2007, Abschnitt B.5.3))
- moderne (approximative) Alternativen:
  INLA (Integrated Nested Laplace Approximation),
  Variational Bayes-Verfahren

MCMC mit blockweisem Gibbs-Sampling

Vorgehen: Teile Parametervektor $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{b}, \boldsymbol{\vartheta})$ in **Teilvektoren zusammengehöriger Parameter**, d.h. üblicherweise $\boldsymbol{\beta}$, $\boldsymbol{b}$ und $\boldsymbol{\vartheta}$ auf.

- Wähle Startwerte $\boldsymbol{\beta}^{(0)}$, $\boldsymbol{b}^{(0)}$, $\boldsymbol{\vartheta}^{(0)}$ und Anzahl der Iterationen T
- Bilde ***vollständig bedingte Dichten*** (full conditionals) gegeben der restlichen Parameter und $\boldsymbol{y}$
- Ziehe **sequentiell Zufallszahlen** $\boldsymbol{\beta}^{(t)}$, $\boldsymbol{b}^{(t)}$, $\boldsymbol{\vartheta}^{(t)}$ aus diesen (geg. jeweils die momentan aktuellen Zustände) bis T erreicht.

Nach einer gewissen Konvergenzphase können die Zufallszahlen als Ziehungen aus den Marginalverteilungen von $\boldsymbol{\beta}|\boldsymbol{y}$, $\boldsymbol{b}|\boldsymbol{y}$ und $\boldsymbol{\vartheta}|\boldsymbol{y}$ angesehen werden.

Vollständig bedingte Dichten

$$
\begin{aligned}
p(\boldsymbol{\beta}|\boldsymbol{b},\boldsymbol{\vartheta},\boldsymbol{y}) &\propto p(\boldsymbol{y}|\boldsymbol{\beta},\boldsymbol{b},\boldsymbol{\vartheta})p(\boldsymbol{\beta}) \\
&\propto \exp\Big(-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{Z}\boldsymbol{b})'\boldsymbol{R}(\boldsymbol{\vartheta})^{-1}(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{Z}\boldsymbol{b}) \\
&\qquad -\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{m})'\boldsymbol{M}^{-1}(\boldsymbol{\beta}-\boldsymbol{m})\Big)
\end{aligned}
$$

$$
\begin{aligned}
p(\boldsymbol{b}|\boldsymbol{\beta},\boldsymbol{\vartheta},\boldsymbol{y}) &\propto \exp\Big(-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{Z}\boldsymbol{b})'\boldsymbol{R}(\boldsymbol{\vartheta})^{-1}(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{Z}\boldsymbol{b}) \\
&\qquad -\frac{1}{2}\boldsymbol{b}'\boldsymbol{G}(\boldsymbol{\vartheta})^{-1}\boldsymbol{b}\Big)
\end{aligned}
$$

$$
\begin{aligned}
p(\boldsymbol{\vartheta}|\boldsymbol{\beta},\boldsymbol{b},\boldsymbol{y}) &\propto p(\boldsymbol{y}|\boldsymbol{\beta},\boldsymbol{b},\boldsymbol{\vartheta})p(\boldsymbol{b}|\boldsymbol{\vartheta})p(\boldsymbol{\vartheta}) \\
&\propto |\boldsymbol{R}(\boldsymbol{\vartheta})|^{-1/2}\,|\boldsymbol{G}(\boldsymbol{\vartheta})|^{-1/2}\exp\Big(-\frac{1}{2}\boldsymbol{b}'\boldsymbol{G}(\boldsymbol{\vartheta})^{-1}\boldsymbol{b} \\
&\qquad -\frac{1}{2}(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{Z}\boldsymbol{b})'\boldsymbol{R}(\boldsymbol{\vartheta})^{-1}(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{Z}\boldsymbol{b})\Big)\,p(\boldsymbol{\vartheta})
\end{aligned}
$$

Vollständig bedingte Dichten von β, b

$$\boldsymbol{\beta}|\cdot \sim N(\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta) \text{ mit } (\textbf{Übung})$$

$$\boldsymbol{\Sigma}_\beta = \left(\boldsymbol{X}'\boldsymbol{R}(\vartheta)^{-1}\boldsymbol{X} + \boldsymbol{M}^{-1}\right)^{-1}$$

$$\boldsymbol{\mu}_\beta = \boldsymbol{\Sigma}_\beta\left(\boldsymbol{M}^{-1}\boldsymbol{m} + \boldsymbol{X}'\boldsymbol{R}(\vartheta)^{-1}(\boldsymbol{y} - \boldsymbol{Z}\boldsymbol{b})\right)$$

$$\boldsymbol{b}|\cdot \sim N(\boldsymbol{\mu}_b, \boldsymbol{\Sigma}_b) \text{ mit (analog)}$$

$$\boldsymbol{\Sigma}_b = \left(\boldsymbol{Z}'\boldsymbol{R}(\vartheta)^{-1}\boldsymbol{Z} + \boldsymbol{G}(\vartheta)^{-1}\right)^{-1}$$

$$\boldsymbol{\mu}_b = \boldsymbol{\Sigma}_b\left(\boldsymbol{Z}'\boldsymbol{R}(\vartheta)^{-1}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})\right)$$

- **nichtinformative Priori** mit $\boldsymbol{M}^{-1} = \boldsymbol{0}$: Erwartungswert $\boldsymbol{\mu}_\beta = (\boldsymbol{X}'\boldsymbol{R}(\vartheta)^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{R}(\vartheta)^{-1}(\boldsymbol{y} - \boldsymbol{Z}\boldsymbol{b})$ ist gewichteter KQ-Schätzer angewandt auf die um $\boldsymbol{Z}\boldsymbol{b}$ bereinigten Daten
- **informative Priori** $\Rightarrow$ Erwartungswert $\boldsymbol{\mu}_\beta$ ist gewichtetes Mittel aus KQ-Schätzer und Priori-Erwartungswert.
- Analog für $\boldsymbol{\mu}_b$.

Vollständig bedingte Dichten von ϑ im Spezialfall

LMM für Clusterdaten mit $\text{Cov}(\varepsilon) = \sigma^2 I_n \Rightarrow \vartheta$ enthält σ^2 und Parameter in $\boldsymbol{D}$.

- nichtinformative Jeffreys Prioris $p(\sigma^2) \propto \sigma^{-2}$, $p(\boldsymbol{D}) \propto |\boldsymbol{D}|^{-\frac{q+1}{2}}$ führen i.A. zu uneigentlichen (d.h. nicht normierbaren) Posteriori-Verteilungen

$\Rightarrow$ Schwach informative inverse Gammaverteilung $\sigma^2 \sim IG(a_\sigma, b_\sigma)$; a_σ, b_σ klein

- Für $\boldsymbol{D}$ oft inverse Wishart-Verteilung. Bei $\boldsymbol{D} = \text{diag}(\tau_1^2, \ldots, \tau_q^2)$ mit unabh. τ_j^2 ergibt diese ein Produkt von IGs mit $\tau_j^2 \sim IG(a_{\tau_j}, b_{\tau_j}), j = 1, \ldots, q$.

Dann full conditionals bei zusätzlich $p(\beta) \propto const$ (nichtinformative Priori):

$$\sigma^2 | \cdot \sim IG(\tilde{a}_\sigma, \tilde{b}_\sigma) \text{ mit } \tilde{a}_\sigma = a_\sigma + \frac{1}{2},\ \tilde{b}_\sigma = b_\sigma + \frac{1}{2} ||\boldsymbol{y} - \boldsymbol{X}\beta - \boldsymbol{Z}\boldsymbol{b}||_2^2$$

$$\tau_j^2 | \cdot \sim IG(\tilde{a}_{\tau_j}, \tilde{b}_{\tau_j}) \text{ mit } \tilde{a}_{\tau_j} = a_{\tau_j} + \frac{m}{2},\ \tilde{b}_{\tau_j} = b_{\tau_j} + \frac{1}{2} \sum_{i=1}^{m} b_{ij}^2$$

Probleme der NV-Annahme für b

NV-Annahme für zufällige Effekte $\boldsymbol{b} \sim N(\mathbf{0}, \boldsymbol{G})$ mathematisch günstig. Aber wie sensitiv ist die Schätzung bei Fehlspezifikation?

- Die Schätzer der festen Effekte $\boldsymbol{\beta}$ und der Kovarianzparameter $\boldsymbol{\vartheta}$ sind meist sehr robust gegenüber Fehlspezifikation der Verteilung der $\boldsymbol{b}$.
  Die Standardfehler können jedoch über/unterschätzt werden.

- Durch den Shrinkage-Effekt können die EBLUPs der zufälligen Effekte $\boldsymbol{b}$ normalverteilt aussehen, selbst wenn die Verteilung der $\boldsymbol{b}$ z.B. bimodal / schief ist / hohe Wahrscheinlichkeitsmasse an den Rändern hat (heavy tails)
  $\Rightarrow$ Schlechte Vorhersagen $\hat{\boldsymbol{b}}$. q-q-Plots der $\hat{\boldsymbol{b}}$ eignen sich nicht zur Diagnose. Diagnose durch Fitten eines flexibleren Modells.

Alternative Prioris für b: Skalenmischungen

- Verwendung von Prioris mit mehr Masse auf den Rändern (heavy-tailed): z.B. t-Verteilung mit niedrigen Freiheitsgraden, Laplace-Verteilung. Diese sind oft darstellbar als Skalenmischung von Normalverteilungen

$$p(b_i) = \int \phi(b_i|\mu, \sigma^2) p(\sigma^2|\theta) d\sigma^2$$

- Darstellbar als Skalenmischung $\Rightarrow$ sehr leicht in Modellhierarchie für LMM einzubauen

- Beispiel: t-Verteilung mit $df = \nu$ ist Skalenmischung aus $N(0, \sigma^2)$ mit $\sigma^{-2} \sim \Gamma(\nu/2, \nu/2)$

Alternative Prioris für b: Finite Mixtures

- Zur Aufdeckung von Clustern (z.B. durch unbeobachtete Kovariablen oder latente Subpopulationen) können multimodale Verteilungen für die zufälligen Effekte verwendet werden, z.B. ***finite Mischverteilungsmodelle***:

$$p(\boldsymbol{b}_i|\boldsymbol{\pi}, \phi) = \sum_{k=1}^{K} \pi_k p(\boldsymbol{b}_i|\phi_k),$$

  mit Gewichten $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_k)$, $\sum_k^K \pi_k = 1$, und parametrischer Verteilungsfamilie $p\,(\boldsymbol{b}_i|\phi_k)$ mit Parametern $\phi = (\phi_1, \ldots, \phi_K)$ (z.B. multivariate NV $p\,(\boldsymbol{b}_i|\boldsymbol{\mu}_k, \boldsymbol{\Sigma})$ mit Erwartungswerten $\boldsymbol{\mu}_k$ und homogenen Kovarianzen $\boldsymbol{\Sigma}$)

- nichtparametrische Erweiterung: K nicht fest/konstant, wird mitgeschätzt.

- Noch flexibler: nichtparametrische Bayes-Ansätze, z.B. **Dirichlet-Prozess**-, Dirichlet-Prozess-Mischungs-Prioris.

- Inferenz für diese Modelle basiert i.d.R. auf MCMC-Techniken.