Introduction to Pólya-Gamma distribution

持橋大地
統計数理研究所 数理・推論研究系
daichi@ism.ac.jp

SVM 2017

2017-9-1 (Fri)

NAIST

Overview

- 離散分布とロジスティック分布
- ロジスティック回帰の完全ベイズ推定
- Pólya-Gamma 分布による Gibbs サンプラー
- ラプラス変換による証明
- 多値ロジスティック回帰, Stick-breaking 過程

Background

- 自然言語処理の離散分布

$$\mathbf{p} = (p_1, p_2, \cdots, p_K) \tag{1}$$

  で, 確率 $p_1, \cdots, p_K$ を直接モデル化するのは限界
  - 共分散が常に一定 (center $\leftrightarrow$ centre, nurse $\leftrightarrow$ hospital)
  - $\sum_k p_k = 1$ の制約 (分布を modulate するのが難しい)
- LDA: トピック分布 $\{p(w|k)\}$ を通じて,「まとめて」共分散を表現
- word2vec: 単語を個々にモデル化できる, 統計的背景が弱い

ロジスティック分布

$$p(k) = \frac{e^{x_k}}{e^{x_1} + e^{x_2} + \cdots + e^{x_K}}, \quad (x_1, \cdots, x_K) \sim \mathrm{N}(0, \Sigma) \tag{2}$$

- アイテム (単語やトピック) 間の共分散を行列 Σ で表現できる
- アイテム k に特徴ベクトル $\mathbf{x}_k = (x_{k1}, \cdots, x_{kD})$ があり, 重み $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_D)$ で回帰される場合は

$$p(k|\mathbf{x}) = \frac{e^{\boldsymbol{\beta}^T \mathbf{x}_k}}{\sum_k e^{\boldsymbol{\beta}^T \mathbf{x}_k}} \quad \cdots \text{ ロジスティック回帰モデル.} \tag{3}$$

回帰係数 $\boldsymbol{\beta}$ の学習

$$p(k|\mathbf{x}) = \frac{\exp\left(\sum_{j=1}^{D} \beta_j x_{kj}\right)}{\sum_k \exp\left(\sum_{j=1}^{D} \beta_j x_{kj}\right)} \tag{4}$$

- Gradient を計算して最適化すればよいのでは?
  $\downarrow$
  教師あり学習なら (大体)Yes, 教師なし学習では No
- 学習途中で, アイテム k の「特徴ベクトル」$\mathbf{x}_k$ 自体が学習すべき潜在変数
  - SGD 等で最適化してしまうと, あっという間に局所解にはまる
  - 正しい $\boldsymbol{\beta}$ の事後分布を求める必要.

例: Continuous-space topic models (NL213)

$$p(w|d) = \frac{p(w)\exp(\boldsymbol{\theta}_d^T \vec{w})}{\sum_v p(v)\exp(\boldsymbol{\theta}_d^T \vec{v})} \tag{5}$$

- 潜在層が連続値 (ガウス分布) の場合のニューラルネット型トピックモデル
- 文書 d の潜在ベクトル $\boldsymbol{\theta}_d$ も, 単語ベクトル $\vec{w}$ も学習前には未知
- $\boldsymbol{\theta}$, $\vec{w}$ の要素ごとに単純なランダムウォーク MCMC で学習 (NL213).
- $\boldsymbol{\theta}$ の「事後分布」を求めたい

ロジスティック回帰のベイズ推定

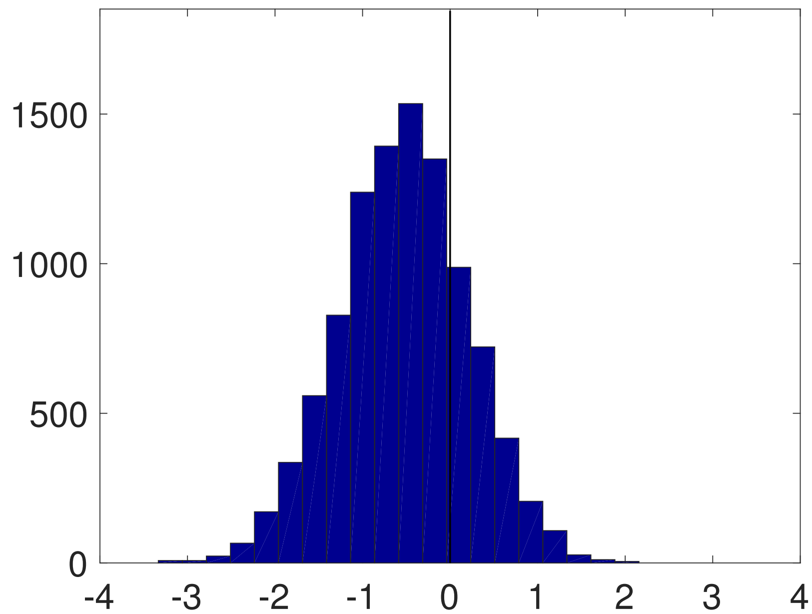
- 最も単純な場合:

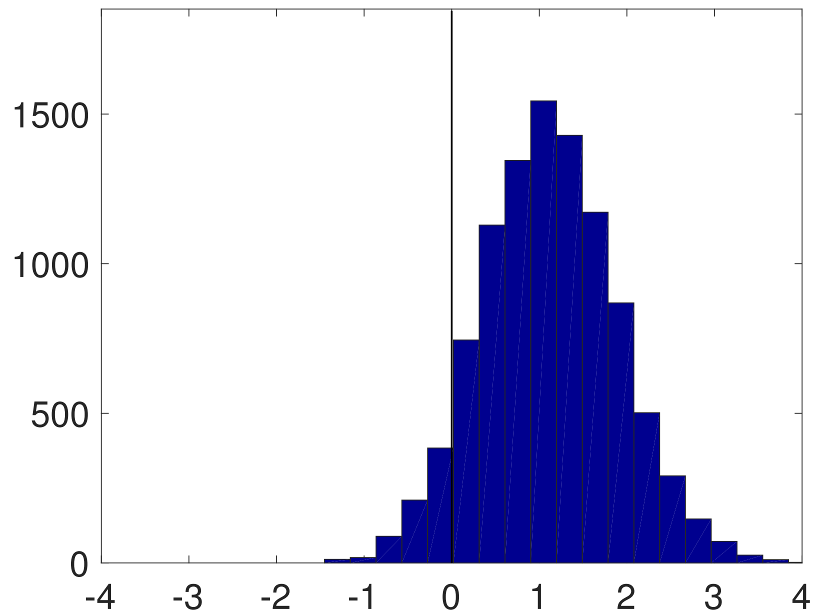
$$\begin{cases} x \sim \mathrm{N}(0,1) \\ y \sim \mathrm{Bernoulli}(\sigma(x)) \end{cases} \qquad \left(\sigma(x) = \frac{1}{1+e^{-x}}\right) \tag{6}$$

- いま, 観測値 $\mathbf{y} = (0,0,1,0)$ だったとする. x の推定値は?
- Gibbs サンプリング: Albert&Chib (1993) の有名な Probit の場合と同様に, Groenewald&Mokgatlhe (2005) などで提案

ロジスティック回帰のベイズ推定 (2)

$\mathbf{y} = [0\ 0\ 1\ 0]$ の場合　　　　$\mathbf{y} = [1\ 1\ 1\ 1]$ の場合





- x を点推定するのは, 大きく問題あり!
- 特に, 少量の観測値からの推測 (よくある), 社会科学・医学等の応用の場面

ロジスティック回帰のベイズ推定 (3)

- x の事後分布を, サンプルではなく分布で推定できないか?
  - $p(x|\mathbf{y})$ はガウス分布に似ているが, 正確には正規分布ではない
- 補助変数 ω を導入することにより, x をガウス分布として推定できる

$$\begin{cases} \text{Draw} \ \ \omega \sim \text{Pólya-Gamma}\,(n, x) \\ \text{Draw} \ \ x \sim \mathrm{N}(\mu(\omega), V(\omega)) \end{cases} \tag{7}$$

- Pólya-Gamma 分布とは?

論文

- Polson&Scott, “Bayesian inference for logistic models using Pólya-Gamma latent variables”,
arXiv (2011), JASA (2013)

[stat.ME] 22 Jul 2013

Bayesian inference for logistic models using
Pólya-Gamma latent variables

Nicholas G. Polson*
University of Chicago

James G. Scott†
Jesse Windle‡
University of Texas at Austin

First Draft: August 2011
This Draft: July 2013

Abstract

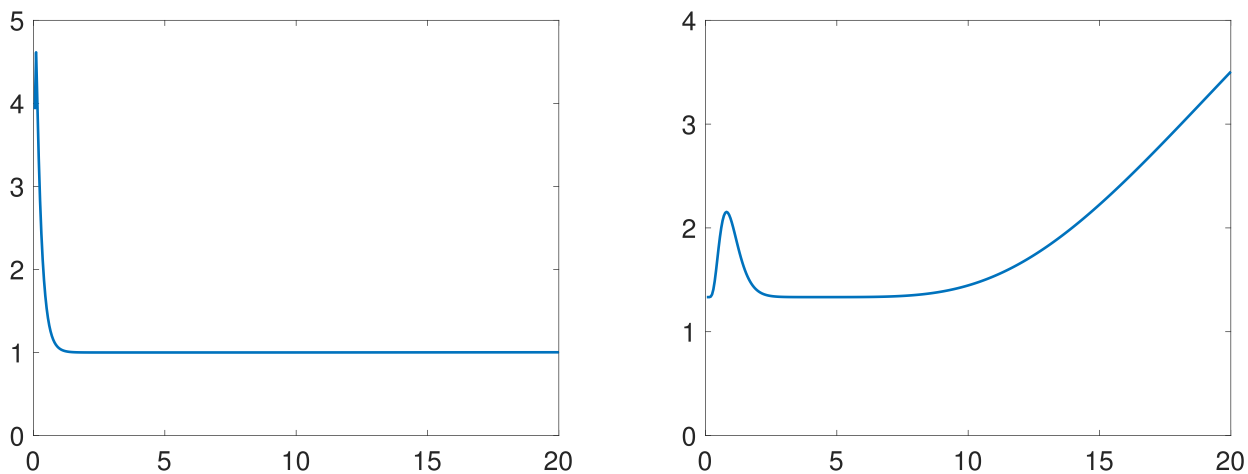
We propose a new data-augmentation strategy for fully Bayesian inference in models with binomial likelihoods. The approach appeals to a new class of Pólya-Gamma distributions, which are constructed in detail. A variety of examples are presented to show the versatility of the method, including logistic regression, negative binomial regression,

- 激しく難解!!

Pólya-Gamma 分布

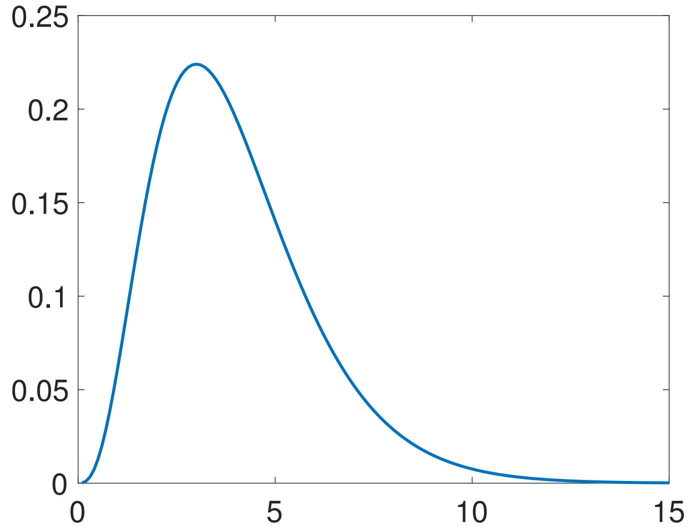
$$\mathrm{PG}(b,c)=\frac{1}{2\pi^2}\sum_{k=1}^{\infty}\frac{g_k}{(k-\frac{1}{2})^2+\frac{c^2}{4\pi^2}},\quad g_k\sim\mathrm{Ga}(b,1)\ . \tag{8}$$

- 無限個のガンマ乱数 $\mathrm{Ga}(b,1)$ の重みつき和
- Pólya の名前は, Pólya frequency function という解析学の概念に由来する (Karlin 1968)
  - 指数乱数の無限和で表現できる分布のクラス

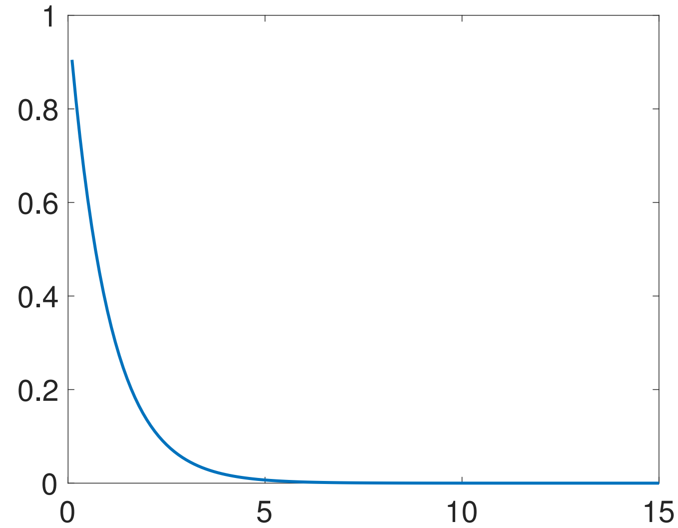


$\mathrm{PG}(1,0)$ $\quad$ $\mathrm{PG}(4,0)$

ガンマ分布



$\mathrm{Ga}(4, 1)$



$\mathrm{Ga}(1, 1) = \mathrm{Exp}(1)$

$$\mathrm{Ga}(x|a, b) = \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx} \tag{9}$$

- 注:

$$\Gamma(a) = \int_0^\infty x^{a-1} e^{-x} dx \ . \tag{10}$$

ガンマ分布 (2)

- これから, ガンマ分布のラプラス変換は

$$
\begin{aligned}
E[e^{-tx}]_{x\sim\mathrm{Ga}(a,b)} &= \int_0^\infty e^{-tx} x^{a-1} e^{-bx} dx && (11)\\
&= \int_0^\infty x^{a-1} e^{-(b+t)x} dx && (12)\\
&= \frac{\Gamma(a)}{(b+t)^a} \, . && (13)
\end{aligned}
$$

- 特に, 確率変数 X が $X \sim \dfrac{1}{a}\mathrm{Ga}(b,1)$ に従うとき,

$$
\begin{aligned}
E[e^{-tx}] &= \int_0^\infty e^{-t\cdot\frac{1}{a}x} \cdot x^{b-1} e^{-x} dx && (14)\\
&= \int_0^\infty x^{b-1} e^{-(\frac{t}{a}+1)x} dx && (15)\\
&= \frac{\Gamma(b)}{\left(1+\frac{t}{a}\right)^b} \, . \quad \square && (16)
\end{aligned}
$$

二値ロジスティック回帰

- 簡単のため, まずは二値 $\{0,1\}$ のロジスティック回帰モデルについて考える.
- いま, N 個の観測値があるとする.

$$\mathbf{y} = (y_1, y_2, \cdots, y_N) \in \{0,1\}^N \tag{17}$$

- このとき, 潜在変数 $x \in \mathbb{R}$ からのロジスティック回帰

$$p(y{=}1|x) = \sigma(x) = \frac{1}{1+e^{-x}} = \frac{e^x}{1+e^x} \tag{18}$$

で y が生成されたとすると,

$$p(\mathbf{y}|x) = \prod_{i=1}^{N} p(y_i|x) = \frac{(e^x)^n}{(1{+}e^x)^N} \tag{19}$$

ただし n は $\mathbf{y}$ の中で 1 が出た回数で, $n := \sum_{i=1}^{N} \mathbb{I}(y_i = 1)$.

二値ロジスティック回帰 (2)

$$p(\mathbf{y}|x) = \prod_{i=1}^{N} p(y_i|x) = \frac{(e^x)^n}{(1+e^x)^N} \tag{20}$$

- このとき $x \sim \mathrm{N}(0, \sigma^2)$ とすれば, ベイズの定理から

$$p(x|\mathbf{y}) \propto p(\mathbf{y}|x)p(x) = e^{-\frac{1}{2\sigma^2}x^2} \frac{(e^x)^n}{(1+e^x)^N} \tag{21}$$

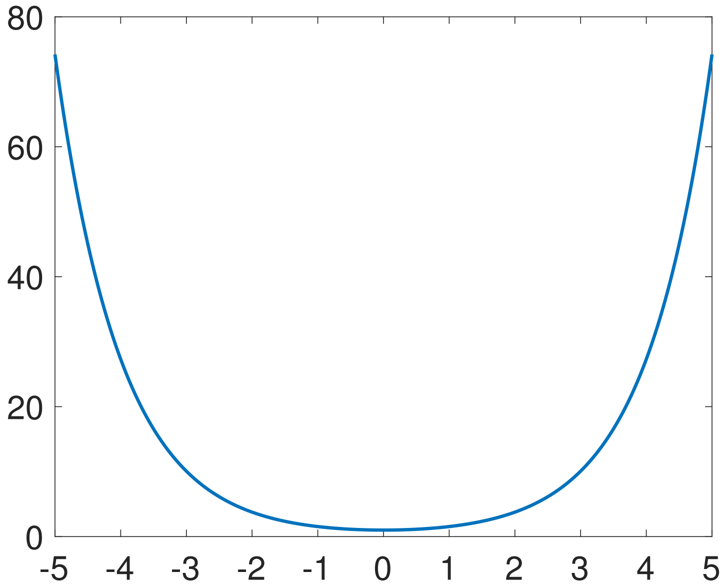
- x の事後分布が解析的な形にならない!!

$\downarrow$

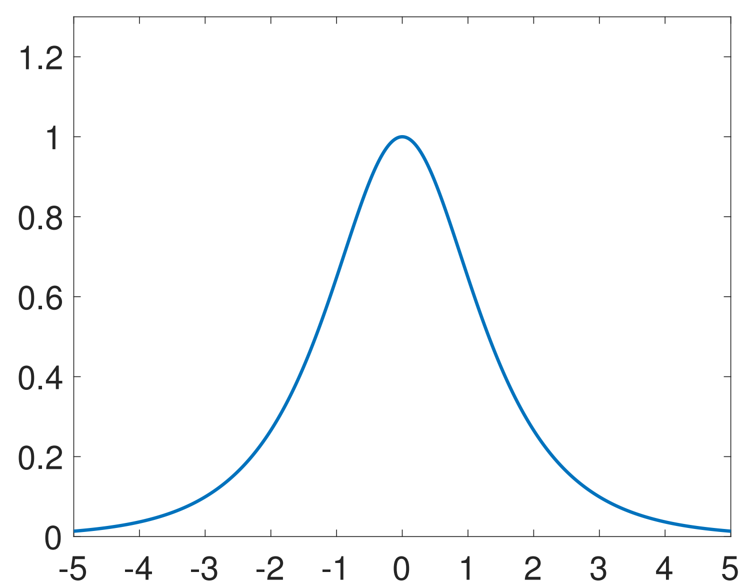
どうするか?

Hyperbolic cosine 関数

$$\cosh x = \frac{e^x + e^{-x}}{2} \tag{22}$$



$y = \cosh(x)$



$y = \cosh^{-1}(x)$

準備 (1)

$$\cosh x = \frac{e^x + e^{-x}}{2} \tag{23}$$

$$= \frac{1}{2}\left(e^x + \frac{1}{e^x}\right) = \frac{1}{2}\frac{e^{2x}+1}{e^x} \tag{24}$$

- よって,

$$\cosh\frac{x}{2} = \frac{1}{2}\frac{e^x+1}{e^{\frac{x}{2}}} \tag{25}$$

$$\therefore \quad \cosh^{-1}\frac{x}{2} = 2\,\frac{e^{\frac{x}{2}}}{e^x+1} \tag{26}$$

- これから,

$$\left(\cosh^{-1}\frac{x}{2}\right)^a = 2^a\frac{e^{\frac{x}{2}a}}{(e^x+1)^a} = 2^a\frac{(e^x)^{\frac{a}{2}}}{(e^x+1)^a} \tag{27}$$

が成り立つ.

準備 (2)

$$\left(\cosh^{-1}\frac{x}{2}\right)^a = 2^a\frac{e^{\frac{x}{2}a}}{(e^x+1)^a} = 2^a\frac{\left(e^x\right)^{\frac{a}{2}}}{(e^x+1)^a} \tag{28}$$

- よって,

$$\frac{(e^x)^a}{(e^x+1)^b} = \cosh^{-b}\frac{x}{2}\cdot\frac{e^{xa}}{e^{\frac{x}{2}b}}\cdot 2^{-b} \tag{29}$$

$$= 2^{-b}e^{(a-\frac{b}{2})x}\cosh^{-b}\frac{x}{2} \tag{30}$$

- これを使えば,

$$p(\mathbf{y}|x) = \frac{(e^x)^n}{(1+e^x)^N} = 2^{-N}e^{\left(n-\frac{N}{2}\right)x}\cosh^{-N}\frac{x}{2}\ . \tag{31}$$

二値ロジスティック回帰 (3)

$$p(\mathbf{y}|x) = 2^{-N} e^{\left(n - \frac{N}{2}\right)x} \cosh^{-N} \frac{x}{2} \tag{32}$$

- 上のままでは Gibbs サンプラーの使える共役形式にならないが，ここで次の魔法のような定理を導入する

Theorem 1 ω が Pólya-Gamma 分布 $\mathrm{PG}(b, 0)$ に従うとき，

$$E[e^{-t\omega}]_{\omega \sim \mathrm{PG}(b,0)} = \cosh^{-b} \sqrt{\frac{t}{2}} \tag{33}$$

$\sqrt{\frac{t}{2}} = \frac{x}{2}$ すなわち $t = \frac{x^2}{2}$ とおけば，

$$E[e^{-\frac{x^2}{2}\omega}]_{\omega \sim \mathrm{PG}(b,0)} = \cosh^{-b} \frac{x}{2}\ . \tag{34}$$

注意

$$E[e^{-\frac{x^2}{2}\omega}]_{\omega\sim\mathrm{PG}(b,0)} = \cosh^{-b}\frac{x}{2}$$

は, ガウス分布の分散 (精度) に関する scale mixture を表している

- $e^{-\frac{x^2}{2}\omega} = e^{-\frac{x^2}{2/\omega}}$ なので, 精度 $\omega = \sigma^{-1}$ を PG 分布に従って様々に変えたときの, ガウス分布 $x \sim \mathrm{N}(0, \sigma^2)$ の確率密度の期待値
- 参考: t 分布

$$\int_0^\infty \mathrm{N}(x|\mu, (\eta\lambda)^{-1})\mathrm{Ga}(\eta|\nu/2, \nu/2)d\eta \tag{35}$$

$$= \mathrm{St}(x|\mu, \lambda, \nu) \tag{36}$$

二値ロジスティック回帰 (4)

(34) 式をもとの式に代入すれば,

$$p(\mathbf{y}|x) = \frac{(e^x)^n}{(1+e^x)^N} = 2^{-N} e^{\left(n-\frac{N}{2}\right)x} \int_0^\infty e^{-\frac{x^2}{2}\omega}\, \mathrm{PG}(\omega|b,0) d\omega \tag{37}$$

すなわち,

$$p(\mathbf{y},\omega|x) \propto \underbrace{e^{\left(n-\frac{N}{2}\right)x} e^{-\frac{x^2}{2}\omega}}_{\text{Gaussian}}\ \mathrm{PG}(\omega|N,0) \tag{38}$$

- ここで, 青い部分で次の定理が成り立つことを利用する.

Theorem 2

$$e^{-\frac{x^2}{2}\omega}\, \mathrm{PG}(\omega|b,0) \sim \mathrm{PG}(\omega|b,x) \tag{39}$$

- ◦ PG 事前分布 $\mathrm{PG}(\omega|b,0)$ は, 「尤度」$e^{-\frac{x^2}{2}\omega}$ によって, 事後分布 $\mathrm{PG}(\omega|b,x)$ になる

二値ロジスティック回帰 (5)

- ゆえに,

$$p(x,\omega|\mathbf{y}) \propto p(x)p(\mathbf{y},\omega|x) \tag{40}$$

$$\propto \overbrace{e^{-\frac{1}{2\sigma^2}x^2} e^{\left(n-\frac{N}{2}\right)x} e^{-\frac{x^2}{2}\omega}}^{\text{Gaussian}} \hspace{-4em} \underbrace{\phantom{e^{-\frac{x^2}{2}\omega}}\mathrm{PG}(\omega|N,0)}_{\mathrm{PG}(\omega|N,x)} \tag{41}$$

- これから, 次の 2 ステップを繰り返す Gibbs サンプラーで, x は正しい分布に収束する. □

  (i) x の下で, ω を $\mathrm{PG}(N,x)$ から draw

  (ii) ω の下で, x を (41) 式の Gaussian から draw

定理 1 (PG 分布のラプラス変換) の証明

$$x \sim \mathrm{PG}(b,0) = \frac{1}{2\pi^2}\sum_{k=1}^{\infty}\frac{g_k}{(k-\frac{1}{2})^2} \quad \text{のとき}, \tag{42}$$

$$E[e^{-tx}] = \int_0^\infty e^{-tx}\,\mathrm{PG}(x|b,0)dx \tag{43}$$

$$= \int_0^\infty e^{-t\left(\frac{1}{2\pi^2}\sum_{k=1}^{\infty}\frac{x}{(k-\frac{1}{2})^2}\right)} x^{b-1}e^{-x}dx \tag{44}$$

$$= \prod_{k=1}^{\infty}\int_0^\infty x^{b-1}e^{-\left(1+\frac{t}{2\pi^2\left(k-\frac{1}{2}\right)^2}\right)x}dx \tag{45}$$

$$= \prod_{k=1}^{\infty}\frac{\Gamma(b)}{\left(1+\frac{t}{2\pi^2\left(k-\frac{1}{2}\right)^2}\right)^b} \quad \text{(準備より)} \tag{46}$$

$$= \cosh^{-b}\sqrt{\frac{x}{2}}\,. \quad \text{(準備より)} \tag{47}$$

Weierstrass の分解定理

$$\cosh x = \prod_{k=1}^{\infty}\left(1 + \frac{4}{\pi^2}\frac{x^2}{(2k-1)^2}\right) \tag{48}$$

$x := \sqrt{\frac{x}{2}}$ を代入して,

$$\cosh\sqrt{\frac{x}{2}} = \prod_{k=1}^{\infty}\left(1 + \frac{4}{\pi^2}\frac{\frac{x}{2}}{(2k-1)^2}\right) \tag{49}$$

$$= \prod_{k=1}^{\infty}\left(1 + \frac{x}{2\pi^2(k-\frac{1}{2})^2}\right) \tag{50}$$

よって,

$$\cosh^{-b}\sqrt{\frac{x}{2}} = \prod_{k=1}^{\infty}\left(1 + \frac{x}{2\pi^2(k-\frac{1}{2})^2}\right)^{-b}. \tag{51}$$

定理 2 (PG 分布の事後分布) の証明

- ガウス分布の分散 (精度) の, $\mathrm{PG}(b,0)$ による scale mixture を考える.

$$f = \frac{e^{-\frac{c^2}{2}x}\,\mathrm{PG}(x|b,0)}{\int_0^\infty e^{-\frac{c^2}{2}x}\,\mathrm{PG}(x|b,0)dx} = \frac{e^{-\frac{c^2}{2}x}\,\mathrm{PG}(x|b,0)}{\cosh^{-b}\left(\frac{c}{2}\right)} \tag{52}$$

- ラプラス変換すると,

$$E[e^{-tx}]_{x\sim f} = \frac{\displaystyle\int_0^\infty e^{-\left(t+\frac{c^2}{2}\right)x}\,\mathrm{PG}(x|b,0)dx}{\cosh^{-b}\left(\frac{c}{2}\right)} \tag{53}$$

$$= \frac{\cosh^b\left(\frac{c}{2}\right)}{\cosh^b\left(\sqrt{\frac{t+\frac{c^2}{2}}{2}}\right)} \tag{54}$$

定理 2 (PG 分布の事後分布) の証明 (2)

- Weierstrass の分解定理より，

$$E[e^{-tx}]_{x\sim f} = \prod_{k=1}^{\infty}\left(\frac{1+\frac{c^2/2}{2\pi^2\left(k-\frac{1}{2}\right)^2}}{1+\frac{c^2/2+t}{2\pi^2\left(k-\frac{1}{2}\right)^2}}\right)^b \tag{55}$$

$$= \prod_{k=1}^{\infty}\left(\frac{1+\frac{c^2/2}{y}}{1+\frac{c^2/2+t}{y}}\right)^b = \prod_{k=1}^{\infty}\left(\frac{y+\frac{c^2}{2}}{y}\cdot\frac{y}{y+\frac{c^2}{2}+t}\right)^b$$

$$= \prod_{k=1}^{\infty}\left(\frac{2\pi^2 z + c^2/2}{2\pi^2 z + c^2/2 + t}\right)^b \quad \left(z = \left(k-\frac{1}{2}\right)^2\right) \tag{56}$$

- ここで

$$\frac{X}{X+t} = \left(\frac{X+t}{X}\right)^{-1} = \left(1+\frac{t}{X}\right)^{-1} \quad \text{なので,} \tag{57}$$

定理 2 (PG 分布の事後分布) の証明 (3)

$$E[e^{-tx}]_{x\sim f} = \prod_{k=1}^{\infty}\left(1+\frac{t}{2\pi^2 z+\frac{c^2}{2}}\right)^{-b} \tag{58}$$

$$= \prod_{k=1}^{\infty}(1+d_k^{-1}t)^{-b} \qquad \left(d_k = 2\pi^2\left(k-\frac{1}{2}\right)^2+\frac{c^2}{2}\right) \tag{59}$$

- ここで
$$(1+d_k^{-1})^{-b} = \frac{1}{\left(\frac{1}{d_k}t+1\right)^b} \tag{60}$$
はガンマ分布 $\frac{1}{d_k}\mathrm{Ga}(b,1)$ のラプラス変換だから, 変換を元に戻せば

$$f = \sum_{k=1}^{\infty}\frac{g_k}{2\pi^2(k-\frac{1}{2})^2+\frac{c^2}{2}}, \quad g_k \sim \mathrm{Ga}(b,1) \tag{61}$$

$$= \frac{1}{2\pi^2}\sum_{k=1}^{\infty}\frac{g_k}{(k-\frac{1}{2})^2+\frac{c^2}{4\pi^2}}, \quad g_k \sim \mathrm{Ga}(b,1) = \mathrm{PG}(b,c)\,. \;\; \square \tag{62}$$

Pólya-Gamma Gibbs sampling: ω の事後分布

- 元に戻って, ロジスティック回帰の学習アルゴリズムを導出する.
- 特徴ベクトル $x_i = (x_{i1}, \cdots, x_{id})$ をもつアイテムについて, n_i 回の試行についての観測値 y_i が

$$y_i \sim \mathrm{Binom}(n_i, \sigma(x_i^T \boldsymbol{\beta})) \tag{63}$$

  で得られたとする. ($i = 1, \cdots, N$)
- このとき, 前の議論から

$$p(y_i | x_i, \boldsymbol{\beta}) \propto e^{(y - \frac{n_i}{2})x} e^{-\frac{x_i^2}{2}\omega_i} \,\mathrm{PG}(\omega_i | n_i, 0) \tag{64}$$

  なので, ω_i の事後分布は

$$\omega_i \,|\, \boldsymbol{\beta} \sim \mathrm{PG}(n_i, x_i^T \boldsymbol{\beta}) \ . \tag{65}$$

Pólya-Gamma Gibbs sampling: β の事後分布

- 同様に $\boldsymbol{\beta}$ については, データ $Y = (y_1, \cdots, y_N)$ 全体に関わるので

$$\prod_{i=1}^{N} p(y_i|x_i, \boldsymbol{\beta}) \propto e^{\sum_{i=1}^{N}(y_i - \frac{n_i}{2})x_i^T \boldsymbol{\beta}} \, e^{\sum_{i=1}^{N} -\frac{(x_i^T \boldsymbol{\beta})^2}{2}\omega_i} \tag{66}$$

- よって, $\kappa_i = y_i - n_i/2$, $z_i = \kappa_i/\omega_i$ とおいて

$$\sum_{i=1}^{N} \log p(y_i|x_i, \boldsymbol{\beta}) \propto \sum_{i=1}^{N} \left[\left(y_i - \frac{n_i}{2} \right) x_i^T \boldsymbol{\beta} - \frac{(x_i^T \boldsymbol{\beta})^2}{2} \omega_i \right] \tag{67}$$

$$= \sum_{i=1}^{N} \left[-\frac{\omega_i}{2} \left((x_i^T \boldsymbol{\beta})^2 - 2\frac{\kappa_i}{\omega_i} x_i^T \boldsymbol{\beta} \right) \right] \tag{68}$$

$$\propto \sum_{i=1}^{N} \left[-\frac{\omega_i}{2} \left(x_i^T \boldsymbol{\beta} - \frac{\kappa_i}{\omega_i} \right)^2 \right] \tag{69}$$

$$\propto -\frac{1}{2}(z - X\boldsymbol{\beta})^T \Omega (z - X\boldsymbol{\beta}) \tag{70}$$

Pólya-Gamma Gibbs sampling: β の事後分布 (2)

$$\log p(\boldsymbol{\beta}|X,\Omega) \propto (z-X\boldsymbol{\beta})^T\Omega(z-X\boldsymbol{\beta}) + \boldsymbol{\beta}^T\Sigma\boldsymbol{\beta} \tag{71}$$

$$= \boldsymbol{\beta}^T(\underbrace{X^T\Omega X+\Sigma}_{V})\boldsymbol{\beta} + 2\boldsymbol{\beta}^T(X^T\Omega z) + z^T\Omega z \tag{72}$$

- 一方,

$$(\boldsymbol{\beta}-m)^T V(\boldsymbol{\beta}-m) = \boldsymbol{\beta}^T V\boldsymbol{\beta} - 2\boldsymbol{\beta}^T(Vm) + m^T Vm \tag{73}$$

と比較して,

$$Vm = X^T\Omega z \tag{74}$$

$$\therefore\ \ m = V^{-1}X^T\Omega z \tag{75}$$

- よって,

$$\boldsymbol{\beta}\,|X,\Omega \sim \mathrm{N}\left(V^{-1}X^T\Omega z, V^{-1}\right). \tag{76}$$

多値ロジスティック回帰

- $\mathbf{x}_i$ から出力カテゴリ j が選ばれる確率は

$$p(j|\mathbf{x}_i) = \frac{e^{\mathbf{x}_i^T\boldsymbol{\beta}_j}}{\sum_{k=1}^K e^{\mathbf{x}_i^T\boldsymbol{\beta}_k}} = \frac{\exp(\mathbf{x}_i^T\boldsymbol{\beta}_j - c_{ij})}{1+\exp(\mathbf{x}_i^T\boldsymbol{\beta}_j - c_{ij})} \tag{77}$$

$$c_{ij} = \log\sum_{k\neq j}\exp(\mathbf{x}_i^T\boldsymbol{\beta}_k)$$

- 上の確率を θ_{ij} とおくと,

$$p(\boldsymbol{\beta}_j|\boldsymbol{\beta}_{-j},\mathbf{y}) \propto \prod_{i=1}^N\prod_{k=1}^K p(j|\mathbf{x}_i)^{\mathbb{I}(y_i=j)}p(\ell\neq j|\mathbf{x}_i)^{\mathbb{I}(y_i\neq j)} \tag{78}$$

$$= \prod_{i=1}^N \theta_{ij}^{\mathbb{I}(y_i=j)}\left(\frac{\sum_{k\neq j}e^{\mathbf{x}_i^T\boldsymbol{\beta}_k}\cdot\frac{e^{\mathbf{x}_i^T\boldsymbol{\beta}_\ell}}{\sum_{k\neq j}e^{\mathbf{x}_i^T\boldsymbol{\beta}_k}}}{\sum_{k=1}^K e^{\mathbf{x}_i^T\boldsymbol{\beta}_k}}\right) \tag{79}$$

$$= \prod_{i=1}^N \theta_{ij}^{\mathbb{I}(y_i=j)}((1-\theta_{ij})w_i)^{\mathbb{I}(y_i\neq j)} \tag{80}$$

多値ロジスティック回帰 (2)

$$p(\boldsymbol{\beta}_j|\boldsymbol{\beta}_{-j},\mathbf{y}) \propto \prod_{i=1}^{N} \theta_{ij}^{\mathbb{I}(y_i=j)}(1-\theta_{ij})^{\mathbb{I}(y_i\neq j)} \tag{81}$$

- したがって尤度項は

$$\prod_{i=1}^{N} e^{\kappa_{ij}\theta_{ij}} e^{-\frac{\theta_{ij}^2}{2}\omega_{ij}} \mathrm{PG}(\omega_{ij}|n_i,0) \tag{82}$$

- これから, 更新式は

$$\begin{cases} \omega_{ij}\,|\,\boldsymbol{\beta}_j & \sim \mathrm{PG}(n_i,\theta_{ij}) \\ \boldsymbol{\beta}_j\,|\,\Omega_j & \sim \mathrm{N}(m_j(\Omega),V_j(\Omega)) \end{cases} . \tag{83}$$

まとめ

- ロジスティック回帰モデルの完全なベイズ推定
  - Pólya-Gamma 分布に従う補助変数 ω の下で, パラメータはガウス分布に従う
- ロジスティック回帰を, 統計モデルのコンポーネントとして使うことができる
  - 教師なし学習でも OK
  - 豊富な共分散構造・連続値モデル
- 多値および Stick-breaking 過程への展開, 様々な modulation