Nonparametric Regression with Random Sampling

E. K. CHICKEN
Department of Statistics, Florida State University, Tallahassee, FL 32306-4303, USA

Abstract

Wavelet methods have been applied to nonparametric regression problems with much success. Usually the Xn are equispaced values. When the number of data points is dyadic, the discrete wavelet transform (DWT) works especially well, providing fast convergence to the unknown underlying function in a computationally efficient manner. This paper presents a method to be applied to the problem when the data are randomly placed in an interval or follow a random process. Specific conditions on the distributions and processes are given that enable the DWT through the use of block thresholding, to come within a constant factor of the optimal rate of convergence.

1 Introduction

When the positions of the sampling points are equipped, wavelets excel in the areas of spatial adaptivity, optimality, and low computational cost. For fixed, non-equispaced sample points, Cai and Brown (1998) investigated wavelet methods via an approximation approach. Hall and Tschern (1997) used simulation to deal with samples with random design. Unfortunately, these methods are more complex from a computational standpoint than their equispaced counterparts.

To maintain computational efficiency, it is desirable to use equispaced methods to analyze the data. Cai and Brown (1999) have examined convergence rates when the positions of the sample points are distributed as independent uniform random variables. Using term-by-term thresholding, they showed that equiparametric methods will be directly applied to the non-equipared data without a loss in the rate of convergence, i.e. to within a logarithmic factor of the optimal minimax convergence rates found by Donoho and Johnstone (1994). This method maintains the computational efficiency and simplicity of the equispaced approach.

Here, the results of Cai and Brown (1999) for nonparametric regression with randomly placed sampling points are improved upon.

1. By using the block thresholding of wavelet coefficients method proposed by Hall et al. (1999) and refined by Cai (1999), the logarithmic penalty in the convergence rate associated with term-by-term thresholding has been removed.

2. The scope of the function space has been enlarged to include not only the Holder functions but many Besov spaces as well.

3. Moment conditions are then defined for a class of random distributions for which these results hold.

4. This fast convergence rate is also attained when the sample points come from a random process. Conditions are given defining the processes for which this method will result in optimal convergence rates.

5. Spatial adaptivity is maintained, and the computational cost remains low since the equispaced algorithm is used.

2 Model

The usual model for nonparametric regression is

\[ y_i = f(x_i) + \sigma \varepsilon_i, \quad i = 1, \ldots, n, \]

where the errors \( \varepsilon_i \) are iid standard normal random variables and \( \sigma \) is known. To use the DWT, we require that \( n = 2^m \). We will modify this to the more general problem

\[ y_i = f(x_i) + \sigma \varepsilon_i, \quad i = 1, \ldots, N. \]

There are two schemes:

1. If the sample points \( X_i \) are randomly placed in the interval, then \( N \) is a constant and we may assume it to be dyadic. The \( X_i \) in this case are iid.

2. If the \( X_i \) follow a random process, then \( N \) is a random variable.

In each case, we assume the distribution of the \( X_i \) (and hence \( N \) in the random process case) are independent of the noise \( \varepsilon_i \).

3 Wavelets and the Estimator

\( f \) and \( \varepsilon \) will represent the father and mother wavelets, respectively. Both are assumed to be compactly supported and periodicized to the interval \( [0, 1] \). The wavelet coefficients of a function \( f \) are the usual inner product\( \langle f; \phi_j \rangle \) in terms of resolution, the \( \phi_j \) coefficients represent the coarsest, smoothed portions of \( f \) and the \( \psi_j \) and \( \psi_{j+k} \) are the coefficients representing the detailed structure of \( f \). Rather than compute the coefficients in the above manner \( f \) is unknown, we see the DWT to estimate them.

By ordering the sample points and performing the equivalent rounding of the \( y_i \)’s, the sample data now looks like

\[ y_i = \langle f; \phi_j \rangle + \sigma \psi_j. \]

where the \( X_j \)’s are either the ordered random variables or the arrival times of the random process. The \( (X_j, y_j) \)’s are random variables and the \( (X_j, \varepsilon_j) \)’s are random variables with zero mean. Using the above reconstruction, and the observed data vector is now

\[ (X_j, y_j) \rightarrow (X_j, \varepsilon_j) \rightarrow (X_j, \varepsilon_j) \rightarrow (X_j, \psi_j). \]

This randomly spaced data points will be treated as equivalent in the wavelet algorithm. Each \( X_j \) will be replaced with \( X_j(\psi_j) \).

Provided that \( \varepsilon_j \) is redundant. These spaces cover a very wide range irregular functions. In simulation, the following functions from Donoho and Johnstone (1994) will be used.

\[ \text{non-equipared wavelet estimator, } \text{V to the term-by-term estimator, } \text{U to the uniform } \]

4 Function Spaces

The functions to be studied reside in Besov and Holder spaces. These may be defined by their wavelet coefficients as in Meyer (1990). If \( f \) and \( \psi \) are wavelet coefficients of \( f \), then \( f = S(\psi) \) for some function \( S \).

5 Rates of Convergence

Denote and Johnstone (1994) showed that the best rate (in terms of minimum) for a diagonal projection estimator such as the one presented here

\[ (f) = \frac{\langle f; \phi_j \rangle}{(f; \phi_j)} \]

where \( x \) is the number of places the function is sampled, and \( \sigma \) is the smoothness parameter of the function space. Using term-by-term thresholding on equispaced samples, the DWT estimate is within a log of the optimal rate. By applying block thresholding, this rate of convergence is improved.

Theorem 1

Let \( X_i \) be iid positive random variables with \( \langle EX_i \rangle = 1/n \), var \( X_i \leq c/n^2 \), and \( X_i \rightarrow X \).

Suppose a sample \( (X_1, y_1, \ldots, X_n, y_n) \) is collected with \( y_i = f(x_i) + \sigma \varepsilon_i \).

Further, let \( h \) be a vanishing moment wavelet. \( S \) and \( U \) are independent.

\[ \text{Optional step: } \frac{\langle f; \phi_j \rangle}{\langle f; \phi_j \rangle} \]

where \( c \) is constructed above, and \( c \) is each interval.

Theorem 2

Let \( X_i \) be iid random variables on \([0, 1]\) whose density is bounded.

\[ \langle E X_i \rangle = 1/n \leq \sigma \langle f; \phi_j \rangle \]

and \( X_i \rightarrow X \).

Suppose a sample \( (X_1, y_1, \ldots, X_n, y_n) \) is collected with \( y_i = f(x_i) + \sigma \varepsilon_i \).

Further, \( h \) has \( r \) vanishing moments; \( \langle f; \phi_j \rangle \leq \sigma \langle f; \phi_j \rangle \) and \( c \); and \( c \leq \sigma \langle f; \phi_j \rangle \).

\[ \text{Optional step: } \frac{\langle f; \phi_j \rangle}{\langle f; \phi_j \rangle} \]

We see that the minimum rate is attained within a constant in each case. Note also that if the space of functions is just \( \langle f; \phi_j \rangle \), this exceeds the performance of previous estimators due to the removal of the log penalty term.

6 Simulations

6.1 Numerical Results - I

Here we specifically look at design points uniformly placed on an interval and design points following a Poisson process. "B" refers to the block thresholded wavelet estimator, "V" to the term-by-term estimator, "U" to the uniform process scheme.

7 Conclusions

- Want to use DWT because of its adaptivity and speed
- Previously shown to be within a log of minimum for uniform design
- Here, have shown to be within a constant of minimum for design points
- Randomly distributed on an interval (theorem 2)
- Following a random process (theorem 1)
- Results hold on more general space of functions than previously considered

References