5.1 Two-Step Conditional Density Estimator

We can write

$$y = g(X) + e$$

where $g(x)$ is the conditional mean function and e is the regression error. Let $f_e(e \mid x)$ be the conditional density of e given $X = x$. Then the conditional density of y is

$$f(y \mid x) = f_e(y - g(x) \mid x)$$

That is, we can write the conditional density of y in terms of the regression function and the conditional density of the error.

This decomposition suggests an alternative two-step estimator of f. First, estimate g. Second, estimate f_e.

The estimator $\hat{g}(x)$ for g can be NW, WNW, or LL.

The residuals are $\hat{e}_i = y_i - \hat{g}(X_i)$.

The second step is a conditional density estimator (NW, WNW or LL) applied to the residuals $\hat{e}_i$ as if they are observed data. This gives an estimator $\hat{f}_e(e \mid x)$.

The estimator for f is then

$$\hat{f}(y \mid x) = \hat{f}_e(y - \hat{g}(x) \mid x)$$

The first-order asymptotic distribution of $\hat{f}$ turns out to be identical to the ideal case where e_i is directly observed. This is because the first step conditional mean estimator $\hat{g}(x)$ converges at a rate faster than the second step estimator (at least if the first step is done with a bandwidth of the optimal order). e.g. if $q = 1$ then $\hat{g}(x)$ is optimally computed with a bandwidth $h \sim n^{-1/5}$, so that $\hat{g}$ converges at the rate $n^{-2/5}$, yet the estimator $\hat{f}_e$ converges at the best rate $n^{-1/3}$, so the error induced by estimation of $\hat{g}$ is of lower stochastic order.

The gain from the two-step estimator is that the conditional density of e typically has less dependence on X than the conditional density of y itself. This is because the conditional mean $g(X)$ has been removed, leaving only the higher-order dependence. The accuracy of nonparametric estimation improves as the estimated function becomes smoother and less dependent on the conditioning variables. Partially this occurs because reduced dependence allows for larger bandwidths, which reduces estimation variance.)

As an extreme case, if $f_e(e \mid x) = f_e(e \mid x)$ does not depend on one of the X variables, the $\hat{f}_e$ can converge at the $n^{-2/(q+4)}$ rate of the conditional mean. In this case the two-step estimator actually has an improved rate of convergence relative to the conventional estimator.

Two-step estimators of this form are often employed in practical applications, but do not seem to have been discussed much in the theoretical literature.

We could also consider a 3-step estimator, based on the expressions

$$\begin{aligned} y &= g(X) + e \\ e^2 &= \sigma^2(X) + \eta \\ \eta &\mid x \sim f_\eta(\eta \mid x) \end{aligned}$$

$$f(y \mid x) = f_\eta\left(\frac{y - g(x)}{\sigma(x)} \mid x\right)$$

The 3-step estimator is: First $\hat{g}(x)$ by nonparametric regression, obtain residuals $\hat{e}_i$. Second, $\hat{\sigma}^2(x)$ by nonparametric regression using $\hat{e}_i^2$ as dependent variable. Obtain rescaled residuals $\hat{\eta}_i = \hat{e}_i / \hat{\sigma}(X_i)$. Third, $\hat{f}_\eta(\eta \mid x)$ as the nonparametric conditional density estimator for $\hat{\eta}_i$. Then we can set

$$\hat{f}(y \mid x) = \hat{f}_\eta\left(\frac{y - \hat{g}(x)}{\hat{\sigma}(x)} \mid x\right)$$

In cases of strong variance effects (such as in financial data) this method may be desireable.

As the variance estimator $\hat{\sigma}^2(x)$ converges at the same rate as the mean $\hat{g}(x)$, the same first-order properties apply to the 3-step estimator as to the 2-step estimator. Namely, f_η should have reduced dependence on x, so it should be relatively well estimated even with large x-bandwidths, resulting in reduced MSE relative to the 1-step and 2-step estimators.

Given these insights, it might seem sensible to apply the 2-step or 3-step idea to conditional distribution estimation. Unfortunately the analysis is not quite as simple. In this setting, the nonparametric conditional mean, conditional variance, and conditional distribution estimators all converge at the same rates. Thus the distribution of the estimate of the CDF of e_i depends on the fact that it is a 2-step estimator, and it is not immediately obvious how this affects the asymptotic distribution. I have not seen an investigation of this issue.

6 Conditional Quantile Estimation

6.1 Quantiles

Suppose Y is univariate with distribution F.

If F is continuous and strictly increasing then its inverse function is uniquely defined. In this case the α'th quantile of Y is

$$q_\alpha = F^{-1}(\alpha).$$

If F is not strictly increasing then the inverse function is not well defined and thus quantiles are not unique but are interval-valued. To allow for this case it is conventional to simply define the quantile as the lower bound of this endpoint. Thus the general definition of the α'th quantile is

$$q_\alpha = \inf \{y : F(y) \geq \alpha\}.$$

Quantiles are functions from probabilities to the sample space, and monotonically increasing in α.

Multivariate quantiles are not well defined. Thus quantiles are used for univariate and conditional settings.

If you know a distribution function F then you know the quantile function q_α. If you have an estimate $\hat{F}(y)$ of $F(y)$ then you can define the estimate

$$\hat{q}_\alpha = \inf \left\{y : \hat{F}(y) \geq \alpha\right\}$$

If $\hat{F}(y)$ is monotonic in y then $\hat{q}_\alpha$ will also be monotonic in α. When a smoothed estimator $\hat{F}(y)$ is used, then we can write the quantile estimator more simply as $\hat{q}_\alpha = \hat{F}^{-1}(\alpha)$.

Suppose that $\hat{F}(y)$ is the (unsmoothed) EDF from a sample of size n. In this case, $\hat{q}_\alpha$ equals $Y_{([n\alpha])}$, the $[\alpha n]$'th order statistic from the sample. If αn is not an integer, $[\alpha n]$ is the greatest integer less than αn. We could also view the interval $\left[Y_{([n\alpha])}, Y_{([n\alpha]+1)}\right]$ as the quantile estimate. We ignore these distinctions in practice.

When $\hat{F}(y)$ is the EDF we can also write the quantile estimator as

$$\hat{q}_\alpha = \underset{q}{\operatorname{argmin}} \sum_{i=1}^{n} \rho_\alpha (Y_i - q)$$

where

$$\begin{aligned} \rho_\alpha(u) &= u\left[\alpha - 1(u \leq 0)\right] \\ &= \begin{cases} -(1-\alpha)u & u < 0 \\ \alpha u & u \geq 0 \end{cases} \end{aligned}$$

is called the "check function".

6.2 Conditional Quantiles

If the conditional distribution of Y given $X = x$ is $F(y \mid x)$ then the conditional quantile of Y given $X = x$ is

$$\begin{aligned} q_\alpha &= \inf\{y : F(y \mid x) \geq \alpha\} \\ &= F^{-1}(\alpha \mid x) \end{aligned}$$

Conditional quantiles are functions from probabilities to the sample space, for a fixed value of the conditioning variables.

One method for nonparametric conditional quantile estimation is to invert an estimated distribution function. Take an estimate $\hat{F}(y \mid x)$ of $F(y \mid x)$. Then we can define

$$\hat{q}_\alpha(x) = \inf\left\{y : \hat{F}(y \mid x) \geq \alpha\right\}$$

When $\hat{F}(y \mid x)$ is smooth in y we can write this as $\hat{q}_\alpha(x) = \hat{F}^{-1}(\alpha \mid x)$.

This method is particularly appropriate for inversion of the smoothed CDF estimators $\tilde{F}(y \mid x)$.

This inversion method requires that $\hat{F}(y \mid x)$ be a distribution function (that it lies in $[0,1]$ and is monotonic), which is not ensured if $\hat{F}(y \mid x)$ is computed by LL. The NW, WNW and smoothed versions are all appropriate. When $\hat{F}(y \mid x)$ is a distribution function then $\hat{q}_\alpha(x)$ will satisfy the properties of a quantile function.

6.3 Check Function Approach

Another estimation method is to define a weighted check function. This can be done using either a locally constant or locally linear specification.

The locally constant (NW) method uses the criterion

$$S_\alpha(q \mid x) = \sum_{i=1}^{n} K\left(H^{-1}(X_i - x)\right) \rho_\alpha(Y_i - q)$$

It is a locally weighted the check function, for observations "close to" $X_i = x$. The nonparametric quantile estimator is

$$\hat{q}_\alpha(x) = \underset{q}{\operatorname{argmin}}\, S_\alpha(q \mid x).$$

The local linear (LL) criterion is

$$S_\alpha(q, \beta \mid x) = \sum_{i=1}^{n} K\left(H^{-1}(X_i - x)\right) \rho_\alpha\left(Y_i - q - (X_i - x)'\beta\right).$$

The estimator is

$$\left\{\hat{q}_\alpha(x), \hat{\beta}_\alpha(x)\right\} = \underset{q,\beta}{\operatorname{argmin}}\, S_\alpha(q, \beta \mid x).$$

The conditional quantile estimator is $\hat{q}_\alpha(x)$, with derivative estimate $\hat{\beta}_\alpha(x)$. Numerically, these problems are identical to weighted linear quantile regression.

6.4 Asymptotic Distribution

The asymptotic distributions of the quantile estimators are scaled versions of the asymptotic distributions of the CDF estimators (see the Li-Racine text for details).

The CDF inversion method and the check function method have the same asymptotic distributions.

The asymptotic bias of the quantile estimators depends on whether a local constant or local linear method was used, and whether smoothing in the y direction is used.

6.5 Bandwidth Selection

Optimal bandwidth selection for nonparametric quantile regression is less well studied than the other methods.

As the asymptotic distributions seem to be scaled versions of the CDF estimators, and the quantile estimator can be viewed as a by-product of CDF estimation, it seems reasonable to select bandwidths by a method optimal for CDF estimation, e.g. cross-validation for conditional distribution function estimation.