The Korean Communications
in Statistics Vol. 7, No. 1, 2000
pp. 129-139

A Review on Nonparametric Density Estimation Using Wavelet Methods[1]

Sungho Lee[2] and Hwa Rak Lee[3]

Abstract

Wavelets constitute a new orthogonal system which has direct application in density estimation. We introduce a brief wavelet density estimation and summarize some asymptotic results. An application to mixture normal distributions is implemented with S-Plus.

1. Introduction

The subject of nonparametric probability density estimation has generated a vast area of research and challenging problems. In addition to the traditional histogram method there are methods based on kernels(Parzen(1962), Rosenblatt(1956)), on Fourier series(Kronmal and Tarter(1968)), on Fourier transforms(Davis(1977)), on orthogonal polynomials (Schwartz(1967)), on splines(Wahba(1975)) and on general delta sequences (Foldes and Revez(1974)). The performances of all these procedures depend strongly on the choice of a smoothing parameter or bandwidth. This choice is in fact by no means an easy task. Different approaches have been considered, generally corresponding to some optimal solution of some well-posed problem. As an example, if the regularity class of the estimated function is assumed to be known, then it is possible to choose the bandwidth so that the estimate attains the minimax rate. Of course, from a practical point of view, this is not entirely satisfactory since it requires some extra knowledge. Various attempts have also been investigated to reduce this knowledge. Among these, the recent appearance of explicit orthonormal bases based on multiresolution analysis has given different opportunities to solve the problem. Indeed, unlike traditional Fourier bases, wavelet bases, since they have localization properties in space as well as in frequency, enable expansions of a function into coefficients which are reliable indicators of its regularity.

Wavelet methods have been introduced to statistics by Donoho(1992), Donoho and Johnstone(1992, 1994a, b) and Kerkyacharian and Picard(1992, 1993). These authors have demonstrated the virtues of wavelet methods from the viewpoint of adaptive smoothing,

---

1) This research was supported (in part) by the Taegu University Research grant, 1999.
2) Professor, Department of Statistics, Taegu University, Korea
3) Graduate student, Department of Statistics,Taegu University,Kyungsan,712-714,Korea

typically in the context of the achievability of very good convergence rates uniformly over exceptionally large function classes. The first use of wavelet bases for density estimation appears in papers by Doukan and Leon(1990), Kerkyacharian and Picard(1992) and Walter(1992).

2. Multiresolution analysis and Besov space

2.1 Multiresolution analysis and wavelets

Let us recall(cf. Meyer(1990)) that one can construct a function φ such that :

(1) The sequence $\{\varphi(x-k), k\in Z\}$ is an orthonormal family of $L^2(R)$. Let us call V_0 the subspace spanned by this sequence.

(2) $\forall j\in Z,\ V_j\subset V_{j+1}$ if V_j denotes the subspaces spanned by $\{\varphi_{j,k},\ k\in Z\}$, $\varphi_{j,k}(x)=2^{j/2}\varphi(2^jx-k)$.

Then we have $\bigcap_{j\in z} V_j=\{0\}$ and, furthermore, if $\int\varphi=1,\ L^2(R)=\overline{\bigcup_{j\in z} V_j}$.

It is possible to require in addition that one of the following conditions holds :

(a) φ is of class C^r, φ and all its derivatives up to the order r are rapidly decreasing. One says in this case that the analysis is r-regular.

(b) $\varphi\in\mathfrak{I}$ (Meyer's wavelet).

(c) φ is of class C^r compactly supported(e.g.Daubechies's wavelet; see Daubechies (1988)).

Under these conditions, let us define the space W_j by the following :

$V_{j+1}=V_j\oplus W_j$.

There exists a function ψ(the "wavelet") such that

- $\{\psi(x-k),\ k\in Z\}$ is an orthonormal basis of W_0,
- ψ has the same regularity properties as φ(i.e. (a), (b) or (c)),
- the family $\{\psi_{j,k},\ k\in Z,\ j\in Z\}$ is an orthonormal basis of $L^2(R)$,where $\psi_{j,k}(x)=2^{j/2}\psi(2^jx-k)$.

The following decomposition is also true :

$L^2(R)=V_{j_0}\oplus W_{j_0}\oplus W_{j_0+1}\oplus\cdots$

That is, $\forall f\in L^2(R),\ f=\sum_{k=Z}\alpha_{j_0,k}\varphi_{j_0,k}+\sum_{j\geq j_0}\sum_{k\in Z}\beta_{j,k}\psi_{j,k}$

where $\alpha_{j_0,k}=\int f(x)\overline{\varphi_{j_0,k}}(x)dx,\ \ \beta_{j,k}=\int f(x)\overline{\psi_{j,k}}(x)dx,\ \ j\geq j_0.$

2.2 Besov spaces

These spaces rather usual in analysis(cf. Bergh and Lostrom(1976)) appear to be easily characterizable in terms of multiresolution analysis.

Definition. For $0<s<1$, $1\le p, q\le +\infty$, set

$$\gamma_{s,p,q}(f) = \left(\int_R \left(\frac{\|\tau_h f - f\|_p}{|h|^s}\right)^q \frac{dh}{|h|}\right)^{1/q},$$

$$\gamma_{s,p,+\infty}(f) = \sup_{h\in R} \frac{\|\tau_h f - f\|_p}{|h|^s}, \qquad \text{where } \tau_h f(x) = f(x-h),$$

In the case s=1, set

$$\gamma_{1,p,q}(f) = \left(\int_R \left(\frac{\|\tau_h f + \tau_{-h} f - 2f\|_p}{|h|}\right)^q \frac{dh}{|h|}\right)^{1/q}, \quad \gamma_{1,p,+\infty}(f) = \sup_{h\in R} \frac{\|\tau_h f + \tau_{-h} f - 2f\|_p}{|h|}.$$

For $0<s\le 1$ and $1\le p, q\le +\infty$, set $B_{s,p,q} = \{f\in L_p \mid \gamma_{s,p,q} < +\infty\}$.

This space is equipped with the norm $\|f\|_{s,p,q} = \|f\|_p + \gamma_{s,p,q}(f)$.

For $s>1$, s writes $n+\alpha$ with $n\in N$ and $0<\alpha\le 1$ and $f\in B_{s,p,q} \Leftrightarrow f^{(m)} \in B_{\alpha,p,q}$ $\forall m\le n$ where $f^{(m)}$ denotes the m-derivative of f. This space is equipped with the norm, $\|f\|_p + \sum_{m\le n} \gamma_{\alpha,p,q}(f^{(m)}) = \|f\|_{s,p,q}$.

Examples. Sobolev space $H_2^s = B_{s,2,2}$, $B_{s,\infty,\infty}$ is the set of bounded, Holderian functions of order s, $B_{1,\infty,\infty}$ is the Zygmund's class.

These spaces may be also characterized in the following way(cf. Meyer(1992)).

Theorem 2.1. If conditions (1) and (a) hold for φ, with $r>s$, then

$$f\in B_{s,p,q} \Leftrightarrow J_{s,p,q}(f) = \|E_0 f\|_{L^p(R)} + \left(\sum_{j\ge 0} (\|D_j f\|_{L^p(R)} 2^{js})^q\right)^{1/q} < \infty$$

(with the usual modification for $q=+\infty$).

Remarks. (1) The norm $J_{s,p,q}$ is equivalent to $\|\ \ \|_{s,p,q}$.

(2) This characterization doesn't depend on the functions φ provided that φ is r-regular with $r>s$.

(3) Using , $E_0 f = \sum_k \alpha_{0,k}\varphi_{0,k}$ $D_j f = \sum_k \beta_{j,k}\psi_{j,k}$, and $\|\beta_j\|_{l^p} = (\sum_k |\beta_{j,k}|^p)^{1/p}$,

the norm $J'_{spq}(f) = \|\alpha_{0.}\|_{l^p} + [\sum_j (2^{j(s+1/2-1/p)} \|\beta_{j.}\|_{l^p})^q]^{1/q}$ is again equivalent to

$J_{spq}(f)$. This is exactly reducing Besov spaces to sequences spaces.

(4) $B_{s'pq'} \subset B_{spq}$ for $s'>s$, or $s'=s$ and $q'\le q$,

$B_{s'p'q} \subset B_{spq}$ for $p>p'$, $s'=s-\frac{1}{p}-\frac{1}{p'}$,

$B_{sp1} \subset H_p^s \subset B_{sp\infty}$.

3. A Review on some main results

In this section we review and discuss some main results. From now on, let us take a function φ verifying (1) and (c) (in 2.1) and let $s>0$, $1\le p, q\le +\infty$. We define the class F depending on s, p, q and a linear estimator f^* as follows:

$$F_{s,p,q}=\{f,\ f\ge 0,\ \int f=1,\ J_{spq}(f)\le M\},\qquad f^*(x)=\sum_{k\in Z}\hat{a}_{j(n),k}\,\varphi_{j(n),k}(x).$$

where $\hat{a}_{j(n),k}=\frac{1}{n}\sum_{i=1}^{n}\varphi_{j(n),k}(X_i)$.

Theorem 3.1. (Kerkyacharian and Picard(1992))

(1) Let $s>0$, $1\le p<+\infty$, $1\le q\le +\infty$ and $s<r$, where r is the regularity of MRA. If d is taken to be either $d_{s'pq'}$, $0<s'<s$, $q'\ge q$ or $d^*_{s'p}$, $0\le s'<s$, s' integer, then for $p\ge 2$, $\exists C_1$ constant such that

$$\forall f\in F_{s,p,q},\ E_f d(f^*,f)\le C_1 n^{(s'-s)p/(1+2s)} \tag{1}$$

where $2^{j(n)}=n^{1/(1+2s)}$, $d_{s',p,q'}(f,g)=(J_{s',p,q}(f-g))^p$, $d^*_{s',p}(f,g)=\|f^{(s')}-g^{(s')}\|_p^p$

(2) Let $R_n=\inf_{\hat{f}}\sup_{f\in F_{spq}} E_f d(\hat{f},f)$. Then for M large enough, for $1\le p<\infty$, there exists a constant C_2 such that $R_n\ge C_2 n^{(s'-s)p/(1+2s)}$.

Remarks. (1) Theorem3.1 indicates that for $p\ge 2$, $E_f\|f^*-f\|_p^p\le c n^{-sp/(1+2s)}$.

(2) For $1\le p<2$, equation(1) is still true if $\exists a\in R$, ω symmetric function of $L^{p/2}(R)$ not increasing on R^+ such that $f(x-a)\le\omega(x)$, $\forall x\in R$.

(3) $2^{-j(n)}$ plays the role of the usual window $h(n)$.

The estimator in Theorem3.1 is linear. If the density has spatially varying properties, the linear estimators can be under-performing. When the density belongs to B_{spq} and errors are measured with L_π norm($1\le\pi<\infty$, $\pi>p$), non-linearity of the estimator becomes essential since the linear estimators are sub-optimal. Hence the unknown density needs to be approximated with some details. That is,

$$\hat{f}=\sum_{k\in Z}\hat{a}_{j_0k}\,\varphi_{j_0k}+\sum_{j=j_0}^{j_1}\sum_{k\in Z}\widehat{\beta_{jk}}\,\psi_{jk}.$$

Donoho et al.(1996) suggested a thresholded estimator obtained by truncating the wavelet series and then thresholding the smallest empirical coefficients. They suggested thresholded wavelet coefficients $\hat{\beta}_{jk}$ using hard or soft threshold techniques. Given a wavelet coefficient $\hat{\beta}_{jk}$ and a threshold $t>0$ the hard threshold value is given by $T_{hard}(\hat{\beta}_{jk}t)=\hat{\beta}_{jk}I\{|\hat{\beta}_{jk}|>t\}$

and the soft threshold value is given by $T_{soft}(\hat{\beta}_{jk,t}) = sign(\hat{\beta}_{jk})(\hat{\beta}_{jk} - t)I\{|\hat{\beta}_{jk}| \geq t\}$,where I is the usual indicator function, $\hat{\alpha}_{j_0k} = \frac{1}{n}\sum_{i=1}^{n}\varphi_{j_0k}(X_i)$ and $\hat{\beta}_{jk} = \frac{1}{n}\sum_{i=1}^{n}\psi_{jk}(X_i)$. From simulation experience, one expects that soft thresholding will better suppress noise artifacts, while hard thresholding will better preserve the visual appearance of peaks and jumps. Donoho et al.(1996) also indicated that the same form, based on simple thresholding of the wavelet coefficients, achieves nearly optimal performance in terms of rate of convergence over a variety of global error measure and over a variety of function spaces. They defined :

$$\hat{\beta}_{jkt} = \hat{\beta}_{jk} I\{|\hat{\beta}_{jk}| > KC(j)n^{-1/2}\} \quad \text{and} \quad \hat{f} = \sum_{k\in Z}\hat{\alpha}_{j_0k}\phi_{j_0k} + \sum_{j=j_0}^{j_1}\sum_{k\in Z}\hat{\beta}_{jk}\psi_{jkt}.$$

Theorem 3.2. (Donoho,Johnstone, Kerkyacharian, and Picard(1996))
Let $s-1/p>0$ and $p\wedge 1 \leq p' \leq \infty$. If $C(j)=\sqrt{j}$, there exist constants $C_5 = C_5(s,p,q,M)$ and $K_0 = K_0(s,p,p';M)$ such that if

$$2^{j_0(n)} \simeq (n(\log n)^{[(p'-p)/p]I\{\varepsilon\geq 0\}})^{1-2\alpha}$$
$$2^{j_1(n)} \simeq (n/\log n)^{\alpha/s'}$$

and $K \geq K_0$, then

$$\sup_{f\in F_{spq}(T)} (E_f\|\hat{f}-f\|_{p'}^{p'})^{1/p'} \leq \begin{cases} C_5(\log n)^{(1-\varepsilon/sp)\alpha}n^{-\alpha}, & \varepsilon > 0, \\ C_5(\log n)^{(1/2-p/qp')_+}(\frac{\log n}{n})^{\alpha}, & \varepsilon = 0, \\ C_5(\frac{\log n}{n})^{\alpha}, & \varepsilon < 0, \end{cases}$$

where $x_+ = \max(x,0)$, $F_{spq}(T) = \{f : f\in F_{spq},\ supp(f)\subset[-T,T]\}$

$$\alpha = \min(\frac{s}{1+2s}, \frac{s-1/p+1/p'}{1+2s-2/p}),\quad \varepsilon = sp - \frac{p'-p}{2},\quad s' = s - \frac{1}{p} + \frac{1}{p'}.$$

Remarks. (1) When $p \geq p'$, optional rate is $n^{-s/(1+2s)}$ and linear estimators attain the optional rate.
(2) When $p < p'$, we have $s' = s - 1/p + 1/p' < s$ and the convergence rate of linear estimators $n^{-s'/(1+2s')}$ is slower than $n^{-s/(1+2s)}$.

Kerkyacharian et al.(1996) considered the global level j in stead of thresholding each coefficient. This different point of view has advantages as well as drawback. But a practical aspect of this method seems to be that one does a good job for a reasonable amount of data.

Let $\hat{\eta}_j^H = I\{\hat{\Theta}_j \geq 2^j/n^{p/2}\}$, $\hat{\eta}_j^S = \frac{\hat{\Theta}_j - 2^j/n^{p/2}}{\hat{\Theta}_j} I\{\hat{\Theta}_j \geq 2^j/n^{p/2}\}$, where $\widehat{\Theta}_j$ is defined as follows: For a even integer p

$$\widehat{\Theta}_j(p) = \binom{n}{p} C^{-1} \sum_{(i_1,\cdots,i_p)\in S_p} \sum_k \psi_{jk}(X_{i1}) \cdots \psi_{jk}(X_{ip})$$

where S_p is the set of p-dimensional vector of $\{1,\ldots,n\}^p$ such that all the coordinates are different. For $p = \alpha p_1 + (1-\alpha)p_2$, where $\alpha \in (0,1)$ and p_1, p_2 are even integers, $\widehat{\Theta}_j = (\widehat{\Theta}_j(p_1))^{\alpha}(\widehat{\Theta}_j(p_2))^{1-\alpha}$.

Theorem 3.3. (Kerkyacharian, Picard and Tribouley (1996)).

Let $p \geq 2$, Let $\hat{f} = \sum_k \hat{\alpha}_{j_0k}\varphi_{j_0k} + \sum_{j=j_0}^{j_1} \hat{\eta}_j \sum_k \hat{\beta}_{j_0k}\psi_{j_0k}$,

where $j_0 = 0$, $j_1 = \log_2(n)$ and $\hat{\eta}_j$ is either $\hat{\eta}_j^H$ or $\hat{\eta}_j^S$.

Then, for $s \in (1/p, r+1)$, $q \in [1, +\infty]$, there exists a constant C such that

$$\sup_{f \in F_{spq}(B)} E_f \| \hat{f}_{j_0,j_1} - f \|_p^p \leq C\, n^{-ps/(1+2s)}$$

Remark. The above Theorem 3.3 can be understood as follows: if for s known, $\sup_{f\in F_{spq}(B)} E_f \| \hat{f}_{j_0,j_1} - \hat{f}_L \|_p^p \leq C n^{-sp/(1+2s)}$, $\hat{f}_L = \sum_k \hat{\alpha}_{j_0k}\varphi_{j_0k} + \sum_{j=j_0}^{j_s}\sum_k \hat{\beta}_{jk}\psi_{jk}$, $2^{j_s} = n^{1/1+2s}$.

Unlike the above works, Hall and Patil (1995) discuss the performance of wavelet-based density estimators for a fixed density, rather than for a very large number of candidates for the density. The estimators that they employ are different from those suggested by Donoho et al. (1996) and other papers. Suppose f admit the wavelet expansion(cf. Meyer(1992))

$$f = \sum_{-\infty<j<\infty} b_j \varphi_j + \sum_{i=0}^{\infty} \sum_{-\infty<j<\infty} b_{ij}\psi_{ij}$$

where $\varphi_j(x) = p^{\frac{1}{2}}\varphi(px - j)$, $\psi_{ij}(x) = p_i^{\frac{1}{2}}\psi(p_i x - j)$, $\log_2 p = \log_2 p(n)$ is the most coarse resolution level of the fitted wavelet, $p_i = 2^i p$, $b_j = \int f\varphi_j$, and $b_{ij} = \int f\psi_{ij}$. Unbiased estimators of b_j and b_{ij} are given by $\hat{b}_j = n^{-1}\sum_{m=1}^{n}\varphi_j(X_m)$ and $\hat{b}_{ij} = n^{-1}\sum_{m=1}^{n}\psi_{ij}(X_m)$. A nonlinear wavelet estimator of f has the form

$$\hat{f}(x) = \sum_j \hat{b}_j\varphi_j(x) + \sum_{i=0}^{q-1}\sum_j \hat{b}_{ij} I(|\hat{b}_{ij}| > \delta)\psi_{ij}(x)$$

where $\delta > 0$ is a threshold and $q \geq 1$ is an integer that is typically chosen so that $p2^q$ is close to n.

Theorem3.4. (Hall and Patil(1995))

Assume that $f^{(r)}$ exists in a piecewise sense and is bounded and piecewise continuous on $(-\infty,\infty)$ with finite and well-defined left-and right-hand limits, and monotone on $(-\infty,-u)$ and on (u,∞) for sufficiently large u. Assume that $p\rightarrow\infty$, $q\rightarrow\infty$, $(\sum_{i=0}^{q-1}p_i)\,\delta^2\rightarrow 0$, $(\sum_{i=0}^{q-1}p_i)^{1+2r}\delta^2\rightarrow\infty$ and $\delta\geq C(n^{-1}\log n)^{1/2}$, where $C>2\{r(\sup f)/(1+2r)\}^{1/2}$. Then if $(\sum_{i=0}^{q-1}p_i)^{1+2r}\delta^2\rightarrow\infty$,

$$\int E(\hat{f}-f)^2\sim n^{-1}p+c_1p^{-2r} \qquad (2)$$

Where "~" means that the ratio of the left and right hand side converges to 1 as $n\rightarrow\infty$.

Remarks. (1) p plays the role of the inverse of bandwidth.

(2) Equation (2) is asymptotically minimized by taking $p\sim const.\,n^{1/1+2r}$ and the minimum size of (2) is $const.\,n^{-2r/1+2r}$.

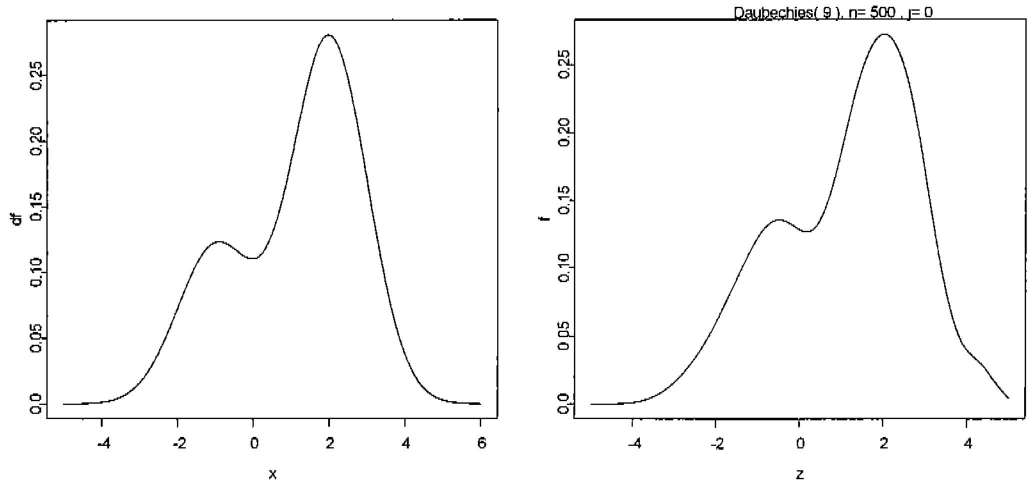
(3) Unlike the analogous situation for kernel density estimators, the failure of the smoothness condition at a finite number of points does not affect Theorem 3.4 (see Theorem 2.2 in Hall and Patil(1995)).

Standard wavelet-based density estimators may not retain non-negativity.

Pinheiro and Vidakovic (1997) estimated the square root of the density. But they did not formulate any asymptotic results in their paper. Penev and Dechevsky(1997) presented another estimator of the square root of the density, which enables us to control the positiveness and retain asymptotic minimax optimality result.

4. Simulation

In this section we show the performance of linear wavelet estimators using normal mixture densities. Notice that in applying wavelets an important choice is the wavelet family. Here Daubechies #9 wavelet function was used with S+ as an example and the program is appended for the reader convenience(S' wavelet can't be directly used for probability density function estimation). Figure 1 shows the linear wavelet estimator with Daub#9 for a sample of 500 observations simulated from 0.3N(-1,1)+0.7N(2,1)(For more simulation results and S+ program, see Lee(1999)).



(a) f(x)=0.3*N(-1,1)+0.7*N(2,1) (b) wavelet linear estimator

<figure 1>

<S$^+$ program>

```
daub.f<-function(n,m=8)
{
   h<-n.select(n)
   phi<-c(1,rep(0,length(h)-1))
   x<-0:(length(h)-1)
   for(j in 1:m) {
      y<-seq(0,by=2^(-j),length=length(phi)*2)
      phi.1<-rep(0,length(y))
      for(i in 1:length(phi)) {
         if(i<=n-1) {
            for(l in 1:i) {
               phi.1[2*i-1]<-phi.1[2*i-1]+sqrt(2)*h[2*l-1]*phi[i-l+1]
               phi.1[2*i]<-phi.1[2*i]+sqrt(2)*h[2*l]*phi[i-l+1]
            }
         }
         else {
            for(l in 1:(length(h)/2)) {
               phi.1[2*i-1]<-phi.1[2*i-1]+sqrt(2)*h[2*l-1]*phi[i-l+1]
               phi.1[2*i]<-phi.1[2*i]+sqrt(2)*h[2*l]*phi[i-l+1]
            }
```

```
            }
        }
        phi<-rep(0,length(y))
        x<-rep(0,length(y))
        phi<-phi.1
        x<-y
    }
    y<-seq((1-n),by=2^(-m),length=length(phi))
    shi<-rep(0,length(y))
    for (i in 1:length(shi)) {
        for(l in 1:length(h)) {
            if(x[1]<=2*y[i]+l-2 && 2*y[i]+l-2<=x[length(x)])
                shi[i]<-shi[i]+sqrt(2)*(-1)^l*h[l]*phi[x==2*y[i]+l-2]
        }
    }
    daub<-data.frame(x,phi,y,shi)
}

dens.f<-function(data,number=2,j=2)
{
    daub<-daub.f(number)
    low<-floor(min(data))
    hi<-ceiling(max(data))
    z<-seq(low,hi,2^(-8))
    low<-low*2^j
    hi<-hi*2^j
    c<-rep(0,(hi-low+1))
    k<-1
    for (l in low:hi) {
        for (i in 1:length(data)) {
            if(daub$x[1]<=2^j*data[i]-l && 2^j*data[i]-l<=daub$x [length(daub$x)])
                c[k]<-c[k]+2^(j/2)*daub$phi[2^j*data[i]-l-2^(-9)<daub$x &
daub$x<2^j*data[i]-l+2^(-9)]/ length(data)
        }
        k<-k+1
    }
    f<-rep(0,length(z))
    for (i in 1:length(z)) {
        k<-1
        for (l in low:hi) {
```

```
            if(daub$x[1]<=2^j*z[i]-1 && 2^j*z[i]-1<=daub$x[length(daub$x)])
               f[i]<-f[i]+c[k]*2^(j/2)*daub$phi[2^j*z[i]-1-2^(-9)<daub$x &
daub$x<2^j*z[i]-1+2^(-9)]
               k<-k+1
         }
      }
      plot(z,f,type="l")
      mtext(side=3,line=0.1,paste("Daubechies(",number,"), n=",length(data),", j=",j))
   }
```

References

[1] Bergh J. and J. Lofstrom(1976), *Interpolation Spaces : An introduction*, Springer, Berlin.

[2] Daubechies I.(1988). Orthogonal bases of compactly supported wavelets. *Comm. Pure Appl. Math.* 41, 909-996.

[3] Davis, K.(1977). Mean integrated squared error properties of density estimates. *Ann. Statist,* 5,530-535.

[4] Donoho, D.L.(1992). Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition. Technical Report 403, Dept. Statistics, Stanford Univ.

[5] Donoho, D.L. and Johnstone, I.M.(1992). Minimax estimation via wavelet shrinkage. Technical Report 402, Dept. Statistics, Stanford Univ.

[6] Donoho, D.L. and Johnstone, I.M.(1994a). Ideal spatial adaptation by wavelet shrinkage. *Biometrika* 81, 425-455.

[7] Donoho, D.L. and Johnstone, I.M.(1994b). Minimax risk over l_p-balls for l_q-error. *Probab. Theory Related Fields* 99, 277 - 303.

[8] Donoho, D.L. Johnstone, I.M. Kerkyacharian, G. and Picard, D.(1995) Wavelet shrinkage : asymptopia? (with discussion). *J. Roy Statist. Soc.* Ser. B. 57, 301-369.

[9] Donoho, D.L. Johnstone, I.M. Kerkyacharian, G. and Picard, D.(1996) Density estimation by wavelet thresholding. *Ann. Statist.* 24, 508-539.

[10] Doukhan, P. and Leon, J.(1990). Deviation quadratique d'estimateurs d'une densite par projection orthogonale.*C.R. Acad . Sci. Paris Ser.* I Math 310. 425-430.

[11] Foldes, A. and Revez, P.(1974). A general method of density estimation. *Studia Sci. Math. Hungar.* 9, 82-92.

[12] Hall, Peter and Patil, Prakash(1995). Formulae for mean integrated squared error of nonlinear wavelet-based density estimators. *Ann. Stat.* 23, 905-928.

[13] Härdle, W. et. al.(1998). *Wavelets, Approximation, and Statistical Applications.* Lecture Notes in Statistics 129, Springer-Verlag, New York.

[14] Kerkyacharian, G. and Picard, D.(1992). Density estimation in Besov spaces. *Statist. Probab. Lett.* 13, 15 - 24.

[15] Kerkyacharian, G. and Picard, D.(1993). Density estimation by kernel and wavelet methods: Optimality in Besov spaces. *Statist. Probab. Lett.* 18, 327-336.

[16] Kerkyacharian, G. Picard, D. and Triboully, K.(1996). L^p adaptive density estimation. *Bernoulli* 2, 229-247.

[17] Kronmal, R. and Tarter, M.(1968). The estimation of probability densities and cumulative by fourier series methods. *J. Amer. Statist. Assoc.* 63, 925-952

[18] Lee,H.(1999).A Study on Wavelet Density Estimation, M.A. thesis,Taegu University

[19] Meyer, Y.(1992). *Wavelets and Operators.* Cambridge Univ. Press Cambrige.

[20] Parzen, E.(1962). On estimation of a probability density function and mode. *Ann. Math. Statist.* 33, 1065-1076.

[21] Penev, S. and Dechevsky, L.(1997). On non-negative wavelet-based density estimators. *Nonpara. Statist.*, 7, 365-394.

[22] Pinheiro, A. and Vidakovic, B.(1997). Estimating the square root of a density via compactly Supporter Wavelets. *Computational Statistics and Data Analysis,* 25, 399-415.

[23] Rosenblatt, M.(1956). Remark on some nonparametric estimates of a density function. *Ann. Math. Statist.* 27, 832 - 837.

[24] Vannucci, M.(1995). Nonparametric density estimation using wavelets, DP 95-26, ISDS, Duke University.

[25] Vidakovic, B.(1999). *Statistical Modeling by Wavelets.* John wiley & Sons,

[26] Wahba, G.(1975). Interpolating spline methods for density estimation I. equi-spaced knots. *Ann. Statist.* 3, 30 - 48.

[27] Water, G.(1992). Approximation of the delta function by wavelets. *J. Approx. Theory.* 71. 329-343.

[28] Walter, G.(1994). *Wavelets and Other Orthogonal Systems with Applications.* CRC Press, Inc.