

## DENSITY ESTIMATION WITH HAAR SERIES

Joachim ENGEL

*Department of Statistics, University of Michigan, 1450 Mason Hall, Ann Arbor, MI 48109-1027, USA*

Received January 1989

Revised March 1989

*Abstract:* Rate of convergence for density estimators based on Haar series are derived under very mild condition: the unknown density has to be of bounded variation. These estimators are histograms on dyadic intervals.

*Keywords:* Orthogonal series density estimator, Haar series, histogram, mean integrated absolute error, bounded variation.

### 1. Introduction

Given  $n$  observations of a random variable  $X$ , say  $X_1, \dots, X_n$ , we estimate the underlying density  $f$ . Our density estimator is an orthogonal series estimator based on the system of Haar functions. Due to the special nature of Haar functions this estimator is a histogram on certain dyadic intervals. Hence this is an approach to analyzing histograms. We derive rate of convergence results for the integrated mean absolute error. The main result (Theorem 3) is based on a lemma by Devroye and Györfi (1985) which decomposes the error into a deterministic part and a large deviation part. The deterministic part is estimated with a result by Ciesielski (1966) regarding the rate of convergence of Haar polynomials. Our density is assumed to belong to a particular Lipschitz class  $\text{Lip}(\alpha, 1)$  of functions.  $\text{Lip}(\alpha, 1)$  is defined as the class of functions whose integral modulus of continuity  $\omega_p(\delta)$  is of the order  $\delta^\alpha$ . Under less restrictive assumptions we obtain the same rate as Abou-Jaoude (1977) for the equal cell width histogram. Our result holds if  $f$  is of bounded variation. The Haar series method can also be applied in the  $L_2$  case where our results are comparable to those of Freedman and Diaconis (1981). The Haar series estimator is attractive because it approximates the equal cell width histogram fairly well and its cell points are recursively refined. This saves computations when updating the estimator after additional observations are obtained.

First we put together some results from abstract harmonic analysis regarding Haar–Fourier polynomials and their rate of convergence. Then we derive rate of convergence ( $L_1$ ) for densities which belong to  $\text{Lip}(\alpha, 1)$ . Since Haar series density estimators are particular histograms we compare our results with other histogram in the literature.

The orthogonal series estimator was first suggested by Čencov (1962). Given a complete orthonormal system  $\{\chi_\nu(t), \nu = 1, 2, \dots\}$  the density  $f$  can be approximated by Fourier polynomials:

$$H_r(f; t) = \sum_{\nu=1}^r a_\nu \chi_\nu(t) \rightarrow f(t) \quad \text{a.e. as } r \rightarrow \infty,$$

where the Fourier coefficients are given by

$$a_\nu = \int_0^1 f(t) \chi_\nu(t) dt = E \chi_\nu.$$

Given a sample  $X_1, \dots, X_n$  drawn from  $f$ ,  $a_\nu$  is estimated without bias by

$$\hat{a}_{n\nu} = \frac{1}{n} \sum_{j=1}^n \chi_\nu(X_j).$$

Hence we estimate  $f(t)$  by

$$\hat{f}_n(t) = \sum_{\nu=1}^r \hat{a}_{n\nu} \chi_\nu(t) = \frac{1}{n} \sum_{j=1}^n \sum_{\nu=1}^r \chi_\nu(X_j) \cdot \chi_\nu(t) = \frac{1}{n} \sum_{j=1}^n K_r(t, X_j),$$

where  $K_r(t, s) = \sum_{\nu=1}^r \chi_\nu(t) \cdot \chi_\nu(s)$ . The upper summation limit  $r = r(n)$  is a smoothing factor for the estimator.

**2. Haar-Fourier series in  $L_p[0, 1]$  for  $1 \leq p < \infty$**

In this section we examine the approximation of  $L_p$  functions ( $1 \leq p < \infty$ ) by their Haar-Fourier series. Let  $\{\chi_n(t)\}$  be the orthonormalized system of Haar functions on the interval  $[0, 1]$ . They are defined as follows:  $\chi_1(t) \equiv 1$  on  $[0, 1]$ , and for  $n = 2^m + k$  ( $k = 1, \dots, 2^m, m = 0, 1, \dots$ ),

$$\chi_n(t) = \chi_m^{(k)}(t) = \begin{cases} \sqrt{2^m} & \text{for } t \in ((2k - 2)/2^{m+1}, (2k - 1)/2^{m+1}), \\ -\sqrt{2^m} & \text{for } t \in ((2k - 1)/2^{m+1}, (2k)/2^{m+1}), \\ 0 & \text{otherwise.} \end{cases}$$

It is essential to note that the system of Haar functions  $\{\chi_n(t)\}$  not only forms a orthonormal system, but that the Haar-Fourier expansion converges almost everywhere for all  $f \in L_p[0, 1]$ ,  $p \geq 1$ . This can be seen as a consequence of the martingale convergence theorem (see, for example, Garsia, 1970, p. 72).

The notation of the  $L_p$  modulus of continuity plays an important role in the results we obtain.

**Definition 1.** (i) The  $L_p$  modulus of continuity of a function  $f \in L_p[a, b]$ ,  $a, b \in \mathbb{R}$ ,  $a < b$ ,  $1 \leq p < \infty$ , is given by

$$\omega_p(\delta; f) = \sup_{0 < h \leq \delta} \left\{ \int_a^{b-h} |f(s+h) - f(s)|^p ds \right\}^{1/p}.$$

(ii) For  $\alpha > 0$ ,  $p \geq 1$  let  $\text{Lip}(\alpha, p) = \{f \in L_p \mid \omega_p(\delta; f) = O(\delta^\alpha)\}$ .

For  $f \in L_p[0, 1]$ ,  $1 \leq p < \infty$ , the Haar-Fourier polynomial of  $f$  is given by  $H_n(f; t) = \sum_{\nu=1}^n a_\nu \chi_\nu(t)$  where  $\chi_\nu$  is the  $\nu$ th Haar function and  $a_\nu$  is the  $\nu$ th Haar-Fourier coefficient,  $a_\nu = \int_0^1 f(s) \chi_\nu(s) ds$ . Then

$$H_n(f; t) = \sum_{\nu=1}^n \left\{ \int_0^1 f(s) \chi_\nu(s) ds \right\} \chi_\nu(t) = \int_0^1 f(s) \sum_{\nu=1}^n \chi_\nu(s) \chi_\nu(t) ds = \int_0^1 f(s) K_n(s, t) ds,$$

where the Haar kernel  $K_n(s, t) = K_m^{(k)}(s, t) = \sum_{\nu=1}^n \chi_\nu(s) \chi_\nu(t)$  can be represented as (see Haar, 1910, for details),

$$K_m^{(k)}(s, t) = \begin{cases} \frac{1}{t_i - t_{i-1}} & \text{if } s, t \text{ in same subinterval } (t_{i-1}, t_i), i = 1, \dots, n, \\ 0, & \text{otherwise,} \end{cases}$$

where

$$t_i = \begin{cases} \frac{i}{2^{m+1}} & \text{for } i = 0, \dots, 2k, \\ \frac{i-k}{2^m} & \text{for } i = 2k+1, \dots, n. \end{cases}$$

For the Haar expansion we have then

$$\begin{aligned} H_n(f; t) &= H_m^{(k)}(f; t) = \int_0^1 f(s) K_m^{(k)}(s, t) \, ds \\ &= \begin{cases} \frac{1}{t_i - t_{i-1}} \int_{t_{i-1}}^{t_i} f(s) \, ds & \text{for } t \in (t_{i-1}, t_i), i = 1, \dots, n, \\ 0 & \text{for } t = t_i. \end{cases} \end{aligned}$$

This means  $H_m^{(k)}(f)$  equals  $f$  averaged over the dyadic intervals  $(t_{i-1}, t_i)$ . The following theorem due to Ciesielski (1966, p. 309) gives us an estimate for the Haar–Fourier approximation in terms of the  $L_p$  modulus of continuity.

**Theorem 1.** *Let  $f \in L_p[0, 1]$ ,  $1 \leq p < \infty$ . Then for  $n \geq 1$ ,*

$$\|f - H_n(f)\|_p \leq 6\omega_p\left(\frac{1}{n}; f\right),$$

*i.e. for  $f \in \text{Lip}(\alpha, p)$  the Haar–Fourier approximation is of the order  $O(n^{-\alpha})$ .  $\square$*

The next proposition follows immediately from the definition of  $\text{Lip}(\alpha, p)$ .

- Proposition 1.** (i) *If  $\alpha_1 < \alpha_2$ , then  $\text{Lip}(\alpha_2, p) \subset \text{Lip}(\alpha_1, p)$ .*  
 (ii) *If  $1 \leq p_1 < p_2$ , then  $\text{Lip}(\alpha, p_2) \subset \text{Lip}(\alpha, p_1)$  for all  $\alpha > 0$ .*  
 (iii) *If  $f \in \text{Lip}(\alpha, p)$  for some  $\alpha > 1$ , then  $f \equiv \text{const. a.e.}$   $\square$*

For us the most interesting case is for  $\alpha = 1, p = 1$ .

**Theorem 2.** *If  $f$  is a function of bounded variation, then  $f \in \text{Lip}(1, 1)$ .*

**Proof.** If  $f$  is of bounded variation on  $[0, 1]$ , we can represent  $f$  as a difference of two nondecreasing functions:  $f(t) = f_1(t) - f_2(t)$ . Noting that  $\omega_1(\delta; f) \leq \omega_1(\delta; f_1) + \omega_1(\delta; f_2)$  we see that it is sufficient to show that  $f \in \text{Lip}(1, 1)$  for nondecreasing  $f(t)$ . But if  $f$  is nondecreasing, then for  $h > 0$  the function  $f(t+h) - f(t)$  is nonnegative on  $[0, 1]$ . Therefore

$$\int_0^1 |f(t+h) - f(t)| \, dt = \int_h^1 f(t) \, dt - \int_0^{1-h} f(t) \, dt = \int_{1-h}^1 f(t) \, dt - \int_0^h f(t) \, dt \leq Mh,$$

where  $M$  is an upper bound on  $f$ . Therefore

$$\omega_1(\delta; f) \leq \sup_{0 \leq h \leq \delta} Mh = M\delta = O(\delta). \quad \square$$

### 3. Haar series density estimation

The following lemma is from Devroye and Györfi (1985, p. 292). It decomposes the mean integrated absolute error into a term determined by the approximation error of the Haar–Fourier series and a large deviation term.

**Lemma 1.** *Let  $\hat{f}_n$  be the Haar series estimate of  $f$  with  $r = r(n)$  terms. Then*

$$E \int_0^1 |\hat{f}_n(t) - f(t)| dt \leq \|H_r(f) - f\|_1 + \frac{1}{\sqrt{n}} \int_0^1 \sqrt{EK_r^2(t, X_1)} dt. \quad \square$$

To obtain a bound for the  $L_1$  error we estimate both terms on the right hand side of this lemma. The expression  $\|H_r(f) - f\|_1$  is estimated by Ciesielski's (1966) theorem. We estimate the large deviation term by the following lemma.

**Lemma 2.**

$$\int_0^1 \sqrt{EK_r^2(t, X_1)} dt \leq \sqrt{2r} \left( \int_0^1 \sqrt{f(t)} dt + o(1) \right).$$

**Proof.** Let  $r = 2^m + k, 1 \leq k \leq 2^m$ . Then

$$EK_r^2(t, X_1) = \int_0^1 K_r^2(t, s) f(s) ds = \frac{1}{(t_i - t_{i-1})^2} \int_{t_{i-1}}^{t_i} f(s) ds$$

for  $t \in (t_{i-1}, t_i)$ . Now  $t_i - t_{i-1} \geq 2^{-(m+1)} \geq (2r)^{-1}$ . Hence

$$EK_r^2(t, X_1) \leq \frac{2r}{t_i - t_{i-1}} \int_{t_{i-1}}^{t_i} f(s) ds, \quad t \in (t_{i-1}, t_i),$$

and

$$\begin{aligned} \int_0^1 \sqrt{EK_r^2(t, X_1)} dt &= \sum_{i=1}^r \int_{t_{i-1}}^{t_i} \sqrt{EK_r^2(t, X_1)} dt \leq \sqrt{2r} \sum_{i=1}^r \int_{t_{i-1}}^{t_i} \sqrt{\frac{1}{t_i - t_{i-1}} \int_{t_{i-1}}^{t_i} f(s) ds} dt \\ &= \sqrt{2r} \sum_{i=1}^r \sqrt{t_i - t_{i-1}} \cdot \sqrt{\int_{t_{i-1}}^{t_i} f(s) ds}, \end{aligned}$$

using  $\int_{t_{i-1}}^{t_i} f(s) ds = f(t_i)(t_i - t_{i-1}) + o(1)$  as  $r \rightarrow \infty$  we obtain

$$\int_0^1 \sqrt{EK_r^2(t, X_1)} dt \leq \sqrt{2r} \sum_{i=1}^r (t_i - t_{i-1}) (\sqrt{f(t_i)} + o(1)) = \sqrt{2r} \left( \int_0^1 \sqrt{f(s)} ds + o(1) \right). \quad \square$$

Now we can derive some rate of convergence results for our density estimator.

**Theorem 3.** *If  $f \in \text{Lip}(\alpha, 1)$  for some  $\alpha > 0$  then an  $L_1$  optimal choice for the smoothing factor is  $r = r(n) \sim n^{1/(2\alpha+1)}$  and the  $L_1$  error is bounded by*

$$E \int_0^1 |\hat{f}_n(t) - f(t)| dt \leq O(n^{-\alpha/(2\alpha+1)}).$$

If  $f$  is a function of bounded variation, then

$$E \int_0^1 |\hat{f}_n(t) - f(t)| dt \leq O(n^{-1/3}).$$

**Proof.** Follows immediately from Lemma 1, Lemma 2, Theorem 1 and Theorem 2.  $\square$

So far we obtained statements regarding the order of the rate of convergence, but no upper bound coefficient. This is due to the general nature of the class  $Lip(\alpha, p)$ . By definition of  $Lip(\alpha, p)$  there is no upper bound coefficient involved in determining if  $f \in Lip(\alpha, p)$  or not. To be able to derive an upper bound coefficient we have to restrict the class of densities further. If  $f$  is twice differentiable,  $f' \in L_1$  and  $f''$  bounded, then we can estimate the modulus of continuity:

$$\begin{aligned} \omega_1\left(\frac{1}{r}; f\right) &= \sup_{0 < h < 1/r} \int_0^{1-h} |f(t+h) - f(t)| dt \leq \sup_{0 < h \leq 1/r} \int_0^{1-h} |f'(t) \cdot h| dt + O(h^2) \\ &\leq \int_0^1 |f'(t)| dt / r + O(r^{-2}). \end{aligned}$$

We obtain the following.

**Theorem 4.** Assume that  $f$  is twice differentiable,  $f' \in L_1$  and  $f''$  is bounded. If we choose

$$r = r(n) = \left(6\sqrt{2} \int_0^1 |f'(t)| dt / \int_0^1 \sqrt{f(t)} dt\right)^{2/3} \cdot n^{1/3},$$

then

$$E \int_0^1 |\hat{f}_n(t) - f(t)| dt \leq 3 \left(3 \int_0^1 |f'(t)| dt\right)^{1/3} \cdot \left(\int_0^1 \sqrt{f(t)} dt\right)^{2/3} \cdot n^{-1/3} + o(n^{-1/3}).$$

**Proof.** With the above estimate of  $\omega_1(1/r; f)$  and Lemma 1, 2 and Theorem 1 we have

$$E \int_0^1 |\hat{f}_n(t) - f(t)| dt \leq 6 \int_0^1 |f'(t)| dt / r + \sqrt{\frac{2r}{n}} \left(\int_0^1 \sqrt{f(t)} dt + o(1)\right) + O(r^{-2}).$$

The first two terms are minimized, if we choose

$$r = r(n) = \left(6\sqrt{2} \int_0^1 |f'(t)| dt / \int_0^1 \sqrt{f(t)} dt\right)^{2/3} \cdot n^{1/3},$$

and we obtain the stated result.  $\square$

A similar approach is possible for the estimation of the mean integrated squared error: First decompose the  $L_2$  error into a deterministic and a large deviation term (see Devroye and Györfi, 1985, p. 292). Then estimate the deterministic term by Ciesielski (1966) theorem and the large deviation term analogous to Lemma 2. From Proposition 1 the most interesting case is when  $f$  is assumed to belong to  $Lip(1, 2)$  which is true if  $f$  is absolutely continuous and  $f' \in L_2$  (see Edwards, 1979, p. 136). Then the  $L_2$  error is of the order  $O(n^{-2/3})$  and an upper bound coefficient can be obtained analogously to Theorem 4.

#### 4. Comparison with histograms

Since Haar series density estimators are histograms on dyadic intervals it is a natural question to ask how this estimator compares to other histograms. Our histograms are almost equal cell width. Our estimator  $\hat{f}_n$  is a histogram whose first  $k$  cells are of width  $2^{-(m+1)}$  where the smoothing factor is represented as  $r = 2^m + k$ ,  $1 \leq k \leq 2^m$ ,  $m = 0, 1, \dots$ . The remaining cells are of width  $2^{-m}$ . Rate of convergence results for equal cell width histograms were obtained in the  $L_1$  case by Abou-Jaoude (1977) and in the  $L_2$  case by Freedman and Diaconis (1981).

In the  $L_1$  case we obtained the same rate of convergence as Abou-Jaoude, but our assumptions on  $f$  are less restrictive: we assume  $f$  to be of bounded variation while Abou-Jaoude assumes  $f$  to be absolutely continuous with a bounded and continuous derivative  $f'$ . In Abou-Jaoude's result the upper bound coefficient has exactly the same functional dependence on  $f$  as in our Theorem 4. However, our upper bound coefficient is not quite as good: a simple comparison shows that our upper bound coefficient is  $2 \cdot (3\pi)^{1/3} \approx 4.22$  times the coefficient of Abou-Jaoude. As explanation of this loss we recall that our histogram is not exactly Abou-Jaoude's equal cell width histogram. Furthermore, our result depends essentially on two inequalities (Ciesielski's theorem and Devroye-Györfi's lemma), none of which is necessarily sharp.

While an improvement on the coefficient is still possible, the rate cannot be improved upon. Even if we consider smoother densities, the rate of Haar series estimators equals the rate of any histograms which is  $O(n^{-1/3})$ .

#### 5. Final remarks

(1) The choice of the sample space as the interval  $[0, 1]$  seems to be quite restrictive. But for the  $L_1$  error this is no restriction at all, since the  $L_1$  error is invariant under continuous and strictly monotone transformations of the coordinate axes (see Devroye and Györfi, 1985, p. 2f, for details).

(2) Rearrangement of the sum  $\hat{f}_n(t) = \sum a_{n\nu} \chi_\nu(t)$  creates histograms of unequal cell width. Applying the Kronmal-Tarter method (Tarter and Kronmal, 1970) of term selection for orthogonal series estimators leads to a locally adaptive histogram.

(3) The results given here generalize to multivariate densities in a straight forward way, once an appropriate multivariate version of the modulus of continuity and the corresponding Lipschitz class  $\text{Lip}^{(d)}(\alpha, p)$  is defined (see Engel, 1988, for details). The applicability of Haar series density estimation for nonparametric regression problems is subject of a forthcoming investigation by the author.

#### Acknowledgements

This work is part of the author's Ph.D. Thesis at the University of Southern California, written under the supervision of Professor Louis Gordon, whose guidance and suggestions are gratefully acknowledged and appreciated. A comment by an anonymous referee helped to improve Theorem 4.

#### References

- Abou-Jaoude, S. (1977), La convergence  $L_1$  et  $L_\infty$  de certains estimateurs d'une densité de probabilité, Thèse à l'Univ. de Paris VI (Paris).
- Čencov, N.N. (1962), Evaluation of an unknown distribution density from observations, *Soviet Math.* **3**, 1559–1562.
- Ciesielski, Z. (1966), Properties of the orthonormal Franklin system, II, *Studia Math.* **27**, 289–323.
- Devroye, L. and L. Györfi, *Nonparametric Density Estimation: The  $L_1$  View* (Wiley, New York).

- Edwards, R.E. (1979), *Fourier Series, Vol 1* (Springer, New York).
- Engel, J. (1988), Density estimation with Haar series, Ph.D. Thesis, Univ. of Southern California (Los Angeles, CA).
- Freedman, D. and P. Diaconis (1981), On the histogram as a density estimator:  $L_2$  theory, *Z. Wahrsch. Verw. Gebiete* **57**, 453–476.
- Garsia, A. (1970), *Topics in Almost Everywhere Convergence* (Markham, Chicago, IL).
- Haar, A. (1910), Zur Theorie der orthogonales Funktionensysteme, *Math. Ann.* **69**, 331–371.
- Tarter, M. and R. Kronmal (1970), On multivariate density estimates based on orthogonal expansions, *Ann. Statist.* **41**, 718–722.

f