Objective priors for the robust Bayesian estimation

中川　智之
東京理科大学 理工学部 情報科学科 *
Tomoyuki Nakagawa
Department of Information Sciences, Faculty of Science & Technology
Tokyo University of Science

1　導入

統計解析においてデータが外れ値を含んでいることが多くある．また近年, 膨大なデータが収集されるようになり, さらに外れ値が含まれる状況が増えている．外れ値を含むデータは通常の推定量ではバイアスが大きくなるなど多くの問題がある．

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, Hampel et al. (1986), Huber (2011) を参照). 最近では, Basu et al. (1998) 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, Fujisawa and Eguchi (2008) では Jones et al. (2001) で提案された γ-divergence を用いることで, 外れ値の割合に依らずに頑健な推定が可能であることを示している．

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．Dawid (1973) 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．さらに Andrade and O'Hagan (2006) 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．Desgagné (2015) では, 対数正則変動関数を用いた t-分布など

* 〒278-8510 千葉県野田市山崎 2641

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多くが 1 変数の場合であり, また外れ値が無い場合に推定精度を大きく損なってしまう. そこで Ghosh and Basu (2016) 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, Nakagawa and Hashimoto (2019) では, γ-divergence を用いて, Ghosh and Basu (2016) 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り, さらに 2 つの事後分布を用いたベイズ推定量は外れ値が入っていない場合にも十分な推定が行えていることが, 数値実験からもわかっている. 特に Nakagawa and Hashimoto (2019) で提案されているベイズ推定量は γ-divergence と同様に外れ値の割合が多くても頑健な推定を可能にしている.

本稿では, Ghosh and Basu (2016) や Nakagawa and Hashimoto (2019) で提案している事後分布に対して, 客観事前分布を与える. 客観事前分布は Jeffery の事前分布などに代表されるような無情報事前分布が有名である. 近年, ベイズ論と頻度論を繋げる方法として注目を集めており, 様々な客観事前分布が提案されている (Ghosh (2011) などを参照). 本稿ではまず, 第 2 節で $R^{(\alpha)}$-posterior と γ-posterior の紹介を行い, その漸近性質を第 3 節で紹介する. そして第 4 節で Reference prior と Moment matching prior について紹介する. 最後に第 5 節でそれぞれの事前分布の精度を数値比較する.

2 Robust bayesian estimation

$x_1, \ldots, x_n$ を密度関数 $g(x) = (1-\varepsilon)f(x) + \varepsilon\delta(x)$ から独立同一に観測されたデータとする. ここで, $f(x)$ はターゲットの分布であり, $\delta(x)$ は汚染分布である. また, $f_\theta(x) = f(x|\theta)$ $(\theta \in \Theta \subset \mathbb{R}^p)$ をモデルとし, $f_{\theta_0}(x) = f(x)$ とする. ここで重要なのは $f(x)$ に興味があり, $g(x)$ には興味がない点である. 通常の推測はデータ発生分布と興味がある分布が一致している. しかしながら, 外れ値の問題に関しては $\delta(x)$ は汚染分布であり, 実際に知りたい情報は $f(x)$ のみであるので $g(x)$ に焦点を当てると間違った推測を与えてしまう. ベイズ法の枠組みで, Ghosh and Basu (2016) と Nakagawa and Hashimoto (2019) はそれぞれロバストな事後分布を構成している.

定義 2.1 ($R^{(\alpha)}$-posterior). $\alpha > 0$ に対して,

$$\pi^{(\alpha)}(\theta|X_n) = \frac{\exp\{-nd_\alpha(\bar{g}, f_\theta)\}\pi(\theta)}{\int_\Theta \exp\{-nd_\alpha(\bar{g}, f_\theta)\}\pi(\theta)d\theta}$$

を $R^{(\alpha)}$-*posterior* という．ここで，

$$-nd_\alpha(\bar{g}, f_\theta) = \sum_{j=1}^n q^{(\alpha)}(x_j;\theta), \qquad q^{(\alpha)}(x;\theta) = \frac{1}{\alpha} f_\theta(x) - \frac{1}{\alpha+1}\int f_\theta^{1+\alpha}(t)dt.$$

注意 2.1. ここで，$d_\alpha(g,f) = \frac{1}{\alpha}\int g(x)f_\theta(x)dx - \frac{1}{\alpha+1}\int f_\theta^{1+\alpha}(t)dt$ は *density-power divergence* の *cross entropy* であり，$R^{(\alpha)}$-*posterior* は *density-power divergence* と非常に関係がある．

定義 2.2 (γ-posterior)**.** $\gamma > 0$ に対して，

$$\pi^{(\gamma)}(\theta|X_n) = \frac{\exp\{-n\tilde{d}_\gamma(\bar{g}, f_\theta)\}\pi(\theta)}{\int_\Theta \exp\{-n\tilde{d}_\gamma(\bar{g}, f_\theta)\}\pi(\theta)d\theta}$$

を γ-*posterior* という．ここで，

$$-n\tilde{d}_\gamma(\bar{g}, f_\theta) = \sum_{j=1}^n q^{(\gamma)}(x_i;\theta) - \frac{n}{\gamma}, \quad q^{(\gamma)}(x;\theta) = \frac{1}{\gamma} f_\theta^\gamma(x)\left\{\int f_\theta^{1+\gamma}(t)dt\right\}^{-\gamma/(1+\gamma)}.$$

注意 2.2. ここで，$d_\gamma(g,f) = \frac{1}{\gamma}\log\int g(x)f_\theta(x)dx - \frac{1}{\gamma+1}\log\int f_\theta^{1+\alpha}(t)dt$ は γ-*divergence* の *cross entropy* であり，$\tilde{d}_\gamma(g,f)$ は $\tilde{d}_\gamma(g,f) = -(1/\gamma)\exp\{-\gamma d_\gamma(g,f)\} + 1/\gamma$ のように γ-*divergence* の *cross entropy* の単調変換にで与えられる．そのため，γ-*posterior* は γ-*divergence* と非常に関係がある．また，$\tilde{d}_\gamma(g,f)$ は擬球スコア *(pseudo-spherical score; Good (1971))* に対応している．

また Ghosh and Basu (2016) と Nakagawa and Hashimoto (2019) では, それぞれの事後分布から推定量の漸近性質とロバスト性を示している．

3 漸近性質

d をダイバージェンスとし，$\hat{\theta}_n = \arg\min_{\theta\in\Theta} d(\bar{g}, f_\theta)$，$\theta_g = \arg\min_{\theta\in\Theta} d(g, f_\theta)$，$\tilde{\theta}_n = \int \theta\pi^{(d)}(\theta|X_n)d\theta$, $Q_n(\theta) = -nd(\bar{g}, f_\theta)$ とする．このとき, Ghosh and Basu (2016) と Nakagawa and Hashimoto (2019) では, いくつかの正則条件のもとで以下の定理が成り立っている．

定理 3.1. 確率 1 で $\partial Q_n^{(d)}(\hat{\theta}_n)/\partial\theta = 0$ を満たし，$\hat{\theta}_n \xrightarrow{p} \theta_g$，$\pi(\theta)$ を連続とする．このとき，

$$\int \left|\pi^{*(d)}(t|X_n) - (2\pi)^{-p/2}\left|J^{(d)}(\theta_g)\right|^{1/2}\exp\left(-\frac{1}{2}t^\top J^{(d)}(\theta_g)t\right)\right| dt \xrightarrow{p} 0$$

が成り立つ．ここで，

$$J^{(\gamma)}(\theta) = -E_{\theta_g}\left[\frac{\partial^2}{\partial\theta\partial\theta^\top} q^{(d)}(\theta; X_1)\right]$$

である．

また以下も成り立っている．

定理 3.2. 確率 1 で $\partial Q_n^{(d)}(\hat{\theta}_n)/\partial\theta = 0$ を満たし，$\hat{\theta}_n \xrightarrow{p} \theta_g$, $\pi(\theta)$ の期待値が有限とする．このとき，$n \to \infty$ に対して

$$\sqrt{n}\left(\tilde{\theta}_n - \hat{\theta}_n\right) \xrightarrow{p} 0,$$
$$\sqrt{n}\left(\tilde{\theta}_n - \theta_g\right) \xrightarrow{d} N_p(0, V^{(d)}(\theta_g))$$

が成り立つ．ここで，

$$V^{(d)}(\theta) = J^{(d)}(\theta)^{-1} I^{(d)}(\theta) J^{(d)}(\theta)^{-1}, \quad I^{(\gamma)}(\theta) = E_{\theta_g}\left[\frac{\partial}{\partial\theta} q^{(d)}(\theta; X_1) \frac{\partial}{\partial\theta^\top} q^{(d)}(\theta; X_1)\right]$$

である．

この定理により，$R^{(\alpha)} - posterior$ と γ-posterior の事後平均はそれぞれの対応するダイバージェンスの推定量と漸近的に同じ性質を持つことがわかる．

4 Objective priors

$R^{(\alpha)}$-posterior と γ-posterior は事前分布の影響を大きく受けることは Nakagawa and Hashimoto (2019) 等でよく知られている．通常のベイズ推定では事前情報などがない場合には Jefferys prior などの客観事前分布や共役事前分布を使うことが多い．しかしながら，$R^{(\alpha)}$-posterior や γ-posterior は事後分布が複雑な形をしているため，共役事前分布を考えることが困難である．そのため，本稿では客観事前分布に着目する．客観事前分布はいくつかあり，代表的なものとして，事後分布と事前分布の距離を最大にする事前分布 (Reference prior) がある (Ghosh (2011) などを参照)．Giummolè et al. (2017) では $R^{(\alpha)}$-posterior と γ-posterior を含む擬似的な事後分布に対して，Reference prior に対応する事前分布が

$$\pi_R^{(d)}(\theta) \propto |J^{(d)}(\theta)|^{1/2}.$$

となることを示している．一方で，Ghosh and Liu (2011) で最尤推定量と高次のオーダーでバイアスが一致する Moment Matching prior が提案されている．本稿では $p = 1$ の場

合に $R^{(\alpha)}$-posterior と γ-posterior に対応する Moment matching prior を導出する．定理 3.1 の条件と以下の条件を仮定する．

- $q(x;\theta)$ が 4 階微分可能とする．
- $E[M(X)] < \infty$ となる関数で, 以下を満たすものが存在する．

$$\left|\frac{\partial^4}{\partial\theta^4}q(x;\theta)\right| \leq M(x).$$

このとき, Giummolè et al. (2017) より以下が成り立つ．

定理 4.1. 確率 1 で $\partial Q_n^{(d)}(\hat{\theta}_n)/\partial\theta = 0$ を満たし, $\hat{\theta}_n \xrightarrow{p} \theta_g$, $\pi(\theta)$ が C^1 級の関数とする．このとき,

$$\tilde{\theta}_n = \hat{\theta}_n + \frac{1}{nJ^{(d)}(\hat{\theta}_n)}\left\{-\frac{1}{2J^{(d)}(\hat{\theta}_n)}\frac{\partial^3}{\partial\theta^3}d\left(\bar{g}, f_{\hat{\theta}_n}\right) + \frac{\pi'\left(\hat{\theta}_n\right)}{\pi\left(\hat{\theta}_n\right)}\right\} + o_p(n^{-1})$$

を得る．

これより,

$$n(\tilde{\theta}-\hat{\theta}) \xrightarrow{p} \frac{1}{J^{(d)}(\theta_g)}\left\{-\frac{g_3^{(d)}(\theta_g)}{2J^{(d)}(\theta_g)} + \frac{\pi'(\theta_g)}{\pi(\theta_g)}\right\}$$

となる．ここで, $g_3(\theta) = E_{\theta_g}[\partial^3 q(X_1;\theta)/\partial\theta^3]$ とする．つまり,

$$\pi_M^{(d)}(\theta) = \exp\left\{-\int^{\theta}\frac{g_3^{(d)}(t)}{J^{(d)}(t)}dt\right\}$$

とすれば, $n(\tilde{\theta}-\hat{\theta}) \xrightarrow{p} 0$ となる． $\pi_M(\theta)$ を事後分布 $\pi^{(d)}(\theta|X_n)$ に対する Moment matching prior という．

5 数値実験

5.1 設定

$R^{(\alpha)}$-posterior と γ-posterior について一様分布, Reference prior と Moment Matching prior に関する数値比較を行う．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．またデータ発生分布を

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

とし, 候補モデルを正規分布 $N(\mu,\sigma^2)$ とする．正規分布の場合の Reference prior は以下のようのに与えられる．

$$\pi_R^{(\alpha)}(\mu,\sigma)\propto\sigma^{-(2+\alpha)}\ (R^{(\alpha)}\text{-posterior}),$$
$$\pi_R^{(\gamma)}(\mu,\sigma)\propto\sigma^{-3+1/(1+\gamma)}\ (\gamma\text{-posterior}).$$

また, Moment Matching prior は $p=1$ の場合,

$$\pi_M^{(\alpha)}(\sigma)\propto\sigma^{A_1/2A_2}\ (R^{(\alpha)}\text{-posterior}),$$
$$\pi_M^{(\gamma)}(\mu,\sigma)\propto\sigma^{-(5+\gamma)/2(1+\gamma)}\ (\gamma\text{-posterior}),$$

となる．ここで,

$$A_1=\pi^{(1+\alpha)/2}(\alpha(1+\alpha)^3(2+\alpha)+(10-\alpha^2(-2+\alpha(5+\alpha(3+\alpha))))),$$
$$A_2=(1+\alpha)(-\alpha(1+\alpha)^2\sqrt{\pi}+(-2+\alpha+\alpha^2+\alpha^3)\pi^{(1+\alpha)/2}).$$

5.2 結果

全ての表から通常の事後分布を用いるよりは $R^{(\alpha)}$-posterior と γ-posterior を用いる方がバイアスが小さくなることがわかる．また表 1 と表 2 の結果と比べ, 表 3 と表 4 の結果ではバイアスは小さくなっている．さらに表 3 と表 4 と表 5 と表 6 を比較するとほぼ変わらないことがわかる．このことから Objective prior を用いた方が推定量のバイアスは小さくなる．一方で α, γ によってこれらの事後分布での推定は影響されており, また外れ値の大きさによってもかなり変動することが表よりわかる．

6 まとめ

客観事前分布を使うことで, バイアスを小さくすることが可能になった．Moment Matching prior に関しては, 多変量への拡張も考えており, より汎用的に用いることができると考えられる．また外れ値がない場合でも十分な標本数があれば, 通常のベイズ推定とバイアスは変わらないことがわかった．また $R^{(\alpha)}$-posterior の場合は仮定するモデルによって Objective prior に外れ値の影響が出てくる．一方で γ-posterior に関する

表 1　The empirical biases of the posterior means for the mean parameter under the uniform prior

		ordinary	$R^{(\alpha)}$-posterior				γ-posterior			
		α, γ	α				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 biases of the posterior means for the variance parameter under the uniform prior

		ordinary	$R^{(\alpha)}$-posterior				γ-posterior			
		α, γ	α				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

Objective prior は Fujisawa and Eguchi (2008) の $\int \delta(x) f_\theta(x) dx \approx 0$ の仮定を置けば, モデルにのみで Objective prior を導出できる. しかしながら, 標本数が小さい場合にはバイアスはあまり改善できていない. 今後は小標本の場合にも有効なものを提案する必要がある.

表 3　The empirical biases of the posterior means for the mean parameter under the reference prior

		ordinary	$R^{(\alpha)}$-posterior				γ-posterior			
		α, γ	α				γ			
ε	n	0.00	0.30	0.50	0.70	1.00	0.30	0.50	0.70	1.00
0.00	20	0.005	0.001	-0.001	-0.003	0.001	0.001	-0.001	-0.003	-0.004
0.00	50	-0.005	0.006	-0.004	-0.003	-0.002	-0.005	-0.004	-0.003	-0.002
0.00	100	-0.001	-0.002	-0.003	-0.003	-0.004	-0.002	-0.003	-0.003	-0.004
0.05	20	0.284	0.028	0.011	0.013	0.069	0.025	0.006	0.002	0.003
0.05	50	0.302	0.013	0.002	0.000	0.003	0.011	0.001	0.000	-0.001
0.05	100	0.296	0.005	-0.002	-0.003	-0.002	0.004	-0.002	-0.003	-0.003
0.20	20	1.214	0.596	0.338	0.357	0.685	0.579	0.270	0.187	0.237
0.20	50	1.194	0.510	0.141	0.085	0.139	0.483	0.087	0.025	0.011
0.20	100	1.217	0.516	0.088	0.035	0.037	0.480	0.044	0.013	0.008

表 4　The empirical biases of the posterior means for the variance parameter under the reference prior

		ordinary	$R^{(\alpha)}$-posterior				γ-posterior			
		α, γ	α				γ			
ε	n	0.00	0.30	0.50	0.70	1.00	0.30	0.50	0.70	1.00
0.00	20	0.025	0.004	-0.007	-0.015	-0.042	-0.002	-0.039	-0.105	-0.213
0.00	50	0.015	0.006	0.000	-0.007	-0.001	0.004	-0.009	-0.032	-0.093
0.00	100	0.004	0.000	-0.002	-0.006	-0.009	0.000	-0.006	-0.016	-0.042
0.05	20	0.610	0.111	0.090	0.130	0.180	0.089	-0.029	-0.007	-0.047
0.05	50	0.643	0.054	0.033	0.036	0.073	0.038	-0.002	-0.019	-0.059
0.05	100	0.638	0.038	0.024	0.026	0.037	0.025	0.000	-0.011	-0.035
0.20	20	1.664	1.032	0.772	0.811	0.788	0.966	0.560	0.453	0.451
0.20	50	1.619	0.947	0.412	0.382	0.567	0.874	0.244	0.153	0.146
0.20	100	1.620	0.967	0.285	0.233	0.313	0.877	0.133	0.080	0.080

謝辞

本研究を行うにあたって, 広島大学の橋本准教授には有益なコメントをいただきました. ここに感謝の意を表します. 本研究は JSPS 科研費 JP19K14597 の助成を受けたものです.

表 5　The mean parameter under the moment matching prior

		ordinary	$R^{(\alpha)}$-posterior				γ-posterior			
		α,γ	α				γ			
ε	n	0.00	0.30	0.50	0.70	1.00	0.30	0.50	0.70	1.00
0.00	20	-0.002	-0.001	-0.001	-0.001	0.001	-0.001	-0.001	0.000	0.000
0.00	50	0.000	0.001	0.001	0.001	-0.002	0.001	0.001	0.001	0.001
0.00	100	0.004	0.003	0.003	-0.003	-0.003	0.003	0.003	0.003	0.003
0.05	20	0.290	0.022	0.007	0.006	0.028	0.020	0.004	0.003	0.014
0.05	50	0.294	0.016	0.008	0.007	0.007	0.015	0.007	0.006	0.006
0.05	100	0.302	0.015	0.008	0.008	0.008	0.014	0.008	0.007	0.008
0.20	20	1.202	0.585	0.349	0.313	0.530	0.576	0.319	0.274	0.362
0.20	50	1.219	0.535	0.174	0.091	0.058	0.514	0.129	0.057	0.045
0.20	100	1.205	0.509	0.079	0.029	0.012	0.477	0.043	0.011	0.008

表 6　The variance parameter under the moment matching prior

		ordinary	density power posterior				γ-posterior			
		α,γ	α				γ			
ε	n	0.00	0.30	0.50	0.70	1.00	0.30	0.50	0.70	1.00
0.00	20	0.031	0.012	0.011	-0.054	-0.454	0.011	0.016	0.029	0.093
0.00	50	0.008	0.003	0.001	-0.022	-0.229	0.003	0.005	0.006	0.000
0.00	100	0.002	0.000	-0.001	-0.013	-0.120	0.000	0.000	0.000	0.000
0.05	20	0.616	0.099	0.101	0.070	-0.303	0.085	0.080	0.130	0.316
0.05	50	0.630	0.051	0.036	0.019	-0.178	0.040	0.020	0.020	0.032
0.05	100	0.636	0.038	0.026	0.020	-0.086	0.027	0.010	0.007	0.006
0.20	20	1.659	1.018	0.814	0.661	0.020	0.975	0.713	0.782	1.221
0.20	50	1.638	0.966	0.450	0.347	0.082	0.906	0.309	0.217	0.259
0.20	100	1.613	0.969	0.278	0.215	0.102	0.887	0.144	0.095	0.110

参考文献

Andrade, J. A. A. and O'Hagan, A. (2006). Bayesian robustness modeling using regularly varying distributions. *Bayesian Analysis*, 1(1):169–188.

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.

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

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.

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.

Ghosh, M. (2011). Objective priors: An introduction for frequentists. *Statistical Science*, 26(2):187–202.

Ghosh, M. and Liu, R. (2011). Moment matching priors. *Sankhya A*, 73(2):185–201.

Giummolè, F., Mameli, V., Ruli, E., and Ventura, L. (2017). Objective bayesian inference with proper scoring rules. *TEST*, pages 1–28.

Good, I. J. (1971). Comment on "measuring information and uncertainty," by R. J. Buehler. In *Foundations of Statis- tical Inference*, V. P. Godambe & D. A. Sprott, eds. Toronto: Holt, Rinehart and Winston.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., and Stahel, W. A. (1986). *Robust statistics*. Wiley Online Library.

Huber, P. J. (2011). *Robust statistics*. Springer.

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.

Nakagawa, T. and Hashimoto, S. (2019). Robust bayesian inference via γ-divergence. *Communications in Statistics-Theory and Methods*, pages 1–18.