# Methodology for nonparametric deconvolution when the error distribution is unknown

Aurore Delaigle,

University of Melbourne

Joint work with

Peter Hall, University of Melbourne

# Model and data

- We observe continuous i.i.d. data $W_1, \ldots, W_n$
- $W_i = X_i + U_i$, $X_i$ and $U_i$ are independent
- $X_i \sim f_X$: variable of interest, $U_i \sim f_U$: measurement errors
- Ex: $X =$ long term saturated fat intake, systolic blood pressure, etc.

# Classical measurement error setting

- Characteristic function of a r.v. $V$:

$$\phi_V(t) = E\big(e^{itV}\big) = \int e^{itv} f_V(v)\, dv$$

- Assume $f_U$ is known and even, $\phi_U(t) \neq 0$ for all $t$, $\phi_X \in L_1$.

- $W_i = X_i + U_i$, $X_i$ and $U_i$ are independent

$$\Rightarrow \phi_W(t) = \phi_X(t)\phi_U(t) \;\Rightarrow\; \phi_X(t) = \phi_W(t)/\phi_U(t).$$

- Fourier inversion theorem implies

$$f_X(x) = \frac{1}{2\pi}\int e^{-itx}\phi_X(t)\, dt = \frac{1}{2\pi}\int e^{-itx}\phi_W(t)/\phi_U(t)\, dt$$

- Goal: estimate

$$f_X(x) = \frac{1}{2\pi} \int e^{-itx} \phi_W(t)/\phi_U(t)\, dt$$

- We know $f_U \Rightarrow$ we know $\phi_U$.
- Data: $W_1, \ldots, W_n \Rightarrow$ estimate $\phi_W(t) = E(e^{itW})$ by $\hat{\phi}_W(t) = n^{-1} \sum_{j=1}^{n} e^{itW_j}$

- Stefanski and Carroll (1990): estimate $f_X$ by

$$\hat{f}_X(x) = \frac{1}{2\pi} \int e^{-itx} \hat{\phi}_W(t) w(t)/\phi_U(t)\, dt$$

where $w$ is a weight function s.t. $w(t) \to 0$ as $|t| \to \infty$.

- Often, $w(t) = \phi_K(ht)$
- $h > 0$ is a param, $K$ is a symm fction, $\phi_K(t) = \int e^{itx} K(x)\, dx$

## Our setting

- Problem: Error density $f_U$ is not always known.
- Is it possible to estimate $f_X$ in this case?
- Assume $f_U$ is symmetric (even) and $\phi_U(t) > 0$ for all $t$.

- Can estimate $\phi_U$ from replicates $W_{ij} = X_i + U_{ij}$ (Li and Vuong, 1998; Delaigle et al., 2008)
- or from sample of $U_i$'s (Diggle and Hall, 1993; Neumann, 1997).
- Other cases: $f_U$ is parametric (Butucea and Matias, 2005, Meister, 2006).
- Dong and Lewbel (2011): prove identification when $X$ is binary and asymmetric.
- No general method can be applied broadly and enjoys good performance.

# Our setting

- Can we estimate $f_X$ without replicates and without param model?
- Not always.

- Our approach is unusual.
- It is the irregularity and unpleasantness of a real world $F_X$ distribution that allows us to do unexpected things.
- If $F_X$ were nice, symmetric and conventional (e.g. Gaussian), we could not recover it from data on $W$ without knowing the distribution of $U$.
- If it is reasonably irregular then we can estimate it consistently.

## When can we identify $f_X$?

- $f_U$ sym $\Rightarrow$ if $f_X$ sym, can't distinguish $f_X$ from $f_U$ knowing only $f_W$.
- Thus $f_X$ has to be asymmetric, but this is not enough.

- Suppose $X = Y + Z$ where $Y$ and $Z$ indep, $f_Z$ symmetric.
- Then $W = X + U = (Y + U) + Z = Y + (U + Z)$.
- Symmetric error could be $Z$, $U$ or $U + Z$ (can't identify which one).

- Thus $f_X$ can't be such that $X = Y + Z$ as above.
- Many non symmetric r.v. can be expressed in this way, but how likely are we to encounter them in real life?

# Real life distributions

- In real applications, $f_X$ can rarely be expected to be "regular".
- Classical symmetric distributions are often convenient for inference, but
- we rarely believe that our data come perfectly from such distributions.

- Data can come from very diverse populations $\Rightarrow f_X$ can be viewed as a member of a very diverse universe.
- Factorisation like $X = Y + Z$ with $Z$ symmetric imposes constraints on the structure of the universe
- Fails to hold with probability 1 for a member of that universe drawn at random.

# Phase function

- Phase function of a r.v. $X$: $\rho_X = \phi_X/|\phi_X|$.
- Since $W = X + U$ , $X$ indep of $U$ and $f_U$ symmetric, then

$$\rho_W = \phi_W/|\phi_W| = \phi_X\,\phi_U/\{|\phi_X|\,\phi_U\} = \rho_X$$

- Let $T = X+V$, $X$ indep of $V$, $f_V$ sym. Then $\rho_X = \rho_T$ and $\mathrm{var}(T) > \mathrm{var}(X)$.

- If $X = Y+Z$ ($Z \perp\!\!\!\perp Y$) with $f_Z$ symmetric, then $\rho_X = \rho_Y$, $\mathrm{var}(Y) < \mathrm{var}(X)$.
- We argued before that the above is unlikely in real life.
- Motivates us to assume that:

$F_X$ is the distr with smallest var among all distr with phase function $\rho_X$

# Estimation method

- Estimate $\rho_X = \phi_W / |\phi_W|$ from the data $W_1, \ldots, W_n$.
- Among all distr with phase fction $\hat{\rho}_X$, find the one with smallest variance.
- Tricky $\Rightarrow$ discretize the problem to make it simpler.

- Approximate $F_X$ by discrete distribution that puts masses $p_1, \ldots, p_m$ at atoms $x_1, \ldots, x_m$.
- Phase function of that distribution:

$$\rho_p(t) = \sum_{j=1}^{m} p_j \, \exp(itx_j) / \Big| \sum_{j=1}^{m} p_j \, \exp(itx_j) \Big|$$

# Estimation method (contd)

- Choose discrete approximation s.t. $\rho_p$ close to $\rho_X$.
- $\rho_p = \rho_X \iff \phi_W(t) - |\phi_W(t)|\, \rho_p(t) = 0$ for all $t$.

- Only have $\hat{\phi}_W$ and $|\hat{\phi}_W|$, and quality of $\hat{\phi}_W(t)$ degrades as $|t|$ increases.
- Let $w$ be a weight function
- Choose discrete distribution that minimises

$$T(p) = \int_{-\infty}^{\infty} \left| \hat{\phi}_W(t) - |\hat{\phi}_W(t)|\, \rho_p(t) \right|^2 w(t)\, dt \tag{1}$$

at the same time minimising the variance of the discrete distribution.

## How to do this in practice?

- We only optimise over the $p_j$'s.
- We draw $x_1, \ldots, x_m$ randomly and uniformly in $[\min_i W_i, \max_i W_i]$.
- We have a rule for choosing $m$, but could be improved.
- Take the weight $w(t) = 1\{|\hat{\phi}_W(t)| > n^{-1/4}\}$.

- Then find $p_1, \ldots, p_m$ that minimises the variance and

$$T(p) = \int_{-\infty}^{\infty} \left| \hat{\phi}_W(t) - |\hat{\phi}_W(t)| \, \rho_p(t) \right|^2 w(t) \, dt \tag{2}$$

under the constraint $\sum p_j = 1$, $p_j \geq 0$.

# Obtain density

- Once we have our discrete distribution with probas $\hat{p}_1, \ldots, \hat{p}_m$ at points $x_1, \ldots, x_m$, turn it into a density.

- Recall that by the Fourier inversion theorem,

$$f_X(x) = \frac{1}{2\pi} \int e^{-itx} \phi_X(t)\, dt.$$

  Take

$$\hat{f}_X(x) = \frac{1}{2\pi} \int e^{-itx} \hat{\phi}_X(t) \phi_K(ht)\, dt.$$

  where $\hat{\phi}_X(t)$ is the char fction of the discrete distribution, $K$ is a kernel function, $h > 0$ is a parameter.

- Choose $h$ as in standard errors-in-variables problems.

# Non decomposability

- Our method is motivated by the assumption that we cannot write

$$X = Y + Z$$

with $Y$ and $Z$ independent and $Z$ symmetric. This is called non decomposability where one component is symmetric.

- We have proved this in the discrete case in various "random universe" settings.
- For example it holds with probability one if $F_X$ is drawn randomly from the space of discrete distributions where the atoms $x_1, \ldots, x_m$ are irregularly spaced in the sense that the joint distribution of any finite number of the atoms is continuous.

# Consistency

- We have proved that $\sup_x |\hat{f}_X(x) - f_X(x)| \xrightarrow{P} 0$ as $n \to \infty$.
- However we don't have convergence rates.
- The problem is particularly difficult.

- Rather inexplicit conditions, but roughly:
- $\exists$ unique distribution with minimum var and phase function $\rho_X$.
- $m \to \infty$ as $n \to \infty$ (discrete approximation gets more precise)
- Smoothness constraints on $f_X$ and $f_U$.

# Real data example 1

- Framingham study (National Heart, Lung, and Blood Institute).

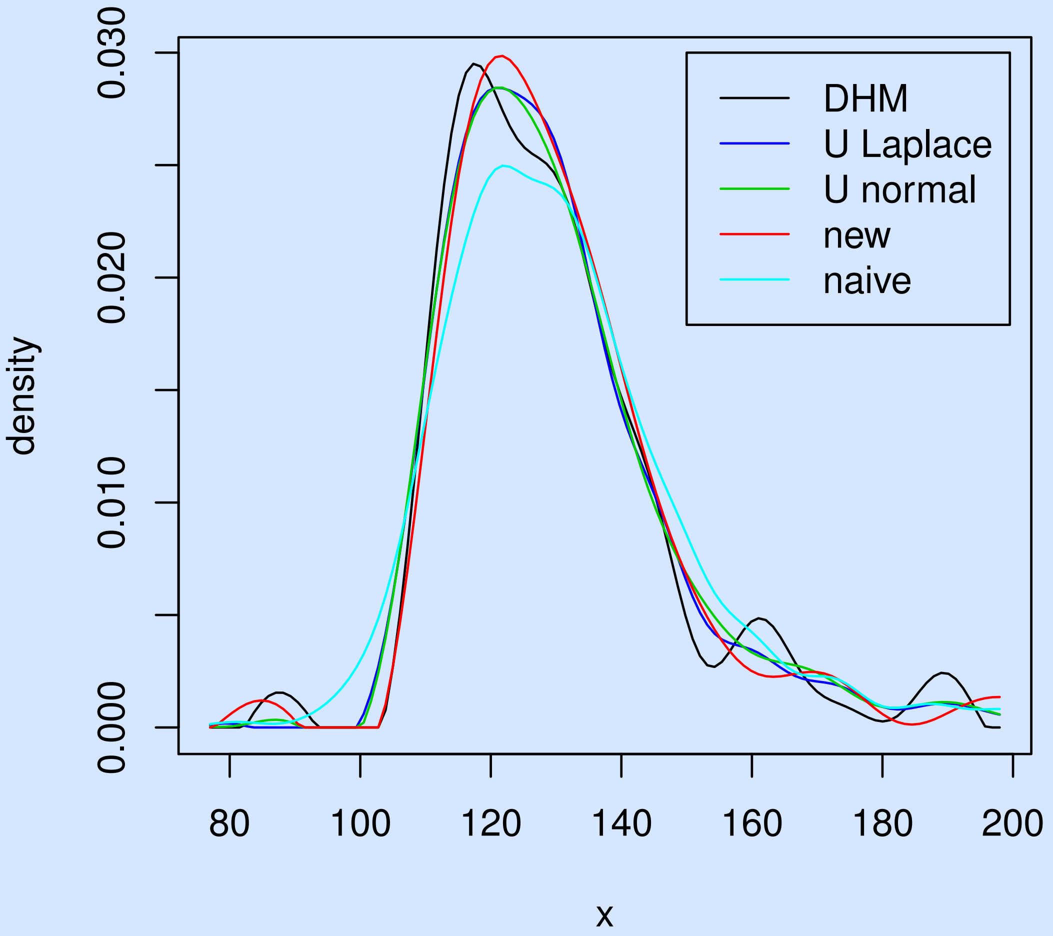
- Long term systolic blood pressure (SBP) is measured with lots of noise for $n = 1615$ patients.

- For each patient $i$, SBP measured twice at two exams. As in Carroll et al. (2006), for each $i$ let $M_{ij}$ be the average of the two measurements at exam $j$, for $j = 1$ and $2$.

- Take $W_{ij} = \log(50 - M_{ij})$ and assume $X_i = \log(50 - \mathrm{SBP}_i)$.

# Real data example 1

- Goal: see if our method works well with real data.
- Apply our method to data $W_{i1}$.
- Compare with method that estimates $f_U$ through $W_{i1} - W_{i2}$ as in Delaigle, Hall and Meister (2008).
- Compare with naive estimator that ignores the error.
- Also computes deconvolution estimator that assumes parametric model for $f_U$ (variance estimated by our method)

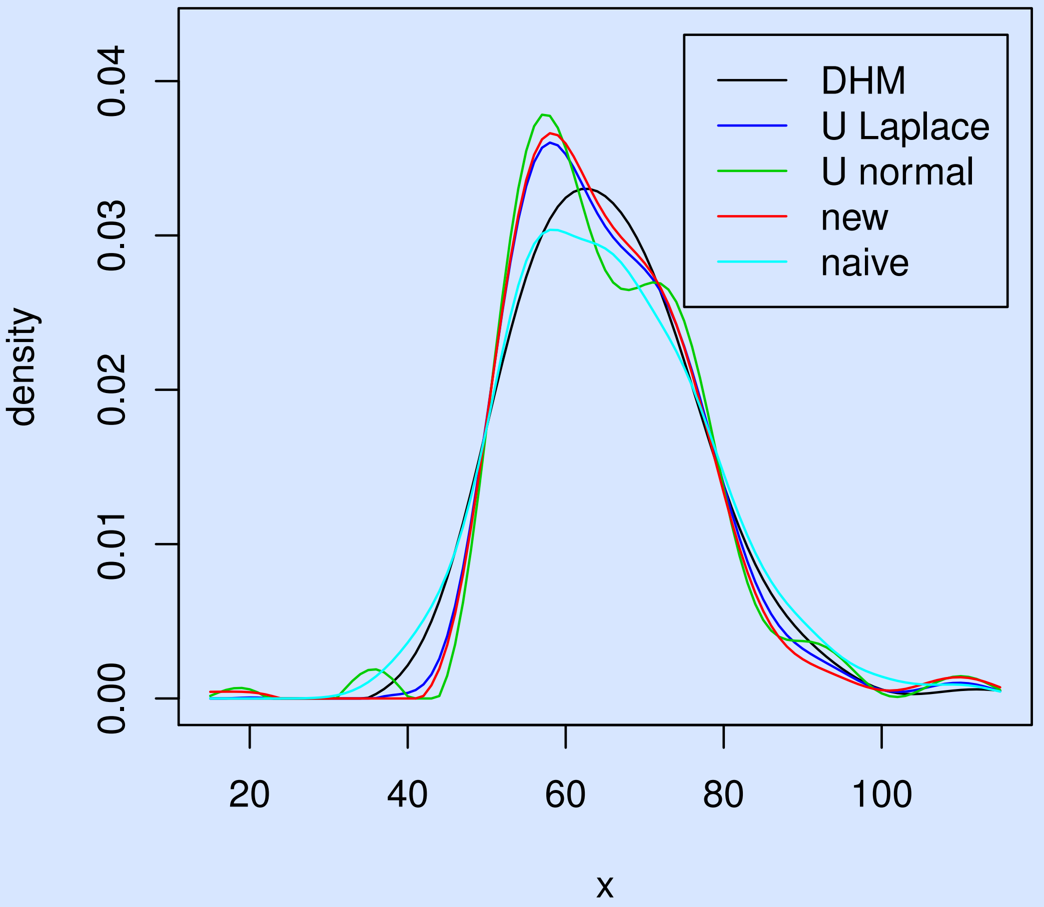

density
0.000
0.010
0.020
0.030
80
100
120
140
160
180
200
x
DHM
U Laplace
U normal
new
naive

# Real data example 2

- Pilot study on coronary heart disease (Morris, Marr and Clayton, 1977).

- We have error-prone measurements $W_{i1}$, $i = 1, \ldots, n$ of the ratio $X_i$ of poly-unsaturated fat to saturated fat intake for $n = 336$ men in a one-week dietary survey.

- For $60$ patients, $W_i$ is measured a second time several months later.
- As in example 1, can compare our method with method that estimates $f_U$ through $W_{i1} - W_{i2}$.


DHM
U Laplace
U normal
new
naive
density
x
0.00
0.01
0.02
0.03
0.04
20
40
60
80
100

# EXTENSION

- From our method we can estimate $f_U$.
- Therefore can apply it to other problems of measurement errors, e.g. regression.
- We have done this.
- Method works surprisingly well

# Some references


- Butucea, C. and Matias, C. (2005). Minimax estimation of the noise level and of the deconvolution density in a semiparametric convolution model. *Bernoulli*, **11**, 309–340.
- Butucea, C., Matias, C. and Pouet, P. (2008). Adaptivity in convolution models with partially known noise distribution. *Electron. J. Statist.*, **2**, 897-915.
- Carroll, R.J., Ruppert, D., Stefanski, L.A. and Crainiceanu, C.M. (2006). *Measurement Error in Nonlinear Models*, 2nd Edn. Chapman and Hall.
- Delaigle, A. and Hall, P. (2016). Methodology for nonparametric deconvolution when the error distribution is unknown. *J. Roy. Statist. Soc. B*, **78**, 231–252.
- Delaigle, A., Hall, P. and Meister, A. (2008). On Deconvolution with repeated measurements. *Ann. Statist.*, **36**, 665–685.
- Dong, Y. and Lewbel, A. (2011). Nonparametric identification of a binary random factor in cross section data. *J. Econometrics*, **163**, 163–171.
- Fan, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.*, **19**, 1257–1272.

# Some references


- Li, T. and Vuong, Q. (1998). Nonparametric estimation of the measurement error model using multiple indicators. *J. Multivariate Anal.*, **65**, 139–165.
- Meister, A. (2006). Density estimation with normal measurement error with unknown variance. *Statist. Sinica*, **16**, 195–211.
- Morris, J.N., Marr, J.W. and Clayton, D.G. (1977). Diet and heart: a postscript. *British Med. J.*, **2**, 1307–1314.
- Stefanski, L.A and Carroll, R.J. (1990). Deconvoluting kernel density estimators. *Statistics*, **21**, 169184.