Available online at www.sciencedirect.com
ScienceDirect
Journal of Statistical Planning and Inference 137 (2007) 2102–2114
journal of statistical planning and inference
www.elsevier.com/locate/jspi

Uniform L_1 posterior consistency in compact Gaussian shift experiments☆

Suman Majumdar*

University of Connecticut, 1 University Place, Stamford, CT 06901-2315, USA

Received 30 November 2004; received in revised form 30 November 2005; accepted 21 May 2006
Available online 18 September 2006

Abstract

For a suitably chosen metric ρ for the topology of weak convergence in the space of prior distributions on the shift parameter of a compact Gaussian shift experiment, the posterior distribution (induced by a full support prior) of the shift parameter given independent and identically distributed observations from the experiment is uniformly (in the shift parameter) L_1 consistent in ρ. The corresponding posterior mean is uniformly L_1 consistent in the norm on the parameter space. ρ-distance between the posterior mean (induced by a full support hyperprior) of the prior (i.e., mixing) distribution given independent and identically distributed observations from the mixture experiment and *the empirical distribution of the parameter sequence in the product experiment* [*the mixing distribution*] goes to zero in L_1 of the product [mixture] experiment, uniformly in parameter sequences [mixing distributions].

MSC: Primary 62F15; 62F12; secondary 62C25; 62C10; 60B10

Keywords: Consistency; GC subset; Hyperprior; Isonormal process; Mixture; Prior; Product experiment; Posterior; Topological support; Weak convergence

1. Introduction

The objective of this paper is to examine the asymptotic behavior of various posterior quantities in the Gaussian shift experiment. Specifically, we consider the posterior mean and distribution of the shift parameter given (conditionally) independent and identically distributed (iid hereafter) observations from the experiment, and the (hyperprior induced) posterior mean of the prior (i.e., mixing) distribution given (again, conditionally) iid observations from the mixture of the experiment. Restricting to a compact sub-experiment, we obtain uniform L_1 consistency of these quantities (in a number of different but related models for appropriate targets), as corollaries to the following technical result: in L_1 of the product of mixtures model, the probability of a Kullback–Leibler neighborhood of the empirical distribution of the mixing sequence, calculated under the (hyperprior induced) posterior distribution of the prior, converges to one, uniformly over mixing sequences satisfying a hyperprior related condition on their empirical distributions.

The results outlined above are best understood within a hierarchical Bayes model we present in the next paragraph. In the remainder of this paragraph, we summarize the notational conventions used in this paper. To denote the integral

☆ This research was partially supported by NSA Grant No. MDA904-96-1-0083.
* Tel.: +1 203 251 8417; fax: +1 203 251 8592.
E-mail address: suman.majumdar@uconn.edu.

0378-3758/$ - see front matter
doi:10.1016/j.jspi.2006.05.011

of a function f with respect to (wrt hereafter) a measure v, we interchangeably use the left operator notation $v(f)$ (or even vf) and the integral notation $\int f\,dv$ with the dummy variable of integration sometimes (partially) displayed. Probabilities are always identified with their induced expectations; consistent with that, sets are identified with their indicator functions. The set-theoretic complement of a set A is denoted by A^c. For a measurable mapping X on a probability space $(.,.,P)$, PX^{-1} denotes the induced probability measure on the range space; the n-fold measure theoretic product of P on the product space is denoted by P^n. The notation $a := b$ means that a and b are equal by definition. $\Re$ stands for the real line.

Let Θ be a Polish space (endowed with its Borel σ-algebra) and $\{P_\theta : \theta \in \Theta\}$ an experiment on a sample space $(\mathscr{X}, \mathscr{F})$ such that $\theta \mapsto P_\theta(A)$ is measurable $\forall\, A \in \mathscr{F}$. Let Ω denote the set of all prior distributions (Borel probabilities) on Θ and $\omega_{\underline{x}}$ the posterior distribution of θ given $\underline{x} := (x_1, \ldots, x_n)$ when $\theta \sim \omega \in \Omega$ and, given θ, $\underline{x} \sim P_\theta^n$, i.e., $(x_1, \ldots, x_n)$ are iid observations from the distribution P_θ. For $\omega \in \Omega$ and $A \in \mathscr{F}$, let $P_\omega(A) := \int P_\theta(A)\,d\omega(\theta)$, so that the experiment $\{P_\omega : \omega \in \Omega\}$, called the mixture (of the) experiment ($\{P_\theta : \theta \in \Theta\}$) in the sequel, is the marginal experiment in the Bayes model when $\theta \sim \omega$ and, given θ, $x \sim P_\theta$. Let $\Lambda_{\underline{x}}$ denote the posterior distribution of ω given $\underline{x} := (x_1, \ldots, x_n)$ when $\omega \sim \Lambda$ (a hyperprior on Ω) and, given ω, $\underline{x} \sim P_\omega^n$ (the so-called Bayes empirical Bayes model). Let $\widehat{\Lambda}(B) := \int_\Omega \omega(B)\,d\Lambda_{\underline{x}}(\omega)$; $\widehat{\Lambda}$ is called the (hyperprior induced) posterior mean of the prior. For $\underline{\theta} := (\theta_1, \ldots, \theta_n) \in \Theta^n$, the n-fold Cartesian product of Θ, let $\boldsymbol{P}_{\underline{\theta}}$ denote the product measure $\times_{\alpha=1}^n P_{\theta_\alpha}$(the so-called compound model), and let $G_n := n^{-1}\sum_{\alpha=1}^n \delta_{\theta_\alpha}(\in \Omega)$ denote the empirical distribution (on Θ) of the parameter sequence $\underline{\theta}$, where δ_θ denotes the unit point mass at θ. Similarly, for $\underline{\omega} := (\omega_1, \ldots, \omega_n) \in \Omega^n$, the n-fold Cartesian product of Ω, let $\boldsymbol{P}_{\underline{\omega}}$ denote the product measure $\times_{\alpha=1}^n P_{\omega_\alpha}$, and let $\bar{\omega} := n^{-1}\sum_{\alpha=1}^n \omega_\alpha$ denote the empirical distribution (on Ω) of the mixing sequence $\underline{\omega}$.

By compounding the mixture model, i.e., considering $\underline{\omega}$, we are able to treat three different models — the Bayes model, the Bayes empirical Bayes model, and the Bayes compound model — as sub-models of our hierarchical Bayes model. Note that a similar hierarchical structure was considered by Datta (1991a), but he neither considered the Bayes model nor compounded the mixture model (and as a consequence had to handle the Bayes empirical Bayes model separately from the Bayes compound model). At the risk of belaboring the obvious, we would note that our sole objective behind creating this edifice, the compound mixture model sitting at the top of the hierarchical Bayes model, is to deal with the various sub-models of statistical importance described above with minimal effort. (That is not to say that the compound mixture model is devoid of any statistical importance, but it has not been studied in the literature and is a straightforward extension of the compounding idea to the mixture experiment.)

Let $\{P_\theta : \theta \in H\}$ be the Gaussian shift experiment (in the sense of LeCam, 1986), where H is a real separable Hilbert space, with p_θ denoting a density of P_θ wrt $P_{\mathbf{0}} := \mu$, where $\mathbf{0}$ is the origin of H. (We present a technical summary of the features of the Gaussian shift experiment necessary for this paper in Section 2. For references to the vast literature on the properties, examples, and applications of the Gaussian shift experiment, see Majumdar, 1996b, Section 2.) Note that for the Gaussian shift experiment, $\theta \mapsto P_\theta$ is continuous in the maximum variation norm (LeCam, 1986, p. 158), implying that $\forall\, A \in \mathscr{F}$, $\theta \mapsto P_\theta(A)$ is continuous, consequently Borel measuarble. Let Θ be a norm compact subset of H and let p_ω denote a density of P_ω wrt μ. Let $\mathscr{I}_\pi(\omega) := P_\pi(\ln(p_\pi/p_\omega))$ denote the Kullback–Leibler divergence between ω and π in Ω. For $\varepsilon > 0$, let $\mathscr{V}_\varepsilon(\pi) := \{\omega \in \Omega : \mathscr{I}_\pi(\omega) < \varepsilon\}$ denote a Kullback–Leibler neighborhood of π. Our main result (Theorem 3.1) asserts uniform (in $\underline{\omega}$ such that $\bar{\omega} \in S_\Lambda$) convergence of $\Lambda_{\underline{x}}((\mathscr{V}_\varepsilon(\bar{\omega}))^c)$ to 0 in $L_1(\boldsymbol{P}_{\underline{\omega}})$, where S_Λ is the smallest closed subset of Ω of Λ measure 1, the so-called topological support of Λ (see Majumdar, 1992).

For $f \in L_q(\mu)$, let $\|f\|_q$ denote its $L_q(\mu)$ norm. Proposition 3.1 shows that $\rho(\omega, \pi) := \|p_\omega - p_\pi\|_1$ defines a metric for weak convergence in Ω. Corollary 3.1 asserts uniform (in $\underline{\omega}$ such that $\bar{\omega} \in S_\Lambda$) convergence of $\rho(\widehat{\Lambda}, \bar{\omega})$ to 0 in $L_1(\boldsymbol{P}_{\underline{\omega}})$. Corollary 3.2 asserts consistency of the posterior distribution — uniform (in $\theta \in \Theta$) convergence of $\rho(\omega_{\underline{x}}, \delta_\theta)$ to 0 in $L_1(P_\theta^n)$, provided that $S_\omega = \Theta$, in which case $\|\widehat{\omega} - \theta\| \to 0$ in $L_1(P_\theta^n)$, uniformly in $\theta \in \Theta$, where the posterior mean $\widehat{\omega}$ of θ is defined as the Pettis integral of identity wrt $\omega_{\underline{x}}$ (Corollary 3.3). Corollary 3.4 obtains $L_1(\boldsymbol{P}_{\underline{\theta}})$ convergence of $\rho(\widehat{\Lambda}, G_n)$ to 0, uniformly in $\underline{\theta} \in \Theta^n$, if the hyperprior Λ satisfies $S_\Lambda = \Omega$, in which case $\rho(\widehat{\Lambda}, \omega) \to 0$ in $L_1(P_\omega^n)$, uniformly in $\omega \in \Omega$ (Corollary 3.5).

Posterior consistency results obtained in Corollaries 3.2, 3.3, and 3.5 can trace their roots back to Doob (1948), who considered a.s. asymptotic behavior. Almost sure posterior consistency in various models, including examples of inconsistency, has been extensively discussed by Diaconis and Freedman (1990, 1986a, 1986b). We consider the L_1 behavior primarily because of the applications of some of these consistency results to *compound decision* problems