



# Cardon Research Papers

---

## in Agricultural and Resource Economics

Research  
Paper  
2004-05  
September  
2004

## Empirical Bayes Nonparametric Kernel Density Estimation

Alan P. Ker  
*University of Arizona*

A. Tolga Ergün  
*Suffolk University*

The University of Arizona is an equal opportunity, affirmative action institution. The University does not discriminate on the basis of race, color, religion, sex, national origin, age, disability, veteran status, or sexual orientation in its programs and activities.

Department of Agricultural and Resource Economics  
College of Agriculture and Life Sciences  
The University of Arizona

This paper is available online at <http://ag.arizona.edu/arec/pubs/workingpapers.html>

---

Copyright ©2004 by the author(s). All rights reserved. Readers may make verbatim copies of this document for noncommercial purposes by any means, provided that this copyright notice appears on all such copies.

---

# Empirical Bayes Nonparametric Kernel Density Estimation

Alan P. Ker\*  
A. Tolga Ergün†

First Draft: May 2003  
This Draft: December 2003

## Abstract

This manuscript proposes using empirical Bayes techniques on estimated density values from nonparametric kernels in attempts to exploit potential similarities among a set of unknown densities. Our asymptotic theory and simulation results suggest that the empirical Bayes nonparametric kernel estimator may be a viable alternative to the standard kernel estimator when a set of possibly similar densities are being estimated. The strengths of the proposed estimator are (i) it allows all types of kernel estimators; and (ii) it does not require specification as to the degree or form of similarity.

**Keywords** : shrinkage estimator, Stein estimator, ensemble estimator

---

\*Department of Economics, University of Arizona. Email: aker@ag.arizona.edu

†Department of Economics, University of Arizona. Email: ergun@u.arizona.edu

# 1 Introduction

It is often necessary to estimate a set of densities  $\{f_1, f_2, \dots, f_Q\}$  that are thought to be of similar structure rather than a single density. In a parametric framework, similarity is imposed by assuming the set of densities belong to a single known parametric family. Although parametric approaches may be relatively efficient given they achieve the  $L_2$  rate  $O(n^{-1/2})$  under the assumption that the parametric family is correctly assumed, an incorrectly assumed parametric family will lead to inconsistent estimates which may result in significantly biased analyses and inferences. Unfortunately, the true parametric family is generally not discernible from the available information. In response, nonparametric density estimators have become mainstream in empirical applications despite slower  $L_2$  rates of convergence  $O(n^{-2/5})$ . Generally, any structural similarities among the densities tend to be left unexploited using nonparametric methods. In special circumstances, a known or estimable transformation may be invoked such that the set of densities collapses to a single density. In empirical analyses, however, such a transformation tends to be unrecoverable since the form or extent of the similarities is rarely known. We wish to construct an estimator that does not require any conditions or specifications on the degree of similarity but rather only a vague notion that the densities could possibly be of similar shape.

This manuscript proposes using empirical Bayes techniques on estimated density values from kernel-type estimators in attempts to exploit any similarities among the set of densities when the form or extent of the similarities is unknown. The estimator makes use of the pointwise limiting distribution of kernel estimators to construct a hierarchical model in an empirical Bayes framework. The proposed estimator lies in stark contrast to Bayesian nonparametric density estimation first proposed in the seminal article of Ferguson (1973). This line of estimators employs Gaussian mixtures and the Dirichlet process prior. With the introduction of Markov Chain Monte Carlo (MCMC) algorithms, empirical applications of Ferguson’s method have started to appear. See, for example, Diebolt and Robert (1994), Escobar and West (1995), and Roeder and Wasserman (1997). Note, however, that this approach has only been used—although not necessarily—to recover a single density and not a set of densities. The estimator proposed here differs in that kernel-type density estimators rather than Gaussian mixtures are employed and empirical Bayes is performed on the estimated density values as opposed to placing a prior on a given parameter space.

The main strengths of the proposed estimator are derived from the kernel estimator. First, since empirical Bayes techniques are employed with the values from the kernel estimator, all the variations of kernel estimators, including higher order kernels, variable kernel methods, and transformation-kernel density estimators, may be employed. Second, in the case when the set of densities are not identical, the empirical Bayes nonparametric kernel estimator converges in probability to the kernel estimator at a rate quicker than the kernel estimator converges in probability to the unknown density of interest. As a result, the asymptotic properties of the kernel estimator are passed along to the empirical Bayes nonparametric kernel estimator. Third, the proposed estimator does not require

any prior knowledge about the degree or form of possible similarities among the set of densities of interest. This is particularly beneficial because in empirical settings this is rarely (if ever) known. The manuscript proceeds by outlining the estimator, proving some asymptotic properties, and presenting our simulation results.

## 2 The Estimator

Consider estimating an unknown univariate density  $f$  based on a random sample  $X_1, \dots, X_n$  from  $f$ . The standard nonparametric kernel estimator of  $f$  at a support point  $x$  is defined as

$$\hat{f}(x) = \frac{1}{nh} \sum_{k=1}^n K\left(\frac{x - X_k}{h}\right)$$

where  $h = h(n)$  is the smoothing parameter and  $K$  is a second order kernel satisfying the usual properties  $\int K(t)dt = 1$ ,  $\int tK(t)dt = 0$ ,  $\int t^2K(t)dt = k_2 \neq 0$ . For a thorough review of kernel density estimators, see either Wand and Jones (1995), Scott (1994) or Silverman (1986). The asymptotic distribution of the kernel estimator is  $N(f(x) + \beta_x, \sigma_x^2)$  where  $\beta_x = \frac{1}{2}h^2 f''(x)k_2 + O(h^4)$  and  $\sigma_x^2 = \frac{1}{nh} f(x) \int K(t)^2 dt + o((nh)^{-1})$ . Parzen (1962) showed that the optimal smoothing parameter  $h_{opt}$  for minimizing the approximate mean integrated squared error is

$$h_{opt} = k_2^{-2/5} \left( \int K(t)^2 dt \right)^{1/5} \left( \int f''(x) dx \right)^{-1/5} n^{-1/5}.$$

Parzen (1962) also showed consistency of the kernel estimator if  $K$  is a bounded Borel function satisfying  $\int |K(t)|dt < \infty$ ,  $\int K(t)dt = 1$ , and  $|tK(t)| \rightarrow 0$  as  $|t| \rightarrow \infty$ . It was necessarily assumed that  $h \rightarrow 0$  and  $nh \rightarrow \infty$  as  $n \rightarrow \infty$ . For the purposes at hand, it is necessary to assume that  $f$  is uniformly continuous (with respect to Lebesgue measure) over some compact set of interest, say  $\mathcal{S}$ , so that the evaluation of  $f$  at points in  $\mathcal{S}$  is a mathematically well-defined operation.

Recall, we wish to consider not only a single density but a set of densities, for example, one for each experimental unit of interest. Suppose we have  $Q$  experimental units with densities  $\{f_1, \dots, f_Q\}$  and we observe random samples  $X_{i1}, \dots, X_{in_i}$  from  $f_i$  for  $i = 1, \dots, Q$  where  $n_i/n_j \rightarrow 1$ . For notational convenience we will assume  $n_i = n_j = n$  throughout. Denote the kernel estimate at  $x$  for experimental unit  $i$  as  $\hat{f}_i(x)$ . For notational convenience we will suppress the support point  $x$  as is commonly done. Based on the limiting distribution of kernel density estimators, the following hierarchical model is proposed:

$$\begin{aligned} \hat{f}_i | \mu_i &\sim N(\mu_i, \sigma_i^2) \\ \mu_i &\sim N(\mu, \tau^2) \end{aligned}$$

where  $\mu_i = f_i + \beta_i$ ,  $f_i$  is the unknown density value for experimental unit  $i$  at  $x$ ,  $\beta_i$  is the bias for  $\hat{f}_i$  at  $x$ ,  $\sigma_i^2$  is the variance of  $\hat{f}_i$  at  $x$ ,  $\mu$  is the mean of the  $\mu_i$ 's at  $x$ , and  $\tau^2$  is the variance of the  $\mu_i$ 's at  $x$ .

The intuition behind the hierarchical model is that even though the  $\mu_i$ 's are mutually independent for a given  $x$ , they are tied together in that there exists a squared error loss function for estimating the  $Q$  densities at support point  $x$ . Thus, in flavor similar to Stein's paradox, an estimator (the posterior mean, median, or mode) which is a function of the  $\{\hat{f}_1, \dots, \hat{f}_Q\}$  can be constructed which *may* be preferable to the kernel estimate. Given the above hierarchical model, the posterior mean is

$$\tilde{f}_i = \hat{f}_i \left( \frac{\tau^2}{\tau^2 + \sigma_i^2} \right) + \mu \left( \frac{\sigma_i^2}{\tau^2 + \sigma_i^2} \right)$$

where the unknowns  $(\mu, \tau^2, \sigma_i^2)$  must be estimated. The asymptotic variance or a bootstrapped estimator of the variance may be used to estimate  $\sigma_i^2$ . Estimators of the mean and variance across experimental units are obtained using the following method of moments estimators:  $\hat{\mu} = \frac{1}{Q} \sum_{i=1}^Q \hat{f}_i$  and  $\hat{\tau} = \hat{s}^2 - \frac{1}{Q} \sum_{i=1}^Q \hat{\sigma}_i^2$  where  $\hat{s}^2 = \frac{1}{Q-1} \sum_{i=1}^Q (\hat{f}_i - \hat{\mu})^2$ . Alternatively, one could use the biased marginal maximum likelihood estimator  $\tilde{s}^2 = \frac{1}{Q} \sum_{i=1}^Q (\hat{f}_i - \hat{\mu})^2$ . Also note that it is necessary to use  $\frac{1}{Q} \sum_{i=1}^Q \hat{\sigma}_i^2$  since it is not assumed that  $\sigma_i = \sigma_k \forall i \neq k$  as is commonly done. Thus, the empirical Bayes nonparametric kernel density estimator at support  $x$  for experimental unit  $i$  is

$$\tilde{f}_i = \hat{f}_i \left( \frac{\hat{\tau}^2}{\hat{\tau}^2 + \hat{\sigma}_i^2} \right) + \hat{\mu} \left( \frac{\hat{\sigma}_i^2}{\hat{\tau}^2 + \hat{\sigma}_i^2} \right).$$

The resulting posterior mean or empirical Bayes nonparametric kernel estimator  $\tilde{f}_i$  is very intuitive. As the estimated variance of the kernel estimates across experimental units increases ( $\hat{\tau}^2 \uparrow$ ), the posterior mean  $\tilde{f}_i$  will shrink less towards the overall mean  $\hat{\mu}$ . Conversely, the larger the estimated variance of the kernel estimate for a given experimental unit ( $\hat{\sigma}_i^2 \uparrow$ ), the posterior mean  $\tilde{f}_i$  will shrink more towards the overall mean  $\hat{\mu}$ . Not surprisingly, as with many shrinkage or Stein type estimators, the greater the estimated variance within the experimental units relative to the estimated variance across the experimental units, the greater the shrinkage and the greater the potential improvements in efficiency. Hence, it might be expected that the empirical Bayes nonparametric kernel estimator may offer the largest improvements in small samples where the estimated variance within experimental units tends to be relatively high as compared to the variance across experimental units. Note that, as usual in empirical Bayes estimators, estimation of a single experimental unit "borrows" information from all experimental units.

## 2.1 Finite-Sample Distribution of Kernel Estimators

The proposed estimator is based on the limiting distribution of kernel estimators being Gaussian. A practical question dealing with implementation of the estimator is, loosely speaking, how quickly the distribution of kernel estimators becomes approximately Gaussian. Note that the pointwise kernel estimator is simply the sum of individual kernels, one for each realization. Also note those realizations are independent. As a result, one might initially expect the distribution would be approximately Gaussian for relatively small samples. However, one cannot invoke the standard

central limit theorem. Consider a kernel function  $K$  that vanishes outside some compact set, say  $\Lambda$ . Generally, as  $n \uparrow$ ,  $h \downarrow$ , and  $\mathcal{L}(\Lambda) \downarrow$  where  $\mathcal{L}$  is the Lebesgue measure. As a result, when  $n$  increases, the number of elements over which the sum is taken does not necessarily increase. However, recognizing that the sum may be couched as a U-statistic, arguments from degenerate Martingale theory can be used to prove asymptotic normality (Hall, 1984). Intuitively though,  $\Lambda$  will influence the rate at which the distribution of the kernel estimator approaches normality because it will directly impact the number of terms in the sum. However, choosing a kernel such as the Gaussian, where  $\Lambda = \mathfrak{R}$ , will result in the distribution of the estimator approaching normality relatively quickly as all realizations will be included in the sum even as  $n$  grows. Hence, it may not be a substantial abuse to assume the distribution of the kernel estimator is Gaussian in many empirical applications even with moderate sized samples. A final note on the distribution of kernel estimators in relation to the empirical Bayes nonparametric kernel density estimator: if the pointwise estimates are approximately Gaussian and independent across experimental units, then their sum, and hence their mean, is also approximately Gaussian. Thus, it is only necessary that  $n \rightarrow \infty$  rather than both  $n \rightarrow \infty$  and  $Q \rightarrow \infty$  to justify the use of the hierarchical model.

## 2.2 Bias Reduction Methods

A second point regarding the proposed estimator is the bias of kernel estimators. In a similar type estimator in a nonparametric regression context, Altman and Casella (1995) were able to explicitly recognize the bias in their hierarchical model because they were able to recover an unbiased estimator of the regression curve. Unfortunately, there does not exist a nonparametric unbiased estimator of the density.

Recently, however, there has been a great deal of literature on improving the bias of kernel estimators from  $O(h^2)$  to  $O(h^4)$  (assuming that  $f$  has 4 continuous derivatives) while keeping the variance  $O((nh)^{-1})$ . Examples include higher order kernels and generalized jackknifing (e.g. Bartlett, 1963; Gajek, 1986; Jones and Foster, 1993; Schucany and Sommers, 1977), variable kernel methods (e.g. Abramson, 1982; Jones, 1990; Samiuddin and el-Sayyad, 1990), transformations (e.g. Ruppert and Cline, 1994), and multiplicative bias-correction approaches (e.g. Jones et al., 1995). One might at first suggest that the use of bias reducing kernel methods in the proposed estimator could result in improvements. However, these estimators, while reducing bias, are accompanied by increases in variance for finite samples. In fact, despite these methods having a lower asymptotic mean integrated squared error, for small to moderate sample sizes, improvements in mean integrated squared error may or may not be realized (Jones and Signorini, 1997). Hence, very little, if any, gain may result in empirical applications with common sample sizes by using the bias reduced kernel estimators within the empirical Bayes nonparametric kernel density estimator. Although the model ignores the bias as is customarily done with kernel estimators, we note that the empirical Bayes nonparametric kernel estimator is a convex combination of two biased estimators and thus bias does

not necessarily increase.

Altman and Casella (1995) pointed out that the bias of the nonparametric estimators did not allow them to recover an empirical Bayes estimator which dominates the standard nonparametric estimator. Instead, they provided a succinct and heuristic argument as to why shrinkage may generally be preferable. One problem with such a statement is the practical determination of when shrinkage is preferable to no shrinkage. Their simulation results suggest that if the variance across curves is low relative to the variance within curves, shrinkage is preferable. Rather than re-iterate a similar heuristic argument, the next section considers some asymptotic properties of the empirical Bayes nonparametric kernel density estimator.

### 3 Some Asymptotic Properties

Recall that the empirical Bayes nonparametric kernel density estimator for experimental unit  $i$  is defined as

$$\tilde{f}_i = \hat{f}_i \left( \frac{\hat{\tau}^2}{\hat{\tau}^2 + \hat{\sigma}_i^2} \right) + \hat{\mu} \left( \frac{\hat{\sigma}_i^2}{\hat{\tau}^2 + \hat{\sigma}_i^2} \right).$$

Deriving the exact form of bias and variance and thus the mean integrated squared error (MISE) of the estimator is not trivial as it is the sum of multiples of different estimators.<sup>1</sup> However, it is possible to consider the asymptotic behavior of the estimator in the following two scenarios: (i)  $f_i = f_k \forall i, k$  and (ii)  $f_i \neq f_k$  for  $i \neq k$ .

**Lemma 1.**  $\hat{\tau}^2$  is  $O_p(n^{-4/5})$  in scenario (i) and  $O_p(1)$  in scenario (ii).

*Proof.* After some algebra, we can show that

$$\begin{aligned} \hat{\tau}^2 &= \frac{1}{Q-1} \sum_{i=1}^Q (\hat{f}_i - \hat{\mu})^2 - \frac{1}{Q} \sum_{i=1}^Q \hat{\sigma}_i^2 \\ &= \frac{1}{Q-1} \left( \sum_{i=1}^Q (\hat{f}_i - f_i)^2 + \sum_{i=1}^Q (\hat{\mu} - f_i)^2 - 2 \sum_{i=1}^Q (\hat{f}_i - f_i)(\hat{\mu} - f_i) \right) - \frac{1}{Q} \sum_{i=1}^Q \hat{\sigma}_i^2 \end{aligned}$$

where

$$\sum_{i=1}^Q (\hat{\mu} - f_i)^2 = \frac{1}{Q^2} \sum_{i=1}^Q \left( (\hat{f}_i - f_i) + \sum_{k \neq i} (\hat{f}_k - f_i) \right)^2$$

and

$$\sum_{i=1}^Q (\hat{f}_i - f_i)(\hat{\mu} - f_i) = \frac{1}{Q} \sum_{i=1}^Q \left( (\hat{f}_i - f_i) \left( (\hat{f}_i - f_i) + \sum_{k \neq i} (\hat{f}_k - f_i) \right) \right).$$

---

<sup>1</sup>We suspect that viewing the estimator as a function of  $\hat{f}_i, \hat{\mu}, \hat{\tau}^2$ , and  $\hat{\sigma}_i^2$  and applying a Taylor expansion around  $f_i, \mu, \tau^2$ , and  $\sigma_i^2$  gives a linear form which can be more manageable.

Under (i),  $\sum_{i=1}^Q (\hat{\mu} - f_i)^2 = O_p(n^{-4/5})$  and  $\sum_{i=1}^Q (\hat{f}_i - f_i)(\hat{\mu} - f_i) = O_p(n^{-4/5})$  so  $\hat{\tau}^2 = O_p(n^{-4/5})$ . Under (ii),

$$\sum_{i=1}^Q (\hat{\mu} - f_i)^2 = \frac{1}{Q^2} \sum_{i=1}^Q \left( \sum_{i=1}^Q (\hat{f}_i - f_i) + \sum_{k \neq i} (f_k - f_i) \right)^2$$

which is  $O_p(1)$  and

$$\sum_{i=1}^Q (\hat{f}_i - f_i)(\hat{\mu} - f_i) = \frac{1}{Q} \sum_{i=1}^Q (\hat{f}_i - f_i) \left( \sum_{i=1}^Q (\hat{f}_i - f_i) + \sum_{k \neq i} (f_k - f_i) \right)$$

which is  $O_p(n^{-2/5})$ . So  $\hat{\tau}^2 = O_p(1)$ .  $\square$

If we define  $\hat{\lambda}_i = \hat{\tau}^2 / (\hat{\tau}^2 + \hat{\sigma}_i^2)$  then  $\tilde{f}_i = \hat{f}_i \hat{\lambda}_i + \hat{\mu}(1 - \hat{\lambda}_i)$ . Thus, given lemma 1,  $\hat{\lambda}_i = O_p(1)$  in both (i) and (ii). In (ii)  $\hat{\lambda}_i - 1 = O_p(n^{-4/5})$ , whereas in (i)  $\hat{\lambda}_i - c = O_p(n^{-4/5})$  where  $c < 1$ . Thus, we have the following theorem.

**Theorem 1.** *The empirical Bayes nonparametric kernel density estimator converges to the standard kernel estimator at a rate of  $O_p(n^{-4/5})$  in scenario (ii).*

*Proof.*

$$\begin{aligned} \tilde{f}_i - \hat{f}_i &= \hat{f}_i \left( \frac{\hat{\tau}^2}{\hat{\tau}^2 + \hat{\sigma}_i^2} \right) + \hat{\mu} \left( \frac{\hat{\sigma}_i^2}{\hat{\tau}^2 + \hat{\sigma}_i^2} \right) - \hat{f}_i \\ &= -\hat{f}_i \left( \frac{\hat{\sigma}_i^2}{\hat{\tau}^2 + \hat{\sigma}_i^2} \right) + \hat{\mu} \left( \frac{\hat{\sigma}_i^2}{\hat{\tau}^2 + \hat{\sigma}_i^2} \right) \\ &= (\hat{\mu} - \hat{f}_i) \left( \frac{\hat{\sigma}_i^2}{\hat{\tau}^2 + \hat{\sigma}_i^2} \right) \\ &= O_p(1) O_p(n^{-4/5}) \\ &= O_p(n^{-4/5}) \end{aligned}$$

$\square$

The intuition is straightforward since  $\hat{\tau}^2$  is  $O_p(1)$  and  $\hat{\sigma}_i^2$  is  $O_p(n^{-4/5})$ . Recall that the empirical Bayes nonparametric kernel density estimator is the convex combination of the kernel estimator and the mean of all kernel estimators. Note, however, that both the weight on the kernel estimator goes to one and the weight on the overall mean goes to zero at exactly the same rate as  $\hat{\sigma}_i^2$  goes to zero. As a result, the empirical Bayes nonparametric kernel density estimator converges to the standard kernel estimator at  $O_p(n^{-4/5})$ . Thus, in case (ii), the empirical Bayes nonparametric kernel density estimator inherits, asymptotically, the properties of the kernel estimator and converges to the unknown density at the optimal rate of  $O_p(n^{-2/5})$  since  $\tilde{f}_i - f_i = \tilde{f}_i - \hat{f}_i + \hat{f}_i - f_i = O_p(n^{-4/5}) + O_p(n^{-2/5}) = O_p(n^{-2/5})$ . The empirical Bayes nonparametric kernel density estimator converges in probability at a quicker rate to the kernel estimator than the kernel estimator converges in probability to the unknown density. As a result, the empirical Bayes nonparametric kernel density estimator converges to the unknown density at the same rate as the kernel estimator.



## 4 Simulation Results

In order to determine the appropriateness of our proposed estimator in empirical settings, we conducted twelve simulations. These simulations imposed three levels of similarity among the underlying densities and four different sample sizes  $\{25, 50, 100, 500\}$ . The first set we term “Low-Similarity” as we attempt to draw realizations from densities which are completely dissimilar. To this end, we use the first nine test densities of Marron and Wand (1992). These densities are Gaussian mixtures and exhibit a wide range of data generating processes; in fact, these densities are often used to evaluate the finite-sample performance of new density estimators. As a result, we feel this aptly represents a “worst-case scenario.” The second set of simulations we term “Identical” as the densities are identical, i.e., perfectly similar. To this end, we assume the nine densities are all standard Gaussian—the first of the test densities in Marron and Wand (1992). Obviously this is congruent to the first set in that this represents a “best-case scenario.” The third set of simulations, which we term “Moderate-Similarity,” attempt to draw realizations from densities that are somewhat similar. To this end, we draw realizations from the following five densities: (1)  $N(0,1)$ , (2)  $0.95 \cdot N(0,1) + 0.05 \cdot N(-2,0.5)$ , (3)  $0.9 \cdot N(0,1) + 0.1 \cdot N(-2,0.5)$ , (4)  $0.85 \cdot N(0,1) + 0.15 \cdot N(-2,0.5)$ , and (5)  $0.8 \cdot N(0,1) + 0.2 \cdot N(-2,0.5)$ . Note that in this last set we are systematically increasing the mass of the secondary distribution in the mixture.

For each set, the number of simulations is 500. In each simulation we calculate the smoothing parameter  $h$  by minimizing the ISE as is customary. In all tables,  $\hat{f}$  is the nonparametric kernel density estimator and  $\tilde{f}$  is the empirical Bayes nonparametric kernel density estimator. We use the standard Gaussian density as our kernel function. For  $\sigma_i^2$ , we use the bootstrapped estimator and for  $\tau^2$ , we use the positive part estimator. Table 1 contains the average MISE results over the nine densities (five for the moderate case). Table 2 contains the results for individual densities in the low-similarity case whereas table 3 contains the results in the moderate-similarity case. In table 1,  $\hat{\lambda}$  is  $\hat{\lambda}_i$  averaged over the grid points at which the densities are evaluated and then over the simulations and finally over the set of densities. In tables 2 and 3,  $\hat{\lambda}$  is  $\hat{\lambda}_i$  averaged only over the grid points at which the densities are evaluated and the simulations. One can loosely interpret  $\hat{\lambda}$  as an average measure of the weight in the empirical Bayes estimator placed on the standard kernel estimator  $\hat{f}$ .

Table 1: MISE\*1000 and  $\hat{\lambda}$

	Low-Similarity			Moderate-Similarity			Identical		
	$\hat{\lambda}$	$\tilde{f}$	$f$	$\hat{\lambda}$	$\tilde{f}$	$f$	$\hat{\lambda}$	$\tilde{f}$	$f$
$n = 25$	0.40	2.45	2.06	0.10	0.64	0.68	0.02	0.47	0.59
$n = 50$	0.48	1.57	1.36	0.10	0.38	0.43	0.04	0.28	0.39
$n = 100$	0.55	0.94	0.85	0.11	0.23	0.27	0.06	0.17	0.25
$n = 500$	0.76	0.28	0.27	0.20	0.08	0.09	0.10	0.05	0.08

Table 2: MISE\*1000 and  $\hat{\lambda}$  — Low-Similarity

		$\hat{\lambda}$	$\hat{f}$	$\hat{f}$
$n = 25$	density1	0.41	0.65	0.75
	density2	0.41	0.94	0.92
	density3	0.40	6.29	5.06
	density4	0.35	4.97	4.23
	density5	0.44	2.71	2.09
	density6	0.40	0.95	0.98
	density7	0.42	3.34	2.26
	density8	0.40	1.02	1.04
	density9	0.40	1.20	1.17
$n = 50$	density1	0.48	0.38	0.44
	density2	0.47	0.53	0.54
	density3	0.48	4.32	3.51
	density4	0.42	3.02	2.76
	density5	0.50	1.66	1.35
	density6	0.48	0.64	0.67
	density7	0.53	2.02	1.42
	density8	0.47	0.73	0.74
	density9	0.48	0.83	0.81
$n = 100$	density1	0.55	0.21	0.25
	density2	0.53	0.31	0.31
	density3	0.56	2.54	2.16
	density4	0.48	1.80	1.74
	density5	0.54	0.99	0.87
	density6	0.58	0.40	0.42
	density7	0.62	1.12	0.86
	density8	0.55	0.51	0.52
	density9	0.58	0.56	0.55
$n = 500$	density1	0.78	0.07	0.08
	density2	0.73	0.10	0.10
	density3	0.74	0.76	0.70
	density4	0.65	0.53	0.53
	density5	0.63	0.25	0.26
	density6	0.82	0.13	0.13
	density7	0.83	0.30	0.26
	density8	0.79	0.17	0.17
	density9	0.82	0.18	0.18

The best-case (identical) results are as expected; since all the densities are the same, the empirical Bayes estimator “borrows” information from the same data generating processes and performs quite favorably against the standard kernel estimator. Recall that the empirical Bayes estimator does not converge to the standard kernel estimator in the best-case scenario (see lemma 1 case (i)).

In the worst-case scenario,  $\hat{\lambda}$  converges to one and does so quite fast since in this case the denominator approaches the numerator at a rate of  $O_p(n^{-4/5})$  (see lemma 1 case (ii)). The standard

Table 3: MISE\*1000 and  $\hat{\lambda}$  — Moderate-Similarity

		$\hat{\lambda}$	$\hat{f}$	$\hat{f}$
$n = 25$	density1	0.09	0.64	0.68
	density2	0.09	0.66	0.69
	density3	0.10	0.67	0.71
	density4	0.10	0.63	0.65
	density5	0.10	0.59	0.69
$n = 50$	density1	0.10	0.38	0.41
	density2	0.10	0.39	0.43
	density3	0.10	0.38	0.42
	density4	0.10	0.37	0.42
	density5	0.11	0.38	0.48
$n = 100$	density1	0.11	0.23	0.25
	density2	0.11	0.24	0.26
	density3	0.11	0.24	0.27
	density4	0.11	0.23	0.28
	density5	0.12	0.23	0.28
$n = 500$	density1	0.18	0.08	0.08
	density2	0.19	0.08	0.09
	density3	0.20	0.08	0.09
	density4	0.21	0.08	0.09
	density5	0.20	0.08	0.09

kernel estimator in the worst-case outperforms the empirical Bayes estimator as expected. Recall, we purposely chose the set of densities to be very dissimilar. However, when working with real data, even though the true data generating processes are unknown, one commonly would have a rough idea about similarity. The extent of dissimilarity that these test densities exhibit would and should not lead someone to use the proposed estimator. Despite this, the empirical Bayes estimator performs quite competitively and, as expected from theorem 1, converges at a faster rate to the standard kernel than does the standard kernel estimator to the true density. This is evident in the simulation results. When  $n = 25$ , the MISE of the empirical Bayes estimator is 19% higher than the standard kernel estimator whereas when  $n = 500$  the MISE is only 4% higher. For each of the nine densities, the difference between the MISE for the standard kernel estimator and the empirical Bayes estimator shrinks as  $n$  increases since  $\hat{\lambda}$  increases with  $n$ . It is interesting that the empirical Bayes estimator outperforms the standard kernel estimator for some of the nine densities. As a reviewer pointed out, the greatest improvement with the empirical Bayes estimator is for densities closest to the mean of all nine densities and the improvement declines as the density departs from the mean.

The somewhat similar (moderate-similarity) simulation shows how the proposed estimator can perform quite favorably in a compromise situation. The MISE of the empirical Bayes estimator is smaller than or equal to the standard kernel estimator for all five test densities at all four sample

sizes. As expected,  $\hat{\lambda}$  falls between the  $\hat{\lambda}$  in the best-case scenario and the  $\hat{\lambda}$  in the worst-case scenario. Also as expected,  $\hat{\lambda}$  increases with  $n$  indicating that as  $n$  increases more weight is given to the standard kernel estimate and less to  $\hat{\mu}$ . Again, the closer the density is to the mean of the five densities, the greater the efficiency gain with the empirical Bayes estimator.

## 5 Conclusions

This manuscript proposed using empirical Bayes techniques on estimated density values from nonparametric kernels in attempts to exploit potential similarities among the set of densities. This is in contrast to Bayesian nonparametric density estimation which employs Gaussian mixtures and the Dirichlet process prior. Our asymptotic and simulation results suggest that the empirical Bayes nonparametric kernel estimator may be a viable alternative to the standard kernel estimator when a set of densities are being estimated. The empirical Bayes nonparametric kernel estimator is forwarded only as an alternative estimator to the standard kernel estimator.

The strengths of the empirical Bayes nonparametric kernel estimator are derived from the kernel estimator. Since the empirical Bayes techniques are employed with the fitted values of the kernel estimators, all the variations of kernel estimators including higher order kernels, variable kernel methods, and transformation-kernel density estimators may be employed. In the worst-case scenario when the data generating processes are all different, the empirical Bayes nonparametric kernel estimator converges in probability to the standard kernel estimator at a rate quicker than the standard kernel estimator converges in probability to the unknown density of interest, and thus the asymptotic properties of the standard kernel are passed along to the empirical Bayes nonparametric kernel estimator. One of the proposed estimator's major strengths is that knowledge about the form or the degree of similarity among the set of densities is not required. Although not discussed, the extension to higher dimensions is trivial in that although the kernel estimator changes, the hierarchical model does not.

## References

- [1] Abramson, I.S. (1982), On bandwidth variation in kernel estimates—a square root law, *Ann. Statist.* 10, 1217-1223.
- [2] Altman, N.S. and G. Casella (1995), Nonparametric empirical Bayes growth curve analysis, *J. Amer. Statist. Assoc.* 90, 508-515.
- [3] Bartlett, M.S. (1963), Statistical estimation of density functions, *Sankhya A* 25, 245-254.
- [4] Diebolt, J. and C. Robert (1994), Estimation of finite mixture distributions through Bayesian sampling, *J. Roy. Statist. Soc. B* 56, 363-375.
- [5] Escobar, M.D. and M. West. (1995), Bayesian density estimation and inference using mixtures, *J. Amer. Statist. Assoc.* 90, 577-588.
- [6] Ferguson, T.S. (1973), A Bayesian analysis of some nonparametric problems, *Ann. Statist.* 1, 209-230.
- [7] Gajek, L. (1986), On improving density estimators which are not bona fide functions, *Ann. Statist.* 14, 1612-1618.
- [8] Hall, P. (1984), Central limit theorem for integrated square error of multivariate density estimators, *J. Multivariate Anal.* 14, 1-16.
- [9] Jones, M.C. (1990), Variable kernel density estimates, *Australian J. Statist.* 32, 361-371.
- [10] Jones, M.C. and P.J. Foster (1993), Generalized jackknifing and higher order kernels, *J. Nonparametric Statist.* 3, 81-94.
- [11] Jones, M.C. and D.F. Signorini (1997), A comparison of higher-order bias kernel density estimators, *J. Amer. Statist. Assoc.* 92, 1063-1073.
- [12] Jones, M.C., O. Linton and J.P. Nielsen (1995), A simple and effective bias reduction method for kernel density estimation, *Biometrika* 82, 327-338.
- [13] Marron, J.S. and M.P. Wand (1992), Exact mean integrated squared error, *Ann. Statist.* 20, 712-736.
- [14] Parzen, E. (1962), On estimation of a probability density function and mode, *Ann. Math. Statist.* 33, 1065-1076.
- [15] Roeder, K. and L. Wasserman (1997), Practical Bayesian density estimation using mixtures of normals, *J. Amer. Statist. Assoc.* 92, 894-902.

- [16] Ruppert, D. and D.B.H. Cline (1994), Bias reduction in kernel density estimation by smoothed empirical transformations, *Ann. Statist.* 22, 185-210.
- [17] Samiuddin, M. and G.M. el-Sayyad (1990), On nonparametric kernel density estimates, *Biometrika* 77, 865-874.
- [18] Scott, D.W. (1994), *Multivariate Density Estimation: Theory, Practice, and Visualization* (Wiley, New York).
- [19] Schucany, W.R. and J.P. Sommers (1977), Improvement of kernel type density estimators, *J. Amer. Statist. Assoc.* 72, 420-423.
- [20] Silverman, B.W. (1986), *Density Estimation for Statistics and Data Analysis* (Chapman and Hall, New York).
- [21] Wand, M.P. and M.C. Jones (1995), *Kernel Smoothing* (Chapman and Hall, New York).