Empirical Bayes estimation by wavelet series

Marianna Pensky*, Mohammed Alotaibi

Received: September 24, 2004; Accepted: September 21, 2005

Summary: In traditional nonparametric EB (empirical Bayes) setting, the paper proposes generalization of the linear EB estimation method which takes advantage of the flexibility of the wavelet techniques. A nonparametric EB estimator is represented as a wavelet series expansion and the coefficients are estimated by minimizing the prior risk of the estimator. Although wavelet series have been used previously for EB estimation, the method suggested in the paper is completely novel since the EB estimator as a whole is represented as a wavelet series rather than its components. Moreover, the method exploits de-correlating property of wavelets which is not instrumental for the former wavelet-based EB techniques. As a result, estimation of wavelet coefficients requires solution of a well-posed sparse system of linear equations. The technique provides asymptotically optimal EB estimators posterior risks of which tend to zero at the optimal rate as the number of observations tends to infinity.

1 Introduction

The traditional empirical Bayes (EB) setting is as follows. One observes independent two-dimensional random vectors \((X_1, \theta_1), \ldots, (X_n, \theta_n)\), where each \(\theta_i\) is distributed according to some unknown prior pdf \(g\) and, given \(\theta_i = \theta\) the observation \(X_i\) has the known conditional density function \(q(x|\theta)\). In each pair the first component is observable, but the second is not. After the \((n+1)\)-th observation \(y \equiv X_{n+1}\) is taken, the goal is to estimate \(\theta \equiv \theta_{n+1}\).

If we knew the prior density \(g(\theta)\), then the Bayes estimator of \(\theta\) under the squared loss and the prior \(g(\theta)\) is given by

\[
t(y) = \int_{-\infty}^{\infty} q(y|\theta) g(\theta) d\theta / \int_{-\infty}^{\infty} q(y|\theta) g(\theta) d\theta.
\] (1.1)

As the prior density is unknown, we construct an EB estimator \(\hat{\theta}(y) = \hat{\theta}(y; X_1, X_2, \ldots, X_n)\) as the estimator of (1.1) from observations \(X_1, X_2, \ldots, X_n\). Denote

* The work was supported in part by National Science Foundation (NSF), grant DMS-0505133.
AMS 1980 subject classification: 62C12, 62G20
Key words and phrases: linear empirical Bayes estimation, wavelets, convergence rate
\[ p(y) = \int_{-\infty}^{\infty} q(y|\theta) g(\theta) d\theta, \]  
(1.2)

\[ \Psi(y) = \int_{-\infty}^{\infty} \theta q(y|\theta) g(\theta) d\theta. \]  
(1.3)

Hence \( t(y) \) can be rewritten as

\[ t(y) = \frac{\Psi(y)}{p(y)}. \]  
(1.4)

After Robbins [22] formulated the EB estimation problem many statisticians have been developing EB methods. The comprehensive list of references as well as numerous examples of applications of EB techniques can be found in [3] or [13].

The EB techniques can be divided into two groups: parametric and nonparametric. Parametric EB methods require that the family to which the mixing distribution belongs is specified a priori. The past data is then used to estimate the values of the unknown parameters, usually using the likelihood maximization technique. In nonparametric EB estimation, the entire mixing distribution is completely unspecified. One of the approaches to nonparametric EB estimation is based on estimation of the numerator and the denominator in the ratio in (1.1). This approach was introduced by Robbins himself and later developed by a number of authors (see e.g. [4], [5], [15]–[17], [23], [25]–[29] among others). The method provides estimators having good convergence rates, however, it requires relatively tedious three-step procedure: estimation of the top and the bottom of the fraction and then the fraction itself. Recent research in this direction seems to be devoted to constructing estimators under somewhat different conditions or for a wider range of distribution families (see e.g. [2], [6], [10]–[12], [31]) rather than developing efficient computational schemes. A new tool – wavelets – provides an opportunity to construct adaptive wavelet–based EB estimators with better computational properties (see e.g. [9], [18]–[20]) but the necessity of estimation of the ratio in (1.1) remains.

Another method of nonparametric EB, linear EB estimation, was introduced by Robbins [24]. Robbins suggested to approximate Bayes estimator \( t(x) \) by a linear function of \( x \) and to determine the coefficients of \( t(x) \) by minimizing the expected squared difference between \( t(x) \) and \( \theta \), with subsequent estimation of the coefficients on the basis of observations \( X_1, \ldots, X_n \). The technique is extremely efficient computationally and was immediately put to practical use, for instance, for prediction of the finite population mean. We shall use this problem as an example to describe development of the linear EB theory. Since the Bayes estimator is linear under normal model on \( \theta \), Ghosh and Meeden [8] suggested EB estimators using normal priors on \( \theta_i \)'s. Later, Ghosh and Lahiri [7] relaxed normality assumptions on the priors assuming that the conditional expectation of \( \theta_i \) given \( X_i \) is linear in \( \theta_i \). In [14] results of [8] are generalized to the case of unequal unknown sample variances but again for a normal model. Arora et al. [1] considers even more complicated normal model with auxiliary variables. Finally, Karunamuni and Zhang [11] construct EB estimators in a true spirit of Robbins’ linear EB technique. However, their estimators are optimal only in the class of estimators linear in \( y \). To summarize, the merit of the linear EB estimator is that it has a very simple form, but the disadvantage is that the linear function may not be able to provide a good approximation to the Bayes
estimator. To overcome this defect, Pensky and Ni [21] generalized approach of Rob-bins [24] extending his technique to approximation of \( t(x) \) by an algebraic polynomial. However, although the polynomial-based EB estimation provides significant improve-ment in the convergence rates in comparison with the linear EB estimation, the system of linear equations resulting from the method is badly conditioned which leads to compu-tational difficulties and loss of precision. Another shortcoming of the polynomial-based estimator of Pensky and Ni [21] is that no assessment of its optimality has been carried out.

Present paper proposes generalization of the linear EB estimation method which takes advantage of the flexibility of the wavelet techniques. We present an EB esti-mator as a wavelet series expansion and estimate coefficients by minimizing the prior risk of the estimator. Although wavelet series have been used previously for EB estima-tion, the method suggested in the paper is completely novel since the EB estimator as a whole is represented as a wavelet series rather than its components. More-over, the method exploits de-correlating property of wavelets which is not instrumen-tal for the former wavelet-based EB techniques. As a result, estimation of wavelet coefficients requires solution of a well-posed sparse system of linear equations. The dimension of the system depends on the size of wavelet support and smoothness of the Bayes estimator. The number of nonzero diagonals of the matrix of the system is equal to the length of the support of the father wavelet, and, furthermore, this matrix tends to identity matrix as the scale tends to zero. The technique proposed in the pa-per is motivated by computational efficiency, nevertheless, it provides asymptotically optimal EB estimators posterior risks of which tend to zero at an optimal rate as \( n \to \infty \).

The rest of the paper is organized as follows. Section 2 introduces EB estimation algorithm. Section 3 assesses estimation error and presents the results on the asymptotic optimality of the EB method suggested in the paper. Section 4 contains examples of construction of EB estimators when \( \theta \) is a location parameter. Section 5 is reserved for discussion. Proofs of all the statements in the paper are placed in Section 6.

2 EB estimation algorithm

In order to construct an estimator of \( t(y) \) defined in (1.4), choose wavelets with bounded support and \( s \) vanishing moments, so that

\[
\text{supp } \varphi \in [M_1, M_2],
\]

\[
\int_{-\infty}^{\infty} x^j \sum_{k \in \mathbb{Z}} \varphi(x - k)\varphi(z - k)dx = z^j, \quad 0 \leq j \leq s - 1.
\]

Approximate \( t(y) \) by a wavelet series

\[
t_m(y) = \sum_{k \in \mathbb{Z}} a_{m,k} \varphi_{m,k}(y)
\]
where \( \psi_{m,k}(y) = 2^m/2\psi(2^my - k) \), and estimate coefficients of \( t(y) \) by minimizing the global mean squared difference

\[
\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[ \sum_{k \in \mathbb{Z}} a_{m,k} \psi_{m,k}(y) - z \right]^2 q(y|z) g(z) dz dy \Rightarrow \min.
\] (2.4)

Taking derivatives of the last expression with respect to \( a_{m,j} \) and equating them to zero, we obtain the system of linear equations

\[
Ba = c
\] (2.5)

with

\[
b_{j,k} = \int_{-\infty}^{\infty} \psi_{m,k}(x) \psi_{m,j}(x) p(x) dx = E[\psi_{m,k}(X) \psi_{m,j}(X)],
\] (2.6)

\[
c_j = \int_{-\infty}^{\infty} \psi_{m,j}(x) \Psi(x) dx.
\] (2.7)

Here and in what follows we use the symbol \( E \) for expectation over the distribution of \( X_1, X_2, \ldots, X_n \). The expectations over any other distributions are represented in integral forms.

System (2.5) is an infinite system of equations. However, since we are interested in estimating \( t(x) \) locally at \( x = y \), we shall keep only indices \( k, j \in K_{m,y} \) where

\[
K_{m,y} = \{ k \in \mathbb{Z} : 2^m y - M_2 - r(M_2 - M_1) \leq k \leq 2^m y - M_1 + r(M_2 - M_1) \}
\] (2.8)

where \( r \) will be determined later. Observe that really expansion (2.3) contains just coefficients \( a_{m,k} \) with \( 2^m y - M_2 \leq k \leq 2^m y - M_1 \), however, for evaluation of these coefficients we need to keep more terms in the system of equations (2.5).

The entries (2.6) of the matrix \( B \) are unknown and can be estimated by sample means

\[
b_{j,k} = n^{-1} \sum_{l=1}^{n} [\psi_{m,k}(X_l) \psi_{m,j}(X_l)]
\] (2.9)

In order to estimate \( c_j \), find functions \( u_{m,j}(x) \) such that for any \( \theta \)

\[
\int_{-\infty}^{\infty} q(x|\theta) u_{m,j}(x) dx = \int_{-\infty}^{\infty} \theta q(x|\theta) \psi_{m,j}(x) dx.
\] (2.10)

Then, multiplying both sides of (2.10) by \( g(\theta) \) and integrating over \( \theta \), we obtain

\[
E u_{m,j}(X) = \int_{-\infty}^{\infty} u_{m,j}(x) p(x) dx = \int_{-\infty}^{\infty} \psi_{m,j}(x) \Psi(x) dx = c_j.
\]

Note that functions \( u_{m,j}(x) \) are the same functions which appear in the wavelet estimator of the numerator \( \Psi(y) \) of the EB estimator (1.4), therefore, the estimator considered herein can be constructed whenever wavelet EB estimation is possible (see...
Empirical Bayes estimation by wavelet series

Solutions of equation (2.10) can be easily obtained, for example, when \( q(x|\theta) \) is a location parameter family, scale parameter family, one-parameter exponential family or a family of uniform distributions (see [18]–[20]). In Section 4 we consider in detail the case when \( \theta \) is a location parameter.

Once functions \( u_{m,j}(x) \) are derived, coefficients \( c_j \) can be estimated by

\[
\hat{c}_j = n^{-1} \sum_{l=1}^{n} u_{m,j}(X_l) \tag{2.11}
\]

and system (2.5) replaced by \( \hat{B}\hat{a} = \hat{c} \). However, though estimators \( \hat{B} \) and \( \hat{c} \) converge in mean squared sense to \( B \) and \( c \), respectively, the estimator \( \hat{a} = \hat{B}^{-1}\hat{c} \) may not even have finite expectation. To understand this fact, note that both \( \hat{B} \) and \( \hat{c} \) are asymptotically normal. In one dimensional case, the ratio of two normal random variables has Cauchy distribution and hence does not have finite mean. In multivariate case the difficulty remains. To ensure that the estimator of \( a \) has finite expectation, we choose \( \delta > 0 \) and construct an estimator of \( a \) of the form

\[
\hat{a}_\delta = (\hat{B} + \delta I)^{-1}\hat{c} \tag{2.12}
\]

where \( I \) is the identity matrix. Observe that matrix \( \hat{B} \) is nonnegative definite, so that \( \hat{B} + \delta I \) is a positive definite matrix and, hence, is nonsingular. Solution \( \hat{a}_\delta \) is used for construction of the EB estimator

\[
\hat{t}(y) = \sum_{k \in K_m} (\hat{a}_\delta)_{m,k} \psi_{m,k}(y). \tag{2.13}
\]

### 3 Estimation error

An EB estimator \( \hat{t}(y) \) may be characterized by the posterior risk

\[
R(y; \hat{t}) = (p(y))^{-1} \int_{-\infty}^{\infty} (\hat{t}(y) - \theta)^2 q(y|\theta) g(\theta) d\theta
\]

which can be partitioned into two components. The first component of this sum is

\[
R(y; t(y)) = \inf_f R(y; f(y)) = (p(y))^{-1} \int_{-\infty}^{\infty} (t(y) - \theta)^2 q(y|\theta) g(\theta) d\theta.
\]

which is independent of \( \hat{t}(y) \) and represents the posterior risk of the Bayes estimator (1.1). Thus we shall judge EB estimators by the second component

\[
\hat{R}_n(y) = E(\hat{t}(y) - t(y))^2. \tag{3.1}
\]

It must be noted that often the quality of the EB estimator is described by

\[
E\hat{R}_n(y) = \int_{-\infty}^{\infty} \hat{R}_n(y) p(y) dy,
\]
which is the difference between the prior risk $E \int_{-\infty}^{\infty} R(y; \hat{t}(y)) p(y) \, dy$ of the EB estimator $\hat{t}(y)$ and the prior risk $\int_{-\infty}^{\infty} R(y; t(y)) p(y) \, dy = \inf_{f} \int_{-\infty}^{\infty} R(y; f(y)) p(y) \, dy$ of the Bayes estimator $t(y)$. However, the risk function (3.1) has several advantages compared with $ER_n(y)$. First, $R_n(y)$ enables one to calculate the mean squared error for the given observation $y$ which is the quantity of interest. Note that the wavelet series (2.13) is local in a sense that coefficients $(\hat{a}_s)_{m,k}$ change whenever $y$ changes, hence, working with a local measure of the risk makes much more sense. Using the prior risk for the estimator which local in nature prevents one from seeing advantages of this estimator. Second, by using the risk function (3.1) we eliminate the influence on the risk function of the observations having very low probabilities. So, the use of $R_n(y)$ provides a way of getting EB estimators with better convergence rates. Third, posterior risk allows one to assess optimality of EB estimators for majority of familiar distribution families via comparison of the convergence rate of the estimator with the lower bounds for the risk derived in [17]. Finally, one can pursue evaluation of the prior risk for the estimator (2.13). The derivation will require assumptions similar to the ones in [18] and can be accomplished by standard methods.

The error (3.1) consists of two components

$$R_n(y) \leq 2(\Delta_1^2 + \Delta_2)$$

where the first component $\Delta_1$ is due to replacement of the Bayes estimator $t(y)$ by its wavelet representation (2.3), while $\Delta_2$ is due to replacement of vector $a = B^{-1}c$ by $\hat{a}_s$ given by (2.12):

$$\Delta_1 = t_m(y) - t(y),$$

$$\Delta_2 = E \left[ \sum_{k \in K_{m,y}} ((\hat{a}_s)_{m,k} - a_{m,k}) \varphi_m(k)(y) \right]^2.$$  

(3.4)

We shall refer to $\Delta_1$ and $\Delta_2$ as the systematic and the random error components, respectively.

The **systematic error component.** For evaluation of the systematic error component $\Delta_1$, let us introduce matrices $U_h$ and $U^{*}_h$ and vectors $D_h$ and $D^{*}_h$ with components

$$\begin{align*}
(U_h)_{k,l} &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} z^h \varphi(z + 2^m y - k) \varphi(z + 2^m y - l) \, dz, \\
(U^{*}_h)_{k,l} &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} z^h \varphi(z + 2^m y - k) \varphi(z + 2^m y - l) p^{(b)}(y + v_2 2^{-m} z) \, dz, \\
(D_h)_{k} &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} z^h \varphi(z + 2^m y - k) \, dz, \\
(D^{*}_h)_{k} &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} z^h \varphi(z + 2^m y - k) \Psi^{(b)}(y + v_2 2^{-m} z) \, dz,
\end{align*}$$

(3.5-3.8)

where $v_1, v_2 \in (0, 1)$. Observe that $U_h$ and $D_h$ are independent of unknown functions $p(x)$ and $\Psi(x)$, and that $U_0 = I$ where $I$ is the identity matrix. Denote

$$\Omega_{m,y} = \{ x : |x - y| \leq 2^{-m}(r + 1)(M_2 - M_1) \}.$$  

(3.9)

Then the following statement is true.
Lemma 3.1 Let functions $p(x)$ and $\Psi(x)$ be $r \leq s$ times continuously differentiable in the neighborhood $\Omega_y$ of $y$ and let $\Omega_{m,y} \subseteq \Omega_y$. Then

$$B = p(y)I + \sum_{h=1}^{r-1}2^{-mh}(h!)^{-1}p^{(h)}(y)U_h + 2^{-mr}(r!)^{-1}U_r^*,$$  \hspace{1cm} (3.10)

$$c = 2^{-m/2}\sum_{l=0}^{r-1}2^{-ml}(l!)^{-1}\psi^{(l)}(y)D_l + 2^{-mr}(r!)^{-1}D_r^*,$$  \hspace{1cm} (3.11)

where $I$ is the identity matrix.

The proofs of this and later statements are placed in the Appendix.

Lemma 3.1 establishes that for large $m$ matrix $B$ is close to identity matrix, so the system (2.5) is well-conditioned. Furthermore, if $m \to \infty$, vector $a$ in (2.3) tends to $2^{-m/2}E_p(y)\psi(y)I_d$ where $2^{-m/2}\sum_k(D_0)\psi_{m,k}(y) = 1$ for any $y$. The latter implies that the systematic error goes to zero as $m \to \infty$. The following lemma shows that it goes to zero at an optimal rate of $O(2^{-mr})$.

Lemma 3.2 Let functions $p(x)$ and $\Psi(x)$ be $r \leq s - 1$ times continuously differentiable in the neighborhood $\Omega_y$ of $y$ and let $\Omega_{m,y} \subseteq \Omega_y$. Then $\Delta_1$ has the following asymptotic expression

$$\Delta_1 = \frac{2^{-mr}}{r!} \int_{-\infty}^\infty z^r Q(2^m y, z + 2^m y) \frac{p(y)}{p(y)} \left[ \psi^{(r)}(y + \frac{v_1 z}{2m}) - \psi^{(r)}(y + \frac{v_2 z}{2m}) \right] \, dz + o(2^{-mr}),$$  \hspace{1cm} (3.12)

where $Q(x, z) = \sum_{k \in \mathbb{Z}} \varphi(x - k)\varphi(z - k)$ and $v_1, v_2 \in (0, 1)$. Hence, $\Delta_1 = O(2^{-mr})$.

The random error component. In order to calculate $\Delta_2$, introduce vectors $\gamma^{(j)}(m)$, $j = 1, 2$, with components

$$\gamma^{(j)}_k(m) = \left[ \int_{-\infty}^\infty u_{m,k}^{(j)}(x) \, dx \right]^{1/2}, \quad k \in \mathbb{K}_{m,y},$$  \hspace{1cm} (3.13)

where $u_{m,k}^{(j)}(x)$ are defined in (2.10). Observe that the values of $\gamma^{(j)}_k(m)$ are independent of the unknown density $g(\theta)$ and can be calculated explicitly.

The functions $\gamma^{(j)}_k(m)$ are similar to functions $\sigma_j(m) = \sup_k \int_{-\infty}^\infty u_{m,k}^{(j)}(x) \, dx$ in Assumption A2 of [18]. The only difference is that supremum over all possible $k$ is not used any more. Hence, the assumptions of the current paper are less restrictive than those of [18]. It is impossible to impose these assumptions in terms of $q(x|\theta)$ and $g(\theta)$ rather than via $\|\gamma^{(j)}(m)\|$ in a general setup. However, for any particular case (e.g., location parameter, scale parameter, one-parameter exponential family), the assumptions will be formulated in terms of $q(x|\theta)$ and $g(\theta)$.
Lemma 3.3 Let \( \delta \sim n^{-1}2^m \). Then, under the assumptions of Lemma 3.2,
\[
\Delta_2 = O(2^m n^{-1} \|\gamma'(m)\|^2 + 2^m n^{-1})
\]
provided \( m \) is such that \( 2^m = o(n) \) and \( \|\gamma''(m)\|^2 2^m = o(n^3) \) as \( n \to \infty \).

Note that the variance \( \Delta_2 \) cannot be asymptotically smaller than \( O(2^m n^{-1}) \) which is the error of estimating pdf \( p(y) \) and is asymptotically larger than \( O(2^m n^{-1}) \) if \( \|\gamma^{(j)}(m)\| \to \infty \) as \( m \to \infty \). Combining Lemmas 3.2 and 3.3, we obtain

Theorem 3.4 Let functions \( p(x) \) and \( \Psi(x) \) be \( r \geq 1/2 \) times continuously differentiable in the neighborhood \( \Omega_x \) of \( y \) and let \( \Omega_{m,y} \subseteq \Omega_y \) where \( \Omega_{m,y} \) is defined in (3.9). Choose \( \delta \sim n^{-1}2^m \) and denote
\[
m_n = \arg \min (n^{-1}2^m \|\gamma'(m)\|^2 + 1 + 2^{-2mr}).
\]
If wavelets possesses \( s \) vanishing moments, \( s \geq r + 1 \), then
\[
R_n(y) = O(n^{-2mr})
\]
provided \( \|\gamma''(m)\|^2 2^m = o(n^3) \) as \( n \to \infty \). In particular, if \( \|\gamma^{(j)}(m)\|, j = 1, 2, \) are bounded as \( m \to \infty \), then \( R_n(y) = O(n^{-2r/(2r+1)}) \).

4 Example: location parameter family

Let us consider the situation when \( \theta \) is a location parameter, i.e. \( q(x|\theta) = q(x - \theta) \). Denote the Fourier transform of \( f \) at the point \( w \) by \( f(\omega) \) or \( \mathcal{F}[f](\omega) \). The EB estimator of \( \theta \) can be represented as
\[
t(y) = y - \Psi(y)/p(y)
\]
where \( \Psi(x) = \int_{-\infty}^{\infty} (x - \theta)q(x - \theta)g(\theta)d\theta \). Consequently, \( u_{m,k}(x) \) in (2.11) are of the form (see [20])
\[
u_{m,k}(x) = 2^{m/2} U_m(2^m x - k).
\]
Here \( U_m(x) \) is the inverse Fourier transform of the function
\[
\hat{U}_m(\omega) = i\hat{q}(-2^m \omega)(\hat{q}(-2^m \omega))^{-1} \hat{\varphi}(\omega),
\]
where \( i \) is the imaginary unit and \( \hat{q}(-2^m \omega) \) is the value of \( \hat{q}(\cdot) \) at the point \( -2^m \omega \). In order for the inverse Fourier transform of the function \( \hat{U}_m(\omega) \) to exist, \( \varphi(x) \) should be chosen so that the right hand side of (4.3) is square integrable. As examples of the above, we shall study the cases when \( q(x) \) is a gamma or a normal density function.

If \( q(x) \) is a pdf of the gamma distribution
\[
q(x) = [\sigma^\beta \Gamma(\beta)]^{-1} x^{\beta-1} \exp(-x/\sigma),
\]

\[\text{This article is protected by German copyright law. You may copy and distribute this article for your personal use only. Other use is only allowed with written permission by the copyright holder.}\]
then $\tilde{q}(\omega) = (1 - i\sigma)^{-\beta}$. Direct calculations yield

$$
\tilde{U}_m(\omega) = 2^{-m} \beta \tilde{\psi}(\omega) \mathcal{F}(\exp(2^{-m}x) I(x < 0))(\omega).
$$

Parseval identity and simple transformations of the integral lead to

$$
U_m(x) = \beta \int_0^{\infty} e^{-z} \varphi(2^m z + x) dz.
$$

In the case of the normal distribution, $q(x) = [2\pi\sigma^2]^{-1/2} \exp(-x^2/(2\sigma^2))$ and $\tilde{q}(\omega) = \sigma \exp(-\omega^2\sigma^2/2)$. Hence, $\tilde{U}_m(\omega) = i2^m\sigma^2 \omega \varphi(\omega)$ and

$$
U_m(x) = 2^m\sigma^2 \varphi'(x).
$$

Note that in the case of the gamma density, the inverse Fourier transform of $\tilde{U}_m(\omega)$ always exists while for the normal case we need to choose differentiable $\varphi(x)$, so that $\omega \varphi(\omega)$ is square integrable.

To assert convergence rates of the EB estimators in the case of the gamma density, verify that

$$
\gamma_k^{(1)}(m) = \int_{-\infty}^{\infty} U_m^2(x) dx
= \beta^2 \int_{-\infty}^{\infty} \int_0^{\infty} \int_0^{\infty} e^{-z} e^{-t} \varphi(2^m z + x) \varphi(2^m t + x) dz dt dx
\leq \beta^2 \int_{-\infty}^{\infty} \varphi^2(x) dx.
$$

$$
\gamma_k^{(2)}(m) = 2^m \int_{-\infty}^{\infty} U_m^4(x) dx
= \beta^4 \int_{-\infty}^{\infty} \left( \int_0^{\infty} e^{-z} \varphi(2^m z + x) dz \right)^4 dx
\leq \beta^4 \int_{-\infty}^{\infty} \varphi^4(x) dx.
$$

Consequently, by Theorem 3.4, $m_n \sim n^{1/(2r+1)}$ and $R_n(y) = O(n^{-2r/(2r+1)})$. Observe that the fraction $|\tilde{q}(-2m\omega)/\tilde{q}(-m\omega)|$ is bounded as $m$ grows, hence both $\|\gamma_k^{(1)}(m)\|$ and $\|\gamma_k^{(2)}(m)\|$ are $O(1)$ as $m \to \infty$.

In the case of the normal distribution,

$$
\gamma_k^{(1)}(m) = 2^m(m-1) \int_{-\infty}^{\infty} U_m^2(x) dx
= \sigma^4 2^m(m-1) \int_{-\infty}^{\infty} [\varphi'(x)]^2 dx, \quad j = 1, 2.
$$

which implies $m_n \sim n^{1/(2r+3)}$ and $R_n(y) = O(n^{-2r/(2r+3)})$. Note that since both $\Psi(y)$ and $p(y)$ are infinitely differentiable, the value of $r$ is arbitrary large. However, the rate of $R_n(y)$ differs from the optimal convergence rate of $O(n^{-1}/(\ln n)^{3/2})$ although it can be made arbitrary close to this rate by choosing large $r$. The optimal convergence rate can be obtained only with infinitely smooth wavelets (see [19]), and slight loss in convergence
rate observed for the procedure suggested herein is the price for computational advantages of the finitely supported wavelets.

5 Discussion

In the present paper we introduce a new wavelet-based method of EB estimation. The method is based on approximating the Bayes estimator corresponding to observation y by wavelet series using finitely supported wavelet family.

Approximating \( t(y) \) as a whole rather than its parts allows much better control on the quality and resolution of approximation. Flexibility of wavelets can be exploited in various ways here. For example, if one wants exploratory analysis of EB estimator, some simple wavelet functions can be used (e.g., Haar or Daubechies with few vanishing moments). Since the smoothing parameter \( r \) should not exceed \( s \), this choice will lead to a small system of linear equations. In the case of Haar wavelets, the system of equations is diagonal, while in the case of small order Daubechies wavelets it is just very sparse. After this initial analysis is done, one can re-evaluate EB estimator using a new wavelet family which possesses larger number of vanishing moments. In comparison, for traditional nonparametric EB estimator, we do not have this immediate control and flexibility since regularized fraction of two orthonormal series is not an orthonormal series itself. In addition, the proposed algorithm can be written in matrix form which makes calculations in a programming language like MatLab really fast and convenient.

As we mentioned before, the methods described above exploits de-correlating property of wavelets. Since \( b_{j,k} \neq 0 \) whenever \( |k-j| > M_2 - M_1 \), the number of nonzero diagonals in the matrix \( B \) of the system of equations is fixed and does not depend on the smoothness of the estimator itself, no matter how large the system is.

Finally, we want to discuss the choice of smoothing parameter \( r \). In general, smoothing parameter can be chosen by cross validation. However, in the case of EB estimation the conditional density \( q(x|\theta) \) is available, hence, in many situations we have a lower bound for \( r \) readily available and this bound can be quite large. For example, if \( q(x|\theta) \) is a family of the gamma distributions (4.4) with the location parameter \( \theta \), then \( r \geq r_\beta \equiv \text{Int}(\beta) + 1 + I(\text{Int}(\beta) = \beta) \) where \( \text{Int}(\beta) \) is the integer part of \( \beta \) and \( I(\cdot) \) is the indicator function (see [20]). The extreme case here is the situation when \( q(x|\theta) \equiv q(x-\theta) \) is the normal family of densities, for which in theory \( r \) can be chosen arbitrary large. However, in practice \( r \) should take moderate value since the improvement of the precision may be minimal for finite sample size while the amount of computations grows significantly.

The reason for the latter is that the larger \( r \) is, the more vanishing moments the wavelet should have, the larger the support of the wavelet is. In general, \( M_2 - M_1 \) grows as \( O(r) \), for instance, \( M_2 - M_1 \approx 2r - 1 \) for Daubechies wavelets. Since the dimension of the system of equations (2.5) is \( (2r+1)(M_2-M_1) \), the number of coefficients in the system of linear equations grows as \( O(r^2) \) while the number of nonzero diagonals of the matrix of the system is equal to \( 2(M_2-M_1) \) and, thus, grows as \( O(r) \). This simple consideration implies that the method would work the best when it is applied to less smooth densities \( q(x) \), e.g. gamma densities (4.4) with relatively small values of \( \beta \). As a result, whenever \( \beta \) in (4.4) is relatively large, one may not want to choose \( r > r_\beta \) and set \( r = r_\beta \).
Appendix

Proof of Lemma 3.1: Without loss of generality, let us prove formula (3.10). Recall representation of \( \phi_{m,k}(x) \) and change variables \( z = 2^m(x - y) \). Then use Taylor series for \( p(x) \) centered at \( y \) and write the result in terms of matrices \( U_h \) and \( U_r^* \):

\[
b_{j,k} = 2^m \int_{-\infty}^{\infty} \phi(2^m x - k) \phi(2^m x - j) p(x) dx
\]

\[
= \int_{-\infty}^{\infty} \phi(z + 2^m y - k) \phi(z + 2^m y - j) p(y + 2^{-m} z) dz
\]

\[
= \sum_{h=0}^{r-1} 2^{-mh} (h!)^{-1} p^{(h)}(y) U_h + 2^{-mr} (r!)^{-1} U_r^*.
\]

Now to complete the proof, just note that \( U_0 = I \), the identity matrix. \( \square \)

Validity of Lemma 3.2 is based on the following statements.

Lemma A.1 Matrix \( B^{-1} \) can be represented as

\[
B^{-1} = \sum_{h=0}^{r-1} 2^{-mh} V_h + 2^{-mr} V_r^* \quad (A.1)
\]

with \( V_j = \sum \alpha_{k_1,k_2,\ldots,k_l} U_{k_1} U_{k_2} \cdots U_{k_l} \). Here, \( l \leq j \) and coefficients \( \alpha_{k_1,k_2,\ldots,k_l} \) are polynomial functions of the derivatives \( p^{(k)}(y) \) with \( \sum_{h=0}^{\ell} k_h = j \) divided by powers of \( p(y) \).

Proof: Write \( B^{-1} \) in the form (A.1) and find \( V_h, 0 \leq h \leq r - 1 \), and \( V_r^* \) such that \( BB^{-1} = I + O(2^{-mr}) \). Multiplying (3.10) and (3.11), introducing a new parameter \( i = l + h \) and eliminating \( O(2^{-mr}) \) terms, we obtain equation

\[
\sum_{i=0}^{2r-1} 2^{-mi} \sum_{h=\max(0,i+1-r)}^{i} \frac{p^{(i-h)}(y)}{(i-h)!} U_{i-h} V_h = I.
\]

Equating matrix coefficients for various powers of \( 2^{-m} \), we derive a system of linear equations

\[
\sum_{h=0}^{r-1} U_{i-h} p^{(i-h)}(y)(i-h)!^{-1} V_h = I(i = 0) I, \quad i = 0, \ldots, r - 1, \quad (A.2)
\]

where \( I(\cdot) \) is the indicator function. Formula (A.2) suggests a recursive procedure to calculate \( V_h, 0 \leq h \leq r - 1 \).

It is straightforward to see that \( V_0 = [p(y)]^{-1} I \) which verifies (A.1) for \( r = 0 \). Let us use mathematical induction to prove Lemma A.1. Assuming that Lemma A.1 is valid...
Lemma A.2 Let \( Q(x, z) \) be defined in Lemma 3.2. Then
\[
\int_{-\infty}^{\infty} z^j Q(2^m y + x, 2^m y + z)dz = x^j, \quad 0 \leq j \leq s - 1.
\]

Validity of Lemma A.2 follows directly from Theorem 3.2 in [30].

Lemma A.3 Let \( 1 \leq v \leq s - 1 \) and \( l_1 + \cdots + l_t = h \), where \( t \leq s-1 \) and \( 0 \leq l_i \leq s-1 \). Then
\[
\sum_{k \in K_{m,v}} (U_{l_1} U_{l_2} \cdots U_{l_t} D_{v-h})k \varphi(2^m y - k) = 0. \tag{A.3}
\]

Proof: Recalling (3.5) and (3.7), we obtain
\[
S = \sum_{k \in K_{m,v}} (U_{l_1} U_{l_2} \cdots U_{l_t} D_{v-h})k \varphi(2^m y - k)
\]
\[
= \int \cdots \int_{k_1, \ldots, k_t, k \in K_{m,v}} z_1^{l_1} \varphi(2^m y + z_1 - k) \varphi(2^m y + z_1 - k_1)
\]
\[
\times z_2^{l_2} \varphi(2^m y + z_2 - k_1) \varphi(2^m y + z_1 - k_2)
\]
\[
\cdots z_t^{l_t} \varphi(2^m y + z_t - k_{t-1}) \varphi(2^m y + z_1 - k_t)
\]
\[
\times z^{v-h} \varphi(2^m y + z - k_t) \varphi(2^m y - k)dz_1 \cdots dz_t dz.
\]

Now, note that the set \( K_{m,v} \) is so large that all the sums over \( k_1 \notin K_{m,v}, \ldots, k_t \notin K_{m,v}, k \notin K_{m,v} \) vanish, so we can replace \( K_{m,v} \) by the set of all integers \( Z \) in the
summations. Using the definition of \( Q(x, z) \) and Lemma A.2 and replacing the multiple integral by the iterated integrals, we derive

\[
S = \int_{\delta_1}^{\delta_2} Q(2^m y + z, 2^m y + z_1)dz_1 \int_{\delta_1}^{\delta_2} Q(2^m y + z_2, 2^m y + z_1)dz_2 \cdots \\
\int_{\delta_1}^{\delta_2} Q(2^m y + z_1, 2^m y + z)dz
\]

\[
= \int_{\delta_1}^{\delta_2} Q(2^m y, 2^m y + z_1)dz_1 = 0, \quad v \leq r - 1.
\]

Proof of Lemma 3.2: Note that by Lemma 3.1,

\[
a = B^{-1}c = 2^{-\frac{m}{2}} \left( \sum_{k=0}^{r-1} \frac{2^{-mh}}{h!} p^{(h)}(y) U_h + \frac{2^{-m_2}}{m!} U_2 \right) \\
\sum_{l=0}^{r-1} \frac{2^{-ml}}{l!} \Psi^{(l)}(y) D_l + \frac{2^{-m_3}}{m!} D_3
\]

\[
= 2^{-m/2} \sum_{v=0}^{r-1} \sum_{h=0}^{m_v} (v - h)! [v - 1] \Psi^{(v-h)}(y) V_h D_{v-h} + O(2^{-mr}),
\]

where the last relation is obtained by introducing a new parameter \( v = l + h \), re-arranging the sums and combining the terms \( O(2^{-mr}) \). Recall that by (2.3) and Lemmas A.1–A.3, since \( V_0 = [p(y)]^{-1}I \) and all the sums are finite,

\[
t_{k}(y) = \sum_{h=0}^{v} \sum_{k=0}^{r-1} \frac{2^{-mh}}{h!} [v - 1] \Psi^{(v-h)}(y) \cdot (V_h D_{v-h})_k \varphi(2^m y - k) + O(2^{-mr})
\]

\[
= \sum_{v=0}^{r-1} \sum_{h=0}^{m_v} \frac{\Psi^{(v-h)}(y)}{(v - h)!} \sum_{\sum_{k=0}^{r-1}} \alpha_{k_1, k_2, \ldots, k_r} \\
\sum_{k=0}^{r-1} (U_{k_1} \cdots U_{k_h} D_{v-h})_k \varphi(2^m y - k) + O(2^{-mr})
\]

\[
= \Psi(y) \sum_{k \in K_0, k_y} (V_0 D_0)_k \varphi(2^m y - k) + O(2^{-mr})
\]

\[
= t(y) + O(2^{-mr}).
\]

Hence, \( \Delta_1 = O(2^{-mr}) \).
Now, let us evaluate the \(2^{-mr}\) term

\[
\Delta_{m,r} = 2^{-mr} \sum_{k \in \mathbb{M}, y} \left[ \sum_{h=1}^{r-1} \frac{\Psi^{(r-h)}(y)}{(r-h)!} (V_h D_{r-h})_k + \frac{(V_0 D_r)_k}{r!} \right] + \Psi(y) (V_r^* D_0)_k \right] (2^{-m} y - k).
\]

(A.4)

Since \(s - 1 \geq r\), the first portion of \(\Delta_{m,r}\) vanishes. The second term can be written explicitly using formula (3.8). To calculate the last term, we need to find asymptotic expression for \(V_r^*\). Using (A.2), we derive

\[
V_r^* = [r! p^2(y)]^{-1} U_r^* - [p(y)]^{-1} \sum_{h=1}^{r-1} V_h U_{r-h}.
\]

Plugging \(V_r^*\) into (A.4), applying Lemma A.3 and recalling (3.6) and (3.8), we obtain (3.12).

Proof of Lemma 3.3 is based on the following lemma.

**Lemma A.4** Let \(B\) and \(\hat{B}\) be symmetric positive definite \(M \times M\) matrices with \(\|B^{-1}\| < \infty\) and \(c\) and \(\hat{c}\) be vectors in \(\mathbb{R}^M\). Let

\[
a = B^{-1} c, \quad \hat{a}_s = (\hat{B} + \delta I)^{-1} \hat{c}
\]

(A.5)

where \(I\) is the \(M \times M\) identity matrix. Then for any \(\delta > 0\) we have \(E \|\hat{a}_s - a\|^2 \leq 3(R_1 + R_2 + R_3)\) where

\[
R_1 \leq \|B^{-1}\|^4 \|c\|^2 [2^3 E \|\hat{B} - B\|^2 + 2^7 \delta^{-2} E \|\hat{B} - B\|^4 + 2\delta^2],
\]

\[
R_2 \leq \|B^{-1}\|^2 \|E\| \|\hat{c} - c\|^2,
\]

\[
R_3 \leq \|B^{-1}\|^4 [2^7 E \|\hat{B} - B\|^4 + 2^{15} \delta^{-4} E \|\hat{B} - B\|^8 + 2^{13} \delta^{-4} \|E\| \|\hat{c} - c\|^{1/2}].
\]

Proof of Lemma A.4: Denote \(B_\delta = B + \delta I\) and \(\hat{B}_\delta = \hat{B} + \delta I\). By properties of the norm,

\[
\|\hat{a}_s - a\| \leq \|\hat{B}_\delta^{-1} - B^{-1}\| \|c\| + \|B^{-1}\| \|\hat{c} - c\| + \|\hat{B}_\delta^{-1} - B^{-1}\| \|\hat{c} - c\|.
\]

(A.6)

Partition \(\hat{B}_\delta^{-1} - B^{-1} = (\hat{B}_\delta^{-1} - B^{-1}) + (B^{-1} - B^{-1})\) and note that \(B_\delta^{-1} - B^{-1} = -\delta B_\delta^{-1} B^{-1}\) and \(\hat{B}_\delta^{-1} - B^{-1} = \hat{B}_\delta^{-1} (B - \hat{B})^{-1} B^{-1}\). Consequently,

\[
\|\hat{B}_\delta^{-1} - B^{-1}\| \leq \|\hat{B}_\delta^{-1}\| \|B - \hat{B}\| \|B^{-1}\|, \quad \|\hat{B}_\delta^{-1} - B^{-1}\| \leq \|\hat{B}_\delta^{-1}\| \|\hat{B} - B\| \|B^{-1}\|.
\]

(A.7)

(A.8)

Formula (A.8) also implies that \(\|\hat{B}_\delta^{-1}\| \leq \|B^{-1}\|/(1 - \|\hat{B} - B\| \|B^{-1}\|)\) provided \(\|\hat{B} - B\| \leq \|B^{-1}\|^{-1}\). Together with (A.8), the latter leads to

\[
\|\hat{B}_\delta^{-1} - B^{-1}\| \leq \|\hat{B}_\delta^{-1}\| \|I(\hat{B} - B)\| \leq (2 \|B^{-1}\|)^{-1} \leq 2 \|B^{-1}\|^2 \|\hat{B} - B\|. \quad \text{(A.9)}
\]
Hence, by (A.9) and Chebyshev’s inequality
\[ E \| \hat{B}_g^{k-1} - B_0^{-1} \|^2_j = E \left[ \| \hat{B}_g^{k-1} - B_0^{-1} \|^2_j I(\| \hat{B} - B \| \leq 2\| B^{-1} \|^{-1}) \right] \\
+ E \left[ \| \hat{B}_g^{k-1} - B_0^{-1} \|^2_j I(\| \hat{B} - B \| > 2\| B^{-1} \|^{-1}) \right] \leq 2^{2j} \| B^{-1} \|^4 j E \| \hat{B} - B \|^2_j + 2^{2j} \delta^{-2j} (2\| B^{-1} \|)^k E \| \hat{B} - B \|^k \] (A.10)

for any \( k \), since \( \| B_0^{-1} \| \leq \delta^{-1} \) and \( \| \hat{B}_g^{k-1} \| \leq \delta^{-1} \). Finally, applying Hölder’s inequality and (A.7), we derive that for any positive \( j \) and \( k \)
\[ E \| \hat{B}_g^{k-1} - B^{-1} \|^2_j \leq 2^{2j-1} \left\{ E \| \hat{B}_g^{k-1} - B^{-1} \|^2_j + \delta^{2j} \| B^{-1} \|^4 j \right\} \leq 2^{2j-1} \left\{ 2^{2j} \| B^{-1} \|^4 j E \| \hat{B} - B \|^2_j \right. \\
+ \left. 2^{2j} \delta^{-2j} (2\| B^{-1} \|)^k E \| \hat{B} - B \|^k + \delta^{2j} \| B^{-1} \|^4 j \right\} \] (A.11)

Now, observe that from (A.6) by Cauchy’s inequality we have \( E \| \hat{a}_s - a \|^2 \leq 3(R_1 + R_2 + R_3) \) where \( R_1 = \| \| \|^2 E \| \hat{B}_g^{k-1} - B^{-1} \|^2 \), \( R_2 = \| \| \|^2 E \| \hat{c} - c \|^2 \), \( R_3 \leq \sqrt{E \| \hat{B}_g^{k-1} - B^{-1} \|^4 E \| \hat{c} - c \|^4} \). To complete the proof, apply (A.11) with \( j = 1 \) and \( k = 4 \) for \( R_1 \) and \( j = 2 \) and \( k = 8 \) for \( R_3 \).

Proof of Lemma 3.3: Introduce vector \( \varphi \) with components \( \varphi_k = \varphi(2^my - k), k \in K_{m,y} \), and note that
\[ \Delta_2 = 2^m E \left\{ (\hat{B}_g^{k-1} \hat{c} - B^{-1} \hat{c})^T \varphi \right\}^2 \leq 2^m \| \varphi \|^2 E \| \hat{a}_s - a \|^2. \] (A.12)

Recall that the set \( K_{m,y} \) contains \( M = 2^r(r + 1)(M_2 - M_1) \) terms and, by Lemma 3.1, \( B = p(y)I + O(2^{-m}), B^{-1} = [p(y)]^{-1}I + O(2^{-m}) \) and \( c = 2^{-m/2} \psi(y)D_0 \) where (see (3.7)) \( D_0 \) is the vector with identical components equal to \( \int_{-\infty}^{\infty} \varphi(x)dx \). Consequently,
\[ \| B^{-1} \| = \| [p(y)]^{-1}I + O(2^{-m}) \| \]
\[ \| B^{-1} \| = \| [p(y)]^{-1}I + O(2^{-m}) \| \]
\[ \| c \| = 2^{-m/2} \psi(y)\sqrt{M} \int_{-\infty}^{\infty} \varphi(x)dx + O(2^{-m}). \] (A.13)

Now, in order to apply Lemma A.4 to the expectation in (A.12), we need to find upper bounds for \( E \| \hat{B} - B \|^{2j}, j = 1, 2, 4, \) and \( E \| \hat{c} - c \|^{2j} \) with \( j = 1, 2 \). Let us start with \( E \| \hat{B} - B \|^{2j}. \) Observe that
\[ E(\hat{B}_{i,j} - B_{i,j})^2 \leq n^{-1} \int_{-\infty}^{\infty} \varphi_{m,i}^2(x)\varphi_{m,j}^2(x)p(x)dx \leq 2^m n^{-1} \int_{-\infty}^{\infty} \varphi^2(z - i)\varphi^2(z - j)dz = O(n^{-1}2^m). \] (A.14)
where \( \| P \|_C = \sup_x |p(x)| \). Similarly,

\[
E \| \hat{B} - B \|_4^4 \leq M^2 \sum_{i,j} E(\hat{B}_{ij} - B_{ij})^4 \\
\leq M^2 \left[ n^{-3} \sum_{i,j} \int_{-\infty}^{\infty} \varphi_m^2(x) \varphi_{m,j}^2(x)p(x)dx \\
+ 3n^{-2} \sum_{i_1,j_1} \sum_{i_2,j_2} \int_{-\infty}^{\infty} \varphi_m^2(x) \varphi_{m,j_1}^2(x)p(x)dx \right. \\
\left. \cdot \int_{-\infty}^{\infty} \varphi_m^2(x) \varphi_{m,j_2}^2(x)p(x)dx \right] \\
= O(n^{-2}2^m + n^{-1}2^m) \quad (A.15)
\]

Continuing in this manner, we obtain

\[
E \| \hat{B} - B \|_8^8 = O(n^{-7}2^m + n^{-6}2^m + n^{-5}2^m + n^{-4}2^m) \\
= O(n^{-4}2^m). \quad (A.16)
\]

For \( E \| \hat{c} - c \|_j^j \) with \( j = 1, 2 \), note that

\[
E \| \hat{c} - c \|_1^2 \leq n^{-1} \sum_k \int_{-\infty}^{\infty} u_{m,k}^2(x)p(x)dx \leq n^{-1} \| \varphi'_{(1)}(m) \|^2 \| P \|_C, \quad (A.17)
\]

\[
E \| \hat{c} - c \|_2^4 \leq M \left[ n^{-3} \sum_k \int_{-\infty}^{\infty} u_{m,k}^2(x)p(x)dx \\
+ 3n^{-2} \left( \sum_k \int_{-\infty}^{\infty} u_{m,k}^2(x)p(x)dx \right)^2 \right] \\
\leq M \| P \|_C \left[ n^{-3} \| \varphi'_{(2)}(m) \|^2 + 3n^{-2} \| \varphi'_{(1)}(m) \|^4 \right]. \quad (A.18)
\]

Plugging (A.14)–(A.18) and \( \delta \sim n^{-1}2^m \) into the expression for \( E \| \hat{a} - a \|_2^2 \) given by Lemma A.4 and finally using (A.12) and assumptions of Lemma 3.3, we arrive at (3.14).

**Proof of Theorem 3.4:** Validity of Theorem 3.4 follows directly from Lemmas 3.2 and 3.3.

\[\square\]

**References**


Empirical Bayes estimation by wavelet series


Marianna Pensky
Dept. of Mathematics
University of Central Florida
Orlando, FL 32816
USA
mpensky@pegasus.cc.ucf.edu

Mohammed Alotaibi
Dept. of Mathematics
University of Central Florida
Orlando, FL 32816
USA