Moderate deviations for Bayes posteriors

Peter Eichelsbacher* and Ayalvadi Ganesh †
University of Bielefeld and BRIMS, Bristol

Abstract

Let $(X_k)_{k \in \mathbb{N}}$ be a sequence of i.i.d. random variables taking values in a set Ω, and consider the problem of estimating the law of X_1 in a Bayesian framework. We prove that under mild conditions on the support of the prior, the sequence of posterior distributions satisfies a moderate deviation principle.

Key words: moderate deviations principle, Bayes posteriors, Fisher-information, exponentially good approximations

1 Introduction

Bayesian methods have become increasingly popular in statistics in recent years and there has been renewed interest in non-parametric Bayesian inference. In turn, this has stimulated much recent work on the consistency of these inference procedures. Freedman [12] and Diaconis and Freedman [8] showed that even if the prior puts positive mass in every weak neighbourhood of the true distribution, it does not follow that the posterior mass of each weak neighbourhood tends to 1 (in fact, it can tend to zero!). Under the stronger condition that the prior puts positive mass in each Kullback-Leibler neighbourhood of the true distribution, Schwartz [27] showed that asymptotically the posterior does concentrate on weak neighbourhoods of this distribution. If, in addition, the relevant space of probability distributions satisfies a 'metric entropy' condition, then Barron *et al.* [4] show

*Fakultät für Mathematik, Universität Bielefeld, Postfach 100131, D–33501 Bielefeld, Germany

†BRIMS, Hewlett-Packard Labs, Bristol, Filton Road, Stoke Gifford, Bristol BS12 QZ, UK. Research supported by ARC Grant 880 from the Anglo-German foundation.

that the posterior concentrates on neighbourhoods defined by the Hellinger metric; these are finer than weak neighbourhoods. (The Hellinger distance between two densities f and g with respect to a reference measure μ is defined by $\int(\sqrt{f}-\sqrt{g})^2 d\mu$. As a distance between probability distribution, it does not depend on the choice of the reference measure.) Rates of convergence have been investigated by Schwartz [27], Johnson [20, 21], Ghosal *et al.* [18] and Shen and Wasserman [28]. Recent research on the consistency of Bayes methods is reviewed by Ghosal *et al.* [17] and Wasserman [29].

There has been comparatively little work on more refined asymptotics for Bayes posteriors. Results on asymptotic normality of the posterior are few and mixed. Le Cam [22] and Ibragimov and Has'minskii [19] prove central limit theorems for parametric models under certain smoothness conditions; the covariance of the limiting normal distribution is the inverse of the Fisher information. Non-parametric problems are studied by Cox [6] and Diaconis and Freedman [9]. Large deviation asymptotics have been studied by Fu and Kass [13] and Ganesh and O'Connell [14, 15]; also see Lynch and Sethuraman [24], who establish a large deviation principle for a sequence of Dirichlet distributions.

In this paper, we obtain a *moderate deviation principle* for a sequence of Bayes posteriors. We begin by deriving the result on a finite sample space and extend it to general state spaces using the concept of exponentially good approximations introduced by Eichelsbacher and Schmock [11].

Let $(\mathcal{X},\tau)$ be a Hausdorff topological space with Borel σ-algebra $\mathcal{B}$, and let $(\mu_n)_{n\in\mathbb{N}}$ be a sequence of probability measures on $(\mathcal{X},\mathcal{B})$. A *rate function* is a non-negative lower semi-continuous function on $\mathcal{X}$. Let $(\lambda_n)_{n\in\mathbb{N}}$ be a positive sequence decreasing to zero. We say that the sequence $(\mu_n)_{n\in\mathbb{N}}$ satisfies a *large deviation principle* (LDP) on $\mathcal{X}$ with rate function I and speed λ_n, if for all $B\in\mathcal{B}$,

$$-\inf_{x\in B^\circ} I(x) \le \liminf_{n\to\infty} \lambda_n \log \mu_n(B) \le \limsup_{n\to\infty} \lambda_n \log \mu_n(B) \le -\inf_{x\in \bar{B}} I(x).$$

Here B° and $\bar{B}$ denote the interior and closure of B, respectively. A rate function I is *good*, when the level sets $\{x : I(x)\le L\}$, $L\ge 0$, are compact in $(\mathcal{X},\tau)$.

First we consider the case that Ω is a finite set and denote by $M_1(\Omega)$ (respectively, $M_b(\Omega)$) the space of probability measures (respectively, finite signed measures) on Ω. Consider a sequence of independent random variables $(X_k)_{k\in\mathbb{N}}$ taking values in Ω, with common law μ. Denote by