Intrinsic posterior regret gamma-minimax estimation for the exponential family of distributions

Mohammad Jafari Jozani[*]

Department of Statistics, University of Manitoba, Winnipeg, MB, Canada, R3T 2N2
e-mail: m_jafari_jozani@umanitoba.ca

and

Nahid Jafari Tabrizi

Department of Mathematics, Islamic Azad University-Karaj Branch, Karaj, Iran
e-mail: njafaritabrizi@yahoo.com

Abstract: In practice, it is desired to have estimates that are invariant under reparameterization. The invariance property of the estimators helps to formulate a unified solution to the underlying estimation problem. In robust Bayesian analysis, a frequent criticism is that the optimal estimators are not invariant under smooth reparameterizations. This paper considers the problem of posterior regret gamma-minimax (PRGM) estimation of the natural parameter of the exponential family of distributions under intrinsic loss functions with Kullback-Leibler distance. We show that under the class of Jeffrey's Conjugate Prior (JCP) distributions, PRGM estimators are invariant to smooth one-to-one reparameterizations. We apply our results to several distributions and different classes of JCP, as well as the usual conjugate prior distributions. We observe that, in many cases, invariant PRGM estimators in the class of JCP distributions can be obtained by some modifications of PRGM estimators in the usual class of conjugate priors. Moreover, when the class of priors are convex or dependant on a hyper-parameter belonging to a connected set, we show that the PRGM estimator under the intrinsic loss function could be Bayes with respect to a prior distribution in the original prior class. Theoretical results are supplemented with several examples and illustrations.

AMS 2000 subject classifications: Primary 62F10, 62F15.
Keywords and phrases: Intrinsic loss function, Bayes estimator, robust Bayesian analysis, posterior risk, posterior regret gamma-minimax.

Received December 2012.

Contents

---

[*]Corresponding author.

1. Introduction

Suppose x is a realization of a random sample X with a sampling model given by a family of densities $\{f(\cdot|\theta) : \theta \in \Theta\}$ with respect to a σ-finite measure ν on a sample space χ where θ is the unknown parameter of interest with $\theta \in \Theta$. Let $\pi(\cdot)$ be a prior distribution on Θ and $\pi(\cdot|x)$ denote the posterior distribution of θ given x. In standard Bayesian analysis, one needs to specify a prior distribution $\pi(\cdot)$. However, in practice, elicitation of the prior distribution can never be done without error. Hence, we usually need to consider a class Γ of prior distributions which reflects (approximately) true prior beliefs, i.e., the prior distribution $\pi(\cdot)$ is an unknown element of Γ. Robust Bayesian analysis is designed to acknowledge such a prior uncertainty by considering the class Γ of plausible prior distributions instead of a single prior distribution π and studying the corresponding range of Bayesian solutions. See [1, 14] for comprehensive overview of different robust Bayesian analysis methods and their applications. One may also attempt to determine an optimal estimator δ by minimizing some measures of robustness. Several criteria have been proposed for the selection of procedures in robust Bayesian studies. In this paper, we study the maximal posterior regret method (e.g., [14, 15]) to obtain the posterior regret gamma-minimax (PRGM) estimator of the unknown parameter for the one-parameter exponential family of distributions. The PRGM criterion has been used recently by many people from both theoretical and practical points of view. For example, [9] investigated the use of PRGM for credibility premium estimation in Actuarial Science, [4, 5] in insurance for collective risk model analysis, and [12] in statistical inference based on record data.

For an observed value x, a prior distribution π and the corresponding posterior distribution $\pi(\cdot|x)$, we denote the posterior risk of an estimate $\delta(x)$ of the unknown parameter θ under $L(\theta, \delta)$ by $r(x, \delta) = \mathbb{E}[L(\theta, \delta(x))|x]$. The Bayes estimator of θ under the loss function $L(\theta, \delta)$ is then given by a $\delta_\pi(X)$ such that $r(x, \delta_\pi) = \inf_\delta r(x, \delta)$.

Definition 1. The PRGM estimator of θ under the loss function $L(\theta, \delta)$ and a class Γ of prior distributions is defined as an estimator δ_{PR} such that

$$\sup_{\pi \in \Gamma} \rho(\delta_\pi, \delta_{PR}) = \inf_\delta \sup_{\pi \in \Gamma} \rho(\delta_\pi, \delta), \qquad (1.1)$$

where $\rho(\delta_\pi, \delta) = r(x, \delta) - r(x, \delta_\pi)$ is the posterior regret, measuring the loss entailed in choosing the action $\delta(x)$ instead of the optimal Bayes action $\delta_\pi(x)$ (under prior π and loss L).

In this paper, we study the construction of PRGM estimators under the so-called *intrinsic loss* functions. These loss functions shift attention from the

distance between the estimator δ and the true parameter value θ, to the more relevant distance between statistical models they label. More specifically, the intrinsic loss of using δ as a proxy for θ is the intrinsic distance between the true model $f(x|\theta)$ and the model $f(x|\delta)$ when $\theta = \delta$, that is

$$L(\theta, \delta) = d\left(f(x|\theta), f(x|\delta)\right), \tag{1.2}$$

where $d(\cdot, \cdot)$ is a suitable distance measure. In practice, intrinsic loss functions could be used as benchmark losses when the utility function related to the underlying statistical problem cannot be obtained by practitioners. A desired property of intrinsic loss functions is that they are invariant under one-to-one smooth reparameterizations. The invariance property of intrinsic loss functions provides a very convenient tool for statistical application.

Given two densities $f(\cdot|\theta)$ and $f(\cdot|\delta)$ with respect to the σ-finite measure ν, there are many choices of the distance $d(f(\cdot|\theta), d(\cdot|\delta))$ in (1.2). In this paper, we consider the Kullback-Leibler distance and we show that, under suitable conditions, it can be used to formulate a unified set of solutions to the problem of PRGM estimation of the unknown parameter of the exponential family of distributions. As mentioned by [2], an estimator intended for general use should surely be invariant under one-to-one transformations, especially when we merely wish to report an estimate for some quantity of interest. For example, suppose one obtains the PRGM estimate of the variance of the normal model. It would be very difficult to sell to a practitioner that he/she can not use the square root of this estimate as the PRGM estimate of the standard deviation. This is a rather obvious requirement, which unfortunately many statistical methods fail to satisfy. We provide a solution to this problem in Section 3.

The outline of this paper is as follows. In Section 2, we obtain the PRGM estimator of the natural parameter θ of the exponential family of distributions under the intrinsic loss function (1.2) when $d(\cdot, \cdot)$ is chosen to be the Kullback-Leibler distance. We consider different classes of conjugate priors on the natural parameter θ and show how to obtain the PRGM estimator of θ in each class. We provide an automated and unified solution to the PRGM estimation of the unknown parameter of the exponential family of distributions under different loss functions, including, but not limited to, quadratic, LINEX, entropy and Stein loss functions.

In Bayesian statistical analysis, as pointed out by [8], transformations of the parameter typically suggest new families of prior distributions. Therefore, the usual robust Bayesian inferences are not invariant under reparameterizations. For example, if $\delta_{PR}(X)$ is the PRGM estimator of θ, then it is not necessarily true that $h(\delta_{PR}(X))$ is the PRGM estimator of $\eta = h(\theta)$, when h is a one-to-one smooth function. A solution to this problem is proposed in Section 3. To this end, we obtain invariant PRGM estimators of θ under the intrinsic loss function and different classes of Jeffrey's Conjugate Prior (JCP) distributions. We show that the resulting PRGM estimates are invariant under one-to-one smooth transformations of θ. Theoretical results are augmented with several examples and illustrations. In Section 4, we provide some general results showing that, under general conditions, PRGM and intrinsic PRGM estimators are

Bayes with respect to prior distributions in the underlying class of priors. We study two cases of convex classes of prior distributions as well as the case where the underlying class of priors depends on a hyper-parameter belonging to a connected set. We provide a sufficient condition under which the PRGM and intrinsic PRGM estimators are Bayes with respect to data independent prior distributions within the underlying class of priors. In Section 5, we present some examples to show how the PRGM and intrinsic PRGM results can be extended to a multi-parameter exponential family of distributions. Finally, in Section 6, we give some concluding remarks.

2. PRGM estimation under intrinsic loss functions

Suppose X is a random variable, where its distribution belongs to the one-parameter exponential family of distributions $\mathcal{F} = \{f(x|\theta) : x \in \chi \subseteq \mathbb{R}, \theta \in \Theta \subseteq \mathbb{R}\}$, with probability density function (pdf)

$$f(x|\theta) = \beta(\theta)t(x)e^{-\theta r(x)}, \tag{2.1}$$

where $r(x) > 0$, $\beta(\theta)t(x) > 0$ and θ is the unknown real-valued natural parameter of the model. The density is considered with respect to the Lebesgue measure for continuous and the counting measure for discrete distributions. Suppose δ is an estimate of θ with both $\theta, \delta \in \Theta$. We define the intrinsic loss function (1.2), using the Kullback-Leibler measure between $f(x|\theta)$ and $f(x|\delta)$, as follows

$$L(\theta, \delta) = \mathbb{E}_\theta \left[\log \left(\frac{f(X|\theta)}{f(X|\delta)} \right) \right] = \int_\chi \log \left(\frac{f(x|\theta)}{f(x|\delta)} \right) f(x|\theta) d\nu(x). \tag{2.2}$$

Loss function (2.2) can be interpreted as the expected log-likelihood ratio in favour of the true model. Thus, the intrinsic loss function (2.2) not only has the desired invariance property but it is also related to the relevant measure of evidence in the Neyman-Pearson Lemma. Note that the intrinsic loss function (2.2) is invariant under reparameterization since the parameters affect the loss function only via the probability distributions they label, which are independent of the particular parameterization. For more details on intrinsic loss functions, we refer to [13] who studied the Bayesian estimation of the unknown parameters of the exponential family of distributions under intrinsic loss functions with Kullback-Leibler and Hellinger distances with respect to conjugate class of priors. Similar results for symmetric intrinsic loss functions under reference priors are given by [2] and references cited therein.

First, we give a lemma which identifies the intrinsic loss function for the exponential family of distributions.

Lemma 1. *For the exponential family of distributions* (2.1)*, the intrinsic loss function* (2.2) *reduces to*

$$L(\theta, \delta) = \log \left(\frac{\beta(\theta)}{\beta(\delta)} \right) + (\delta - \theta) \frac{\beta'(\theta)}{\beta(\theta)}, \tag{2.3}$$

where $\beta'(\theta) = \frac{d}{d\theta}\beta(\theta)$.

Let $H(t) := \beta'(t)/\beta(t)$. A straightforward calculation shows that the posterior risk associated with δ, under the loss function (2.3), is

$$r(x, \delta) = \mathbb{E}(\log \beta(\theta)|x) - \log \beta(\delta(x)) + \delta(x)\mathbb{E}\left(H(\theta)\big|x\right) - \mathbb{E}\left(\theta H(\theta)\big|x\right). \quad (2.4)$$

The Bayes estimator of θ can therefore be obtained by minimizing (2.4) in δ as follows

$$\delta_\pi(X) = H^{-1}\{\mathbb{E}\left(H(\theta)\big|X\right)\}. \quad (2.5)$$

Following the decreasing monotone likelihood ratio property of densities (2.1) in $r(X)$, and since $\mathbb{E}[r(X)] = H(\theta)$, $H(\cdot)$ is a decreasing function. Therefore, the Bayes estimator $\delta_\pi(X)$ is unique. Furthermore, the posterior regret for estimating θ using δ instead of the optimal estimator δ_π is obtained by

$$\rho(\delta_\pi, \delta) = \log \frac{\beta(\delta_\pi)}{\beta(\delta)} + (\delta - \delta_\pi)H(\delta_\pi). \quad (2.6)$$

Note that $\rho(\delta_\pi, \delta)$, as a function of δ_π, decreases then increases with a unique minimum at $\delta_\pi = \delta$. The main result of this section is given in the following theorem which obtains the PRGM estimator of θ under the intrinsic loss function (2.3).

Theorem 1. *Let $\underline{\delta}(x) = \inf_{\pi \in \Gamma} \delta_\pi(x)$ and $\overline{\delta}(x) = \sup_{\pi \in \Gamma} \delta_\pi(x)$ and suppose that $\underline{\delta}(x)$ and $\overline{\delta}(x)$ are finite almost everywhere. The PRGM estimator of θ in the exponential family (2.1) under the loss function (2.3) and in the class of prior distributions Γ is given by*

$$\delta_{PR}(X) = \frac{\overline{\delta}(X)\, H(\overline{\delta}(X)) - \underline{\delta}(X)\, H(\underline{\delta}(X)) - \log \frac{\beta(\overline{\delta}(X))}{\beta(\underline{\delta}(X))}}{H(\overline{\delta}(X)) - H(\underline{\delta}(X))}. \quad (2.7)$$

Proof. First, note that

$$\inf_\delta \sup_{\pi \in \Gamma} \rho(\delta_\pi, \delta) = \min\left\{\inf_{\delta \leq \underline{\delta}} \sup_{\pi \in \Gamma} \rho(\delta_\pi, \delta), \inf_{\underline{\delta} < \delta < \overline{\delta}} \sup_{\pi \in \Gamma} \rho(\delta_\pi, \delta), \inf_{\delta \geq \overline{\delta}} \sup_{\pi \in \Gamma} \rho(\delta_\pi, \delta)\right\}.$$

So, we consider the following three cases:

Case 1. When $\delta \leq \underline{\delta}$, we have $\sup_{\pi \in \Gamma} \rho(\delta_\pi, \delta) = \rho(\overline{\delta}, \delta)$. Let $f_1(\delta) = \rho(\overline{\delta}, \delta) = \log \frac{\beta(\overline{\delta})}{\beta(\delta)} + (\delta - \overline{\delta})H(\overline{\delta})$ with $f_1'(\delta) = H(\overline{\delta}) - H(\delta) < 0$, following the decreasing property of $H(\cdot)$. Hence, $f_1(\delta)$ is a decreasing function of δ for $\delta \leq \underline{\delta}$ and $\inf_{\delta \leq \underline{\delta}} f_1(\delta) = f_1(\underline{\delta})$. Therefore,

$$\inf_{\delta \leq \underline{\delta}} \sup_{\pi \in \Gamma} \rho(\delta_\pi, \delta) = \rho(\overline{\delta}, \underline{\delta}).$$

Case 2. For $\delta \geq \overline{\delta}$, we have $\sup_{\pi \in \Gamma} \rho(\delta_\pi, \delta) = \rho(\underline{\delta}, \delta)$. Let $f_2(\delta) = \rho(\underline{\delta}, \delta) = \log \frac{\beta(\underline{\delta})}{\beta(\delta)} + (\delta - \underline{\delta})H(\underline{\delta})$ with $f_2'(\delta) = H(\underline{\delta}) - H(\delta) > 0$. Hence, $f_2(\delta)$ is an increasing function of δ for $\delta \geq \overline{\delta}$ and $\inf_{\delta \geq \overline{\delta}} f_2(\delta) = f_2(\overline{\delta})$. Therefore,

$$\inf_{\delta \geq \overline{\delta}} \sup_{\pi \in \Gamma} \rho(\delta_\pi, \delta) = \rho(\underline{\delta}, \overline{\delta}).$$

Case 3. If $\underline{\delta} < \delta < \overline{\delta}$, then $\sup_{\pi \in \Gamma} \rho(\delta_\pi, \delta) = \max\{\rho(\overline{\delta}, \delta), \rho(\underline{\delta}, \delta)\}$. Let $f_3(\delta) = f_1(\delta) - f_2(\delta)$ where $f_3'(\delta) = H(\overline{\delta}) - H(\underline{\delta}) < 0$. Since $f_3(\delta)$ is a decreasing function of δ with $f_3(\overline{\delta}) < 0$ and $f_3(\underline{\delta}) > 0$, there exists a unique $\delta^* \in (\underline{\delta}, \overline{\delta})$ (as the root of $f_3(\delta) = 0$) such that $\rho(\underline{\delta}, \delta^*) = \rho(\overline{\delta}, \delta^*)$. Hence, for $\underline{\delta} < \delta < \delta^*$, $\sup_{\pi \in \Gamma} \rho(\delta_\pi, \delta) = \rho(\overline{\delta}, \delta)$ and for $\delta^* < \delta < \overline{\delta}$, $\sup_{\pi \in \Gamma} \rho(\delta_\pi, \delta) = \rho(\underline{\delta}, \delta)$. Note that, for $\underline{\delta} < \delta < \overline{\delta}$, $\rho(\overline{\delta}, \delta)$ is a decreasing function in δ with $\inf_{\underline{\delta} < \delta < \delta^*} \sup_{\pi \in \Gamma} \rho(\delta_\pi, \delta) = \rho(\overline{\delta}, \delta^*)$ and $\rho(\underline{\delta}, \delta)$ is an increasing function in δ with $\inf_{\delta^* < \delta < \overline{\delta}} \sup_{\pi \in \Gamma} \rho(\delta_\pi, \delta) = \rho(\underline{\delta}, \delta^*)$. Therefore,

$$\inf_{\underline{\delta} < \delta < \overline{\delta}} \sup_{\pi \in \Gamma} \rho(\delta_\pi, \delta) = \rho(\underline{\delta}, \delta^*) = \rho(\overline{\delta}, \delta^*).$$

Following the above cases, we conclude that

$$\inf_{\delta \in D} \sup_{\pi \in \Gamma} \rho(\delta_\pi, \delta) = \inf_{\underline{\delta} < \delta < \overline{\delta}} \sup_{\pi \in \Gamma} \rho(\delta_\pi, \delta) = \rho(\underline{\delta}, \delta^*) = \rho(\overline{\delta}, \delta^*).$$

That is, the PRGM estimator of θ is given by $\delta_{PR} = \delta^* \in (\underline{\delta}, \overline{\delta})$, as the solution of

$$\log \frac{\beta(\overline{\delta})}{\beta(\underline{\delta})} + \delta_{PR}\big(H(\overline{\delta}) - H(\underline{\delta})\big) + \underline{\delta}H(\underline{\delta}) - \overline{\delta}H(\overline{\delta}) = 0,$$

in δ_{PR} which results in the estimator (2.7). $\qquad \square$

We give some applications of Theorem 1.

Example 1 (Normal distribution). Suppose $X \sim N(\mu, 1)$ is a normally distributed random variable with unknown parameter $\mu \in \mathbb{R}$ and pdf $f(x|\mu) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-\mu)^2}$, $-\infty < x < \infty$. The pdf $f(x|\mu)$ belongs to the exponential family (2.1) with $\theta = \mu$, and $\beta(\theta) = e^{-\frac{\theta^2}{2}}$. Also, $H(\theta) = -\theta$, and the intrinsic loss function (2.3) reduces to $L(\theta, \delta) = \frac{1}{2}(\delta - \theta)^2$ which is proportional to the usual squared error loss function. Let $\underline{\delta}$ and $\overline{\delta}$ be defined as in Theorem 1. Using (2.7), subject to the existence of $\overline{\delta}$ and $\underline{\delta}$, the PRGM estimator of θ in the class Γ of prior distributions is given by

$$\delta_{PR}(X) = \frac{1}{2}(\overline{\delta}(X) + \underline{\delta}(X)),$$

which is also obtained in [1, 15]. [1] provides an excellent overview of robust Bayesian analysis, including the PRGM estimation under squared error loss function.

Example 2 (Exponential distribution). Suppose $X \sim Exp(\sigma)$ is an exponential random variable with pdf $f(x|\sigma) = \frac{1}{\sigma} e^{-x/\sigma}$, $x > 0$, where $\sigma > 0$ is the unknown parameter. The pdf $f(x|\sigma)$ belongs to the exponential family (2.1) with $\theta = \frac{1}{\sigma}$, and $\beta(\theta) = \theta$. In this case, $H(\theta) = \theta^{-1}$, and the intrinsic loss function (2.3) reduces to the Stein loss

$$L(\theta, \delta) = \frac{\delta}{\theta} - \log \frac{\delta}{\theta} - 1.$$

Using (2.7), subject to the existence of $\overline{\delta}$ and $\underline{\delta}$, the PRGM estimator of θ under the Stein loss function is given by

$$\delta_{PR}(X) = \frac{\log \frac{1}{\overline{\delta}(X)} - \log \frac{1}{\underline{\delta}(X)}}{\frac{1}{\overline{\delta}(X)} - \frac{1}{\underline{\delta}(X)}}.$$

The PRGM estimator of σ is also obtained in Example 5.

Example 3 (Binomial distribution). Suppose $X \sim Bin(n, p)$ is a binomial random variable with probability mass function (pmf) $f(x|p) = \binom{n}{x} p^x (1-p)^{n-x}$, where n is known, $x = 0, 1, \ldots, n$, and $p \in [0, 1]$ is the unknown parameter. The pmf $f(x|p)$ is a member of the exponential family (2.1) with $\theta = \log(\frac{1-p}{p})$ and $\beta(\theta) = (1 + e^{-\theta})^{-n}$. We also have $H(\theta) = \frac{n}{1+e^\theta}$ which results in the intrinsic loss function

$$L(\theta, \delta) = n \left\{ \log \left(\frac{e^\theta}{e^\delta} \cdot \frac{1 + e^\delta}{1 + e^\theta} \right) + \frac{\delta - \theta}{1 + e^\theta} \right\}. \tag{2.8}$$

Using (2.7), subject to the existence of $\overline{\delta}$ and $\underline{\delta}$, the PRGM estimator of θ is given by

$$\delta_{PR}(X) = \frac{\frac{\overline{\delta}(X)}{1+e^{\overline{\delta}(X)}} - \frac{\delta(X)}{1+e^{\underline{\delta}(X)}} - \log \left\{ \frac{e^{\overline{\delta}(X)}}{e^{\underline{\delta}(X)}} \frac{1+e^{\underline{\delta}(X)}}{1+e^{\overline{\delta}(X)}} \right\}}{\frac{1}{1+e^{\overline{\delta}(X)}} - \frac{1}{1+e^{\underline{\delta}(X)}}}. \tag{2.9}$$

In Example 7, we obtain the PRGM estimator of p.

We now consider the PRGM estimation of θ under conjugate classes of prior distributions. For the exponential family (2.1) and a conjugate prior distribution

$$\pi_{\alpha,\lambda}(\theta) \propto \{\beta(\theta)\}^\alpha e^{-\theta \lambda}, \tag{2.10}$$

the posterior distribution is given by $\pi(\theta|x) \propto \{\beta(\theta)\}^{1+\alpha} e^{-(\lambda + r(x))\theta}$, and $\pi(\theta|x) = \pi_{\alpha+1, \lambda+r(x)}(\theta)$. Also, as established by [6], $\mathbb{E}[H(\theta)|x] = \frac{\lambda + r(x)}{\alpha + 1}$. Now, the Bayes estimator of θ under the intrinsic loss function (2.3) is obtained by (e.g., [3, 10, 13])

$$\delta_\pi(X) = H^{-1} \left(\frac{\lambda + r(X)}{\alpha + 1} \right). \tag{2.11}$$

Furthermore, the posterior regret for estimating θ with $\delta(x)$ is $\rho(\delta_\pi, \delta) = \log \frac{\beta(\delta_\pi(x))}{\beta(\delta(x))} + (\delta(x) - \delta_\pi(x)) \frac{\lambda + r(x)}{\alpha + 1}$. Now, suppose that the prior distribution $\pi_{\alpha,\lambda}$ belongs to the following class of conjugate prior distributions:

$$\Gamma = \{\pi_{\alpha,\lambda}(\theta) \; : \; \alpha \in [\alpha_1, \alpha_2], \; \lambda \in [\lambda_1, \lambda_2]\},$$

with suitable choices of $\alpha_1 < \alpha_2$ and $\lambda_1 < \lambda_2$ leading to proper posterior distributions for θ. A straightforward calculation shows that $H(\underline{\delta}(x)) = \frac{\lambda_2 + r(x)}{\alpha_1 + 1}$ and $H(\overline{\delta}(x)) = \frac{\lambda_1 + r(x)}{\alpha_2 + 1}$. Hence, we can state the following result.

Corollary 1. *Suppose* $H(t) = \beta'(t)/\beta(t)$. *The PRGM estimate of θ for the exponential family* (2.1) *under the intrinsic loss function* (2.3) *and in the class Γ of prior distributions is given by*

$$\delta_{PR}^{\Gamma}(x) = \frac{\frac{\lambda_1+r(x)}{\alpha_2+1}H^{-1}\left(\frac{\lambda_1+r(x)}{\alpha_2+1}\right) - \frac{\lambda_2+r(x)}{\alpha_1+1}H^{-1}\left(\frac{\lambda_2+r(x)}{\alpha_1+1}\right) - \log\left(\frac{\beta(H^{-1}(\frac{\lambda_1+r(x)}{\alpha_2+1}))}{\beta(H^{-1}(\frac{\lambda_2+r(x)}{\alpha_1+1}))}\right)}{\frac{\lambda_1+r(x)}{\alpha_2+1} - \frac{\lambda_2+r(x)}{\alpha_1+1}}.$$

(2.12)

Remark 1. One can also consider other classes of conjugate priors such as $\Gamma_1 = \{\pi_{\alpha,\lambda_0}(\theta) : \alpha \in [\alpha_1, \alpha_2], \ \lambda_0 \text{ is fixed}\}$ or $\Gamma_2 = \{\pi_{\alpha_0,\lambda}(\theta) : \alpha = \alpha_0 \text{ is fixed}, \lambda \in [\lambda_1, \lambda_2]\}$. The PRGM estimator of θ in Γ_1 or Γ_2 can be obtained using (2.12) and by letting $\lambda_1 = \lambda_2 = \lambda_0$ or $\alpha_1 = \alpha_2 = \alpha_0$, respectively.

Example 4. In Example 2, let $\pi_{\alpha,\lambda}(\theta) \propto \theta^{\alpha-1}e^{-\theta\lambda}$ with the posterior distribution $\pi(\theta|x) = \pi_{\alpha+1,\lambda+x}(\theta)$, and $\delta_{\pi}(x) = \frac{\alpha+1}{\lambda+x}$. Using (2.12), the PRGM estimator of θ under the Stein loss function $L(\theta,\delta) = \frac{\delta}{\theta} - \log\frac{\delta}{\theta} - 1$ in $\Gamma = \{\pi_{\alpha,\lambda}(\theta) : \alpha \in [\alpha_1, \alpha_2], \ \lambda \in [\lambda_1, \lambda_2]\}$, with $0 < \alpha_1 < \alpha_2$ and $0 < \lambda_1 < \lambda_2$ is given by

$$\delta_{PR}^{\Gamma}(X) = \log\left(\frac{\alpha_1+1}{\alpha_2+1}\frac{\lambda_1+X}{\lambda_2+X}\right) \Big/ \left(\frac{\lambda_1+X}{\alpha_2+1} - \frac{\lambda_2+X}{\alpha_1+1}\right).$$

In Γ_1, as defined in Remark 1, we have

$$\delta_{PR}^{\Gamma_1}(X) = \frac{(\alpha_1+1)(\alpha_2+1)}{\alpha_1-\alpha_2}\log\left(\frac{\alpha_1+1}{\alpha_2+1}\right)\frac{1}{\lambda_0+X}.$$

Similarly, in Γ_2, we have

$$\delta_{PR}^{\Gamma_2}(X) = \left(\frac{\alpha_0+1}{\lambda_2-\lambda_1}\right)\log\left(\frac{\lambda_2+X}{\lambda_1+X}\right).$$

3. Intrinsic PRGM estimation

In Section 2, we obtained the PRGM estimator of the natural parameter θ of the exponential family under the intrinsic loss function. In some applications, there may be interest in finding PRGM estimation of the original parameter of the underlying model rather than the natural parameter θ. Unfortunately, like many other estimators, PRGM estimators are not necessarily invariant under reparameterization. Although results of this nature, that are not invariant under reparameterization, can sometimes be interesting in theory, they tend to be less useful in practice. Indeed, it is difficult to sell to a practitioner that the PRGM estimator of $h(\theta)$ is not necessarily $h(\delta_{PR})$. In this section, we obtain PRGM estimators that are invariant under one-to-one smooth reparameterizations, hence the name intrinsic PRGM estimators.

For the exponential family (2.1), as opposed to the well known and commonly used conjugate prior (2.10), consider the following conjugate prior distribution

for θ

$$\pi_{\alpha,\lambda}^J(\theta) \propto \{\beta(\theta)\}^\alpha \, e^{-\lambda\theta} \, \sqrt{I_\theta(\theta)}, \tag{3.1}$$

where $I_\theta(\theta)$ is the Fisher information for θ. [7] introduced (3.1) and referred to it as the *Jeffrey's Conjugate Prior* (JCP). It is easy to see that the JCP is invariant under smooth reparameterizations, and the necessary conditions on α and λ in (3.1), leading to proper posterior distributions, do not depend on the choice of reparameterizations. The invariance property of JCP under any smooth and one-to-one reparameterization $\eta = h(\theta)$ can be shown by the following relationship

$$I_\eta(\eta) = I_\theta(h^{-1}(\eta)) \times \Big|\frac{dh^{-1}(\eta)}{d\eta}\Big|^2.$$

Remark 2. For the exponential family (2.1), since $I_\theta(\theta) = -H'(\theta)$, the JCP is given by $\pi_{\alpha,\lambda}^J(\theta) \propto \{\beta(\theta)\}^\alpha \, e^{-\lambda\theta} \sqrt{-H'(\theta)}$.

First, we give the following result.

Lemma 2. *Suppose δ_π^J is the Bayes estimator of the natural parameter θ of the exponential family (2.1) under the intrinsic loss function (2.2) with respect to the JCP distribution (3.1). For every one-to-one smooth transformation $h(\theta)$, the Bayes estimator of $h(\theta)$ is $h(\delta_\pi^J)$.*

Proof. The proof is similar to the proof of Lemma 6.2 of [13] and hence omitted. $\square$

Now, we state the main result of this section which can easily be proved using the invariance property of both the class of JCP distributions and the intrinsic loss functions under smooth reparameterization of θ.

Theorem 2. *Suppose $\delta_{IPR}^{\Gamma^J}(X)$ is the PRGM estimator of the unknown parameter θ for the exponential family (2.1) under the intrinsic loss function (2.3) with respect to a class Γ^J of JCP distributions for θ. Then, for any one-to-one smooth transformation $h(\theta)$, the PRGM estimator of $h(\theta)$ is $h(\delta_{IPR}^{\Gamma^J}(X))$.*

Proof. By definition, the PRGM estimator of $h(\theta)$ in the class Γ^J of JCP distributions is given by the solution of

$$\inf_\delta \sup_{\pi \in \Gamma^J} \rho(\delta_\pi^h, \delta) = \inf_\delta \sup_{\pi \in \Gamma^J} \left\{ \log \frac{\beta(\delta_\pi^h)}{\beta(\delta)} + (\delta - \delta_\pi^h)H(\delta_\pi^h) \right\},$$

where δ_π^h is the Bayes estimator of $h(\theta)$. Note that $\rho(\delta_\pi^h, \delta) = L(\delta_\pi^h, \delta)$ where L is defined in (2.3). Now, using the invariance property of L and Lemma 2, since $\delta_\pi^h = h(\delta_\pi)$, with δ_π being the Bayes estimator of θ, we have

$$\inf_\delta \sup_{\pi \in \Gamma^J} \rho(\delta_\pi^h, \delta) = \inf_\delta \sup_{\pi \in \Gamma^J} \rho(h(\delta_\pi), \delta)$$

$$= \inf_{t:h(t)=\delta} \sup_{\pi \in \Gamma^J} \rho(h(\delta_\pi), h(t))$$

$$= \inf_{t:h(t)=\delta} \sup_{\pi \in \Gamma^J} \rho(\delta_\pi, t).$$

Therefore, if $\delta_{IPR}^{\Gamma^J}(X)$ is the PRGM estimator of θ, i.e., $\delta_{IPR}^{\Gamma^J}$ minimizes (in t) $\sup_{\pi \in \Gamma^J} \rho(\delta_\pi, t)$, then, the transform $h(\delta_{IPR}^{\Gamma^J}(X))$ is the PRGM estimator of $h(\theta)$, that is, $h(\delta_{IPR}^{\Gamma^J})$ minimizes (in δ) $\sup_{\pi \in \Gamma^J} \rho(\delta_\pi^h, \delta)$ and this completes the proof. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad$ $\square$

Example 5. Suppose $X \sim Exp(\sigma)$ with $\sigma, x > 0$. In Example 2, we showed that the intrinsic loss for estimating $\theta = \sigma^{-1}$ by δ reduces to the Stein loss function

$$L(\theta, \delta) = \frac{\delta}{\theta} - \log \frac{\delta}{\theta} - 1.$$

Under the JCP distribution $\pi_{\alpha,\lambda}^J(\theta) \propto \theta^{\alpha-2} e^{-\theta\lambda}$, $\alpha > 1$, the posterior distribution is a $Gamma(\alpha, \frac{1}{\lambda+x})$ with $\pi^J(\theta|x) \propto \theta^{\alpha-1} e^{-(\lambda+x)\theta}$ which results in the Bayes estimator of θ as $\delta_\pi(X) = \frac{\alpha}{\lambda+X}$. Also, the intrinsic PRGM estimator of θ under $L(\theta, \delta)$ is given by

$$\delta_{IPR}^{\Gamma^J}(X) = \frac{\log \frac{1}{\overline{\delta}(X)} - \log \frac{1}{\underline{\delta}(X)}}{\frac{1}{\overline{\delta}(X)} - \frac{1}{\underline{\delta}(X)}}.$$

Now, for the estimation of $\eta = \sigma = \frac{1}{\theta}$ using $\tilde{\delta}$, it is easy to see that the Bayes estimator of η under the entropy loss function

$$L(\eta, \tilde{\delta}) = \frac{\eta}{\tilde{\delta}} - \log \frac{\eta}{\tilde{\delta}} - 1,$$

is given by $\tilde{\delta}_\pi(X) = \frac{\lambda+X}{\alpha} = \frac{1}{\delta_\pi(X)}$. To see this, note that $\pi^J(\eta) \propto \eta^{-\alpha} e^{-\lambda/\eta}$ with $\pi^J(\eta|x) \propto \eta^{-(\alpha+1)} e^{-\frac{\lambda+x}{\eta}}$ and $\tilde{\delta}_\pi(x) = \mathbb{E}[\eta|x]$. Also, the intrinsic PRGM estimator of η is given by

$$\tilde{\delta}_{IPR}^{\Gamma^J}(X) = \frac{\overline{\tilde{\delta}}(X) - \underline{\tilde{\delta}}(X)}{\log \overline{\tilde{\delta}}(X) - \log \underline{\tilde{\delta}}(X)}$$

$$= \frac{\frac{1}{\overline{\delta}(X)} - \frac{1}{\underline{\delta}(X)}}{\log \frac{1}{\overline{\delta}(X)} - \log \frac{1}{\underline{\delta}(X)}}$$

$$= \frac{1}{\delta_{IPR}^{\Gamma^J}(X)}.$$

For the PRGM estimation of θ under the entropy loss function and its application to record data analysis we refer to [12]. Similarly, if $\eta^* = -\frac{1}{a} \log \theta$, $a \neq 0$, then the intrinsic PRGM estimator of η^* under the LINEX loss function

$$L(\eta^*, \delta^*) = e^{a(\eta^* - \delta^*)} - a(\eta^* - \delta^*) - 1,$$

is given by

$$\delta^{*\Gamma^J}_{IPR}(X) = \underline{\delta^*}(X) + \frac{1}{a} \log \left\{ \frac{e^{a(\overline{\delta^*}(X) - \underline{\delta^*}(X))} - 1}{a(\overline{\delta^*}(X) - \underline{\delta^*}(X))} \right\}$$

$$= \frac{1}{a} \log \delta_{IPR}^{\Gamma^J}(X),$$

which is the PRGM estimator obtained in [5].

For the exponential family (2.1), suppose that the prior distribution belongs to the following class of JCP distributions:

$$\Gamma^J = \{\pi^J_{\alpha,\lambda}(\theta) \ : \ \alpha \in [\alpha_1, \alpha_2], \ \lambda \in [\lambda_1, \lambda_2]\}, \tag{3.2}$$

for suitable choices of $\alpha_1 < \alpha_2$ and $\lambda_1 < \lambda_2$. We continue with some applications of Theorem 2 under the above class of priors. Similar results can be obtained in other classes of JCP distributions (see Remark 1), which we do not present here. In view of Theorem 2, and to obtain an intrinsic PRGM estimator, the critical condition is that the elements of the underlying class of prior distributions are in the form of (3.1) and the underlying loss function is intrinsic. We observe that, in many cases (see Examples 6 and 7) intrinsic PRGM estimators under Γ^J can be obtained using the PRGM estimators under the usual class Γ of conjugate priors with modified values of α_is and λ_is in Γ, $i = 1, 2$. One can easily check that this will happen whenever the mean-value parameter is conjugate for the natural parameter in the sense of [11]. In the one-parameter case, a sufficient condition for this is that the exponential family have a quadratic variance function (see Section 3.3 of [11]).

Example 6. In Example 5, we showed that $\pi^J_{\alpha,\lambda}(\theta) \propto \theta^{\alpha-2} e^{-\theta\lambda}$ and $\delta_\pi(x) = \frac{\alpha}{\lambda+x}$. Since $\pi^J_{\alpha,\beta}(\theta|x)$ is equal to $\pi(\theta|x)$, the posterior distribution of θ, given the usual conjugate prior $\pi_{\alpha-1,\lambda+x}(\theta)$, the intrinsic PRGM estimator of θ under the Stein loss function and the class of JCP distributions can be obtained using the PRGM estimator of θ under the usual class of conjugate priors (as in Example 4), by replacing α_i with $\alpha_i - 1$, $i = 1, 2$. For example, the intrinsic PRGM estimator of θ in Γ^J with $0 < \alpha_1 < \alpha_2$ and $0 < \lambda_1 < \lambda_2$ is given by

$$\delta^{\Gamma^J}_{IPR}(X) = \log\left(\frac{\alpha_1}{\alpha_2}\frac{\lambda_1 + X}{\lambda_2 + X}\right) \bigg/ \left(\frac{\lambda_1 + X}{\alpha_2} - \frac{\lambda_2 + X}{\alpha_1}\right).$$

Let $\Gamma^J_1 = \{\pi^J_{\alpha,\beta}(\theta) : \alpha \in [\alpha_1, \alpha_2] \text{ and } \lambda = \lambda_0\}$. Then, the intrinsic PRGM estimator of θ in Γ^J_1, with $0 < \alpha_1 < \alpha_2$ and $\lambda_0 > 0$, is given by

$$\delta^{\Gamma^J_1}_{IPR}(X) = \left(\frac{\alpha_1 \, \alpha_2}{\alpha_1 - \alpha_2}\right) \log\left(\frac{\alpha_1}{\alpha_2}\right) \frac{1}{\lambda_0 + X}.$$

Similarly, in $\Gamma^J_2 = \{\pi^J_{\alpha,\beta}(\theta) : \alpha = \alpha_0 \text{ fixed and } \lambda \in [\lambda_1, \lambda_2]\}$, the intrinsic PRGM estimator of θ in Γ^J_2, with $\alpha_0 > 0$ and $0 < \lambda_1 < \lambda_2$, is given by

$$\delta^{\Gamma^J_2}_{IPR}(X) = \frac{\alpha_0}{\lambda_2 - \lambda_1} \log\left\{\frac{\lambda_2 + X}{\lambda_1 + X}\right\}.$$

Similar results can be obtained for estimating any smooth and one-to-one function of θ under corresponding class of JCP distributions.

Example 7 (Binomial Distribution)**.** In Example 3, we showed that pmf of X can be written as $f(x|\theta) = \binom{n}{x}(\frac{e^\theta}{1+e^\theta})^n e^{-x\theta}$ with $\theta = \log(\frac{1-p}{p})$. Here $I_\theta(\theta) = \frac{e^\theta}{(1+e^\theta)^2}$ and the JCP for θ is obtained as $\pi^J_{\alpha,\lambda}(\theta) \propto (\frac{e^\theta}{1+e^\theta})^\alpha e^{-\lambda\theta}\frac{e^{\theta/2}}{1+e^\theta}$. This results

in the posterior distribution $\pi^J(\theta|x) \propto (\frac{e^\theta}{1+e^\theta})^{\alpha+n+1}e^{-(x+\lambda+\frac{1}{2})\theta}$. Since $\pi^J(\theta|x)$ is equal to $\pi(\theta|x)$, the posterior distribution of θ, given the usual conjugate prior $\pi_{\alpha+1,\lambda+\frac{1}{2}}(\theta)$, the intrinsic PRGM estimator $\delta^{\Gamma^J}_{IPR}(X)$ of θ in Γ^J can be obtained using (2.9) and by replacing α_i and λ_i with α_i+1 and $\lambda_i+\frac{1}{2}$, $i=1,2$, respectively. Also, the intrinsic PRGM estimator of $p = \frac{1}{1+e^\theta}$ under the loss function

$$L(p, \tilde{\delta}) = p \log\left(\frac{p}{\tilde{\delta}}\right) + (1-p) \log\left(\frac{1-p}{1-\tilde{\delta}}\right),$$

is given by $\tilde{\delta}^{\Gamma^J}_{IPR}(X) = \{1 + e^{\delta^{\Gamma^J}_{IPR}(X)}\}^{-1}$.

Remark 3. It is possible to generalize the results of Sections 2 and 3 to another class of intrinsic loss functions using the *reverse Kullback-Leibler* loss, which is defined by

$$L_r(\theta, \delta) = \mathbb{E}_\delta\left[\log\left(\frac{f(X|\theta)}{f(X|\delta)}\right)\right] = \int_\chi \log\left(\frac{f(x|\theta)}{f(x|\delta)}\right) f(x|\delta)\, d\nu(x). \qquad (3.3)$$

Using (2.1), the loss function (3.3) reduces to the following class of intrinsic loss functions

$$L_r(\theta, \delta) = \log\left(\frac{\beta(\theta)}{\beta(\delta)}\right) + (\delta - \theta)\frac{\beta'(\delta)}{\beta(\delta)},$$

which includes many interesting loss functions (such as the Stein loss function, etc.) as its special case.

4. PRGM, Intrinsic PRGM and Bayes estimators

In this section, we provide some general results concerning the Bayesianity of the PRGM and intrinsic PRGM estimators of θ for the exponential family distribution (2.1) under the intrinsic loss function (2.3) with respect to priors in the underlying class of prior distributions. The results are only presented for PRGM estimators of θ, but they can also be used for intrinsic PRGM estimators by simple modifications. Our framework in this section closely resembles the one introduced by [15], who considered similar problem for the quadratic loss function. Results of this nature are also obtained by [16] under the quadratic loss function for the binomial distribution. Several of the following preliminary results are reported here for the sake of completeness. The idea is to check the continuity of the underlying Bayes estimator with respect to the prior. Similar to [15] we study two cases, when (a) the class of prior distributions is convex, or (b) the underlying class of prior distributions depends on a hyper-parameter belonging to a connected set.

First, consider the situation where the class Γ of priors is convex. That is, if $\pi_0, \pi_1 \in \Gamma$, then $\pi_t = t\pi_0 + (1-t)\pi_1$ belongs to Γ, for any $t \in [0,1]$. Suppose that X is a random variable whose density belongs to the family of distributions (2.1). Let $\psi(t) = H(\delta_{\pi_t}(x))$ which is a decreasing function of δ_{π_t} for any $t \in [0,1]$.

In the next lemma we show that $\psi(t)$ is a continuous function in its domain $t \in [0, 1]$.

Lemma 3. *Suppose $\psi(t)$, the posterior expectation of $H(\theta) = \frac{\beta'(\theta)}{\beta(\theta)}$ when $\pi_t = t\pi_0 + (1-t)\pi_1$, $t \in [0, 1]$, is finite. Then, $\psi(t)$ is continuous in t, $t \in [0, 1]$.*

Now, we use the continuity of $\psi(t)$ to prove that, under the conditions of Lemma 3, the PRGM estimator δ_{PR} is Bayes if the class of priors is convex.

Proposition 1. *Suppose Γ is a convex class of prior distributions on the unknown parameter θ of the exponential family of distributions (2.1). Then, there exists a prior distribution $\pi \in \Gamma$ such that $\delta_{PR} = \delta_\pi$, where δ_{PR} is defined in (2.7).*

A shortcoming of the result in Theorem 1 is that it is not applicable to the cases where the class of prior distributions depends on a hyper-parameter whose range is connected. For this case, we need Lemma 4 and Proposition 2 which are simple extensions of Lemma 3.2 and Proposition 3.2 of [15]. The proofs of Lemma 4 and Proposition 2 are essentially similar to the proofs of Lemma 3.2 and Proposition 3.2 of [15] and hence they are omitted. Let

$$\psi(\pi) = \frac{\int_\Theta H(\theta) f(x|\theta) \pi(\theta) d\theta}{\int_\Theta f(x|\theta) \pi(\theta) d\theta} = \frac{r(\pi)}{s(\pi)}. \tag{4.1}$$

Consider $d(\pi, \pi') = \sup_\Theta |\pi(\theta) - \pi'(\theta)|$ to be the usual l_∞ distance between prior densities π and π', where $H(t) = \beta'(t)/\beta(t)$ is defined as before.

Lemma 4. *Suppose that $\int_\Theta |H(\theta)| f(x|\theta) d\theta$ exists and it is finite. Then, $\psi(\pi)$ is continuous in π, in the topology generated by the l_∞ distance.*

Proposition 2. *Let $\Gamma = \{\pi_\alpha : \alpha \in \Lambda\}$, where Λ is a connected set and π_α's are densities. Under the conditions of Lemma 4 and the assumption that $\alpha_n \to \alpha$ implies $d(\pi_{\alpha_n}, \pi_\alpha) \to 0$, there exists a prior distribution $\pi \in \Gamma$ such that $\delta_{PR} = \delta_\pi$, that is, the PRGM estimator (2.7) is Bayes.*

In the following lemma, we provide a sufficient condition under which the PRGM (or intrinsic PRGM) estimator is Bayes with respect to the same prior in the underlying class of prior distribution, regardless of the observed value of x.

Lemma 5. *Let $\Gamma = \{\pi_\alpha : \alpha \in [\alpha_1, \alpha_2]\}$ be the class of prior distributions. Suppose the Bayes estimator $\Psi(\alpha, x) = H^{-1}\{\mathbb{E}[H(\theta)|x]\}$ is a differentiable function of the hyper-parameter α and the observed value x. Assume that we are under the conditions of Proposition 2. If*

$$\frac{\partial}{\partial x} \Psi(\alpha, x) = \frac{\partial}{\partial x} \left\{ \frac{\Psi(\alpha_1, x) H(\Psi(\alpha_1, x)) - \Psi(\alpha_2, x) H(\Psi(\alpha_2, x)) - \log \frac{\beta(\Psi(\alpha_1, x))}{\beta(\Psi(\alpha_2, x))}}{H(\Psi(\alpha_1, x)) - H(\Psi(\alpha_2, x))} \right\}, \tag{4.2}$$

has a constant solution in α, then there is a data independent prior $\pi_\alpha \in \Gamma$ resulting in the PRGM estimate as the Bayes estimate of the natural parameter θ of the exponential family (2.1) *under the intrinsic loss function* (2.2).

Proof. Under the conditions of Proposition 2, there exists a solution $\alpha(x)$ such that the PRGM estimator (2.7) is Bayes with respect to the prior $\pi_{\alpha(x)} \in \Gamma$ under the intrinsic loss function (2.2). That is,

$$\Psi(\alpha(x), x) = \frac{\Psi(\alpha_1, x)\, H(\Psi(\alpha_1, x)) - \Psi(\alpha_2, x)\, H(\Psi(\alpha_2, x)) - \log \frac{\beta(\Psi(\alpha_1, x))}{\beta(\Psi(\alpha_2, x))}}{H(\Psi(\alpha_1, x)) - H(\Psi(\alpha_2, x))}.$$

Now, differentiating the equation with respect to x leads to

$$\frac{\partial}{\partial \alpha} \Psi(\alpha(x), x) \frac{d\alpha(x)}{dx} + \frac{\partial}{\partial x} \Psi(\alpha(x), x)$$
$$= \frac{\partial}{\partial x} \left\{ \frac{\Psi(\alpha_1, x)\, H(\Psi(\alpha_1, x)) - \Psi(\alpha_2, x)\, H(\Psi(\alpha_2, x)) - \log \frac{\beta(\Psi(\alpha_1, x))}{\beta(\Psi(\alpha_2, x))}}{H(\Psi(\alpha_1, x)) - H(\Psi(\alpha_2, x))} \right\}.$$

If $\alpha(x)$ is data independent, i.e., $\alpha(x) = \alpha$, then $\frac{d\alpha(x)}{dx} = 0$. Now, the desired value for α is the constant solution to the equation (4.2) leading to a data independent prior for the PRGM estimator to be Bayes. □

Example 8. In Example 1, the condition (4.2) reduces to the condition (5) in Proposition 3.3 of Ríos Insua et al. (1995) as follows

$$2 \frac{\partial}{\partial x} \Psi(\alpha, x) = \frac{\partial}{\partial x} \Psi(\alpha_1, x) + \frac{\partial}{\partial x} \Psi(\alpha_2, x).$$

Now, consider the class $\Gamma = \{\pi_{\alpha, \lambda_0} : \alpha \in [\alpha_1, \alpha_2], \lambda_0 \text{ is fixed}\}$ of conjugate priors where π_{α, λ_0} is given by (2.10) with $\theta = \mu$ and $\beta(\theta) = e^{-\theta^2/2}$. Here, the Bayes estimator of θ is given by $\Psi(\alpha, X) = \delta_{\pi_{\alpha, \lambda}}(X) = \frac{X - \lambda_0}{\alpha + 1}$. It is easy to see that, the PRGM estimator of θ given by

$$\delta_{PR}(X) = \frac{1}{2} \left\{ \frac{X - \lambda_0}{\alpha_1 + 1} + \frac{X - \lambda_0}{\alpha_2 + 1} \right\},$$

is Bayes with respect to the data independent prior $\pi_{\alpha^*, \lambda_0} \in \Gamma$ where α^* is the solution to the following equation

$$\frac{2}{\alpha^* + 1} = \frac{1}{\alpha_1 + 1} + \frac{1}{\alpha_2 + 1}.$$

That is, $\alpha^* = \frac{\alpha_1 + \alpha_2 + 2\alpha_1 \alpha_2}{\alpha_1 + \alpha_2 + 2} \in [\alpha_1, \alpha_2]$ and $\delta_{PR}(X) = \frac{X - \lambda_0}{\alpha^* + 1} = \delta_{\pi_{\alpha^*, \lambda}}(X)$.

Example 9. In Example 4, the condition (4.2) reduces to

$$\frac{\partial}{\partial x} \Psi(\alpha, x) = \frac{\partial}{\partial x} \left\{ \frac{\log \frac{1}{\Psi(\alpha_1, x)} - \log \frac{1}{\Psi(\alpha_2, x)}}{\frac{1}{\Psi(\alpha_1, x)} - \frac{1}{\Psi(\alpha_2, x)}} \right\}.$$

Now, consider the class $\Gamma_1 = \{\pi_{\alpha,\lambda_0}(\theta) : \alpha \in [\alpha_1, \alpha_2], \ \lambda_0 \text{ is fixed}\}$ of conjugate priors on θ. Here, the Bayes estimator of θ with respect to the prior $\pi_{\alpha,\lambda_0}(\theta)$ is $\Psi(\alpha, X) = \delta_{\pi_{\alpha,\lambda_0}}(X) = \frac{\alpha+1}{\lambda_0+X}$. The PRGM estimator of θ is then Bayes with respect to a data independent prior $\pi_{\alpha^*,\lambda_0}(\theta) \in \Gamma_1$, if there exists a data independent solution α^* to the equation

$$-\frac{\alpha^* + 1}{(\lambda_0 + X)^2} = -\log\left(\frac{\alpha_1 + 1}{\alpha_2 + 1}\right) \frac{(\alpha_1 + 1)(\alpha_2 + 1)}{\alpha_1 - \alpha_2} \frac{1}{(\lambda_0 + X)^2}.$$

A straightforward calculation shows that

$$\alpha^* = \frac{(\alpha_1 + 1)(\alpha_2 + 1)}{\alpha_1 - \alpha_2} \log\left(\frac{\alpha_1 + 1}{\alpha_2 + 1}\right) - 1 \in [\alpha_1, \alpha_2].$$

Therefore, the PRGM estimator of θ under the Stein loss function can be obtained as the Bayes estimator of θ with respect to the prior distribution $\pi_{\alpha^*,\lambda_0}(\theta) \in \Gamma_1$ as follows

$$\delta_{PR}^{\Gamma_1}(X) = \frac{(\alpha_1 + 1)(\alpha_2 + 1)}{\alpha_1 - \alpha_2} \log\left(\frac{\alpha_1 + 1}{\alpha_2 + 1}\right) \frac{1}{\lambda_0 + X} = \frac{\alpha^* + 1}{\lambda_0 + X} = \delta_{\pi_{\alpha^*,\lambda_0}}(X).$$

Similarly, in Example 6, one can easily show that the intrinsic PRGM estimator $\delta_{IPR}^{\Gamma_1^J}(X)$ is the Bayes estimator of θ under the Stein loss function with respect to the prior distribution $\pi_{\alpha^{**},\lambda_0}^J \in \Gamma_1^J$, when $\alpha^{**} = \frac{\alpha_1 \alpha_2}{\alpha_1 - \alpha_2} \log(\frac{\alpha_1}{\alpha_2})$. Note that $1/\alpha^{**}$ is the logarithmic mean of $1/\alpha_1$ and $1/\alpha_2$, and $\alpha^{**} \in [\alpha_1, \alpha_2]$.

5. PRGM and intrinsic PRGM estimation in multi-parameter case

It is straightforward to generalize the results of Sections 2 and 3 to the multi-parameter exponential family of distributions. In this section, we study the PRGM and intrinsic PRGM estimation of the unknown parameters of a multivariate normal distribution under intrinsic loss functions as two examples of such problems. For the PRGM estimation of the regression coefficients in the canonical normal regression model see [15]. For the case studied in [15], similar to Example 11 (below), one can easily check that the PRGM and intrinsic PRGM estimators of the regression coefficients are equivalent.

Example 10. Suppose $\underset{\sim}{\mathbf{X}} = (\mathbf{X}_1, \ldots, \mathbf{X}_n)$ is a sample of size n from a p-variate normal distribution $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with the joint pdf

$$f(\underset{\sim}{\mathbf{x}}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = (2\pi)^{-\frac{np}{2}} |\boldsymbol{\Sigma}|^{-\frac{n}{2}} \exp\{-\frac{1}{2}\sum_{i=1}^{n}(\mathbf{x}_i - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}_i - \boldsymbol{\mu})\}, \qquad (5.1)$$

where $\boldsymbol{\mu}$ is the p-dimensional vector of the population means, $\boldsymbol{\Sigma}$ is a $p \times p$ nonsingular covariance matrix and $|\boldsymbol{\Sigma}|$ is the determinant of $\boldsymbol{\Sigma}$. Suppose $\boldsymbol{\mu} = \boldsymbol{\mu}_0$ is known and $\boldsymbol{\Sigma}$ is unknown. It is easy to see that the intrinsic loss with the Kullback-Leibler distance for estimating $\boldsymbol{\Sigma}$ is given by

$$L_1(\boldsymbol{\Sigma}, \widehat{\boldsymbol{\Sigma}}) \propto tr\left(\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}\right) - \log|\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}| - p. \qquad (5.2)$$

Consider the conjugate class of prior distributions $\Gamma_1 = \{\pi_{\alpha,\mathbf{V}}(\boldsymbol{\Sigma}^{-1}) : \alpha_0 \leq \alpha \leq \alpha_1, \mathbf{V}_0 \leq \mathbf{V} \leq \mathbf{V}_1\}$, where

$$\pi_{\alpha,\mathbf{V}}(\boldsymbol{\Sigma}^{-1}) \propto |\boldsymbol{\Sigma}|^{-\frac{1}{2}(\alpha-2)} \exp\{-\frac{1}{2}tr(\boldsymbol{\Sigma}^{-1}\mathbf{V})\},$$

$\mathbf{V}$ is a $p \times p$ positive definite and symmetric matrix, $\mathbf{V}_0 \leq \mathbf{V}_1$ whenever $\mathbf{V}_1 - \mathbf{V}_0$ is nonnegative definite and $\alpha_0, \alpha_1 > 0$ are scalars. In other word, we assume that $\boldsymbol{\Sigma}^{-1}$ is distributed according to the Wishart distribution, $W(\alpha + p - 1, \mathbf{V}^{-1})$. It is easy to show that the posterior distribution of $\boldsymbol{\Sigma}^{-1}$ given $\underset{\sim}{\mathbf{x}}$ is $W(n + \alpha + p - 1, (\mathbf{U}(\underset{\sim}{\mathbf{x}}) + \mathbf{V})^{-1})$ where $\mathbf{U}(\underset{\sim}{\mathbf{x}}) = \sum_{i=1}^{n}(\mathbf{x}_i - \boldsymbol{\mu}_0)'(\mathbf{x}_i - \boldsymbol{\mu}_0)$. Now, the Bayes estimator of $\boldsymbol{\Sigma}$ with respect to $\pi_{\alpha,\mathbf{V}}$ and under the loss function (5.2) is given by

$$\widehat{\boldsymbol{\Sigma}}_{\alpha,\mathbf{V}}(\underset{\sim}{\mathbf{x}}) = \mathbb{E}\left[\boldsymbol{\Sigma}|\underset{\sim}{\mathbf{x}}\right] = \frac{\mathbf{U}(\underset{\sim}{\mathbf{x}}) + \mathbf{V}}{n + \alpha - 2}.$$

Let

$$\overline{\overline{\boldsymbol{\Sigma}}}(\underset{\sim}{\mathbf{x}}) = \inf_{\pi_{\alpha,\mathbf{V}} \in \Gamma_1} \widehat{\boldsymbol{\Sigma}}_{\alpha,\mathbf{V}}(\underset{\sim}{\mathbf{x}}) = \frac{\mathbf{U}(\underset{\sim}{\mathbf{x}}) + \mathbf{V}_0}{n + \alpha_1 - 2}$$

and

$$\underline{\widehat{\boldsymbol{\Sigma}}}(\underset{\sim}{\mathbf{x}}) = \sup_{\pi_{\alpha,\mathbf{V}} \in \Gamma_1} \widehat{\boldsymbol{\Sigma}}_{\alpha,\mathbf{V}}(\underset{\sim}{\mathbf{x}}) = \frac{\mathbf{U}(\underset{\sim}{\mathbf{x}}) + \mathbf{V}_1}{n + \alpha_0 - 2}.$$

The PRGM estimator of $\boldsymbol{\Sigma}$ under the loss function (5.2) is now given as the solution of

$$L_1(\overline{\overline{\boldsymbol{\Sigma}}}, \widehat{\boldsymbol{\Sigma}}_{PR}) = L_1(\underline{\widehat{\boldsymbol{\Sigma}}}, \widehat{\boldsymbol{\Sigma}}_{PR}), \tag{5.3}$$

or equivalently $tr(\widehat{\boldsymbol{\Sigma}}_{PR}^{-1}(\overline{\overline{\boldsymbol{\Sigma}}} - \underline{\widehat{\boldsymbol{\Sigma}}})) = \log|\overline{\overline{\boldsymbol{\Sigma}}}| - \log|\underline{\widehat{\boldsymbol{\Sigma}}}|$ in $\widehat{\boldsymbol{\Sigma}}_{PR}$. To obtain intrinsic PRGM estimator of $\boldsymbol{\Sigma}$, we first note that the Fisher information matrix of $\boldsymbol{\Sigma}$ is given by $I(\boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-(p+1)}$ and so the Jeffrey's conjugate prior distribution is given by

$$\pi_{\alpha,\mathbf{V}}^{J}(\boldsymbol{\Sigma}^{-1}) \propto |\boldsymbol{\Sigma}|^{-\frac{1}{2}(\alpha+p-3)} \exp(-\frac{1}{2}tr(\boldsymbol{\Sigma}^{-1}\mathbf{V})),$$

which is equivalent to a $\pi_{\alpha+p-1,\mathbf{V}}(\boldsymbol{\Sigma}^{-1}) \in \Gamma_1$. Here, the posterior distribution of $\boldsymbol{\Sigma}^{-1}$ given $\underset{\sim}{\mathbf{x}}$ is a $W(n + \alpha + 2p - 2, (\mathbf{U}(\underset{\sim}{\mathbf{x}}) + \mathbf{V})^{-1})$ and the intrinsic PRGM estimator of $\boldsymbol{\Sigma}$ is obtained by replacing α_i with $\alpha_i + p - 1$, $i = 0, 1$, in the PRGM estimator obtained from (5.3).

Example 11. In Example 10, suppose that $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_0$ is known and $\boldsymbol{\mu}$ is the unknown parameter of interest. The intrinsic loss function for estimating $\boldsymbol{\mu}$ by $\widehat{\boldsymbol{\mu}}$ is given by

$$L_2(\boldsymbol{\mu}, \widehat{\boldsymbol{\mu}}) \propto (\boldsymbol{\mu} - \widehat{\boldsymbol{\mu}})'\boldsymbol{\Sigma}_0^{-1}(\boldsymbol{\mu} - \widehat{\boldsymbol{\mu}}), \tag{5.4}$$

where the Bayes estimator of $\boldsymbol{\mu}$ is $\mathbb{E}[\boldsymbol{\mu}|\underset{\sim}{\mathbf{x}}]$. Consider the following class Γ_2 of conjugate prior distributions for $\boldsymbol{\mu}$

$$\Gamma_2 = \{\pi_{\boldsymbol{\theta},\mathbf{V}_0}(\boldsymbol{\mu}) = N_p(\boldsymbol{\theta}, \mathbf{V}_0) \mid \boldsymbol{\theta}_0 \leq \boldsymbol{\theta} \leq \boldsymbol{\theta}_1, \;\; \boldsymbol{\theta}_0, \boldsymbol{\theta}_1 \in \mathbb{R}^p\},$$

where $\mathbf{V}_0$ is a known positive definite covariance matrix (e.g., $\mathbf{V}_0 = \sigma_0^2 \, \mathbb{I}_{p \times p}$, with known $\sigma_0 > 0$). It is easy to see that the posterior distribution of $\boldsymbol{\mu}$ given $\underset{\sim}{\mathbf{x}}$ is a p-variate normal distribution $N_p(\boldsymbol{\mu}_{\boldsymbol{\theta}, \mathbf{V}_0}(\underset{\sim}{\mathbf{x}}), \mathbf{V}^*)$ with

$$\boldsymbol{\mu}_{\boldsymbol{\theta}, \mathbf{V}_0}(\underset{\sim}{\mathbf{x}}) = (n\boldsymbol{\Sigma}_0^{-1} + \mathbf{V}_0^{-1})^{-1}(n\boldsymbol{\Sigma}_0^{-1}\bar{\mathbf{x}} + \mathbf{V}_0^{-1}\boldsymbol{\theta}) \quad \text{and} \quad \mathbf{V}^* = (n\boldsymbol{\Sigma}_0^{-1} + \mathbf{V}_0^{-1})^{-1}.$$

Now, the PRGM estimation of $\boldsymbol{\mu}$ under L_2 with respect to the class Γ_2 of prior distribution is given by

$$\begin{aligned}
\widehat{\boldsymbol{\mu}}_{PR}(\underset{\sim}{\mathbf{x}}) &= \frac{1}{2}\left(\inf_{\pi_{\boldsymbol{\theta}, \mathbf{V}_0} \in \Gamma_2} \boldsymbol{\mu}_{\boldsymbol{\theta}, \mathbf{V}_0}(\underset{\sim}{\mathbf{x}}) + \sup_{\pi_{\boldsymbol{\theta}, \mathbf{V}_0} \in \Gamma_2} \boldsymbol{\mu}_{\boldsymbol{\theta}, \mathbf{V}_0}(\underset{\sim}{\mathbf{x}}) \right) \\
&= \frac{1}{2}\left(\boldsymbol{\mu}_{\boldsymbol{\theta}_0, \mathbf{V}_0}(\underset{\sim}{\mathbf{x}}) + \boldsymbol{\mu}_{\boldsymbol{\theta}_1, \mathbf{V}_0}(\underset{\sim}{\mathbf{x}}) \right) \\
&= (n\boldsymbol{\Sigma}_0^{-1} + \mathbf{V}_0^{-1})^{-1}\left(n\boldsymbol{\Sigma}_0^{-1}\bar{\mathbf{x}} + \mathbf{V}_0^{-1}(\frac{\boldsymbol{\theta}_0 + \boldsymbol{\theta}_1}{2}) \right).
\end{aligned}$$

Note that, in this case, since the Jeffrey's conjugate prior distribution for $\boldsymbol{\mu}$ is the same as the usual conjugate prior (the Fisher information $I(\boldsymbol{\mu})$ is a constant), then $\widehat{\boldsymbol{\mu}}_{PR}(\underset{\sim}{\mathbf{x}})$ is also an intrinsic PRGM estimator of $\boldsymbol{\mu}$ under the loss function L_2 and within the class Γ_2 of prior distributions.

In Examples 10 and 11, when both $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are unknown, the intrinsic loss with the Kullback-Leibler distance for estimating $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ is given by

$$L_3(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\Sigma}}) \propto tr\left(\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}\right) + (\boldsymbol{\mu} - \widehat{\boldsymbol{\mu}})'\widehat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{\mu} - \widehat{\boldsymbol{\mu}}) - p - \log|\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}|. \quad (5.5)$$

We note that the PRGM results of Sections 2 and 3 are not directly applicable in this case. However, one can obtain the PRGM and intrinsic PRGM estimators of the nature parameters $\boldsymbol{\theta}_1 = \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$ and $\boldsymbol{\theta}_2 = -\frac{1}{2}\boldsymbol{\Sigma}^{-1}$ by extending Theorems 1 and 2 to the multi-parameter exponential family of distributions

$$f(x|\boldsymbol{\theta}) = \beta(\boldsymbol{\theta})\, t(x)\, e^{-\sum_{i=1}^{p} \boldsymbol{\theta}_i r_i(x)},$$

under the corresponding intrinsic loss function, when $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_p)$ is the vector of natural parameters.

6. Concluding remarks

Invariant estimators are usually demanding in practice. In this paper, we have provided general results concerning the PRGM estimation of the natural parameter of the one-parameter exponential family of distributions under intrinsic loss functions. The PRGM estimators are shown to be invariant to one-to-one smooth reparameterizations under intrinsic loss functions and the class of Jeffrey's conjugate prior distributions. Moreover, when the class of priors are convex or dependant on a hyper-parameter belonging to a connected set, we show that the obtained PRGM estimators are Bayes with respect to prior distributions in the underlying class of priors. Examples are provided to show how the PRGM and intrinsic PRGM results can be extended to a multi-parameter exponential family of distributions.

Acknowledgements

References

[1] BERGER, J.O. (1994). *An overview of robust Bayesian analysis.* Test, **3**, 5–124. MR1293110

[2] BERNARDO, J.M. (2011). *Integrated objective Bayesian estimation and hypothesis testing (with discussion).* In Bayesian Analysis 9 (eds. J.M. Bernardo, M.J. Bayarri, J.O. Berger, A.P. Dawid, D. Heckerman, A.F.M. Smith and M. West). Oxford University Press, 1–68.

[3] BERNARDO, J.M. and SMITH, A.F.M. (1994). *Bayesian Theory.* Chichester: Wiley. MR1274699

[4] BORATYŃSKA, A. (2002). *Posterior regret gamma-minimax estimation in a normal model with asymmetric loss function.* Acta Mathematicae, **29**, 7–13. MR1906423

[5] Boratyńska, A. (2006). *Robust Bayesian prediction with asymmetric loss function in Poisson model of insurance risk.* Acta Universitatis Lodziensis, Folia Oeconomica, **196**, 123–138.

[6] DIACONIS, P. and YLVISAKER, D. (1979). *Conjugate priors for exponential families.* Annals of Statistics, **7**, 269–281. MR0520238

[7] DRUILHET, P. and POMMERET, D. (2012). *Invariant conjugate analysis for exponential families.* Bayesian Analysis, **7**, 235–248. MR3000019

[8] GELMAN, A. (2004). *Parameterization and Bayesian modelling.* Journal of the American Statistical Association, **99**, 537–545. MR2109315

[9] GÓMEZ-DÉNIZ, E. (2009). *Some Bayesian credibility premiums obtained by using posterior regret gamma-minimax methodology.* Bayesian Analysis, **4**, 223–242. MR2507362

[10] Gutiérrez-Peña, E. (1992). *Expected Logarithmic Divergence for Exponential Families.* In Bayesian Statistics 4 (J.M. Bernardo, J.O. Berger, A.P. Dawid y A.F.M. Smith, eds.) Oxford: University Press, 669–674. MR1380301

[11] GUTIÉRREZ-PEÑA, E. and SMITH, A.F.M. (1995). *Conjugate Parametrizations for Natural Exponential Families.* Journal of the American Statistical Association, **90**, 1347–1356. MR1379477

[12] JAFARI JOZANI, M., and PARSIAN, A. (2008). *Posterior regret Γ-minimax estimation and prediction based on k-record data under entropy loss function.* Communications in Statistics: Theory and Methods, 37, **14**, 2202–2212. MR2526674

[13] ROBERT, C.P. (1996). *Intrinsic loss functions.* Theory and Decision, **40**, 192–214. MR1385186

[14] RÍOS INSUA, D., and RUGGERI, F. (2000). *Robust Bayesian analysis.* Lecture Notes in Statistics **152**, Springer-Verlag, New York. MR1795206

[15] RÍOS INSUA, D., RUGGERI, F., and VIDAKOVIC, B. (1995). *Some results on posterior regret* Γ-*minimax estimation.* Statistics & Decisions, **13**, 315–351. MR1380252

[16] ZEN, M., and DASGUPTA, A. (1993). *Estimating a binomial parameter: Is robust Bayes real Bayes?* Statistics & Decisions, **11**, 37–60. MR1220436