Bayesianische Regression

lineare und logistische Modelle

Lona Koers

LMU

25. Juli 2025

Motivation und Intuition

TODO: gutes Intro-Beispiel

- Generalisierte Lineare Modelle (GLMs)
- Punktvorhersage vs. Verteilung vorhersagen
- warum reicht uns ein CI / PI

Outline

Frequentistisches $\rightarrow$ bayesianisches lineares Modell

Annahmen:

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

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

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

Bayesianisches lineares Modell:

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

Prior-Verteilungen für das bayesianische LM

Prior-Annahme für $\boldsymbol{\theta}$ (und evtl. σ^2) notwendig $\rightarrow$ sehr vielseitige Modell-Anpassung möglich

Normal-Invers-Gamma Prior:

$$\begin{aligned} \boldsymbol{\theta} \mid \sigma^2 &\sim \mathcal{N}(\breve{\boldsymbol{\mu}}, \sigma^2 \breve{\Sigma}) \\ \sigma^2 &\sim \mathsf{IG}(\breve{a}, \breve{b}) \\ \boldsymbol{\theta}, \sigma^2 &\sim \mathsf{NIG}(\breve{\boldsymbol{\mu}}, \sigma^2 \breve{\Sigma}, \breve{a}, \breve{b}) \end{aligned}$$

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

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

TODO: Bild

Uninformative Prior als Spezialfall der NIG-Prior

(TODO: relevant?)

Uninformative Prior

z.B. mit NIG-Prior mit Prior Parametern

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

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

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

TODO: Bild

Regularisierung: frequentistisch $\rightarrow$ bayesianisch

Erinnerung: Regularisierung im *frequentistischen* LM durch Minimierung von

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

mit Regularisierungs-Parameter $\lambda > 0$.

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

Regularisierung durch Wahl der Prior-Verteilung

Ridge Regularisierung

- Frequentistisch [15, 16]: $\text{pen}(\boldsymbol{\theta}) = \|\boldsymbol{\theta}\|_2^2$
- Bayesianisch [19]: $\boldsymbol{\theta} \sim \mathcal{N}(\mathbf{0}, \tau^2 \boldsymbol{I})$ mit $\tau^2 \propto \frac{1}{\lambda}$

Lasso Regularisierung

- Frequentistisch [26]: $\text{pen}(\boldsymbol{\theta}) = \|\boldsymbol{\theta}\|_1$
- Bayesianisch [23]:

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

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

Problem: keine Variablenselektion (im Gegensatz zu frequentistischem Lasso)

→ Alternative Priors für Variablenselektion: Spike and Slab [20], Horseshoe [4], u.v.m.

TODO: Bild regularization Priors + update

Regularisierung in Anwendung

(TODO: relevant? -> Anhang bzw. Diskussion)

Vorteile von bayesianischer Regularisierung sind u.a.:

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

Outline

Bayesianisches LM $\rightarrow$ **GLM**

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

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

Bayesianisches GLM → **logistisches** Modell

Bayesianisches logistisches Modell

$$
\begin{aligned}
\boldsymbol{y}_i \mid \boldsymbol{\theta} &\sim \mathsf{Bin}(1, g^{-1}(\boldsymbol{x}_i\boldsymbol{\theta})), \quad i = 1, \dots, n \\
g^{-1}(\boldsymbol{x}_i\boldsymbol{\theta}) &= \sigma(\boldsymbol{x}_i\boldsymbol{\theta})
\end{aligned}
$$

Für Beobachtungen $\boldsymbol{x}_i = (1, x_{i1}, \dots, x_{ip})^\top$ 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) addressieren Separation und fördern Shrinkage [11, 13]
- Für Regularisierung können dieselben Priors wie im LM verwendet werden [22, 8, 7]

TODO: Bild mti decision boundard? Lieber nicht ...

Outline

Inferenz der Parameter Posterior

Erinnerung

- Bayes-Regel zur Ermittlung der Parameter Posterior

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

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

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

bayesianisches LM: exakte Inferenz mit konjugierten Prioris

Inferenz mit konjugierten Priors

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

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

mit

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

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

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

- **Ridge**: Spezialfall der NIG-Prior $\rightarrow$ konjugierte Berechnung der Parameter Posterior
- **Lasso**: Prior hat keine geschlossene Form, aus der Posterior kann aber einfach mit Gibbs-Sampling simuliert werden [23]

bayesianisches GLM: approximative Inferenz

Problem: Inferenz mit konjugierten Priors ist nur sehr selten möglich [**polson_bayesian_2013**]

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

- Markov chain Monte Carlo Methoden
  - **Metropolis-Hastings Algorithmus** [14]
  - Gibbs Sampling (bedingte Konjugiertheit) [6]
  - Hamiltonian Monte Carlo (v.a. hochdimensionale Posterior) [21]
- Data Augmentation mit latenten Variablen, um künstlich Konjugiertheit zu induzieren [1, 17, 9, 25]
- **Laplace Approximation** [**tierney_accurate_1986**]

Metropolis Hastings Algorithmus [14]

- **Idee**: Aus der Posterior $p(\boldsymbol{\theta} \mid \boldsymbol{y})$ 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 $\int p(\boldsymbol{y} \mid \boldsymbol{\theta})\, p(\boldsymbol{\theta}) d\boldsymbol{\theta}$) bekannt sind
- **Inputs**:
  - Anzahl der Ziehungen $K \rightarrow$ frei wählbar
  - Likelihood $p(\boldsymbol{\theta} \mid \boldsymbol{y})$ und Prior $p(\boldsymbol{\theta}) \rightarrow$ bekannt
  - Proposal Verteilung $q \rightarrow$ muss sorgfältig gewählt werden

Metropolis Hastings: Wahl der Proposal Verteilung q

Effizienz des Algorithmus ist stark abhängig von der Proposal Verteilung.

$\rightarrow$ **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 $p(\boldsymbol{\theta} \mid \boldsymbol{y})$ wird meist mit IWLS geschätzt [10, 18, 25]

- Scott [25] schlägt **Verteilungen mit schweren Rändern** vor $\rightarrow$ mehr Mixing, kürzerer Burn-in und schnellere Konvergenz

Laplace Approximation (LA) [**tierney_accurate_1986**]

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

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

  mit Maximum-a-Posteriori Schätzer $\hat{\boldsymbol{\theta}}_{MAP}$ und Hesse-Matrix H von $p(\boldsymbol{\theta} \mid \boldsymbol{y})$
- Erweiterung für hierarchische Modelle: Integrated Nested Laplace Appriximation (INLA) [24]

Beispiel 2: Vergleich von approximativen Inferenz Methoden

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

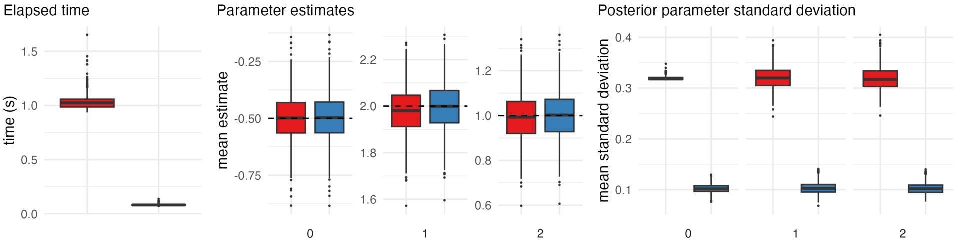
$$\boldsymbol{X} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}), \quad \boldsymbol{\theta} = (-0.5, 2, 1)$$
$$\text{linear: } \boldsymbol{y} \mid \boldsymbol{\theta} \sim \mathcal{N}(\boldsymbol{X\theta}, \boldsymbol{I})$$
$$\text{logistic: } \boldsymbol{y} \mid \boldsymbol{\theta} \sim \text{Ber}(\sigma(\boldsymbol{X\theta}))$$

Experiment: für jeden Datensatz wurde angepasst:

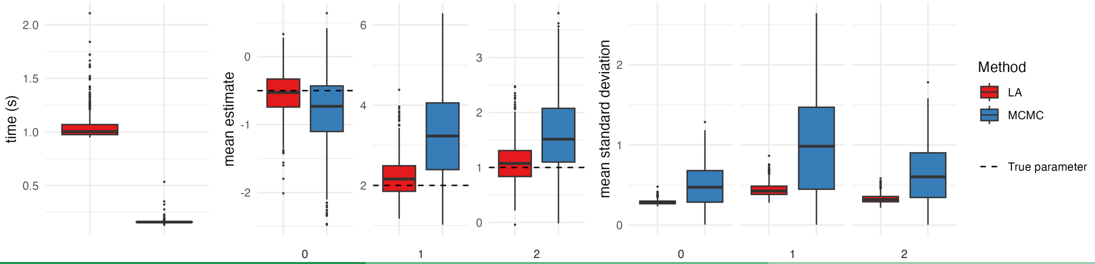
- Ein lineares und ein logsistisches Regressionsmodell mit $\boldsymbol{\theta} \sim \mathcal{N}(0, 10 \cdot \boldsymbol{I})$ und $\sigma^2 = 10$
- Mit Laplace Approximation und Metropolis-Hastings (K = 5.000, Burn-in = 500, Thinning Intervall = 10)

Beispiel 2: LA vs. Metropolis-Hastings

Lineares Modell



Logistisches Modell



Vorhersagen mit bayesianischen Modellen

Aus der Bayes Regel:

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

Posterior Predictive Distribution

$\rightarrow$ Vorhersagen für $\tilde{\boldsymbol{X}}$ im Bayesianischen GLM [3, 2]

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

Für das bayesianische **logistische Modell** berechnet man

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

mit $y_i \in \{0 \text{ (negative)}, 1 \text{ (positive)}\}$

TODO: Bild

Berechnung der Posterior Predictive Distribution

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

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

Alternativ (und viel wichtiger): **Approximation**

- für Metropolis-Hastings: $p(\tilde{\boldsymbol{y}} = 1 \mid \boldsymbol{\theta}, \boldsymbol{y}) \approx \frac{1}{K} \sum_{k=1}^{K} \sigma(\tilde{\boldsymbol{X}}\boldsymbol{\theta}^{(k)})$
- für Laplace Approximation:
  - LA-approximierte PPD analytisch berechnen:

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

  - Samples $\boldsymbol{\theta}^{(s)} \sim \mathcal{N}(\hat{\boldsymbol{\theta}}_{MAP}, H^{-1}(\hat{\boldsymbol{\theta}}_{MAP}))$ mit $s = 1, \ldots, S$ aus der LA-approximierten Parameter Posterior ziehen und wie bei Metropolis-Hastings vorgehen

Literatur Empfehlungen

- Bayesianische Regression (v.a. für praktische Anwendung): Gelman et al. [12]
- Prior Verteilungen (v.a. Shrinkage): Erp, Oberski, and Mulder [7] und Celeux et al. [5]
- Software: z.B. `brms` in `R`, `PyMC` in `python`

Outline

Outline

Herleitung der kojugierten Priori

.

Ergebnisse des Regularisierungsexperiments

.

Posterior Inference: exakter Metropolis (-Hastings) Algorithmus

1. Initialize $\boldsymbol{\theta}^{(1)}$
2. For $k = 1, \ldots, K$
   1. Draw $\boldsymbol{\theta}^{(*)}$ from the *proposal distribution* $q(\boldsymbol{\theta}^{(*)} \mid \boldsymbol{\theta}^{(k)})$
   2. calculate the *acceptance probably*

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

   3. Accept or discard the proposal $\boldsymbol{\theta}^{(*)}$ (for $u \sim$ Uni$[0, 1]$)

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

Und z.B.

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

mit $H(\boldsymbol{\theta}) = \nabla^2_{\boldsymbol{\theta}} \log\Big(p(\boldsymbol{\theta}^{(k)} \mid \boldsymbol{y})\, p(\boldsymbol{\theta}^{(k)})\Big)$

Posterior Inference: Mittelwert und Varianz für Laplace Approximation

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

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

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

Experiment-Setup: Regularisierung

Beispiel 1: Regularisierung in data-sparse Szenarien

Experiment-Setup: approximate Inference

.

[1] James H. Albert and Siddhartha Chib. “Bayesian Analysis of Binary and Polychotomous Response Data”. In: *Journal of the American Statistical Association* 88.422 (1993), pp. 669–679. ISSN: 0162-1459. DOI: 10.2307/2290350.

[2] 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.

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] Sylvia Frühwirth-Schnatter and Rudolf Frühwirth. “Auxiliary mixture sampling with applications to logistic models”. In: *Computational Statistics & Data Analysis* 51.7 (Apr. 2007), pp. 3509–3528. ISSN: 0167-9473. DOI: 10.1016/j.csda.2006.10.006.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] CC Holmes and Leonhard Knorr-Held. "Efficient simulation of Bayesian Logistic Regression Models". In: (Jan. 2003).
[18] 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.
[19] 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.
[20] 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.
[21] Radford M Neal. "Probabilistic inference using Markov chain Monte Carlo methods". In: *Department of Computer Science, University of Toronto Toronto, Ontario, Canada* (1993).
[22] 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.

[23] 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.

[24] 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.

[25] 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.

[26] 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.