Measurement Errors: Connections and Solutions

Andrey Feuerverger
University of Toronto
andrey@fisher.utstat.toronto.edu

Jiayang Sun
Case Western Reserve University
jiayang@sun.cwru.edu

__________________________

Workshop on Developments and Challenges in Mixture Models,
Bump Hunting and Measurement Error Models,
Cleveland, June 2-4, 2002

Outline

• Introduction
• Connections
• Challenges
• Naive Solutions
• Better Solutions
• Discussion and Remarks
Introduction

Many interesting problems can be formulated as studies about data with measurement errors:

- **Signal processing**: output is the original signal coupled with the filter’s impulse and errors
- **(Microarray) Image analysis**: observable is a blurred image

Ref: http://www.niac.man.ac.uk/projects/mcellum.html

- **Astronomy**: data are often subject to measurement errors

<table>
<thead>
<tr>
<th>velocity km/sec</th>
<th>id</th>
<th>fiber</th>
<th>peak width</th>
<th>tdr</th>
<th>error</th>
</tr>
</thead>
<tbody>
<tr>
<td>133.8519</td>
<td>318___15.</td>
<td>81</td>
<td>0.09</td>
<td>102.07</td>
<td>3.08</td>
</tr>
<tr>
<td>221.1162</td>
<td>337___15.</td>
<td>32</td>
<td>0.12</td>
<td>120.29</td>
<td>3.19</td>
</tr>
<tr>
<td>-120.3765</td>
<td>343___14.</td>
<td>88</td>
<td>0.15</td>
<td>129.35</td>
<td>4.35</td>
</tr>
<tr>
<td>144.1023</td>
<td>371___15.</td>
<td>84</td>
<td>0.20</td>
<td>114.98</td>
<td>5.82</td>
</tr>
<tr>
<td>-119.0284</td>
<td>202___14.</td>
<td>65</td>
<td>0.20</td>
<td>149.91</td>
<td>5.75</td>
</tr>
</tbody>
</table>

...
DO NOT ignore measurement errors

especially in bump hunting problems.

---

Road Map - connections

- $Y = m(X) + \epsilon$  
  **Observe** $Y$ and $W = X + U$. **What** is $m(x)$?

- $Y = X + U$  
  **Observe** $Y$. **What** is density $f_X$ of $X$?

- $Y(t) = K(x(t)) + U(t)$  
  related  
  **Observe** $Y$ at $t$ and know $k$. **What** is $x(t)$?  
  - inverse problems such as signal/image analyses
Challenges

Observe: $Y_j = X_j + U_j$, independently,
where: $X_j \sim f_X$, $U_j \sim f_j$.

1. If $f_j = f_0$, homogeneous; otherwise
   nonhomogeneous, e.g. $f_j = U(-h_j, h_j)$ or $N(0, h_j)$.

2. If $f_j$ is Uniform, it is a blessing and a challenge!

$\implies$ implications for other error distributions.

Deconvolution

• Homogeneous case: $Y_i = X_i + U_i$, $X_i \stackrel{iid}{\sim} f_X, U_i \stackrel{iid}{\sim} f_U$

Density:

$$f_Y = f_X * F_U$$

Characteristic function:

$$\varphi_Y = \varphi_X \cdot \varphi_U$$

Naive estimate:

$$\hat{f}_X(x) = \frac{1}{2\pi} \int \frac{\hat{\varphi}_Y(t)}{\varphi_U(t)} e^{-itx} dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{1}{n} \sum_j e^{itY_j} \varphi_U(t) e^{-itx} dt$$

Modified estimate:

$$\hat{f}_X(x) = \frac{1}{2\pi n} \int_{-\infty}^{\infty} \sum_j e^{itY_j} W(a_n t) e^{-itx} dt, \quad (1)$$

where $W(t)$ is a ‘window’ function, and $a_n \downarrow 0$. 
Note: If $W = \varphi_K$ where $K$ is a kernel, then (1) is the kernel deconvolution estimate studied by many authors, Carrol and Hall(88), Fan(91), Zhang(90) ...

**Problem:** The convergence rate of $\hat{f}_X$ is extremely slow and gets worse as $f_U$ gets smoother. The uniform distribution is neither ordinary nor super smooth as defined by Fan (1991). However, note that $
abla_U(t) = \sin(ht)/(ht)$ if $U \sim U(-h,h)$, so,

$$\hat{f}_X(x) = \frac{1}{2\pi n} \int_{-\infty}^{\infty} \sum_j e^{itY_j} \frac{W(a_n t) e^{-itx}}{\sin(t)/t} \, dt,$$

where WOLG we took $h=1$. There are singularities at $k\pi$! The numerator is not zero when the denominator is though $E(\text{numerator}) = 0$ then.

---

**Solutions**

1. **Abandon** the characteristic functions. This opens a completely new horizon.

   Sun, Morrison, Harding and Woodroffe (2002)

2. **Smooth out** the singularity points. There are several possibilities, including ones using splines and Shannon sampling formula. Some lead to optimal estimators!

   Both work for the nonhomogeneous case and have implications for non-uniform errors.
Solution 1
Abandon Ch functions - SMHW

\[ Y_j = X_j + U_j \text{ has a density} \]

\[ g_j(y) = \frac{F(y + h_j) - F(y - h_j)}{2h_j}, \quad F = F_X \]

\[ \Rightarrow \quad F(x) = 2 \sum_{k=0}^{\infty} \frac{1}{n} \sum_{j=1}^{n} h_j g_j(x - kh_j) \]

\[ \Rightarrow \quad \hat{F}_-(x) = 2 \sum_{k=0}^{\infty} \frac{1}{n} \sum_{i=1}^{n} h_i w_b(x - kh_i - Y_i) \]

\[ \hat{f}_-(x) = 2 \sum_{k=0}^{\infty} \frac{1}{n} \sum_{i=1}^{n} h_i w'_b(x - kh_i - Y_i) \]

where

\[ w_b(x) = \frac{1}{b} w\left(\frac{x}{b}\right). \]
Notes: $\hat{f}_-(x)$ is much better than the $\hat{f}_{naive}(x)$, where the true density is $f(x)$ in black.
Improvements: look at another side

Similarly

\[ g_j(y) = \frac{1 - F(y - h_j) - (1 - F(y + h_j))}{2h_j} \]

\[ \implies F(x) = 1 - 2 \sum_{k=\text{odd}} \frac{1}{n} \sum_{j=1}^{n} h_j g_j(x + kh_j) \]

\[ \implies \hat{F}_+(x) = 1 - 2 \sum_{k=\text{odd}} \frac{1}{n} \sum_{i=1}^{n} h_i w_b(x + kh_i - Y_i) \]

\[ \hat{f}_+(x) = -2 \sum_{k=\text{odd}} \frac{1}{n} \sum_{i=1}^{n} h_i w'_b(x + kh_i - Y_i), \]

---

**Proposition 1.** Let \( H = \sum h_j \). Then under modest conditions on \( b, h_j \) and \( f \),
\( \hat{F}^+, \hat{F}^- \) are asymptotically independent, and

\[ \hat{F}^- (x) \sim N \left( F(x), F(x) \frac{2H\|w\|^2}{n^2b} \right) \]

\[ \hat{F}^+ (x) \sim N \left( F(x), (1 - F(x)) \frac{2H\|w\|^2}{n^2b} \right) \]

**Idea:** use the Combined Estimator:

\[ \tilde{F}(x) = [1 - p(x)]\hat{F}^-(x) + p(x)\hat{F}^+(x), \]

where \( p \) is a distribution function.
Choices of $p$

Adhoc: $p = e^x/(1 + e^x)$.

$$\tilde{F}(x) = [1 - \frac{e^x}{1 + e^x}]\hat{F}_-(x) + \frac{e^x}{1 + e^x}\hat{F}_+(x)$$

MVUE: $p = F$

$$\hat{F}(x) = [1 - \hat{F}(x)]\hat{F}_-(x) + \hat{F}(x)\hat{F}_+(x)$$

$$\hat{F}(x) = \frac{\hat{F}_-(x)}{\hat{F}_-(x) + 1 - \hat{F}_+(x)},$$

“Optimal” Estimates of Densities

$$\tilde{f}(x) = [1 - p(x)]\tilde{f}_-(x) + p(x)\tilde{f}_+(x) + p'(x)[\hat{F}_+(x) - \hat{F}_+(x)]$$

$$\hat{f}(x) = \frac{[1 - \hat{F}(x)]\hat{f}_-(x) + \hat{F}(x)\hat{f}_+(x)}{\hat{F}_-(x) + 1 - \hat{F}_+(x)}$$
**MSE**

**Proposition 2.** Under the same conditions of Proposition 1, let $\bar{h} = H/n = O(1)$, then

\[
E\hat{F}_+(x) = E\hat{F}_-(x) = w_b * F(x),
\]

\[
E\hat{f}_+(x) = E\hat{f}_-(x) = w_b * f(x),
\]

\[
\text{MSE}(\hat{F}) = w_2^2 \frac{b^4}{4} f''^2(x) + \frac{2\bar{h}}{nb}||w||^2 F(x)[1 - F(x)] + o(b^4) + o\left(\frac{1}{nb}\right),
\]

\[
\text{MSE}(\hat{f}) = w_2^2 \frac{b^4}{4} f''''^2(x) + \frac{2\bar{h}}{n b^3}||w'||^2 F(x)[1 - F(x)] + o(b^4) + o\left(\frac{1}{nb^3}\right).
\]

---

**Optimal b**

The **optimal window widths** minimizing MISE of $\hat{F}$ and $\hat{f}$ are:

\[
b_F = n^{-1/5} \frac{2\bar{h}||w||^2 c}{w_2^2 ||f'||^2}^{1/5},
\]

\[
b_f = n^{-1/7} \left(\frac{6\bar{h}||w'||^2 c}{w_2^2 ||f''''||^2}\right)^{1/7}
\]

respectively, where $c = c(F) = \int F(x) (1 - F(x)) dx$.

<table>
<thead>
<tr>
<th>Rates</th>
<th>$b = O(n^{-1/5})$</th>
<th>$b = O(n^{-1/5})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{F}$</td>
<td>$n^{-2/5}$</td>
<td>$n^{-1/7}$</td>
</tr>
<tr>
<td>$\hat{f}$</td>
<td>$n^{-1/5}$</td>
<td>$n^{-1/7}$</td>
</tr>
</tbody>
</table>
Practical choices of $b$

We suggest 2 ways of computing $b_f$. The 1st is to estimate $f''$ and $c(F)$ based on a pilot kernel estimate of $f$ with a larger window width, say $b = n^{-1/7 + 1/5} b_w$. The justification of this method can be based on that of bootstrap (Delaigle A. and Gijbels I, (2001)). The 2nd can be obtained by tracking similar arguments to that of Silverman’s rule of thumb:

$$\hat{b}_F = n^{-1/5} 1.0324 (R^4 \bar{h})^{1/5}, \quad \hat{b}_f = n^{-1/7} 0.9763 (R^6 \bar{h})^{1/7},$$

where

$$R = \min\{[\text{Var}(Y) - (\bar{h})^2 / 3]^{1/2}, \frac{IQR(Y)}{1.34}\}.$$
Solution 2
Smooth out the singularity points

Naive deconvolution
→ Crude singularity adjusted deconvolution
→ Cosine bell adjusted deconvolution
→ Shannon singularity correction!

Two Examples using Solution 2

• Example 1:
  \[ X \sim N(0, 0.2^2), U \sim U(0, 1) \]

to motivate the final solution
Characteristic Functions of X and U

Uniform Char. Function

Normal Char. Function

Characteristic Function of Y = X + U

Naive Deconvolution
Example 2

- Example 2:

\[ X \sim 0.5 \cdot N(0.2, 0.2^2) + 0.5 \cdot N(-0.3, 0.1^2), \ U \sim U(0, 1) \]

to check how the final estimate work
Characteristic Function of Y = X + U, with Shannon-Fourier Singularity Correction

Densities: true = solid line, estimate = symbol
Discussion and Remarks

1. Connection to Regression:

\[ \varphi_{Y,n}(t) = \varphi(t) + e(t) = \varphi_X(t)\varphi_U(t) + e(t) \quad (2) \]

where the multiplier function \( \varphi_U(t) \) is known and the covariance of \( e(t) \) is easy to estimate. It has a connection to image analysis - see p6:

2. Splines - two folded:
   - Model \( \varphi_X(t) \) in (2), \( \varphi_X(t) = \beta^T s(t) \).
   - Smooth \( \hat{\varphi}_Y/\varphi_U \) - see p9 and jhf.ps.

3. Generalization to non-homogeneous errors:

\[
\hat{f}_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{1}{n} \sum_j \frac{e^{itY_j}}{\varphi_{U_j}(t)} e^{-itx} \, dt
\]

4. Comparisons of modified deconvolution and non-deconvolution estimates.

5. In non-homogeneous case, there is one data point for each error distribution. Are the final estimates (plots) much different under the normal and uniform error models (with the uniform and normal variances matched in some ways)?

6. Application in deblurring optical images that have been subjected to uniform motion over a finite interval of time, and other errors.