

Polya-Gamma Distribution and Data Augmentation

Toni Luhdo

March 3, 2025

Introduction

- Logistic regression is a fundamental model for classification problems.
- Bayesian inference requires efficient sampling methods.
- The Polya-Gamma distribution provides an elegant solution for the Gibbs sampling strategy.

Polya-Gamma Distribution

Definition: A random variable ω is called Polya-Gamma distributed with parameters b und c if it has the following density function:

$$p(\omega|b, c) = \frac{\exp(-c^2\omega/2)p(\omega|b, 0)}{E_{PG(b,0)}[\exp(-c^2\omega/2)]} \quad (1)$$

where $p(\omega|b, 0)$ is the standard Polya-Gamma density.

- $b \in \mathbb{R}^+$, $b > 0$ Steuert die Form der Verteilung (Shape-Parameter)
- $c \in \mathbb{R}$, $b > 0$ Beeinflusst die Skalierung/Neigung (Skalenparameter)
- $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$, für $x > 0$

Laplace-Transformations-Properties

Die Laplace-Transformation der Polya-Gamma Distribution ist gegeben durch:

$$E[\exp(-\omega t)] = \frac{\cosh^b(c/2)}{\cosh^b(\sqrt{\frac{c^2/2+t}{2}})} \quad (2)$$

- Nützlich für analytische Berechnungen und Simulationen.
- Basis für die MCMC-Methodik in der Bayesschen Inferenz.

Alternative Darstellung der Polya-Gamma Distribution

- Die Verteilung kann als unendliche Summe von Gamma-verteilten Variablen dargestellt werden:

$$\omega \sim \sum_{k=1}^{\infty} \frac{G_k}{(k - 1/2)^2 + c^2/(4\pi^2)}, \quad G_k \sim \text{Gamma}(b, 1) \quad (3)$$

- Diese Darstellung verdeutlicht die Beziehung zur Gamma-Verteilung und eröffnet alternative Berechnungsmöglichkeiten.

Properties der Polya-Gamma-Klasse

- Geschlossen unter Faltung: Sind $\omega_1 \sim PG(b_1, z)$ und $\omega_2 \sim PG(b_2, z)$ unabhängig, dann gilt:

$$\omega_1 + \omega_2 \sim PG(b_1 + b_2, z) \quad (4)$$

- Expectation:

$$E(\omega) = \frac{b}{2c} \tanh(c/2) \quad (5)$$

Bayessche Inferenz für Logistische Modelle

Die Likelihood-Funktion eines logistischen Modells ist:

$$P(y_i = 1|x_i, \beta) = \frac{1}{1 + e^{-x_i^T \beta}} \quad (6)$$

- Introduction eines latenten Polya-Gamma Parameters ω_i ermöglicht eine Gibbs-Sampling Methode.
- Die posteriori Verteilung ist normal für β und Polya-Gamma für ω . Umformulieren oben umbenennen in omega und direkt in Gleichung

Transformation der Likelihood mit Polya-Gamma

Die Likelihood-Funktion eines logistischen Modells ist ursprünglich gegeben durch:

$$P(y_i = 1|x_i, \beta) = \frac{1}{1 + e^{-x_i^T \beta}} \quad (7)$$

Durch die Polya-Gamma Data Augmentation nutzen wir die Identität:

$$\frac{e^{y_i x_i^T \beta}}{1 + e^{x_i^T \beta}} = \int_0^\infty e^{-\omega_i (x_i^T \beta)^2 / 2} p(\omega_i | 1, 0) d\omega_i \quad (8)$$

Dadurch kann die Likelihood so umgeschrieben werden, dass sie eine bedingte Normalverteilung für β ergibt, was Gibbs-Sampling ermöglicht.

Gibbs-Sampler für das logistische Modell

- 1 Ziehe $(\omega_i|\beta) \sim PG(n_i, x_i^T \beta)$ für alle i .
- 2 Ziehe $(\beta|z, \omega) \sim N(m_\omega, V_\omega)$, where:

$$m_\omega = V_\omega X^T \kappa, \quad V_\omega = (X^T \Omega X + B^{-1})^{-1} \quad (9)$$

mit $\kappa = (y_1 - n_1/2, \dots, y_N - n_N/2)$ und Ω ist die Diagonalmatrix der ω_i .

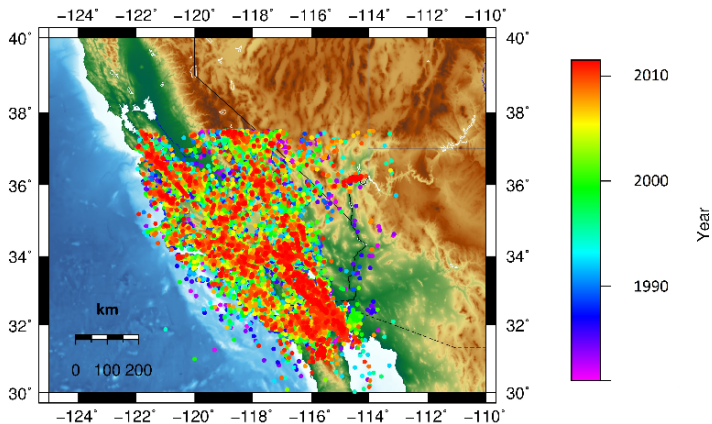
B ist die a-priori Kovarianzmatrix für die Koeffizienten β

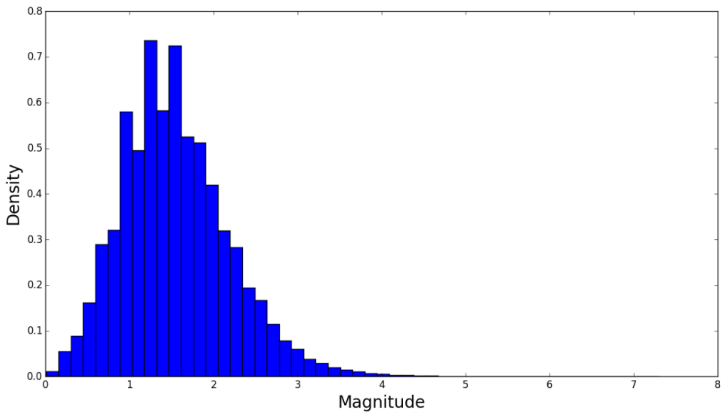
Das Gutenberg-Richter-Gesetz für Erdbeben

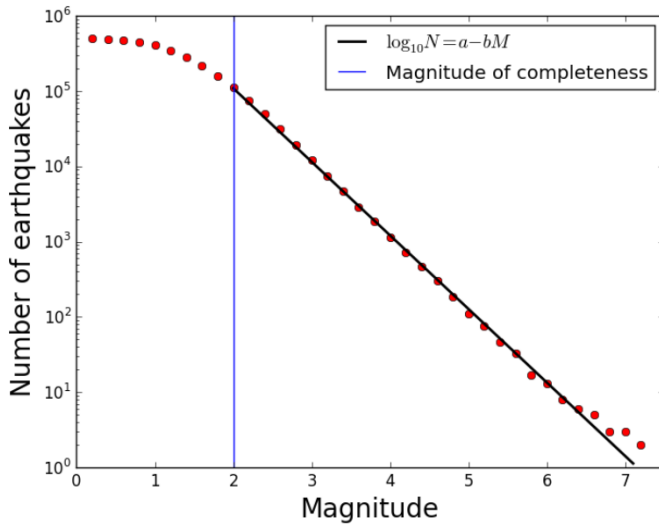
- empirisches Gesetz, welches die Beziehung der Magnitude zur Häufigkeit von Erdbeben angibt
- vorgeschlagen von Charles Francis Richter und Beno Gutenberg (1956)
- Gutenberg-Richter Gesetz

$$\log_{10}(N) = a - bM$$

- kumulierte jährliche Anzahl N von Erdbeben mit einer Magnitude größer als M
- a Gesamtseismizitätsrate des betrachteten Gebiets
- b drückt den relativen Anteil von kleinen zu größeren Erdbeben aus







Schätzung von b mit Polya-Gamma Data Augmentation

- Die Bayesianische Schätzung von b kann mit einer Polya-Gamma Data Augmentation erfolgen.
- Durch Umformung der Likelihood-Funktion in ein logistisches Modell:

$$P(y_i = 1|m_i, b) = \frac{1}{1 + e^{-m_i b}} \quad (10)$$

kann die Polya-Gamma-Methode für effizientes Gibbs Sampling genutzt werden.

- Dies könnte helfen, robuste Schätzungen für seismische Aktivitätsparameter zu erhalten.

Schätzung von b mit Polya-Gamma Data Augmentation

- Die Bayessche Schätzung des Parameters b kann durch eine Polya-Gamma Data Augmentation erleichtert werden.
- Durch Umformung der Gutenberg-Richter-Gleichung in ein logistisches Modell:

$$P(y_i = 1 | m_i, b) = \frac{1}{1 + e^{-m_i b}} \quad (11)$$

kann die Polya-Gamma-Methode genutzt werden, um eine Gibbs-Sampling-Strategie zur Bestimmung von b zu entwickeln.

- Vorteile dieser Methode:
 - Sie ermöglicht eine robuste Schätzung von b , selbst bei kleinen oder verrauschten Daten.
 - Durch die Bayesianische Herangehensweise können Unsicherheiten in der Schätzung direkt quantifiziert werden.
 - Sie erlaubt eine natürliche Einbindung zusätzlicher a-priori

Warum Polya-Gamma für die Schätzung?

- Modellierung der Erdbebenhäufigkeit als Poisson-Prozess:

$$N(M) \sim \text{Poisson}(\lambda(M))$$

- Log-Transformation führt zu einem linearen Modell:

$$\log N(M) = a - bM$$

- Durch die Polya-Gamma Methode wird Gibbs-Sampling zur Schätzung ermöglicht.

Summary

- Die Polya-Gamma Distribution vereinfacht die Bayessche Inferenz für logistische Modelle erheblich.
- Die Gibbs-Sampling-Strategie ermöglicht effiziente numerische Berechnungen.
- Die Methode ist gut geeignet für großskalige Datenanalysen und maschinelles Lernen.

Referenzen

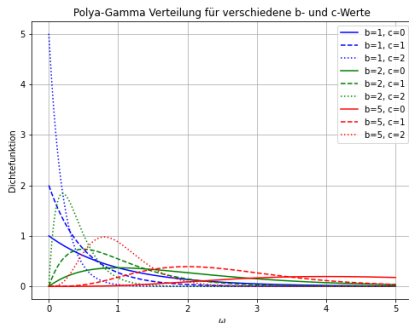
Polson, N. G., Scott, J. G., Windle, J. (2013). Bayesian inference for logistic models using Polya-Gamma latent variables. *Journal of the American Statistical Association*, 108(504), 1339-1349.

Anhang: Bedeutung der Parameter b und c in der Polya-Gamma Distribution

- ****Parameter b (Shape-Parameter):****
 - Gibt an, wie stark die Verteilung skaliert ist.
 - Höhere Werte machen die Verteilung schmaler und konzentrierter.
 - Wird oft aus der Anzahl der Trials im Modell abgeleitet.
- ****Parameter c (Skalenparameter):****
 - Steuert die exponentielle Dämpfung.
 - Höhere Werte von c verschieben die Verteilung nach links.
 - In logistischen Modellen ist $c = x^T \beta$.

Anhang: Auswirkungen von b und c auf die Verteilung

- Falls $b = 1$ und $c = 0$: Die Verteilung ähnelt einer Gamma-Verteilung.
- Falls c steigt: Die Verteilung wird nach links verschoben.
- Falls b steigt: Die Verteilung wird schmaler und konzentrierter.



Anhang: Warum ist $p(\omega_i|1, 0)$ in der Likelihood-Transformation?

- In der logistischen Regression ergibt sich die Likelihood:

$$P(y_i = 1|x_i, \beta) = \frac{1}{1 + e^{-x_i^T \beta}} \quad (12)$$

- Durch die Polya-Gamma-Transformation wird dies umgeschrieben als:

$$P(y_i = 1|x_i, \beta) = \int_0^\infty e^{-\omega_i(x_i^T \beta)^2/2} p(\omega_i|1, 0) d\omega_i \quad (13)$$

- ****Warum $PG(1, 0)$?****

- $b = 1$ weil es sich um eine binäre logistische Regression handelt (Bernoulli-Modell).
- $c = 0$ weil keine exponentielle Dämpfung benötigt wird.