

A Note on Bayesian Density Estimation

Una nota sulla stima bayesiana di densità

Catia Scricciolo¹

Istituto di Metodi Quantitativi,
Università Commerciale “L. Bocconi”, e-mail: catia.scricciolo@uni-bocconi.it

Riassunto: In questa nota si considera il problema della stima di una funzione di densità su $[0, 1]$ dotata di derivata prima hölderiana di ordine incognito. Si adotta come stimatore la densità attesa *a posteriori* corrispondente ad una misura iniziale che seleziona quasi certamente distribuzioni con densità di forma poligonale. L’attenzione è rivolta a fornire una valutazione dell’accuratezza asintotica della suddetta procedura di stima bayesiana non parametrica, quale espressa dalla velocità di convergenza a zero del rischio associato alla funzione di perdita indotta dal quadrato della norma L_1 . Si dimostra che, puntualmente, lo stimatore in esame raggiunge la velocità di convergenza ottima (secondo il criterio del minimax), a meno di un eventuale fattore logaritmico, adattandosi automaticamente all’incognito livello di regolarità della densità che genera le osservazioni campionarie.

Keywords: Adaptive density estimation, polygonally smoothed prior, posterior rate of convergence.

1. Introduction and Main Results

In this note we consider the problem of estimating a density function on the unit interval from a random sample of observations. The sampling density is assumed to belong to a Hölder class of unknown level of smoothness. The interest is in the asymptotic behaviour of the Bayes’ estimator arising from a polygonally smoothed prior. We provide an evaluation of the accuracy of this estimation procedure in terms of the rate of convergence of the mean squared L_1 -error as the sample size increases.

Let us now proceed to state the problem in more formal terms. The observations $X_1, \dots, X_n$ are an i.i.d. sample from a density f_0 on $[0, 1]$ which is assumed to belong to some Hölder class $\mathcal{F}_\alpha$ of unknown smoothness level α . We recall that a density f_0 is α -Hölderian, $\alpha > 0$, if there exists a constant $L_0 > 0$ such that $|f_0^{(\lfloor \alpha \rfloor)}(x) - f_0^{(\lfloor \alpha \rfloor)}(y)| \leq L_0 |x - y|^{\alpha - \lfloor \alpha \rfloor}$ for all $x, y \in [0, 1]$. A Bayesian nonparametric approach to this problem suggests adopting the Bayes’ estimate

$$\hat{f}_n(\cdot) := \int f_P(\cdot) \Pi(dP|X_1, \dots, X_n),$$

namely, the posterior expected density arising from a prior Π . We take Π to be a *polygonally smoothed prior*. This is a hierarchical prior selecting probability measures with polygon-shaped densities of the form

$$p_{\mathbf{w}_k}(x) = w_{1,k} k \mathbb{1}_{[0, 1/2k]}(x) + \sum_{j=1}^{k-1} p_{j,k}(x) + w_{k,k} k \mathbb{1}_{(1-1/2k, 1]}(x), \quad x \in [0, 1],$$

¹Address of correspondence: Catia Scricciolo, Istituto di Metodi Quantitativi, Università Commerciale “L. Bocconi”, Viale Isonzo 25, I-20135 Milano, Italy.

where $p_{j,k}(x) = [w_{j,k}k(c_{j+1,k} - x) + w_{j+1,k}k(x - c_{j,k})]k\mathbb{1}_{(c_{j,k}, c_{j+1,k}]}(x)$ with $c_{j,k} = (j - 1/2)/k$ for $j = 1, \dots, k - 1$, the number k of equally spaced bins has probability mass function $\rho(\cdot)$ on the positive integers and, given k , the vector of weights $\mathbf{w}_k = (w_{1,k}, \dots, w_{k,k})$ has a Dirichlet distribution on the k -dimensional simplex, see Scricciolo (2007) for details.

We investigate one aspect of the asymptotic behaviour of the Bayes' estimator from the frequentist perspective. Specifically, we assess the accuracy of the estimation procedure in terms of the speed at which the expected error measured by the squared Hellinger or the equivalent L_1 -distance between the estimate and the true density f_0 converges to zero as n tends to infinity. This aim mathematically translates in finding the asymptotic size order of the squared Hellinger risk, that is, a sequence $\epsilon_n \downarrow 0$ such that

$$\mathbb{E}_0^n[d_H^2(\hat{f}_n, f_0)] = \mathbb{E}_0^n \left[\int (\hat{f}_n^{1/2} - f_0^{1/2})^2 \right] = O(\epsilon_n^2).$$

Since the Bayes' estimator is an inferential summary of the posterior distribution, indications on its rate of convergence can be drawn from the rate of convergence of the posterior itself, that is, the “fastest” sequence ϵ_n such that for a sufficiently large constant $M > 0$, $\mathbb{E}_0^n[\Pi(\{P : d_H(f_P, f_0) > M\epsilon_n\} | X_1, \dots, X_n)] \rightarrow 0$ as $n \rightarrow \infty$. This convergence indicates that the posterior distribution asymptotically concentrates its mass on Hellinger balls of shrinking radius ϵ_n around the true density f_0 . We prove that if f_0 is α -Hölderian, $\alpha \in (1, 2]$, bounded away from zero on $[0, 1]$ and with the first derivative satisfying the conditions

$$f'_0(x) = a_0x^p + o(x^p), \quad x \downarrow 0, \quad f'_0(x) = b_0(1 - x)^q + o((1 - x)^q), \quad x \uparrow 1,$$

for constants $a_0, b_0 \in \mathbb{R}$ and $1/2 \leq p, q < \infty$, then the posterior of a polygonally smoothed prior converges at rate $n^{-\alpha/(2\alpha+1)}(\log n)^{(2\alpha+1/2)/(2\alpha+1)}$. Consequently, the Bayes' estimator satisfies

$$\mathbb{E}_0^n[d_H^2(\hat{f}_n, f_0)] = O(n^{-2\alpha/(2\alpha+1)}(\log n)^{(4\alpha+1)/(2\alpha+1)}),$$

where $n^{-2\alpha/(2\alpha+1)}$ is known to be the optimal minimax rate for the present problem. Thus, for α -Hölderian densities with $\alpha \in (1, 2]$, the Bayes' estimator arising from a polygonally smoothed prior achieves the optimal rate up to a (negligible) logarithmic factor. Moreover, the rate automatically adapts to the unknown smoothness level α of the true density, by this meaning that even though the estimator is independent of α , the correct rate arises whatever may be the true value of α . This result complements that of Theorem 4 in Scricciolo (2007), where only the case that $\alpha \in (0, 1]$ was considered. Comparing these rates with those found by Kruijer and van der Vaart (2005) for estimating α -smooth densities using mixtures of beta densities, it seems that the Bayes' estimator arising from a polygonally smoothed prior is more accurate than the one arising from a Bernstein-Dirichlet prior. We believe this may be attributed to the poor approximation quality of mixtures of beta densities.

References

- Kruijer W. and van der Vaart A. (2005) Posterior convergence rates for Dirichlet mixtures of beta densities, in *Stochastics section: selected preprints, WS2005-9*. Department of Mathematics, Faculty of Sciences, Vrije Universiteit Amsterdam. Available at <http://www.math.vu.nl/sto/publications-en.html/stochastics/2005/WS2005-9.pdf>.
- Scricciolo C. (2007) On rates of convergence for Bayesian density estimation, *Scandinavian Journal of Statistics* (to appear).