Polya-Gamma Verteilung und Data Augmentation

Toni Luhdo

March 3, 2025

Einführung

- Logistische Regression ist ein fundamentales Modell für Klassifikationsprobleme.
- Die Bayessche Inferenz erfordert effiziente Sampling-Methoden.
- Die Polya-Gamma-Verteilung ermöglicht eine elegante Lösung zur Gibbs-Sampling-Strategie.

Polya-Gamma-Verteilung

Definition: Eine Zufallsvariable ω heißt PolyaGamma verteilt mit den Parametern b und c wenn sie die folgende Dichtefunktion besitzt:

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

wobei $p(\omega|b,0)$ die Standard-Polya-Gamma-Dichte ist.

- $b \in \mathbb{R}^+, \quad b > 0$ Steuert die Form der Verteilung (Shape-Parameter)
- $c \in \mathbb{R}, \quad b > 0$ Beeinflusst die Skalierung/Neigung (Skalenparameter)

- $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt, \quad$ für $x > 0$

Laplace-Transformations-Eigenschaften

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

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

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

Alternative Darstellung der Polya-Gamma-Verteilung

- 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.

Eigenschaften 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:

$$\omega_1 + \omega_2 \sim PG(b_1 + b_2, z) \tag{4}$$

- Erwartungswert:

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

Bayessche Inferenz für Logistische Modelle

Die Likelihood-Funktion eines logistischen Modells ist:

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

- Einführung 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 umbennen in omega und direkt in Gleichung

Transformation der Likelihood mit Polya-Gamma

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

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

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

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

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 $(\beta|z,\omega) \sim N(m_\omega, V_\omega)$, wobei:

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

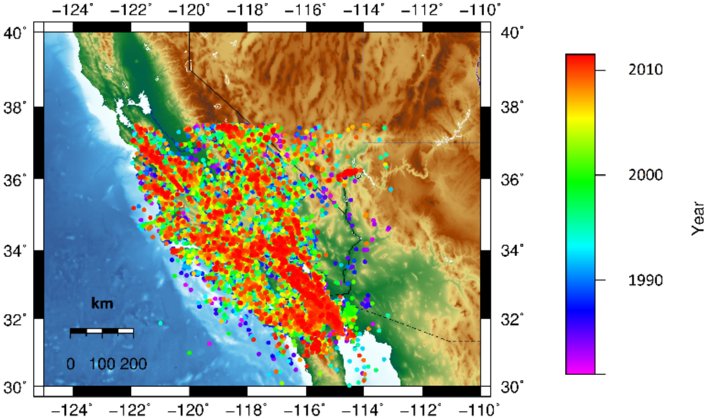
mit $\kappa = (y_1 - n_1/2, ..., y_N - n_N/2)$ und Ω ist die Diagonalmatrix der ω_i.
B ist die a-priori Kovarianzmatrix für die Koeffizienten β

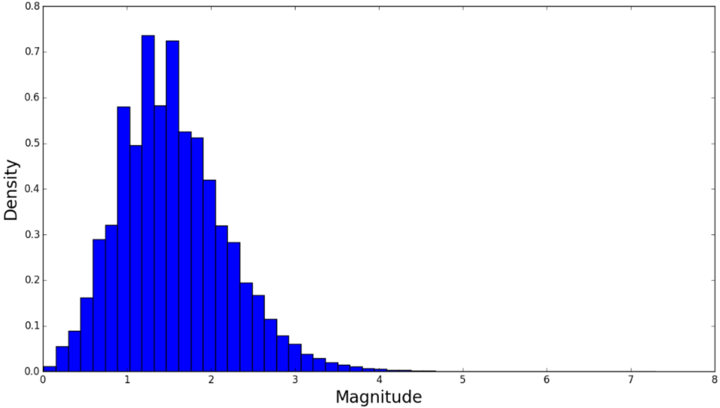
Das Gutenberg-Richter-Gesetz für Erdbeben

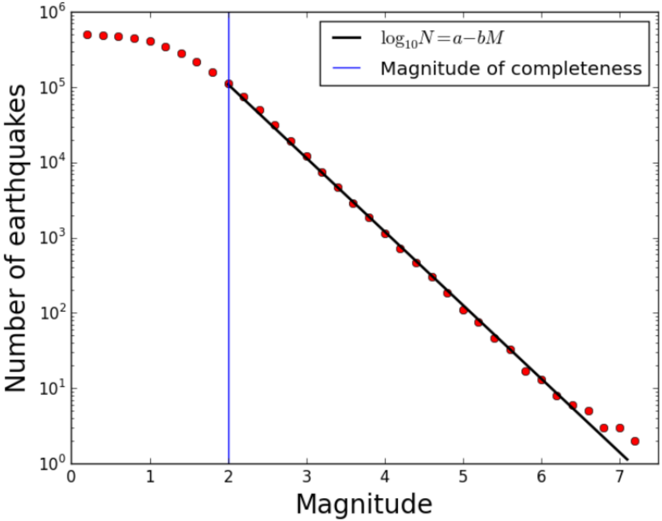
- empirisches Gesetz, welches die Beziehung der Magnitude zur Häufigkeit von Erdbeben angibt
- vorgeschlagen von Charles Francis Richter und Beno Gutenberg (1956)
- Gutenberg-Richter Gesetz

$$log10(N) = a - bM$$

- 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


−124° −122° −120° −118° −116° −114° −112° −110°
40° 38° 36° 34° 32° 30°
km
0 100 200
2010
2000
1990
Year


0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Density
0
1
2
3
4
5
6
7
8
Magnitude


10^6
10^5
10^4
10^3
10^2
10^1
10^0
Number of earthquakes
0
1
2
3
4
5
6
7
Magnitude
$\log_{10} N = a - bM$
Magnitude of completeness

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:

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

  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:

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

  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.

Zusammenfassung

- Die Polya-Gamma-Verteilung 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

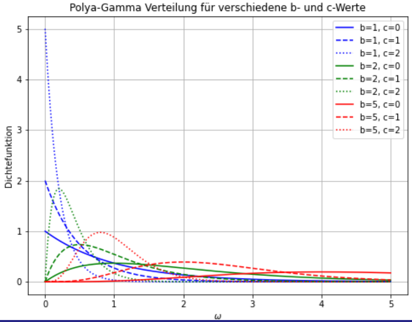
Polson, N. G., Scott, J. G., Windle, J. (2013). Bayesian inference for logistic models using Polya-Gamma latent variables. *Journal of the American Statistical Association*, 108(504), 1339-1349.

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

- **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:

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

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

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

- **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.