Deconvolution in Nonparametric Statistics

K. De Brabanter[1]

[1]Department of Statistics
Department of Industrial and Manufacturing Systems Engineering
Iowa State University

November, 15 2019

Outline

1. Introduction
   - What is deconvolution?
   - Where do such problems occur?
   - Notations & assumptions
2. Density deconvolution
   - Naive estimator
   - Kernel density deconvolution estimator
3. Nonparametric regression with errors-in-variables
   - Errors-in-variables problem formulation
   - Kernel regression with errors-in-variables

What is deconvolution?

- Reverse of convolution

What is deconvolution?

- Reverse of convolution

Deconvolution

Goal: Estimate a function f while access is restricted to the quantity

$$z = f * G = \int f(x-y)\, dG(y),$$

i.e., the convolution of f and some probability distribution G.

What is deconvolution?

- Reverse of convolution

Deconvolution

Goal: Estimate a function f while access is restricted to the quantity

$$z = f * G = \int f(x-y)\,dG(y),$$

i.e., the convolution of f and some probability distribution G.

- SOLUTION
  1. Estimate $z \Rightarrow \hat{z}$
  2. Apply deconvolution procedure to $\hat{z} \Rightarrow \hat{f}$

Where do such problems occur?

- Density estimation based on contaminated data (Delaigle & Gijbels, 2006)
- Nonparametric regression with errors-in-variables (Delaigle & Meister, 2011)
- Image and signal deblurring (Qiu, 2005)
- Econometrics (Meister, 2009)

Notations & assumptions

Deconvolution

$$z = f * G = \int f(x-y)\, dG(y)$$

- For now, assume G is fully known
- G has a density function g in Lebesgue sense:

$$z = f * G = \int f(x-y) g(y)\, dy$$

- Main tool: Fourier transform (FT)

$$\mathcal{G}(t) = \int \exp(itx)\, dG(x), \quad t \in \mathbb{R}$$

- Result via FT: $\mathcal{Z} = \mathcal{F} \cdot \mathcal{G}$

Simple algorithm

$$\mathcal{Z} = \mathcal{F} \cdot \mathcal{G}$$

1. Estimate $\mathcal{Z}$ based on empirical information, denoted by $\hat{\mathcal{Z}}$.
2. Calculate $\hat{\mathcal{Z}}(t)$ and divide it by $\mathcal{G}(t)$, leading to $\hat{\mathcal{F}}(t)$.
3. Regularize $\hat{\mathcal{F}}(t)$ so that the inverse Fourier transform $\hat{f}$ exists.

Simple algorithm

$$\mathcal{Z} = \mathcal{F} \cdot \mathcal{G}$$

1. Estimate $\mathcal{Z}$ based on empirical information, denoted by $\hat{\mathcal{Z}}$.
2. Calculate $\hat{\mathcal{Z}}(t)$ and divide it by $\mathcal{G}(t)$, leading to $\hat{\mathcal{F}}(t)$.
3. Regularize $\hat{\mathcal{F}}(t)$ so that the inverse Fourier transform $\hat{f}$ exists.

Simple algorithm, theoretical effort must not be underestimated

Assumptions & general estimation procedure

- Direct data not always available in practice because of measurement error
- Observe contaminated data $Y_1, \dots, Y_n$ instead of the true data $X_1, \dots, X_n$
- Assume the following model

$$Y_j = X_j + \varepsilon_j, \quad j \in \{1, \dots, n\}$$

- Assumptions
  - X_j and ε_j are real valued and independent
  - $\mathbf{E}(\varepsilon_j | X_j) = 0$
  - $\mathbf{Var}(\varepsilon_j | X_j) < \infty$
- Goal: estimate density f of the **UNOBSERVED** random variable X

Estimation procedure

- Basic statistics: density of the sum of two independent random variables $\Rightarrow$ deconvolution of both addends
- **STEP 1**: estimate density z given observations Y_j

$$\mathcal{Z}(t) = \int \exp(itx) z(x)\, dx = \mathbf{E} \exp(itY) = \Psi_Y(t)$$

- **STEP 2**: define estimator of $\Psi_Y(t)$

$$\hat{\Psi}_Y(t) = \frac{1}{n} \sum_{j=1}^{n} \exp(itY_j)$$

Estimation procedure

- Basic statistics: density of the sum of two independent random variables $\Rightarrow$ deconvolution of both addends
- **STEP 1**: estimate density z given observations Y_j

$$\mathcal{Z}(t) = \int \exp(itx) z(x)\, dx = \mathbf{E} \exp(itY) = \Psi_Y(t)$$

- **STEP 2**: define estimator of $\Psi_Y(t)$

$$\hat{\Psi}_Y(t) = \frac{1}{n} \sum_{j=1}^{n} \exp(itY_j)$$

$$\mathcal{Z}(t) \quad = \quad \Psi_Y(t) = \mathbf{E} \exp(it(X+\varepsilon)) = \mathbf{E}[\exp(itX) \exp(it\varepsilon)]$$

Estimation procedure

- Basic statistics: density of the sum of two independent random variables $\Rightarrow$ deconvolution of both addends
- **STEP 1**: estimate density z given observations Y_j

$$\mathcal{Z}(t) = \int \exp(itx) z(x)\, dx = \mathbf{E} \exp(itY) = \Psi_Y(t)$$

- **STEP 2**: define estimator of $\Psi_Y(t)$

$$\hat{\Psi}_Y(t) = \frac{1}{n} \sum_{j=1}^{n} \exp(itY_j)$$

$$\begin{aligned} \mathcal{Z}(t) &= \Psi_Y(t) = \mathbf{E} \exp(it(X+\varepsilon)) = \mathbf{E}[\exp(itX) \exp(it\varepsilon)] \\ &= \mathbf{E} \exp(itX) \cdot \mathbf{E} \exp(it\varepsilon) \end{aligned}$$

Estimation procedure

- Basic statistics: density of the sum of two independent random variables $\Rightarrow$ deconvolution of both addends
- **STEP 1**: estimate density z given observations Y_j

$$\mathcal{Z}(t) = \int \exp(itx) z(x)\, dx = \mathbf{E} \exp(itY) = \Psi_Y(t)$$

- **STEP 2**: define estimator of $\Psi_Y(t)$

$$\hat{\Psi}_Y(t) = \frac{1}{n} \sum_{j=1}^{n} \exp(itY_j)$$

$$\begin{aligned} \mathcal{Z}(t) &= \Psi_Y(t) = \mathbf{E} \exp(it(X+\varepsilon)) = \mathbf{E}[\exp(itX)\exp(it\varepsilon)] \\ &= \mathbf{E} \exp(itX) \cdot \mathbf{E} \exp(it\varepsilon) = \Psi_X(t) \cdot \Psi_\varepsilon(t) \end{aligned}$$

Estimation procedure

- Basic statistics: density of the sum of two independent random variables $\Rightarrow$ deconvolution of both addends
- **STEP 1**: estimate density z given observations Y_j

$$\mathcal{Z}(t) = \int \exp(itx) z(x)\, dx = \mathbf{E} \exp(itY) = \Psi_Y(t)$$

- **STEP 2**: define estimator of $\Psi_Y(t)$

$$\hat{\Psi}_Y(t) = \frac{1}{n} \sum_{j=1}^{n} \exp(itY_j)$$

$$\begin{aligned} \mathcal{Z}(t) &= \Psi_Y(t) = \mathbf{E} \exp(it(X+\varepsilon)) = \mathbf{E}[\exp(itX)\exp(it\varepsilon)] \\ &= \mathbf{E} \exp(itX) \cdot \mathbf{E} \exp(it\varepsilon) = \Psi_X(t) \cdot \Psi_\varepsilon(t) = \mathcal{F}(t) \cdot \mathcal{G}(t) \end{aligned}$$

Estimation procedure (cont'd)

$$\mathcal{Z}(t) = \Psi_X(t) \cdot \Psi_\varepsilon(t) = \mathcal{F}(t) \cdot \mathcal{G}(t) = \hat{\Psi}_Y(t) = \frac{1}{n} \sum_{j=1}^{n} \exp(itY_j)$$

- **STEP 3**: define estimator for $\mathcal{F}(t)$

$$\hat{\Psi}_X(t) = \frac{1}{n} \frac{\sum_{j=1}^{n} \exp(itY_j)}{\mathcal{G}(t)}$$

- This estimator is unbiased and consistent. A naive estimator for f is

$$\boxed{\hat{f}_{\text{naive}}(x) = \frac{1}{2\pi} \int \exp(-itx) \hat{\Psi}_X(t)\, dt}$$

Estimation procedure (cont'd)

$$\mathcal{Z}(t) \quad = \quad \Psi_X(t) \cdot \Psi_\varepsilon(t) = \mathcal{F}(t) \cdot \mathcal{G}(t) = \hat{\Psi}_Y(t) = \frac{1}{n} \sum_{j=1}^{n} \exp(itY_j)$$

- **STEP 3**: define estimator for $\mathcal{F}(t)$

$$\hat{\Psi}_X(t) = \frac{1}{n} \frac{\sum_{j=1}^{n} \exp(itY_j)}{\mathcal{G}(t)}$$

- This estimator is unbiased and consistent. A naive estimator for f is

$$\boxed{\hat{f}_{\text{naive}}(x) = \frac{1}{2\pi} \int \exp(-itx) \hat{\Psi}_X(t)\, dt}$$

This estimator has some drawbacks $\Rightarrow$ regularization needed!!

Kernel density deconvolution estimator

- Use the following estimator for the density z

$$\hat{z}(x) = \frac{1}{nh} \sum_{j=1}^{n} K\left(\frac{x - Y_j}{h}\right)$$

- $K : \mathbb{R} \to \mathbb{R}^+$
- bandwidth parameter $h > 0$

Kernel density deconvolution estimator

- Use the following estimator for the density z

$$\hat{z}(x) = \frac{1}{nh} \sum_{j=1}^{n} K\left(\frac{x - Y_j}{h}\right)$$

- $K : \mathbb{R} \to \mathbb{R}^+$
- bandwidth parameter $h > 0$

Existence condition

If $K \in L_1(\mathbb{R}) \cap L_2(\mathbb{R})$, in Lebesgue sense, the estimator $\hat{z}$ also lies $L_1(\mathbb{R}) \cap L_2(\mathbb{R})$ almost surely so that its Fourier transform exist.

Kernel density deconvolution estimator

- Use the following estimator for the density z

$$\hat{z}(x) = \frac{1}{nh} \sum_{j=1}^{n} K\left(\frac{x - Y_j}{h}\right)$$

- $K : \mathbb{R} \to \mathbb{R}^+$
- bandwidth parameter $h > 0$

Existence condition

If $K \in L_1(\mathbb{R}) \cap L_2(\mathbb{R})$, in Lebesgue sense, the estimator $\hat{z}$ also lies $L_1(\mathbb{R}) \cap L_2(\mathbb{R})$ almost surely so that its Fourier transform exist.

$$\mathcal{Z}(t) \quad = \quad \frac{1}{nh} \int \sum_{j=1}^{n} \exp(itx) K\left(\frac{x - Y_j}{h}\right) dx = \ldots = \hat{\Psi}_Y(t) \cdot \mathcal{K}(th)$$

Kernel density deconvolution estimator (cont'd)

- $\mathcal{Z}(t) = \hat{\Psi}_Y(t) \cdot \mathcal{K}(th)$
- $\mathcal{Z}(t) = \mathcal{F}(t) \cdot \mathcal{G}(t)$

$$\Downarrow$$

$$\hat{\Psi}_X(t) = \frac{\hat{\Psi}_Y(t)\mathcal{K}(th)}{\mathcal{G}(t)}$$

$$\Downarrow$$

$$\boxed{\hat{f}(x) = \frac{1}{2\pi} \int \exp(-itx)\mathcal{K}(th) \frac{\frac{1}{n}\sum_{j=1}^{n} \exp(itY_j)}{\mathcal{G}(t)}\, dt}$$

Some examples

- Consider two densities

$$X \sim 0.5N(-3,1^2)+0.5N(2,1^2) \text{ and } X \sim 0.5N(0,1^2)+0.5N(3,(1/2)^2)$$

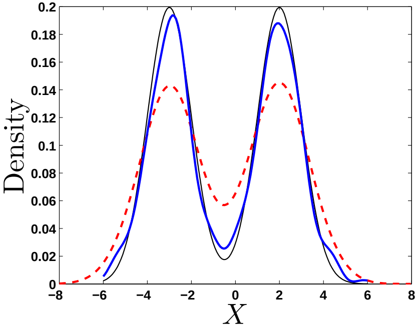
- Error density g is $\mathcal{L}(0,0.5) \to \mathcal{G}(t) = 4/(4+t^2)$
- Kernel: $K(u) = (1/2\pi)[\sin(\frac{1}{2}u)/\frac{1}{2}u]^2 \Rightarrow \mathcal{K}(t) = \max(1-|t|,0)$

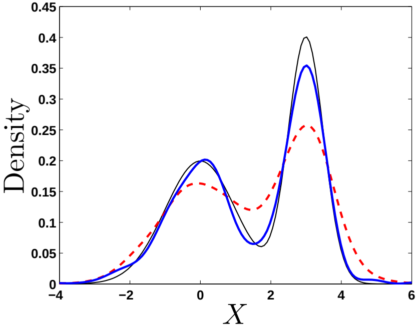
Some examples

- Consider two densities

$$X \sim 0.5N(-3,1^2)+0.5N(2,1^2) \text{ and } X \sim 0.5N(0,1^2)+0.5N(3,(1/2)^2)$$

- Error density g is $\mathcal{L}(0,0.5) \to \mathcal{G}(t) = 4/(4+t^2)$
- Kernel: $K(u) = (1/2\pi)[\sin(\frac{1}{2}u)/\frac{1}{2}u]^2 \Rightarrow \mathcal{K}(t) = \max(1-|t|,0)$





Errors-in-variables problem formulation

1. **Classical view**
    - Covariates X can only be observed with some additive independent noise
    - We observe the i.i.d. data $(W_1, Y_1), \ldots, (W_n, Y_n)$
    - $W_j = X_j + \delta_j \qquad \text{and} \qquad Y_j = m(X_j) + \varepsilon_j \quad \text{for } j = 1, \ldots, n$

Errors-in-variables problem formulation

1. **Classical view**
   - Covariates X can only be observed with some additive independent noise
   - We observe the i.i.d. data $(W_1, Y_1), \ldots, (W_n, Y_n)$
   - $W_j = X_j + \delta_j \quad$ and $\quad Y_j = m(X_j) + \varepsilon_j \quad$ for $j = 1, \ldots, n$
2. **Berkson regression** (Berkson, 1950)
   - Main difference: covariate is affected by additive noise **after** it was measured
   - We observe the i.i.d. data $(X_1, Y_1), \ldots, (X_n, Y_n)$
   - $Y_j = m(X_j + \delta_j) + \varepsilon_j, \quad j = 1, \ldots, n$

Kernel regression with errors-in-variables

- Adapt Nadaraya-Watson regression estimator for errors-in-variables

$$\hat{m}(x) = \sum_{j=1}^{n} \frac{K\left(\frac{x-X_j}{h}\right) Y_j}{\sum_{j=1}^{n} K\left(\frac{x-X_j}{h}\right)}$$

- Using similar techniques as for density estimation (Fan & Truong, 1993)

$$\Downarrow$$

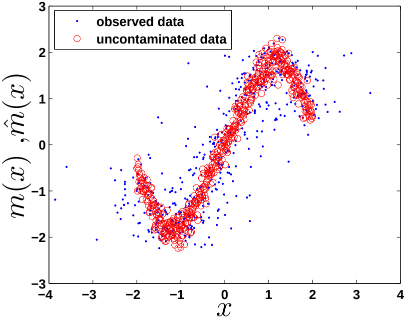
$$\boxed{\hat{m}(x) = \frac{\frac{1}{nh}\sum_{j=1}^{n} \frac{1}{2\pi}\left[\int \exp\left(-i\left(\frac{x-W_j}{h}\right)u\right)\frac{\mathcal{K}(u)}{\mathcal{G}(\frac{u}{h})}\,du\right] Y_j}{\frac{1}{2n\pi}\sum_{j=1}^{n}\int \exp(-itx)\mathcal{K}(th)\frac{\exp(itW_j)}{\mathcal{G}(t)}\,dt}}$$

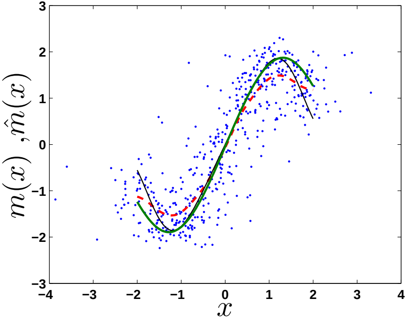
Some examples

- $W_j = X_j + \delta_j \quad \text{and} \quad Y_j = m(X_j) + \varepsilon_j \quad \text{for } j = 1, \ldots, n$
- $m(x) = 2x \exp(-10x^4/81) \quad \text{with} \quad x \in [-2, 2]$
- $\varepsilon \sim N(0, 0.2^2)$ and $\delta \sim \mathcal{L}(0, 0.2)$
- $\mathbf{E}[\varepsilon | X] = 0$ and the δ_j are independent of the (X_j, Y_j)

Some examples

- $W_j = X_j + \delta_j$ and $Y_j = m(X_j) + \varepsilon_j$ for $j = 1, \ldots, n$
- $m(x) = 2x \exp(-10x^4/81)$ with $x \in [-2, 2]$
- $\varepsilon \sim N(0, 0.2^2)$ and $\delta \sim \mathcal{L}(0, 0.2)$
- $\mathbf{E}[\varepsilon|X] = 0$ and the δ_j are independent of the (X_j, Y_j)





Unknown noise distribution

- **Additional data**: error density is estimated from separate independent experiment (e.g. Golden standard - normal measurement)

Unknown noise distribution

- **Additional data**: error density is estimated from separate independent experiment (e.g. Golden standard - normal measurement)
- **Replicated measurements**: independent measurements (affected by error) are done several times

Unknown noise distribution

- **Additional data**: error density is estimated from separate independent experiment (e.g. Golden standard - normal measurement)
- **Replicated measurements**: independent measurements (affected by error) are done several times
- These are topics of research: see e.g. Efromovich (1997), Delaigle *et al.* (2009), Meister (2009), Delaigle & Meister (2011)

Thank you for your attention

Any questions...?

References

A. Delaigle & I. Gijbels. Data-driven boundary estimation in deconvolution problems. *Computational Statistics & Data Analysis*, 50(8):19651994, 2006.

A. Delaigle & A. Meister. Rate-optimal nonparametric estimation in classical and Berkson errors-in-variables. *Journal of Statistical Planning and Inference*, 141(1):102 114, 2011.

P. Qiu. *Image Processing and Jump Regression Analysis*. John Wiley & Sons, New York, 2005.

A. Meister. *Deconvolution Problems in Nonparametric Statistics*. Springer-Verlag, Berlin Heidelberg, 2009.

J. Berkson. Are there two regression problems? *Journal of the American Statistical Association*, 45:164180, 1950.

S. Efromovich. Density estimation for the case of supersmooth measurement error. *Journal of the American Statistical Association*, 92(438):526535, 1997.

A. Delaigle, J. Fan, & R.J. Carroll. A design-adaptive local polynomial estimator for the errors-in-variables problem. *Journal of the American Statistical Association*, 104(485):348359, 2009.

J. Fan & Y.K. Truong. Nonparametric regression with errors in variables. *The Annals of Statistics*, 21(4):19001925, 1993.