Frequentist Standard Errors of Bayes Estimators

DongHyuk Lee, Raymond J. Carroll, and Samiran Sinha
Department of Statistics, Texas A&M University, College Station, TX 77843-3143, USA
dhyuklee@stat.tamu.edu, carroll@stat.tamu.edu, and sinha@stat.tamu.edu

Summary

Frequentist standard errors are a measure of uncertainty of an estimator, and the basis for statistical inferences. Frequestist standard errors can also be derived for Bayes estimators. However, except in special cases, the computation of the standard error of Bayesian estimators requires bootstrapping, which in combination with Markov chain Monte Carlo (MCMC) can be highly time consuming. We discuss an alternative approach for computing frequentist standard errors of Bayesian estimators, including importance sampling. Through several numerical examples we show that our approach can be much more computationally efficient than the standard bootstrap.

Key Words: Bootstrap; Importance sampling; Markov chain; Posterior distribution; Standard error; Tail probability.

Running title: Frequentist Uncertainties of Bayes Estimators

1 Introduction

Suppose that $f(\bullet|\theta)$ is the data generating density and $\pi(\theta)$ is the prior distribution for the parameter θ. Let $\boldsymbol{D}$ be the observed data. Then the posterior distribution of θ is

$$\pi(\theta|\boldsymbol{D}) = K_\pi f(\boldsymbol{D}|\theta)\pi(\theta),$$

where K_π denotes the normalizing constant. There are several posterior summaries, such as the mean, $m(\boldsymbol{D}) = E(\theta|\boldsymbol{D}) = \int \theta\pi(\theta|\boldsymbol{D})d\theta$, the posterior median $\widetilde{m}(\boldsymbol{D})$, which satisfies $\int_{-\infty}^{\widetilde{m}(\boldsymbol{D})} \pi(\theta|\boldsymbol{D})d\theta = 0.5$, the α^{th} quantile $q_\alpha(\boldsymbol{D})$, that satisfies $\int_{-\infty}^{q_\alpha(\boldsymbol{D})} \pi(\theta|\boldsymbol{D})d\theta = \alpha$ for any $\alpha \in (0, 1)$, and the posterior mode $m_o(\boldsymbol{D}) = \arg\max_\theta \pi(\theta|\boldsymbol{D})$. Suppose that by $s(\boldsymbol{D})$ we refer to any summary of the posterior distribution. Throughout this article we assume that $\boldsymbol{D}$ consists of $(X_1, \ldots, X_n)$ independently and identically distributed observations. The goal of this paper is to discuss approaches of computing the frequentist standard error of $s(\boldsymbol{D})$.

Under a large sample, the observed data dominates the prior information in a Bayesian framework, and under standard regularity conditions, the posterior distribution of finite dimensional model parameters converges to the Gaussian distribution with the maximum likelihood estimator and the inverse of the Fisher Information matrix as the asymptotic mean and asymptotic variance, respectively. This asymptotic connection indicates that the Bayesian philosophy of integrating the observed data and the prior knowledge can be seen as a general procedure that encompasses the frequentist procedure as a special case. Therefore, as pointed out by a referee, frequentist standard error of a Bayes estimator is a way of assessing uncertainty of the general procedure. Particularly, for a large sample, the frequentist variance of the posterior mean converges to the inverse of the Fisher's information matrix. From the Bayesian perspective, frequentist standard errors can be used for comparing uncertainty of estimators under different priors (Efron, 2015). Although the posterior standard error (the standard deviation of the posterior distribution) is a measure of uncertainty of the posterior distribution, Efron (2015) argued that in a Bayesian paradigm, accuracy of a Bayes estimator, such as posterior mean could be judged based on the posterior distribution given that the prior distribution of the parameter reflects the truth in some degree. Therefore, finding accuracy of a Bayes estimator in an objective way among different subjective and objective priors is important. In fact, Berger (2006) discussed that the "pseudo-Bayes procedures" where subjective, objective, or a mixture of subjective and objective priors are used, often fail to provide any guidance on the performance of true subjective or objective Bayesian analysis. He then pointed out the necessity of validating these Bayesian approaches, and frequentist standard error of a posterior summary can be seen as a measure of such validation. Although Bayes factor is a way of comparing Bayesian procedures, many practitioners still want to compare estimators based on a frequentist uncertainty measure. Therefore, despite an apparent lack of coherence for incorporating a frequentist comparisons among Bayes procedures, it provides a measure of comparing uncertainties of the estimators. Of course, we should not use this measure solely to elicit the optimal prior for a Bayes procedure as for a proper comparison one should consider consistency, posterior convergence rate, along with the uncertainty of the estimator.

Efron (2015) proposed methods for computing frequentist standard errors of the posterior mean of a function of a parameter. In particular, he derived the approximate frequentist standard deviation of the posterior mean of a parameter based on the delta method. Suppose that T is the sufficient statistic. Following our notations, his formula for the approximate standard deviation of $\widehat{t} = E\{t(\theta)|\boldsymbol{D}\} = E\{t(\theta)|T\}$, the posterior mean of $t(\theta)$, a function of θ, is $[\text{cov}\{t(\theta), \alpha_T(\theta)|T\}^T V_\theta \text{cov}\{t(\theta), \alpha_T(\theta)|T\}]^{1/2}$, where $\alpha_T(\theta) = \partial \text{log}\{f_\theta(T)\}/\partial T$ denotes the gradient of $\log\{f_\theta(T)\}$ with respect to T, the sufficient statistic for θ, $f_\theta(T)$ is the density for the sufficient statistic T, and V_θ denotes the variance of the sufficient statistic. For application of this method it is critical that V_θ is readily available. Secondly, one key component of the delta method is the gradient of $\widehat{t}$ with respect to T, and here this gradient is expressed as the posterior covariance $\text{cov}\{t(\theta), \alpha_T(\theta)|T\}$. Expressing $\partial\widehat{t}/\partial T$ as $\text{cov}\{t(\theta), \alpha_T(\theta)|T\}$ critically relies on the fact that $\widehat{t}$ is a posterior expectation. This posterior covariance is easy to estimate from a sample from the posterior distribution of θ. Therefore, when V_θ is available and $\widehat{t}$ is a posterior expectation, then Efron's formula is easy to apply and it is computationally fast.

In a special case with the exponential family of distributions where θ is considered to be the natural or canonical parameter vector, along with an uninformative prior for θ, he showed that the standard error of the posterior mean of $t(\theta) = \theta$ can be computed without running the MCMC step to generate posterior samples for computing $\text{cov}\{\theta, \alpha_T(\theta)|T\}$. In lieu of the MCMC sampling, he used a parametric bootstrap resampling technique (Efron, 2012) to compute the posterior covariance term. Although the proposed method is applicable to only posterior means and when V_θ is easily available, the main advantage is that this method, when it is applicable, is much faster than the regular bootstrap procedure.

Inspired by this work we propose a general method of efficiently computing the frequentist standard error not only of the posterior mean but also of any posterior summary, $s(\boldsymbol{D})$. Our method is applicable for data generated from any parametric model, not necessarily from an exponential family of distributions. The proposed method relies on the bootstrap idea. Usually, the standard error of an estimator can be computed by the bootstrap method (Efron and Tibshirani, 1986), where the standard error is estimated by the standard deviation of the Bayes estimators obtained from a large number of bootstrap samples. On the other hand, the Bayes estimator for a bootstrap sample is usually calculated by drawing a large number of Markov chain Monte Carlo (MCMC) samples, which is often time consuming, and consequently drawing posterior samples for each of the bootstrap data can be a prohibitively time consuming task.

The main aim of this article is to reduce this computation time. To do so, the MCMC method will be used once to draw samples from the posterior distribution of the parameters given the original data. Then use these posterior samples along with the importance sampling idea to compute the posterior summary for each bootstrap data. The details are discussed in the following sections. To make it clearer, we want to re-state that in the proposed method, we do need bootstrap sampling, but we bypass the MCMC sampling for each bootstrap data by a clever use of the importance sampling method. Here is a brief description of the importance sampling method in a few words. Suppose that we are interested in estimating $\theta = \int g(x)f(x)dx$, where $f(x)$ is a density. With another density $h(x)$, we can re-write $\theta = \int g(x)\omega(x)h(x)dx$, where $\omega(x) = f(x)/h(x)$ is called the importance weight. Then the importance sampling estimator of θ is $\widehat{\theta} = m^{-1}\sum_{i=1}^{m} g(x_i)\omega(x_i)$, where $x_1, \ldots, x_m$ are iid from $h(x)$. This technique is quite useful for efficient estimation of tail probabilities, and is used for drawing bootstrap samples, specially for estimating standard error of small probabilities (Efron and Tibshirani, 1994, pp. 349). However, in this paper we are using importance sampling technique to compute estimators based on a bootstrap re-sampled data. Basically in the proposed approach bootstrap samples are drawn using standard bootstrap resampling technique and then importance sampling is used to compute the Bayes estimators. Although importance sampling idea has been used in many other contexts, including but not limited to the simulated maximum likelihood estimation, computer graphics, modelling stock market data, modelling linear and nonlinear dynamic processes (Liang, 2002), to the best of our knowledge, the use of this technique in the present context seems to be the new.

A brief outline of the remainder of the manuscript is as follows. In Section 2 we provide the background information. The main idea related to the posterior mean is discussed in Section 3, while Section 4 considers posterior quantiles and the posterior mode. Section 5 describes the results of two simulation studies and a real data examples. Section 6 contains conclusions.

2 Background

To motivate this research first we consider three commonly used models.

Logistic regression model: Suppose that $Y_1, \ldots, Y_n$ are independently drawn from the Bernoulli(p_i) distribution, where $p_i = \text{pr}(Y_i = 1|X_i) = \{1 + \exp(-\alpha - \beta X_i)\}^{-1}$ with a scalar covariate X_i. Assume priors $\alpha \sim \text{Normal}(a, \sigma^2)$ and $\beta \sim \text{Normal}(b, \tau^2)$, and let $\boldsymbol{D}$ denote the observed data $\{(X_i, Y_i), i = 1, \ldots, n\}$. Then the posterior distribution of α and β is

$$\begin{aligned}\pi(\alpha, \beta|\boldsymbol{D}) \;\propto\; & \prod_{i=1}^{n} \left\{ \frac{1}{1 + \exp(-\alpha - \beta X_i)} \right\}^{Y_i} \left\{ \frac{\exp(-\alpha - \beta X_i)}{1 + \exp(-\alpha - \beta X_i)} \right\}^{(1-Y_i)} \times \\ & \frac{\exp\{-(\alpha - a)^2/2\sigma^2\}}{\sqrt{2\pi\sigma^2}} \times \frac{\exp\{-(\beta - b)^2/2\tau^2\}}{\sqrt{2\pi\tau^2}}.\end{aligned}$$

For computing any posterior summary for $\pi(\alpha, \beta|\boldsymbol{D})$, usually we draw posterior samples from $\pi(\alpha, \beta|\boldsymbol{D})$ using the MCMC method. So, a numerical method is must for computing frequentist standard errors of any summary of the posterior distribution. In the simulation section, for illustration, we apply the proposed method on this model.

Linear measurement error model: Now, we consider the following simple linear regression problem, where using the observed data $\boldsymbol{D} = \{(Y_i, W_i), i = 1, \ldots, n\}$, we want to fit $Y_i = \alpha + X_i\beta + \epsilon_i$, where X_i is unobserved but we observed its surrogate variable W_i, and $\epsilon_i \sim \text{Normal}(0, \sigma_\epsilon^2)$. The observed surrogate W_i is associated with the true X_i through the classical additive measurement error model $W_i = X_i + U_i$, where $U_i \sim \text{Normal}(0, \sigma_u^2)$ and σ_u^2 is considered to be known for simplicity. We further assume that measurement error is nondifferential such that Y_i is conditionally independent of W_i given the true X_i (Carroll et al., 2006, pp. 36), and $X_i \sim \text{Normal}(\mu_x, \sigma_x^2)$.

It is well-known that the simple linear regression of Y on W will cause an attenuation towards 0 by the multiplicative factor $\sigma_x^2/(\sigma_x^2 + \sigma_u^2)$. One of the corrections for attenuation is the method of moments. That is, the resulting estimator $\widehat{\beta} = \widehat{\beta}_w \widehat{\sigma}_w^2/(\widehat{\sigma}_w^2 - \sigma_u^2)$, where $\widehat{\beta}_w$ is the OLS estimator ignoring measurement error, $\widehat{\sigma}_w^2$ is the sample variance of the observed W, and σ_u^2 is the variance of U (Carroll et al., 2006, Section 3.4.1; Fuller, 1987, Section 2.5). In addition, it is well-known that $\widehat{\beta}$ has no finite moments, because the denominator term $\widehat{\sigma}_w^2 - \sigma_u^2$ can get arbitrarily close to zero (Fuller, 1987). Therefore, Bayesian calculations are an attractive alternative.

We attempt to use a Bayesian inference for the parameters $\theta = (\alpha, \beta, \mu_x, \sigma_x^2, \sigma_\epsilon^2)$ in which α and β are the main parameters of interest. Assigning normal priors, Normal$(0, \sigma_\alpha^2)$, Normal$(0, \sigma_\beta^2)$, Normal$(0, \sigma_\mu^2)$ for α, β, μ_x, respectively and inverse gamma priors IG(δ_x, λ_x), IG$(\delta_\epsilon, \lambda_\epsilon)$ for σ_x^2, σ_ϵ^2, respectively (Carroll et al., 2006, Section 9.4), the joint posterior distribution of θ and the latent variable $X = (X_1, \ldots, X_n)$ is

$$\begin{aligned}\pi(\theta, X|\boldsymbol{D}) \;\; \propto \;\; & (\sigma_\epsilon^2)^{-n/2-\delta_\epsilon-1}(\sigma_x^2)^{-n/2-\delta_x-1}\exp\Bigg\{ -\frac{\sum_{i=1}^n (Y_i - \alpha - X_i\beta)^2/2 + \lambda_\epsilon}{\sigma_\epsilon^2} \\ & -\frac{\sum_{i=1}^n (W_i - X_i)^2/2 + \lambda_u}{\sigma_u^2} - \frac{\sum_{i=1}^n (X_i - \mu_x)^2/2 + \lambda_x}{\sigma_x^2} - \frac{\alpha^2}{\sigma_\alpha^2} - \frac{\beta^2}{\sigma_\beta^2} - \frac{\mu_x^2}{\sigma_\mu^2}\Bigg\}.\end{aligned}$$

Due to the conjugacy of the prior distributions, it is easy to apply the Gibbs sampler to draw posterior samples from $\pi(\theta, X|\boldsymbol{D})$. Specifically, the conditional posterior distributions of α and β given other parameters and the latent variable X are normal distributions so that we can easily obtain their posterior summaries. However, it is not an easy problem to find the variances of their posterior summaries mainly because they are dependent on the unobserved X. Thus a numerical method is required.

Weibull regression model: Suppose that $T_1, \ldots, T_n$ are independently drawn from the Weibull(α, λ_i) distribution whose density is $g(t|\alpha, \lambda) = \alpha t^{\alpha-1}\exp\{\lambda - \exp(\lambda)t^\alpha\}$ (Ibrahim et al., 2001, eq. 2.2.1). Let $C_1, \ldots, C_n$ be the corresponding censoring times whose distribution does not include any information about parameters α and λ_i (non-informative censoring) and $\Delta_1, \ldots, \Delta_n$ be the censoring indicator where $\Delta_i = 1$ if $T_i \leq C_i$ (observed) and $\Delta_i = 0$ if $T_i > C_i$ (censored). In this example, let $\boldsymbol{D} = \{Y_i, \Delta_i, X_i, i = 1, \ldots, n\}$, where $Y_i = \min(T_i, C_i)$, and X_i is the covariate for the i^{th} individual. We regress the parameter λ_i on covariates X_i, i.e., $\lambda_i = X_i'\beta$. Assigning a normal prior, Normal(μ_0, Σ_0), for β and a gamma prior, Gamma(α_0, κ_0), for α, the posterior distribution of α and β is

$$\pi(\alpha,\beta|\boldsymbol{D}) \propto \alpha^{\alpha_0+d-1}\exp\left[\sum_{i=1}^{n}\{\Delta_i X_i'\beta+\Delta_i(\alpha-1)\log(Y_i)-Y_i^{\alpha}\exp(X_i'\beta)\}\right.$$
$$\left. -\kappa_0\alpha-\frac{1}{2}(\beta-\mu_0)\Sigma^{-1}(\beta-\mu_0)\right],$$

where $d=\sum_{i=1}^{n}\Delta_i$ (Ibrahim et al., 2001, eq. 2.2.4). Likewise in the logistic regression example, we need not only to draw posterior samples from $\pi(\alpha,\beta|\boldsymbol{D})$ using the MCMC method to evaluate any posterior summary, we also necessitate a numerical procedures for computing frequentist standard errors of those posterior summaries. We use this model to analyze the Melanoma data set in Section 5.3, and compute uncertainty measures using the proposed approach.

These examples show that even for these well researched models, posterior summaries may not have an explicit expression that is easy to compute. Additionally, the computation of the standard error of the posterior summaries requires extra numerical work.

3 Standard errors of posterior means

In this section we concentrate only on the posterior mean and its standard error calculations. In Section 4, we provide recipes for efficiently calculating frequentist standard errors of other types of Bayes estimators. For any generic vector a, we shall use $a^{\otimes 2}$ to denote aa^T.

The frequentist standard error of the posterior mean of θ, $\widehat{\theta}=m^*(\boldsymbol{D})=E(\theta|\boldsymbol{D})=E_{\pi(\cdot|\boldsymbol{D})}(\theta)$, where $\pi(\cdot|\boldsymbol{D})$ is the posterior distribution, is

$$\sqrt{\mathrm{var}_F(\widehat{\theta})}=\sqrt{\int\{m^*(\boldsymbol{D})\}^{\otimes 2}dF(\boldsymbol{D}|\theta)-\left\{\int m^*(\boldsymbol{D})dF(\boldsymbol{D}|\theta)\right\}^{\otimes 2}}.$$

Suppose that one draws B random samples each of size m from the posterior distribution $\pi(\theta|\boldsymbol{D})$. Denote the b^{th} sample as $(\theta_{b1},\ldots,\theta_{bm})$, $b=1,\cdots,B$. Define $\widehat{\theta}_b=\sum_{j=1}^{m}\theta_{bj}/m$, and $\overline{\theta}.=\sum_{b=1}^{B}\widehat{\theta}_b/B$. It is obvious that the variance among $\widehat{\theta}_1,\ldots,\widehat{\theta}_B$ does not estimate $\mathrm{var}_F(\widehat{\theta})$ as $(B-1)^{-1}\sum_{b=1}^{B}(\widehat{\theta}_b-\overline{\theta}.)^2\to(1/m)\mathrm{var}_{\pi(\cdot|\boldsymbol{D})}(\theta)$ almost surely as $B\to\infty$, where $\mathrm{var}_{\pi(\cdot|\boldsymbol{D})}(\theta)$ denotes the posterior variance of θ. One obvious approach to estimate $\text{var}_F(\widehat{\theta})$ is to adopt the bootstrap idea. In the bootstrap world, instead of $\text{var}_F(\widehat{\theta})$ we target estimating $\text{var}_{\widehat{F}}(\widehat{\theta})$, where the observed data are treated as the entire population. In the bootstrap method, we draw B bootstrap samples with replacement from the original data, calculate the posterior mean for each bootstrap sample, and then take the variance of the B posterior means. Let $\boldsymbol{D}^{(b)}$ be the b^{th} bootstrap data, and $\pi(\theta|\boldsymbol{D}^{(b)})$ be the corresponding posterior distribution. Define $\widehat{\theta}^{(b)} = E(\theta|\boldsymbol{D}^{(b)}) = E_{\pi(\cdot|\boldsymbol{D}^{(b)})}(\theta)$ as the posterior mean of θ for the b^{th} bootstrap data. Further define $\overline{\theta}^{(\cdot)} = \sum_{b=1}^{B} \widehat{\theta}^{(b)}/B$. Then

$$(B-1)^{-1}\sum_{b=1}^{B}(\widehat{\theta}^{(b)} - \overline{\theta}^{(\cdot)})^2 \rightarrow \text{var}_{\widehat{F}}(\widehat{\theta}) \text{ as } B \rightarrow \infty.$$

In practice, $\widehat{\theta}^{(b)}$ is estimated by the Monte Carlo estimator $\widehat{\theta}^{(b)}_{\text{mc}} = \sum_{j=1}^{M} \theta_j^{(b)}/M$, where $\theta_1^{(b)}, \ldots, \theta_M^{(b)}$ are M random draws from $\pi(\theta|\boldsymbol{D}^{(b)})$, and $\widehat{\theta}^{(b)}_{\text{mc}} \rightarrow \widehat{\theta}^{(b)}$ almost surely as $M \rightarrow \infty$. Also, define $\overline{\theta}^{(\cdot)}_{\text{mc}} = B^{-1}\sum_{b=1}^{B} \widehat{\theta}^{(b)}_{\text{mc}}$. Then as $M \rightarrow \infty$,

$$(B-1)^{-1}\sum_{b=1}^{B}(\widehat{\theta}^{(b)}_{\text{mc}} - \overline{\theta}^{(\cdot)}_{\text{mc}})^2 \rightarrow (B-1)^{-1}\sum_{b=1}^{B}(\widehat{\theta}^{(b)} - \overline{\theta}^{(\cdot)})^2.$$

Hence $\sum_{b=1}^{B}(\widehat{\theta}^{(b)}_{\text{mc}} - \overline{\theta}^{(\cdot)}_{\text{mc}})^2/(B-1)$ will be used as the estimator of $\text{var}_{\widehat{F}}(\widehat{\theta})$. In the following paragraph we describe how we estimate $\widehat{\theta}^{(1)}, \ldots, \widehat{\theta}^{(B)}$ without having numerically computing B posterior distributions using B MCMC chains thereby saving lots of computation time.

Suppose that using MCMC method we have drawn $\theta_1, \ldots, \theta_M$ from $\pi(\theta|\boldsymbol{D})$, the posterior distribution of θ given the entire data $\boldsymbol{D}$. Suppose that in the b^{th} bootstrap sample, X_i occurs $r_i^{(b)}$ times, where $0 \leq r_i^{(b)} \leq n$, but $\sum_{i=1}^{n} r_i^{(b)} = n$. Then the posterior distribution of θ given the b^{th} bootstrap data $\boldsymbol{D}^{(b)}$ is

$$\pi(\theta|\boldsymbol{D}^{(b)}) = \frac{\prod_{i=1}^{n} f^{r_i^{(b)}}(X_i|\theta)\pi(\theta)}{\int \prod_{i=1}^{n} f^{r_i^{(b)}}(X_i|\theta)\pi(\theta)d\theta},$$

so

$$\widehat{\theta}^{(b)} = \int \theta\pi(\theta|\boldsymbol{D}^{(b)})d\theta = \frac{\int \theta \prod_{i=1}^{n} f^{r_i^{(b)}}(X_i|\theta)\pi(\theta)d\theta}{\int \prod_{i=1}^{n} f^{r_i^{(b)}}(X_i|\theta)\pi(\theta)d\theta} = \frac{G_1^{(b)}}{G_0^{(b)}},$$

where $G_s^{(b)} = \int \theta^s \prod_{i=1}^n f^{r_i^{(b)}}(X_i|\theta)\pi(\theta)d\theta$ for $s = 0$ and 1. Next, we can re-write

$$\begin{aligned} G_s^{(b)} &= \frac{1}{K_\pi}\int \theta^s \frac{\prod_{i=1}^n f^{r_i^{(b)}}(X_i|\theta)\pi(\theta)}{\prod_{i=1}^n f(X_i|\theta)\pi(\theta)} K_\pi \prod_{i=1}^n f(X_i|\theta)\pi(\theta)d\theta \\ &= \frac{1}{K_\pi}\int \theta^s \omega^{(b)}(\theta) K_\pi \prod_{i=1}^n f(X_i|\theta)\pi(\theta)d\theta, \end{aligned}$$

where the importance weight $\omega^{(b)}(\theta) = \prod_{i=1}^n f^{r_i^{(b)}}(X_i|\theta)/\prod_{i=1}^n f(X_i|\theta) = \prod_{i=1}^n f^{(r_i^{(b)}-1)}(X_i|\theta)$. Hence $\widehat{\theta}^{(b)}$ can be estimated by

$$\widehat{\theta}_{\text{is}}^{(b)} = \frac{\sum_{j=1}^M \theta_j \omega^{(b)}(\theta_j)}{\sum_{j=1}^M \omega^{(b)}(\theta_j)}, \tag{1}$$

where $\theta_1, \cdots, \theta_M$ are M MCMC samples drawn from $\pi(\theta|\boldsymbol{D})$, the posterior distribution of θ given the original data $\boldsymbol{D}$. Importantly, under regularity conditions, $\widehat{\theta}_{\text{is}}^{(b)} \to \widehat{\theta}^{(b)}$ almost surely as $M \to \infty$ (see the Appendix). Next define $\overline{\theta}_{\text{is}}^{(\cdot)} = B^{-1}\sum_{b=1}^B \widehat{\theta}_{\text{is}}^{(b)}$. As M gets large,

$$(B-1)^{-1}\sum_{b=1}^B (\widehat{\theta}_{\text{is}}^{(b)} - \overline{\theta}_{\text{is}}^{(\cdot)})^2 \to (B-1)^{-1}\sum_{b=1}^B (\widehat{\theta}^{(b)} - \overline{\theta}^{(\cdot)})^2.$$

Hence we use $\sum_{b=1}^B (\widehat{\theta}_{\text{is}}^{(b)} - \overline{\theta}_{\text{is}}^{(\cdot)})^2/(B-1)$ to estimate $\text{var}_{\widehat{F}}(\widehat{\theta})$. The above procedure can be summarized in the following steps.

Step 1. Draw M MCMC samples from $\pi(\theta|\boldsymbol{D})$, and call them $(\theta_1, \cdots, \theta_M)$.
Step 2. Draw B bootstrap samples with replacement from $\boldsymbol{D}$, and each bootstrap sample consists of n observations. For the b^{th} sample we obtain $(r_1^{(b)}, \cdots, r_n^{(b)})$, with $0 \le r_i^{(b)} \le n$ and $\sum_{i=1}^n r_i^{(b)} = n$, where $r_i^{(b)}$ is the number of times X_i appears in the b^{th} bootstrap sample, $b = 1, \ldots, B$.
Step 3. Compute $\widehat{\theta}_{\text{is}}^{(b)} = \sum_{j=1}^M \theta_j \omega^{(b)}(\theta_j)/\sum_{j=1}^M \omega^{(b)}(\theta_j)$ with $\omega^{(b)}(\theta_j) = \prod_{i=1}^n f^{(r_i^{(b)}-1)}(X_i|\theta_j)$ for $b = 1, \ldots, B$, and $\overline{\theta}_{\text{is}}^{(\cdot)} = \sum_{b=1}^B \widehat{\theta}_{\text{is}}^{(b)}/B$.
Step 4. Compute $(B-1)^{-1}\sum_{b=1}^B (\widehat{\theta}_{\text{is}}^{(b)} - \overline{\theta}_{\text{is}}^{(\cdot)})^2$.

One of the main concerns of importance sampling is the behavior of the importance weights that have influence on the efficiency of the estimator. The following remark gives an intuitive justification that our choice $\pi(\theta|\boldsymbol{D})$ as the trial distribution provides a bounded importance weight with high probability.

Remark 1. Note that $\omega^{(b)}(\theta) = \exp[\sum_{i=1}^{n}(r_i^{(b)} - 1)\log\{f(X_i|\theta)\}] = \exp\{\ell^{(b)}(\theta) - \ell(\theta)\}$, where $\ell^{(b)}(\theta) = \sum_{i=1}^{n} r_i^{(b)}\log f(X_i|\theta) + \log\{\pi(\theta)\}$ and $\ell(\theta) = \sum_{i=1}^{n}\log\{f(X_i|\theta)\} + \log\{\pi(\theta)\}$, and θ is drawn from the posterior distribution $\pi(\theta|\boldsymbol{D})$. Now,

$$\ell^{(b)}(\theta) - \ell(\widetilde{\theta}) \leq \ell^{(b)}(\theta) - \ell(\theta) \leq \ell^{(b)}(\widetilde{\theta}_{(b)}) - \ell(\theta),$$

where $\widetilde{\theta}_{(b)}$ is the posterior mode based on the b^{th} bootstrap data set and $\widetilde{\theta}$ is the posterior mode based on the original data. Then under certain regularity conditions, posterior distribution $\pi(\theta|\boldsymbol{D})$ has the asymptotic normal distribution having mean $\widetilde{\theta}$ and the variance is minus the inverse Hessian of the log posterior evaluated at $\widetilde{\theta}$ for large n (Theorem 3.1 of Carlin and Louis, 2008).

4 Other Bayes estimators

4.1 Posterior quantile

Here we broadly discuss the standard error calculation of posterior quantiles that include the posterior median and credible intervals as special cases. The α^{th} quantile is defined as $q_\alpha(\boldsymbol{D}) = F^{-1}_{\pi(\theta|\boldsymbol{D})}(\alpha)$, where $F_{\pi(\theta|\boldsymbol{D})}(r) = \int_{-\infty}^{r}\pi(\theta|\boldsymbol{D})d\theta$. To estimate the frequentist standard error of $q_\alpha(\boldsymbol{D})$, we may apply the regular bootstrap method by calculating the α^{th} quantile for each of the B posterior distributions, that means one needs to draw posterior samples from $\pi(\theta|\boldsymbol{D}^{(b)})$ using MCMC technique for each $b = 1, \ldots, B$. Instead of doing this for multiple bootstrap data sets, here we can also apply the importance sampling idea. For a trial density $h(\theta)$, we have

$$\begin{aligned} F_{\pi(\theta|\boldsymbol{D}^{(b)})}(r) = \int_{-\infty}^{\infty} I(\theta \leq r)\pi(\theta|\boldsymbol{D}^{(b)})d\theta &= \int_{-\infty}^{\infty} I(\theta \leq r)\frac{\pi(\theta|\boldsymbol{D}^{(b)})}{h(\theta)}h(\theta)d\theta \\ &= \int_{-\infty}^{\infty} I(\theta \leq r)\omega^{(b)}(\theta)h(\theta)d\theta, \end{aligned}$$

where $\omega^{(b)}(\theta) = \pi(\theta|\boldsymbol{D}^{(b)})/h(\theta)$. The distribution function can be estimated by

$$\widehat{F}_{\pi(\theta|\boldsymbol{D}^{(b)})}(r) = \frac{\sum_{j=1}^{M} I(\theta_j \leq r)\omega^{(b)}(\theta_j)}{\sum_{j=1}^{M} \omega^{(b)}(\theta_j)}, \tag{2}$$

where $\theta_1, \ldots, \theta_M$ are drawn from $h(\theta)$. We shall evaluate $\widehat{F}_{\pi(\theta|\boldsymbol{D}^{(b)})}(r)$ for a grid of values of r. Next, the estimated α^{th} quantile is defined as $q_{\alpha,\text{is}}^{(b)} = \inf\{r : \widehat{F}_{\pi(\theta|\boldsymbol{D}^{(b)})}(r) \geq \alpha\}$. Note that we shall use the same set of $\theta_1, \ldots, \theta_M$ drawn from $h(\theta)$, for each bootstrap data set thereby saving considerable computation time.

When α takes a moderate value in the range of 0.2 to 0.8, the importance sampling estimates are reasonable if $\pi(\theta|\boldsymbol{D})$ is used as the trial distribution. For more extreme values of α, (smaller than 0.2 or larger than 0.8), we recommend the following trial distribution for efficient estimation of the α^{th} quantile. To be more specific, without any loss of generality, write $\theta = (\theta_1, \theta_2^T)^T$, and suppose that we are interested in estimating the α^{th} quantile of θ_1 based on the b^{th} bootstrap data. Take $h(\theta) = h_1(\theta_1)h_2(\theta_2)$, where h_1 denotes the uniform density over $[l, u]$ for given values of l and u, and h_2 is taken as the posterior distribution of θ_2 given the data $\boldsymbol{D}$, that means, $h_2(\theta_2) = \int \pi(\theta|\boldsymbol{D})d\theta_1$. Although there is no optimum choice of l or u, based on our computing experiences, we recommend $l = q_{0.5}(\boldsymbol{D}) - 6 \times sd_{\theta_1}(\boldsymbol{D})$ and $u = q_{0.5}(\boldsymbol{D}) + 6 \times sd_{\theta_1}(\boldsymbol{D})$, where $q_\alpha(\boldsymbol{D})$ and $sd_{\theta_1}(\boldsymbol{D})$ denote the α^{th} quantile and the posterior standard deviation of θ_1 given the entire data $\boldsymbol{D}$.

Suppose that $(\theta_{11}, \ldots, \theta_{1M})$ are M random draws from $h_1(\theta_1)$, and $(\theta_{21}, \ldots, \theta_{2M})$ are M random draws from $\pi(\theta_2|\boldsymbol{D})$. The later sample is obtained by simply discarding the first component from each of the M MCMC samples drawn from $\pi(\theta|\boldsymbol{D}) \equiv \pi(\theta_1, \theta_2|\boldsymbol{D})$. Computation of the importance weight $\omega^{(b)}(\theta)$ at $\theta = \theta_j = (\theta_{1j}, \theta_{2j}^T)^T$, for any $j = 1, \ldots, M$, requires $h_2(\theta_{2j}) = \int \pi(\theta_1^*, \theta_{2j}^T|\boldsymbol{D})d\theta_1^* = \kappa^{-1} \int \prod_{i=1}^{n} f(X_i|\theta_1^*, \theta_{2j}^T)\pi(\theta_1^*, \theta_{2j}^T)d\theta_1^*$, where κ is the normalizing constant that does not depend on θ_j. In order to save computation time, instead of targeting to evaluate $h_2(\theta_{2j})$ separately, we consider directly evaluating $\omega^{(b)}(\theta_j)$, and

$$\begin{aligned}\omega^{(b)}(\theta_j) &= \frac{\pi(\theta_{1j},\theta_{2j}^T|\boldsymbol{D}^{(b)})}{h_1(\theta_{1j})\int_{\theta_{1,\min}-\varepsilon}^{\theta_{1,\max}+\varepsilon}\pi(\theta_1^*,\theta_{2j}^T|\boldsymbol{D})d\theta_1^*} = \frac{\kappa_b^{-1}\prod_{i=1}^n f^{r_i^{(b)}}(X_i|\theta_{1j},\theta_{2j}^T)\pi(\theta_{1j},\theta_{2j}^T)}{h_1(\theta_{1j})\kappa^{-1}\int_{\theta_{1,\min}-\varepsilon}^{\theta_{1,\max}+\varepsilon}\prod_{i=1}^n f(X_i|\theta_1^*,\theta_{2j}^T)\pi(\theta_1^*,\theta_{2j}^T)d\theta_1^*} \\ &= \left[h_1(\theta_{1j})\frac{\kappa^{-1}}{\kappa_b^{-1}}\int_{\theta_{1,\min}-\varepsilon}^{\theta_{1,\max}+\varepsilon}\left\{\prod_{i=1}^n\frac{f(X_i|\theta_1^*,\theta_{2j}^T)}{f^{r_i^{(b)}}(X_i|\theta_{1j},\theta_{2j}^T)}\right\}\left\{\frac{\pi(\theta_1^*,\theta_{2j}^T)}{\pi(\theta_{1j},\theta_{2j}^T)}\right\}d\theta_1^*\right]^{-1}, \qquad (3)\end{aligned}$$

where κ_b is the normalizing constant for the b^{th} bootstrap data $\boldsymbol{D}^{(b)}$, and $\theta_{1,\min}$ and $\theta_{1,\max}$ denote the observed minimum and maximum values of θ_1 in the posterior samples drawn from $\pi(\theta_1,\theta_2|\boldsymbol{D})$. To cover the entire domain of θ_1, we extend the range of the integration by adding and subtracting a small number $\varepsilon > 0$. In all our computations, we used $\varepsilon = 0.1 \times \text{IQR}$, where IQR stands for the inter quartile range of the posterior distribution of θ_1 given the original data $\boldsymbol{D}$. Importantly, we do not need to evaluate κ and κ_b for estimating $F_{\pi(\theta|\boldsymbol{D}^{(b)})}(r)$ as they are independent of θ_j, so they get canceled from the normalized weight. Finally, we recommend to use Gauss-Legendre quadrature to determine the above integral in (3). Also to reduce the computational burden, once $\omega^{(b)}(\theta_j)$ is calculated for some b, then we compute $\omega^{(b')}(\theta_j)$ using the following formula $\omega^{(b')}(\theta_j) = \omega^{(b)}(\theta_j)\pi(\theta_{1j},\theta_{2j}^T|\boldsymbol{D}^{(b')})/\pi(\theta_{1j},\theta_{2j}^T|\boldsymbol{D}^{(b)})$, for any $b' \neq b$ as

$$\omega^{(b')}(\theta_j) = \frac{\pi(\theta_{1j},\theta_{2j}^T|\boldsymbol{D}^{(b')})}{h_1(\theta_{1j})\int_{\theta_{1,\min}-\varepsilon}^{\theta_{1,\max}+\varepsilon}\pi(\theta_1^*,\theta_{2j}^T|\boldsymbol{D})d\theta_1^*} = \underbrace{\frac{\pi(\theta_{1j},\theta_{2j}^T|\boldsymbol{D}^{(b)})}{h_1(\theta_{1j})\int_{\theta_{1,\min}-\varepsilon}^{\theta_{1,\max}+\varepsilon}\pi(\theta_1^*,\theta_{2j}^T|\boldsymbol{D})d\theta_1^*}}_{\omega^{(b)}(\theta_j)} \times \frac{\pi(\theta_{1j},\theta_{2j}^T|\boldsymbol{D}^{(b')})}{\pi(\theta_{1j},\theta_{2j}^T|\boldsymbol{D}^{(b)})}. \qquad (4)$$

4.2 Posterior mode

Here we do not apply the importance sampling idea but use another approach for time efficient computation. The posterior mode is defined as $\widehat{\theta}_{\text{mode}} = \arg\max_\theta \pi(\theta|\boldsymbol{D})$. The variance of $\widehat{\theta}_{\text{mode}}$, $\text{var}_F(\widehat{\theta}_{\text{mode}})$ can be estimated by $\sum_{b=1}^B(\widehat{\theta}_{\text{mode}}^{(b)} - \overline{\theta}_{\text{mode}}^{(\cdot)})^2/(B-1)$, where $\widehat{\theta}_{\text{mode}}^{(b)}$ denotes the posterior mode for the b^{th} bootstrap sample, and $\overline{\theta}_{\text{mode}}^{(\cdot)} = \sum_{b=1}^B \widehat{\theta}_{\text{mode}}^{(b)}/B$. Since this standard bootstrap method could be time consuming as it requires to solve a set of gradient equations for each of the B bootstrap data sets, we propose the following alternative approach of estimating that variance.

Under sufficient smoothness conditions, $\widehat{\theta}_{\text{mode}}$ will satisfy $S(\widehat{\theta}_{\text{mode}}|\boldsymbol{D}) = 0$, where $S(\theta|\boldsymbol{D}) = \partial \text{log}\{\pi(\theta|\boldsymbol{D})\}/\partial\theta = \partial\text{log}\{f(\boldsymbol{D}|\theta)\}/\partial\theta + \partial\text{log}\{\pi(\theta)\}/\partial\theta = 0$. Suppose that as $n \to \infty$, $\widehat{\theta}_{\text{mode}} \to \theta_{\text{mode}}$. Then

$$\begin{aligned} 0 = S(\widehat{\theta}_{\text{mode}}|\boldsymbol{D}) &= \frac{\partial}{\partial\theta}\text{log}\{f(\boldsymbol{D}|\widehat{\theta}_{\text{mode}})\} + \frac{\partial}{\partial\theta}\text{log}\{\pi(\widehat{\theta}_{\text{mode}})\} \\ &\approx [\frac{\partial}{\partial\theta}\text{log}\{f(\boldsymbol{D}|\theta_{\text{mode}})\} + \frac{\partial}{\partial\theta}\text{log}\{\pi(\theta_{\text{mode}})\}] + \\ &\quad [\frac{\partial^2}{\partial\theta^2}\text{log}\{f(\boldsymbol{D}|\theta_{\text{mode}})\} + \frac{\partial^2}{\partial\theta^2}\text{log}\{\pi(\theta_{\text{mode}})\}](\widehat{\theta}_{\text{mode}} - \theta_{\text{mode}}). \end{aligned}$$

Thus, with $A = E[\partial^2\text{log}\{f(\boldsymbol{D}|\theta_{\text{mode}})\}/\partial\theta^2 + \partial^2\text{log}\{\pi(\theta_{\text{mode}})\}/\partial\theta^2]$, we have $(\widehat{\theta}_{\text{mode}} - \theta_{\text{mode}}) \approx A^{-1}[\partial\text{log}\{f(\boldsymbol{D}|\theta_{\text{mode}})\}/\partial\theta + \partial\text{log}\{\pi(\theta_{\text{mode}})\}/\partial\theta]$, and consequently the variance can be obtained by the sandwich formula,

$$\text{var}_F(\widehat{\theta}_{\text{mode}}) = A^{-1}\text{var}[\frac{\partial}{\partial\theta}\text{log}\{f(\boldsymbol{D}|\theta_{\text{mode}})\} + \frac{\partial}{\partial\theta}\text{log}\{\pi(\theta_{\text{mode}})\}]A^{-T}.$$

Here A can be estimated by $\widehat{A} = \partial^2\text{log}\{f(\boldsymbol{D}|\widehat{\theta}_{\text{mode}})\}/\partial\theta^2 + \partial^2\text{log}\{\pi(\widehat{\theta}_{\text{mode}})\}/\partial\theta^2$. The middle term of the variance formula is $\text{var}[\partial\text{log}\{f(\boldsymbol{D}|\theta_{\text{mode}})\}/\partial\theta]$ that can be estimated by

$$\widehat{\text{var}}[\frac{\partial}{\partial\theta}\text{log}\{f(\boldsymbol{D}|\theta_{\text{mode}})\}] = (B-1)^{-1}\sum_{b=1}^{B}\left[\frac{\partial}{\partial\theta}\text{log}\{f(\boldsymbol{D}^{(b)}|\widehat{\theta}_{\text{mode}})\} - \frac{1}{B}\sum_{b'=1}^{B}\frac{\partial}{\partial\theta}\text{log}\{f(\boldsymbol{D}^{(b')}|\widehat{\theta}_{\text{mode}})\}\right]^2,$$

and in particular, for fast computation we use $\partial\text{log}\{f(\boldsymbol{D}^{(b)}|\widehat{\theta}_{\text{mode}})\}/\partial\theta = \sum_{i=1}^{n} r_i^{(b)}\partial\text{log}\{f(X_i |\widehat{\theta}_{\text{mode}})\}/\partial\theta$. Finally, $\text{var}_F(\widehat{\theta}_{\text{mode}})$ is estimated by $\widehat{A}^{-1}\widehat{\text{var}}[\partial\text{log}\{f(\boldsymbol{D}|\theta_{\text{mode}})\}/\partial\theta]\widehat{A}^{-T}$.

5 Numerical studies

In order to assess and compare the performances of the methods, we conducted simulation studies and real data analysis for the motivating examples described in Section 2. Specifically, we provide simulation results for the logistic regression model. Next, the linear measurement error model is illustrated using a simulated data set. Third, we present an analysis of real data set using the Weibull regression model. Finally, following a refree's request, we consider an application of the proposed method to a vector autoregressive (VAR) model.

5.1 Logistic regression model

We generated 500 data sets, and each simulated data set consists of $n = 500$ observations, denoted by $\{(X_i, Y_i), i = 1, \ldots, n\}$. We drew X from Normal$(0, 1)$ distribution and the response variable Y was simulated from a Bernoulli distribution with the success probability $\text{pr}(Y = 1|X) = \exp(\alpha + \beta X)/\{1 + \exp(\alpha + \beta X)\}$. The true values of α and β were -2.5 and 1, respectively. That makes the proportion of success around 10%. For the Bayesian inference of the parameters α and β we used the same Normal$(0, 2)$ priors for both of them. Then for the MCMC computation, we used $15,000$ iterations with the first $5,000$ samples were used as burn-in samples.

For each data set, we estimated the posterior mean of α and β. We also calculated standard errors of the posterior means for each data set. Let $\widehat{\alpha}_j$ and $\widehat{\beta}_j$ be the posterior mean based on the j^{th} data set, for $j = 1, \ldots, 500$. For each data set, we computed the frequentist standard error of the estimator based on 1) the regular bootstrap method and 2) the proposed importance sampling based approach. For the j^{th} data set, we drew $B = 500$ bootstrap samples with replacement. Suppose that $(\widehat{\alpha}^{(b)}_{mcmc,j}, \widehat{\beta}^{(b)}_{mcmc,j})$ denotes the posterior means for the b^{th} bootstrap data, for $b = 1, \ldots, 500$, and these posterior means were calculated by applying the MCMC method to each bootstrap data separately. The regular bootstrap standard error for $\widehat{\alpha}_j$ and $\widehat{\beta}_j$ are now expressed as $sd_{1,j}(\alpha) = \sqrt{(1/499)\sum_{b=1}^{500}(\widehat{\alpha}^{(b)}_{mcmc,j} - \sum_{b'=1}^{500}\widehat{\alpha}^{(b')}_{mcmc,j}/500)^2}$ and $sd_{1,j}(\beta) = \sqrt{(1/499)\sum_{b=1}^{500}(\widehat{\beta}^{(b)}_{mcmc,j} - \sum_{b'=1}^{500}\widehat{\beta}^{(b')}_{mcmc,j}/500)^2}$, respectively. Next, we computed the proposed importance sampling based standard error, $sd_{2,j}(\alpha) = \sqrt{(1/499)\sum_{b=1}^{500}(\widehat{\alpha}^{(b)}_{is,j} - \sum_{b'=1}^{500}\widehat{\alpha}^{(b')}_{is,j}/500)^2}$ and $sd_{2,j}(\beta) = \sqrt{(1/499)\sum_{b=1}^{500}(\widehat{\beta}^{(b)}_{is,j} - \sum_{b'=1}^{500}\widehat{\beta}^{(b')}_{is,j}/500)^2}$, where $(\widehat{\alpha}^{(b)}_{is,j}, \widehat{\beta}^{(b)}_{is,j})$ denotes the posterior means for the b^{th} bootstrap data based on the importance sampling idea. Our goal is to illustrate that instead of using the regular bootstrap idea that is way more time consuming, one can simply use the importance sampling based method to estimate the frequentist standard error of the Bayes estimators. We wanted to show that proposed method is computationally far more time efficient, and on the other hand, the standard error calculated using the proposed method is close to the standard error calculated based on the regular bootstrap method. We, once again, point out that the regular bootstrap approach requires enumeration of B MCMC chains, one for each of the B bootstrap data sets, while the proposed approach requires enumeration of only one MCMC chain. In the appendix, we compare the computational complexity of the two approaches.

Figure 1 shows a scatter plot of two standard errors (sd_1 and sd_2) for 500 data sets for the intercept and slope parameter. The figure reveals that the two estimates of the standard error are in good agreement as the points are well dispersed around the 45 degree line. Table 1 shows the computation time (in sec) for the two methods, and clearly the proposed importance sampling based approach is computationally far more superior than the regular bootstrap method.

Since the logistic regression belongs to the class of the generalized linear models, we are able to apply Efron (2015)'s method to evaluate the standard deviation of the posterior mean for the intercept and slope parameters. Let $\theta = (\alpha, \beta)^T$. From Equation (3.1) of Efron (2015), $f_\theta(T) = \exp[\theta^T T - \sum_{j=1}^n \log\{1 + \exp(\alpha + \beta X_j)\}]$, where $T = (\sum_{j=1}^n Y_j, \sum_{j=1}^n X_j Y_j)^T$ is the sufficient statistic for θ. Then, $E(T) = (\sum_{j=1}^n p_j, \sum_{j=1}^n X_j p_j)^T$ and $\text{var}(T) = V_\theta = \sum_{j=1}^n p_j(1-p_j)(1, X_j)^T(1, X_j)$, where $p_j = \text{P}(Y = 1|X_j) = \exp(\alpha+\beta X_j)/\{1+\exp(\alpha+\beta X_j)\}$, the success probability given $X = X_j$. Due to the numerical instability of the "conversion factors" we are not able to apply his method that completely avoids MCMC sampling, and this issue has been acknowledged in Efron (2015). However, we apply his general approach for calculating the standard deviation of the posterior mean that is summarized in the following steps.

Step 1. Draw M MCMC samples $(\theta_1, \ldots, \theta_M)$ from $\pi(\theta|\boldsymbol{D})$.

Step 2. Estimate $\text{cov}(\theta, \theta|T)$ by $\widehat{\text{cov}} = \sum_{j=1}^M (\theta_j - \bar{\theta})(\theta_j - \bar{\theta})^T/M$, where $\bar{\theta} = \sum_{j=1}^M \theta_j/M$. Then we obtain $sd_3 = [\widehat{\text{cov}}^T V_{\widehat{\theta}} \widehat{\text{cov}}]^{(1/2)}$.

Now we compare sd_3 with the gold standard approach, sd_1, in Figure 2. In terms of computation time, Efron's approach is much much faster than any other procedure (Table 1). However, Efron's approach is applicable when V_θ is easily available, and his method can compute standard error for posterior mean only, not for any quantiles.

Next we calculated the standard error of the first quartile, third quartile, the 2.5^{th} percentile, and the 97.5^{th} percentile of the posterior distribution of α and β based on 1) the regular bootstrap method and 2) the importance sampling based method. We particularly considered 2.5^{th} and 97.5^{th} percentiles as they are often used for constructing credible intervals. Figures 3 and 4 show the standard errors computed using the two approaches for each of these summary statistics for the simulated data sets. We want to point out that for the 2.5^{th} and 97.5^{th} percentiles we used the trial distribution that is described in Section 4.1 and it involves with a slightly more computation than the scenario where $\pi(\alpha, \beta|\boldsymbol{D})$ is used as a trial distribution (see Table 1). However, despite of being more computationally involved, overall this approach is more time efficient (see Table 1) than the regular bootstrap method where one needs to run MCMC method on each bootstrap data set separately. We also need to keep in mind that this time comparison is heavily depended on the number of MCMC iterations used in the computation, and the time gain will be more if more MCMC iterations are used for the posterior inference. For a fair comparison, every core computation was conducted using FORTRAN 90 within an R script. That is, generation of random samples from the posterior distribution $\pi(\theta|\boldsymbol{D})$ and evaluation of the importance weight $\omega^{(b)}(\theta)$ in Sections 3, 4.1 were programmed in FORTRAN. Although there are a number of presumably optimized programs or R packages for Bayesian computing, we decide to write our own code for fair comparison across the methods.

The computational complexity of the proposed method and the regular bootstrap method using MCMC simulations are of the same order, and according to the Bachman-Landau notation it is $O(BMn)$, where B, M, n denote the number of bootstrap samples, the number of MCMC iterations, and the sample size, respectively. In Appendix B, we have explained the computational complexity for this example through algorithms, and similar algorithms can be written for other examples. Although the computational complexity of the regular bootstrap method and the proposed method are of the same order, by avoiding MCMC simulations the computation of posterior summary is much faster in the latter method than the former approach.

5.2 Linear measurement error model

Next, we revisit the linear measurement error model. We first note that the joint distribution of the observed Y and W, $f_{Y,W}(y,w)$ is an exponential family. Since $f_{Y,W}(y,w) = \int f(w,x,y)dx$, where $f(y,w,x)$ is the joint density of W, X, Y,

$$\begin{aligned} f_{Y,W}(Y,W) &= h(Y,W)c(\theta)\exp\Big[-\frac{1}{2}\{\frac{1}{\sigma_\epsilon^2} - \frac{\beta^2/\sigma_\epsilon^4}{\beta^2/\sigma_\epsilon^2 + 1/\sigma_u^2 + 1/\sigma_x^2}\}Y^2 \\ &\quad -\{\frac{\alpha\beta^2/\sigma_\epsilon^4 + \beta\mu_x/(\sigma_\epsilon^2\sigma_x^2)}{\beta^2/\sigma_\epsilon^2 + 1/\sigma_u^2 + 1/\sigma_x^2} - \frac{\alpha}{\sigma_\epsilon^2}\}Y + \frac{1/\sigma_u^4}{2(\beta^2/\sigma_\epsilon^2 + 1/\sigma_u^2 + 1/\sigma_x^2)}W^2 \\ &\quad -\frac{\alpha\beta/(\sigma_\epsilon^2\sigma_u^2) - \mu_x/(\sigma_u^2\sigma_x^2)}{\beta^2/\sigma_\epsilon^2 + 1/\sigma_u^2 + 1/\sigma_x^2}W + \frac{\beta/(\sigma_\epsilon^2\sigma_u^2)}{\beta^2/\sigma_\epsilon^2 + 1/\sigma_u^2 + 1/\sigma_x^2}YW\Big], \end{aligned}$$

where $h(Y,W) = \exp(-W^2/2\sigma_u^2)$ does not depend on θ since σ_u^2 is known and $c(\theta) = (2\pi)^{-1}\{\sigma_\epsilon^2\sigma_u^2\sigma_x^2(\beta^2/\sigma_\epsilon^2 + 1/\sigma_u^2 + 1/\sigma_x^2)\}^{-1/2}\exp\{-\alpha^2/2\sigma_\epsilon^2 - \mu_x^2/2\sigma_x^2 + (\alpha^2\beta^2/2\sigma_\epsilon^4 + \mu_x^2/2\sigma_x^4 - \alpha\beta\mu_x/2\sigma_x^2\sigma_\epsilon^2)/(\beta^2/\sigma_\epsilon^2 + 1/\sigma_u^2 + 1/\sigma_x^2)\}$ is a function of θ. Therefore, $T = (Y^2, Y, W^2, W, YW)$ is a sufficient statistic for the natural parameter $\eta = (\eta_1, \ldots, \eta_5)$, where $\eta_1 = 1/\sigma_\epsilon^2 - (\beta^2/\sigma_\epsilon^4)/(\beta^2/\sigma_\epsilon^2 + 1/\sigma_u^2 + 1/\sigma_x^2)$, $\eta_2 = \{\alpha\beta^2/\sigma_\epsilon^4 + \beta\mu_x/(\sigma_\epsilon^2\sigma_x^2)\}/(\beta^2/\sigma_\epsilon^2 + 1/\sigma_u^2 + 1/\sigma_x^2) - \alpha/\sigma_\epsilon^2$, $\eta_3 = 1/\sigma_u^2 - (1/\sigma_u^4)/(\beta^2/\sigma_\epsilon^2 + 1/\sigma_u^2 + 1/\sigma_x^2)$ $\eta_4 = \{\alpha\beta/(\sigma_\epsilon^2\sigma_u^2) - \mu_x/(\sigma_u^2\sigma_x^2)\}/(\beta^2/\sigma_\epsilon^2 + 1/\sigma_u^2 + 1/\sigma_x^2)$, and $\eta_5 = \{\beta/(\sigma_\epsilon^2\sigma_u^2)\}/(\beta^2/\sigma_\epsilon^2 + 1/\sigma_u^2 + 1/\sigma_x^2)$. In order to apply Efron (2015)'s method, we need to find the variance covariance matrix V_η of T, which is a very difficult if not impossible task. Therefore, we applied our approach to compute the frequentist standard error for the posterior summaries of α and β.

We generated a single data set comprising of $\boldsymbol{D} = \{(Y_i, W_i, X_i), i = 1, \ldots, n = 1{,}000\}$ under the true model $Y_i = \alpha + \beta X_i + \epsilon_i$, $\alpha = 0.23$, $\beta = 0.47$, and $W_i = X_i + U_i$, where $\epsilon_i \sim \text{Normal}[0, (\sqrt{0.5})^2]$, $U_i \sim \text{Normal}[0, (\sqrt{0.5})^2]$ and $X_i \sim \text{Normal}(0.5, 1)$. We analyzed the data according to the method described in Sections 3 and 4, without using X in the analysis. We applied Gibbs sampling to draw samples from the posterior distribution of the parameters, and used $M = 10,000$ iterations after the first $5,000$ samples as burn-in samples. For the prior distributions, we set $\sigma_\alpha^2 = \sigma_\beta^2 = \sigma_\mu^2 = 10,000$ and $\delta_x = \delta_\epsilon = \lambda_x = \lambda_\epsilon = 1$. Then we drew $B = 500$ bootstrap samples with replacement and we evaluated sd_1 and sd_2 as described in Section 5.1. Table 2 shows the frequentist standard errors corresponding to the posterior summaries of α and β, along with the computation time. The results show the advantages of the proposed method over the regular bootstrap method in terms of computational time.

5.3 Weibull regression model

We now analyze a subset of the E1684 melanoma clinical trial data (Example 1.2 and 2.2 of Ibrahim et al., 2001) to determine the frequentist standard errors of posterior summaries from the Weibull model. This was a phase III clinical trial conducted by Eastern Cooperative Oncology Group (ECOG) with chemotherapy of interferon alpha-2b in melanoma patient and can be found at "`http://merlot.stat.uconn.edu/~mhchen/survbook/`". The data set contains observed time measured in year, (right) censoring indicator and chemotherapy treatment indicator for each of 255 patients. The purpose of this clinical study was to examine the treatment effect on the survival times (Y). Among the possible models for this objective, we fit a Weibull regression model on the survival times (Y) using chemotherapy as a covariate (X) according to Example 2.2 in Ibrahim et al. (2001). Following Ibrahim et al. (2001), we used a Gamma(1, 0.001) prior for α and a Normal($(0, 0)^T$, $10^4 I_2$) prior for β, where I_2 denotes the 2×2 identity matrix, for the Weibull regression model described in Section 2. Here we also generated $B = 500$ bootstrap data sets to calculate standard errors for the posterior summaries of parameters.

Table 3 shows the posterior estimates of β_0, β_1 and α, corresponding frequentist standard errors, and computing times. Instead of presenting only posterior means as done in Table 2.2

of Ibrahim et al. (2001), we extend that table to include other posterior summaries and the frequentist uncertainty of the estimates. Moreover, following the method described in Section 5, we are able to calculate the standard errors more time efficiently.

Furthermore, it is worth to note that it is difficult to apply Efron (2015)'s approach for calculating frequentist standard deviation of posterior mean to the Weibull model because it is not an exponential family of distributions. Secondly, the joint density of the above model is $f(\boldsymbol{D}|\alpha, \beta_0, \beta_1) = \exp[\sum_{i=1}^{n}\{\Delta_i \log\alpha + \Delta_i(\beta_0 + X_i\beta_1) + \Delta_i(\alpha-1)\log(Y_i) - Y_i^{\alpha}\exp(\beta_0 + X_i\beta_1)\}]$ so that it is also hard to calculate V_θ the variance of the sufficient statistic, where $\theta = (\alpha, \beta_0, \beta_1)$. Hence, we are not able to apply his method in this context.

5.4 Vector autoregressive model (VAR)

In the previous examples, we discussed the frequentist standard errors of posterior summaries for parameters themselves. We now discuss a more complicated case where the main interest is a function of parameters. Suppose that we have a p-dimensional time series data $\boldsymbol{y}_s$, $s = 1, \ldots, S$, and assume that the data follows a vector autoregression (VAR) model. The VAR model with lag L is $\boldsymbol{y}'_s = \boldsymbol{\mu} + \sum_{j=1}^{L} \boldsymbol{y}'_{s-j}\boldsymbol{B}_j + \boldsymbol{\epsilon}'_s$, where $\boldsymbol{\mu}$ is an $1 \times p$ vector, $\boldsymbol{B}_j$ is a $p \times p$ coefficient matrix, $\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_S$ are iid $N(0, \boldsymbol{\Sigma})$, and the covariance $\boldsymbol{\Sigma}$ is an unknown $p \times p$ positive definite matrix. Instead of focusing our attention on the elements of parameter matrices $\boldsymbol{B} = (\boldsymbol{B}'_1, \ldots, \boldsymbol{B}'_L)'$ and $\boldsymbol{\Sigma}$, it is more of interest to estimate the impact of changing an element of $\boldsymbol{y}_s$ on the future value $\boldsymbol{y}_{s+k}$. These effects are called impulse responses (Stock and Watson, 2001), and they are defined as nonlinear functions of the parameter matrices $\boldsymbol{B}$ and $\boldsymbol{\Sigma}$.

The likelihood function of $(\boldsymbol{\mu}, \boldsymbol{B}, \boldsymbol{\Sigma})$ is $L(\boldsymbol{\Phi}, \boldsymbol{\Sigma}) = (2\pi)^{-Sp/2}|\boldsymbol{\Sigma}|^{-S/2}\exp[-tr\{(\boldsymbol{Y} - \boldsymbol{X\Phi})\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y} - \boldsymbol{X\Phi})'\}/2]$, where $\boldsymbol{Y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_S)'$, $\boldsymbol{\Phi} = (\boldsymbol{\mu}', \boldsymbol{B}')'$, $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_S)'$, and $\boldsymbol{x}_s = (1, \boldsymbol{y}'_{s-1}, \ldots, \boldsymbol{y}'_{s-L})'$. Note that $\boldsymbol{Y}$ is $S \times p$ matrix, $\boldsymbol{X}$ is $S \times (Lp+1)$ matrix, and $\boldsymbol{\Phi}$ is $(Lp+1) \times p$ matrix. Here we consider the impulse response to orthogonalized errors $U = \boldsymbol{\epsilon}'\boldsymbol{\Psi}^{-1}$, where $\boldsymbol{\Psi}$ is the Cholesky matrix for $\boldsymbol{\Sigma}$, i.e., $\boldsymbol{\Sigma} = \boldsymbol{\Psi}'\boldsymbol{\Psi}$. That is, the impulse responses $\boldsymbol{Z}_k$ of $\boldsymbol{y}_{s+k}$ based on the structural shock $\boldsymbol{\epsilon}_s'\boldsymbol{\Psi}^{-1}$ is $\boldsymbol{Z}_k = \boldsymbol{\Psi}\boldsymbol{H}_k$, where $\boldsymbol{H}_j = \sum_{i=1}^{j} \boldsymbol{B}_j \boldsymbol{H}_{j-i}$, and $\boldsymbol{B}_i = 0$ for i larger than lag L and $\boldsymbol{B}_0 = I$ (Sims, 1980; Ni et al., 2007).

For the computational purpose, we consider conjugate priors for $(\boldsymbol{\Phi}, \boldsymbol{\Sigma})$. That is, $\boldsymbol{\pi}(\boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-(p+1)/2}$, the Jeffreys prior, and $\boldsymbol{\pi}(\boldsymbol{\phi}) \propto |M_0|^{-1/2} \exp\{-(\boldsymbol{\phi} - \boldsymbol{\phi}_0) M_0^{-1} (\boldsymbol{\phi} - \boldsymbol{\phi}_0)'/2\}$, where $\boldsymbol{\phi} = \mathrm{vec}(\boldsymbol{\Phi})$. Next, following Ni et al. (2007), the conditional density of $\boldsymbol{\phi}$ given $\boldsymbol{\Sigma}, \boldsymbol{D}$ is $N(\boldsymbol{m}, \boldsymbol{V})$ and the conditional density of $\boldsymbol{\Sigma}$ given $\boldsymbol{\Phi}, \boldsymbol{D}$ is inverse Wishart $(\boldsymbol{S}(\boldsymbol{\Phi}), M)$, where $\boldsymbol{m} = \widehat{\boldsymbol{\phi}}_{mle} + \{M_0^{-1} + \boldsymbol{\Sigma}^{-1} \otimes (\boldsymbol{X}'\boldsymbol{X})\}^{-1} M_0^{-1} (\boldsymbol{\phi}_0 - \widehat{\boldsymbol{\phi}}_{mle})$, $\boldsymbol{V} = \{M_0^{-1} + \boldsymbol{\Sigma}^{-1} \otimes (\boldsymbol{X}'\boldsymbol{X})\}^{-1}$, $\widehat{\boldsymbol{\phi}}_{mle} = \mathrm{vec}(\widehat{\boldsymbol{\Phi}}_{mle})$, $\widehat{\boldsymbol{\Phi}}_{mle} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$, $\boldsymbol{S}(\boldsymbol{\Phi}) = (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\Phi})'(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\Phi})$, and $\boldsymbol{D} = \{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_S\}$ is the observed data. Since the impulse response is a function of $\boldsymbol{B}$ and $\boldsymbol{\Sigma}$, we rewrite $\boldsymbol{Z}_k = \boldsymbol{Z}(\boldsymbol{B}, \boldsymbol{\Sigma}, k)$. We take the posterior mean as a Bayes estimator of the impulse response, and it is $(\widehat{\boldsymbol{Z}}_k)_{(i,j)} = \int \{\boldsymbol{Z}(\theta, k)\}_{(i,j)} \pi(\theta|\boldsymbol{D}) d\theta$, where $\theta = (\boldsymbol{B}, \boldsymbol{\Sigma})$, and $(\boldsymbol{Z}_s)_{(i,j)}$ is the (i,j) element of $\boldsymbol{Z}_s$. Now, we apply the proposed method to calculate the frequentist standard error of $\widehat{\boldsymbol{Z}}_k$. In this complex example, it is nearly impossible to find the variance-covariance matrix of the sufficient statistics of θ, therefore it is not possible to apply Efron's approach.

For illustration purpose, we generated a data set from the following VAR(1) model with $p = 2$,

$$\boldsymbol{y}_s' = \begin{bmatrix} -0.7 & 1.3 \end{bmatrix} + \begin{bmatrix} 0.7 & 0.3 \\ 0.2 & 0.6 \end{bmatrix} \boldsymbol{y}_{s-1}' + \boldsymbol{\epsilon}_s,\ \boldsymbol{\epsilon}_s \overset{iid}{\sim} N(0, \boldsymbol{\Sigma}),\ \boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix},\ s = 1, \ldots, S,$$

with $S = 1,000$. Since the time series data is no more independent, we used the moving block bootstrap (MBB), where we divided the series into N overlapping blocks of length ℓ to preserve the dependence structure of the original dataset (Kreiss and Lahiri, 2012). Then we chose b blocks out of N blocks to make the bootstrap observations $\boldsymbol{y}_1^*, \ldots, \boldsymbol{y}_S^*$.

We fit a VAR(2) model to the simulated dataset. As in previous examples, we used $M = 10000$ iterations after the burn-in samples. We imposed noninformative priors for $\boldsymbol{\Phi}$, where $\boldsymbol{\phi}_0 = \boldsymbol{0}$

and $M_0 = 20\boldsymbol{I}$. Then we drew $B = 500$ MBB samples with 15% of the total dataset as a block length (ℓ).

Figure 5 show the point estimate (posterior mean) and the 95% confidence band based on the frequentist standard error of the posterior mean for the impulse responses of $\boldsymbol{y}_2$ to $\boldsymbol{y}_1$ and $\boldsymbol{y}_1$ to $\boldsymbol{y}_2$, respectively. The confidence bands based on sd_1 and sd_2 are similar, but computationally the second approach (sd_2) was about 5.6 times faster than the first approach (sd_1). In Table 4, we also report the numerical values of the standard errors at each time lag, and the results do not show any appreciable difference between sd_1 and sd_2.

6 Conclusions

In this paper we have discussed numerical approaches for efficient computation of standard errors for posterior summaries. The main theme of the paper is to use bootstrap samples but avoid using full blown MCMC based inference for each of the bootstrap data. The methods rely on the importance sampling idea, and are broadly applicable. The R code for our computation is available at `https://stat.tamu.edu/~sinha/research.html`.

It is well-known that the presence of outliers results in a poor performance in a bootstrap approach because they are more frequent in bootstrap samples than the original dataset if we consider the classical nonparametric bootstrap (Salibian-Barrera and Zamar, 2002; Willems and Van Aelst, 2005; Huber and Ronchetti, 2009). Therefore, the performance of our proposed method can be affected by the outliers in the data as we have used the classical nonparametric bootstrap with replacement. However, this may be overcome by considering robust bootstrap methods for drawing samples (Singh, 1998; Hu and Hu, 2000; Salibian-Barrera and Zamar, 2002), or a combination of a robust bootstrap method and a robust Bayesian method, possibly with a flat-tailed prior (Berger et al., 1994; Marín, 2000).

Acknowledgments

The authors would like to thank two referees for their thoughtful comments that lead to a much improved manuscript. Carroll's research was supported by a grant from the National Cancer Institute (U01-CA057030). Lee and Sinha's research was supported by NIH grant R03CA176760.

References

Berger, J. (2006). The case for objective bayesian analysis. *Bayesian analysis*, 1, 385–402.

Berger, J. O., Moreno, E., Pericchi, L. R., Bayarri, M. J., Bernardo, J. M., Cano, J. A., De la Horra, J., Martín, J., Ríos-Insúa, D., Betrò, B., et al. (1994). An overview of robust bayesian analysis. *Test*, 3, 5–124.

Carlin, B. P. and Louis, T. A. (2008). *Bayesian Methods for Data Analysis.* CRC Press.

Carroll, R., Ruppert, D., Stefanski, L., and Crainiceanu, C. (2006). *Measurement Error in Nonlinear Models: A Modern Perspective, Second Edition.* CRC Press.

Efron, B. (2012). Bayesian inference and the parametric bootstrap. *Annals of Applied Statistics*, 6, 1971–1997.

Efron, B. (2015). Frequentist accuracy of Bayesian estimates. *Journal of the Royal Statistical Society, Series B*, 77, 617–646.

Efron, B. and Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science*, 1, 54–75.

Efron, B. and Tibshirani, R. (1994). *An introduction to the bootstrap.* CRC press.

Fuller, W. (1987). *Measurement Error Models.* John Wiley & Sons, New York.

Hu, F. and Hu, J. (2000). A note on breakdown theory for bootstrap methods. *Statistics & probability letters*, 50, 49–53.

Huber, P. and Ronchetti, E. (2009). *Robust Statistics.* John Wiley & Sons, New York.

Ibrahim, J., Chen, M., and Sinha, D. (2001). *Bayesian Survival Analysis.* Springer.

Jones, G. L. (2004). On the markov chain central limit theorem. *Probability Surveys*, 1, 299–320.

Kreiss, J.-P. and Lahiri, S. (2012). Bootstrap methods for time series. *Handbook of Statistics: Time Series Analysis: Methods and Applications*, 30.

Liang, F. (2002). Dynamically weighted importance sampling in Monte Carlo computation. *Journal of the American Statistical Association*, 97, 807–821.

Marín, J. M. (2000). A robust version of the dynamic linear model with an economic application. In *Robust Bayesian Analysis*, pages 373–383. Springer.

Ni, S., Sun, D., and Sun, X. (2007). Intrinsic bayesian estimation of vector autoregression impulse responses. *Journal of Business & Economic Statistics*, 25, 163–176.

Robert, C. and Casella, G. (2005). *Monte Carlo Statistical Methods*. Springer New York.

Salibian-Barrera, M. and Zamar, R. H. (2002). Bootstrapping robust estimates of regression. *The Annals of Statistics*, 30, 556–582.

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica*, 48, 1–48.

Singh, K. (1998). Breakdown theory for bootstrap quantiles. *The annals of Statistics*, 26, 1719–1732.

Stock, J. H. and Watson, M. W. (2001). Vector autoregressions. *The Journal of Economic Perspectives*, 15, 101–115.

Willems, G. and Van Aelst, S. (2005). Fast and robust bootstrap for LTS. *Computational Statistics & Data Analysis*, 48, 703–715.

Appendix

A Proof of the convergence result of Section 3

Here we discuss the convergence of (1). Suppose that $\omega^{(b)}(\theta)$ and $\theta\omega^{(b)}(\theta)$ are integrable functions of θ with respect to the posterior distribution of the original data $\pi(\theta|\boldsymbol{D})$ so that $G_s^{(b)} = \int \theta^s \omega^{(b)}(\theta)\pi(\theta|\boldsymbol{D})d\theta/K_\pi = E_{\pi(\cdot|\boldsymbol{D})}\{\theta^s\omega^{(b)}(\theta)\}/K_\pi$ is finite for all b and $s = 0, 1$. Therefore, as $M \to \infty$, from the ergodic theorem (Jones, 2004; Robert and Casella, 2005), with probability 1,

$$\begin{aligned} \frac{1}{M}\sum_{j=1}^{M} \omega^{(b)}(\theta_j) \quad &\to \quad E_{\pi(\cdot|\boldsymbol{D})}\{\omega^{(b)}(\theta)\} = K_\pi G_0^{(b)}, \\ \frac{1}{M}\sum_{j=1}^{M} \theta_j\omega^{(b)}(\theta_j) \quad &\to \quad E_{\pi(\cdot|\boldsymbol{D})}\{\theta\omega^{(b)}(\theta)\} = K_\pi G_1^{(b)}. \end{aligned}$$

From Remark 1 in Section 3, $\omega^{(b)}(\theta) = \exp\{\ell^{(b)}(\theta) - \ell(\theta)\}$ implies $\omega^{(b)}(\theta)$ is positive for all θ. Therefore, $\sum_{j=1}^{M} \omega^{(b)}(\theta_j) > 0$ and $G_0^{(b)} > 0$, and consequently

$$\widehat{\theta}_{\text{is}}^{(b)} = \frac{\sum_{j=1}^{M} \theta_j \omega^{(b)}(\theta_j)}{\sum_{j=1}^{M} \omega^{(b)}(\theta_j)} \to \frac{G_1^{(b)}}{G_0^{(b)}} = \widehat{\theta}^{(b)}$$

with probability 1 as $M \to \infty$.

B Computational complexity of the two approaches for the logistic regression example

Algorithm 1 Full Bootstrap method for the logistic regression model in Section 5.1

for $b = 1$ to B **do**
  Draw a bootstrap sample
  Initialize $\alpha_0^{(b)}$ and $\beta_0^{(b)}$
  for $m = 1$ to $M + burn$ **do**
    Propose $\alpha^{cand}, \beta^{cand} \sim q(\alpha, \beta | \alpha_{m-1}^{(b)}, \beta_{m-1}^{(b)})$
    Calculate
    $r = \min\{1, \frac{\pi(\alpha^{cand}, \beta^{cand} | \boldsymbol{D}^{(b)}) q(\alpha_{m-1}^{(b)}, \beta_{m-1}^{(b)} | \alpha^{cand}, \beta^{cand})}{\pi(\alpha_{m-1}^{(b)}, \beta_{m-1}^{(b)} | \boldsymbol{D}^{(b)}) q(\alpha^{cand}, \beta^{cand} | \alpha_{m-1}^{(b)}, \beta_{m-1}^{(b)})}\}$
    Generate $u \sim U(0, 1)$
    if $u < r$ **then**
      $\alpha_m^{(b)} = \alpha^{cand}$
      $\beta_m^{(b)} = \beta^{cand}$
    else
      $\alpha_m^{(b)} = \alpha_{m-1}^{(b)}$
      $\beta_m^{(b)} = \beta_{m-1}^{(b)}$
    end if
  end for
  Find averages $\widehat{\alpha}^{(b)}, \widehat{\beta}^{(b)}$ for the b^{th} bootstrap sample:
  $\widehat{\alpha}^{(b)} = \sum_{j=1}^{M} \alpha_j^{(b)} / M$ and $\widehat{\beta}^{(b)} = \sum_{j=1}^{M} \beta_j^{(b)} / M$
end for
Evaluate standard deviations $sd_1(\alpha)$ and $sd_1(\beta)$ as in Section 5.1

Algorithm 2 Proposed method for the logistic regression model in Section 5.1

for $b = 1$ to B **do**
  Draw a bootstrap sample, and obtain $(r_1^{(b)}, \ldots, r_n^{(b)})$
  for $m = 1$ to M **do**
    evaluate $\omega^{(b)}(\alpha_m, \beta_m) = \prod_{i=1}^n f^{(r_i^{(b)}-1)}(X_i, Y_i|\alpha_m, \beta_m)$ †
  end for
  Find averages $\widehat{\alpha}^{(b)}, \widehat{\beta}^{(b)}$ for the b^{th} bootstrap sample *:
  $\widehat{\alpha}^{(b)} = \sum_{j=1}^M \alpha_j \omega^{(b)}(\alpha_m, \beta_m) / \sum_{j=1}^M \omega^{(b)}(\alpha_m, \beta_m)$ and
  $\widehat{\beta}^{(b)} = \sum_{j=1}^M \beta_j \omega^{(b)}(\alpha_m, \beta_m) / \sum_{j=1}^M \omega^{(b)}(\alpha_m, \beta_m)$
end for
Evaluate standard deviations $sd_2(\alpha)$ and $sd_2(\beta)$ as in Section 5.1

† $(\alpha_1, \beta_1), \ldots, (\alpha_M, \beta_M)$ are from $\pi(\alpha, \beta|\boldsymbol{D})$.

*If we are interested in the q^{th} quantile we will compute $\widehat{\alpha}_q^{(b)}$, $\widehat{\beta}_q^{(b)}$ based on equation 2 in Section 4.1 at this step.

Table 1: Average computing time ($\pm$ standard deviation of 500 simulated data sets) measured in seconds for calculating standard errors of posterior summaries in logistic regression model from Section 5.1 based on the 1) regular bootstrap method, 2) the importance sampling based approach, and 3) the method proposed in Efron (2015). Here Q_1 and Q_3 denote the first and third quartiles, and 2.5^{th} and 97.5^{th} denote the 2.5^{th} percentile and 97.5^{th} percentile of the posterior distribution, respectively.

Method	Time to calculate			Computational
	Mean	Q_1 (Q_3)	2.5^{th} (97.5^{th})	complexity
1	247.78 ± 6.70	247.78 ± 6.70	247.78 ± 6.70	$O(BMn)$
2	46.65 ± 0.41	51.66 ± 0.40	120.44 ± 0.93	$O(BMn)$
3	4.18 ± 0.6			$O(Mn)$

Table 2: The frequentist standard errors and computing times for α and β of the linear measurement error model in Section 5.2. Here sd_1 and sd_2 denote the standard errors based on the regular bootstrap method and the importance sampling based approach.

Parameter		Posterior			
		Mean	Q_2	2.5^{th}	97.5^{th}
α	sd_1	0.028	0.027	0.028	0.027
	sd_2	0.027	0.029	0.028	0.030
β	sd_1	0.034	0.034	0.033	0.036
	sd_2	0.030	0.032	0.035	0.031
Computation time in second	sd_1	233.06	233.06	233.06	233.06
	sd_2	22.99	26.36	24.16	24.16
Computational complexity	sd_1	$O(BMn)$	$O(BMn)$	$O(BMn)$	$O(BMn)$
	sd_2	$O(BMn)$	$O(BMn)$	$O(BMn)$	$O(BMn)$

Table 3: Posterior summaries and the corresponding frequentist standard errors of β_0, β_1, and α used in the Weibull model for analyzing the E1684 melanoma data given in Section 5.3. Here sd_1 and sd_2 denote the standard errors based on the regular bootstrap method and the importance sampling based approach.

Parameter		Posterior Mean	Q_2	2.5^{th}	97.5^{th}
β_0		-1.103	-1.101	-1.710	-0.586
	sd_1	0.278	0.278	0.295	0.266
	sd_2	0.255	0.265	0.252	0.261
β_1		-0.256	-0.256	-0.585	0.090
	sd_1	0.177	0.178	0.180	0.179
	sd_2	0.169	0.176	0.163	0.183
α		0.791	0.793	0.688	0.891
	sd_1	0.038	0.039	0.035	0.043
	sd_2	0.037	0.038	0.034	0.038
Computation time in sec	sd_1	151.77	151.77	151.77	151.77
	sd_2	37.31	43.59	85.75	85.75
Computational complexity	sd_1	$O(BMn)$	$O(BMn)$	$O(BMn)$	$O(BMn)$
	sd_2	$O(BMn)$	$O(BMn)$	$O(BMn)$	$O(BMn)$

Table 4: The frequentist standard errors of the estimated the impulse responses at each time lag. Here sd_1 and sd_2 denote the standard errors based on the regular bootstrap method and the importance sampling based approach.

Time lag	$\boldsymbol{y}_2$ to $\boldsymbol{y}_1$		$\boldsymbol{y}_1$ to $\boldsymbol{y}_2$	
	sd_1	sd_2	sd_1	sd_2
1	0.0241	0.0222	0.0404	0.0403
2	0.0388	0.0374	0.0338	0.0334
3	0.0442	0.0434	0.0304	0.0298
4	0.0431	0.0428	0.0279	0.0272
5	0.0401	0.0402	0.0262	0.0256
6	0.0369	0.0372	0.0253	0.0248
7	0.0339	0.0342	0.0247	0.0244
8	0.0311	0.0315	0.0243	0.0242
9	0.0286	0.0290	0.0239	0.0240
10	0.0263	0.0267	0.0234	0.0236
11	0.0242	0.0246	0.0227	0.0232
12	0.0222	0.0226	0.0220	0.0225
13	0.0204	0.0208	0.0212	0.0218
14	0.0188	0.0191	0.0203	0.0210
15	0.0173	0.0176	0.0194	0.0201

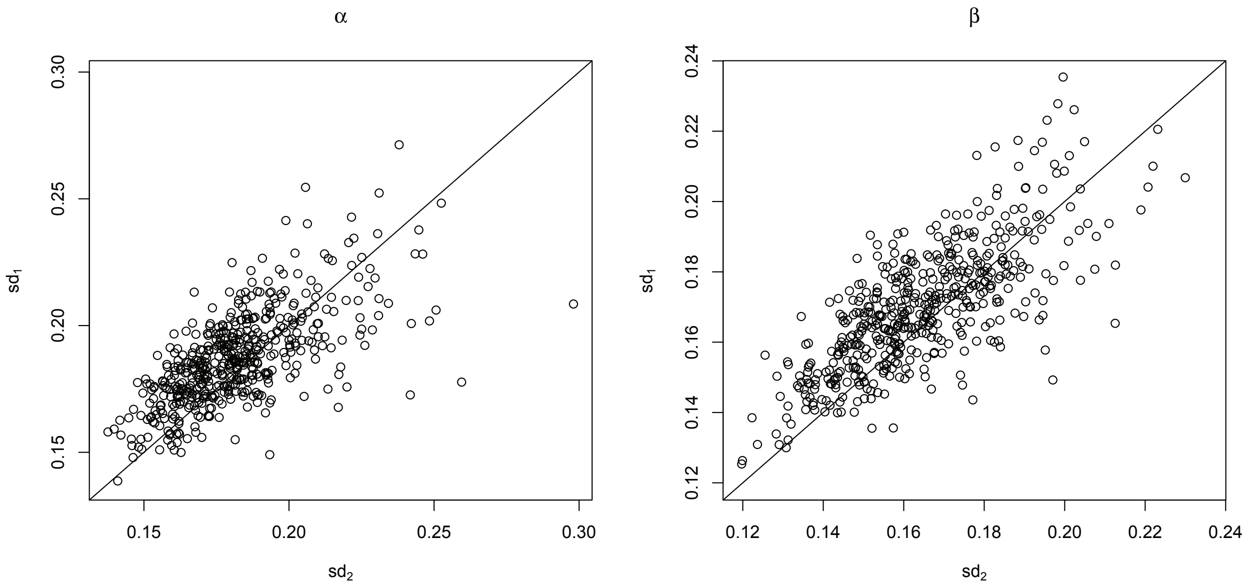

Figure 1: Frequentist standard errors of posterior means of the intercept (α) and the slope (β) of the logistic regression model from the 500 simulated data sets in Section 5.1 based on the regular bootstrap method (Y-axis) and the proposed importance sampling based method (X-axis).

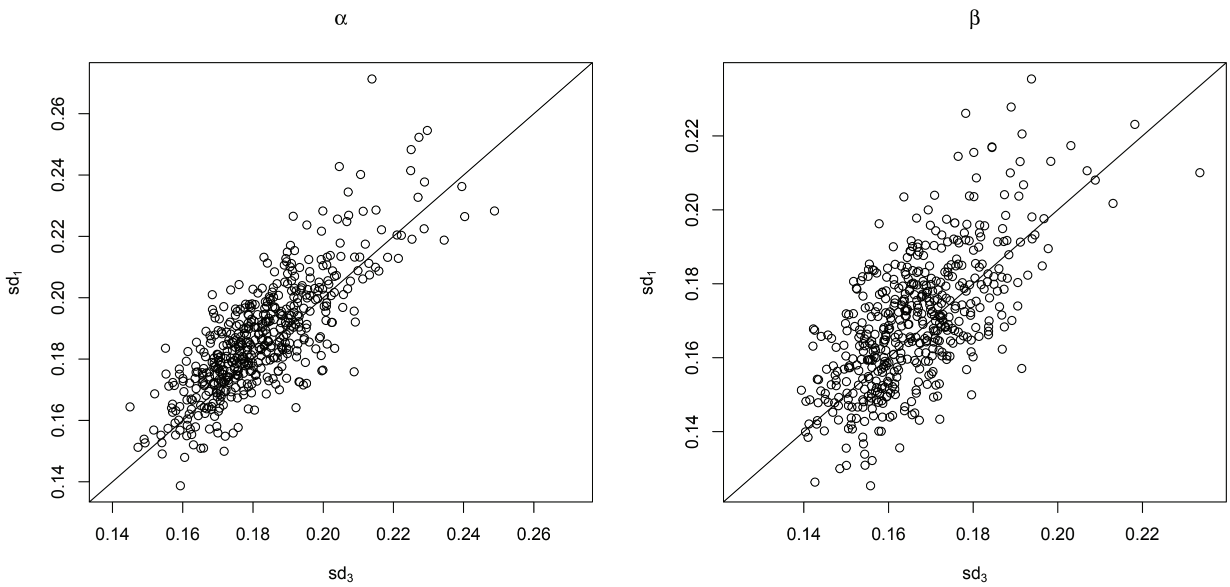

Figure 2: Frequentist standard errors of posterior means of the intercept (α) and the slope (β) of the logistic regression model from the 500 simulated data sets in Section 5.1 based on the regular bootstrap method (Y-axis) and the approach proposed in Efron (2015) (X-axis).

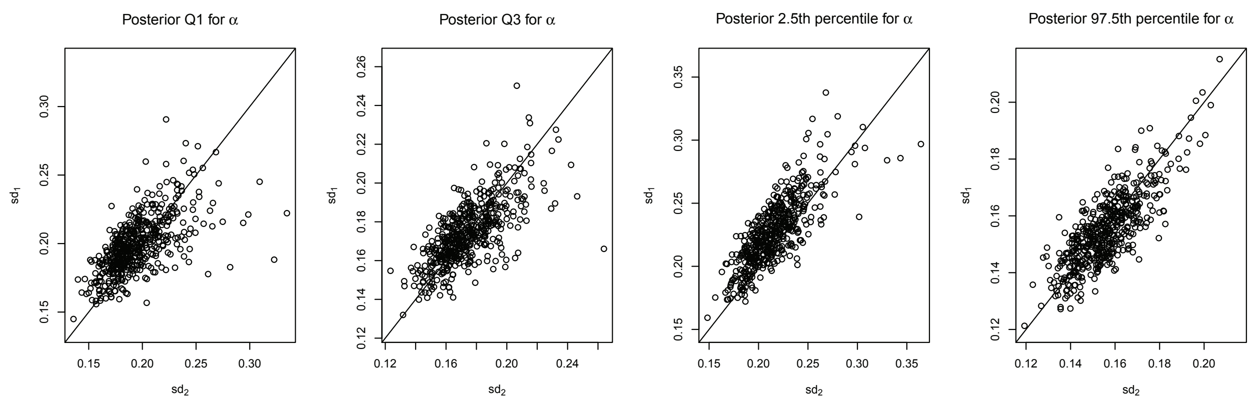

Figure 3: Frequentist standard errors of Q_1, Q_3, 2.5^{th} percentile, and 97.5^{th} percentile of the posterior distribution of α in the logistic regression model from the 500 simulated data set in Section 5.1. Regular bootstrap standard errors are presented along the Y-axis while importance sampling based standard errors are presented along the X-axis.

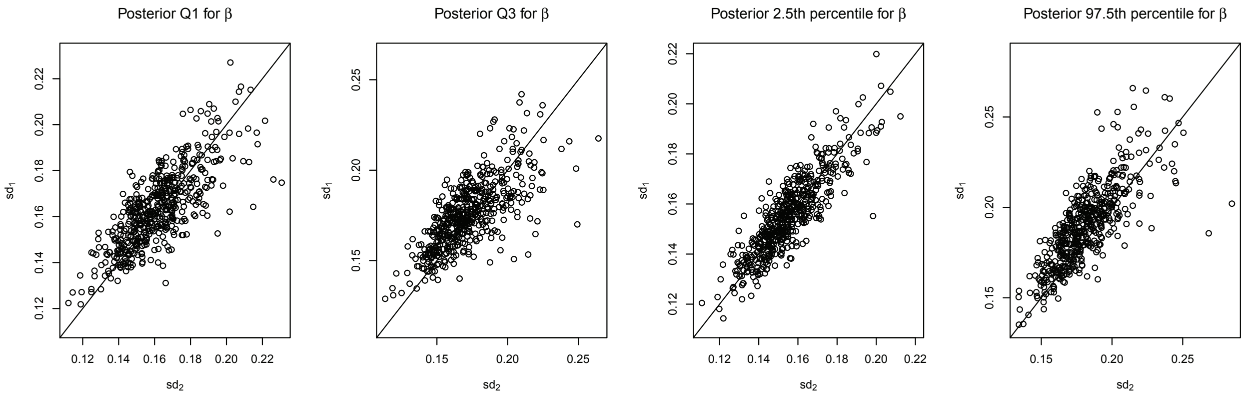

Figure 4: Frequentist standard errors of Q_1, Q_3, 2.5^{th} percentile, and 97.5^{th} percentile of the posterior distribution of β in the logistic regression model from the 500 simulated data set in Section 5.1. Regular bootstrap standard errors are presented along the Y-axis while importance sampling based standard errors are presented along the X-axis.

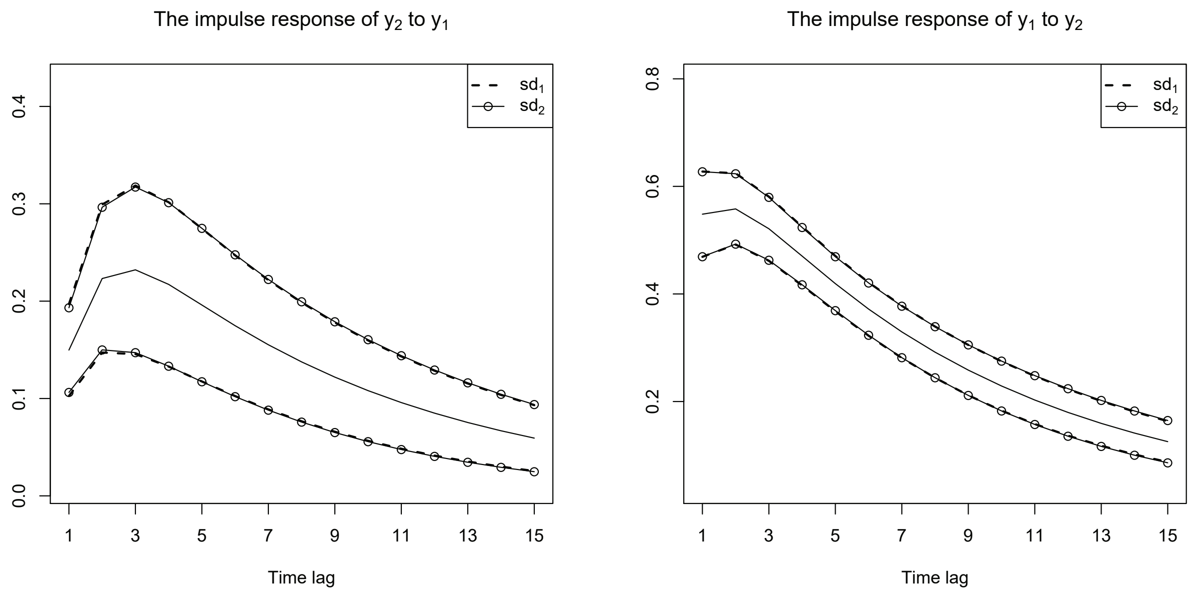

Figure 5: The estimated impulse responses (solid line) and its 95% (pointwise) confidence band of $\boldsymbol{y}_2$ to a shock in $\boldsymbol{y}_1$ (left panel) and vice versa (right panel) referenced in Section 5.4. The bold dotted line is based on regular bootstrap approach (sd_1) while the circled solid line is based on importance sampling based approach (sd_2).