Estimation in nonparametric regression with discrete errors

Wolfgang Wefelmeyer (University of Cologne)

based on joint work with
Uschi Müller (Texas A&M University)
Anton Schick (Binghamton University)

mailto:wefelm@math.uni-koeln.de
http://www.mi.uni-koeln.de/∼wefelm/

EMS, Amsterdam, 6 July 2015

Direct approach

Consider the nonparametric regression model

$$Y = r(X) + \varepsilon.$$

Let Y have density h. We want to estimate h at a point y.
The *direct approach* uses the responses only, say a kernel estimator

$$\hat{h}(y) = \frac{1}{n} \sum_{i=1}^{n} K_b(y - Y_i) \quad \text{with} \quad K_b(y) = \frac{1}{b} K\Big(\frac{y}{b}\Big).$$

If K is bounded with bounded support,

$$nb \operatorname{Var} \hat{h}(y) \to h(y) \int K^2(u)\, du \quad \text{for} \quad nb \to \infty.$$

If h is s times differentiable at y and K is of order s,

$$b^{-s}\Big(E\hat{h}(y) - h(y)\Big) \to h^{(s)}(y) \frac{(-1)^s}{s!} \int u^s K(u)\, du \quad \text{for} \quad b \to 0.$$

If h is s times differentiable at y and K is of order s, then the optimal rate of the bandwidth is $b = n^{-1/(2s+1)}$, and the optimal rate of the kernel estimator is $n^{-s/(2s+1)}$.

For this bandwidth, $n^{s/(2s+1)}(\widehat{h}(y) - h(y))$ is asymptotically normal with mean

$$h^{(s)}(y)\frac{(-1)^s}{s!}\int u^s K(u)\,du$$

and variance

$$h(y)\int K^2(u)\,du.$$

Local von Mises statistic

Let $Y = r(X) + \varepsilon$ with X, ε *independent*. A better estimator for the response density h than the direct $\widehat{h}$ is a *local von Mises statistic*

$$\frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} K_b(y - \widehat{r}(X_i) - \widehat{\varepsilon}_j)$$

with *residuals* $\widehat{\varepsilon}_j = Y_j - \widehat{r}(X_j)$ and a local polynomial smoother $\widehat{r}$.
Several cases can be distinguished.
If both $r(X)$ and ε have *densities*, say f and e, and $f(X)$ and $e(\varepsilon)$ have finite second moments, then the local von Mises statistic has rate $n^{-1/2}$, see Schick/W. 2012, 2013 and Giné/Mason 2007.
If $r(X)$ or ε are *discrete*, then the convolution h is just a linear combination of densities, and the local von Mises statistic is not faster than the kernel estimator based on the responses (even though discrete distributions can be estimated at faster rates than densities).
For discrete $r(X)$ see Müller/Schick/W. 2015.

Regression with discrete errors

Let $Y = r(X) + \varepsilon$ with X, ε *independent* and r *increasing*.
Let ε have support $t_1 < \cdots < t_m$ with $P(\varepsilon = t_k) = p_k > 0$.
Let $r(X)$ have density f. Then Y has convolution density

$$h(y) = \sum_{k=1}^{m} f(y - t_k) p_k.$$

Let X have density g with $r'(r^{-1}(z)) \neq 0$. Then

$$f(z) = \frac{g(r^{-1}(z))}{r'(r^{-1}(z))}.$$

Hence h is s times differentiable at y if (and only if) g is s times differentiable and r is $s+1$ times differentiable at $r^{-1}(y - t_k)$ for $k = 1, \ldots, m$.
We estimate $h(y)$ by a plug-in-estimator of the form

$$\hat{\hat{h}}(y) = \sum_{k=1}^{m} \hat{f}(y - \hat{t}_k) \hat{p}_k.$$

First estimator of $f(z) = g(r^{-1}(z))/r'(r^{-1}(z))$

Estimate $f(z)$ by the plug-in-estimator

$$\widehat{f}(z) = \frac{\widehat{g}(\widehat{r}^{-1}(z))}{\widehat{r}'(\widehat{r}^{-1}(z))}.$$

Assume that g is s times and r is $s+1$ times differentiable at $r^{-1}(z)$.
Then g can be estimated at the *same* rate $n^{-s/(2s+1)}$ as h,
and r at the *faster* rate $n^{-(s+1)/(2(s+1)+1)}$,
but r' only at the *slower* rate $n^{-s/(2(s+1)+1)}$.

It follows that $\widehat{f}(z)$ has the slower rate $n^{-s/(2(s+1)+1)}$.

The corresponding estimator $\widehat{\widehat{h}}(y) = \sum_{k=1}^{m} \widehat{f}(y-\widehat{t}_k)\widehat{p}_k$ of the response density also has this rate and is *slower* than the kernel estimator $\widehat{h}(y)$.

Second estimator of $f(z) = g(r^{-1}(z))/r'(r^{-1}(z))$

Estimate $f(z)$ by the kernel estimator

$$\hat{f}(z) = \frac{1}{n} \sum_{i=1}^{n} K_b(z - \hat{r}(X_i)).$$

We assume that g is s times and r is $s+1$ times differentiable at z.
Then r can be estimated at the rate $n^{-(s+1)/(2(s+1)+1)}$.
The density $f(z)$ of $r(X)$ is s times differentiable.
Take K of order s and $b = n^{-1/(2s+1)}$.
Then the $\hat{r}(X_i)$ enter $\hat{f}(z)$ asymptotically like the true $r(X_i)$.

Hence $n^{s/(2s+1)}(\hat{f}(z) - f(z))$ is asymptotically normal with

$$\text{mean } f^{(s)}(z)\frac{(-1)^s}{s!}\int u^s K(u)\,du \text{ and variance } f(z)\int K^2(u)\,du.$$

Estimator of the regression function

We may take a *local polynomial smoother* $\widehat{r}(x)$ of order $s+1$.

Take $(\widehat{r}(x), \dots, \widehat{r}^{(s+1)}(x)) = (\vartheta_0, \dots, (s+1)!\vartheta_{s+1})$ minimizing

$$\sum_{i=1}^{n} \Big(Y_i - \sum_{j=0}^{s+1} \vartheta_j (X_i - x)^j \Big)^2 w_b(X_i - x).$$

Her $w_b(x) = w(x/b)/b$ and w is a density.

Estimator of the response density

We estimate $h(y)$ by the plug-in-estimator

$$\widehat{\widehat{h}}(y) = \sum_{k=1}^{m} \widehat{f}(y - \widehat{t}_k)\widehat{p}_k,$$

where

$$\widehat{f}(z) = \frac{1}{n}\sum_{i=1}^{n} K_b(z - \widehat{r}(X_i))$$

with kernel K of order s and bandwidth $b = n^{-1/(2s+1)}$.

We will show that t_k and p_k can be estimated at faster rates than f. Hence these estimators will not influence the asymptotic distribution of $\widehat{\widehat{h}}(y)$.

$\hat{t}_k$, $\hat{p}_k$ enter like t_k, p_k

Let $\hat{\varepsilon}_i = Y_i - \hat{r}(X_i)$ denote the residuals. The *residual-based distribution function* is $\hat{F}(z) = \frac{1}{n}\sum_{i=1}^n \mathbf{1}(\hat{\varepsilon}_i \le z)$. With $\mathbf{t} = (t_1, \dots, t_m)$ and $\mathbf{p} = (p_1, \dots, p_m)$, the distribution function of the error ε is

$$F_{\mathbf{tp}}(z) = \sum_{k=1}^m p_k \mathbf{1}(t_k \le z).$$

The *least squares estimator* $\hat{\mathbf{t}}$, $\hat{\mathbf{p}}$ of $\mathbf{t}$, $\mathbf{p}$ minimizes

$$\int \Big(\hat{F}(z) - F_{\mathbf{tp}}(z)\Big)^2 dz.$$

Then $\hat{\mathbf{t}}$ has rate n^{-1} and $\hat{\mathbf{p}}$ has rate $n^{-1/2}$. (This is similar to estimating regression functions with jumps; see Koul/Qian/Surgailis 2003 and Ciuperca 2009.) We obtain

$$n^{s/(2s+1)}(\hat{\hat{h}}(y) - h(y)) = \sum_{k=1}^m n^{s/(2s+1)}(\hat{f} - f)(y - t_k)p_k + o_p(1).$$

Main result

We estimate the response density by

$$\widehat{\widehat{h}}(y) = \sum_{k=1}^{m} \widehat{f}(y - \widehat{t}_k)\widehat{p}_k \text{ with } \widehat{f}(z) = \frac{1}{n}\sum_{i=1}^{n} K_b(z - \widehat{r}(X_i)),$$

where K is of order s and $b = n^{-1/(2s+1)}$. We write

$$\tilde{f}(z) = \frac{1}{n}\sum_{i=1}^{n} K_b(z - r(X_i))$$

and obtain

$$n^{s/(2s+1)}(\widehat{\widehat{h}}(y) - h(y)) = \sum_{k=1}^{m} n^{s/(2s+1)}(\tilde{f} - f)(y - t_k)p_k + o_p(1).$$

Hence $n^{s/(2s+1)}(\widehat{\widehat{h}}(y) - h(y))$ is asymptotically normal with mean

$$\sum_{k=1}^{m} f^{(s)}(y - t_k)p_k \frac{(-1)^s}{s!} \int u^s K(s)\, ds = h^{(s)}(y)\frac{(-1)^s}{s!} \int u^s K(s)\, ds$$

and variance

$$\sum_{k=1}^{m} f(y - t_k)p_k^2 \int K^2(u)\, du.$$

The mean is the same as for the kernel estimator $\widehat{h}(y)$, but the variance now has p_k^2 in place of p_k. This is a (considerable) reduction.

A fast estimator of the regression function

Decompose the real line into intervals $\widehat{I}_1, \ldots, \widehat{I}_m$ that contain $\widehat{t}_1, \ldots, \widehat{t}_m$ in their interiors, using midpoints of $\widehat{t}_1, \ldots, \widehat{t}_m$. Define

$$\overline{r}(X_i) = Y_i - t_k \quad \text{if} \quad \widehat{r}(X_i) \in \widehat{I}_k.$$

Then $\overline{r}(X_i) = r(X_i) + O_p(n^{-1})$.

Hence $\overline{r}$ and $\overline{r}'$ converge faster than $n^{-1/2}$.

First estimator of $f(z) = g(r^{-1}(z))/r'(r^{-1}(z))$, again

Estimate $f(z)$ by the plug-in-estimator

$$\widehat{f}(z) = \frac{\widehat{g}(\overline{r}^{-1}(z))}{\overline{r}'(\overline{r}^{-1}(z))} \quad \text{with} \quad \widehat{g}(x) = \frac{1}{n}\sum_{i=1}^{n} K_b(x - X_i).$$

We assume that g is s times and r is $s+1$ times differentiable at z. Take K of order s and $b = n^{-1/(2s+1)}$. Then

$$n^{s/(2s+1)}(\widehat{f}(z) - f(z)) = \frac{1}{r'(r^{-1}(z))} n^{s/(2s+1)}(\widehat{g}(z) - g(z)) + o_p(1).$$

Estimate $h(y)$ by

$$\widehat{\widehat{h}}(y) = \sum_{k=1}^{m} \widehat{g}(y - \widehat{t}_k)) \frac{\widehat{p}_k}{\overline{r}'(\overline{r}^{-1}(y - \widehat{t}_k))}.$$

This is not always better than the kernel $\widehat{h}(y)$:

$$\begin{aligned} & n^{s/(2s+1)}(\widehat{\widehat{h}}(y) - h(y)) \\ &= \sum_{k=1}^{m} n^{s/(2s+1)}\Big(\widehat{g} - g)(y - t_k) \frac{p_k}{r'(r^{-1}(y - t_k))} + o_p(1). \end{aligned}$$