Ann Inst Stat Math (2013) 65:703–719
DOI 10.1007/s10463-012-0389-2

Large deviations for posterior distributions on the parameter of a multivariate AR(p) process

Claudio Macci · Stefano Trapani

Received: 24 November 2011 / Revised: 2 July 2012 / Published online: 22 December 2012

Abstract We prove the large deviation principle for the posterior distributions on the (unknown) parameter of a multivariate autoregressive process with i.i.d. Normal innovations. As a particular case, we recover a previous result for univariate first-order autoregressive processes. We also show that the rate function can be expressed in terms of the divergence between two spectral densities.

Keywords Large deviation principle · Spectral density · Divergence · Relative entropy

1 Introduction

There is a wide literature in the topic of large deviations, namely, the asymptotic computation of small probabilities on an exponential scale. Some references concern posterior distributions in Bayesian Statistics: an old reference is Fu and Kass (1988); more recent references with large deviation principles for posterior distributions are Ganesh and O'Connell (1999, 2000), Paschalidis and Vassilaras (2001), Macci (2010a,b, 2011); other references with finite mixtures of conjugate prior distributions

CM gratefully acknowledges the financial support of the Research Grant PRIN 2008 *Probability and Finance*.

C. Macci (✉) · S. Trapani
Dipartimento di Matematica, Università di Roma Tor Vergata, Via della Ricerca Scientifica, 00133 Rome, Italy
e-mail: macci@mat.uniroma2.it

S. Trapani
e-mail: trapani@mat.uniroma2.it

are Macci and Petrella (2006, 2009, 2010); finally, we cite Eichelsbacher and Ganesh (2002a,b) with moderate deviation results.

In this paper, we consider a m-dimensional AR(p) process with i.i.d. Normal innovations, and the autoregressive parameters $\theta_1, \ldots, \theta_p$ are assumed to be unknown (a more rigorous definition can be found below). The aim is to prove the large deviation principle for the posterior distributions on the vector of autoregressive parameters. In particular, by setting $m = p = 1$, we recover Proposition 3.2 in Macci (2010b) for the univariate first-order autoregressive processes (see Remark 2 below). For completeness, we recall that one can find large and moderate deviation results for infinite-dimensional autoregressive processes in Mas and Menneteau (2003), but those results do not concern the Bayesian setting.

The divergence $I(f; g)$ of a spectral density f with respect to another one g (see Eq. (6) below for a precise definition) is an important tool in the asymptotic theory of stationary processes; for instance, as pointed out in several parts of Chapter 7 in Taniguchi and Kakizawa (2000), it plays a crucial role in the discriminant analysis. In this paper, we also show that the large deviation rate function can be expressed in terms of the function $I(f; g)$, where the spectral densities f and g concern m-dimensional AR(p) processes. This is not surprising if we take into account the relationship between $I(f; g)$ and the concept of relative entropy (see, e.g., section 2.3 in Cover and Thomas (1991); it is also known as the Kullback Leibler divergence) because, as illustrated for the examples presented in Varadhan (2003), large deviation rate functions are commonly expressed in terms of a relative entropy. Actually, this relationship is illustrated by the following limit for normalized relative entropies: under suitable hypotheses, by Eq. (7.6.6) in Theorem 7.6.5 in Taniguchi and Kakizawa (2000), we have

$$\lim_{n\to\infty} \frac{1}{n}\mathbb{E}_{P_n(f)}\left[\log\left(\frac{P_n(f)}{P_n(g)}\right)\right] = I(f; g),$$

where $P_n(f)$ (respectively $P_n(g)$) is the joint density of n random variables of a m-dimensional stationary process with spectral density f (respectively g) and $\mathbb{E}_{P_n(f)}[\cdot]$ is the expected value under the law with density $P_n(f)$.

The outline of the paper is the following. In Sect. 2, we recall some preliminaries. In Sect. 3, we prove the large deviation principle. In Sect. 4, we show that the rate function can be expressed in terms of the divergence between two spectral densities. The Appendix at the end of the paper gives the proofs of some properties of the set Θ defined by (2).

We conclude by recalling some notation used throughout the paper. We denote the open unit disc centered at the origin in the complex plane by D, i.e., $D := \{z \in \mathbb{C} : |z| < 1\}$. As far as the matrices are concerned, the families of real and complex square matrices of order m are denoted by $\mathcal{M}(m, \mathbb{R})$ and $\mathcal{M}(m, \mathbb{C})$, respectively; moreover, I_m is the identity matrix of order m and 0_m is the square null matrix of order m; finally, given a matrix $T = (t_{jh})_{j,h\in\{1,\ldots,m\}} \in \mathcal{M}(m, \mathbb{C})$, we denote its trace by $\operatorname{tr}(T)$, its transpose by T^t, its transpose and complex conjugate by T^*, and finally, we set $\|T\| := \sqrt{\sum_{j,h=1}^m |t_{jh}|^2}$. Throughout this paper, we always consider column vectors.

2 Preliminaries

2.1 Large deviations

We recall some basic definitions in Dembo and Zeitouni (1998). Let $\mathcal{Z}$ be a Hausdorff topological space with Borel σ-algebra $\mathcal{B}_{\mathcal{Z}}$. A lower semi-continuous function $I : \mathcal{Z} \to [0, \infty]$ is called rate function. A sequence of probability measures $\{\nu_n : n \geq 1\}$ on $(\mathcal{Z}, \mathcal{B}_{\mathcal{Z}})$ satisfies the *large deviation principle* (LDP for short), as $n \to \infty$, with rate function I if

$$\limsup_{n\to\infty} \frac{1}{n} \log \nu_n(F) \leq -\inf_{z\in F} I(z) \quad \text{for all closed sets } F$$

and

$$\liminf_{n\to\infty} \frac{1}{n} \log \nu_n(G) \geq -\inf_{z\in G} I(z) \quad \text{for all open sets } G.$$

A rate function I is said to be good if all the level sets $\{\{z \in \mathcal{Z} : I(z) \leq \gamma\} : \gamma \geq 0\}$ are compact. In what follows, we use condition (b) with equation (1.2.8) in Dembo and Zeitouni (1998), which is equivalent to the lower bound for open sets:

$$\liminf_{n\to\infty} \frac{1}{n} \log \nu_n(G) \geq -I(z) \quad \begin{array}{l}\text{for all } z \in \mathcal{Z} \text{ such that } I(z) < \infty \text{ and} \\ \text{for all open sets } G \text{ such that } z \in G.\end{array} \tag{1}$$

2.2 Multivariate autoregressive processes with i.i.d. Normal innovations

Let us consider $m, p \geq 1$ and $\theta_1, \ldots, \theta_p \in \mathcal{M}(m, \mathbb{R})$. Furthermore, we consider the matrix

$$\tilde{\theta}(z) := I_m - \sum_{j=1}^{p} \theta_j z^j$$

for each fixed $z \in \mathbb{C}$, and the set

$$\Theta := \{\theta = (\theta_1, \ldots, \theta_p)^t \in (\mathcal{M}(m, \mathbb{R}))^p : \det \tilde{\theta}(z) \neq 0 \text{ for all } z \in \overline{D}\}; \tag{2}$$

thus, $\theta = (\theta_1, \ldots, \theta_p)^t \in \Theta$ if and only if we have equation (1.2.25) in Example 1.2.17 in Taniguchi and Kakizawa (2000) with $q = 0$. One can check that Θ is bounded if and only if $m = 1$ and, moreover, Θ is open (some details of the proofs of these properties of Θ is given in the Appendix). In particular, we have $\Theta = (-1, 1)$ if $m = p = 1$.

It is known that, if $\theta = (\theta_1, \ldots, \theta_p)^t \in \Theta$, we can consider a stationary m-dimensional AR(p) process $\{X_n : n \geq 0\}$ defined by

$$X_{k+p} = \sum_{j=1}^{p} \theta_j X_{k+p-j} + Z_{k+p}$$

where $\{Z_k : k \geq 0\}$ are i.i.d. random variables with centered normal distribution and invertible covariance matrix Σ_Z^2, and X_0 has a suitable centered Normal distribution which depends on θ. In such a case, the spectral density matrix f_θ and the autocovariance matrix function Γ_θ are defined by

$$\begin{cases} \Gamma_\theta(r) = \int_{-\pi}^{\pi} e^{ir\lambda} f_\theta(\lambda) d\lambda \\ f_\theta(\lambda) = \frac{1}{2\pi}(\tilde{\theta}(e^{i\lambda}))^{-1} \Sigma_Z^2((\tilde{\theta}(e^{-i\lambda}))^t)^{-1} = \frac{1}{2\pi}(\tilde{\theta}(e^{i\lambda}))^{-1} \Sigma_Z^2((\tilde{\theta}(e^{i\lambda}))^*)^{-1}; \end{cases}$$

see, e.g., Example 1.2.17 in Taniguchi and Kakizawa (2000) with $q = 0$ for the first equality for f_θ. Moreover, if we consider the random variables $A_n^{(h,j)} := \frac{1}{n}\sum_{i=0}^{n-1} X_{i+p-h} X_{i+p-j}^t$ and $B_n^{(h)} := \frac{1}{n}\sum_{i=0}^{n-1} X_{i+p-h} X_{i+p}^t$, we have

$$\begin{cases} \lim_{n\to\infty} B_n^{(h)} = \Gamma_\theta(h) \\ \lim_{n\to\infty} A_n^{(h,j)} = \Gamma_\theta(h-j) \end{cases} \quad \text{with probability } 1. \tag{3}$$

In what follows, we consider a slightly different model, i.e. we require that the joint distribution of $(X_0, \ldots, X_{p-1})$ does not depend on θ. Actually in this case, the joint distribution of $(X_0, \ldots, X_{p-1})$ has no influence and the posterior distributions can be easily handled. We also note that we are changing the joint distribution of a finite subset of random variables of a stationary sequence, and this does not change the asymptotic behavior of the empirical means considered in the paper. In conclusion, all the limits of the empirical means considered below remain valid (as for the stationary case) and can be formulated in terms of the spectral density and the autocovariance function presented above.

2.3 Complex analysis

Some known results on complex analysis are needed in some proofs below. They are summarized in the following lemma.

Lemma 1 *(i) Let $\varphi : \overline{D} \to \mathbb{C}$ be a continuous function on $\overline{D}$ which is harmonic on D. Then $\frac{1}{2\pi}\int_{-\pi}^{\pi} \varphi(e^{i\lambda})d\lambda = \varphi(0)$. (ii) Let $\varphi : \overline{D} \to \mathbb{C}$ be a continuous function on $\overline{D}$ which is holomorphic on D. Then φ is harmonic on D. (iii) Let $\varphi : \overline{D} \to \mathbb{C} \setminus \{0\}$ be a continuous function on $\overline{D}$ which is holomorphic on D. Then* $\log|\varphi|$ *is harmonic on D.*

Proof (i) See, e.g., Gilbarg and Trudinger (1983), Theorem 2.1, Eq. (2.5) with $n = 2$ and $R = 1$. (ii) See, e.g., Rudin (1986), Theorem 11.4, noting that (as pointed out some lines above) φ is harmonic if and only if both the real part and the imaginary part of φ are harmonic. (iii) It is a consequence of Theorem 17.3 in Rudin (1986) applied to f and to $\frac{1}{f}$. □

3 LDP and remarks

For each fixed $n \geq 1$, we consider the log-likelihood based on $x_0, \ldots, x_{p+n-1}$, i.e.

$$\frac{1}{2}\sum_{i=0}^{n-1}\left(x_{i+p}-\sum_{j=1}^{p}\theta_j x_{i+p-j}\right)^t(\Sigma_Z^2)^{-1}\left(x_{i+p}-\sum_{h=1}^{p}\theta_h x_{i+p-h}\right)$$
$$=-\frac{1}{2}\sum_{i=0}^{n-1}x_{i+p}^t(\Sigma_Z^2)^{-1}x_{i+p}+\tilde{L}_n(\theta),$$

where

$$\tilde{L}_n(\theta) = \sum_{h=1}^{p}\sum_{i=0}^{n-1}x_{i+p}^t(\Sigma_Z^2)^{-1}\theta_h x_{i+p-h} -\frac{1}{2}\sum_{j=1}^{p}\sum_{h=1}^{p}\sum_{i=0}^{n-1}x_{i+p-j}^t\theta_j^t(\Sigma_Z^2)^{-1}\theta_h x_{i+p-h}.$$

Actually, since the addendum $-\frac{1}{2}\sum_{i=0}^{n-1}x_{i+p}^t(\Sigma_Z^2)^{-1}x_{i+p}$ does not depend on θ, the essential part of the log-likelihood is $\tilde{L}_n(\theta)$. Moreover, we introduce $a_n^{(h,j)} := \frac{1}{n}\sum_{i=0}^{n-1}x_{i+p-h}x_{i+p-j}^t$ and $b_n^{(h)} := \frac{1}{n}\sum_{i=0}^{n-1}x_{i+p-h}x_{i+p}^t$, i.e., the sampled values of the random variables $A_n^{(h,j)}$ and $B_n^{(h)}$, respectively. Then we have $\tilde{L}_n(\theta) = nL_n(\theta)$, where

$$L_n(\theta) = \sum_{h=1}^{p}\operatorname{tr}\left(b_n^{(h)}(\Sigma_Z^2)^{-1}\theta_h\right) - \frac{1}{2}\sum_{h=1}^{p}\sum_{j=1}^{p}\operatorname{tr}\left(a_n^{(h,j)}\theta_j^t(\Sigma_Z^2)^{-1}\theta_h\right)$$
$$= -\frac{1}{2}\left\{\sum_{h=1}^{p}\sum_{j=1}^{p}\operatorname{tr}\left(a_n^{(h,j)}\theta_j^t(\Sigma_Z^2)^{-1}\theta_h\right) - 2\sum_{h=1}^{p}\operatorname{tr}\left(b_n^{(h)}(\Sigma_Z^2)^{-1}\theta_h\right)\right\}.$$

Furthermore, for a prior distribution π_0 on the parameter space Θ, we consider the posterior distributions $\{\pi_n : n \geq 1\}$ defined by

$$\pi_n(E) := \pi_0(E|x_0, \ldots, x_{p+n-1}) = \frac{\int_E e^{nL_n(\theta)}\pi_0(d\theta)}{\int_\Theta e^{nL_n(\theta)}\pi_0(d\theta)} \text{ for all Borel sets } E \subset \mathcal{B}_\Theta.$$

We often deal with the support of the prior distribution, and we refer to Sect. 2.2 in Parthasarathy (1967). The support of π_0 will be denoted by $\mathbb{S}(\pi_0)$; thus, it 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 $\pi_0(G) > 0$ for all open sets G containing θ.

The aim of this section is to prove Proposition 1 which provides the LDP for posterior distributions. The proof of Proposition 1 will be an immediate consequence of Lemmas 3 and 4 presented below; in these two lemmas, we refer to the following notation $L_\infty(\theta;\theta^{(0)})$, where $\theta,\theta^{(0)}\in\Theta$:

$$L_\infty(\theta;\theta^{(0)}) := -\frac{1}{2}\left\{\sum_{h=1}^{p}\sum_{j=1}^{p}\mathrm{tr}\left(\Gamma_{\theta^{(0)}}(h-j)\theta_j^t(\Sigma_Z^2)^{-1}\theta_h\right)\right.$$
$$\left.-2\sum_{h=1}^{p}\mathrm{tr}\left(\Gamma_{\theta^{(0)}}(h)(\Sigma_Z^2)^{-1}\theta_h\right)\right\}'. \tag{4}$$

Proposition 1 *Assume the two following conditions: $\mathbb{S}(\pi_0)$ is a compact subset of $(\mathcal{M}(m,\mathbb{R}))^p$; there exists $\theta^{(0)}\in\mathbb{S}(\pi_0)$ such that, for all $j,h\in\{1,\ldots,p\}$, $a_n^{(h,j)}\to\Gamma_{\theta^{(0)}}(h-j)$ and $b_n^{(h)}\to\Gamma_{\theta^{(0)}}(h)$ (as $n\to\infty$). Then $\{\pi_n : n\geq 1\}$ satisfies the LDP with good rate function $I(\cdot|\theta^{(0)})$ defined by*

$$I(\theta|\theta^{(0)}) = \begin{cases} \frac{1}{2}\sum_{h=1}^{p}\sum_{j=1}^{p}\mathrm{tr}\Big(\Gamma_{\theta^{(0)}}(h-j) & \\ \quad\cdot(\theta_j-\theta_j^{(0)})^t(\Sigma_Z^2)^{-1}(\theta_h-\theta_h^{(0)})\Big) & \text{if } \theta\in\mathbb{S}(\pi_0) \\ \infty & \text{otherwise.}\end{cases}$$

Remark 1 (i) We already remarked that the set Θ is bounded if $m=1$; thus, in such a case, $\mathbb{S}(\pi_0)$ is bounded and, therefore, we require that $\mathbb{S}(\pi_0)$ is a closed subset of $(\mathcal{M}(m,\mathbb{R}))^p$.
(ii) Assume that $\{\tilde{X}_n : n\geq 1\}$ is a version of the model where θ is known, and denote by $\theta^{(0)}$ the true value of the parameter; then, by (3), we have $\tilde{B}_n^{(h)} := \frac{1}{n}\sum_{i=0}^{n-1}\tilde{X}_{i+p-h}\tilde{X}_{i+p}^t \to \Gamma_{\theta^{(0)}}(h)$ and $\tilde{A}_n^{(h,j)} := \frac{1}{n}\sum_{i=0}^{n-1}\tilde{X}_{i+p-h}\tilde{X}_{i+p-j}^t \to \Gamma_{\theta^{(0)}}(h-j)$ with probability 1. So, by plugging $(\tilde{A}_n^{(h,j)},\tilde{B}_n^{(h)})$ into $(a_n^{(h,j)},b_n^{(h)})$ for all $j,h\in\{1,\ldots,p\}$, $\theta^{(0)}$ can be actually interpreted as the true value of the parameter.
(iii) It is easy to check that, for any fixed $\sigma^2>0$, the rate function $I(\theta|\theta^{(0)})$ does not change if we consider $\sigma^2\Sigma_Z^2$ in place of Σ_Z^2. In some sense, this is the analogous of Remark 3.3(ii) in Macci (2010b) for the case $m=p=1$.

We start with Lemma 2 which provides a standard result; this lemma is useful for the proof of Lemma 3. An interested reader can see Lemma 3.1 in Macci (2010b) for its proof.

Lemma 2 *Let $f : C_1\times C_2\to\mathbb{R}$ be a continuous function and let C_2 be a compact set. Then, if $\lim_{n\to\infty}x_n = x$, we have $\lim_{n\to\infty}\inf_{\theta\in C_2} f(x_n,\theta) = \inf_{\theta\in C_2} f(x,\theta)$.*

Lemma 3 *Assume the same hypotheses of Proposition 1. Then $\{\pi_n : n\geq 1\}$ satisfies the LDP with good rate function $I(\cdot|\theta^{(0)})$ defined by*

$$I(\theta|\theta^{(0)}) = \begin{cases} L_\infty(\theta^{(0)};\theta^{(0)}) - L_\infty(\theta;\theta^{(0)}) & \text{if } \theta\in\mathbb{S}(\pi_0) \\ \infty & \text{otherwise.}\end{cases} \tag{5}$$

Proof The goodness of the rate function is guaranteed by the compactness of $\mathbb{S}(\pi_0)$. Throughout this proof, we think $(\mathcal{M}(m, \mathbb{R}))^p$ equipped with the standard Euclidean distance and we use the symbol $Q_\varepsilon(\theta)$ for the closed ball of the point $\theta \in \Theta$ of radius $\varepsilon > 0$. The proof of the LDP consists of two parts.

Proof of the upper bound for closed sets. Let C be a closed set. The upper bound trivially holds if $\pi_0(C) = 0$, because we have $\pi_n(C) = 0$ for all $n \geq 1$. Thus, from now on we assume $\pi_0(C) > 0$. Then, since, $\pi_0(Q_\varepsilon(\theta^{(0)})) > 0$ (indeed $\theta^{(0)} \in \mathbb{S}(\pi_0)$) for any $\varepsilon > 0$, we have

$$\pi_n(C) \leq \frac{e^{\sup_{\theta \in C \cap \mathbb{S}(\pi_0)} nL_n(\theta)} \pi_0(C)}{\int_{Q_\varepsilon(\theta^{(0)})} e^{nL_n(\theta)} \pi_0(d\theta)} \leq \frac{e^{-\inf_{\theta \in C \cap \mathbb{S}(\pi_0)} -nL_n(\theta)}}{e^{\inf_{\theta \in Q_\varepsilon(\theta^{(0)})} nL_n(\theta)} \pi_0(Q_\varepsilon(\theta^{(0)}))},$$

whence we obtain

$$\frac{1}{n} \log \pi_n(C) \leq -\inf_{\theta \in C \cap \mathbb{S}(\pi_0)} (-L_n(\theta)) - \inf_{\theta \in Q_\varepsilon(\theta^{(0)})} L_n(\theta) - \frac{1}{n} \log \pi_0(Q_\varepsilon(\theta^{(0)})).$$

Moreover, $L_n(\theta) \to L_\infty(\theta; \theta^{(0)})$ as $n \to \infty$ and, since, $C \cap \mathbb{S}(\pi_0)$ is a compact set, by Lemma 2 we get

$$\limsup_{n \to \infty} \frac{1}{n} \log \pi_n(C) \leq -\inf_{\theta \in C \cap \mathbb{S}(\pi_0)} (-L_\infty(\theta; \theta^{(0)})) - \inf_{\theta \in Q_\varepsilon(\theta^{(0)})} L_\infty(\theta; \theta^{(0)});$$

finally, if we let ε go to zero, we have

$$\begin{aligned}
\limsup_{n \to \infty} \frac{1}{n} \log \pi_n(C) &\leq -\inf_{\theta \in C \cap \mathbb{S}(\pi_0)} (-L_\infty(\theta; \theta^{(0)})) - L_\infty(\theta^{(0)}; \theta^{(0)}) \\
&= -\inf_{\theta \in C \cap \mathbb{S}(\pi_0)} (L_\infty(\theta^{(0)}; \theta^{(0)}) - L_\infty(\theta; \theta^{(0)})) \\
&= -\inf_{\theta \in C} I(\theta | \theta^{(0)}).
\end{aligned}$$

Proof of the lower bound for open sets. We want to check condition (1). Then let $\theta^{(*)} \in \mathbb{S}(\pi_0)$ be such that $I(\theta^{(*)} | \theta^{(0)}) < \infty$. Moreover, let G be an open set such that $\theta^{(*)} \in G$. Then, for ε small enough to have $Q_\varepsilon(\theta^{(*)}) \subset G$ (we recall that $\pi_0(Q_\varepsilon(\theta^{(*)})) > 0$, since, $\theta^{(*)} \in \mathbb{S}(\pi_0)$), we have

$$\pi_n(G) \geq \frac{\int_{Q_\varepsilon(\theta^{(*)})} e^{nL_n(\theta)} d\pi_0(\theta)}{e^{\sup_{\theta \in \Theta} nL_n(\theta)}} \geq \frac{e^{\inf_{\theta \in Q_\varepsilon(\theta^{(*)})} nL_n(\theta)} \pi_0(Q_\varepsilon(\theta^{(*)}))}{e^{nL_n(\hat{\theta}^{(n)})}},$$

where $\hat{\theta}^{(n)}$ is the maximum likelihood estimator, i.e. the value such that $\sup_{\theta \in \Theta} L_n(\theta) = L_n(\hat{\theta}^{(n)})$; hence, we obtain

$$\frac{1}{n} \log \pi_n(G) \geq \inf_{\theta \in Q_\varepsilon(\theta^{(*)})} L_n(\theta) + \frac{1}{n} \log \pi_0(Q_\varepsilon(\theta^{(*)})) - L_n(\hat{\theta}^{(n)}).$$

Moreover, by the hypotheses of Proposition 1 and the consistency of the maximum likelihood estimator, we have $L_n(\hat{\theta}^{(n)}) \to L_\infty(\theta^{(0)}; \theta^{(0)})$ as $n \to \infty$. Then, by Lemma 2, we get

$$\liminf_{n\to\infty} \frac{1}{n} \log \pi_n(G) \geq \inf_{\theta \in Q_\varepsilon(\theta^{(*)})} L_\infty(\theta; \theta^{(0)}) - L_\infty(\theta^{(0)}; \theta^{(0)})$$

and, if we let ε go to zero, we have

$$\begin{aligned}\liminf_{n\to\infty} \frac{1}{n} \log \pi_n(G) &\geq L_\infty(\theta^{(*)}; \theta^{(0)}) - L_\infty(\theta^{(0)}; \theta^{(0)}) \\ &= -I(\theta^{(*)}|\theta^{(0)}).\end{aligned}$$

□

Lemma 4 *For all $\theta, \theta^{(0)} \in \Theta$ we have*

$$\begin{aligned}&L_\infty(\theta^{(0)}; \theta^{(0)}) - L_\infty(\theta; \theta^{(0)}) \\ &\quad = \frac{1}{2}\sum_{h=1}^{p}\sum_{j=1}^{p} \mathrm{tr}\left(\Gamma_{\theta^{(0)}}(h-j)(\theta_j - \theta_j^{(0)})^t (\Sigma_Z^2)^{-1}(\theta_h - \theta_h^{(0)})\right).\end{aligned}$$

Proof As far as the equality in the statement of the lemma is concerned, we start with the following explicit expressions for the left-hand side and the right-hand side, denoted by Q_1 and Q_2, respectively:

$$\begin{aligned}Q_1 &= -\frac{1}{2}\sum_{h=1}^{p}\sum_{j=1}^{p} \mathrm{tr}\left(\Gamma_{\theta^{(0)}}(h-j)(\theta_j^{(0)})^t (\Sigma_Z^2)^{-1}\theta_h^{(0)}\right) \\ &\quad + \sum_{h=1}^{p} \mathrm{tr}\left(\Gamma_{\theta^{(0)}}(h)(\Sigma_Z^2)^{-1}\theta_h^{(0)}\right) \\ &\quad + \frac{1}{2}\sum_{h=1}^{p}\sum_{j=1}^{p} \mathrm{tr}\left(\Gamma_{\theta^{(0)}}(h-j)\theta_j^t (\Sigma_Z^2)^{-1}\theta_h\right) \\ &\quad - \sum_{h=1}^{p} \mathrm{tr}\left(\Gamma_{\theta^{(0)}}(h)(\Sigma_Z^2)^{-1}\theta_h\right); \\ Q_2 &= \frac{1}{2}\sum_{h=1}^{p}\sum_{j=1}^{p} \mathrm{tr}\left(\Gamma_{\theta^{(0)}}(h-j)\theta_j^t (\Sigma_Z^2)^{-1}\theta_h\right) \\ &\quad \underbrace{-\frac{1}{2}\sum_{h=1}^{p}\sum_{j=1}^{p} \mathrm{tr}\left(\Gamma_{\theta^{(0)}}(h-j)(\theta_j^{(0)})^t (\Sigma_Z^2)^{-1}\theta_h\right)}_{=:R_1}\end{aligned}$$

$$
\begin{aligned}
&\underbrace{-\frac{1}{2}\sum_{h=1}^{p}\sum_{j=1}^{p}\mathrm{tr}\left(\Gamma_{\theta^{(0)}}(h-j)\theta_j^t(\Sigma_Z^2)^{-1}\theta_h^{(0)}\right)}_{=:R_2}\\
&+\frac{1}{2}\sum_{h=1}^{p}\sum_{j=1}^{p}\mathrm{tr}\left(\Gamma_{\theta^{(0)}}(h-j)(\theta_j^{(0)})^t(\Sigma_Z^2)^{-1}\theta_h^{(0)}\right).
\end{aligned}
$$

Moreover, we can check that R_1 and R_2 in the latter expression coincide and, in view of this, we recall that $\Gamma_{\theta^{(0)}}(r) = (\Gamma_{\theta^{(0)}}(-r))^t$:

$$
\begin{aligned}
R_1 &= \sum_{h=1}^{p}\sum_{j=1}^{p}\mathrm{tr}\left(\theta_h^t(\Sigma_Z^2)^{-1}\theta_j^{(0)}(\Gamma_{\theta^{(0)}}(h-j))^t\right)\\
&= \sum_{h=1}^{p}\sum_{j=1}^{p}\mathrm{tr}\left(\theta_h^t(\Sigma_Z^2)^{-1}\theta_j^{(0)}\Gamma_{\theta^{(0)}}(j-h)\right)\\
&= \sum_{h=1}^{p}\sum_{j=1}^{p}\mathrm{tr}\left(\Gamma_{\theta^{(0)}}(j-h)\theta_h^t(\Sigma_Z^2)^{-1}\theta_j^{(0)}\right) = R_2.
\end{aligned}
$$

Thus,

$$
\begin{aligned}
Q_2 &= \frac{1}{2}\sum_{h=1}^{p}\sum_{j=1}^{p}\mathrm{tr}\left(\Gamma_{\theta^{(0)}}(h-j)\theta_j^t(\Sigma_Z^2)^{-1}\theta_h\right)\\
&\quad -\sum_{h=1}^{p}\sum_{j=1}^{p}\mathrm{tr}\left(\Gamma_{\theta^{(0)}}(h-j)(\theta_j^{(0)})^t(\Sigma_Z^2)^{-1}\theta_h\right)\\
&\quad +\frac{1}{2}\sum_{h=1}^{p}\sum_{j=1}^{p}\mathrm{tr}\left(\Gamma_{\theta^{(0)}}(h-j)(\theta_j^{(0)})^t(\Sigma_Z^2)^{-1}\theta_h^{(0)}\right),
\end{aligned}
$$

whence we obtain

$$
\begin{aligned}
Q_2 &= \frac{1}{2}\sum_{h=1}^{p}\sum_{j=1}^{p}\mathrm{tr}\left(\Gamma_{\theta^{(0)}}(h-j)\theta_j^t(\Sigma_Z^2)^{-1}\theta_h\right)\\
&\quad -\sum_{h=1}^{p}\sum_{j=1}^{p}\mathrm{tr}\left(\Gamma_{\theta^{(0)}}(h-j)(\theta_j^{(0)})^t(\Sigma_Z^2)^{-1}\theta_h\right)\\
&\quad -\frac{1}{2}\sum_{h=1}^{p}\sum_{j=1}^{p}\mathrm{tr}\left(\Gamma_{\theta^{(0)}}(h-j)(\theta_j^{(0)})^t(\Sigma_Z^2)^{-1}\theta_h^{(0)}\right)\\
&\quad +\sum_{h=1}^{p}\sum_{j=1}^{p}\mathrm{tr}\left(\Gamma_{\theta^{(0)}}(h-j)(\theta_j^{(0)})^t(\Sigma_Z^2)^{-1}\theta_h^{(0)}\right).
\end{aligned}
$$

Thus, by comparing the latter equality for Q_2 and the expression for Q_1 at the beginning of the proof, the equality $Q_1 = Q_2$ will be proved if we can show that

$$\sum_{h=1}^{p} \operatorname{tr}\left(\Gamma_{\theta^{(0)}}(h)(\Sigma_Z^2)^{-1}\theta_h^{(0)}\right) - \sum_{h=1}^{p} \operatorname{tr}\left(\Gamma_{\theta^{(0)}}(h)(\Sigma_Z^2)^{-1}\theta_h\right)$$

coincides with

$$\sum_{h=1}^{p}\sum_{j=1}^{p} \operatorname{tr}\left(\Gamma_{\theta^{(0)}}(h-j)(\theta_j^{(0)})^t(\Sigma_Z^2)^{-1}\theta_h^{(0)}\right)$$
$$-\sum_{h=1}^{p}\sum_{j=1}^{p} \operatorname{tr}\left(\Gamma_{\theta^{(0)}}(h-j)(\theta_j^{(0)})^t(\Sigma_Z^2)^{-1}\theta_h\right);$$

this is equivalent to the following equality

$$\sum_{h=1}^{p} \operatorname{tr}\left(\Gamma_{\theta^{(0)}}(h)(\Sigma_Z^2)^{-1}(\theta_h^{(0)} - \theta_h)\right)$$
$$= \sum_{h=1}^{p} \operatorname{tr}\left(\sum_{j=1}^{p} \Gamma_{\theta^{(0)}}(h-j)(\theta_j^{(0)})^t(\Sigma_Z^2)^{-1}(\theta_h^{(0)} - \theta_h)\right),$$

which will be proved by showing that

$$\sum_{j=1}^{p} \Gamma_{\theta^{(0)}}(h-j)(\theta_j^{(0)})^t = \Gamma_{\theta^{(0)}}(h) \text{ for all } h \in \{1, \ldots, p\}.$$

This can be checked as follows. First of all, we have

$$\begin{aligned}
\sum_{j=1}^{p} \Gamma_{\theta^{(0)}}(h-j)(\theta_j^{(0)})^t &= \sum_{j=1}^{p} \int_{-\pi}^{\pi} e^{i(h-j)\lambda} f_{\theta^{(0)}}(\lambda)d\lambda(\theta_j^{(0)})^t \\
&= \int_{-\pi}^{\pi} e^{ih\lambda} f_{\theta^{(0)}}(\lambda) \sum_{j=1}^{p} e^{-ij\lambda}(\theta_j^{(0)})^t d\lambda \\
&= \int_{-\pi}^{\pi} e^{ih\lambda} f_{\theta^{(0)}}(\lambda)(-\tilde{\theta}^{(0)}(e^{-i\lambda}) + I_m)^t d\lambda \\
&= \int_{-\pi}^{\pi} e^{ih\lambda} f_{\theta^{(0)}}(\lambda)d\lambda \\
&\quad - \int_{-\pi}^{\pi} e^{ih\lambda} f_{\theta^{(0)}}(\lambda)(\tilde{\theta}^{(0)}(e^{-i\lambda}))^t d\lambda \\
&= \Gamma_{\theta^{(0)}}(h) - \int_{-\pi}^{\pi} e^{ih\lambda} \frac{1}{2\pi}
\end{aligned}$$

$$\cdot(\tilde{\theta}^{(0)}(e^{i\lambda}))^{-1}\Sigma_Z^2((\tilde{\theta}^{(0)}(e^{-i\lambda}))^t)^{-1}(\tilde{\theta}^{(0)}(e^{-i\lambda}))^t d\lambda$$
$$= \Gamma_{\theta^{(0)}}(h) - \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{ih\lambda}(\tilde{\theta}^{(0)}(e^{i\lambda}))^{-1}d\lambda\Sigma_Z^2;$$

moreover, since the function $z \mapsto z^h(\tilde{\theta}^{(0)}(z))^{-1}$ is continuous on $\overline{D}$ and holomorphic on D, and by noting that $z^h(\tilde{\theta}^{(0)}(z))^{-1}\Big|_{z=0} = 0_m$, we have

$$\int_{-\pi}^{\pi} e^{ih\lambda}(\tilde{\theta}^{(0)}(e^{i\lambda}))^{-1}d\lambda = 0_m$$

by Lemma 1(i–ii) applied to each component of the matrix. □

Remark 2 (Proposition 3.2 in Macci (2010b) as a particular case) Here, we consider Proposition 1 with $m = p = 1$, i.e. the case of univariate first-order autoregressive processes. Actually, Σ_Z^2 and $\Gamma_{\theta^{(0)}}(r)$ are numbers, and Σ_Z^2 is positive; moreover, $\theta^{(0)} \in \Theta = (-1, 1)$. Here, for simplicity, we write σ^2, $\gamma_{\theta_0}(r)$ and θ_0 in place of Σ_Z^2, $\Gamma_{\theta^{(0)}}(r)$ and $\theta^{(0)}$, respectively; moreover, we recall that $\gamma_{\theta_0}(0) = \frac{\sigma^2}{1-\theta_0^2}$ and $\gamma_{\theta_0}(1) = \frac{\sigma^2\theta_0}{1-\theta_0^2}$. Then, we have the same rate function in Proposition 3.2 in Macci (2010b) noting that

$$\frac{1}{2}\gamma_{\theta_0}(0)(\theta-\theta_0)\frac{1}{\sigma^2}(\theta-\theta_0) = \frac{1}{2}\frac{\sigma^2}{1-\theta_0^2}\frac{(\theta-\theta_0)^2}{\sigma^2} = \frac{(\theta-\theta_0)^2}{2(1-\theta_0^2)}.$$

Note that we can meet the rate function in Proposition 3.2 in Macci (2010b) in a different way: by (4) we have

$$\begin{aligned} L_\infty(\theta;\theta_0) &= -\frac{1}{2}\left\{\gamma_{\theta_0}(0)\frac{\theta^2}{\sigma^2} - 2\gamma_{\theta_0}(1)\frac{\theta}{\sigma^2}\right\} \\ &= -\frac{1}{2}\left\{\frac{\theta^2}{1-\theta_0^2} - 2\frac{\theta_0\theta}{1-\theta_0^2}\right\} = -\frac{\theta^2-2\theta_0\theta}{2(1-\theta_0^2)} \end{aligned}$$

and, by (5), we obtain

$$L_\infty(\theta_0;\theta_0) - L_\infty(\theta;\theta_0) = -\frac{(\theta_0)^2-2(\theta_0\theta_0)}{2(1-\theta_0^2)} + \frac{\theta^2-2\theta_0\theta}{2(1-\theta_0^2)} = \frac{(\theta-\theta_0)^2}{2(1-\theta_0^2)}.$$

4 The rate function as the divergence between spectral densities

We recall the definition of the divergence $I(f;g)$ of a spectral density matrix f with respect to another one g, and some properties; see, e.g., Eq. (7.3.7) in Taniguchi and Kakizawa (2000) and the statements after that formula. We have

$$I(f;g) = \frac{1}{4\pi}\int_{-\pi}^{\pi}\left\{\mathrm{tr}\{g^{-1}(\lambda)f(\lambda)\} - m - \log\det\{g^{-1}(\lambda)f(\lambda)\}\right\}d\lambda; \tag{6}$$

moreover, $I(f;g) \geq 0$ and the equality $I(f;g) = 0$ holds if and only if $f(\lambda) = g(\lambda)$ almost everywhere in $[-\pi, \pi]$.

The aim of this section is to prove Proposition 2 which provides an interesting equality; as a consequence of this we can say that, if θ belongs to $\mathbb{S}(\pi_0)$ (i.e. the support of the prior distribution π_0), the rate function for the LDP of the family of posterior distributions $I(\theta|\theta^{(0)})$ is equal to the divergence between spectral densities $I(f_{\theta^{(0)}}; f_\theta)$.

Proposition 2 *For all $\theta, \theta^{(0)} \in \Theta$ we have*

$$I(f_{\theta^{(0)}}; f_\theta) = \frac{1}{2} \sum_{h=1}^{p} \sum_{j=1}^{p} \mathrm{tr}\left(\Gamma_{\theta^{(0)}}(h-j)(\theta_j - \theta_j^{(0)})^t (\Sigma_Z^2)^{-1}(\theta_h - \theta_h^{(0)})\right).$$

The proof of Proposition 2 is presented below after the following lemma.

Lemma 5 *For all $\theta, \theta^{(0)} \in \Theta$ we have*

$$\int_{-\pi}^{\pi} \log\det\{f_\theta^{-1}(\lambda) f_{\theta^{(0)}}(\lambda)\} d\lambda = 0.$$

Proof First of all, we have

$$\begin{aligned} f_\theta^{-1}(\lambda) f_{\theta^{(0)}}(\lambda) &= \left(\frac{1}{2\pi}(\tilde{\theta}(e^{i\lambda}))^{-1}\Sigma_Z^2((\tilde{\theta}(e^{i\lambda}))^*)^{-1}\right)^{-1} \\ &\quad \cdot \frac{1}{2\pi}(\tilde{\theta}^{(0)}(e^{i\lambda}))^{-1}\Sigma_Z^2((\tilde{\theta}^{(0)}(e^{i\lambda}))^*)^{-1} \\ &= (\tilde{\theta}(e^{i\lambda}))^*(\Sigma_Z^2)^{-1}\tilde{\theta}(e^{i\lambda})(\tilde{\theta}^{(0)}(e^{i\lambda}))^{-1}\Sigma_Z^2((\tilde{\theta}^{(0)}(e^{i\lambda}))^*)^{-1}, \end{aligned}$$

whence we obtain

$$\begin{aligned} \det\left(f_\theta^{-1}(\lambda) f_{\theta^{(0)}}(\lambda)\right) &= \det\left((\tilde{\theta}(e^{i\lambda}))^*\right) \det\tilde{\theta}(e^{i\lambda}) \\ &\quad \cdot \det\left((\tilde{\theta}^{(0)}(e^{i\lambda}))^{-1}\right) \det\left(((\tilde{\theta}^{(0)}(e^{i\lambda}))^*)^{-1}\right) \\ &= \overline{\det\tilde{\theta}(e^{i\lambda})} \det\tilde{\theta}(e^{i\lambda}) \det\left(((\tilde{\theta}^{(0)}(e^{i\lambda}))^*\tilde{\theta}^{(0)}(e^{i\lambda}))^{-1}\right) \\ &= \overline{\det\tilde{\theta}(e^{i\lambda})} \det\tilde{\theta}(e^{i\lambda}) \left(\overline{\det\tilde{\theta}^{(0)}(e^{i\lambda})} \det\tilde{\theta}^{(0)}(e^{i\lambda})\right)^{-1} \\ &= |\det\tilde{\theta}(e^{i\lambda})|^2 \left(|\det\tilde{\theta}^{(0)}(e^{i\lambda})|^2\right)^{-1} \end{aligned}$$

and

$$\log\det\left(f_\theta^{-1}(\lambda) f_{\theta^{(0)}}(\lambda)\right) = \log|\det\tilde{\theta}(e^{i\lambda})|^2 - \log\det|\tilde{\theta}^{(0)}(e^{i\lambda})|^2.$$

We complete the proof showing that the integral $\int_{-\pi}^{\pi} \log|\det\tilde{\theta}(e^{i\lambda})|^2 d\lambda$ does not depend on $\theta \in \Theta$; indeed, in such a case, we have the same value if we take $\theta^{(0)} \in \Theta$

in place of θ, and the integral on $[-\pi, \pi]$ of the difference in the latter right-hand side is zero. This can be done noting that, since, $\det\tilde{\theta}(z) \neq 0$ for all $z \in D$, the function $z \mapsto \det\tilde{\theta}(z)$ is continuous on $\overline{D}$ and holomorphic on D; therefore,

$$\int_{-\pi}^{\pi} \log|\det\tilde{\theta}(e^{i\lambda})|^2 d\lambda = 2\pi \log|\det\tilde{\theta}(0)|^2 = 2\pi$$

by Lemma 1(i–iii). □

Proof of Proposition 2 First of all, by (6) and Lemma 5, we have

$$I(f_{\theta^{(0)}}; f_\theta) = \frac{1}{4\pi}\int_{-\pi}^{\pi}\left\{\mathrm{tr}\{f_\theta^{-1}(\lambda) f_{\theta^{(0)}}(\lambda)\} - m\right\} d\lambda. \tag{7}$$

It is known that, there exists a symmetric and positive definite matrix Σ_Z such that $\Sigma_Z^2 = \Sigma_Z \cdot \Sigma_Z$. Throughout this proof, we use the notation

$$T_{\theta,\theta^{(0)}}(z) := \Sigma_Z^{-1}\tilde{\theta}(z)(\Sigma_Z^{-1}\tilde{\theta}^{(0)}(z))^{-1};$$

therefore, we have

$$\begin{aligned}(T_{\theta,\theta^{(0)}}(e^{i\lambda}))^* &= (\Sigma_Z^{-1}\tilde{\theta}(e^{i\lambda})(\Sigma_Z^{-1}\tilde{\theta}^{(0)}(e^{i\lambda}))^{-1})^* \\ &= ((\Sigma_Z^{-1}\tilde{\theta}^{(0)}(e^{i\lambda}))^{-1})^*(\Sigma_Z^{-1}\tilde{\theta}(e^{i\lambda}))^* \\ &= ((\Sigma_Z^{-1}\tilde{\theta}^{(0)}(e^{i\lambda}))^*)^{-1}(\Sigma_Z^{-1}\tilde{\theta}(e^{i\lambda}))^*.\end{aligned}$$

Then, by considering an equality presented at the beginning of the proof of Lemma 5, we have

$$\begin{aligned}f_\theta^{-1}(\lambda) f_{\theta^{(0)}}(\lambda) &= \underbrace{(\tilde{\theta}(e^{i\lambda}))^*\Sigma_Z^{-1}}_{=(\Sigma_Z^{-1}\tilde{\theta}(e^{i\lambda}))^*}\Sigma_Z^{-1}\tilde{\theta}(e^{i\lambda}) \\ &\quad \cdot \underbrace{(\tilde{\theta}^{(0)}(e^{i\lambda}))^{-1}\Sigma_Z}_{=(\Sigma_Z^{-1}\tilde{\theta}^{(0)}(e^{i\lambda}))^{-1}}\underbrace{\Sigma_Z((\tilde{\theta}^{(0)}(e^{i\lambda}))^*)^{-1}}_{=((\tilde{\theta}^{(0)}(e^{i\lambda}))^*\Sigma_Z^{-1})^{-1}} \\ &= (\Sigma_Z^{-1}\tilde{\theta}(e^{i\lambda}))^*\Sigma_Z^{-1}\tilde{\theta}(e^{i\lambda}) \\ &\quad \cdot (\Sigma_Z^{-1}\tilde{\theta}^{(0)}(e^{i\lambda}))^{-1}((\Sigma_Z^{-1}\tilde{\theta}^{(0)}(e^{i\lambda}))^*)^{-1},\end{aligned}$$

and, therefore,

$$\begin{aligned}\mathrm{tr}\{f_\theta^{-1}(\lambda) f_{\theta^{(0)}}(\lambda)\} &= \mathrm{tr}\{(\Sigma_Z^{-1}\tilde{\theta}(e^{i\lambda}))^*\Sigma_Z^{-1}\tilde{\theta}(e^{i\lambda}) \\ &\quad \cdot(\Sigma_Z^{-1}\tilde{\theta}^{(0)}(e^{i\lambda}))^{-1}((\Sigma_Z^{-1}\tilde{\theta}^{(0)}(e^{i\lambda}))^*)^{-1}\} \\ &= \mathrm{tr}\{\Sigma_Z^{-1}\tilde{\theta}(e^{i\lambda})(\Sigma_Z^{-1}\tilde{\theta}^{(0)}(e^{i\lambda}))^{-1} \\ &\quad \cdot((\Sigma_Z^{-1}\tilde{\theta}^{(0)}(e^{i\lambda}))^*)^{-1}(\Sigma_Z^{-1}\tilde{\theta}(e^{i\lambda}))^*\} \\ &= \mathrm{tr}(T_{\theta,\theta^{(0)}}(e^{i\lambda})(T_{\theta,\theta^{(0)}}(e^{i\lambda}))^*) = \|T_{\theta,\theta^{(0)}}(e^{i\lambda})\|^2;\end{aligned}$$

thus, by (7), we obtain

$$I(f_{\theta^{(0)}}; f_\theta) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \left\{ \| T_{\theta,\theta^{(0)}}(e^{i\lambda}) \|^2 - m \right\} d\lambda. \tag{8}$$

Now, for $z \in \overline{D}$, we consider the function $\theta \mapsto \|T_{\theta,\theta^{(0)}}(z)\|^2 - m$; it is a quadratic function of θ (because $\theta \mapsto T_{\theta,\theta^{(0)}}(z)$ is linear, and, therefore, $\theta \mapsto \|T_{\theta,\theta^{(0)}}(z)\|^2$ is quadratic) and we trivially have $\|T_{\theta^{(0)},\theta^{(0)}}(z)\|^2 - m = 0$. Then, noting that $\theta \mapsto \int_{-\pi}^{\pi} \left\{ \|T_{\theta,\theta^{(0)}}(e^{i\lambda})\|^2 - m \right\} d\lambda$ is a quadratic function which vanishes at $\theta = \theta^{(0)}$ and is nonnegative (because $\int_{-\pi}^{\pi} \left\{ \|T_{\theta,\theta^{(0)}}(e^{i\lambda})\|^2 - m \right\} d\lambda = 4\pi I(f_{\theta^{(0)}}; f_\theta)$ and $I(f_{\theta^{(0)}}; f_\theta) \geq 0$), we have

$$\begin{aligned} &\int_{-\pi}^{\pi} \left\{ \|T_{\theta,\theta^{(0)}}(e^{i\lambda})\|^2 - m \right\} d\lambda \\ &= \frac{1}{2} \frac{d^2}{dt^2} \int_{-\pi}^{\pi} \left\{ \|T_{\theta^{(0)}+t\eta,\theta^{(0)}}(e^{i\lambda})\|^2 - m \right\} d\lambda \Big|_{t=0,\eta=\theta-\theta^{(0)}}. \end{aligned} \tag{9}$$

In view of the computation of this second derivative, we remark that $T_{\theta,\theta^{(0)}}(z) = \Sigma_Z^{-1}\tilde{\theta}(z)(\tilde{\theta}^{(0)}(z))^{-1}\Sigma_Z$, whence we obtain

$$T_{\theta^{(0)}+t\eta,\theta^{(0)}}(z) = \Sigma_Z^{-1}(\widetilde{\theta^{(0)} + t\eta})(z)(\tilde{\theta}^{(0)}(z))^{-1}\Sigma_Z;$$

moreover, we have

$$(\widetilde{\theta^{(0)} + t\eta})(z) = I_m - \sum_{j=1}^{p} (\theta_j^{(0)} + t\eta_j) z^j = \tilde{\theta}^{(0)}(z) + t\eta^{\#}(z),$$

where

$$\eta^{\#}(z) := -\sum_{j=1}^{p} \eta_j z^j,$$

whence we obtain

$$\begin{aligned} T_{\theta^{(0)}+t\eta,\theta^{(0)}}(z) &= \Sigma_Z^{-1}(\tilde{\theta}^{(0)}(z) + t\eta^{\#}(z))(\tilde{\theta}^{(0)}(z))^{-1}\Sigma_Z \\ &= I_m + t\, \underbrace{\Sigma_Z^{-1}\eta^{\#}(z)(\tilde{\theta}^{(0)}(z))^{-1}\Sigma_Z}_{=:B_{\theta^{(0)},\eta}(z)}. \end{aligned}$$

Thus,

$$\|T_{\theta^{(0)}+t\eta,\theta^{(0)}}(z)\|^2 = \|I_m\|^2 + t^2\|B_{\theta^{(0)},\eta}(z)\|^2 + 2t \cdot \mathrm{tr}(\mathrm{Re}\{B_{\theta^{(0)},\eta}(z)\}),$$

where the matrix $\mathrm{Re}\{B_{\theta^{(0)},\eta}(z)\}$ is the real part of $B_{\theta^{(0)},\eta}(z)$, and

$$\frac{1}{2}\frac{d^2}{dt^2}\|T_{\theta^{(0)}+t\eta,\theta^{(0)}}(z)\|^2 = \|B_{\theta^{(0)},\eta}(z)\|^2 \text{ for all } t \in \mathbb{R}. \tag{10}$$

Finally, by (8), (9) and (10), we have

$$\begin{aligned}
I(f_{\theta^{(0)}}; f_\theta) &= \frac{1}{4\pi}\int_{-\pi}^{\pi}\left\{\|T_{\theta,\theta^{(0)}}(e^{i\lambda})\|^2 - m\right\}d\lambda \\
&= \frac{1}{4\pi}\int_{-\pi}^{\pi}\|B_{\theta^{(0)},\theta-\theta^{(0)}}(e^{i\lambda})\|^2 d\lambda \\
&= \frac{1}{4\pi}\int_{-\pi}^{\pi}\mathrm{tr}(B_{\theta^{(0)},\theta-\theta^{(0)}}(e^{i\lambda})(B_{\theta^{(0)},\theta-\theta^{(0)}}(e^{i\lambda}))^*)d\lambda \\
&= \frac{1}{4\pi}\int_{-\pi}^{\pi}\mathrm{tr}\left(\Sigma_Z^{-1}(\theta-\theta^{(0)})^{\#}(e^{i\lambda})(\tilde{\theta}^{(0)}(e^{i\lambda}))^{-1}\Sigma_Z\right. \\
&\quad \left.\cdot(\Sigma_Z^{-1}(\theta-\theta^{(0)})^{\#}(e^{i\lambda})(\tilde{\theta}^{(0)}(e^{i\lambda}))^{-1}\Sigma_Z)^*\right)d\lambda
\end{aligned}$$

and, by noting that $(\theta-\theta^{(0)})^{\#}(z) = -\sum_{j=1}^{p}(\theta_j - \theta_j^{(0)})z^j$ for all $z \in \mathbb{C}$, we obtain

$$\begin{aligned}
I(f_{\theta^{(0)}}; f_\theta) &= \frac{1}{4\pi}\int_{-\pi}^{\pi}\mathrm{tr}\left(\Sigma_Z^{-1}\sum_{j=1}^{p}(\theta_j-\theta_j^{(0)})e^{i\lambda j}(\tilde{\theta}^{(0)}(e^{i\lambda}))^{-1}\Sigma_Z\right. \\
&\quad \left.\cdot\left(\Sigma_Z^{-1}\sum_{h=1}^{p}(\theta_h-\theta_h^{(0)})e^{i\lambda h}(\tilde{\theta}^{(0)}(e^{i\lambda}))^{-1}\Sigma_Z\right)^*\right)d\lambda \\
&= \frac{1}{4\pi}\int_{-\pi}^{\pi}\sum_{j=1}^{p}\sum_{h=1}^{p}\mathrm{tr}\left(\Sigma_Z^{-1}(\theta_j-\theta_j^{(0)})e^{i\lambda j}(\tilde{\theta}^{(0)}(e^{i\lambda}))^{-1}\right. \\
&\quad \left.\cdot\Sigma_Z^2((\tilde{\theta}^{(0)}(e^{i\lambda}))^*)^{-1}e^{-i\lambda h}(\theta_h-\theta_h^{(0)})^t\Sigma_Z^{-1}\right)d\lambda \\
&= \frac{1}{4\pi}\sum_{j=1}^{p}\sum_{h=1}^{p}\mathrm{tr}\left(\Sigma_Z^{-1}(\theta_j-\theta_j^{(0)})\int_{-\pi}^{\pi}e^{i\lambda(j-h)}(\tilde{\theta}^{(0)}(e^{i\lambda}))^{-1}\right. \\
&\quad \left.\cdot\Sigma_Z^2((\tilde{\theta}^{(0)}(e^{i\lambda}))^*)^{-1}d\lambda(\theta_h-\theta_h^{(0)})^t\Sigma_Z^{-1}\right) \\
&= \frac{1}{4\pi}\sum_{j=1}^{p}\sum_{h=1}^{p}\mathrm{tr}\left(\Sigma_Z^{-1}(\theta_j-\theta_j^{(0)})\right. \\
&\quad \left.\cdot\int_{-\pi}^{\pi}e^{i\lambda(j-h)}2\pi f_{\theta^{(0)}}(\lambda)d\lambda(\theta_h-\theta_h^{(0)})^t\Sigma_Z^{-1}\right)
\end{aligned}$$

$$= \frac{1}{2}\sum_{j=1}^{p}\sum_{h=1}^{p} \mathrm{tr}\left(\Sigma_Z^{-1}(\theta_j - \theta_j^{(0)})\Gamma_{\theta^{(0)}}(j-h)(\theta_h - \theta_h^{(0)})^t \Sigma_Z^{-1}\right)$$
$$= \frac{1}{2}\sum_{j=1}^{p}\sum_{h=1}^{p} \mathrm{tr}\left(\Gamma_{\theta^{(0)}}(j-h)(\theta_h - \theta_h^{(0)})^t (\Sigma_Z^2)^{-1}(\theta_j - \theta_j^{(0)})\right).$$

This completes the proof (it is enough to exchange the roles of j and h in the latter expression). □

Appendix: Some properties of the set Θ

We start proving that the set Θ is bounded if $m = 1$. Let $r \in \{1, \dots, p\}$ be the degree of the polynomial $P_1(z) = 1 - \sum_{j=1}^r \theta_j z^j$ which depends on the choice of $\theta = (\theta_1, \dots, \theta_p) \in \mathbb{R}^p$, and let us consider the polynomial $P_2(z) = z^r P_1(1/z) = z^r - \sum_{j=1}^r \theta_j z^{r-j}$. Then

$$\begin{aligned} P_2(z) &= \prod_{j=1}^{r}(z - w_j) = z^r + \sum_{j=1}^{r}\left(\sum_{\{h_1,\dots,h_j\}\subset\{1,\dots,r\}} (-w_{h_1})\cdots(-w_{h_j})\right) z^{r-j} \\ &\quad +(-w_1)\cdots(-w_r) \end{aligned}$$

for some $w_1, \dots, w_r \in D$, i.e. $|w_1|, \dots, |w_r| < 1$, because $P_1(z) \neq 0$ for all $z \in \overline{D}$. Note that, for each fixed $j \in \{0, 1, \dots, r\}$, the absolute value of the coefficient of z^{r-j} of P_2 is bounded by $\binom{r}{j}$; moreover, the coefficients of P_2 consists of a rearrangement in a different order of the coefficients of P_1. Thus, the coefficients of P_1 is bounded and this completes the proof.

We can also prove that the set Θ is unbounded if $m \geq 2$. We consider a sequence $\{\theta^{(n)} = (\theta_1^{(n)}, \dots, \theta_p^{(n)}) : n \geq 1\}$, where $\theta_j^{(n)} = (\theta_j^{(n)}(h,k))_{h,k\in\{1,\dots,m\}}$, which is defined as follows:

$$\theta_j^{(n)}(h,k) := \begin{cases} n & \text{if } j = 1 \text{ and } (h,k) = (1,m), \\ 0 & \text{otherwise.} \end{cases}$$

Then $\{\theta^{(n)} : n \geq 1\} \subset \Theta$, because $\det(\tilde{\theta}^{(n)}(z)) = 1$ for all $z \in \mathbb{C}$, and, therefore, $\det(\tilde{\theta}^{(n)}(z)) \neq 0$ for all $z \in \overline{D}$; indeed we have

$$\tilde{\theta}^{(n)}(z) = (\tilde{\theta}^{(n)}(z; h,k))_{h,k\in\{1,\dots,m\}}$$

where

$$\tilde{\theta}^{(n)}(z; h,k) = \begin{cases} 1 & \text{if } h = k \\ nz & \text{if } (h,k) = (1,m) \\ 0 & \text{otherwise.} \end{cases}$$

Thus, Θ is unbounded because $\{\theta^{(n)} : n \geq 1\}$ is unbounded.

Finally, we prove that the set Θ is open. First of all, we have

$$\Theta = \left\{\theta = (\theta_1, \dots, \theta_p)^t \in (\mathcal{M}(m,\mathbb{R}))^p : \min_{z\in\overline{D}} |\det \tilde{\theta}(z)| > 0\right\}.$$

Then Θ is open because the function $\theta \mapsto \min_{z\in\overline{D}} |\det\tilde{\theta}(z)|$ is continuous on $(\mathcal{M}(m,\mathbb{R}))^p$ because one can check that $\min_{z\in\overline{D}} |\det\tilde{\theta}^{(n)}(z)|$ converges to $\min_{z\in\overline{D}} |\det\tilde{\theta}^{(\infty)}(z)|$ whenever $\theta^{(n)}$ converges to $\theta^{(\infty)}$ (the details are omitted).

Acknowledgments We thank Domenico Marinucci for suggesting the idea of studying the large deviation principle in this paper and some properties of the multivariate AR(p) processes useful for the proofs. We also thank the referee for the careful reading of the paper.

References

Cover, T. M., Thomas, J. A. (1991). *Elements of information theory*. New York: John Wiley and Sons.
Dembo, A., Zeitouni, O. (1998). *Large deviations techniques and applications* (2nd ed.). New York: Springer.
Eichelsbacher, P., Ganesh, A. (2002a). Moderate deviations for Bayes posteriors. *Scandinavian Journal of Statistics*, *29*, 153–167.
Eichelsbacher, P., Ganesh, A. (2002b). Bayesian inference for Markov chains. *Journal of Applied Probability*, *39*, 91–99.
Fu, J. C., Kass, R. E. (1988). The exponential rate of convergence of posterior distributions. *Annals of the Institute of Statistical Mathematics*, *40*, 683–691.
Ganesh, A., O'Connell, N. (1999). An inverse of Sanov's Theorem. *Statistics & Probability Letters*, *42*, 201–206.
Ganesh, A., O'Connell, N. (2000). A large deviation principle for Dirichlet posteriors. *Bernoulli*, *6*, 1021–1034.
Gilbarg, D., Trudinger, N. S. (1983). *Elliptic partial differential equations of second order* (2nd ed.). Berlin: Springer.
Macci, C. (2010a). Large deviations for estimators of some threshold parameters. *Statistical Methods Applications*, *19*, 63–77.
Macci, C. (2010b). Large deviations for Bayesian estimators in first-order autoregressive processes. *Journal of Statistical Planning and Inference*, *140*, 2790–2797.
Macci, C. (2011). Large deviations for estimators of unknown probabilities, with applications in risk theory. *Statistics Probability Letters*, *81*, 16–24.
Macci, C., Petrella, L. (2006). Mixtures of conjugate prior distributions and large deviations for level crossing probabilities. *Sankhyā*, *68*, 61–89.
Macci, C., Petrella, L. (2009). Censored exponential data: large deviations for MLEs and posterior distributions. *Communications in Statistics Theory and Methods*, *38*, 2435–2452.
Macci, C., Petrella, L. (2010). Large deviation results on some estimators for stationary Gaussian processes. *Statistics*, *44*, 129–144.
Mas, A., Menneteau, L. (2003). Large and moderate deviations for infinite-dimensional autoregressive processes. *Journal of Multivariate Analysis*, *87*, 241–260.
Parthasarathy, K. R. (1967). *Probability measures on metric spaces*. London: Academic Press.
Paschalidis, I. C., Vassilaras, S. (2001). On the estimation of buffer overflow probabilities from measurements. *IEEE Transactions on Information Theory*, *47*, 178–191.
Rudin, W. (1986). *Real and complex analysis* (3rd ed.). New York: McGraw Hill.
Taniguchi, M., Kakizawa, Y. (2000). *Asymptotic theory of statistical inference for time series*. New York: Springer.
Varadhan, S. R. S. (2003). Large deviations and entropy. In A. Greven, G. Keller, G. Warnecke (Eds.), *Entropy* (pp. 199–214). Princeton: Princeton University Press.