Einführung in die Bayes Statistik- Slides Begleitung zum Online Kurs

Regression und Anwendung

Martje Rave

Sommersemester 2025

ACHTUNG!

Diese Folien sind **NUR** Kurs begleitend und können noch viele Fehler beinhalten. Für die Klausurvorbereitung benutzt bitte den Online-Kurs und die Übungsblätter.
Bitte gebt mir Bescheid, wenn ihr Fehler findet.

Multiple Regression

Wir erweitern unser lineares Regressionsmodell für y_i mit mehreren Kovariablen $x_1, \ldots, x_p$:

$$\begin{aligned} y_i &= \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon_i, \quad i = 1, \ldots, n \\ \varepsilon_i &\sim \mathcal{N}(0, \sigma^2), \\ \mathrm{Cov}(\varepsilon_i, \varepsilon_j) &= 0 \end{aligned}$$

Mit $\beta_0 := \alpha$ und $x_{i0} = 1$ schreiben wir:

$$y_i = \sum_{j=0}^{p} \beta_j x_{ij} + \varepsilon_i$$

Datendichte

Aus dem Modell folgt:

$$y_i \mid \boldsymbol{\beta}, \sigma^2 \sim \mathcal{N}\left(\sum_{j=0}^{p} \beta_j x_{ij}, \sigma^2\right)$$

Die gemeinsame Dichte aller $\boldsymbol{y} = (y_1, \ldots, y_n)$:

$$\begin{aligned} f(\boldsymbol{y}|\boldsymbol{\beta}, \sigma^2) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}\left(y_i - \sum_{j=0}^{p} \beta_j x_{ij}\right)^2\right) \\ &= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} \left(y_i - \sum_{j=0}^{p} \beta_j x_{ij}\right)^2\right) \end{aligned}$$

Priori der Regressionsparameter

- $\beta_j \sim \mathcal{N}(m_j, v_j^2), \quad j = 0, \ldots, p$
- a priori seien die β_j alle voneinander unabhängig.

Priori der Varianz

$$\sigma^2 \sim IG(a, b)$$

Gemeinsame Priori

$$p(\beta, \sigma^2) = \prod_{j=0}^{p} p(\beta_j) \cdot p(\sigma^2)$$

Posteriori

$$p(\beta, \sigma^2|\boldsymbol{y}) \propto f(\boldsymbol{y}|\beta, \sigma^2)p(\beta, \sigma^2)$$

$$\propto \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(y_i - \sum_{j=0}^{p}\beta_j x_{ij}\right)^2\right) \cdot \prod_{j=0}^{p}\exp\left(-\frac{1}{2v_j^2}(\beta_j - m_j)^2\right)$$
$$\cdot (\sigma^2)^{-(a+1)}\exp\left(-\frac{b}{\sigma^2}\right)$$

Vollständig bedingte Posterioris

Zur Erinnerung:

$$f(\theta_i|\theta_{-i}) \propto \frac{f(\theta)}{f(\theta_{-i})} \propto f(\theta)$$

Full conditional von β_j

$$p(\beta_j|\boldsymbol{\beta}_{-j},\sigma^2,\boldsymbol{y}) \propto \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(y_i-\sum_{k\neq j}\beta_k x_{ik}-\beta_j x_{ij}\right)^2\right)\cdot\exp\left(-\frac{1}{2v_j^2}(\beta_j-m_j)^2\right)$$

Dies ergibt eine Normalverteilung:

- Varianz: $v_j^{*2} = \left(\frac{1}{v_j^2}+\frac{1}{\sigma^2}\sum_{i=1}^{n} x_{ij}^2\right)^{-1}$
- Erwartungswert: $m_j^* = v_j^{*2}\left(\frac{m_j}{v_j^2}+\frac{1}{\sigma^2}\sum_{i=1}^{n} x_{ij}(y_i-\sum_{k\neq j}\beta_k x_{ik})\right)$

Full conditional von σ^2

$$
\begin{aligned}
a^* &= a + n/2 \\
b^* &= b + 0.5 \sum_{i=1}^{n} \left(y_i - \sum_{j=0}^{p} \beta_j x_{ij} \right)^2
\end{aligned}
$$

- Der Full conditional ist wieder eine inverse-Gamma-Verteilung.
- Strukturell identisch zur einfachen Regression.

Modellierung mit JAGS

- Die Datendichte jeder Beobachtung $i = 1, \ldots, n$ ist:

$$\text{Fertility}_i \sim \mathcal{N}(\mu_i, \tau), \quad \mu_i = \sum_j \beta_j x_{ij}$$

- Die Priors:

$$\beta_j \sim \mathcal{N}(0, 10^6), \quad \tau \sim \text{Gamma}(0{,}001, 0{,}001)$$

- Varianz: $\sigma^2 = 1/\tau$

JAGS Modell (rjags)

```
linmodell <- "model{
  for (i in 1:n) {
    Fertility[i] ~ dnorm(mu[i], tau)
    mu[i] <- beta[1] + beta[2] * Agriculture[i] + beta[3] * Examina
             beta[4] * Education[i] + beta[5] * Catholic[i] +
             beta[6] * Infant.Mortality[i]
  }
  # Priors
  for (j in 1:p) {
    beta[j] ~ dnorm(0, 0.001)
  }
  tau ~ dgamma(0.001, 0.001)
  sigma2 <- 1/tau
}"
```

Daten vorbereiten

```
daten <- as.list(swiss)
daten$p <- dim(swiss)[2]
daten$n <- dim(swiss)[1]

startwerte <-
  list(beta = c(60, 0.2, 0, 0, 0, 0), tau = 1/144)
```

Kompilieren des Modells

```
model <- jags.model(file = textConnection(linmodell),
                    data = daten, inits = startwerte)
```

MCMC Durchführung

Burn-in (10.000 Iterationen):

```
update(model, 10000, progress.bar="none")
```

Ziehe 20.000 Samples:

```
samp <- jags.samples(model,
  variable.names=c("beta","sigma2"),
  n.iter=20000, progress.bar="none")
```

```r
summary(samp$beta$mean)
summary(samp$beta$median)
summary(samp$beta$quantile, c(0.025, 0.975))
```

Posterior-Ergebnisse

	Intercept	Agriculture	Examination	Education	Catholic	InfantMort.
Mittelwert	58.319	-0.134	-0.159	-0.853	0.107	1.310
Median	58.323	-0.134	-0.158	-0.854	0.107	1.319
2.5%	37.178	-0.275	-0.650	-1.215	0.035	0.577
97.5%	78.241	0.008	0.320	-0.470	0.175	2.036

Interpretation der Ergebnisse

- **Infant.Mortality:** positiver Einfluss auf Fruchtbarkeit.
- **Catholic:** leichter positiver Effekt.
- **Education:** negativer Effekt – je mehr Bildung, desto weniger Fruchtbarkeit.
- **Examination:** kein signifikanter Effekt (95% Intervall enthält 0).

Hierarchische Modelle

Bisher haben wir nur Regressionsmodelle behandelt, die man auch frequentistisch problemlos inferieren kann. Jetzt gehen wir zu komplexeren Modellen über.

Beispiel: Zeitreihe mit Saison-Effekt

Als Beispiel sehen wir uns folgende Daten an:

Anzahl von getöteten oder schwer verletzten Autofahrern in England von Januar 1969 bis Dezember 1984

Beobachtungen

Wir sehen folgende Trends in den Daten:

- Es gibt einen Saisoneffekt, im Winter gibt es mehr Fälle als im Sommer
- Anfang 1983 sinkt die Anzahl der Fälle drastisch (zu diesem Zeitpunkt wurde die Gurtpflicht in Großbritannien eingeführt)
- Zwischen 1974 und 1983 bleibt die Anzahl der Fälle relativ konstant, davor schwankt sie stärker

Diese Trends wollen wir nun im Modell einbauen.

Datenmodell

Wir nehmen y_t als normalverteilt an (für Fallzahlen wäre auch Poissonverteilung denkbar):

$$\sqrt{y_t} \sim \mathcal{N}(\mu_t, \sigma^2), \quad t = 1, \ldots, T = 192$$

Nun teilen wir den Erwartungswert μ_t in verschiedene additive Effekte auf:

$$\mu_t = \alpha + \beta x_t + \gamma_t + \delta_t$$

- t ist der Monat, x_t Dummyvariable Gurtpflicht ja/nein.
- α der Intercept
- β der Einfluss der Gurtpflicht
- γ_t der allgemeine Zeittrend
- δ_t der Saisoneffekt

Priori-Modell – Zeittrend

Wir wollen die zeitliche Entwicklung modellieren.

Problematik:

- Eine lineare Entwicklung ist vermutlich nicht passend.
- Eine parametrisierte Modellierung wäre denkbar – aber welches Modell sollten wir wählen?

Idee: Wenn wir saisonale Effekte und den Effekt der Gurtpflicht herausrechnen, sollten benachbarte Monate ähnliche Fallzahlen aufweisen.

- Dummykodierung: ein Parameter γ_i pro Monat.
- Monat i hat ähnliche Fallzahlen wie Monat $i - 1$.

Random-Walk-Prior für Zeittrend

Formalisierung als Verteilung:

$$\gamma_i \mid \gamma_{i-1}, \tau_c \sim \mathcal{N}(\gamma_{i-1}, \tau_c^{-1}), \quad i = 2, \dots, T$$

- Für Monat i erwarten wir die gleiche Fallzahl wie in Monat $i-1$.
- τ_c^{-1} ist die Varianz (je größer die Präzision τ_c, desto ähnlicher sind sich die Monate).
- Für den ersten Monat: nicht-informative Priori: $p(\gamma_1) \propto \text{const}$

Begriffe:

- *Random Walk (RW1)*: jeder Schritt ist zufällige Abweichung vom vorherigen Punkt.
- Höhere Ordnung (z.B. RW2) bezieht mehrere vergangene Zeitpunkte ein.

Hyperpriors auf die Präzision

Wir haben mit τ_c einen *Hyperparameter* eingeführt – die Präzision unserer Annahme.

- Je größer τ_c, desto ähnlicher sind sich aufeinanderfolgende Monate.
- Je kleiner τ_c, desto größer darf die Varianz zwischen Monaten sein.

Statt τ_c festzulegen, behandeln wir ihn als unbekannt:

- Im bayesianischen Ansatz geben wir eine Priorverteilung für τ_c an.
- Bisher haben wir für Präzisionen (inverse Varianzen) stets die Gamma-Verteilung verwendet:

$$\tau_c \sim \text{Ga}(a_c, b_c)$$

Hierarchisches Bayes-Modell (HBM)

Wir erhalten ein **hierarchisches Modell** mit mehreren Ebenen:

- **Level 1:** Datenmodell – Likelihood $p(y \mid \theta)$
- **Level 2:** Prioris für die unbekannten Parameter θ (z.B. γ_i)
- **Level 3:** Hyperprioris für die Prioriparameter (z.B. τ_c)

Dieses Schichtenmodell ist typisch für viele realistische bayesianische Modelle.

Priorimodell – Saisontrend

- Ziel: Modellierung des **Saisontrends**
- Parametrische Modelle (z.B. Sinus) wären möglich, aber zu unflexibel.
- Stattdessen nutzen wir **Vorwissen**: Monats-Effekte δ_i gleichen sich über das Jahr aus.

Annahme:

$$\sum_{j=1}^{12} \delta_{i+j} \sim \mathcal{N}(0, \tau_d^{-1}), \quad \text{für } i = 0, \ldots, T-12$$

Gemeinsame Priori:

$$p(\boldsymbol{\delta} \mid \tau_d) \propto (\tau_d)^{(T-12+1)/2} \cdot \exp\left(-\frac{\tau_d}{2} \sum_{i=0}^{T-12} \left(\sum_{j=1}^{12} \delta_{i+j}\right)^2\right)$$

Hyperprior für Saisoneffekt

Auch τ_d (Präzision des Saisoneffekts) soll nicht festgelegt werden, sondern geschätzt.

Bayesianisch:

- τ_d ist ein unbekannter Parameter $\Rightarrow$ bekommt eine **Priorverteilung**
- Übliche Wahl: Gamma-Verteilung

$$\tau_d \sim \mathrm{Ga}(a_d, b_d)$$

Priorimodell – Intercept und Kovariableneffekt

Es fehlt noch eine Priori für:

- Intercept α
- Kovariableneffekt β

Nicht-informative Priori:

$$p(\alpha) \propto \text{const}, \quad p(\beta) \propto \text{const}$$

Prior für die Varianz der Fehler:

$$\sigma^2 \sim \mathrm{IG}(a_s, b_s)$$

Posteriori – Gesamtes Modell

Sei $y^* = \sqrt{y}$. Dann:

$$p(\alpha, \beta, \boldsymbol{\gamma}, \boldsymbol{\delta}, \tau_c, \tau_d, \sigma^2 \mid y^*) \propto f(y^* \mid \alpha, \beta, \boldsymbol{\gamma}, \boldsymbol{\delta}, \sigma^2) \cdot p(\alpha, \beta, \boldsymbol{\gamma}, \boldsymbol{\delta}, \tau_c, \tau_d, \sigma^2)$$

Likelihood unabhängig von τ_c, τ_d:

$$f(y^* \mid \alpha, \beta, \boldsymbol{\gamma}, \boldsymbol{\delta}, \tau_c, \tau_d, \sigma^2) = f(y^* \mid \alpha, \beta, \boldsymbol{\gamma}, \boldsymbol{\delta}, \sigma^2)$$

Zerlegung der gemeinsamen Priori

Annahme: Unabhängigkeit der Parameter:

$$p(\alpha, \beta, \boldsymbol{\gamma}, \boldsymbol{\delta}, \tau_c, \tau_d, \sigma^2) = p(\alpha) \cdot p(\beta) \cdot p(\boldsymbol{\gamma} \mid \tau_c) \cdot p(\tau_c) \cdot p(\boldsymbol{\delta} \mid \tau_d) \cdot p(\tau_d) \cdot p(\sigma^2)$$

Gesamte Posteriori:

$$p(\alpha, \beta, \boldsymbol{\gamma}, \boldsymbol{\delta}, \tau_c, \tau_d, \sigma^2 \mid y^*) \propto f(y^* \mid \alpha, \beta, \boldsymbol{\gamma}, \boldsymbol{\delta}, \sigma^2) \cdot p(\alpha)p(\beta)p(\boldsymbol{\gamma} \mid \tau_c)p(\tau_c)p(\boldsymbol{\delta} \mid \tau_d)p(\tau_d)p(\sigma^2)$$

Full Conditional für τ_c

Zur MCMC-Simulation benötigen wir vollständige bedingte Verteilungen (Full Conditionals).

Für τ_c:

$$p(\tau_c \mid \cdot) \propto p(\gamma \mid \tau_c) \cdot p(\tau_c)$$

Beobachtungen:

- $p(\tau_c \mid \cdot)$ hängt nicht direkt von den Daten y^* ab.
- Indirekt jedoch über γ.
- Hierarchische Struktur: jede Ebene hängt nur von der über- und untergeordneten ab.

Folgerung:

- $p(\tau_c \mid \cdot)$ ist proportional zu Gamma-Verteilung
- Die Priori ist semi-konjugiert

Priorimodell – Saisontrend

- Wir wollen den **Saisontrend** modellieren.
- Parametrische Modelle (z. B. Sinus) wären zu unflexibel.
- Stattdessen verwenden wir Vorwissen über die glatte Struktur des Saisontrends:

$$p(\boldsymbol{\delta} \mid \tau_d) \propto (\tau_d)^{(T-12+1)/2} \cdot \exp\left(-\frac{\tau_d}{2} \sum_{i=0}^{T-12} (\delta_{i+1} + \delta_{i+2} + \cdots + \delta_{i+12})^2\right)$$

Matrixform:

$$\boldsymbol{\delta} \mid \tau_d \sim \mathcal{N}_T(0, (\tau_d Q_d)^{-1})$$

Saisonmatrix Q_d

- Die Matrix Q_d entsteht durch Blocksummen von 12 aufeinanderfolgenden Monaten.
- Struktur: Bandmatrix mit Hauptdiagonale und 11 Nebendiagonalen.

In R erzeugbar durch:

```
Qd <- Matrix::sparseMatrix(i=T, j=T, x=0, symmetric=TRUE)
EinsM <- Matrix::sparseMatrix(i=rep(1:12, 12),
                              j=rep(1:12, each=12), x=1)
for (i in 1:(T-12))
  Qd[i+(0:11), i+(0:11)] <- Qd[i+(0:11), i+(0:11)] + EinsM
image(Qd)
```

Folge: Q_d ist positiv semidefinit mit klarer Bandstruktur.

Priors für Intercept und Kovariableneffekt

- Keine spezielle Prior-Information $\Rightarrow$ flache Prior:

$$p(\alpha) \propto \text{const}, \quad p(\beta) \propto \text{const}$$

- Interpretierbar als:

$$\alpha \sim \mathcal{N}(0, 0^{-1}), \quad \beta \sim \mathcal{N}(0, 0^{-1})$$

- Alle Effekte a priori normalverteilt mit unbekannter Präzision.

Gemeinsame Priori über $\boldsymbol{\theta} = (\alpha, \beta, \boldsymbol{\gamma}, \boldsymbol{\delta})$:

$$\boldsymbol{\theta} \mid \boldsymbol{\tau}, Q(\boldsymbol{\tau}) \sim \mathcal{N}_{2+2T}\left(0, \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & \tau_c Q_c & 0 & 0 \\ 0 & 0 & \tau_d Q_d & \end{pmatrix}^{-1}\right)$$

Hyperpriors

- Die Präzisionsparameter τ_c und τ_d regeln die Glattheit der Effekte.
- Wir möchten sie **nicht fest vorgeben**, sondern aus den Daten lernen.
- Dafür: Hyperpriors

$$\tau_c \sim \mathrm{Ga}(a_c, b_c), \quad \tau_d \sim \mathrm{Ga}(a_d, b_d), \quad \sigma^2 \sim \mathrm{IG}(a_s, b_s)$$

Interpretation: Umso größer τ_c, desto glatter ist der Zeittrend.

Posterioriverteilung

Sei $y^* = \sqrt{y}$

Gesamte Posteriori:

$$p(\boldsymbol{\theta}, \boldsymbol{\tau} \mid y^*) \propto f(y^* \mid \boldsymbol{\theta}) \cdot p(\boldsymbol{\theta} \mid \boldsymbol{\tau}) \cdot p(\boldsymbol{\tau})$$

$$\propto \prod_{i=1}^{T} \sigma^{-1} \exp\left(-\frac{1}{2\sigma^2}(y_i^* - \alpha - \beta x_i - \gamma_i + \delta_i)^2\right) \cdot \tau_c^{(T-1)/2} \exp\left(-\frac{\tau_c}{2}\gamma^T Q_c \gamma\right) \cdot \tau_d^{(T-m+1)/2} \exp\left(-\frac{\tau_d}{2}\delta^T Q_d$$

$$\cdot \tau_c^{a_c - 1} e^{-b_c \tau_c} \cdot \tau_d^{a_d - 1} e^{-b_d \tau_d} \cdot (\sigma^2)^{-a_s - 1} e^{-b_s/\sigma^2}$$

Full Conditionals – Zeitlicher Trend

Full conditional für γ:

$$p(\gamma \mid \cdot) \propto \exp\left(-\frac{1}{2\sigma^2}\sum_i (y_i^* - \gamma_i - \delta_i - \beta x_i - \alpha)^2\right) \cdot \exp\left(-\frac{\tau_c}{2}\gamma^T Q_c \gamma\right)$$

$$\propto \exp\left(-\frac{1}{2}\gamma^T \left(\frac{T}{\sigma^2} I + \tau_c Q_c\right)\gamma + \frac{1}{\sigma^2}\gamma^T \boldsymbol{m}_c\right)$$

mit:

$$Q_c^* = \frac{1}{\sigma^2} I + \tau_c Q_c, \quad \boldsymbol{m}_c = \frac{1}{\sigma^2}\left(y_i^* - \alpha - \beta x_i - \delta_i\right)$$

Ergebnis:

$$\boldsymbol{\gamma} \mid \cdot \sim \mathcal{N}\left((Q_c^*)^{-1}\boldsymbol{m}_c, (Q_c^*)^{-1}\right)$$

Full Conditional – Saison-Effekt

Analog für δ:

$$\boldsymbol{\delta} \mid \cdot \sim \mathcal{N}\left((Q_d^*)^{-1}\boldsymbol{m}_d, (Q_d^*)^{-1}\right)$$

mit:

$$Q_d^* = \frac{1}{\sigma^2} I + \tau_d Q_d, \quad \boldsymbol{m}_d = \frac{1}{\sigma^2}\left(y_i^* - \alpha - \beta x_i - \gamma_i\right)$$

Full Conditional – Kovariableneffekt und Intercept

Für den Kovariableneffekt β:

$$p(\beta \mid \cdot) \propto \exp\left(-\frac{1}{2\sigma^2}\sum_i (y_i^* - \alpha - \gamma_i - \delta_i - \beta x_i)^2\right)$$

$$\Rightarrow \beta \mid \cdot \sim \mathcal{N}(q_b^{-1} m_b, q_b^{-1})$$

mit:

$$q_b = \sum x_i^2/\sigma^2, \quad m_b = \frac{1}{\sigma^2}\sum x_i(y_i^* - \alpha - \gamma_i - \delta_i)$$

Analog für Intercept α:

$$\alpha \mid \cdot \sim \mathcal{N}(q_a^{-1} m_a, q_a^{-1})$$

mit:

$$q_a = \frac{T}{\sigma^2}, \quad m_a = \frac{1}{\sigma^2}\sum (y_i^* - x_i\beta - \gamma_i - \delta_i)$$

Full Conditional – Präzisionsparameter

Für τ_c:

$$p(\tau_c \mid \cdot) \propto \tau_c^{(T-1)/2+a_c-1} \exp\left(-\tau_c\left(b_c + \frac{1}{2}\gamma^T Q_c \gamma\right)\right) \Rightarrow \tau_c \mid \cdot \sim \mathrm{Gamma}\left(a_c + \frac{T-1}{2}, b_c + \frac{1}{2}\gamma^T Q_c \gamma\right)$$

Für τ_d:

$$\tau_d \mid \cdot \sim \mathrm{Gamma}\left(a_d + \frac{T-m+1}{2}, b_d + \frac{1}{2}\boldsymbol{\delta}^T Q_d \boldsymbol{\delta}\right)$$

Für σ^2:

$$\sigma^2 \mid \cdot \sim \mathrm{IG}\left(a_s + \frac{T}{2},\ b_s + \sum_i (y_i^* - \alpha - \beta x_i - \gamma_i - \delta_i)^2\right)$$

Implementierung: Vorarbeiten

- Transformation der Zielvariable: `y <- sqrt(Drivers$y)`
- Gurtpflicht-Covariate: `belt <- Drivers$belt`
- Hyperparameter:

```
a.c <- a.d <- 1
b.c <- 0.0005; b.d <- 0.1
a.s <- 1; b.s <- 1

alpha <- mean(y); beta <- 0
gamma <- rep(0,T); delta <- rep(0,T)
tau.c <- a.c/b.c; tau.d <- a.d/b.d
sigma2 <- b.s/a.s
sumx2 <- sum(belt)

burnin <- 2000; nr.it <- 10000
I <- burnin + nr.it
```

Speicherstruktur und Initialisierung

```
alpha.save <- beta.save <- sigma2.save <- rep(0, nr.it)
gamma.save <- array(0, c(T, nr.it))
delta.save <- array(0, c(T, nr.it))
tau.save <- array(0, c(2, nr.it))
iter <- 0
```

Gibbs-Sampler (1/2)

```
while (iter < I) {
  iter <- iter + 1
  gamma <- gamma - mean(gamma)
  delta <- delta - mean(delta)

  m <- sum(y - gamma - delta - belt * beta) / sigma2
  Q <- T / sigma2
  alpha <- rnorm(1, m / Q, 1 / Q)

  m <- sum(belt * (y - gamma - delta - alpha)) / sigma2
  Q <- sumx2 / sigma2
  beta <- rnorm(1, m / Q, 1 / Q)
```

Gibbs-Sampler (2/2)

```
m <- (y - delta - beta * belt - alpha) / sigma2
Q <- tau.c * Qc + Matrix::Diagonal(T) / sigma2
gamma <- rwc::rnorm.Q(Q, as.matrix(m), canon = TRUE)[, 1]

m <- (y - gamma - beta * belt - alpha) / sigma2
Q <- tau.d * Qd + Matrix::Diagonal(T) / sigma2
delta <- rwc::rnorm.Q(Q, as.matrix(m), canon = TRUE)[, 1]

tau.c <- rgamma(1, a.c + (T - 1) / 2, b.c + (t(gamma) %*% Qc %*% gamma / 2)
tau.d <- rgamma(1, a.d + (T - 12 + 1) / 2, b.d + (t(delta) %*% Qd %*% delta
sigma2 <- 1 / rgamma(1, a.s + T / 2,
           b.s + sum((y - alpha - gamma - delta - beta * belt)^2))
```

Ergebnisse speichern

```
  if (iter > burnin) {
    alpha.save[iter - burnin] <- alpha
    beta.save[iter - burnin] <- beta
    gamma.save[, iter - burnin] <- gamma
    delta.save[, iter - burnin] <- delta
    tau.save[, iter - burnin] <- c(tau.c, tau.d)
    sigma2.save[iter - burnin] <- sigma2
  }
}
```

Variablenselektion: Motivation

- Viele Kovariablen, Ziel: Auswahl relevanter Variablen.
- Beispiel: 442 Diabetes-Patienten, mit standardisierten Variablen:
  - Alter, Geschlecht, BMI, Blutdruck
  - Sechs Blutwerte: tc, ldl, hdl, tch, ltg, glu
- Zielgröße: Maß y_i für Krankheitsfortschritt

Bayessche Ridge-Regression

- Modell: $y_i \sim \mathcal{N}(x_i^T \beta, \sigma^2)$
- Priori: $\beta_j \sim \mathcal{N}(0, \tau^{-1})$ für alle j
- Hyperpriori: $\tau \sim \text{Gamma}(a, b)$
- Regularisierung drückt β_j Richtung Null

Posteriori-Verteilungen

- $\beta \mid \tau, \sigma^2 \sim \mathcal{N}(\hat{\beta}, \Sigma)$ mit $\hat{\beta} = (X^T X + \tau I)^{-1} X^T y$
- $\tau \mid \beta \sim \text{Gamma}(a + p/2, b + \sum \beta_j^2 / 2)$
- $\sigma^2 \mid \beta, y \sim IG(a_0 + n/2, b_0 + \sum \epsilon_i^2)$

Bayesscher Lasso (Laplace-Priori)

- Priori: $\beta_j \sim \text{Laplace}(0, \lambda^{-1})$
- Laplace-Dichte: hohe Dichte nahe Null, dickere Tails als Normalverteilung
- Auxiliary Augmentation (Park & Casella 2008):
  - $\beta_j \mid \tau_j^2 \sim \mathcal{N}(0, \tau_j^2)$
  - $\tau_j^2 \sim \text{Exp}(\lambda^2/2)$
- Ermöglicht Gibbs-Sampling

Spike and Slab Priori

- Idee: **Indikatorvariable** $I_j \in \{0, 1\}$ bestimmt Relevanz
- $\beta_j = I_j \cdot \tilde{\beta}_j$
- **Spike**: kleine Varianz bei $I_j = 0$, **Slab**: große Varianz bei $I_j = 1$
- Hyperpriors: $\tau^2 \sim IG(a, b), \quad w \sim \text{Beta}(a_w, b_w)$
- Ermöglicht bayessche Variablenselektion mit Posterior-Wahrscheinlichkeiten $p(\beta_j = 0 \mid y)$

Zusammenfassung der Verfahren

- **Ridge-Regression**: schrumpft β_j Richtung Null, aber nie exakt Null
- **Bayesscher Lasso**: stärkere Schrumpfung kleiner Werte, d.h. sparsamer
- **Spike-and-Slab**: erlaubt explizit Nullsetzen von β_j
- **Alle**: erlauben Regularisierung und helfen gegen Overfitting
- **Bayessche Modellierung**: liefert Posterior-Verteilung über Relevanz von Kovariablen

Modellwahl: Motivation

- Mehr Parameter verbessern die Anpassung, führen aber zu Overfitting.
- **Informationskriterien** (z.B. AIC, BIC) kombinieren Anpassung mit Modellkomplexität.
- Bei hierarchischen Modellen ist die effektive Anzahl der Parameter relevant.
- Idee: Wieviel "Freiheit" hat das Modell wirklich?

Effektive Anzahl der Parameter

- Beispiel: Zeiteffekte γ_t mit Random-Walk-Priori
- Bei $\tau_c \to \infty$: alle γ_t gleich $\Rightarrow$ nur 1 Parameter
- Bei $\tau_c \to 0$: alle γ_t a priori unabhängig $\Rightarrow$ T Parameter
- Realität: effektive Parameteranzahl zwischen 1 und T

Devianz und DIC

- Devianz: $D(y) := -2(\log p(y|\hat{\theta}) - \log p(y|\hat{\theta}^s))$
- Spiegelhalter et al. (2002):
  - Effektive Anzahl Parameter: $p_D = \overline{D(\theta)} - D(\bar{\theta})$
  - $\text{DIC} = D(\bar{\theta}) + p_D$
- Bei unimodaler Posteriori ist DIC sinnvoll interpretierbar
- Bei MCMC:
  - $D(\theta^{(k)})$ für jede Iteration berechnen
  - Mittelwert ergibt $\overline{D(\theta)}$, Punkt-Schätzer für $D(\bar{\theta})$

Zusammenfassung

- Bayessche Modellwahl kombiniert Modell-Fit mit Komplexitätsstrafe
- Effektive Parameteranzahl ist entscheidend bei hierarchischen Modellen
- DIC als bayessches Informationskriterium etabliert
- Einfach zu berechnen aus MCMC-Samples