Biometrika (2023), **110**, 1, *pp.* 69–81
https://doi.org/10.1093/biomet/asac019
Advance Access publication 14 March 2022

On *F*-modelling-based empirical Bayes estimation of variances

BY YEIL KWON

Department of Mathematics, University of Central Arkansas, 201 Donaghey Avenue, Conway, Arkansas 72035, U.S.A.

ykwon1@uca.edu

AND ZHIGEN ZHAO

Department of Statistics, Operations, and Data Science, Temple University, 1801 Liacouras Walk, Philadelphia, Pennsylvania 19122, U.S.A.

zhaozhg@temple.edu

SUMMARY

We consider the problem of empirical Bayes estimation of multiple variances when provided with sample variances. Assuming an arbitrary prior on the variances, we derive different versions of the Bayes estimators using different loss functions. For one particular loss function, the resulting Bayes estimator relies on the marginal cumulative distribution function of the sample variances only. When replacing it with the empirical distribution function, we obtain an empirical Bayes version called the F-modelling-based empirical Bayes estimator of variances. We provide theoretical properties of this estimator, and further demonstrate its advantages through extensive simulations and real data analysis.

Some key words: Empirical distribution function; Selective inference; Uniform convergence.

1. INTRODUCTION

The empirical Bayes approach was introduced as a compound decision procedure in Robbins (1951) and has been widely studied thereafter (Dvoretzky et al., 1956; Robbins, 1956; Efron & Morris, 1972, 1973, 1975; Laird & Louis, 1987; Jiang & Zhang, 2009; Koenker & Gu, 2017). This approach plays an important role in the kinds of data analysis conducted during gene expression experiments, which often involve a large number of parallel inference problems.

The core idea of the empirical Bayes approach is to estimate the prior distribution either directly or indirectly using the available data, wherein the final inference is based on the posterior distribution when using this estimated prior. Efron (2014) classified empirical Bayes approaches as pursuing one of two strategies: (i) f-modelling, which is modelling on the data scale; and (ii) g-modelling, which is modelling on the parameter scale. Under f-modelling, the resulting empirical Bayes rule usually depends on the prior indirectly via the marginal probability density function; under g-modelling, the prior distribution is estimated and then plugged into the posterior calculation. It is further commented in that paper that the g-modelling approach has been widely used in theoretical investigations (Morris, 1983; Laird & Louis, 1987; Jiang & Zhang, 2009), whereas the f-modelling approaches are more prevalent in applications (Robbins, 1956; Brown & Greenshtein, 2009; Efron, 2011).

The simultaneous estimation of variances and the covariance matrix have a long history, dating back to James & Stein (1961). Haff (1980) provided a parametric empirical Bayes estimator of the covariance matrix by assuming an inv-Wishart prior distribution on the covariance matrix. Efron & Morris (1976) proposed an estimator to dominate the sample covariance. Wild (1980) considered simultaneous estimation of the variances under different loss functions. Robbins (1982) discussed a parametric empirical Bayes method for the scale mixture of Gaussians. Champion (2003) considered the shrinkage estimator of variances based on the Kullback–Leibler distance.

Heteroskedasticity is prevalent in many applications, such as microarray experiments, rendering the simultaneous estimation of variances even more important. There have been many attempts to estimate these parameters with different approaches (Tusher et al., 2001; Lönnstedt & Speed, 2002; Lin et al., 2003; Storey & Tibshirani, 2003; Tong & Wang, 2007; Koenker & Gu, 2017). Among these, there are a few parametric empirical Bayes estimators that are widely used. When assuming an inverse gamma prior, Smyth (2004) developed a parametric empirical Bayes estimator of the variances. Cui et al. (2005) approximated both the chi-square distribution and the inverse gamma prior by log-normal random variables and derived the exponential Lindley–James–Stein estimator. Lu & Stephens (2016) assumed that the prior of the variances follows a mixture of inverse gamma distributions to derive a flexible empirical Bayes estimator. These parametric empirical Bayes methods have the advantage of providing the full posterior distribution of the variances for further inference such as constructing credible intervals and performing hypothesis testing. Koenker & Gu (2017) took the g-modelling approach by estimating the probability density function of the prior distribution using a nonparametric maximum likelihood estimator (Kiefer & Wolfowitz, 1956; Koenker & Mizera, 2014).

In this work, we assume an arbitrary prior distribution $g(\sigma^2)$ for the variances to produce a nonparametric empirical Bayes estimator. When assuming some commonly used loss functions, we derive empirical Bayes estimators for the variances by modelling on the data scale. For a particular loss function, the resulting Bayes estimator depends only on the marginal cumulative distribution function of the sample variances, $F(s^2)$. To the best of our knowledge, this is the first estimator for the variances that relies on the marginal cumulative distribution function rather than the marginal probability density function. To differentiate our method from the terminology used in Efron (2014), we call this estimator an F-modelling-based estimator. The advantage of the F-modelling-based estimator is that one can simply replace the marginal cumulative distribution function with the empirical distribution function to obtain the proposed empirical Bayes version, which we call the F-modelling-based empirical Bayes estimator for the variances. The computation of the proposed method is instantaneous without any tuning parameters.

It is known that the empirical distribution function converges to the true distribution function uniformly (Dvoretzky et al., 1956). As shown in § 3, the proposed empirical Bayes estimator converges to the Bayes version uniformly over a set $\mathcal{D}_\delta = (0, D_\delta)$, where D_δ is a large value and tends to infinity when δ goes to zero. We impose this condition for technical reasons, so as to prevent the denominator of the Bayes estimator being arbitrarily small. It causes little practical concern because most often one would be interested in parameters corresponding to the small and moderate sample variances that fall in $\mathcal{D}_\delta$. We have also derived the estimator of the variances for postselection inference and finite Bayes inference (Efron, 2019).

2. Empirical Bayes estimator for variances

Let $\sigma^2_{[1:N]} = (\sigma^2_1, \sigma^2_2, \ldots, \sigma^2_N)$ be the parameters of interest and $s^2_{[1:N]} = (s^2_1, s^2_2, \ldots, s^2_N)$ be the corresponding sample variances. In this paper, we consider the following model:

$$\begin{aligned} s_i^2 \mid \sigma_i^2 &\overset{\text{iid}}{\sim} p(s_i^2 \mid \sigma_i^2) \sim \sigma_i^2 \frac{\chi_k^2}{k}, \\ \sigma_i^2 &\overset{\text{iid}}{\sim} g(\sigma_i^2). \end{aligned} \tag{1}$$

Here, χ^2_k denotes the random variable that follows a chi-squared distribution with k degrees of freedom. We assume an arbitrary prior $g(\sigma^2_i)$ on the variances. When integrating the variance out, the marginal probability density function of the sample variances is $f(s^2_i) = \int_0^\infty p(s^2_i \mid \sigma^2_i) g(\sigma^2_i) \mathrm{d}\sigma^2_i$. Let

$$F(s_i^2) = \int_0^{s_i^2} f(s^2) \mathrm{d}s^2$$

be the corresponding marginal cumulative distribution function of s^2_i.

To derive the Bayes rule $\hat{\sigma}^2_{[1:N]} = (\hat{\sigma}^2_1, \hat{\sigma}^2_2, \ldots, \hat{\sigma}^2_N)$, a loss function must be specified. In their 1985 University of Pittsburgh technical report, B. K. Sinha and M. Ghosh summarized many commonly used loss functions as follows:

$$\begin{aligned} L_0(\sigma^2_{[1:N]}, \hat{\sigma}^2_{[1:N]}) &= \sum_{i=1}^N (\sigma_i^2 - \hat{\sigma}_i^2)^2, \\ L_1(\sigma^2_{[1:N]}, \hat{\sigma}^2_{[1:N]}) &= \sum_{i=1}^N \left(\frac{\sigma_i^2}{\hat{\sigma}_i^2} - 1 \right)^2, \\ L_1'(\sigma^2_{[1:N]}, \hat{\sigma}^2_{[1:N]}) &= \sum_{i=1}^N \left(\frac{\hat{\sigma}_i^2}{\sigma_i^2} - 1 \right)^2, \\ L_2(\sigma^2_{[1:N]}, \hat{\sigma}^2_{[1:N]}) &= \sum_{i=1}^N \left(\frac{\hat{\sigma}_i^2}{\sigma_i^2} - \ln \frac{\hat{\sigma}_i^2}{\sigma_i^2} - 1 \right). \end{aligned}$$

The squared error loss function, $L_0(\cdot)$, is not scale invariant. The other three loss functions are scale invariant. The loss function $L'_1(\cdot)$ is used in the 1964 Stanford University PhD thesis by J. B. Selliah and Ghosh & Sinha (1987). The loss function $L'_1(\cdot)$ is equivalent to using $L_1(\cdot)$ for estimating the precision parameters (Ghosh & Sinha, 1987). The loss function $L'_1(\cdot)$ by nature favours underestimation because 'underestimation has only a finite penalty, while overestimation has an infinite penalty'(Casella & Berger, 2002). This could lead to an estimator that works extremely poorly when focusing on the parameter with the smallest sample variance. On the contrary, both the loss function $L_1(\cdot)$ and Stein's loss function $L_2(\cdot)$ have an infinite penalty for the underestimation. In addition, the loss function $L_2(\cdot)$ is tied to the Kullback–Leibler divergence and the entropy loss (Haff, 1977, 1980; Wild, 1980; Ghosh & Sinha, 1987). A potential drawback of the loss function $L_1(\cdot)$ is that it imposes a finite penalty on the overestimation.

In this article, we derive empirical Bayes estimators with respect to the scale-invariant loss functions $L_1'(\cdot)$, $L_1(\cdot)$ and $L_2(\cdot)$ by modelling on the data scale. We start with the loss function $L_1'(\cdot)$, where $\hat{\sigma}'^2_{B,[1:N]} = (\hat{\sigma}'^2_{1,B}, \hat{\sigma}'^2_{2,B}, \ldots, \hat{\sigma}'^2_{N,B})$ is the corresponding Bayes rule.

THEOREM 1. *Assume that* (1) *holds, and consider the loss function* $L_1'(\cdot)$. *Then*

$$\hat{\sigma}'^2_{i,B} = \frac{k(k-2)s_i^2 f(s_i^2) - 2ks_i^4 f'(s_i^2)}{4s_i^4 f''(s_i^2) - 4(k-2)s_i^2 f'(s_i^2) + k(k-2)f(s_i^2)}. \tag{2}$$

Equation (2) could be viewed as generalizing Tweedie's formula (Efron, 2011) to the simultaneous estimation of variances. The estimator $\hat{\sigma}'^2_{i,B}$ depends on the marginal probability density function $f(s_i^2)$, and its first and second derivatives. We can get an empirical Bayes version by replacing $f(s_i^2)$ and its derivatives with the corresponding estimators using the kernel density estimator (Brown & Greenshtein, 2009), or Lindsey's method (Efron, 2010, 2019). We call this method the f-modelling-based empirical Bayes estimator for variances:

$$\hat{\sigma}'^2_{i,\text{f-EBV}} = \frac{k(k-2)s_i^2 \widehat{f(s_i^2)} - 2ks_i^4 \widehat{f'(s_i^2)}}{4s_i^4 \widehat{f''(s_i^2)} - 4(k-2)s_i^2 \widehat{f'(s_i^2)} + k(k-2)\widehat{f(s_i^2)}}. \tag{3}$$

Next, consider Stein's loss $L_2(\cdot)$ and let $\hat{\sigma}^2_{\text{Stein},[1:N]} = (\hat{\sigma}^2_{1,\text{Stein}}, \hat{\sigma}^2_{2,\text{Stein}}, \ldots, \hat{\sigma}^2_{N,\text{Stein}})$ be the corresponding Bayes rule. Then we have the following theorem.

THEOREM 2. *Assume that* (1) *holds, and consider Stein's loss function* $L_2(\cdot)$. *Then*

$$\hat{\sigma}^2_{i,\text{Stein}} = \left\{ \frac{k-2}{ks_i^2} - \frac{2}{k}\frac{f'(s_i^2)}{f(s_i^2)} \right\}^{-1}.$$

When replacing $f(s^2)$ and $f'(s^2)$ with the corresponding estimators, we have the following f-modelling-based empirical Bayes estimator of the variances when assuming Stein's loss:

$$\hat{\sigma}^2_{i,\text{f-EBVS}} = \left\{ \frac{k-2}{ks_i^2} - \frac{2}{k}\frac{\widehat{f'(s_i^2)}}{\widehat{f(s_i^2)}} \right\}^{-1}. \tag{4}$$

When assuming Stein's loss, the empirical Bayes estimator does not require the estimation of the second derivative of the marginal probability density function. However, it still relies on the marginal density function and its first-order derivative. The nonparametric estimation of the density function and its derivatives is a challenging problem, not to mention that the estimation accuracy on the tail becomes even worse. Additionally, the commonly used approaches such as kernel density estimation relies on the choice of tuning parameters, which are difficult to choose in practice.

Next, we consider the loss function $L_1(\cdot)$ and the corresponding Bayes decision rule $\hat{\sigma}^2_{B,[1:N]} = (\hat{\sigma}^2_{1,B}, \hat{\sigma}^2_{2,B}, \ldots, \hat{\sigma}^2_{N,B})$. We have the following theorem.

THEOREM 3. *Assume that* (1) *holds, and consider the loss function* $L_1(\cdot)$. *If*

$$\int_0^\infty (s^2)^{-(k/2-2)}\, \mathrm{d}F(s^2) < \infty \quad \textit{and} \quad \int_0^\infty (s^2)^{-(k/2-1)}\, \mathrm{d}F(s^2) < \infty,$$

then

$$\hat{\sigma}^2_{i,B} = \frac{k}{2}\left\{\frac{\int_{s_i^2}^{\infty}(s^2)^{-(k/2-2)}\,\mathrm{d}F(s^2)}{\int_{s_i^2}^{\infty}(s^2)^{-(k/2-1)}\,\mathrm{d}F(s^2)} - s_i^2\right\}.$$

According to model (1), we know that

$$\int_0^{\infty}(s^2)^{-(k/2-j)}\,\mathrm{d}F(s^2)$$
$$= \int_0^{\infty}\int_0^{\infty} C_k \frac{(s^2)^{j-1}}{(\sigma^2)^{k/2}}\exp\left(-\frac{ks^2}{2\sigma^2}\right)g(\sigma^2)\,\mathrm{d}\sigma^2\,\mathrm{d}s^2, \qquad j = 1, 2,$$

where $C_k = k^{k/2}/\{\Gamma(k/2)2^{k/2}\}$. When assuming an inverse gamma prior (Smyth, 2004) and a mixture of inverse gamma priors (Lu & Stephens, 2016), basic arithmetic calculations show that the conditions in the theorem hold.

Our F-modelling approach constructs a Bayes estimator of the variances that relies on $F(s^2)$, the cumulative distribution function of the sample variances. The advantage of using an F-modelling-based estimator is that one can avoid the daunting task of estimating the marginal probability density function and its derivatives, which usually requires some kind of assumptions. Instead, to obtain an empirical Bayes version of the Bayes rule, we simply replace $F(s^2)$ with the empirical distribution function $F_N(s^2) = (1/N)\sum_i I(s_i^2 \leqslant s^2)$. After the substitution, we have the following proposed empirical Bayes estimator, which we refer to as the F-modelling-based empirical Bayes estimator of the variances:

$$\hat{\sigma}^2_{i,\text{F-EBV}} = \begin{cases} s_i^2 & \text{if } s_i^2 = \max_{1\leqslant j\leqslant N} s_j^2, \\ \dfrac{k}{2}\left\{\dfrac{\sum_{s_j^2\geqslant s_i^2}(s_j^2)^{-(k/2-2)}}{\sum_{s_j^2\geqslant s_i^2}(s_j^2)^{-(k/2-1)}} - s_i^2\right\} & \text{otherwise}. \end{cases} \tag{5}$$

The proposed estimator is calculated instantaneously and does not involve any tuning parameters.

We now return to (1) with $g(\sigma^2)$ being arbitrary. Assume that one additional sample variance s_0^2 that is independent of $s^2_{[1:N]}$ has been observed. Let σ_0^2 be the corresponding variance that is assumed to be generated from $g(\sigma^2)$ and $s_0^2 \sim \sigma_0^2\chi_k^2/k$. The goal is to estimate σ_0^2 based on the posterior distribution $\sigma_0^2 \mid s_0^2$. When N goes to infinity, the prior distribution $g(\sigma^2)$ could be fully recovered and this reduces to the standard Bayes approach. For a finite N, this problem is called the finite Bayes inference (Efron, 2019). Assume the loss function

$$L_1^{FB}(\hat{\sigma}_0^2, \sigma_0^2) = \left(\frac{\sigma_0^2}{\hat{\sigma}_0^2} - 1\right)^2. \tag{6}$$

Based on the proof of Theorem 3, we know that the Bayes rule is

$$\hat{\sigma}^2_{0,B} = \frac{k}{2}\left\{\frac{\int_{s_0^2}^{\infty}(s^2)^{-(k/2-2)}\,\mathrm{d}F(s^2)}{\int_{s_0^2}^{\infty}(s^2)^{-(k/2-1)}\,\mathrm{d}F(s^2)} - s_0^2\right\}.$$

Consequently, we propose to estimate σ_0^2 by

$$\hat{\sigma}^2_{0,\text{F-EBV}} = \begin{cases} s_0^2 & \text{if } s_0^2 \geqslant \max_{1\leqslant j\leqslant N} s_j^2, \\ \dfrac{k}{2}\left\{\dfrac{\sum_{s_j^2 \geqslant s_0^2}(s_j^2)^{-(k/2-2)}}{\sum_{s_j^2 \geqslant s_0^2}(s_j^2)^{-(k/2-1)}} - s_0^2\right\} & \text{otherwise.}\end{cases}$$

Similarly, we estimate σ_0^2 based on f-modelling methods by

$$\hat{\sigma}'^2_{0,\text{f-EBV}} = \frac{k(k-2)s_0^2\hat{f}(s_0^2) - 2ks_0^4\hat{f}'(s_0^2)}{4s_0^4\hat{f}''(s_0^2) - 4(k-2)s_0^2\hat{f}'(s_0^2) + k(k-2)\hat{f}(s_0^2)}$$

and

$$\hat{\sigma}'^2_{0,\text{f-EBVS}} = \left(\frac{k-2}{ks_0^2} - \frac{2}{k}\frac{\hat{f}'(s_0^2)}{\hat{f}(s_0^2)}\right)^{-1}.$$

We can similarly construct estimators for variances relating to a set of indices, even if the indices have been chosen using the data. Given the data $s^2_{[1:N]} = (s_1^2, s_2^2, \ldots, s_N^2)$, let $\mathcal{C}$ be the set of indices selected using a certain procedure. Our target is to estimate σ_i^2 for all $i \in \mathcal{C}$ under the loss function

$$L_1^{\mathcal{PS}}(\hat{\sigma}^2, \sigma^2) = \sum_{i\in\mathcal{C}}\left(\frac{\sigma_i^2}{\hat{\sigma}_i^2} - 1\right)^2. \tag{7}$$

As an example, we might be interested in the variances corresponding to the K smallest sample variances. In other words, order the sample variances s_i^2 increasingly as $s^2_{(1)} \leqslant s^2_{(2)} \leqslant \cdots \leqslant s^2_{(N)}$. Let $\sigma^2_{(i)}$ be the parameter corresponding to $s^2_{(i)}$. Set $\mathcal{C} = \{i\colon s_i^2 \leqslant s^2_{(K)}\}$.

For any $i \in \mathcal{C}$,

$$\pi(\sigma_i^2 \mid s^2_{[1:N]}, i \in \mathcal{C}) = \pi(\sigma_i^2 \mid s^2_{[1:N]}).$$

This implies that the posterior distribution of σ_i^2 when conditioning on both the data and the selection set is the same as the posterior distribution of σ^2 conditioning on the data. Consequently, the Bayes rule based on the selection remains the same and it is immune to the selection (Dawid, 1994). We therefore propose to estimate $\sigma_i^2, i \in \mathcal{C}$, according to (5) without adjustment. This argument is true because the full dataset is available for the postselection inference. Otherwise, the Bayes rule might be affected by the selection. For instance, if only the data post the selection are available for further inference, then the Bayes rule needs to be corrected for such a selection rule. See Yekutieli (2012) for a full discussion on this issue.

3. THEORETICAL PROPERTIES

In this section, we study the theoretical properties of the proposed method. To ease our notation, we define two functions $l_1(s^2, u) = (s^2)^{-(k/2-2)}\mathbb{I}(s^2 \geqslant u)$ and $l_2(s^2, u) = (s^2)^{-(k/2-1)}\mathbb{I}(s^2 \geqslant u)$, where $\mathbb{I}(\cdot)$ is an indicator function. Then the Bayes decision rule and the proposed method can be respectively written as

$$\hat{\sigma}^2_{i,B} = \frac{k}{2}\left\{\frac{\int_0^\infty l_1(s^2, s_i^2)\,\mathrm{d}F(s^2)}{\int_0^\infty l_2(s^2, s_i^2)\,\mathrm{d}F(s^2)} - s_i^2\right\} \quad \text{and} \quad \hat{\sigma}^2_{i,\text{F-EBV}} = \frac{k}{2}\left\{\frac{\int_0^\infty l_1(s^2, s_i^2)\,\mathrm{d}F_N(s^2)}{\int_0^\infty l_2(s^2, s_i^2)\,\mathrm{d}F_N(s^2)} - s_i^2\right\}.$$

First, we study the numerator and denominator separately.

THEOREM 4. *Assume that the same conditions as in Theorem 3 hold and that $F(s^2)$ is continuous with the support of $(0,\infty)$. Then*

$$\sup_u \left| \int_0^\infty l_1(s^2, u)\,\mathrm{d}F_N(s^2) - \int_0^\infty l_1(s^2, u)\,\mathrm{d}F(s^2)\right| \to 0 \quad \textit{almost surely}$$

and

$$\sup_u \left| \int_0^\infty l_2(s^2, u)\,\mathrm{d}F_N(s^2) - \int_0^\infty l_2(s^2, u)\,\mathrm{d}F(s^2)\right| \to 0 \quad \textit{almost surely.}$$

This theorem implies that both the numerator and the denominator of the proposed empirical Bayes estimator converge to those of the Bayes rule uniformly. However, it does not guarantee that the ratio converges uniformly. The reason is that the denominator $\int_0^\infty l_2(s^2, u)\,\mathrm{d}F(s^2)$ converges to zero when u goes to ∞. To prove that the proposed method converges to the Bayes estimator uniformly, we consider the set such that the denominator of the Bayes rule is greater than some positive number. Namely, for a number $\delta > 0$, let $\mathcal{D}^\delta$ be a set defined as

$$\mathcal{D}^\delta \equiv \left\{ u \;\middle|\; \int_u^\infty (s^2)^{-(k/2-1)}\,\mathrm{d}F(s^2) > \delta \right\}.$$

Since $\int_0^\infty (s^2)^{-(k/2-1)}\,\mathrm{d}F(s^2) < \infty$, then $\mathcal{D}^\delta = (0, D_\delta)$ for some positive number D_δ. We then have the following theorem.

THEOREM 5. *Assume that the same conditions as in Theorem 4 hold. Then*

$$\sup_{s_i^2 \in \mathcal{D}^\delta} |\hat{\sigma}^2_{i,\text{F-EBV}} - \hat{\sigma}^2_{i,B}| \to 0 \quad \textit{almost surely.}$$

The constant D_δ is a quantity that depends on the marginal distribution function of the sample variances only and D_δ tends to infinity when δ tends to 0. For any $0 < \tau < 1$, let $s^2_{[1:N]}$ be a random sample consisting of N sample variances. Let s^2_τ be the τth sample quantile. We can always choose sufficiently small δ such that $\{s_i^2, s_i^2 \leqslant s_\tau^2\} \in \mathcal{D}^\delta$ with large probability. For a sample variance that does not fall in $\mathcal{D}^\delta$, one could estimate the corresponding parameter with these sample variances. Namely, we could modify the proposed estimator to

$$\hat{\sigma}^2_{i,\text{mF-EBV}} = \begin{cases} s_i^2 & \text{if } s_i^2 \geqslant s^2_{(\lceil N\tau \rceil)}, \\ \dfrac{k}{2}\left\{ \dfrac{\sum_{s_j^2 \geqslant s_i^2} (s_j^2)^{-(k/2-2)}}{\sum_{s_j^2 \geqslant s_i^2} (s_j^2)^{-(k/2-1)}} - s_i^2 \right\} & \text{otherwise.} \end{cases}$$

In practice, especially when focusing on parameters with small sample variances, this modification does not make much difference.

We can extend the result to the postselection inference and finite Bayes inference.

Corollary 1. *Assume that the same conditions as in Theorem 4 hold. Then*

$$\sup_{s_i^2 \in \mathcal{D}^\delta, i \in \mathcal{C}} |\hat{\sigma}^2_{i,\text{F-EBV}} - \hat{\sigma}^2_{i,B}| \to 0 \quad \textit{almost surely.}$$

As commented in § 2, the Bayes estimator is immune to the selection rule $\mathcal{C}$, and the empirical Bayes estimator could be a good approximation of the Bayes estimator. However, the discrepancy between these two widens when focusing on the selected case (Pan et al., 2017), and some correction is needed (Hwang & Zhao, 2013). On the other hand, Corollary 1 indicates that the proposed F-modelling-based empirical Bayes estimator converges to the corresponding Bayes version if $s_i^2 \in \mathcal{D}^\delta, i \in \mathcal{C}$. In other words, we do not need to make further corrections for the selection.

Similarly, when considering the finite Bayes inference, the uniform convergence of the proposed estimator guarantees a good estimation as long as $s_0^2 \in \mathcal{D}^\delta$.

Corollary 2. *Assume that the same conditions as in Theorem 4 hold. Then*

$$\sup_{s_0{}^2 \in \mathcal{D}^\delta} |\hat{\sigma}^2_{0,\text{F-EBV}} - \hat{\sigma}^2_{0,B}| \to 0 \quad \textit{almost surely.}$$

4. Numerical studies

In this section, we compare the numerical performances of the proposed methods with existing methods, including the sample variance (s^2), the exponential Lindley–James–Stein, ELJS, estimator (Cui et al., 2005), the method of Tong & Wang (2007), TW, the method of Smyth (2004), the variance adaptive shrinkage, vash method (Lu & Stephens, 2016) and the REBayes method (Koenker & Gu, 2017). As suggested by a referee, we consider two more estimators based on the Smyth method and variance adaptive shrinkage method by considering the loss function $L_1(\cdot)$. Assume that the prior distribution $g(\sigma_i^2)$ in (1) is inverse gamma (a_0, b_0). Then the posterior distribution of σ_i^2 is inverse gamma (a_1, b_1), where $a_1 = a_0 + k/2$, $b_1 = b_0 + ks_i^2/2$. The hyperparameters a_0 and b_0 are estimated by using the method of moments (Smyth, 2004). The Smyth method, which minimizes $EL_1'(\cdot)$, is given as b_1/a_1. The modified Smyth method, which minimizes $EL_1(\cdot)$, is given as

$$\sigma^2_{i,\text{mSmyth}} = \frac{E(\sigma_i^4 \mid s_i^2)}{E(\sigma_i^2 \mid s_i^2)} = \frac{b_1}{a_1 - 2}.$$

Similarly, we include two versions of variance adaptive shrinkage estimators: the original version, vash, and the modified version, mvash, in our simulation studies.

Let $(\sigma_i^2, s_i^2), i = 1, 2, \ldots, N$ be the parameters, and let the sample variances be generated according to (1), where the degree of freedom k is chosen as 5 and the prior $g(\sigma^2)$ is chosen from the following settings.

Setting I. Let $\sigma_i^2 \sim$ inverse gamma distribution: $\textsc{ig}(a, 1)$ with $a = 10$ and 6.

Table 1. *The* $\log_{10}(risk)$ *associated with the loss function (7) of the different estimators for the variances under different simulation settings. For each setting, we consider three selection rules: (i) the parameters corresponding to the 1% smallest sample variances, (ii) the parameters corresponding to the 5% smallest sample variances and (iii) all the parameters*

Setting	a	Selection rule	s^2	ELJS	TW	Smyth	mSmyth	vash	mvash	REBayes	Proposed
		1%	2.60	−0.48	−0.72	−0.90	−1.06	−0.87	−1.06	−0.65	−1.06
I	10	5%	2.00	−0.70	−0.87	−0.89	−1.05	−0.88	−1.05	−0.92	−1.05
		All	0.77	−0.94	−0.98	−0.91	−1.05	−0.92	−1.05	−0.97	−1.03
		1%	2.34	1.05	0.45	−0.14	−0.21	0.87	−0.10	−0.05	−0.22
II	10	5%	1.79	0.62	0.17	−0.10	−0.20	0.74	−0.11	−0.06	−0.22
		All	0.75	0.01	0.00	0.14	−0.43	0.26	−0.48	−0.38	−0.52
		1%	2.22	1.15	0.88	−0.28	−0.48	−0.26	−0.49	−0.50	−0.60
III	4	5%	1.72	0.74	0.53	−0.06	−0.36	−0.05	−0.37	−0.22	−0.39
		All	0.69	0.10	0.16	0.26	−0.35	0.27	−0.35	−0.32	−0.58
		1%	2.28	1.26	0.97	−0.08	−0.28	−0.06	−0.28	−0.13	−0.28
IV	4	5%	1.73	0.77	0.53	−0.13	−0.28	−0.11	−0.29	−0.22	−0.32
		All	0.72	0.14	0.20	0.29	−0.34	0.30	−0.34	−0.30	−0.56

Setting II. Let $\sigma_i^2 \sim$ mixture of inverse gamma distributions: $0.2\textsc{ig}(a, 1) + 0.4\textsc{ig}(8, 6) + 0.4\textsc{ig}(9, 19)$, where $a = 10$ and 6.

Setting III. Let $\sigma_i^2 = a$ with 0.4 probability and $1/a$ with 0.6 probability, where $a = 3$ and 4.

Setting IV. Let $\sigma_i^2 \sim$ mixture of inverse Gaussian distributions: $0.4\text{InvGauss}(1/a, 1) + 0.6\text{InvGauss}(a, a^4)$, where $a = 3$ and 4.

For all simulations, we set $N = 1000$ and the number of replications as 500. For each replication, we generate the data (σ_i^2, s_i^2) and order them according to the sample variances increasing. We consider three different selection rules: (i) the parameters corresponding to the 1% smallest sample variances, (ii) the parameters corresponding to the 5% smallest sample variances and (iii) all the parameters. We calculate the estimated values based on the aforementioned methods. The risks associated with the loss function (7) are calculated and reported in Table 1 and the Supplementary Material. In our numerical studies, it is shown that the two f-modelling estimators defined in (3) and (4) perform poorly, and the results are not reported in the tables. The proposed F-modelling-based empirical Bayes estimator performs the best among all the estimators considered. The modified Smyth method and modified variance adaptive shrinkage method perform similarly under these settings. Under Setting I, when the prior of the variance is an inverse gamma distribution, the proposed method, the modified Smyth method and modified variance adaptive shrinkage method are essentially the same. However, for Settings II to IV, when the prior distribution is not an inverse gamma distribution, the proposed method outperforms all other competing methods, including the modified Smyth method and the modified variance adaptive shrinkage method.

Next, we consider the finite Bayes inference problem. Namely, for each generated dataset $s^2_{[1:N]}$ and a new observation s_0^2, we calculate the estimated values based on different approaches and calculate the risk according to the loss function (6). The risks are reported in Table 2 and the Supplementary Material. Overall, the proposed F-modelling-based empirical Bayes estimator performs the best among all the estimators considered. The modified Smyth method and modified variance adaptive shrinkage method are essentially the same. Under Setting I, when the prior of the variance is an inverse gamma distribution, the proposed method, the modified Smyth method and modified variance adaptive shrinkage method perform similarly with negligible differences.

Table 2. *The* $\log_{10}$*(risk) associated with the loss function (6) of the different estimators for the finite Bayes inference problem*

Setting	(a)	s^2	ELJS	TW	Smyth	mSmyth	vash	mvash	REBayes	Proposed
I	10	0.38	0.16	−1.05	−0.96	−1.06	−0.96	−1.07	−1.02	−1.03
II	10	0.36	0.14	−0.11	0.01	−0.48	−0.02	−0.50	−0.51	−0.55
III	4	0.92	0.72	0.23	0.23	−0.36	0.25	−0.36	−0.31	−0.47
IV	4	0.70	0.49	0.25	0.37	−0.30	0.38	−0.29	−0.10	−0.51

However, for Settings II to IV, when the prior distribution is not an inverse gamma distribution, the proposed method outperforms all other competing methods.

5. Real data analysis

In this section, we apply different variance estimators to two microarray datasets: colon cancer (Alon et al., 1999) and leukemia data (Golub et al., 1999). The colon cancer data include the expressions of genes ($N = 2000$) for 22 patients and 40 healthy people. The leukemia data include the expressions of genes ($N = 7128$) extracted from 72 patients with two types of leukemia: acute lymphoblastic leukemia (47 patients) and acute myeloid leukemia (25 patients). For the leukemia dataset, we first randomly split the subjects into two subgroups, such that both subgroups contain similar numbers of subjects from the acute lymphoblastic leukemia patients and acute myeloid leukemia patients. For each subgroup, we then constructed $1-\gamma$ ($\gamma = 0.05$) confidence intervals for θ_i, the mean parameter of the ith gene, following the work of Hwang et al. (2009) by considering

$$\mathrm{CI}_i = \hat{\theta}_i \pm \sqrt{\hat{M}_i\hat{\sigma}_i^2} \cdot \sqrt{z_{\gamma/2}^2 - \log \hat{M}_i}, \qquad \hat{\theta}_i = \hat{M}_i X_i + (1-\hat{M}_i)\bar{X}, \qquad \hat{M}_i = \frac{\hat{\tau}^2}{\hat{\sigma}_i^2 + \hat{\tau}^2},$$

and

$$\hat{\tau}^2 = \max\left\{\frac{1}{N}\sum_{i=1}^{N}(X_i - \hat{\mu})^2 - \frac{1}{N}\sum_{i=1}^{N}\sigma_i^2, \tau_0^2\right\}.$$

We declare the ith gene, where $i = 1, 2, \ldots, N$, to be significant if the corresponding interval does not enclose zero. We do the same for the other subgroup. We call the decision of the ith gene discordant if the interval based on the first subgroup does (does not) enclose zero while the interval based on the second subgroup does not (does) enclose zero. If a decision is discordant, this implies that a significant conclusion based on one subgroup cannot be replicated by the other. We repeat these steps 500 times to calculate the average proportions of discordant decisions. We perform the same calculation for the colon cancer data by splitting the data into the patient group and healthy group.

In Fig. 1, we plot the box plots of the rate of discordant decisions. The average percentage of discordant decisions are reported in Table 3. It is seen that the proposed method, the modified Smyth and modified variance adaptive shrinkage estimator produce a similar number of discordance decisions. This number is substantially smaller than all the other competing methods.

To further investigate why these three methods perform similarly, we test the hypothesis that the distribution of the sample variances is the convolution of a scaled chi-square distribution and

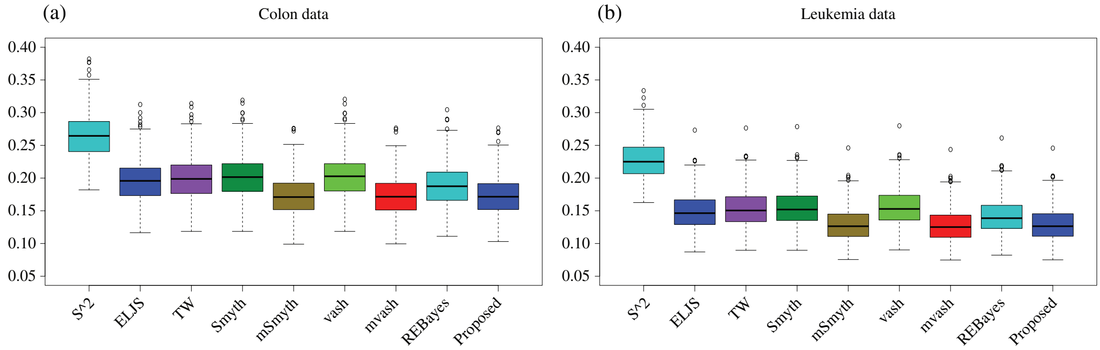

Fig. 1. Box plots of the percentage of discordant decisions for (a) colon cancer data and (b) leukemia data based on 500 replications.

Table 3. *The average percentages of discordant decisions of different intervals when applied to the colon cancer data and leukemia data based on 500 replications*

Data	s^2	ELJS	TW	Smyth	mSmyth	vash	mvash	REBayes	Proposed
Colon	0.27	0.20	0.20	0.20	0.17	0.20	0.17	0.19	0.17
Lukemia	0.23	0.15	0.15	0.15	0.13	0.16	0.13	0.14	0.13

an inverse gamma distribution. The Kolmogorov–Smirnov test statistics for the colon dataset and leukemia dataset are 0.014 and 0.017, respectively. The resulting p-values are 0.80 and 0.031, respectively. In other words, there is no evidence to reject the null hypothesis that states that the prior is an inverse gamma distribution for the colon data, and there is only moderate evidence to reject the null hypothesis for the leukemia data. We expect to see similar performances for these three methods.

The code for the simulations and real data analysis is available on github: `https://github.com/zhaozhg81/FEBV`.

6. Conclusion

The proposed method is developed under (1), assuming a scaled chi-square distribution with equal degrees of freedom. The Bayes estimator in Theorem 4 still applies when the degrees of freedom are different. However, the estimation of the cumulative distribution function requires that the sample variances are identically distributed. Therefore, the proposed method could not be directly applied to cases with unequal degrees of freedom. In practice, we take a slightly conservative approach by considering the smallest degrees of freedom as the common one. Many parametric empirical Bayesian approaches based on g-modelling estimate the prior distribution explicitly and can handle unequal degrees of freedom.

In the real data analysis, we use the estimator of the variances as a plug-in estimator for inferring the mean parameters. One natural follow-up challenge to address is how to obtain a nonparametric empirical Bayes estimator of the means, assuming arbitrary priors for both the means and the variances. Given the observed advantages of the F-modelling-based approach, we would like to further extend this framework to broader settings in future research. We will further study the properties of the F-modelling-based approach under the decision theoretical framework.

ACKNOWLEDGEMENT

The authors were supported by the National Science Foundation. The authors thank the associate editor and the reviewers for comments that substantially helped improve the quality of the paper, and Mr. Matthew P. MacNaughton for editing the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material includes technical details and additional simulations.

REFERENCES

ALON, U., BARKAI, N., NOTTERMAN, D. A., GISH, K., YBARRA, S., MACK, D. & LEVINE, A. J. (1999). Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. In *Proc. Nat. Acad. Sci.*, vol. 96. United States: The National Academy of Sciences, pp. 6745–50.

BROWN, L. D. & GREENSHTEIN, E. (2009). Nonparametric empirical Bayes and compound decision approaches to estimation of a high-dimensional vector of normal means. *Ann. Statist.* **37**, 1685–704.

CASELLA, G. & BERGER, R. (2002). *Statistical Inference*, 2nd ed. Boston, MA: Cengage Learning.

CHAMPION, C. J. (2003). Empirical Bayesian estimation of normal variances and covariances. *J. Mult. Anal.* **87**, 60–79.

CUI, X., HWANG, J. T., QIU, J., BLADES, N. J. & CHURCHILL, G. A. (2005). Improved statistical tests for differential gene expression by shrinking variance components estimates. *Biostatistics* **6**, 59–75.

DAWID, A. P. (1994). Selection paradoxes of Bayesian inference. In *Institute of Mathematical Statistics Lecture Notes - Monograph Series*, vol. 24, T. W. Anderson, K. T. Fang & I. Olkin, eds. Hayward, CA: Institute of Mathematical Statistics, pp. 211–20.

DVORETZKY, A., KIEFER, J. & WOLFOWITZ, J. (1956). Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *Ann. Math. Statist.* **27**, 642–69.

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

EFRON, B. (2011). Tweedie's formula and selection bias. *J. Am. Statist. Assoc.* **106**, 1602–14.

EFRON, B. (2014). Two modeling strategies for empirical Bayes estimation. *Statist. Sci.* **29**, 285–301.

EFRON, B. (2019). Bayes, oracle Bayes and empirical Bayes. *Statist. Sci.* **34**, 177–201.

EFRON, B. & MORRIS, C. N. (1972). Limiting the risk of Bayes and empirical Bayes estimators. II. The empirical Bayes case. *J. Am. Statist. Assoc.* **67**, 130–9.

EFRON, B. & MORRIS, C. N. (1973). Stein's estimation rule and its competitors—an empirical Bayes approach. *J. Am. Statist. Assoc.* **68**, 117–30.

EFRON, B. & MORRIS, C. N. (1975). Data analysis using Stein's estimator and its generalizations. *J. Am. Statist. Assoc.* **70**, 311–19.

EFRON, B. & MORRIS, C. (1976). Multivariate empirical Bayes and estimation of covariance matrices. *Ann. Statist.* **4**, 22–32.

GHOSH, M. & SINHA, B. K. (1987). Inadmissibility of the best equivariant estimators of the variance-covariance matrix, the precision matrix, and the generalized variance under entropy loss. *Statist. Risk Modeling* **5**, 201–28.

GOLUB, T. R., SLONIM, D. K., TAMAYO, P., HUARD, C., GAASENBEEK, M., MESIROV, J. P., COLLER, H., LOH, M. L., DOWNING, J. R., CALIGIURI, M. A. et al. (1999). Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* **286**, 531–7.

HAFF, L. (1977). Minimax estimators for a multinormal precision matrix. *J. Mult. Anal.* **7**, 374–85.

HAFF, L. (1980). Empirical Bayes estimation of the multivariate normal covariance matrix. *Ann. Statist.* **8**, 586–97.

HWANG, J. T., QIU, J. & ZHAO, Z. (2009). Empirical Bayes confidence intervals shrinking both means and variances. *J. R. Statist. Soc.* B **71**, 265–85.

HWANG, J. T. & ZHAO, Z. (2013). Empirical Bayes confidence intervals for selected parameters in high dimension with application to microarray data analysis. *J. Am. Statist. Assoc.* **108**, 607–18.

JAMES, W. & STEIN, C. (1961). Estimation with quadratic loss. In *Proc. 4th Berkeley Symp. Math. Statist. Prob.*, vol. 4. Berkeley, CA: University of California Press, pp. 361–79.

JIANG, W. & ZHANG, C. H. (2009). General maximum likelihood empirical Bayes estimation of normal means. *Ann. Statist.* **37**, 1647–84.

KIEFER, J. & WOLFOWITZ, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *Ann. Math. Statist.* **27**, 887–906.

KOENKER, R. & GU, J. (2017). REBayes: an R package for empirical Bayes mixture methods. *J. Statist. Software* **82**, 1–26.

KOENKER, R. & MIZERA, I. (2014). Convex optimization, shape constraints, compound decisions, and empirical Bayes rules. *J. Am. Statist. Assoc.* **109**, 674–85.

LAIRD, N. M. & LOUIS, T. A. (1987). Empirical Bayes confidence intervals based on bootstrap samples. *J. Am. Statist. Assoc.* **82**, 739–57.

LIN, Y., NADLER, S. T., ATTIE, A. D. & YANDELL, B. S. (2003). Adaptive gene picking with microarray data: detecting important low abundance signals. In *The Analysis of Gene Expression Data: Statistics for Biology and Health*, G. Parmigiani, E. S. Garrett, R. A. Irizarry & S. L. Zeger, eds. New York: Springer, pp. 291–312.

LÖNNSTEDT, I. & SPEED, T. (2002). Replicated microarray data. *Statist. Sinica* **12**, 31–46.

LU, M. & STEPHENS, M. (2016). Variance adaptive shrinkage (vash): flexible empirical Bayes estimation of variances. *Bioinformatics* **32**, 3428–34.

MORRIS, C. N. (1983). Parametric empirical Bayes confidence intervals. In *Scientific Inference, Data Analysis, and robustness*, G. E. P. Box, T. Leonard & C.-F. Wu, eds. New York: Academic Press, pp. 25–50.

PAN, J., HUANG, Y. & HWANG, J. G. (2017). Estimation of selected parameters. *Comp. Statist. Data Anal.* **109**, 45–63.

ROBBINS, H. (1951). Asymptotically subminimax solutions of compound statistical decision problems. In *Proc. 2nd Berkeley Symp. Math. Statist. Prob.*, vol. 2. Berkeley, CA: University of California Press, pp. 131–49.

ROBBINS, H. (1956). An empirical Bayes approach to statistics. In *Proc. 3rd Berkeley Symp. Math. Statist. Prob.*, vol. 3. Berkeley, CA: University of California Press, pp. 157–63.

ROBBINS, H. (1982). Estimating many variances. In *Statistical Decision Theory and Related Topics III*, S. S. Gupta, ed. New York: Academic Press, pp. 251–61.

SMYTH, G. K. (2004). Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Statist. Appl. Genet. Molec. Biol.* **3**, doi: 10.2202/1544-6115.1027.

STOREY, J. & TIBSHIRANI, R. (2003). SAM thresholding and false discovery rates for detecting differential gene expression in DNA microarrays. In *The Analysis of Gene Expression Data: Methods and Software*, Parmigiani, G., Garrett, E. S., Irizarry, R. A., & Zeger, S. L. eds. New York: Springer, pp. 272–90.

TONG, T. & WANG, Y. (2007). Optimal shrinkage estimation of variances with applications to microarray data analysis. *J. Am. Statist. Assoc.* **102**, 113–22.

TUSHER, V. G., TIBSHIRANI, R. & CHU, G. (2001). Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Nat. Acad. Sci.* **98**, 5116–21.

WILD, C. (1980). Loss functions and admissibility of normal variance estimators. *Can. J. Statist.* **8**, 95–101.

YEKUTIELI, D. (2012). Adjusted Bayesian inference for selected parameters. *J. R. Statist. Soc.* B **74**, 515–41.

[*Received on* 6 *November* 2019. *Editorial decision on* 21 *February* 2022]