Journal of Statistical Research *ISSN 0256 - 422 X*
2006, Vol. 40, No. 1, pp. 73-76
Bangladesh

L-DIVERGENCE CONSISTENCY FOR A DISCRETE PRIOR

MARIAN GRENDAR
Department of Mathematics, FPV UMB
Tajovskeho 40, 974 01 Banska Bystrica, Slovakia
Institute of Mathematics and CS, Banska Bystrica, Slovakia
Institute of Measurement Sciences, Bratislava, Slovakia
Email: marian.grendar@savba.sk

SUMMARY

Posterior distribution over a countable set $\mathcal{M}$ of continuous data-sampling distributions piles up at L-projection of the true distribution r on $\mathcal{M}$, provided that the L-projection is unique. If there are several L-projections of r on $\mathcal{M}$, then the posterior probability splits among them equally.

Keywords and phrases: Bayesian consistency, L-divergence, multiple L-projections

AMS Classification: 60F10, 60F15

1 Introduction

Walker [6] has recently considered consistency of posterior distribution in Hellinger distance, for strictly positive prior over a countable set of continuous data-sampling distributions. By means of his martingale approach [7], Walker developed a sufficient condition for the Hellinger consistency of posterior density in the above mentioned setting. Via a simple large-deviations approach we show that in this setting posterior density is always consistent in L-divergence. The consistency holds also under misspecification. If there are multiple 'concentration points' (L-projections) the posterior spreads among them equally.

2 Bayesian nonparametric consistency

Let there be countable set $\mathcal{M} = \{q_1, q_2, \dots\}$ of probability density functions with respect to the Lebesgue measure; sources, for short. On the set a Bayesian puts his strictly positive prior probability mass function $\pi(\cdot)$. Let r be the true source of a random sample $X^n \triangleq X_1, X_2, \dots, X_n$. Provided that $r \in \mathcal{M}$, as the sample size grows to infinity, the posterior distribution $\pi(\cdot|X^n = x^n)$ over $\mathcal{M}$ is expected to concentrate in a neighborhood of the true source r. Whether and under what conditions this indeed happens is a subject of Bayesian

© Institute of Statistical Research and Training (ISRT), University of Dhaka, Dhaka 1000, Bangladesh.

nonparametric consistency investigations. Surveys of the subject can be found at [2], [8] among others.

Ghosal, Ghosh and Ramamoorthi [2] define consistency of a sequence of posteriors with respect to a metric or discrepancy measure d as follows: The sequence $\{\pi(\cdot|X^n), n \geq 1\}$ is said to be d-consistent at r, if there exists a $\Omega_0 \subset \mathbb{R}^\infty$ with $r(\Omega_0) = 1$ such that for $\omega \in \Omega_0$, for every neighborhood U of r, $\pi(U|X^n) \to 1$ as n goes to infinity. If a posterior is d-consistent for any $r \in \mathcal{M}$ then it is said to be d-consistent. There, two modes of convergence are usually considered: convergence in probability and almost sure convergence.

Obviously, in the definition the set of sources is not restricted to be countable. The present work is concerned with the countable $\mathcal{M}$ case.

3 Sanov's Theorem for Sources, L-consistency

Let $\mathcal{M}^e \triangleq \{q : q \in \mathcal{M}, \pi(q) > 0\}$ be support of the prior pmf. In what follows, r is not necessarily from $\mathcal{M}^e$. Thus we are interested also in Bayesian consistency under misspecification; i.e., when $\pi(r) = 0$. The problem is the same as in the case of standard Bayesian consistency (cf. Sect. 2): to find the source(s) upon which the posterior concentrates.

For two densities p, q with respect to the Lebesgue measure[1] λ, the I-divergence $I(p||q) \triangleq \int p \log(p/q)$. The L-divergence $L(q||p)$ of q with respect to p is defined as $L(q||p) \triangleq -\int p \log q$. The L-projection $\hat{q}$ of p on $\mathcal{Q}$ is $\hat{q} \triangleq \arg\inf_{q\in\mathcal{Q}} L(q||p)$. There $\mathcal{Q}$ is a set of probability densities defined on the same support. The value of L-divergence at an L-projection of p on $\mathcal{Q}$ is denoted by $L(\mathcal{Q}||p)$.

The following Sanov's Theorem for Sources (LST) will be needed for establishing the consistency in L-divergence. The Theorem provides rate of the exponential decay of the posterior probability.

***L*ST** *Let* $\mathcal{N} \subset \mathcal{M}^e$. As $n \to \infty$,

$$\frac{1}{n} \log \pi(q \in \mathcal{N}|x^n) \to -\{L(\mathcal{M}^e||r) - L(\mathcal{N}||r)\},$$

with probability one.

Proof Let $l_n(q) \triangleq \exp(\sum_{l=1}^n \log q(X_l))$, $l_n(A) \triangleq \sum_{q\in A} l_n(q)$, and $\rho_n(q) \triangleq \pi(q) l_n(q)$, $\rho_n(A) \triangleq \sum_{q\in A} \rho_n(q)$. In this notation $\pi(q \in \mathcal{N}|x^n) = \frac{\rho_n(\mathcal{N})}{\rho_n(\mathcal{M}^e)}$. The posterior probability is bounded above and below as follows:

$$\frac{\hat{\rho}_n(\mathcal{N})}{\hat{l}_n(\mathcal{M}^e)} \leq \pi(q \in \mathcal{N}|x^n) \leq \frac{\hat{l}_n(\mathcal{N})}{\hat{\rho}_n(\mathcal{M}^e)},$$

where $\hat{l}_n(A) \triangleq \sup_{q\in A} l_n(q)$, $\hat{\rho}_n(A) \triangleq \sup_{q\in A} \rho_n(q)$.

$\frac{1}{n}(\log \hat{l}_n(\mathcal{N}) - \log \hat{\rho}_n(\mathcal{M}^e))$ converges with probability one to $L(\mathcal{N}||r) - L(\mathcal{M}^e||r)$. The same is the 'point' of a.s. convergence of $\frac{1}{n}\log$ of the lower bound. □

[1] Any σ-finite measure, in general.

Let for $\epsilon > 0$, $\mathcal{N}_\epsilon^C(\mathcal{M}^e) \triangleq \{q : L(\mathcal{M}^e||r) - L(q||r) > \epsilon, q \in \mathcal{M}^e\}$. Let $\mathcal{N}_\epsilon(\mathcal{M}^e) \triangleq \mathcal{M}^e \backslash \mathcal{N}_\epsilon^C$.

Corollary *Let there be a finite number of L-projections of r on $\mathcal{M}^e$. As $n \to \infty$, $\pi(q \in \mathcal{N}_\epsilon^C(\mathcal{M}^e)|x^n) \to 0$, with probability one.*

Standard Bayesian consistency follows as a special $\pi(r) > 0$ case of the Corollary.

4 Posterior Equi-concentration of Sources

If there is more than one L-projection of r on $\mathcal{M}^e$, how is the posterior probability asymptotically spread among them? This issue is 'in probability' answered by the next Theorem. Let $\mathcal{N}_\epsilon^1 \subset \mathcal{N}_\epsilon(\mathcal{M}^e)$ contain (among other sources) just one L-projection of r on $\mathcal{M}^e$.

Theorem *Let there be* k *L-projections of r on $\mathcal{M}^e$. Then for n going to infinity, $\pi(q \in \mathcal{N}_\epsilon^1|x^n) \to \frac{1}{\mathrm{k}}$, in probability.*

Proof For any $\epsilon > 0$, there exists such n_0 that for $n > n_0$, $r\{x^n : S(\hat{q}_\lambda) = S(\hat{q}_L)\} = 1$, where $\hat{q}_\lambda \triangleq \arg\sup_{q \in \mathcal{M}^e} \pi(q|x^n)$, $\hat{q}_L$ is L-projection of r on $\mathcal{M}^e$, and $S(\cdot)$ stands for 'set of all'. Consequently, $\pi(\hat{q}_L|x^n) \geq \pi(q|x^n)$ for all $q \in \mathcal{M}^e$. Posterior $\pi(q \in \mathcal{N}_\epsilon^1|x^n)$ can be expressed as $(1-A)/(\mathrm{k}(1-B))$, where $A \triangleq \sum_{\sigma_1} \pi(q|x^n)/\pi(\hat{q}_L|x^n)$, $B \triangleq \sum_{\sigma_2} \pi(q|x^n)/\mathrm{k}\pi(q|x^n)$; $\sigma_1 \triangleq \mathcal{N}_\epsilon^1 \backslash \hat{q}_L$, $\sigma_2 \triangleq \mathcal{M}^e \backslash \bigcup_{j=1}^k \hat{q}_L^j$. Markov's inequality implies that $\pi(q|x^n)/\pi(\hat{q}_L|x^n)$ converges to zero, in probability. Slutsky's Theorem then implies that A, B converges to zero, in probability. □

5 EndNotes

In order to place this note in context let us make a few comments.

1) An inverse of Sanov's Theorem has been established by Ganesh and O'Connell [1] for the case of sources with finite alphabet, by means of formal large-deviations approach. Unaware of their work, the present author developed in [3] an inverse of Sanov's Theorem for n-sources, for both discrete and continuous alphabet and applied it to conditioning by rare sources problem and criterion choice problem; cf. also [4].

2) At [3] the concepts of L-divergence and L-projection were introduced. See [3] for a short discussion on why or why not the 'new' divergence.

3) The present form of Sanov's Theorem for Sources (LST) as well as its proof are new.

4) Bayesian consistency under misspecification has already been studied by Kleijn and van der Vaart [5] for general setting of continuous prior on a set of continuous sources, using a different technique. The authors developed *sufficient* conditions for somewhat related consistency (cf. Corollary 2.1 and Lemma 6.4 of [5]) as well as rates of convergence. The equi-concentration was not considered there.

5.1 Acknowledgements

Supported by VEGA grant 1/3016/06.

References

[1] Ganesh, A. and N. O'Connell (1999). An inverse of Sanov's Theorem. *Stat. & Prob. Letters,* **42***, 201–206.*

[2] Ghosal, A., Ghosh, J. K. and R. V. Ramamoorthi (1999). Consistency issues in Bayesian Nonanparametrics. *Asymptotics, Nonparametrics and Time Series: A Tribute to Madan Lal Puri*, 639–667, Dekker.

[3] Grendar. M. (2005). Conditioning by rare sources. *Acta Univ. Belii, Math.,* **12** *19-29.* Online at http://actamath.savbb.sk

[4] Grendar, M. and G. Judge (2006). Large deviations theory and Empirical Estimator choice. *preprint, UC Berkeley.*

[5] Kleijn, B. J. K. and A. W. van der Vaart (2002). Misspecification in infinite-dimensional Bayesian statistics. *preprint.*

[6] Walker, S. (2004). Hellinger consistency for a discrete prior. *Pak. J. Stat.,* **20** *321–327.*

[7] Walker. S. (2004). New appraoches to Bayesian consistency. *Ann. Statist.,* **32** *2028–2043.*

[8] Walker, S., Lijoi, A. and I. Prünster (2004). Contibutions to the understanding of Bayesian consistency. *working paper.*