# ロバストダイバージェンスを用いたベイズ推論について

中川 智之 [1]

[1] 東京理科大学理工学部情報科学科

## 1 導入

近年, 計算機の発達に伴い, 大規模で複雑なデータが多くなってきており, 解析の機会も多くなっている. このような大規模で複雑なデータには, しばしば外れ値や異常値が混在したり, 仮定したモデルが真のデータ生成モデルになっていない可能性がある. 仮定したモデルが誤特定している場合は, ノンパラメトリックな手法を用いることが多い. 一方で, 外れ値は入力ミスなどの実際のデータとは大きく異なるデータが含まれることになるため, 最尤法やノンパラメトリックな手法はデータ解析の結果にバイアスが生じ, 誤った推論を与えることになる.

ベイズ統計においても, 通常の事後分布は外れ値に対しては頑健ではないため, 外れ値の問題は古くから研究されてきた. 特に位置尺度母数モデルに関して, $t$ 分布のような裾が正則変動 (Regularly varying) している裾の重いモデルを用いると外れ値に影響されにくくなることがわかっている (O'Hagan and Pericchi, 2012, などを参照). さらに近年では正則変動よりさらに裾の重い対数正則変動 (log-Regularly varying) するモデルに関するベイズ推定も提案され, 頑健性の 1 つである Posterior robustness が示されている (Desgagné, 2015; Gagnon et al., 2020; Hamura et al., 2020, など). ここで Posterior Robustness は, 非常に大きな値のデータに対しては事後分布は影響されないという性質である (詳細は第 3 節で説明する.).

一方で, 裾の重い分布の構成は位置尺度母数モデルよりも一般的なモデルへの拡張は容易ではなく, それぞれのモデルやデータ構造によって構成が必要になる. そこで, 本研究では, 一般のモデルに拡張が容易であるロバストなダイバージェンスを用いたベイズ推論を考える. ロバストなダイバージェンスを用いたベイズ推定は, Hooker and Vidyashankar (2014) や Ghosh and Basu (2016), Nakagawa and Hashimoto (2020) などによって提案されている. これらの方法は, データ $Y=(y_1,\dots,y_n)$ とパラメータ $\theta$ に対するあるロス関数 $D(y;\theta)$ を用いて, 事後分布を

$$\pi(\theta \mid Y) \coloneqq \frac{\pi(\theta)\exp(\sum_{i=1}^n D(y_i;\theta))}{\int_\Theta \pi(\theta)\exp(\sum_{i=1}^n D(y_i;\theta))\mathrm{d}\theta}$$

と与え, ベイズ推論を行う方法であり, この事後分布のことを一般化事後分布 (General posterior distribution) と呼ぶ. 一般化事後分布は, 擬似事後分布 (quasi-posterior distribution, pseudo-posterior distribution) などとも呼ばれ, 古くから研究が行われている. また, 一般化事後分布を用いてベイズ推論を行うことを一般化ベイズ推論 (General Bayesian Inference) と呼ぶ. 近年, 一般化ベイズ推論は, Bissiri et al. (2016) によって, ターゲットのパラメータ

$$\theta^* = \arg\min_{\theta\in\Theta} \int_{\mathcal{Y}} D(y,\theta)\mathrm{d}G(y)$$

に関する合理的かつ妥当なベイズ更新を与えることが示されている．ここで, $G$ はデータ生成分布である．つまり, 一般化ベイズ推論は, ロス関数を最小化にするパラメータ $\theta^*$ に関して, 合理的かつ妥当なベイズ推論が可能である．これにより, ダイバージェンスを最小化する推定量の性質が一般化ベイズ推論を通して, ベイズ推定にも適用できるようになった．また, Hooker and Vidyashankar (2014) や Ghosh and Basu (2016), Nakagawa and Hashimoto (2020) などでは, 影響関数などを用いてロバスト性を調べている．しかしながら, Posterior Robustness のようなベイズ統計におけるロバスト性に関しては, まだ知られていない.

本講演では, ロバストなダイバージェンスに基づいて一般化事後分布が Posterior Robustness を満たす条件について考察する．また, その条件を $\gamma$-divergence に基づく一般化事後分布が満たすことを示す．さらに, シミュレーションを用いて外れ値がそれぞれの事後分布にどの程度影響するのかを調べる.

## 2　ロバストなダイバージェンス

本節では, 本講演で扱う $D(y,\theta)$ を紹介する．$f_\theta(y)$ を仮定した統計モデルとする．以下は, それぞれのダイバージェンスに基づく $D(y,\theta)$ である.

- Kullback-Leibler divergence

$$D_{KL}(y;\theta) = \log f_\theta(y).$$

- Density power divergence (Basu et al. (1998))

$$D_\alpha(y;\theta) = \frac{1}{\alpha} f_\theta(y)^\alpha - \frac{1}{\alpha+1}\int f_\theta(x)^{\alpha+1}\mathrm{d}x \quad (\alpha > 0).$$

- $\gamma$-divergence (Jones et al., 2001; Fujisawa and Eguchi, 2008)

$$D_\gamma(y;\theta) = \frac{1}{\gamma} f_\theta(y)^\gamma \left\{\int f_\theta(x)^{\gamma+1}\mathrm{d}x\right\}^{-\gamma/(\gamma+1)} - \frac{1}{\gamma} \quad (\gamma > 0).$$

- $\alpha$-divergence

$$D_\alpha(y;\theta) = \frac{1}{\alpha(1-\alpha)} g_n(y)^{\alpha-1} f_\theta(y)^{1-\alpha}.$$

  ただし, $g_n(y)$ は $Y$ に対するカーネル密度推定である.

Density power divergence, $\gamma$-divergence, $\alpha$-divergence に関してはそれぞれロバストのダイバージェンスとして知られており, 様々な観点から外れ値に対する頑健性が示されている．しかしながら, $\alpha$-divergence はカーネル密度推定が必要であり, 実用上使うのが非常に難しい．本講演では, Density power divergence, $\gamma$-divergence についてのみ扱うことにする.

## 3　一般化事後分布の頑健性

本節の研究は, 広島大学の橋本真太郎先生, 東京大学の菅澤翔之助先生との共同研究である．まず, Posterior robustness を説明するために, $i$ 番目のデータを $y_i = a_i + \omega b_i$ とする．このとき, $Y_k = (y_1,\dots,y_k)$ が外れ値でないとするならば, $b_1 = \cdots = b_k = 0$ である．つまり, $y_{k+1},\dots,y_n$ が

外れ値である．このとき, posterior robustness は $\omega \to \infty$ の場合に $Y$ を用いた事後分布と $Y_k$ を用いた事後分布が漸近的に一致することである．Desgagné (2015) では, 対数正則変動関数を用いたモデルは Posterior robustness が成り立つことが示されている．本講演では, ロス関数に以下の仮定を考える．

(A1) 次を満たす関数 $h$ が存在する．

$$\lim_{|z|\to\infty} \frac{\exp\{D(z;\theta)\}}{h(z)} = 1 \ (\forall \theta \in \Theta)$$

(A2) 任意の $\theta \in \Theta$ で $\prod_{i=k+1}^{n} \left[\exp\left\{D(y_i;\theta)\right\}/h(y_i)\right] < \infty$ が成り立つ．

このとき, $D(y;\theta)$ が仮定 (A1), (A2) を満たすもとで以下の定理が成り立つ．

**Theorem 1** (Posterior robustness)**.** 仮定 *(A1), (A2)* と正則条件の下で, 以下が成り立つ．

*(a)* $\lim_{\omega\to\infty} \frac{m(Y)}{\prod_{i=k+1}^{n} h(y_i)} = m(Y_k)$.

*(b)* $\lim_{\omega\to\infty} \pi(\theta \mid Y) = \pi(\theta \mid Y_k)$ $(\forall \theta \in \Theta)$.

*(c)* $\lim_{\omega\to\infty} \int_{\Theta} |\pi(\theta \mid Y) - \pi(\theta \mid Y_k)|\, d\theta = 0$.

*(d)* $\theta \mid Y \xrightarrow{d} \theta \mid Y_k$ *as* $\omega \to \infty$.

Theorem 1 は一般化事後分布が Posterior Robustness を満たすことを意味している．実際には, 以下の仮定 (A1), (A2) が Posterior robustness を満たすための重要な条件であり, Desgagné (2015) では, 位置尺度母数に対して, 裾が対数正則変動 (log-Regularly varying) するモデルを用いることで, 仮定 (A1), (A2) を満たすことが示すことができる．また, 仮定したモデルが

$$\lim_{\omega\to\infty} f_\theta(y) = 0 (\forall \theta \in \Theta) \tag{1}$$

を満たすとき, $\gamma$-divergence に基づく一般化事後分布も仮定 (A1), (A2) が成り立つことを示すことができる．しかしながら, Density power divergence に基づく一般化事後分布は条件 (1) が成り立っていても, 仮定 (A1), (A2) は成り立たない場合がある．また条件 (1) は Fujisawa and Eguchi (2008) や Hooker and Vidyashankar (2014) などでロバスト性を示す際に用いられる条件の特別な場合になっており, 比較的妥当な条件と言える．次の節では, これらの詳細について, 様々なモデルついて紹介する．

# 4 応用例

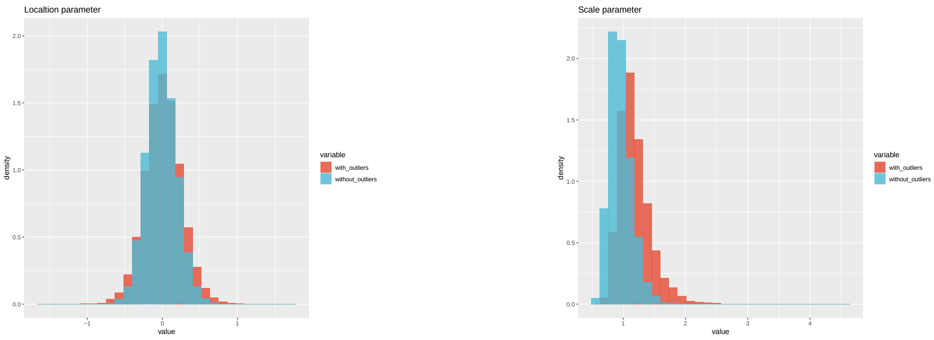
本節では応用例として, 正規分布の場合と順序付き回帰の場合についての数値実験を紹介する．

## 4.1 応用: 正規分布

本節では, 正規分布 $\mathcal{N}(\mu,\sigma^2)$ について考える．図 1 と 2 は $(1-\varepsilon)\mathcal{N}(0,1) + \varepsilon\mathcal{N}(6,1)$ から 200 個サンプリングしたデータに対して, それぞれの平均 $\mu$ と標準偏差 $\sigma$ の事後分布のヒストグラムを描いたものである．図 1 と図 2 は, それぞれ正規分布 $\mathcal{N}(\mu,\sigma^2)$ モデルに対する Density power divergence と $\gamma$-divergence に基づく一般化事後分布である．青色のヒストグラムが外れ値が含まれ

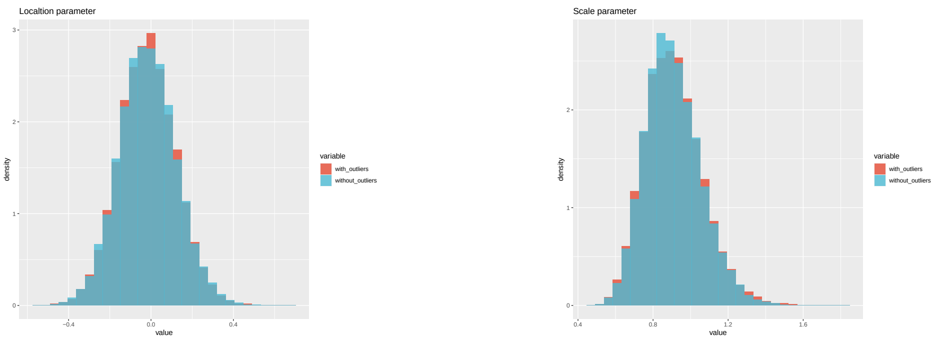
ていない ($\varepsilon = 0$) ときの事後分布であり, 赤色のヒストグラムが外れ値が含まれている ($\varepsilon = 0.20$) ときの事後分布である.

図 1 や図 2 のダイバージェンスを用いた事後分布は外れ値の影響をほとんど受けていないことがわかる. 特に $\gamma$-divergence を用いた方法は $\sigma$ に関しても全く外れ値の影響を受けていない.




$\mu$ に関する事後分布　　　　$\sigma$ に関する事後分布

図 1: DPD を用いた一般化事後分布




$\mu$ に関する事後分布　　　　$\sigma$ に関する事後分布

図 2: $\gamma$-divergence を用いた一般化事後分布

### 4.2 応用: 順序付き回帰

2 つ目は順序付き回帰モデルについて考える. なお, この研究は東京理科大学大学院の桃﨑智隆氏との共同研究である. 本研究では, カテゴリ数 $M$ の順序付きカテゴリカルデータ $y_i (i = 1, \ldots, n)$ を目的変数とし, 説明変数 $\boldsymbol{x}_i$ での回帰を以下のように考える. 潜在変数 $z_i (i = 1, \ldots, n)$ を仮定し, $y_i = m \Leftrightarrow \delta_{m-1} < z_i \leq \delta_m (m = 1, \ldots, M)$ が成り立つとし,

$$z_i = \boldsymbol{x}_i^\top \boldsymbol{\beta} + \varepsilon_i$$

となるような回帰を考える. このとき, 係数パラメータ $\boldsymbol{\beta}$ とカットオフパラメータ $-\infty = \delta_0 < \delta_1 < \delta_2 < \cdots < \delta_{M-1} < \delta = \infty$ のベイズ推定を行う. このモデルに対して, Huber 型の推定量が Iannario et al. (2017) や Scalera et al. (2021) において用いられており, さらにダイバージェンス最小化推定量は Pyne et al. (2022) や Momozaki and Nakagawa (2022) によって性質が調べられているが, ベイズ推定に関してはロバストなものはない. 本研究では, Momozaki and Nakagawa (2022) をもとに一般化事後分布を構成し, Posterior robustness を導出した. また, Desgagné (2015);

Gagnon et al. (2020); Hamura et al. (2020) などのように誤差分布 $\varepsilon_i$ に裾の重い分布を仮定しても, Posterior robustness が成り立たないことも示した.

図 3 と図 4 は, $M = 5$ とし, データを $z_i = x_{i1}\beta_1 + x_{i2}\beta_2 + \varepsilon_i$ から 200 個サンプリングし, 事後分布を描いたものである. ここで, 今日変量はそれぞれ $x_{i1}$ を $(1-\rho)\mathcal{N}(0,1) + \rho\mathcal{N}(20,1)$, $x_{i2}$ を $\mathcal{N}(0,1)$ から生成したものとし, 誤差分布 $\varepsilon_i \sim \mathcal{N}(0,1)$ とする. 図 3 と図 4 では, 緑色が外れ値が含まれていない $(\rho = 0)$ ときの事後分布であり, ピンク色が外れ値が含まれている $(\rho = 0.05)$ ときの事後分布である.

図 3 と図 4 からわかる通り, それぞれのパラメータに関して, 通常の事後分布は影響受けているのに対し, ダイバージェンスを用いた方法は外れ値の影響を受けていないことがわかる.

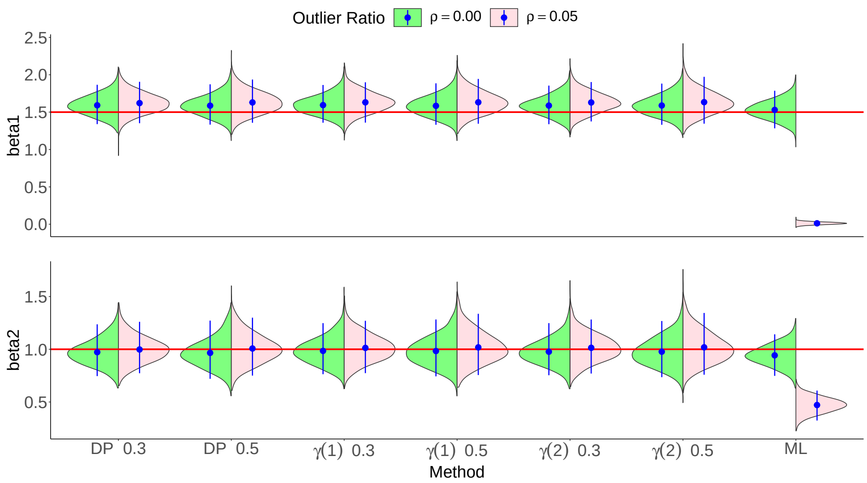



図 3: 回帰係数 $\boldsymbol{\beta}$ に関する事後分布

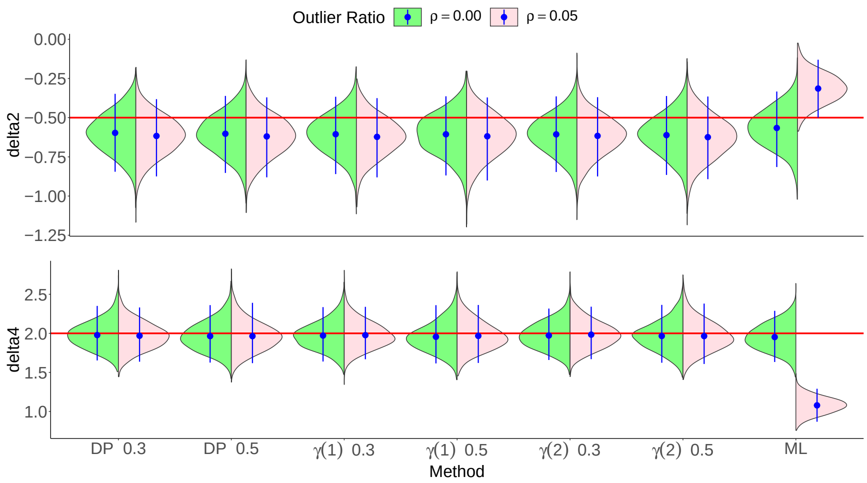



図 4: カットオフポイント $\boldsymbol{\delta}$ に関する事後分布

## 参考文献


Basu, A., Harris, I. R., Hjort, N. L., and Jones, M. (1998). Robust and efficient estimation by minimising a density power divergence. *Biometrika*, 85(3):549–559.

Bissiri, P. G., Holmes, C. C., and Walker, S. G. (2016). A general framework for updating belief distributions. *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, 78(5):1103.

Desgagné, A. (2015). Robustness to outliers in location–scale parameter model using log-regularly varying distributions. *The Annals of Statistics*, 43(4):1568–1595.

Fujisawa, H. and Eguchi, S. (2008). Robust parameter estimation with a small bias against heavy contamination. *Journal of Multivariate Analysis*, 99(9):2053–2081.

Gagnon, P., Desgagné, A., Bédard, M., et al. (2020). A new bayesian approach to robustness against outliers in linear regression. *Bayesian Analysis*, 15(2):389–414.

Ghosh, A. and Basu, A. (2016). Robust bayes estimation using the density power divergence. *Annals of the Institute of Statistical Mathematics*, 68(2):413–437.

Hamura, Y., Irie, K., and Sugasawa, S. (2020). Log-regularly varying scale mixture of normals for robust regression. *arXiv preprint arXiv:2005.02800*.

Hooker, G. and Vidyashankar, A. N. (2014). Bayesian model robustness via disparities. *Test*, 23(3):556–584.

Iannario, M., Monti, A. C., Piccolo, D., and Ronchetti, E. (2017). Robust inference for ordinal response models. *Electronic Journal of Statistics*, 11(2):3407–3445.

Jones, M., Hjort, N. L., Harris, I. R., and Basu, A. (2001). A comparison of related density-based minimum divergence estimators. *Biometrika*, 88(3):865–873.

Momozaki, T. and Nakagawa, T. (2022). Robustness against outliers in ordinal response model via divergence approach. *arXiv preprint arXiv:2209.11965*.

Nakagawa, T. and Hashimoto, S. (2020). Robust bayesian inference via $\gamma$-divergence. *Communications in Statistics-Theory and Methods*, 49(2):343–360.

O'Hagan, A. and Pericchi, L. (2012). Bayesian heavy-tailed models and conflict resolution: A review. *Brazilian Journal of Probability and Statistics*, 26(4):372–401.

Pyne, A., Roy, S., Ghosh, A., and Basu, A. (2022). Robust and efficient estimation in ordinal response models using the density power divergence. *arXiv preprint arXiv:2208.14011*.

Scalera, V., Iannario, M., and Monti, A. C. (2021). Robust link functions. *Statistics*, 55(4):963–977.