Probability in the Engineering and Informational Sciences, **34**, 2020, 1–13.
doi:10.1017/S026996481800027X

APPLICATIONS OF LIKELIHOOD RATIO ORDER IN BAYESIAN INFERENCES

KAI HUANG and JIE MI
Department of Mathematics and Statistics,
Florida International University, Miami, FL 33199, USA
E-mail: khuang@fiu.edu; mi@fiu.edu

The present paper studies the likelihood ratio order of posterior distributions of parameter when the same order exists between the corresponding prior of the parameter, or when the observed values of the sufficient statistic for the parameter differ. The established likelihood order allows one to compare the Bayesian estimators associated with many common and general error loss functions analytically. It can also enable one to compare the Bayes factor in hypothesis testing without using numerical computation. Moreover, using the likelihood ratio (LR) order of the posterior distributions can yield the LR order between marginal predictive distributions, and posterior predictive distributions.

Keywords: bayesian estimators, bayes factor, likelihood ratio order, predictive distribution, prior distribution, posterior distribution

1. INTRODUCTION

Nowadays Bayesian inference receives more and more attention due to its merits and the rapidly growing computational power of computers. Suppose that a statistical model for the data from a population is given, and one is interested in some numerical characteristics of the population expressed in terms of certain parameters. These parameters are unknown so Bayesian approach tries to describe their uncertainty by the so-called prior distribution. In the case when the shape of the prior distribution is determined, then it becomes necessary to further determine the hyper-parameters appearing in the prior distribution.

Except for the empirical Bayesian inference, the determination of hyper-parameters should be done independently without using the data from the underlying population. The independent information about the hyper-parameters could be obtained from the previous similar experiments, experts' opinion, or literature. However, it could also happen that either these sources provide only very limited information or none of these sources are available. It is certainly an extreme challenge to deal with this situation. Taking this difficulty into consideration, the present paper tries to cast some lights on the choice of the hyper-parameters. Precisely, the results of our study establish the monotone increasing/decreasing relationship between the Bayesian estimators of population parameter associated with different loss functions and the hyper-parameter under certain conditions. Hence the monotonicity can give guidance for selecting the hyper-parameter when researchers feel the interested

© Cambridge University Press 2018 0269-9648/18 $25.00

https://doi.org/10.1017/S026996481800027X Published online by Cambridge University Press

population parameter is overestimated or underestimated. That is, these results can, at the least, provide a certain idea for choosing the value of a hyper-parameter.

In order to achieve this goal, we applied one of the various stochastic orders – the likelihood ratio (LR) order. Stochastic orders have been used in many areas such as economics, insurance, operations research, reliability theory, and survival analysis among others.

Let X and Y be two random variables with cumulative distribution functions (cdf) $F(t)$ and $G(t)$, and probability density functions (pdf) $f(t)$ and $g(t)$, respectively. If the ratio $g(t)/f(t)$ increases in t over the union of the supports of X and Y, then we say that X is smaller than Y in the LR order. For convenience, we also express this order by saying F (or f) is smaller than G (or g) in the LR order throughout this paper. The above order is denoted by $X \overset{\text{lr}}{\leq} Y$, or $F \overset{\text{lr}}{\leq} G$, or $f \overset{\text{lr}}{\leq} g$. More details about the LR order can be found in Shaked and Shanthikumar [7].

The application of the LR order is natural since both the prior and posterior distributions in Bayesian inference are described by density functions. Here the usage of probability density function is in its broader sense. That is, it covers both the discrete and continuous cases. In Section 2 we study the LR order of the posterior distributions. Various applications of the results derived in Section 2 are shown in Section 3. Four examples are given in Section 4. Brief summarization is made in Section 5.

2. LR ORDER OF POSTERIOR DISTRIBUTION

Consider a population represented by a random variable X. Suppose that the pdf $f(x|\theta)$ of X is indexed by parameter $\theta \in \Omega$, and we denote this distribution as $X \sim f(x|\theta)$. Here by pdf we mean probability density function when X is a continuous random variable or probability mass function when X is discrete. We will only consider these two cases in this paper. Let $X_1, \ldots, X_n$ be a random sample from the population, that is, $X_1, \ldots, X_n$ $iid \sim f(x|\theta)$ and x_i be the observed value of $X_i, 1 \leq i \leq n$. Denote $\mathbf{X} = (X_1, \ldots, X_n)$ and $\mathbf{x} \equiv (x_1, \ldots, x_n)$.

According to Bayesian approach θ is treated as a realization of a random variable, say Θ, which has pdf $\pi(\theta)$ named as its prior. We first study the effect of the prior of Θ on the posterior of Θ. In this paper, we always assume Θ is a continuous random variable though similar studies can be conducted when Θ is discrete.

The result in the following theorem is mentioned in Spizzichino [8] and Shaked and Shanthikumar [7]. We just state it in different notations.

THEOREM 1: *Let $\pi_1(\theta)$ and $\pi_2(\theta)$ be two prior pdfs of Θ. If $\pi_1(\theta) \overset{\text{lr}}{\leq} \pi_2(\theta)$, then $\pi_1(\theta|\mathbf{x}) \overset{\text{lr}}{\leq} \pi_2(\theta|\mathbf{x})$ where $\pi_j(\theta|\mathbf{x}), j = 1, 2$ are the posterior pdfs of Θ given $\mathbf{X} = \mathbf{x}$.*

Often, its corollary shown below is more useful.

COROLLARY 1: *Let $T = T(\mathbf{x})$ be a sufficient statistic for the parameter θ. If $\pi_1(\theta) \overset{\text{lr}}{\leq} \pi_2(\theta)$, then $\pi_1(\theta|T = t) \overset{\text{lr}}{\leq} \pi_2(\theta|T = t), \forall t = T(\mathbf{x})$, where $\pi_j(\theta|T = t)$ is the posterior pdf of Θ given $T(\mathbf{x}) = t$.*

PROOF: It is well known that the classical and Bayesian definitions of sufficiency are equivalent. (See, for instance, Zacks [10] or Berger [2]).

https://doi.org/10.1017/S026996481800027X Published online by Cambridge University Press

Hence $T = T(\mathbf{x})$ is also a sufficient statistic in the Bayesian sense, that is

$$\pi_j(\theta|\mathbf{X} = \mathbf{x}) = \pi_j(\theta|T = t), \quad j = 1, 2.$$

and so the result follows. ■

Now we study the effect of the observed value of a sufficient statistic $T = T(\mathbf{X})$ on the posterior pdf $\pi(\theta|T = t)$ of Θ for a fixed prior distribution. We need the following concept of increasing (decreasing) LR.

DEFINITION 1: *Let $\{g(y|\theta) : \theta \in \Omega\}$ be a family of pdf. We say that this family has an increasing (decreasing) LR if $g(y|\theta')/g(y|\theta)$ is an increasing (decreasing) function on $\{y : g(y|\theta') > 0$ or $g(y|\theta) > 0\}$ whenever $\theta \leq \theta' \in \Omega$. Without specifying increasing or decreasing it is common just to say that $\{g(y|\theta) : \theta \in \Omega\}$ has monotone LR (MLR).*

Let $[Y|\Theta = \theta]$ denote any random variable that has the conditional distribution of Y given $\Theta = \theta$ as its distribution. Then the requirement for Y to have an increasing LR can also be expressed as

$$[Y|\Theta = \theta] \overset{\text{lr}}{\leq} [Y|\Theta = \theta']$$

or simply $[Y|\theta] \overset{\text{lr}}{\leq} [Y|\theta']$ whenever $\theta \leq \theta' \in \Omega$.

THEOREM 2: *Suppose that $[X|\Theta = \theta]$ has pdf $f(x|\theta)$ for each $\theta \in \Omega$ and Θ has prior distribution $\pi(\theta)$. Let $T = T(\mathbf{X})$ be a sufficient statistics for θ. If T has an increasing LR, that is,*

$$[T|\theta] \overset{\text{lr}}{\leq} [T|\theta']$$

whenever $\theta \leq \theta'$ then

$$[\Theta|T = t] \overset{\text{lr}}{\leq} [\Theta|T = t']$$

whenever $t \leq t'$, or the posterior pdf

$$\pi(\theta|t) \overset{\text{lr}}{\leq} \pi(\theta|t').$$

PROOF: By the equivalence of the classical and Bayesian sufficiency the posterior pdf of Θ given $\mathbf{X} = \mathbf{x}$ equals to $\pi(\theta|T(\mathbf{x}) = t)$. We thus have

$$\pi(\theta|T = t) = \frac{g(t|\theta)\pi(\theta)}{\int_\Omega g(t|\theta)\pi(\theta)d\theta} = \frac{g(t|\theta)\pi(\theta)}{h(t)}$$

where $g(t|\theta)$ is the pdf of T given $\Theta = \theta$ and $h(t) \equiv \int_\Omega g(t|\theta)\pi(\theta)d\theta$ is the value of the marginal pdf of T at t.

https://doi.org/10.1017/S026996481800027X Published online by Cambridge University Press

It gives the ratio

$$\frac{\pi(\theta|T=t')}{\pi(\theta|T=t)} = \frac{h(t)}{h(t')} \cdot \frac{g(t'|\theta)}{g(t|\theta)}.$$

The assumption $[T|\theta] \overset{\text{lr}}{\leq} [T|\theta']$ implies that

$$\frac{g(t|\theta')}{g(t|\theta)} \leq \frac{g(t'|\theta')}{g(t'|\theta)}, \quad \forall t \leq t'$$

or

$$\frac{g(t'|\theta)}{g(t|\theta)} \leq \frac{g(t'|\theta')}{g(t|\theta')}, \quad \forall t \leq t'.$$

Note that

$$\frac{\pi(\theta'|T=t')}{\pi(\theta'|T=t)} = \frac{h(t)}{h(t')} \cdot \frac{g(t'|\theta')}{g(t|\theta')},$$

so it follows that

$$\frac{\pi(\theta|T=t')}{\pi(\theta|T=t)} \leq \frac{\pi(\theta'|T=t')}{\pi(\theta'|T=t)},$$

that is,

$$[\Theta|T=t] \overset{\text{lr}}{\leq} [\Theta|T=t']$$

whenever $t' \leq t$. ■

The following corollary is an important special case of Theorem 2.

COROLLARY 2: *Let* $\{f(x|\theta), \theta \in \Omega\}$ *be an exponential family of pdfs and*

$$f(x|\theta) = h(x)c(\theta)\exp\{-a(\theta)b(x)\}.$$

Suppose that $a(\theta)$ *and* $b(x)$ *are both increasing (decreasing) functions of* θ *and* x*, respectively. Then the sufficient statistic* $T \equiv T(\mathbf{X}) = \sum_{i=1}^{n} b(X_i)$ *has MLR and so the result of Theorem 2 holds in this case.*

Remark 2.1: A very simple case of Theorem 2 is when there is only one observation $X = x$ was considered before, for example, in Suen [9]. It is trivial since if there is only one observation then X is certainly the sufficient statistic for θ.

3. APPLICATIONS

We will keep the notations introduced in Section 2.

3.1. Point Estimation

The results in Section 2 state that the posterior distributions inherit the likelihood ratio order existing among the associated prior distributions. For certain loss functions these results can then be applied to compare the Bayesian estimators of θ corresponding to different prior distributions.

https://doi.org/10.1017/S026996481800027X Published online by Cambridge University Press

Consider the loss function $L^{(w)}(a,\theta) \equiv w(\theta)(\theta - a)^2$ where $w(\theta)$ is a given non-negative function. Let $\hat{\theta}_B^{(w)}$ denote the Bayesian estimator of θ associated with $L^{(w)}(a,\theta)$. It is known that

$$\hat{\theta}_B^{(w)} = \frac{\int_\Omega \theta w(\theta)\pi(\theta|\mathbf{x})d\theta}{\int_\Omega w(\theta)\pi(\theta|\mathbf{x})d\theta}$$

where $\pi(\theta)$ is the prior of Θ and Ω is the parameter space of θ.

THEOREM 3: *Suppose that two posterior distributions* $\pi_i(\theta|\mathbf{x}), i = 1,2$ *satisfy* $\pi_1(\theta|\mathbf{x}) \overset{\text{lr}}{\leq} \pi_2(\theta|\mathbf{x})$. *Then it holds that* $\hat{\theta}_{B_1}^{(w)} \leq \hat{\theta}_{B_2}^{(w)}$.

PROOF: It suffices to show that

$$\frac{\int_\Omega \theta w(\theta)\pi_1(\theta|\mathbf{x})d\theta}{\int_\Omega w(\theta)\pi_1(\theta|\mathbf{x})d\theta} \leq \frac{\int_\Omega \xi w(\xi)\pi_2(\xi|\mathbf{x})d\xi}{\int_\Omega w(\xi)\pi_2(\xi|\mathbf{x})d\xi},$$

$$\iint\limits_{\Omega\times\Omega} \theta w(\theta)w(\xi)\pi_1(\theta|\mathbf{x})\pi_2(\xi|\mathbf{x})d\theta\, d\xi \leq \iint\limits_{\Omega\times\Omega} \xi w(\theta)w(\xi)\pi_1(\theta|\mathbf{x})\pi_2(\xi|\mathbf{x})d\theta\, d\xi,$$

$$\begin{aligned}
0 &\leq \iint\limits_{\Omega\times\Omega} (\xi - \theta)w(\theta)w(\xi)\pi_1(\theta|\mathbf{x})\pi_2(\xi|\mathbf{x})d\theta\, d\xi \\
&= \iint\limits_{\theta<\xi} (\xi - \theta)w(\theta)w(\xi)\pi_1(\theta|\mathbf{x})\pi_2(\xi|\mathbf{x})d\theta\, d\xi \\
&\quad + \iint\limits_{\theta>\xi} (\xi - \theta)w(\theta)w(\xi)\pi_1(\theta|\mathbf{x})\pi_2(\xi|\mathbf{x})d\theta\, d\xi \\
&= \iint\limits_{\xi>\theta} (\xi - \theta)w(\theta)w(\xi)[\pi_1(\theta|\mathbf{x})\pi_2(\xi|\mathbf{x}) - \pi_1(\xi|\mathbf{x})\pi_2(\theta|\mathbf{x})]d\theta\, d\xi.
\end{aligned}$$

Hence it suffices to show

$$\pi_1(\theta|\mathbf{x})\pi_2(\xi|\mathbf{x}) - \pi_1(\xi|\mathbf{x})\pi_2(\theta|\mathbf{x}) \geq 0, \quad \forall \xi > \theta,$$

or

$$\frac{\pi_2(\xi|\mathbf{x})}{\pi_1(\xi|\mathbf{x})} \geq \frac{\pi_2(\theta|\mathbf{x})}{\pi_1(\theta|\mathbf{x})} \quad \forall \xi > \theta. \tag{3.1}$$

The inequality (3.1) is true since it is assumed that

$$\pi_1(\theta|\mathbf{x}) \overset{\text{lr}}{\leq} \pi_2(\theta|\mathbf{x}).$$

■

Remark 3.1: The similar results associated with squared/absolute error loss functions were mentioned in Arias-Nicolás, Ruggeri and Suárez-Llorens [1]. Clearly, the squared error loss function is a special case of $L^{(w)}(a,\theta)$ with $w(\theta) = 1, \forall \theta \in \Omega$.

THEOREM 4: *Let* $T = T(\mathbf{X})$ *be a sufficient statistic for* θ. *If* $[T|\theta] \overset{\text{lr}}{\leq} [T|\theta']$ *whenever* $\theta \leq \theta' \in \Omega$, *then both* $\hat{\theta}_B^{(w)}[t]$ *and* $\hat{\theta}_B^{(a)}[t]$ *are increasing in* t *where* $\hat{\theta}_B^{(w)}[t]$ *and* $\hat{\theta}_B^{(a)}[t]$ *are the*

https://doi.org/10.1017/S026996481800027X Published online by Cambridge University Press

Bayesian estimators of θ associated with the loss function $L^{(w)}(a,\theta)$ and the absolute error loss function $L^{(a)}(a,\theta)=|a-\theta|$, respectively.

PROOF: With the assumption $[T|\theta] \overset{\text{lr}}{\leq} [T|\theta']$ whenever $\theta \leq \theta' \in \Omega$ it follows from Theorem 2 that

$$[\Theta|T=t] \overset{\text{lr}}{\leq} [\Theta|T=t'] \quad \text{whenever} \quad t \leq t'. \tag{3.2}$$

This immediately implies that $\hat{\theta}_B^{(w)}[t]$ is increasing in t by Theorem 3.

Note that $\hat{\theta}_B^{(a)}[t]$ is just the median of the $[\Theta|T=t]$, and the LR order $[\Theta|T=t] \overset{\text{lr}}{\leq} [\Theta|T=t']$ in (3.2) implies $[\Theta|T=t] \overset{st}{\leq} [\Theta|T=t']$ whenever $t \leq t'$. The result about $\hat{\theta}_B^{(a)}[t]$ is certainly true. ■

3.2. Hypothesis Testing

A monotonicity property of the Bayes factor, which is important in Bayesian hypothesis testing is shown below.

Suppose that we want to test

$$H_0 : \theta \leq \theta_0 \quad vs \quad H_1 : \theta > \theta_0.$$

The posterior odds ratio of H_0 relative to H_1 is determined by the Bayes factor of H_0 relative to H_1 defined as

$$BF_{01} = \frac{\int_{-\infty}^{\theta_0} f(\mathbf{x}|\theta)\pi(\theta)d\theta}{\int_{\theta_0}^{\infty} f(\mathbf{x}|\theta)\pi(\theta)d\theta}. \tag{3.3}$$

The smaller BF_{01} is, the stronger evident against H_0. For more details, the readers are referred to Ghosh, Delampady and Samanta [5].

THEOREM 5: *Let $\pi_j(\theta), j=1,2$ be two prior pdfs of Θ and $\Omega=\Omega_0\cup\Omega_1$ where $\Omega_0=(\theta_l,\theta_0]$ and $\Omega_1=(\theta_0,\theta_u)$. Denote the Bayes factor of H_0 relative to H_1 of testing*

$$H_0 : \theta \in \Omega_0, \quad vs \quad H_1 : \theta \in \Omega_1$$

as $BF_{01}^{(j)}$ when Θ has prior $\pi_j(\theta), j=1,2$. If $\pi_1(\theta) \overset{\text{lr}}{\leq} \pi_2(\theta)$, then $BF_{01}^{(1)} \geq BF_{01}^{(2)}$.

PROOF: We have

$$BF_{01}^{(j)} = \frac{\int_{\Omega_0} f(x|\theta)\pi_j(\theta)d\theta}{\int_{\Omega_1} f(x|\mu)\pi_j(\mu)d\mu}, \quad j=1,2 \tag{3.4}$$

The desired inequality is equivalent to

$$\int_{\Omega_0} f(x|\theta)\pi_1(\theta)d\theta \int_{\Omega_1} f(x|\mu)\pi_2(\mu)d\mu \geq \int_{\Omega_0} f(x|\theta)\pi_2(\theta)d\theta \int_{\Omega_1} f(x|\mu)\pi_1(\mu)d\mu$$

$$\int_{\theta\in\Omega_0}\int_{\mu\in\Omega_1} f(x|\theta)f(x|\mu)\pi_1(\theta)\pi_1(\mu)\left[\frac{\pi_2(\theta)}{\pi_1(\theta)}-\frac{\pi_2(\mu)}{\pi_1(\mu)}\right]d\mu\, d\theta \leq 0. \tag{3.5}$$

The assumption $\pi_1(\theta) \overset{\text{lr}}{\leq} \pi_2(\theta)$ implies

$$\frac{\pi_2(\theta)}{\pi_1(\theta)} \leq \frac{\pi_2(\mu)}{\pi_1(\mu)}, \quad \forall\theta\in\Omega_0, \quad \text{and} \quad \mu\in\Omega_1 \tag{3.6}$$

https://doi.org/10.1017/S026996481800027X Published online by Cambridge University Press

since $\Omega_0 = (\theta_l,\ \theta_0]$ and $\Omega_1 = (\theta_0,\ \theta_u)$. The inequality (3.6) shows that (3.5) is true and consequently $BF_{01}^{(1)} \geq BF_{01}^{(2)}$. ■

Remark 3.2: The result says that if the prior order $\pi_1(\theta) \overset{\text{lr}}{\leq} \pi_2(\theta)$ holds, then Θ tends to be "bigger" under prior $\pi_2(\theta)$ than under prior $\pi_1(\theta)$ given the same data $\mathbf{x}$. To see the meaning of "bigger", we refer the readers to Billingsley [3] and Ross [6]. Hence, Theorem 5 indicates that the Bayes factor can be interpreted as the extent to which the data support H_1.

3.3. Predictive Distribution

One of the key purposes of statistical inference is to make a prediction about the random variable of interest. The present subsection will deal with the LR order of prediction distributions.

THEOREM 6: *Let the family of pdfs* $\{f(x|\theta), \theta \in \Omega\}$ *has an increasing LR, that is,* $[X|\Theta = \theta] \overset{\text{lr}}{\leq} [X|\Theta = \theta']$ *whenever* $\theta \leq \theta' \in \Omega$. *Consider two priors* $\pi_j(\theta), j = 1, 2$ *of* Θ. *Suppose that* $\pi_1(\theta) \overset{\text{lr}}{\leq} \pi_2(\theta)$. *Then the following holds:*

(a) $m_1(x) \overset{\text{lr}}{\leq} m_2(x)$ *where* $m_j(x)$ *is the marginal distribution (or the prior predictive distribution) of* X.

(b) Denote the posterior predictive function of the new unobserved X_{n+1} *as* $f_j(x_{n+1}|\mathbf{x})$ *when* Θ *has prior* $\pi_j(\theta), j = 1, 2$. *Then*

$$f_1(x_{n+1}|\mathbf{x}) \overset{\text{lr}}{\leq} f_2(x_{n+1}|\mathbf{x}). \tag{3.7}$$

PROOF: The prior predictive distribution of X has pdf

$$m_j(x) = \int_\Omega f(x|\theta)\pi_j(\theta)d\theta \tag{3.8}$$

when Θ has prior pdf $\pi_j(\theta)$.

By the assumption $[X|\theta] \overset{\text{lr}}{\leq} [X|\theta']$ whenever $\theta \leq \theta' \in \Omega$ we have $f(x|\theta) \overset{\text{lr}}{\leq} f(x|\theta')$. From this fact and Eq. (3.8) it follows that

$$m_1(x) = \int_\Omega f(x|\theta)\pi_1(\theta)d\theta \overset{\text{lr}}{\leq} \int_\Omega f(x|\theta)\pi_2(\theta)d\theta = m_2(x)$$

by Theorem 1.C.17 of Shaked and Shanthikumar [7]. This shows result (a).

To show result (b) note that the posterior predictive function of the unobserved new X_{n+1} given $\mathbf{x}$ has pdf

$$f_j(x_{n+1}|\mathbf{x}) = \int_\Omega f(x_{n+1}|\theta)\pi_j(\theta|\mathbf{x})d\theta. \tag{3.9}$$

From the assumption $\pi_1(\theta) \overset{\text{lr}}{\leq} \pi_2(\theta)$ we see that $\pi_1(\theta|\mathbf{x}) \overset{\text{lr}}{\leq} \pi_2(\theta|\mathbf{x})$ by Theorem 1. Combing this with the fact that $f(x|\theta)$ has an increasing LR we can apply Theorem 1.C.17 of Shaked and Shanthikumar [7] again and obtain

$$f_1(x_{n+1}|\mathbf{x}) = \int f(x_{n+1}|\theta)\pi_1(\theta|\mathbf{x})d\theta \overset{\text{lr}}{\leq} \int f(x_{n+1}|\theta)\pi_2(\theta|\mathbf{x})d\theta = f_2(x_{n+1}|\mathbf{x}).$$ ■

https://doi.org/10.1017/S026996481800027X Published online by Cambridge University Press

THEOREM 7: *Let $T = T(\mathbf{X})$ be a sufficient statistic for θ. Denote the pdf of T as $g(t|\theta)$. Suppose that $\{g(x|\theta), \theta \in \Omega\}$ has an increasing LR. Then the following result is true:*

$$f(x_{n+1}|\mathbf{x}) \overset{\text{lr}}{\leq} f(x_{n+1}|\mathbf{x}')$$

whenever $T(\mathbf{x}) = t \leq t' = T(\mathbf{x}')$.

PROOF: Due to the equivalence of the classical and Bayesian sufficiency it holds that the posterior predictive distribution of the unobserved new X_{n+1} has pdf

$$f(x_{n+1}|\mathbf{x}) = \int_\Omega f(x_{n+1}|\theta)\pi(\theta|\mathbf{X} = \mathbf{x})d\theta = \int_\Omega f(x_{n+1}|\theta)\pi(\theta|T = t)d\theta$$

where $t = T(\mathbf{x})$. By Theorem 2 it holds that $\pi(\theta|T = t) \overset{\text{lr}}{\leq} \pi(\theta|T = t')$ whenever $t \leq t'$. Therefore, $f(x_{n+1}|\mathbf{x}) \overset{\text{lr}}{\leq} f(x_{n+1}|\mathbf{x}')$ whenever $t = T(\mathbf{x}) \leq T(\mathbf{x}') = t'$. ■

4. EXAMPLES

We will exhibit four examples in this section. In some of these examples the Bayesian estimators mentioned above have closed-form expressions, hence their monotonicity with respect to hyper-parameters can be checked easily. In the other examples these estimators do not have closed-forms and thus it is difficult to see the monotonicity directly. Nevertheless, the monotonicity can be shown easily by using the results derived above.

Example 4.1: Let $X \sim \text{Poisson}(\theta)$ with mean $\theta \in \Omega = (0, \infty)$. Consider conjugate prior $\Theta \sim \text{Gamma}(\alpha, \beta)$ with pdf

$$\pi(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\beta\theta}, \quad \theta > 0.$$

It is well known that the posterior distribution of Θ given $\mathbf{X} = \mathbf{x}$ is $\text{Gamma}(\alpha + t, \beta + n)$, where $t = \sum_{i=1}^n x_i$. Note that the statistic $T = T(\mathbf{X}) \equiv \sum_{i=1}^n X_i \sim \text{Poisson}(n\theta)$ is sufficient for θ. Thus the posterior distribution of Θ can be written as

$$\pi(\theta|\mathbf{x}) = \pi(\theta|T = t) = \frac{\beta^{\alpha+t}}{\Gamma(\alpha + t)}\theta^{\alpha+t-1}e^{-(\beta+n)\theta}, \quad \theta > 0.$$

It is easy to see that $[T|\theta] \overset{\text{lr}}{\leq} [T|\theta']$ whenever $\theta \leq \theta' \in \Omega$. By Theorem 2 we know that $[\Theta|T = t] \overset{\text{lr}}{\leq} [\Theta|T = t']$ whenever $t \leq t'$. Theorem 4 further implies $\hat{\theta}_B^{(s)}[t] \leq \hat{\theta}_B^{(s)}[t']$ and $\hat{\theta}_B^{(a)}[t] \leq \hat{\theta}_B^{(a)}[t']$ whenever $t \leq t'$. If we notice that $[\Theta|\mathbf{x}] \sim \text{Gamma}(\alpha + t, \beta + n)$ and the posterior mean of Θ given $\mathbf{X} = \mathbf{x}$ is

$$\hat{\theta}_B^{(s)}[t] = \frac{\alpha + nt}{\beta + n},$$

then, of course, we see that $\hat{\theta}_B^{(s)}[t]$ is increasing in t. But it is not easy to see that the posterior median $\hat{\theta}_B^{(a)}[t]$ is also increasing in t without applying Theorem 4.

https://doi.org/10.1017/S026996481800027X Published online by Cambridge University Press

In addition, due to Theorem 7 we immediately conclude that in this example

$$f(x_{n+1}|\mathbf{x}) \overset{\text{lr}}{\leq} f(x_{n+1}|\mathbf{x}')$$

whenever $t = \sum_i^n x_i \leq t' = \sum_i^n x_i'$, where $f(x_{n+1}|\mathbf{x})$ is the posterior predictive distribution of a new unobserved X_{n+1} given $\mathbf{X} = \mathbf{x}$. This can be verified directly by noting that $f(x_{n+1}|\mathbf{x})$ actually has negative-binomial pdf

$$f(x_{n+1}|\mathbf{x}) = \binom{\alpha + t + x_{n+1} - 1}{x_{n+1}} \left(\frac{\beta + n}{\beta + n + 1}\right)^{\alpha+t} \left(\frac{1}{\beta + n + 1}\right)^{x_{n+1}}. \tag{4.1}$$

Moreover, let us consider two different values β_1 and β_2 of the hyper-parameter β in the prior distribution Gamma(α, β). That is, we consider two priors $\pi_1(\theta) = \text{Gamma}(\alpha, \beta_1)$ and $\pi_2(\theta) = \text{Gamma}(\alpha, \beta_2)$ with $0 < \beta_1 \leq \beta_2$. Under this situation Theorem 3 implies that $\hat{\theta}_B^{(s)}(\beta_1) \geq \hat{\theta}_B^{(s)}(\beta_2)$ and $\hat{\theta}_B^{(a)}(\beta_1) \geq \hat{\theta}_B^{(a)}(\beta_2)$. The results of Theorem 6 imply that the prior and posterior predictive distributions corresponding to $\pi_1(\theta)$ and $\pi_2(\theta)$ satisfy

$$m_2(x) \overset{\text{lr}}{\leq} m_1(x) \quad \text{and} \quad f_2(x_{n+1}|\mathbf{x}) \overset{\text{lr}}{\leq} f_1(x_{n+1}|\mathbf{x}).$$

It is easy to obtain the inequality about the posterior means and prior predictive distributions directly, however, it is not so straight forward to show the results about the posterior medians and predictive distributions without applying Theorem 3 and Theorem 6.

Example 4.2: Let X follow normal distribution $N(0, \sigma^2)$. Suppose that Σ^2 has Inv-$\chi^2(\nu_0, \sigma^2)$ as its prior, that is,

$$\Sigma^2 \sim \pi(\sigma^2|\nu, \sigma_0^2) \propto (\sigma^2)^{-(\nu_0/2+1)} \exp\left(-\frac{\nu_0\sigma_0^2}{2\sigma^2}\right), \quad \sigma^2 > 0 \tag{4.2}$$

where $\nu_0 > 0$ and σ_0^2 are hyper-parameters.

It is easy to see that $[X|\sigma_1^2] \overset{\text{lr}}{\leq} [X|\sigma_2^2]$ whenever $0 < \sigma_1^2 \leq \sigma_2^2$ and $\pi(\sigma^2|\nu_0, \sigma_{01}^2) \overset{\text{lr}}{\leq} \pi(\sigma^2|\nu_0, \sigma_{02}^2)$ whenever $0 < \sigma_{01}^2 \leq \sigma_{02}^2$.

Therefore, from the results in Section 3 we immediately conclude that the posterior mean of Σ^2 satisfy $\widehat{\sigma^2}_B^{(s)}(\sigma_{01}^2) \leq \widehat{\sigma^2}_B^{(s)}(\sigma_{02}^2)$ and the same inequality for the posterior median $\widehat{\sigma^2}_B^{(a)}(\sigma_{01}^2) \leq \widehat{\sigma^2}_B^{(a)}(\sigma_{02}^2)$, $\forall 0 < \sigma_{01}^2 \leq \sigma_{02}^2$.

Obviously, the statistic $T \equiv T(\mathbf{X}) = \sum_{i=1}^n X_i^2$ is sufficient for σ^2. It can be readily verified that for any $n \geq 2$, $[T|\sigma_1^2] \overset{\text{lr}}{\leq} [T|\sigma_2^2]$ whenever $0 < \sigma_1^2 \leq \sigma_2^2$ since $T/\sigma^2 \sim \chi^2(n)$. Hence, by Theorem 4 we know that $\widehat{\sigma^2}_B^{(s)}[t]$ and $\widehat{\sigma^2}_B^{(a)}[t]$ are increasing in $t \equiv \sum_{i=1}^n x_i^2$. Furthermore, from Theorem 7 it follows that $f(x_{n+1}|\mathbf{x}) \overset{\text{lr}}{\leq} f(x_{n+1}|\mathbf{x}')$ whenever $t = \sum_i^n x_i^2 \leq t' = \sum_i^n (x_i')^2$ holds for the posterior predictive distribution of a new unobserved X_{n+1}.

It is known that the posterior of Σ^2 given $\mathbf{X} = \mathbf{x}$ is the scaled Inv-$\chi^2(\nu_n, \sigma_n^2)$ distribution where

$$\nu_n = \nu_0 + n, \quad \sigma_n^2 = \frac{\nu_0\sigma_0^2 + nv}{\nu_n} \quad \text{and} \quad v = \frac{\sum_{i=1}^n x_i^2}{n}.$$

That is

$$\pi(\sigma^2|\mathbf{x}) \propto (\sigma^2)^{-(\nu_n/2+1)} \exp(-\frac{\nu_n\sigma_n^2}{2\sigma^2}), \quad \sigma^2 > 0. \tag{4.3}$$

https://doi.org/10.1017/S026996481800027X Published online by Cambridge University Press

From (4.3) it follows that the Bayes estimator of σ^2 with the squared loss function is given by the posterior mean

$$\widehat{\sigma^2}_B^{(s)} = \frac{\nu_n \sigma_n^2}{\nu_n - 2} = \frac{\nu_0 \sigma_0^2 + \sum_{i=1}^n x_i^2}{\nu_0 + n - 2} \tag{4.4}$$

based on (4.1).

From (4.4) it can be seen that $\widehat{\sigma^2}_B^{(s)}(\sigma_0^2)$ is increasing in σ_0^2 and $\widehat{\sigma^2}_B^{(s)}[t]$ is increasing in $t \equiv \sum_{i=1}^n x_i^2$. Nevertheless, it is not so simple to verify the similar inequality for the posterior medians. It is even more difficult to obtain the LR order of the posterior predictive distributions of X_{n+1} given $\mathbf{X} = \mathbf{x}$.

Example 4.3: We continue the discussion in Example 4.1 and still assume X has Poisson distribution with mean θ. But this time we consider a hierarchical prior as follows. Let p be a fixed positive integer. We assume that Θ follows a chi-square distribution with $p + 2j$ degrees of freedom (df) when the hyper-parameter $J = j$, that is, $[\Theta | J = j] \sim \chi^2_{p+2j}(\theta)$, which is actually a Gamma$(p/2 + j, 1/2)$ distribution. We further assume $J \sim$ Poisson(λ) for a known constant $\lambda > 0$. According to this hierarchical model the marginal distribution $\pi(\theta)$ of Θ is a non-central chi-square distribution with pdf of p and non-central parameter λ, which we denote as $\Theta \sim \chi^2_p(\theta|\lambda)$. It is known that the non-central chi-square distribution $\chi^2_p(\theta|\lambda)$ is, in fact, the mixture of central chi-square distributions $\{\chi^2_{p+2j}, j = 0, 1, \ldots\}$ with mixing distribution Poisson(λ) and thus has pdf given by

$$\pi(\theta) = \sum_{j=0}^{\infty} \frac{\theta^{p/2+j-1} e^{-\theta/2}}{\Gamma(p/2 + j) 2^{p/2+j}} \cdot \frac{e^{-\lambda} \lambda^j}{j!}, \quad \theta > 0.$$

If we denote the cumulative distribution function of a random variable J as $G(j|\lambda)$, then the prior pdf of Θ can also be expressed as

$$\pi(\theta) = \int_0^{\infty} \chi^2_{p+2j}(\theta) dG(j|\lambda)$$

where $\chi^2_{p+2j}(\theta)$ stands for the pdf of χ^2_{p+2j} distribution.

Let $0 < \lambda_1 \le \lambda_2$ be two given constants and $\pi_k(\theta) = \int_0^{\infty} \chi^2_{p+2j}(\theta) dG(j|\lambda_k)$, $k = 1, 2$. Clearly it holds that $\pi_1(\theta) \overset{\text{lr}}{\le} \pi_2(\theta)$ by Theorem 1.C.17 in Shaked and Shanthikumar [7] since Poisson$(\lambda_1) \overset{\text{lr}}{\le}$ Poisson(λ_2). Therefore, from the results of our Theorem 1, Theorem 3, and Theorem 6, we immediately conclude that both $\hat{\theta}_B^{(s)}(\lambda)$ and $\hat{\theta}_B^{(a)}(\lambda)$ are increasing in $\lambda > 0$, and the posterior predictive distributions also satisfy the LR order $f_1(x_{n+1}|\mathbf{x}) \overset{\text{lr}}{\le} f_2(x_{n+1}|\mathbf{x})$. These facts can hardly, if not impossible, be shown analytically without using the results derived in this paper because the posterior distribution $\pi(\theta|\mathbf{x})$ of Θ given $\mathbf{X} = \mathbf{x}$ is expressed in terms of infinite series.

Figure 1 is based on our numerical computation and shows the Bayesian estimates $\hat{\theta}_B^{(s)}(\lambda)$ and $\hat{\theta}_B^{(a)}(\lambda)$ corresponding to square loss function and absolute error loss function, respectively. The horizon axis displays λ. Our numerical simulation used $\theta = 5, n = 15$, and $p = 2$. In both the left and right panel of Figure 1 there are three curves associated with three different values of $t = \sum_{i=1}^n x_i$. These simulation results clearly verified that both $\hat{\theta}_B^{(a)}$ and $\hat{\theta}_B^{(s)}$ increase in λ and t as claimed by the analytical argument.

https://doi.org/10.1017/S026996481800027X Published online by Cambridge University Press

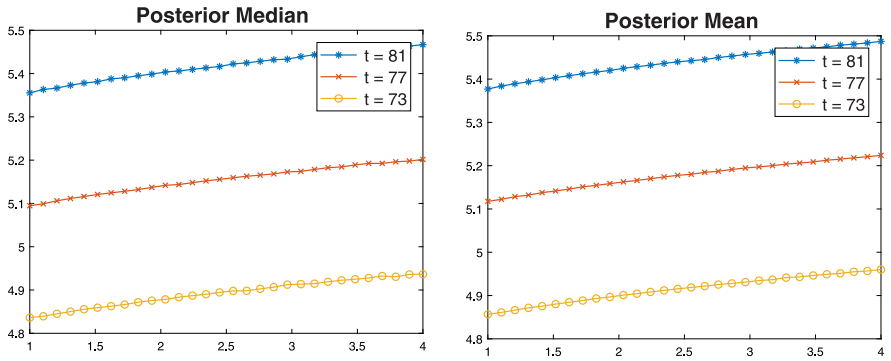


FIGURE 1. Example 3.

Example 4.4: Suppose that Bernoulli random variable X has a probability of success $P(X = 1) = \theta$ and a sample $\mathbf{X} = \mathbf{x}$ is available. The usual way of obtaining the Bayesian estimator of θ is to use Beta(α, β) as the prior distribution of Θ. This conjugate prior certainly gives a great convenience. But if it was believed a priori that Θ had lower and upper bounds a and b where $0 \le a < b \le 1$ are two known constants, then using Beta(α, β) on the interval $(0, 1)$ as the prior of Θ in this case, of course, is no longer appropriate. Instead, it would be more reasonable to consider the beta-distribution restricted to the interval (a, b) as the prior of Θ. For this, the readers are referred to Christensen et al. [4]. Assume that the Beta(α, β) distribution restricted to interval $(0, c)$ is used as the prior of Θ, where $c \in (0, 1]$ is a given constant. For the brevity of notation, here we denote this prior distribution by $\pi_c(\theta)$ and the associated posterior distribution as $\pi_c(\theta|t)$ where $t = \sum_{i=1}^n x_i$ is the observed value of the sufficient statistic $T = \sum_{i=1}^n X_i$ for θ. Clearly,

$$\pi_c(\theta) = \frac{1}{B(\alpha, \beta)c} \left(\frac{\theta}{c}\right)^{\alpha-1} \left(1 - \frac{\theta}{c}\right)^{\beta-1} I_{(0,c)}(\theta)$$

where $I_{(0,c)}(\theta)$ is the indicator function of the interval $(0, c)$. Denote the Bayesian estimator of θ under the absolute error lose function as $\hat{\theta}_B^{(a)}(c)$. It is easy to see that $\pi_{c_1}(\theta) \overset{\text{lr}}{\le} \pi_{c_2}(\theta|)$ whenever $0 < c_1 < c_2 \le 1$. Hence for the two corresponding posterior distributions, it is true that $\pi_{c_1}(\theta|t) \overset{\text{lr}}{\le} \pi_{c_2}(\theta|t)$. Hence $\hat{\theta}_B^{(a)}(c)$ increases in c and t.

Consider the loss function $L^{(w)}(a, \theta) \equiv w(\theta)(\theta - a)^2$ with $w(\theta) = \theta^{-1}$ and denote the corresponding Bayesian estimator as $\hat{\theta}_B^{(w)}(c)$. It is easy to see that

$$\hat{\theta}_B^{(w)}(c) = \frac{1}{\int_0^1 \theta^{-1}\pi_c(\theta|t)d\theta}.$$

By Theorem 3 and Theorem 4 it is straightforward to conclude that and $\hat{\theta}_B^{(w)}(c)$ increase in t and c.

On the other hand, the posterior distribution of Θ given $T = t$ does not possess a closed-form expression and thus it is difficult to derive the same conclusions without applying the results in the present paper. Again, Figure 2 based on simulation studies exhibits the validity of the theoretical results claimed above.

Similar to Figure 1, the Bayesian estimates $\hat{\theta}_B^{(a)}(c)$ and $\hat{\theta}_B^{(w)}(c)$ corresponding to three different values of $t = \sum_{i=1}^n x_i$ when Θ has prior distribution $\pi_c(\theta)$ are shown in Figure 2.

https://doi.org/10.1017/S026996481800027X Published online by Cambridge University Press

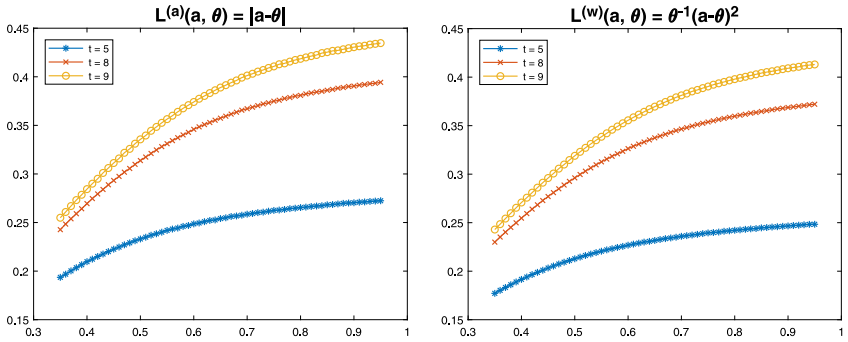


FIGURE **2.** Example 4.

The horizontal line is c-axis. In this example, $\theta = 0.4, \alpha = 2, \beta = 3$, and $n = 20$ were used. Once more, numerical computation validates that both $\hat{\theta}_B^{(a)}(c)$ (left panel) and $\hat{\theta}_B^{(w)}(c)$ (right panel) are increasing in c and t as the theoretical results asserted.

5. SUMMARY

In the Bayesian approach, Bayesian actions depend on posterior distribution. Hence studying the behavior of posterior distributions is necessary. It was shown in Spizzichino [8] and Shaked and Shanthikumar [7] that posterior distributions inherit the LR order existing among the associated prior distributions. In this paper, we prove that if the sufficient statistic for the parameter of interest has an increasing (or decreasing) LR, then there exists LR order among the posterior distributions corresponding to different observed values of the sufficient statistic.

As for the LR order among prior distributions, commonly it can be determined by comparing the values of the hyper-parameter appearing in a family of prior distributions. For any specific prior, paper Arias-Nicolas et al. [1] proposed a new way to obtain a prior which is greater (less) than the specific prior according to the LR order by using a convex (concave) distortion function on $[0, 1]$.

It was mentioned in Arias-Nicolas et al. [1] that the Bayesian estimators for a parameter of interest associated with the squared and absolute error loss functions can be ordered, that is there is a monotonicity if the estimators were based on two posterior distributions ordered by the LR. The present paper extends the result to the estimators $\hat{\theta}_B^{(w)}$ associated with the widely used loss function $L^{(w)}(a, \theta) = w(\theta)(\theta - a)^2$. It is worthwhile to notice that the class of loss function $\mathbb{L}_{sm}$ considered in Arias-Nicolas et al. [1] and the class $\{L^{(w)}(a, \theta) : \text{ all } w(\cdot) \geq 0\}$ are not the same.

In the Bayesian approach, it is common to consider a family of prior distributions that can be indexed by a hyper-parameter. In this case, the monotonicity result about the Bayesian estimators can provide ideas on choosing the hyper-parameters in the family. Additionally, the similar monotonicity is also established for Bayes factors in hypothesis testing, and the posterior predictive function of the new unobserved data with respect to the hyper-parameter. The monotonicity results developed in this paper are particularly useful when these Bayesian quantities do not have closed-form expressions. Some examples have been provided to illustrate the results. Intensive simulation has been carried out to verify the analytical argument which gives convincing evidence.

https://doi.org/10.1017/S026996481800027X Published online by Cambridge University Press

The determination of the hyper-parameters in Bayesian inference is always an issue. There is no any algorithm for determining its value except for the case of the Empirical Bayesian inference. This paper cannot provide any "formula" for making the determination either. Nevertheless, if there is a certain evidence showing either overestimation or underestimation of the parameter with the currently used hyper-parameter, then the results of this paper can definitely determine the direction towards which the hyper-parameter should be adjusted.

We considered a possible extension for the vector parameter. However, in this case, the (multivariate) LR order, in general, does not hold between posterior distributions even this order is true for their corresponding prior distributions unless $f(x|\theta)$ is MTP_2. To address this would need much more space and was thus omitted here. It certainly deserves further investigation in the future.

Acknowledgments

The authors would like to thank the anonymous reviewers for their valuable comments and suggestions.

References

1. Arias-Nicolás, J.P., Ruggeri, F., & Suárez-Llorens, A. (2016). New classes of priors based on stochastic orders and distortion functions. *Bayesian Analysis* 11(4): 1107–1136.
2. Berger, J.O. (1985). *Statistical decision: theory and Bayesian analysis*, 2nd ed. New York, USA: Springer-Verlag.
3. Billingsley, P. (1986). *Probability and measure*, 2nd ed. New York, USA: Wiley.
4. Christensen, R., Johnson, W., Branscum, A., & Hanson, T.E. (2011). *Bayesian ideas and data analysis: An introduction for scientists and statisticians*. Boca Raton, FL, USA: CRC Press Taylor & Francis Group.
5. Ghosh, J., Delampady, M., & Samanta, T. (2006). *An introduction to Bayesian analysis theory and methods*. New York, USA: Springer.
6. Ross, S.M. (1983). *Stochastic processes*. New York: John Williams & Sons.
7. Shaked, M. & Shanthikumar, J.G. (2007). *Stochastic orders*. New York, USA: Springer.
8. Spizzichino, F. (2001). *Subjective Probability Models for Lifetimes*. Boca Raton, FL, USA: Chapman and Hall/CRC.
9. Suen, W. Monotone likelihood ratio property. Available at http://www.sef.hku.hk/wsuen/teaching/uncertainty/mlrp.pdf
10. Zacks, S. (1971). *The theory of statistical inference*. New York, USA: Wiley & Sons, INC.

https://doi.org/10.1017/S026996481800027X Published online by Cambridge University Press