Australian Journal of Basic and Applied Sciences, 3(4): 3173-3178, 2009
ISSN 1991-8178

On the Behavior of Bayes' Estimators, under Squared-error Loss Function

[1]Amir Teimour PayandehNajafabadi

Department of Mathematical Sciences, Shahid Beheshti University, G.C., Tehran, IRAN

Abstract: The Bayes estimator updates our belief about a parameter μ in the exist of new evidence about θ which summarized in prior (θ)It is well known that the behavior of a Bayes estimator varies from prior selection. In many cases, we would like to shrink (or stretch) a given Bayes estimator such that a desire condition meets by its corresponding Bayes estimator. In this paper, we consider estimation of location parameter μ under the square-error loss function. Then (i) for given scale parameter we derive a sufficient condition on prior 2 such that 2 shrinks (or stretches) with resect to given$_1$ (ii) for misspecified scale parameter we also similar condition on prior 2 such that 2 shrinks (or stretches) with resect to given$_1$.

Key words: Bayes estimator, prior distribution, monotone increasing likelihood ratio property, unimodal and symmetric density functions, square-error loss function, minimaxity, admissibility.

INTRODUCTION

Since the Bayes estimator updats our belief about a parameter μ in the exist of new evidence about μ which summarized in prior (θ)It have been received considerable attention form almost all field of sciences. In the lack of new evidence about θ many statisticians recommend using of the noninformative priors which they believe the noninformative Bayes estimator still preforms better than (from several point of views, such as admissibility, minimaxity, and etc) other nonBayesian estimators. But, it is important to keep in mind that when using noninformative priors (or even informative), it is necessary to perform sensitivity analysis to ensure that the prior beliefs are not unduly affecting the results. From this point of view, in the literature, there is a bulk of papers dealing with robustness or sensitivity of Bayes estimator with respect to the likelihood, the prior, and the loss function. A good review can be found in Berger (1984, 1990, 1994). Also, a considerable amount of the literature focussed on local sensitivity, study sensitivity of Bayes estimator with respect to an infinitesimal changes in the prior. Sensitivity of inferences to the choice of model, has been examined by Sivaganesan (1993), Gustafson, Srinivasan, and Wasserman (1994), Gustafson (1996), Basu (1994), and Tsou & Royall (1995). Chao (1970), Fabius (1964) and Abraham & Cadre (2004) studied asymptotic behavior of Bayes estimator with resect to prior selection. But surprisingly, as long as we know, a few studies has been done on behavior of Bayes estimator, in general not only for asymptotic case, in choice of prior. This paper tries to do so.

Suppose a random variable X has unimodal and symmetric density function proportional to

$$f_0(x-\theta,\sigma) \propto \frac{1}{\sigma}\exp\{-h(\frac{x-\theta}{\sigma})\}, \tag{1}$$

Where h is a function in class of functions

$H^* := \{h\colon h$ increasing and convex and $h$concave$\}$

Moreover, suppose that we are interested to estimate an unknown parameter μ under square-error loss function

$$L_2(\delta,\theta,\sigma) := (\frac{\delta(x)-\theta}{\sigma})^2,$$

Corresponding Author: Amir Teimour Payandeh Najafabadi, Department of Mathematical Sciences, Shahid Beheshti University, G. C., Tehran, IRAN
E-mail: amirtpayandeh@sbu. ac. ir

based on an observation XInteresting question pertain to the frequentist performance of Bayesian estimators, such as the determination or description of Bayesian estimators that shrink or stretch upon a benchmark Bayes estimator such as fully uniform Bayes estimator, i.e., a Bayes estimator with respect to a noninformative prior, $FU(\theta) = 1$for all μAn importance of the study may be viewed in two directions: (i): As many authors such as Moors (1981, 1985), Charras & van Eeden (1991a, 1991b, 1992), among others, pointed out that under square-error loss function the maximum likelihood estimator (mle) is not an admissible one, whenever X has density function (1) and parameter space restricted in some senses. So if one obtain a dominator Bayes estimator for the mle all of shrunk Bayes estimator will dominate the mle as well. (ii): In many situations, we are interested on such estimators which shrink (or stretch) a given estimator. For instance, consider an insurance firm whose uses the Bayesian producer to calculate: (a) the mount of its reimbursement to policyholders who has some claims; (b) the premium of policyholders based upon their prior information. The insurance company interests on such priors which lead to shrunk and stretched estimators, respectively for (a) and (b), compare to an usual estimator.

Section 2 collects some useful elements which are required for other sections. Section 3 considers the problem for given scale parameter then derives some conditions on the prior distribution, such that their corresponding Bayes estimators shrink or stretch a given Bayes estimator. Namely, we show that when is σ given, sufficient condition on π_2 such that $\delta_{\pi 2}$ shrinks with respect to $\delta_{\pi 1}$ is ratio $\pi_1(\theta)/\pi_2(\theta)$ be nondecreasing in $|\theta|$ Section 4 generalizes of results of section 3, whenever the scale parameter misspecified. We find that sufficient condition to shrink $\delta_{\pi 1}$ is to find prior $\pi_2(\theta, \sigma)$ such that $\pi_2(\theta, .)$ is symmetric about zero and ratio $\pi_1(\theta, .)/\pi_2(\theta, .)$ be nondecreasing in $|\theta|$ and $\pi_2(\,.,\sigma)/\pi_2(\,.\,,\sigma)$ nondecreasing in $1/\sigma$.

Preliminaries:

It is useful to recall that, family of densities function $\{p\theta(.) : \theta \in \Theta\}$ is said to have monotone likelihood ratio (mlr) in $T(.)$, such that for all $\theta_1 > \theta_2$, the densities $p\theta_{ii}(.)$, for $i = 1,2$, are distinct, and the ratio $p\theta_1(x)/p\theta_2(x)$ is a nondecreasing function of $T(x)$The following from Lehmann (1997) recalls an important property of a class of density functions which have the mlr property.

Lemma 1. *(Kline-Rubin's lemma) Suppose* $\{p\theta(.) : \theta \in \Theta\}$ *is a family of density functions with the mlr in x. Moreover, suppose that* $\psi(x)$ *is a nondecreasing function in xThen* $E_\theta(\psi\ X))$ *is nondecreasing function in* θ.

The following generalizes Lemma (1) to handel more complicated situation.

Lemma 2. *Suppose X and Y are two random variables with joint density function* $f_\theta(x, y)$*where for any fixed* θ $f_\theta(x, y)$*has the mlr property in x (or y) whenever y (or x) viewed as a parameter.*

i) *If ratio* $f_\theta(x, y_1)/f_\theta(x, y_2)$ *increasing in x for fixed* θ *and all* $y_1 > y_2$*Then, conditional density function of random variable Y given X = x has the mlr property in ywhen x viewed as a parameter;*

ii) *If ratio* $k(x,y,\theta_1\ \theta_2) := f_{\theta 1}(x, y)/f_{\theta 2}(x, y)$ *increasing in x for fixed y and all* $\theta_1 > \theta_2$ *Then, marginal density function of random variable X has the mlr property in x when* μ *observed as a parameter.*

Proof: For part (i), observe that for $x_1 > x_2$

$$\frac{f_\theta(y|x_1)}{f_\theta(y|x_2)} \propto \frac{f_\theta(x_1, y)}{f_\theta(x_2, y)}.$$

Desire result arrives from the assumption on joint density. For part (ii) observes that R

$$\frac{f_{\theta 1}(x)}{f_{\theta 2}(x)} = \frac{\int_{\mathbb{R}} f_{\theta 1}(x,y)dy}{\int_{\mathbb{R}} f_{\theta 2}(x,y)dy} = E(k(x,T,\theta_1,\theta_2)),$$

where T is a random variable with density function proportional to $_{f\theta 2}$ (x, t) which has the mlr property in t as x viewed as a parameter. Rest of proof follows from Lemma (1) along assumption on joint density function.□

The following is useful for the next section.

Lemma 3: *Suppose nonnegative and continuous random variables U has density function propor- tional to* $(x,u) := k(x,u)d\mu(\pi_i(u)),\ where\ k(x,y) := \left[f_0(x-u) + f_0(x+u)\right]I_{[0,\infty}(u)\ and\ \mu(\pi_i(u))$ *nondecreasing positive measure in uThen, density function of U has the mlr properties in u whenever x viewed as a parameter.*

Proof: From Lehmann (1997), it is sufficient (and necessary) to show that the mixed derivative $\frac{\partial^2}{\partial x \partial u} \ln k(x,u) \geq 0$, ›r all x and uRest of proof follows from Marchand, Ouassou, Payandeh, & Perron (2007).□

Study Behavior of Bayes Estimator with Respect to Prior Selection, under Given Scale Parameter:

Suppose X has density function with form (1) and location parameter θ is unknown. Moreover, suppose that the scale parameter σ is given, without loss of generality let $\sigma = 1$. This section develops conditions on prior distributions, such that their corresponding Bayes estimators shrink or stretch with respect to a given Bayes one. Following theorem as the first main result of this paper develops condition on the prior distribution, which its corresponding Bayes estimator shrinks with respect to a given Bayes estimator.

Theorem 1. : *Suppose* μ $(\pi_i(\theta))i = 1,2$, *are two positive and symmetric measures, about zero,where* $\frac{\mu(\pi_1(\theta))}{\mu(\pi_2(\theta))}$ *is an increasing function in* $|\delta_{\pi 1}(x)| \geq |\delta_{\pi 2}(x)|$ under square-error loss.

Proof: Since $\delta_{\pi i}(-x) = -\delta_{\pi i}(x)$ (invariant of model) it suffices to work with $x \geq 0$. From symmetricity of density and prior Bayes estimator with respect to $\pi i(\theta)$ may be represented as

$$\delta_{\pi i}(x) = \frac{\int_{-\infty}^{\infty} \theta f_0(x-\theta) d\mu(\pi_i(\theta))}{\int_{-\infty}^{\infty} \theta f_0(x-\theta) d\mu(\pi_i(\theta))}$$

$$= \frac{\int_0^{\infty} \theta\left[f_0(x-\theta) - f_0(x+\theta)\right] d\mu(\pi_i(\theta))}{\int_0^{\infty} \theta\left[f_0(x-\theta) + f_0(x+\theta)\right] d\mu(\pi_i(\theta))}.$$

given that the $d\mu(\pi_i(\theta,.))$ are symmetric about zero. Set $H_1(u) := [f_0(x\ u)\ f_0(x+u)][f_0(x\ u) + f_0(x + u)]$, so Bayes estimator will be $E_{\pi i}(UH_1(U))$, where U is a random variable with density function proportional to $[f_0(x\ u) + f_0(x + u)]\ I_{[0,\infty)}(u)du(\pi_i(u))$ for any fixed positive x. Density function of u has the mlr in u as x viewed as a parameter (see Lemma 3) ,because $\mu(\pi i(\theta))$ is nondecreasing in $|\theta|$ and $f_0(x-\theta)$ is the mlr, by using properties of the mlr family of distribution, it is sufficient to show that $uH_1(u)$ is nondecreasing in u for any fixed positive x. From the definition of $H_1(u)$ and $f_0(.)$, it is easy to observe that

$$uH_1(u) = u\tanh\left(\frac{h(x+u) - h(x-u)}{2}\right),$$

thus $uH_1(u)$ is nondecreasing in u, because x and u are positive so $h(x+u) > h(x-u)$. Consequently from properties of the mlr family of distribution

$$E_{\pi 1}(UH_1(U)) \geq E_{\pi 1}(UH_1(U))\ for x \geq 0$$

Which the desire result arrives.□

It would be worth to mention the following.

Remark 1. *Result of Theorem (1) is valid, whenever parameter space restricted into an interval* $\Theta = [a,b]$.

Now, we present our result for some important cases. One of the most popular Bayes estimator in statistical inference is fully uniform Bayes estimator, $\delta_{FU}(x)$, which obtained from the noninformative prior $\pi_{FU}(\theta) = 1$, for all μ in parameter space.Θ

Example 1. : *(Fully uniform prior) Suppose X has density from(1) and $\pi(\theta)$ is a symmetric prior about zero. Then Bayes estimator δ_π with respect to π, under square-error loss satisfies*

$$a) |\delta_\pi(x)| \geq |\delta_{FU}(x)|, \text{ wheneuer } \pi(\theta) \text{ nondecreasing in } |\theta|;$$
$$b) |\delta_{FU}(x)| \geq |\delta_\pi(x)|, \text{ wheneuer } \pi(\theta) \text{ nondecreasing in } |\theta|;$$

Marchand & Perron (2001) and Hartigan (2004) considered the problem of improving the mle of location parameter of $N(\theta, 1)$under square-error loss, whenever $|\theta| \leq m$. They showed that the δ_{FU} dominates the mle.[2] Part (b) of Example (1) along Remark (1) present a class of Bayes estimators which improve the mle. This observation also pointed out by Marchand & Perron (2001).

To find a minimax estimator, most of statisticians prefer to use two-point prior (or boundary uniform prior), i.e., $\pi_{BU}(\theta) = 05$for $\theta = \pm m$. The next example considers such prior.

Example 2. : *(Boundary uniform prior) Suppose X has density from(1). Then Bayes estimator with respect to the boundary uniform prior π_{BU} say δ_{BU} satisfies $|\delta_{BU}(x)| \geq |\delta_{FU}(x)|$.*

In an influential paper, Casella & Strawderman (1981) considered problem of finding a minimax estimator for θ based an observation X which sampled from $N(\theta\ 1)$and $\theta \in [a, b]$They established that the boundary uniform Bayes rule is minimax one, as long as $b\text{-}a \leq 2$. Example (2) along part (b) of Example (1) and Remark (1) suggest a class of minimax estimators for their problem as long as $b\text{-}a \leq 2$.

Efron(1996) introduced a class of priors, so-called Spike and slab priors, which gives a vital role to boundary points according to counting measure μC and mass of g to rest of parameter space according to Lebesgue measure μ_L Such kind of prior useful in some problems in medicine and genetic, for more see Ishwaran & Rao (2005). More precisely the Spike and slab prior defends as

$\pi^*(\theta) = \alpha\mu_C(\theta) + (1-\alpha)g(\theta)\mu_L(\theta), \text{ for } \theta \in [-m, m]$, where α is a given value in [0, 1] The next example explores behavior of Spike and slab Bayescompare to δ_{FU}.

Example 3. : *Suppose X has density from (1) and location parameter θ is restricted into an interval $[m, m]$ Moreover, suppose that g(.) in the Spike and slab prior $\delta_{\pi^*}(x) = \alpha\mu_C(\theta) + (1\text{-}\alpha)g(\theta)\mu_L(\theta)$ is a symmetric, about 0, decreasing, and $g(m) \geq \alpha/(1\text{-}\alpha)$. Then the Spike and slab Bayes estimators shrinks with respect to the fully uniform Bayes estimator (i.e.,· $|\delta_{\pi^*}| \leq |\delta_{FU}|$).*

[2]Indeed, they established the result for a p-variate normal distribution, which valid for univariate normal as well.

4. Study Behavior of Bayes Estimator with Respect to Prior Selection, under Misspecified Scale Parameter

For the estimation of positive normal mean based on $X \sim N(\theta, 1)$Katz (1961) showed admissibility and minimaxity of the fully uniform Bayes rule, δ_{FU} Maruyama & Iwasaki (2005) studied the robustness properties of these results to misspecified of the scale parameter σ^2 They showed that, when $\sigma^2 \neq 1$, δ_{FU} is minimax if and only if $\sigma^2 > 1$Concerning the admissibility when $\sigma^2 \neq 1$, they showed that δ_{FU} is admissible if and only if σ^{-2} is a positive integer. This is an example of extreme non-robustness of estimator with respect to misspecified scale parameter. This section studies derive some mild conditions on the scale parameter such that results of previous section are hold. The following represents more general version of Theorem (1)

Theorem 2. : *Suppose $\mu(\pi_i(\theta, \sigma))i = 1, 2$, are two positive measures, where $\mu(\pi_i(\theta,.)),i = 1,2$, are symmetric about zero; and $k(\theta)$ and $k(.,\sigma)$ are nondecreasing function in $|\theta|$ and* $1/\sigma$, respectively, where $k(\theta, \sigma) := \dfrac{\mu(\pi_1(\theta,\sigma))}{\mu(\pi_2(\theta,\sigma))}$. *Then Bayes estimators, with respect to i $i = 12$under square-error loss function, given by (2), satisfies* $|\delta_{\pi 1}(x)| \geq |\delta_{\pi 2}(x)|$.

Proof: Since $\delta_{\pi i}(-x) = -\delta_{\pi i}(x)$ (invariant of model) it suffices to work with the case $x \geq 0$. The Bayes estimator with respect to $\pi_i(\theta, \sigma)$ under square error loss may be represented as:

$$\delta_{\pi i}(x)=\frac{\int_0^\infty\int_{-\infty}^\infty \theta f_0\left(\frac{x-\theta}{\sigma}\right)/\sigma^3 d\mu(\pi_i(\theta,\sigma))}{\int_0^\infty\int_{-\infty}^\infty \theta f_0\left(\frac{x-\theta}{\sigma}\right)/\sigma^3 d\mu(\pi_i(\theta,\sigma))}$$

$$=\frac{\int_0^\infty\int_0^\infty \theta\left[f_0\left(\frac{x-\theta}{\sigma}\right)-f_0\left(\frac{x-\theta}{\sigma}\right)\right]/\sigma^3 d\mu(\pi_i(\theta,\sigma))}{\int_0^\infty\int_0^\infty \theta\left[f_0\left(\frac{x-\theta}{\sigma}\right)+f_0\left(\frac{x-\theta}{\sigma}\right)\right]/\sigma^3 d\mu(\pi_i(\theta,\sigma))}$$

given that $d\mu(\pi i(\theta,.))$s are symmetric about zero. Set $H_2(u, \upsilon) := \left[f_0\left(\frac{x-u}{\upsilon}\right)-f_0\left(\frac{x-u}{\upsilon}\right)\right]/\left[f_0\left(\frac{x-u}{\upsilon}\right)+f_0\left(\frac{x-u}{\upsilon}\right)\right]$

so the Bayes estimator will be $E_{\pi i}(UH_2(U.V))$ where U and V are two random variables with joint density function proportional to

$$\frac{1}{\upsilon^3}\left[f_0\left(\frac{x-u}{\upsilon}\right)+f_0\left(\frac{x+u}{\upsilon}\right)\right]I_{[0,\infty)}(u)I_{[0,\infty)}(\upsilon)d\mu(u,\upsilon)$$

for any fixed positive x. From the definition of $H(u, \upsilon)$ and $f_0(.)$, one can write

$$uH_2(u,\upsilon)=u\ \tanh\left(\frac{h\left(\frac{x+u}{\upsilon}\right)-h\left(\frac{x-u}{\upsilon}\right)}{2}\right).$$

Thus $uH_2(u,\upsilon)$ and $H_2(u, \upsilon)$ are nondecreasing in u and $1/\upsilon$, whenever the second argument is fixed. One other hand, from the Wald's identity,

$$\delta_{\pi i}(x)=E(UE(H_2(U,V)|U)).$$

Lemma (2) along Lemma (1) lead to the following, which completes the desire proof.

$$E_{\pi 1}(UH_2(U,V))\geq E_{\pi 2}(UH_2(U,V))\quad for\ x\ \geq 0.\ \square$$

An immediate consequence of previous theorem can be the following example.

Example 4. : *Suppose X has density function given by (1). The Bayes estimators under two priors* $\pi^{c}_{FU}(\theta,\sigma)=1/\sigma^2$ *and* $\pi^{c0}_{FU}(\theta,\sigma=1)=1$ *satisfy* $\left|\delta^{c}_{FU}(x)\right|\geq\left|\delta^{c}_{FU}(x)\right|$.

Above observation says minimaxity and admissibility of the fully uniform Bayes estimator, when the scale parameter is given, do not retain whenever the scale parameter misspecified. Indeed, this result confirmed by Maruyama & Iwasaki (2006) for positive normal mean.

ACKNOWLEDGEMENTS

The support of Shahid Beheshti University is gratefully acknowledged by author. The author thanks Professor Éric Marchand who brought my attention to the problem and also for his contractive suggestions which led to several improvements in the paper

REFERENCES

Abraham, C., B. Cadre, 2004. Asymptotic global robustness in Bayesian decision theory. Ann. Statist., 32: 1341–1366.

Basu, S., 1994. Posterior sensitivity to the sampling distribution and the prior: more than one observation. Technical Report 66, Department of Mathematical Sciences, University of Arkansas.

Berger, J.O., 1984. The robust Bayesian viewpoint. Stud. Bayesian Econometrics, 4: 63–144.

Berger, J.O., 1990. Robust Bayesian analysis: sensitivity to the prior. J. Statist. Plann. Inference, 25: 303–328.

Berger, J.O., 1994. An overview of robust Bayesian analysis. Test, 3: 5–124.

Casella, G., & W.E. Strawderman, 1981. Estimating a bounded normal mean. Ann. Statist., 9: 870–878.

Chao, M.T., 1970. The asymptotic behavior of Bayes' estimators. Ann. Math. Statist., 41: 60 1–608.

Charras, A. and van C. Eeden, 1991a. Bayes and admissibility properties of estimators in truncated parameter spaces. Canad. J. Statist., 19: 121–134.

Charras, A. and van C. Eeden, 1991b. Limits of Bayes estimators in convex, truncated pa- rameter spaces. Statist. Probab. Lett., 11: 479–4 83.

Charras, A. and van C. Eeden, 1992. Bayes properties of estimators of location parameters in truncated parameter spaces Statist. Decisions, 10: 81–86.

Efron, B., 1996. Empirical Bayes Methods for Combining Likelihood. J. Amer. Statist. Assoc., 91: 538–565.

Gustafson, P., 1996. Local sensitivity of inferences to prior marginals. J. Amer. Statist. Assoc., 91: 774–781.

Fabius, J., 1964. Asymptotic behavior of Bayes's estimates. The Ann. Math. Statist., 35: 846–856.

Gustafson, P., C. Srinivasan and L. Wasserman, 1994. Local sensitivity analysis. Bayesian statistics, 5: 197–210.

Hartigan, J.A., 2004. Uniform priors on convex sets improve risk. Statist. Probab. Lett., 67: 285–288.

Ishwaran, H. and J.S. Rao, 2005. Spike and slab gene selection for multigroup microarray data. J. Amer. Statist. Assoc., 100: 764–780.

Katz, M.W., 1961. Admissible and minimax estimatoes of parameters in truncated spaces. Ann. Math. Statist., 34: 967–972.

Lehmann, E.L., 1997. Testing statistical hypotheses. 2nd edition. Springer Texts in Statistics. Springer-Verlag, New York.

Marchand, E., I. Ouassou, A.T. Payandeh and F. Perron, 2007. On the estimation of restricted location parameter for symmetric distributions. Technical report number 46, University de Sherbrooke.

Marchand, E. and F. Perron, 2001. Improving on the MLE of a bounded normal mean. Ann. Statist., 29: 1078–1093.

Maruyama, Y., K. Iwasaki, 2005. Sensitivity of minimaxity and admissibility in the estimation of a positive normal mean. Ann. Inst. Statist. Math., 57: 145–156.

Moors, J J.A., 1981. Inadmissibility of linearly estimators in truncated parameter spaces. J. Amer. Statist. Assoc., 76: 910–915.

Moors, J.J.A., 1985. Estimation in Truncated parameter spaces. Ph.D. thesis, Tilburg University, Tilburg, The Netherlands.

Sivaganesan, S., 1993. Robust Bayesian diagnostics. J. Statist. Plann. Inference, 35: 171–188.

Tsou, T.S. and R.M. Royall, 1995. Robust likelihoods. J. Amer. Statist. Assoc., 90: 316–320.