Empirical Likelihood Based Posterior Expectation: from nonparametric posterior means via double empirical Bayesian estimators to nonparametric versions of the James-Stein estimator

BY A. VEXLER, G. TAO AND A. D. HUTSON

Department of Biostatistics, University at Buffalo, State University of New York, Buffalo, New York 14214, USA

avexler@buffalo.edu getao@buffalo.edu ahutson@buffalo.edu

SUMMARY

Posterior expectation is a well-accepted method for data analysis via Bayesian inference based on parametric likelihoods. In this paper we propose utilizing empirical likelihood (EL) methodology to develop novel nonparametric posterior expectation. The parametric Bayesian methodology contains the empirical Bayes approach for the purpose of using the observed data to estimate parameters, or even functional forms, of prior distributions. We adapt this approach to nonparametric posterior estimation based on EL's, defining *double empirical posterior estimators*. The proposed methodology yields a nonparametric analog of the well-known James-Stein estimation that has been well addressed in the literature dealing with multivariate-normal observations. Calculation of the parametric posterior expectation is often intractable due to the complex integral forms. The classical Bayesian literature carries out relevant approximations to the expectation. We also establish accurate asymptotic approximations to the proposed EL posterior expectations. These approximations are shown to be similar to those related to parametric Bayesian point estimators. We show that in many cases the proposed estimators are more efficient than the classic nonparametric procedures, even when non-informative priors are utilized. When forms of the nonparametric posterior estimators use informative priors the proposed nonparametric estimation generally outperforms the relevant maximum likelihood estimation. Extensive Monte Carlo evaluations confirm the efficiency of the proposed methodology, especially when the underlying data distribution is skewed. We apply the proposed method to analyze thiobarbituric acid reaction substances data from a case-control myocardial infarction study, thus showing the excellent applicability of the developed technique.

Some key words: Empirical Bayes methods; Empirical likelihood; James-Stein estimator; Nonparametric estimation; Posterior expectation.

1. INTRODUCTION

Bayesian posterior expectation is a powerful approach used by researchers to characterize an important dimension of posterior and predictive distributions (e.g., Tierney et al. 1989). The posterior expectation serves as a Bayesian analogue for the commonly used frequentist techniques of point estimation (e.g., Carlin & Louis 2000). The posterior expectation efficiently incorporates information from prior distributions and likelihood functions based on the observed data. The traditional Bayesian based methodology assumes a parametric form for the likelihood based on data. It may be desirable in the Bayesian framework to develop a nonparametric approach relative to the traditional likelihood construction. In this case we should require that the posterior distribution, based on a nonparametric likelihood, should obey the laws of probability in a context that corresponds to statements derived from Bayes' rule (for details, see Monahan & Boos 1992).

Lazar (2003) showed that the empirical likelihood (EL) technique (e.g., Owen 2001) can provide a proper likelihood that can serve as the basis for robust and accurate Bayesian inference. The key idea is to substitute the parametric likelihood (PL) with the EL in the Bayesian likelihood construction relative to the component of the likelihood used to model the observed data. This approach can provide a robust nonparametric data-driven alternative to the more classical Bayesian procedures.

In this paper we apply the general theoretical framework of Lazar (2003) to propose and examine a distribution free approach for obtaining the posterior expectation. This approach relaxes the need to assume a parametric form for the underlying data distribution and provides posterior estimators incorporating information from prior distributions and observed data. The proposed method is shown to produce nonparametric estimators that are generally more efficient, in the context of corresponding variances comparisons, than the classic nonparametric procedures.

The statistical literature displays that in the general case parametric-based posterior expectations are difficult to calculate analytically (e.g., Newton & Raftery 1994, DiCiccio et al. 1997, Sweeting 1995, Lieberman 1994, Polson 1991, Tierney & Kadane 1986, Kass & Vaidyanathan 1992). In some elementary cases, e.g. the exponential family of data distributions given a set of conjugate priors, the integrals used in the posterior expectation calculations might be evaluable analytically. However, this is typically not the case. In general, the relevant integrals of interest are intractable and therefore need to be evaluated using numerical methods (e.g., DiCiccio et al. 1997, Tierney et al. 1989, Erkanli 1994, Miyata 2004). In the case involving integrals of posterior distributions that incorporate EL functions the integrands have no analytical forms and must be computed using numerical methods at each value of the functions' arguments. (Regarding the EL functional forms, see, e.g., Owen 2001, Lazar & Mykland 1998, Vexler et al. 2009, Vexler et al. 2012, Yu et al. 2011). This increases the complexity of calculations related to the proposed estimators, especially when the nonparametric procedures are based on relatively large samples.

Tierney & Kadane (1986) developed an easily computable asymptotic approximation for the parametric posterior expectation using the Laplace method. Another key piece of the research developed in this note is the derivation of asymptotic approximations to the proposed nonparametric posterior expectations. We demonstrate the asymptotic propositions are very accurate and have a direct analogy to those of parametric posterior-based procedures.

In this paper we show that the corresponding variances of the proposed estimation procedures incorporating non-informative priors are generally smaller than those of traditional nonparametric estimators, especially when the underlying data distributions are skewed, e.g., when the data follows a log-normal distribution. When informative priors are incorporated into the EL-based form of the posterior likelihood the proposed distributions-free estimators generally have variances smaller than those of their classic MLE counterparts.

In various Bayesian scenarios, prior functions are known up to a given set of parameters. The empirical Bayes method uses the observed data to estimate the prior parameters, e.g., by maximizing the marginal distributions, and then proceeds as though the prior were known (e.g., Carlin & Louis 2000). In this paper, we propose to use EL's as substitutes for PL's in the empirical Bayesian posterior procedure. The distribution-free estimators obtained via this manner are denoted as *double empirical Bayesian point estimators*.

In the case of multivariate normally distributed data Stein (1956) proved that when the dimension of the observed vectors is greater than or equal to three, the MLE's are inadmissible estimators of the corresponding parameters. James & Stein (1961) provided another estimator that yields the frequentist risk (MSE) no larger than that of the MLE's. Efron & Morris (1972) showed that the James-Stein estimator belongs to a class of parametric empirical Bayes (PEB) point estimators in the Gaussian/Gaussian model. In this context, we infer and illustrate in this note that the proposed double empirical Bayesian point estimators can lead to nonparametric versions of the James-Stein estimators when normal priors with unknown parameters are utilized.

This paper is organized as follows: In Section 2 we define and evaluate the nonparametric posterior expectations of simple functionals. In Section 3 we extend the results of Section 2 to evaluate the nonparametric posterior expectations of general functionals. The nonparametric version of the James-Stein estimator is proposed in Section 4. In Section 5 we carry out a Monte Carlo (MC) study to demonstrate the relative efficiency of the proposed methods. In Section 6 we apply the proposed estimators to a real data study of myocardial infarction death. In Section 6 we demonstrate the applicability of the proposed nonparametric estimation procedure. We conclude with remarks in Section 7. Proofs corresponding to the theoretical results presented in this paper are outlined in the Appendix.

2. NONPARAMETRIC POSTERIOR EXPECTATIONS OF SIMPLE FUNCTIONALS

Let $X_1, \ldots, X_n$ be independent identically distributed observations from a distribution function $F(x \mid \theta)$, where θ is the parameter to be evaluated. For convenience of exposition and without loss of generality we assume the parameter θ is one-dimensional. The Bayesian point estimator of θ can be defined as the posterior expectation

$$\hat{\theta} = \frac{\int \theta \prod_{i=1}^{n} f(X_i \mid \theta)\pi(\theta)d\theta}{\int \prod_{i=1}^{n} f(X_i \mid \theta)\pi(\theta)d\theta}, \tag{1}$$

where f is the density function of X_i, $i = 1, \ldots, n$ and $\pi(\theta)$ is the prior distribution. The estimator at (1) utilizes the PL, $\prod_{i=1}^{n} f(X_i \mid \theta)$, provided that the form of f is known.

Lazar (2003) showed that the EL technique can provide proper non-parametric likelihoods that can serve as the basis for Bayesian inference, supplying robustness to relative to the choice of prior. In this paper we propose using the relevant EL function instead of the PL at (1) in order to obtain the nonparametric posterior expectation. We start with an example of this approach with the straightforward case of the mean. The analysis presented here is relatively clear and has the basic ingredients for more general cases.

Following the EL literature (e.g., Owen 1988, Lazar & Mykland 1998, Vexler et al. 2009, Yu et al. 2011) we define the simple EL function with respect to the mean of $X_1, \ldots, X_n$ as

$$EL_1(\theta) = \max_{0<p_1,\cdots,p_n<1} \{\prod_{i=1}^{n} p_i : \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i X_i = \theta\}.$$

Thus the nonparametric posterior expectation has the form of

$$\hat{\theta} = \frac{\int_{X_{(1)}}^{X_{(n)}} \theta e^{\log EL_1(\theta)}\pi(\theta)d\theta}{\int_{X_{(1)}}^{X_{(n)}} e^{\log EL_1(\theta)}\pi(\theta)d\theta} = \frac{\int_{X_{(1)}}^{X_{(n)}} \theta e^{\log ELR_1(\theta)}\pi(\theta)d\theta}{\int_{X_{(1)}}^{X_{(n)}} e^{\log ELR_1(\theta)}\pi(\theta)d\theta}, \tag{2}$$

where $X_{(1)}, \ldots, X_{(n)}$ are the order statistics based on the sample $X_1, \ldots, X_n$, $ELR_1(\theta) = EL_1(\theta)n^n$ is the EL ratio (e.g., Vexler et al. 2009).

The posterior expectation based on using the PL approach is well addressed in the statistical literature (e.g., DasGupta 2008, Evans & Swartz 1995, Johnson 1970, Tierney & Kadane 1986, Yee et al. 2002). In some elementary cases, e.g. the exponential family of data distributions with conjugate priors, the integrals used in the posterior expectation may be evaluated analytically (e.g., Consonni & Veronese 1992). In general, the integrals used for calculating the posterior mean are intractable and need to be evaluated numerically (e.g., Newton & Raftery 1994, DiCiccio et al. 1997, Sweeting 1995, Lieberman 1994, Polson 1991, Tierney & Kadane 1986, Kass & Vaidyanathan 1992). A useful and accurate approximation for analyzing integrals necessary for Bayesian calculations can be achieved by assuming the posterior density is unimodal or at least dominated by a single mode, such that it is highly peaked about its maximum, which is the posterior mode. In this instance we can expand the log-PL, $\log \prod f(X_i \mid \theta)$, as a quadratic about the MLE of θ. Then, exponentiating it yields approximations to the integrands at (1) that have the normal density-type forms. This method is based on the Laplace method (e.g., Bleistein & Handelsman 2010, Tierney & Kadane 1986).

In our application, the integrands at (2) involve the EL function that has no analytical form and henceforth must be computed using numerical methods at each value of the function's argument (e.g., Owen 2001). This analytical shortcoming increases the complexity of calculations related to the proposed estimator.

In this article we show that nonparametric marginal distributions based on the EL approach behave similarly to those based on parametric likelihoods, i.e., $EL_1(\theta)$ is highly peaked about its maximum value. That is, we can approximate integrals of the forms of $\int \theta EL_1(\theta)\pi(\theta)d\theta$ and $\int EL_1(\theta)\pi(\theta)d\theta$ in a similar manner to the approximations related to the parametric posterior expectations. Towards this end we introduce the following lemma that can be considered as a non-asymptotic alternative to Lemma 1 presented in Qin & Lawless (1994).

LEMMA 1. *Define θ_M to be a root of the equation $n^{-1}\sum_{i=1}^{n} G(X_i, \theta_M) = 0$, where $\partial G(X_i, \theta)/\partial\theta < 0$ (or $\partial G(X_i, \theta)/\partial\theta > 0$), for all $i = 1, \ldots, n$. Then the argument θ_M is a global maximum of the function*

$$W(\theta) = \max\{\prod_{i=1}^{n} p_i : 0 < p_i < 1, \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i G(X_i, \theta) = 0\}$$

that increases and decreases monotonically for $\theta < \theta_M$ and $\theta > \theta_M$, respectively.

For example, when $G(u, \theta) = u - \theta$ we obtain $\theta_M = \bar{X} = n^{-1}\sum_{i=1}^{n} X_i$ and the function $W(\theta) = EL_1(\theta)$. Now, we can obtain the following results that are analogues to the asymptotic propositions that are well addressed in the parametric literatures (e.g., DasGupta 2008, Carlin & Louis 2000).

PROPOSITION 1. *Assume $E \mid X_1 \mid^4 < \infty$, $\int \mid \theta \mid \pi(\theta)d\theta < \infty$ and $\pi(\theta)$ is twice continuously differentiable in a neighborhood of $\bar{X} = n^{-1}\sum_{i=1}^{n} X_i$, then the proposed estimator* (2) *satisfies*

$$\hat{\theta} = \frac{\int \theta \exp\left[-\frac{n(\bar{X}-\theta)^2}{2\sigma_n^2}\right]\pi(\theta)d\theta}{\int \exp\left[-\frac{n(\bar{X}-\theta)^2}{2\sigma_n^2}\right]\pi(\theta)d\theta} + \frac{M_n^3}{\sigma_n^2\, n} + g_n,$$

where $\sigma_n^2 = n^{-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$, $M_n^3 = n^{-1}\sum_{i=1}^{n}(X_i - \bar{X})^3$, $g_n = O_p(n^{-3/2+\varepsilon})$ for all $\varepsilon > 0$ as $n \to \infty$.

COROLLARY 1. *Let $\pi(\theta) = (2\pi\sigma_\pi^2)^{-1/2}\exp[-(\theta - \mu_\pi)^2/2\sigma_\pi^2]$, where μ_π and σ_π^2 are known hyperparameters, and the conditions of Proposition* 1 *hold. Then the posterior expectation at* (2) *can be approximated as*

$$\hat{\theta} = \tilde{\theta} + \frac{M_n^3}{\sigma_n^2 n} + O_p(n^{-3/2+\varepsilon}),$$
$$\tilde{\theta} = \frac{(\mu_\pi \sigma_n^2 + \bar{X}\sigma_\pi^2 n)}{(n\sigma_\pi^2 + \sigma_n^2)} = \frac{(\sigma_\pi^2)^{-1}\mu_\pi}{(\sigma_\pi^2)^{-1} + n(\sigma_n^2)^{-1}} + \frac{n(\sigma_n^2)^{-1}\bar{X}}{(\sigma_\pi^2)^{-1} + n(\sigma_n^2)^{-1}}.$$

The estimator $\tilde{\theta}$ is equivalent to the form of the parametric posterior expectation derived under the Normal/Normal model (e.g., Carlin & Louis 2000). The integral mentioned in Proposition 1 can be sometimes obtained analytically depending upon the form of $\pi(\theta)$. However, following the process of the asymptotic evaluation of the parametric posterior expectations we can easily show the following:

COROLLARY 2. *Under the conditions of Proposition* 1*, let* $\pi(\theta)$ *be a prior function with* $\mid d^3 \log(\pi(\theta))/d\theta^3 \mid < \infty$*, for all* θ*. Then we have the following result:*

$$\hat{\theta} = \frac{n\bar{X} + \sigma_n^2\{\log \pi(\bar{X})\}' - \sigma_n^2\{\log \pi(\bar{X})\}''\bar{X}}{n - \sigma_n^2\{\log \pi(\bar{X})\}''} + \frac{M_n^3}{\sigma_n^2 n} + O_p(n^{-3/2+\varepsilon}), \varepsilon > 0, as\ n \to \infty.$$

Now, we consider the normal prior, $\pi(\theta)$, when μ_π and σ_π^2 are unknown. Following the empirical Bayes concepts (e.g., Carlin & Louis 2000) the unknown hyperparameters can be estimated by, e.g., maximizing the respective marginal distributions. This method can be applied to the nonparametric posterior expectation yielding double empirical posterior estimation. In this case, we define

$$\hat{\theta}_E = \frac{\int \theta \exp\{\log EL_1(\theta)\}\exp\{-(\theta - \hat{\mu}_\pi)^2/2\hat{\sigma}_\pi^2\}d\theta}{\int \exp\{\log EL_1(\theta)\}\exp\{-(\theta - \hat{\mu}_\pi)^2/2\hat{\sigma}_\pi^2\}d\theta}, \tag{3}$$

where $(\hat{\mu}_\pi, \hat{\sigma}_\pi^2) = \arg\max_{\mu,\sigma}[(2\pi\sigma^2)^{-1/2}\int_{-\infty}^{\infty}\exp\{\log EL_1(\theta)\}\exp\{-(\theta-\mu)^2/2\sigma^2\}]$. The next result implies a simple asymptotic form of $\hat{\theta}_E$.

COROLLARY 3. *Assume* $E \mid X_1 \mid^4 < \infty$*, then the posterior expectation* $\hat{\theta}_E$ *satisfies*

$$\hat{\theta}_E = \frac{\hat{\mu}_\pi\sigma^2 + \bar{X}\hat{\sigma}_\pi^2 n}{n\hat{\sigma}_\pi^2 + \sigma^2} + \frac{M_n^3}{\sigma_n^2 n} + O_p(n^{-3/2+\varepsilon})$$
$$= \frac{(\hat{\sigma}_\pi^2)^{-1}\mu_\pi}{(\hat{\sigma}_\pi^2)^{-1} + n(\hat{\sigma}_n^2)^{-1}} + \frac{n(\hat{\sigma}_n^2)^{-1}\bar{X}}{(\hat{\sigma}_\pi^2)^{-1} + n(\hat{\sigma}_n^2)^{-1}} + \frac{M_n^3}{\sigma_n^2\, n} + O_p(n^{-3/2+\varepsilon}),$$

where $\hat{\mu}_\pi = \bar{X}$*,* $\hat{\sigma}_\pi^2 - \max\{0, \sigma_n^2 - \sigma^2\} \to 0$*,* $\sigma^2 = Var(X_1)$*,* $\varepsilon > 0$ *as* $n \to \infty$*.*

The proof of Corollary 3 is technical and follows directly from the proof scheme of Proposition 1. Thus the proof is omitted.

Remark 1. Note that, according to the Central Limit Theorem it follows that $\sqrt{n}(\bar{X} - \theta) \sim N(0, \sigma^2)$, as $n \to \infty$. Then under the conditions of Proposition 1 the nonparametric posterior expectations $\hat{\theta}$ and $\hat{\theta}_E$ have the following asymptotic distributions, respectively, given as $\sqrt{n}[\{\hat{\theta} - M_n^3(\sigma_n^2 n)^{-1}\}(\sigma_n^2 + \sigma_\pi^2 n)(\sigma_\pi^2 n)^{-1} - \mu_\pi\sigma_n^2(\sigma_\pi^2 n)^{-1} - \theta] \sim N(0, \sigma^2)$ and $\sqrt{n}[\{\hat{\theta}_E - M_n^3(\sigma_n^2 n)^{-1}\}(\sigma_n^2 + \hat{\sigma}_\pi^2 n)(\hat{\sigma}_\pi^2 n)^{-1} - \mu_\pi\sigma_n^2(\hat{\sigma}_\pi^2 n)^{-1} - \theta] \sim N(0, \sigma^2)$.

To extend the above results to more general situations, we assume that $D(\theta)$ defines a function of θ and denote the nonparametric posterior expectation of $D(\theta)$ to be

$$\hat{D} = \int D(\theta) e^{EL_1(\theta)} \pi(\theta) d\theta \{ \int e^{EL_1(\theta)} \pi(\theta) d\theta \}^{-1}.$$

PROPOSITION 2. *Under the conditions that* $D(\theta) > 0$, $\int \mid D(\theta) \mid \pi(\theta) d\theta < \infty$, $\mid \{\log D(\theta)\}''' \mid < \infty$, *and* $\mid \{\log \pi(\theta)\}''' \mid < \infty$, *for all* θ, *it can be shown that the nonparametric posterior expectation of* $D(\theta)$, *satisfies, for all* $\varepsilon > 0$,

$$\hat{D} = \int D(\theta) e^{-(\sum X_i - n\theta)^2 / 2n\sigma_n^2} \pi(\theta) d\theta \left\{ \int e^{-(\sum X_i - n\theta)^2 / 2n\sigma_n^2} \pi(\theta) d\theta \right\}^{-1} + \frac{D'(\bar{X}) M_n^3}{\sigma_n^2 n} + O_p(n^{-3/2+\varepsilon}),$$

where $\sigma_n^2 = n^{-1} \sum_{i=1}^n (X_i - \bar{X})^2$, $M_n^3 = n^{-1} \sum_{i=1}^n (X_i - \bar{X})^3$ *as* $n \to \infty$.

The proof of this proposition is similar to that of Proposition 1.

Now, in a similar manner to Tierney et al. (1989) under the assumptions stated above we apply the proof strategies utilized for Proposition 1 and Corollary 2 to show that the posterior expectation of $D(\theta)$ can be approximated by

$$\begin{aligned}
\hat{D} &= D(\bar{X}) \left(\frac{n - \sigma_n^2 \{\log \pi(\bar{X})\}''}{n - \sigma_n^2 \{\log D(\bar{X})\}'' - \sigma_n^2 \{\log \pi(\bar{X})\}''} \right)^{1/2} \\
&\times \exp \Bigg([n\bar{X} + \sigma_n^2 \{\log D(\bar{X})\}' - \sigma_n^2 \{\log D(\bar{X})\}'' \bar{X} + \sigma_n^2 \{\log \pi(\bar{X})\}' - \sigma_n^2 \{\log \pi(\bar{X})\}'' \bar{X}]^2 \\
&\times [n - \sigma_n^2 \{\log D(\bar{X})\}'' - \sigma_n^2 \{\log \pi(\bar{X})\}'']^{-2} \\
&- \left[\frac{n\bar{X} + \sigma_n^2 \{\log \pi(\bar{X})\}' - \sigma_n^2 \{\log \pi(\bar{X})\}'' \bar{X}}{n - \sigma_n^2 \{\log \pi(\bar{X})\}''} \right]^2 - \{\log D(\bar{X})\}' \bar{X} + \frac{\{\log D(\bar{X})\}'' \bar{X}^2}{2} \Bigg) \\
&+ \frac{D'(\bar{X}) M_n^3}{\sigma_n^2 n} + O_p(n^{-3/2+\varepsilon}).
\end{aligned}$$

3. NONPARAMETRIC POSTERIOR EXPECTATIONS OF GENERAL FUNCTIONALS

In order to consider the more general case and extend the results found in Section 2, we begin with the definition of the EL function presented in the form of

$$EL_2(\theta) = \max\{ \prod_{i=1}^n p_i : 0 < p_i < 1, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i G(X_i, \theta) = 0 \},$$

where we assume for simplicity that $\partial G(u, \theta)/\partial\theta > 0$ or $\partial G(u, \theta)/\partial\theta < 0$, for all u, and $E \mid G(X_1, \theta) \mid^4 < \infty$. In this framework the posterior expectation takes the form

$$\hat{\theta} = \frac{\int_{X_{(1)}}^{X_{(n)}} \theta e^{\log EL_2(\theta)} \pi(\theta) d\theta}{\int_{X_{(1)}}^{X_{(n)}} e^{\log EL_2(\theta)} \pi(\theta) d\theta} = \frac{\int_{X_{(1)}}^{X_{(n)}} \theta e^{\log ELR_2(\theta)} \pi(\theta) d\theta}{\int_{X_{(1)}}^{X_{(n)}} e^{\log ELR_2(\theta)} \pi(\theta) d\theta}, \; ELR_2(\theta) = n^n EL_2(\theta). \quad (4)$$

According to Lemma 1, $EL_2(\theta)$ increases and decreases monotonically for $\theta < \theta_M$ and $\theta > \theta_M$, respectively, where θ_M is a root of $n^{-n}\sum_{i=1}^{n} G(X_i, \theta_M) = 0$. Then, utilizing the same technique used in the proof of Proposition 1, we can derive the following result:

PROPOSITION 3. *If* $\int \mid \theta \mid \pi(\theta)d\theta < \infty$ *and* $\pi(\theta)$ *is twice continuously differentiable in a neighborhood of* θ_M*, then the estimator defined at* (4) *has the asymptotic form given by*

$$\hat{\theta} = \frac{\int \theta \exp\left[-\{\sum_{i=1}^{n} G(X_i,\theta)\}^2/2n\sigma_{Gn}^2\right]\pi(\theta)d\theta}{\int \exp\left[-\{\sum_{i=1}^{n} G(X_i,\theta)\}^2/2n\sigma_{Gn}^2\right]\pi(\theta)d\theta} + \frac{M_{Gn}^3}{\sigma_{Gn}^2 n} + g_n,$$

where $\sigma_{Gn}^2 = n^{-1}\sum_{i=1}^{n} G(X_i,\theta)^2$, $M_n^3 = n^{-1}\sum_{i=1}^{n} G(X_i,\theta)^3$, $g_n = O_p(n^{-3/2+\varepsilon})$.

Moreover, if $\pi(\theta)$ is a prior distribution function with $\mid \{\log \pi(\theta)\}''' \mid < \infty$, and $\mid \partial^2 G(X_i,\theta)/\partial\theta^2 \mid < \infty$, for all θ, we have that

$$\hat{\theta} = [\{\sum_{i=1}^{n} \partial G(X_i,\theta_M)/\partial\theta_M\}^2 - \{\log\pi(\theta_M)\}''\sum_{i=1}^{n} G(X_i,\theta_M)^2]^{-1}\Bigg[\theta_M\{\sum_{i=1}^{n}\frac{\partial G(X_i,\theta_M)}{\partial\theta_M}\}^2$$

$$+ \{\log\pi(\theta_M)\}'\sum_{i=1}^{n} G(X_i,\theta_M)^2 - \sum_{i=1}^{n} G(X_i,\theta_M)\sum_{i=1}^{n}\frac{\partial G(X_i,\theta_M)}{\partial\theta_M}$$

$$- \{\log\pi(\theta_M)\}''\sum_{i=1}^{n} G(X_i,\theta_M)^2\theta_M\Bigg] + \frac{M_{Gn}^3}{\sigma_{Gn}^2 n} + O_p(n^{-3/2+\varepsilon}),\ for\ all\ \varepsilon > 0,\ as\ n\to\infty.$$

The proof of this proposition is similar to that of Proposition 1 and Corollary 2.

Remark 2. The nonparametric posterior expectation of $D(\theta)$ defined earlier and given in the more general form as

$$\hat{D} = \frac{\int_{X_{(1)}}^{X_{(n)}} D(\theta)e^{\log EL_2(\theta)}\pi(\theta)d\theta}{\int_{X_{(1)}}^{X_{(n)}} e^{\log EL_2(\theta)}\pi(\theta)d\theta}$$

can be analyzed in a similar manner to that used in Propositions 2 and 3.

Generally, we can define the EL function as

$$EL_3(\theta_1,\ldots,\theta_K) = \max\{\prod_{i=1}^{n} p_i : 0 < p_i < 1, \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i G_k(X_i,\theta_k) = 0, k = 1,\ldots,K\}$$

to propose the estimation

$$\hat{D}_G = \frac{\underbrace{\int\cdots\int}_{K-times} D(\theta_1,\ldots,\theta_K)e^{\log ELR_3(\theta_1,\ldots,\theta_K)}\pi(\theta_1,\ldots,\theta_K)d\theta_1\cdots d\theta_K}{\underbrace{\int\cdots\int}_{K-times} e^{\log ELR_3(\theta_1,\ldots,\theta_K)}\pi(\theta_1,\ldots,\theta_K)d\theta_1\cdots d\theta_K},\ ELR_3 = n^n EL_3.$$

Sections 2 and 3 provide the basic ingredients in order to analyze this complex estimator. For example, without loss of generality and for ease of presentation, we consider $K = 2$, $G_k(X_i,\theta_k) = X_i^k - \theta_k$, $k = 1, 2$ and

$$\hat{D}_G = \frac{\int_{X_{(1)}}^{X_{(n)}}\int_{X_{(1)}^2}^{X_{(n)}^2} D(\theta_1,\theta_2)e^{\log ELR_3(\theta_1,\theta_2)}\pi(\theta_1,\theta_2)d\theta_1 d\theta_2}{\int_{X_{(1)}}^{X_{(n)}}\int_{X_{(1)}^2}^{X_{(n)}^2} e^{\log ELR_3(\theta_1,\theta_2)}\pi(\theta_1,\theta_2)d\theta_1 d\theta_2}.$$

If we assume that $\int\int \mid D(\theta_1,\theta_2) \mid \pi(\theta_1,\theta_2)d\theta_1 d\theta_2 < \infty$ exists, D and π are twice continuously differentiable in neighborhoods of $(\bar{X},\overline{X^2})$, then the following proposition yields the relevant asymptotic result:

PROPOSITION 4. *Assume $E \mid X_1 \mid^4 < \infty$, then the asymptotic approximation to the proposed posterior expectation of $D(\theta_1,\theta_2)$ is given by*

$$
\begin{aligned}
\hat{D}_G = &\int\int D(\theta_1,\theta_2)\exp\Bigg\{-\frac{0.5n(\theta_1-\bar{X})^2}{\sigma_n^2-(\sigma_{XX^2n})^2/\sigma_{X^2n}^2}-\frac{0.5n(\theta_2-\overline{X^2})^2}{\sigma_{X^2n}^2-(\sigma_{XX^2n})^2/\sigma_n^2}\\
&+\frac{n(\theta_1-\bar{X})(\theta_2-\overline{X^2})}{\sigma_{X^2n}^2\sigma_n^2/\sigma_{XX^2n}-\sigma_{XX^2n}}\Bigg\}\pi(\theta_1,\theta_2)d\theta_1 d\theta_2\\
&\times\Bigg[\int\int\exp\Bigg\{-\frac{0.5n(\theta_1-\bar{X})^2}{\sigma_n^2-(\sigma_{XX^2n})^2/\sigma_{X^2n}^2}-\frac{0.5n(\theta_2-\overline{X^2})^2}{\sigma_{X^2n}^2-(\sigma_{XX^2n})^2/\sigma_n^2}\\
&+\frac{n(\theta_1-\bar{X})(\theta_2-\overline{X^2})}{\sigma_{X^2n}^2\sigma_n^2/\sigma_{XX^2n}-\sigma_{XX^2n}}\Bigg\}\pi(\theta_1,\theta_2)d\theta_1 d\theta_2\Bigg]^{-1}\\
&+\frac{J_n}{n}+O_p(n^{-3/2+\varepsilon}),\ J_n=O_p(1),
\end{aligned}
$$

for all $\varepsilon>0$, as $n\to\infty$, where $\overline{X^2}=n^{-1}\sum_{i=1}^n X_i^2$, $\sigma_{X^2n}^2=n^{-1}\sum_{i=1}^n(X_i^2-\overline{X^2})^2$, $\sigma_{XX^2n}=n^{-1}\sum_{i=1}^n(X_i-\bar{X})(X_i^2-\overline{X^2})$ and the term J_n has a complicated form depicted in the equation (A15) *of the Appendix.*

4. NONPARAMETRIC ANALOG OF JAMES-STEIN ESTIMATION

Let us begin by outlining the classic James-Stein estimation process assuming the observations $X_1,\ldots,X_n$, are independent and identically distributed as multivariate normal with corresponding mean vector $\boldsymbol{\theta}=(\theta_1,\ldots,\theta_K)$ and covariance matrix Σ, i.e., $X_i=(X_{i1},\ldots,X_{iK})^T\sim N((\theta_1,\ldots,\theta_K)^T,\Sigma)$, $i=1,\ldots,n$. In this case, Stein (1956) proved that for $K\geq 3$, the MLE of $\boldsymbol{\theta}$ is inadmissible, i.e., there exists another estimator with frequentist risk (MSE) that is less than or equal to that of the MLE. Through the analysis of the quadratic loss function one such dominating estimator was derived by James & Stein (1961). Efron & Morris (1972) showed that the James-Stein estimator belongs to a class of the PEB point estimators related to the Gaussian/Gaussian model.

In Section 2, we showed that when $K=1$ and the prior function is a normal density function the proposed nonparametric posterior expectation is asymptotically equivalent to the parametric posterior expectation derived under assumptions of the Gaussian/Gaussian model. In this section, we assume $X_1,\ldots,X_n$ are independent random vectors, $X_i=(X_{i1},\ldots,X_{iK})^T$, with an unknown distribution, and $E\mid X_{ij}\mid^4<\infty$, $j=1,\ldots,K$, $i=1,\ldots,n$. Under these set of assumptions we propose a nonparametric estimate of the mean $(\theta_1,\ldots,\theta_K)^T$ using the double empirical posterior estimation, in the form of

$$\hat{\theta}_{Ej}=\frac{\int\theta\exp\{\log EL_{4j}(\theta)\}\exp(-\theta^2/2\tilde{\sigma}_\pi^2)d\theta}{\int\exp\{\log EL_{4j}(\theta)\}\exp(-\theta^2/2\tilde{\sigma}_\pi^2)d\theta} \quad (5)$$

with

$$\tilde{\sigma}_{\pi}^2 = \arg\max_{\sigma^2} \sum_{j=1}^{K} \log[(2\pi\sigma^2)^{-1/2} \int \exp\{\log EL_{4j}(\theta)\} \exp\{-(\theta^2/2\sigma^2\} d\theta],$$

where $EL_{4j}(\theta_j) = \max\{\prod_{i=1}^n p_{ij} : 0 < p_{ij} < 1, \sum_{i=1}^n p_{ij} = 1, \sum_{i=1}^n p_{ij} X_{ij} = \theta_j\}$, $j = 1, \ldots, K$. In the following proposition we will show the proposed distribution free estimation is asymptotically equivalent to the parametric version of the James-Stein estimator.

PROPOSITION 5. *For all $\varepsilon > 0$ and as $n \to \infty$ the double empirical posterior estimator* (5) *has the following asymptotic form:*

$$\hat{\theta}_{Ej} = (1 - \frac{K-2}{\sum_{r=1}^K \bar{X}_{.r}^2})\bar{X}_{.j} + \frac{1}{n} \sum_{r=1}^{K} \frac{\sum_{i=1}^n (X_{ir} - \bar{X}_{.r})^3}{\sum_{i=1}^n (X_{ir} - \bar{X}_{.r})^2} + O_p(n^{-3/2+\varepsilon}), \tag{6}$$

where $\bar{X}_{.j} = n^{-1} \sum_{i=1}^n X_{ij}$, $j = 1, \ldots, K$.

The proof of Proposition 5 is technical and follows directly from the steps used to prove Propositions 1 and 4, respectively. Thus the proof is omitted.

5. MONTE CARLO RESULTS

We begin with MC comparisons between the proposed nonparametric posterior expectation (2), its asymptotic forms stated in Proposition 1 and Corollary 1, the classical nonparametric estimator $\bar{X}$ and the corresponding MLE's of $\theta = EX_1$. Towards this end 15,000 MC samples of size $n = 10, 20, 30, 50$ and 75 were generated from both a $N(1,1)$ and $LogNorm(0,1)$ distribution. We focus on normal and lognormal distributed data, since one can show the EL approach is in general very efficient in analyzing normally distributed observations, whereas applications of EL methods can be inaccurate when skewed data are utilized (e.g., see Vexler et al. 2009, Yu et al. 2010). Note that the MLE of $\theta = EX_1$ given data from a lognormal distribution has the form $\hat{\theta} = \exp(\bar{X} + \sigma_n^2/2)$, whereas $\bar{X}$ is the MLE when data follow a normal distribution. Let $\mu = E(X_1)$ be equal to 1 or $\exp(1/2)$ when normal or lognormal samples are used, respectively. We considered the following prior distributions to be utilized for the Bayesian estimator (2): $N(0,1)$, $N(\mu-1, \sigma_\pi^2)$, $N(\mu, \sigma_\pi^2)$, $N(\mu+1, \sigma_\pi^2)$, $\{N(-\mu, \sigma_\pi^2) + N(\mu, \sigma_\pi^2)\}/2$, $U(0, \mu+0.5)$, $U(\mu-0.5, \mu+0.5)$, $U(\mu-0.25, \mu+0.25)$, where $\sigma_\pi = 0.5, 0.1, 0.05$ represents different scenarios depicting our "relative confidence" with respect to our prior information pertaining to the unknown parameter. Note that, to examine posterior distributions based on EL functions, Lazar (2003) used priors of $N(\mu, 1/n)$-type forms for Monte Carlo evaluations, where n denotes the corresponding sample size. In the interest of economy of space the Monte Carlo evaluations obtained using the prior distributions $N(\mu, 0.05^2)$, $N(\mu-1, 0.1^2)$, $N(\mu-1, 0.05^2)$, $N(\mu+1, \sigma_\pi^2)$, $\{N(-\mu, 0.05^2) + N(\mu, 0.05^2)\}/2$ and $U(\mu-0.5, \mu+0.5)$ are not presented in this paper. The results of these experiments confirmed conclusions that are shown in this section.

The MC estimates of the means and variances of the estimators $\bar{X}$ and the MLEs are presented in Table 1. Table 2 provides the MC estimated mean and variance values for the proposed estimator (2) and the relevant asymptotic form from Proposition 1, $(\hat{\theta}\_P1)$ and Corollary 1, $(\hat{\theta}\_C1)$. Regarding the selected prior distributions, we remark that the $N(0,1)$ and $N(0, 0.5^2)$ distributions are supposed to contain "no correct information" about the true values of θ (i.e., this distribution functions are not centered around the true values of the parameter); the priors $N(\mu, 0.5^2)$ and $N(\mu, 0.1^2)$ are centered near the true values of the parameter,displaying

”correct information” about locations of the true values of the parameter with the relatively large and small variances respectively; the prior distributions $\{N(-1, 0.5^2) + N(1, 0.5^2)\}/2$ and $\{N(-1, 0.1^2) + N(1, 0.1^2)\}/2$ can reflect information that the target parameter can be 1 unit $+/-$ within the standard deviations 0.5 and 0.1, respectively. Table 2 shows that when the data are normally distributed and the $N(0, 1)$-prior distribution is used then the variances of $\hat{\theta}$ are comparable to those of $\bar{X}$, the MLE of EX_1 in these scenarios of experiments. The variances of the asymptotic forms $\hat{\theta}\_P1$ and $\hat{\theta}\_C1$ are very close to those of $\hat{\theta}$, from (2) even for the moderately small sample size setting at $n = 20$. When using the prior $N(1, 0.5^2)$, the proposed estimator $\hat{\theta}$ from (2) performs significantly better than $\bar{X}$. When $n = 10$, the variance of $\hat{\theta}$ is 42% smaller than that of $\bar{X}$. As n increases, the variance of $\hat{\theta}$ becomes close to that of $\bar{X}$. The above conclusions are magnified when the prior $N(1, 0.1^2)$ is utilized. When using the improper prior, $N(0, 0.5^2)$, the variance of $\hat{\theta}$ is about 27% greater than that of $\bar{X}$ for samples of size $n = 10$. However, when n is large, the variances of $\hat{\theta}$ are comparable to those of $\bar{X}$. When the prior distribution $\{N(-1, 0.5^2) + N(1, 0.5^2)\}/2$ is utilized, the variance of $\hat{\theta}$ is about 35% smaller than that of $\bar{X}$ for samples of size $n = 10$. As the sample size increases the variance of $\hat{\theta}$ is comparable to that of $\bar{X}$. These conclusions, relative to the gains in efficiency, are strongly observed when $\{N(-1, 0.1^2) + N(1, 0.1^2)\}/2$ is utilized as the prior distribution. When a non-informative uniform prior distribution is used, e.g., $U(0, 1.5)$, the variance of $\hat{\theta}$ is about 38% smaller than that of $\bar{X}$ for samples of size $n = 10$. When n increases, the variance of $\hat{\theta}$ becomes close to that of $\bar{X}$. When an uniform prior centered near the true parameter value is used, e.g., $U(0.75, 1.25)$, the variance of $\hat{\theta}$ is about 94% smaller than that of $\bar{X}$ for samples of size $n = 10$.

In the case where data have the assumed lognormal distribution and the proposed estimator is based on the prior distribution, $N(0, 1)$, we have that the variance of $\hat{\theta}$ is about 61% less than that of $\bar{X}$ and about 71% less than that of the MLE based on samples of size $n = 10$. When using the prior $N(\exp(1/2), 0.5^2)$, the proposed estimator $\hat{\theta}$ performs much better than $\bar{X}$ and the MLE, e.g., when $n = 10$, the variance of $\hat{\theta}$ is about 76% smaller than that of $\bar{X}$ and about 82% less than that of the MLE. As n increases, the variances of $\hat{\theta}$ become close to those of $\bar{X}$ and the MLE. The asymptotic forms $\hat{\theta}\_P1$ and $\hat{\theta}\_C1$, perform similarly to $\hat{\theta}$ even when $n = 20$. These results are significantly shown when the prior distribution $N(\mu, 0.1^2)$ is used. When using an improper prior distribution, e.g., $N(\mu - 1, 0.5^2)$, the performance of the proposed estimator is still better than that of $\bar{X}$ and the MLE, e.g., when $n = 10$, the variance of $\hat{\theta}$ is about 53% smaller than that of $\bar{X}$ and about 65% smaller than that of the MLE. When n is large, the variances of $\hat{\theta}$ are comparable to those of $\bar{X}$ and the MLE. In addition, the estimators $\hat{\theta}\_P1$ and $\hat{\theta}\_C1$ are still very close to the proposed estimator $\hat{\theta}$ from (2). When we have information that the target parameter can be μ or $-\mu$, e.g., $\{N(-\mu, 0.5^2) + N(\mu, 0.5^2)\}/2$, the variance of $\hat{\theta}$ is about 76% smaller than that of $\bar{X}$ and about 82% less than that of the MLE, for samples of size $n = 10$. As n increases, the variances of $\hat{\theta}$ become comparable to those of $\bar{X}$ and the MLE. These results are highlighted when $\{N(-\mu, 0.1^2) + N(\mu, 0.5^2)\}/2$ is utilized as the prior distribution. When a non-informative uniform prior is used, e.g., $U(0, \mu + 0.5)$, the variance of $\hat{\theta}$ is about 74% smaller than that of $\bar{X}$ and about 80% less than that of the MLE, for samples of size $n = 10$. When n is large, the variances of $\hat{\theta}$ are close to those of $\bar{X}$ and the MLE. When an uniform prior centered near the true parameter value is used, e.g., $U(\mu - 0.25, \mu + 0.25)$, the variance of $\hat{\theta}$ is about 99% smaller than that of $\bar{X}$ and the MLE for samples of size $n = 10$. One can also note that the use of the normal prior distributions with mean 0 to estimate the parameter $\theta = 1$ (or $\theta = \exp(0.5)$) leads to negative biases of the estimations. These biases are relatively small and vanished when the sample size increases.

-TABLE 1, 2-

From the MC study based on data sampled from a lognormal distribution we observe that the proposed estimator outperforms $\bar{X}$ and the MLE even when using priors that are supposed to contain "no information" about the true values of θ. The efficiency of the proposed estimator is clearly demonstrated in the case of skewed data. It has been discussed in the literature that the traditional estimation of the mean of a lognormal distribution is inaccurate due to the non-quadratic and asymmetric shape of the likelihood profile (e.g., Wu et al. 2003). In this case the proposed approach can serve as valid alternatives to the traditional techniques.

The performance of the asymptotic forms $\hat{\theta}\_P1$ and $\hat{\theta}\_C1$ are observed to be similar to that of $\hat{\theta}$ from (2) across a wide range of scenarios. Note that in additional MC evaluations, which were omitted from this paper, we consistently observed that the estimators $\hat{\theta}\_P1$ and $\hat{\theta}\_C1$ provided accurate and efficient approximations to $\hat{\theta}$. We also numerically evaluated the double empirical Bayesian estimator $\hat{\theta}_E$ given at equation (3) and the corresponding asymptotic form from Corollary 3. We concluded that these proposed estimators are comparable to $\bar{X}$, the MLE, when data are normally distributed, e.g., when $n = 20$, the variances of $\bar{X}$ and $\hat{\theta}_E$ were 0.0499 and 0.0505, respectively. However, when data were generated from a lognormal distribution the proposed estimator demonstrated an improvement efficiency as compared with the classical nonparametric estimator $\bar{X}$. For example, when $n = 75$, the variance of $\hat{\theta}_E$ was 7% smaller than that of $\bar{X}$.

Monte Carlo evaluations of the nonparametric James-Stein estimator. In this part of the experimental study, we carried out MC evaluations of the nonparametric James-Stein estimator $\hat{\theta}_{Ej}$ and compared it to the classical nonparametric estimator $\bar{X}_{.j}$ in terms of relative bias and efficiency. For simplicity and without loss of generality we assumed the dimension $K = 3$ for the underlying multivariate distributions used within this MC study. The independent samples were generated from either a $MVN\{(1,1,1)^T, I\}$ or a $MVLogNorm\{(0,0,0)^T, I\}$, where we used the covariance structure

$$I = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

In addition, correlated samples were generated from either $MVN\{(1,1,1)^T, \Sigma\}$ or a $MVLogNorm\{(0,0,0)^T, \Sigma\}$, with covariance structure given as

$$\Sigma = \begin{pmatrix} 1 & 0.5 & 0.5 \\ 0.5 & 1 & 0.5 \\ 0.5 & 0.5 & 1 \end{pmatrix}.$$

The sample sizes n were $10, 20, 30, 50, 75$, respectively. We compared our estimator $(\hat{\theta}_{E1}, \hat{\theta}_{E2}, \hat{\theta}_{E3})$ given at (6) with the classical nonparametric estimator $\bar{X} = (\bar{X}_{.1}, \bar{X}_{.2}, \bar{X}_{.3})$. The MC variance estimates for the respective estimators are defined as $V(\hat{\theta}_{Ej}) = \sum_{i=1}^{T}(\hat{\theta}_{Eij} - \theta_{ij})^2/T$ and $V(\bar{X}_j) = \sum_{i=1}^{T}(\bar{X}_{ij} - \theta_{ij})^2/T$, respectively, where $j = 1, 2, 3$ and $T = 15000$ is the number of the MC replications. The MC means and variances of the estimators are presented in Tables 3-6. In the cases where samples were generated from a $MVN\{(1,1,1)^T, I\}$ distribution the proposed James-Stein estimator was more efficient than $\bar{X}$ (the MLE in these cases) for small sample sizes, e.g. when $n = 10$ the variances of $(\hat{\theta}_{E1}, \hat{\theta}_{E2}, \hat{\theta}_{E3})$ were $(0.062, 0.065, 0.064)$, respectively, while the variances of $\bar{X}$ were $(0.096, 0.100, 0.099)$. As the sample size increased, the variance of the nonparametric James-Stein estimators were observed to be close to those of each component of $\bar{X}$. In the case where samples were generated from $MVLogNorm\{(0,0,0)^T, I\}$ the James-Stein estimator had a smaller component-wise variance as compared to the corresponding estimators for each element of $\bar{X}$. For example, when the sample size was set to $n = 10$, the variance of each element of $(\hat{\theta}_{E1}, \hat{\theta}_{E2}, \hat{\theta}_{E3})$ was

$(0.398, 0.531, 0.395)$, while the variance of each element of $\bar{X}$ was $(0.476, 0.591, 0.485)$. In the case where correlated data was generated, we observed similar results in that the performance of $(\hat{\theta}_{E1}, \hat{\theta}_{E2}, \hat{\theta}_{E3})$ was better than that of $\bar{X}$ in the sense of the relative efficiency. We conclude that for fixed sample sizes, ranging from small to large, the nonparametric James-Stein estimator consistently outperforms the classical nonparametric estimator, $\bar{X}$, in the multivariate setting.

6. Application

Thiobarbituric acid reaction substances (TBARS) is a common biomarker used in the study of oxidative stress (e.g., Schisterman et al. 2001). Many epidemiological studies have been carried forth for the purpose of examining the association between TBARS and myocardial infarction (MI) disease. The difficulty in analyzing TBARS is that values of this biomarker have been illustrated to have a non-normal distribution. In this section, we demonstrate the utility of the proposed nonparametric Bayesian approach by applying it to data from a population-based case-control study. The sample of 300 MI cases and 300 healthy controls was derived from randomly selected set of residents of Erie and Niagara counties, 35 to 79 years of age. The data was collected from two sources. First, a sample of residents between the ages of 35 and 65 years was randomly selected using the New York State Department of Motor Vehicles drivers' license rolls. A second sample of elderly residents between the ages of 65 to 79 years old was randomly selected from the Health Care Financing Administration database. In terms of constricting the proposed posterior expectations we utilized information from an earlier study by Schisterman et al. (2001), which had a similar design. They reported that the mean and standard deviation of TBARS was 1.84 and 0.80 for the case group, 1.44 and 0.48 for the control group, respectively. Therefore, we can reasonably consider $N(1.84, 0.80^2)$ and $N(1.44, 0.48^2)$ as possible prior distributions for TBARS for the case and control groups, respectively. For the purpose of illustration we also considered prior distribution functions such as $N(1.84, 0.1^2)$ and $N(1.44, 0.1^2)$, $N(1.84, (1/300)^2)$ and $N(1.44, (1/300)^2)$, $N(1, 1)$ and $N(1, 1)$ for the case and the control groups, respectively. We evaluated the mean and confidence intervals (CIs) for each group using the nonparametric Bayesian estimator $\hat{\theta}$ defined at (2). We also examine the double empirical Bayesian estimator $\hat{\theta}_E$ defined at (3), as well as the classical nonparametric moment estimator $\bar{X}$ based on the data. The results are shown in Table 7.

Note that across a variety of prior distributions, ranging from less informative to more informative, the proposed 95% CI's for TBARS, based on $\hat{\theta}$ and $\hat{\theta}_E$, respectively, do not overlap between the cases and controls. By contrast, the 95% CI's corresponding to the classical nonparametric moment estimator, $\bar{X}$, do not provide this conclusion. (These CI's based on $\bar{X}$ were calculated using the Central Limit Theorem approximation.)

-Table 7-

Bootstrap-Type Simulation Study. In addition to the above data example we conducted a bootstrap-type study for the purpose of examining the behavior of the estimators $\hat{\theta}$, $\hat{\theta}_E$ and $\bar{X}$ in terms of relative efficiency using TBARS MI case data as the underlying theoretical pseudo-population. For this study we set the number of bootstrap resamples to be $B = 5000$. For each bootstrap resample the dataset was divided into a sample set, of size $n = 50$ and $n = 70$, respectively, and a pseudo-population set at size $n = 300 - 50 = 250$ and $n = 300 - 70 = 230$, respectively. For each resample we calculated $\hat{\theta}_i$, $\hat{\theta}_{Ei}$, $\bar{X}_i$ and the pseudo-population mean using the relatively large samples. Define the obtained estimator based on the large samples as $\tilde{\mu}_i$, $i = 1, \ldots, B$. (The subscript i of $\hat{\theta}_i$, $\hat{\theta}_{Ei}$, $\bar{X}_i$ and $\tilde{\mu}_i$ indicates corresponding estimator's values obtained at ith bootstrap repetition.) Our measures of variance used to examine the relative efficiency between $\hat{\theta}$, $\hat{\theta}_E$ and $\bar{X}$ take the forms $\sum_{i=1}^{B}(\hat{\theta}_i - \tilde{\mu}_i)^2/B$, $\sum_{i=1}^{B}(\hat{\theta}_{Ei} - \tilde{\mu}_i)^2/B$,

$\sum_{i=1}^{B}(\bar{X}_i - \tilde{\mu}_i)^2/B$, respectively. In this simulation study, we used the prior distribution functions $N(1,1)$, $N(1.1, 0.1^2)$, $N(1.3, 0.1^2)$, $N(1.4, 0.1^2)$, and $N(1.84, 0.1^2)$ for construction of the estimator $\hat{\theta}$ defined at (2). The results of this study are show in Tables 8 and 9. When we use the $N(1,1)$ prior supposed to contain "no correct information" about the true values of θ, the proposed estimator $\hat{\theta}$ provides the bootstrap type variances comparable to those of $\bar{X}$, e.g., when $n = 50$, the variance of $\bar{X}$ is 0.00359, while the variance of $\hat{\theta}$ is 0.00351. When the priors were chosen to center around the bootstrap mean given as $\mu = \sum_{i=1}^{B} \tilde{\mu}_i/B$ then the bootstrap-type variances of $\hat{\theta}$ are smaller than those of $\bar{X}$, e.g., in the case with the prior $N(1.4, 0.1^2)$, the bootstrap-type variance of $\hat{\theta}$ is about 52% smaller than that of $\bar{X}$ when $n = 70$. The double empirical Bayesian estimator $\hat{\theta}_E$ provides similar bootstrap-type variances to those of $\bar{X}$.

-Table 8,9-

7. Discussion

In this paper, we proposed and examined a novel approach for developing the nonparametric Bayesian posterior expectation fashion by incorporating the EL methodology into the posterior likelihood construction. The asymptotic approximations to the new distribution-free posterior expectations were developed and shown to be quite accurate, even in the finite sample setting at $n = 20$. The asymptotic forms are similar to those derived in the well-known parametric Bayesian and Frequentist statistical literature. In the case when the prior distribution function depends on unknown hyper-parameters we proposed a nonparametric version of the empirical Bayesian method. This yielded double empirical Bayesian estimators. The extensive MC study showed that when prior distributions contained "no information" about the true values of the estimated parameter are assumed and the underlying distribution is skewed the proposed distribution-free posterior mean estimator outperforms the classical nonparametric estimator $\bar{X}$ in terms of relative efficiency. When the data are normally distributed the proposed estimator is comparable to the MLE. When proper priors are used and the data is generated from either a normal or lognormal distribution the proposed estimator provides significantly smaller variances as compared to the classical estimator $\bar{X}$ and the MLE. This in turn yields much narrower confidence intervals.

In the multivariate setting, with prior functions defined to be normal distributions with unknown hyper-parameters, the double empirical Bayesian estimation yields a nonparametric version of the well-known James-Stein estimator. The MC study in the multivariate setting confirmed that the proposed nonparametric James-Stein estimator has smaller variances than the classical nonparametric estimator $\bar{X}$ for data generated from MVN and $MVLogNorm$ distributions. The data example demonstrated the applicability of the proposed methodology in a real-world setting.

Acknowledgment

This research is supported by NIH grant 1R03DE020851-01A1 (the National Institute of Dental and Craniofacial Research).

Appendix

Proof of Lemma 1

It is clear that the argument θ_M, a root of $n^{-1}\sum_{i=1}^n G(X_i,\theta_M)=0$, maximizes the function $W(\theta)$, since in this case $W(\theta_M)=n^{-n}$ with $p_i=n^{-1}$, $i=1,\dots,n$, that maximize $\prod_{i=1}^n p_i$ given the sole constraint $\sum_{i=1}^n p_i=1$, $0\le p_i\le 1$, $i=1,\dots,n$.

Using the Lagrange method, one can represent $W(\theta)$ as

$$W(\theta)=\prod_{i=1}^n p_i, 0<p_i=\frac{1}{n+\lambda G(X_i,\theta)}<1,\ i=1,\dots,n,$$

where the Lagrange multiplier λ is a root of the equation $\sum G(X_i,\theta)\{n+\lambda G(X_i,\theta)\}^{-1}=0$ (e.g., Owen 2001). This then yields the following expression

$$\frac{d\log\{W(\theta)\}}{d\theta}=-\lambda\sum_{i=1}^n\frac{\partial G(X_i,\theta)/\partial\theta}{n+\lambda G(X_i,\theta)}-\sum_{i=1}^n\frac{G(X_i,\theta)}{n+\lambda G(X_i,\theta)}\frac{\partial\lambda}{\partial\theta} \quad (A1)$$
$$=-\lambda\sum_{i=1}^n\frac{\partial G(X_i,\theta)/\partial\theta}{n+\lambda G(X_i,\theta)},$$

where without loss of generality we assume $\partial G(X_i,\theta)/\partial\theta>0$, $i=1,\dots,n$.

Now define the function $L(\lambda)=\sum_{i=1}^n G(X_i,\theta)\{n+\lambda G(X_i,\theta)\}^{-1}$. Since $dL(\lambda)/d\lambda<0$ the function $L(\lambda)$ decreases with respect to λ and has just one root relative to solving $L(\lambda)=0$. Consider the scenario with $\theta>\theta_M$. In this case when $\lambda_0=0$ we can conclude that

$$L(\lambda_0)=\sum_{i=1}^n G(X_i,\theta)(n)^{-1}\ge\sum_{i=1}^n G(X_i,\theta_M)(n)^{-1}=0,$$

since $G(X_i,\theta)$ increases with respect to θ ($\partial G(X_i,\theta)/\partial\theta>0$).

The function $L(\lambda)$ decreases. This implies that the root of $L(\lambda)=0$ should be located on the right side from $\lambda_0=0$ and then this root is positive. For a graphical representation of this case see Figure 1(a) below. Thus, by virtue of (A1), we prove that the function $W(\theta)$ decreases, when $\theta>\theta_M$.

Taking the same approach, one can show that the root of $L(\lambda)=0$ should be to the left of $\lambda_0=0$, when $\theta<\theta_M$. For a graphical representation of this case see Figure 1(b). This result combined with (A1) completes the proof of Lemma 1.

-Figure 1-

Proof of Proposition 1

To prove the proposition, we first show that

$$\int_{X_{(1)}}^{X_{(n)}}\theta^{\upsilon}e^{\log ELR_1(\theta)}\pi(\theta)d\theta\cong\int_{\bar{X}-\varphi_n n^{-1/2}}^{\bar{X}+\varphi_n n^{-1/2}}\theta^{\upsilon}e^{\log ELR_1(\theta)}\pi(\theta)d\theta,\ \upsilon=0,1,$$

where a positive sequence $\varphi_n n^{-1/2}\to\infty$, $\varphi_n\to\infty$, as $n\to\infty$. This approximation allows us to analyze the numerator ($\upsilon=1$) and the denominator ($\upsilon=0$) defined at (2). Let us rewrite the function $EL_1(\theta)$ in the form of $\log EL_1(\theta)=\sum_{i=1}^n\log p_i$, where p_i can be defined by maximizing the Lagrangian

$$\Lambda=\sum_{i=1}^n\log p_i+\lambda_1(1-\sum_{i=1}^n p_i)+\lambda_2(\theta-\sum_{i=1}^n p_iX_i),$$

where λ_1 and λ_2 are Lagrange multipliers. Thus one can show that $p_i=\{n+\lambda(X_i-\theta)\}^{-1}$, where λ is a root of the equation $\sum_{i=1}^n(X_i-\theta)/\{n+\lambda(X_i-\theta)\}=0$.

Now define the function

$$L(\lambda) = \sum_{i=1}^{n} (X_i - \theta)/\{n + \lambda(X_i - \theta)\}. \tag{A2}$$

According to Lemma 1, when $\theta < \bar{X}$ then the function $\log ELR_1(\theta)$ is strictly increasing and when $\theta > \bar{X}$ then the function $\log ELR_1(\theta)$ is strictly decreasing. This implies that the function $\log ELR_1(\theta)$ is maximized at the point $\theta = \bar{X}$. Now, denote $a = \bar{X} - \varphi_n n^{-1/2}$ and $b = \bar{X} + \varphi_n n^{-1/2}$, where $\varphi_n = n^{1/6-\beta}$ and $\beta \in (0, 1/6)$. Then it follows that

$$\int_{X_{(1)}}^{X_{(n)}} e^{\log ELR_1(\theta)} \pi(\theta) d\theta = \int_{X_{(1)}}^{a} e^{\log ELR_1(\theta)} \pi(\theta) d\theta + \int_{a}^{b} e^{\log ELR_1(\theta)} \pi(\theta) d\theta + \int_{b}^{X_{(n)}} e^{\log ELR_1(\theta)} \pi(\theta) d\theta.$$

By virtue of the above considerations we can bound the remainder term

$$\int_{X_{(1)}}^{a} e^{\log ELR_1(\theta)} \pi(\theta) d\theta \le e^{\log ELR_1(a)} \int_{X_{(1)}}^{X_{(n)}} \pi(\theta) d\theta \le e^{\log ELR_1(a)}.$$

In order to arrive at an expression for the value of $\log ELR_1(a)$, taking into account the definition of ELR_1 in (2), we evaluate (A2) at $\theta = a$ such that

$$\begin{aligned} L(\lambda) &= \sum_{i=1}^{n} \frac{(X_i - \bar{X} + \varphi_n n^{-1/2})}{n + \lambda(X_i - \bar{X} + \varphi_n n^{-1/2})} \\ &= \frac{1}{n} \sum_{i=1}^{n} \frac{(X_i - \bar{X} + \varphi_n n^{-1/2})[\{1 + \lambda n^{-1}(X_i - \bar{X} + \varphi_n n^{-1/2})\} - \lambda n^{-1}(X_i - \bar{X} + \varphi_n n^{-1/2})]}{1 + \lambda n^{-1}(X_i - \bar{X} + \varphi_n n^{-1/2})} \\ &= \frac{1}{n} \{ \sum_{i=1}^{n} (X_i - \bar{X} + \varphi_n n^{-1/2}) - \lambda n^{-1} \sum_{i=1}^{n} \frac{(X_i - \bar{X} + \varphi_n n^{-1/2})^2}{1 + \lambda n^{-1}(X_i - \bar{X} + \varphi_n n^{-1/2})} \}. \end{aligned} \tag{A3}$$

Defining $\lambda_c = n^{2/3}\tau_n^{-1}$, where $\tau_n = n^{\gamma}$, $0 < \gamma < \beta < 1/6$, and substituting it into (A3) yields

$$\sqrt{n} L(\lambda_c) = \varphi_n - \sqrt{n} \frac{n^{2/3-1}}{\tau_n} \frac{1}{n} \sum_{i=1}^{n} \frac{(X_i - \bar{X} + \varphi_n n^{-1/2})^2}{1 + n^{-1/3}\tau_n^{-1}(X_i - \bar{X} + \varphi_n n^{-1/2})},$$

Since $(X_i - \bar{X})/(n^{1/3}\tau_n) = O_p(1)$ (e.g., Owen 1988), we have

$$\sqrt{n} L(\lambda_c) = \varphi_n - \frac{n^{1/6}}{\tau_n} \frac{1}{n} \sum_{i=1}^{n} \frac{(X_i - \bar{X} + \varphi_n n^{-1/2})^2}{1 + O_p(1)}.$$

Now, it follows that $\sqrt{n} L(\lambda_c) \to -\infty$, as $n \to \infty$. In a similar manner, $\sqrt{n} L(-\lambda_c) \to \infty$, as $n \to \infty$. Thus, the solution, λ_0, of equation $\sqrt{n} L(\lambda_0) = 0$ belongs to the interval $(-\lambda_c, \lambda_c)$, i.e. $\lambda_0 = O_p(n^{2/3}\tau_n^{-1})$.

Let us now derive the approximate value corresponding to λ_0 as $n \to \infty$. Since $L(\lambda_0) = 0$,

$$\sum_{i=1}^{n} (X_i - \bar{X} + \varphi_n n^{-1/2}) \frac{1}{1 + \lambda_0 n^{-1}(X_i - \bar{X} + \varphi_n n^{-1/2})} = 0. \tag{A4}$$

Applying a Taylor series expansion to (A4) we then obtain

$$\sum_{i=1}^{n} (X_i - \bar{X} + \varphi_n n^{-1/2}) \{1 - \lambda_0 n^{-1}(X_i - \bar{X} + \varphi_n n^{-1/2}) + \frac{\lambda_0^2 n^{-2}(X_i - \bar{X} + \varphi_n n^{-1/2})^2}{(1 + \omega_i)^2}\} = 0, \tag{A5}$$

where $0 < \omega_i < \lambda_0 n^{-1}(X_i - \bar{X} + \varphi_n n^{-1/2})$. Since $\lambda_0 = O_p(n^{2/3}\tau_n^{-1})$, we can re-express (A5) as

$$\sum_{i=1}^{n}(X_i - \bar{X} + \frac{\varphi_n}{n^{1/2}}) - \frac{\lambda}{n}\sum_{i=1}^{n}(X_i - \bar{X} + \frac{\varphi_n}{n^{1/2}})^2 + \frac{O(n^{1/3})}{\tau_n^2}\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X} + \frac{\varphi_n}{n^{1/2}})^3 = 0. \quad \text{(A6)}$$

Then it follows that the approximate solution based on solving (A6) is given by

$$\lambda_0 = \frac{\varphi_n n^{1/2}}{n^{-1}\sum_{i=1}^{n}(X_i - \bar{X} + \varphi_n n^{-1/2})^2} + \frac{O(n^{1/3})}{\tau_n^2}. \quad \text{(A7)}$$

Applying a Taylor series expansion to $\log ELR_1(\theta)$ by (2) with $\theta = a$ yields the following expression

$$\begin{aligned}\log ELR_1(a) &= -\sum_{i=1}^{n}\log\{1 + \frac{\lambda_0}{n}(X_i - \bar{X} + \varphi_n n^{-1/2})\}\\ &= -\sum_{i=1}^{n}\frac{\lambda_0}{n}(X_i - \bar{X} + \varphi_n n^{-1/2}) + \frac{1}{2}\sum_{i=1}^{n}\frac{\lambda_0^2}{n^2}(X_i - \bar{X} + \varphi_n n^{-1/2})^2\\ &\quad - \frac{1}{3}\sum_{i=1}^{n}\frac{\lambda_0^3}{n^3}\frac{(X_i - \bar{X} + \varphi_n n^{-1/2})^3}{(1+\omega_i^*)^3},\end{aligned}$$

where $0 < \omega_i^* < \lambda_0 n^{-1}(X_i - \bar{X} + \varphi_n n^{-1/2})$. By virtue of (A7) and the fact that $\lambda_0 = O(n^{2/3}/\tau_n)$ we then have

$$\begin{aligned}&\log ELR_1(a) = -\frac{\lambda}{n}\varphi_n n^{1/2} + \frac{1}{2}\sum_{i=1}^{n}\frac{\lambda^2}{n^2}(X_i - \bar{X} + \varphi_n n^{-1/2})^2 - O(n^{-3\gamma})\\ &= -\frac{\varphi_n^2 n}{n n^{-1}\sum_{i=1}^{n}(X_i - \bar{X} + \varphi_n n^{-1/2})^2} - \frac{O(n^{4/3})}{\tau_n^2 n^2}\varphi_n n^{1/2}\\ &+ \frac{1}{2}\Bigg[\frac{\varphi_n^2 n}{\{n^{-1}\sum_{i=1}^{n}(X_i - \bar{X} + \varphi_n n^{-1/2})^2\}^2} + 2\frac{O(n^{4/3})}{\tau_n^2 n}\frac{\varphi_n^2 n^{1/2}}{n^{-1}\sum_{i=1}^{n}(X_i - \bar{X} + \varphi_n n^{-1/2})^2}\\ &+ \frac{O(n^{8/3})}{\tau_n^4 n^2}\Bigg]\frac{1}{n^2}\sum_{i=1}^{n}(X_i - \bar{X} + \varphi_n n^{-1/2})^2 - O(n^{-3\gamma})\\ &= -\frac{1}{2}\frac{\varphi_n^2}{n^{-1}\sum_{i=1}^{n}(X_i - \bar{X} + \varphi_n n^{-1/2})^2} - O(n^{4/3-2-2\gamma+1/6-\beta+1/2})\\ &+ O(n^{4/3-1-2\gamma+1/6-\beta+1/2-1}) + O(n^{8/3-2-4\gamma-1}) - O(n^{-3\gamma})\\ &= -\frac{1}{2}\frac{\varphi_n^2}{n^{-1}\sum_{i=1}^{n}(X_i - \bar{X} + \varphi_n n^{-1/2})^2} - O(n^{-3\gamma}) \to \infty\end{aligned}$$

as $n \to \infty$, where $\varphi_n^2 = n^{1/3-2\beta} \to \infty$ and $0 < \gamma < \beta < 1/6$. Thus, we arrive at the result that $\int_{X_{(1)}}^{a}\exp\{\log ELR_1(\theta)\}\pi(\theta)d\theta \le \exp\{\log ELR_1(a)\} = O\{\exp(-wn^{1/3-2\beta})\} \to 0$ as $n \to \infty$, where w is a positive constant. It follows similarly that $\int_{b}^{X_{(n)}}\exp\{\log ELR_1(\theta)\pi(\theta)d\theta\} \le \exp\{\log ELR_1(b)\} = O\{\exp(-w_1 n^{1/3-2\beta})\} \to 0$ as well as $\int_{X_{(1)}}^{a}\theta e^{\log ELR_1(\theta)}\pi(\theta)d\theta \le O(e^{-w_2 n^{1/3-2\beta}}) \to 0$, $\int_{b}^{X_{(n)}}\theta e^{\log ELR_1(\theta)}\pi(\theta)d\theta \le O(e^{-w_3 n^{1/3-2\beta}}) \to 0$, where w_1, w_2, w_3 are positive constants and $n \to \infty$.

Now we consider the main term $\int_a^b \theta e^{\log ELR_1(\theta)}\pi(\theta)d\theta$ of the marginal distribution defined at (2). This integral consists of $\log ELR_1(\theta)$ that, by virtue of the Taylor theorem and (A1), is

$$\log ELR_1(\theta) = \log ELR_1(\bar{X}) + (\theta - \bar{X})\lambda(\bar{X}) + \frac{1}{2}(\theta - \bar{X})^2\left(\frac{d\lambda(u)}{du}\Big|_{u=\bar{X}}\right) \quad \text{(A8)}$$

$$+\frac{1}{6}(\theta-\bar{X})^3\left(\frac{d^2\lambda(u)}{du^2}\Big|_{u=\bar{X}}\right)+\frac{1}{24}(\theta-\bar{X})^4\left(\frac{d^3\lambda(u)}{du^3}\Big|_{u=\theta+\varpi(\bar{X}-\theta)}\right), \varpi\in(0,1).$$

Since the function $\lambda(u)$ is defined by $\sum(X_i-u)/\{n+\lambda(u)(X_i-u)\}$, one can show that

$$\frac{d\lambda(\theta)}{d\theta}=-\frac{n\sum_{i=1}^n p_i^2}{\sum_{i=1}^n(X_i-\theta)^2p_i^2},$$

$$\frac{d^2\lambda(\theta)}{d\theta^2}=\frac{2(d\lambda(\theta)/d\theta)^2\sum_{i=1}^n(X_i-\theta)^3p_i^3+4n(d\lambda(\theta)/d\theta)\sum_{i=1}^n(X_i-\theta)p_i^3-2n\lambda(\theta)\sum_{i=1}^n p_i^3}{\sum_{i=1}^n(X_i-\theta)^2p_i^2},$$

$$\frac{d^3\lambda(\theta)}{d\theta^3}=\left\{\sum_{i=1}^n(X_i-\theta)^2p_i^2\right\}^{-1}\left[6\frac{d\lambda(\theta)}{d\theta}\frac{d^2\lambda(\theta)}{d\theta^2}\sum_{i=1}^n(X_i-\theta)^3p_i^3+6n\frac{d^2\lambda(\theta)}{d\theta^2}\sum_{i=1}^n(X_i-\theta)p_i^3\right.$$
$$-6\left(\frac{d\lambda(\theta)}{d\theta}\right)^3\sum_{i=1}^n(X_i-\theta)^4p_i^4-18n\left(\frac{d\lambda(\theta)}{d\theta}\right)^2\sum_{i=1}^n(X_i-\theta)^2p_i^4+12n\left(\frac{d\lambda(\theta)}{d\theta}\right)\sum_{i=1}^n p_i^3$$
$$\left.-\left\{18n^2\left(\frac{d\lambda(\theta)}{d\theta}\right)+6n(\lambda(\theta))^2\right\}\sum_{i=1}^n p_i^4\right],$$

where $p_i=\{n+\lambda(\theta)(X_i-\theta)\}^{-1}$. Noting that, the argument $\bar{X}$ maximizes the function $\log ELR_1(\theta)$, $\log ELR_1(\bar{X})=0$ and $\lambda(\bar{X})=0$, we have

$$\frac{d\lambda(\theta)}{d\theta}\Big|_{\theta=\bar{X}}=-\frac{n}{n^{-1}\sum_{i=1}^n(X_i-\bar{X})^2}=-\frac{n}{\sigma_n^2},\ \frac{d^2\lambda(\theta)}{d\theta^2}\Big|_{\theta=\bar{X}}=\frac{2n\sum_{i=1}^n(X_i-\bar{X})^3/n}{\left\{n^{-1}\sum_{i=1}^n(X_i-\bar{X})^2\right\}^3}=\frac{2nM_n^3}{(\sigma_n^2)^3}$$

as well as $d^3\lambda(\theta)/d\theta^3=O(n)$, for $\theta\in(a,b)$, since, in this case, using the same techniques applied to the previous proofs and utilizing results found in Owen (1988) and Lazar & Mykland (1998) one can derive the following expressions:

$$\lambda=\frac{\sum_{i=1}^n(X_i-\bar{X})}{n^{-1}\sum_{i=1}^n(X_i-\bar{X})^2}+\frac{O(n^{1/3})}{\tau_n^2}=O(n^{2/3-\beta}),\ \frac{\lambda(\theta)}{n}(X_i-\theta)=O(1),$$
$$p_i=\frac{1}{n}\left\{1+\frac{\lambda(\theta)}{n}(X_i-\theta)\right\}^{-1}=O(n^{-1}),$$

when $\mid\bar{X}-\theta\mid\le\varphi_n n^{-1/2}=n^{-1/3-\beta}$, $0<\beta<1/6$. The above asymptotic results, (A8) and a Taylor expansion imply

$$\int_a^b e^{\log ELR_1(\theta)}\pi(\theta)d\theta=\int_a^b\exp\left\{-\frac{n}{2\sigma_n^2}(\theta-\bar{X})^2+\frac{nM_n^3}{3(\sigma_n^2)^3}(\theta-\bar{X})^3+O(n)(\theta-\bar{X})^4\right\}\pi(\theta)d\theta$$
$$=\int\exp\left\{-\frac{n}{2\sigma_n^2}(\theta-\bar{X})^2\right\}\pi(\theta)d\theta+\frac{nM_n^3}{3(\sigma_n^2)^3}\int(\theta-\bar{X})^3\exp\left\{-\frac{n}{2\sigma_n^2}(\theta-\bar{X})^2\right\}\pi(\theta)d\theta$$
$$+O(n)\int_a^b(\theta-\bar{X})^4\exp\left\{-\frac{n}{2\sigma_n^2}(\theta-\bar{X})^2\right\}\pi(\theta)d\theta. \quad \text{(A9)}$$

It follows similarly that

$$\int_a^b(\theta-\bar{X})e^{\log ELR_1(\theta)}\pi(\theta)d\theta=\int(\theta-\bar{X})\exp\left\{-\frac{n}{2\sigma_n^2}(\theta-\bar{X})^2\right\}\pi(\theta)d\theta$$
$$+\frac{nM_n^3}{3(\sigma_n^2)^3}\int(\theta-\bar{X})^4\exp\left\{-\frac{n}{2\sigma_n^2}(\theta-\bar{X})^2\right\}\pi(\theta)d\theta$$
$$+O(n)\int_a^b(\theta-\bar{X})^5\exp\left\{-\frac{n}{2\sigma_n^2}(\theta-\bar{X})^2\right\}\pi(\theta)d\theta. \quad \text{(A10)}$$

By virtue of the definition (2), the nonparametric posterior expectation $\hat{\theta}$ can be represented in the form of

$$\hat{\theta} = \frac{\int \theta \exp\left\{-\frac{n}{2\sigma_n^2}(\theta - \bar{X})^2\right\}\pi(\theta)d\theta}{\int \exp\left\{-\frac{n}{2\sigma_n^2}(\theta - \bar{X})^2\right\}\pi(\theta)d\theta} + Q_n,$$

where

$$\begin{aligned}
Q_n &\equiv \frac{\int_a^b \theta e^{\log ELR_1(\theta)}\pi(\theta)d\theta \int e^{-\frac{n}{2\sigma_n^2}(\theta-\bar{X})^2}\pi(\theta)d\theta - \int \theta e^{-\frac{n}{2\sigma_n^2}(\theta-\bar{X})^2}\pi(\theta)d\theta \int_a^b e^{\log ELR_1(\theta)}\pi(\theta)d\theta}{\int e^{-\frac{n}{2\sigma_n^2}(\theta-\bar{X})^2}\pi(\theta)d\theta \int_a^b e^{\log ELR_1(\theta)}\pi(\theta)d\theta} \\
&= \left\{\int e^{-\frac{n}{2\sigma_n^2}(\theta-\bar{X})^2}\pi(\theta)d\theta \int_a^b e^{\log ELR_1(\theta)}\pi(\theta)d\theta\right\}^{-1} \\
&\times \left\{\int_a^b (\theta - \bar{X})e^{\log ELR_1(\theta)}\pi(\theta)d\theta \int e^{-\frac{n}{2\sigma_n^2}(\theta-\bar{X})^2}\pi(\theta)d\theta \right. \\
&\left. - \int (\theta - \bar{X})e^{-\frac{n}{2\sigma_n^2}(\theta-\bar{X})^2}\pi(\theta)d\theta \int_a^b e^{\log ELR_1(\theta)}\pi(\theta)d\theta\right\}.
\end{aligned}$$

It is clear that, taking into account the results (A9), (A10), the facts $\pi(\theta) = \pi(\bar{X}) + (\theta - \bar{X})\pi'(\bar{X}) + 0.5(\theta - \bar{X})^2\pi''(\bar{X} + q(\theta - \bar{X}))$, $q \in (0, 1)$, $\int(\theta - \bar{X})e^{-\frac{n}{2\sigma_n^2}(\theta-\bar{X})^2}d\theta = 0$ and $b - a = n^{1/6-\beta}/\sqrt{n}$, we obtain

$$\begin{aligned}
Q_n &= \left\{\frac{nM_n^3}{3(\sigma_n^2)^3}\int(\theta - \bar{X})^4 e^{-\frac{n}{2\sigma_n^2}(\theta-\bar{X})^2}\pi(\theta)d\theta \int e^{-\frac{n}{2\sigma_n^2}(\theta-\bar{X})^2}\pi(\theta)d\theta \right. \\
&+ O(n)\int_a^b (\theta - \bar{X})^5 e^{-\frac{n}{2\sigma_n^2}(\theta-\bar{X})^2}\pi(\theta)d\theta \int e^{-\frac{n}{2\sigma_n^2}(\theta-\bar{X})^2}\pi(\theta)d\theta \\
&- \frac{nM_n^3}{3(\sigma_n^2)^3}\int(\theta - \bar{X})^3 e^{-\frac{n}{2\sigma_n^2}(\theta-\bar{X})^2}\pi(\theta)d\theta \int(\theta - \bar{X})e^{-\frac{n}{2\sigma_n^2}(\theta-\bar{X})^2}\pi(\theta)d\theta \\
&\left. - O(n)\int_a^b (\theta - \bar{X})^4 e^{-\frac{n}{2\sigma_n^2}(\theta-\bar{X})^2}\pi(\theta)d\theta \int(\theta - \bar{X})e^{-\frac{n}{2\sigma_n^2}(\theta-\bar{X})^2}\pi(\theta)d\theta\right\} \\
&\times \left[\left\{\int e^{-\frac{n}{2\sigma_n^2}(\theta-\bar{X})^2}\pi(\theta)d\theta\right\}^2 + \frac{nM_n^3}{3(\sigma_n^2)^3}\int(\theta - \bar{X})^3 e^{-\frac{n}{2\sigma_n^2}(\theta-\bar{X})^2}\pi(\theta)d\theta \int e^{-\frac{n}{2\sigma_n^2}(\theta-\bar{X})^2}\pi(\theta)d\theta \right. \\
&\left. + O(n)\int_a^b (\theta - \bar{X})^4 e^{-\frac{n}{2\sigma_n^2}(\theta-\bar{X})^2}\pi(\theta)d\theta \int e^{-\frac{n}{2\sigma_n^2}(\theta-\bar{X})^2}\pi(\theta)d\theta\right]^{-1} \\
&= \left\{\frac{nM_n^3}{3(\sigma_n^2)^3}\int(\theta - \bar{X})^4 e^{-\frac{n}{2\sigma_n^2}(\theta-\bar{X})^2}d\theta \int e^{-\frac{n}{2\sigma_n^2}(\theta-\bar{X})^2}d\theta \right. \\
&\left. + O(n)\int_a^b (\theta - \bar{X})^5 e^{-\frac{n}{2\sigma_n^2}(\theta-\bar{X})^2}d\theta \int e^{-\frac{n}{2\sigma_n^2}(\theta-\bar{X})^2}d\theta\right\} \\
&\times \left\{\int e^{-\frac{n}{2\sigma_n^2}(\theta-\bar{X})^2}d\theta\right\}^{-2} + O(n^{-1/2-6\beta}).
\end{aligned}$$

Computing the definite integrals written above, we deduces that

$$\hat{\theta} = \frac{\int \theta \exp\left\{-\frac{n}{2\sigma_n^2}(\theta - \bar{X})^2\right\}\pi(\theta)d\theta}{\int \exp\left\{-\frac{n}{2\sigma_n^2}(\theta - \bar{X})^2\right\}\pi(\theta)d\theta}$$

$$+ \frac{\frac{2nM_n^3}{3(\sigma_n^2)^3}\frac{4!(\pi)^{1/2}(2\sigma_n^2)^{5/2}}{2!n^{5/2}2^5}\frac{(2\pi\sigma_n^2)^{1/2}}{n^{1/2}} + O\{n(n^{1/6-\beta-1/2})^6\}\frac{(2\pi\sigma_n^2)^{1/2}}{n^{1/2}}}{(2\pi\sigma_n^2)n^{-1}} + O(n^{-1/2-6\beta}).$$

Making use of $\beta = 1/6 - \varepsilon/6$, $\varepsilon > 0$ we complete the proof of Proposition 1.

Proof of Corollary 1

Corollary 1 can be proven by directly applying the result of Proposition 1.

Proof of Corollary 2

To prove this corollary, we can use the result

$$\hat{\theta} = \frac{\int_a^b \theta \exp\left\{-\frac{1}{2n}\frac{(\sum X_i - n\theta)^2}{\sigma_n^2}\right\}\pi(\theta)d\theta}{\int_a^b \exp\left\{-\frac{1}{2n}\frac{(\sum X_i - n\theta)^2}{\sigma_n^2}\right\}\pi(\theta)d\theta} + \frac{M_n^3}{\sigma_n^2 n} + O_p(n^{-3/2+\varepsilon}),$$

where $a = \bar{X} - \varphi_n n^{-1/2}$, $b = \bar{X} + \varphi_n n^{-1/2}$, $\varphi_n = n^{1/6-\beta}$, $0 < \gamma < \beta < 1/6$, $\beta = 1/6 - \varepsilon/6$, $\varepsilon > 0$.

This approximation was obtained above via the process related to the proof of Proposition 1. Applying the Taylor expansion $\pi(\theta) = \pi(\bar{X}) + (\theta - \bar{X})\pi'(\bar{X}) + (\theta - \bar{X})^2\pi''(\bar{X})/2 + (\theta - \bar{X})^3\pi'''(\tilde{X})/6$, $\tilde{X} \in (\theta, \bar{X})$ to the asymptotic form of $\hat{\theta}$, and in a similar manner of the Laplace method (e.g., Bleistein & Handelsman 2010, p.180), we compete the proof.

Proof of Proposition 4

We begin with the asymptotic analysis related to the numerator of the definition of the nonparametric posterior expectation $\hat{D}_G$. To approximate the double integral $\int\int D(\theta_1, \theta_2)e^{\log ELR_3(\theta_1,\theta_2)}\pi(\theta_1, \theta_2)d\theta_1 d\theta_2$, we first show that the main term of the integral is

$$\int_a^b \int_{a_1}^{b_1} D(\theta_1, \theta_2)e^{\log ELR_3(\theta_1,\theta_2)}\pi(\theta_1, \theta_2)d\theta_1 d\theta_2,$$

where $a = \bar{X} - \varphi_n n^{-1/2}$, $b = \bar{X} + \varphi_n n^{-1/2}$, $a_1 = \overline{X^2} - \varphi_n n^{-1/2}$, $b_1 = \overline{X^2} + \varphi_n n^{-1/2}$, $\varphi_n = n^{1/6-\beta}$, $0 < \beta < 1/6$ and $\overline{X^2} = \sum_{i=1}^n X_i^2/n$. Since

$$\int_{X_{(1)}}^a \int_{X_{(1)}^2}^{X_{(n)}^2} D(\theta_1, \theta_2)e^{\log ELR_3(\theta_1,\theta_2)}\pi(\theta_1, \theta_2)d\theta_1 d\theta_2$$

$$\leq \int_{X_{(1)}}^a \int_{X_{(1)}^2}^{X_{(n)}^2} D(\theta_1, \theta_2)\pi(\theta_1, \theta_2)d\theta_2 e^{\log ELR_1(\theta_1)}d\theta_1,$$

in a similar manner to the proof of Proposition 1, we conclude

$$\int_{X_{(1)}}^a \int_{X_{(1)}^2}^{X_{(n)}^2} D(\theta_1, \theta_2)\pi(\theta_1, \theta_2)d\theta_2 e^{\log ELR_1(\theta_1)}d\theta_1$$

$$\leq \int_{X_{(1)}}^a \int_{X_{(1)}^2}^{X_{(n)}^2} D(\theta_1, \theta_2)\pi(\theta_1, \theta_2)d\theta_2 d\theta_1 e^{\log ELR_1(a)} = O(e^{-wn^{1/3-2\beta}}) \to 0,$$

where w is a positive constant and $n \to \infty$.

Likewise, we have $\int_b^{X_{(n)}} \int_{X_{(1)}^2}^{X_{(n)}^2} D(\theta_1,\theta_2)e^{\log ELR_3(\theta_1,\theta_2)}\pi(\theta_1,\theta_2)d\theta_1 d\theta_2 = O(e^{-wn^{1/3-2\beta}}) \to 0$, as $n \to \infty$.

Now, we define $ELR_5(\theta) = n^n \max_{0<p_1,\dots,p_n<1} \{\prod_{i=1}^n p_i : \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i X_i^2 = \theta\}$ It is clear that $ELR_5(\theta) \geq ELR_3(\theta_1,\theta)$ for all (θ_1,θ) and hence

$$\int_a^b \int_{X_{(1)}^2}^{a_1} D(\theta_1,\theta_2)e^{\log ELR_3(\theta_1,\theta_2)}\pi(\theta_1,\theta_2)d\theta_1 d\theta_2 \leq \int_a^b \int_{X_{(1)}^2}^{a_1} D(\theta_1,\theta_2)\pi(\theta_1,\theta_2)d\theta_1 e^{\log ELR_5(\theta_2)} d\theta_2$$

$$\leq \int_a^b \int_{X_{(1)}^2}^{a_1} D(\theta_1,\theta_2)\pi(\theta_1,\theta_2)d\theta_1 d\theta_2 e^{\log ELR_5(a_1)} = O(e^{-wn^{1/3-2\beta}}) \to 0,$$

and $\int_a^b \int_{b_1}^{X_{(n)}^2} D(\theta_1,\theta_2)e^{\log ELR_3(\theta_1,\theta_2)}\pi(\theta_1,\theta_2)d\theta_1 d\theta_2 \to 0$, $n \to \infty$.

In order to apply almost directly the proof scheme of Proposition 1, we note that

$$\log ELR_3(\theta_1,\theta_2) = -\sum_{i=1}^n \log\{1 + \frac{\lambda_1}{n}(X_i - \theta_1) + \frac{\lambda_2}{n}(X_i^2 - \theta_2)\},$$

where the Lagrange multipliers λ_1 and λ_2 satisfy

$$L_1(\theta_1,\theta_2) \equiv \sum_{i=1}^n \frac{X_i - \theta_1}{n + \lambda_1(X_i - \theta_1) + \lambda_2(X_i^2 - \theta_2)} = 0 \text{ and} \quad \text{(A11)}$$
$$L_2(\theta_1,\theta_2) \equiv \sum_{i=1}^n \frac{X_i^2 - \theta_2}{n + \lambda_1(X_i - \theta_1) + \lambda_2(X_i^2 - \theta_2)} = 0$$

(e.g., Owen 2001). Since (A11), one can show that

$$\frac{\partial \log ELR_3(\theta_1,\theta_2)}{\partial \theta_1} = \lambda_1(\theta_1,\theta_2) \text{ and } \frac{\partial \log ELR_3(\theta_1,\theta_2)}{\partial \theta_2} = \lambda_2(\theta_1,\theta_2) \quad \text{(A12)}$$

Then the fact $\lambda_1(\bar{X}, \overline{X^2}) = 0$, $\lambda_2(\bar{X}, \overline{X^2}) = 0$ and a Taylor expansion argument yield

$$\log ELR_3(\theta_1,\theta_2) = \frac{1}{2}(\theta_1 - \bar{X})^2 \frac{\partial \lambda_1}{\partial \theta_1}\Big|_{\theta_1=\bar{X},\theta_2=\overline{X^2}} + (\theta_1 - \bar{X})(\theta_2 - \overline{X^2})\frac{\partial \lambda_1}{\partial \theta_2}\Big|_{\theta_1=\bar{X},\theta_2=\overline{X^2}} \quad \text{(A13)}$$
$$+ \frac{1}{2}(\theta_2 - \overline{X^2})^2 \frac{\partial \lambda_2}{\partial \theta_2}\Big|_{\theta_1=\bar{X},\theta_2=\overline{X^2}} + \frac{1}{3!}\Bigg\{(\theta_1 - \bar{X})^3 \frac{\partial^2 \lambda_1}{\partial \theta_1^2}\Big|_{\theta_1=\bar{X},\theta_2=\overline{X^2}}$$
$$+ 3(\theta_1 - \bar{X})^2(\theta_2 - \overline{X^2})\frac{\partial^2 \lambda_1}{\partial \theta_1 \partial \theta_2}\Big|_{\theta_1=\bar{X},\theta_2=\overline{X^2}} + 3(\theta_1 - \bar{X})(\theta_2 - \overline{X^2})^2 \frac{\partial^2 \lambda_2}{\partial \theta_1 \partial \theta_2}\Big|_{\theta_1=\bar{X},\theta_2=\overline{X^2}}$$
$$+ (\theta_2 - \overline{X^2})^3 \frac{\partial^2 \lambda_2}{\partial \theta_2^2}\Big|_{\theta_1=\bar{X},\theta_2=\overline{X^2}}\Bigg\} + O(n^{-1/3-4\beta}),$$

when $\theta_1 \in (a,b)$, $\theta_2 \in (a_1,b_1)$, $0 < \beta < 1/6$ and where

$$\frac{\partial \lambda_1}{\partial \theta_1}\Big|_{\theta_1=\bar{X},\theta_2=\overline{X^2}} = -\frac{n}{\sigma_n^2 - (\sigma_{XX^2n})^2/\sigma_{X^2n}^2}, \; \frac{\partial \lambda_2}{\partial \theta_2}\Big|_{\theta_1=\bar{X},\theta_2=\overline{X^2}} = -\frac{n}{\sigma_{X^2n}^2 - (\sigma_{XX^2n})^2/\sigma_n^2}, \quad \text{(A14)}$$
$$\frac{\partial \lambda_1}{\partial \theta_2}\Big|_{\theta_1=\bar{X},\theta_2=\overline{X^2}} = \frac{\partial \lambda_2}{\partial \theta_1}\Big|_{\theta_1=\bar{X},\theta_2=\overline{X^2}} = \frac{n}{\sigma_{X^2n}^2\sigma_n^2/\sigma_{XX^2n} - \sigma_{XX^2n}},$$
$$\frac{\partial^2 \lambda_1}{\partial \theta_k^2}\Big|_{\theta_1=\bar{X},\theta_2=\overline{X^2}} = \frac{2n^{-2}\sigma_{XX^2n}\sum_{i=1}^n (X_i^2 - \overline{X^2})\Psi_{ki} - \sigma_{X^2n}^2 \sum_{i=1}^n (X_i - \bar{X})\Psi_{ki}}{(\sigma_{XX^2n})^2 - \sigma_{X^2n}^2\sigma_n^2},$$
$$\frac{\partial^2 \lambda_2}{\partial \theta_k^2}\Big|_{\theta_1=\bar{X},\theta_2=\overline{X^2}} = \frac{2n^{-2}\sigma_n^2 \sum_{i=1}^n (X_i^2 - \overline{X^2})\Psi_{ki} - \sigma_{XX^2n}\sum_{i=1}^n (X_i - \bar{X})\Psi_{ki}}{\sigma_{X^2n}^2\sigma_n^2 - (\sigma_{XX^2n})^2},$$

$$\frac{\partial^2 \lambda_1}{\partial \theta_1 \partial \theta_2} = \frac{\partial^2 \lambda_2}{\partial \theta_1^2} \text{ and } \frac{\partial^2 \lambda_2}{\partial \theta_1 \partial \theta_2} = \frac{\partial^2 \lambda_1}{\partial \theta_2^2}$$

that can be obtained by utilizing (A11) and (A12) with the definitions

$$\Psi_{ki} = \left\{ \frac{\partial \lambda_1}{\partial \theta_k}\Big|_{\theta_1=\bar{X},\theta_2=\overline{X^2}} (X_i - \bar{X}) + \frac{\partial \lambda_2}{\partial \theta_k}\Big|_{\theta_1=\bar{X},\theta_2=\overline{X^2}} (X_i^2 - \overline{X^2}) \right\}^2, \ k = 1, 2,$$

$$\sigma^2_{X^2 n} = \frac{1}{n}\sum_{i=1}^{n} (X_i^2 - \overline{X^2})^2, \ \sigma_{XX^2 n} = \frac{1}{n}\sum_{i=1}^{n} (X_i - \bar{X})(X_i^2 - \overline{X^2}).$$

The validity of the Proposition 4 follows by arguments similar to those of the proof of Proposition 1 (see the proof scheme from (A8) to the end of the Proposition 1's proof) where the Taylor expansion for

$$D(\theta_1, \theta_2) = D(\bar{X}, \overline{X^2}) + (\theta_1 - \bar{X}) \frac{\partial D(\theta_1, \theta_2)}{\partial \theta_1}\Big|_{\theta_1=\bar{X},\theta_2=\overline{X^2}} + (\theta_2 - \overline{X^2}) \frac{\partial D(\theta_1, \theta_2)}{\partial \theta_2}\Big|_{\theta_1=\bar{X},\theta_2=\overline{X^2}} + ...$$

is applied evaluating a Q_n-type remainder term (see the remainder term Q_n and its analysis in the proof of Proposition 1). In this case, we present the remainder term, J_n, which appears in the expansion of Proposition 4, in the integral form

$$\begin{aligned} J_n = \Bigg[\int\int & \left\{ (\theta_1 - \bar{X}) \frac{\partial D(t_1, t_2)}{\partial t_1}\Big|_{t_1=\bar{X},t_2=\overline{X^2}} + (\theta_2 - \overline{X^2}) \frac{\partial D(t_1, t_2)}{\partial t_2}\Big|_{t_1=\bar{X},t_2=\overline{X^2}} \right\} \\ & \times \frac{1}{6}\left\{ (\theta_1 - \bar{X})^3 \frac{\partial^2 \lambda_1}{\partial t_1^2}\Big|_{t_1=\bar{X},t_2=\overline{X^2}} + 3(\theta_1 - \bar{X})^2(\theta_2 - \overline{X^2}) \frac{\partial^2 \lambda_1}{\partial t_1 \partial t_2}\Big|_{t_1=\bar{X},t_2=\overline{X^2}} \right. \\ & \left. + 3(\theta_1 - \bar{X})(\theta_2 - \overline{X^2})^2 \frac{\partial^2 \lambda_2}{\partial t_1 \partial t_2}\Big|_{t_1=\bar{X},t_2=\overline{X^2}} + (\theta_2 - \overline{X^2})^3 \frac{\partial^2 \lambda_2}{\partial t_2^2}\Big|_{t_1=\bar{X},t_2=\overline{X^2}} \right\} \\ & \times e^{-\frac{0.5n(\theta_1 - \bar{X})^2}{\sigma_n^2 - (\sigma_{XX^2 n})^2/\sigma^2_{X^2 n}} - \frac{0.5n(\theta_2 - \overline{X^2})^2}{\sigma^2_{X^2 n} - (\sigma_{XX^2 n})^2/\sigma_n^2} + \frac{n(\theta_1 - \bar{X})(\theta_2 - \overline{X^2})}{\sigma^2_{X^2 n}\sigma_n^2/\sigma_{XX^2 n} - \sigma_{XX^2 n}}} \pi(\theta_1, \theta_2) d\theta_1 d\theta_2 \Bigg] \\ & \times \left\{ \int\int e^{-\frac{0.5n(\theta_1 - \bar{X})^2}{\sigma_n^2 - (\sigma_{XX^2 n})^2/\sigma^2_{X^2 n}} - \frac{0.5n(\theta_2 - \overline{X^2})^2}{\sigma^2_{X^2 n} - (\sigma_{XX^2 n})^2/\sigma_n^2} + \frac{n(\theta_1 - \bar{X})(\theta_2 - \overline{X^2})}{\sigma^2_{X^2 n}\sigma_n^2/\sigma_{XX^2 n} - \sigma_{XX^2 n}}} \pi(\theta_1, \theta_2) d\theta_1 d\theta_2 \right\}^{-1}, \end{aligned} \tag{A15}$$

where the corresponding derivatives of λ_1 and λ_2 are defined in (A14).

REFERENCES

BLEISTEIN, N. & HANDELSMAN, R. A. (2010). *Asymptotic expansions of integrals*. New York: Courier Dover Publications.

CARLIN, B. P. & LOUIS, T. A. (2000). *Bayes and Empirical Bayes Methods for Data Analysis*. New York: Chapman and Hall/CRC.

CONSONNI, G. & VERONESE, P. (1992). Conjugate priors for exponential families having quadratic variance functions. *Journal of the American Statistical Association* **87**, 1123–1127.

DASGUPTA, A. (2008). *Asymptotic theory of statistics and probability*. New York: Springer.

DICICCIO, T. J., KASS, R. E., RAFTERY, A. & WASSERMAN, L. (1997). Computing bayes factors by combining simulation and asymptotic approximations. *Journal of the American Statistical Association* **92**, 903–915.

EFRON, B. & MORRIS, C. (1972). Limiting the risk of bayes and empirical bayes estimatorspart ii: The empirical bayes case. *Journal of the American Statistical Association* **67**, 130–139.

ERKANLI, A. (1994). Laplace approximations for posterior expectations when the mode occurs at the boundary of the parameter space. *Journal of the American Statistical Association* **89**, 250–258.

EVANS, M. & SWARTZ, T. (1995). Methods for approximating integrals in statistics with special emphasis on bayesian integration problems. *Statistical Science* **10**, 254–272.

JAMES, W. & STEIN, C. (1961). Estimation with quadratic loss. In *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability*, vol. 1.

Johnson, R. A. (1970). Asymptotic expansions associated with posterior distributions. *The annals of mathematical statistics* **41**, 851–864.

Kass, R. E. & Vaidyanathan, S. K. (1992). Approximate bayes factors and orthogonal parameters, with application to testing equality of two binomial proportions. *Journal of the Royal Statistical Society. Series B (Methodological)* , 129–144.

Lazar, N. & Mykland, P. A. (1998). An evaluation of the power and conditionality properties of empirical likelihood. *Biometrika* **85**, 523–534.

Lazar, N. A. (2003). Bayesian empirical likelihood. *Biometrika* **90**, 319–326.

Lieberman, O. (1994). A laplace approximation to the moments of a ratio of quadratic forms. *Biometrika* **81**, 681–690.

Miyata, Y. (2004). Fully exponential laplace approximations using asymptotic modes. *Journal of the American Statistical Association* **99**, 1037–1049.

Monahan, J. F. & Boos, D. D. (1992). Proper likelihoods for bayesian analysis. *Biometrika* **79**, 271–278.

Newton, M. A. & Raftery, A. E. (1994). Approximate bayesian inference with the weighted likelihood bootstrap. *Journal of the Royal Statistical Society. Series B (Methodological)* , 3–48.

Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* **75**, 237–249.

Owen, A. B. (2001). *Empirical likelihood*, vol. 92. New York: Chapman and Hall/CRC.

Polson, N. G. (1991). A representation of the posterior mean for a location model. *Biometrika* **78**, 426–430.

Qin, J. & Lawless, J. (1994). Empirical likelihood and general estimating equations. *The Annals of Statistics* , 300–325.

Schisterman, E. F., Faraggi, D., Browne, R., Freudenheim, J., Dorn, J., Muti, P., Armstrong, D., Reiser, B. & Trevisan, M. (2001). Tbars and cardiovascular disease in a population-based sample. *Journal of cardiovascular risk* **8**, 219–225.

Stein, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *Proceedings of the Third Berkeley symposium on mathematical statistics and probability*, vol. 1.

Sweeting, T. J. (1995). A framework for bayesian and likelihood approximations in statistics. *Biometrika* **82**, 1–23.

Tierney, L. & Kadane, J. B. (1986). Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association* **81**, 82–86.

Tierney, L., Kass, R. E. & Kadane, J. B. (1989). Fully exponential laplace approximations to expectations and variances of nonpositive functions. *Journal of the American Statistical Association* **84**, 710–716.

Vexler, A., Liu, S., Kang, L. & Hutson, A. D. (2009). Modifications of the empirical likelihood interval estimation with improved coverage probabilities. *Communications in StatisticsSimulation and Computation®* **38**, 2171–2183.

Vexler, A., Tsai, W.-M. & Malinovsky, Y. (2012). Estimation and testing based on data subject to measurement errors: from parametric to non-parametric likelihood methods. *Statistics in Medicine* **31**, 2498–2512.

Wu, J., Wong, A. & Jiang, G. (2003). Likelihood-based confidence intervals for a log-normal mean. *Statistics in medicine* **22**, 1849–1860.

Yee, J. L., Johnson, W. O. & Samaniego, F. J. (2002). Asymptotic approximations to posterior distributions via conditional moment equations. *Biometrika* **89**, 755–767.

Yu, J., Vexler, A., Kim, S.-E. & Hutson, A. D. (2011). Two-sample empirical likelihood ratio tests for medians in application to biomarker evaluations. *Canadian Journal of Statistics* **39**, 671–689.

Yu, J., Vexler, A. & Tian, L. (2010). Analyzing incomplete data subject to a threshold using empirical likelihood methods: an application to a pneumonia risk study in an icu setting. *Biometrics* **66**, 123–130.

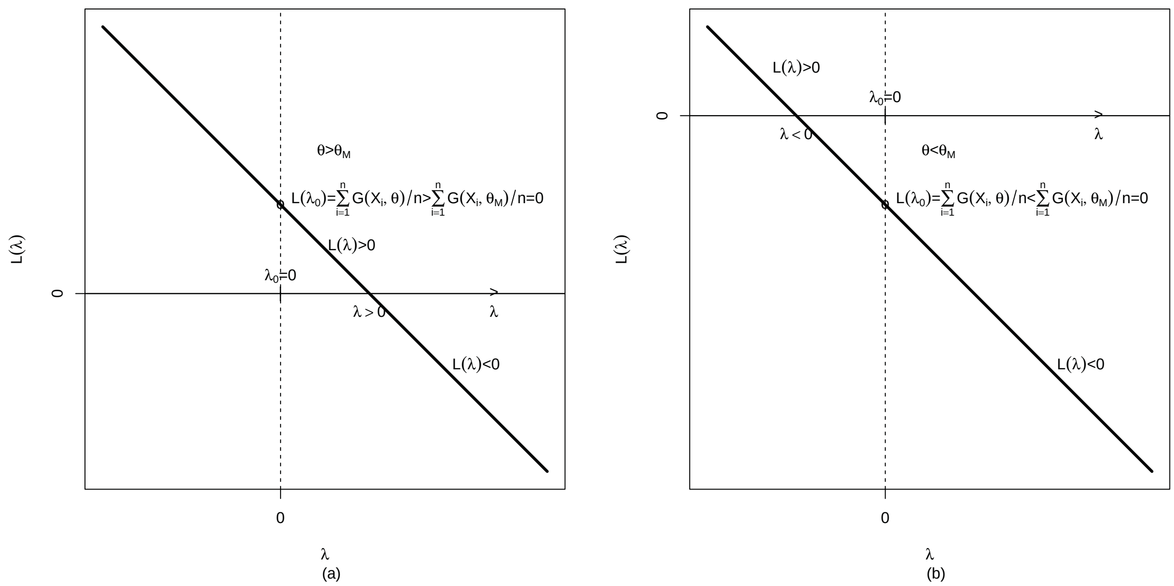

Fig. 1: The schematic behaviors of $L(\lambda)$ plotted against λ (the axis of abscissa), when (a): $\theta > \theta_M$ and (b): $\theta < \theta_M$, respectively.

Table 1: Monte Carlo means and variances of $\bar{X}$ and the MLE.

	$X_1, \ldots, X_n \sim N(1,1)$		$X_1, \ldots, X_n \sim \log N(0,1)$			
	$\bar{X}$		$\bar{X}$		MLE	
n	mean	var	mean	var	mean	var
10	1.0025	0.0997	1.6511	0.4563	1.7964	0.6137
20	0.9971	0.0499	1.6447	0.2383	1.7056	0.2480
30	0.9993	0.0339	1.6530	0.1539	1.6880	0.1534
50	1.0023	0.0204	1.6516	0.0921	1.6729	0.0870
75	0.9988	0.0138	1.6586	0.0627	1.6647	0.0573

Table 2: Monte Carlo means and variances of the estimator $\hat{\theta}$ by (2) and its asymptotic forms $\hat{\theta}\_P1$ and $\hat{\theta}\_C1$ obtained by Proposition 1 and Corollary 1, respectively.

	$X_1, \ldots, X_n \sim N(1,1)$						$X_1, \ldots, X_n \sim \log N(0,1)$					
	Prior: $\pi \sim N(0,1)$						Prior: $\pi \sim N(0,1)$					
	$\hat{\theta}$		$\hat{\theta}\_P1$		$\hat{\theta}\_C1$		$\hat{\theta}$		$\hat{\theta}\_P1$		$\hat{\theta}\_C1$	
n	mean	var	mean	var	mean	var	mean	var	mean	var	mean	var
10	0.9161	0.0965	0.9157	0.0952	0.9157	0.0952	1.4109	0.1773	1.2376	0.2383	1.2008	0.2760
20	0.9487	0.0489	0.9517	0.0483	0.9517	0.0483	1.5278	0.1094	1.3709	0.1362	1.3607	0.1454
30	0.9618	0.0324	0.9636	0.0321	0.9636	0.0321	1.5641	0.0828	1.4341	0.0970	1.4303	0.1001
50	0.9805	0.0197	0.9815	0.0197	0.9815	0.0197	1.6003	0.0609	1.5033	0.0634	1.5023	0.0640

75	0.9867	0.0132	0.9871	0.0132	0.9871	0.0132	1.6198	0.0452	1.5467	0.0450	1.5465	0.0452
	Prior: $\pi \sim N(1, 0.5^2)$						Prior: $\pi \sim N(\exp(1/2), 0.5^2)$					
	$\hat{\theta}$		$\hat{\theta}\_P1$		$\hat{\theta}\_C1$		$\hat{\theta}$		$\hat{\theta}\_P1$		$\hat{\theta}\_C1$	
n	mean	var	mean	var	mean	var	mean	var	mean	var	mean	var
10	1.0003	0.0579	1.0001	0.0569	1.0001	0.0569	1.5792	0.1110	1.5141	0.1110	1.5138	0.1110
20	1.0004	0.0359	1.0004	0.0364	1.0004	0.0364	1.6426	0.0762	1.5734	0.0727	1.5734	0.0727
30	1.0000	0.0268	0.9999	0.0271	0.9999	0.0271	1.6594	0.0648	1.5929	0.0600	1.5929	0.0600
50	0.9998	0.0167	0.9996	0.0168	0.9996	0.0168	1.6752	0.0521	1.6152	0.0466	1.6152	0.0466
75	1.0000	0.0123	1.0001	0.0123	1.0001	0.0123	1.6857	0.0417	1.6328	0.0362	1.6328	0.0362
	Prior: $\pi \sim N(1, 0.1^2)$						Prior: $\pi \sim N(\exp(1/2), 0.1^2)$					
	$\hat{\theta}$		$\hat{\theta}\_P1$		$\hat{\theta}\_C1$		$\hat{\theta}$		$\hat{\theta}\_P1$		$\hat{\theta}\_C1$	
n	mean	var	mean	var	mean	var	mean	var	mean	var	mean	var
10	1.0003	0.0024	1.0002	0.0019	1.0001	0.0018	1.6191	0.0081	1.6115	0.0080	1.6119	0.0076
20	1.0002	0.0019	1.0002	0.0019	1.0002	0.0019	1.6341	0.0027	1.6205	0.0047	1.6205	0.0047
30	0.9996	0.0020	0.9995	0.0021	0.9995	0.0021	1.6383	0.0019	1.6252	0.0036	1.6252	0.0036
50	0.9994	0.0023	0.9994	0.0024	0.9994	0.0024	1.6422	0.0018	1.6300	0.0029	1.6300	0.0029
75	1.0008	0.0025	1.0008	0.0026	1.0008	0.0026	1.6439	0.0019	1.6327	0.0026	1.6327	0.0026
	Prior: $\pi \sim N(0, 0.5^2)$						Prior: $\pi \sim N(\exp(1/2) - 1, 0.5^2)$					
	$\hat{\theta}$		$\hat{\theta}\_P1$		$\hat{\theta}\_C1$		$\hat{\theta}$		$\hat{\theta}\_P1$		$\hat{\theta}\_C1$	
n	mean	var	mean	var	mean	var	mean	var	mean	var	mean	var
10	0.7546	0.1271	0.7522	0.1259	0.7519	0.1259	1.2500	0.2126	1.0942	0.3380	1.0794	0.3563
20	0.8344	0.0635	0.8425	0.0610	0.8425	0.0610	1.3610	0.1268	1.2099	0.2188	1.2066	0.2227
30	0.8774	0.0418	0.8839	0.0404	0.8839	0.0404	1.4238	0.0878	1.2900	0.1524	1.2888	0.1537
50	0.9232	0.0228	0.9269	0.0224	0.9269	0.0224	1.4945	0.0550	1.3898	0.0885	1.3896	0.0887
75	0.9484	0.0149	0.9503	0.0148	0.9503	0.0148	1.5281	0.0419	1.4467	0.0604	1.4466	0.0605
	Prior: $\pi \sim \{N(-1, 0.5^2) + N(1, 0.5^2)\}/2$						Prior: $\pi \sim \{N(-e^{1/2}, 0.5^2) + N(e^{1/2}, 0.5^2)\}/2$					
	$\hat{\theta}$		$\hat{\theta}\_P1$		$\hat{\theta}\_C1$		$\hat{\theta}$		$\hat{\theta}\_P1$		$\hat{\theta}\_C1$	
n	mean	var	mean	var	mean	var	mean	var	mean	var	mean	var
10	0.9903	0.0649	0.9908	0.0639	0.9987	0.0577	1.5821	0.1098	1.5163	0.1083	1.5161	0.1083
20	0.9945	0.0344	0.9947	0.0346	0.9959	0.0340	1.6395	0.0766	1.5697	0.0730	1.5697	0.0730
30	1.0000	0.0262	1.0000	0.0264	1.0005	0.0263	1.6629	0.0626	1.5967	0.0577	1.5967	0.0577
50	0.9948	0.0173	0.9948	0.0174	0.9949	0.0173	1.6721	0.0508	1.6127	0.0461	1.6127	0.0461
75	0.9984	0.0121	0.9984	0.0121	0.9984	0.0121	1.6860	0.0412	1.6338	0.0361	1.6338	0.0361
	Prior: $\pi \sim \{N(-1, 0.1^2) + N(1, 0.1^2)\}/2$						Prior: $\pi \sim \{N(-e^{1/2}, 0.1^2) + N(e^{1/2}, 0.1^2)\}/2$					
	$\hat{\theta}$		$\hat{\theta}\_P1$		$\hat{\theta}\_C1$		$\hat{\theta}$		$\hat{\theta}\_P1$		$\hat{\theta}\_C1$	
n	mean	var	mean	var	mean	var	mean	var	mean	var	mean	var
10	0.9962	0.0062	0.9979	0.0036	0.9998	0.0017	1.6180	0.0094	1.6099	0.0093	1.6105	0.0087
20	0.9988	0.0018	0.9988	0.0019	0.9988	0.0019	1.6335	0.0028	1.6195	0.0051	1.6195	0.0051
30	0.9999	0.0020	1.0000	0.0021	1.0000	0.0021	1.6391	0.0019	1.6267	0.0034	1.6267	0.0034
50	0.9996	0.0023	0.9996	0.0024	0.9996	0.0024	1.6425	0.0018	1.6306	0.0028	1.6306	0.0028
75	0.9995	0.0024	0.9994	0.0025	0.9994	0.0025	1.6429	0.0019	1.6317	0.0027	1.6317	0.0027
	Prior: $\pi \sim U(0, 1.5)$						Prior: $\pi \sim U(0, \exp(1/2) + 0.5)$					
	$\hat{\theta}$		$\hat{\theta}\_P1$				$\hat{\theta}$		$\hat{\theta}\_P1$			
n	mean	var	mean	var			mean	var	mean	var		
10	0.9424	0.0613	0.9423	0.0606			1.4692	0.1204	1.3625	0.1540		
20	0.9744	0.0376	0.9758	0.0378			1.5564	0.0680	1.4624	0.0865		
30	0.9913	0.0279	0.9924	0.0279			1.5949	0.0521	1.5155	0.0628		
50	0.9996	0.0192	1.0000	0.0193			1.6341	0.0388	1.5712	0.0428		
75	0.9992	0.0132	0.9992	0.0132			1.6511	0.0324	1.5997	0.0339		
	Prior: $\pi \sim U(0.75, 1.25)$						Prior: $\pi \sim U(\exp(1/2) - 0.25, \exp(1/2) + 0.25)$					
	$\hat{\theta}$		$\hat{\theta}\_P1$				$\hat{\theta}$		$\hat{\theta}\_P1$			

n	mean	var	mean	var	mean	var	mean	var
10	0.9991	0.0054	0.9992	0.0050	1.6190	0.0064	1.6068	0.0072
20	1.0007	0.0056	1.0006	0.0058	1.6289	0.0051	1.6121	0.0068
30	0.9987	0.0061	0.9986	0.0063	1.6336	0.0048	1.6164	0.0065
50	0.9985	0.0065	0.9984	0.0066	1.6390	0.0052	1.6220	0.0063
75	1.0010	0.0067	1.0011	0.0068	1.6405	0.0059	1.6235	0.0066

Table 3: The MC means and variances of $\bar{X}_{.j}$ and the estimator $\hat{\theta}_{Ej}$ by (6) based on $MVN\{(1,1,1)^T, I\}$.

n	$\bar{X}_{.1}$	$\bar{X}_{.2}$	$\bar{X}_{.3}$	$V(\bar{X}_{.1})$	$V(\bar{X}_{.2})$	$V(\bar{X}_{.3})$	$\hat{\theta}_{E1}$	$\hat{\theta}_{E2}$	$\hat{\theta}_{E3}$	$V(\hat{\theta}_{E1})$	$V(\hat{\theta}_{E2})$	$V(\hat{\theta}_{E3})$
10	(0.995	0.995	0.989)	(0.096	0.100	0.099)	(0.994	0.995	0.990)	(0.062	0.065	0.064)
20	(1.002	0.993	1.000)	(0.052	0.050	0.048)	(1.001	0.995	0.999)	(0.033	0.032	0.031)
30	(0.999	1.002	1.000)	(0.033	0.034	0.034)	(0.999	1.001	1.000)	(0.021	0.021	0.022)
50	(0.999	1.000	1.000)	(0.020	0.020	0.020)	(0.999	1.000	1.000)	(0.013	0.013	0.013)
75	(1.003	1.000	0.999)	(0.014	0.014	0.014)	(1.003	1.000	1.000)	(0.009	0.009	0.009)

Table 4: The MC means and variances of $\bar{X}_{.j}$ and the estimator $\hat{\theta}_{Ej}$ by (6) based on $MVLogNorm\{(0,0,0)^T, I\}$.

n	$\bar{X}_{.1}$	$\bar{X}_{.2}$	$\bar{X}_{.3}$	$V(\bar{X}_{.1})$	$V(\bar{X}_{.2})$	$V(\bar{X}_{.3})$	$\hat{\theta}_{E1}$	$\hat{\theta}_{E2}$	$\hat{\theta}_{E3}$	$V(\hat{\theta}_{E1})$	$V(\hat{\theta}_{E2})$	$V(\hat{\theta}_{E3})$
10	(1.630	1.656	1.638)	(0.446	0.581	0.445)	(1.631	1.655	1.638)	(0.398	0.531	0.395)
20	(1.657	1.654	1.654)	(0.235	0.232	0.239)	(1.657	1.654	1.654)	(0.206	0.204	0.211)
30	(1.644	1.654	1.640)	(0.161	0.162	0.148)	(1.645	1.653	1.640)	(0.140	0.141	0.128)
50	(1.650	1.642	1.640)	(0.093	0.088	0.085)	(1.650	1.642	1.640)	(0.080	0.076	0.073)
75	(1.649	1.645	1.653)	(0.061	0.059	0.063)	(1.649	1.645	1.653)	(0.052	0.051	0.054)

Table 5: The MC means and variances of $\bar{X}_{.j}$ and the estimator $\hat{\theta}_{Ej}$ by (6) based on $MVN\{(1,1,1)^T, \Sigma\}$.

n	$\bar{X}_{.1}$	$\bar{X}_{.2}$	$\bar{X}_{.3}$	$V(\bar{X}_{.1})$	$V(\bar{X}_{.2})$	$V(\bar{X}_{.3})$	$\hat{\theta}_{E1}$	$\hat{\theta}_{E2}$	$\hat{\theta}_{E3}$	$V(\hat{\theta}_{E1})$	$V(\hat{\theta}_{E2})$	$V(\hat{\theta}_{E3})$
10	(0.995	0.993	0.993)	(0.101	0.099	0.101)	(0.994	0.993	0.994)	(0.083	0.082	0.082)
20	(0.996	0.997	0.998)	(0.049	0.049	0.049)	(0.997	0.997	0.998)	(0.040	0.040	0.039)
30	(1.001	1.000	1.000)	(0.033	0.035	0.035)	(1.001	1.000	1.000)	(0.027	0.028	0.028)
50	(1.000	1.003	0.997)	(0.019	0.020	0.020)	(0.999	1.002	0.998)	(0.016	0.016	0.016)
75	(1.002	0.998	0.999)	(0.013	0.014	0.014)	(1.001	0.999	1.000)	(0.011	0.011	0.011)

Table 6: The MC means and variances of $\bar{X}_{.j}$ and the estimator $\hat{\theta}_{Ej}$ by (6) based on $MVLogNorm\{(0,0,0)^T, \Sigma\}$.

n	$\bar{X}_{.1}$	$\bar{X}_{.2}$	$\bar{X}_{.3}$	$V(\bar{X}_{.1})$	$V(\bar{X}_{.2})$	$V(\bar{X}_{.3})$	$\hat{\theta}_{E1}$	$\hat{\theta}_{E2}$	$\hat{\theta}_{E3}$	$V(\hat{\theta}_{E1})$	$V(\hat{\theta}_{E2})$	$V(\hat{\theta}_{E3})$
10	1.663	1.637	1.650	0.551	0.471	0.480	1.663	1.637	1.650	0.520	0.441	0.449
20	1.651	1.658	1.652	0.237	0.232	0.218	1.651	1.658	1.652	0.218	0.215	0.202
30	1.651	1.652	1.641	0.148	0.161	0.149	1.650	1.651	1.642	0.136	0.148	0.137
50	1.652	1.648	1.644	0.094	0.095	0.092	1.652	1.648	1.645	0.086	0.087	0.084
75	1.648	1.648	1.645	0.064	0.061	0.061	1.648	1.648	1.645	0.058	0.056	0.056

Table 7: The proposed mean and 95% CI estimations compared with those based on averages $\bar{X}$.

Case	$\hat{\theta}$ by (2)	95% CI of $\hat{\theta}$	$\bar{X}$	95% CI of $\bar{X}$
Prior: $\pi \sim N(1,1)$	1.410476	(1.356624, 1.464328)	1.407113	(0.4740198, 2.3402069)
Prior: $\pi \sim N(1.84, 0.8^2)$	1.411334	(1.357493 , 1.465174)		
Prior: $\pi \sim N(1.84, 0.1^2)$	1.44765	(1.395704 , 1.499596)		
Prior: $\pi \sim N(1.84, (1/300)^2)$	1.833769	(1.827283 , 1.840255)		
Double empirical Bayesian method	$\hat{\theta}_E$ by (3)	95% CI of $\hat{\theta}_E$		
	1.40706	(1.40693 , 1.40719)		
Control	$\hat{\theta}$ by (2)	95% CI of $\hat{\theta}$	$\bar{X}$	95% CI of $\bar{X}$
Prior: $\pi \sim N(1,1)$	1.307575	(1.268270, 1.346879)	1.305487	(0.6245805, 1.9863928)
Prior: $\pi \sim N(1.44, 0.48^2)$	1.307944	(1.268666, 1.347222)		
Prior: $\pi \sim N(1.44, 0.1^2)$	1.31313	(1.274586,1.351675)		
Prior: $\pi \sim N(1.44, (1/300)^2)$	1.43842	(1.431975 , 1.444865)		
Double empirical Bayesian method	$\hat{\theta}_E$ by (3)	95% CI of $\hat{\theta}_E$		
	1.30554	(1.30541, 1.30567)		

Table 8: The Bootstrap type mean and estimators of the variances of $\hat{\theta}$ by (2) and $\bar{X}$.

$n_1 = 50$	$\bar{X}$	Variance of $\bar{X}$	$\hat{\theta}$	Variance of $\hat{\theta}$	$\mu = \sum_{i=1}^{B} \tilde{\mu}_i / B$
Prior: $\pi \sim N(1,1)$	1.465098	0.003590306	1.480343	0.003760704	1.46829
Prior: $\pi \sim N(1.1, 0.1^2)$	1.465774	0.006393996	1.370295	0.002919261	1.41322
Prior: $\pi \sim N(1.3, 0.1^2)$	1.467083	0.004116504	1.420281	0.005468906	1.4843
Prior: $\pi \sim N(1.4, 0.1^2)$	1.465391	0.003834	1.448914	0.002129113	1.47422
Prior: $\pi \sim N(1.84, 0.1^2)$	1.466738	0.003594508	1.644271	0.032628722	1.47131
$n_1 = 70$	$\bar{X}$	Variance of $\bar{X}$	$\hat{\theta}$	Variance of $\hat{\theta}$	$\mu = \sum_{i=1}^{B} \tilde{\mu}_i / B$
Prior: $\pi \sim N(1,1)$	1.465068	0.00213376	1.476581	0.00236309	1.461262
Prior: $\pi \sim N(1.1, 0.1^2)$	1.465698	0.004950603	1.388167	0.001362499	1.412425
Prior: $\pi \sim N(1.3, 0.1^2)$	1.46639	0.002382526	1.429889	0.003684917	1.48225
Prior: $\pi \sim N(1.4, 0.1^2)$	1.465945	0.00362532	1.453863	0.001748157	1.426625
Prior: $\pi \sim N(1.84, 0.1^2)$	1.465883	0.002038279	1.612017	0.022057167	1.469325

Table 9: The Bootstrap type mean and estimators of the variances of $\hat{\theta}_E$ by (3) and $\bar{X}$.

Double empirical	$n_1 = 50$	$\bar{X}$	Variance of $\bar{X}$	$\hat{\theta}_E$	Variance of $\hat{\theta}_E$	$\mu = \sum_{i=1}^{B} \tilde{\mu}_i / B$
Bayesian method		1.465026	0.003575563	1.464817	0.003617737	1.46839
Double empirical	$n_1 = 70$	$\bar{X}$	Variance of $\bar{X}$	$\hat{\theta}_E$	Variance of $\hat{\theta}_E$	$\mu = \sum_{i=1}^{B} \tilde{\mu}_i / B$
Bayesian method		1.46535	0.0021899	1.465335	0.002191446	1.4593