数理解析研究所講究録
第2047巻 2017年 55-66

Comparison of two robust Bayes estimations using the divergence under heavy contamination

中川 智之, 橋本 真太郎
広島大学大学院・理学研究科数学専攻 *
Tomoyuki Nakagawa and Shintaro Hashimoto
Department of Mathematics, Graduate School of Science
Hiroshima University

1 導入

本稿は外れ値に頑健なベイズ推定を行う. 外れ値を含むデータの場合, 最尤推定量や最小二乗推定量などはバイアスが大きくなる. また, 事後期待値や最大事後確率 (MAP) 推定量などのベイズ推定も外れ値の影響を受けてしまう. また近年, 膨大かつ高次元なデータが多くなり, 外れ値を除くことが困難になっている. そのため, 外れ値の影響を受けにくい推定方法が必要になってくる.

外れ値に頑健な推定方法は, これまでも多くの研究がなされている (例えば, [7], [8] を参照). 最近では, [2] や [9] で density-power divergence に基づく外れ値に頑健な推定方法が提案されている. density-power divergence は外れ値に対して頑健であるが, 外れ値の割合が多くなると影響を受けてバイアスが大きくなってしまう. また, 尺度母数に対する推定が悪いことも数値的に知られている. その改善策として, [5] では外れ値の割合に依らずに頑健な推定が可能な γ-divergence を提案している. [5] では, 外れ値の割合に仮定をするのではなく, 外れ値を発生させる汚染分布に仮定を入れている.

また, ベイズ推定に関しても外れ値の問題は多く研究されている. [3] では位置母数の推定の際に実際のモデルに正規分布ではなく裾の重い t-分布を使うことで, 推定している.

* 〒739-8526 広島県東広島市鏡山 1 丁目 3 番 1 号

さらに [1] では事前分布にも裾の重い分布を仮定して, 位置母数だけでなく尺度母数の推定も行なっている. しかし, 尺度母数に対しては推定がうまくいっていない. [4] では, 対数正則変動関数を用いた t-分布などよりもさらに裾の重い分布を用いることで, 尺度母数の精度を上げている. しかし, これらのロバストなベイズ推定は 1 変量の場合のみである. そこで [6] では, 汚染分布を用いた外れ値の定義から density-power divergence を用いて, 頑健な事後分布を構成する方法を提案している. また, [10] では, γ-divergence を用いて, [6] と同様に頑健な事後分布を構成している. これらのように divergence を用いることで, 多変量への拡張も容易である.

本稿では, 通常の事後分布と [6] で提案された擬似的な事後分布 (DP-posterior), [10] で提案された擬似的事後分布 (γ-posterior) の比較を数値実験を用いて比較する. 第 2 節では, divergence を用いた事後分布の構成について行う. 第 3 節では, [10] で示されている γ-posterior における漸近性質を記述する. 第 4 節で, 数値実験を用いてそれぞれの事後分布, 事後平均を比較する.

2 外れ値に頑健な事後分布の構成

ベイズ推定において, 事後分布は重要な役割を持っている. この章では, 通常の事後分布の形から divergence を用いた事後分布の構成方法を紹介し, [6] と [10] で提案された外れ値に頑健な事後分布 DP-posterior と γ-posterior の定義を述べる.

まず, divergence による推定について紹介する. f, g を密度関数とする. このとき, divergence $\mathcal{D}$ は cross entropy d を用いて次のように表せる.

$$\mathcal{D}(g,f) = -d(g,g) + d(g,f)$$

また, divergence と cross entropy は次の性質を満たす.

(a) $\mathcal{D}(g,f) \geq 0$ で, 等号成立は $g = f$ のときに限る.
(b) $d(g,f) \geq d(g,g)$ で, 等号成立は $g = f$ のときに限る.

そのため、divergence は g からの f への擬似的な距離と考えることができる. ここで, g を真の分布とすれば, divergence を最小にするような f を選べば良いことがわかる. しかし, g は実際には未知であるから, g をデータの経験分布 $\bar{g}$ で置き換え, 候補モデルの中で cross entropy を最小にするように推定する. 特に, 本稿で取り扱う divergence は以下の 3 つである. ここでは, 各 divergence に対応する cross entropy を示す.

- K-L divergence

$$d_{KL}(g,f) = -\int g \log f d\boldsymbol{x}$$

- density power divergence

$$d_\beta(g,f) = -\int g f^\beta d\boldsymbol{x} + \int f^{1+\beta} d\boldsymbol{x}$$

- γ-divergence

$$d_\gamma(g,f) = -\log \int g f^\gamma d\boldsymbol{x} + \log \int f^{1+\gamma} d\boldsymbol{x}$$

ただし, $\beta, \gamma > 0$ である. density-power divergence は [2], γ-divergence は [5] でそれぞれ外れ値に頑健な divergence として提案されている.

次に, ベイズ推論について議論を行う. データ $\boldsymbol{x}_1, \ldots \boldsymbol{x}_n$ が密度関数 g の分布に独立同一に従うとし, f を知りたい分布とする. 通常のベイズ推論は候補モデルを $f_{\boldsymbol{\theta}}(\boldsymbol{\theta} \in \Theta)$ に対して, 次の事後分布を考える.

$$\begin{aligned}\pi(\boldsymbol{\theta}|\boldsymbol{X}_n) &\propto \prod_{i=1}^n f_{\boldsymbol{\theta}}(\boldsymbol{x}_i)\pi(\boldsymbol{\theta}) \\ &= \exp\{\log f_{\boldsymbol{\theta}}(\boldsymbol{x}_i)\}\pi(\boldsymbol{\theta}) \\ &= \exp\{-d_{KL}(\bar{g}, f_{\boldsymbol{\theta}})\}\pi(\boldsymbol{\theta}).\end{aligned}$$

ここで, $\bar{g}$ は $\boldsymbol{X}_n = (\boldsymbol{x}_1, \ldots \boldsymbol{x}_n)$ の経験分布である. この形から, 事前分布が一様分布の場合の MAP 推定量と divergence 最小化による推定量が同等であることがすぐにわかる. 通常の事後分布は $g \in \{f_{\boldsymbol{\theta}}|\boldsymbol{\theta} \in \Theta\}$ を仮定しているが, 今データの分布は f に従うデータに汚染分布 δ に従うデータが混ざっている. すなわち $g = (1-\varepsilon)f + \varepsilon\delta$ のような分布に従う. そのため, 通常のベイズ推定では大きなバイアスを持ってしまう. そこで, [6] では density-power divergecnce を用いて, 擬似的事後分布を構成している. 通常の事後分布の $d_{KL}(\cdot,\cdot)$ の部分を $d_\beta(\cdot,\cdot)$ に変えることで, 次のような擬似的な事後分布を構成している.

定義 2.1 (DP-posterior). $\beta > 0$ とする. $\pi(\boldsymbol{\theta})$ を事前分布, $f_{\boldsymbol{\theta}}$ を候補のモデルとしたと

き, *DP-posterior* を次で定義する.

$$\begin{aligned}\pi^{(\beta)}(\boldsymbol{\theta}|\boldsymbol{X}_n) &= \frac{\exp\{Q_n^{(\beta)}(\boldsymbol{\theta})\}\pi(\boldsymbol{\theta})}{\int \exp\{Q_n^{(\beta)}(\boldsymbol{\theta})\}\pi(\boldsymbol{\theta})d\boldsymbol{\theta}} \\ &= \frac{\prod_{i=1}^n \exp(q_{\boldsymbol{\theta}}^{(\beta)}(\boldsymbol{x}_i))\pi(\boldsymbol{\theta})}{\int \prod_{i=1}^n \exp(q_{\boldsymbol{\theta}}^{(\beta)}(\boldsymbol{x}_i))\pi(\boldsymbol{\theta})d\boldsymbol{\theta}} \\ &\propto \prod_{i=1}^n \exp(q_{\boldsymbol{\theta}}^{(\beta)}(\boldsymbol{x}_i))\pi(\boldsymbol{\theta}),\end{aligned} \tag{1}$$

ただし,

$$q_{\boldsymbol{\theta}}^{(\beta)}(x) = \frac{1}{\beta} f_{\boldsymbol{\theta}}^{\beta}(x) - \frac{1}{1+\beta}\int f_{\boldsymbol{\theta}}^{1+\beta} d\boldsymbol{x}.$$

[6] では DP-posterior に関する漸近性質や頑健性が述べられている. そこで, [10] では denstiy-power divergence に変えて, 次のような γ-divergence を用いた擬似的な事後分布の構成を行い, その性質を導出している.

定義 2.2 (γ-posterior)**.** $\gamma > 0$ とする. $\pi(\boldsymbol{\theta})$ を事前分布, $f_{\boldsymbol{\theta}}$ を候補のモデルとしたとき, *γ-posterior* を次で定義する.

$$\begin{aligned}\pi^{(\gamma)}(\boldsymbol{\theta}|\boldsymbol{X}_n) &= \frac{\exp\{Q_n^{(\gamma)}(\boldsymbol{\theta})\}\pi(\boldsymbol{\theta})}{\int \exp\{Q_n^{(\gamma)}(\boldsymbol{\theta})\}\pi(\boldsymbol{\theta})d\boldsymbol{\theta}} \\ &= \frac{\prod_{i=1}^n \exp(q_{\boldsymbol{\theta}}^{(\gamma)}(\boldsymbol{x}_i))\pi(\boldsymbol{\theta})}{\int \prod_{i=1}^n \exp(q_{\boldsymbol{\theta}}^{(\gamma)}(\boldsymbol{x}_i))\pi(\boldsymbol{\theta})d\boldsymbol{\theta}} \\ &\propto \prod_{i=1}^n \exp(q_{\boldsymbol{\theta}}^{(\gamma)}(\boldsymbol{x}_i))\pi(\boldsymbol{\theta}),\end{aligned} \tag{2}$$

ただし,

$$q_{\boldsymbol{\theta}}^{(\gamma)}(\boldsymbol{x}) = \frac{1}{\gamma} f_{\boldsymbol{\theta}}^{\gamma}(\boldsymbol{x}) \left\{\int f_{\boldsymbol{\theta}}^{1+\gamma} d\boldsymbol{x}\right\}^{-\gamma/(1+\gamma)} - \frac{1}{\gamma}.$$

3 γ-posterior に基づくベイズ推定量の漸近性質

この節では [10] で示されている γ-posterior に関する漸近性質の結果を紹介する. まず, 以下の条件を考える. $\boldsymbol{\theta}_g$ を $d_\gamma(g, f_{\boldsymbol{\theta}})$ を最小にする点とする.

(A1) $f_{\boldsymbol{\theta}}$ の台は未知パラメータに依らない. $f_{\boldsymbol{\theta}}$ は $\boldsymbol{\theta}_g$ の近傍 U で 3 回微分可能である.

(A2) 微分と積分の交換は可能とし, $E_g[\partial_i q^{(\gamma)}(\boldsymbol{\theta}_g; X_1)]$ と $E_g[\partial_i\partial_j q^{(\gamma)}(\boldsymbol{\theta}_g; X_1)]$ は存在し, さらにすべての $i, j, k = 1, \ldots, p$ に対して, 次を満たす $M_{ijk}(x)$ が存在する.

$$\sup_{\boldsymbol{\theta}\in U}\left|\partial_i\partial_j\partial_k q^{(\gamma)}(\boldsymbol{\theta}; \boldsymbol{x})\right| \leq M_{ijk}(\boldsymbol{x}) \text{ and } E_g\left[M_{ijk}(X_1)\right] < \infty,$$

ここで, $\partial_i = \partial/\partial\theta_i$, $\partial = \partial/\partial\boldsymbol{\theta}$ である.

(A3) 任意の $\delta > 0$ に対して, 十分大きな n すべてに対して, 確率 1 で次を満たす $\varepsilon > 0$ が存在する.

$$\sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_g\|>\delta} n^{-1}\left\{Q_n^{(\gamma)}(\boldsymbol{\theta}) - Q_n^{(\gamma)}(\boldsymbol{\theta}_g)\right\} < -\varepsilon.$$

注意 3.1. *(A1), (A2)* の条件はモデルの正則条件である. また, *(A3)* の条件は n が十分大きいとき, $Q_n(\boldsymbol{\theta})$ の最大化点がただ一つ $\boldsymbol{\theta}_g$ であるという条件である. これは $\boldsymbol{\theta}_g$ の定義から自然である.

次に, $Q_n^{(\gamma)}(\boldsymbol{\theta})$ を最大化する推定量を $\tilde{\boldsymbol{\theta}}_n^{(\gamma)}$, 事後期待値を $\hat{\boldsymbol{\theta}}_n^{(\gamma)}$ とする. また, 対称行列 $I(\boldsymbol{\theta})$ と $J(\boldsymbol{\theta})$ を次で定義する.

$$\begin{aligned} I(\boldsymbol{\theta}) &= E_g\left[\partial q^{(\gamma)}(\boldsymbol{\theta}; X_1)\partial^\top q^{(\gamma)}(\boldsymbol{\theta}; X_1)\right], \\ J(\boldsymbol{\theta}) &= -E_g\left[\partial\partial^\top q^{(\gamma)}(\boldsymbol{\theta}; X_1)\right]. \end{aligned}$$

さらに, $I(\boldsymbol{\theta})$ と $J(\boldsymbol{\theta})$ は正定値行列とする. このとき, 次の定理が成り立つ.

定理 3.1. 仮定 *(A1)-(A3)* と $\tilde{\boldsymbol{\theta}}_n$ は $\partial Q_n^{(\gamma)}(\tilde{\boldsymbol{\theta}}_n^{(\gamma)}) = \mathbf{0}$ と $\tilde{\boldsymbol{\theta}}_n^{(\gamma)} \xrightarrow{p} \boldsymbol{\theta}_g (n \to \infty)$ を満たすとする. このとき, 任意の事前分布の密度関数 $\pi(\theta)$ が連続で $\boldsymbol{\theta}_g$ で正ならば, 次が成り立つ.

$$\int\left|\pi^{*(\gamma)}(\boldsymbol{t}|\boldsymbol{X}_n) - (2\pi)^{-p/2}\left|J(\boldsymbol{\theta}_0)\right|^{1/2}\exp\left(-\frac{1}{2}\boldsymbol{t}^\top J(\boldsymbol{\theta}_0)\boldsymbol{t}\right)\right|d\boldsymbol{t} \xrightarrow{p} 0 \quad (n \to \infty)。 \tag{3}$$

ここで, $\pi^{*(\gamma)}(\boldsymbol{t}|\boldsymbol{X}_n)$ は $\boldsymbol{t} = \sqrt{n}(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}_n)$ の γ*-posterior* である.

定理 3.2. 定理 *3.1* の仮定に加えて, 事前分布 $\pi(\boldsymbol{\theta})$ が期待値を持つとする. このとき, $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \tilde{\boldsymbol{\theta}}_n) \xrightarrow{p} \mathbf{0}(n \to \infty)$ が成り立つ.

系 3.1. 定理 *3.2* の仮定を満たすとする. もし

$$\sqrt{n}(\tilde{\boldsymbol{\theta}}_n^{(\gamma)} - \boldsymbol{\theta}_g) \xrightarrow{d} N_p(\mathbf{0}, V(\boldsymbol{\theta}_0)),$$

ならば, 次が成り立つ.

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n^{(\gamma)} - \boldsymbol{\theta}_g) \xrightarrow{d} N_p(\mathbf{0}, V(\boldsymbol{\theta}_0)).$$

以上のことから, γ-posterior は n が大きいところでは事前分布の影響を受けない. また, γ-posterior に関する事後平均は, 漸近的に γ-divergence 最小化の推定量と一致し, さらに漸近正規性を持つこともわかる.

4 数値実験

この章では, 第 2 章で紹介した DP-posterior と γ-posterior の数値比較を行う. 基本的な設定としては, 1000 回のモンテカルロシミュレーションを用いて, 標本数 $n = 20, 50, 100$, 汚染の割合 $\varepsilon = 0.00, 0.05, 0.20$ の場合について, 事後平均のバイアスを見ていく. また, 事後平均は重点サンプリングで導出する.

4.1 正規分布

この節では, 正規分布:$N(\mu, \sigma^2)$ の場合を考える. データ発生分布を

$$(1-\varepsilon)N(0,1) + \varepsilon N(6,1)$$

とし, 候補モデルを $N(\mu, \sigma^2)$ とする. 事前分布としては 一様分布と Jeffreys の事前分布, 正規分布の共役事前分布である正規-逆ガンマ分布を用いる.

すべての表から通常の事後平均よりも DP-posterior と γ-posterior はバイアスが小さいことがわかる. また, 表 1 と表 3, 表 5 からわかるように DP-posterior と γ-posterior は平均の推定に関しては, ほとんど変わらず, ε が 0.20 と大きい時に少し γ-posterior が良くなっている. 一方で, 分散の推定に対しては, 表 2 と表 4 は, γ-posterior が DP-posterior より良くなっていることがわかる. しかし, 表 6 の場合, DP-posterior の方が良くなっている. これは事前分布の影響が大きいと考えられる.

表 1　The empirical bias of the Bayes estimators for mean parameter under uniform prior

		Bayes	DP-posterior				γ-posterior			
		β,γ	β				γ			
ε	n	0.00	0.30	0.50	0.70	1.00	0.30	0.50	0.70	1.00
0.00	20	0.001	0.001	0.004	0.005	0.005	0.001	0.002	0.003	0.004
0.00	50	0.001	0.001	0.001	0.001	0.002	0.001	0.001	0.001	0.001
0.00	100	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.002
0.05	20	0.302	0.094	0.076	0.059	0.028	0.083	0.059	0.053	0.046
0.05	50	0.300	0.026	0.015	0.026	0.039	0.022	0.008	0.006	0.007
0.05	100	0.302	0.013	0.005	0.004	0.013	0.012	0.004	0.002	0.002
0.20	20	1.191	0.724	0.490	0.328	0.166	0.730	0.507	0.390	0.292
0.20	50	1.194	0.678	0.413	0.342	0.242	0.643	0.280	0.164	0.131
0.20	100	1.202	0.614	0.209	0.154	0.218	0.574	0.091	0.021	0.011

表 2　The empirical bias of the Bayes estimators for variance parameter under uniform prior

		Baye	DP-posterior				γ-posterior			
		β,γ	β				γ			
ε	n	0.00	0.30	0.50	0.70	1.00	0.30	0.50	0.70	1.00
0.00	20	0.266	1.249	4.444	9.155	12.755	0.955	2.187	4.175	7.280
0.00	50	0.090	0.259	0.515	1.447	7.275	0.230	0.361	0.547	1.020
0.00	100	0.041	0.111	0.192	0.336	1.292	0.101	0.150	0.208	0.319
0.05	20	2.441	2.766	6.143	10.225	13.075	2.126	3.367	5.338	8.173
0.05	50	1.943	0.523	0.821	2.366	8.700	0.425	0.464	0.668	1.275
0.05	100	1.833	0.233	0.288	0.489	2.125	0.189	0.189	0.236	0.354
0.20	20	7.504	9.700	11.373	12.993	13.934	8.793	8.933	9.702	11.014
0.20	50	6.333	6.017	5.944	8.379	12.484	5.358	3.285	2.844	3.682
0.20	100	6.054	4.631	2.423	3.003	7.895	4.180	0.957	0.496	0.596

表 3 The empirical bias of the Bayes estimators for mean parameter under the Jeffreys prior

		Bayes	DP-posterior				γ-posterior			
		β,γ	β				γ			
ε	n	0.00	0.30	0.50	0.70	1.00	0.30	0.50	0.70	1.00
0.00	20	-0.004	-0.004	-0.004	-0.005	-0.008	-0.004	-0.004	-0.004	-0.004
0.00	50	-0.001	-0.001	-0.001	-0.002	-0.002	-0.001	-0.001	-0.002	-0.002
0.00	100	0.000	0.000	0.001	0.001	0.001	0.000	0.001	0.001	0.001
0.05	20	0.293	0.046	0.029	0.028	0.017	0.039	0.017	0.012	0.009
0.05	50	0.298	0.018	0.007	0.007	0.015	0.016	0.004	0.002	0.001
0.05	100	0.299	0.011	0.003	0.002	0.004	0.010	0.003	0.001	0.001
0.20	20	1.189	0.614	0.350	0.238	0.139	0.599	0.297	0.180	0.119
0.20	50	1.197	0.578	0.241	0.170	0.153	0.542	0.140	0.051	0.029
0.20	100	1.205	0.549	0.121	0.064	0.090	0.507	0.048	0.010	0.005

表 4 The empirical bias of the Bayes estimators for variance parameter under the Jeffreys prior

		Bayes	DP-posterior				γ-posterior			
		β,γ	β				γ			
ε	n	0.00	0.30	0.50	0.70	1.00	0.30	0.50	0.70	1.00
0.00	20	0.116	0.410	1.051	2.500	4.971	0.334	0.556	0.871	1.434
0.00	50	0.045	0.123	0.218	0.440	1.811	0.107	0.150	0.195	0.276
0.00	100	0.018	0.053	0.089	0.149	0.389	0.046	0.063	0.080	0.105
0.05	20	2.031	1.002	1.716	3.265	5.520	0.792	0.883	1.200	1.759
0.05	50	1.811	0.277	0.341	0.677	2.478	0.219	0.188	0.225	0.312
0.05	100	1.757	0.148	0.159	0.246	0.631	0.113	0.084	0.095	0.121
0.20	20	6.535	5.775	5.246	5.948	6.936	5.221	3.543	3.027	3.010
0.20	50	6.041	4.392	2.756	3.219	5.554	3.940	1.363	0.782	0.761
0.20	100	5.910	3.846	1.294	1.220	2.799	3.452	0.440	0.180	0.177

表 5 The empirical bias of the Bayes estimators for mean parameter under the normal and inverse-gamma prior

		Bayes	DP-posterior				γ-posterior			
		β,γ	β				γ			
ε	n	0.00	0.30	0.50	0.70	1.00	0.30	0.50	0.70	1.00
0.00	20	0.006	0.006	0.005	0.004	0.003	0.006	0.005	0.003	0.004
0.00	50	0.000	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	0.000
0.00	100	0.000	0.000	0.001	0.001	0.001	0.000	0.001	0.001	0.001
0.05	20	0.293	0.046	0.029	0.028	0.017	0.039	0.017	0.012	0.009
0.05	50	0.298	0.018	0.007	0.007	0.015	0.016	0.004	0.002	0.001
0.05	100	0.299	0.011	0.003	0.002	0.004	0.010	0.003	0.001	0.001
0.20	20	0.912	0.080	0.004	0.001	0.001	0.073	0.002	0.001	0.001
0.20	50	1.060	0.202	0.016	0.004	0.001	0.186	0.008	0.001	-0.001
0.20	100	1.127	0.294	0.014	0.003	0.000	0.264	0.005	0.000	-0.001

表 6 The empirical bias of the Bayes estimators for variance parameter under the normal and inverse-gamma prior

		Bayes	DP-posterior				γ-posterior			
		β,γ	β				γ			
ε	n	0.00	0.30	0.50	0.70	1.00	0.30	0.50	0.70	1.00
0.00	20	-0.135	-0.310	-0.486	-0.621	-0.707	-0.313	-0.532	-0.726	-0.827
0.00	50	-0.058	-0.134	-0.215	-0.302	-0.410	-0.134	-0.222	-0.345	-0.579
0.00	100	-0.028	-0.067	-0.108	-0.156	-0.226	-0.066	-0.110	-0.169	-0.288
0.05	20	2.031	1.002	1.716	3.265	5.520	0.792	0.883	1.200	1.759
0.05	50	1.811	0.277	0.341	0.677	2.478	0.219	0.188	0.225	0.312
0.05	100	1.757	0.148	0.159	0.246	0.631	0.113	0.084	0.095	0.121
0.20	20	4.901	0.244	-0.450	-0.594	-0.672	0.122	-0.621	-0.780	-0.847
0.20	50	5.384	1.283	-0.007	-0.143	-0.251	1.087	-0.218	-0.422	-0.673
0.20	100	5.590	1.959	0.138	0.065	0.030	1.667	-0.094	-0.204	-0.365

4.2 指数分布

次に, 以下の指数分布 $Ex(\lambda)$ の場合を考える.

$$f(x;\lambda) = \lambda e^{-\lambda x}.$$

データ発生分布を

$$(1-\varepsilon)Ex(1) + \varepsilon Ex(0.1)$$

とし, 候補モデルを $Ex(\lambda)$ とする. 事前分布としては 一様分布と Jeffreys の事前分布を用いる.

表 7 から ε が大きくなるにつれて, λ の通常の事後平均はバイアスが大きくなっていることがわかる. また, DP-posterior は全体的に安定していることがわかる. 一方で, γ-posteior は n が小さいときにバイアスが大きくなっている. n が十分大きければ, 改善されており, 特に $\varepsilon = 0.20$ では DP-posterior よりも良くなっている. 表 8 では, 通常の事後平均や DP-posterior の結果対しては一様分布のときとほぼ同様の値を取っているが, γ-posterior に関してはバイアス小さくなっていることがわかる.

5 まとめと今後の課題

DP-posterior と γ-posterior の数値的な比較を行った. DP-posterior よりも γ-posterior 方が全体的にパフォーマンスは良かったが, 事前分布の選び方に依ることも確認できた. 今後の課題としては, γ-posterior の外れ値における影響を調べるとともにロバスト性についての理論的な保証を与えていく. また, γ-posterior を回帰分析や判別分析などに応用していく.

謝辞

本研究を行うに際して, 若木 宏文教授 (広島大・理) に貴重な御意見を頂いたことを感謝申し上げます.

表 7 The empirical bias of the Bayes estimators for λ under uniform prior

		Bayes	DP-posterior				γ-posterior			
		β,γ	β				γ			
ε	n	0.00	0.30	0.50	0.70	1.00	0.30	0.50	0.70	1.00
0.00	20	0.104	0.169	0.214	0.230	0.212	0.188	0.322	0.645	1.565
0.00	50	0.041	0.066	0.085	0.094	0.091	0.072	0.113	0.176	0.347
0.00	100	0.020	0.034	0.042	0.046	0.044	0.369	0.054	0.078	0.126
0.05	20	-0.146	0.103	0.163	0.182	0.161	0.133	0.316	0.706	1.700
0.05	50	-0.231	0.010	0.040	0.049	0.042	0.025	0.089	0.160	0.381
0.05	100	-0.272	-0.027	-0.009	-0.006	-0.012	-0.016	0.020	0.050	0.104
0.20	20	-0.534	-0.087	0.009	0.027	0.001	-0.026	0.268	0.816	1.967
0.20	50	-0.602	-0.196	-0.122	-0.109	-0.120	-0.158	-0.013	0.101	0.376
0.20	100	-0.621	-0.221	-0.158	-0.147	-0.155	-0.189	-0.073	-0.004	0.088

表 8 The empirical bias of the Bayes estimators for λ under Jeffreys prior

		Bayes	DP-posterior				γ-posterior			
		β,γ	β				γ			
ε	n	0.00	0.30	0.50	0.70	1.00	0.30	0.50	0.70	1.00
0.00	20	0.055	0.068	0.073	0.055	-0.050	0.081	0.137	0.244	0.598
0.00	50	0.016	0.020	0.023	0.018	-0.001	0.025	0.042	0.068	0.133
0.00	100	0.011	0.011	0.012	0.011	0.002	0.013	0.021	0.033	0.060
0.05	20	-0.183	0.007	0.025	0.007	-0.107	0.030	0.125	0.273	0.669
0.05	50	-0.250	-0.030	-0.018	-0.024	-0.048	-0.017	0.019	0.054	0.138
0.05	100	-0.279	-0.039	-0.026	-0.028	-0.041	-0.028	0.000	0.020	0.051
0.20	20	-0.546	-0.198	-0.141	-0.164	-0.313	-0.150	0.024	0.231	0.683
0.20	50	-0.614	-0.235	-0.183	-0.185	-0.221	-0.200	-0.091	-0.023	0.100
0.20	100	-0.627	-0.249	-0.193	-0.189	-0.206	-0.217	-0.114	-0.060	0.001

参考文献

[1] Andrade, J. A. A. and O'Hagan, A. (2006). Bayesian robustness modelling regularly varying distributions. *Bayesian Anal.*, **1** 169–188.

[2] Basu, A., Harris, I., Hjort, N. and Jones, M.C. (1998). Robust and efficient estimation by minimising a density power divergence. *Biometrika*, **85**, 549–559.

[3] Dawid, A. P. (1973). Posterior expectations for large observations. *Biometrika*, **60**, 664–667.

[4] Desgagné, A. (2015). Robustness to outliers in location-scale parameter model using log-regularly varying distributions. *Ann. Statist.*, **43**, 1568–1595.

[5] Fujisawa, H and Eguchi, S. (2008). Robust parameter estimation with a small bias against heavy contamination. *J. Multivariate Anal.*, **99**, 2053–2081.

[6] Ghosh, A. and Basu, A. (2016). Robust Bayes estimation using the density power divergence. *Ann. Inst. Statist. Math.*, **68**, 413–437.

[7] Hampel, F. R., Rousseeuw, P. J., Ronchetti, E. M. and Stahel, W. A. (1986). *Robust Statistics. The Approach Based on Influence Functions.* Wiley, New York.

[8] Huber, P. J. (1986). *Robust Statistics.* Wiley, New York.

[9] Jones, M. C., Hjort, N. L., Harris, I. R. and Basu, A. (2001), A comparison of related density-based minimum divergence estimatiors. *Biometrika*, **88**, 865–873.

[10] Nakagawa, T. and Hashimoto, S. (2017). Robust Bayesian inference based on quasi-posterior under heavy contamination. *Hiroshima Research Group Technical Report*, TR17–5.