Statistica Sinica **25** (2015), 725-741
doi:http://dx.doi.org/10.5705/ss.2013.099

A BAYESIAN APPROACH TO CONSTRUCTING MULTIPLE CONFIDENCE INTERVALS OF SELECTED PARAMETERS WITH SPARSE SIGNALS

Zhigen Zhao and Sanat K. Sarkar

Temple University

Abstract: Selective inference using multiple confidence intervals is an emerging area of statistical research whose importance is being realized very recently. We consider making such inference in the context of analyzing data with sparse signals in a Bayesian framework. Although the traditional posterior credible intervals are immune to selection, they can have low power in detecting the true signals because of covering no-signal too often if the sparse nature of the data is not properly taken into account. We demonstrate this phenomenon using a canonical Bayes model with the parameters of interest following a zero-inflated mixture prior. We propose a new method of constructing multiple intervals for any given selection rule taking a Bayesian decision theoretic approach under such a model. It involves the local fdr, the posterior probability of a parameter being null which is commonly used in multiple testing. It controls an overall measure of error rate, the Bayes or posterior false coverage rate, at a desired level among the selected intervals. We apply this method to the regression problem and demonstrate via simulations as well as data analyses that it is much more powerful in terms of enclosing zero less frequently than the traditional and some alternative methods.

Key words and phrases: FCR, multiple intervals, selection.

1. Introduction

In modern statistical problems involving a large number of parameters, inference is often made on parameters that are selected based on the data; see, for instance, Rossouw et al. (2002), Giovannucci et al. (1995), Qiu and Hwang (2007), Benjamini, Heller and Yekutieli (2009), Zhao and Hwang (2012), and Hwang and Zhao (2013). The statistical challenges associated with such selective inference have not been realized until recently. For constructing confidence intervals, scientists often proceed with the naive approach of constructing standard confidence intervals for the selected parameters, pretending there was no selection process (see, for instance, Giovannucci et al. (1995) and Rossouw et al. (2002)). However, the bias introduced by the selection, the so-called "winner's curse", can cause the usual confidence interval to have an extremely low coverage probability. One can see this phenomenon from the following:

Toy Example: We generated a random sample of pairs (Y_i, β_i), $i = 1, \ldots, 1{,}000$, as $Y_i|\beta_i \overset{\text{ind}}{\sim} N(\beta_i, 1)$ where $\beta_i = 0$ with probability 0.8, and $\beta_i \sim N(0,1)$ with probability 0.2. Let $Y_{(1)} = \max_{1\le i\le p} Y_i$ and $\beta_{(1)}$ be the corresponding parameter. Construct the usual 95% confidence interval for $\beta_{(1)}$, $CI_{(1)} = Y_{(1)} \pm 1.96$. We repeated this experiment 10,000 times to simulate the coverage probability. It was 42.4%. Here, of course, $Y_{(1)}|\beta_{(1)}$ is no longer normal. Indeed, it is easy to see that $EY_{(1)} \ge \beta_{(1)}$. When using such a biased estimator as the center of the confidence interval, the margin of error should account for this bias as well as the randomness.

This example illustrates the importance of developing a confidence interval for a selected parameter that is statistically sound. This paper addresses the problem of constructing confidence intervals for multiple selected parameters subject to controlling an overall measure of false coverage. There are pioneering works in this direction, Benjamini and Yekutieli (2005), Qiu and Hwang (2007), Benjamini, Heller and Yekutieli (2009), Efron (2011), Zhao and Hwang (2012), and Hwang and Zhao (2013). We address the problem in the context of a Bayes model with a zero inflated mixture prior (ZIMP) that is relevant for data arising in many modern applications.

There is some belief that the Bayes rule is immune to selection bias (Efron (2011), Senn (2008), and Dawid (1994)). We argue that the usual posterior credible intervals may lack good inferential property in the sense of enclosing zero very often, particularly under the ZIMP model

$$\boldsymbol{Y}|\boldsymbol{\beta} \sim f(\boldsymbol{y}|\boldsymbol{\beta}), \tag{1.1}$$

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)$ and, marginally,

$$\beta_i \sim \pi(\beta_i) = \pi_0 1(\beta_i = 0) + (1 - \pi_0)\psi(\beta_i). \tag{1.2}$$

Here, π_0 is the prior probability of β_i being zero, and $\psi(\beta_i)$ is the distribution of β_i given $\beta_i \ne 0$. This model is useful in genetic experiments where many of the genes are believed to be non-differentially expressed, and in regressions with sparsity structure (Chen and Dunson (2003), Rodriguez, Dunson and Taylor (2009)). If π_0 is large, the posterior probability of β_i being 0 can also be large. This is problematic for equal-tail credible intervals or highest-posterior-density (HPD) regions: equal-tail credible intervals can contain zero a high proportion of times and HPD regions always include zero due to the existence of a point mass at zero. The same problem occurs under the regression model $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $\boldsymbol{X}$ is an $n \times p$ matrix and the β_i's follow ZIMP.

With selection, we need to properly account for ZIMP while obtaining Bayes credible intervals. Here, we adopt the Bayesian decision theoretic approach to constructing such intervals (see Faith (1976), Casella and Hwang (1983), He (1992), and Hwang, Qiu and Zhao (2009)), but with loss functions adjusted for ZIMP.

We consider a loss function that penalizes the inclusion of zero when β_i is indeed non-zero. The Bayesian decision interval is then forced to include zero if there is overwhelming evidence that β_i is 0, or equivalently, if the local fdr score $P(\beta_i = 0|\boldsymbol{Y})$ (see Efron et al. (2001), Efron (2005, 2010)) is large. The local fdr score is compared to a tuning parameter k_2 used in the loss function, and the zero component is included in the interval if the local fdr score is greater than k_2. The choice of k_2 is critical, and difficult to determine. Some ad-hoc values for k_2 have been suggested in the literature, such as 0.2 (Efron (2008, 2010) and Zhao and Hwang (2012)). For a given selection rule, we determine k_2 so that the posterior false coverage rate (PFCR) or the Bayes false coverage rate (BFCR, Zhao and Hwang (2012)) is controlled at a desired level. Such k_2 is often larger than α, implying that the proposed interval doesn't necessarily include zero even if $P(\beta_i = 0|\boldsymbol{Y}) > \alpha$. The proposed intervals account for the selection via the choice of k_2 and meet the goal of controlling an overall measure of false coverage. Their usefulness is confirmed in simulations and in data analysis.

Here is how the article is organized. In Section 2, we discuss why the traditional Bayes intervals can be unreliable under ZIMP, particularly when π_0 is large, and recall the concepts of PFCR and the BFCR. In Section 3, we introduce our newly proposed loss function and present the derivation of the corresponding Bayesian decision intervals under the model (1-2). The choice of the tuning parameter k_2 guaranteeing control of the PFCR and the BFCR is also discussed in this section. In Section 4, we present the results of simulation studies and a real data analysis we conducted to compare our proposed intervals for regression coefficients under a hierarchical Bayesian lasso model using ZIMP with the usual equal tail credible intervals and the intervals that Park and Casella (2008) derived for the original Bayesian lasso model without assuming ZIMP. The technical proof and the steps of Gibbs sampling are put in the supplementary document.

2. Traditional Bayes Credible Intervals: Some Issues under ZIMP

There are two measures of false coverage that we consider from a Bayesian point of view. These are the Posterior False Coverage Rate (PFCR) and the Bayes False Coverage Rate (BFCR), defined as follows. Let $\mathcal{R}(\boldsymbol{Y})$ be the set of indices of the parameters selected based on the observation $\boldsymbol{Y}$, and $R = \#\mathcal{R}(\boldsymbol{Y})$. Given the credible intervals CI_i for β_i, $i \in \mathcal{R}(\boldsymbol{Y})$, let $\mathcal{V}$ consist of $i \in \mathcal{R}(\boldsymbol{Y})$ such that $\beta_i \notin CI_i$, and $V = \#\mathcal{V}$. Let Q be the proportion of the selected parameters that are not covered by their respective intervals: $Q = V/R$ if $R > 0$, and $= 0$ if $R = 0$. Benjamini and Yekutieli (2005) proposed the False Coverage Rate (FCR),

FCR $= E(Q|\boldsymbol{\beta})$, as a measure of false coverage among the selected parameters in a frequentist sense. A Bayesian analog of this measure is the Posterior False Coverage Rate (PFCR),

$$\text{PFCR}(\boldsymbol{Y}) = E(Q|\boldsymbol{Y}) = \begin{cases} \frac{1}{R}\sum_{i\in\mathcal{R}(\boldsymbol{Y})} P(\beta_i \notin CI_i|\boldsymbol{Y}) & \text{if } R > 0, \\ 0 & \text{if } R = 0. \end{cases} \tag{2.1}$$

Zhao and Hwang (2012) considered averaging the FCR with respect to the prior distribution of $\boldsymbol{\beta}$ to define the BFCR,

$$\text{BFCR} = \int \text{PFCR}(\boldsymbol{Y}) m(\boldsymbol{Y}) \boldsymbol{dY}, \tag{2.2}$$

where $m(\boldsymbol{Y})$ is the marginal density of $\boldsymbol{Y}$.

Clearly, if $P(\beta_i \notin CI_i|\boldsymbol{Y}) \leq \alpha$, for all $i = 1, 2, \ldots, p$, then both PFCR and BFCR are less than or equal to α for any selection rule $\mathcal{R}(\boldsymbol{Y})$. Thus, $100(1-\alpha)\%$ credible intervals obtained directly from the posterior distributions of β_i's can avoid adjusting for the selection rule, but such intervals do not have good inferential properties.

Let $\psi(\beta_i|\boldsymbol{Y}, \beta_i \neq 0)$ be the posterior distribution of β_i conditional on $\beta_i \neq 0$. Then,

$$\psi(\beta_i|\boldsymbol{Y}) = fdr_i(\boldsymbol{Y})1(\beta_i = 0) + (1 - fdr_i(\boldsymbol{Y}))\psi(\beta_i|\boldsymbol{Y}, \beta_i \neq 0),$$

where $fdr_i(\boldsymbol{Y}) = P(\beta_i = 0|\boldsymbol{Y})$ is the local fdr score that is widely used in the multiple testing literature. Generally, calculations of $fdr_i(\boldsymbol{Y})$ and $\psi(\beta_i|\boldsymbol{Y}, \beta_i \neq 0)$ require something like MCMC. But, our main interest is the post inference assuming that the posterior draws of all the parameters are available via certain methods. Whether or not we construct these intervals based on the posterior draws, there might be problems with $100(1-\alpha)\%$ credible intervals directly based on this posterior distribution, particularly when π_0 is large.

Theorem 1. *Let CI_i be a posterior interval for β_i such that $P(\beta_i \notin CI_i|\boldsymbol{Y}) \leq \alpha$. If $fdr_i(\boldsymbol{Y}) > \alpha$, then $0 \in CI_i$.*

Thus, whenever $fdr_i(\boldsymbol{Y}) > \alpha$, the posterior $100(1-\alpha)\%$ credible interval appears to enclose zero. In fact, when π_0 is large, most of the time the local fdr score will exceed α. To see this, consider the following:

Toy Example: Assume that $Y_i|\beta_i \overset{\text{ind}}{\sim} N(\beta_i, 1)$ and β_i's are i.i.d. draws from the population that follows (1.2) with $\pi_0 = 0.8$ and $\psi(\beta_i) \sim N(0, 1)$. The dimension is $p = 10,000$. We randomly generated the i.i.d. random vectors (Y_i, β_i)'s, $i = 1, 2, \ldots, p$, and calculated $fdr_i(\boldsymbol{Y})$ for each β_i. We then plotted the histogram of $fdr_i(\boldsymbol{Y})$, and present it in Figure 1. As seen from this figure, the

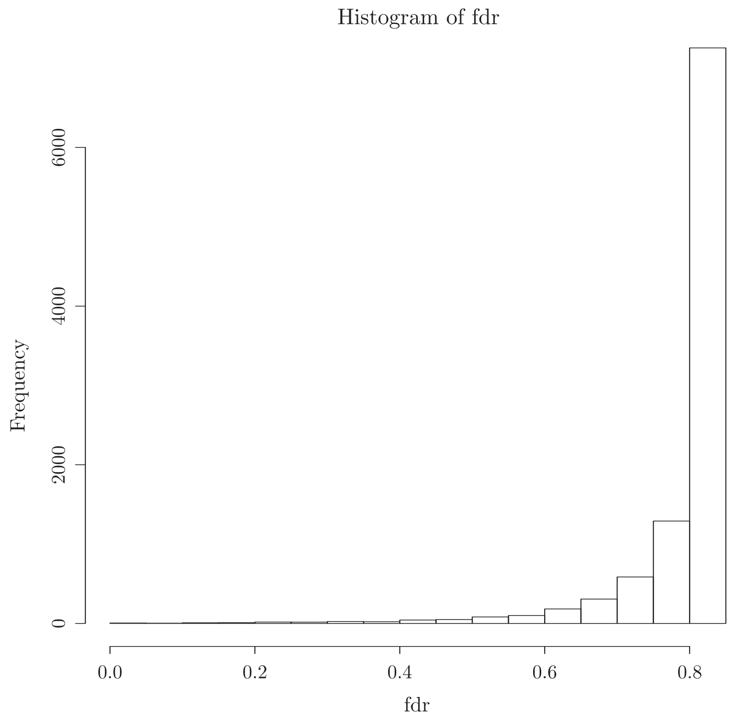

Figure 1. The histogram of $P(\beta_i = 0|\boldsymbol{Y})$ based on 10,000 random pairs (β_i, y_i) generated from the following distribution: $Y_i|\beta_i \overset{\text{ind}}{\sim} N(\beta_i, 1)$, where $\beta_i = 0$ with probability 0.8 and $\beta_i \sim N(0,1)$ with probability 0.2.

proportion of the $fdr_i(\boldsymbol{Y})$'s exceeding 0.10 can be overwhelmingly large. One can theoretically determine how large this proportion could be, since for this example,

$$fdr_i(\boldsymbol{Y}) = P(\beta_i = 0|\boldsymbol{Y}) = \frac{\pi_0\phi(y_i)}{\pi_0\phi(y_i) + \pi_1\phi(y_i/\sqrt{2})/\sqrt{2}},$$

where $\phi(x)$ is the density of the standard normal distribution. After some calculations, we find that $P(fdr_i(\boldsymbol{Y}) \geq 0.10) \geq P(|Y_i| \leq 3.96) \geq 0.999$. Thus, more than 99.9% of the time the 90% credible intervals will include zero.

A similar phenomenon has been seen in hypothesis testing. It appears to be conservative if we accept a null hypothesis when the local fdr is larger than α. Efron (2008) suggested the threshold of the local fdr for rejecting a hypothesis as 0.2. Scott and Berger (2006) argued that it is important to separate the zero component and $P(\beta_i|\boldsymbol{Y}, \beta_i \neq 0)$ for the posterior distribution $\psi(\beta_i|\boldsymbol{Y})$. They further applied decision theory to derive the testing procedure with a loss function involving a tuning parameter. Sarkar, Zhou and Ghosh (2008) considered the optimal testing procedure which controls the Bayes FDR. None of these papers provide an answer to the construction of credible intervals.

3. Bayesian Decision Theoretic Confidence Intervals under ZIMP

There have been many attempts to apply the Bayesian decision theoretic approach to the construction of confidence sets/intervals. Faith (1976) considered a loss function for confidence set CS of the vector $\boldsymbol{\beta}$ as $L(\boldsymbol{\beta}, CS) = kVolume(CS) - I_{CS}(\boldsymbol{\beta})$, where $I_{CS}(\boldsymbol{\beta})$ equals 1 if and only if the vector $\boldsymbol{\beta}$ is in the confidence set CS. Casella and Hwang (1983) used the same loss where the tuning parameter k was determined so that the usual $100(1-\alpha)\%$ confidence set would be minimax. When assuming $Y_i|\beta_i \overset{\text{ind}}{\sim} N(\beta_i, \sigma^2)$, He (1992) used $L(\beta_i, CI_i) = kLen(CI_i) - I_{CI_i}(\beta_i)$ as the loss function for the interval estimator CI_i of the parameter β_i. Hwang, Qiu and Zhao (2009) modified this loss function as $L(\beta_i, CI_i) = kLen(CI_i)/\sigma_i - I_{CI_i}(\beta_i)$, assuming unknown and unequal variances, and constructed confidence intervals that shrink both means and variances. These loss functions are not appropriate for the model (1.1-1.2).

As these loss functions have risks reduced by the addition of zero to the confidence intervals, we need to penalize the inclusion of zero when β_i is indeed non-zero. We take the loss function

$$L(\beta_i, CI_i) = \{(k_1^i Len(CI_i) - I_{CI_i}(\beta_i))1(\beta_i \neq 0) + I_{CI_i}(0)(k_2 - 1(\beta_i = 0))\}, \quad (3.1)$$

where $0 \leq k_2 \leq 1$.

The first term $(k_1^i Len(CI_i) - I_{CI_i}(\beta_i))1(\beta_i \neq 0)$ balances length and true coverage. The second term $I_{CI_i}(0)(k_2 - 1(\beta_i = 0))$ affects the loss function only when the corresponding interval does include zero. If $0 \in CI_i$ and β_i is indeed zero, then $k_2 - 1(\beta_i = 0) = k_2 - 1 \leq 0$, and including zero reduces the loss and is beneficial. On the other hand, if $0 \in CI_i$ but β_i is non-zero, this term is positive and becomes a penalty. Unlike existing loss functions, the inclusion of zero is not always beneficial in reducing the risk. When β_i is nonzero, the tuning parameter k_2 decides the amount of penalty for including zero in the interval. When $k_2 = 1$, the interval is forced to exclude zero. We allow the tuning parameter k_1^i to depend on the observation $\boldsymbol{Y}$. Otherwise, it leads to a paradox, as demonstrated in Casella, Hwang and Robert (1993).

We want to construct the Bayes intervals CI_i^{BD}'s that minimize $E(L(\beta_i, CI_i)|\boldsymbol{Y}))$ for any observation $\boldsymbol{Y}$, assuming the mixture model (1.1)−(1.2) and the loss function (3.1).

Theorem 2. *Assume the model* (1.1)−(1.2) *and the loss function* (3.1)*. Then the Bayesian decision theoretic interval for* β_i *is*

$$CI_i^{BD} = \begin{cases} \{\beta_i : k_1^i < \psi(\beta_i|\boldsymbol{Y}, \beta_i \neq 0)\} \setminus \{0\} \text{ if } fdr_i(\boldsymbol{Y}) < k_2, \\ \{\beta_i : k_1^i < \psi(\beta_i|\boldsymbol{Y}, \beta_i \neq 0)\} \cup \{0\} \text{ if } fdr_i(\boldsymbol{Y}) \geq k_2. \end{cases} \quad (3.2)$$

Here, if the fdr score is smaller than the tuning parameter k_2, implying strong evidence that the corresponding parameter β_i is non-zero, zero is excluded from the interval. If the score is larger than k_2, we force the interval to include zero.

In (3.2), the component $\{\beta_i : k_1^i < \psi(\beta_i|\boldsymbol{Y}, \beta_i \neq 0)\}$ relies on the tuning parameters k_1^i's, and the posterior density of β_i, conditional on $\beta_i \neq 0$, given $\boldsymbol{Y}$. One can thus choose the k_1^i in such a way that the resulting interval is a $100(1-\alpha)\%$ HPD region $CI_i(\alpha)$ based on $\psi(\beta_i|\boldsymbol{Y}, \beta_i \neq 0)$. With this choice of $CI_i(\alpha)$, we define our Bayesian decision interval as

$$CI_i^{BD} = \begin{cases} CI_i(\alpha) \setminus \{0\}, \text{ if } fdr_i(\boldsymbol{Y}) < k_2, \\ CI_i(\alpha) \cup \{0\}, \text{ if } fdr_i(\boldsymbol{Y}) \geq k_2. \end{cases} \quad (3.3)$$

Since we mix zero with another connected set, the derived intervals could be disconnected, which is in agreement with a comment from Efron (2008), stating that "this kind of disconnected description is natural to the two groups models". We will still use the word "interval" in our discussion. In the intervals (3.3), k_2 is the same for all β_i's, the choice of which should take into account both the multiplicity and selection issues when we are interested in multiple selected parameters.

Theorem 3. *Assume the model* (1.1)−(1.2), *and that the intervals are constructed according to* (3.3), *where the* k_1^i*'s are chosen such that* $P(\beta_i \notin CI_i|\boldsymbol{Y}, \beta_i \neq 0) \leq \alpha$. *Let*

$$k_2 = argmax_k\{k : \frac{1}{R} \sum_{i \in \mathcal{R}(\boldsymbol{Y})} fdr_i(\boldsymbol{Y})(I(fdr_i(\boldsymbol{Y}) < k) - \alpha) \leq 0\}. \quad (3.4)$$

Then, the PFCR and BFCR for these intervals is controlled at α.

The choice of k_2 in (3.4) thus guarantees that the PFCR is less than or equal to α. This criterion is different from that of the posterior coverage probability $P(\beta_i \in CI_i|\boldsymbol{Y})$ being greater than or equal to $1-\alpha$ for all $i \in \mathcal{R}$. Also, k_2 is larger than α in general. For the toy example we considered in Section 2, our numerical calculation shows that k_2 is 0.179 if we choose $\alpha = 0.1$ and select the top 1% of the most important parameters according to the magnitude of the observation. The proposed intervals thus enclose zero less frequently than the traditional approaches.

4. Application to Variable Selection

In this section, we apply (3.3) to the problem of making multiple inference about the components of $\boldsymbol{\beta}$ in the regression model: $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, although the application does not have to be restricted to this particular setup.

Park and Casella (2008) considered the Bayes version of the lasso (Tibshirani (1996)) and derived the Gibbs sampler to construct credible intervals based on the posterior draw of the parameters. Kyung et al. (2010) further extended this work to other methods such as Group Lasso (Yuan and Lin (2006)), Fused Lasso (Tibshirani et al. (2005)), and Elastic Net (Zou and Hastie (2005)). In these work, Park and Casella (2008) and Kyung et al. (2010) did not assume ZIMP for the parameters β_i's. Since it is common to assume a sparsity structure for the β_i's when the dimension p is much larger than n, we consider a hierarchical Bayesian lasso model with ZIMP for the β_i's:

$$\begin{cases} \boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \text{where } \boldsymbol{\epsilon} \sim N_n(\mathbf{0}_n, \sigma^2 \boldsymbol{I}_n), \\ \beta_i | \tau_1^2, \ldots, \tau_p^2, \pi_0 \overset{\text{ind}}{\sim} \pi_0 1(\beta_i = 0) + (1-\pi_0) N(0_p, \sigma^2 \tau_i^2), \\ \pi_0 \sim Beta(k\eta, k(1-\eta)), \sigma^2 \sim (\sigma^2)^{-1} d\sigma^2, \\ \tau_1^2, \ldots, \tau_p^2 \sim \prod_{j=1}^{p} \frac{\lambda^2}{2} \exp\Big(-\frac{\lambda^2 \tau_j^2}{2} \Big) d\tau_j^2, \\ \lambda^2 \sim \frac{\delta^r}{\Gamma(r)} (\lambda^2)^{r-1} e^{-\delta\lambda^2}. \end{cases} \tag{4.1}$$

The steps for the Gibbs sampler are in the supplementary document. The code is available at `http://astro.temple.edu/~zhaozhg/software.html`.

4.1. Simulation

We conducted extensive simulation studies to investigate how the confidence intervals (3.3) compare to the Bayes equal-tail $100(1-\alpha)\%$ credible intervals (abbreviated as Equal Tail) and the equal-tail credible intervals for the Bayesian lasso without ZIMP (abbreviated as No ZIMP).

Three simulation settings were considered differing on how the design matrix $\boldsymbol{X}$ and the parameters $(\boldsymbol{\beta}, \sigma)$ were chosen before generating the data from the model:

$$Y_i = \boldsymbol{X}_i' \boldsymbol{\beta} + \epsilon_i, \; i = 1, \ldots, n, \text{ where } \epsilon_i \overset{\text{i.i.d.}}{\sim} N(0, \sigma^2),$$

with $\boldsymbol{X}_i'$ being the ith row of $\boldsymbol{X}$.

Setting 1. We set $n = 100$, $p = 500$, $\pi_0 = 0.95$, and generated each $\boldsymbol{X}_i$ from $N_p(\mathbf{0}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} = (\sigma_{ij})_{i,j=1,2,\ldots,p}$ with $\sigma_{ij} = 0.5^{|i-j|}$. Of the 500 β_i's, $p(1-\pi_0) = 25$ were randomly selected prior to generating them according to (4.1) with the hyper parameters of the Gamma distribution for λ^2 chosen as $r = 1$ and $\delta = 10$, the rest were set at 0, and $\sigma = 1, 3$, and 5.

Setting 2. We set $n = 100$ and $p = 40$, generated each $\boldsymbol{X}_i$ from $N_p(\mathbf{0}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} = (1-\rho)\boldsymbol{I}_p + \rho \mathbf{1}_p \mathbf{1}_p^T$ (with $\mathbf{1}_p = (1, \ldots, 1)'$) and $\rho = 0.5$, and considered

$\boldsymbol{\beta} = (\mathbf{0}, \mathbf{2}, \mathbf{0}, \mathbf{0})$, with both $\mathbf{0}$ and $\mathbf{2}$ ten-dimensional. For σ, the values 1, 3 and 5 were chosen.

Setting 3. This was similar to Setting 1 except that, out of the 25 non-zero β_i's, 15 were chosen randomly with their values set to 2 and the rest set to -1.5.

We fit the hierarchical Bayesian model with ZIMP assuming unknown π_0 and setting the hyper parameters at $k = 2$ and $\eta = 1/2$ for the Beta distribution. We took $r = 1$ and $\delta = 10$. The Gibbs sampler was run 11,000 times. The first 1,000 of these iterations were considered as burn-in and every 10th value obtained from the remaining 10,000 was chosen to lessen the sequential dependence between iterations. We also fit this model by using the Bayesian lasso (Park and Casella (2008)) without assuming ZIMP. All the simulation codes were written in R except the Gibbs sampler was implemented using C with gnu scientific library. The simulation was performed on a desktop with Debian Wheezy x86_64 operation system. The CPU was Intel (R) Core(TM) 2 Quad Core Q9650 @ 3.00 GHz. The total memory was 8GB. It took 5 minutes and 55 seconds to finish one simulation in Setting 1; it took 14 seconds to finish one simulation in Setting 2, and the simulation time in Setting 3 was comparable to that of Setting 1.

Two selection rules were considered in each simulation setting. One of these, a multiple testing procedure abbreviated as MTP, selected the parameters using a method for controlling the Bayes FDR at 15% (Sarkar, Zhou and Ghosh (2008)): found $R = max\{1 \leq k \leq p : (1/k)\sum_{i=1}^{k} fdr_{(i)}(\boldsymbol{Y}, \boldsymbol{X}) \leq 0.15\}$, where $fdr_{(1)}(\boldsymbol{Y}, \boldsymbol{X}) \leq \cdots \leq fdr_{(p)}(\boldsymbol{Y}, \boldsymbol{X})$ are the ordered versions of the local fdr scores $fdr_i(\boldsymbol{Y}, \boldsymbol{X}) = P(\beta_i = 0|\boldsymbol{Y}, \boldsymbol{X})$, and then selected the parameters corresponding to $fdr_{(1)}(\boldsymbol{Y}, \boldsymbol{X}), \ldots, fdr_{(R)}(\boldsymbol{Y}, \boldsymbol{X})$. The second rule, a screening method abbreviated as SCR, selected the top $d = \min(\lceil 3n/\log n \rceil, p)$ predictors after sorting them in descending order of the magnitude of correlation between each of them and $\boldsymbol{Y}$ (Fan and Lv (2008)).

The proposed intervals according to (3.3), the Bayes equal-tail $100(1-\alpha)\%$ credible intervals, and the equal-tail credible intervals for the Bayesian lasso without ZIMP were then constructed. We set $\alpha = 0.1$ in all the simulations. For each construction, we calculated the false coverage proportion $V/(R \vee 1)$ to simulate BFCR based on 100 replications. Also calculated were the simulated values of two quantities measuring inferential properties of the intervals produced by a particular method: $ETP = E(\#\{i : \theta_i \neq 0, \theta_i \text{ is selected}, 0 \notin CI_i\}/\#\{i : \theta_i \neq 0, \theta_i \text{ is selected}\})$, the expected proportion of intervals not containing zero among all intervals for the selected non-zero parameters, and $EFP = E(\#\{i : \theta_i = 0, \theta_i \text{ is selected}, 0 \notin CI_i\}/\#\{i : \theta_i = 0, \theta_i \text{ is selected}\})$, the expected proportion of intervals not containing zero among all intervals for the selected zero parameters. The simulated values of the average length (Leng) of the selected parameters

Table 1. The results of simulation studies

			Proposed				Equal Tail			No ZIMP		
	σ	Sel	k_2	BFCR	Leng	ETP/EFP(%)	BFCR	Leng	ETP/EFP(%)	BFCR	Leng	ETP/EFP(%)
1	1	MTP	0.19	0.069	0.841	51/0.021	0.063	0.996	45/0.0042	0.66	1.16	16/0.023
		SCR	0.71	0.026	0.787	69/0.81	0.019	0.322	45/0.0042	0.13	0.99	16/0.023
	3	MTP	0.14	0.076	2.4	39/0.025	0.071	2.81	33/0.0063	0.59	2.83	15/0.013
		SCR	0.63	0.047	2.1	59/0.9	0.03	1.05	33/0.0063	0.12	2.33	15/0.013
	5	MTP	0.12	0.074	3.96	32/0.0063	0.061	4.63	27/0.0042	0.57	4.37	13/0.015
		SCR	0.61	0.051	3.39	56/0.78	0.031	1.73	27/0.0042	0.11	3.51	13/0.015
2	1	MTP	0.13	0.21	1.68	84/0.033	0.2	1.7	68/0	0.26	1.66	95/1.8
		SCR	0.3	0.17	1.55	100/8.6	0.082	1.4	68/0	0.11	1.52	95/1.8
	3	MTP	0.15	0.11	1.93	87/2	0.059	1.94	76/0.33	0.13	1.94	95/4.7
		SCR	0.34	0.14	1.76	98/12	0.026	1.56	76/0.33	0.06	1.75	95/4.7
	5	MTP	0.11	0.092	2.47	56/1.9	0.058	2.53	42/0.57	0.12	2.43	69/3.8
		SCR	0.3	0.13	2.22	85/9.4	0.045	1.97	42/0.57	0.071	2.16	69/3.8
3	1	MTP	0.11	0.057	1.15	71/0.013	0.079	1.47	60/0	0.59	1.66	14/0.0042
		SCR	0.69	0.026	1.27	100/0.72	0.025	0.669	60/0	0.15	1.53	14/0.0042
	2	MTP	0.1	0.025	2.39	16/0.0063	0.021	2.66	12/0	0.45	2.03	11/0.0021
		SCR	0.46	0.052	1.81	50/0.57	0.038	1.24	12/0	0.14	1.46	11/0.0021
	3	MTP	0.1	0.19	2.27	4.5/0.025	0.11	2.36	1.8/0	0.4	1.86	4.8/0.025
		SCR	0.27	0.19	1.68	28/0.85	0.12	1.47	1.8/0	0.17	1.39	4.8/0.025

were also calculated for the different approaches. The simulated values of BFCR, Leng, ETP, and EFP for the different procedures, selection rules, and simulation settings are presented in Table 1.

As anticipated, the BFCR of the Equal Tail method, being selection-free, is well controlled under both selection rules. However, the ETP for this method is uniformly smaller than that for our proposed method. In Setting 3 when $\sigma = 1$, the proposed method is seen to identify all of the non-zero parameters under SCR, a significant improvement over the existing methods. The EFP of the proposed method is larger than that of the equal-tail credible intervals. This usually translates into a larger ETP as long as the BFCR is controlled. In all studies, the number of proposed intervals for the non-zero parameters selected by MTP and not containing zero is uniformly smaller than that for SCR. One explanation is that the MTP tends to select fewer parameters for further investigation, missing a large portion of the non-zero parameters in the selection step.

The maximum value of the BFCR for the proposed method is 21%; it occurs because the dataset is not generated from (4.1). The BFCR is controlled very well in Setting 1 with the dataset generated according to (4.1), except σ and π_0 are considered fixed. This raises the issue of robustness of the model, as pointed out by one of the referees but left for the future investigation.

In Table 1, we include the simulated values of k_2 for our method as defined in (3.4). These values are all seen to be greater than α, a good feature as explained in Section 3.

Remark 1. We checked that the usual frequentist t-intervals, not adjusted for the selection rule, did not work when used for the selected parameters.

The proposed intervals generally have less shrinkage effect than the equal-tail intervals. To demonstrate this, we took two selected parameters and plotted the intervals in Figure 2. As seen from these figures, when the selected parameter is zero, with larger shrinkage effect, the traditional approach appears to produce much shorter intervals than the proposed one; when the parameter is non-zero, large shrinkage effect causes the traditional $1-\alpha$ credible intervals to enclose zero while the proposed intervals avoid this.

Small shrinkage appears to be an appealing feature for the proposed intervals, especially for the non-zero β_i's about which statisticians care most, because they tend not to enclose zero. One side effect, of course, is that the proposed intervals can be longer than the traditional ones, especially for those zero parameters.

By taking a Bayesian decision theoretic approach, we have balanced the length, a measure of true coverage, and statistically meaningful inferential properties of the intervals. This, we would argue, is an appropriate way to construct credible intervals for sparse signals.

4.2. Diabetes data analysis

We applied our proposed intervals to the diabetes data used by Efron et al. (2004) that has 442 ($=n$) subjects and 10 baseline variables. We included all the baseline variables, squares of the baseline variables except the dichotomous variable "sex", and all the first-order interactions; so there were 64 ($=p$) predictors in total. We had each column of the design matrix $\boldsymbol{X}$ with mean 0 and unit length through appropriate scale and location transformations. We then fit the regression model

$$\boldsymbol{Y} - \frac{1}{n}\mathbf{1}_n\bar{Y} = \boldsymbol{X\beta} + \boldsymbol{\epsilon},$$

where $\bar{Y} = (1/n)\sum_{i=1}^{n} Y_i$, along with the Bayesian lasso model with ZIMP with hyper parameters $k=2$, $\eta = 1/2$, $r=1$, and $\delta = 10$. As in the simulations, we ran the Gibbs sampler 11,000 times, with the first 1,000 of these being the burn-in and every 10th generated value from the remaining 10,000 iterations being chosen to avoid sequential dependence.

We considered the selection rules MTP and SCR, as described in Section 4.1. Among the 64 parameters, 27 were selected by the MTP whereas all 64 were selected by the SCR.

We constructed the intervals for the corresponding regression coefficients using (3.3) and the equal-tail credible intervals, both under the assumption of Bayesian lasso with ZIMP for all the regression coefficients, and the credible intervals based on the Bayesian lasso without ZIMP for these regression coefficients

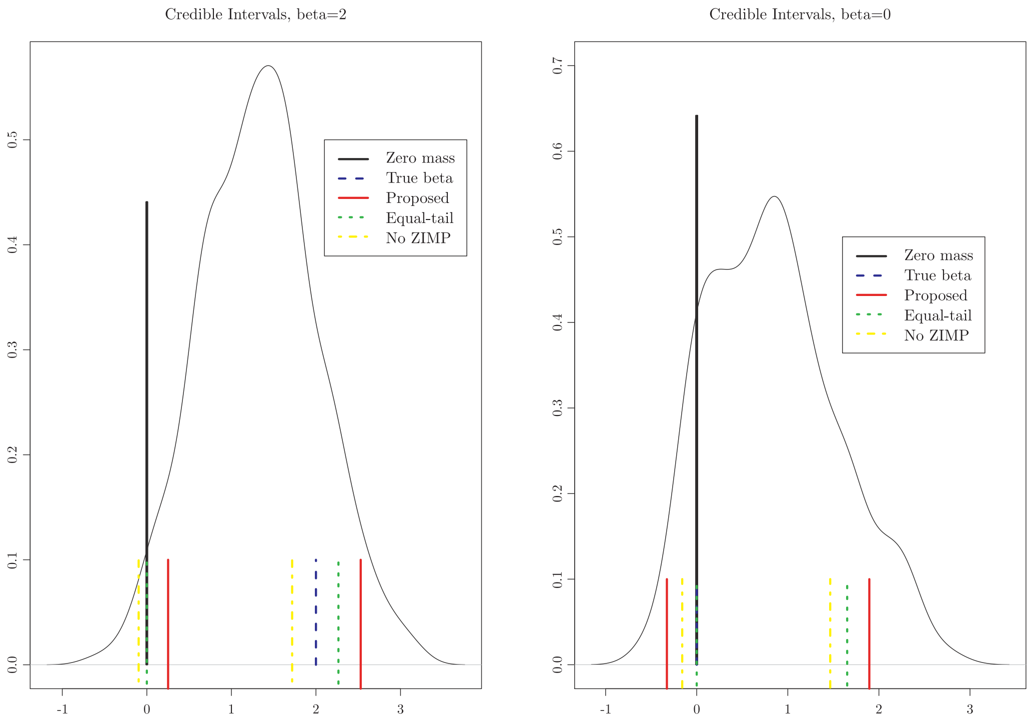

Figure 2. Different confidence intervals for two selected parameter in one simulation. In the left panel, the parameter β is non-zero and the parameter β in the right panel is zero. The curve is the kernel density estimator of $\psi(\beta_i|\boldsymbol{Y}, \boldsymbol{X}, \beta_i \neq 0)$.

(Park and Casella (2008)), considering $\alpha = 0.05$. These intervals are displayed in Figure 3 for the MPT, and in Figure 4 for the SCR. Table 2 reports the average lengths of the selected parameters.

The proposed method identified seven predictors with confidence intervals excluding zero, for each of the two selection rules we have considered. Specifically, bmi, map, ltg, hdl, sex, and the interactions between age and sex and between bmi and map were the seven predictors. The traditional equal-tail credible intervals under Bayesian lasso with ZIMP identified only three predictors: bmi, map, and ltg, and the credible intervals based on the Bayesian lasso without ZIMP identified five predictors: bmi, map, ltg, sex, and age:sex.

We ran the *lars* algorithm of Efron et al. (2004) and the solution path indicated that the variables bmi, ltg, map, hdl, bmi:map, age:sex, entered into the model consecutively. Two predictors, hdl and bmi:map, enter into the model early but were only detected by the proposed method.

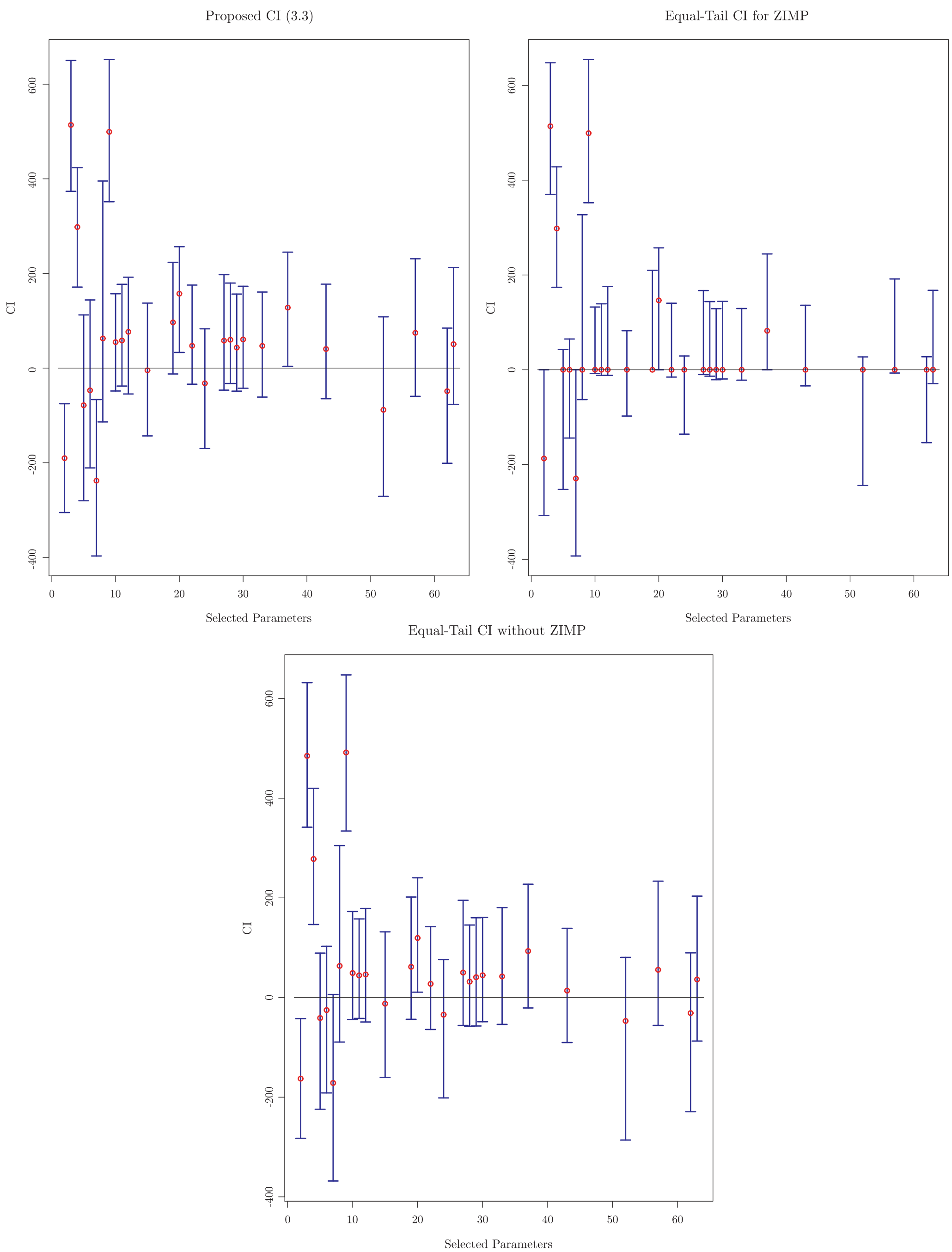

Figure 3. Confidence intervals for the 27 parameters selected out of the total 64 parameters using MPT constructed under Bayesian lasso with ZIMP using (3.3) (Panel 1), the traditional equal-tail credible intervals (Panel 2), and under Bayesian lasso without ZIMP using equal-tail credible intervals (Panel 3).

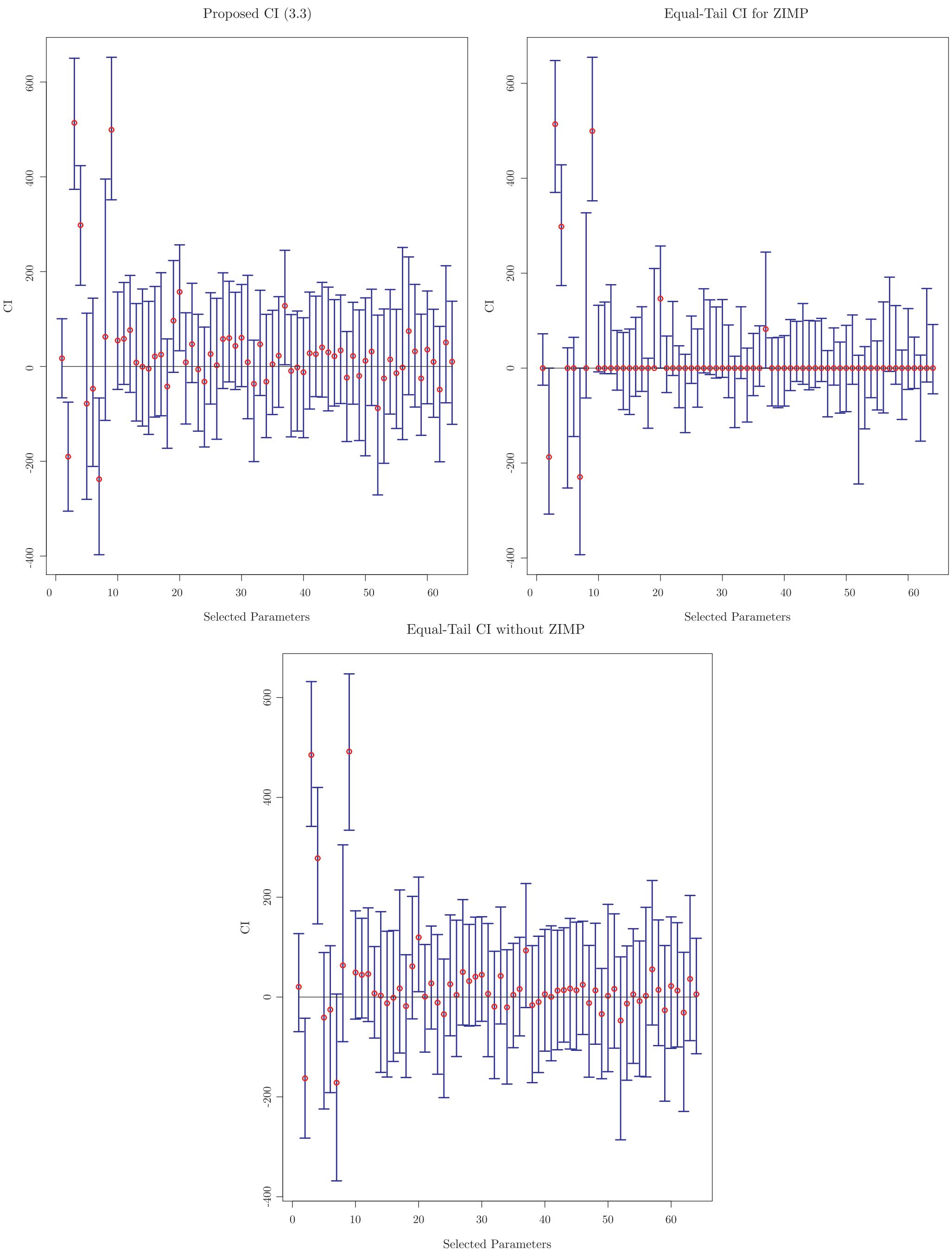

Figure 4. Confidence intervals for all the 64 parameters selected upon screening the parameters according to the marginal correlation under Bayesian lasso with ZIMP using (3.3) (Panel 1), the traditional equal-tail credible intervals (Panel 2), and under Bayesian lasso without ZIMP using equal-tail credible intervals (Panel 3).

Table 2. Confidence intervals for selected parameters of the diabetes data analysis.

	Proposed	Equal Tail	No ZIMP
Ave. length based on MTP	272.0	220.0	268.4
Ave. length based on SCR	263.5	178.6	262.9
Number of intervals excluding zero	7	3	5

5. Conclusion

We have been motivated by the need for a method of constructing confidence intervals for multiple parameters selected from the data in a Bayesian framework under ZIMP. One usually aims at maintaining *a high frequency of true coverage,* and *short length or small volume.* When these parameters are selected from the data, maintaining these properties may not be possible without accounting for the selection. A way out of this difficulty is to consider Bayes credible intervals that are immune to this selection. Under ZIMP, particularly when π_0 is high, *low frequency of enclosing zero* for those selected parameters needs to be taken into account in traditional Bayes intervals.

A decision theoretic formulation with a loss function penalizing wrong inclusions of zero has led us to a method that allows us to maintain this low frequency of enclosing zero for the selected parameters via the local fdr, the posterior probability of the parameter being null. The method turns out to have better performance than its alternatives, as we have seen through an application to the regression problem. Our proposed intervals work for any given selection rule, although we have considered just the MTP and the SCR in our numerical studies.

Acknowledgements

We thank Professor Malay Ghosh at University of Florida for his helpful ideas and comments. We also thank the referees for their constructive comments that led to a much improved article, and the Editor for the effort spent on this submission. Sarkar's research is supported by NSF Grant DMS-1006344 and DMS-1208735; Zhao's research is supported by NSF Grant DMS-1208735.

References

Benjamini, Y., Heller, R. and Yekutieli, D. (2009). Selective inference in complex research. *Philosophical Transactions of the Royal Society A. Mathematical, Physical and Engineering Sciences* **367**(1906), 4255.

Benjamini, Y. and Yekutieli, D. (2005). False discovery rate-adjusted multiple confidence intervals for selected parameters. *J. Amer. Statist. Assoc.* **100**(469), 71-93.

Casella, G. and Hwang, J. T. (1983). Empirical Bayes confidence sets for the mean of a multivariate normal distribution. *J. Amer. Statist. Assoc.* **78**(383), 688-698.

Casella, G., Hwang, J. T. and Robert, C. (1993). A paradox in decision-theoretic interval estimation. *Statist. Sinica* **3**, 141-155.

Chen, Z. and Dunson, D. B. (2003). Random effects selection in linear mixed models. *Biometrics* **59**, 762-769.

Dawid, A. P. (1994). Selection paradoxes of Bayesian inference. *Lecture Notes-Monograph Series*, 211-220.

Efron, B. (2005). Local false discovery rates. Available at: `http://statistics.stanford.edu/~ckirby/techreports/BIO/BIO%20234.pdf`.

Efron, B. (2008). Microarrays, empirical Bayes and the two-groups model. *Statist. Sci.* **23**, 1-22.

Efron, B. (2010). *Large-scale Inference, Empirical Bayes Methods for Estimation, Testing, and Prediction.* Cambridge University Press.

Efron, B. (2011). Tweedie's formula and selection bias. *J. Amer. Statist. Assoc.* **106**, 1602-1614.

Efron, B., Hastie, T., Johnstone, I. and Tibshirani, R. (2004). Least angle regression. *Ann. Statist.* **32**, 407-451.

Efron, B., Tibshirani, R., Storey, J. D. and Tusher, V. (2001). Empirical Bayes analysis of a microarray experiment. *J. Amer. Statist. Assoc.* **96**1151-1160.

Faith, R. E. (1976). Minimax Bayes point and set estimators of a multivariate normal mean. Unpublished Ph.D. dissertation, Department of Statistics, University of Michigan.

Fan, J. and Lv, J. (2008). Sure independence screening for ultrahigh dimensional feature space. *J. Roy. Statist. Soc. Ser. B* **70**, 849-911.

Giovannucci, E., Ascherio, A., Rimm, E. B., Stampfer, M. J., Colditz, G. A. and Willett, W. C. (1995). Intake of carotenoids and retino in relation to risk of prostate cancer, *J. National Cancer Institute* **87**, 1767-1776.

He, K. (1992). Parametric empirical Bayes confidence intervals based on James-Stein estimator. *Statist. Decisions* **10**, 121-132.

Hwang, J. T., Qiu, J. and Zhao, Z. (2009). Empirical Bayes confidence intervals shrinking both means and variances. *J. Roy. Statist. Soc. Ser. B* **71**, 265-285.

Hwang, J. T. and Zhao, Z. (2013). Empirical Bayes confidence intervals for selected parameters in high dimension with application to microarray data analysis. *J. Amer. Statist. Assoc.* **108**, 607-618.

Kyung, M., Gill, J., Ghosh, M. and Casella, G. (2010). Penalized regression, standard errors, and Bayesian LASSOs. *Bayesian Analysis* **5**, 369-412.

Park, T. and Casella, G. (2008). The Bayesian Lasso. *J. Amer. Statist. Assoc.* **103**, 681-686.

Qiu, J. and Hwang, J. T. (2007). Sharp simultaneous intervals for the means of selected populations with application to microarray data analysis. *Biometrics* **63**, 767-776.

Rodriguez, A., Dunson, D. B. and Taylor, J. (2009). Bayesian hierarchically weighted finite mixture models for samples of distributions. *Biostatistics* **10**, 155.

Rossouw, J. E., Anderson, G. L., Prentice, R. L. and et al. (2002). Risks and benefits of Estrogen plus Progestin in healthy postmenopausal women: principal results from the women's health initiative randomized controlled trial. *J. Amer. Statist. Assoc.* **288**, 321-333.

Sarkar, S. K., Zhou, T. and Ghosh, D. (2008). A general decision theoretic formulation of procedures controlling fdr and fnr from a Bayesian perspective. *Statist. Sinica* **18**, 925-945.

Scott, J. G. and Berger, J. O. (2006). An exploration of aspects of Bayesian multiple testing. *J. Statist. Plann. Inference* **136**, 2144-2162.

Senn, S. (2008). A note concerning a selection paradox of Dawid's. *Amer. Statist.* **62**, 206-210.

Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *J. Roy. Statist. Soc. Ser. B* **58**, 267-288.

Tibshirani, R., Saunders, M., Rosset, S., Zhu, J. and Knight, K. (2005). Sparsity and smoothness via the fused lasso. *J. Roy. Statist. Soc. Ser. B* **67**, 91-108.

Yuan, M. and Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *J. Roy. Statist. Soc. Ser. B* **68**, 49-67.

Zhao, Z. and Hwang, J. T. (2012). Empirical Bayes false coverate rate controlling confidence interval. *J. Roy. Statist. Soc. Ser. B* **74**, 871-891.

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *J. Roy. Statist. Soc. Ser. B* **67**, 301-320.

Department of Statistics, Temple University, Philadelphia, PA 19122, USA.

E-mail: zhaozhg@temple.edu

Department of Statistics, Temple University, Philadelphia, PA 19122, USA.

E-mail: sanat@temple.edu

(Received April 2013; accepted April 2014)