Published in final edited form as:

J Nonparametr Stat. 2012 January 1; 24(1): 153–167. doi:10.1080/10485252.2011.647024.

The effects of error magnitude and bandwidth selection for deconvolution with unknown error distribution

Xiao-Feng Wang^{a,*} and Deping Ye^b

^aDepartment of Quantitative Health Sciences/Biostatistics Section, Cleveland Clinic Foundation, Cleveland, OH 44195, USA

^bDepartment of Mathematics and Statistics, Memorial University of Newfoundland, St. John's, NL A1C 5S7, Canada

Abstract

The error distribution is generally unknown in deconvolution problems with real applications. A separate independent experiment is thus often conducted to collect the additional noise data in those studies. In this paper, we study the nonparametric deconvolution estimation from a contaminated sample coupled with an additional noise sample. A ridge-based kernel deconvolution estimator is proposed and its asymptotic properties are investigated depending on the error magnitude. We then present a data-driven bandwidth selection algorithm with combining the bootstrap method and the idea of simulation extrapolation. The finite sample performance of the proposed methods and the effects of error magnitude are evaluated through simulation studies. A real data analysis for a gene Illumina BeadArray study is performed to illustrate the use of the proposed methods.

Keywords

deconvolution; unknown error; ridge-based approach; bandwidth selection; SIMEX

1. Introduction

Deconvolution problems have attracted considerable attention in the past two decades. The interest of studying these problems is mainly due to a great number of medical, chemical, astronomical and financial studies in which data are measured with error. The classical deconvolution problem can be described as follows. Suppose that the random variables $X_1, \dots, X_n$ are identically distributed as X . However, X_j 's cannot be observed directly. Instead, we observe

$$W_j = X_j + U_j, \quad 1 \leq j \leq n, \quad (1)$$

where the measurement errors U_j 's are identically distributed as U , and X_j 's and U_j 's are independent. The goal is to recover the probability density function f_X of X from W_j 's.

The popular approach to estimate the density is known as the *deconvolution kernel method*, where Fourier transform and the kernel smoothing technique are applied to obtain an estimator. Let $K(\cdot)$ be a symmetric probability kernel with a finite variance $\int x^2 K(x) < \infty$,

$\varphi_K(t) = \int e^{itx} K(x) dx$ be its Fourier transform with $\varphi_K(0) = 1$, and φ_U be the known characteristic function of the error variable U . The deconvolution kernel density estimator (Stefanski and Carroll 1990) is defined as

$$\hat{f}_X(x) = \frac{1}{2\pi} \int e^{-itx} \varphi_K(ht) \frac{\widehat{\varphi}_W(t)}{\varphi_U(t)} dt, \quad (2)$$

where $\widehat{\varphi}_W = n^{-1} \sum_{j=1}^n \exp(itW_j)$ is the empirical characteristic function of W , and $h = h(n) > 0$ is a smoothing parameter. Under the common assumption that φ_K is compactly supported and φ_U does not vanish on the real line, the estimator $\hat{f}_X$ in (2) is well defined and finite.

There is a large amount of literature on deconvolution problems. Early contributions include Carroll and Hall (1988), Stefanski and Carroll (1990), Zhang (1990), Fan (1991a,b), Fan (1992), Neumann (1997) among others. Recent related works include discussions of heteroscedastic measurement errors (Delaigle and Meister 2008; Wang et al. 2010), repeated measurements (Delaigle et al. 2008), estimation of distributions, moments and quantiles (Hall and Lahiri 2008), and density estimation with unknown error distribution (Comte and Lacour 2011; Johannes 2009). Deconvolution problems in image analysis have been extensively studied by Hall and Qiu (2007a,b) and Qiu (2008). Fourier-type deconvolution kernel methods have also been applied to stochastic volatility models in finance (Comte 2004; Van Es et al. 2003). A new R software package *decon* has been recently developed by Wang and Wang (2011), which contains a collection of functions to deal with deconvolution problems for practical use.

Classical deconvolution problems usually assume that the density function of the measurement error is perfectly known. This is often an unrealistic assumption in practice. In real studies, one could have an additional noise sample $U'_1, \dots, U'_m$ from a separate independent experiment. The data U'_k 's are indeed direct observations from the error distribution, i.e., U'_k 's are independent and identically distributed as U . If one does not impose any parametric distribution assumption to U , φ_U can be estimated by its empirical counterpart, that is,

$$\widehat{\varphi}_U(t) = \sum_{k=1}^m \exp(itU'_k) / m. \quad (3)$$

However, replacing $\varphi_U(t)$ in (2) by $\widehat{\varphi}_U(t)$ without any regularization is risky, since it is in the denominator of the formula. Neumann (1997) introduced the truncated kernel estimator

$$\tilde{f}_X(x) = \frac{1}{2\pi} \int e^{-itx} \varphi_K(th) \frac{\widehat{\varphi}_W(t)}{\widehat{\varphi}_U(t)} \chi(|\widehat{\varphi}_U(t)| > m^{-1/2}) dt,$$

where $\chi(A)$ is the indicator function of the set A . Thresholding in the Fourier domain by using this indicator function accounts for the uncertainty caused by estimating φ_U .

Johannes (2009) studied another spectral cut-off estimator that depended on two bandwidth-type parameters. The rate of convergence of the estimator was derived, while the selection of the parameters was not investigated. Recently, Comte and Lacour (2011) presented an interesting approach of data-driven density estimation in presence of error with unknown

distribution. There was no kernel smooth function in their estimator. Instead, a regularized parameter in the integral was introduced, which played a bandwidth-type role in the estimation. The estimator is defined as

$$\tilde{f}_x(x) = \frac{1}{2\pi} \int_{-\pi g}^{\pi g} e^{-itx} \frac{\widehat{\varphi}_w(t)}{\widehat{\varphi}_u(t)} \chi(|\widehat{\varphi}_u(t)| > m^{-1/2}) dt,$$

where g is the parameter to be selected. A penalization approach was proposed to choose g in their paper.

Hall and Qiu (2007a) investigated a ridge-based estimator for deconvolution in multivariate problems and applied it to image deblurring. A ridge function was applied to regularize the empirical function of the Fourier transform of $f(x)$, which prevents the denominator of the estimator from being too close to zero in a direct way. Motivated from their method, in this paper, we propose a ridge-based kernel deconvolution estimator for contaminated data with unknown error distribution. We study the effects of error magnitude on the proposed estimator, since it plays an important role in deconvolution problems. A data-driven bandwidth selection method is developed, which is not restricted to be applicable to the proposed estimator only but can be exploited in other deconvolution density estimators.

The rest of the paper is organized in the following way. In Section 2, we present the estimator and explore its asymptotic properties depending on the error magnitude. In Section 3, a data-driven bandwidth selection method is proposed by combining the bootstrap approach and the idea of simulation extrapolation (SIMEX). In Section 4, we investigate the numerical properties of the estimator. Simulation studies are conducted to examine both the effects of error magnitude and the performance of the bandwidth selection method. Our method is also compared to the penalized method by Comte and Lacour (2011). A real data analysis for a gene Illumina BeadArray study is performed to illustrate the use of the proposed methods. The proofs of Theorems are provided in Section 5.

2. The estimator and its asymptotic properties

Under the additive error model (1), we assume that two samples are observed: a contaminated sample $W_1, \dots, W_n$ and an additional noise sample $U'_1, \dots, U'_m$. By combining the kernel smoothing approach and the ridge-parameter technique, we consider the following regularized deconvolution density estimator for contaminated data with unknown error distribution,

$$\widehat{f}_{x, \rho_m}(x) = \frac{1}{2\pi} \int e^{-itx} \varphi_K(th) \widehat{\varphi}_w(t) \frac{\widehat{\varphi}_u(-t)}{\max\{|\widehat{\varphi}_u(t)|^2, \rho_m\}} dt, \quad (4)$$

where h is the smoothing parameter, $\rho_m = \rho(m) > 0$ is an additional ridge regularization parameter depending on m , φ_K is the characteristic function of the kernel K , $\widehat{\varphi}_w$ and $\widehat{\varphi}_u$ are the empirical characteristic functions of W and U , respectively.

The parameters h and ρ_m need to be appropriately selected to lead to a bias-variance compromise. Note that φ_U can be estimated by (3) at each point t with the rate $m^{-1/2}$. $\widehat{\varphi}_U$ is an appropriate estimator as $|\widehat{\varphi}_U| \gg m^{-1/2}$, while it becomes unstable as $|\widehat{\varphi}_U| \ll m^{-1/2}$. Therefore, it is reasonable to take the ridge parameter as,

$$\rho_m = m^{-1}, \quad (5)$$

in (4) in order to avoid the difficulty of selecting two parameters simultaneously. The ridge-based kernel method was also discussed in Meister (2009), where a different ridge parameter was used. Throughout this paper, the ridge function is defined by (5).

The rates of convergence for the conventional deconvolution kernel estimator (2) have been thoroughly studied by Fan (1991b), who characterized the error distributions into two classes: the ordinary smooth and the supersmooth. The difficulty of deconvolution depends heavily on the smoothness of the error distribution. For instance, when the error distribution is normal, the convergence rate is only at $O((\log n)^{-1/2})$. Such a slow convergence rate seems to indicate that the deconvolution kernel method is infeasible in practice. However, reasonable estimation results could be reached in both simulations and real data analysis. The phenomenon was explained by Delaigle (2008) through studying the asymptotic behavior of the estimator (2) with a “double asymptotical approach”, where the asymptotic properties were derived by considering that both the error variance goes to zero and the sample size goes to infinity.

Here we also derive the asymptotic properties of the ridge-based kernel estimator (4) depending on the error magnitude. Following the discussion by Delaigle (2008), we assume that the error U can be standardized by a variable that has unit variance. That is, $U = \sigma Z$, where σ is a scale parameter and the variance of Z is one. However, there is no parametric assumption imposed on Z . Thus, the model (1) is rewritten as

$$W_j = X_j + \sigma Z_j, \quad 1 \leq j \leq n, \quad (6)$$

where f_Z is unknown but $\text{Var}(Z) = 1$. The additional data $U'_1, \dots, U'_m$ are independent and identically distributed as $U = \sigma Z$.

The common smoothness condition imposed on the unknown density f_X is that f_X is in the set

$$\mathcal{F}_{\nu, \alpha, B} = \left\{ f : \left(\int |f^{(\nu)}(x) - f^{(\nu)}(x+\delta)|^2 dx \right)^{\frac{1}{2}} \leq B|\delta|^\alpha, |f| \leq B \right\},$$

where $\nu \in \mathbb{N}$ is the smoothness degree of f_X , $\alpha \in [0, 1)$ and $B > 0$ are known constants. As commented in Fan (1991a), the class $F_{\nu, \alpha, B}$ is larger than the commonly-used class that formulated in Stone (1982). It also contains the class formulated in Delaigle (2008), namely, f satisfies that both $\|f^{(\nu)}\|_\infty$ and $\int |f^{(\nu)}(x)|^2 dx$ are bounded from above by a constant. Indeed, the class $F_{\nu, \alpha, B}$ contains many commonly-used densities. For instance, normal, normal mixture, and Cauchy are in $F_{\nu, \alpha, B}$ for all $\nu \in \mathbb{N}$, $\alpha = 0$ and some $B > 0$; Gamma(k, θ) ($k \in \mathbb{N}$) is also in $F_{\nu, \alpha, B}$ for $\nu = k - 2$.

We always assume that $\varphi_Z(t) \neq 0$ for all $t \in \mathbb{R}$. The asymptotic results of f_{X, ρ_m} are derived based on two classes of error distributions as in the classical literature (Fan 1991b): *ordinary smooth* error Z of order β if

$$d_0|t|^{-\beta} \leq |\varphi_Z(t)| \leq d_1|t|^{-\beta}, \quad \text{for all } |t| > Q,$$

with d_0, d_1, β, Q some positive constants; and *supersmooth* error Z of order β if

$$d_0 |t|^{\beta_0} \exp(-|t|^\beta / \gamma) \leq |\varphi_Z(t)| \leq d_1 |t|^{\beta_1} \exp(-|t|^\beta / \gamma), \quad \text{for all } |t| > Q,$$

with $d_0, d_1, \gamma, \beta, Q$ some positive constants and β_0, β_1 some constants. Ordinary smooth distributions include, for example, gamma, symmetric gamma and Laplacian distributions, while supersmooth distributions include, such as, normal, mixture normal, Cauchy distributions, etc.

We assume that the kernel K satisfies:

- (A) K is bounded, continuous, and $\int |y|^{\nu+2\alpha} / K(y) dy < \infty$. Moreover, its characteristic function φ_K is symmetric and satisfies $\varphi_K(t) = 1 + O(t/|t|^{\nu+\alpha})$ as $t \rightarrow 0$.

This condition basically asserts that K is a kernel function with order $\nu + \alpha$. We use $c(n) \gg b(n)$ (resp. $c(n) \ll b(n)$) to represent $b(n) = o(c(n))$ (resp. $c(n) = o(b(n))$). Similarly, $h \sim a(n)$ means that there exist constants $a_1, a_2 > 0$ such that $a_1 a(n) \leq h \leq a_2 a(n)$.

For the ordinary smooth f_Z , the following theorem gives the rate of convergence of the estimator $\hat{f}_{X,\rho m}$. We need the following additional assumption on the kernel K :

- (B) $\int |t|^{2\beta} |\varphi_K(t)|^2 dt < \infty$.

Theorem 2.1

Assume that Z is ordinary smooth of order β but unknown. Under conditions (A), (B), one has,

- i. if $\sigma = O((\min\{n, m\})^{-\frac{1}{2\nu+2\alpha+1}})$ and $h \sim (\min\{n, m\})^{-\frac{1}{2\nu+2\alpha+1}}$, then

$$\sup_{f_X \in \mathcal{F}_{\nu, \alpha, B}} \mathbb{E} \int |\hat{f}_{X,\rho m}(x) - f_X(x)|^2 dx = O((\min\{n, m\})^{-\frac{2\nu+2\alpha}{2\nu+2\alpha+1}});$$

- ii. if $\sigma \gg (\min\{n, m\})^{-\frac{1}{2\nu+2\alpha+1}}$ and $h \sim \sigma^{\frac{2\beta}{2\nu+2\alpha+2\beta+1}} (\min\{n, m\})^{-\frac{1}{2\nu+2\alpha+2\beta+1}}$, then

$$\sup_{f_X \in \mathcal{F}_{\nu, \alpha, B}} \mathbb{E} \int |\hat{f}_{X,\rho m}(x) - f_X(x)|^2 dx = O(\sigma^{\frac{4\beta(\nu+\alpha)}{2\nu+2\alpha+2\beta+1}} (\min\{n, m\})^{-\frac{2(\nu+\alpha)}{2\nu+2\alpha+2\beta+1}}).$$

The rate of convergence of the estimator $\hat{f}_{X,\rho}$ for the supersmooth error is given by the following theorem. We need another assumption for the kernel K :

- (C) $\varphi_K(t) = 0$ for all $|t| > 1$, that is, $\varphi_K(t)$ has support on $[-1, 1]$. Moreover, $\int |\varphi_K(t)|^2 [t^{-2\beta_0} + |t|^{-2\beta_1}] dt < \infty$.

Theorem 2.2

Assume that Z is supersmooth of order β but unknown. Under conditions (A), (C), one has,

- i. if $\sigma = O((\min\{n, m\})^{-\frac{1}{2\nu+2\alpha+1}})$ and $h \sim (\min\{n, m\})^{-\frac{1}{2\nu+2\alpha+1}}$, then

$$\sup_{f_X \in \mathcal{F}_{\nu, \alpha, B}} \mathbb{E} \int |\widehat{f_{X, \rho m}}(x) - f_X(x)|^2 dx = O(\min\{n, m\}^{-\frac{2\nu+2\alpha}{2\nu+2\alpha+1}});$$

ii.

if $\sigma = (\min\{n, m\})^{-\frac{1}{2\nu+2\alpha+1}} a(n, m)$ and $h = \left(\frac{2}{\gamma D}\right)^{\frac{1}{\beta}} \sigma \{\log a(n, m)\}^{-\frac{1}{\beta}}$ with $1 \ll a(n, m) \ll (\min\{n, m\})^{\frac{1}{2(\nu+\alpha)+1}}$ and $D < 2\nu + 2\alpha + 1$, then

$$\sup_{f_X \in \mathcal{F}_{\nu, \alpha, B}} \mathbb{E} \int |\widehat{f_{X, \rho m}}(x) - f_X(x)|^2 dx = O\left(\sigma^{2\nu+2\alpha} \{\log a(n, m)\}^{-\frac{2(\nu+\alpha)}{\beta}}\right).$$

Conditions (B), and (C) give the restrictions on φ_K to ensure integrability of the estimator, for the ordinary smooth error and supersmooth error, respectively. In supersmooth error case, Condition (C) imposes a much stronger restriction on the tail behavior of φ_K ; see Fan (1992); Stefanski and Carroll (1990) for further details. A widely used kernel function K for practice in deconvolution literature is the one with $\varphi_K(t) = (1 - t^2)^3 \chi(-1 \leq t \leq 1)$, which satisfies the conditions (A), (B) and (C) (for $\beta_0, \beta_1 < 1/2$).

The quality of a sample does not only depend on its size but also relates to the magnitude of measurement error in deconvolution problems. Under the model (6), Theorems 2.1 and 2.2 provide a better interpretation of asymptotic behaviors for the estimator $\hat{f}_{X, \rho m}$ compared to from a classical asymptotic view of the problem. There are two different rates of convergence depending on the error level. In the case of ordinary smooth error, Part (i) of Theorem 2.1 shows that if $\sigma = O(h) = O((\min\{n, m\})^{-\frac{1}{2\nu+2\alpha+1}})$, the rate of convergence for $\hat{f}_{X, \rho m}$ can be as good as the conventional (error-free) kernel density estimation, while part (ii) of Theorem 2.1 gives a better rate of convergence for $\hat{f}_{X, \rho m}$ than that of the classical deconvolution estimator $\hat{f}_X$, if σ is small enough.

In the case of supersmooth error, Theorem 2.2 says that the convergence rate varies from the convergence rate of the error-free kernel density estimation to the very slow rate of the classical deconvolution depending on the error level. If $\sigma \rightarrow 0$, the rate of convergence for $\hat{f}_{X, \rho m}$ is better than that of the classical deconvolution. The results help us to explain why the estimator $\hat{f}_{X, \rho m}$ for supersmooth unknown error can work well in practice, even with moderate sample sizes. We will also see that it is not necessary to have a small error variance for this theory to be appropriate from the section of the simulation study.

Theorems 2.1 and 2.2 generalize the asymptotic results of Delaigle (2008), who derived the double asymptotic properties of the deconvolution estimator (2) with a known error density.

It is noticed from Theorem 2.1 and 2.2 that the sample size of the additional data U'_k 's contributes to the rate of convergence for $\hat{f}_{X, \rho m}$ only if $m \rightarrow n$.

3. A SIMEX-type bootstrap bandwidth selection method

The selection of bandwidth plays a crucial role in the implementation of practical estimation using deconvolution kernel techniques. This has been extensively studied in the classical deconvolution problems (Delaigle and Gijbels 2004; Hesse 1999). Although several kernel-type regularized deconvolution estimators have been proposed for density deconvolution with unknown error distribution (Johannes 2009; Neumann 1997), the bandwidth selection of them was not investigated. Here we propose a data-driven bandwidth selection algorithm for the ridge-based kernel estimator $\hat{f}_{X, \rho m}$ which combines the bootstrap bandwidth

selection and the SIMEX idea. The algorithm can be also applied to the bandwidth selection for other deconvolution kernel estimators.

A natural idea to implement a resampling-based bootstrap method for bandwidth selection in deconvolution with unknown error distribution can be described as follows. We first obtain an initial density estimate to generate bootstrap samples. If the measurement error is ignored, a *naïve* estimate of $f_X(x)$ is the ordinary kernel estimate of

$f_W: \widehat{f}_{X,naive}(x;g_1) = \widehat{f}_W(x;g_1) \triangleq (ng_1)^{-1} \sum_{j=1}^n K((x - W_j)/g_1)$. This provides a reasonable but overly smoothed estimate of $f_X(x)$. The bootstrap data is generated as $W_j^* = X_j^* + U_j^*$ ($j=1, \dots, n$), where X_j^* is generated from $\widehat{f}_{X,naive}$, and U_j^* is generated from $\widehat{f}_U(x;g_2) = (mg_2)^{-1} \sum_{k=1}^m K((x - U'_k)/g_2)$, the ordinary kernel density estimate from the additional data U'_k 's.

Then, we construct bootstrap deconvolution kernel estimates $\widehat{f}_{X,\phi_m}^{b*}(x;h)$, $b = 1, \dots, B$, where B is the number of bootstrap samples to be taken. The bootstrap estimate of the optimal bandwidth for the deconvolution kernel estimate is the value h that minimizes over h the bootstrap mean integrated squared error (MISE),

$$\text{MISE}^*(h) = \frac{1}{B} \sum_{b=1}^B \int \left[\widehat{f}_{X,\phi_m}^{b*}(x;h) - \widehat{f}_{X,naive}(x) \right]^2 dx.$$

This naive approach, however, tends to select an overly large bandwidth in practice. This is not surprising since we can not observe X_j 's and the naive bootstrap sample $X_1^*, \dots, X_n^*$ above is not a “true” bootstrap sample from $X_1, \dots, X_n$, but a bootstrap sample with error. The bootstrap sample $W_1^*, \dots, W_n^*$ above is a contaminated sample with a higher error level. Hence, we need an appropriate shrinking factor to adjust $\widehat{h}$ from the naive method. Our algorithm is motivated from Delaigle and Hall (2008) who used SIMEX for the cross-validation bandwidth selection in nonparametric regression with errors-in-variables. MISE^* in the above naive method is the bootstrap MISE of the contaminated data W_j 's rather than that of the “true” data X_j 's. Using the SIMEX concept, we may develop two versions of bootstrap MISEs for data with higher error levels, *i.e.*, MISE_1^* and MISE_2^* for the variables $W^{(1)} = W + U$ and $W^{(2)} = W^{(1)} + U$, respectively. Then we estimate the optimal bandwidths for data with the higher error levels, and back-extrapolate to yield the final bandwidth of interest. Our SIMEX-type bootstrap bandwidth selection algorithm is described as follows.

Algorithm 1

SIMEX-type bootstrap bandwidth selection for deconvolution with unknown error distribution.

1. Generate two bootstrap error-inflated samples $W_j^{*(1)} = W_j^* + U_j^{*(1)}$, and $W_j^{*(2)} = W_j^{*(1)} + U_j^{*(2)}$, $j = 1, \dots, n$, where W_j^* is generated from $\widehat{f}_W$, and $U_j^{*(1)}, U_j^{*(2)}$ are generated from $\widehat{f}_U$ independently.
2. Construct the deconvolution estimates $\widehat{f}_{W,\phi_m}^*(x, h_1)$ and $\widehat{f}_{W^{(1)},\phi_m}^*(x, h_2)$, with given bandwidths h_1, h_2 , based on the bootstrap samples $W_j^{*(1)}$ and $W_j^{*(2)}$.

3. Repeat the steps (1) and (2) B times to obtain $\widehat{f}_{W,\varphi_m}^{b*}, \widehat{f}_{W^{(1)},\varphi_m}^{b*}, b = 1, \dots, B$.
4. Estimate $\text{MISE}_1^*(h_1)$ for the variable W by

$$\text{MISE}_1^*(h_1) = \frac{1}{B} \sum_{b=1}^B \int \left[\widehat{f}_{W,\varphi_m}^{b*}(x; h_1) - \widehat{f}_W(x) \right]^2 dx,$$

and $\text{MISE}_2^*(h_2)$ for the variable $W^{(1)}$ by

$$\text{MISE}_2^*(h_2) = \frac{1}{B} \sum_{b=1}^B \int \left[\widehat{f}_{W^{(1)},\varphi_m}^{b*}(x; h_2) - \widehat{f}_{W^{(1)}}(x) \right]^2 dx,$$

where $\widehat{f}_W$ is the ordinary kernel density estimate of f_W , and $\widehat{f}_{W^{(1)}}$ is the kernel density estimate of $f_{W^{(1)}}$, given by

$$\widehat{f}_{W^{(1)}}(x) = (\widehat{f}_W * \widehat{f}_U)(x) = \int \widehat{f}_W(y) \widehat{f}_U(x - y) dy.$$

5. Obtain the estimated optimal bandwidths $\hat{h}_1$ and $\hat{h}_2$ by minimizing $\text{MISE}_1^*(h_1)$ and $\text{MISE}_2^*(h_2)$ over h_1 and h_2 .
6. Select the bandwidth $\hat{h}$ for $f_{X,\rho_m}(x, h)$ by using the linear back-extrapolation from the pair $(\log \hat{h}_1, \log \hat{h}_2)$. This suggests $\hat{h} = \widehat{h}_1^2 / \hat{h}_2$.

The key rationale of this algorithm remains the same as that of in SIMEX, which is to determine the effect of measurement error on the bandwidth experimentally via simulation. The effect of measurement error on a statistic of interest can be studied with a simulation experiment in which additional measurement error is added to the measured data and then the statistic recalculated. In the algorithm, $W_j^{*(2)}$ measures $W_j^{*(1)}$ in the same way that $W_j^{*(1)}$ measures W_j and that W_j measures X_j . So, we expect that the relationship between $\hat{h}_2$ and $\hat{h}_1$ is close to that between $\hat{h}_1$ and $\hat{h}$.

The second step of the classical SIMEX method is *extrapolation* (Stefanski and Cook 1995). Typically, a regression approach is used to fit an extrapolant function to the “pseudo” statistics. Then, extrapolation to the case of no measurement error yields the final estimate. Note that here an assumption of the extrapolant function needs to be imposed in order to fit the regression curve. One of a few simple functional forms, such as a quadratic function, is often used to make SIMEX an approximate method in practice. In our algorithm, we use the simple linear back-extrapolation function on the logarithm of the bandwidths. That is, the relation $\log \hat{h}_2 - \log \hat{h}_1 \approx \log \hat{h}_1 - \log \hat{h}$ is applied here, which leads to $\hat{h} = \widehat{h}_1^2 / \hat{h}_2$. In such a case, the shrinking factor to adjust the naive method is $\hat{h}_1 / \hat{h}_2$. The back-extrapolation function we use is the same as in Delaigle and Hall (2008). However, our algorithm is essentially different from theirs. We consider the bootstrap bandwidth selection for density deconvolution, while they studied the cross-validation bandwidth selection in nonparametric regression with errors-in-variables.

Certainly, like all other SIMEX approaches, using this linear back-extrapolation function is a fairly strong assumption. Extrapolating on the logarithm of the bandwidths is based on our experiences from numerical studies. Our numerical simulations indicate that the extrapolant function is a reasonable choice and the algorithm works quite well in practice for

deconvolution with unknown error distribution. One may consider a more computationally intensive SIMEX algorithm and/or another extrapolant function.

4. Numerical Studies

4.1. Simulation examples

We illustrate the performance of the proposed methods through simulations in this subsection. Throughout the studies, we take the commonly-used second order kernel for deconvolution problems,

$$\varphi_K(t) = (1 - t^2)^3 \chi(-1 \leq t \leq 1),$$

which corresponds to the kernel function

$$K(x) = \begin{cases} 48(1 - 15/x^2)\cos(x)/(\pi x^4) - 144(2 - 5/x^2)\sin(x)/\pi x^5, & \text{as } x \neq 0, \\ 16/(35\pi), & \text{as } x=0. \end{cases}$$

We use the integrated squared error (ISE), defined by $\text{ISE}(\hat{f}) = \int \{\hat{f}_X(x) - f_X(x)\}^2 dx$, to evaluate the performance of deconvolution estimators.

Our first simulation study was to examine the averaged ISEs (also denoted as “MISE”) as a function of the noise level σ . We considered three target densities of X corresponding to (1) $X \sim N(0, 1)$, (2) $X \sim 0.5N(-2, 1) + 0.5N(2, 1)$, (3) $X \sim \text{Gamma}(2, 1)$. The measurement errors were generated from $U \sim N(0, \sigma^2)$, where three error levels were considered ($\sigma = 0.4, 0.6$, or 0.8). In each case, we set the sample size $n = 100, 250$, or 1000 , and the size of the additional noise data $m = 100$, or 500 , respectively. In the study, we performed a grid search to find the bandwidth h that minimizes ISE under the assumption that the true density function is known. Although it is not a practical bandwidth selection method, it provides an optimistic view of the ISE performance of the proposed estimator. The grid size was set to 25. The numerical integration of ISE was evaluated from $\min_{1 \leq j \leq n} X_j$ to $\max_{1 \leq j \leq n} X_j$.

The MISEs averaged over 250 replications are reported in Table 1. For the purpose of comparison, the MISEs of the ordinary kernel density estimates from uncontaminated data are listed, which corresponds to $\sigma = 0$ in the table. The MISEs with $\sigma = 0.4$ are close to those with no measurement error, which supports our asymptotic conclusion that the deconvolution is not more difficult than ordinary density estimation when the error level is small. For all three target densities, MISEs increase when the error levels increase, and MISEs decrease when the sample sizes increase. The quality of the deconvolution estimates not only depends on the sample size but also is affected by the level of the measurement error variance. We have shown that there are two very different convergence rates depending on the error magnitude in Section 2. Although it is difficult to tell which rate applies in practice, the proposed estimator seems feasible for a moderate sample size even though the error level is not very small in this simulation study.

Next we illustrate our SIMEX-type bootstrap method for bandwidth selection through one of typical simulated examples. In this example, X_j 's were generated from $0.5N(-2, 1) + 0.5N(2, 1)$. The errors U_j 's and U_k 's were generated from $N(0, 0.5^2)$. The sample size $n = 1000$, the size of the additional noise data $m = 500$, and the bootstrap size $B = 200$. The analysis results are displayed in Figure 1. The left panel of the figure shows the three

estimated ISEs as the functions of the bandwidth parameter h : the solid line denotes the ISE calculated using the true density function from the original contaminated data; the dashed line denotes the $MISE_1^*$ from the first level error-inflated bootstrap samples; and the dotted line denotes the $MISE_2^*$ from the second level error-inflated bootstrap samples. The corresponding estimated optimal bandwidths are $\hat{h}_{opt} = 0.188$, $\hat{h}_1 = 0.218$, and $\hat{h}_2 = 0.239$, respectively. Note that the bandwidth increases as the level of measurement error increases. Using the proposed linear back-extrapolation function on the logarithm of bandwidths, our final selected bandwidth $\hat{h} = \hat{h}_1^2 / \hat{h}_2 = 0.199$. It is very close to $\hat{h}_{opt}$, the bandwidth assuming that the true density is known. The right panel of Figure 1 shows the density estimates: the ordinary kernel estimate from the uncontaminated sample X_j 's (solid line), the ridge-based kernel deconvolution estimate with the bandwidth $\hat{h}$ from the contaminated sample W_j 's (dashed line), and the ordinary kernel estimate from W_j 's (dotted line). The true density is denoted by the solid line. The deconvolution estimate with the selected bandwidth works quite well to recover the true density, which is very close to the kernel estimate from the uncontaminated data.

We finally compare the performance of our estimator with Comte and Lacour (2011)'s penalized estimator. The true model we considered was $X \sim 0.5N(-2, 1) + 0.5N(2, 1)$. The measurement error U was from $N(0, \sigma^2)$. Four different cases of sample sizes were studied, and two levels of measurement error variance were considered: $\sigma = 0.3$ or 0.6 . The bootstrap size $B = 200$ for the bandwidth selection. Table 2 displays the MISEs and their associated standard errors from 100 replications. The results show that the performance of our estimator is comparable to that of the estimator of Comte and Lacour (2011). As the measurement error variance is small, the method of Comte and Lacour (2011) is slightly better than ours. As the error variance becomes large and sample size increases, our method appears to become a little better. In general, our kernel estimator seems to be a good alternative to the penalized estimator under the simulation setting.

4.2. A case study with Illumina BeadArray data

We illustrate the deconvolution method using the Illumina BeadArray data from a leukemia study (Xie et al. 2009). The objective of the leukemia study was to investigate the pathogenesis of leukemia. Irradiated mice who subsequently developed acute myeloid leukemia (AML) were involved to study the leukemogenic process. Illumina Mouse-6 V1 BeadChip mouse whole-genome expression arrays were used to obtain the gene expression profiles of AML samples. More details of the biology experiment can be found in Xie et al. (2009).

Illumina BeadArray has become more and more popular in recent years. The BeadArray technology from Illumina Inc. makes its preprocessing and quality control different from other microarray technologies. BeadArrays are arrays of randomly positioned silica beads of 3 micron diameter. In this leukemia study, the pre-processing of the raw BeadArray image data was routinely carried out using the *BeadStudio* software developed by Illumina inc. For each bead, the foreground intensity was calculated as a weighted average of signals. The local background, an average of the five dimmest pixels (unsharpened intensities) within the 17×17 pixel area around each bead centre, was then subtracted to produce *bead summary data* from the software (Dunning et al. 2008). Other image processing options (Qiu and Sun 2007; Wang and Ye 2010) may be also feasible here.

One of distinctive features of Illumina technology is that more than one thousand control bead types in addition to gene sequences are allocated in each array. These control beads do not correspond to any expressed sequences in the genome, which are not expected to

hybridize to any genes in the RNA samples. This unique design enable us to obtain additional noise data for the non-specific binding effect on the array in an experiment.

The first step of the microarray data analysis is sometimes named as *background correction*, which is a process of correcting the measurement error effects of the observed gene intensities on an array using information only on that array. It should note that “background” here means the noise from, such as, the non-specific binding on the chip rather than the local image background. The popular method for background correction is the “normexp” method based on an additive measurement error model (Irizarry et al. 2003; Silver et al. 2009; Wang et al. 2011). The model imposes the parametric assumptions that the observed intensity (W) is the sum of the true signal (X , assumed exponentially distributed) and the noise (U , assumed normally distributed).

In the leukemia study, 46120 genes and 1655 negative controls were randomly allocated on each array. Here we only report the analysis of the data from Array # 4 in the leukemia study, as a demonstrating example, to assess whether the normal-exponential assumption is valid. Denote the observed (local image-background subtracted) gene intensity from the array as $W_j, j = 1, \dots, 46120$ and the observed (local image-background subtracted) negative control intensity as $U'_k, k = 1, \dots, 1655$. Consider the additive error model, $W_j = X_j + U_j$, where X_j is the true signal and U_j is the noise. We recovered the density of the true signal, f_X , without imposing any distribution assumptions from the data W_j 's coupled with U'_k 's. Figure 2 displays the analysis results. The left panel shows the ordinary kernel density estimates of contaminated gene intensities (solid line) and negative control intensities (dashed line). Both data exhibit long-tailed and right-skewed distributions. A Kolmogorov-Smirnov test was performed to test the normality of the negative control data ($p < 0.0001$). Hence, there is no evidence to support that error is from a normal distribution. The right panel shows the ridge-based kernel deconvolution estimate $\hat{f}_{X, pm}$. The SIMEX-type bootstrap method was used for the bandwidth selection, which suggested that the smoothing parameter was 8.14. Note that the location of the recovered density function is shifted to the left about 100 units. This is due to the non-zero mean noise of the non-specific binding. The recovered density shows a “gamma-like” shape. A higher peak is obtained in the recovered function (0.0121 as $X = 10.07$), comparing with that in the naive estimate from contaminated gene intensities (0.0115 as $X = 112.72$). The analysis provides some evidences that the normal-exponential model may impose too strong parametric assumptions and may not be appropriate to the Illumina BeadArray background correction.

5. Proofs

Parseval's identity and Fubini theorem imply that

$$\mathbb{E} \int |\widehat{f}_{X, pm} - \widehat{f}_X|^2 dx = \frac{1}{2\pi} \int |\varphi_K(th)|^2 (\mathbb{E}_W |\widehat{\varphi}_W(t)|^2) \mathbb{E}_U \left| \frac{\widehat{\varphi}_U(-t)}{\max\{|\widehat{\varphi}_U(t)|^2, m^{-1}\}} - \frac{1}{\varphi_U(t)} \right|^2 dt.$$

Recall that $\mathbb{E}_W |\widehat{\varphi}_W(t)|^2 = \mathbb{E} |n^{-1} \sum_{j=1}^n \exp(itW_j)|^2 = n^{-1} (1 - |\varphi_W(t)|^2) + |\varphi_W(t)|^2$. From Proposition 31.8 in Port (1994), one has

$$\mathbb{E}_U |\widehat{\varphi}_U(t)^{-1} - \varphi_U(t)^{-1}|^2 = (1 - |\varphi_U(t)|^2) m^{-1} |\varphi_U(t)|^{-4} + o(m^{-1}). \quad (7)$$

We now show that

$$\mathbb{E}_U \left| \frac{\widehat{\varphi}_U(-t)}{\max\{|\widehat{\varphi}_U(t)|^2, m^{-1}\}} - \frac{1}{\varphi_U(t)} \right|^2 = I_1 + I_2 = O(m^{-1} |\varphi_U(t)|^{-4}),$$

where

$$I_1 = \mathbb{E}_U \left(|\widehat{\varphi}_U(t)^{-1} - \varphi_U(t)^{-1}|^2 \chi(|\widehat{\varphi}_U(t)| \geq m^{-1/2}) \right),$$

$$I_2 = \mathbb{E}_U \left(|m\widehat{\varphi}_U(-t) - \varphi_U(t)^{-1}|^2 \chi(|\widehat{\varphi}_U(t)| < m^{-1/2}) \right).$$

To this end, by formula (7), one gets $I_1 = O(m^{-1} |\varphi_U(t)|^{-4})$. Following the same line as the proof of Lemma 2.1 in Neumann (1997), one gets $I_2 = O(\min\{m^{-1} |\varphi_U(t)|^{-4}, |\varphi_U(t)|^{-2}\})$

Therefore, as $\varphi_W(t) = \varphi_U(t)\varphi_X(t)$, $|\varphi_X(t)| \leq 1$, and $\varphi_U(t) = \varphi_Z(\sigma t)$, one has

$$\begin{aligned} \mathbb{E} \int |\widehat{f}_{X\varphi_m} - \widehat{f}_X|^2 dx &= \left(\frac{1}{2\pi} \int |\varphi_K(th)|^2 |\varphi_W(t)|^2 O(m^{-1} |\varphi_U(t)|^{-4}) dt \right) (1 + o(n^{-1})) \\ &= O\left(\frac{1}{2\pi} \int |\varphi_K(th)|^2 m^{-1} |\varphi_Z(\sigma t)|^{-2} dt \right). \end{aligned}$$

Proof of Theorem 2.1

Let Z have ordinary smooth error f_Z . Together with assumptions (A) and (B), one has

$$\mathbb{E} \int |\widehat{f}_{X\varphi_m} - \widehat{f}_X|^2 dx = O\left(\frac{1}{mh} + \frac{\sigma^{2\beta}}{mh^{2\beta+1}} \right).$$

In fact, by letting $\tau = th$, one has

$$\begin{aligned} \frac{1}{2\pi} \int \frac{|\varphi_K(th)|^2}{m|\varphi_Z(\sigma t)|^2} dt &= \frac{1}{2\pi hm} \int \frac{|\varphi_K(\tau)|^2}{|\varphi_Z(\sigma\tau/h)|^2} d\tau \\ &= \frac{1}{2\pi hm} \int_{|\sigma\tau/h| < Q} \frac{|\varphi_K(\tau)|^2}{|\varphi_Z(\sigma\tau/h)|^2} d\tau + \frac{1}{2\pi hm} \int_{|\sigma\tau/h| \geq Q} \frac{|\varphi_K(\tau)|^2}{|\varphi_Z(\sigma\tau/h)|^2} d\tau \\ &= O\left(\frac{1}{mh} \right) + \frac{\sigma^{2\beta}}{2\pi d_0^2 mh^{2\beta+1}} \int_{|\sigma\tau/h| \geq Q} |\varphi_K(\tau)|^2 |\tau|^{2\beta} d\tau = O\left(\frac{1}{mh} + \frac{\sigma^{2\beta}}{mh^{2\beta+1}} \right). \end{aligned} \quad (8)$$

Following the same rational as the proof of Theorem 2.1 in Delaigle (2008), one can check that,

$$\sup_{f_X \in \mathcal{F}_{\nu, \alpha, B}} \mathbb{E} \int |\widehat{f}_X - f_X|^2 dx = O(h^{2\nu+2\alpha}) + O\left(\frac{\sigma^{2\beta}}{nh^{2\beta+1}} \right) + O\left(\frac{1}{nh} \right).$$

Cauchy-Schwarz inequality implies that

$$\begin{aligned} \text{MISE}(\widehat{f}_{X\varphi_m}) &= \mathbb{E} \int |\widehat{f}_{X\varphi_m} - f_X|^2 dx \leq 2\mathbb{E} \int |\widehat{f}_{X\varphi_m} - \widehat{f}_X|^2 dx + 2\mathbb{E} \int |\widehat{f}_X - f_X|^2 dx \\ &= O\left(\frac{1}{\min\{n, m\}h}\right) + O\left(\frac{\sigma^{2\beta}}{\min\{n, m\}h^{2\beta+1}}\right) + O(h^{2\nu+2\alpha}). \end{aligned}$$

i. Let $\sigma = O(h)$. We have

$$\sup_{f_X \in \mathcal{F}_{\nu, \alpha, B}} \text{MISE}(\widehat{f}_{X\varphi_m}) = O(h^{2\nu+2\alpha}) + O\left(\frac{1}{\min\{n, m\}h}\right).$$

Clearly, by choosing $h \sim (\min\{n, m\})^{-\frac{1}{2\nu+2\alpha+1}}$, one gets the desired result.

ii. Let $\sigma \gg h$. As long as $\frac{\sigma^{2\beta}}{\min\{n, m\}h^{2\beta}} \rightarrow 0$ as $\min\{n, m\} \rightarrow \infty$ and $h \rightarrow 0$, one has

$$\sup_{f_X \in \mathcal{F}_{\nu, \alpha, B}} \text{MISE}(\widehat{f}_{X\varphi_m}) = O(h^{2\nu+2\alpha}) + O\left(\frac{\sigma^{2\beta}}{\min\{n, m\}h^{2\beta+1}}\right).$$

As in (i), one can then choose $h \sim \sigma^{\frac{2\beta}{2\nu+2\alpha+2\beta+1}} (\min\{n, m\})^{-\frac{1}{2\nu+2\alpha+2\beta+1}}$, which of course verifies $\frac{\sigma^{2\beta}}{\min\{n, m\}h^{2\beta}} \rightarrow 0$. Combining $\sigma \gg h$ with $h \sim \sigma^{\frac{2\beta}{2\nu+2\alpha+2\beta+1}} (\min\{n, m\})^{-\frac{1}{2\nu+2\alpha+2\beta+1}}$, one can get $\sigma \gg (\min\{n, m\})^{-\frac{1}{2\nu+2\alpha+1}}$. The desired conclusion of (ii) follows by the choice of h .

Proof of Theorem 2.2

Let Z have supersmooth error f_Z . Under the assumptions (A), (C), similar to the estimation of in (8), we have

$$\sup_{f_X \in \mathcal{F}_{\nu, \alpha, B}} \text{MISE}(\widehat{f}_{X\varphi_m}) = O(h^{2\nu+2\alpha}) + O\left(\frac{1}{\min\{n, m\}h}\right) + O\left(\frac{1}{\min\{n, m\}h} \left(\frac{\sigma}{h}\right)^l \exp\left(\frac{2\sigma^\beta}{\gamma h^\beta}\right)\right),$$

where $l = 0$ if $\beta_0 = 0$ and $l = -2\beta_0$ if $\beta_0 < 0$.

i. if $\sigma = O(h)$, the proof is identical to that of (i) in Theorem 2.1.

ii. Let $\sigma \gg h$. As long as $\frac{1}{\min\{n, m\}} \left(\frac{\sigma}{h}\right)^l \exp\left(\frac{2\sigma^\beta}{\gamma h^\beta}\right) \rightarrow 0$ as $\min\{n, m\} \rightarrow \infty$ and $h \rightarrow 0$, one has

$$\sup_{f_X \in \mathcal{F}_{\nu, \alpha, B}} \text{MISE}(\widehat{f}_{X\varphi_m}) = O(h^{2\nu+2\alpha}) + O\left(\frac{1}{\min\{n, m\}h} \left(\frac{\sigma}{h}\right)^l \exp\left(\frac{2\sigma^\beta}{\gamma h^\beta}\right)\right). \quad (9)$$

Now let us choose $\sigma = (\min\{n, m\})^{-\frac{1}{2\nu+2\alpha+1}} a(n, m)$ and $h = \left(\frac{2}{\gamma D}\right)^{\frac{1}{\beta}} \sigma \{\log a(n, m)\}^{-\frac{1}{\beta}}$, with $1 \ll a(n, m) \ll (\min\{n, m\})^{\frac{1}{2\nu+2\alpha+1}}$ and $D < 2\nu+2\alpha+1$. Clearly our choice of h verifies the condition $\frac{1}{\min\{n, m\}} \left(\frac{\sigma}{h}\right)^l \exp\left(\frac{2\sigma^\beta}{\gamma h^\beta}\right) \rightarrow 0$ and the assumption $\sigma \gg h$. The desired conclusion of (ii) is then an immediate consequence by putting the choice of h into (9).

Acknowledgments

We are grateful to the Associate Editor and two reviewers for their valuable suggestions which substantially improved the paper. The research of XFW is supported in part by NIH UL1 RR024989. The research of DY has been initiated and completed by the support of NSF-FRG DMS grants 0652571 and 0652684 (from University of Missouri, Columbia).

References

- Carroll RJ, Hall P. Optimal Rates of Convergence for Deconvolving a Density. *Journal of the American Statistical Associations*. 1988; 83:1184–1186.
- Comte F. Kernel Deconvolution of Stochastic Volatility Models. *Journal of Time Series Analysis*. 2004; 25:563–582.
- Comte F, Lacour C. Data-driven Density Estimation in the Presence of Additive Noise with Unknown Distribution. *Journal of the Royal Statistical Society Series B-Methodological*. 2011; 73:601–627.
- Delaigle A, Gijbels I. Practical Bandwidth Selection in Deconvolution Kernel Density Estimation. *Computational Statistics and Data Analysis*. 2004; 45:249–267.
- Delaigle A, Meister A. Density Estimation with Heteroscedastic Error. *Bernoulli*. 2008; 14:562–579.
- Delaigle A. An Alternative View of the Deconvolution Problem. *Statistica Sinica*. 2008; 18:1025–1045.
- Delaigle A, Hall P. Using SIMEX for Smoothing-parameter Choice in Errors-in-variables Problems. *Journal of the American Statistical Association*. 2008; 103:280–287.
- Delaigle A, Hall P, Meister A. On Deconvolution with Repeated Measurements. *Annals of Statistics*. 2008; 36:665–685.
- Dunning MJ, Barbosa-Morais NL, Lynch AG, Tavaré S, Ritchie ME. Statistical Issues in the Analysis of Illumina Data. *BMC Bioinformatics*. 2008; 9:85. [PubMed: 18254947]
- Fan J. Global Behavior of Deconvolution Kernel Estimates. *Statistica Sinica*. 1991a; 1:541–551.
- Fan J. On the Optimal Rates of Convergence for Nonparametric Deconvolution Problems. *Annals of Statistics*. 1991b; 19:1257–1272.
- Fan J. Deconvolution with Supersmooth Distributions. *The Canadian Journal of Statistics*. 1992; 20:155–169.
- Hall P, Lahiri SN. Estimation of Distributions, Moments and Quantiles in Deconvolution Problems. *Annals of Statistics*. 2008; 36:2110–2134.
- Hall P, Qiu P. Nonparametric Estimation of a Point-Spread Function in multivariate Problems. *Annals of Statistics*. 2007a; 35:1512–1534.
- Hall P, Qiu P. Blind Deconvolution and Deblurring in Image Analysis. *Statistica Sinica*. 2007b; 17:1483–1509.
- Hesse C. Data-Driven Deconvolution. *Journal of Nonparametric Statistics*. 1999; 10:343–373.
- Irizarry R, Hobbs B, Collin F, Beazer-Barclay Y, Antonellis K, Scherf U, Speed T. Exploration, Normalization, and Summaries of High Density Oligonucleotide Array Probe Level Data. *Biostatistics*. 2003; 4:249–264. [PubMed: 12925520]
- Johannes J. Deconvolution with Unknown Error Distribution. *Annals of Statistics*. 2009; 37:2301–2323.
- Meister, A. *Lecture Notes in Statistics*. Springer; New York: 2009. Deconvolution Problems in Nonparametric Statistics.
- Neumann MH. On the Effect of Estimating the Error Density in Nonparametric Deconvolution. *Journal of Nonparametric Statistics*. 1997; 7:307–330.
- Port, SC. *Lecture Notes in Statistics*. John Wiley; New York: 1994. Theoretical Probability for Applications.
- Qiu P. A Nonparametric Procedure for Blind Image Deblurring. *Computational Statistics & Data Analysis*. 2008; 52:4828–4841.
- Qiu P, Sun J. Local Smoothing Image Segmentation for Spotted Microarray Images. *Journal of the American Statistical Association*. 2007; 102:1129–1144.

- Silver JD, Ritchie ME, Smyth GK. Microarray Background Correction: Maximum Likelihood Estimation for the Normal-exponential Convolution. *Biostatistics*. 2009; 10:352–363. [PubMed: 19068485]
- Stefanski LA, Carroll RJ. Deconvoluting Kernel Density Estimators. *Statistics*. 1990; 21:169–184.
- Stefanski LA, Cook JR. Simulation Extrapolation: The Measurement Error Jackknife. *Journal of the American Statistical Association*. 1995; 90:1247–1256.
- Stone CJ. Optimal Global Rates of Convergence for Nonparametric Regression. *Annals of Statistics*. 1982; 10:1040–1053.
- Van Es B, Spreij P, Van Zanten H. Nonparametric Volatility Density Estimation. *Bernoulli*. 2003; 9:451–465.
- Wang B, Wang XF, Xi Y. Normalizing Bead-based MicroRNA Expression Data: A Measurement Error Model-Based Approach. *Bioinformatics*. 2011; 27:1506–1512. [PubMed: 21498399]
- Wang XF, Fan Z, Wang B. Estimating Smooth Distribution Function in the Presence of Heteroscedastic Measurement Errors. *Computational Statistics and Data Analysis*. 2010; 54:25–36. [PubMed: 20160998]
- Wang XF, Wang B. Deconvolution Estimation in Measurement Error Models: The R Package decon. *Journal of Statistical Software*. 2011; 39:1–24. [PubMed: 21572908]
- Wang XF, Ye D. On Nonparametric Comparison of Images and Regression Surfaces. *Journal of Statistical Planning and Inference*. 2010; 140:2875–2884. [PubMed: 20543891]
- Xie Y, Wang X, Story M. Statistical Methods of Background Correction for Illumina BeadArray Data. *Bioinformatics*. 2009; 25:751–757. [PubMed: 19193732]
- Zhang CH. Fourier Methods for Estimating Mixing Densities and Distributions. *Annals of Statistics*. 1990; 18:806–831.

Figure 1.

A simulated example of density deconvolution with unknown error distribution using the SIMEX-type bootstrap bandwidth selection method. The left panel shows the different estimated ISEs as functions of h : ISE with the true density (solid line), $MISE_1^*$ (dashed line), and $MISE_2^*$ (dotted line). The right panel shows the different density estimates: the ordinary kernel estimate from X_j 's (solid line), the deconvolution estimate from W_j 's (dashed line), and the ordinary kernel estimate from W_j 's (dash-dotted line). The dotted line denotes the true density function.

Figure 2.

A case study of Illumina BeadArray data: the left panel displays the ordinary kernel density estimates of contaminated gene intensities (solid line) and negative controls (dashed line). The right panel displays the ridge-based kernel deconvolution estimate with the SIMEX-type bootstrap bandwidth selection method.

Table 1

MISE (expressed in 0.01 units) for the ridge-based kernel deconvolution estimate as a function of the error level σ . The “optimal” bandwidths of the estimates are estimated by assuming that f_X is known. The column of $\sigma = 0$ corresponds to the MISEs of the kernel estimates from the uncontaminated data.

Density	n	MISE						
		$m=100$			$m=500$			
		$\sigma=0$	$\sigma=0.4$	$\sigma=0.6$	$\sigma=0.8$	$\sigma=0.4$	$\sigma=0.6$	$\sigma=0.8$
Normal	100	0.508	0.816	1.145	1.632	0.703	0.993	1.334
	250	0.269	0.532	0.831	1.259	0.442	0.692	1.115
	1000	0.097	0.296	0.501	0.857	0.232	0.405	0.682
Normal mixture	100	0.409	0.657	0.851	1.092	0.565	0.747	1.026
	250	0.224	0.413	0.617	0.862	0.349	0.506	0.741
	1000	0.103	0.208	0.363	0.562	0.168	0.279	0.466
Gamma	100	0.609	0.961	1.272	1.656	0.835	1.059	1.454
	250	0.311	0.618	0.899	1.269	0.528	0.768	1.096
	1000	0.131	0.394	0.639	0.964	0.315	0.526	0.797

Table 2

Comparison of our estimator with the SIMEX-type bootstrap bandwidth selection and the penalized estimator of Comte and Lacour (2011). Table entries without parentheses are the MISEs and entries with parentheses are their associated standard errors (both are expressed in 0.01 units).

Case	MISE and its standard error					
	Comte & Lacour's method			The proposed method		
	$\sigma=0.3$	$\sigma=0.6$	$\sigma=0.3$	$\sigma=0.6$	$\sigma=0.3$	$\sigma=0.6$
$n=250$						
$m=100$	0.432 (0.033)	0.711 (0.037)	0.493 (0.038)	0.727 (0.041)		
$n=250$						
$m=500$	0.387 (0.035)	0.587 (0.033)	0.401 (0.037)	0.584 (0.040)		
$n=1000$						
$m=100$	0.247 (0.016)	0.423 (0.018)	0.261 (0.013)	0.416 (0.017)		
$n=1000$						
$m=500$	0.207 (0.017)	0.332 (0.014)	0.201 (0.013)	0.313 (0.015)		