7-2012

Bayesian Analysis of Hypothesis Testing Problems for General Population: A Kullback–Leibler Alternative

Naveen K. Bansal
Marquette University, naveen.bansal@marquette.edu

Gholamhossein Hamedani
Marquette University, gholamhoss.hamedani@marquette.edu

Ru Sheng
Marquette University

Journal of Statistical Planning and Inference, Vol. 142, No. 7 (July 2012): 1991-1998. DOI.

Bayesian Analysis of Hypothesis Testing Problems for General Population: A Kullback–Leibler Alternative

Naveen K. Bansal
Department of Mathematics, Statistics and Computer Science, Marquette University, Milwaukee, WI
G.G. Hamedani
Department of Mathematics, Statistics and Computer Science, Marquette University, Milwaukee, WI
Ru Sheng
Finance Department, School of Business, University of Connecticut, Storrs, CT

Abstract

We consider a hypothesis problem with directional alternatives. We approach the problem from a Bayesian decision theoretic point of view and consider a situation when one side of the alternatives is more important or more probable than the other. We develop a general Bayesian framework by specifying a mixture prior structure and a loss function related to the Kullback–Leibler divergence. This Bayesian decision method is applied to

Normal and Poisson populations. Simulations are performed to compare the performance of the proposed method with that of a method based on a classical *z*-test and a Bayesian method based on the “0–1” loss.

Keywords

Directional alternatives, Bayes decision rule, Type-III error

1. Introduction

A three-decision problem is a special case of the general decision problem stated by Wald (1951). An example of this type of problem is

(1.1)

$$H_0\colon \theta = \theta_0 \text{vs.} H_-\colon \theta < \theta_0 \text{or} H_+\colon \theta > \theta_0,$$

where ϑ_0 is some known value. Many methods have been considered in the literature to test this hypothesis; see, for example, Jones and Tukey (2000), Lehmann (1950) and Kaiser (1960). All of these methods, however, give equal preferences to both alternative hypotheses H_- and H_+.

We consider a skewness in the alternatives and consider the problem from a Bayesian decision theoretic point of view. We develop alternative statistical procedures by specifying a skewed prior structure and a loss function. For the prior density, we consider the mixture structure

(1.2)

$$\pi(\theta) = p_0 I(\theta = \theta_0) + p_- \pi_-(\theta) + p_+ \pi_+(\theta),$$

where π_- and π_+ are the left and the right tail densities with supports $\{\vartheta:\vartheta<\vartheta_0\}$ and $\{\vartheta:\vartheta>\vartheta_0\}$ respectively. $p_0 = P(\theta = \theta_0), p_- = P(\theta < \theta_0)$ and $p_+=P(\vartheta>\vartheta_0)$ are the prior probabilities of H_0, H_-, and H_+ respectively which can be defined subjectively or estimated from the data. Note that p_0, p_- and p_+ reflect the skewness in the alternatives. So, a prior belief that one tail is more likely than the other can be reflected through appropriate choice of p_- and p_+.

For the loss function we use Kullback–Leibler divergence which was first introduced in the famous paper “On Information and Sufficiency” (Kullback and Leibler, 1951). Kullback–Leibler divergence measures the dissimilarity between two probability distributions. It is defined by

(1.3)

$$D(f_1, f_2) = \int f_1(x) \log \frac{f_1(x)}{f_2(x)} dx$$

where $f_1(x)$ and $f_2(x)$ are the two pdfs with the same support. For discrete distributions, the integral above is replaced by a sum.

The Kullback–Leibler divergence has been used for many aspects of statistical inferences. Primarily, it is used for goodness of fit test; see, for example, Ebrahimi et al. (1992) and Arizono and Ohta (1989). For some other applications, see Burnham and Anderson (2001) and Reschenhofer (1999). Kullback–Leibler divergence has been also used in Bayesian setting for constructing prior and loss functions. Walker et al. (2004) used it to construct prior by assigning positive mass to Kullback–Leibler neighborhoods of certain densities. Hall (1987) and George et al. (2006) used it for defining loss functions. In this paper, we use it to construct a loss function for the hypothesis problem (1.1).

The rest of the paper is structured as follows. In Section 2, we give a general formulation of the Bayesian decision theoretical approach to hypothesis testing problem (1.1) with skewed alternatives and develop a Bayesian methodology under a Kullback–Leibler loss. In Section 3, we consider the normal population, and give a frequentist's comparison through simulation of the Bayes rule under Kullback–Leibler loss, Bayes rule under "0–1" loss and a rule based on a classical test statistic. In Section 4, we develop Bayesian decision theoretic methodology for the Poisson population under Kullback–Leibler loss. Section 5 deals with summary and conclusion.

2. Bayesian decision theoretical formulation

Let $\boldsymbol{X} = (X_1, \dots, X_n)'$ be a random sample from a population having pdf (or pmf) $f(x|\theta, \eta)$, where $\theta \in \mathcal{R}$ is the parameter of interest and η is a nuisance parameter. The prior on ϑ is given by (1.2), where π_- and π_+ might depend on the nuisance parameter η. This prior would then be the conditional prior of ϑ given η. Let the action space be denoted by $A = \{-1, 0, 1\}$,, where the actions $a = 0, a = -1$ and a=1 mean acceptance of H_0, H_-, and H_+ respectively. If $L_\eta(\theta, a)$ denotes the loss for taking action $a \in A$, then it can be seen that the Bayes rule is given by

$$\delta^B(\mathbf{X}) = i \text{ifandonlyif} E[L_\eta(\theta, i)|\mathbf{X}] = \min_{j=-1,0,1} E[L_\eta(\theta, j)|\mathbf{X}].$$

Writing it in the context of hypothesis testing, the above rule can be restated as: Reject H_0 if

(2.1)

$$\max\{E[L_\eta(\theta, -1)|\boldsymbol{x}], E[L_\eta(\theta, 1)|x]\} < E[L_\eta(\theta, 0)|\boldsymbol{x}];$$

and upon rejecting H_0, select H_- or H_+ according to the smaller of $E[L_\eta(\vartheta,-1)|x]$ and $E[L_\eta(\vartheta,1)|x]$. We allow the loss function to depend on the nuisance parameter η which makes sense in some cases as it can be seen through the example of normal distribution. If ϑ is the mean and σ the standard deviation of the normal distribution, then the distances between ϑ values make more sense when they are defined in relation to the standard deviation σ. Note that two different ϑ values that are considered small for a large σ will be considered large for a small σ.

In the presence of a nuisance parameter η, the Bayes rule can be computed by first computing the posterior expected loss with respect to the posterior distribution of ϑ given X=x assuming η is known, and then by computing the posterior expectation with respect to the posterior distribution of η given X=x. Thus, we first discuss the posterior expected loss assuming η known. Later we discuss the unknown η case and recommend replacing η by its posterior mode instead of further computing the posterior expectation. For simplicity of notation, we suppress the symbol η hereafter unless it is necessary.

It is easy to see that the posterior distribution of ϑ given X=x and η is given by

(2.2)

$$\pi(\theta|\boldsymbol{x}) = \pi(H_0|\boldsymbol{x}) I(\theta = \theta_0) + \pi(H_-|\boldsymbol{x}) \pi(\theta|H_-, \boldsymbol{x}) + \pi(H_+|\boldsymbol{x}) \pi(\theta|H_+, \boldsymbol{x}),$$

where $\pi(\vartheta|H_-,x)$ and $\pi(\vartheta|H_+,x)$ are the posterior densities of ϑ with respect to the priors π_(ϑ) and $\pi_+(\vartheta)$ respectively, and

(2.3)

$$\pi(H_0|\boldsymbol{x}) \propto p_0 f(\boldsymbol{x}|\theta_0), \pi(H_-|\boldsymbol{x}) \propto p_- f(\boldsymbol{x}|H_-), \text{and } \pi(H_+|\boldsymbol{x}) \propto p_+ f(\boldsymbol{x}|H_+),$$

where $f(x|H_-)$ and $f(x|H_+)$ are the marginal densities under the priors $\pi_-(\vartheta)$ and $\pi_+(\vartheta)$ respectively, keeping η fixed. Note that the proportionality constant is the inverse of $[p_- f(x|H_-)+p_0 f(x|\vartheta_0)+p_+ f(x|H_+)]$.

Based on the Kullback–Leibler divergence, we consider the following loss function:

(2.4)

$$L(\theta,0)=\begin{cases}0, & \theta=\theta_0\\ q(\theta,\theta_0), & \theta\neq\theta_0\end{cases}, L(\theta,-1)=\begin{cases}0, & \theta<\theta_0\\ l_-+q(\theta,\theta_0), & \theta\geq\theta_0\end{cases}, L(\theta,1)=\begin{cases}0, & \theta>\theta_0\\ l_++q(\theta,\theta_0), & \theta\leq\theta_0\end{cases}$$

where l_- and l_+ are some positive constants and $q(\vartheta,\vartheta_0)$ is the Kullback–Leibler divergence given by

(2.5)

$$q(\theta,\theta_0)=E^{\theta}[\log f(X|\theta)/\log f(X|\theta_0)]$$

The expectation above is with respect to $X\sim f(x|\vartheta)$.

The motivation behind the above loss is that the loss for taking a wrong action should depend on how far is the true value of ϑ. In the case of action $a=-1$ or $a=1$, there is a pre-assigned loss l_- or l_+ when H_0 is true, but the loss increases as the true ϑ moves away from ϑ_0 in the opposite direction. Note also that there is no minimum loss for taking action $a=0$ when H_- or H_+ is true. This way if the true ϑ is close to ϑ_0, we are keeping the possibility that the loss for taking action $a=0$ may not be much.

From (2.2), (2.4), it can be seen that the posterior expected loss functions are given by

(2.6)

$$E[L(\theta,0)|\boldsymbol{x}]=\pi(H_-|\boldsymbol{x})E[q(\theta,\theta_0)|H_-,\boldsymbol{x}]++\pi(H_+|\boldsymbol{x})E[q(\theta,\theta_0)|H_+,\boldsymbol{x}]$$

(2.7)

$$E[L(\theta,-1)|\boldsymbol{x}]=l_-\pi(H_0|\boldsymbol{x})+\pi(H_+|\boldsymbol{x})(l_-+E[q(\theta,\theta_0)|H_+,\boldsymbol{x}])$$

and

(2.8)

$$E[L(\theta,1)|\boldsymbol{x}]=l_+\pi(H_0|\boldsymbol{x})+\pi(H_-|\boldsymbol{x})(l_++E[q(\theta,\theta_0)|H_-,\boldsymbol{x}])$$

If the nuisance parameter η is unknown, then a further expectation is required with respect to the posterior distribution of η given $X=x$ on the right hand sides of the above expressions. This is a difficult task. So, we recommend (at least for large sample size n) under some condition replacing η by its posterior mode rather than taking further expectation. To see why this is reasonable, assume that the density of $X=(X_1,\ldots,X_n)'$ has the following decomposition:

(2.9)

$$f(\boldsymbol{x}|\theta,\eta)=f_1(t|\theta,\eta)f_2(s|\eta),$$

where f_1 is the density of $T=T(X)$ and f_2 is the density of $S=S(X)$.

For illustration, we consider a general case of computing E[g(θ,η)|x], for some measurable function $g(\vartheta,\eta)$. For our purpose, $g(\vartheta,\eta)$ is $L_\eta(\theta,0)$, $L_\eta(\theta,-1)$ or $L_\eta(\vartheta,1)$. Note that, when η is kept fixed, the computation of $E[g(\theta,\eta)|\boldsymbol{x}]$ is based on the posterior distribution of ϑ given η and x, which only requires the conditional prior of ϑ given η and the density of T, f_1. Suppose the resulting expression is $h(\eta|t)$. For an unknown η, a further integration is required with respect to $f_{1m}(t|\eta)f_2(s|\eta)\pi_2(\eta)/f_m(t,s)d\eta$, where f_{1m} is the marginal density of T given η, f_m is the marginal density of (T,S), and π_2 is the prior density of η. If $f_{2m}(s)$ denotes the marginal density of S then $E[g(\vartheta,\eta)|x]$ can be written as

(2.10)

$$E[g(\theta,\eta)|\boldsymbol{x}] = \frac{f_{2m}(s)}{f_m(t,s)}\int h(\eta|t)f_{1m}(t|\eta)\pi_2(\eta|s)d\eta$$

If the posterior $\pi_2(\eta|s)$ is unimodal and $\hat{\eta}$ is its posterior mode, then intuitively $E[g(\theta,\eta)|\boldsymbol{x}]$ can be approximated by

(2.11)

$$E[g(\theta,\eta)|\boldsymbol{x}] \approx \frac{f_{2m}(s)}{f_m(t,s)}f_{1m}(t|\hat{\eta})h(\hat{\eta}|t)$$

provided $|\eta-\hat{\eta}|$ is significantly small with very high posterior probability. Of course, for large sample size n this would be the case as the posterior distribution will converge to the distribution degenerated at $\eta=\hat{\eta}$. For a more formal treatment of this, see Theorem A of Appendix A.

Note that since the Bayes rule is based on comparing the posterior expected loss functions, the expression $((f_{2m}(s))/(f_m(t,s)))f_{1m}(t|\hat{\eta})$ in (2.11) can be ignored, and thus the comparison of the loss functions can be based on just $h(\hat{\eta}|t)$.

From now on, we will assume that if the nuisance parameter η is present and unknown, it will be replaced by the mode of the posterior distribution of η given S=s. Thus, in the expressions in (2.6), (2.7), (2.8), we will substitute $\hat{\eta}$ for η if it is unknown. The Theorem below follows directly by comparing the loss functions given by (2.6), (2.7), (2.8) as suggested in (2.1) with its subsequent arguments. Note that a simple interpretation of the Bayes rule as stated in the Theorem below would not be possible if further expectation in (2.6), (2.7), (2.8) was taken with respect to the posteror distribution of η instead of replacing it by its posterior mode.

Theorem 2.1

Under the prior structure (1.2) *and the loss function* (2.4), *the Bayes rule rejects* H_0 *if*

(2.12)

$$\frac{\pi(H_-|\boldsymbol{x})}{1-\pi(H_-|\boldsymbol{x})} > \frac{l_-}{E[q(\theta,\theta_0)|H_-,\boldsymbol{x}]}$$

or

(2.13)

$$\frac{\pi(H_+|\boldsymbol{x})}{1-\pi(H_+|\boldsymbol{x})} > \frac{l_+}{E[q(\theta,\theta_0)|H_+,\boldsymbol{x}]}$$

and after rejecting H_0, it selects H_- or H_+ according to the smaller of $l_-(1-\pi(H_-|x))-\pi(H_-|x)(E[q(\vartheta,\vartheta_0)|H_-,x])$ and $l_+(1-\pi(H_+|x))-\pi(H_+|x)(E[q(\vartheta,\vartheta_0)|H_+,x])$ respectively.

Note that the left hand sides of (2.6), (2.7) are the posterior odds in favor of H_- and H_+ respectively. To interpret the above theorem, we assume that the posteriors of ϑ with respect to the priors π_- and π_+ are unimodal and peaked at $\hat{\theta}_-$ and $\hat{\theta}_+$ respectively. Then, $E[q(\theta,\theta_0)|H_-,x]$ and $E[q(\vartheta,\vartheta_0)|H_+,x]$ can be intuitively approximated by $q(\hat{\theta}_-,\theta_0)$ and $q(\hat{\theta}_+,\theta_0)$ respectively. Theorem 2.1 thus implies that if $\hat{\theta}_-$ or $\hat{\theta}_+$ is close to ϑ_0, then the posterior odds in favor of H_- or H_+ must be significantly high in order to reject H_0.

A more familiar form of the above Bayes rule is possible if in addition to the decomposition (2.9) we assume that for each fixed η, $f_1(t|\theta,\eta)$ has MLR property in t. Note that, from (2.6), (2.7),

$$l_-[\pi(H_0|\boldsymbol{x})+\pi(H_+|\boldsymbol{x})]<\pi(H_-|\boldsymbol{x})E[q(\theta,\theta_0)|H_-,\boldsymbol{x}],$$

which, from (2.3), can be written asl−[p0f(x|θ0)+p+f(x|H+)]<p−∫−∞0q(θ,θ0)f(x|θ)π−(θ)dθ.

Simplifying this, using (2.9), yields

(2.14)

$$l_-[p_0 f(\boldsymbol{x}|\theta_0)+p_+ f(\boldsymbol{x}|H_+)]<p_-\int_{-\infty}^{0}q(\theta,\theta_0)f(\boldsymbol{x}|\theta)\pi_-(\theta)d\theta.$$

Since for $\vartheta<\vartheta_0$,$f_1(t|\vartheta,\eta)/f_1(t|\vartheta_0,\eta)$ is monotonically increasing in t and for $\vartheta>\vartheta_0$ it is monotonically decreasing in t, the left hand side of (2.14) is monotonically increasing in t. Thus, (2.14) can be equivalently written as $t>k_+$ for some k_+ depending on θ_0,η and (p_-,p_0,p_+). Similarly, it can be shown that $E[L(\vartheta,1)|x]<E[L(\vartheta,0)|x]$ if $t<k_-$ for some k_- depending on θ0,η and (p_-,p_0,p_+). This proves the following result.

Theorem 2.2

Under the same conditions as in Theorem 2.1*, if for each fixed η, $f_1(t|\vartheta,\eta)$ has MLR property in t, then the Bayes rule rejects H_0 if $t<k_-$ or if $t>k_+$; and if H_0 is rejected, it selects H_- if $t<k_-$ or selects H_+ if $t>k_+$. Here, k_- and k_+ are the unique solutions of*

(2.15)

$$p_-\int_{-\infty}^{\theta_0}q(\theta,\theta_0)\frac{f_1(k|\theta,\eta)}{f_1(k|\theta_0,\eta)}\pi_-(\theta)d\theta-p_+\int_{\theta_0}^{\infty}\frac{f_1(k|\theta,\eta)}{f_1(k|\theta_0,\eta)}\pi_+(\theta)d\theta=p_0 l_-$$

and

(2.16)

$$p_+\int_{\theta_0}^{\infty}q(\theta,\theta_0)\frac{f_1(k|\theta,\eta)}{f_1(k|\theta_0,\eta)}\pi_+(\theta)d\theta-p_-\int_{-\infty}^{\theta_0}\frac{f_1(k|\theta,\eta)}{f_1(k|\theta_0,\eta)}\pi_-(\theta)d\theta=p_0 l_+$$

for k, respectively.

A further simplification is possible if we take π_+ and π_- to be truncated densities of a symmetric distribution as we will see in the next section.

3. Normally distributed population

Suppose that $\mathbf{X} = (X_1, \cdots, X_n)'$ and $X_1, \cdots, X_n i.i.d. \sim N(\theta, \sigma^2)$. It can be seen that

$$E[\log \frac{f(X,\theta_0)}{f(X,\theta)}] = \frac{(\theta-\theta_0)^2}{2\sigma^2}.$$

For simplicity, we assume that ϑ_0=0. Thus the Kullback–Leibler loss function is the following:

(3.1)

$$L(\theta,0) = \begin{cases} 0, & \theta = 0 \\ \frac{\theta^2}{2\sigma^2}, & \theta \neq 0 \end{cases}, L(\theta,-1) = \begin{cases} 0, & \theta < 0 \\ l_- + \frac{\theta^2}{2\sigma^2}, & \theta \geq 0 \end{cases}, L(\theta,1) = \begin{cases} 0, & \theta > 0 \\ l_+ + \frac{\theta^2}{2\sigma^2}, & \theta \leq 0 \end{cases}.$$

The joint density of *X* can be decomposed in the form of (2.9) with $t = \bar{x}$, and *s* as the sample standard deviation. Note that the density f_1 of $\bar{X}$ has MLR property in $\bar{x}$. Thus, from Theorem 2.2, the Bayes rule rejects H_0 if $\bar{x} < k_-$ or $\bar{x} > k_+$; and upon rejecting H_0 it selects H_- if $\bar{x} < k_-$ or selects H_+ if $\bar{x} > k_+$. k_- and k_+ are determined by solving (2.15), (2.16) respectively.

Let $\pi(\vartheta)$ be a density symmetric around 0 and let $\pi_-(\vartheta)=2\pi(\vartheta)I(\vartheta<0)$ and $\pi_+(\vartheta)=2\pi(\vartheta)I(\vartheta>0)$.
Then (2.15), (2.16) simplify to

(3.2)

$$p_- \int_{-\infty}^{0} \frac{\theta^2}{\sigma^2} \frac{f_1(k_-|\theta,\sigma^2)}{f_1(k_-|0,\sigma^2)} \pi(\theta) d\theta - 2p_+ \int_0^{\infty} \frac{f_1(k_-|\theta,\sigma^2)}{f_1(k_-|0,\sigma^2)} \pi(\theta) d\theta = p_0 l_-$$

and

(3.3)

$$p_+ \int_0^{\infty} \frac{\theta^2}{\sigma^2} \frac{f_1(k_+|\theta,\sigma^2)}{f_1(k_+|0,\sigma^2)} \pi(\theta) d\theta - 2p_- \int_{-\infty}^{0} \frac{f_1(k_+|\theta,\sigma^2)}{f_1(k_+|0,\sigma^2)} \pi(\theta) d\theta = p_0 l_+$$

If k_- and k_+ are written as functions of p−,p+ and *l*, then due to symmetry of π and f_1 it can be seen that $k_-(p_-, p_+, l_-) = -k_+(p_+, p_-, l_-)$. Thus, only the algorithm to solve k_+ from (3.3) is needed. The solution for k_- can be obtained by switching p_- and p_+ and replacing l_+ by l_-.

If $\pi(\vartheta)$ is the $N(0, \sigma^2/\omega_0)$ density for some known ω_0, (3.3) simplifies to

(3.4)

$$\sqrt{\frac{\omega_0}{n+\omega_0}}\exp\left(\frac{\tilde{k}_+^2}{2\tilde{\sigma}_n^2}\right)\left[\frac{p_+(\tilde{k}_+^2+\tilde{\sigma}_n^2)}{\sigma^2}\Phi\left(\frac{\tilde{k}_+}{\tilde{\sigma}_n}\right)+\frac{p_+\tilde{k}_+\tilde{\sigma}_n}{\sigma^2}\varphi\left(\frac{\tilde{k}_+}{\tilde{\sigma}_n}\right)-2p_-\Phi\left(-\frac{\tilde{k}_+}{\tilde{\sigma}_n}\right)\right]=p_0 l_+,$$

where φ and Φ are the *pdf* and the *cdf* of the standard normal distribution, $\tilde{k}_+ = nk_+/(n+\omega_0)$ and $\tilde{\sigma}_n^2 = \sigma^2/(n+\omega_0)$.

Theorem 3.1

Let $X_1, X_2, \ldots, X_n$ be i.i.d. with $X_i \sim N(\vartheta, \sigma^2)$, where σ^2 is known. Let the prior on ϑ be given by (1.2) *with π_- and π_+ as the truncated densities of* $N(0, \sigma^2/\omega_0)$ distribution. *Then the Bayes rule under Kullback–Leibler loss* (3.1) *rejects H_0 if $\bar{X} < k_-$ or $\bar{X} > k_+$; and upon rejecting H_0, it selects H_- if $\bar{X} < k_-$ or selects H_+ if $\bar{X} > k_+$, where $k_+ = k_+(p_-, p_+, l_+)$ is the unique solution of* (3.4) *and $k_- = -k_+(p_+, p_-, l_-)$.*

3.1. Simulation study

In this simulation study, we perform a frequentist's comparisons of the power of the Bayes rule given in Theorem 3.1 with that of Bayes rule under the "0–1" type loss (Bansal and Sheng, 2010) and the test proposed by Jones and Tukey (2000). We consider the hypothesis problem H_0: $\vartheta=0$ vs H_-: $\vartheta<0$ or H_+: $\vartheta>0$. Jones and Tukey (2000) formulated this problem as testing H_1: $\vartheta\le 0$ vs. H_{1a}: $\vartheta>0$, and H_2: $\vartheta\ge 0$ vs. H_{2a}: $\vartheta<0$ each at level $\alpha/2$. They proved that the resulting test is of Type-III error α. So, if a classic Z-test is used as a test statistic, then Jones and Tukey's method would reject H_0 if $|Z|>z_{\alpha/2}$, and upon rejecting H_0, accept H_- if $Z<-z_{\alpha/2}$ or accept H_+ if $Z>z_{\alpha/2}$.

We simulate 10,000 samples each of size n=100 from Normal (ϑ, 1). Then we compute the number of rejections of H_0 for different ϑ values. For the Bayesian method given in Theorem 3.1, we consider two choices of (p_-, p_0, p_+): $(0.4, 0.4, 0.2)$ and (0.2, 0.4, 0.4). l_- and l_+ are chosen such that the Type-III error of the Bayesian test is the same as α=0.05. Similar adjustment is made for the Bayes rule of Bansal and Sheng (2010). The priors π_- and π_+ are chosen as the truncated $N(0, \omega_0\sigma^2)$ with ω_0=1.

Fig. 1 presents the power comparisons for $(p_-, p_0, p_+) = (0.4, 0.4, 0.2)$ and $(p_-, p_0, p_+) = (0.2, 0.4, 0.4)$ respectively for the sample size of n=100. In both cases, Bayes rule based on Kullback–Leibler loss perform better than the Bayes rule under "0–1" type loss or Jones and Tukey's test in the regions of ϑ which are more probable. When $p_->p_+$, then ϑ with negative values are more probable, and the simulation shows that the Bayes rule under Kullback–Leibler loss is more powerful in the negative region of ϑ. The same is observed for the case $p_+>p_-$.

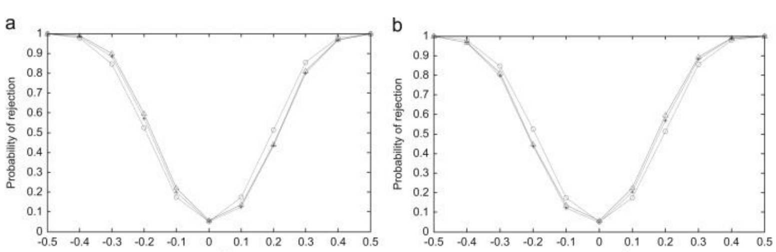

Fig. 3.1. Power comparison for the sample size n=100 when (a) $p_- = 0.4, p_0 = 0.4, p_+ = 0.2$, (b) $p_- = 0.2, p_0 = 0.4, p_+ = 0.4$. Solid line with triangle ($-\Delta$): Power curve of Bayes rule with Kullback–Leibler loss; Dotted line with asterisk ($\cdot\cdot\ast$): Power curve of Bayes rule with "0-1" type loss; Dotted line with circle ($\cdot\cdot\circ$): Power curve of the classical Z-test.

We repeat the above comparison for a smaller sample size of n=30. The results are presented in Fig. 2. We still find that when $p_->p_+$, both Bayesian approaches perform better than classical Z-test in terms of power in the more probable region of ϑ<0. However, unlike the large sample size case, the Bayesian method for “0–1” loss yields higher powers than the Kullback–Leibler loss in this left region. The same is observed for $p_+ > p_-$.

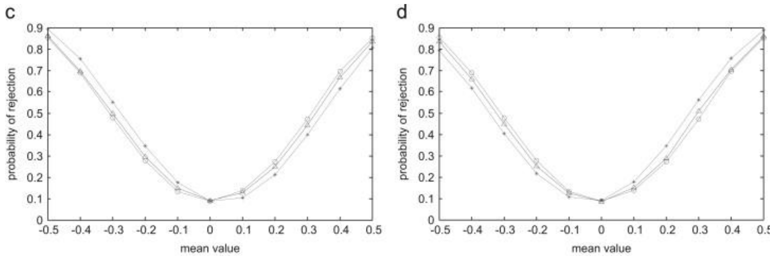

Fig. 3.2. Power comparison for the sample size *n*=30 when (c) $p_- = 0.4, p_0 = 0.4, p_+ = 0.2$, (d) $p_- = 0.2, p_0 = 0.4, p_+ = 0.4$. Solid line with triangle (−*Δ*): Power curve of Bayes rule with Kullback–Leibler loss; Dotted line with asterisk (··∗): Power curve of Bayes rule with “0–1” type loss; Dotted line with circle (··○): Power curve of the classical *Z*-test.

4. Poisson population

Consider the Poisson population as another example. Suppose that $X = (X_1, \cdots, X_n)'$ and $X_1, \cdots, X_n iid \sim P(\lambda)$. The three-decision hypothesis problem is H_0: $\lambda=\lambda_0$ vs. H_-: $\lambda<\lambda_0$ or H_+: $\lambda>\lambda_0$, where λ_0 is some known constant. After reparametrizing the parameter λ in terms of $\vartheta=\log(\lambda/\lambda_0)$, the above hypothesis problem is restated as H_0:$\vartheta=0$ vs. H_-:$\vartheta<0$ or H_+:$\vartheta>0$. The Kullback–Leibler divergence function $q(\vartheta,0)$ is given by

(4.1)

$$q(\theta, 0) = E[\log \frac{f(X|\theta)}{f(X|0)}] = \lambda_0[(\theta - 1)\exp(\theta) + 1]$$

The density of $X = (X_1, \dots, X_n)'$ can be decomposed in the form of (2.9) with $T=\Sigma X_i$. The density of T is given by

(4.2)

$$f_1(t|\theta) = \frac{\exp(t\theta - n\lambda_0\exp(\theta))}{t!}, t = 0,1, \dots$$

For the prior, we choose π_- and π_+ as the truncated distributions of the double exponential distribution symmetric about 0. Thus, $\pi_-(\theta) = \beta\exp(\beta\theta)I(\theta < 0)$ and $\pi_+(\theta) = \beta\exp(-\beta\theta)I(\theta > 0)$, for some constant $\beta > 0$.

Since $f_1(t|\vartheta)$ has MLR property in $t=\Sigma x_i$, Theorem 2.2 can be applied. From (4.1), (4.2), it can be seen that

$$\int_0^\infty q(\theta,\theta_0)\frac{f_1(k|\theta,\eta)}{f_1(k|\theta_0,\eta)}\pi_+(\theta)d\theta = \lambda_0\beta\exp(n\lambda_0)\int_1^\infty (u\log u - u + 1)u^{k-\beta-1}e^{-n\lambda_0 u}du$$

and

$$\int_{-\infty}^0 \frac{f_1(k|\theta,\eta)}{f_1(k|\theta_0,\eta)}\pi_-(\theta)d\theta = \beta\exp(n\lambda_0)\int_0^1 u^{k+\beta-1}e^{-n\lambda_0 u}du.$$

$k_+=k_+(p_-,p_+,l_+)$ now can be solved by (2.16), and k_- can be solved similarly using (2.15). The following theorem gives the complete procedure.

Theorem4.1

Let $X_1, X_2, \ldots, X_n$ be i.i.d. with $X_i \sim P(\lambda)$. If the problem is to test $H_0{:}\lambda=\lambda_0$ vs. $H_-{:}\lambda<\lambda_0$ or $H_+{:}\lambda>\lambda_0$ and if the prior density is (1.2) with $\pi_-(\lambda)=(\beta/\lambda_0)(\lambda/\lambda_0)^{\beta-1}I(\lambda\le\lambda_0)$ and $\pi_+(\lambda)=(\beta/\lambda_0)(\lambda/\lambda_0)^{-\beta-1}I(\lambda>\lambda_0)$ for some $\beta>0$, then the Bayes rule under the Kullback–Leibler loss rejects H_0 if $\sum X_i<k_-$ or $\sum X_i>k_+$; and upon rejecting H_0, it selects H_- if $\sum X_i<k_-$ or selects H_+ if $\sum X_i>k_+$, where k_+ and k_- are the unique solutions of

$$\beta\exp(n\lambda_0)[\lambda_0 p_+\int_1^\infty(u\mathrm{log}u-u+1)u^{k-\beta-1}\mathrm{e}^{-n\lambda_0\mathrm{u}}\mathrm{du}-\mathrm{p}_-\int_0^1 u^{k+\beta-1}e^{-n\lambda_0 u}du] = p_0 l_+$$

and

$$\beta\exp(n\lambda_0)[\lambda_0 p_-\int_0^1(u\mathrm{log}u-u+1)u^{k-\beta-1}\mathrm{e}^{-n\lambda_0\mathrm{u}}\mathrm{du}-\mathrm{p}_+\int_1^\infty u^{k+\beta-1}e^{-n\lambda_0 u}du] = p_0 l_-$$

respectively.

4.1. Simulation study

To illustrate the performance of the Bayesian approach for different loss functions, we conduct the similar power comparisons as in Section 3. We perform a frequentist's comparisons of the power of the Bayes rule given in Theorem 4.1 with that of Bayes rule under the “0–1” type loss (Bansal and Sheng, 2010) and the test proposed by Jones and Tukey (2000). For Poisson population, we consider the hypothesis problem H_0: $\lambda=\lambda_0$ vs. H_-: $\lambda<\lambda_0$ or H_+: $\lambda>\lambda_0$. The procedures of these two methods are the same as in Section 3.

We simulate 1000 samples each of size n=100 from Poisson (λ). Then we compute the number of rejections of H_0 for different λ values, from 0.4 to 1.4. For the Bayesian method given in Theorem 4.1, we choose $(p_-, p_0, p_+) as (0.8, 0.1, 0.1)$. l_- and l_+ are chosen such that the Type-III error of the Bayesian test is the same as α=0.05. Similar adjustment is also made for the Bayes rule of Bansal and Sheng (2010). The prior densities are $\pi_-(\lambda)=(\beta/\lambda_0)(\lambda/\lambda_0)^{\beta-1}I(\lambda\le\lambda_0)$ and $\pi_+(\lambda)=(\beta/\lambda_0)(\lambda/\lambda_0)^{-\beta-1}I(\lambda>\lambda_0)$ for β=0.1.

Fig. 3 presents the power comparisons for $(p_-, p_0, p_+)=(0.8, 0.1, 0.1)$. It indicates that Bayes rule based on Kullback–Leibler loss perform better than the Bayes rule under “0–1” type loss or Jones and Tukey's classical test in the region of λ which is more probable. Here $p_->p_+$, therefore λ with values less than λ_0=0.9 are more probable, and the simulation shows that the Bayes rule under Kullback–Leibler loss is more powerful in the negative region of ϑ.

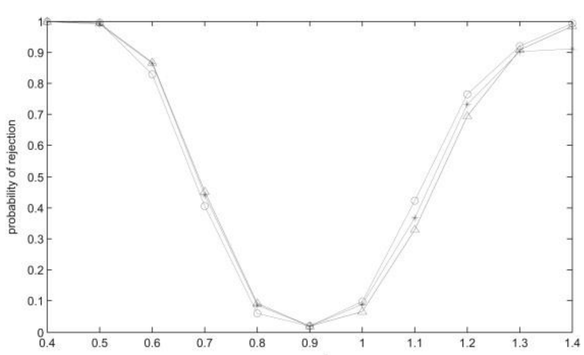

Fig. 4.1. Power comparison for the sample size *n*=100 when $p_- = 0.8, p_0 = 0.1, p_+ = 0.1$. Solid line with triangle (−*Δ*): Power curve of Bayes rule with Kullback–Leibler loss; Dotted line with asterisk (··∗): Power curve of Bayes rule with "0–1" type loss; Dotted line with circle (··○): Power curve of the classical *Z*-test.

5. Conclusion

The goal of this paper is to develop a new statistical methodology for hypothesis testing problems with skewed alternatives. In three-decision problems, the skewness in alternatives is manifested in many practical situations where one side of the alternative hypotheses is more likely to occur when the null is not true. We built a general formulation of this problem in a Bayesian decision theoretic framework. The skewness is expressed by specifying a mixture prior structure.

Since the "0–1" loss is not reasonable, we propose a new loss function based on Kullback–Leibler divergence. 3 Normally distributed population, 4 Poisson population demonstrate this methodology for Normal and Poisson distributions. By comparing the proposed Bayesian method with classical tests through simulation in Section 3 and Section 4, we have shown that the proposed method performs better than classical tests in terms of power in the more probable side of the parameter. When we further compare the Kullback–Leibler loss function and the "0–1" type loss function, the Bayes rule with the Kullback–Leibler loss function yield a better power in the more probable region of the parameter.

Acknowledgment

The authors wish to thank an anonymous referee for careful reading of the paper and for making some valuable suggestions that helped us to improve its original version.

Appendix A. Estimation of $E[g(\vartheta,\eta)]$

From (2.10),

$$E[g(\theta,\eta)|\boldsymbol{x}] = \frac{f_{2m}(s)}{f_m(t,s)} \int h(\eta|t) f_{1m}(t|\eta)\pi_2(\eta|s)d\eta$$

We assume that the density of $\pi_2(\eta|s)$ is unimodel and peaked at the posterior mode $\hat{\eta}_n$. If $h(\eta|t)$ and $f_{1m}(t|\eta)$ are continuous in η, $h(\eta|t)f_{1m}(t|\eta)$ is bounded for every t, and if for every $\epsilon > 0$, the posterior probability $P(||\eta - \hat{\eta}_n| < \epsilon|s)$ converges to 1 as $n \to \infty$, then

(A.1)

$$\int h(\eta|t) f_{1m}(t|\eta)\pi_2(\eta|s)d\eta - h(\hat{\eta}_n|t) f_{1m}(t|\hat{\eta}_n) \to 0$$

as $n \to \infty$. If this is true, then $E[g(\vartheta,\eta)|x]$ can be approximated by plugging in $\hat{\eta}_n$ for η, i.e.

$$E[g(\theta,\eta)|\boldsymbol{x}] \approx \frac{f_{2m}(s)}{f_m(t,s)} f_{1m}(t|\hat{\eta})h(\hat{\eta}|t)$$

Note that the conditions stated above are satisfied for the problem considered in Section 3.

Since $h(\eta|t)$ and $f_{1m}(t|\eta)$ are continuous in η, for every $\delta > 0$, there is an $\epsilon > 0$ which may depend on t such that $|\eta - \hat{\eta}_n| < \epsilon$ implies that

(A.2)

$$h(\eta|t)f_{1m}(t|\eta) - h(\hat{\eta}_n|t)f_{1m}(t|\hat{\eta}_n)| < \delta$$

From (A.1),

$$\begin{aligned}|\int h(\eta|t)f_{1m}(t|\eta)\pi_2(\eta|s)d\eta - h(\hat{\eta}_n|t)f_{1m}(t|\hat{\eta}_n)| \\ \leq \int_{|\eta-\hat{\eta}_n|<\epsilon} |h(\eta|t)f_{1m}(t|\eta) - h(\hat{\eta}_n|t)f_{1m}(t|\hat{\eta}_n)|\pi_2(\eta|s)d\eta \\ + \int_{|\eta-\hat{\eta}_n|\geq\epsilon} |h(\eta|t)f_{1m}(t|\eta) - h(\hat{\eta}_n|t)f_{1m}(t|\hat{\eta}_n)|\pi_2(\eta|s)d\eta\end{aligned}$$

From (A.2), the first part of the right hand side is less than $\delta P(|\eta - \hat{\eta}_n| < \epsilon)$; and due to boundedness of $h(\eta|t)f_{1m}(t|\eta)$, the second part of the right hand side is less than or equal to $2C(t)P(|\eta - \hat{\eta}_n| \geq \epsilon)$, where $C(t)$ is an upper bound of $h(\eta|t)f_{1m}(t|\eta)$. Thus

$$|\int h(\eta|t)f_{1m}(t|\eta)\pi_2(\eta|s)d\eta - h(\hat{\eta}_n|t)f_{1m}(t|\hat{\eta}_n)| \leq \delta P(|\eta - \hat{\eta}_n| < \epsilon) + 2C(t)P(|\eta - \hat{\eta}_n| \geq \epsilon)$$

Since $P(|\eta - \hat{\eta}_n| \geq \epsilon) \to 0$ as $n \to \infty$, as assumed, we get for every $\delta>0$

$$|\int h(\eta|t)f_{1m}(t|\eta)\pi_2(\eta|s)d\eta - h(\hat{\eta}_n|t)f_{1m}(t|\hat{\eta}_n)| \leq \delta \text{as} n \to \infty.$$

This proves the desired result.

References

Arizono and Ohta, 1989. I. Arizono, H. Ohta. **A Test for normality Based on Kullback–Leibler information.** *The American Statistician*, 43 (1) (1989), pp. 20-22. (February)

Bansal and Sheng, 2010. N.K. Bansal, R. Sheng. **Bayesian decision theoretic approach to hypothesis problems with skewed alternatives.** *Journal of Statistical Planning and Inference*, 140 (10) (2010), pp. 2894-2903. (October)

Burnham and Anderson, 2001. K.P. Burnham, D.R. Anderson. **Kullback–Leibler information as a basis for strong inference in ecological studies.** *Wildlife Research*, 28 (2001), pp. 111-119

Ebrahimi et al., 1992. N. Ebrahimi, M. Habibullah, E.S. Soofi. **Testing exponentiality based on Kullback–Leibler information.** *Journal of the Royal Statistical Society. Series B (Methodological),* 54 (3) (1992), pp. 739-748

George et al., 2006. W.I. George, F. Liang, X. Xu. **Improved minimax predictive densities under Kullback–Leibler loss.** *The Annals of Statistics*, 34 (1) (2006), pp. 78-91. (February)

Hall, 1987. P. Hall. **On Kullback–Leibler loss and density estimation.** *The Annals of Statistics*, 15 (4) (1987), pp. 1491-1519

Jones and Tukey, 2000. L.V. Jones, J.W. Tukey. **A sensible formulation of the significant test.** *Psychological Methods*, 5 (2000), pp. 411-414

Kaiser, 1960. H.F. Kaiser. **Directional statistical decision.** *Psychological Review*, 67 (1960), pp. 160-167

Kullback and Leibler, 1951. S. Kullback, R.A. Leibler. **On information and sufficiency.** *The Annals of Mathematical Statistics*, 22 (1) (1951), pp. 79-86

Lehmann, 1950. E.L. Lehmann. **Some principles of the theory of testing hypotheses.** *The Annals of Mathematical Statistics*, 20 (1) (1950), pp. 1-26. (March)

Reschenhofer, 1999. E. Reschenhofer. **Improved estimation of the expected Kullback–Leibler discrepancy in case of misspecification.** *Econometric Theory*, 15 (3) (1999), pp. 377-387. (June)

Wald, 1951. A. Wald. **Statistical decision functions.** *Mathematics Magazine*, 24 (5) (1951), pp. 283-285. (May–June)

Walker et al., 2004. S. Walker, P. Damien, P. Lenk. **On priors with a Kullback–Leibler property.** *Journal of the American Statistical Association*., 99 (466) (2004), pp. 404-408. (April)