Appendix

Appendix

The following paragraphs contain additional information about the **simulation studies** that were covered in chapter **??**.

Section ??: Correlated Predictor Variables

The simulation design was chosen in the following way:

- The design matrix $\mathbf{X} = \begin{pmatrix}\mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3\end{pmatrix}$ is simulated from a three dimensional normal distribution $\mathcal{N}_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with mean vector $\boldsymbol{\mu} = \begin{pmatrix}-5 & 2 & 0\end{pmatrix}^T$ and covariance matrix $\begin{pmatrix}1 & \rho & \rho \\ \rho & 3 & \rho \\ \rho & \rho & 5\end{pmatrix}$. Hence, the dependence among the regressors is fully determined by the parameter ρ.
- The design matrix $\mathbf{Z} = \begin{pmatrix}\mathbf{z}_1 & \mathbf{z}_2\end{pmatrix}$ consists of linear combinations of the regressors $\mathbf{x}_1$ up to $\mathbf{x}_3$, more specifically $\mathbf{z}_1 = 0.8 \cdot \mathbf{x}_1 + 0.2 \cdot \mathbf{x}_2$ and $\mathbf{z}_2 = \mathbf{x}_2 - 0.5 \cdot \mathbf{x}_3$. In both design matrices intercept columns are added for estimation purposes.
- The true coefficient vectors are given by $\boldsymbol{\beta} = \begin{pmatrix}\beta_0 & \beta_1 & \beta_2 & \beta_3\end{pmatrix}^T = \begin{pmatrix}0 & 3 & -1 & 1\end{pmatrix}^T$ and $\boldsymbol{\gamma} = \begin{pmatrix}\gamma_0 & \gamma_1 & \gamma_2\end{pmatrix}^T = \begin{pmatrix}0 & 2 & 0\end{pmatrix}^T$.
- The outcome variable $\mathbf{y}$ is generated according to the correctly specified location-scale model $y_i \overset{iid}{\sim} \mathcal{N}\left(\mathbf{x}_i^T\boldsymbol{\beta}, \exp\left(\mathbf{z}_i^T\boldsymbol{\gamma}\right)^2\right)$ for $i = 1, \dots, n$ with sample size $n = 50$. Before data generation all columns in $\mathbf{X}$ and $\mathbf{Z}$ are standardized, i.e. mean-centered around 0 and scaled to unit variance.
- Three different values were chosen for $\rho \in \{0, -0.5, 0.9\}$ to compare the 'nice' case of uncorrelated predictors with the performance for negative and positive dependence. For each covariance structure the three models `mcmc_ridge()`, `mcmc()` and `lmls()` are fitted to the standardized covariates, where each Posterior Mean estimate from both of the Markov Chain Monte Carlo samplers is based on 10.000 samples.

Section ??: Sample Size

Section ??: Redundant Covariates

We again state the conditions that the simulation study is based on:

- The design matrix $\mathbf{X} = \begin{pmatrix}\mathbf{1}_n & \mathbf{x}_1 & \cdots & \mathbf{x}_{20}\end{pmatrix}$ consists of one intercept column plus 10 *pairs* of successive regressors, starting with the pair $(\mathbf{x}_1, \mathbf{x}_2)$. Each pair $(\mathbf{x}_i, \mathbf{x}_{i+1})$ for $i \in \{1, 3, \dots, 19\}$ is (independently from all remaining pairs) drawn from a bivariate normal distribution with mean vector $\boldsymbol{\mu} = \begin{pmatrix}0 & 0\end{pmatrix}^T$ and correlation matrix $\begin{pmatrix}1 & 0.9 \\ 0.9 & 1\end{pmatrix}$.

- The design matrix $\mathbf{Z} = \begin{pmatrix}\mathbf{1}_n & \mathbf{x}_1 & \mathbf{x}_3\end{pmatrix}$ is of minor interest in this case and consists of an intercept column plus two uncorrelated columns chosen from $\mathbf{X}$.
- The true coefficients of $\boldsymbol{\beta}$ are determined by the pattern $\beta_i = 0$, if i is even and $\beta_i = 1$, if i is odd. Thus, all covariates with even subscript are redundant, whereas those with odd subscript contribute to $\mathbf{y}$. The true $\boldsymbol{\gamma}$, again of minor interest here, is given by $\boldsymbol{\gamma} = \begin{pmatrix}0 & 1 & 1\end{pmatrix}^T$.
- The outcome variable $\mathbf{y}$ is generated according to the correctly specified location-scale model $y_i \overset{iid}{\sim} \mathcal{N}\left(\mathbf{x}_i^T\boldsymbol{\beta}, \exp\left(\mathbf{z}_i^T\boldsymbol{\gamma}\right)^2\right)$ for $i = 1, \ldots, n$, where the covariates are used on their original scale.
- The sample size $n = 50$ is deliberately chosen small compared to the number of regressors. Before fitting each of the three models, all columns of $\mathbf{X}$ except the intercept column is standardized to zero mean and unit variance. Both of the Bayesian models generate 10.000 values for each coefficient.

Section ??: Challenging the Model Assumptions

The data for this simulation study is generated by the following conventions:

- The design matrix $\mathbf{X} = \begin{pmatrix}\mathbf{1}_n & \mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3 & \mathbf{x}_4\end{pmatrix}$ contains four independently sampled regressor variables plus one intercept column:
  - $\mathbf{x}_1 \overset{iid}{\sim} \mathcal{N}(5, 16)$,
  - $\mathbf{x}_2 \overset{iid}{\sim} \mathrm{Exp}(5)$,
  - $\mathbf{x}_3 \overset{iid}{\sim} \mathcal{U}([-2,\ 12])$,
  - $\mathbf{x}_4 \overset{iid}{\sim} \mathrm{Ber}(0.3)$.
- The design matrix $\mathbf{Z} = \begin{pmatrix}\mathbf{1}_n & \mathbf{x}_1 & \mathbf{x}_2 & \mathbf{z}_3\end{pmatrix}$ contains the additional regressor variable $\mathbf{z}_3 \overset{iid}{\sim} t_{10}$, which is independently sampled from all other columns.
- The true coefficient vectors are given by $\boldsymbol{\beta} = \begin{pmatrix}\beta_0 & \beta_1 & \beta_2 & \beta_3 & \beta_4\end{pmatrix}^T = \begin{pmatrix}0 & -10 & -5 & -3 & -1\end{pmatrix}^T$ and $\boldsymbol{\gamma} = \begin{pmatrix}\gamma_0 & \gamma_1 & \gamma_2 & \gamma_3\end{pmatrix}^T = \begin{pmatrix}0 & 3 & 5 & 10\end{pmatrix}^T$.
- Three different specifications for the outcome distribution were chosen:
  - $y_i \sim \mathcal{N}\left(\mu, \sigma^2\right)$,
  - $y_i \sim \mu + \left(\sigma \cdot \sqrt{\frac{3}{5}}\right) T$, where $T \sim t_5$,
  - $y_i \sim \mu + \sigma \cdot U$, where $U \sim \mathcal{U}([0,\ 1])$. In all cases, the outcome vectors are generated with the covariates on their original (unstandardized) scale.
- In order to isolate the impact of the different shapes of the three probability distributions from the effect of varying moment structures, the mean $\mu = \mathbf{x}_i^T\boldsymbol{\beta}$ and the variance $\sigma^2 = \exp\left(\mathbf{z}_i^T\boldsymbol{\gamma}\right)^2$ are held constant across the models.
- All three models `mcmc_ridge()`, `mcmc()` and `lmls()` are fitted with standardized covariates. The sample size is set to $n = 50$ and the result of both Bayesian samplers are based on 10.000 simulations.

Section ??: Hyperparameters - Impact on Estimation Accuracy

The simulation study of the impact of the Hyperparameters on the coefficients is constructed as follows:

- The design matrix $\mathbf{X} = \begin{pmatrix}\mathbf{x}_1 & \mathbf{x}_2\end{pmatrix}$ is simulated from a two dimensional normal distribution $\mathcal{N}_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with mean vector $\boldsymbol{\mu} = \begin{pmatrix}1 & 2\end{pmatrix}^T$ and identity covariance matrix $\boldsymbol{\Sigma} = \mathbf{I}_2$. The same holds true for the design matrix $\mathbf{Z} = \begin{pmatrix}\mathbf{z}_1 & \mathbf{z}_2\end{pmatrix}$ with mean vector $\boldsymbol{\mu} = \begin{pmatrix}5 & 3\end{pmatrix}^T$ and identity covariance matrix. However, after simulating $\mathbf{X}$ and $\mathbf{Z}$, both are standardized to a zero mean and a unit variance.
- In both design matrices intercept columns are added for estimation purposes. The true coefficient vectors are given by $\boldsymbol{\beta} = \begin{pmatrix}\beta_0 & \beta_1 & \beta_2\end{pmatrix}^T = \begin{pmatrix}0 & -1 & 4\end{pmatrix}^T$ and $\boldsymbol{\gamma} = \begin{pmatrix}\gamma_0 & \gamma_1 & \gamma_2\end{pmatrix}^T = \begin{pmatrix}0 & -2 & 1\end{pmatrix}^T$.
- The outcome variable $\mathbf{y}$ is generated according to the correctly specified location-scale model $y_i \overset{iid}{\sim} \mathcal{N}\left(\mathbf{x}_i^T\boldsymbol{\beta}, \exp\left(\mathbf{z}_i^T\boldsymbol{\gamma}\right)^2\right)$ for $i = 1, \ldots, n$ with sample size $n = 50$ as well as standardized $\mathbf{X}$ and $\mathbf{Z}$.
- For sampling the location parameter, the full conditional multivariate normal distribution of $\boldsymbol{\beta}$ is chosen, i.e. `mcmc_ridge(..., mh_location = FALSE)` is used. Therefore, the location estimate is directly affected by the hyperparameters.
- For simulating the influence of the hyperparameters, nine different values are chosen: $a_\tau, b_\tau, a_\xi, b_\xi \in \{-1, 0, 0.5, 1, 2, 10, 50, 100, 200\}$. Since for statistical properties like the mean of an Inverse Gamma distribution $\frac{b}{a-1}$ the condition $a > 1$ is required, particular attention is given to larger values. However, it is an aim to inspect the performance of the sampler for smaller hyperparameter values than 1 as well.

Section ??: Hyperparameters - Impact on Ridge Penalty

This simulation study is conducted in the following way:

- The column vectors of the design matrices $\mathbf{X} = \begin{pmatrix}\mathbf{x}_1 & \mathbf{x}_2\end{pmatrix}$ and $\mathbf{Z} = \begin{pmatrix}\mathbf{z}_1 & \mathbf{z}_2\end{pmatrix}$ are independently drawn from normal distributions with variance $\sigma^2 = 1$ and varying means $\mu \in (1, 2, 5, 3)$. Then both design matrices are standardized and intercept columns are added.
- The true coefficient vectors are given by $\boldsymbol{\beta} = \begin{pmatrix}\beta_0 & \beta_1 & \beta_2\end{pmatrix}^T = \begin{pmatrix}0 & 8 & 2\end{pmatrix}^T$ and $\boldsymbol{\gamma} = \begin{pmatrix}\gamma_0 & \gamma_1 & \gamma_2\end{pmatrix}^T = \begin{pmatrix}0 & 3 & 3\end{pmatrix}^T$.
- The outcome variable $\mathbf{y}$ is generated based on the standardized covariates according to the correctly specified location-scale model $y_i \overset{iid}{\sim} \mathcal{N}\left(\mathbf{x}_i^T\boldsymbol{\beta}, \exp\left(\mathbf{z}_i^T\boldsymbol{\gamma}\right)^2\right)$ for $i = 1, \ldots, n$ with sample size $n = 50$.
- The pairs of hyperparameters (a_τ, b_τ) and (a_ξ, b_ξ) take values on a grid, which is constructed by all combinations of the sequence $\left(\frac{1}{16}, \frac{1}{8}, \frac{1}{4}, \frac{1}{2}, 1, 2, 4, 8, 16, 32, 64, 128, 256\right)$. While one pair is varied, the other pair is fixed on the values $(1, 1)$.
- The `mcmc_ridge()` function does not use the `lmls()` function as basis in this case, but rather use the standardized data directly as input. The number of simulations `num_sim` is chosen as 1000 and the starting values `beta_start` and `gamma_start` are set to $\begin{pmatrix}1 & 1\end{pmatrix}$, respectively.