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: wenhua.jiang@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 \(((\log n)^5/\sqrt{n})^{1/(p/2)}\). 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 ........................................ 2273
2 Main results ....................................... 2275
  2.1 An oracle inequality for the GMLEB .......... 2276
  2.2 Consequences of the oracle inequality ....... 2277
3 A large deviation inequality for the generalized MLE .... 2279
4 Other elements of the oracle inequality .......... 2282
  4.1 Connection between the GMLEB and the regularized rule .... 2283
  4.2 Some properties of the regularized Bayes estimator .... 2283
  4.3 Regret of a regularized Bayes estimator with a misspecified prior 2285
  4.4 An entropy bound for regularized Bayes rules .... 2286
5 Numerical studies ................................... 2286
6 Proofs ............................................ 2287
Acknowledgement ..................................... 2296
References ........................................... 2296
1. Introduction

In this paper we consider empirical Bayes for heteroscedastic data:

\[ X_i | (\theta_i, \sigma_i^2) \sim N(\theta_i, \sigma_i^2), \quad i = 1, \ldots, n, \]  

(1.1)

where \( \sigma_i^2 \) are known. The problem is to estimate \( \theta = (\theta_1, \ldots, \theta_n) \) under the average squared loss

\[ L_n(\theta, \hat{\theta}) = n^{-1} \| \hat{\theta} - \theta \|^2 = n^{-1} \sum_{i=1}^{n} (\hat{\theta}_i - \theta_i)^2. \]  

(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 \( \theta_i \) are typically treated as constants in the compound approach [13]. In a homoscedastic compound decision problem, the average risk is written as

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

(1.3)

where \( f(x|\theta, \sigma) \) is the density of \( N(\theta, \sigma^2) \), and \( G_n \) is the empirical distribution of \( \theta_i \). Robbins [13, 14] observed that the optimal solution of the above problem is the Bayes rule \( t^*_n G_n(X) = \mathbb{E}_{G_n}(\theta|X = x, \sigma) \). 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 \( \sigma_i \), we may write

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

(1.4)

where \( G_n \) is the empirical distribution of \( (\theta_i, \sigma_i) \). This still connects the compound problem to Bayes. However, the fundamental theorem fails in the presence of heteroscedasticity with observable \( \sigma_i \) 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 \( \sigma \)-value to allow sufficiently accurate estimation of a nonparametric unknown prior.

One plausible way is to take empirical Bayes view that \( \theta_i \) 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 \( \mathcal{D} \), with oracular knowledge of \( G \), the oracle benchmark is \( R_D(G) = \)
\[ \inf_{t \in \mathcal{D}} n^{-1} \mathbb{E}_G \sum_{i=1}^{n} (t(X_i, \sigma_i) - \theta_i)^2. \]

The regret of an estimator \( \hat{t}_n \) is

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

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

\[ r_{G,\mathcal{D}}(\hat{t}_n) \rightarrow 0, \quad n \rightarrow \infty. \tag{1.6} \]

Let \( G \) be a normal distribution with mean \( \mu \) and variance \( \tau^2 \). With \( \mathcal{D} \) being the class of all linear estimators, the optimal estimator in \( \mathcal{D} \) is \( t_{\mathcal{D}}^*(x) = \mu + (1 - B)(x - \mu) \) where \( B = \sigma^2/(\sigma^2 + \tau^2) \). In the homoscedastic case, \( \sigma_i^2 \equiv \sigma^2 \), the James-Stein estimator \( \hat{\theta}_i^{JS} = \overline{X} + (1 - B_n)(X_i - \overline{X}) \) with \( B_n = (n - 3)/\sum_i (X_i - \overline{X})^2 \) approximates the optimal linear rule \( t_{\mathcal{D}}^*(x) \) in the sense of (1.6). In the heteroscedastic case, Xie, Kou and Brown [18] proposed to select an estimator from the class \( \{ \tau^2 X_i/(\sigma_i^2 + \tau^2) + \sigma_i^2 \mu/\sigma_i^2 + \tau^2) : \mu \in \mathbb{R}, \tau^2 > 0 \} \). The parameters \( \mu \) and \( \tau^2 \) 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 \( \sigma_i^2 \). Both SURE estimators satisfy the asymptotic optimality (1.6). Since \( \sigma_i^2/\sigma_i^2 + \tau^2) \) is monotone increasing in \( \sigma_i^2 \), 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 \( \sigma \) by

\[ f_{G,\sigma}(x) = \int \frac{1}{\sigma} \varphi \left( \frac{x - u}{\sigma} \right) dG(u), \tag{1.7} \]

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]

\[ t_{G}^*(X_i, \sigma_i) = \mathbb{E}_G(\theta_i | X_i, \sigma_i) = X_i + \sigma_i^2 f_{G,\sigma}(X_i)/f_{G,\sigma}(X_i), \tag{1.8} \]

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

\[ R_n^*(G) = n^{-1} \sum_{i=1}^{n} R_{\theta_i}^*(G), \tag{1.9} \]

where \( R_{\theta_i}^*(G) = \sigma^2 \{ 1 - \sigma^2 \{ f_{G,\sigma}/f_{G,\sigma} \}^2 f_{G,\sigma} \} \) 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 \( \theta \) 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]

\[
\hat{G}_n = \arg \max_{G \in \mathcal{G}} \prod_{i=1}^{n} f_{G,\sigma_i}(X_i),
\]

where \( f_{G,\sigma}(x) \) is the mixture density as in (1.7) and \( \mathcal{G} \) 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. \( \theta_i \). 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 \( \theta_i \) 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 \( \theta_i \) are drawn from is denoted by \( G_n^* \). We assume that the variances are uniformly bounded, i.e., there exist constants \( \sigma_l \) and \( \sigma_u \) such that \( \sigma_l \leq \inf_n \min_i \sigma_i \leq \sup_n \max_i \sigma_i \leq \sigma_u \). In our analyses, we allow approximate solutions to (1.10). For definiteness and notation simplicity, the generalized MLE is any solution of

\[
\prod_{i=1}^{n} f_{\hat{G}_n,\sigma_i}(X_i) \geq q_n \sup_{G \in \mathcal{G}} \prod_{i=1}^{n} f_{G,\sigma_i}(X_i),
\]

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

\[
\hat{\theta}_i = t^*_{\hat{G}_n}(X_i, \sigma_i) = X_i + \sigma^2_i \frac{f'_{\hat{G}_n,\sigma_i}(X_i)}{f_{\hat{G}_n,\sigma_i}(X_i)}, \quad i = 1, \ldots, n,
\]
where \( \hat{G}_n \) 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 \( \mu_p(G) = \{ \int |u|^p dG(u) \}^{1/p} \) be the \( p \)-th absolute moment of a distribution function \( G \). The convergence rate \( \varepsilon_n \), 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 \( P_{G^*}, \theta_1, \ldots, \theta_n \) are i.i.d. random variables from a distribution \( G_n^* \), and given \( \theta_i \)'s, \( X_i \sim N(\theta_i, \sigma_i^2) \) are independent observations with known variances. Let \( \hat{\theta}_i = t^*_{G_n}(X_i, \sigma_i) \) be the GMLEB estimator in (2.2) with an approximate generalized MLE \( \hat{G}_n \) satisfying (2.1). Then, there exists a universal constant \( M_0 \) such that for all \( \log n > 1/p \),

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

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 \( f_{\hat{G}_n, \sigma} \) in definition (2.2) could be arbitrarily small. In order to rule out that possibility, we define a regularized rule \( t^*_{\hat{G}_n}(X_i, \sigma_i; \rho_n) \) which replaces this denominator with \( f_{\hat{G}_n, \sigma} \sqrt{\rho_n} / \sigma_i \), and in Theorem 5 we show that this rule relates to the GMLEB as

\[
\sum_{i=1}^{n} \left( t^*_{\hat{G}_n}(X_i, \sigma_i; \rho_n) - \theta_i \right)^2 = \sum_{i=1}^{n} \left( t^*_{\hat{G}_n}(X_i, \sigma_i; \rho_n) - \theta_i \right)^2, \quad \rho_n = \frac{q_n}{\sqrt{2\pi cn}}. \tag{2.5}
\]

Let \( A_n = \{ d(\hat{G}_n, G_n^*) \leq (x_\ast \vee 1)\varepsilon_n \} \) where \( x_\ast \) is the constant as in Theorem 4, and \( d(\cdot, \cdot) \) 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 \( \mathbb{E}_{G_n^*} \xi^2_{A_n} \) where

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

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

\[
\left\{ (t^*_{\hat{G}}(\cdot, \sigma_1; \rho_n), \ldots, t^*_{\hat{G}}(\cdot, \sigma_n; \rho_n)) : d(G, G_n^*) \leq \eta \varepsilon_n \right\} \tag{2.7}
\]
under \(\|\cdot\|_{\infty,M}\), where \(\eta^*\) will be manifested in Theorem 7. By the entropy bound in Theorem 7, there exists a collection of distributions \(\{H_j, j \leq N\}\) of manageable size \(N\) such that

\[
\zeta_{2n} = \left\{ \sum_{i=1}^{n} \left( t_{G_n}^* (X_i, \sigma; \rho_n) - \theta_i \right)^2 I_{A_n} \right\}^{1/2}
- \max_{j \leq N} \left\{ \sum_{i=1}^{n} \left( t_{H_j}^* (X_i, \sigma; \rho_n) - \theta_i \right)^2 \right\}^{1/2}
\]

(2.8)
is small. Since the collection \(\{H_j, j \leq N\}\) is of manageable size, a Gaussian isoperimetric inequality yields that

\[
\zeta_{3n} = \max_{j \leq N} \left\{ \left\{ \sum_{i=1}^{n} \left( t_{H_j}^* (X_i, \sigma; \rho_n) - \theta_i \right)^2 \right\}^{1/2}
- \mathbb{E}_{G_n} \left\{ \sum_{i=1}^{n} \left( t_{H_j}^* (X_i, \sigma; \rho_n) - \theta_i \right)^2 \right\}^{1/2} \right\}
\]

(2.9)
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

\[
\zeta_{4n} = \max_{j \leq N} \left\{ \mathbb{E}_{G_n} \sum_{i=1}^{n} \left( t_{H_j}^* (X_i, \sigma; \rho_n) - \theta_i \right)^2 \right\}^{1/2} \leq \left\{ n R_n^* (G_n^*) \right\}^{1/2} + \sum_{j=1}^{4} \left( \mathbb{E}_{G_n} \zeta_{2j}^2 \right)^{1/2}.
\]

(2.10)

is small. These upper bounds for individual pieces \(\mathbb{E}_{G_n} \zeta_{2j}^2\) are put together via

\[
\left\{ \mathbb{E}_{G_n} \sum_{i=1}^{n} \left( t_{G_n}^* (X_i, \sigma_i) - \theta_i \right)^2 \right\}^{1/2} \leq \left\{ n R_n^* (G_n^*) \right\}^{1/2} + \sum_{j=1}^{4} \left( \mathbb{E}_{G_n} \zeta_{2j}^2 \right)^{1/2}.
\]

(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 \(\theta_i\)'s, \(X_i \sim \mathcal{N}(\theta_i, \sigma_i^2)\) are independent observations with known variances. Let \(\hat{\theta}_i = t_{G_n}^* (X_i, \sigma_i)\) be the GMLEB estimator in (2.2) with an approximate generalized MLE \(\hat{G}_n\) satisfying (2.1). Then,

\[
\limsup_{n \to \infty} \frac{\mathbb{E}_{G_n} \sum_{i=1}^{n} \left( t_{G_n}^* (X_i, \sigma_i) - \theta_i \right)^2 / n}{R_n^* (G_n^*)} = 1,
\]

(2.12)

provided that \(\mu_\infty (G_n^*) = O(\sqrt{\log n})\) and \(n R_n^* (G_n^*) / (\log n)^5 \to \infty\).

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

\[
\mathcal{R}_n (\mathcal{G}) = \inf_{\tilde{t} G \in \mathcal{G}} \sup_{G \in \mathcal{G}} \frac{1}{n} \sum_{i=1}^{n} \left( t(X_i, \sigma_i) - \tilde{\theta}_i \right)^2,
\]

(2.13)
where the infimum is taken over all bivariate Borel functions. An estimator is adaptive minimax if
\[
\sup_{G \in \mathcal{G}_n} \mathbb{E}_G \frac{\sum_{i=1}^n (\hat{\theta}_i - \theta_i)^2/n}{\mathcal{R}_n(\mathcal{G}_n)} \to 1
\] (2.14)
holds uniformly for a range of sequences \(\{\mathcal{G}_n, n \geq 1\}\) of distribution classes. For positive \(p\) and \(C\), the \(L_p\) balls of distribution functions are defined as
\[
\mathcal{G}_{p,C} = \left\{ G : \int |u|^p dG(u) \leq C^p \right\}.
\] (2.15)

**Theorem 3.** Suppose that under \(P_{G_n^*}\), \(\theta_1, \ldots, \theta_n\) are i.i.d. random variables from a distribution \(G_n^*\), and given \(\theta_i\)'s, \(X_i \sim N(\theta_i, \sigma_i^2)\) are independent observations. Let \(t_{G_n}^* (X_i, \sigma_i)\) be the GMLEB estimator in (2.2) with an approximate generalized MLE \(\hat{G}_n\) satisfying (2.1). Then, the adaptive minimaxity (2.14) holds in \(L_p\) balls \(\mathcal{G}_{p,C_n}\) in (2.15), provided that \(C_n \to 0\) and \(\sqrt{n}C_n^{p/2}/(\log n)^{5/2} \to \infty\).

**Proof of Theorem 3.** By definition of minimax risk in (2.13), we have
\[
\mathcal{R}_n(\mathcal{G}_{p,C_n}) \geq \sup_{G \in \mathcal{G}_{p,C_n}} \inf_t \mathbb{E}_G \frac{1}{n} \sum_{i=1}^n (t(X_i, \sigma_i) - \theta_i)^2
\]
\[
= \sup_{G \in \mathcal{G}_{p,C_n}} \mathbb{E}_G \frac{1}{n} \sum_{i=1}^n (t_{G_n}^* (X_i, \sigma_i) - \theta_i)^2
\]
\[
= \sup_{G \in \mathcal{G}_{p,C_n}} R_n^*(G).
\] (2.16)

By Theorem 1, (2.16) and \(\sqrt{n}C_n^{p/2}/(\log n)^{5/2} \to \infty\), there exists a universal constant \(M_1\) such that
\[
\sup_{G \in \mathcal{G}_{p,C_n}} \mathbb{E}_G \frac{1}{n} \sum_{i=1}^n (t_{G_n}^* (X_i, \sigma_i) - \theta_i)^2 \leq \sup_{G \in \mathcal{G}_{p,C_n}} R_n^*(G) + M_1 (\log n)^{5/2}/\sqrt{n}
\]
\[
\leq \mathcal{R}_n(\mathcal{G}_{p,C_n}) + o(1)C_n^{p/2}.
\] (2.17)

Donoho and Johnstone [3] proved that as \(C_n \to 0\),
\[
\mathcal{R}_n(\mathcal{G}_{p,C_n}) = O(1)C_n^{p/2} \left\{ 2 \log(1/C_n) \right\}^{(1-p/2)}.
\] (2.18)

Thus, (2.17) and (2.18) lead to that
\[
\sup_{G \in \mathcal{G}_{p,C_n}} \mathbb{E}_G \frac{1}{n} \sum_{i=1}^n (t_{G_n}^* (X_i, \sigma_i) - \theta_i)^2 \leq \left( 1 + o(1) \right) \mathcal{R}_n(\mathcal{G}_{p,C_n}).
\]

This is the adaptive minimaxity in \(\mathcal{G}_{p,C_n}\).
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 \( \{ d(f_{G_n}, f_{G^*}) \leq x\varepsilon_n \} \), the other is inside this neighborhood. An essential ingredient is a large deviation inequality for \( d(f_{G_n}, f_{G^*}) \). In the heteroscedastic case, it seems that certain omnibus distance between \( f_{G_n,\sigma_i} \) and \( f_{G^*,\sigma_i} \) should be used. We use the average Hellinger distance \( \overline{d}(G_n, G^*_n) \) as defined in (3.2) below. We provide a large deviation inequality for \( \overline{d}(G_n, G^*_n) \). 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

\[
\mathcal{F}_n = \left\{ (f_{G,\sigma_1}(x), \ldots, f_{G,\sigma_n}(x)), G \in \mathcal{G} \right\},
\]

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

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

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,

\[
\|h\|_{\infty, M} = \max_{i \leq n} \|h_i\|_{\infty, M} = \max_{i \leq n} \sup_{|x| \leq M} |h_i(x)|,
\]

where \( h = (h_1(x), \ldots, h_n(x)) \) is an \( n \)-dimensional vector of functions.

**Theorem 4.** Suppose that under \( P_{G^*_n} \), \( \theta_1, \ldots, \theta_n \) are i.i.d. random variables from a distribution \( G^*_n \), and given \( \theta_i \)'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 \( G_n \) be certain approximate generalized MLE satisfying (2.1). Then, there exists a universal constant \( x_0 \) such that for all \( t \geq x_0 \) and \( \log n > 1/p \),

\[
\mathbb{P}_{G^*_n} \left\{ \overline{d}(G_n, G^*_n) \geq \varepsilon_n \right\} \leq \exp \left( -\frac{t^2 n \varepsilon_n^2}{2 \log n} \right) \leq e^{-t^2 \log n},
\]

where \( \varepsilon_n = \varepsilon(n, G^*_n, p) \) 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 \( M = 2\sigma_n \varepsilon_n^2/(\log n)^{3/2} \). Define

\[
h^*(x) = \eta I\{|x| \leq M\} + \frac{\eta M^2}{x^2} I\{|x| > M\}.
\]

We consider any approximate generalized MLE satisfying

\[
\prod_{i=1}^{n} \frac{f_{\hat{G}_n,\sigma_i}(X_i)}{f_{G^*_n,\sigma_i}(X_i)} \geq e^{-4t^2 n \varepsilon_n^2/15}.
\]
Let \( \{ (f_{H_j, \sigma_i}(x), \ldots, f_{H_j, \sigma_n}(x)) \}, j \leq N \) be an \( \eta \)-net of \( F_n \) under the semi-norm \( \| \cdot \|_{\infty,M} \), with \( N = N(\eta, F_n, \| \cdot \|_{\infty,M}) \). Let \( H_{0,j} \) be distributions satisfying

\[
\overline{d}(H_{0,j}, G_n^*) \geq t \varepsilon_n, \quad \max_{i \leq n} \| f_{H_{0,j}, \sigma_i} - f_{H_j, \sigma_i} \|_{\infty,M} \leq \eta, \tag{3.7}
\]

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

\[
f_{G, \sigma_i}(x) \leq \left\{ f_{H_{0,j}, \sigma_i}(x) + 2\eta = f_{H_{0,j}, \sigma_i}(x) + 2h^*(x), \quad |x| < M, \right.
\]

\[
\left. f_{H_{0,j}, \sigma_i}(x) + 2h^*(x), \quad |x| \geq M. \right\}
\]

It follows that when \( \overline{d}(\hat{G}_n, G_n^*) \leq t \varepsilon_n \),

\[
\prod_{i=1}^n \frac{f_{\hat{G}_n, \sigma_i}(X_i)}{f_{G_n^*, \sigma_i}(X_i)} = \prod_{|X_i|<M} \frac{f_{\hat{G}_n, \sigma_i}(X_i)}{f_{G_n^*, \sigma_i}(X_i)} \prod_{|X_i|\geq M} \frac{f_{\hat{G}_n, \sigma_i}(X_i)}{f_{G_n^*, \sigma_i}(X_i)} \leq \sup_{j \in J} \prod_{|X_i|<M} \frac{f_{H_{0,j}, \sigma_i}(X_i) + 2h^*(X_i)}{f_{G_n^*, \sigma_i}(X_i)} \prod_{|X_i|\geq M} \frac{1/(\sqrt{2\pi\sigma_i})}{2h^*(X_i)} \leq \sup_{j \in J} \prod_{i=1}^n \frac{f_{H_{0,j}, \sigma_i}(X_i) + 2h^*(X_i)}{f_{G_n^*, \sigma_i}(X_i)} \prod_{|X_i|\geq M} \frac{1/(\sqrt{2\pi\sigma_i})}{2h^*(X_i)}. \]

Thus, by (3.6),

\[
\mathbb{P}_{G_n^*} \left\{ \overline{d}(\hat{G}_n, G_n^*) \geq t \varepsilon_n \right\} \leq \mathbb{P}_{G_n^*} \left\{ \sup_{j \in J} \prod_{i=1}^n \frac{f_{H_{0,j}, \sigma_i}(X_i) + 2h^*(X_i)}{f_{G_n^*, \sigma_i}(X_i)} \prod_{|X_i|\geq M} \frac{1/(\sqrt{2\pi\sigma_i})}{2h^*(X_i)} \geq e^{-4t^2n\varepsilon_n^2/15} \right\} \leq \mathbb{P}_{G_n^*} \left\{ \sup_{j \in J} \prod_{i=1}^n \frac{f_{H_{0,j}, \sigma_i}(X_i) + 2h^*(X_i)}{f_{G_n^*, \sigma_i}(X_i)} \geq e^{-8t^2n\varepsilon_n^2/5} \right\} + \mathbb{P}_{G_n^*} \left\{ \prod_{|X_i|\geq M} \frac{1/(\sqrt{2\pi\sigma_i})}{2h^*(X_i)} \geq e^{4t^2n\varepsilon_n^2/3} \right\}. \tag{3.8}
\]

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

\[
\mathbb{P}_{G_n^*} \left\{ \prod_{i=1}^n \frac{f_{H_{0,j}, \sigma_i}(X_i) + 2h^*(X_i)}{f_{G_n^*, \sigma_i}(X_i)} \geq e^{-8t^2n\varepsilon_n^2/5} \right\} \leq \exp \left( \frac{4t^2n\varepsilon_n^2}{5} \right) \prod_{i=1}^n \int \sqrt{f_{H_{0,j}, \sigma_i} + 2h^*} \sqrt{f_{G_n^*, \sigma_i}} \leq \exp \left\{ \frac{4t^2n\varepsilon_n^2}{5} + \sum_{i=1}^n \left( \int \sqrt{f_{H_{0,j}, \sigma_i} + 2h^*} \sqrt{f_{G_n^*, \sigma_i}} - 1 \right) \right\}. \tag{3.9}
\]
By Jensen’s inequality, 
\[
\sum_{i=1}^{n} \left( \int f_{H_{0,j}, \sigma_{i}} + 2h^{*}(f_{G_{n}^{*}, \sigma_{i}}) + (2 \int h^{*})^{1/2} \right) 
\leq -t^{2} n \sigma_{n}^{2} + n \sqrt{\eta M}.
\] (3.10)

Since \(|J| \leq N\), (3.9) and (3.10) yield
\[
\mathbb{P}_{G_{n}^{*}} \left\{ \sup_{j} \prod_{i=1}^{n} \frac{f_{H_{0,j}, \sigma_{i}}(X_{i}) + 2h^{*}(X_{i})}{f_{G_{n}^{*}, \sigma_{i}}(X_{i})} \geq e^{-8t^{2} n \sigma_{n}^{2}/5} \right\} 
\leq \exp \left\{ \log N - \frac{t^{2} n \sigma_{n}^{2}}{5} + n \sqrt{\eta M} \right\}. \tag{3.11}
\]

Since \(\eta = 1/n^{2}\) and \(M = 2\sigma_{n} n \sigma_{n}^{2}/(\log n)^{3/2} \geq 4\sigma_{a} \sqrt{\log n}\), by Lemma 4,
\[
\log N + n \sqrt{\eta M} \leq C(2 \log n)^{2} \max \left\{ \frac{M}{\sqrt{2 \log n}}, 1 \right\} + \sqrt{8M} 
\leq \left\{ \left(\frac{t^{*}}{20} \right)^{2} M(\log n)^{3/2} \leq \frac{t^{2} n \sigma_{n}^{2}}{100} \right\}
\]
for \(t^{*} = t\). Thus, by (3.11),
\[
\mathbb{P}_{G_{n}^{*}} \left\{ \sup_{j} \prod_{i=1}^{n} \frac{f_{H_{0,j}, \sigma_{i}}(X_{i}) + 2h^{*}(X_{i})}{f_{G_{n}^{*}, \sigma_{i}}(X_{i})} \geq e^{-8t^{2} n \sigma_{n}^{2}/5} \right\} \leq \frac{e^{-t^{2} n \sigma_{n}^{2}/10}}{100}. \tag{3.12}
\]

By (3.5), \(1/h^{*}(x) = x^{2}/(\eta M^{2}) = (nx/M)^{2}\) for \(|x| \geq M\). So that
\[
\mathbb{P}_{G_{n}^{*}} \left\{ \prod_{|X_{i}| \geq M} \frac{1/(\sqrt{2 \pi} \sigma_{i})}{2h^{*}(X_{i})} \geq e^{4t^{2} n \sigma_{n}^{2}/10} \right\} 
\leq \exp \left( - \frac{2t^{2} n \sigma_{n}^{2}}{3 \log n} \right) \mathbb{E}_{G_{n}^{*}} \left\{ \prod_{|X_{i}| \geq M} \frac{nX_{i}}{\sqrt{\sigma_{i} M}} \right\}^{1/\log n}. \tag{3.13}
\]

Since \(M = 2\sigma_{n} n \sigma_{n}^{2}/(\log n)^{3/2} \geq 4\sigma_{a} \sqrt{\log n}\), Lemma 5 is applicable with \(a_{i} = n/(\sqrt{\sigma_{i} M})\) and \(\lambda = 1/\log n < 1\). Because \(a_{i} \leq n/(\sqrt{\sigma_{i} M})\),
\[
\mathbb{E}_{G_{n}^{*}} \left\{ \prod_{|X_{i}| \geq M} \frac{nX_{i}}{\sqrt{\sigma_{i} M}} \right\}^{1/\log n} \leq \exp \left\{ \left( e^{1/2 \log n} \right) \left( \frac{1}{\sqrt{2 \pi} \log n} + n \left( \frac{2\mu_{p}(G_{n}^{*})}{M} \right)^{p} \right) \right\}. \tag{3.14}
\]
By the definition of $\varepsilon_n$,
\[
\frac{n\varepsilon_n^2}{\log n} \geq 1.
\]
Therefore, (3.13) and (3.14) give
\[
\mathbb{P}_{G_n^*} \left\{ \prod_{|X_i| \geq M} \frac{1}{2h^*(X_i)} \geq e^{4t^2n\varepsilon_n^2/3} \right\} \leq \exp \left\{-\left( \frac{2t^2}{3} - \frac{e}{\sigma_i^{1/(2\log n)}} \right) \frac{n\varepsilon_n^2}{\log n} + \frac{e}{\sqrt{2\pi\log n}} \right\}. \tag{3.15}
\]
Inserting (3.12) and (3.15) into (3.8), we find that for large $n$ and $t \geq x_*$,
\[
\mathbb{P}_{G_n^*} \left\{ d(\widehat{G}_n, G_n^*) \geq t\varepsilon_n \right\} \leq \exp \left(-\frac{t^2n\varepsilon_n^2}{2\log n}\right) \leq e^{-t^2\log n} = n^{-t^2}.
\]
This completes the proof of Theorem 4. \qed

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
\[
t^*_G(x,\sigma;\rho) = x + \frac{\sigma^2 f'_{G,\sigma}(x)}{(\rho/\sigma) \lor f_{G,\sigma}(x)}. \tag{4.1}
\]
Denote $t^*_G(x) = x + f'_G(x)/f_G(x)$ and $t^*_G(x) = x + f'_G(x)/(\rho \lor f_G(x))$ 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 $G$:
\[
\int h(u)dF(u) = \int h(u/\sigma)dG(u). \tag{4.2}
\]
With $y = x/\sigma$, by the condition on $F$, we have $t^*_G(x,\sigma)/\sigma = t^*_F(y)$ and
\[
\frac{t^*_G(x,\sigma;\rho)}{\sigma} = y + \frac{f'_F(y)}{\rho \lor f_F(y)} = t^*_F(y;\rho). \tag{4.3}
\]
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

\[ t^*_{\hat{G}_n}(X_i, \sigma_i) = t^*_{\hat{G}_n}(X_i, \sigma_i; \rho_n), \quad \rho_n = q_n/(\sqrt{2\pi}en), \]  

where \(0 < q_n \leq 1\). This is a consequence of the following theorem.

**Theorem 5.** Let \( \hat{G}_n \) be an approximate generalized MLE satisfying

\[ \prod_{i=1}^{n} f_{\hat{G}_n,\sigma_i}(X_i) \geq q_n \sup_{\theta} \prod_{i=1}^{n} f_{\theta,\sigma_i}(X_i) \]  

for certain \(0 < q_n \leq 1\). Then, for all \(j = 1, \ldots, n\),

\[ f_{\hat{G}_n,\sigma_j}(X_j) \geq \frac{q_n}{\sqrt{2\pi}en\sigma_j}. \]  

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

\[ \prod_{i=1}^{n} f_{\hat{G}_{n,j},\sigma_i}(X_i) \geq q_n \prod_{i=1}^{n} f_{\hat{G}_{n,j},\sigma_i}(X_i) \geq q_n(1 - \varepsilon)^{n-1}(\varepsilon/\sqrt{2\pi}\sigma_j)^{n-1} \prod_{i \neq j} f_{\hat{G}_n,\sigma_i}(X_i). \]

Thus, \( f_{\hat{G}_{n,j},\sigma_j}(X_j) \geq q_n(1 - \varepsilon)^{n-1}\varepsilon/\sqrt{2\pi}\sigma_j \) after the cancelation of \( f_{\hat{G}_n,\sigma_i}(X_i) \) for \( i \neq j \). 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

\[ \tilde{L}(y) = \sqrt{-\log(2\pi y^2)}, \quad y \geq 0. \]  

Since \( \max_{1 \leq i \leq n} |X_i - \theta_i| \leq \sigma_u \sqrt{2\log n} \) with large probability, we expect that \( \max_{1 \leq i \leq n} |t^*_{\hat{G}_n}(x, \sigma; \rho) - x| \leq c_0 \sqrt{\log n} \) 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 \( \tilde{L}(y) = \sqrt{-\log(2\pi y^2)} \). Then,

\[ \left( \frac{f^{''}_{G,\sigma}(x)}{f_{G,\sigma}(x)} \right)^2 \leq \frac{f^{''}_{G,\sigma}(x)}{f_{G,\sigma}(x)} + \frac{1}{\sigma^2} \leq \frac{1}{\sigma^2} \tilde{L}^2(\sigma f_{G,\sigma}(x)) = \frac{1}{\sigma^2} \log \left( \frac{1}{2\pi \sigma^2 f^2_{G,\sigma}(x)} \right). \]  

(4.8)
Proof of Lemma 1. Let $Y|\xi \sim N(\xi, \sigma^2)$ and $\xi \sim G$ under $P_G$. Then,

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

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

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

This completes the proof. \hfill \Box

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

$$\begin{cases} |t_G^*(x, \sigma; \rho) - x| \leq \sigma \tilde{L}(\rho), & 0 < \rho < (2\pi e)^{-1/2}, \\ 0 \leq (\partial/\partial x)t_G^*(x, \sigma; \rho) \leq \tilde{L}^2(\rho), & 0 < \rho < (2\pi e)^{-1/2}. \end{cases} \quad (4.9)$$

Proof of Lemma 2. By Lemma 1,

$$|t_G^*(x, \sigma; \rho) - x| = \sigma^2 \frac{f_{G, \sigma}(x)}{(\rho/\sigma) \vee f_{G, \sigma}(x)} \left| \frac{f''_{G, \sigma}(x)}{f_{G, \sigma}(x)} \right| \leq \sigma \left( \frac{f_{G, \sigma}(x)}{(\rho/\sigma) \vee f_{G, \sigma}(x)} \right) \tilde{L}(\sigma f_{G, \sigma}(x)). \quad (4.10)$$

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

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

$$\frac{\partial t_G^*(x, \sigma; \rho)}{\partial x} = \begin{cases} 1 + \sigma^2 \frac{f''_{G, \sigma}(x)}{f_{G, \sigma}(x)} \frac{f_{G, \sigma}(x)}{(\rho/\sigma) \vee f_{G, \sigma}(x)} - \sigma^2 \left( \frac{f''_{G, \sigma}(x)}{f_{G, \sigma}(x)} \right)^2, & f_{G, \sigma}(x) \geq \rho/\sigma, \\ 1 + \sigma^2 \left( \frac{f''_{G, \sigma}(x)}{f_{G, \sigma}(x)} \right)^2, & f_{G, \sigma}(x) < \rho/\sigma. \end{cases} \quad (4.11)$$

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

$$\frac{\partial t_G^*(x, \sigma; \rho)}{\partial x} \leq 1 + \sigma^2 \frac{f''_{G, \sigma}(x)}{f_{G, \sigma}(x)} \leq \tilde{L}^2(\sigma f_{G, \sigma}(x)) \leq \tilde{L}^2(\rho).$$

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

$$0 \leq 1 - \frac{f_{G, \sigma}(x)}{\rho/\sigma} \leq 1 + \sigma^2 \frac{f''_{G, \sigma}(x)}{f_{G, \sigma}(x)} \leq 1 + \frac{f_{G, \sigma}(x)}{\rho/\sigma} (\tilde{L}^2(\sigma f_{G, \sigma}(x)) - 1). \quad (4.12)$$
Theorem 6. For any $0 < \rho \leq (2\pi e^2)^{-1/2}$ and $x_0 > 0$, 

$$
\frac{1}{n} \mathbb{E}_{G^n} \sum_{i=1}^{n} (t^*_G(X_i, \sigma_i; \rho) - \theta_i)^2 - R^*_n(G^n)
\leq \frac{M_0}{n} \sum_{i=1}^{n} \sigma_i^2 \max \{ ||\log \rho||^3, ||\log d(f_{G,\sigma_i}, f_{G^n,\sigma_i})|| \} d^2(f_{G,\sigma_i}, f_{G^n,\sigma_i}) \\
+ \frac{2}{n} \sum_{i=1}^{n} \sigma_i^2 \mathbb{P}_{G^n} \{ ||\theta_i/\sigma_i|| > x_0 \} + 2x_0\rho \tilde{L}^2(\rho) \\
+ 2\rho(\tilde{L}^2(\rho) + 2)^{1/2}.
$$

(4.15)
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 $E_G^* \zeta_n^2$ in (2.9) with a Gaussian isoperimetric inequality. For any family $\mathcal{H}$ of functions and semi-distance $d$, the $\eta$-covering number is

$$N(\eta, \mathcal{H}, d) \equiv \inf \{ N : \mathcal{H} \subseteq \bigcup_{j=1}^{N} \text{Ball}(h_j, \eta, d) \}$$

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

$$\mathcal{F}_\rho = \{ (t^*_G(\cdot, \sigma_1; \rho), \ldots, t^*_G(\cdot, \sigma_n; \rho)) : G \in \mathcal{G} \}. \quad (4.16)$$

where $\mathcal{G}$ 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 $\tilde{L}(y) = \sqrt{-\log(2\pi y^2)}$ be the inverse of $y = \varphi(x)$. Then, for all $0 < \eta \leq \sigma_1 \rho \leq (2\pi e)^{-1/2}$,

$$\log N(\eta^*, \mathcal{F}_\rho, \| \cdot \|_{\infty, M}) \leq \{ 4(6\tilde{L}^2(\eta) + 1)(2M/\tilde{L}(\eta) + 3) + 2 \} |\log \eta|, \quad (4.17)$$

where

$$\eta^* = \frac{\eta}{\rho} \left\{ \tilde{L}(\eta) \left( \frac{\sigma_u^2}{\sigma_l} + \frac{\sigma_u^3}{\sigma_l^3} + \frac{\sigma_u^2}{2\sqrt{2\pi} \sigma_l} + \frac{\sigma_u^2}{\sqrt{2\pi} \sigma_l} \right) \right. \right.$$  

$$+ \frac{2\sigma_u^3}{\sqrt{12} \pi \sigma_l^3} + \frac{\sigma_u}{\sqrt{12} \pi} + \frac{2\sigma_u^3}{\sqrt{2\pi} e^{3/2} \sigma_l^3} + \frac{\sigma_u^3}{\sqrt{2\pi} e \sigma_l^3} \} \right. \right. \quad (4.18)$$

5. Numerical studies

In order to investigate the adaptivity of the GMLEB to different heteroscedastic mean vectors, we carry out a simulation study. In Table 1, $\theta_i$ are drawn from
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 σ_i ∼ 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 \( t^*_G(\cdot; \sigma_i) \) 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 \( \theta_i \) and \( \sigma_i^2 \). The means are generated by \( \theta_i \sim (1 - p)\delta_0 + pN(3, \tau^2) \) where \( \delta_u \) 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 \( q_n = (\sqrt{2\pi}/n^2) \wedge 1 \) in (4.5) so that (4.4) holds with \( \rho_n = 1/n^3 \). Let \( \xi_n \) and \( x_* \) be as in Theorem 4 and \( \hat{L}(\rho) = -\log(2\pi\rho^2) \) be as in (4.7). Let \( \eta^* \) be as in (4.18) and

\[
\eta = \frac{\rho_n}{n} = \frac{1}{n^3}, \quad M = \frac{2\sigma_u n \xi_n^2}{(\log n)^{3/2}}. \tag{6.1}
\]

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

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

Step 1. Let \( A_n = \{ \theta \in \hat{G}_n, G_n^* \leq x^* \} \) with \( x^* = \max(x_*, 1) \) and \( \zeta_{1n} \) be as in (2.6). Since \( x^* \geq 1 \) and \( n \xi_n^2 \geq 2(\log n)^2 \) by (2.3), it follows from Theorem 4 that \( \mathbb{P}_{G_n^*} \{ A_n^c \} \leq \exp \left\{ - (x^*)^2 n \xi_n^2 / (2 \log n) \right\} \leq 1/n \). Thus, since
Since

\[ (t_{G_n}^* (X_i, \sigma_i; \rho_n) - X_i) + (X_i - \theta_i) \]

By (6.2) and (6.3),

\[ \sum_{i=1}^{n} \sigma_i^2 \bar{L}^2(\rho_n) \mathbb{P}(A_n^c) + 2 \mathbb{E}_{G_n} \sum_{i=1}^{n} (X_i - \theta_i)^2 I_{A_n^c} \]

\[ \leq M_0 \log n + 2 \sum_{i=1}^{n} \int_{0}^{\infty} \min \{P\{||N(0, \sigma_i^2)| > x\}, 1/n\} dx^2. \]  

(6.2)

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

\[ \sum_{i=1}^{n} \int_{0}^{\infty} \min \{P\{||N(0, \sigma_i^2)| > x\}, 1/n\} dx^2 \]

\[ \leq \int_{0}^{\infty} \min \{2ne^{-x^2/(2\sigma_i^2)}, 1\} dx^2 = 2\sigma_n^2 + 2\sigma_n^2 \log(2n). \]  

(6.3)

By (6.2) and (6.3),

\[ \mathbb{E}_{G_n} \zeta_{1n}^2 \leq M_0 \log n \leq M_0 n^{q_n^2}. \]  

(6.4)

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

\[ \zeta_{2n}^2 \leq \min_{j \leq N} \sum_{i=1}^{n} (t_{G_n}^*(X_i, \sigma_i; \rho_n) - t_{H_j}^*(X_i, \sigma_i; \rho_n))^2 I_{A_n} \]

\[ \leq (2\eta^*)^2 \# \{i: |X_i| \leq M\} + \{2\sigma_u \bar{L}(\rho_n)\}^2 \# \{i: |X_i| > M\}. \]

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

\[ \int_{|u| \geq M/2} dG_n^*(u) \leq \left( \frac{\mu_p(G_n^*)}{M/2} \right)^p \leq \left( \frac{2n^{q_n^2}}{M(\log n)^{3/2}} \right)^p \frac{\varepsilon_n}{\log n} = \left( \frac{1}{\sigma_n} \right)^p \frac{\varepsilon_n^2}{\log n}. \]  

(6.5)

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

\[ \mathbb{E}_{G_n} \zeta_{2n}^2 \leq n(2\eta^*)^2 + 4\sigma_n^2 \bar{L}^2(n^{-3}) \mathbb{E}_{G_n} \# \{i: |X_i| > M\} \]

\[ \leq M_0 (\log n) \left( \frac{1}{n} + \int_{|u| \geq M/2} dG_n(u) + \sum_{i=1}^{n} P\{|N(0, \sigma_i^2)| > 2\sigma_u \sqrt{\log n}\} \right) \]

\[ \leq M_0 (\log n) \left( \frac{1}{n} + \frac{\varepsilon_n^2}{\log n} + \sum_{i=1}^{n} P\{|N(0,1)| > 2\sqrt{\log n}\} \right) \]

\[ \leq M_0 (\log n) \left( \frac{1}{n} + \frac{\varepsilon_n^2}{\log n} + \frac{2}{n} \right). \]
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 \( h(x) = \left( \sum_{i=1}^n (t_G^*(x_i, \sigma_i; \rho) - \theta_i) \right)^{1/2} \). It follows from Lemma 2 that for \( 0 < \rho \leq (2\pi e^3)^{-1/2} \),
\[
|h(x) - h(y)| \leq \left\| x - y \right\| \max_{1 \leq i \leq n} \left| (\partial/\partial x) t_G^*(x, \sigma_i; \rho) \right|.
\]
Thus, \( h(x)/\tilde{L}^2(\rho) \) has the unit Lipschitz norm. The Gaussian isoperimetric inequality (e.g., [16]) gives that for any deterministic distribution \( G \) and \( x > 0 \),
\[
\mathbb{P}_{G_n^*} \left\{ \left( \sum_{i=1}^n (t_G^*(X_i, \sigma_i; \rho) - \theta_i) \right)^{1/2} \geq \mathbb{E}_{G_n^*} \left( \sum_{i=1}^n (t_G^*(X_i, \sigma_i; \rho) - \theta_i) \right)^{1/2} + x \right\}
\leq \exp \left( - \frac{x^2}{2\tilde{L}^4(\rho)} \right). \tag{6.7}
\]
This and (2.9) imply that
\[
\mathbb{E} G_n^* \zeta_{3n}^2 = \int_0^\infty \mathbb{P}_{G_n^*} \{ \zeta_{3n} > x \} \, dx^2
\leq \int_0^\infty \min \left\{ 1, N \exp \left( - x^2/(2\tilde{L}^4(\rho_n)) \right) \right\} \, dx^2
\leq 2\tilde{L}^4(\rho_n)(1 + \log N). \tag{6.8}
\]
The entropy bound for regularized Bayes rules in Theorem 7 and (6.1) give that
\[
\log N \leq M_0 (\log n)^{3/2} M/2 \leq M_0 n \varepsilon_n^2. \tag{6.9}
\]
Hence by (6.7) and (6.8),
\[
\mathbb{E} G_n^* \zeta_{3n}^2 \leq M_0 n \varepsilon_n^2 (\log n)^2. \tag{6.10}
\]

**Step 4.** In this step, we bound \( \mathbb{E} G_n^* \zeta_{4n}^2 \). First of all, it follows from (2.10) that
\[
\zeta_{4n}^2 \leq \max_{j \leq N} \left\{ \mathbb{E} G_n^* \sum_{i=1}^n (t_{H_j}^*(X_i, \sigma_i; \rho_n) - \theta_i)^2 - n R_n^*(G_n^*) \right\}. \tag{6.11}
\]
By Theorem 6, for any \(0 < \rho_n \leq (2\pi e^2)^{-1/2}\) and \(x_0 > 0\),

\[
\mathbb{E}_{G_n^*} \sum_{i=1}^{n} (t_{H_j}(X_i, \sigma_i; \rho_n) - \theta_i)^2 - nR_n^*(G_n^*) \\
\leq M_0 \sum_{i=1}^{n} \sigma_i^2 \max \{|\log \rho_n|^3, |\log d(f_{H_j, \sigma_i}, f_{G_n^*, \sigma_i})|\} d^2(f_{H_j, \sigma_i}, f_{G_n^*, \sigma_i}) \\
+ 2 \sum_{i=1}^{n} \sigma_i^2 \left\{ \mathbb{P}_{G_n^*} \{\theta_i/\sigma_i > x_0\} + 2x_0\rho_n \tilde{L}^2(\rho_n) \\
+ 2\rho_n (\tilde{L}^2(\rho_n) + 2)^{1/2} \right\}. \tag{6.11}
\]

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

\[
\frac{\mathbb{P}_{G_n^*} \{\theta_i/\sigma_i > x_0\}}{|\log \rho_n|^3(x^*\varepsilon_n)^2} \leq \frac{\int_{|u| \geq M/2} dG_n^*(u)}{|\log \rho_n|^3(x^*\varepsilon_n)^2} \leq \frac{(1/\sigma_u)^p(\varepsilon_n^2/\log n)}{(\log n)^2\varepsilon_n^2} \leq M_0. \tag{6.12}
\]

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

\[
\frac{2(M/(2\sigma_l) + 1)\rho_n \tilde{L}^2(\rho_n)}{(\log \rho_n)^3(x^*\varepsilon_n)^2} \leq \frac{M_0(n\varepsilon_n^2/(\log n)^{3/2} + 1)/n^3}{(\log n)^2\varepsilon_n^2} \leq \frac{M_0}{n^2(\log n)^{7/2}} \leq M_0. \tag{6.13}
\]

Thus, by (6.10)–(6.13),

\[
\zeta_n^2 \leq M_0 n(\log \rho_n)^3/3\varepsilon_n^2 = M_0 n\varepsilon_n^2 (\log n)^3. \tag{6.14}
\]

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

\[
\sum_{j=1}^{4} \left( \mathbb{E}_{G_n^*} \zeta_{sj} \right)^{1/2} \leq M_0 n^{1/2}\varepsilon_n (\log n)^{3/2}.
\]

This and (2.11) complete the proof. \(\square\)

**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 \(m = (2[6a^2] + 1)[2M/(a\sigma_l) + 2] + 1\) atoms, such that

\[
\max_{i \leq n} \|f_{G, \sigma_i} - f_{G_m, \sigma_i}\|_{\infty, M} \leq \left(1 + \frac{1}{\sqrt{2\pi}}\right) \frac{\eta}{\sigma_l}. \tag{6.15}
\]

**Proof of Lemma 3.** Let \(j^* = [2M/(a\sigma_l) + 2]\) and \(k^* = [6a^2]\). Let

\[
I_j = (-M + (j - 2)a\sigma_l, -M + (j - 1)a\sigma_l) \land (M + a\sigma_l), \quad j = 1, \ldots, j^*.
\]
be a partition of \((-M - a\sigma_t, M + a\sigma_t]\). 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 \([-M - a\sigma_t, M + a\sigma_t]\) 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 \(t^2/2 \leq k^* + 2\),

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

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

\[
|f_{G, \sigma_t}(x) - f_{G_m, \sigma_t}(x)| \leq \left| \int_{(I_{j-1} \cup I_j \cup I_{j+1})} \frac{1}{\sigma_t} \varphi\left(\frac{x - u}{\sigma_t}\right) d(G(u) - G_m(u)) \right|
+ \left| \int_{(I_{j-1} \cup I_j \cup I_{j+1})} \frac{1}{\sigma_t} \text{Rem}\left(\frac{x - u}{\sigma_t}\right) d(G(u) - G_m(u)) \right|
\leq \frac{1}{\sigma_t} \varphi\left(\frac{a\sigma_t}{\sigma_u}\right) + \frac{(2a^2)^{k^*+1}}{\sigma_t \sqrt{2\pi} (k^*+1)!}
\leq \frac{1}{\sigma_t} \varphi\left(\frac{a\sigma_t}{\sigma_u}\right) + \frac{(e/3)^{k^*+1}}{\sigma_t \sqrt{2\pi} (k^*+1)^{1/2}}. \tag{6.16}
\]

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

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

\[
\log N(\eta, \mathcal{F}_n, \|\cdot\|_{\infty, M}) \leq C |\log \eta|^2 \max\left(\frac{M}{\sqrt{|\log \eta|}}, 1\right), \tag{6.17}
\]

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

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

\[
m = (2[6a^2] + 1)[2M/(a\sigma_t) + 2] + 1 \leq C |\log \eta| \max\left(\frac{M}{\sqrt{|\log \eta|}}, 1\right). \tag{6.18}
\]

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

\[
\max_{i \leq n} \|f_{G, \sigma_t} - f_{G_m, \sigma_t}\|_{\infty, M} \leq \left(1 + \frac{1}{\sqrt{2\pi}}\right) \frac{\eta}{\sigma_t}. \tag{6.19}
\]
The next step is to approximate the \( f_{G_{m,\sigma}} \) in (6.19) by \( f_{\tilde{G}_{m,\sigma}} \), where \( G_{m,\eta} \) is supported in a lattice and has no more than \( m \) atoms. Let \( \xi \sim G_m \) and \( \xi_\eta = \eta \operatorname{sgn}(\xi)\|\xi\|/\eta \). Define \( G_{m,\eta} \) as the distribution of \( \xi_\eta \). The support of \( G_{m,\eta} \) is in the grid \( \Omega_{\eta,M} = \{0, \pm \eta, \pm 2\eta, \ldots\} \cap [-M - a\sigma, M + a\sigma] \). Since \(|\xi - \xi_\eta| \leq \eta \) and \( \sup_x (\partial/\partial x)(1/\sigma_\eta)(x/\sigma_\eta) = 1/(\sqrt{2\pi\sigma_\eta^2}) \),

\[
\max_{i \leq n} \left\| f_{G_{m,\sigma}, i} - f_{G_{m,\eta}, i} \right\|_\infty \leq \frac{1}{\sqrt{2\pi\sigma_\eta^2}} \eta. \tag{6.20}
\]

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

\[
\inf_{w^{m,\eta} \in \mathcal{P}_m,\eta} \left\| w - w^{m,\eta} \right\|_1 \leq \eta, \quad \forall w \in \mathcal{P}_m,
\]

with \( N(\eta, \mathcal{P}_m, \| \cdot \|_1) \) elements. Let \( \{u_j, j = 1, \ldots, m\} \) be the support of \( G_{m,\eta} \) and \( w^{m,\eta} \) be a probability vector in \( \mathcal{P}_m,\eta \) satisfying \( \sum_{j=1}^m |G_{m,\eta}(\{u_j\}) - w^{m,\eta}_j| \leq \eta \). Denote \( G_{m,\eta} = \sum_{j=1}^m w^{m,\eta}_j \delta_{u_j} \). Then,

\[
\max_{i \leq n} \left\| f_{G_{m,\sigma}, i} - f_{G_{m,\eta}, i} \right\|_\infty \leq \frac{1}{\sqrt{2\pi\sigma_\eta}} \eta, \tag{6.21}
\]

since \( \sup_x (1/\sigma_\eta)(x/\sigma_\eta) = 1/(\sqrt{2\pi\sigma_\eta}) \).

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

\[
\max_{i \leq n} \left\| f_{G,\sigma, i} - f_{G_{m,\eta}, i} \right\|_{\infty,M} \leq \left(1 + \frac{1}{\sqrt{2\pi}} + \frac{1}{\sqrt{2\pi}}\right) \eta \| f \|_1 = \eta^{**}, \tag{6.22}
\]

The support of \( G_{m,\eta} \) is also in \( \Omega_{\eta,M} \). Counting the number of ways to realize \( \{u_j\} \) and \( w^{m,\eta} \), we find

\[
N(\eta^{**}, \mathcal{F}_n, \| \cdot \|_{\infty,M}) \leq \left[\frac{\Omega_{\eta,M}}{m}\right] N(\eta, \mathcal{P}_m, \| \cdot \|_1), \tag{6.23}
\]

with \( m \) satisfying (6.18) and \( |\Omega_{\eta,M}| = 1 + 2[(M + a\sigma)/\eta] \).

Since \( \mathcal{P}_m \) is in the \( \ell_1 \) unit-sphere of \( \mathbb{R}^m \), \( N(\eta, \mathcal{P}_m, \| \cdot \|_1) \) is no greater than the maximum number of disjoint \( \operatorname{Ball}(v_j, \eta/2, \| \cdot \|_1) \) with \( \| v_j \| = 1 \). Here is the argument. Suppose there exists \( w \in \mathcal{P}_m \) such that \( w \notin \bigcup_j \operatorname{Ball}(v_j, \eta, \| \cdot \|_1) \). Then \( \operatorname{Ball}(w, \eta/2, \| \cdot \|_1) \cap \operatorname{Ball}(v_j, \eta/2, \| \cdot \|_1) = \emptyset, \forall j \). This is a contradiction with the maximum number of disjoint balls. Hence \( \mathcal{P}_m \subseteq \bigcup_j \operatorname{Ball}(v_j, \eta, \| \cdot \|_1) \).

Since all these disjoint \( \operatorname{Ball}(v_j, \eta/2, \| \cdot \|_1) \) are inside the \((1 + \eta/2)\) \( \ell_1 \)-ball, volume comparison yields \( N(\eta, \mathcal{P}_m, \| \cdot \|_1) \leq (2/\eta+1)^m \). This, (6.23) and another application of the Stirling formula yield

\[
N(\eta^{**}, \mathcal{F}_n, \| \cdot \|_{\infty,M}) \leq \frac{(2/\eta+1)^m |\Omega_{\eta,M}|^m}{m!}.
\]
Lemma 5. Suppose that under $P_{G^*_n}$, $\theta_1, \ldots, \theta_n$ are i.i.d. random variables from a distribution $G^*_n$, and given $\theta_i$’s, $X_i \sim N(\theta_i, \sigma_i^2)$ are independent observations with known variances. Then for all constants $M \geq \sigma_u \sqrt{8 \log n}$, $0 < \lambda < \min(1, p)$, and $a_1, \ldots, a_n > 0$,

$$
\underset{G^*_n}{E}\left\{ \prod_{i=1}^{n} |a_iX_i|^\lambda I\{|X_i| \geq M\} \right\}^\lambda \leq \exp\left\{ \sum_{i=1}^{n} (a_iM)^\lambda \left( \frac{4\sigma_u}{Mn\sqrt{2\pi}} + \left( \frac{2\mu_p(G^*_n)}{M} \right)^p \right) \right\}.
$$

Proof of Lemma 5. It follows that

$$
\underset{G^*_n}{E}\left\{ \prod_{i=1}^{n} |a_iX_i|^\lambda I\{|X_i| \geq M\} \right\}^\lambda 
\leq \prod_{i=1}^{n} \left( 1 + a_i^\lambda E|X_i|^\lambda I\{|X_i| \geq M\} \right)
\leq \exp\left\{ \sum_{i=1}^{n} a_i^\lambda \int_{|x| \geq M} |x|^\lambda f_{G^*_n, \sigma_i}(x) dx \right\}.
$$

Let $Z \sim N(0, \sigma_2^2)$ and $\theta \sim G^*_n$. Since $Z + \theta \sim f_{G^*_n, \sigma}$, and $0 < \lambda < 1$,

$$
\int_{|x| \geq M} |x|^\lambda f_{G^*_n, \sigma_i}(x) dx 
= E|Z + \theta|^\lambda I\{|Z + \theta| \geq M\}
\leq E|2Z|^\lambda I\{|Z| \geq M/2\} + E|2\theta|^\lambda I\{|\theta| \geq M/2\}
\leq 2M^{\lambda-1}E|Z|I\{|Z| \geq M/2\} + \int_{|x| \geq M/2} (2|x|)^\lambda dG^*_n(x).
$$

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

$$
\int_{|x| \geq M/2} (2|x|)^\lambda dG^*_n(x) \leq 2\lambda p \mu_p(G^*_n) \left( \mathbb{P}_{G^*_n}\{|x| \geq M/2\} \right)^{1-\lambda/p} 
\leq M^{\lambda} \left( \frac{2\mu_p(G^*_n)}{M} \right)^p.
$$

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

$$
2M^{\lambda-1}E|Z|I\{|Z| \geq M/2\} \leq 2M^{\lambda-1}\sigma_u E|Z/\sigma_i|I\{|Z/\sigma_i| \geq M/2\sigma_u\}.
$$
Proof of Theorem 7. It follows from (4.1) and Lemma 2 that

\[ |t'_G(x, \sigma; \rho) - t'_H(x, \sigma; \rho)| \leq \frac{\sigma^2\tilde{L}(\rho)}{\rho} |f_{G,\sigma}(x) - f_{H,\sigma}(x)|. \]  

(6.27)

Let \( \eta = \varphi(a_{\sigma_1}/\sigma_u) \) so that \( a_{\sigma_1}/\sigma_u = \tilde{L}(\eta) \). Let \( j^* = \lceil 2M/(a_{\sigma_1}) + 2 \rceil \) and \( k^* = \lceil 6a^2 \rceil \). Let

\[ I_j = (-M + (j-2)a_{\sigma_1}, -M + (j-1)a_{\sigma_1}) \cap (M + a_{\sigma_1}], \quad j = 1, \ldots, j^*, \]

be a partition of \((-M - a_{\sigma_1}, M + a_{\sigma_1}]\). It follows from the Carathéodory’s theorem that for each distribution function \( G \) there exists a discrete distribution function \( G_m \) with support \([-M - a_{\sigma_1}, M + a_{\sigma_1}]\) and no more than \( m = (2k^* + 2)j^* + 1 \) support points such that

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

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

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

Thus, since \( k^* + 1 \geq 6a^2 \) and \( a_{\sigma_1}/\sigma_u \geq 1 \), for \( x \in I_j \cap [-M, M] \), the Stirling formula yields

\[ \left| f'_{G,\sigma}(x) - f'_{G_m,\sigma}(x) \right| \leq \left| \int_{I_{j-1}\cup I_j\cup I_{j+1}} \left( \frac{u-x}{\sigma_i^2} \right) \varphi \left( \frac{x-u}{\sigma_i} \right) d\left(G(u) - G_m(u)\right) \right| + \left| \int_{I_{j-1}\cup I_j\cup I_{j+1}} \left( \frac{u-x}{\sigma_i^2} \right) \text{Rem} \left( \frac{x-u}{\sigma_i} \right) d\left(G(u) - G_m(u)\right) \right| \leq \frac{1}{\sigma_i^2} \max_{t \geq a_{\sigma_1}/\sigma_u} t \varphi(t) + \frac{2a((2a)^2/2)^{k^*+1}}{\sigma_i^2 \sqrt{2\pi}(k^*+1)!} \]
Similarly, for $|x| \leq M$,

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

(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),

$$
\max_{i \leq n} \|t^*_G(\cdot, \sigma_i; \rho) - t^*_G(\cdot, \sigma_i; \rho)\|_{\infty,M}
\leq \frac{\sigma_u^3}{\sigma_l^2 \rho} \left( \frac{\eta \bar{L}(\eta)}{2\pi \sqrt{6a^2}} + \frac{\sigma_u^2 \bar{L}(\rho)}{\sigma_l \rho} \right)
\leq \frac{\eta}{\rho} \left( \frac{\sigma_u^2}{\sigma_l} + \frac{\sigma_u^3}{\sqrt{12\pi \sigma_l^2}} + \frac{\sigma_u}{\sqrt{12\pi}} \right).
$$

(6.30)

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

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

This and (6.27) imply

$$
\max_{i \leq n} \|t^*_G(\cdot, \sigma_i; \rho) - t^*_G(\cdot, \sigma_i; \rho)\|_{\infty} \leq \frac{\eta}{\rho} \left( \frac{2\sigma_u^3}{\sqrt{2\pi e\sigma_i^2}} + \frac{\sigma_u^2 \bar{L}(\eta)}{\sqrt{2\pi e\sigma_i^2}} \right).
$$

(6.31)

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

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

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

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

$$
\|f_{G_{m,\eta,\sigma_i}} - f_{G_{m,\eta,\sigma_i}}\|_{\infty} \leq \frac{1}{\sqrt{2\pi \sigma_i}} \eta, \quad \|f'_{G_{m,\eta}} - f'_{G_{m,\eta,\sigma_i}}\|_{\infty} \leq \frac{1}{\sqrt{2\pi \sigma_i}} \eta,
$$

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

$$
\max_{i \leq n} \|t^*_G(\cdot, \sigma_i; \rho) - t^*_G(\cdot, \sigma_i; \rho)\|_{\infty} \leq \frac{\eta}{\rho} \left( \frac{\sigma_u^3}{\sqrt{2\pi e\sigma_i^2}} + \frac{\sigma_u^2 \bar{L}(\eta)}{\sqrt{2\pi e\sigma_i^2}} \right).
$$

(6.32)
The support of $G_{m,\eta}$ and $\tilde{G}_{m,\eta}$ is $\Omega_{\eta,M} = \{0, \pm \eta, \pm 2\eta, \ldots\} \cap [-M - a\sigma_1, M + a\sigma_1]$.

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

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

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

$$N(\eta^*, \mathcal{F}_p, \| \cdot \|_{\infty, M}) \leq \left( \frac{|\Omega_{\eta,M}|}{m} \right) N(\eta, \mathcal{F}^{m,\eta}, \| \cdot \|_1),$$

(6.33)

with $m = (2k^* + 2)j^* + 1$, $|\Omega_{\eta,M}| = 1 + 2[(M + a\sigma_1)/\eta]$, $\eta = \varphi(a\sigma_1/\sigma_u)$, $j^* = [2M/(a\sigma_1) + 2]$ and $k^* = [6a^2]$. The rest of the proofs follow the same line in Lemma 4. The bound is $N(\eta^*, \mathcal{F}_p, \| \cdot \|_{\infty, M}) \leq \eta^{-2m}$ where $m \leq 2(6a^2 + 1)(2M/(a\sigma_1) + 3) + 1$.}

\textbf{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.

\textbf{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 $\ell_p$-balls for $\ell_q$-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