Journal of the Korean Statistical Society 43 (2014) 189–200

Contents lists available at ScienceDirect

Journal of the Korean Statistical Society

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

Extension of some large deviation results for posterior distributions ☆

Claudio Macci *

Dipartimento di Matematica, Università di Roma Tor Vergata, Via della Ricerca Scientifica, I-00133 Roma, Italy

ARTICLE INFO

Article history:
Received 9 November 2012
Accepted 15 July 2013
Available online 7 August 2013

AMS 2000 subject classifications:
60F10
62F12
62F15
60G10
60G15

Keywords:
Large deviations
Stationary Gaussian process
Short-range and long-range dependence

ABSTRACT

We consider a family of statistical models with positive unknown parameter (which includes some well-known models for censored exponential data) and some statistical models for samples from stationary Gaussian processes. We prove large deviation results for posterior distributions and, in some cases, also for maximum likelihood estimators.

1. Introduction

The theory of large deviations gives an asymptotic computation of small probabilities on exponential scale. There is a wide range of applications where this theory plays a crucial role in solving problems of interest in several fields. Some applications have interest in statistics; here we recall the hypothesis testing problems studied in Sections 3.4 and 3.5 in Dembo and Zeitouni (1998), and the problems connected with applications in risk theory studied in Ganesh and O'Connell (1999, 2000) and Macci (2011).

In this paper we prove large deviation results for sequences of posterior distributions and, in some cases, also for maximum likelihood estimators. The results for posterior distributions follow the same lines of the ones in Ganesh and O'Connell (1999, 2000) and in Paschalidis and Vassilaras (2001); see also Eichelsbacher and Ganesh (2002a,b) with moderate deviation results. More recent references are Macci and Petrella (2006, 2009, 2010), where several results concern parametric models and finite mixtures of conjugate prior distributions (such a restriction allows to prove the results using the Gärtner Ellis Theorem).

A part of the results in this paper concerns stationary Gaussian processes; throughout this paper we use the symbol $\mathtt{GAUSS}(\mu, \Sigma)$ for the distribution of a stationary Gaussian process with constant mean μ and invertible covariance matrices $\Sigma = (\Sigma_n)$.

A contribution of this paper is the generalization of some results in Macci and Petrella (2009, 2010) without the restriction of finite mixtures of conjugate prior distributions; this has some analogies with what happens in Macci (2011) for the results in Macci and Petrella (2006). More precisely we consider the statistical model $\mathtt{GAUSS}(\mu, \Sigma)$ with unknown μ studied in

☆ The financial support of the Research Grant PRIN 2008 *Probability and Finance* is gratefully acknowledged.

* Tel.: +39 0672594728.
E-mail addresses: macci@mat.uniroma2.it, claudio.macci@gmail.com.

1226-3192/$ – see front matter © 2013 The Korean Statistical Society. Published by Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jkss.2013.07.004

Macci and Petrella (2010), and the statistical models with a quite general likelihood

$$L_n(\theta) = \exp\left(f_n\left\{\log\theta - \frac{\theta}{\widetilde{\theta}_n}\right\}\right) \quad (\theta \in (0,\infty)), \tag{1}$$

where f_n does not depend on θ (i.e. f_n is a constant or depends only on data), and $\widetilde{\theta}_n$ is the maximum likelihood estimator. Actually the likelihood (1) allows to recover the statistical models for censored data in Macci and Petrella (2009) and the statistical model $\mathtt{GAUSS}(\mu, r^{-1}\Sigma)$ with unknown r (which plays the role of θ) in Macci and Petrella (2010). In this paper we also present a result for the statistical model $\mathtt{GAUSS}(\mu, r^{-1}\Sigma)$ with unknown (μ, r), which was not considered in Macci and Petrella (2010).

The outline of the paper is the following. We start with some preliminaries in Section 2. The results for the statistical models with the likelihood (1) and for the statistical models for samples from stationary Gaussian processes are presented in Sections 3 and 4, respectively; their proofs (except the one of Proposition 4.2, which is an immediate consequence of Proposition 3.1) are presented in Section 6, after some concluding remarks in Section 5.

Finally we introduce some notation used throughout this paper. The family of all Borel subsets of a set S will be denoted by $\mathcal{B}(S)$. The neighborhood of a point x_0 (in some $\mathbb{R}^d$) and radius $\varepsilon > 0$ will be denoted by $B_\varepsilon(x_0)$; thus we have $B_\varepsilon(x_0) := [x_0 - \varepsilon, x_0 + \varepsilon]$ if $d = 1$. We use the symbol $N[\mu, \sigma^2]$ for the univariate Normal distribution with mean μ and variance σ^2 and the symbol $G[\alpha, \beta]$ for the Gamma distribution with continuous density $g_{\alpha,\beta}(y) = \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\beta y} 1_{(0,\infty)}(y)$ (where $\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y} dy$ is the Gamma function).

2. Preliminaries

We start with some preliminaries on large deviations; see Dembo and Zeitouni (1998) as a reference on this topic. Let Ω be a Hausdorff topological space with Borel σ-algebra $\mathcal{B}(\Omega)$; a lower semi-continuous function $I : \Omega \to [0, \infty]$ is called a rate function. Then a sequence of probability measures $\{\xi_n : n \geq 1\}$ on $(\Omega, \mathcal{B}(\Omega))$ satisfies the *large deviation principle* (LDP for short) with rate function I and speed v_n if $v_n \to \infty$ as $n \to \infty$,

$$\limsup_{n\to\infty} \frac{1}{v_n} \log \xi_n(C) \leq -\inf_{\omega\in C} I(\omega) \quad \text{for all closed sets } C \subset \Omega$$

and

$$\liminf_{n\to\infty} \frac{1}{v_n} \log \xi_n(G) \geq -\inf_{\omega\in G} I(\omega) \quad \text{for all open sets } G \subset \Omega.$$

In Proposition 4.3 below we prove a *weak large deviation principle*, i.e. the weaker version of LDP where the upper bound holds for all compact sets C only. The lower bound for the open sets is equivalent to the following condition:

$$\liminf_{n\to\infty} \frac{1}{v_n} log \xi_n(G) \geq -I(\omega) \quad \begin{array}{l}\text{for all } \omega \in \Omega \text{ such that } I(\omega) < \infty \text{ and}\\ \text{for all open sets } G \text{ such that } \omega \in G.\end{array}$$

The rate function I is said to be good if all its level sets $\{\{\omega \in \Omega : I(\omega) \leq \eta\} : \eta \geq 0\}$ are compact; throughout this paper we always have good rate functions. We also say that a sequence $\{Y_n : n \geq 1\}$ of Ω-valued random variables satisfies the LDP if the sequence $\{\xi_n : n \geq 1\}$ defined by $\xi_n(\cdot) = P(Y_n \in \cdot)$ satisfies the LDP; we refer to this definition when we consider LDPs for maximum likelihood estimators (MLEs from now on).

Throughout this paper we refer to the following known large deviation results (see e.g. Dembo & Zeitouni, 1998): Cramér Theorem (Section 2.2), Gärtner Ellis Theorem (Section 2.3), the contraction principle (Theorem 4.2.1).

On the LDPs for posterior distributions in this paper. In general we have a statistical model $\{P_\phi : \phi \in \Phi\}$ and the sequence of posterior distributions $\{\pi_n : n \geq 1\}$ on ϕ where, for each fixed $n \geq 1$, π_n depends on sampled data $y_1, \ldots, y_{s(n)}$, and $s(n)$ goes to infinity as $n \to \infty$. Moreover we always assume that the sampled values of consistent estimators $c_n(y_1, \ldots, y_{s(n)})$, say, converge to some value $\tilde{\phi}$ in the support of prior distribution. Then, if we think to have random variables $\{Y_n : n \geq 1\}$ with distribution $P_{\tilde{\phi}}$, and if we plug in $Y_1, \ldots, Y_{s(n)}$ into the data $y_1, \ldots, y_{s(n)}$, $\tilde{\phi}$ can be interpreted as the true value of the parameter ϕ.

On the support of a probability measure. Here we refer to Parthasarathy (1967, Section 2.2). The support $\mathbb{S}(\pi_0)$ of a probability measure π_0 is the smallest closed set having probability 1 with respect to π_0; moreover $\mathbb{S}(\pi_0)$ is the set of all the points θ such that $\pi_0(U) > 0$ for all open sets U containing θ. Throughout this paper π_0 is always a prior distribution concentrated on the parameter space; the parameter space is often an open set and we shall see that all the rate functions for posterior distributions are equal to infinity outside the intersection between the support of the prior distribution and the parameter space.

3. Statistical models with the likelihood (1)

In this section we consider a class of statistical models with likelihood (1) and we present the LDP for the sequence of posterior distributions. This is a wide class of statistical models and some examples will be presented at the end of this section.