

# Deconvolution estimation of mixture distributions with boundaries

Mihee Lee\* and Peter Hall

*Department of Mathematics and Statistics, University of Melbourne  
Parkville, VIC, Australia*

*e-mail: miheel@unimelb.edu.au; halpstat@ms.unimelb.edu.au*

Haipeng Shen\*, J. S. Marron\* and Jon Tolle

*Department of Statistics and Operations Research, University of North Carolina  
Chapel Hill, NC 27599, USA*

*e-mail: haipeng@email.unc.edu; marron@email.unc.edu; tolle@email.unc.edu*

Christina Burch†

*Department of Biology, University of North Carolina  
Chapel Hill, NC 27599, USA*

*e-mail: cburch@bio.unc.edu*

**Abstract:** In this paper, motivated by an important problem in evolutionary biology, we develop two sieve type estimators for distributions that are mixtures of a finite number of discrete atoms and continuous distributions under the framework of measurement error models. While there is a large literature on deconvolution problems, only two articles have previously addressed the problem taken up in our article, and they use relatively standard Fourier deconvolution. As a result the estimators suggested in those two articles are degraded seriously by boundary effects and negativity. A major contribution of our article is correct handling of boundary effects; our method is asymptotically unbiased at the boundaries, and also is guaranteed to be nonnegative. We use roughness penalization to improve the smoothness of the resulting estimator and reduce the estimation variance. We illustrate the performance of the proposed estimators via our real driving application in evolutionary biology and two simulation studies. Furthermore, we establish asymptotic properties of the proposed estimators.

**AMS 2000 subject classifications:** Primary 62G08, 62H25; secondary 65F30.

**Keywords and phrases:** Boundary effect, maximum likelihood, measurement error, mixture distribution, penalization, sieve method.

Received March 2012.

## Contents

1	Introduction . . . . .	324
2	Description of the virus lineage data . . . . .	326

\*Supported in part by NSF grants DMS-0606577, DMS-08054908, CMMI-0800575, DMS-1106912, and The Xerox Foundation UAC Award.

†Supported in part by NIH grant R01-GM067940.

3	Model and methodology . . . . .	327
3.1	Model . . . . .	327
3.2	A standard sieve estimator . . . . .	328
3.3	A penalized sieve estimator . . . . .	330
3.4	Parameter selection . . . . .	331
3.4.1	Selection of $r$ . . . . .	331
3.4.2	Selection of $\lambda$ . . . . .	331
4	Simulation studies . . . . .	332
4.1	Simulation Study I . . . . .	332
4.2	Simulation Study II . . . . .	334
5	Application to the virus lineage data . . . . .	335
5.1	Estimation results . . . . .	335
5.2	Validation of the exponential assumption . . . . .	336
6	Consistency of the proposed estimators . . . . .	337
	Acknowledgements . . . . .	340
	References . . . . .	340

## 1. Introduction

The research described in this paper is motivated primarily by an important application in evolutionary biology, where biologists are interested in estimating the distribution of virus mutation effects (Burch et al., 2007; Lee et al., 2010). Biological prior information suggests that there are two kinds of mutations: no or silent mutations which have no effects on virus fitness; and deleterious mutations whose effects are continuous and positive in reducing virus fitness. Hence, the target distribution of the mutation effect is a mixture of a pointmass at zero and a positively-supported continuous distribution, which has a non-smooth left boundary at the origin, i.e. the density is discontinuous at the left boundary of the support. The observations are also contaminated with measurement errors. For more details on the biological experiment and the relevant data, see Section 2.

To the best of our knowledge, this is the first paper considering the aforementioned two special features: (1) discrete and continuous mixtures, and (2) non-smooth boundaries. There are only two recent works (van Es et al., 2008; Lee et al., 2010) which consider mixtures of one discrete atom and one continuous component in the context of measurement error models. (The two estimators are essentially the same, whose convergence rates were recently derived by Gugushvili et al. (2011).) However, they do not consider boundaries, hence the proposed estimators give poor performance near non-smooth boundaries. Moreover, since they use conventional Fourier deconvolution methods, their estimators of nonnegative densities suffer from negativity. On the other hand, while the deconvolution problem with known boundaries has been studied by Pensky (2002), Hall and Qiu (2005), Meister (2007), and Zhang and Karunamuni (2009), these authors consider only continuous target distributions. An interesting direction for future work is to extend these methods, for example the

boundary kernels of Zhang and Karunamuni (2009), to the problem of estimating discrete/continuous mixture distributions which is the focus of this paper.

In the context of purely continuous distributions, existing deconvolution methods mainly fall into two groups. The first group contains Fourier-based methods that use Fourier and Inverse Fourier transformations along with nonparametric smoothing. For detailed description of various methods and their asymptotic properties, see van Es et al. (2008), Lee et al. (2010) and references therein. The second group consists of non-Fourier based deconvolution methods. Many such methods first employ basis functions such as B-splines or wavelets to expand the target density (or distribution) function, and then estimate the basis coefficients using various approaches. For example, see Johnstone et al. (2004), Staudenmayer et al. (2008) and references therein. In addition, several other alternatives for deconvolution have been proposed, such as NPMLE, SIMEX, and TAYLEX, which are well reviewed in, for example, Carroll et al. (2006), Wagner and Stadtmüller (2008), and Wang et al. (2009). Staudenmayer et al. (2008) recently proposed a B-spline Bayesian deconvolution approach, considering only continuous target variables.

The statistical deconvolution setting starts with some observations of a variable  $Y$ , which is the sum of two variables: one is the *unobservable* variable whose distribution we are interested in estimating, say  $X$ , and the other is an error variable,  $Z$ . In particular, to estimate the target distribution of  $X$ , only the error contaminated observations  $Y = X + Z$  are available. Density estimation of  $X$  in this case, called the deconvolution problem, has been studied widely. Nevertheless, as discussed above, most earlier work restricts  $X$  to be continuous. In this paper, we consider practically important cases with two special features:

- (1) the target distribution of  $X$  is a mixture of a finite number of discrete atoms and a continuous distribution;
- (2) the continuous mixture component has non-smooth boundaries.

This combination of discrete and continuous mixtures and non-smooth boundary makes this problem challenging. To solve the problem, we employ three main ideas: *discretization*, *maximum likelihood* and *penalization*. First, we approximate the distribution of  $X$  using discretization, which gives a *sieve* of the distribution family. The mixture structure, and any information known about the distribution such as its atoms and boundaries, are reflected in the construction of the sieve. Then, we estimate the distribution using maximum likelihood within each sieve. The measurement error problem is naturally addressed via computation of the likelihood function. In addition, when smoothness of the target distribution is assumed, we improve the proposed basic sieve estimator using a roughness penalty function, which results in our penalized sieve estimator.

Sieve type estimators have been proposed for deconvolution problems by Cordy and Thomas (1997), whose technique can be extended to our sieve estimator (Section 3.2) when degenerate distributions are used to approximate the continuous mixture component. However, our approach has several advantages: (i) we introduce penalization to improve the estimation performance for smooth target functions; (ii) our ready-to-be-distributed optimization implementation

is more direct and efficient, and (iii) theoretical properties of our estimators are accessible. In the simple error-free case, Ruppert et al. (2007) proposed a sieve type density estimator for certain special distributions with known boundaries.

Although this work is motivated by the evolutionary biology application, it is also applicable to other areas. For example, another potential application is the nonparametric empirical Bayes approach to estimation of a high dimensional vector of normal means (Brown and Greenshtein, 2009; Greenshtein and Park, 2009; Raykar and Zhao, 2011). For this application, the distribution of interest is the prior distribution used for the empirical Bayes approach. For sparse scenarios when some of the normal means are assumed to be exactly zero, the prior distribution would be a mixture of an atom at zero and another continuous component; for the non-sparse scenarios, the prior would be just a continuous component. With or without the sparsity assumption, our methods can be used to estimate the prior distribution nonparametrically and enable posterior inference.

The remainder of the paper is organized as follows. Section 2 describes our motivating virus application and the virus lineage data provided by Burch et al. (2007). In Section 3, we explicitly state our problem of interest and the model, and then propose a standard sieve estimator and a penalized sieve estimator, along with some estimation algorithms, including procedures for data-driven parameter selection. Section 4 illustrates the performance of the estimators via two simulation studies. We analyze the virus lineage data in Section 5. Section 6 studies asymptotic properties of the proposed estimators. The online supplement (Lee et al., 2013) contains technical details of the optimization algorithm for our estimators and the proofs for the theorems.

## 2. Description of the virus lineage data

Evolutionary biologists performed experiments on 10 virus lineages to propagate the accumulation of mutations over time (Burch and Chao, 2004; Burch et al., 2007). During the experiments, the viruses were plated onto a lawn of some standard host, and formed plaques as they evolved. The final scientific goal was to estimate the distribution of mutation effects on virus fitness, denoted as  $S$ .

Unfortunately, virus fitness is not directly measurable in the experiments. Burch and Chao (2004) experimentally determined that  $S$  is related with the mutation effect on plaque size  $X$  through the following equation

$$X = 22.73 \log(1 + S). \quad (2.1)$$

Hence, to estimate the distribution of mutation effect on fitness, one normally first estimates the distribution of mutation effect on plaque size, and then transforms the resulting distribution to the distribution of fitness through inverting the above relationship.

Due to the biological reasons given above, for each lineage, the plaque sizes were measured at 40 different time points sequentially. Define  $Y$  as the reduction between two consecutive plaque size measurements. The reduction  $Y$  can be considered as the sum of two components: one is the real mutation effect on

plaque sizes, i.e.  $X$ , and the other is a measurement error  $Z$  which comes from the technical difficulty in measuring plaque sizes. Experimental details suggest that it is reasonable to assume  $Z$  to be normally distributed.

The virus lineage data have three important intrinsic features that need to be taken into account:

- (1) a mixture structure of the mutation effect distribution;
- (2) the known boundary information;
- (3) the existence of measurement errors.

As will be seen in Section 3, our estimators are developed to incorporate all three features. Among previous analyses of the data, Burch et al. (2007) considered only the mixture structure, and Lee et al. (2010) did not properly incorporate the boundary information. Below we elaborate on these three features.

According to Burch et al. (2007), the virus considered in the experiments is fairly advanced in evolution, and the correspondingly biological model gives only two mutation possibilities during a given time interval:

- no mutation or a silent mutation occurs;
- a deleterious mutation occurs.

A silent mutation is defined as a mutation that has no effect on virus fitness or plaque size, hence the theoretical mutation effect of this first case is 0. As for the second case, a deleterious mutation reduces virus fitness or the plaques sizes, so its effect is positive in terms of the plaque size reduction. In addition, it is usually assumed that the effect of the deleterious mutation is continuous. As a result, the distribution of mutation effects on plaque size should be modeled as a mixture of a pointmass at 0 and a continuous distribution supported on the positive half-line. Moreover, the observations  $Y$  are contaminated with measurement errors that are usually not negligible. This makes it necessary to consider the measurement error model on top of the mixture structure.

As discussed above, the observations  $Y$  are either mutation effects contaminated with measurement errors, if deleterious mutations occurred during the measured time points; or purely measurement errors, if no mutations or only silent mutations occurred. Hence our interest is to estimate (i) the relative frequency of the deleterious mutations, and (ii) the distribution or density of the *continuous* deleterious mutation effects, which is supported only on the positive real line.

### 3. Model and methodology

In this section, we first describe the model that we are interested in, and then develop two sieve estimation procedures in Sections 3.2 and 3.3. Parameter selection methods are discussed in Section 3.4.

#### 3.1. Model

Suppose we observe independent and identically distributed data  $Y_1, \dots, Y_n$  generated from

$$Y = X + Z, \tag{3.1}$$

where  $X$  and  $Z$  are independent,  $Z$  has a known density  $f_Z$ , and we wish to estimate the distribution of  $X$ . We assume that, for an integer  $\nu \geq 0$ ,

$$F_X(x) = P(X \leq x) = \pi_1 I(a_1 \leq x) + \cdots + \pi_\nu I(a_\nu \leq x) + \pi_{\nu+1} P(X_c \leq x), \quad (3.2)$$

where the nonnegative quantities  $\pi_1, \dots, \pi_{\nu+1}$  add up to 1, and the random variable  $X_c$  has a continuous distribution. Without loss of generality, we assume that  $a_l < a_{l+1}$  for any  $l$ .

The case studied in the literature corresponds to  $\nu = 0$ , where  $X$  is continuous, or, in effect,  $X = X_c$ . It will be assumed that the number of discrete atoms  $\nu$  and the atoms  $a_1, \dots, a_\nu$ , although not their masses  $\pi_1, \dots, \pi_\nu$ , are known, which is true for our virus lineage application. We estimate  $\pi_1, \dots, \pi_{\nu+1}$  and the distribution of  $X_c$ . By reflecting the data-generating mechanism in this way, we give our method an opportunity to respond correctly to the model that produced the data; alternative approaches, e.g. van Es et al. (2008) and Lee et al. (2010), generally restrict  $X_c$  to have a continuous density and produce estimators that do not respond well to boundary jumps and take negative values beyond the boundaries.

Indeed, the sieve techniques developed below do not suffer from these difficulties. Even in cases where there is no atom and the density of  $X_c$  is continuous on the real line, our methodology is still attractive, because it produces distribution and density estimators which are bona fide distribution and density functions, respectively. This is of substantial value in many practical problems. In particular, if one of our aims in estimating the distribution of  $X$  is to enable bootstrap simulation from the estimated distribution, then it is essential that the estimator is a proper distribution function. Another advantage of our method is the reduction of computational cost: the computation time is much shorter than that of the standard Fourier deconvolution estimators of van Es et al. (2008) and Lee et al. (2010); in addition, the specially tailored optimization procedure is more efficient than the built-in optimizers in standard statistical packages such as MATLAB and R.

The practical evolutionary biology problem summarized in the Introduction and discussed in detail in Section 2, which motivated our work, involved  $\nu = 1$  and  $a_1 = 0$ , and required  $X_c$  to be distributed on the positive real line. Biological considerations indicated that the density of  $X_c$  would likely have a jump discontinuity at the origin, but otherwise be continuous on  $(0, \infty)$ .

### 3.2. A standard sieve estimator

In this subsection, we develop a basic sieve estimation procedure. Our proposed method is of the *sieve* type (Grenander, 1981); we first consider a sieve, which is a sequence of classes of specific distributions, and restrict the problem to estimation in the sieve. Then the sieve is extended to the entire distribution space as the sample size increases, and hence we obtain a solution to the original estimation problem.

We establish a lattice on the support of  $X_c$ , taking it to be the sequence of equally spaced points  $\{x_j : 1 \leq j \leq r\}$  where  $x_{j+1} - x_j = h > 0$  plays a role similar to that of a bandwidth, or of a binwidth in histogram estimation. For example, when  $X_c$  is nonnegative, we can take  $x_j = (j - 0.5)h$ . The distribution of  $X_c$  in (3.2) is then approximated by the distribution of a discrete random variable  $\tilde{X}_c$  with potential atoms at each of the points  $x_j$ :

$$P(\tilde{X}_c = x_j) = \theta_j \quad \text{for } j = 1, \dots, r. \quad (3.3)$$

Here, each  $\theta_j$  is nonnegative, and  $\sum_j \theta_j = 1$ , i.e.  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_r)^T$  denotes an  $r$ -dimensional parameter vector for a univariate discrete distribution.

To estimate the parameter  $\boldsymbol{\theta}$ , we proceed as follows. Suppose that the measurement error  $Z$  has a known density function  $f_Z$ , and let  $f_Y$  be the density of  $Y$ . In view of (3.1) and (3.3), our approximation to  $f_Y$  is given by

$$f_Y(y|\boldsymbol{\pi}, \boldsymbol{\theta}) = \sum_{l=1}^{\nu} \pi_l f_Z(y - a_l) + \pi_{\nu+1} \sum_{j=1}^r \theta_j f_Z(y - x_j), \quad (3.4)$$

where  $\boldsymbol{\pi} = (\pi_1, \dots, \pi_{\nu+1})^T$ . We use maximum likelihood to estimate the parameters  $\boldsymbol{\pi}$  and  $\boldsymbol{\theta}$ . Theoretical properties of the estimation procedure are investigated in Section 6.

According to (3.4), the log-likelihood function of  $(\boldsymbol{\pi}, \boldsymbol{\theta})$  can be written as

$$\ell(\boldsymbol{\pi}, \boldsymbol{\theta}) = \sum_{i=1}^n \log \left[ \sum_{l=1}^{\nu} \pi_l f_Z(Y_i - a_l) + \pi_{\nu+1} \sum_{j=1}^r \theta_j f_Z(Y_i - x_j) \right]. \quad (3.5)$$

The *standard sieve* (SS) estimator for  $(\boldsymbol{\pi}, \boldsymbol{\theta})$  is then defined as the maximizer of the above log-likelihood function, i.e.,

$$(\hat{\boldsymbol{\pi}}^{SS}, \hat{\boldsymbol{\theta}}^{SS}) = \operatorname{argmax}_{\boldsymbol{\pi}, \boldsymbol{\theta}} \left\{ \ell(\boldsymbol{\pi}, \boldsymbol{\theta}) : \pi_l \geq 0, \theta_j \geq 0, \sum_{l=1}^{\nu+1} \pi_l = 1, \text{ and } \sum_{j=1}^r \theta_j = 1 \right\}.$$

The parameters can be estimated using an EM algorithm, where the discretized version of  $X$  is considered as a missing variable. However, each maximization step still involves a nontrivial constrained optimization. We first tried to use standard built-in optimizers in Matlab and R, but none led to stable results. We then developed our own optimizer by carefully combining the Karush-Kuhn-Tucker condition in optimization (Bertsekas, 2005) with Newton's iteration algorithm. Details of our optimization algorithm can be found in the online supplement (Lee et al., 2013).

Once we obtain the estimator  $(\hat{\boldsymbol{\pi}}^{SS}, \hat{\boldsymbol{\theta}}^{SS})$ , we take  $\hat{\boldsymbol{\pi}}^{SS}$  to be our estimator of  $\boldsymbol{\pi}$ . To estimate the density of  $X_c$ , we construct a continuous function on  $(x_1 - 0.5h, x_r + 0.5h)$  by interpolating the  $(x_j, \hat{\theta}_j)$ , using linear interpolation as follows. Notice that

$$\begin{aligned} \theta_j &= P(\tilde{X}_c = x_j) = P(x_j - 0.5h \leq \tilde{X}_c < x_j + 0.5h) \\ &\approx P(x_j - 0.5h \leq X_c < x_j + 0.5h) \approx h f_c(x_j). \end{aligned}$$

These approximations motivate defining

$$\tilde{f}_c^{SS}(x) = \frac{\hat{\theta}_{j-1}}{h} + \frac{\hat{\theta}_j - \hat{\theta}_{j-1}}{h(x_j - x_{j-1})}(x - x_{j-1}) \quad \text{for } x \in [x_{j-1}, x_j] \quad (3.6)$$

for any  $1 \leq j \leq r+1$ , and 0 otherwise. We set  $x_0 = x_1 - 0.5h$ ,  $x_{r+1} = x_r + 0.5h$ ,  $\hat{\theta}_0 = \hat{\theta}_1$  and  $\hat{\theta}_{r+1} = \hat{\theta}_r$ . It can easily be shown that  $\tilde{f}_c^{SS}$  is a proper continuous density function. The corresponding distribution estimator  $\tilde{F}_c^{SS}$  is obtained from  $\tilde{f}_c^{SS}$  by integration. Higher order interpolations can similarly be defined if required.

### 3.3. A penalized sieve estimator

The standard sieve estimation described above in Section 3.2 involves discretizing the distribution of the continuous component  $X_c$ . As a result of this discretization, the resulting density estimator  $\hat{f}_c^{SS}$  can be rather rough for finite samples. To overcome this problem, in this section we introduce a roughness penalty on  $\theta$ .

More specifically, we consider the following *penalized* log-likelihood function,

$$\ell_\lambda(\boldsymbol{\pi}, \boldsymbol{\theta}) = \ell(\boldsymbol{\pi}, \boldsymbol{\theta}) - \lambda P(\boldsymbol{\theta}), \quad (3.7)$$

where  $\ell(\boldsymbol{\pi}, \boldsymbol{\theta})$  is the log-likelihood in (3.5),  $P(\cdot)$  is a roughness penalty on  $\boldsymbol{\theta}$ , and  $\lambda$  is a penalty parameter that balances the effects of the log-likelihood and the penalty term. We then propose the *penalized sieve* (PS) estimator for  $(\boldsymbol{\pi}, \boldsymbol{\theta})$  as the maximizer of the above penalized log-likelihood (3.7), i.e.,

$$(\hat{\boldsymbol{\pi}}^{PS}, \hat{\boldsymbol{\theta}}^{PS}) = \underset{\boldsymbol{\pi}, \boldsymbol{\theta}}{\operatorname{argmax}} \left\{ \ell_\lambda(\boldsymbol{\pi}, \boldsymbol{\theta}) : \pi_l \geq 0, \theta_j \geq 0, \sum_{l=1}^{\nu+1} \pi_l = 1, \text{ and } \sum_{j=1}^r \theta_j = 1 \right\}.$$

The roughness penalty  $P(\cdot)$  in (3.7) can be any function that decreases as  $\boldsymbol{\theta}$  gets smoother. In this paper, we choose the sum of squared first order differences, i.e.  $P(\boldsymbol{\theta}) = \sum_{j=2}^r (\theta_j - \theta_{j-1})^2$ . This function achieves the minimum value of 0 when all the  $\theta_j$ s are the same.

Furthermore, we can combine the above two sieve estimation procedures. The following hybrid approach results in a discrete distribution estimator  $\hat{\boldsymbol{\pi}}$  with small bias, and also produces a smoother density estimator. To implement the hybrid method, first estimate  $\boldsymbol{\pi}$  by the standard sieve estimator  $\hat{\boldsymbol{\pi}}^{SS}$ . After that, apply the penalized sieve estimator to obtain a smoother estimator of the density  $f_c$ . Specifically, define  $\hat{\boldsymbol{\theta}}^{PS}$  to be the maximizer of the following penalized conditional log-likelihood function,

$$\hat{\boldsymbol{\theta}}^{PS} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \left\{ \ell(\boldsymbol{\theta} | \hat{\boldsymbol{\pi}}^{SS}) - \lambda P(\boldsymbol{\theta}) : \theta_j \geq 0 \text{ and } \sum \theta_j = 1 \right\}.$$

After obtaining  $\hat{\boldsymbol{\theta}}^{PS}$ , we can use linear interpolation in the manner as discussed in Section 3.2.



### 3.4. Parameter selection

The sieve estimators involve two tuning parameters: the number of bins used to approximate the continuous density,  $r$ , and the penalty parameter for the penalized sieve estimator,  $\lambda$ . In this subsection we discuss the issue of parameter selection, and propose selecting the two parameters in a sequential manner: first choose an “optimal”  $r$  to derive our SS estimator, and then choose an “optimal”  $\lambda$  for the fixed  $r$  in the PS estimation. The sequential selection avoids simultaneous choice of the two parameters, which will involve a computationally more demanding two-dimensional grid search.

#### 3.4.1. Selection of $r$

The number of bins,  $r$ , for our sieve methods plays a role similar to (the inverse of) the bandwidth in kernel estimation. More bins, i.e. smaller binwidth or bandwidth, reduce the bias of the estimator, but increase its variance. To derive a data-driven selection procedure for this parameter, we make use of the connection between our sieve approach and finite mixture modeling (McLachlan and Peel, 2000). After the discretization of  $X_c$ , our estimation problem can be regarded as a finite mixture proportion estimation problem with known components.

For any fixed  $r$ , the approximated density of  $Y$  given in (3.4) is a mixture density with known components  $f_Z(\cdot - m_i)$  where  $m_i \in \{a_1, \dots, a_\nu, x_1, \dots, x_r\}$ . Moreover, there exists a simple one-to-one correspondence between the parameters  $(\boldsymbol{\pi}, \boldsymbol{\theta})$  and the vector of mixture proportions. Hence, the estimation of  $(\boldsymbol{\pi}, \boldsymbol{\theta})$  is equivalent to estimating the mixture proportion vector. For example, when the error variable  $Z$  has a normal distribution with mean 0 and standard deviation  $\sigma$ , the sieve approximation of  $Y$  can be represented as a finite normal mixture with  $\nu + r$  components with means  $a_1, \dots, a_\nu, x_1, \dots, x_r$  and a common standard deviation  $\sigma$ .

This connection suggests that we can choose  $r$ , the number of discrete atoms used to approximate  $X_c$ , using similar techniques for selecting the number of mixture components. For more details, see McLachlan and Peel (2000). In the numerical studies reported in Sections 4 and 5, we consider the Akaike Information Criterion (AIC, Akaike (1974)). In particular, when deriving the SS estimator we define the selection criterion as

$$\text{AIC}(r) = -2 \ell(\hat{\boldsymbol{\pi}}^{SS}, \hat{\boldsymbol{\theta}}^{SS}) + 2(\nu + r - 1),$$

where  $\ell(\hat{\boldsymbol{\pi}}^{SS}, \hat{\boldsymbol{\theta}}^{SS})$  is the maximized log-likelihood in (3.5), evaluated at the corresponding standard sieve estimator  $(\hat{\boldsymbol{\pi}}^{SS}, \hat{\boldsymbol{\theta}}^{SS})$ ; and  $\nu + r - 1$  is the corresponding number of free parameters.

#### 3.4.2. Selection of $\lambda$

On the other hand, the penalty parameter  $\lambda$  plays the role of a smoothing parameter for our penalized sieve method. When  $\lambda = 0$ , the penalty term disappears,

and  $\widehat{\theta}^{PS}$  reduces to the standard sieve estimator  $\widehat{\theta}^{SS}$ . As discussed earlier, the resulting density estimator may not be smooth in this case. On the other hand, as  $\lambda$  increases to infinity, the penalty term dominates the log-likelihood. As a result, the estimator is flat in the limit when the sum of squared first order differences is used as the penalty function.

The selection of the penalty parameter is a challenging problem. One reason is that the density  $f_c$  is an unknown target. To ease the difficulty, we consider the following simulation-based approach:

- (a) start with a certain family of parametric models for the density;
- (b) use the method of moments to estimate the model parameters, and consider the estimated model as the “true” target;
- (c) simulate data from the estimated model, and add the known level of measurement error;
- (d) apply our penalized sieve estimator to the simulated noisy data to derive the density estimator for a grid of the tuning parameter  $\lambda$ ;
- (e) select  $\lambda$  to minimize the mean integrated squared error (MISE) of the density estimate, relative to the “true” target;
- (f) use the “optimal”  $\lambda$  found above to construct our penalized sieve estimator computed from the data at hand.

#### 4. Simulation studies

In this section, we perform simulation studies to investigate the performance of the proposed standard sieve (SS) and penalized sieve (PS) estimators, and compare them with the direct deconvolution (DC) estimator of van Es et al. (2008) and Lee et al. (2010). In Section 4.1 we consider distributions that are compactly supported, and in Section 4.2 we treat distributions supported on the positive real lines.

##### 4.1. Simulation Study I

This first simulation study investigates how the estimators perform for compactly-supported distributions. In particular, we consider the following simulation model for  $X$  in (3.1):

$$\text{Model Beta : } X \sim \begin{cases} \text{Beta}(1, 5) & \text{with probability } 0.25, \\ \text{Beta}(6, 6) & \text{with probability } 0.25, \\ 0 & \text{with probability } 0.5. \end{cases} \quad (4.1)$$

In Model (4.1), Beta(1, 5) stands for a beta distribution with shape parameters 1 and 5; the other beta distribution in (4.1) can be interpreted similarly. The continuous component  $X_c$  has compact support  $[0, 1]$  under the model, and its density is plotted in Figure 1. The measurement error  $Z$  is then generated from a normal distribution with mean 0 and standard deviation  $\sigma = 0.05$ . One hundred random samples of size 350 are simulated as in (3.1).

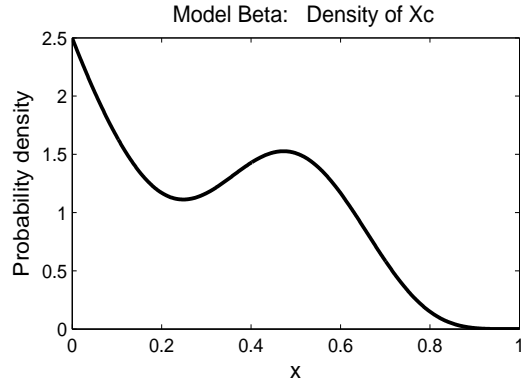


FIG 1. Simulation Study I. Density plot of the continuous component in Model (4.1).

TABLE 1  
Simulation Studies: Comparison of the pointmass and density estimation

			pointmass $\pi$			density $f_c$		
			Bias <sup>2</sup>	Var	MSE	Bias <sup>2</sup>	Var	MISE
Simulation I	Model (4.1)	DC	.0030	.0042	.0072	.1909	.0882	.2791
		SS	.0001	.0019	<b>.0020</b>	.0168	.1267	.1435
		PS	.0001	.0019	<b>.0020</b>	.0182	.0504	<b>.0686</b>
Simulation II	Model (4.2)	DC	.0020	.0033	.0233	.0305	.0024	.0329
		SS	.0031	.0012	<b>.0043</b>	.0039	.0016	<b>.0055</b>
		PS	.0031	.0012	<b>.0043</b>	.0044	.0012	<b>.0055</b>
	Model (4.3)	DC	.0001	.0016	.0018	.0645	.0049	.0694
		SS	.0000	.0005	<b>.0005</b>	.0053	.0103	<b>.0156</b>
		PS	.0000	.0005	<b>.0005</b>	.0053	.0103	<b>.0156</b>

We then apply the DC estimator, the SS estimator and the PS estimator to the simulated datasets to estimate both the pointmass  $\pi(= 0.5)$  and the continuous density  $f_c$ . The hybrid method described in Section 3.3 is used to derive the PS estimator: First we estimate the pointmass by the standard sieve method, and then we estimate  $\theta$  by the maximizer of the penalized log-likelihood.

In the top part of Table 1 we summarize the estimation results in terms of squared bias, variance and mean squared error (MSE). (For the continuous density, we calculate the integrated measures over the support.) The estimators that minimize MSE are highlighted in boldface. We use the procedure suggested by Lee et al. (2010) to select the tuning parameter for the DC estimator, and the procedures described in Section 3.4 to select the turning parameters  $r$  and  $\lambda$  for the sieve estimators. When selecting  $\lambda$ , we use a single beta distribution to generate the initial density target, although the true target is a mixture of two beta distributions. Our experience suggests that our parameter selection procedure is rather robust to the choice of the parametric target, as long as the parametric family covers a range of flexible shapes.

As one can see, our sieve estimators have much smaller bias and MSE than the DC estimators of van Es et al. (2008) and Lee et al. (2010), for both the pointmass and the continuous density. Based on plots not shown here, the DC estimator has serious bias near the boundaries, and puts positive probability mass beyond the compact support. When estimating the continuous density, due to the boundary effect, the DC estimator also has large bias around the second mode near 0.5. Compared with the SS estimator, the PS estimator further reduces the estimation variance and leads to a smaller MSE than SS, which shows the benefit of the penalization.

#### 4.2. Simulation Study II

The second simulation study investigates the effect of the pointmass and the performance of the estimators for distributions that are supported on the positive-half line. To make the simulation scientifically more relevant, this simulation scheme reflects the biological context of the virus lineage application described in Section 2. As discussed there, the mutation effect on virus fitness ( $S$ ) cannot be directly measured in experiments. Instead, evolutionary biologists measure it through the mutation effect on plaque size ( $X$ ), which is defined as the virus plaque size reduction between two consecutive measurement times. It has been established experimentally that  $X$  and  $S$  are related according to (2.1).

In our simulation study, we consider two models for  $S$  that are mixtures of a pointmass at 0 and an exponential distribution with different mixing probabilities. More specifically, the two models are

$$\text{Model Exp1 :} \quad S \sim \text{Exp}(0.12) \quad (4.2)$$

$$\text{Model Exp2 :} \quad S \sim \begin{cases} \text{Exp}(0.12) & \text{with probability 0.1,} \\ 0 & \text{with probability 0.9.} \end{cases} \quad (4.3)$$

In the above two models,  $\text{Exp}(0.12)$  denotes an exponential distribution with mean 0.12. We then define  $X$  according to (2.1). To generate the measurement error perturbed observations  $Y = X + Z$ , we simulate the measurement error  $Z$  from a normal distribution with mean 0 and standard deviation  $\sigma = 0.48$ . This value of  $\sigma$  is chosen based on the analysis of the real virus lineage data of Burch et al. (2007). We then simulate 100 random samples of size  $n = 350$  following the above simulation scheme. The exponential distribution is chosen because it is the most common model for virus mutation effect.

The continuous component of  $X$  has a density  $f_c$  that is supported only on the positive real line, and has a jump discontinuity at 0. Once the distribution of  $X$  is estimated, the distribution estimate of  $S$  can be easily obtained from (2.1) via a simple change of variables, if there is interest as is the case in the real application of Section 5.

The estimation results of the DC/SS/PS estimators are reported in Panel (b) of Table 1. The estimators that minimize the MSE are again highlighted in

boldface. When selecting  $\lambda$ , we assume an exponential distribution to generate the initial density target.

Our results suggest that, regardless of the presence of the pointmass ( $\pi = 0$  vs.  $\pi = 0.9$ ), our sieve estimators work better than the DC estimator, and have significantly smaller MSE for both  $\pi$  and  $f_c$ . In the case of the continuous density  $f_c$ , the bias of the sieve estimators is much smaller than DC, which can not appropriately incorporate the discontinuity at the origin. In plots not shown here, the boundary problem of the DC estimator is apparent near  $x = 0$ , where it is seriously biased, and it puts positive probability mass on the negative half-line. Moreover, the continuous density estimator takes negative values, which contradicts the basic nonnegative property of probability density functions.

The two sieve estimators perform similarly in the current setup, and penalization has little effect, which is different from the previous case described in Section 4.1. We think the reason is that the true density  $f_c$  in the current setup is simpler and smoother, and the standard sieve estimates are already sufficiently smooth. As the pointmass increases from  $\pi = 0$  to  $\pi = 0.9$ , the amount of information available for estimating the continuous density decreases, which explains why all three estimators perform worse when estimating  $f_c$ , but better for estimating the pointmass.

## 5. Application to the virus lineage data

### 5.1. Estimation results

We now apply our estimators to the virus lineage data described in Section 2, and compare the results with the commonly assumed parametric exponential fit, and the results of Burch et al. (2007) and Lee et al. (2010). For the pointmass  $\pi_1$  corresponding to the proportion of no (or silent) mutations, our standard sieve (SS) estimator in Section 3.2 gives  $\hat{\pi}_1^{SS} = 0.9204$  while the direct deconvolution estimate of Lee et al. (2010) gives 0.9363. Note that Burch et al. (2007) gives 0.9027 based on the virus genotype sequencing data.

As an illustration, Figure 2(a) plots the various estimators of the deleterious mutation effect density. One can clearly see the boundary problem of the direct deconvolution estimator (the grey dash-dotted curve), which puts positive density on the negative real line although the mutation effect distribution is supported only on the positive real line. On the other hand, our sieve estimators have positive support. In addition, the penalized sieve (PS) estimator described in Section 3.3 (the black solid curve) is smoother than the standard sieve estimator (the grey solid curve). The sum of squared first order differences is used as the penalty function in deriving the penalized sieve estimator.

Using the parameter selection procedure discussed in Section 3.4, we choose the number of bins needed for our sieve estimators to be  $r = 4$ , and the penalty parameter  $\lambda$  for the PS estimator to be 10 (Panel (b) of Figure 2). The mean parameter of the exponential fit (the black dash-dotted curve) is estimated using the method of moments. The tuning parameter  $M$  for the direct deconvolution estimator is selected using a procedure suggested by Lee et al. (2010).

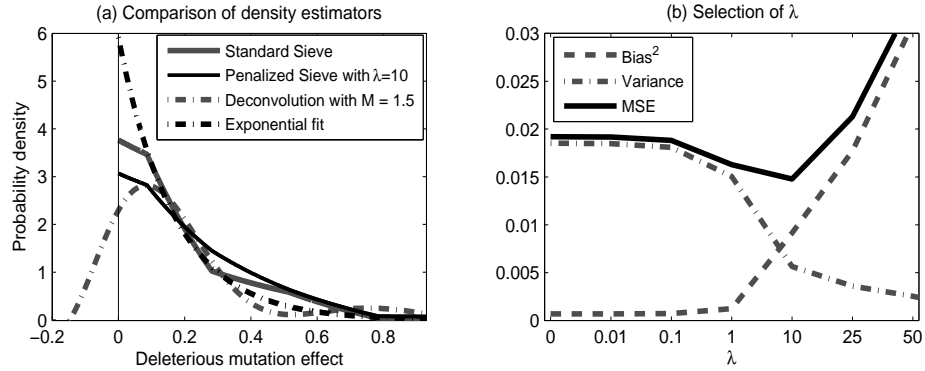


FIG 2. The motivating virus lineage application. Panel (a) shows various density estimators of the deleterious mutation effects. Panel (b) shows the selection for the penalty parameter  $\lambda$ .

## 5.2. Validation of the exponential assumption

One more statistical issue we want to address is the validation of the exponential assumption of the mutation effect distribution. Exponential distributions have been traditionally used to fit the mutation effects, but no serious work has been done to validate this parametric assumption. Figure 2 shows that the overall trends of our sieve estimators are similar to that of an exponential distribution. To formally check the validity of the exponential assumption, we propose using a graphical tool, the *density-envelope plot*. (The usage of the density-envelope plot is much more general than just for checking exponential distributions.)

Figure 3 provides the density-envelope for the sieve estimator with different penalty parameters. (Note that  $\lambda = 0$  corresponds to the standard sieve estimator.) To obtain the envelope, we first fit an exponential distribution to our data, then generate 100 random samples from the mixture of a pointmass at 0 with probability 0.9204, and the fitted exponential distribution with probability 0.0796. The simulated samples are of the same size as our original data. We then obtain the penalized sieve density estimator for each sample, using the corresponding penalty parameter  $\lambda$ . Finally these 100 estimators are overlaid to form the envelope, which represents the natural variation of the sieve estimator. In addition to the envelope, the black dash-dotted curve shows the exponential density that generates the simulation samples, the black solid curve is the penalized sieve estimator obtained from the data, and the dark grey solid curve is the average of the 100 light grey curves, respectively.

As can be seen from Figure 3, the sieve estimator obtained from the data lies nicely within the density envelope for every panel, and it almost overlaps with the average estimator. This implies that the difference between the sieve estimator and the fitted exponential density can be explained as being due to natural variation. Equivalently, we can state that an exponential distribution is a reasonable model for deleterious mutation effects.

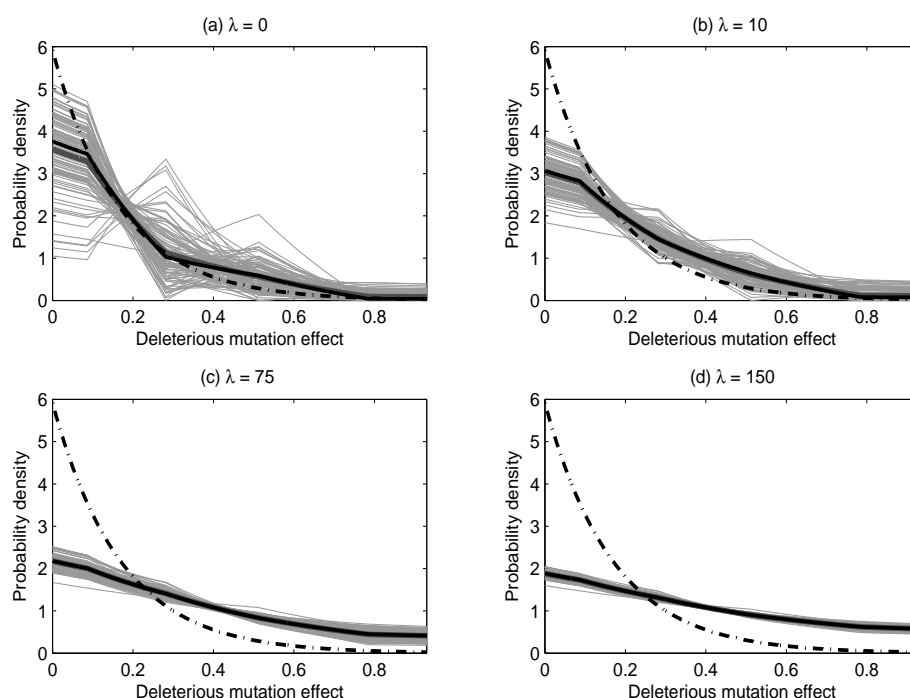


FIG 3. Validation of the exponential assumption: the density-envelope plots for various values of the roughness penalty parameter  $\lambda$ . The light grey curves form the envelope to show the natural estimation variation, the black dash-dotted curve shows the exponential density which generates the simulation samples, the black solid curve is the penalized sieve density estimator obtained from the data, and the dark grey solid curve is the average of the 100 grey curves.

Comparing the various panels, we can study the effect of the penalty parameter  $\lambda$  in the sieve estimation. As  $\lambda$  increases, the estimator gets smoother, so a larger  $\lambda$  is preferred when the target density is known to be smooth; at the same time, the estimation bias increases. In particular, the standard sieve estimator in Figure 3(a) is almost unbiased, but the estimation variance is large, as shown by the density envelope; on the other hand, Figure 3(d) shows the penalized sieve estimator with  $\lambda = 150$ , which is clearly over-smoothed: the bias of the estimator is quite large, but the estimation variance is very small. This way of studying the estimation result over multiple penalty parameters is similar to the scale-space approach for smoothing parameter selection proposed by Chaudhuri and Marron (2000).

## 6. Consistency of the proposed estimators

In this section we establish the consistency of our sieve estimators in Theorems 1 and 2. Technical details of the proofs are provided in the online supplement (Lee

et al., 2013). Our aim is to show that, with either approach, a consistent estimator of the distribution can be obtained under the minimal assumption that the binwidth  $h$  is of larger order than  $n^{-1}$ . That is, under some regularity conditions,  $\widehat{F}_X(x)$  converges to  $F_X(x)$  with probability 1, where

$$\widehat{F}_X(x) = \widehat{\pi}_1 I(a_1 \leq x) + \cdots + \widehat{\pi}_\nu I(a_\nu \leq x) + \widehat{\pi}_{\nu+1} \sum_{j=1}^r \widehat{\theta}_j I(x_j \leq x), \quad (6.1)$$

and  $F_X$  is the true distribution of  $X$  as defined in (3.2). If continuity of  $f_c$  is assumed, then it can be proved that the interpolated estimator in (3.6) is a consistent estimator of  $f_c$ .

In order to make our results simple to frame, we assume that the distribution of  $X_c$  has a bounded density  $f_c$  supported in a compact interval  $[a, b]$ ; the values of  $a$  and  $b$  need not be known, unless they are atoms of  $F_X$ . We construct the distribution estimator in a potentially larger interval  $[a^*, b^*]$ , where  $-\infty < a^* \leq a < b \leq b^* < \infty$ . Let  $c = \min\{a_1, a^*\}$  and let  $d = \max\{a_\nu, b^*\}$ . Define

$$g(y) = g(y|c, d) = \sup_{c \leq u_1, u_2 \leq d} \frac{f_Z(y - u_1)}{f_Z(y - u_2)}, \quad (6.2)$$

and write  $\mathcal{F}$  for the set of all distributions which have a mixture structure with atoms at  $a_1, \dots, a_\nu$  and a continuous distribution supported in  $[a, b]$ .

We also assume that for some  $x_0 \in [c, d]$ , and a constant  $B > 0$  not depending on  $s$ ,

$$\text{for all integers } s \geq 1, \quad E|\log g(Y)|^s + E|\log f_Z(Y - x_0)|^s \leq s! B^s, \quad (6.3)$$

$$\inf_{F \in \mathcal{F}} \text{Var} \left[ \log \left\{ \int f_Z(Y - x) dF(x) \right\} \right] > 0. \quad (6.4)$$

Property (6.3) can be verified by direct calculation for a large class of densities  $f_Z$ . For example, the normal and Laplace distributions are both particular cases of the set of Subbotin distributions  $\text{SD}_\gamma(\mu, \sigma^2)$ , with probability density

$$f(x|\gamma, \mu, \sigma) = \frac{C_\gamma}{\sigma} \exp \left( - \frac{|x - \mu|^\gamma}{\gamma \sigma^\gamma} \right),$$

where  $\gamma, \sigma > 0$ ,  $-\infty < \mu < \infty$  and  $C_\gamma^{-1} = 2\Gamma(1/\gamma)\gamma^{(1/\gamma)-1}$ . See Donoho and Jin (2004) for discussion of this class. Property (6.3) holds for these and many other distributions, for example distributions for which  $f_Z$  has regularly varying tails, as in the case of Student's  $t$ -distribution. Property (6.4) can be established by contradiction, noting that if it fails then there is an  $F \in \mathcal{F}$  for which the variance equals zero, and that this is not possible for the classes of error distributions just mentioned.

Next, taking  $f_X(x) = \sum_{l=1}^\nu \pi_l \delta_{a_l}(x) + \pi_{\nu+1} f_c(x)$  to be the generalized density of  $X$ , as defined in Cuevas and Walter (1992), we impose the following conditions:



- (R1)  $f_c$  is supported in the interval  $[a, b]$  and bounded above by a constant  $C > 0$ ;
- (R2) we construct the histogram approximation  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_r)^T$  to  $f_c$  in the interval  $[a^*, b^*]$ , where  $-\infty < a^* \leq a < b \leq b^* < \infty$ , in which case  $rh \leq b^* - a^*$ ;
- (R3) the error density  $f_Z$  satisfies the conditions (6.3) and (6.4);
- (R4) we restrict each  $\theta_j$  by insisting that  $\theta_j \leq C_1 h$ , where  $C_1 \geq C$ , the latter constant is as in (R1), and  $C_1$  is arbitrarily large;
- (R5)  $r = o(n)$  and  $r \rightarrow \infty$ ;
- (R6) the distribution of  $X$  is uniquely identifiable from its convolution with the distribution of  $Z$ .

Assumptions (R1)–(R3) merely formalize constraints discussed earlier; (R4) requires that our construction of the estimator reflects the boundedness assumption, but permits our prior impression of the bound to be arbitrarily large; (R5) is the key assumption, and, since it requires only  $r = o(n)$ , or equivalently that the binwidth  $h$  be of larger order than  $n^{-1}$ , it is the weakest possible condition of this type; and (R6) is a necessary condition, and holds if (for example) the characteristic function of  $Z$  vanishes only on a set of measure zero. When  $X_c$  has a smooth density  $f_c$ , the constraint (R4) is not necessary in practice, since estimators of the masses  $\theta_j$  are only very rarely much larger than the probabilities associated with the corresponding histogram blocks. Note that we do not assume continuity of  $f_c$ .

We first state the consistency results for the standard sieve estimator in the following Theorem 1.

**Theorem 1.** *Suppose that (R1)–(R6) hold. Then, with probability 1, the distribution estimator given in (6.1), where  $\hat{\boldsymbol{\pi}} = \hat{\boldsymbol{\pi}}^{SS}$  and  $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}^{SS}$  are obtained by the standard sieve estimation, converges to the true distribution of  $X$ .*

The next theorem establishes the parallel consistency result for the penalized sieve distribution estimator. For that, we need some extra conditions on the roughness penalty function and the penalty parameter.

**Theorem 2.** *In addition to (R1)–(R6), suppose that the roughness penalty  $P(\cdot)$  is asymptotically bounded. If the penalty parameter  $\lambda$  increases slower than  $n$ , i.e.  $\lambda = o(n)$  as  $n \rightarrow \infty$ , then  $\hat{F}_X(x)$  given by the penalized sieve estimation converges to the true distribution of  $X$  with probability 1.*

Proofs of the theorems can be found in the online supplement (Lee et al., 2013). Similar arguments can be given in cases where  $f_c$  is not assumed to be compactly supported, but they require a much longer discussion and proof, and additional regularity conditions. These include assumptions about the rate at which  $f_c$  decreases to zero in the left- and right-hand tails, the distances to the far left- and right-hand bins, the total number of bins  $r$ , and sometimes also the binwidth  $h$ , unless we deliberately link  $h$  to  $r$  through the other conditions. None of these conditions is particularly onerous or unrealistic, but together they make the theory less transparent and less elegant, and of course the technical

arguments become more elaborate. By way of contrast, of all these assumptions only the condition on  $r$  is needed in the compactly supported case; it is given in (R5).

### Acknowledgements

The authors want to thank the reviewers for their thoughtful comments.

### References

- AKAIKE, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control* **19**, 716–723. [MR0423716](#)
- BERTSEKAS, D. P. (2005). *Nonlinear Programming*. Athena Scientific, second edition.
- BROWN, L. AND GREENSHTEIN, E. (2009). Non parametric empirical Bayes and compound decision approaches to estimation of a high dimensional vector of normal means. *Ann. Statist* **37**, 1685–1704. [MR2533468](#)
- BURCH, C. L. AND CHAO, L. (2004). Epistasis and its relationship to canalization in the RNA virus  $\phi 6$ . *Genetics* **167**, 559–567.
- BURCH, C. L., GUYADER, S., SAMAROV, D., AND SHEN, H. (2007). Experimental estimate of the abundance and effects of nearly neutral mutations in the RNA virus  $\phi 6$ . *Genetics* **176**, 467–476.
- CARROLL, R. J., RUPPERT, D., STEFANSKI, L. A., AND CRAINICEANU, C. (2006). *Measurement error in nonlinear models: a modern perspective*. CRC Press, 2 edition. [MR2243417](#)
- CHAUDHURI, P. AND MARRON, J. S. (2000). Scale space view of curve estimation. *The Annals of Statistics* **28**, 408–428. [MR1790003](#)
- CORDY, C. B. AND THOMAS, D. R. (1997). Deconvolution of a distribution function. *Journal of the American Statistical Association* **92**, 1459–1465. [MR1615256](#)
- CUEVAS, A. AND WALTER, G. G. (1992). On estimation of generalized densities. *Communications in Statistics-Theory and Methods* **21**, 1807–1821. [MR1173496](#)
- DONOHO, D. L. AND JIN, J. (2004). Higher criticism for detecting sparse heterogeneous mixtures. *The Annals of Statistics* **32**, 962–994. [MR2065195](#)
- GREENSHTEIN, E. AND PARK, J. (2009). Application of Non Parametric Empirical Bayes Estimation to High Dimensional Classification. *Journal of Machine Learning Research* **10**, 1687–1704. [MR2534875](#)
- GRENANDER, U. (1981). *Abstract Inference*. Wiley & Sons. [MR0599175](#)
- GUGUSHVILI, S. AND VAN ES, B. AND SPREIJ, P. (2011). Deconvolution for an atomic distribution: rates of convergence. *Journal of Nonparametric Statistics*. **23**, 1003–1029. [MR2854252](#)
- HALL, P. AND QIU, P. (2005). Discrete-transform approach to deconvolution problems. *Biometrika* **92**, 135–148. [MR2158615](#)

- JOHNSTONE, I., KERKYACHARIAN, G., PICARD, D., AND RAIMONDO, M. (2004). Wavelet deconvolution in a periodic setting. *Journal of the Royal Statistical Society. Series B* **66**, 547–573. [MR2088290](#)
- LEE, M., HALL, P., SHEN, H., MARRON, J. S., TOLLE, J., AND BURCH, C. (2013). Deconvolution estimation of mixture distributions with boundaries: supplementary materials. Available online at <http://www.unc.edu/~haipeng/research/sieve-supp.pdf>.
- LEE, M., SHEN, H., BURCH, C., AND MARRON, J. S. (2010). Direct deconvolution density estimation of a mixture distribution motivated by mutation effects distribution. *Journal of Nonparametric Statistics* **22**, 1–22. [MR2598951](#)
- MCLACHLAN, G. J. AND PEEL, D. (2000). *Finite Mixture Models*. John Wiley & Sons. [MR1789474](#)
- MEISTER, A. (2007). Deconvolving compactly supported densities. *Mathematical Methods of Statistics* **16**, 63–76. [MR2319471](#)
- PENSKY, M. (2002). Density deconvolution based on wavelets with bounded supports. *Statistics & Probability Letters* **56**, 261–269. [MR1892987](#)
- RAYKAR, V. C. AND ZHAO, L. H. (2011). Empirical bayesian thresholding for sparse signals using mixture loss functions. *Statistica Sinica* **21**, 449–474. [MR2796871](#)
- RUPPERT, D., NETTLETON, D., AND HWANG, J. T. G. (2007). Exploring the information in  $p$ -values for the analysis and planning of multiple-test experiments. *Biometrics* **63**, 483–495. [MR2370807](#)
- STAUDENMAYER, J., RUPPERT, D., AND BUONACCORSI, J. P. (2008). Density estimation in the presence of heteroscedastic measurement error. *Journal of the American Statistical Association* **103**, 726–736. [MR2524005](#)
- VAN ES, B., GUGUSHVILI, S., AND SPREIJ, P. (2008). Deconvolution for an atomic distribution. *Electronic Journal of Statistics* **2**, 265–297. [MR2399196](#)
- WAGNER, C. AND STADTMÜLLER, U. (2008). Asymptotics for taylex and simex estimators in deconvolution of densities. *Journal of Nonparametric Statistics* **20**, 507–522. [MR2446440](#)
- WANG, X.-F., SUN, J., AND FAN, Z. (2009). Deconvolution density estimation with heteroscedastic error using SIMEX. Submitted to *Electronic Journal of Statistics*.
- ZHANG, S. AND KARUNAMUNI, R. J. (2009). Deconvolution boundary kernel method in nonparametric density estimation. *Journal of Statistical Planning and Inference* **139**, 2269–2283. [MR2507989](#)