

Bayesianische Regression

Lineare und logistische Modelle

Lona Koers

LMU

25. Juli 2025

Frequentistisch: ein Modell + Konfidenzintervalle

Bayesianisch: ganze Verteilung von möglichen Modellen

→ Wie können wir Vorwissen (korrekt) modellieren?

→ Wie können wir Unsicherheit im Modell darstellen?

- 1 Bayesianisches **lineares** Modell (LM)
- 2 Bayesianisches **generalisiertes** lineares Modell (GLMs)
- 3 Inferenz im bayesianischen (G)LM
- 4 Referenzen

Annahmen:

- ① i.i.d. Daten $\mathbf{D} = (\mathbf{y}, \mathbf{X})$
- ② Kondition auf \mathbf{X} (implizit)

Frequentistisches lineares Modell: $\mathbf{y} \sim \mathcal{N}(\mathbf{X}\boldsymbol{\theta}, \sigma^2 \mathbf{I})$

- ③ Gewichtsparameter $\boldsymbol{\theta}$ als Zufallsvariable interpretieren

Bayesianisches lineares Modell:

$$\mathbf{y} \mid \boldsymbol{\theta}, \sigma^2 \sim \mathcal{N}(\mathbf{X}\boldsymbol{\theta}, \sigma^2 \mathbf{I})$$

Prior-Annahme für θ (und evtl. σ^2) notwendig \rightarrow sehr vielseitige Modell-Anpassung möglich

Normal-Invers-Gamma Prior:

$$\theta \mid \sigma^2 \sim \mathcal{N}(\check{\mu}, \sigma^2 \check{\Sigma})$$

$$\sigma^2 \sim \text{IG}(\check{a}, \check{b})$$

$$\theta, \sigma^2 \sim \text{NIG}(\check{\mu}, \sigma^2 \check{\Sigma}, \check{a}, \check{b})$$

mit Prior Parametern: $\check{\mu}$, $\check{\Sigma}$, \check{a} und \check{b}

Vorteil: NIG-Prior ist mit Normalverteilungs-Likelihood konjugiert \rightarrow exakte Inferenz möglich (mehr dazu später)

Uninformative Prior

z.B. mit NIG-Prior mit Prior Parametern

$$\check{\boldsymbol{\mu}} = \mathbf{0}, \quad \check{\Sigma}^{-1} = \mathbf{0} \text{ i.e. } \check{\Sigma} \rightarrow \infty, \quad \check{a} = -\frac{p}{2}, \quad \check{b} = 0$$

\implies flache (und damit uninformative) Prior und maximaler Einfluss der Daten auf die Posterior:

$$\boldsymbol{\theta} \mid \sigma^2 \stackrel{a}{\sim} \mathcal{N}(\check{\boldsymbol{\mu}}, \sigma^2 \infty) \implies p(\boldsymbol{\theta} \mid \sigma^2) \propto 1$$

TODO: Bild

Erinnerung: Regularisierung im *frequentistischen* LM durch Minimierung von

$$\text{PLS}(\boldsymbol{\theta}) = (\mathbf{y} - \mathbf{X}\boldsymbol{\theta})^\top (\mathbf{y} - \mathbf{X}\boldsymbol{\theta}) + \lambda \text{pen}(\boldsymbol{\theta})$$

mit Regularisierungs-Parameter $\lambda > 0$.

Bayesianische Regularisierung durch Wahl der Prior-Verteilung für $\boldsymbol{\theta}$

Ridge Regularisierung

- Frequentistisch [14, 15]: $\text{pen}(\boldsymbol{\theta}) = \|\boldsymbol{\theta}\|_2^2$
- Bayesianisch [17]: $\boldsymbol{\theta} \sim \mathcal{N}(\mathbf{0}, \tau^2 \mathbf{I})$ mit $\tau^2 \propto \frac{1}{\lambda}$

Lasso Regularisierung

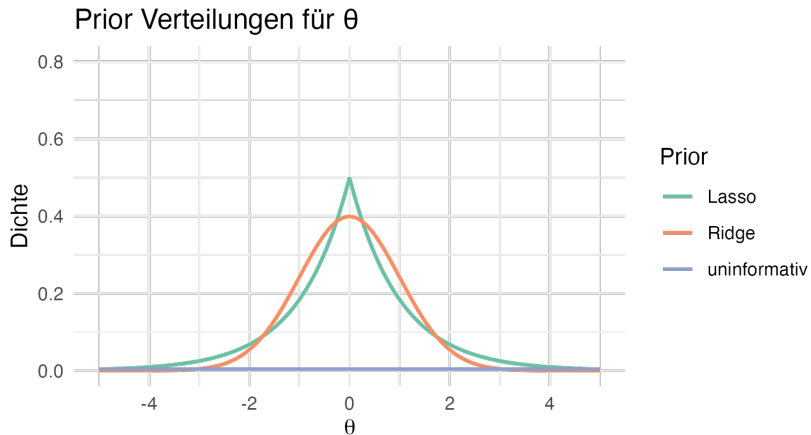
- Frequentistisch [25]: $\text{pen}(\boldsymbol{\theta}) = \|\boldsymbol{\theta}\|_1$
- Bayesianisch [21]:

$$\boldsymbol{\theta} \mid \tau^2 \sim \mathcal{N}(\mathbf{0}, \tau^2 \mathbf{I})$$

$$\tau_j^2 \stackrel{\text{i.i.d.}}{\sim} \text{Exp}(0.5\lambda^2), \quad j = 1, \dots, p$$

Problem: keine Variablenselektion (im Gegensatz zu frequentistischem Lasso)

→ Alternative Priors für Variablenselektion: Spike and Slab [18], Horseshoe [3], u.v.m.



- 1 Bayesianisches **lineares** Modell (LM)
- 2 Bayesianisches **generalisiertes** lineares Modell (GLMs)
- 3 Inferenz im bayesianischen (G)LM
- 4 Referenzen

$$\text{LM: } \mathbf{y} \mid \boldsymbol{\theta}, \sigma^2 \sim \mathcal{N}(\mathbf{X}\boldsymbol{\theta}, \sigma^2 \mathbf{I}) \quad \rightarrow \quad \text{GLM: } \mathbf{y} \mid \boldsymbol{\theta} \sim F(g^{-1}(\mathbf{X}\boldsymbol{\theta}))$$

- Verteilungsannahme von \mathbf{y} wird (äquivalent zum frequentistischen GLM) auf alle Verteilungen F der Exponentialfamilie ausgeweitet
- Skala des linearen Prädiktors $\mathbf{X}\boldsymbol{\theta}$ wird mit der Link-Funktion g^{-1} angepasst

→ Bayesianisches logistisches Modell

$$\mathbf{y}_i \mid \boldsymbol{\theta} \sim \text{Bin}(1, g^{-1}(\mathbf{x}_i \boldsymbol{\theta})), \quad i = 1, \dots, n$$
$$g^{-1}(\mathbf{x}_i \boldsymbol{\theta}) = \sigma(\mathbf{x}_i \boldsymbol{\theta})$$

Für Beobachtungen $\mathbf{x}_i = (1, x_{i1}, \dots, x_{ip})^\top$ und Sigmoid-Link $\sigma(y) = \frac{\exp(y)}{1+\exp(y)}$

Prior Wahl

- I.A. äquivalent zum LM möglich, z.B. Normalverteilungs-Prior
- Verteilungen mit schweren Rändern (z.B. t-Verteilung, Cauchy Verteilung) adressieren Separation und fördern Shrinkage [10, 12]
- Für Regularisierung können dieselben Priors wie im LM verwendet werden [20, 7, 6]

- 1 Bayesianisches **lineares** Modell (LM)
- 2 Bayesianisches **generalisiertes** lineares Modell (GLMs)
- 3 Inferenz im bayesianischen (G)LM
- 4 Referenzen

Erinnerung

- Inferenz im frequentistischen LM: z.B. *kleinste Quadrate Schätzung* mit

$$\hat{\boldsymbol{\theta}}_{KQ} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y} \quad \text{mit} \quad \hat{\boldsymbol{\theta}}_{KQ} \overset{a}{\sim} \mathcal{N}(\boldsymbol{\theta}, \sigma^2 (\mathbf{X}^\top \mathbf{X})^{-1})$$

- Inferenz im bayesianischen LM: *Bayes-Regel* zur Ermittlung der Parameter Posterior

$$p(\boldsymbol{\theta} \mid \mathbf{y}) = \frac{p(\mathbf{y} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta})}{\int p(\mathbf{y} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta}) d\boldsymbol{\theta}} = \frac{\mathcal{L}(\boldsymbol{\theta}) p(\boldsymbol{\theta})}{\int \mathcal{L}(\boldsymbol{\theta}) p(\boldsymbol{\theta}) d\boldsymbol{\theta}}$$

Inferenz mit konjugierten Priors

Bayesianisches LM: aus der Konjugiertheit von $\mathbf{y} \mid \boldsymbol{\theta}, \sigma^2 \sim \mathcal{N}$ und $\boldsymbol{\theta}, \sigma^2 \sim \text{NIG}$ ergibt sich

$$\boldsymbol{\theta}, \sigma^2 \mid \mathbf{y} \sim \text{NIG}(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}, \hat{a}, \hat{b})$$

mit

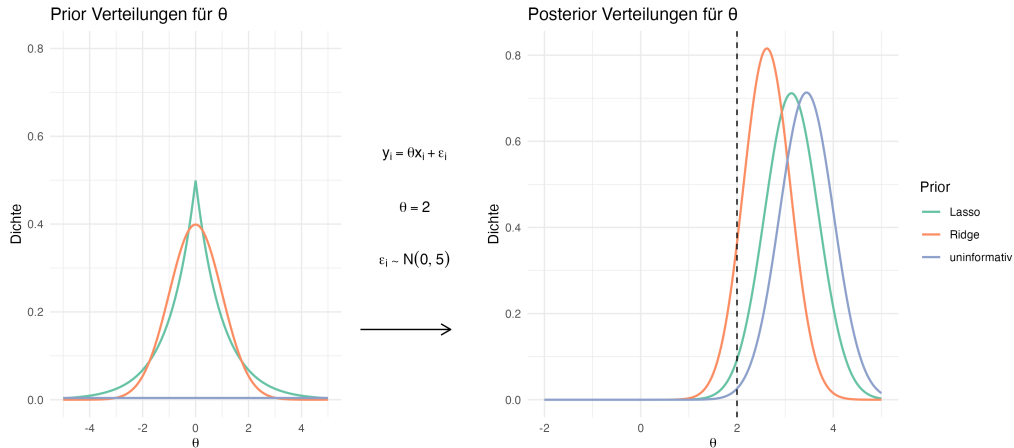
$$\hat{\boldsymbol{\mu}} = \hat{\boldsymbol{\Sigma}}(\check{\boldsymbol{\Sigma}}^{-1}\check{\boldsymbol{\mu}} + \mathbf{X}^\top \mathbf{y}), \quad \hat{\boldsymbol{\Sigma}} = (\mathbf{X}^\top \mathbf{X} + \check{\boldsymbol{\Sigma}}^{-1})^{-1}$$

- **uninformative NIG-Prior:** Posterior Mean und KQ-Schätzer sind äquivalent:

$$\check{\boldsymbol{\mu}} = \mathbf{0}, \quad \check{\boldsymbol{\Sigma}}^{-1} = \mathbf{0} \implies \hat{\boldsymbol{\mu}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}, \quad \hat{\boldsymbol{\Sigma}} = (\mathbf{X}^\top \mathbf{X})^{-1}$$

- **Ridge:** Spezialfall der NIG-Prior \rightarrow konjugierte Berechnung der Parameter Posterior
- **Lasso:** Prior hat keine geschlossene Form, aus der Posterior kann aber mit Gibbs-Sampling simuliert werden [21]

Posterior Update für Prior-Verteilungen



Problem: Inferenz mit konjugierten Priors ist nur sehr selten möglich [22]

→ **Approximative bayesianische Inferenz**, um aus der Posterior zu ziehen, mit z.B.:

- Sampling Methoden (Markov chain Monte Carlo Methoden)
 - ▶ **Metropolis-Hastings Algorithmus** [13]
 - ▶ Gibbs Sampling (bedingte Konjugiertheit) [5]
 - ▶ Hamiltonian Monte Carlo (v.a. hochdimensionale Posterior) [19]
- Deterministische Approximation, z.B. **Laplace Approximation** [26]
- U.v.m.

- **Idee:** Aus der Posterior $p(\boldsymbol{\theta} \mid \mathbf{y})$ ziehen, ohne Annahmen über ihre exakte Form machen zu müssen
- **Problem:** Ergebnisse sind am Besten, wenn die Posterior bis auf eine Konstante (meist die Normalisierungskonstante $\int p(\mathbf{y} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta}) d\boldsymbol{\theta}$) bekannt sind
- **Inputs:**
 - ▶ Anzahl der Ziehungen $K \rightarrow$ frei wählbar
 - ▶ Likelihood $p(\boldsymbol{\theta} \mid \mathbf{y})$ und Prior $p(\boldsymbol{\theta}) \rightarrow$ bekannt
 - ▶ Proposal Verteilung $q \rightarrow$ muss sorgfältig gewählt werden

Effizienz des Algorithmus ist stark abhängig von der Proposal Verteilung.

→ **Optionen:**

- **Normalverteilung**

$$q(\boldsymbol{\theta}^{(*)} \mid \boldsymbol{\theta}^{(k)}) \sim \mathcal{N}(\boldsymbol{\theta}^{(k)} \mid -H^{-1}(\boldsymbol{\theta}^{(k)}))$$

mit Mittelwert beim letzten Sample. Die Hesse-Matrix H von $p(\boldsymbol{\theta} \mid \mathbf{y})$ wird meist mit IWLS geschätzt [8, 16, 24]

- Scott [24] schlägt **Verteilungen mit schweren Rändern** vor → mehr Mixing, kürzerer Burn-in und schnellere Konvergenz

- **Idee:** Approximation der Posterior mit einer Normalverteilung

$$p(\boldsymbol{\theta} \mid \mathbf{y}) \approx \mathcal{N}(\hat{\boldsymbol{\theta}}_{MAP}, H^{-1}(\hat{\boldsymbol{\theta}}_{MAP}))$$

mit Maximum-a-Posteriori Schätzer $\hat{\boldsymbol{\theta}}_{MAP}$ und Hesse-Matrix H von $p(\boldsymbol{\theta} \mid \mathbf{y})$

- Erweiterung für hierarchische Modelle: Integrated Nested Laplace Approximation (INLA) [23]

Setup: 1.000 synthetische Datensätze mit $n = 100$ und

$$\mathbf{X} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \quad \boldsymbol{\theta} = (-0.5, 2, 1)$$

linear: $\mathbf{y} \mid \boldsymbol{\theta} \sim \mathcal{N}(\mathbf{X}\boldsymbol{\theta}, \mathbf{I})$

logistic: $\mathbf{y} \mid \boldsymbol{\theta} \sim \text{Ber}(\sigma(\mathbf{X}\boldsymbol{\theta}))$

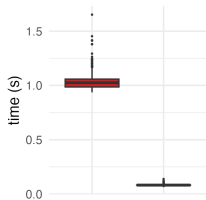
Experiment: für jeden Datensatz wurde angepasst:

- Ein lineares und ein logistisches Regressionsmodell mit $\boldsymbol{\theta} \sim \mathcal{N}(0, 10 \cdot \mathbf{I})$ und $\sigma^2 = 10$
- Mit Laplace Approximation und Metropolis-Hastings ($K = 5.000$, Burn-in = 500, Thinning Intervall = 10)

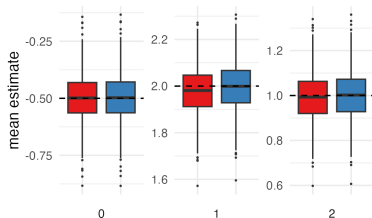
Beispiel: LA vs. Metropolis-Hastings

Lineares Modell

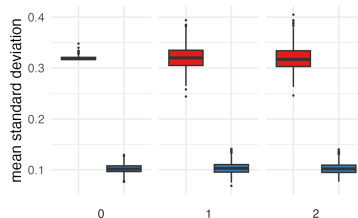
Elapsed time



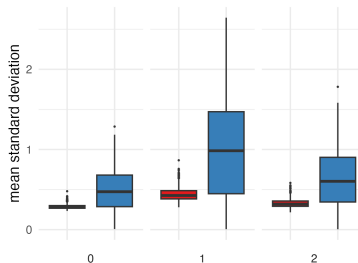
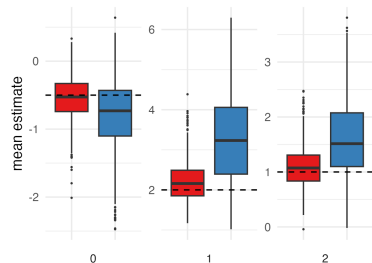
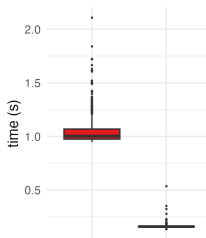
Parameter estimates



Posterior parameter standard deviation



Logistisches Modell



Method
LA
MCMC
-- True parameter

Aus der Bayes Regel:

$$p(\mathbf{y}) = \int p(\mathbf{y}, \boldsymbol{\theta}) d\boldsymbol{\theta} = \int p(\mathbf{y} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta}) d\boldsymbol{\theta}$$

Posterior Predictive Distribution

→ Vorhersagen für neue Daten $\tilde{\mathbf{X}}$ im bayesianischen GLM: [2, 1]

$$p(\tilde{\mathbf{y}} \mid \mathbf{y}) = \int p(\tilde{\mathbf{y}}, \boldsymbol{\theta} \mid \mathbf{y}) d\boldsymbol{\theta} = \int p(\tilde{\mathbf{y}} \mid \boldsymbol{\theta}, \mathbf{y}) p(\boldsymbol{\theta}) d\boldsymbol{\theta} \stackrel{\tilde{\mathbf{y}} \perp \mathbf{y} \mid \boldsymbol{\theta}}{=} \int p(\tilde{\mathbf{y}} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta}) d\boldsymbol{\theta}$$

Für das bayesianische **logistische Modell** berechnet man

$$p(\tilde{\mathbf{y}} = 1 \mid \mathbf{y})$$

mit $y_i \in \{0 \text{ (negativ)}, 1 \text{ (positiv)}\}$

Analytische Berechnung: z.B. für die NIG-Prior ergibt sich

$$\tilde{\mathbf{y}} \mid \boldsymbol{\theta}, \sigma^2, \mathbf{y} \sim \mathcal{T}(2\hat{a}, \tilde{\mathbf{X}}\boldsymbol{\theta}, \frac{\hat{b}}{\hat{a}}(\mathbf{I} + \tilde{\mathbf{X}}\hat{\Sigma}\tilde{\mathbf{X}}^\top))$$

Approximation

- für Metropolis-Hastings: $p(\tilde{\mathbf{y}} = 1 \mid \mathbf{y}) \approx \frac{1}{K} \sum_{k=1}^K \sigma(\tilde{\mathbf{X}}\boldsymbol{\theta}^{(k)})$
- für Laplace Approximation:
 - ▶ Samples $\boldsymbol{\theta}^{(s)} \sim \mathcal{N}(\hat{\boldsymbol{\theta}}_{MAP}, H^{-1}(\hat{\boldsymbol{\theta}}_{MAP}))$ mit $s = 1, \dots, S$ aus der LA-approximierten Parameter Posterior ziehen und wie bei Metropolis-Hastings vorgehen oder
 - ▶ LA-approximierte PPD analytisch berechnen:

$$p(\tilde{\mathbf{y}} = 1 \mid \mathbf{y}) = \int \sigma(\tilde{\mathbf{X}}\boldsymbol{\theta}) \mathcal{N}_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}_{MAP}, H^{-1}(\hat{\boldsymbol{\theta}}_{MAP})) d\boldsymbol{\theta}$$

Fazit

- Anwendungsmöglichkeiten: kleine Stichproben, explizite Nutzung von Vorwissen, einfache Regularisierung + Erhalt der Interpretierbarkeit

Literatur Empfehlungen

- Bayesianische Regression (v.a. für praktische Anwendung): Gelman et al. [11]
- Prior Verteilungen (v.a. Shrinkage): Erp, Oberski, and Mulder [6] und Celeux et al. [4]
- Software: z.B. brms in R, PyMC in python

- 1 Bayesianisches **lineares** Modell (LM)
- 2 Bayesianisches **generalisiertes** lineares Modell (GLMs)
- 3 Inferenz im bayesianischen (G)LM
- 4 Referenzen

Vorteile von bayesianischer Regularisierung sind u.a.:

- Probabilistisches Modell trotz Regularisierung
- Regularisierungs-Parameter muss nicht als Hyperparameter optimiert werden (z.B. durch Prior auf τ^2)
- Mehr Anpassungsmöglichkeiten durch Prior-Spezifikation

Frequentistische vs. bayesianische Lasso Regularisierung

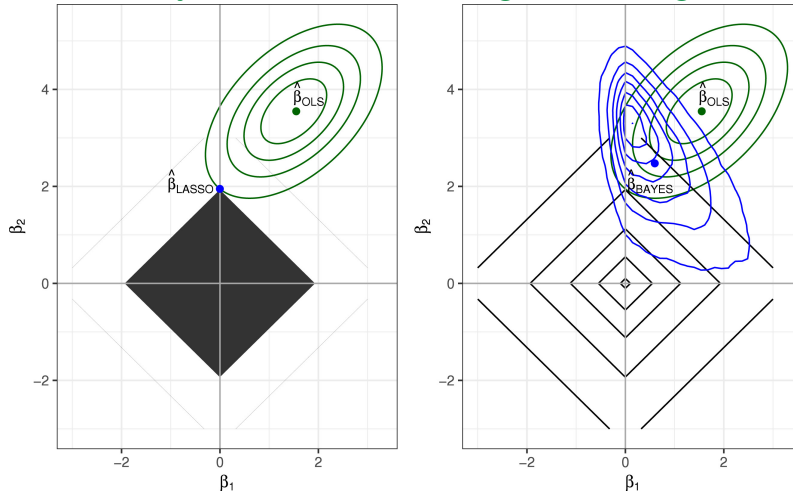


Figure 1: Contour plot representing the sum of squared residuals, classical lasso constraint region (left), bivariate lasso prior and posterior distribution (right), and the classical and Bayesian penalized point estimates. Aus Erp, Oberski, and Mulder [6]

Beispiel: Vergleich von Regularisierungs-Priors in linearer und logistischer Regression

Setup: zwei synthetische Datensätze mit $n = 150$

- **Szenario A:** Sparses Setting ohne Kollinearität oder Überparametrisierung
 - ▶ $n = 150$, $n_{train} = 100$, $n_{test} = 50$
 - ▶ $\theta = (2, 1.5, 0, 0, 0)$
 - ▶ $\mathbf{X} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
 - ▶ linear: $\mathbf{y} \mid \theta \sim \mathcal{N}(\mathbf{X}\theta, \mathbf{I})$, logistisch: $\mathbf{y} \mid \theta \sim \text{Ber}(\sigma(\mathbf{X}\theta))$
- **Szenario B:** uninformatives Setting mit $n \approx p$ und Kollinearität zwischen informativen und uninformativen Koeffizienten
 - ▶ $n = 150$, $n_{train} = 30$, $n_{test} = 120$
 - ▶ $\theta = (2, 1.5, 0, \overset{26 \text{ times}}{\dots}, 0)$
 - ▶ $\mathbf{X} \sim \mathcal{N}(\mathbf{0}, \Sigma)$, $\Sigma = \begin{pmatrix} 1 & & & \\ & S_3 & & 0 \\ & & I_{26} & \\ & 0 & & \end{pmatrix}$, $S_3 = \begin{pmatrix} 1 & 0.8 & 0.8 \\ 0.8 & 1 & 0.8 \\ 0.8 & 0.8 & 1 \end{pmatrix}$
 - ▶ linear: $\mathbf{y} \mid \theta \sim \mathcal{N}(\mathbf{X}\theta, \mathbf{I})$, logistisch: $\mathbf{y} \mid \theta \sim \text{Ber}(\sigma(\mathbf{X}\theta))$

Experiment: Für jeden Datensatz wurde angepasst:

- bayesianisches lineares und logistisches Modell (MCMC mit $K = 20000$, Burn-in = 1000, Thinning = 10) mit
- uninformative, Ridge und Lasso Priori

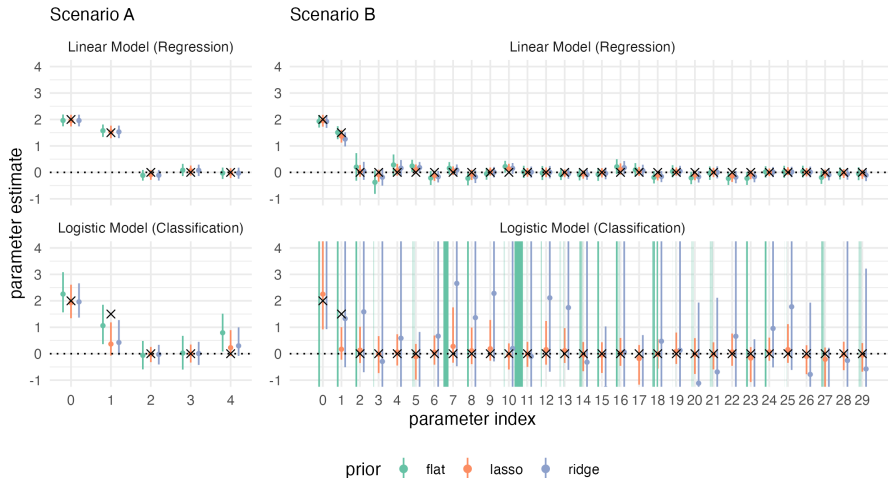
Evaluation

- Anzahl korrekt als einflussreich identifizierter Kovariablen (Hits)
- Anzahl fälschlich als einflussreich identifizierter Kovariablen (FP)
- Mean log posterior predictive density (MLPPD) [9]

Ergebnis

Model	Prior	Scenario A			Scenario B		
		Hits (of 2)	FP (of 3)	MLPPD	Hits (of 2)	FP (of 28)	MLPPD
Linear	flat	2	0	-1.425	2	1	-1.605
Linear	LASSO	2	0	-1.424	2	0	-1.464
Linear	ridge	2	0	-1.427	2	0	-1.575
Logit	flat	2	1	-0.390	0	21	$-\infty$
Logit	LASSO	1	0	-0.463	1	0	-0.485
Logit	ridge	1	0	-0.455	1	0	-0.493

Ergebnis: Regularisierung in bayesianischen Modellen (II)



- Im frequentistischen Modell:

$$\hat{\theta}_{KQ} \stackrel{a}{\sim} \mathcal{N}(\theta, \sigma^2(\mathbf{X}^\top \mathbf{X})^{-1})$$

- Im bayesianischen Modell (NIG-Prior):

$$\theta \sim \mathcal{T}(2\hat{a}, \hat{\mu}, \hat{b}/\hat{a}\hat{\Sigma}) \approx \mathcal{N}(\hat{\mu}, \hat{b}/\hat{a}\hat{\Sigma})$$

für großes p

- ➊ $\theta^{(1)}$ initialisieren
- ➋ Für $k = 1, \dots, K$
 - ➊ $\theta^{(*)}$ aus der *Proposal Verteilung* $q(\theta^{(*)} | \theta^{(k)})$ ziehen
 - ➋ Akzeptanzwahrscheinlichkeit berechnen

$$\alpha = \min\left(1, \frac{p(\theta^{(*)} | \mathbf{y}) p(\theta^{(*)}) q(\theta^{(k)} | \theta^{(*)})}{p(\theta^{(k)} | \mathbf{y}) p(\theta^{(k)}) q(\theta^{(*)} | \theta^{(k)})}\right)$$

- ➌ Vorschlag $\theta^{(*)}$ akzeptieren oder verwerfen (für $u \sim \text{Uni}[0, 1]$)

$$\begin{cases} u \leq \alpha & \theta^{(k+1)} = \theta^{(*)} \\ u > \alpha & \theta^{(k+1)} = \theta^{(k)} \end{cases}$$

Und z.B.

$$q(\theta^{(*)} | \theta^{(k)}) \sim \mathcal{N}(\theta^{(k)} | -H^{-1}(\theta^{(k)}))$$

mit $H(\theta) = \nabla_{\theta}^2 \log(p(\theta^{(k)} | \mathbf{y}) p(\theta^{(k)}))$

Exemplarische Berechnung für ein bayesianisches logistisches Modell mit der Prior $\boldsymbol{\theta} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$:

Es gilt dass mit Laplace Approximation $p(\boldsymbol{\theta} \mid \mathbf{y}) \approx \mathcal{N}(\hat{\boldsymbol{\theta}}_{MAP}, H^{-1}(\hat{\boldsymbol{\theta}}_{MAP}))$ mit

$$\begin{aligned}\hat{\boldsymbol{\theta}}_{MAP} &= \arg \max_{\boldsymbol{\theta}} p(\boldsymbol{\theta} \mid \mathbf{y}) \stackrel{\text{Bayes' rule}}{=} \arg \max_{\boldsymbol{\theta}} p(\mathbf{y} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta}) d\boldsymbol{\theta} \\ &= \arg \max_{\boldsymbol{\theta}} \sum_{i=1}^n \log(\sigma(y_i \mathbf{x}_i \boldsymbol{\theta})) - \frac{1}{2\sigma^2} \boldsymbol{\theta}^\top \boldsymbol{\theta} \\ H(\boldsymbol{\theta}) &= -\nabla_{\boldsymbol{\theta}}^2 \log p(\boldsymbol{\theta} \mid \mathbf{y}) = \frac{1}{\sigma^2} \mathbf{I} + \sum_{i=1}^n \sigma(y_i \mathbf{x}_i \boldsymbol{\theta}) (1 - \sigma(y_i \mathbf{x}_i \boldsymbol{\theta})) \mathbf{x}_i \mathbf{x}_i^\top.\end{aligned}$$

- [1] Maria Maddalena Barbieri. “Posterior Predictive Distribution”. In: *Wiley StatsRef: Statistics Reference Online*. John Wiley & Sons, Ltd, 2015, pp. 1–6. ISBN: 978-1-118-44511-2. DOI: 10.1002/9781118445112.stat07839.
- [2] George E. P. Box. “Sampling and Bayes’ Inference in Scientific Modelling and Robustness”. In: *Journal of the Royal Statistical Society. Series A (General)* 143.4 (1980), pp. 383–430. ISSN: 0035-9238. DOI: 10.2307/2982063.
- [3] Carlos M. Carvalho, Nicholas G. Polson, and James G. Scott. “The horseshoe estimator for sparse signals”. English. In: *BIOMETRIKA* 97.2 (June 2010), pp. 465–480. ISSN: 0006-3444. DOI: 10.1093/biomet/asq017.
- [4] Gilles Celeux et al. “Regularization in Regression: Comparing Bayesian and Frequentist Methods in a Poorly Informative Situation”. In: *Bayesian Analysis* 7.2 (June 2012), pp. 477–502. ISSN: 1936-0975, 1931-6690. DOI: 10.1214/12-BA716.
- [5] P. Dellaportas and A. F. M. Smith. “Bayesian Inference for Generalized Linear and Proportional Hazards Models via Gibbs Sampling”. In: *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 42.3 (1993), pp. 443–459. ISSN: 0035-9254. DOI: 10.2307/2986324.

- [6] Sara van Erp, Daniel L. Oberski, and Joris Mulder. “Shrinkage priors for Bayesian penalized regression”. In: *Journal of Mathematical Psychology* 89 (Apr. 2019), pp. 31–50. ISSN: 0022-2496. DOI: 10.1016/j.jmp.2018.12.004.
- [7] Ludwig Fahrmeir, Thomas Kneib, and Susanne Konrath. “Bayesian regularisation in structured additive regression: a unifying perspective on shrinkage, smoothing and predictor selection”. In: *Statistics and Computing* 20.2 (2010), pp. 203–219. ISSN: 1573-1375. DOI: 10.1007/s11222-009-9158-3.
- [8] Dani Gamerman. “Markov chain Monte Carlo for dynamic generalised linear models”. In: *Biometrika* 85.1 (Mar. 1998), pp. 215–227. ISSN: 0006-3444. DOI: 10.1093/biomet/85.1.215.
- [9] Andrew Gelman, Jessica Hwang, and Aki Vehtari. *Understanding predictive information criteria for Bayesian models*. July 2013. DOI: 10.48550/arXiv.1307.5928.
- [10] Andrew Gelman et al. “A weakly informative default prior distribution for logistic and other regression models”. In: *The Annals of Applied Statistics* 2.4 (Dec. 2008), pp. 1360–1383. ISSN: 1932-6157, 1941-7330. DOI: 10.1214/08-AOAS191.
- [11] Andrew Gelman et al. *Bayesian Data Analysis*. 3rd ed. New York: Chapman and Hall/CRC, 2013. ISBN: 978-0-429-11307-9. DOI: 10.1201/b16018.

- [12] Joyee Ghosh, Yingbo Li, and Robin Mitra. *On the Use of Cauchy Prior Distributions for Bayesian Logistic Regression*. Feb. 2017. DOI: 10.48550/arXiv.1507.07170.
- [13] W. K. Hastings. "Monte Carlo Sampling Methods Using Markov Chains and Their Applications". In: *Biometrika* 57.1 (1970), pp. 97–109. ISSN: 0006-3444. DOI: 10.2307/2334940.
- [14] Arthur E. Hoerl and Robert W. Kennard. "Ridge Regression: Applications to Nonorthogonal Problems". In: *Technometrics* 12.1 (1970), pp. 69–82. ISSN: 0040-1706. DOI: 10.2307/1267352.
- [15] Arthur E. Hoerl and Robert W. Kennard. "Ridge Regression: Biased Estimation for Nonorthogonal Problems". In: *Technometrics* 12.1 (1970), pp. 55–67. ISSN: 0040-1706. DOI: 10.2307/1267351.
- [16] Peter J. Lenk and Wayne S. DeSarbo. "Bayesian Inference for Finite Mixtures of Generalized Linear Models with Random Effects". en. In: *Psychometrika* 65.1 (Mar. 2000), pp. 93–119. ISSN: 0033-3123, 1860-0980. DOI: 10.1007/BF02294188.
- [17] David J. C. MacKay. "Bayesian Interpolation". In: *Neural Computation* 4.3 (May 1992), pp. 415–447. ISSN: 0899-7667. DOI: 10.1162/neco.1992.4.3.415.

- [18] Tj Mitchell and Jj Beauchamp. “Bayesian Variable Selection in Linear-Regression”. English. In: *Journal of the American Statistical Association* 83.404 (Dec. 1988), pp. 1023–1032. ISSN: 0162-1459. DOI: 10.2307/2290129.
- [19] Radford M Neal. “Probabilistic inference using Markov chain Monte Carlo methods”. In: *Department of Computer Science, University of Toronto Toronto, Ontario, Canada* (1993).
- [20] R. B. O’Hara and M. J. Sillanpää. “A review of Bayesian variable selection methods: what, how and which”. In: *Bayesian Analysis* 4.1 (Mar. 2009), pp. 85–117. ISSN: 1936-0975, 1931-6690. DOI: 10.1214/09-BA403.
- [21] Trevor Park and George Casella. “The Bayesian Lasso”. English. In: *Journal of the American Statistical Association* 103.482 (June 2008), pp. 681–686. ISSN: 0162-1459. DOI: 10.1198/016214508000000337.
- [22] Nicholas G. Polson, Scott James G., and Jesse Windle. “Bayesian Inference for Logistic Models Using Pólya–Gamma Latent Variables”. In: *Journal of the American Statistical Association* 108.504 (2013), pp. 1339–1349. ISSN: 0162-1459. DOI: 10.1080/01621459.2013.829001.

- [23] Håvard Rue, Sara Martino, and Nicolas Chopin. “Approximate Bayesian Inference for Latent Gaussian models by using Integrated Nested Laplace Approximations”. In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 71.2 (Apr. 2009), pp. 319–392. ISSN: 1369-7412. DOI: 10.1111/j.1467-9868.2008.00700.x.
- [24] Steven L. Scott. “Data augmentation, frequentist estimation, and the Bayesian analysis of multinomial logit models”. en. In: *Statistical Papers* 52.1 (Feb. 2011), pp. 87–109. ISSN: 1613-9798. DOI: 10.1007/s00362-009-0205-0.
- [25] Robert Tibshirani. “Regression Shrinkage and Selection via the Lasso”. In: *Journal of the Royal Statistical Society. Series B (Methodological)* 58.1 (1996), pp. 267–288. ISSN: 0035-9246.
- [26] Luke Tierney and Joseph B. Kadane. “Accurate Approximations for Posterior Moments and Marginal Densities”. In: *Journal of the American Statistical Association* 81.393 (1986), pp. 82–86. ISSN: 0162-1459. DOI: 10.1080/01621459.1986.10478240.