

BIBLIOTECA DI SCIENZE STATISTICHE
SERVIZIO BIBLIOTECARIO NAZIONALE
810 PUV0966570 BID
ACQ. 366 / 104 INV. 84486
COLL. 5-coll. WP. 13/2003

**Asymptotics for Bayesian
histograms**

C. Scricciolo

2003.13

**Dipartimento di Scienze Statistiche
Università degli Studi
Via C. Battisti 241-243
35121 Padova**

Luglio 2003

BIBLIOTECA DI SCIENZE STABILIMENTO
SERVIZIO BIBLIOTECARIO CARICATO
110 010 010
ACQ. SECT. IIIA
BOL. CLASS. 110

Associazione per l'istruzione

Elementare

C. 1000000

110

Associazione per l'istruzione

Elementare

C. 1000000

110

110

ASYMPTOTICS FOR BAYESIAN HISTOGRAMS

BY CATIA SCRICCIOLO

University of Padova

We consider the problem of consistently estimating an unknown probability density function on a bounded interval from a sample of n independent and identically distributed univariate random variables. Adopting a Bayesian nonparametric approach, as a first approximation, a hierarchical prior whose weak support comprises all absolutely continuous distribution functions, is considered that selects piecewise constant densities. The prior measure is constructed by putting a prior on the number of equal length bins and a Dirichlet distribution on the bin values. Convergence rate for the Hellinger loss of the Bayes' estimator is deduced from the posterior rate, which is studied for various densities generating the data. This rate is comparable up to a logarithmic factor to that of the frequentist histogram estimator. Smoothing the Bayesian histogram, we get a continuous, piecewise linear competitor which possesses a faster rate of convergence.

1. Introduction. Given a random sample of n independent observations X_1, \dots, X_n from an unknown probability distribution P_0 absolutely continuous with respect to (w.r.t.) Lebesgue measure λ on a bounded interval, which, for convenience, is taken to be $[0, 1]$, we consider the problem of nonparametrically estimating the distribution's density function f_0 adopting a Bayesian approach. Let \mathcal{F} denote the class of all probability measures on $[0, 1]$ that are absolutely continuous w.r.t. λ . A prior measure with full topological support on \mathcal{F} may be constructed exploiting the approximating property of the "theoretical histogram", see Abou-Jaoude (1976).

Received July 2003

AMS 2000 subject classifications. Primary 62G07, 62G20.

Key words and phrases. Bayesian histogram, density estimation, polygon, posterior distribution, rate of convergence.

We introduce some notation. For any integer $k > 0$, let $\{A_{1,k}, \dots, A_{k,k}\}$ be a partition of $[0, 1]$ defined by

$$A_{1,k} = [0, t_{1,k}], \quad A_{j,k} = (t_{j-1,k}, t_{j,k}], \quad 2 \leq j \leq k,$$

where $t_{j,k} = j/k$, for $0 \leq j \leq k$. Defining $c_{j,k} = (t_{j-1,k} + t_{j,k})/2$, for $1 \leq j \leq k$,

$$A_{1,k} = \left[0, c_{1,k} + \frac{1}{2k}\right], \quad A_{j,k} = \left(c_{j,k} - \frac{1}{2k}, c_{j,k} + \frac{1}{2k}\right], \quad 2 \leq j \leq k.$$

For later use, we define also

$$A_{1,k} = A_{1,k}^- \cup A_{1,k}^+, \quad \text{where } A_{1,k}^- = [0, c_{1,k}] \text{ and } A_{1,k}^+ = (c_{1,k}, t_{1,k}],$$

$$A_{j,k} = A_{j,k}^- \cup A_{j,k}^+, \quad \text{where } A_{j,k}^- = (t_{j-1,k}, c_{j,k}] \text{ and } A_{j,k}^+ = (c_{j,k}, t_{j,k}], \quad 2 \leq j \leq k.$$

Uncertainty in the density may be thus modeled:

- (a) let K be a random variable (r.v.) with probability mass function $\rho(k) \geq 0$ on the positive integers, $\sum_{k=1}^{\infty} \rho(k) = 1$, determining the number of bins;
- (b) given $K = k$, let $\mathbf{W}_k = (W_{1,k}, \dots, W_{k,k})$ denote a random vector distributed over the $(k-1)$ -dimensional simplex S_{k-1} in \mathbb{R}^k ,

$$S_{k-1} = \left\{ (y_1, \dots, y_k) \in \mathbb{R}^k : 0 \leq y_j \leq 1, j = 1, \dots, k, \sum_{j=1}^k y_j = 1 \right\}.$$

It is assumed that conditionally on the value of K , $\mathbf{W}_k | k \sim \text{Dir}(\alpha_{1,k}, \dots, \alpha_{k,k})$;

- (c) given $K = k$ and $\mathbf{W}_k = \mathbf{w}_k$, the r.v.'s X_1, \dots, X_n are distributed according to the piecewise constant density

$$h_{k, \mathbf{w}_k}(x) = \sum_{j=1}^k w_{j,k} k \mathbb{I}_{A_{j,k}}(x), \quad x \in [0, 1],$$

where the $w_{j,k}$'s play the role of the mixing weights for the densities $k \mathbb{I}_{A_{j,k}}(x)$, with $\mathbb{I}_{A_{j,k}}(\cdot)$ the indicator function of $A_{j,k}$.

Let $\mathcal{H}_k = \{P : f_P = h_{k, \mathbf{w}_k} = \sum_{j=1}^k w_{j,k} k \mathbb{I}_{A_{j,k}}, \mathbf{w}_k \in S_{k-1}\}$. Then, $\mathcal{H} = \cup_{k=1}^{\infty} \mathcal{H}_k$ is the parameter space and a prior on it is induced by

$$\pi = \sum_{k=1}^{\infty} \rho(k) \pi_k = \sum_{k=1}^{\infty} \rho(k) \text{Dir}(\alpha_{1,k}, \dots, \alpha_{k,k}),$$

which selects stepwise constant density functions. Still, as shown in Proposition 1, if for all k , $\rho(k) > 0$ and \mathbf{W}_k has full support on S_{k-1} , then every absolutely

continuous distribution on $[0, 1]$ is in the weak support. This makes such a prior suitable for the purpose of nonparametric density estimation and the point-wise posterior expectation of the random density,

$$\hat{f}_n(x) = \sum_{k=1}^{\infty} \rho(k|x_1, \dots, x_n) E[W_{j,k}|k, x_1, \dots, x_n] k \mathbb{I}_{A_{j,k}}(x), \quad x \in [0, 1],$$

which needs not be a histogram, provides the Bayes' estimate under quadratic loss.

Strong consistency of the Bayes' estimator follows from posterior consistency w.r.t. the Hellinger metric $d_H(f, g) = \{\int(\sqrt{f} - \sqrt{g})^2\}^{1/2}$, the L_2 -distance between square rooted densities, which can be established generalizing the argument used by Barron, Schervish and Wasserman (1999) for histograms over binary partitions.

The focus here is on determining the rate of convergence of the Bayes' estimator. It suffices to study the rate of the posterior, namely the speed at which the posterior mass concentrates on shrinking Hellinger neighborhoods of f_0 , because the entailed estimator will then converge at least as fast. For the purpose, we assume throughout that for some constant $M > 0$, $\alpha_{j,k} \leq M$ for all j, k . This assumption is fulfilled by specifying $\alpha_{j,k} = M_0 \alpha_0(A_{j,k})$, with α_0 a probability measure on $[0, 1]$, or $\alpha_{j,k} = 1$ for all j, k . Issues concerning the choice of the $\alpha_{j,k}$'s are well discussed in Petrone (1999).

Applying the general theory developed by Ghosal, Ghosh and van der Vaart (2000), Shen and Wasserman (2001), which provides sufficient conditions for assessing posterior rates, we show that if f_0 is indeed piecewise constant, then the posterior converges at rate $n^{-1/2}(\log n)$. The convergence cannot be so fast in general and under mild conditions, the rate is equivalent up to a logarithmic factor to the frequentist rate $n^{-1/3}$ of the histogram estimator for integrated mean squared error.

Smoothing the histogram by joining the ordinates at mid-bin points with straight lines, we get the density

$$p_{k, \mathbf{w}_k}(x) = \begin{cases} h_{k, \mathbf{w}_k}(c_{1,k}), & 0 \leq x \leq c_{1,k}, \\ k [h_{k, \mathbf{w}_k}(c_{j,k})(c_{j+1,k} - x) + h_{k, \mathbf{w}_k}(c_{j+1,k})(x - c_{j,k})], & c_{j,k} < x \leq c_{j+1,k}, \\ h_{k, \mathbf{w}_k}(c_{k,k}), & c_{k,k} < x \leq 1, \end{cases}$$

where $1 \leq j \leq k-1$, which unlike the histogram is continuous and, moreover, has a faster posterior rate of convergence. It is worth noting that the density thus defined is different from the frequency polygon of Scott (1985, 1992).

The problem of computing Bayes' estimates may be circumvented by taking a sequence of priors depending on the sample-size. This allows to truncate the sum to a number of terms varying with n .

Various aspects concerning Bayesian histograms have been studied by Andreev and Arjas (1996), Gasparini (1996), Hartigan (1996), Leonard (1973), but to our knowledge, the issue of rates of convergence has not been addressed yet.

The paper is organized as follows. In the next section, we present some results on the rate of convergence for the posterior. Closing remarks are in Section 3. Some proofs are deferred to the Appendix.

2. Convergence rate of the posterior. When studying rates of convergence, it is sensible to consider also sampling densities other than piecewise constant, because the prior has full weak support on \mathcal{F} provided that for all k , $\rho(k) > 0$ and \mathbf{W}_k is supported on the simplex.

PROPOSITION 1. *Let $\rho(k) > 0$ for $k = 1, 2, \dots$, and the conditional density of \mathbf{W}_k given k , $g(\mathbf{w}_k|k) > 0$ for all $\mathbf{w}_k \in S_{k-1}$. Then, π has full weak support on \mathcal{F} .*

To study posterior rates, we appeal to a general theorem by computing the prior concentration rate on Kullback-Leibler type neighborhoods of f_0 , which accounts for the approximation error of a density by histograms, and by estimating the Hellinger packing number of sieve sets. The first result deals with the case when f_0 is piecewise constant. The rate $n^{-1/2}(\log n)$ is obtained.

THEOREM 1. *Suppose f_0 is piecewise constant, i.e. $f_0 = k_0 \sum_{j=1}^{k_0} w_{j,k_0}^0 \mathbb{I}_{A_{j,k_0}}$ for some k_0 and $\mathbf{w}_{k_0}^0 \in S_{k_0}$. Let $0 < \rho(k) \leq Be^{-\beta k}$ for constants $B, \beta > 0$. Then, for a constant M large enough,*

$$\pi \left(\left\{ P : d_H(f_0, f_P) \geq M \frac{\log n}{\sqrt{n}} \right\} \middle| X_1, \dots, X_n \right) \rightarrow 0$$

in P_0^n -probability.

PROOF. We appeal to Theorem 2.1 in Ghosal (2001), page 1268. Condition (2.2) is shown to be fulfilled for $\bar{\varepsilon}_n = n^{-1/2} \log n$. Let k_n be a sequence such that for constants $C_2 > C_1 > 0$, $C_1 \log(1/\bar{\varepsilon}_n) \leq k_n \leq C_2 \log(1/\bar{\varepsilon}_n)$. Let $\mathcal{F}_n = \cup_{k=1}^{k_n} \mathcal{H}_k$.

Proceeding as in the proof of Theorem 2.2 of Ghosal (2001), pages 1268-1270, we get that $\log D(\varepsilon, \mathcal{F}_n, d_H) \leq 2 \log k_n + k_n \log(3\sqrt{2\pi\varepsilon}/\varepsilon)$. For all large n and $c_1 = C_2/2$,

$$\log D(\tilde{\varepsilon}_n, \mathcal{F}_n, d_H) \leq 2k_n \log \frac{1}{\tilde{\varepsilon}_n} \leq 2C_2 \left(\log \frac{1}{\tilde{\varepsilon}_n} \right)^2 \leq c_1 n \tilde{\varepsilon}_n^2.$$

Condition (2.3) is seen to be verified for $\tilde{\varepsilon}_n = n^{-1/2}(\log n)^{1/2}$. More precisely, for $c_3 = B(1 - e^{-\beta})^{-1}$ and some $0 < \alpha < 1/2$,

$$\pi(\mathcal{F}_n^c) \leq \sum_{k=k_n}^{\infty} B e^{-\beta k} \leq B(1 - e^{-\beta})^{-1} e^{-\beta k_n} \leq c_3 e^{-\beta C_1 (\frac{1}{2} - \alpha) n \tilde{\varepsilon}_n^2}.$$

Now, we check condition (2.4). For any $\varepsilon > 0$, let

$$N(\varepsilon^2) = \left\{ P : \max \left\{ \int f_0 \log(f_0/f_P), \int f_0 (\log(f_0/f_P))^2 \right\} \leq \varepsilon^2 \right\}.$$

Recall that if $P \in \mathcal{H}_{k_0}$, then $f_P = h_{k_0, \mathbf{w}_{k_0}}$ for some $\mathbf{w}_{k_0} \in S_{k_0}$. Note that for any $P \in \mathcal{H}_{k_0}$ such that $\sum_{j=1}^{k_0} |w_{j,k_0} - w_{j,k_0}^0| \leq \varepsilon$, it results $d_H(f_P, f_0) \leq \sqrt{\varepsilon}$. If furthermore $\min_{1 \leq j \leq k_0} w_{j,k_0} > (\varepsilon/2)^4$, then for any $\delta \in (0, 1]$,

$$M_\delta^2 = \int_{\{f_0/f_P \geq e^{1/\delta}\}} f_0 \left(\frac{f_0}{f_P} \right)^\delta < \frac{16k_0 (\max_{1 \leq j \leq k_0} w_{j,k_0}^0)}{\varepsilon^4} = (C_0/\varepsilon^2)^2,$$

having set $C_0^2 = 16k_0 (\max_{1 \leq j \leq k_0} w_{j,k_0}^0)$. By taking $K = \max\{e^{1/\delta}, (C_0/\varepsilon^{5/2})^{1/\delta}\}$ in the proof of Theorem 5 of Wong and Shen (1995), page 357-358, the conclusion goes through. Thus, if

$$N_{k_0} = \left\{ P \in \mathcal{H}_{k_0} : \sum_{j=1}^{k_0} |w_{j,k_0} - w_{j,k_0}^0| \leq \varepsilon, \min_{1 \leq j \leq k_0} w_{j,k_0} > (\varepsilon/2)^4 \right\},$$

for any $P \in N_{k_0}$, it results $\max\{\int f_0 \log(f_0/f_P), \int f_0 (\log(f_0/f_P))^2\} \lesssim \varepsilon (\log \frac{1}{\varepsilon})^2$.

Therefore,

$$N_{k_0} \subseteq N \left(c_0 \varepsilon \left(\log \frac{1}{\varepsilon} \right)^2 \right),$$

for some c_0 depending on f_0 . The prior probability of N_{k_0} can be estimated by appealing to Lemma A.1. of Ghosal (2001), pages 1278-1279. It implies that $\pi(N(c_0 \varepsilon (\log(1/\varepsilon))^2)) \geq D_1 \exp\{-d_1 \log \frac{1}{\varepsilon}\}$ for constants $d_1, D_1 > 0$. Let $\varepsilon' = (c_0 \varepsilon)^{1/2} \log(1/\varepsilon)$. Noting that $\log(1/\varepsilon') \sim \log(1/\varepsilon)$, we get

$$\pi(N(\tilde{\varepsilon}_n^2)) \gtrsim e^{-c_2 n \tilde{\varepsilon}_n^2},$$

for some $c_2 > 0$. By choosing $C_1 = (c_2 + 4)/[\beta(\frac{1}{2} - \alpha)]$, we have $\pi(\mathcal{F}_n^c) \leq c_3 e^{-(c_2+4)n\tilde{\varepsilon}_n^2}$. Finally, $\varepsilon_n = \max\{\tilde{\varepsilon}_n, \tilde{\varepsilon}_n\} = n^{-1/2} \log n$. \square

The result is analogous to Theorem 5.1 in Ghosal and van der Vaart (2001) and Theorem 2.2 in Ghosal (2001): when the “true” density is of the same form as the statistical model, the “nearly parametric rate” $n^{-1/2}(\log n)$ is obtained; otherwise, the rate will be slower. For positively lower bounded Lipschitz densities, we find $n^{-1/3}(\log n)^{5/6}$. The following statement is analogous to Theorem 2.3 of Ghosal (2001) for Bernstein densities.

THEOREM 2. *Suppose f_0 is a Lipschitz function bounded away from zero. Let $\rho(k) = Be^{-\beta k}$ for fixed $B, \beta > 0$. Then, for a constant M large enough,*

$$\pi \left(\left\{ P : d_H(f_0, f_P) \geq M \frac{(\log n)^{5/6}}{n^{1/3}} \right\} \middle| X_1, \dots, X_n \right) \rightarrow 0$$

P_0^∞ -almost surely.

PROOF. It is along the same lines as that of Theorem 1. Let $\tilde{\varepsilon}_n = n^{-1/3}(\log n)^{1/3}$ and $\bar{\varepsilon}_n = n^{-1/3}(\log n)^{5/6}$. Also, for $D_2 > D_1 > 0$ to be suitably chosen, let $D_1(1/\tilde{\varepsilon}_n) \log(1/\tilde{\varepsilon}_n) \leq s_n \leq D_2(1/\bar{\varepsilon}_n) \log(1/\bar{\varepsilon}_n)$. Define $\mathcal{F}_n = \cup_{k=1}^{s_n} \mathcal{H}_k$. Then, for all large n ,

$$\log D(\bar{\varepsilon}_n, \mathcal{F}_n, d_H) \leq 2s_n \log \frac{1}{\bar{\varepsilon}_n} \leq 2D_2 \frac{1}{\bar{\varepsilon}_n} \left(\log \frac{1}{\bar{\varepsilon}_n} \right)^2 \leq d_1 n \bar{\varepsilon}_n^2,$$

with $d_1 = 2D_2/9$. For $d_3 = B/(1 - e^{-\beta})$ and some $0 < \alpha < 1$,

$$\pi(\mathcal{F}_n^c) \leq d_3 e^{-\beta s_n} \leq d_3 e^{-\beta D_1 \frac{1}{\bar{\varepsilon}_n} \log \frac{1}{\bar{\varepsilon}_n}} \leq d_3 e^{-\frac{\beta D_1}{3}(1-\alpha)n\bar{\varepsilon}_n^2}.$$

In order to obtain an almost-sure assertion, we employ condition (2.5) in Ghosal, Ghosh and van der Vaart (2000), page 505. Note that f_0 is continuous and bounded on $[0, 1]$. Let $\|f_0\|_\infty = \max_{0 \leq x \leq 1} f_0(x)$ and $L_0 = \min_{0 \leq x \leq 1} f_0(x) > 0$. Define $h_{k, \mathbf{w}_k^0} = k \sum_{j=1}^k w_{j,k}^0 \mathbb{I}_{A_{j,k}}$, with $w_{j,k}^0 = \int_{(j-1)/k}^{j/k} f_0(x) dx$, $j = 1, \dots, k$. Using the mean value theorem and a Lipschitz argument, it can be seen that $\|f_0 - h_{k, \mathbf{w}_k^0}\|_\infty \leq M_0 k^{-1}$, with M_0 the Lipschitz constant of f_0 . If $b_1 \varepsilon^{-1} \leq k \leq b_2 \varepsilon^{-1}$, then for any $\mathbf{w}_k \in S_k$ satisfying $\sum_{j=1}^k |w_{j,k}^0 - w_{j,k}| \leq \varepsilon^2$, we have

$$\|f_0 - h_{k, \mathbf{w}_k}\|_\infty \leq M_0 k^{-1} + k \sum_{j=1}^k |w_{j,k}^0 - w_{j,k}| \leq C_1 \varepsilon,$$

for some $C_1 > 0$. Consequently, for k large enough or ε sufficiently small, the histogram h_{k, \mathbf{w}_k} is bounded away from 0 and $d_H^2(f_0, h_{k, \mathbf{w}_k}) \|f_0/h_{k, \mathbf{w}_k}\|_\infty \leq C_2 \varepsilon^2$,

with $C_2 > 0$. Therefore,

$$(1) \quad H(C_2\varepsilon^2) \equiv \left\{ P : d_{\text{H}}^2(f_0, f_P) \left\| \frac{f_0}{f_P} \right\|_{\infty} \leq C_2\varepsilon^2 \right\} \\ \supseteq \left\{ P \in \mathcal{H}_k : \sum_{j=1}^k |w_{j,k}^0 - w_{j,k}| \leq \varepsilon^2 \right\}.$$

By Lemma 6.1 of Ghosal, Ghosh and van der Vaart (2000), pages 518-519, for constants $d_4, D > 0$, $\pi(H(C_2\varepsilon^2)) \geq d_4 e^{-\beta k} \exp\{-Dk \log \frac{1}{\varepsilon}\}$, provided that $\varepsilon^2/2 \leq k^{-1}$. Taking $\tilde{\varepsilon}_n^{-1} \leq k_n \leq 2\tilde{\varepsilon}_n^{-1}$, for all large n ,

$$\pi(H(C_2\tilde{\varepsilon}_n^2)) \geq d_4 e^{-d_2 n \tilde{\varepsilon}_n^2},$$

with $d_2 > 0$. Putting $D_1 = 3(d_2 + 4)/[(1 - \alpha)\beta]$, we have $\pi(\mathcal{F}_n^c) \leq d_3 e^{-(d_2+4)n\varepsilon_n^2}$. The assertion follows. \square

Except for the log-factor, the rate is equivalent to that at which the integrated mean squared error (IMSE) for the frequentist estimator defined by the relative frequency of the number of observations falling into each cell converges to zero. The IMSE is asymptotically $O(n^{-2/3})$, see Scott (1979).

Ghosal (2001), Theorem 2.3, obtains the same rate $n^{-1/3}(\log n)^{5/6}$ for the Bernstein polynomial prior, even though assuming more smoothness in the true density. In what follows, we show that under slightly stronger regularity conditions on f_0 than in Theorem 2 and Theorem 2.3 of Ghosal (2001), the rate can be improved by adopting a prior on the polygons, which also has weak full support.

PROPOSITION 2. *Under the same assumptions as in Proposition 1, π has full weak support on \mathcal{F} .*

THEOREM 3. *Suppose f_0 is bounded away from zero and twice continuously differentiable with f_0' and f_0'' uniformly bounded. Let $\rho(k) = Be^{-\beta k}$, for constants $B, \beta > 0$. Then, for M sufficiently large,*

$$\pi \left(\left\{ P : d_{\text{H}}(f_0, f_P) \geq M \frac{(\log n)^{7/8}}{n^{3/8}} \right\} \middle| X_1, \dots, X_n \right) \rightarrow 0$$

P_0^∞ -almost surely.

PROOF. Similar to the proof of Theorem 1 (Theorem 2). Let $\tilde{\varepsilon}_n = n^{-3/8}(\log n)^{3/8}$ and $\bar{\varepsilon}_n = n^{-3/8}(\log n)^{7/8}$. For constants $D_2 > D_1 > 0$, define

$$D_1 \left(\frac{1}{\tilde{\varepsilon}_n} \right)^{2/3} \log \left(\frac{1}{\tilde{\varepsilon}_n} \right) \leq s_n \leq D_2 \left(\frac{1}{\tilde{\varepsilon}_n} \right)^{2/3} \log \left(\frac{1}{\tilde{\varepsilon}_n} \right),$$

where $D_1 > 0$ will be shortly determined for the purpose. Note that for suitable constants $b_1, b_2 > 0$ and all large n ,

$$b_1 n^{1/4} (\log n)^{3/4} \leq s_n \leq b_2 n^{1/4} (\log n)^{3/4}.$$

Take $\mathcal{F}_n = \cup_{k=1}^{s_n} \mathcal{P}_k$. Using the relationship between packing and bracketing numbers, from Lemma 1, it follows

$$\begin{aligned} D(\bar{\varepsilon}_n, \mathcal{F}_n, d_H) &\leq \sum_{k=1}^{s_n} D(\bar{\varepsilon}_n, \mathcal{P}_k, d_H) \leq \sum_{k=1}^{s_n} N_{[]}(\bar{\varepsilon}_n, \mathcal{P}_k, d_H) \\ &\leq \sum_{k=1}^{s_n} \frac{k(2\pi e)^{k/2}}{\bar{\varepsilon}_n^{k-1}} \\ &\leq s_n^2 \left(\frac{\sqrt{2\pi e}}{\bar{\varepsilon}_n} \right)^{s_n}. \end{aligned}$$

Therefore, for all n large enough and $d_1 = 9D_2/16$,

$$\log D(\bar{\varepsilon}_n, \mathcal{F}_n, d_H) \leq 4s_n \log \frac{1}{\bar{\varepsilon}_n} \leq d_1 n \bar{\varepsilon}_n^2.$$

Also, for $d_3 = B(1 - e^{-\beta})^{-1}$,

$$\pi(\mathcal{F}_n^c) \leq d_3 e^{-\beta s_n} \leq d_3 e^{-\frac{3\beta D_1}{8}(1-\alpha)n\bar{\varepsilon}_n^2}.$$

Next, we check condition (2.4). We denote by p_k^0 the polygon-shaped density p_{k, w_k^0} based on P_0 . For any $x \in [0, 1]$,

$$(2) \quad |f_0(x) - p_k(x)| \leq |f_0(x) - p_k^0(x)| + |p_k^0(x) - p_k(x)| \equiv T_1(x) + T_2(x).$$

We claim that

$$(3) \quad \|T_1\|_\infty = O(k^{-1}).$$

Let $\Delta_{j,k}^0(x) = |f_0(x) - p_k^0(x)| \mathbb{I}_{A_{j,k}^+ \cup A_{j+1,k}^-}(x)$ for $1 \leq j \leq k-1$. Write

$$T_1(x) = |f_0(x) - kw_{1,k}^0| \mathbb{I}_{A_{1,k}^-}(x) + \sum_{j=1}^{k-1} \Delta_{j,k}^0(x) + |f_0(x) - kw_{k,k}^0| \mathbb{I}_{A_{k,k}^+}(x), \quad x \in [0, 1].$$

Now, using a second order Taylor expansion, for each $1 \leq j \leq k$,

$$kw_{j,k}^0 = f_0(c_{j,k}) + R(c_{j,k}),$$

where, for a point $\xi_{j,k}$ lying in between t and $c_{j,k}$,

$$R(c_{j,k}) = \frac{k}{2} \int_{c_{j,k} - \frac{1}{2k}}^{c_{j,k} + \frac{1}{2k}} (t - c_{j,k})^2 f_0''(\xi_{j,k}) dt = O(k^{-2}),$$

because f_0'' is bounded. For any $x \in A_{1,k}^-$, $|f_0(x) - kw_{1,k}^0| \leq \int_x^{c_{1,k}} |f_0'(t)| dt + |R(c_{1,k})|$. Since f_0' is bounded, $\sup_{x \in A_{1,k}^-} |f_0(x) - kw_{1,k}^0| = O(k^{-1})$. Analogously, $\sup_{x \in A_{k,k}^+} |f_0(x) - kw_{k,k}^0| = O(k^{-1})$. Reasoning as in Beirlant, Berinet, Biau and Vajda (2002), page 213, $\sup_{x \in A_{j,k}^+ \cup A_{j+1,k}^-} \Delta_{j,k}^0(x) = O(k^{-2})$ uniformly over $1 \leq j \leq k-1$. Thus, (3) follows. Also, it can be seen that

$$(4) \quad \|T_2\|_\infty = \|p_k^0 - p_k\|_\infty \leq 2k \sum_{j=1}^k |w_{j,k}^0 - w_{j,k}|.$$

If $\sum_{j=1}^k |w_{j,k}^0 - w_{j,k}| \leq \varepsilon^2$, taking $c_1 \varepsilon^{-2/3} \leq k \leq c_2 \varepsilon^{-2/3}$ for some $c_1, c_2 > 0$, then (2), (3) and (4) imply that $\|f_0 - p_k\|_\infty \leq C_0 \varepsilon^{2/3}$, with C_0 a constant depending on f_0 . Thus, for ε sufficiently small, p_k is bounded away from 0.

Let us now consider $d_{\mathbb{H}}^2(f_0, p_k)$. It is known from Beirlant, Berinet, Biau and Vajda (2002), page 211, that under the assumptions on f_0 and its first two derivatives, $\chi^2(f_0 \| p_k^0) = \int_0^1 (f_0^2 / p_k^0) d\lambda - 1 = O(k^{-3})$, whence

$$\begin{aligned} d_{\mathbb{H}}^2(f_0, p_k) &\leq 2(d_{\mathbb{H}}^2(f_0, p_k^0) + d_{\mathbb{H}}^2(p_k^0, p_k)) \\ &\leq 2\chi^2(f_0 \| p_k^0) + 2\|p_k^0 - p_k\|_1 \\ &\leq O(k^{-3}) + 2 \sum_{j=1}^k |w_{j,k}^0 - w_{j,k}| \\ &\leq C_1 \varepsilon^2. \end{aligned}$$

Consequently, for ε sufficiently small, $d_{\mathbb{H}}^2(f_0, p_k) \|f_0/p_k\|_\infty \leq C_2 \varepsilon^2$ for some $C_2 > 0$, and for a constant $d_4 > 0$,

$$\pi(H(C_2 \varepsilon^2)) \geq \pi \left(\left\{ P \in \mathcal{P}_k : \sum_{j=1}^k |w_{j,k}^0 - w_{j,k}| \leq \varepsilon^2 \right\} \right) \geq d_4 \exp \left\{ -Dk \log \frac{1}{\varepsilon} \right\},$$

if only $\varepsilon^2/2 \leq k^{-1}$. Taking $\tilde{\varepsilon}_n^{-2/3} \leq k_n \leq 2\tilde{\varepsilon}_n^{-2/3}$, where $\tilde{\varepsilon}_n^{2/3} = n^{-1/4}(\log n)^{1/4}$, clearly $\tilde{\varepsilon}_n^2/2 \leq k_n^{-1}$. So, for $d_2, d_4 > 0$,

$$\pi(H(C_2\tilde{\varepsilon}_n^2)) \geq d_4 e^{-d_2 n \tilde{\varepsilon}_n^2}.$$

Finally, putting $D_1 = 8(d_2 + 4)/[3\beta(1 - \alpha)]$, it results $\pi(\mathcal{F}_n^c) \leq d_3 e^{-(d_2+4)n\tilde{\varepsilon}_n^2}$ and the proof is completed. \square

REMARK 1. The rate turns out to be equivalent up to a logarithmic factor to the rate $n^{-3/4}$ in expected χ^2 -divergence obtained by Beirlant, Berlinet, Biau and Vajda (2002) for the corresponding frequentist estimator. The rate is driven by the approximation rate in χ^2 -divergence of the histogram part near the endpoints 0 and 1 which is less accurate than elsewhere, being only proportional to k^{-3} .

REMARK 2. A better rate may be obtained by considering a sequence of sample-size-dependent priors supported on $\mathcal{F}_n = \cup_{r=1}^{k_n} \mathcal{P}_r$. The second condition thus becomes trivial. Then, $\pi_n = \sum_{r=1}^{k_n} \rho_n(r) \text{Dir}(\alpha_{1,r}^{(n)}, \dots, \alpha_{r,r}^{(n)})$ with $k_n \rightarrow \infty$, as $n \rightarrow \infty$, and $\sum_{r=1}^{k_n} \rho_n(r) = 1$ for all n , where $\alpha_{j,r}^{(n)} \leq M$ for $j = 1, \dots, r$ and sufficiently large n . Under the same assumptions as in Theorem 2, for $c_1(n/\log n)^{1/4} \leq k_n \leq c_2(n/\log n)^{1/4}$, with $c_2 > c_1 > 0$, the rate is slightly improved to $n^{-3/8}(\log n)^{3/8}$. It also becomes possible to compute the Bayes' estimate.

3. Closing remarks. All rates obtained differ for a logarithmic factor from those of the corresponding frequentist estimators. This is typical when using Dirichlet priors. Indeed, as remarked by Ghosal, Ghosh and van der Vaart (2000), it is not said that this is due to a prior deficiency, rather it is possible that the approximation used is not accurate enough. Following Ghosal (2001), the Bayesian bootstrap method gives a proxy for the posterior distribution and has a rate at par with that of the frequentist estimators.

Previous results only provide upper bounds on the actual rates. Even though far from being able to state that these rates are sharp, we may say that the difference between $n^{-1/3}(\log n)^{5/6}$ and $n^{-3/8}(\log n)^{7/8}$ is an indication of a weaker rate, as if smoothing the histogram by means of the Bernstein polynomial approximation does not produce any positive effect as, instead, smoothing by means of polygons seems to do. But, it is still too early to draw general conclusions on this matter and further investigations are necessary.

APPENDIX

PROOF OF PROPOSITION 1. If every set in a base for the weak topology of \mathcal{F} has positive π -measure, then the assertion follows. Let $\varepsilon > 0$ be fixed. For $r \geq 1$, let f_1, \dots, f_r be bounded, continuous, real-valued functions on $[0, 1]$. For any fixed $P \in \mathcal{F}$, let

$$U_\varepsilon(P) \equiv \left\{ Q \in \mathcal{F} : \left| \int f_i dQ - \int f_i dP \right| < \varepsilon, i = 1, \dots, r \right\}$$

be a neighborhood of P of the usual base. Clearly,

$$\pi(U_\varepsilon(P)) = \pi \left(\bigcup_{k=1}^{\infty} \left\{ Q \in \mathcal{H}_k : \left| \int f_i(h_k^Q - f_P) \right| < \varepsilon, i = 1, \dots, r \right\} \right)$$

and it suffices to show that $\pi(U_\varepsilon(P)) > 0$.

Let $\|f\|_\infty = \max_{1 \leq i \leq r} \|f_i\|_\infty$. We denote by h_k^P the histogram-shaped density based on P . For each $i \in \{1, \dots, r\}$ and any integer $k \geq 1$,

$$\begin{aligned} \left| \int f_i(h_k^Q - f_P) \right| &\leq \|f\|_\infty \left(\|h_k^Q - h_k^P\|_1 + \|h_k^P - f_P\|_1 \right) \\ &= \|f\|_\infty \|h_k^Q - h_k^P\|_1 + \|f\|_\infty \|h_k^P - f_P\|_1 \equiv T_{1,k} + T_{2,k}. \end{aligned}$$

By Theorem 1 of Abou-Jaoude (1976), page 219, and Csiszár (1973), page 168, $\|h_k^P - f_P\|_1 \rightarrow 0$ as $k \rightarrow \infty$. There thus exists $\bar{k} \equiv \bar{k}(\varepsilon)$ such that $T_{2,k} < \varepsilon/2$ for all $k \geq \bar{k}$. If $\mathbf{w}_k^Q \in N_\varepsilon(\mathbf{w}_k^P) \equiv \{\mathbf{w}_k \in S_{k-1} : \sum_{j=1}^k |w_{j,k} - w_{j,k}^P| < \varepsilon/(2\|f\|_\infty)\}$, it results $T_{1,k} \leq \|f\|_\infty \sum_{j=1}^k |w_{j,k}^Q - w_{j,k}^P| < \varepsilon/2$. Then, $|\int f_i(h_k^Q - f_P)| < \varepsilon$ for all $1 \leq i \leq r$. Hence,

$$\pi(U_\varepsilon(P)) \geq \sum_{k \geq \bar{k}} \rho(k) \int_{N_\varepsilon(\mathbf{w}_k^P)} g(\mathbf{w}_k|k) d\mathbf{w}_k > 0,$$

because $\rho(k) > 0$ for all k and $g(\mathbf{w}_k|k) > 0$ for all $\mathbf{w}_k \in S_{k-1}$. \square

PROOF OF PROPOSITION 2. It is just outlined because it proceeds along the same lines as the proof of the last proposition. For each $k \geq 1$ and $1 \leq i \leq r$,

$$(5) \quad \left| \int f_i(p_k^Q - f_P) \right| \leq \left| \int f_i(p_k^Q - h_k^Q) \right| + \left| \int f_i(h_k^Q - f_P) \right|.$$

To study the first term, write

$$p_k^Q(x) = \begin{cases} h_k^Q(c_{1,k}), & x \in A_{1,k}^-, \\ kw_{j,k}^Q + k^2(x - c_{j,k})(w_{j+1,k}^Q - w_{j,k}^Q), & x \in A_{j,k}^+ \cup A_{j+1,k}^-, \quad 1 \leq j \leq k-1, \\ h_k^Q(c_{k,k}), & x \in A_{k,k}^+. \end{cases}$$

Taking into account that $j/k = c_{j,k} + 1/(2k)$ as well as $j/k = c_{j+1,k} - 1/(2k)$ and using the generalized mean value theorem, yields

$$\begin{aligned} & \left| \int f_i(p_k^Q - h_k^Q) \right| \\ &= \left| \sum_{j=1}^{k-1} \int_{A_{j,k}^+ \cup A_{j+1,k}^-} f_i(p_k^Q - h_k^Q) \right| \\ &= k^2 \left| \sum_{j=1}^{k-1} (w_{j+1,k} - w_{j,k}) \left[\int_{c_{j,k}}^{j/k} f_i(x)(x - c_{j,k}) dx - \int_{j/k}^{c_{j+1,k}} f_i(x)(c_{j+1,k} - x) dx \right] \right| \\ &\leq \frac{1}{8} \sum_{j=1}^{k-1} |w_{j+1,k} - w_{j,k}| |f_i(\xi_i) - f_i(\xi'_i)|, \end{aligned}$$

with $\xi_i \in (c_{j,k}, j/k]$ and $\xi'_i \in (j/k, c_{j+1,k}]$. Note that $\xi'_i - \xi_i < 1/k$. Each f_i is uniformly continuous. Given $\varepsilon > 0$, there exists $\delta_{i,\varepsilon} > 0$ such that for any $x, x' \in [0, 1]$ with $|x - x'| < \delta_{i,\varepsilon}$, it is $|f_i(x) - f_i(x')| < \varepsilon$. Define $\delta_\varepsilon = \min_{1 \leq i \leq r} \delta_{i,\varepsilon}$. If k is large enough so that $k^{-1} < \delta_\varepsilon$, then $\xi'_i - \xi_i < k^{-1} < \delta_\varepsilon \leq \delta_{i,\varepsilon}$ and

$$\left| \int f_i(p_k^Q - h_k^Q) \right| \leq \frac{1}{8} \sum_{j=1}^{k-1} |w_{j+1,k} - w_{j,k}| |f_i(\xi_i) - f_i(\xi'_i)| < \frac{\varepsilon}{4}, \quad \forall i = 1, \dots, r.$$

The second term in (5) can be bounded reasoning as in the proof of Proposition 1. For any $\mathbf{w}_k^Q \in N_{\varepsilon/2}(\mathbf{w}_k^P)$, it turns out $|\int f_i(h_k^Q - f_P)| < \frac{3}{4}\varepsilon$ for all $1 \leq i \leq r$. Hence, for a suitable \bar{k} ,

$$\pi(U_\varepsilon(P)) \geq \sum_{k \geq \bar{k}} \rho(k) \int_{N_{\varepsilon/2}(\mathbf{w}_k^P)} g(\mathbf{w}_k|k) d\mathbf{w}_k > 0,$$

which completes the proof. \square

In order to prove Theorem 3, we state the following lemma.

LEMMA 1. *If $\varepsilon \leq 1$, then*

$$N_{[\cdot]}(\varepsilon, \mathcal{P}_k, d_H) \leq \frac{k(2\pi e)^{k/2}}{\varepsilon^{k-1}}.$$

PROOF. The idea is to relate the bracketing number of \mathcal{P}_k with that of S_{k-1} .

Consider $p_{k, \mathbf{w}_k} \in \mathcal{P}_k$ and let

$$\begin{aligned} p_{k, \mathbf{w}_k}(x) &= kw_{1,k} \mathbb{I}_{A_{1,k}^-}(x) \\ &\quad + \sum_{j=1}^{k-1} k^2 [w_{j,k}(c_{j+1,k} - x) + w_{j+1,k}(x - c_{j,k})] \mathbb{I}_{A_{j,k}^+ \cup A_{j+1,k}^-}(x) \\ &\quad + kw_{k,k} \mathbb{I}_{A_{k,k}^+}(x) \\ &\equiv f_{w_{1,k}}(x) + \sum_{j=1}^{k-1} g_{w_{j,k}, w_{j+1,k}}(x) + h_{w_{k,k}}(x), \quad x \in [0, 1]. \end{aligned}$$

Let $\mathbf{a} = (a_1, \dots, a_k)$ and $\mathbf{b} = (b_1, \dots, b_k)$ be an ε -Hellinger bracket for \mathbf{w}_k . Define $l = p_{k, \mathbf{a}}$ and $u = p_{k, \mathbf{b}}$. Clearly, $l(x) \leq p_{k, \mathbf{w}_k}(x) \leq u(x)$ for every $x \in [0, 1]$, and $d_{\text{H}}(l, u) \leq \varepsilon$ because

$$\begin{aligned} d_{\text{H}}^2(l, u) &\leq d_{\text{H}}^2(f_{a_1}, f_{b_1}) + \sum_{j=1}^{k-1} d_{\text{H}}^2(g_{a_j, a_{j+1}}, g_{b_j, b_{j+1}}) + d_{\text{H}}^2(h_{a_k}, h_{b_k}) \\ &\leq \sum_{j=1}^k (\sqrt{a_j} - \sqrt{b_j})^2 = d_{\text{H}}^2(\mathbf{a}, \mathbf{b}) \leq \varepsilon^2. \end{aligned}$$

The minimal number of such brackets is bounded above by the bracketing number of S_{k-1} ,

$$N_{[]}(\varepsilon, \mathcal{P}_k, d_{\text{H}}) \leq N_{[]}(\varepsilon, S_{k-1}, d_{\text{H}}) \leq \frac{k(2\pi e)^{k/2}}{\varepsilon^{k-1}},$$

see Lemma 2 in Genovese and Wasserman (2000), pages 1110-1111. \square

REFERENCES

- Andreev, A. and Arjas, E. (1996). A note on histogram approximation in Bayesian density estimation. In *Bayesian Statistics 5*, Bernardo, J. M., Berger, J. O., Dawid, A. P. & Smith, A. F. M. (Eds.), 487-490. Oxford Univ. Press, Oxford.
- Abou-Jaoude, S. (1976). Conditions nécessaires et suffisantes de convergence L1 en probabilité de l'histogramme pour une densité. *Ann. Inst. Henri Poincaré, Section B: Calcul des Probabilités et Statistique* **12** 213-231.
- Barron, A., Schervish, M. J. and Wasserman, L. (1999). The consistency of posterior distributions in nonparametric problems. *Ann. Statist.* **27** 536-561.

- Beirlant, J., Berliet, A., Biau, G. and Vajda, I. (2002). Divergence-type errors of smooth Barron-type density estimators. *Test* **11** 191–217.
- Csiszár, I (1973). Generalized entropy and quantization problems, in *Trans. Sixth Prague Conf. Information Theory, Statistical Decision Functions, Random Processes*. Prague: Academia, 159–174.
- Gasparini, M. (1996). Bayesian density estimation via Dirichlet density processes. *Nonparametric Statist.* **6** 355–366.
- Genovese, C. R. and Wasserman, L. (2000). Rates of convergence for the Gaussian mixture sieve. *Ann. Statist.* **28** 1105–1127.
- Ghosal, S. (2001). Convergence rates for density estimation with Bernstein polynomials. *Ann. Statist.* **29** 1264–1280.
- Ghosal, S., Ghosh, J. K. and van der Vaart, A. W. (2000). Convergence rates of posterior distributions. *Ann. Statist.* **28** 500–531.
- Ghosal, S. and van der Vaart, A. W. (2001). Entropies and rates of convergence for maximum likelihood and Bayes estimation for mixtures of normal densities. *Ann. Statist.* **29** 1233–1263.
- Hartigan, J. A. (1996). Bayesian histograms. In *Bayesian Statistics 5*, Bernardo, J. M., Berger, J. O., Dawid, A. P. & Smith, A. F. M. (Eds.), 211–222. Oxford Univ. Press, Oxford.
- Leonard, T. (1973). Bayesian method of histograms. *Biometrika* **60** 297–308.
- Petrone, S. (1999). Random Bernstein polynomials. *Scand. J. Statist.* **26** 373–393.
- Scott, D. W. (1979). On optimal and data-based histograms *Biometrika* **66** 605–610.
- Scott, D. W. (1985). Frequency polygons: theory and application. *J. Amer. Statist. Assoc.* **80** 348–354.
- Scott, D. W. (1992). *Multivariate Density Estimation - Theory, Practice and Visualization*. Wiley, New York.
- Shen, X. and Wasserman, L. (2001). Rates of convergence of posterior distributions. *Ann. Statist.* **29** 687–714.
- Wong, W. H. and Shen, X. (1995). Probability inequalities for likelihood ratios and convergence rates of sieve MLEs. *Ann. Statist.* **23** 339–362.

C. SCRICCILO
DIPARTIMENTO DI SCIENZE STATISTICHE
UNIVERSITÀ DEGLI STUDI DI PADOVA
VIA C. BATTISTI 241/243
I-35121 PADOVA
ITALY
E-MAIL: catia@stat.unipd.it

