Journal of Statistical Planning and Inference 141 (2011) 3382–3390

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Convergence rates of nonparametric posterior distributions

Yang Xing

Centre for Mathematical Sciences, Lund University, Box 118, SE-22100 Lund, Sweden

ARTICLE INFO

Article history:
Received 9 March 2009
Received in revised form
21 August 2009
Accepted 19 October 2010
Available online 28 October 2010

Keywords:
Hellinger consistency
Posterior distribution
Rate of convergence
Sieve
Infinite-dimensional model

ABSTRACT

We study the asymptotic behavior of posterior distributions for i.i.d. data. We present general posterior convergence rate theorems which extend several known results on rates of posterior convergence. Our main tools are the Hausdorff $\alpha-$entropy introduced by Xing and Ranneby (2009) and a new notion of prior concentration. Our results are applied to several statistical models.

1. Introduction

Recently, a major theoretical advance has occurred in the theory of Bayesian consistency for infinite-dimensional models. Schwartz (1965) first proved that, if the true density is in the Kullback–Leibler support of the prior distribution, then the sequence of the posterior distributions accumulates in each given weak neighborhood of the true density. It is known that the condition of positivity of prior mass on each Kullback–Leibler neighborhood in Schwartz's theorem is not a necessary condition. When one considers problems like density estimation, it is natural to ask for the almost sure consistency of Bayesian procedures. Sufficient conditions for the almost sure consistency and for evaluating consistency rates have been currently developed by many authors. In this paper we study the problem of determining whether the posterior distributions accumulate in Hellinger neighborhoods of the true density. The rate of thus a convergence can be measured by the size of the smallest shrinking Hellinger balls around the true density on which posterior masses tend to one as the sample size increases to infinity. It is known that the convergence rate of posterior distributions can be determined by two quantities: the rate of a metric entropy and the prior concentration rate. Roughly speaking, the rate of the metric entropy describes how large the model is, and the prior concentration rate depends on prior masses near the true density. Since the true density is unknown, the later assumption actually requires that the prior distribution spreads its mass more or less uniformly over the whole density space. Another elegant approach for determination of the convergence rate of posterior distributions was provided by Walker (2004), who obtained a sufficient condition for the almost sure consistency by using summability of square root of prior probability instead of the metric entropy method. In this paper, in dealing with the rate of metric entropies we shall apply the Hausdorff $\alpha-$entropy introduced by Xing and Ranneby (2009), which is smaller than widely used metric entropies and the bracketing entropy. For some important prior distributions of statistical models the Hausdorff $\alpha-$entropies of all sieves are uniformly bounded, whereas it is generally impossible to get uniform boundedness of metric entropies of large sieves. The application of the Hausdorff $\alpha-$entropy leads to refinements of several theorems on posterior convergence rates,

E-mail address: yang.xing@math.lth.se

0378-3758/$ - see front matter © 2011 Published by Elsevier B.V.
doi:10.1016/j.jspi.2010.10.009

for instance, the well known assumptions on metric entropies and summability of square root of prior probability have been weakened. To handle the prior concentration rate, we shall introduce a new notion of prior concentration which is easier to use. Finally, we present a result to deal with the posterior convergence at the optimal rate $1/\sqrt{n}$.

An outline of this paper is as follows. In Section 2 we define the Hausdorff $\alpha-$entropy with respect to a given prior and then present general theorems for rates of convergence of posterior distributions. We also give a new approach to compute the concentration rates. In Section 3 we apply our results to priors based on uniform distribution on finite subsets, log spline models and finite-dimensional models, which leads to some improvements on known results for these models. The proofs of the main results are contained in Section 4.

2. Notations and theorems

We consider a family of probability measures dominated by a $\sigma-$finite measure μ in $\mathbb{X}$, a Polish space endowed with a $\sigma-$algebra $\mathcal{X}$. Let $X_1, X_2, \ldots, X_n$ stand for an independent identically distributed (i.i.d.) sample of n random variables, taking values in $\mathbb{X}$ and having a common density f_0 with respect to the measure μ. Denote by F_0^∞ the infinite product distribution of the probability distribution F_0 associated with f_0. For two probability densities f and g we denote the Hellinger distance $H(f,g) = (\int_{\mathbb{X}}(\sqrt{f(x)} - \sqrt{g(x)})^2 \mu(dx))^{1/2}$ and the Kullback–Leibler divergence $K(f,g) = \int_{\mathbb{X}} f(x) \log(f(x)/g(x))\mu(dx)$. Assume that the space $\mathbb{F}$ of densities is separable with respect to the Hellinger metric and that $\mathcal{F}$ is the Borel $\sigma-$algebra of $\mathbb{F}$. Given a prior distribution Π on $\mathbb{F}$, the posterior distribution Π_n is a random probability measure with the following expression:

$$\Pi_n(A) = \Pi(A|X_1, X_2, \ldots, X_n) = \frac{\int_A \prod_{i=1}^n f(X_i)\ \Pi(df)}{\int_{\mathbb{F}} \prod_{i=1}^n f(X_i) \Pi(df)} = \frac{\int_A R_n(f) \Pi(df)}{\int_{\mathbb{F}} R_n(f) \Pi(df)}$$

for all measurable subsets $A \subset \mathbb{F}$, where $R_n(f) = \prod_{i=1}^n \{f(X_i)/f_0(X_i)\}$ is the likelihood ratio. In other words, the posterior distribution Π_n is the conditional distribution of Π given the observations $X_1, X_2, \ldots, X_n$. If the posterior distribution Π_n concentrates on arbitrarily small neighborhoods of the true density f_0 almost surely, then it is said to be consistent at f_0 almost surely. Throughout this paper, the almost sure convergence should be understood as to be with respect to the infinite product distribution F_0^∞ of F_0.

Our aim of this article is to present general theorems on posterior convergence rates at f_0. By the posterior convergence rate theorems of Shen and Wasserman (2001) or Ghosal et al. (2000), we know that the prior concentration rate together with the rate of metric entropy can determine the convergence rate of posterior distributions. More specifically, a key inequality to determine almost sure convergence rates of posterior distributions is that for each $\varepsilon > 0$,

$$\int_{\mathbb{F}} R_n(f) \Pi(df) \geq e^{-3n\varepsilon^2}\ \Pi(f : H(f_0,f)^2 \|f_0/f\|_\infty < \varepsilon^2)$$

almost surely for all sufficiently large n, where $\|g\|_\infty$ stands for the supremum norm of the function g on $\mathbb{X}$. This inequality was obtained by Ghosal et al. (2000, Lemma 8.4), under mild assumptions. Handling rates of almost sure convergence of posterior distributions appear in many papers. The reason is that in order to get the convergence rate of posterior distributions one needs to find a suitable lower bound for the denominator in the expression of posterior distributions, which is true if the prior Π puts sufficient amount of mass around the true density f_0 in the sense: $\Pi(f : H(f_0,f)^2\ \|f_0/f\|_\infty < \tilde{\varepsilon}_n^2) \geq e^{-n\tilde{\varepsilon}_n^2 c}$ for some fixed constant c. Such a sequence $\{\tilde{\varepsilon}_n\}$ is referred to as the concentration rate of the prior Π around f_0. Here we give a weak condition instead. We use the following premetric:

$$H_*(f_0,f) = \left(\int_{\mathbb{X}} \left(\sqrt{f_0(x)} - \sqrt{f(x)}\right)^2 \left(\frac{2}{3}\sqrt{\frac{f_0(x)}{f(x)}} + \frac{1}{3}\right)\mu(dx)\right)^{1/2}.$$

It is clear that the inequality $\|f_0/f\|_\infty \geq 1$ holds for all densities f and f_0 such that the supremum is well-defined, and the quality holds if and only if f=f_0 almost surely. Observe also that $H_*(f_0,f) \neq H_*(f,f_0)$ and $3^{-1/2}H(f_0,f) \leq H_*(f_0,f)$. Moreover, we have

$$H_*(f_0,f) \leq H(f_0,f)\|\tfrac{2}{3}\sqrt{f_0/f} + \tfrac{1}{3}\|_\infty^{1/2} \leq H(f_0,f)\|f_0/f\|_\infty^{1/4} \leq H(f_0,f)\|f_0/f\|_\infty^{1/2}$$

which yields

$$\{f \in \mathbb{F} : H_*(f_0,f) \leq \tilde{\varepsilon}_n\} \supset \{f \in \mathbb{F} : H(f_0,f)^2 \sqrt{\|f_0/f\|_\infty} < \tilde{\varepsilon}_n^2\}$$
$$\supset \{f \in \mathbb{F} : H(f_0,f)^2 \|f_0/f\|_\infty < \tilde{\varepsilon}_n^2\}.$$

The following simple lemma shows that, for $\tilde{\varepsilon}_n$ to be a prior concentration rate, it is enough to assume $\Pi(W_{\tilde{\varepsilon}_n}) \geq e^{-n\,\tilde{\varepsilon}_n^2 c_3}$, where $W_\varepsilon = \{f \in \mathbb{F} : H_*(f_0,f) \leq \varepsilon\}$.

Lemma 1. *Let $\varepsilon > 0$ and $c > 0$. Then the inequality*

$$F_0^\infty\left(\int_{\mathbb{F}} R_n(f)\Pi(df) \leq e^{-n\varepsilon^2(3+2c)} \Pi(W_\varepsilon)\right) \leq e^{-n\varepsilon^2 c}$$

holds for all n.