On Non-parametric Function Fitting

Peter M. Robinson

London School of Economics

Commemoration Day

In Honour of Maurice Priestley, 18 December 2013

Time Series: 1940's-1960's:

Bartlett

Tukey

Anderson

Grenander

Rosenblatt

Hannan

Durbin

Akaike

Parzen

Priestley

Nonparametric spectral density estimation was a major theme.

E.g. lag window kernel estimate of spectral density $s(\lambda)$ of short memory time series x_i, $i = 1, ..., n$:

$$\widehat{s}(\lambda) = \frac{1}{2\pi} \sum_{u=1-n}^{n-1} k\,(hu)\, c\,(u) \cos\,(u\lambda)\,,$$

where $c\,(u)$ is lag-u sample autocovariance of x_i, kernel $k\,(u)$ satisfies

$$k\,(0) = 1$$

and bandwidth h satisfies $h \to 0$, $nh \to \infty$ as $n \to \infty$.

Choice of k was a popular topic, maybe more than h.

Consistency, MSE and CLT were discussed.

e.g.

M Priestley (1962) Basic considerations in the estimtion of spectra. Technometrics

M Priestley (1962) The role of bandwidth in spectral analysis. Appl Statist.

focussed on relatively practical issues.

More theoretical contributions included

E Parzen (1957) On consistent estimates of the spectrum of a stationary time series. Ann. Math. Statist.

E Parzen (1958) On asymptotically efficient consistent estimates of the spectral density function of a stationary time series. JRSSB.

Asymptotics used local smoothness and properties of kernel to get MSE rate that is slower than parametric, but can get abitrarily close to parametric rate.

Much more recently, for long memory time series, where

$$s(\lambda) \sim \lambda^{-2d}, \text{ as } \lambda \to 0_+, \ \ |d| < 1/2,$$

analogous theory was developed for estimating d under analogous local smoothness conditions on ratio $s(\lambda)/\lambda^{-2d}$.

Also extensions to higher-order spectra etc.

Interestingly nonparametric spectral density estimation largely developed before, and influenced, nonparametric probability density and regression etsimation.

e.g Parzen modified his earlier spectral work in

E Parzen (1962) On estimation of a probability density function and mode. Ann. Math. Statist.

Kernel estimate of pdf $f(x)$ of x_i given iid sequence $x_i = 1, ..., n$:

$$\widehat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right),$$

where the kernel $K(u)$ satisfies

$$\int K(u)\, du = 1$$

and may be thought of as the Fourier inverse of the previous lag window kernel k used in spectral estimation, and bandwidth h is essentially as before, so satisfies $h \to 0$, $nh \to \infty$ as $n \to \infty$.

Vast literature on these (and other) nonparametric probability density estimates, with many consistency, MSE and CLT results, using similar ideas to those used in the spectral theory, but with otherwise quite different proofs.

Choice of bandwidth h, eg by rule of thumb, cross-validation etc has been a much bigger theme than in the spectral analysis literature.

Much of the methods and theory for pdf estimation extend relatively straightforwardly to estimating $m(x)$ in stochastic-design nonparametric regression

$$y_i = m(x_i) + e_i, \; i = 1, ..., n,$$

where most simply x_i and e_i, and thus y_i, form iid sequences.

E.g. Nadaraya-Watson estimate

$$\widehat{m}(x) = \frac{\sum_{i=1}^{n} y_i K\left(\frac{x - x_i}{h}\right)}{\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right)}.$$

Much of pdf and regression theory has been extended to short memory time series case, e.g. Markov, mixing series.

Due to local properties of estimate, first-order asymptotic results (i.e. CLT, MSE) are the same as in iid case.

But with sufficient degree of long memory, quite different asymptotics results.

Also extensions to nonstationary series, eg unit roots.

Fixed-design nonparametric regression model is

$$y_i = m\left(\frac{i}{n}\right) + e_i,\ i = 1, ..., n,$$

where $m(x)$ is now defined on $(0, 1)$ and most simply the e_i are iid.

Note that the $y_i = y_{in}$ form a triangular array.

The division by n in $m\left(\frac{i}{n}\right)$ is needed to ensure sufficient accumulation of information to achieve consistent estimation.

M Priestley and M Chao (1972) On non-parametric function fitting. JRSSB.

Kernel estimate

$$\widehat{m}(x) = \frac{1}{nh}\sum_{i=1}^{n} y_i K\left(\frac{x - i/n}{h}\right).$$

where $K(x)$ and h are as in pdf and stochastic design regression.

Essentially formalizes old idea of a moving mean, e.g. take $K(x)$ to be uniform kernel on $(-1, 1)$

An alternative is

$$\widehat{m}(x) = \frac{\sum_{i=1}^{n} y_i K\left(\frac{x-i/n}{h}\right)}{\sum_{i=1}^{n} K\left(\frac{x-i/n}{h}\right)}.$$

.

Consistency, MSE, CLT results established, analogous to those for spectrum, pdf and stochastic-design regression estimates.

Though the model assumes an ordering of data y_i,

$$y_i = m\left(\frac{i}{n}\right) + e_i,\ i = 1, ..., n,$$

is not fully a 'time series model' if e_i are iid.

Extension to short memory e_i : same rate of convergence as in iid case, but asymptotic variance different.

Extension to short memory e_i : rate of convergence slower.

Nonstationary time series extension: e.g.

$$(1 - L)^d y_i = m\left(\frac{i}{n}\right) + e_i,\ i = 1, ..., n,$$

where nonstationary values of d, i.e. $d > 1/2$, are possible.