Electronic Journal of Statistics
Vol. 14 (2020) 2272–2297
ISSN: 1935-7524
https://doi.org/10.1214/20-EJS1717

On general maximum likelihood empirical Bayes estimation of heteroscedastic IID normal means

Wenhua Jiang

School of Mathematical Sciences, Fudan University and Shanghai Center for Mathematical Sciences, 220 Handan Road, Shanghai 200433, China
e-mail: wenhuajiang@yahoo.com

Abstract: We propose a general maximum likelihood empirical Bayes (GMLEB) method for the heteroscedastic normal means estimation with known variances. The idea is to plug the generalized maximum likelihood estimator in the oracle Bayes rule. From the point of view of restricted empirical Bayes, the general empirical Bayes aims at a benchmark risk smaller than the linear empirical Bayes methods when the unknown means are i.i.d. variables. We prove an oracle inequality which states that under mild conditions, the regret of the GMLEB is of smaller order than $(\log n)^5/n$. The proof is based on a large deviation inequality for the generalized maximum likelihood estimator. The oracle inequality leads to the property that the GMLEB is adaptive minimax in L_p balls when the order of the norm of the ball is larger than $((\log n)^{5/2}/\sqrt{n})^{1/(p\wedge 2)}$. We demonstrate the superb risk performance of the GMLEB through simulation experiments.

MSC 2010 subject classifications: 62C12, 62G05, 62G20.
Keywords and phrases: Empirical Bayes, heteroscedasticity, generalized MLE, adaptive minimaxity.

Received July 2019.

Contents

1. Introduction

In this paper we consider empirical Bayes for heteroscedastic data:

$$X_i|(\theta_i, \sigma_i^2) \overset{\text{ind}}{\sim} N(\theta_i, \sigma_i^2), \quad i = 1, \dots, n, \tag{1.1}$$

where σ_i^2 are known. The problem is to estimate $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n)$ under the average squared loss

$$L_n(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}) = n^{-1}\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|^2 = n^{-1}\sum_{i=1}^{n}(\hat{\theta}_i - \theta_i)^2. \tag{1.2}$$

This problem has been considered by many in the literature, including recent studies by [18] and [17]. However, while the existing studies are typically based on the shrinkage approach, our focus is on the general empirical Bayes [13, 15], or equivalently nonparametric empirical Bayes [12].

In general empirical Bayes, the unknowns θ_i are typically treated as constants in the compound approach [13]. In a homoscedastic compound decision problem, the average risk is written as

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_{\theta_i,\sigma}\big(t(X_i) - \theta_i\big)^2 = \int\left[\int \big(t(x) - \theta\big)^2 f(x|\theta,\sigma)dx\right]dG_n(\theta), \tag{1.3}$$

where $f(x|\theta, \sigma)$ is the density of $N(\theta, \sigma^2)$, and G_n is the empirical distribution of θ_i. Robbins [13, 14] observed that the optimal solution of the above problem is the Bayes rule $t^*_{G_n,\sigma}(x) = \mathbb{E}_{G_n}(\theta|X = x, \sigma)$. This can be viewed as fundamental theorem of compound decisions as it connects the compound problem to the Bayes approach. The idea is to plug-in estimated G_n to mimic the Bayes rule or its performance. In the presence of heteroscedasticity, the same calculation as in (1.3) will not go through as $X_i - \theta_i$ do not have the same distribution. In the heteroscedastic case with known σ_i, we may write

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_{\theta_i,\sigma_i}\big(t(X_i,\sigma_i) - \theta_i\big)^2 = \int\left[\int \big(t(x,\sigma) - \theta\big)^2 f(x|\theta,\sigma)dx\right]dG_n(\theta,\sigma), \tag{1.4}$$

where G_n is the empirical distribution of (θ_i, σ_i). This still connects the compound problem to Bayes. However, the fundamental theorem fails in the presence of heteroscedasticity with observable σ_i in general as the meaning and implication of putting a known quantity in the prior G_n is unclear. Moreover, there may not be sufficient sample size at each σ-value to allow sufficiently accurate estimation of a nonparametric unknown prior.

One plausible way is to take empirical Bayes view that θ_i are i.i.d. variables with an unknown common prior G. Empirical Bayes methods can be understood from the point of view of restricted empirical Bayes. Given a class of decision functions $\mathscr{D}$, with oracular knowledge of G, the oracle benchmark is $R_{\mathscr{D}}(G) =$

$\inf_{t\in\mathscr{D}} n^{-1}\mathbb{E}_G \sum_{i=1}^n (t(X_i,\sigma_i)-\theta_i)^2$. The regret of an estimator $\hat{t}_n$ is

$$r_{G,\mathscr{D}}(\hat{t}_n) = \frac{1}{n}\mathbb{E}_G \sum_{i=1}^n \big(\hat{t}_n(X_i,\sigma_i)-\theta_i\big)^2 - R_{\mathscr{D}}(G). \tag{1.5}$$

The aim of restricted empirical Bayes is to seek $\hat{t}_n \in \mathscr{D}$ satisfying the asymptotic optimality

$$r_{G,\mathscr{D}}(\hat{t}_n) \to 0, \quad \text{as } n\to\infty. \tag{1.6}$$

Let G be a normal distribution with mean μ and variance τ^2. With $\mathscr{D}$ being the class of all linear estimators, the optimal estimator in $\mathscr{D}$ is $t^*_{\mathscr{D}}(x) = \mu + (1-B)(x-\mu)$ where $B = \sigma^2/(\sigma^2+\tau^2)$. In the homoscedastic case, $\sigma_i^2 \equiv \sigma^2$, the James-Stein estimator $\hat{\theta}_i^{\mathrm{JS}} = \bar{X} + (1-B_n)(X_i - \bar{X})$ with $B_n = (n-3)\sigma^2/\sum_i (X_i - \bar{X})^2$ approximates the optimal linear rule $t^*_{\mathscr{D}}(x)$ in the sense of (1.6). In the heteroscedastic case, Xie, Kou and Brown [18] proposed to select an estimator from the class $\{\tau^2 X_i/(\sigma_i^2+\tau^2) + \sigma_i^2\mu/(\sigma_i^2+\tau^2)\colon \mu\in\mathbb{R}, \tau^2>0\}$. The parameters μ and τ^2 are estimated by minimizing a Stein's unbiased risk estimate (SURE) function. Xie, Kou and Brown [18] also suggested a semiparametric shrinkage estimator of the form $(1-b_i)X_i + b_i\mu$ where b_i is nondecreasing in σ_i^2. Both SURE estimators satisfy the asymptotic optimality (1.6). Since $\sigma_i^2/(\sigma_i^2+\tau^2)$ is monotone increasing in σ_i^2, any estimator of the previous form is also of the latter form. Hence, the semiparametric SURE aims at a smaller benchmark risk than the parametric SURE.

Denote the density of the normal location mixture by distribution G with scale σ by

$$f_{G,\sigma}(x) = \int \frac{1}{\sigma}\varphi\Big(\frac{x-u}{\sigma}\Big) dG(u), \tag{1.7}$$

where $\varphi(x)$ is the standard normal density. It is well known that for any prior G, the Bayes rule is given by Tweedie's formula [14, 1, 4]

$$t^*_G(X_i,\sigma_i) = \mathbb{E}_G(\theta_i|X_i,\sigma_i) = X_i + \sigma_i^2 \frac{f'_{G,\sigma_i}(X_i)}{f_{G,\sigma_i}(X_i)}, \tag{1.8}$$

where $f_{G,\sigma}(x)$ is as in (1.7). The Bayes risk under (1.2) is

$$R^*_n(G) = n^{-1}\sum_{i=1}^n R^*_{\sigma_i}(G), \tag{1.9}$$

where $R^*_\sigma(G) = \sigma^2\{1 - \sigma^2\int (f'_{G,\sigma}/f_{G,\sigma})^2 f_{G,\sigma}\}$ is the Bayes risk for univariate estimation. The general empirical Bayes approach assumes no knowledge about the unknown prior G but still aims to mimic the Bayes rule $t^*_G(\cdot,\sigma_i)$ in (1.8) or approximately achieve the risk benchmark $R^*_n(G)$. Compared with the parametric and semiparametric methods, the general empirical Bayes is greedier since it aims at the optimal estimator among all the rules. There are two main strategies to approximate the Bayes rule in (1.8): modeling on the θ space, called "g-modeling", and modeling on the x space, called "f-modeling". Efron

[5] provided examples and summarized some advantages of both strategies. As demonstrated in [7] and [10], compound decision problem is a favorable case for nonparametric g-modeling. Nonparametric g-modeling refers to estimating the unknown prior by the generalized MLE [10]

$$\widehat{G}_n = \arg\max_{G\in\mathscr{G}} \prod_{i=1}^{n} f_{G,\sigma_i}(X_i), \tag{1.10}$$

where $f_{G,\sigma}(x)$ is the mixture density as in (1.7) and $\mathscr{G}$ is the family of all distribution functions. The calculation of the generalized MLE is usually difficult. Recently, Koenker and Mizera [10] proposed a convex optimization approach to computing the generalized MLE, which is proven to be efficient and accurate. The heteroscedastic option in the REBayes package [9] facilitates our research. Fu, James and Sun [6] also considered the general empirical Bayes method for the heteroscedastic normal mean problem (1.1)–(1.2) with i.i.d. θ_i. They suggested an f-modeling procedure to mimic the Bayes rule in (1.8) and proved its optimality in the sense (1.6). Still, the heart of the question is whether the gain by aiming at the smaller benchmark risk is large enough to offset the additional cost of the nonparametric estimation. Our results affirm that when θ_i are drawn from a common prior G, the proposed general maximum likelihood empirical Bayes (GMLEB) estimator realizes risk reduction over linear methods.

The rest of this paper is organized as follows. In Section 2 we provide an oracle inequality that gives non-asymptotic upper bounds for the regret of the GMLEB. Some implications are given. In Section 3 we prove a large deviation inequality for the generalized MLE under the average Hellinger distance, which is a key element for the oracle inequality. Other elements leading to the oracle inequality are provided in Section 4. In Section 5 we present some simulation results. Mathematical proofs of theorems and lemmas are given either right after their statements or in Section 6.

2. Main results

In the remaining part of the paper, the unknown prior where θ_i are drawn from is denoted by G_n^*. We assume that the variances are uniformly bounded, i.e., there exist constants σ_l and σ_u such that $\sigma_l \le \inf_n \min_i \sigma_i \le \sup_n \max_i \sigma_i \le \sigma_u$. In our analyses, we allow approximate solutions to (1.10). For definiteness and notation simplicity, the generalized MLE is any solution of

$$\prod_{i=1}^{n} f_{\widehat{G}_n,\sigma_i}(X_i) \ge q_n \sup_{G\in\mathscr{G}} \prod_{i=1}^{n} f_{G,\sigma_i}(X_i), \tag{2.1}$$

where $q_n = (e\sqrt{2\pi}/n^2) \wedge 1$. The GMLEB estimator is defined as

$$\hat{\theta}_i = t^*_{\widehat{G}_n}(X_i,\sigma_i) = X_i + \sigma_i^2 \frac{f'_{\widehat{G}_n,\sigma_i}(X_i)}{f_{\widehat{G}_n,\sigma_i}(X_i)}, \quad i = 1,\dots,n, \tag{2.2}$$

where $\widehat{G}_n$ is any approximate generalized MLE (2.1) for prior G_n^* and $f_{G,\sigma}(x)$ is as in (1.7).

2.1. An oracle inequality for the GMLEB

Let $\mu_p(G) = \big\{\int |u|^p dG(u)\big\}^{1/p}$ be the p-th absolute moment of a distribution function G. The convergence rate ε_n, as a function of the sample size n, the mixing distribution G, and the power p of the absolute moment, is defined as

$$\varepsilon(n, G, p) = \max\left\{\sqrt{2\log n}, \left\{n^{1/p}\sqrt{\log n}\mu_p(G)\right\}^{p/(2+2p)}\right\}\sqrt{\frac{\log n}{n}}. \tag{2.3}$$

Theorem 1. *Suppose that under $P_{G_n^*}$, $\theta_1, \dots, \theta_n$ are i.i.d. random variables from a distribution G_n^*, and given θ_i's, $X_i \sim N(\theta_i, \sigma_i^2)$ are independent observations with known variances. Let $\hat{\theta}_i = t^*_{\widehat{G}_n}(X_i, \sigma_i)$ be the GMLEB estimator in (2.2) with an approximate generalized MLE $\widehat{G}_n$ satisfying (2.1). Then, there exists a universal constant M_0 such that for all $\log n > 1/p$,*

$$\left\{\frac{1}{n}\mathbb{E}_{G_n^*}\sum_{i=1}^n \big(t^*_{\widehat{G}_n}(X_i, \sigma_i) - \theta_i\big)^2\right\}^{1/2} - \big\{R_n^*(G_n^*)\big\}^{1/2} \le M_0 \varepsilon_n (\log n)^{3/2}, \tag{2.4}$$

where $R_n^*(G_n^*)$ is the Bayes risk as in (1.9), and $\varepsilon_n = \varepsilon(n, G_n^*, p)$ is as in (2.3).

Here is an outline of the proof of Theorem 1. First of all, one problem with analyzing the GMLEB is that the denominator $f_{\widehat{G}_n,\sigma_i}$ in definition (2.2) could be arbitrarily small. In order to rule out that possibility, we define a regularized rule $t^*_{\widehat{G}_n}(X_i, \sigma_i; \rho_n)$ which replaces this denominator with $f_{\widehat{G}_n,\sigma_i} \vee (\rho_n/\sigma_i)$, and in Theorem 5 we show that this rule relates to the GMLEB as

$$\sum_{i=1}^n \big(t^*_{\widehat{G}_n}(X_i, \sigma_i) - \theta_i\big)^2 = \sum_{i=1}^n \big(t^*_{\widehat{G}_n}(X_i, \sigma_i; \rho_n) - \theta_i\big)^2, \quad \rho_n = \frac{q_n}{\sqrt{2\pi e n}}. \tag{2.5}$$

Let $A_n = \big\{\overline{d}(\widehat{G}_n, G_n^*) \le (x_* \vee 1)\varepsilon_n\big\}$ where x_* is the constant as in Theorem 4, and $\overline{d}(\cdot,\cdot)$ is the average Hellinger distance defined in (3.2). The large deviation inequality in Theorem 4 and the analytical properties of the regularized Bayes rule in Lemma 2 provides an upper bound for $\mathbb{E}_{G_n^*}\zeta_{1n}^2$ where

$$\zeta_{1n} = \left\{\sum_{i=1}^n \big(t^*_{\widehat{G}_n}(X_i, \sigma_i; \rho_n) - \theta_i\big)^2 I_{A_n^c}\right\}^{1/2}. \tag{2.6}$$

Because the generalized MLE is based on the same data, $\hat{\theta}_i = t^*_{\widehat{G}_n}(X_i, \sigma_i; \rho_n)$ is not separable. We use the following strategy. Let $\big\{(t^*_{H_j}(\cdot, \sigma_1; \rho_n), \dots, t^*_{H_j}(\cdot, \sigma_n; \rho_n)), j \le N\big\}$ be a set of approximated regularized Bayes rules in the sense that it is a $(2\eta^*)$-net of

$$\Big\{\big(t^*_G(\cdot, \sigma_1; \rho_n), \dots, t^*_G(\cdot, \sigma_n; \rho_n)\big) : \overline{d}(G, G_n^*) \le x^* \varepsilon_n\Big\} \tag{2.7}$$

under $\|\cdot\|_{\infty,M}$, where η^* will be manifested in Theorem 7. By the entropy bound in Theorem 7, there exists a collection of distributions $\{H_j, j \le N\}$ of manageable size N such that

$$\begin{aligned}\zeta_{2n} \;=\; & \left|\Big\{\sum_{i=1}^{n}\big(t^*_{\widehat{G}_n}(X_i,\sigma_i;\rho_n)-\theta_i\big)^2 I_{A_n}\Big\}^{1/2}\right.\\ & \left. -\max_{j\le N}\Big\{\sum_{i=1}^{n}\big(t^*_{H_j}(X_i,\sigma_i;\rho_n)-\theta_i\big)^2\Big\}^{1/2}\right| \qquad (2.8)\end{aligned}$$

is small. Since the collection $\{H_j, j \le N\}$ is of manageable size, a Gaussian isoperimetric inequality yields that

$$\begin{aligned}\zeta_{3n} \;=\; & \max_{j\le N}\bigg\{\Big\{\sum_{i=1}^{n}\big(t^*_{H_j}(X_i,\sigma_i;\rho_n)-\theta_i\big)^2\Big\}^{1/2}\\ & -\mathbb{E}_{G^*_n}\Big\{\sum_{i=1}^{n}\big(t^*_{H_j}(X_i,\sigma_i;\rho_n)-\theta_i\big)^2\Big\}^{1/2}\bigg\}_+ \qquad (2.9)\end{aligned}$$

is small. Finally, Theorem 6 provides an upper bound of the regret due to the lack of the knowledge of G^*_n, which implies that

$$\zeta_{4n} = \max_{j\le N}\bigg\{\mathbb{E}_{G^*_n}\sum_{i=1}^{n}\big(t^*_{H_j}(X_i,\sigma_i;\rho_n)-\theta_i\big)^2\bigg\}^{1/2} - \big\{nR^*_n(G^*_n)\big\}^{1/2} \qquad (2.10)$$

is small. These upper bounds for individual pieces $\mathbb{E}_{G^*_n}\zeta^2_{jn}$ are put together via

$$\bigg\{\mathbb{E}_{G^*_n}\sum_{i=1}^{n}\big(t^*_{\widehat{G}_n}(X_i,\sigma_i)-\theta_i\big)^2\bigg\}^{1/2} \le \big\{nR^*_n(G^*_n)\big\}^{1/2} + \sum_{j=1}^{4}\big(\mathbb{E}_{G^*_n}\zeta^2_{jn}\big)^{1/2}. \qquad (2.11)$$

2.2. Consequences of the oracle inequality

Theorem 2. *Suppose that under $P_{G^*_n}$, $\theta_1,\ldots,\theta_n$ are i.i.d. random variables from a distribution G^*_n, and given θ_i's, $X_i \sim N(\theta_i,\sigma_i^2)$ are independent observations with known variances. Let $\hat\theta_i = t^*_{\widehat{G}_n}(X_i,\sigma_i)$ be the GMLEB estimator in* (2.2) *with an approximate generalized MLE $\widehat{G}_n$ satisfying* (2.1)*. Then,*

$$\limsup_{n\to\infty}\frac{\mathbb{E}_{G^*_n}\sum_{i=1}^{n}\big(t^*_{\widehat{G}_n}(X_i,\sigma_i)-\theta_i\big)^2/n}{R^*_n(G^*_n)} = 1, \qquad (2.12)$$

provided that $\mu_\infty(G^*_n) = O(\sqrt{\log n})$ and $nR^*_n(G^*_n)/(\log n)^5 \to \infty$.

For a class of distributions $\mathscr{G}$, the minimax risk for the average squared loss (1.2) is

$$\mathscr{R}_n(\mathscr{G}) = \inf_t \sup_{G\in\mathscr{G}} \mathbb{E}_G \frac{1}{n}\sum_{i=1}^{n}\big(t(X_i,\sigma_i)-\theta_i\big)^2, \qquad (2.13)$$

where the infimum is taken over all bivariate Borel functions. An estimator is adaptive minimax if

$$\frac{\sup_{G\in\mathscr{G}_n}\mathbb{E}_G\sum_{i=1}^n(\hat\theta_i-\theta_i)^2/n}{\mathscr{R}_n(\mathscr{G}_n)}\to 1 \tag{2.14}$$

holds uniformly for a range of sequences $\{\mathscr{G}_n, n\ge 1\}$ of distribution classes. For positive p and C, the L_p balls of distribution functions are defined as

$$\mathscr{G}_{p,C}=\Big\{G\colon \int |u|^p dG(u)\le C^p\Big\}. \tag{2.15}$$

Theorem 3. *Suppose that under $P_{G_n^*}$, $\theta_1,\dots,\theta_n$ are i.i.d. random variables from a distribution G_n^*, and given θ_i's, $X_i\sim N(\theta_i,\sigma_i^2)$ are independent observations with known variances. Let $\hat\theta_i=t^*_{\widehat G_n}(X_i,\sigma_i)$ be the GMLEB estimator in (2.2) with an approximate generalized MLE $\widehat G_n$ satisfying (2.1). Then, the adaptive minimaxity (2.14) holds in L_p balls $\mathscr{G}_{p,C_n}$ in (2.15), provided that $C_n\to 0$ and $\sqrt{n}C_n^{p\wedge 2}/(\log n)^{5/2}\to\infty$.*

Proof of Theorem 3. By definition of minimax risk in (2.13), we have

$$\begin{aligned}\mathscr{R}_n(\mathscr{G}_{p,C_n}) &\ge \sup_{G\in\mathscr{G}_{p,C_n}}\inf_t \mathbb{E}_G\frac1n\sum_{i=1}^n\big(t(X_i,\sigma_i)-\theta_i\big)^2\\ &= \sup_{G\in\mathscr{G}_{p,C_n}}\mathbb{E}_G\frac1n\sum_{i=1}^n\big(t^*_G(X_i,\sigma_i)-\theta_i\big)^2\\ &= \sup_{G\in\mathscr{G}_{p,C_n}} R_n^*(G). \end{aligned}\tag{2.16}$$

By Theorem 1, (2.16) and $\sqrt{n}C_n^{p\wedge 2}/(\log n)^{5/2}\to\infty$, there exists a universal constant M_1 such that

$$\begin{aligned}\sup_{G\in\mathscr{G}_{p,C_n}}\mathbb{E}_G\frac1n\sum_{i=1}^n\big(t^*_{\widehat G_n}(X_i,\sigma_i)-\theta_i\big)^2 &\le \sup_{G\in\mathscr{G}_{p,C_n}} R_n^*(G)+M_1(\log n)^{5/2}/\sqrt{n}\\ &\le \mathscr{R}_n(\mathscr{G}_{p,C_n})+o(1)C_n^{p\wedge 2}. \end{aligned}\tag{2.17}$$

Donoho and Johnstone [3] proved that as $C_n\to 0$,

$$\mathscr{R}_n(\mathscr{G}_{p,C_n})=O(1)C_n^{p\wedge 2}\big\{2\log(1/C_n^p)\big\}^{(1-p/2)_+}. \tag{2.18}$$

Thus, (2.17) and (2.18) lead to that

$$\sup_{G\in\mathscr{G}_{p,C_n}}\mathbb{E}_G\frac1n\sum_{i=1}^n\big(t^*_{\widehat G_n}(X_i,\sigma_i)-\theta_i\big)^2\le\big(1+o(1)\big)\mathscr{R}_n(\mathscr{G}_{p,C_n}).$$

This is the adaptive minimaxity in $\mathscr{G}_{p,C_n}$. ∎

3. A large deviation inequality for the generalized MLE

In [7], the analysis of risk is divided into two parts. One is outside a Hellinger neighborhood $\{d(f_{\widehat{G}_n}, f_{G_n^*}) \le x\varepsilon_n\}$, the other is inside this neighborhood. An essential ingredient is a large deviation inequality for $d(f_{\widehat{G}_n}, f_{G_n^*})$. In the heteroscedastic case, it seems that certain omnibus distance between $f_{\widehat{G}_n,\sigma_i}$ and $f_{G_n^*,\sigma_i}$ should be used. We use the average Hellinger distance $\overline{d}(\widehat{G}_n, G_n^*)$ as defined in (3.2) below. We provide a large deviation inequality for $\overline{d}(\widehat{G}_n, G_n^*)$. This result plays a crucial role in the oracle inequality stated in Theorem 1.

Define the collection of n-dimensional vectors of marginal densities as

$$\mathscr{F}_n = \big\{\big(f_{G,\sigma_1}(x), \dots, f_{G,\sigma_n}(x)\big), G \in \mathscr{G}\big\}, \tag{3.1}$$

where G is the family of all distribution functions. For two vectors $(f_{G,\sigma_1}(x), \dots, f_{G,\sigma_n}(x))$, $(f_{H,\sigma_1}(x), \dots, f_{H,\sigma_n}(x)) \in \mathscr{F}_n$, define the average Hellinger distance

$$\overline{d}(G,H) = \left\{\frac{1}{n}\sum_{i=1}^n d^2(f_{G,\sigma_i}, f_{H,\sigma_i})\right\}^{1/2}, \tag{3.2}$$

where $d^2(f,g) = (1/2)\int(\sqrt{f} - \sqrt{g})^2$ is the square of the Hellinger distance between probability densities f and g. Define the supreme norm in bounded intervals,

$$\|\boldsymbol{h}\|_{\infty,M} = \max_{i\le n} \big\|h_i\big\|_{\infty,M} = \max_{i\le n} \sup_{|x|\le M} \big|h_i(x)\big|, \tag{3.3}$$

where $\boldsymbol{h} = (h_1(x), \dots, h_n(x))$ is an n-dimensional vector of functions.

Theorem 4. *Suppose that under $P_{G_n^*}$, $\theta_1, \dots, \theta_n$ are i.i.d. random variables from a distribution G_n^*, and given θ_i's, $X_i \sim N(\theta_i, \sigma_i^2)$ are independent observations with known variances. Let $f_{G,\sigma}$ be as in (1.7). Let $\widehat{G}_n$ be certain approximate generalized MLE satisfying (2.1). Then, there exists a universal constant x_* such that for all $t \ge x_*$ and $\log n > 1/p$,*

$$\mathbb{P}_{G_n^*}\Big\{\overline{d}(\widehat{G}_n, G_n^*) \ge t\varepsilon_n\Big\} \le \exp\Big(-\frac{t^2 n\varepsilon_n^2}{2\log n}\Big) \le e^{-t^2\log n}, \tag{3.4}$$

where $\varepsilon_n = \varepsilon(n, G_n^*, p)$ is as in (2.3) and $\overline{d}(G,H)$ is the average Hellinger distance (3.2).

Proof of Theorem 4. Let $\eta = 1/n^2$ and $M = 2\sigma_u n\varepsilon_n^2/(\log n)^{3/2}$. Define

$$h^*(x) = \eta I\big\{|x| \le M\big\} + \frac{\eta M^2}{x^2} I\big\{|x| > M\big\}. \tag{3.5}$$

We consider any approximate generalized MLE satisfying

$$\prod_{i=1}^n \frac{f_{\widehat{G}_n,\sigma_i}(X_i)}{f_{G_n^*,\sigma_i}(X_i)} \ge e^{-4t^2 n\varepsilon_n^2/15}. \tag{3.6}$$

Let $\{(f_{H_j,\sigma_1}(x),\ldots,f_{H_j,\sigma_n}(x)), j\le N\}$ be an η-net of $\mathscr{F}_n$ under the semi-norm $\|\cdot\|_{\infty,M}$, with $N=N(\eta,\mathscr{F}_n,\|\cdot\|_{\infty,M})$. Let $H_{0,j}$ be distributions satisfying

$$\overline{d}(H_{0,j},G_n^*)\ge t\varepsilon_n,\quad \max_{i\le n}\big\|f_{H_{0,j},\sigma_i}-f_{H_j,\sigma_i}\big\|_{\infty,M}\le\eta, \tag{3.7}$$

if they exist, and $J=\{j\le N\colon H_{0,j}\text{ exists}\}$. For any distribution G with $\overline{d}(G,G_n^*)\ge t\varepsilon_n$, there exists $j\in J$ such that for $i=1,\ldots,n$,

$$f_{G,\sigma_i}(x)\le\begin{cases} f_{H_{0,j},\sigma_i}(x)+2\eta=f_{H_{0,j},\sigma_i}(x)+2h^*(x), & |x|<M,\\ 1/(\sqrt{2\pi}\sigma_i), & |x|\ge M.\end{cases}$$

It follows that when $\overline{d}(\widehat{G}_n,G_n^*)\ge t\varepsilon_n$,

$$\begin{aligned}\prod_{i=1}^n\frac{f_{\widehat{G}_n,\sigma_i}(X_i)}{f_{G_n^*,\sigma_i}(X_i)} &= \prod_{|X_i|<M}\frac{f_{\widehat{G}_n,\sigma_i}(X_i)}{f_{G_n^*,\sigma_i}(X_i)}\prod_{|X_i|\ge M}\frac{f_{\widehat{G}_n,\sigma_i}(X_i)}{f_{G_n^*,\sigma_i}(X_i)}\\ &\le \sup_{j\in J}\prod_{|X_i|<M}\frac{f_{H_{0,j},\sigma_i}(X_i)+2h^*(X_i)}{f_{G_n^*,\sigma_i}(X_i)}\prod_{|X_i|\ge M}\frac{1/(\sqrt{2\pi}\sigma_i)}{f_{G_n^*,\sigma_i}(X_i)}\\ &\le \sup_{j\in J}\prod_{i=1}^n\frac{f_{H_{0,j},\sigma_i}(X_i)+2h^*(X_i)}{f_{G_n^*,\sigma_i}(X_i)}\prod_{|X_i|\ge M}\frac{1/(\sqrt{2\pi}\sigma_i)}{2h^*(X_i)}.\end{aligned}$$

Thus, by (3.6),

$$\begin{aligned}&\mathbb{P}_{G_n^*}\Big\{\overline{d}(\widehat{G}_n,G_n^*)\ge t\varepsilon_n\Big\}\\ \le\;& \mathbb{P}_{G_n^*}\Big\{\sup_{j\in J}\prod_{i=1}^n\frac{f_{H_{0,j},\sigma_i}(X_i)+2h^*(X_i)}{f_{G_n^*,\sigma_i}(X_i)}\prod_{|X_i|\ge M}\frac{1/(\sqrt{2\pi}\sigma_i)}{2h^*(X_i)}\ge e^{-4t^2n\varepsilon_n^2/15}\Big\}\\ \le\;& \mathbb{P}_{G_n^*}\Big\{\sup_{j\in J}\prod_{i=1}^n\frac{f_{H_{0,j},\sigma_i}(X_i)+2h^*(X_i)}{f_{G_n^*,\sigma_i}(X_i)}\ge e^{-8t^2n\varepsilon_n^2/5}\Big\}\\ &+\mathbb{P}_{G_n^*}\Big\{\prod_{|X_i|\ge M}\frac{1/(\sqrt{2\pi}\sigma_i)}{2h^*(X_i)}\ge e^{4t^2n\varepsilon_n^2/3}\Big\}.\end{aligned}\tag{3.8}$$

We derive large deviation inequalities for the right hind side of (3.8). For $j\in J$ in the first term,

$$\begin{aligned}&\mathbb{P}_{G_n^*}\Big\{\prod_{i=1}^n\frac{f_{H_{0,j},\sigma_i}(X_i)+2h^*(X_i)}{f_{G_n^*,\sigma_i}(X_i)}\ge e^{-8t^2n\varepsilon_n^2/5}\Big\}\\ \le\;& \exp\left(4t^2n\varepsilon_n^2/5\right)\prod_{i=1}^n\int\sqrt{f_{H_{0,j},\sigma_i}+2h^*}\sqrt{f_{G_n^*,\sigma_i}}\\ \le\;& \exp\Big\{\frac{4t^2n\varepsilon_n^2}{5}+\sum_{i=1}^n\Big(\int\sqrt{f_{H_{0,j},\sigma_i}+2h^*}\sqrt{f_{G_n^*,\sigma_i}}-1\Big)\Big\}.\end{aligned}\tag{3.9}$$

By Jensen's inequality, $\overline{d}(H_{0,j}, G_n^*) \geq t\varepsilon_n$ and $\int h^* = 4\eta M$,

$$\begin{aligned} & \sum_{i=1}^n \Big(\int \sqrt{f_{H_{0,j},\sigma_i} + 2h^*} \sqrt{f_{G_n^*,\sigma_i}} - 1 \Big) \\ \leq \;& \sum_{i=1}^n \Big(-d^2(f_{H_{0,j},\sigma_i}, f_{G_n^*,\sigma_i}) + \big(2 \int h^* \big)^{1/2} \Big) \\ \leq \;& -t^2 n\varepsilon_n^2 + n\sqrt{8\eta M}. \end{aligned} \tag{3.10}$$

Since $|J| \leq N$, (3.9) and (3.10) yield

$$\begin{aligned} & \mathbb{P}_{G_n^*}\Big\{ \sup_{j\in J} \prod_{i=1}^n \frac{f_{H_{0,j},\sigma_i}(X_i) + 2h^*(X_i)}{f_{G_n^*,\sigma_i}(X_i)} \geq e^{-8t^2 n\varepsilon_n^2/5} \Big\} \\ \leq \;& \exp\Big\{ \log N - \frac{t^2 n\varepsilon_n^2}{5} + n\sqrt{8\eta M} \Big\}. \end{aligned} \tag{3.11}$$

Since $\eta = 1/n^2$ and $M = 2\sigma_u n\varepsilon_n^2/(\log n)^{3/2} \geq 4\sigma_u\sqrt{\log n}$, by Lemma 4,

$$\begin{aligned} \log N + n\sqrt{8\eta M} \;&\leq\; C(2\log n)^2 \max\Big(\frac{M}{\sqrt{2\log n}}, 1 \Big) + \sqrt{8M} \\ &\leq\; \Big\{ \frac{(t^*)^2}{20} \Big\} M(\log n)^{3/2} \leq \frac{t^2 n\varepsilon_n^2}{10} \end{aligned}$$

for $t^* \leq t$. Thus, by (3.11),

$$\mathbb{P}_{G_n^*}\Big\{ \sup_{j\in J} \prod_{i=1}^n \frac{f_{H_{0,j},\sigma_i}(X_i) + 2h^*(X_i)}{f_{G_n^*,\sigma_i}(X_i)} \geq e^{-8t^2 n\varepsilon_n^2/5} \Big\} \leq e^{-t^2 n\varepsilon_n^2/10}. \tag{3.12}$$

By (3.5), $1/h^*(x) = x^2/(\eta M^2) = (nx/M)^2$ for $|x| \geq M$. So that

$$\begin{aligned} & \mathbb{P}_{G_n^*}\Big\{ \prod_{|X_i|\geq M} \frac{1/(\sqrt{2\pi}\sigma_i)}{2h^*(X_i)} \geq e^{4t^2 n\varepsilon_n^2/3} \Big\} \\ \leq \;& \exp\Big(-\frac{2t^2 n\varepsilon_n^2}{3\log n} \Big) \mathbb{E}_{G_n^*}\Big\{ \prod_{|X_i|\geq M} \Big| \frac{nX_i}{\sqrt{\sigma_i}M} \Big| \Big\}^{1/\log n}. \end{aligned} \tag{3.13}$$

Since $M = 2\sigma_u n\varepsilon_n^2/(\log n)^{3/2} \geq 4\sigma_u\sqrt{\log n}$, Lemma 5 is applicable with $a_i = n/(\sqrt{\sigma_i}M)$ and $\lambda = 1/\log n < 1$. Because $a_i \leq n/(\sqrt{\sigma_l}M)$,

$$\begin{aligned} & \mathbb{E}_{G_n^*}\Big\{ \prod_{|X_i|\geq M} \Big| \frac{nX_i}{\sqrt{\sigma_l}M} \Big| \Big\}^{1/\log n} \\ \leq \;& \exp\Big\{ \Big(e/\sigma_l^{1/(2\log n)} \Big) \Big(\frac{1}{\sqrt{2\pi\log n}} + n\Big(\frac{2\mu_p(G_n^*)}{M} \Big)^p \Big) \Big\}. \end{aligned} \tag{3.14}$$

By the definition of ε_n,

$$\frac{n\varepsilon_n^2/\log n}{n\big(2\mu_p(G_n^*)/M\big)^p} \ge 1.$$

Therefore, (3.13) and (3.14) give

$$\begin{aligned} & \mathbb{P}_{G_n^*}\bigg\{\prod_{|X_i|\ge M}\frac{1/(\sqrt{2\pi}\sigma_i)}{2h^*(X_i)} \ge e^{4t^2 n\varepsilon_n^2/3}\bigg\} \\ \le\quad & \exp\bigg\{-\Big(\frac{2t^2}{3}-e/\sigma_l^{1/(2\log n)}\Big)\frac{n\varepsilon_n^2}{\log n}+\frac{e/\sigma_l^{1/(2\log n)}}{\sqrt{2\pi\log n}}\bigg\}. \qquad (3.15)\end{aligned}$$

Inserting (3.12) and (3.15) into (3.8), we find that for large n and $t \ge x_*$,

$$\mathbb{P}_{G_n^*}\Big\{\overline{d}(\widehat{G}_n, G_n^*) \ge t\varepsilon_n\Big\} \le \exp\Big(-\frac{t^2 n\varepsilon_n^2}{2\log n}\Big) \le e^{-t^2\log n} = n^{-t^2}.$$

This completes the proof of Theorem 4. □

4. Other elements of the oracle inequality

In this section we provide other elements of the oracle inequality in Theorem 1. We divide this section into four subsections to study: (1) the connection between the GMLEB and the regularized rule, (2) some analytical properties of the regularized Bayes estimator, (3) regret of a regularized Bayes estimator with a misspecified prior, and (4) an entropy bound for regularized Bayes rules.

For the Bayes rule $t_G^*(x,\sigma) = x + \sigma^2 f'_{G,\sigma}(x)/f_{G,\sigma}(x)$, we may want to avoid dividing by a near-zero quantity. Define regularized Bayes rule as

$$t_G^*(x,\sigma;\rho) = x + \frac{\sigma^2 f'_{G,\sigma}(x)}{(\rho/\sigma)\vee f_{G,\sigma}(x)}. \qquad (4.1)$$

Denote $t_G^*(x) = x + f'_G(x)/f_G(x)$ and $t_G^*(x) = x + f'_G(x)/(\rho \vee f_G(x))$ as the Bayes and regularized Bayes rules for the unit-variance normal mean problem with prior G respectively, where $f_G(x) = \int \varphi(x-u)dG(u)$. Let F be a scale change of G:

$$\int h(u)dF(u) = \int h(u/\sigma)dG(u). \qquad (4.2)$$

With $y = x/\sigma$, by the condition on F, we have $t_G^*(x,\sigma)/\sigma = t_F^*(y)$ and

$$\frac{t_G^*(x,\sigma;\rho)}{\sigma} = y + \frac{f'_F(y)}{\rho\vee f_F(y)} = t_F^*(y;\rho). \qquad (4.3)$$

This is a scale invariance of the Bayes and regularized Bayes rules.

4.1. Connection between the GMLEB and the regularized rule

The connection between the GMLEB estimator (2.2) and the regularized Bayes rule in (4.1) is provided by

$$t^*_{\widehat{G}_n}(X_i,\sigma_i) = t^*_{\widehat{G}_n}(X_i,\sigma_i;\rho_n), \quad \rho_n = q_n/(\sqrt{2\pi}en), \tag{4.4}$$

where $0 < q_n \le 1$. This is consequence of the following theorem.

Theorem 5. *Let $\widehat{G}_n$ be an approximate generalized MLE satisfying*

$$\prod_{i=1}^n f_{\widehat{G}_n,\sigma_i}(X_i) \ge q_n \sup_G \prod_{i=1}^n f_{G,\sigma_i}(X_i) \tag{4.5}$$

for certain $0 < q_n \le 1$. Then, for all $j = 1,\dots,n$,

$$f_{\widehat{G}_n,\sigma_j}(X_j) \ge \frac{q_n}{\sqrt{2\pi}en\sigma_j}. \tag{4.6}$$

Proof of Theorem 5. Define $\widehat{G}_{n,j} = (1-\varepsilon)\widehat{G}_n + \varepsilon\delta_{X_j}$, where δ_u is the unit mass at u. Since $f_{\widehat{G}_{n,j},\sigma_i}(X_i) \ge (1-\varepsilon)f_{\widehat{G}_n,\sigma_i}(X_i)$ and $f_{\widehat{G}_{n,j},\sigma_j}(X_j) \ge \varepsilon/(\sqrt{2\pi}\sigma_j)$, so that

$$\prod_{i=1}^n f_{\widehat{G}_n,\sigma_i}(X_i) \ge q_n \prod_{i=1}^n f_{\widehat{G}_{n,j},\sigma_i}(X_i) \ge q_n(1-\varepsilon)^{n-1}\big(\varepsilon/(\sqrt{2\pi}\sigma_j)\big) \prod_{i\ne j} f_{\widehat{G}_n,\sigma_i}(X_i).$$

Thus, $f_{\widehat{G}_n,\sigma_j}(X_j) \ge q_n(1-\varepsilon)^{n-1}\varepsilon/(\sqrt{2\pi}\sigma_j)$ after the cancelation of $f_{\widehat{G}_n,\sigma_i}(X_i)$ for $i \ne j$. The conclusion follows by taking $\varepsilon = 1/n$. □

Remark 3. In the proof of Theorem 1, for notation simplicity, we set $q_n = (e\sqrt{2\pi}/n^2) \wedge 1$ so that $\rho_n = 1/n^3$.

4.2. Some properties of the regularized Bayes estimator

In this subsection we give some analytical properties of the regularized Bayes estimator. Denote the inverse function of $y = \varphi(x)$ by

$$\widetilde{L}(y) = \sqrt{-\log(2\pi y^2)}, \quad y \ge 0. \tag{4.7}$$

Since $\max_{i\le n}|X_i - \theta_i| \le \sigma_u\sqrt{2\log n}$ with large probability, we expect that $\max_{i\le n}|t^*_G(x,\sigma;\rho) - x| \le c_0\sqrt{\log n}$ for some constant c_0. This is established in the following lemmas.

Lemma 1. *Let $f_{G,\sigma}(x)$ be as in (1.7) and $\widetilde{L}(y) = \sqrt{-\log(2\pi y^2)}$. Then,*

$$\Big(\frac{f'_{G,\sigma}(x)}{f_{G,\sigma}(x)}\Big)^2 \le \frac{f''_{G,\sigma}(x)}{f_{G,\sigma}(x)} + \frac{1}{\sigma^2} \le \frac{1}{\sigma^2}\widetilde{L}^2(\sigma f_{G,\sigma}(x)) = \frac{1}{\sigma^2}\log\Big(\frac{1}{2\pi\sigma^2 f^2_{G,\sigma}(x)}\Big). \tag{4.8}$$

Proof of Lemma 1. Let $Y|\xi \sim N(\xi, \sigma^2)$ and $\xi \sim G$ under P_G. Then,

$$\mathbb{E}_G\Big[\frac{\xi - Y}{\sigma^2}\Big|Y = x\Big] = \frac{f'_{G,\sigma}(x)}{f_{G,\sigma}(x)}, \quad \mathbb{E}_G\Big[\frac{(\xi - Y)^2}{\sigma^4}\Big|Y = x\Big] = \frac{f''_{G,\sigma}(x)}{f_{G,\sigma}(x)} + \frac{1}{\sigma^2}.$$

This gives the first inequality of (4.8). Let $h(x) = e^{\sigma^2 x/2}$. The second inequality of (4.8) follows from Jensen's inequality,

$$h\Big(\frac{f''_{G,\sigma}(x)}{f_{G,\sigma}(x)} + \frac{1}{\sigma^2}\Big) \leq \mathbb{E}_G\Big[h\Big(\frac{(\xi - Y)^2}{\sigma^4}\Big)\Big|Y = x\Big] = \frac{1}{\sqrt{2\pi}\sigma f_{G,\sigma}(x)}.$$

This completes the proof. □

Lemma 2. *Let $t^*_G(x, \sigma; \rho)$ be the regularized Bayes estimator in (4.1). Let $\widetilde{L}(y) = \sqrt{-\log(2\pi y^2)}$ be the inverse of $y = \varphi(x)$ as in (4.7). Then, for all $x \in \mathbb{R}$,*

$$\begin{cases} \big|t^*_G(x, \sigma; \rho) - x\big| \leq \sigma\widetilde{L}(\rho), & 0 < \rho < (2\pi e)^{-1/2}, \\ 0 \leq (\partial/\partial x)t^*_G(x, \sigma; \rho) \leq \widetilde{L}^2(\rho), & 0 < \rho < (2\pi e^3)^{-1/2}. \end{cases} \tag{4.9}$$

Proof of Lemma 2. By Lemma 1,

$$\begin{aligned} \big|t^*_G(x, \sigma; \rho) - x\big| &= \sigma^2 \frac{f_{G,\sigma}(x)}{(\rho/\sigma) \vee f_{G,\sigma}(x)} \Big|\frac{f'_{G,\sigma}(x)}{f_{G,\sigma}(x)}\Big| \\ &\leq \sigma \frac{f_{G,\sigma}(x)}{(\rho/\sigma) \vee f_{G,\sigma}(x)} \widetilde{L}\big(\sigma f_{G,\sigma}(x)\big). \end{aligned} \tag{4.10}$$

If $f_{G,\sigma}(x) \geq \rho/\sigma$, since $\widetilde{L}(y)$ is decreasing in $y > 0$, $|t^*_G(x, \sigma; \rho) - x| \leq \sigma\widetilde{L}(\rho)$ by (4.10). If $f_{G,\sigma}(x) < \rho/\sigma$, since $y\widetilde{L}(y)$ is increasing in $0 < y \leq (2\pi e)^{-1/2}$, $|t^*_G(x, \sigma; \rho) - x| \leq \sigma\widetilde{L}(\rho)$. This is the first line of (4.9).

By the definition of $t^*_G(x, \sigma; \rho)$,

$$\frac{\partial t^*_G(x, \sigma; \rho)}{\partial x} = \begin{cases} 1 + \sigma^2 \frac{f''_{G,\sigma}(x)}{f_{G,\sigma}(x)} - \sigma^2 \big(\frac{f'_{G,\sigma}(x)}{f_{G,\sigma}(x)}\big)^2, & f_{G,\sigma}(x) \geq \rho/\sigma, \\ 1 + \sigma^2 \frac{f''_{G,\sigma}(x)}{\rho/\sigma}, & f_{G,\sigma}(x) < \rho/\sigma. \end{cases} \tag{4.11}$$

If $f_{G,\sigma}(x) \geq \rho/\sigma$, by (4.11) and Lemma 1,

$$\frac{\partial t^*_G(x, \sigma; \rho)}{\partial x} \leq 1 + \sigma^2 \frac{f''_{G,\sigma}(x)}{f_{G,\sigma}(x)} \leq \widetilde{L}^2(\sigma f_{G,\sigma}(x)) \leq \widetilde{L}^2(\rho).$$

If $f_{G,\sigma}(x) < \rho/\sigma$, by Lemma 1,

$$0 \leq 1 - \frac{f_{G,\sigma}(x)}{\rho/\sigma} \leq 1 + \sigma^2 \frac{f''_{G,\sigma}(x)}{\rho/\sigma} \leq 1 + \frac{f_{G,\sigma}(x)}{\rho/\sigma}\big(\widetilde{L}^2(\sigma f_{G,\sigma}(x)) - 1\big). \tag{4.12}$$

Because $y(\widetilde{L}^2(y)-1)$ is increasing in $0 \le y \le (2\pi e^3)^{-1/2}$, for $\sigma f_{G,\sigma}(x) \le \rho \le (2\pi e^3)^{-1/2}$,

$$\sigma f_{G,\sigma}(x)\big(\widetilde{L}^2(\sigma f_{G,\sigma}(x))-1\big) \le \rho\big(\widetilde{L}^2(\rho)-1\big). \tag{4.13}$$

Putting (4.12) and (4.13) together, we have

$$0 \le 1+\sigma^2\frac{f''_{G,\sigma}(x)}{\rho/\sigma} \le \widetilde{L}^2(\rho).$$

This gives the second line of (4.9). □

4.3. Regret of a regularized Bayes estimator with a misspecified prior

Let F_i and F_i^* be scale changes of G and G_n^* under parameter σ_i according to (4.2), respectively. Let $Y_i = X_i/\sigma_i$ and $\xi_i = \theta_i/\sigma_i$. It follows from (4.3) that

$$\begin{aligned} & \mathbb{E}_{G_n^*}\big(t_G^*(X_i,\sigma_i;\rho)-\theta_i\big)^2/\sigma_i^2 - \mathbb{E}_{G_n^*}\big(t_{G_n^*}^*(X_i,\sigma_i)-\theta_i\big)^2/\sigma_i^2 \\ = \;& \mathbb{E}_{F_i^*}\big(t_{F_i}^*(Y_i;\rho)-\xi_i\big)^2 - \mathbb{E}_{F_i^*}\big(t_{F_i^*}^*(Y_i)-\xi_i\big)^2. \end{aligned} \tag{4.14}$$

Then, by Theorem 3 of [7] and Lemma 6.1 of [19], for all $0<\rho\le(2\pi e^2)^{-1/2}$ and $x_0>0$,

$$\begin{aligned} & \mathbb{E}_{F_i^*}\big(t_{F_i}^*(Y_i;\rho)-\xi_i\big)^2 - \mathbb{E}_{F_i^*}\big(t_{F_i^*}^*(Y_i)-\xi_i\big)^2 \\ \le\;& M_0\max\big\{|\log\rho|^3, |\log d(f_{F_i}, f_{F_i^*})|\big\} d^2(f_{F_i}, f_{F_i^*}) \\ & +2\Big\{\mathbb{P}_{F_i^*}\big\{|\xi_i|>x_0\big\} + 2x_0\rho\widetilde{L}^2(\rho) + 2\rho\big(\widetilde{L}^2(\rho)+2\big)^{1/2}\Big\}, \end{aligned}$$

where M_0 is a universal constant. Note that the Hellinger distance is invariant under scale change: $d(f_{F_i}, f_{F_i^*}) = d(f_{G,\sigma_i}, f_{G_n^*,\sigma_i})$. Thus we have the following risk bound for the regularized Bayes rule for misspecified prior, which will be used to bound ζ_{4n}^2 in (2.10).

Theorem 6. *For any $0<\rho\le(2\pi e^2)^{-1/2}$ and $x_0>0$,*

$$\begin{aligned} & \frac{1}{n}\mathbb{E}_{G_n^*}\sum_{i=1}^n\big(t_G^*(X_i,\sigma_i;\rho)-\theta_i\big)^2 - R_n^*(G_n^*) \\ \le\;& \frac{M_0}{n}\sum_{i=1}^n\sigma_i^2\max\big\{|\log\rho|^3, |\log d(f_{G,\sigma_i}, f_{G_n^*,\sigma_i})|\big\} d^2(f_{G,\sigma_i}, f_{G_n^*,\sigma_i}) \\ & +\frac{2}{n}\sum_{i=1}^n\sigma_i^2\Big\{\mathbb{P}_{G_n^*}\big\{|\theta_i/\sigma_i|>x_0\big\} + 2x_0\rho\widetilde{L}^2(\rho) \\ & +2\rho\big(\widetilde{L}^2(\rho)+2\big)^{1/2}\Big\}. \end{aligned} \tag{4.15}$$

TABLE 1
Average of $\sum_i(\hat\theta_i-\theta_i)^2$ based on 100 replications: $n=1000$, $\sigma_i\sim\mathrm{Unif}(0.5,1.5)$, $\theta_i\in\{0,\mu\}$, $\#\{i\colon \theta_i\neq 0\}=5$, 50 or 500.

$\#\{\theta_i\neq 0\}$	5				50				500			
μ	1.5	2	2.5	3	1.5	2	2.5	3	1.5	2	2.5	3
James-Stein	**13**	23	33	46	101	168	242	318	387	538	665	744
SURE-M	**13**	22	32	44	97	157	221	287	349	482	601	681
SURE-SG	16	25	35	45	99	159	220	284	357	492	615	695
Group-linear	27	37	46	57	110	170	232	293	364	494	611	690
NEST	119	123	133	130	182	213	238	248	400	499	563	559
GMLEB	**13**	**21**	**26**	**28**	**87**	**121**	**141**	**144**	**333**	**425**	**473**	**454**
Oracle	10	16	21	24	82	117	136	138	326	418	468	448

4.4. An entropy bound for regularized Bayes rules

We now provide an entropy bound for collections of regularized Bayes rules. It is used to bound $\mathbb{E}_{G_n^*}\zeta_{3n}^2$ in (2.9) with a Gaussian isoperimetric inequality. For any family $\mathscr{H}$ of functions and semi-distance d, the η-covering number is

$$N(\eta,\mathscr{H},d)\equiv\inf\Big\{N\colon \mathscr{H}\subseteq\cup_{j=1}^N \mathrm{Ball}(h_j,\eta,d)\Big\}$$

with $\mathrm{Ball}(h,\varepsilon,d)\equiv\big\{f\colon d(f,h)<\eta\big\}$. For each fixed $\rho>0$ define the collection of the regularized Bayes rules $t_G^*(x;\rho)$ in (4.1) as

$$\mathscr{T}_\rho=\big\{\big(t_G^*(\cdot,\sigma_1;\rho),\ldots,t_G^*(\cdot,\sigma_n;\rho)\big)\colon G\in\mathscr{G}\big\}. \tag{4.16}$$

where $\mathscr{G}$ is the family of all distribution functions. The following theorem provides an entropy bound for (4.16) under the seminorm $\|\cdot\|_{\infty,M}$ defined in (3.3).

Theorem 7. *Let $\widetilde L(y)=\sqrt{-\log(2\pi y^2)}$ be the inverse of $y=\varphi(x)$. Then, for all $0<\eta\le\sigma_l\rho\le(2\pi e)^{-1/2}$,*

$$\begin{aligned}&\log N(\eta^*,\mathscr{T}_\rho,\|\cdot\|_{\infty,M})\\ \le\ &\big\{4\big(6\widetilde L^2(\eta)+1\big)\big(2M/\widetilde L(\eta)+3\big)+2\big\}|\log\eta|,\end{aligned} \tag{4.17}$$

where

$$\begin{aligned}\eta^* \ =\ &\frac{\eta}{\rho}\bigg\{\widetilde L(\eta)\Big(\frac{\sigma_u^2}{\sigma_l}+\frac{\sigma_u^3}{\sigma_l^2}+\frac{\sigma_u^2}{\sqrt{2\pi e}\sigma_l^2}+\frac{\sigma_u^2}{\sqrt{2\pi}\sigma_l}\Big)\\ &+\frac{2\sigma_u^3}{\sqrt{12\pi}\sigma_l^2}+\frac{\sigma_u}{\sqrt{12\pi}}+\frac{2\sigma_u^3}{\sqrt{2\pi}e^{3/2}\sigma_l^3}+\frac{\sigma_u^3}{\sqrt{2\pi e}\sigma_l^2}\bigg\}.\end{aligned} \tag{4.18}$$

5. Numerical studies

In order to investigate the adaptivity of the GMLEB to different heteroscedastic mean vectors, we carries out a simulation study. In Table 1, θ_i are drawn from

TABLE 2
Average of $\sum_i(\hat{\theta}_i-\theta_i)^2$ based on 100 replications: $n=1000$, $\sigma_i\sim\mathrm{Unif}(0.5,1.5)$, $\theta_i\sim(1-p)\delta_0+pN(3,\tau^2)$.

τ^2	0.1				1				10			
p	0.2	0.4	0.6	0.8	0.2	0.4	0.6	0.8	0.2	0.4	0.6	0.8
James-Stein	640	749	750	651	669	782	801	748	846	935	967	982
SURE-M	578	683	687	588	606	718	739	686	782	890	931	946
SURE-SG	583	696	701	595	608	729	751	695	769	890	935	950
Group-linear	587	693	696	597	613	725	746	694	781	892	934	949
NEST	459	576	588	492	458	618	689	673	446	650	811	919
GMLEB	**355**	**481**	**499**	**407**	**374**	**557**	**646**	**649**	**372**	**607**	**790**	**915**
Oracle	349	474	493	401	366	549	638	640	360	595	775	898

two points: 0 or μ. The number of nonzero θ_i is 5, 50 or 500. The values of μ are 1.5, 2, 2.5 and 3. The scales are generated by $\sigma_i\sim\mathrm{Unif}(0.5,1.5)$ independently. We report the sum of squared loss $\sum_i(\hat{\theta}_i-\theta_i)^2$ for $n=1000$ based on average of 100 replications. We display our simulation results for five estimators: the extended James-Stein [2], the shrinkage estimator SURE-M and the semiparametric shrinkage estimator SURE-SG [18], the group-linear method [17], the NEST [6] and the GMLEB. We also display Oracle as the risk of the oracle Bayes rule $t^*_{G^*_n}(\cdot,\sigma_i)$ in (1.8). In each column, boldface entry represents the best performer. The sum of squared loss of the GMLEB happens to be the smallest among the reported estimators and tracks the oracle risk very well. Indeed, here the oracle Bayes rule in (1.8) is nonlinear.

In Table 2, we report another simulation for independent θ_i and σ_i^2. The means are generated by $\theta_i\sim(1-p)\delta_0+pN(3,\tau^2)$ where δ_u is the degenerate distribution at u. We set $p=0.2$ to 0.8 with an increment of 0.2, and $\tau^2=0.1$, 1 or 10. The GMLEB is the best throughout all combinations.

6. Proofs

Proof of Theorem 1. We use M_0 to denote a universal real constant which may take a different value on each occurrence. For simplicity, we take $q_n=(e\sqrt{2\pi}/n^2)\wedge 1$ in (4.5) so that (4.4) holds with $\rho_n=1/n^3$. Let ε_n and x_* be as in Theorem 4 and $\widetilde{L}(\rho)=\sqrt{-\log(2\pi\rho^2)}$ be as in (4.7). Let η^* be as in (4.18) and

$$\eta=\frac{\rho_n}{n}=\frac{1}{n^4},\quad M=\frac{2\sigma_u n\varepsilon_n^2}{(\log n)^{3/2}}. \tag{6.1}$$

Let $x^*=\max(x_*,1)$ and $\big\{(t^*_{H_j}(\cdot,\sigma_1;\rho_n),\ldots,t^*_{H_j}(\cdot,\sigma_n;\rho_n)),j\le N\big\}$ be a $(2\eta^*)$-net of (2.7) under $\|\cdot\|_{\infty,M}$.

As we have described in the outline, we divide the proof into four steps.

STEP 1. Let $A_n=\big\{\overline{d}(\widehat{G}_n,G^*_n)\le x^*\varepsilon_n\big\}$ with $x^*=\max(x_*,1)$ and ζ_{1n} be as in (2.6). Since $x^*\ge 1$ and $n\varepsilon_n^2\ge 2(\log n)^2$ by (2.3), it follows from Theorem 4 that $\mathbb{P}_{G^*_n}\{A_n^c\}\le\exp\big\{-(x^*)^2n\varepsilon_n^2/(2\log n)\big\}\le 1/n$. Thus, since

$\widetilde{L}^2(\rho_n) = -\log(2\pi/n^6) \le M_0 \log n$, Lemma 2 gives

$$\begin{aligned}
\mathbb{E}_{G_n^*}\zeta_{1n}^2 &= \mathbb{E}_{G_n^*}\sum_{i=1}^n \Big\{\big(t^*_{\widehat{G}_n}(X_i,\sigma_i;\rho_n) - X_i\big) + \big(X_i - \theta_i\big)\Big\}^2 I_{A_n^c} \\
&\le 2\sum_{i=1}^n \sigma_i^2 \widetilde{L}^2(\rho_n)\mathbb{P}_{G_n^*}\{A_n^c\} + 2\mathbb{E}_{G_n^*}\sum_{i=1}^n (X_i-\theta_i)^2 I_{A_n^c} \\
&\le M_0 \log n + 2\sum_{i=1}^n \int_0^\infty \min\big(P\{|N(0,\sigma_i^2)| > x\}, 1/n\big)dx^2. \quad (6.2)
\end{aligned}$$

Since $P\{N(0,1) > x\} \le e^{-x^2/2}$, we have

$$\begin{aligned}
&\sum_{i=1}^n \int_0^\infty \min\big(P\{|N(0,\sigma_i^2)| > x\}, 1/n\big)dx^2 \\
\le\; & \int_0^\infty \min\big(2ne^{-x^2/(2\sigma_u^2)}, 1\big)dx^2 = 2\sigma_u^2 + 2\sigma_u^2 \log(2n). \qquad (6.3)
\end{aligned}$$

By (6.2) and (6.3),

$$\mathbb{E}_{G_n^*}\zeta_{1n}^2 \le M_0 \log n \le M_0 n\varepsilon_n^2. \qquad (6.4)$$

Step 2. In this step, we bound $\mathbb{E}_{G_n^*}\zeta_{2n}^2$. Since $\big\{(t^*_{H_j}(\cdot,\sigma_1;\rho_n),\ldots,t^*_{H_j}(\cdot,\sigma_n;\rho_n)), j \le N\big\}$ form a $(2\eta^*)$-net of (2.7) under $\|\cdot\|_{\infty,M}$, it follows from Lemma 2 and (2.8) that

$$\begin{aligned}
\zeta_{2n}^2 &\le \min_{j\le N}\sum_{i=1}^n \big(t^*_{\widehat{G}_n}(X_i,\sigma_i;\rho_n) - t^*_{H_j}(X_i,\sigma_i;\rho_n)\big)^2 I_{A_n} \\
&\le (2\eta^*)^2 \#\big\{i\colon |X_i| \le M\big\} + \big\{2\sigma_u \widetilde{L}(\rho_n)\big\}^2 \#\big\{i\colon |X_i| > M\big\}.
\end{aligned}$$

By (2.3), $(n\varepsilon_n^2/\log n)^{p+1} \ge n\big\{\sqrt{\log n}\,\mu_p(G_n^*)\big\}^p$, so that by (6.1),

$$\int_{|u|\ge M/2} dG_n^*(u) \le \Big(\frac{\mu_p(G_n^*)}{M/2}\Big)^p \le \Big(\frac{2n\varepsilon_n^2}{M(\log n)^{3/2}}\Big)^p \frac{\varepsilon_n^2}{\log n} = \Big(\frac{1}{\sigma_u}\Big)^p \frac{\varepsilon_n^2}{\log n}. \quad (6.5)$$

Thus, since $\eta^* = n^{-1}\big\{c_1\widetilde{L}(n^{-4}) + c_2\big\}$ by (4.18) and $M \ge 4\sigma_u\sqrt{\log n}$ by (6.1) and (2.3),

$$\begin{aligned}
\mathbb{E}_{G_n^*}\zeta_{2n}^2 &\le n(2\eta^*)^2 + 4\sigma_u^2\widetilde{L}^2(n^{-3})\mathbb{E}_{G_n^*}\#\big\{i\colon |X_i| > M\big\} \\
&\le M_0(\log n)\Big(\frac{1}{n} + n\int_{|u|\ge M/2} dG_n(u) + \sum_{i=1}^n P\big\{|N(0,\sigma_i^2)| > 2\sigma_u\sqrt{\log n}\big\}\Big) \\
&\le M_0(\log n)\Big(\frac{1}{n} + \frac{n\varepsilon_n^2}{\log n} + \sum_{i=1}^n P\big\{|N(0,1)| > 2\sqrt{\log n}\big\}\Big) \\
&\le M_0(\log n)\Big(\frac{1}{n} + \frac{n\varepsilon_n^2}{\log n} + \frac{2}{n}\Big).
\end{aligned}$$

Since $n\varepsilon_n^2 \geq 2(\log n)^2$, we find

$$\mathbb{E}_{G_n^*}\zeta_{2n}^2 \leq M_0 n\varepsilon_n^2. \tag{6.6}$$

Step 3. In this step, we bound $\mathbb{E}_{G_n^*}\zeta_{3n}^2$. Let $h(\boldsymbol{x}) = \big(\sum_{i=1}^n (t^*_{\widehat{G}_n}(x_i,\sigma_i;\rho) - \theta_i)^2\big)^{1/2}$. It follows from Lemma 2 that for $0 < \rho \leq (2\pi e^3)^{-1/2}$,

$$\begin{aligned}
|h(\boldsymbol{x}) - h(\boldsymbol{y})| &\leq \Big\{\sum_{i=1}^n \big(t^*_G(x_i,\sigma_i;\rho) - t^*_G(y_i,\sigma_i;\rho)\big)^2\Big\}^{1/2} \\
&\leq \|\boldsymbol{x}-\boldsymbol{y}\| \max_{i\leq n}\sup_x |(\partial/\partial x)t^*_G(x,\sigma_i;\rho)| \\
&\leq \widetilde{L}^2(\rho)\|\boldsymbol{x}-\boldsymbol{y}\|.
\end{aligned}$$

Thus, $h(\boldsymbol{x})/\widetilde{L}^2(\rho)$ has the unit Lipschitz norm. The Gaussian isoperimetric inequality (e.g., [16]) gives that for any deterministic distribution G and $x > 0$,

$$\begin{aligned}
\mathbb{P}_{G_n^*}\Big\{\Big(\sum_{i=1}^n \big(t^*_G(X_i,\sigma_i;\rho) - \theta_i\big)^2\Big)^{1/2} &\geq \mathbb{E}_{G_n^*}\Big(\sum_{i=1}^n \big(t^*_G(X_i,\sigma_i;\rho) - \theta_i\big)^2\Big)^{1/2} + x\Big\} \\
&\leq \exp\Big(-\frac{x^2}{2\widetilde{L}^4(\rho)}\Big).
\end{aligned}$$

This and (2.9) imply that

$$\begin{aligned}
\mathbb{E}_{G_n^*}\zeta_{3n}^2 &= \int_0^\infty \mathbb{P}_{G_n^*}\big\{\zeta_{3n} > x\big\}dx^2 \\
&\leq \int_0^\infty \min\Big\{1, N\exp\big(-x^2/(2\widetilde{L}^4(\rho_n))\big)\Big\}dx^2 \\
&= 2\widetilde{L}^4(\rho_n)(1+\log N).
\end{aligned} \tag{6.7}$$

The entropy bound for regularized Bayes rules in Theorem 7 and (6.1) give that

$$\log N \leq M_0(\log n)^{3/2}M/2 \leq M_0 n\varepsilon_n^2. \tag{6.8}$$

Hence by (6.7) and (6.8),

$$\mathbb{E}_{G_n^*}\zeta_{3n}^2 \leq M_0 n\varepsilon_n^2(\log n)^2. \tag{6.9}$$

Step 4. In this step, we bound $\mathbb{E}_{G_n^*}\zeta_{4n}^2$. First of all, it follows from (2.10) that

$$\zeta_{4n}^2 \leq \max_{j\leq N}\Big\{\mathbb{E}_{G_n^*}\sum_{i=1}^n \big(t^*_{H_j}(X_i,\sigma_i;\rho_n) - \theta_i\big)^2 - nR_n^*(G_n^*)\Big\}. \tag{6.10}$$

By Theorem 6, for any $0 < \rho_n \le (2\pi e^2)^{-1/2}$ and $x_0 > 0$,

$$\begin{aligned}
& \mathbb{E}_{G_n^*} \sum_{i=1}^n \big(t^*_{H_j}(X_i, \sigma_i; \rho_n) - \theta_i\big)^2 - nR_n^*(G_n^*) \\
\le \;& M_0 \sum_{i=1}^n \sigma_i^2 \max\big\{|\log \rho_n|^3, |\log d(f_{H_j,\sigma_i}, f_{G_n^*,\sigma_i})|\big\} d^2(f_{H_j,\sigma_i}, f_{G_n^*,\sigma_i}) \\
& +2\sum_{i=1}^n \sigma_i^2 \Big\{\mathbb{P}_{G_n^*}\big\{|\theta_i/\sigma_i| > x_0\big\} + 2x_0\rho_n \widetilde{L}^2(\rho_n) \\
& +2\rho_n\big(\widetilde{L}^2(\rho_n) + 2\big)^{1/2}\Big\}. \qquad (6.11)
\end{aligned}$$

Let $x_0 = M/(2\sigma_l)$ and $\varepsilon_0 = x^*\varepsilon_n \ge \overline{d}(H_j, G_n^*)$. It follows from (6.5) that

$$\frac{\mathbb{P}_{G_n^*}\big\{|\theta_i/\sigma_i| > x_0\big\}}{|\log \rho_n|^3 (x^*\varepsilon_n)^2} \le \frac{\int_{|u|\ge M/2} dG_n^*(u)}{|\log \rho_n|^3 (x^*\varepsilon_n)^2} \le \frac{(1/\sigma_u)^p(\varepsilon_n^2/\log n)}{(\log n)^3\varepsilon_n^2} \le M_0. \qquad (6.12)$$

Since $M = 2\sigma_u n\varepsilon_n^2/(\log n)^{3/2}$ and $\widetilde{L}^2(\rho_n) \le M_0 \log n$,

$$\begin{aligned}
\frac{2(M/(2\sigma_l)+1)\rho_n\widetilde{L}^2(\rho_n)}{(\log\rho_n)^3(x^*\varepsilon_n)^2} &\le \frac{M_0(n\varepsilon_n^2/(\log n)^{3/2}+1)/n^3}{(\log n)^2\varepsilon_n^2} \\
&\le \frac{M_0}{n^2(\log n)^{7/2}} \le M_0. \qquad (6.13)
\end{aligned}$$

Thus, by (6.10)–(6.13),

$$\zeta_{4n}^2 \le M_0 n |(\log \rho_n)/3|^3 \varepsilon_n^2 = M_0 n \varepsilon_n^2 (\log n)^3. \qquad (6.14)$$

Adding (6.4), (6.6), (6.9) and (6.14) together, we have

$$\sum_{j=1}^4 \big(\mathbb{E}_{G_n^*}\zeta_{jn}^2\big)^{1/2} \le M_0 n^{1/2}\varepsilon_n(\log n)^{3/2}.$$

This and (2.11) complete the proof. □

Lemma 3. *Let $a > 0$, $\eta = \varphi(a\sigma_l/\sigma_u)$ and $M > 0$. Given any mixing distribution G, there exists a discrete mixing distribution G_m with support $[-M - a\sigma_l, M + a\sigma_l]$ and at most $m = (2\lfloor 6a^2\rfloor + 1)\lceil 2M/(a\sigma_l) + 2\rceil + 1$ atoms, such that*

$$\max_{i\le n} \big\|f_{G,\sigma_i} - f_{G_m,\sigma_i}\big\|_{\infty,M} \le \Big(1 + \frac{1}{\sqrt{2\pi}}\Big)\frac{\eta}{\sigma_l}. \qquad (6.15)$$

Proof of Lemma 3. Let $j^* = \lceil 2M/(a\sigma_l) + 2\rceil$ and $k^* = \lfloor 6a^2\rfloor$. Let

$$I_j = \big(-M + (j-2)a\sigma_l, (-M + (j-1)a\sigma_l) \wedge (M + a\sigma_l)\big],\ j = 1, \dots, j^*,$$

be a partition of $(-M - a\sigma_l, M + a\sigma_l]$. It follows from the Carathéodory's theorem (e.g., [11]) that for each distribution function G there exists a discrete distribution function G_m with support $[-M - a\sigma_l, M + a\sigma_l]$ and no more than $m = (2k^* + 1)j^* + 1$ support points such that

$$\int_{I_j} u^k dG(u) = \int_{I_j} u^k dG_m(u),\ k = 0, 1, \ldots, 2k^*,\ j = 1, \ldots, j^*.$$

Since the Taylor expansion of $e^{-t^2/2}$ has alternating signs, for $t^2/2 \leq k^* + 2$,

$$0 \leq \mathrm{Rem}(t) = \left|\varphi(t) - \sum_{k=0}^{k^*} \frac{(-t^2/2)^k}{\sqrt{2\pi}k!}\right| \leq \frac{(t^2/2)^{k^*+1}}{\sqrt{2\pi}(k^*+1)!}.$$

Thus, since $k^* + 1 \geq 6a^2$, for $x \in I_j \cap [-M, M]$, the Stirling formula yields

$$\begin{aligned} \big|f_{G,\sigma_i}(x) - f_{G_m,\sigma_i}(x)\big| &\leq \left|\int_{(I_{j-1}\cup I_j \cup I_{j+1})^c} \frac{1}{\sigma_i}\varphi\Big(\frac{x-u}{\sigma_i}\Big) d\big(G(u) - G_m(u)\big)\right| \\ &\quad + \left|\int_{I_{j-1}\cup I_j \cup I_{j+1}} \frac{1}{\sigma_i}\mathrm{Rem}\Big(\frac{x-u}{\sigma_i}\Big) d\big(G(u) - G_m(u)\big)\right| \\ &\leq \frac{1}{\sigma_l}\varphi\Big(\frac{a\sigma_l}{\sigma_u}\Big) + \frac{((2a)^2/2)^{k^*+1}}{\sigma_l\sqrt{2\pi}(k^*+1)!} \\ &\leq \frac{1}{\sigma_l}\varphi\Big(\frac{a\sigma_l}{\sigma_u}\Big) + \frac{(e/3)^{k^*+1}}{\sigma_l 2\pi (k^*+1)^{1/2}}. \end{aligned} \tag{6.16}$$

Furthermore, since $(e/3)^6 \leq e^{-1/2}$ and $k^* + 1 \geq 6a^2 \geq 6(a\sigma_l/\sigma_u)^2$, we have $(e/3)^{k^*+1} \leq e^{-(a\sigma_l/\sigma_u)^2/2}$. Hence (6.15) follows from (6.16), $(e/3)^{k^*+1} \leq e^{-(a\sigma_l/\sigma_u)^2/2}$ and $\eta = \varphi(a\sigma_l/\sigma_u)$. □

Lemma 4. *There exists a universal constant C such that*

$$\log N(\eta, \mathscr{F}_n, \|\cdot\|_{\infty,M}) \leq C|\log\eta|^2 \max\Big(\frac{M}{\sqrt{|\log\eta|}}, 1\Big), \tag{6.17}$$

for all $0 < \eta \leq (2\pi)^{-1/2}$ and $M > 0$.

Proof of Lemma 4. Let a be the value such that $\eta = \varphi(a\sigma_l/\sigma_u)$ and

$$m = (2\lfloor 6a^2\rfloor + 1)\lceil 2M/(a\sigma_l) + 2\rceil + 1 \leq C|\log\eta| \max\Big(\frac{M}{\sqrt{|\log\eta|}}, 1\Big). \tag{6.18}$$

It follows from Lemma 3 that there exists a discrete distribution G_m with support $[-M - a\sigma_l, M + a\sigma_l]$ and at most m atoms such that

$$\max_{i\leq n} \big\|f_{G,\sigma_i} - f_{G_m,\sigma_i}\big\|_{\infty,M} \leq \Big(1 + \frac{1}{\sqrt{2\pi}}\Big)\frac{\eta}{\sigma_l}. \tag{6.19}$$

The next step is to approximate the f_{G_m,σ_i} in (6.19) by $f_{G_{m,\eta},\sigma_i}$ where $G_{m,\eta}$ is supported in a lattice and has no more than m atoms. Let $\xi \sim G_m$ and $\xi_\eta = \eta\,\mathrm{sgn}(\xi)\lfloor|\xi|/\eta\rfloor$. Define $G_{m,\eta}$ as the distribution of ξ_η. The support of $G_{m,\eta}$ is in the grid $\Omega_{\eta,M} = \{0, \pm\eta, \pm 2\eta, \ldots\} \cap [-M - a\sigma_l, M + a\sigma_l]$. Since $|\xi - \xi_\eta| \le \eta$ and $\sup_x (\partial/\partial x)|(1/\sigma_i)\varphi(x/\sigma_i)| = 1/(\sqrt{2\pi e}\sigma_i^2)$,

$$\max_{i\le n} \left\| f_{G_m,\sigma_i} - f_{G_{m,\eta},\sigma_i} \right\|_\infty \le \frac{1}{\sqrt{2\pi e}\sigma_l^2}\eta. \tag{6.20}$$

The last step is to bound the covering number of the collection of all $f_{G_{m,\eta},\sigma_i}$. Let $\mathscr{P}^m$ be the set of all vectors $\boldsymbol{w} = (w_1, \ldots, w_m)$ satisfying $w_j \ge 0$ and $\sum_{j=1}^m w_j = 1$. Let $\mathscr{P}^{m,\eta}$ be an η-net of $\mathscr{P}^m$:

$$\inf_{\boldsymbol{w}^{m,\eta}\in\mathscr{P}^{m,\eta}} \left\| \boldsymbol{w} - \boldsymbol{w}^{m,\eta} \right\|_1 \le \eta, \quad \forall\, \boldsymbol{w} \in \mathscr{P}^m,$$

with $N(\eta, \mathscr{P}^m, \|\cdot\|_1)$ elements. Let $\{u_j, j = 1, \ldots, m\}$ be the support of $G_{m,\eta}$ and $\boldsymbol{w}^{m,\eta}$ be a probability vector in $\mathscr{P}^{m,\eta}$ satisfying $\sum_{j=1}^m |G_{m,\eta}(\{u_j\}) - w_j^{m,\eta}| \le \eta$. Denote $\widetilde{G}_{m,\eta} = \sum_{j=1}^m w_j^{m,\eta}\delta_{u_j}$. Then,

$$\max_{i\le n} \left\| f_{G_{m,\eta},\sigma_i} - f_{\widetilde{G}_{m,\eta},\sigma_i} \right\|_\infty \le \frac{1}{\sqrt{2\pi}\sigma_l}\eta, \tag{6.21}$$

since $\sup_x |(1/\sigma_i)\varphi(x/\sigma_i)| = 1/(\sqrt{2\pi}\sigma_i)$.

Summing (6.19), (6.20) and (6.21) together, we have

$$\begin{aligned}\max_{i\le n} \left\| f_{G,\sigma_i} - f_{\widetilde{G}_{m,\eta},\sigma_i} \right\|_{\infty,M} &\le \Big(1 + \frac{1}{\sqrt{2\pi}} + \frac{1}{\sqrt{2\pi e}\sigma_l} + \frac{1}{\sqrt{2\pi}}\Big)\frac{\eta}{\sigma_l} \\ &\equiv \eta^{**}.\end{aligned} \tag{6.22}$$

The support of $\widetilde{G}_{m,\eta}$ is also in $\Omega_{\eta,M}$. Counting the number of ways to realize $\{u_j\}$ and $\boldsymbol{w}^{m,\eta}$, we find

$$N(\eta^{**}, \mathscr{F}_n, \|\cdot\|_{\infty,M}) \le \binom{|\Omega_{\eta,M}|}{m} N(\eta, \mathscr{P}^m, \|\cdot\|_1), \tag{6.23}$$

with m satisfying (6.18) and $|\Omega_{\eta,M}| = 1 + 2\lfloor(M + a\sigma_l)/\eta\rfloor$.

Since $\mathscr{P}^m$ is in the ℓ_1 unit-sphere of $\mathbb{R}^m$, $N(\eta, \mathscr{P}^m, \|\cdot\|_1)$ is no greater than the maximum number of disjoint $\mathrm{Ball}(\boldsymbol{v}_j, \eta/2, \|\cdot\|_1)$ with $\|\boldsymbol{v}_j\| = 1$. Here is the argument. Suppose there exists $\boldsymbol{w} \in \mathscr{P}^m$ such that $\boldsymbol{w} \notin \cup_j \mathrm{Ball}(\boldsymbol{v}_j, \eta, \|\cdot\|_1)$. Then $\mathrm{Ball}(\boldsymbol{w}, \eta/2, \|\cdot\|_1) \cap \mathrm{Ball}(\boldsymbol{v}_j, \eta/2, \|\cdot\|_1) = \emptyset$, $\forall j$. This is a contradiction with the maximum number of disjoint balls. Hence $\mathscr{P}^m \subset \cup_j \mathrm{Ball}(\boldsymbol{v}_j, \eta, \|\cdot\|_1)$.

Since all these disjoint $\mathrm{Ball}(\boldsymbol{v}_j, \eta/2, \|\cdot\|_1)$ are inside the $(1 + \eta/2)$ ℓ_1-ball, volume comparison yields $N(\eta, \mathscr{P}^m, \|\cdot\|_1) \le (2/\eta + 1)^m$. This, (6.23) and another application of the Stirling formula yield

$$\begin{aligned}&N(\eta^{**}, \mathscr{F}_n, \|\cdot\|_{\infty,M}) \\ \le\ &\frac{(2/\eta + 1)^m |\Omega_{\eta,M}|^m}{m!}\end{aligned}$$

$$
\begin{aligned}
&\le \Big\{\Big(1+\frac{2}{\eta}\Big)\Big(1+\frac{2(M+a\sigma_l)}{\eta}\Big)\Big\}^m \big(m^{m+1/2}e^{-m}\sqrt{2\pi}\big)^{-1} \\
&\le \Big\{\frac{(\eta+2)\big(\eta+2(M+a\sigma_l)\big)e}{m}\Big\}^m \eta^{-2m}(2\pi m)^{-1/2}. \qquad (6.24)
\end{aligned}
$$

For $\eta = o(1)$, $a \to \infty$, so that $m \ge (1+o(1))24a(M/\sigma_l + a) \to \infty$ and $(\eta+2)(\eta+2(M+a\sigma_l))e = (1+o(1))4e(M+a\sigma_l) \le m$. Hence, $N(\eta^{**}, \mathscr{F}_n, \|\cdot\|_{\infty,M}) \le \eta^{-2m}$ by (6.24). This and the definition of η^{**} in (6.22) give (6.17). $\square$

Lemma 5. *Suppose that under $P_{G_n^*}$, $\theta_1, \dots, \theta_n$ are i.i.d. random variables from a distribution G_n^*, and given θ_i's, $X_i \sim N(\theta_i, \sigma_i^2)$ are independent observations with known variances. Then for all constants $M \ge \sigma_u\sqrt{8\log n}$, $0 < \lambda < \min(1,p)$, and $a_1, \dots, a_n > 0$,*

$$
\mathbb{E}_{G_n^*}\Big\{\prod_{i=1}^n |a_iX_i|^{I\{|X_i|\ge M\}}\Big\}^\lambda \le \exp\Big\{\sum_{i=1}^n (a_iM)^\lambda\Big(\frac{4\sigma_u}{Mn\sqrt{2\pi}} + \Big(\frac{2\mu_p(G_n^*)}{M}\Big)^p\Big)\Big\}.
$$

Proof of Lemma 5. It follows that

$$
\begin{aligned}
&\mathbb{E}_{G_n^*}\Big\{\prod_{i=1}^n |a_iX_i|^{I\{|X_i|\ge M\}}\Big\}^\lambda \\
&\le \prod_{i=1}^n \Big(1 + a_i^\lambda E|X_i|^\lambda I\big\{|X_i| \ge M\big\}\Big) \\
&\le \exp\Big\{\sum_{i=1}^n a_i^\lambda \int_{|x|\ge M} |x|^\lambda f_{G_n^*,\sigma_i}(x)dx\Big\}. \qquad (6.25)
\end{aligned}
$$

Let $Z \sim N(0, \sigma_i^2)$ and $\theta \sim G_n^*$. Since $Z + \theta \sim f_{G_n^*,\sigma_i}$ and $0 < \lambda \le 1$,

$$
\begin{aligned}
&\int_{|x|\ge M} |x|^\lambda f_{G_n^*,\sigma_i}(x)dx \\
&= E|Z+\theta|^\lambda I\big\{|Z+\theta| \ge M\big\} \\
&\le E|2Z|^\lambda I\Big\{|Z| \ge \frac{M}{2}\Big\} + E|2\theta|^\lambda I\Big\{|\theta| \ge \frac{M}{2}\Big\} \\
&\le 2M^{\lambda-1}E|Z|I\Big\{|Z| \ge \frac{M}{2}\Big\} + \int_{|x|\ge\frac{M}{2}} (2|x|)^\lambda dG_n^*(x). \qquad (6.26)
\end{aligned}
$$

Since $\lambda < p$, it follows from the Hölder and Markov inequalities that

$$
\begin{aligned}
\int_{|x|\ge\frac{M}{2}} (2|x|)^\lambda dG_n^*(x) &\le 2^\lambda \mu_p^\lambda(G_n^*)\Big(\mathbb{P}_{G_n^*}\big\{|x| \ge M/2\big\}\Big)^{1-\lambda/p} \\
&\le M^\lambda\Big(\frac{2\mu_p(G_n^*)}{M}\Big)^p.
\end{aligned}
$$

Moreover, since $M \ge \sigma_u\sqrt{8\log n}$,

$$
2M^{\lambda-1}E|Z|I\Big\{|Z| \ge \frac{M}{2}\Big\} \le 2M^{\lambda-1}\sigma_u E|Z/\sigma_i|I\Big\{|Z/\sigma_i| \ge \frac{M}{2\sigma_u}\Big\}
$$

$$
\begin{aligned}
&= 4M^{\lambda-1}\sigma_u \int_{\frac{M}{2\sigma_u}}^{\infty} x\varphi(x)dx \\
&\le \frac{4M^{\lambda}\sigma_u}{Mn\sqrt{2\pi}}.
\end{aligned}
$$

Inserting the above inequalities into (6.26) yields

$$
\int_{|x|\ge M} |x|^{\lambda} f_{G_n^*}(x)dx \le M^{\lambda}\Big\{\frac{4\sigma_u}{Mn\sqrt{2\pi}} + \Big(\frac{2\mu_p(G_n^*)}{M}\Big)^p\Big\}.
$$

This and (6.25) imply the conclusion. □

Proof of Theorem 7. It follows from (4.1) and Lemma 2 that

$$
\begin{aligned}
&\big|t_G^*(x,\sigma_i;\rho) - t_H^*(x,\sigma_i;\rho)\big| \\
\le\ & \frac{\sigma_i^3}{\rho}\big|f'_{G,\sigma_i}(x) - f'_{H,\sigma_i}(x)\big| + \frac{\sigma_i^2\widetilde{L}(\rho)}{\rho}\big|f_{G,\sigma_i}(x) - f_{H,\sigma_i}(x)\big|. \qquad (6.27)
\end{aligned}
$$

Let a satisfying $\eta = \varphi(a\sigma_l/\sigma_u)$ so that $a\sigma_l/\sigma_u = \widetilde{L}(\eta)$. Let $j^* = \lceil 2M/(a\sigma_l) + 2\rceil$ and $k^* = \lfloor 6a^2\rfloor$. Let

$$
I_j = \big(-M + (j-2)a\sigma_l, (-M + (j-1)a\sigma_l) \wedge (M + a\sigma_l)\big],\ j = 1,\dots,j^*,
$$

be a partition of $(-M-a\sigma_l, M+a\sigma_l]$. It follows from the Carathéodory's theorem that for each distribution function G there exists a discrete distribution function G_m with support $[-M-a\sigma_l, M+a\sigma_l]$ and no more than $m = (2k^*+2)j^*+1$ support points such that

$$
\int_{I_j} u^k dG(u) = \int_{I_j} u^k dG_m(u),\ k = 0,1,\dots,2k^*+1,\ j = 1,\dots,j^*.
$$

Since the Taylor expansion of $e^{-t^2/2}$ has alternating signs, for $t^2/2 \le k^*+2$,

$$
0 \le \mathrm{Rem}(t) = \Big|\varphi(t) - \sum_{k=0}^{k^*}\frac{(-t^2/2)^k}{\sqrt{2\pi}k!}\Big| \le \frac{(t^2/2)^{k^*+1}}{\sqrt{2\pi}(k^*+1)!}.
$$

Thus, since $k^*+1 \ge 6a^2$ and $a\sigma_l/\sigma_u \ge 1$, for $x \in I_j \cap [-M,M]$, the Stirling formula yields

$$
\begin{aligned}
&\big|f'_{G,\sigma_i}(x) - f'_{G_m,\sigma_i}(x)\big| \\
\le\ & \Big|\int_{(I_{j-1}\cup I_j\cup I_{j+1})^c}\Big(\frac{u-x}{\sigma_i^3}\Big)\varphi\Big(\frac{x-u}{\sigma_i}\Big)d\big(G(u)-G_m(u)\big)\Big| \\
&+\Big|\int_{I_{j-1}\cup I_j\cup I_{j+1}}\Big(\frac{u-x}{\sigma_i^3}\Big)\mathrm{Rem}\Big(\frac{x-u}{\sigma_i}\Big)d\big(G(u)-G_m(u)\big)\Big| \\
\le\ & \frac{1}{\sigma_l^2}\max_{t\ge a\sigma_l/\sigma_u} t\varphi(t) + \frac{2a\big((2a)^2/2\big)^{k^*+1}}{\sigma_l^2\sqrt{2\pi}(k^*+1)!}
\end{aligned}
$$

$$\leq \quad \frac{1}{\sigma_l^2}\eta\widetilde{L}(\eta) + \frac{2a(e/3)^{k^*+1}}{\sigma_l^2 2\pi(k^*+1)^{1/2}}. \tag{6.28}$$

Similarly, for $|x| \leq M$,

$$\big|f_{G,\sigma_i}(x) - f_{G_m,\sigma_i}(x)\big| \leq \frac{\eta}{\sigma_l} + \frac{(e/3)^{k^*+1}}{\sigma_l 2\pi(k^*+1)^{1/2}}. \tag{6.29}$$

Furthermore, since $(e/3)^6 \leq e^{-1/2}$ and $k^*+1 \geq 6a^2 \geq 6(a\sigma_l/\sigma_u)^2$, we have $(e/3)^{k^*+1} \leq e^{-(a\sigma_l/\sigma_u)^2/2}$. So by (6.27), (6.28) and (6.29),

$$\begin{aligned}
&\max_{i\leq n}\big\|t^*_G(\cdot,\sigma_i;\rho) - t^*_{G_m}(\cdot,\sigma_i;\rho)\big\|_{\infty,M} \\
\leq \quad & \frac{\sigma_u^3}{\sigma_l^2\rho}\Big(\eta\widetilde{L}(\eta) + \frac{2ae^{-(a\sigma_l/\sigma_u)^2/2}}{2\pi\sqrt{6a^2}}\Big) + \frac{\sigma_u^2\widetilde{L}(\rho)}{\sigma_l\rho}\Big(\eta + \frac{e^{-(a\sigma_l/\sigma_u)^2/2}}{2\pi\sqrt{6a^2}}\Big) \\
\leq \quad & \frac{\eta}{\rho}\Big\{\widetilde{L}(\eta)\Big(\frac{\sigma_u^2}{\sigma_l} + \frac{\sigma_u^3}{\sigma_l^2}\Big) + \frac{2\sigma_u^3}{\sqrt{12\pi}\sigma_l^2} + \frac{\sigma_u}{\sqrt{12\pi}}\Big\}.
\end{aligned} \tag{6.30}$$

Let $\xi \sim G_m$, $\xi_\eta = \eta\,\mathrm{sgn}(\xi)\lfloor|\xi|/\eta\rfloor$ and $G_{m,\eta} \sim \xi_\eta$. Since $\sup_x(\partial/\partial x)|(1/\sigma_i)\varphi(x/\sigma_i)| = 1/(\sqrt{2\pi e}\sigma_i^2)$ and $\sup_x(\partial^2/\partial x^2)|(1/\sigma_i)\varphi(x/\sigma_i)| = \sqrt{2/\pi}/(e^{3/2}\sigma_i^3)$ and $|\xi - \xi_\eta| \leq \eta$,

$$\big\|f_{G_m,\sigma_i} - f_{G_{m,\eta},\sigma_i}\big\|_\infty \leq \frac{1}{\sqrt{2\pi e}\sigma_l^2}\eta, \quad \big\|f'_{G_m,\sigma_i} - f'_{G_{m,\eta},\sigma_i}\big\|_\infty \leq \frac{2}{\sqrt{2\pi}e^{3/2}\sigma_l^3}\eta.$$

This and (6.27) imply

$$\max_{i\leq n}\big\|t^*_{G_m}(\cdot,\sigma_i;\rho) - t^*_{G_{m,\eta}}(\cdot,\sigma_i;\rho)\big\|_\infty \leq \frac{\eta}{\rho}\Big\{\frac{2\sigma_u^3}{\sqrt{2\pi}e^{3/2}\sigma_l^3} + \frac{\sigma_u^2\widetilde{L}(\eta)}{\sqrt{2\pi e}\sigma_l^2}\Big\}. \tag{6.31}$$

Moreover, $G_{m,\eta}$ has at most m support points.

Let $\mathscr{P}^m$ be the set of all vectors $\boldsymbol{w} = (w_1,\ldots,w_m)$ satisfying $w_j \geq 0$ and $\sum_{j=1}^m w_j = 1$. Let $\mathscr{P}^{m,\eta}$ be an η-net of $N(\eta,\mathscr{P}^m,\|\cdot\|_1)$ elements in $\mathscr{P}^m$:

$$\inf_{\boldsymbol{w}^{m,\eta}\in\mathscr{P}^{m,\eta}}\big\|\boldsymbol{w} - \boldsymbol{w}^{m,\eta}\big\|_1 \leq \eta, \quad \forall\,\boldsymbol{w}\in\mathscr{P}^m.$$

Let $\{u_j, j=1,\ldots,m\}$ be the support of $G_{m,\eta}$ and $\boldsymbol{w}^{m,\eta}$ be a probability vector in $\mathscr{P}^{m,\eta}$ with $\sum_{j=1}^m|G_{m,\eta}(\{u_j\}) - w_j^{m,\eta}| \leq \eta$. Set $\widetilde{G}_{m,\eta} = \sum_{j=1}^m w_j^{m,\eta}\delta_{u_j}$. Then,

$$\big\|f_{G_{m,\eta},\sigma_i} - f_{\widetilde{G}_{m,\eta},\sigma_i}\big\|_\infty \leq \frac{1}{\sqrt{2\pi}\sigma_l}\eta, \quad \big\|f'_{G_{m,\eta}} - f'_{\widetilde{G}_{m,\eta}}\big\|_\infty \leq \frac{1}{\sqrt{2\pi e}\sigma_l^2}\eta,$$

since $\sup_x|(1/\sigma_i)\varphi(x/\sigma_i)| = 1/(\sqrt{2\pi}\sigma_i)$. This and (6.27) imply

$$\max_{i\leq n}\big\|t^*_{G_{m,\eta}}(\cdot,\sigma_i;\rho) - t^*_{\widetilde{G}_{m,\eta}}(\cdot,\sigma_i;\rho)\big\|_\infty \leq \frac{\eta}{\rho}\Big\{\frac{\sigma_u^3}{\sqrt{2\pi e}\sigma_l^2} + \frac{\sigma_u^2\widetilde{L}(\eta)}{\sqrt{2\pi}\sigma_l}\Big\}. \tag{6.32}$$

The support of $G_{m,\eta}$ and $\widetilde{G}_{m,\eta}$ is $\Omega_{\eta,M} = \{0, \pm\eta, \pm 2\eta, \ldots\} \cap [-M - a\sigma_l, M + a\sigma_l]$.

Summing (6.30), (6.31) and (6.32) together, we find

$$\left\| t^*_G(\cdot, \sigma_i; \rho) - t^*_{\widetilde{G}_{m,\eta}}(\cdot, \sigma_i; \rho) \right\|_{\infty,M} \leq \eta^*,$$

where η^* is as in (4.18). Counting the number of ways to realize $\{u_j\}$ and $\boldsymbol{w}^{m,\eta}$, we find

$$N(\eta^*, \mathscr{T}_\rho, \|\cdot\|_{\infty,M}) \leq \binom{|\Omega_{\eta,M}|}{m} N(\eta, \mathscr{P}^m, \|\cdot\|_1), \tag{6.33}$$

with $m = (2k^* + 2)j^* + 1$, $|\Omega_{\eta,M}| = 1 + 2\lfloor (M + a\sigma_l)/\eta \rfloor$, $\eta = \varphi(a\sigma_l/\sigma_u)$, $j^* = \lceil 2M/(a\sigma_l) + 2 \rceil$ and $k^* = \lfloor 6a^2 \rfloor$.

The rest of the proofs follow the same line in Lemma 4. The bound is $N(\eta^*, \mathscr{T}_\rho, \|\cdot\|_{\infty,M}) \leq \eta^{-2m}$ where $m \leq 2(6a^2 + 1)(2M/(a\sigma_l) + 3) + 1$. $\square$

Acknowledgement

The author would like to thank Professor Cun-Hui Zhang for helpful discussions. The author is also grateful to an Associate Editor and a referee for valuable comments and suggestions.

References

[1] Brown, L. D. (1971). Admissible estimators, recurrent diffusions, and insoluble boundary value problems. *Ann. Math. Statist.* **42** 855–903. MR0286209

[2] Brown, L. D. (2008). In-season prediction of batting averages: A field test of empirical Bayes and Bayes methodologies. *Ann. Appl. Statist.* **2** 113–152. MR2415597

[3] Donoho, D. L. and Johnstone, I. M. (1994). Minimax risk over ℓ_p-balls for ℓ_q-error. *Probab. Theory Related Fields* **99** 277–303. MR1278886

[4] Efron, B. (2011). Tweedie's formula and selection bias. *J. Amer. Statist. Assoc.* **106** 1602–1614. MR2896860

[5] Efron, B. (2014). Two modeling strategies for empirical Bayes estimation. *Statist. Science* **29** 285–301. MR3264543

[6] Fu, L., James, G. and Sun, W. (2018). Nonparametric empirical bayes estimation on heterogeneous data. Arxiv preprint. Available at `https://arxiv.org/abs/2002.12586`.

[7] Jiang, W. and Zhang, C.-H. (2009). General maximum likelihood empirical Bayes estimation of normal means. *Ann. Statist.* **37** 1647–1684. MR2533467

[8] Kiefer, J. and Wolfowitz, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *Ann. Math. Statist.* **27** 887–906. MR0086464

[9] Koenker, R., Gu, J. and Mizera, I. (2015). REBayes: empirical Bayes estimation and inference, R package version 1.3. Available at http://cran.r-project.org. MR3631958

[10] Koenker, R. and Mizera, I. (2014). Convex optimization, shape constraints, compound decisions and empirical Bayes rules. *J. Amer. Statist. Assoc.* **109** 674–685. MR3223742

[11] Lindsay, B. G. (1995). *Mixture Models: Theory, Geometry, and Applications.* IMS, Hayward, CA.

[12] Morris, C. N. (1983). Parametric empirical Bayes inference: theory and applications. *J. Amer. Statist. Assoc.* **78** 47–55. MR0696849

[13] Robbins, H. (1951). Asymptotically subminimax solutions of compound statistical decision problems. *Proc. Second Berkeley Symp. Math. Statist. Probab.* **1** 131–148. University of California Press, Berkeley. MR0044803

[14] Robbins, H. (1956). An empirical Bayes approach to statistics. *Proc. Third Berkeley Symp. Math. Statist. Probab.* **1** 157–163. University of California Press, Berkeley. MR0084919

[15] Robbins, H. (1983). Some thoughts on empirical Bayes estimation. *Ann. Statist.* **11** 713–723. MR0707923

[16] van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes.* Springer, New York. MR1385671

[17] Weinstein, A., Ma, Z., Brown, L. D. and Zhang, C.-H. (2018). Group-linear empirical Bayes estimates for a heteroscedastic normal mean. *J. Amer. Statist. Assoc.* **113** 698–710. MR3832220

[18] Xie, X., Kou, S. and Brown, L. D. (2012). SURE estimates for a heteroscedastic hierarchical model. *J. Amer. Statist. Assoc.* **107** 1465–1479. MR3036408

[19] Zhang, C.-H. (2005). General empirical Bayes wavelet methods and exactly adaptive minimax estimation. *Ann. Statist.* **33** 54–100. MR2157796

[20] Zhang, C.-H. (2009). Generalized maximum likelihood estimation of normal mixture densities. *Statist. Sinica* **19** 1297–1318. MR2536157