ESAIM: Probability and Statistics February 2000, Vol. 4, p. 1–24
URL: http://www.emath.fr/ps/

SHARP LARGE DEVIATIONS FOR GAUSSIAN QUADRATIC FORMS WITH APPLICATIONS

Bernard Bercu[1], Fabrice Gamboa[2] and Marc Lavielle[3]

Abstract. Under regularity assumptions, we establish a sharp large deviation principle for Hermitian quadratic forms of stationary Gaussian processes. Our result is similar to the well-known Bahadur-Rao theorem [2] on the sample mean. We also provide several examples of application such as the sharp large deviation properties of the Neyman-Pearson likelihood ratio test, of the sum of squares, of the Yule-Walker estimator of the parameter of a stable autoregressive Gaussian process, and finally of the empirical spectral repartition function.

Résumé. Sous des hypothèses de régularité convenables, on établit un principe de grandes déviations précises pour des formes quadratiques de processus gaussiens stationnaires. Notre résultat est l'analogue du théorème de Bahadur-Rao [2] sur la moyenne empirique. Nous proposons également plusieurs exemples d'application comme les propriétés de grandes déviations précises pour le test du rapport de vraisemblance de Neyman-Pearson, pour la somme des carrés, pour l'estimateur de Yule-Walker du paramètre d'un processus gaussien autorégressif stable, et finalement pour la fonction de répartition spectrale empirique.

AMS Subject Classification. 60F10, 11E25, 60G15, 47B35.

Received January 30, 1998. Revised December 21, 1998 and November 3, 1999.

1. Introduction

The basic common way to study the rate of convergence in the law of large numbers is given by the Central Limit theorem (CLT). Let $(Z_1, Z_2, \dots, Z_n)$ be a sample of independent and identically distributed random variables with zero mean and unit variance. The rate of convergence of the empirical mean

$$\overline{Z}_n = \frac{1}{n}\sum_{i=1}^{n} Z_i$$

towards 0 is evaluated rescaling $\overline{Z}_n$ by $\sqrt{n}$. The CLT tells us that for any $c \in \mathbb{R}$, $\mathbb{P}(\sqrt{n}\overline{Z}_n \geq c)$ converges to $1 - F(c)$ where F is the standard normal distribution. Under the well-known Cramer's condition [16], one could refine the previous convergence *via* an Edgeworth expansion of the density of $\sqrt{n}\overline{Z}_n$ (see *e.g.* [27]). On the one hand, as the rescaling factor is $\sqrt{n}$, this tool is well adapted if one wishes to evaluate $\mathbb{P}(\overline{Z}_n \geq c)$

Keywords and phrases: Large deviations, Gaussian processes, quadratic forms, Toeplitz matrices.

[1] Université Paris-Sud, bâtiment 425, 91405 Orsay Cedex, France; e-mail: Bernard.Bercu@math.u-psud.fr

[2] Université Paul Sabatier, Toulouse, France; e-mail: Gamboa@cict.fr

[3] Université René Descartes and Université Paris-Sud, France; e-mail: Marc.Lavielle@math.u-psud.fr

© EDP Sciences, SMAI 2000

where c is in the normal range for $\overline{Z}_n$, namely $c = \mathcal{O}(n^{-1/2})$. On the other hand, if one wants to estimate $\mathbb{P}(\overline{Z}_n \geq c)$ where c is bounded away from zero, the CLT or one of its refinement does not give an accurate equivalent for this probability (see the introduction of [27]). Indeed, under some regularity assumptions on the cumulant generating function of Z_i, $\mathbb{P}(\overline{Z}_n \geq c)$ goes exponentially fast to 0. Moreover, this exponential behavior depends on the distribution of Z_i and could be quite different from the Gaussian one given by $\exp(-nc^2/2)$. Roughly speaking, typical values of $\overline{Z}_n$ are of order $n^{-1/2}$ but with exponentially small probabilities, $\overline{Z}_n$ takes relatively large values. This is well-known (see *e.g.* [18] and references therein) and formalized by saying that $(\overline{Z}_n)$ satisfies a Large Deviations Principle (LDP). More precisely, there exists a positive lower semicontinuous function I having compact level sets such that for any closed set A of $\mathbb{R}$

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\overline{Z}_n \in A\right) \leq -\inf_{x\in A} I(x),$$

while for any open set B of $\mathbb{R}$

$$-\inf_{x\in B} I(x) \leq \liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\overline{Z}_n \in B\right).$$

Actually, the rate function I could be easily calculated as the Legendre transform of the cumulant generating function of Z_i. Of course, for estimating $\mathbb{P}(\overline{Z}_n \geq c)$ with $c > 0$ in a practical situation, the LDP is not so helpful since it only gives a logarithmic equivalent for this probability. Bahadur and Rao [2] overcome this difficulty by proving a full expansion of this tail probability. They established a sharp large deviation principle for $(\overline{Z}_n)$ in the spirit of the following definition.

Definition 1.1. Let (V_n) be a sequence of real random variables converging almost surely to some real number v. We say that (V_n) satisfies a Local Sharp Large Deviation Principle (LSLDP) of order $p \in \mathbb{N}$ at point $c \in \mathbb{R}$ whenever the following expansion holds

$$\mathbb{P}(V_n \geq c) = \frac{a_0 \exp(-nb)}{\sqrt{n}} \left(1 + \frac{a_1}{n} + \ldots + \frac{a_p}{n^p} + o\left(\frac{1}{n^p}\right)\right) \quad (c > v) \tag{1.1}$$

or

$$\mathbb{P}(V_n \leq c) = \frac{a_0 \exp(-nb)}{\sqrt{n}} \left(1 + \frac{a_1}{n} + \ldots + \frac{a_p}{n^p} + o\left(\frac{1}{n^p}\right)\right) \quad (c < v) \tag{1.2}$$

where $a_0, a_1, \ldots, a_p \in \mathbb{R}$ and $b \in \mathbb{R}^+ \cup \{+\infty\}$ only depend on c. We say that (V_n) satisfies a Sharp Large Deviation Principle (SLDP) of order $p \in \mathbb{N}$ if it satisfies an LSLDP of order p at any point $c \in \mathbb{R}$.

After some refinements, the expansion remains valid in the range of middle or small deviations. This is the so-called saddlepoint approximation used in statistics. For further references, the interested reader should check Barndorff-Nielsen and Cox [3] and Jensen [27].

In this paper, our purpose is to establish an LSLDP for quadratic forms of Gaussian processes and to illustrate the interest of such tractable expansions by some simple examples arising from statistics. Let us begin by introducing the frame of the paper. Let (X_n) be a centered stationary real Gaussian process having spectral density g. The function g is nonnegative, integrable on the torus $\mathbb{T} = [-\pi, \pi[$ and the auto-covariances of the process (X_n) are the Fourier coefficients of g ($\hat{g}_i = E(X_n X_{n+i})$ with $n, i \in \mathbb{Z}$). For a given sequence of Hermitian matrices (M_n), we are interested in the asymptotic properties of the random variable

$$W_n = \frac{1}{n} X^{(n)*} M_n X^{(n)} \tag{1.3}$$

where $X^{(n)*} = (X_1, \dots, X_n)$. Under some assumptions on the sequence of matrices (M_n), (W_n) converges almost surely to a constant $\overline{\varphi}$. Hereafter, it is an interesting problem to study the rate of convergence. In [7,11], it was proved that under some regularity assumptions (W_n) satisfies an LDP. In the present paper, we study the sharp large deviation properties of (W_n). As for the sequence of empirical means $(\overline{Z}_n)$, one cannot deduce from the LDP an asymptotic equivalent for $\mathbb{P}(W_n \geq c)$. Therefore, it is necessary to establish an LSLDP for (W_n) as in Definition 1.1. We will use the same approach as the one developed by Bahadur and Rao for establishing the LSLDP for (W_n). Actually, the main idea to study the LSLDP for (W_n) is to express W_n as a weighted sum of i.i.d. random variables. Some extensions of Bahadur and Rao theorem were proposed in [8]. However, these extensions will not be helpful for us since they cannot integrate the very particular behavior of the sequence of weights. This behavior is related to the asymptotic theory of Toeplitz operators. Moreover, the difficulty in our framework is that the expansion of the characteristic function needs more analysis. This will be achieved *via* the strong Szegö theorem (see [22, 23, 25, 28]).

The paper is organized as follows. The results concerning LDP for quadratic forms (see [7] and [11]) are recalled in Section 2. In Section 3, we establish under suitable assumptions, the LSLDP for (W_n). Section 4 is devoted to some alternatives to our approach which are essentially the Edgeworth and saddlepoint approximations [2,27,32]. Roughly speaking, on the one hand, Edgeworth expansion cannot give an accurate approximation for large deviation probabilities. On the other hand, saddlepoint approximation is generally untractable in our present frame. As a matter of fact, the exponential part is generally very intrigated. Nevertheless, our result may also be seen as approximated saddlepoint expansions in the spirit of Whittle approximation of a Gaussian likelihood [1]. In Section 5, we provide several statistical applications of our general result: we obtain an LSLDP for the likelihood ratio statistic, for the estimated variance of a sequence of Gaussian random variables, for the empirical spectral repartition function and finally for the Yule-Walker estimator of the parameter of a stable autoregressive Gaussian process. Notice also that this kind of expansion has been extensively studied for the autoregressive process, partly for its importance in econometrics (see [27] and the references therein). One of the interest of this paper is to present general results, obtained in the Large Deviations framework. These results are new and of statistical interest. Some numerical comparisons between the sharp large deviation estimates and Edgeworth or saddlepoint expansions are proposed. The theoretical results are shown to be of practical interest since small tailed probabilities (smaller than 10^{-2}) are very well approximated by using the LSLDP. The Edgeworth expansion provides very good results for small and medium deviations (larger than 10^{-3}) and the saddlepoint approximation is shown to be very efficient for approximating any probability, when it can be computed. In Appendices A and B, we recall some useful known and less known results on the asymptotic behavior of the eigenvalues of Toeplitz forms. These asymptotic results are widely used in Section 5.

2. Large deviations for quadratic forms

Let $T_n(g)$ be the covariance matrix of the vector $X^{(n)}$ which is also called the order n Toeplitz matrix associated with g, $(T_n(g) = (\hat{g}_{i-j})_{1\leq i,j\leq n})$. Denote by $\lambda_1^n \leq \ldots \leq \lambda_n^n$ the eigenvalues of $T_n(g)^{\frac{1}{2}} M_n T_n(g)^{\frac{1}{2}}$ which are also the eigenvalues of $M_n T_n(g)$. After an orthogonal change of basis, one can rewrite (1.3) as

$$W_n = \frac{1}{n} \sum_{i=1}^{n} \lambda_i^n Z_i^n \tag{2.1}$$

where $Z_1^n, \dots, Z_n^n$ are i.i.d. random variables with $\chi^2(1)$ distribution. We are now in position to study the tail properties of the sequence (W_n). A classical tool to investigate the large deviations of a sequence of random variables is the normalized cumulant generating function

$$L_n(\tau) = \frac{1}{n} \log \mathbb{E}\left(e^{n\tau W_n}\right). \tag{2.2}$$

It clearly follows from the decomposition (2.1) that

$$L_n(\tau) = \begin{cases} -\frac{1}{2n}\sum_{i=1}^n \log(1-2\tau\lambda_i^n) & \text{if } \tau \in \Delta_n \\ +\infty & \text{otherwise} \end{cases} \tag{2.3}$$

where

$$\Delta_n = \left\{\tau \in \mathbb{R} \ / \ \max(\lambda_1^n\tau, \lambda_n^n\tau) < \frac{1}{2}\right\}.$$

First, as we want to apply some classical results such as the Gärtner-Ellis theorem (see *e.g.* [18]), we are brought to make one of the major assumptions of the paper.

There exists a function $\varphi \in L^\infty(\mathbb{T})$ which is not identically zero such that, if $m_\varphi = \text{essinf}\ \varphi$ and $M_\varphi = \text{esssup}\ \varphi$, then for all $n \geq 1$,

$$(H1) \qquad m_\varphi \leq \lambda_1^n \leq \lambda_n^n \leq M_\varphi \quad \text{and} \quad \frac{1}{n}\sum_{i=1}^n \delta_{\lambda_i^n} \Longrightarrow P_\varphi$$

where the limit P_φ is the image probability of the uniform measure on the torus $\mathbb{T}$ by the application φ.

We clearly have from $(H1)$, $\lim \lambda_1^n = m_\varphi$ and $\lim \lambda_n^n = M_\varphi$. Therefore, under $(H1)$, the asymptotic cumulant generating function is

$$L(\tau) = \begin{cases} -\dfrac{1}{4\pi}\displaystyle\int_{\mathbb{T}} \log\left(1-2\tau\varphi(x)\right)dx & \text{if } \tau \in \Delta \\ +\infty & \text{otherwise} \end{cases} \tag{2.4}$$

where

$$\Delta = \left\{\tau \in \mathbb{R} \ / \ \max(m_\varphi\tau, M_\varphi\tau) < \frac{1}{2}\right\}.$$

Using this limit function, it was shown in [7] that an LDP is available for (W_n) (see also [11] for the special case where M_n is the identity matrix). Denote by $L^\star$ the Fenchel-Legendre dual of L

$$L^\star(c) = \sup_{\tau\in\mathbb{R}}\left(c\tau + \frac{1}{4\pi}\int_{\mathbb{T}} \log\left(1-2\tau\varphi(x)\right)dx\right). \tag{2.5}$$

The following theorem was proved in [7]:

Theorem 2.1. *Assume that $(H1)$ holds. Then, the sequence (W_n) satisfies an LDP with rate function $L^\star$. In particular,*

$$\begin{cases} \lim\limits_{n\to\infty} \dfrac{1}{n}\log \mathbb{P}(W_n \geq c) = -L^\star(c), \ (c > \overline{\varphi}) \\ \lim\limits_{n\to\infty} \dfrac{1}{n}\log \mathbb{P}(W_n \leq c) = -L^\star(c), \ (c < \overline{\varphi}) \end{cases} \tag{2.6}$$

where $\overline{\varphi} = \dfrac{1}{2\pi}\displaystyle\int_{\mathbb{T}} \varphi(x)dx$.

In [7], we also investigated the case where $(H1)$ is not satisfied.

3. Local sharp large deviations for quadratic forms

Let $\mathcal{C}_\varphi$ be either $[0,+\infty[$ if $m_\varphi \geq 0$, $]-\infty,0]$ if $M_\varphi \leq 0$ or $\mathbb{R}$ otherwise. Actually, $\mathcal{C}_\varphi$ is, for n large enough the support of the distribution of (W_n). The LSLDP for (W_n) will hold for any point c lying in

$$\mathcal{D}_\varphi = \left\{ c \in \mathbb{R} \ / \ \exists \tau \in \Delta \text{ such that } c = \frac{1}{2\pi} \int_{\mathbb{T}} \frac{\varphi(x)}{1-2\tau\varphi(x)} dx \right\}.$$

The following lemma will be useful in the sequel. It is a direct consequence of the work on ill posed inverse problems developed in [20] and [21].

Lemma 3.1.

a) *For any $c \in Int\, \mathcal{C}_\varphi$ (the interior of $\mathcal{C}_\varphi$), there exists a unique τ_c in Δ such that $L^\star(c) = c\tau_c - L(\tau_c)$.*

b) *If c lies in $\mathcal{D}_\varphi$, then $c = L'(\tau_c)$ so that*

$$L'(\tau_c) = \frac{1}{2\pi} \int_{\mathbb{T}} \frac{\varphi(x)}{1-2\tau_c\varphi(x)} dx. \tag{3.1}$$

c) *Finally, assume that φ is differentiable on $\mathbb{T}$. Then, $Int\, \mathcal{C}_\varphi = \mathcal{D}_\varphi$ and consequently, (3.1) holds for any $c \in Int\, \mathcal{C}_\varphi$.*

Roughly speaking, the parameter τ_c gives the exponential change of probability which allows us to sharpen Theorem 2.1, establishing the LSLDP for (W_n) at point $c \in \mathcal{D}_\varphi$. From assumption $(H1)$, the sequence (L_n) converges simply to L. In order to get the LSLDP, we need to specify this convergence assuming that for some $p \geq 0$, the following conditions are satisfied.

Case $p = 0$: there exists a function H such that,

$$L_n(\tau_c) = L(\tau_c) + \frac{1}{n} H(\tau_c) + o\left(\frac{1}{n}\right). \tag{$H2(0)$}$$

Case $p > 0$: there exists a function H differentiable $2p+3$ times, such that, for any $0 \leq k \leq 2p+3$,

$$L_n^{(k)}(\tau_c) = L^{(k)}(\tau_c) + \frac{1}{n} H^{(k)}(\tau_c) + \mathcal{O}\left(n^{-p-2}\right) \tag{$H2(p)$}$$

where $L^{(k)}$ is the k'th derivative of L.

Assumption $(H2(p))$ is not really restrictive. Indeed, we shall show in Section 5 that $(H2(p))$ is fulfilled in several statistical applications for any $p \geq 0$. We now propose the main result of the paper, giving the LSLDP theorem for (W_n).

Theorem 3.2. *Let $c \in \mathcal{D}_\varphi$ and assume that there exists $p \geq 0$ such that $(H1)$ and $(H2(p))$ hold. If $c > \overline{\varphi}$, there exists a sequence $d_{c,1}, d_{c,2}, \ldots, d_{c,p}$, such that, for n large enough*

$$\mathbb{P}(W_n \geq c) = \frac{e^{-nL^\star(c)+H(\tau_c)}}{\sigma_c \tau_c \sqrt{2\pi n}} \big[1 + o(1)\big], \qquad \text{if } p = 0, \tag{3.2}$$

$$\mathbb{P}(W_n \geq c) = \frac{e^{-nL^\star(c)+H(\tau_c)}}{\sigma_c \tau_c \sqrt{2\pi n}} \left[1 + \sum_{k=1}^{p} \frac{d_{c,k}}{n^k} + \mathcal{O}(n^{-p-1})\right], \text{ if } p > 0. \tag{3.3}$$

Here, $\sigma_c^2 = L''(\tau_c)$ and the coefficients $d_{c,1}, d_{c,2}, \dots, d_{c,p}$ may be explicitly computed as functions of the derivatives of L and H at point τ_c. For example, the first coefficient $d_{c,1}$ is given by

$$d_{c,1} = \frac{1}{\sigma_c^2}\left(-\frac{h_2}{2} - \frac{h_1^2}{2} + \frac{l_4}{8\sigma_c^2} + \frac{l_3 h_1}{2\sigma_c^2} - \frac{5l_3^2}{24\sigma_c^4} + \frac{h_1}{\tau_c} - \frac{l_3}{2\tau_c\sigma_c^2} - \frac{1}{\tau_c^2}\right) \tag{3.4}$$

with $l_k = L^{(k)}(\tau_c)$ and $h_k = H^{(k)}(\tau_c)$. In addition, relations similar to (3.2) or (3.3) are also valid for $\mathbb{P}(W_n \le c)$ when $c < \overline{\varphi}$. In other words, (W_n) satisfies an LSLDP of order $p \ge 0$ at point $c \in \mathcal{D}_\varphi$.

Remark 3.3.

a) If φ is differentiable on $\mathbb{T}$ and if $(H1)$ holds and $(H2(p))$ is globally satisfied for $p \ge 0$ and for all $c \in \text{Int } \mathcal{C}_\varphi$, then Theorem 3.2 gives an SLDP for the sequence (W_n).
b) For sack of simplicity, we prefer to express locally the assumption. If $(H2)$ is satisfied for any c in a non empty open set it implies $(H1)$.

Proof. Recall that W_n may be expressed as a weighted sum of i.i.d. random variables with $\chi^2(1)$ distribution

$$W_n = \frac{1}{n}\sum_{i=1}^{n} \lambda_i^n Z_i^n.$$

In order to prove (3.3), we follow the same approach as Bahadur and Rao [2] for the sample mean. The main difference is that in the i.i.d. case, no approximation between $L_n^\star$ and $L^\star$ has to be evaluated, since the large deviation functional is directly the Cramer transform of the sample and consequently does not depend on n. We assume that $m_\varphi < 0$ and $M_\varphi > 0$. The other case can be tackled identically. Consider the exponential change of probability

$$\frac{d\mathbb{Q}_n}{d\mathbb{P}} = \exp\Big(\tau_c n W_n - n L_n(\tau_c)\Big).$$

Then, we have the decomposition $\mathbb{P}(W_n \ge c) = A_n B_n$ with

$$A_n = \exp\Big[n(L_n(\tau_c) - c\tau_c)\Big], \tag{3.5}$$

$$B_n = \mathbb{E}_n\Big(\exp\Big[-\tau_c n(W_n - c)\Big]\mathbb{I}_{W_n \ge c}\Big), \tag{3.6}$$

where $\mathbb{E}_n$ is the expectation over the new probability $\mathbb{Q}_n$. As $L^\star(c) = \tau_c c - L(\tau_c)$, we clearly have $L_n(\tau_c) - c\tau_c = L_n(\tau_c) - L(\tau_c) - L^\star(c)$. Therefore, from assumption $(H2(p))$ we have

$$A_n = \exp\Big[-nL^\star(c) + H(\tau_c)\Big]\left(1 + \mathcal{O}(n^{-p-1})\right). \tag{3.7}$$

We must now evaluate B_n. For this end, we use the classical normalization of the CLT

$$U_n = \frac{\sqrt{n}(W_n - c)}{\sigma_c}. \tag{3.8}$$

Then, we immediately have from (3.6)

$$B_n = \mathbb{E}_n\Big(\exp\Big[-\tau_c\sigma_c\sqrt{n}U_n\Big]\mathbb{I}_{U_n \ge 0}\Big).$$

We shall now prove that, for $c > \overline{\varphi}$, the distribution of U_n under $\mathbb{Q}_n$ converges to the standard Gaussian distribution. Denote by Φ_n the characteristic function of U_n over the probability $\mathbb{Q}_n$

$$\Phi_n(u) = \mathbb{E}_n\left(\exp\left[\frac{iu\sqrt{n}(W_n - c)}{\sigma_c}\right]\right). \tag{3.9}$$

It is easy to see that

$$\Phi_n(u) = \exp\Big[-\frac{iu\sqrt{n}c}{\sigma_c} + nL_n(\tau_c + \frac{iu}{\sigma_c\sqrt{n}}) - nL_n(\tau_c)\Big]. \tag{3.10}$$

Consequently, from (2.3)

$$|\Phi_n(u)|^2 = \prod_{j=1}^{n}\left(1 + \frac{4u^2(\lambda_j^n)^2}{\sigma_c^2 n(1-2\tau_c\lambda_j^n)^2}\right)^{-\frac{1}{2}}. \tag{3.11}$$

Hence, for $n \geq 2$, $\Phi_n \in L^2(\mathbb{R})$ and we may use Parseval's formula to calculate B_n. Thus, we obtain

$$B_n = \frac{1}{2\pi}\int_{\mathbb{R}}\left(\frac{1}{\tau_c\sigma_c\sqrt{n} + iu}\right)\Phi_n(u)\,du = \frac{C_n}{\tau_c\sigma_c\sqrt{2\pi n}}, \tag{3.12}$$

$$C_n = \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}}\left(1 + \frac{iu}{\tau_c\sigma_c\sqrt{n}}\right)^{-1}\Phi_n(u)\,du. \tag{3.13}$$

Let ε be some positive constant such that $P_\varphi([m_\varphi, M_\varphi]\backslash[-\varepsilon;\varepsilon]) = \delta_\varepsilon > 0$ together with $1 - 2\tau_c\varepsilon > 0$. Denote by $q_n = \#\{\lambda_i^n\,;\ |\lambda_i^n| > \varepsilon\}$. Then $(H1)$ implies that $q_n \sim n\delta_\varepsilon$. In addition, we also have from (3.11),

$$|\Phi_n(u)|^2 \leq \left(1 + \frac{(\lambda_c^\varepsilon u)^2}{n}\right)^{-\frac{q_n}{2}} \quad \text{where} \quad \lambda_c^\varepsilon = \frac{2\varepsilon}{\sigma_c(1+2\tau_c\varepsilon)}. \tag{3.14}$$

First, consider the simple case $p = 0$. In order to prove (3.2), it only remains to show that C_n converges to 1. From (3.14), we have for all $v \in \mathbb{R}$,

$$\left|\Phi_n\left(\sqrt{\frac{n}{q_n}}v\right)\right| \leq \left(1 + \frac{(\lambda_c^\varepsilon v)^2}{q_n}\right)^{-\frac{q_n}{4}}. \tag{3.15}$$

For all $a \in \mathbb{R}$, let $h_a(x) = (1 + a^2x^{-1})^{-x}$. The function h_a is non-increasing on $\mathbb{R}^+$. In addition, for n large enough, $q_n \geq 4$. Consequently, it follows from (3.15) that

$$\left|\Phi_n\left(\sqrt{\frac{n}{q_n}}v\right)\right| \leq \left(1 + \frac{(\lambda_c^\varepsilon v)^2}{4}\right)^{-1}. \tag{3.16}$$

The function at the right hand side of (3.16) belongs to $L^1(\mathbb{R})$. Therefore, as $\lim \Phi_n(u) = \exp(-u^2/2)$, we obtain from (3.13) together with the Lebesgue dominated convergence theorem that

$$\lim_{n\to+\infty} C_n = \frac{1}{\sqrt{2\pi\delta_\varepsilon}}\int_{\mathbb{R}}\exp\left(-\frac{v^2}{2\delta_\varepsilon}\right)\,dv = 1. \tag{3.17}$$

We now return to the general case $p > 0$. Let C be some positive constant, explicitly given at the end of the proof, and set $s_n = Cn^{\frac{1}{6}}$. One can rewrite $C_n = I_n + J_n$ where

$$I_n = \frac{1}{\sqrt{2\pi}} \int_{|u|\leq s_n} \left(1 + \frac{iu}{\tau_c \sigma_c \sqrt{n}}\right)^{-1} \Phi_n(u)\, du, \tag{3.18}$$

$$J_n = \frac{1}{\sqrt{2\pi}} \int_{|u|> s_n} \left(1 + \frac{iu}{\tau_c \sigma_c \sqrt{n}}\right)^{-1} \Phi_n(u)\, du. \tag{3.19}$$

On the one hand, we derive from (3.14) that

$$\begin{aligned} |J_n| &\leq \frac{2\tau_c\sigma_c\sqrt{n}}{\sqrt{2\pi}} \int_{s_n}^{+\infty} \frac{1}{u} |\Phi_n(u)|\, du \leq \frac{2\tau_c\sigma_c\sqrt{n}}{\sqrt{2\pi} s_n^2} \int_{s_n}^{+\infty} u \left(1 + \frac{(\lambda_c^\varepsilon u)^2}{n}\right)^{-\frac{q_n}{4}} du \\ &\leq \frac{4\tau_c\sigma_c n\sqrt{n}}{\sqrt{2\pi}(q_n - 4)(\lambda_c^\varepsilon s_n)^2} \left(1 + \frac{(\lambda_c^\varepsilon s_n)^2}{n}\right)^{-\frac{q_n}{4}+1}. \end{aligned}$$

Consequently,

$$|J_n| = \mathcal{O}\left(\frac{\sqrt{n}}{s_n^2}\Big(1 + \frac{(\lambda_c^\varepsilon s_n)^2}{n}\Big)^{-\frac{n\delta_\varepsilon}{4}+1}\right)$$

so that we can always find, for n large enough, some constant $0 < \theta < 1$ such that $|J_n| = \mathcal{O}(\theta^n)$. Now, from assumption $(H2(p))$, for any $0 \leq k \leq 2p+3$

$$L_n^{(k)}(\tau_c) = l_k + \frac{h_k}{n} + \mathcal{O}(n^{-p-2}) \tag{3.20}$$

with $l_k = L^{(k)}(\tau_c)$ and $h_k = H^{(k)}(\tau_c)$. Therefore, from (3.10), there exists $\xi \in \mathbb{R}$ such that, for all $|u| < \sigma_c\sqrt{n}$

$$\log \Phi_n(u) = -\frac{iu\sqrt{n}c}{\sigma_c} + n \sum_{k=1}^{2p+3} \left(\frac{iu}{\sigma_c\sqrt{n}}\right)^k \left(\frac{l_k}{k!} + \frac{h_k}{n\,k!}\right) + \frac{n}{(2p+4)!}\left(\frac{iu}{\sigma_c\sqrt{n}}\right)^{2p+4} L_n^{(2p+4)}(\tau_c + i\xi) + \mathcal{O}(n^{-p-1}).$$

From (2.3), we have for all $\xi \in \mathbb{R}$, $\mid L_n^{(2p+4)}(\tau_c + i\xi) \mid \leq L_n^{(2p+4)}(\tau_c)$. Consequently, we find that

$$\log \Phi_n(u) = -\frac{u^2}{2} + n \sum_{k=3}^{2p+3} \left(\frac{iu}{\sigma_c\sqrt{n}}\right)^k \frac{l_k}{k!} + \sum_{k=1}^{2p+1} \left(\frac{iu}{\sigma_c\sqrt{n}}\right)^k \frac{h_k}{k!} + \frac{(1 + u^{2p+4})}{n^{p+1}}\mathcal{O}(1)$$

where the remainder $\mathcal{O}(1)$ is uniform in u. Then, for $|u| < Cn^{\frac{1}{6}}$ with $C = \sigma_c \mid l_3 \mid^{-\frac{1}{3}}$

$$\frac{\Phi_n(u)}{1 + \frac{iu}{\tau_c\sigma_c\sqrt{n}}} = e^{-\frac{u^2}{2}} \left(1 + \sum_{k=1}^{2p+1} \frac{P_{p,k}(u)}{n^{\frac{k}{2}}} + \frac{(1 + u^{6(p+1)})}{n^{p+1}}\mathcal{O}(1)\right) \tag{3.21}$$

where the $P_{p,k}$ are polynomials in odd powers of u for k odd, and polynomials in even powers of u for k even. The same expansion was established in the i.i.d. case in [16] and [19]. Finally, from (3.18) and (3.21) together with standard calculus about the moments of $\mathcal{N}(0,1)$ distribution, we obtain relation (3.3). For example, in the particular case $p = 1$, we have

$$P_{1,2}(u) = -\frac{l_3 u^4}{6\tau_c\sigma_c^4} + \frac{h_1 u^2}{\tau_c\sigma_c^2} + \frac{l_4 u^4}{24\sigma_c^4} - \frac{h_2 u^2}{2\sigma_c^2} - \frac{l_3^2 u^6}{72\sigma_c^6} - \frac{h_1^2 u^2}{2\sigma_c^2} + \frac{l_3 h_1 u^4}{6\sigma_c^4} - \frac{u^2}{\tau_c^2\sigma_c^2}.$$

Therefore, we find that

$$I_n = 1 + \frac{1}{\sigma_c^2 n}\left(-\frac{h_2}{2} - \frac{h_1^2}{2} + \frac{l_4}{8\sigma_c^2} + \frac{l_3 h_1}{2\sigma_c^2} - \frac{5l_3^2}{24\sigma_c^4} + \frac{h_1}{\tau_c} - \frac{l_3}{2\tau_c\sigma_c^2} - \frac{1}{\tau_c^2}\right) + \mathcal{O}\left(\frac{1}{n^2}\right)$$

which completes the proof of Theorem 3.2. □

4. Edgeworth and saddlepoint approximations

We shall discuss in this section some alternatives to our approach, which are the Edgeworth and saddlepoint approximations [3, 27]. We first need to introduce some notations. Denote by f and F the standard normal density and distribution, respectively. By the Fourier inversion formula, we have

$$f(x) = \frac{1}{2\pi}\int_{\mathbb{R}} \exp\left(-\frac{t^2}{2}\right)\exp(-itx)\, dt$$

so that, for all $n \in \mathbb{N}$,

$$f^{(n)}(x) = \frac{1}{2\pi}\int_{\mathbb{R}} (-it)^n \exp\left(-\frac{t^2}{2}\right)\exp(-itx)\, dt.$$

The Hermite polynomials (H_n) are given, for all $n \in \mathbb{N}$, by

$$H_n(x) = (-1)^n f^{-1}(x) f^{(n)}(x) \tag{4.1}$$

which implies that for all $a > 0$,

$$\int_a^{+\infty} f(x) H_n(x)\, dx = f(a) H_{n-1}(a).$$

Furthermore, the Esscher functions (B_n) are given, for all $n \in \mathbb{N}$ and $a > 0$, by

$$B_n(a) = \frac{1}{2\pi}\int_{\mathbb{R}} \exp\left(-\frac{t^2}{2}\right)\frac{(it)^n}{1 + it/a}\, dt. \tag{4.2}$$

It is possible to show (see *e.g.* [27], p. 24) that the Hermite polynomials and the Esscher functions are associated by the relation

$$B_n(a) = a\int_0^{+\infty} f(x)\exp(-ax) H_n(x)\, dx. \tag{4.3}$$

The first six Hermite polynomials and Esscher functions are well-known and can be found in [27], p. 19 and p. 24.

4.1. **Edgeworth expansion**

Set $m_n = E(W_n) = L_n^{'}(0)$, $\sigma_n^2 = V(W_n) = L_n^{''}(0)/n$ and

$$U_n = \frac{W_n - m_n}{\sigma_n}.$$

Denote by f_n the density of U_n. Then, we have

$$\mathbb{P}(W_n \geq c) = \int_{d_n}^{+\infty} f_n(x)\, dx$$

with $d_n = (c - m_n)/\sigma_n$. Consequently, using an Edgeworth expansion for f_n, we obtain *via* the Hermite polynomials (4.1) the following expansion.

Approximation 1. *For any $c \in \mathbb{R}$, we have*

$$\mathbb{P}(W_n \geq c) \approx 1 - F(d_n) + f(d_n) \sum_{j=3}^{6} \frac{k_{n,j}}{j!} H_{j-1}(d_n) + \frac{1}{72} k_{n,3}^2 H_5(d_n) \tag{4.4}$$

where $k_{n,j} = \dfrac{n}{(n\sigma_n)^j} L_n^{(j)}(0)$.

Remark 4.1. As $d_n \to \infty$, we have for any $p \geq 0$,

$$1 - F(d_n) = \frac{f(d_n)}{d_n} \left\{ 1 - \frac{1}{d_n^2} + \frac{3}{d_n^4} + \cdots + \frac{(-1)^p (2p)!}{d_n^{2p} 2^p p!} + \mathcal{O}\left(\frac{1}{d_n^{2p+2}} \right) \right\}.$$

The Edgeworth approximation is suitable whenever the distribution of (W_n) is close to a Gaussian distribution. Thus (4.4) is competitive only when c belongs to the normal range for W_n, namely $c = \mathcal{O}(n^{-1/2})$. On the other hand, this approximation gives poor results for small tail probabilities. A numerical example will illustrate this behavior below.

4.2. **Saddlepoint expansion**

As in the proof of our main result in Section 3, we have the decomposition $\mathbb{P}(W_n \geq c) = A_n B_n$ with

$$\begin{aligned} A_n &= \exp\Big[n(L_n(\tau_n) - c\tau_n) \Big], \\ B_n &= \mathbb{E}_n\Big(\exp\Big[-\tau_n n (W_n - c) \Big] \mathbb{I}_{W_n \geq c} \Big), \end{aligned}$$

where $\mathbb{E}_n$ is the expectation over the usual exponential change of probability $\mathbb{Q}_n$ associated with the saddlepoint τ_n such that $L_n^{'}(\tau_n) = c$. Set $\sigma_n^2 = L_n^{''}(\tau_n)$ and

$$U_n = \frac{\sqrt{n}(W_n - c)}{\sigma_n}.$$

Denote by Φ_n the characteristic function of U_n over $\mathbb{Q}_n$. By Parseval's formula, we have

$$B_n = \frac{1}{\tau_n \sigma_n \sqrt{n}} \frac{1}{2\pi} \int_{\mathbb{R}} \left(1 + \frac{iu}{\tau_n \sigma_n \sqrt{n}} \right)^{-1} \Phi_n(u)\, du.$$

Consequently, using an Edgeworth expansion for Φ_n, we find via the Esscher functions (4.3) the following expansion.

Approximation 2. *For any $c \in \mathbb{R}$, we have*

$$\mathbb{P}(W_n \geq c) \approx \frac{e^{n(L_n(\tau_n) - c\tau_n)}}{\tau_n \sigma_n \sqrt{n}} \left[B_0(p_n) + \sum_{j=3}^{6} \frac{k_{n,j}}{j!} B_j(p_n) + \frac{1}{72} k_{n,3}^2 B_6(p_n) \right] \tag{4.5}$$

where $p_n = \tau_n \sigma_n \sqrt{n}$ *and* $k_{n,j} = \dfrac{n}{(\sqrt{n}\sigma_n)^j} L_n^{(j)}(\tau_n)$.

Remark 4.2. The saddlepoint approximation is very sharp if we can explicitly calculate the normalized cumulant generating function L_n and all its derivatives. Consequently, except for very special cases, L_n is very complicated and (4.5) is untractable in a general frame. Now, from a practical point of view, it is always possible to replace L_n by $L + H/n$ and compute the approximation given above. A theoretical study of this approach is precisely the main purpose of this paper. Finally, we want to point out that the remainders in Edgeworth and Saddlepoint expansions were not properly evaluated in the literature in a general frame.

5. Examples of application

5.1. **Likelihood ratio test**

Denote by $\mathbb{P}_g$ the probability distribution of (X_n). Let g_0 and g_1 be two spectral densities which differ on a positive Lebesgue measure set. If we wishes to test $H_0 : g = g_0$ *versus* $H_1 : g = g_1$ on the basis of the observation $X_1, \dots, X_n$, then the Neyman-Pearson theorem tells us that the optimal strategy is the likelihood ratio test (see *e.g.* [30], Th. 1, p. 65). Here, the likelihood ratio statistics only depends on the random variable

$$W_n = \frac{1}{n} X^{(n)*} \left[T_n^{-1}(g_0) - T_n^{-1}(g_1) \right] X^{(n)}. \tag{5.1}$$

The rejecting set for the optimal strategy at level $\alpha \in]0,1[$ is $[c_{n,\alpha}, +\infty[$ where the threshold $c_{n,\alpha}$ is such that $\mathbb{P}_{g_0}(W_n \geq c_{n,\alpha}) = \alpha$.

5.1.1. *General sharp large deviations*

We will provide here, for large values of n, the sharp behavior of $\mathbb{P}_{g_0}(W_n \geq c)$ for appropriate values of c. The asymptotic behavior of $\mathbb{P}_{g_0}(W_n \geq c)$ using the large deviation theory was first studied by Dacunha-Castelle in [17] (see also [10, 15]). Nevertheless, in these papers, no full LDP is provided. First of all, we recall the result recently obtained in [7] which establishes the LDP for (W_n) under the classical following assumption: (H3)

i) *The spectral density* g_0 *is in the Szegö class, i.e.* $\log g_0 \in L^1(\mathbb{T})$.
ii) *The ratio* $\dfrac{g_0}{g_1} \in L^\infty(\mathbb{T})$.

Proposition 5.1. *Assume that* $(H3)$ *holds. Then, under the null hypothesis,* (W_n) *satisfies an LDP with rate function* $L^\star$ *which is the Fenchel-Legendre dual of*

$$L(\tau) = \begin{cases} -\dfrac{1}{4\pi} \displaystyle\int_{\mathbb{T}} \log\left(1 - 2\tau\varphi(x)\right) dx & \text{if } \tau \in \Delta \\ +\infty & \text{otherwise} \end{cases} \tag{5.2}$$

where $\varphi = 1 - \dfrac{g_0}{g_1}$.

We are now in position to sharpen Proposition 5.1, in a general framework, *via* the strong Szegö theorem (see [23, 25, 28]). For $f \in L^1(\mathbb{T})$, denote by $\hat{f}_n$ the Fourier coefficient of order $n \in \mathbb{Z}$ of f. Let

$$E(\mathbb{T}) = \left\{ f \in L^2(\mathbb{T}) \text{ such that } \sum_{n=1}^{+\infty} n |\hat{f}_n|^2 < +\infty \right\}. \tag{5.3}$$

For any complex valued function $f \in L^1(\mathbb{T})$ such that $\log f$ has a meaning together with $\log f \in E(\mathbb{T})$, set

$$\psi(f) = \sum_{n=1}^{\infty} n \left| \widehat{\log f}_n \right|^2 . \tag{5.4}$$

The two following corollaries give the local sharp asymptotic behavior for the likelihood ratio test.

Corollary 5.2. *Assume that $(H3)$ holds. In addition, assume that both $\log g_1$ and $\log g_{\tau_c}$ belong to $E(\mathbb{T})$ where $g_{\tau_c} = (1-2\tau_c)g_1 + 2\tau_c g_0$ and $c \in \mathcal{D}_\varphi$. Then, (W_n) satisfies the LSLDP of order 0 at point $c \in \mathcal{D}_\varphi$, given in Theorem 3.2, with*

$$L(\tau_c) = -\frac{1}{4\pi} \int_{\mathbb{T}} \log(1 - 2\tau_c \varphi(x))\, dx, \tag{5.5}$$

$$\sigma_c^2 = \frac{1}{\pi} \int_{\mathbb{T}} \left(\frac{\varphi(x)}{1 - 2\tau_c \varphi(x)} \right)^2 dx, \tag{5.6}$$

$$H(\tau_c) = \frac{1}{2} \Big(\psi(g_1) - \psi(g_{\tau_c}) \Big). \tag{5.7}$$

Remark 5.3. First, observe that if $(H3)$ holds, then g_1 also lies in the Szegö class. Next, if g_0 and g_1 are continuously derivable with Lipzsichian derivative, then we deduce from Lemma 3.1 that $\mathcal{D}_\varphi = \text{Int}\, \mathcal{C}_\varphi$ and theorem p. 76 of [23] implies that both g_1 and g_{τ_c} belong to $E(\mathbb{T})$. Therefore, we find that (W_n) satisfies an SLDP of order 0.

Proof. In order to apply Theorem 3.2, we have to evaluate the asymptotic behavior of $Z_n(c) = n(L_n(\tau_c) - L(\tau_c))$. We clearly obtain from standard calculus on Gaussian distribution that

$$L_n(\tau) = \frac{1}{n} \log \mathbb{E}_{\mathbb{P}_{g_0}} \left[e^{n\tau W_n} \right]$$

$$L_n(\tau) = \begin{cases} \dfrac{1}{2n} \log \det T_n(g_1) - \dfrac{1}{2n} \log \det T_n(g_\tau) & \text{if } \tau \in \Delta_n \\ +\infty \quad \text{otherwise.} \end{cases} \tag{5.8}$$

Consequently, it follows from (5.2) together with Theorem A.1 and the discussion in Appendix A that, for all $c \in \mathcal{D}_\varphi$

$$Z_n(c) = \frac{1}{2} \Big(\psi(g_1) - \psi(g_{\tau_c}) \Big) + \mathcal{O}(n^{-1})$$

which completes the proof of Corollary 5.2. □

Corollary 5.4. *Assume that g_0 and g_1 both admit analytical extensions on the annulus $\{z \in \mathbb{C} \,/\, \rho < |z| < \rho^{-1}\}$ with $0 < \rho < 1$. Moreover, assume that neither g_0 and g_1 vanish on $\mathbb{T}$. Then, for any $p \geq 0$, (W_n) satisfies an SLDP of order p with L, H and σ^2 given in Corollary 5.2.*

Proof. First, the positivity and the analyticity of g_0 and g_1 imply $(H3)$. Next, it follows from Lemma 3.1 that $\mathcal{D}_\varphi = \text{Int}\, \mathcal{C}_\varphi$. Hence, it remains to show that assumption $(H2(p))$ holds for any $p \geq 0$. From the proof of Corollary 5.2, it holds for $p = 0$. For $p > 0$, we will prove the following stronger property: for all $\nu \in \Delta$ and $k > 0$

$$\frac{\partial^k}{\partial \tau^k} \log \det T_n(g_\nu) = \frac{n}{2\pi} \frac{\partial^k}{\partial \tau^k} \int_{\mathbb{T}} \log(g_\nu(x)) dx + \frac{\partial^k}{\partial \tau^k} \psi(g_\nu) + \mathcal{O}(\theta^{2n})$$

with $0 < \theta < \rho$. Let $\nu \in \Delta$ and $\varepsilon > 0$ with $]\nu - \varepsilon, \nu + \varepsilon[\subset \Delta$. Denote by $D(\nu, \varepsilon)$ the complex open disk centered at ν with radius ε. Set, for all $z \in D(\nu, \varepsilon)$

$$S_n(z) = \log \det(T_n(g_z)) - \frac{n}{2\pi} \int_{\mathbb{T}} \log(g_z(x)) dx.$$

As g_ν is real, we can choose ε such that both the widing number of g_z is 0 and g_z does not vanish on $\mathbb{T}$ (see the App. A and also point b, p. 71 of [9]). Hence, from relation (A.5) of Theorem A.1, we obtain that, for all $z \in D(\nu, \varepsilon)$

$$S_n(z) = \psi(g_z) + R_n(z)$$

where $R_n(z) = o(\theta^{2n})$. In addition, we can rewrite $\psi(g_z)$ as (see formula (1.19), p. 193 of [25])

$$\psi(g_z) = \frac{1}{32\pi} \int_{\mathbb{T}^2} \left[\frac{\log g_z(x) - \log g_z(y)}{\sin \frac{1}{2}(x - y)} \right]^2 dxdy.$$

Therefore, the function $\psi(g_z)$ is analytical on $D(\nu, \varepsilon)$. Finally, using the Cauchy estimates for the derivatives of an analytical function (see *e.g.* [31], Th. 10.26, p. 213), we find that, for any $k > 0$, $R_n^{(k)}(z) = o(\theta^{2n})$. It implies that $(H2(p))$ holds for any $p > 0$, which completes the proof of Corollary 5.4. □

5.1.2. *Application to ARMA processes*

Recently, Barone *et al.* [4] proved, in the ARMA framework (*i.e.* when both g_0 and g_1 are quotients of positive trigonometric polynomials), an LDP for the likelihood ratio statistics given by (5.1). Hence, they calculated, for any appropriate value of c, the limit of $n^{-1} \log \mathbb{P}_{g_0}(W_n \geq c)$, and proposed a numerical way to estimate this asymptotic probability. Actually, the large deviation properties of the sequence (W_n) allows the authors of [4] to estimate $\mathbb{P}_{g_0}(W_n \geq c)$ by using an importance sampling technique. We will complete their results showing an SLDP for the likelihood-ratio statistics by use of the following useful Lemma on Szegö theorems for rational functions (see [23] and App. A and B).

Lemma 5.5. *Assume that g can be written as $g(x) = \sigma^2 |B(e^{ix})/A(e^{ix})|^2$, where A and B are polynomials of degree p and q:*

$$A(e^{ix}) = \prod_{j=1}^{p} (1 - \alpha_j e^{ix}), \qquad B(e^{ix}) = \prod_{k=1}^{q} (1 - \beta_k e^{ix}).$$

Moreover, assume that the roots of A and B are all outside the unit circle (i.e. $|\alpha_j| < 1$ and $|\beta_k| < 1$ for any j and any k). Then, we have

$$\frac{1}{2\pi} \int_{\mathbb{T}} \log g(x) dx = \log \sigma^2.$$

In addition, there exists $0 \leq \gamma < 1$ such that

$$\frac{\det T_n(g)}{(\sigma^2)^n} = \exp \psi(g) + \mathcal{O}(\gamma^n), \tag{5.9}$$

where

$$\exp \psi(g) = \frac{\prod_{j=1}^{p} \prod_{k=1}^{q} |1 - \alpha_j \overline{\beta}_k|^2}{\prod_{j=1}^{p} \prod_{j'=1}^{p} (1 - \alpha_j \overline{\alpha}_{j'}) \prod_{k=1}^{q} \prod_{k'=1}^{q} (1 - \beta_k \overline{\beta}_{k'})} \cdot \tag{5.10}$$

If $q = 0$ (case of an AR(p) process), then, for $n \geq p$, we have the remarkable equality

$$\frac{\det T_n(g)}{(\sigma^2)^n} = \frac{1}{\prod_{j=1}^p \prod_{j'=1}^p (1 - \alpha_j \overline{\alpha}_{j'})}.$$

Proof. For any $|\alpha| < 1$ and any $n > 0$, we have

$$\int_{\mathbb{T}} \log |1 - \alpha e^{ix}|^2 \, dx = 0, \qquad \int_{\mathbb{T}} e^{inx} \log |1 - \alpha e^{ix}|^2 \, dx = -\frac{\alpha^n}{n}. \tag{5.11}$$

Consequently, we find that

$$\widehat{\log g}(n) = \frac{1}{n} \left(\sum_{j=1}^p \alpha_j^n - \sum_{k=1}^q \beta_k^n \right). \tag{5.12}$$

Therefore, we obtain from (5.4) and (5.12) that

$$\psi(g) = \sum_{j=1}^p \sum_{k=1}^q \log \left|1 - \alpha_j \overline{\beta}_k\right|^2 - \frac{1}{2} \sum_{j=1}^p \sum_{j'=1}^p \log |1 - \alpha_j \overline{\alpha}_{j'}|^2 - \frac{1}{2} \sum_{k=1}^q \sum_{k'=1}^q \log \left|1 - \beta_k \overline{\beta}_{k'}\right|^2.$$

As all the roots of A and B are outside the unit circle (5.9) immediately follows from Theorem A.1. □

Next, let g_0 and g_1 be two spectral densities of ARMA processes:

$$g_0(x) = \sigma_0^2 \left| \frac{B_0(e^{ix})}{A_0(e^{ix})} \right|^2, \qquad g_1(x) = \sigma_1^2 \left| \frac{B_1(e^{ix})}{A_1(e^{ix})} \right|^2 \tag{5.13}$$

where A_0, B_0, A_1 and B_1 are polynomials of degree p_0, q_0, p_1 and q_1, whose roots are all outside the unit circle. Recall that here the function L is finite on $\Delta = \{\tau \in \mathbb{R} \ / \ (1 - 2\tau) g_1 + 2\tau g_0 > 0\}$. For any $\tau \in \Delta$, we consider the polynomial F_τ of degree $r = \max(p_0 + q_1, p_1 + q_0)$ defined on $\mathbb{C}$ by

$$F_\tau(z) = (1 - 2\tau) \left|A_0(z) B_1(z)\right|^2 + 2\tau \frac{\sigma_0^2}{\sigma_1^2} \left|A_1(z) B_0(z)\right|^2. \tag{5.14}$$

Thus, there exists $f_0(\tau), f_1(\tau), \ldots, f_r(\tau)$ and $z_1(\tau), z_2(\tau), \ldots, z_r(\tau)$ such that

$$\begin{aligned} F_\tau(z) &= f_0(\tau) + \sum_{j=1}^r f_j(\tau)(z^j + \overline{z}^j) && (5.15) \\ &= \frac{(-1)^r f_r(\tau)}{\prod_{j=1}^r z_j(\tau)} \prod_{j=1}^r (1 - z_j(\tau) z) \left(1 - \frac{z_j(\tau)}{z}\right). && (5.16) \end{aligned}$$

Here, the $z_1(\tau), \ldots, z_r(\tau)$ are the r roots of F_τ having a modulus lesser than 1. Then, for $\tau \in \Delta$, we have from (5.2) together with (5.11)

$$\begin{aligned} L(\tau) &= -\frac{1}{4\pi} \int_{\mathbb{T}} \log(1 - 2\tau + 2\tau \frac{g_0(x)}{g_1(x)}) \, dx = -\frac{1}{4\pi} \int_{\mathbb{T}} \log \frac{F_\tau(e^{ix})}{|A_0(e^{ix}) B_1(e^{ix})|^2} \, dx \\ &= -\frac{1}{2} \log \left(\frac{(-1)^r f_r(\tau)}{\prod_{j=1}^r z_j(\tau)} \right). && (5.17) \end{aligned}$$

Now, for $\tau \in \Delta$, set $g_\tau = (1-2\tau)g_1 + 2\tau g_0$. We clearly have from (5.13) and (5.14)

$$g_\tau(x) = \frac{\sigma_1^2 F_\tau(e^{ix})}{|A_0(e^{ix})A_1(e^{ix})|^2}. \tag{5.18}$$

Consequently, for $\tau \in \Delta$, g_τ is the spectral density of an ARMA(r, p_0+p_1) stationary process and $(H3)$ is satisfied. We can use Lemma 5.5 to compute $\psi(g_\tau)$ and equation (5.7) to calculate $H(\tau)$ and its derivatives. We are now in position to precise the sharp asymptotic behavior of the likelihood-ratio statistics.

Corollary 5.6. *In the ARMA framework, for any $p \geq 0$, (W_n) satisfies an SLDP of order p, with*

$$L(\tau_c) = -\frac{1}{2}\log\left(\frac{(-1)^r f_r(\tau_c)}{\prod_{j=1}^r z_j(\tau_c)}\right), \tag{5.19}$$

$$\sigma_c^2 = \frac{1}{\pi}\int_{\mathbb{T}}\left(\frac{\varphi(x)}{1-2\tau_c\varphi(x)}\right)^2 dx, \tag{5.20}$$

$$H(\tau_c) = \frac{1}{2}\Big(\psi(g_1) - \psi(g_{\tau_c})\Big) \tag{5.21}$$

where ψ is given in relation (5.10).

5.1.3. *Numerical experiments*

As a simple example, assume that we want to test H_0: *X is a white noise versus* H_1: *X is a AR(1) process with parameter* $\alpha = 0.1$. More precisely, assume that the spectral density of X has the form $g(x) = |1-\alpha e^{ix}|^{-2}$ and we want to test $H_0 :\ \alpha = 0$ *versus* $H_1 :\ \alpha = 0.1$. For a given value of c and a given n, we first estimate the unknown probability $\mathbb{P}_{g_0}(W_n \geq c)$ by a Monte-Carlo procedure, using the importance sampling scheme proposed by Barone *et al.* [4]. The size of the Monte-Carlo sample is large enough, in order to get a very good accuracy in the estimation and to assume that the true value of $\mathbb{P}_{g_0}(W_n \geq c)$ is known. The aim of this section is to compute an approximation of $\mathbb{P}_{g_0}(W_n \geq c)$ by use of Theorem 3.2, and to study numerically its behavior. The functions L and H are explicitely given in Corollary 5.6. The order p is the only parameter which must be chosen arbitrarily. Unfortunately, there is no general rule for an optimal practical choice of p. Indeed, the accuracy of the LSLDP approximation, and then the choice of p, strongly depends on the sequence (W_n) together with n and c. Nevertheless, we can mention that, in our examples, very good results are obtained with small values of p.

Figure 1 shows the approximations of $\mathbb{P}_{g_0}(W_n \geq c)$ with $n = 100$, obtained with $p = 0$ in (3.2) and $p = 1$ in (3.3). We clearly see that both approximations are very accurate for probabilities smaller than 0.01.

5.2. **Sum of squares**

The large deviation properties of the sum of squares

$$W_n = \frac{1}{n}\sum_{j=1}^n X_j^2 \tag{5.22}$$

was extensively studied by Bryc and Dembo [11] (see also [12] for the autoregressive process and [13] for a heuristic approach). Bucklew and Sadowsky [14] also established LSLDP of order 0 for (W_n). Their approach is different from ours, they did not estimate the normalized cumulant generating function L_n of W_n by functions L and H. Using the same arguments as in the proofs of Corollaries 5.2 and 5.4 with $\varphi = g$ where g is the spectral density of (X_n), we obtain the next two corollaries which precise the local sharp asymptotic behavior for the sum of squares.

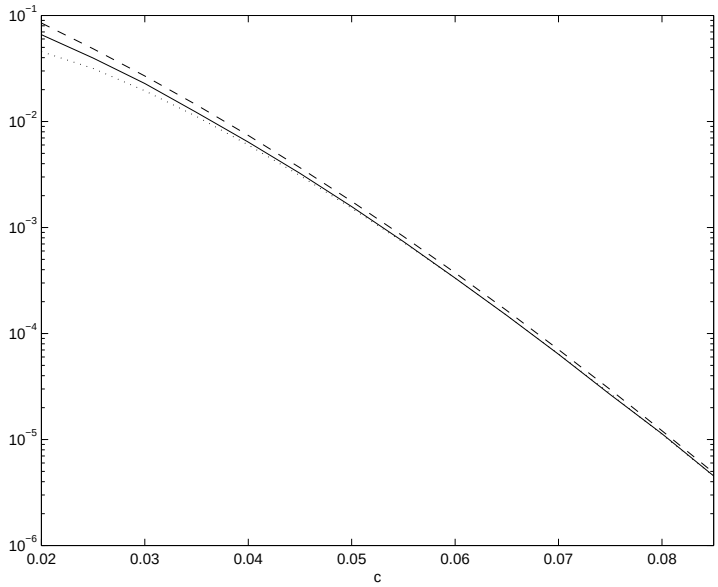

FIGURE 1. Application of the SLDP for the likelihood-ratio test. Comparison of different approximations of $\mathbb{P}(W_n > c)$ obtained with $n = 100$ for different values of c. —— : the true probability $\mathbb{P}(W_n > c)$, $--$: the LSLDP approximation obtained with $p = 0$, $\cdot$: the LSLDP approximation obtained with $p = 1$.

Corollary 5.7. *Let*

$$L(\tau) = -\frac{1}{4\pi} \int_{\mathbb{T}} \log(1 - 2\tau g(x))\, dx \tag{5.23}$$

and $c \in \mathcal{D}_\varphi$ such that τ_c satisfies $\log(1 - 2\tau_c g) \in \mathbb{E}(\mathbb{T})$. Then, (W_n) satisfies the LSLDP of order 0 at c, given in Theorem 3.2, with

$$\sigma_c^2 = \frac{1}{\pi} \int_{\mathbb{T}} \left(\frac{g(x)}{1 - 2\tau_c g(x)} \right)^2 dx, \tag{5.24}$$

$$H(\tau_c) = \frac{1}{2} \psi(1 - 2\tau_c g). \tag{5.25}$$

Corollary 5.8. *Assume that g admits an analytical extension on the annulus $\{z \in \mathbb{C}\ ;\ \rho < |z| < \rho^{-1}\}$ with $0 < \rho < 1$. Moreover, assume that g does not vanish on $\mathbb{T}$. Then, for any $p \geq 0$, (W_n) satisfies an SLDP of order p with L, H and σ^2 given in Corollary 5.7.*

Similar calculus from those of Section 5.1.2 can be easily achieved for the sum of squares when (X_n) is an ARMA(p, q) process with spectral density g. In this framework, $g_\tau = 1 - 2\tau g$ is the spectral density of an ARMA(r, q) process with $r = \max(p, q)$.

5.3. Yule-Walker estimator

Consider the autoregressive process

$$X_{n+1} = \theta X_n + \varepsilon_{n+1}, \qquad |\, \theta \,| < 1 \tag{5.26}$$

where (ε_n) is i.i.d. with $\mathcal{N}(0,\sigma^2)$ distribution. Assume that X_0 is independent of (ε_n) with $\mathcal{N}(0,\sigma^2/(1-\theta^2))$ distribution. (X_n) is a centered stationary Gaussian process with spectral density defined for all $x \in \mathbb{T}$ by $g(x) = \sigma^2(1+\theta^2-2\theta\cos x)^{-1}$. Let $\tilde{\theta}_n$ be the Yule-Walker estimator of the parameter θ

$$\tilde{\theta}_n = \frac{\sum_{i=1}^{n} X_i X_{i-1}}{\sum_{i=0}^{n} X_i^2}. \tag{5.27}$$

5.3.1. *Sharp large deviations for the Yule-Walker estimator*

The large deviation properties of $(\tilde{\theta}_n)$ are related to the ones of

$$W_n(c) = \frac{1}{n}\left(\sum_{i=1}^{n} X_i X_{i-1} - c\sum_{i=0}^{n} X_i^2\right) \tag{5.28}$$

with $c \in \mathbb{R}$ since $\mathbb{P}(\tilde{\theta}_n \geq c) = \mathbb{P}(W_n(c) \geq 0)$. One has to keep in mind that the threshold c for $\tilde{\theta}_n$ appears like a parameter for $W_n(c)$. Thus, we only have to consider the threshold zero for $W_n(c)$. This key-point allows us to establish the following result [7] and also to calculate the sharp large deviation behavior for $(\tilde{\theta}_n)$.

Proposition 5.9. $(\tilde{\theta}_n)$ *satisfies an LDP with rate function*

$$L^\star(c) = \begin{cases} \frac{1}{2}\log\left(\frac{1+\theta^2-2\theta c}{1-c^2}\right) & \text{if } \mid c \mid < 1 \\ +\infty & \text{otherwise.} \end{cases} \tag{5.29}$$

Corollary 5.10. *For any $p \geq 0$, $(\tilde{\theta}_n)$ satisfies an SLDP of order p with, for $\mid c \mid < 1$, $L(\tau_c) = -L^\star(c)$ and*

$$\tau_c = \frac{\theta(1+c^2) - c(1+\theta^2)}{c^2-1}, \qquad \sigma_c^2 = \frac{1-c^2}{(1+\theta^2-2\theta c)^2} \tag{5.30}$$

$$H(\tau_c) = -\frac{1}{2}\log\left(\frac{(1-c\theta)^4}{(1-\theta)^2(1+\theta^2-2\theta c)(1-c^2)^2}\right). \tag{5.31}$$

Proof. We can rewrite (5.28) as

$$W_n(c) = \frac{1}{n} X^{(n)*} T_n(f_c) X^{(n)} \tag{5.32}$$

where $T_n(f_c)$ is the Toeplitz matrix associated to $f_c(x) = \cos x - c$, $x \in \mathbb{T}$. Set $\rho = 1+\theta^2+2c\tau$, $q = -\theta-\tau$, $r = \rho - \theta^2$ and, for $\rho > 2 \mid q \mid$

$$\alpha = \frac{\rho + \sqrt{\rho^2-4q^2}}{2} \quad \text{and} \quad \beta = \frac{\rho - \sqrt{\rho^2-4q^2}}{2}. \tag{5.33}$$

Exactly as in [7], the normalized cumulant generating function of $W_n(c)$ is

$$L_n(\tau) = -\frac{1}{2n}\log\Big[(1-\theta^2)^{-1}(\alpha-\beta)^{-1}\left((r-\beta)^2\alpha^n - (r-\alpha)^2\beta^n\right)\Big] \tag{5.34}$$

if $\tau \in \Delta_n$ and $L_n(\tau) = +\infty$ otherwise. The asymptotic cumulant generating function is

$$L(\tau) = \begin{cases} -\dfrac{1}{2}\log\left(\dfrac{\rho + \sqrt{\rho^2 - 4q^2}}{2}\right) & \text{if } \tau \in \Delta \\ +\infty \quad \text{otherwise.} \end{cases} \tag{5.35}$$

Let τ_c be the unique point of Δ such that $L^{'}(\tau_c) = 0$ *i.e.*

$$\tau_c = \frac{\theta(1+c^2) - c(1+\theta^2)}{c^2 - 1}. \tag{5.36}$$

First of all, we have $L^\star(c) = -L(\tau_c)$ and $\sigma_c^2 = L^{''}(\tau_c)$. Next, from (5.34) together with (5.35), for any $\tau \in \Delta$

$$\lim_{n\to+\infty} n\Big(L_n(\tau) - L(\tau)\Big) = H(\tau) = -\frac{1}{2}\log\Big[(1-\theta^2)^{-1}(\alpha-\beta)^{-1}(r-\beta)^2\Big].$$

It is easy to check that

$$H(\tau_c) = -\frac{1}{2}\log\left(\frac{(1-c\theta)^4}{(1-\theta)^2(1+\theta^2-2\theta c)(1-c^2)^2}\right). \tag{5.37}$$

Finally, for any $\tau \in \Delta$, we can explicitly calculate

$$L_n(\tau) - L(\tau) - \frac{1}{n}H(\tau) = -\frac{1}{2n}\log\left(1 - \frac{(r-\alpha)^2}{(r-\beta)^2}\left(\frac{\beta}{\alpha}\right)^n\right). \tag{5.38}$$

Consequently, for all $p \geq 0$, the condition $(H2(p))$ is satisfied since we always have $0 \leq \beta < \alpha$. In particular, if

$$\delta = \left(\frac{c^2(1+\theta^2-2\theta c) - (1-c^2)\theta^2}{(1+\theta^2-2\theta c) - (1-c^2)\theta^2}\right)^2,$$

we have from (5.38)

$$L_n(\tau_c) - L(\tau_c) - \frac{1}{n}H(\tau_c) = -\frac{1}{2n}\log\left(1 - \delta c^{2n}\right) \tag{5.39}$$

and this remainder vanishes exponentially fast to zero since $\mid c \mid < 1$. □

5.3.2. *Numerical experiments*

We consider the autoregressive process given in (5.26) with $\theta = 0.5$. This parameter is estimated by the Yule-Walker estimator $\tilde{\theta}_n$.

Table 1 presents different approximations of $\mathbb{P}(\tilde{\theta}_n > c)$ for $c \in [0.55\,,\,0.95]$ and $n = 50$. As it was already mentioned by Jensen [27], the normal approximation breaks down from small probabilities (smaller than 10^{-2}). The Edgeworth expansion slightly improves the results for medium deviations, but it is not of much help for small tail probabilities. On the other hand, the approximation obtained from the SLDP is very poor for large probabilities and very accurate for small tail probabilities. The inclusion of the correction terms (with $p \geq 1$ in (3.3)) is not of great importance here. It is possible in this example to compute the saddlepoint approximation given in (4.5), since the function L_n is explicitly known. For a practical point of view, it is interesting to remark that this approximation is very accurate for all table entries.

TABLE 1. Different approximations of $\mathbb{P}(\tilde{\theta}_n > c)$ obtained for different values of c with $n = 50$ and $\theta = 0.5$.

c	$\mathbb{P}(\tilde{\theta}_n > c)$	CLT	Edgeworth	SLDP		Saddle
				$p = 0$	$p = 1$	
0.55	.277	.282	.272	.800	76.8	.298
0.60	.142	.150	.137	.259	1.40	.155
0.70	$.58\ 10^{-1}$	$.58\ 10^{-1}$	$.55\ 10^{-1}$	$.83\ 10^{-1}$	.116	$.61\ 10^{-1}$
0.75	$.16\ 10^{-1}$	$.13\ 10^{-1}$	$.15\ 10^{-1}$	$.20\ 10^{-1}$	$.18\ 10^{-1}$	$.16\ 10^{-1}$
0.80	$.24\ 10^{-2}$	$.75\ 10^{-3}$	$.21\ 10^{-2}$	$.29\ 10^{-2}$	$.25\ 10^{-2}$	$.25\ 10^{-2}$
0.85	$.15\ 10^{-3}$	$-.49\ 10^{-3}$	$-.88\ 10^{-4}$	$.19\ 10^{-3}$	$.16\ 10^{-3}$	$.16\ 10^{-3}$
0.90	$.27\ 10^{-5}$	$-.22\ 10^{-3}$	$-.13\ 10^{-3}$	$.29\ 10^{-5}$	$.24\ 10^{-5}$	$.25\ 10^{-5}$
0.95	$.21\ 10^{-8}$	$-.68\ 10^{-4}$	$-.48\ 10^{-4}$	$.29\ 10^{-8}$	$.24\ 10^{-8}$	$.25\ 10^{-8}$

5.4. **Empirical spectral repartition function**

We will now study the sharp large deviation property of the empirical spectral repartition function. First of all, let $G(t) = \int_{-t}^{t} g(x)dx/2\pi$ be the spectral repartition function over the interval $[-t, t]$ with $t \in [0, \pi]$. $G(t)$ can be estimated, on the basis of the observation $X_1, \dots, X_n$, by

$$W_n(t) = \frac{1}{2\pi} \int_{-t}^{t} \mathcal{I}_n(x)dx \tag{5.40}$$

where $\mathcal{I}_n$ is the empirical periodogram

$$\mathcal{I}_n(x) = \frac{1}{n} \left| \sum_{j=1}^{n} e^{ijx} X_j \right|^2 . \tag{5.41}$$

One can rewrite $W_n(t)$ as expression (1.3) with $M_n = T_n(\mathbb{I}_{[-t,t]})$. The sequence $(W_n(t))$ can be used in several statistical applications (see *e.g.* [29] for the spectral change point problem). We first propose an LDP for $(W_n(t))$. To our knowledge, this result is new.

Proposition 5.11. *Assume that $g \in L^\infty(\mathbb{T})$ and let $t \in [0, \pi]$ such that*

$$\| g \|_\infty = \sup_{x \in [-t,t]} g(x). \tag{5.42}$$

Then, $(W_n(t))$ satisfies an LDP with rate function $L^\star$ given in (2.5) with $\varphi = g\mathbb{I}_{[-t,t]}$.

Proof. By (5.42), the proof immediately follows from Corollary 2 of [7]. □

Unfortunately, the function φ is not in general smooth enough in order to obtain, *via* the Toeplitz operators theory (see App. A and B), a refinement through an SLDP for $(W_n(t))$. For this reason, we will consider a regularized version of $(W_n(t))$. Actually, instead of using the function $\mathbb{I}_{[-t,t]}$, we will now study the tapered empirical spectral function

$$V_n(t) = \frac{1}{2\pi} \int_{\mathbb{T}} \chi_t(x)\mathcal{I}_n(x)dx \tag{5.43}$$

where the function χ_t is such that $\chi_t(x) = 1$ on $[-t, t]$, $0 \leq \chi_t(x) < 1$ if $x \notin [-t, t]$ and there exists a small $\epsilon > 0$ such that $\chi_t(x)$ vanishes outside of $]t - \epsilon, t + \epsilon[$. Taperization is widely used in statistics of time series (see *e.g.* [24], Chap. 4). To state the main result of this section, we need some additional notations. For any function $f \in L^\infty(\mathbb{T})$, define the infinite Toeplitz matrix $T(f) = (\hat{f}_{i-j})_{i,j\geq 0}$. For $a \in \mathbb{R}$ and for $f, g \in L^\infty(\mathbb{T})$, denote whenever it makes sense (see App. B)

$$\exp \psi_a(f, g) = \Gamma_a(f, g) \det \left[T(1 + afg)T((1 + afg)^{-1})\right] \tag{5.44}$$

where

$$\Gamma_a(f, g) = \det \left[T(1 + afg)^{-1}\Big(I + aT(f)T(g)\Big)\Big(I + aT(g)T(f)\Big)T(1 + afg)^{-1}\right]$$

and I is the infinite identity matrix. The following Corollary gives the sharp asymptotic behavior for the tapered empirical spectral function.

Corollary 5.12. *Assume that $g \in L^\infty(\mathbb{T})$ and let $t \in [0, \pi]$ such that (5.42) is satisfied. Moreover, assume that both χ_t and g are two times continuously differentiable on $\mathbb{T}$. Then, $(V_n(t))$ satisfies an SLDP of order 0, with*

$$L(\tau_c) = -\frac{1}{4\pi}\int_{\mathbb{T}} \log\left(1 - 2\tau_c\chi_t(x)g(x)\right) dx, \tag{5.45}$$

$$\sigma_c^2 = \frac{1}{\pi}\int_{\mathbb{T}} \left(\frac{g(x)\chi_t(x)}{1 - 2\tau_c g(x)\chi_t(x)}\right)^2 dx, \tag{5.46}$$

$$H(\tau_c) = \psi_{-2\tau_c}(\chi_t, g). \tag{5.47}$$

Proof. Since both g and χ_t are assumed to be two times continuously differentiable on $\mathbb{T}$, Lemma 3.1 gives $\mathrm{Int}\, \mathcal{C}_\varphi = \mathcal{D}_\varphi$. So, it remains to show that assumption $(H2(0))$ is satisfied and to calculate H. This is achieved by applying Theorem B.2. □

Appendix A. Asymptotic behavior of the eigenvalues of a Toeplitz form

In this appendix, we will recall sharp results on the asymptotic behavior of the eigenvalues of Toeplitz matrices. To be more precise, we will focus our attention on the remainder in the strong Szegö theorem. For $f \in L^1(\mathbb{T})$, denote by $\hat{f}_n$ the Fourier coefficient of order $n \in \mathbb{Z}$ of f. Let

$$E(\mathbb{T}) = \left\{ f \in L^2(\mathbb{T}) \text{ such that } \sum_{n=1}^{+\infty} n|\hat{f}_n|^2 < +\infty \right\}. \tag{A.1}$$

For any non negative function $f \in L^1(\mathbb{T})$ with $\log f \in E(\mathbb{T})$, set

$$\Theta(f) = \exp\left(\frac{1}{2\pi}\int_{\mathbb{T}} \log f(x)dx\right), \qquad \psi(f) = \sum_{n=1}^{\infty} n|\widehat{\log f}_n|^2. \tag{A.2}$$

If f is the spectral density of an autoregressive process of order p then, for all $n \geq p$, we have the remarkable equality (see [23] formulas (18) and (22), pp. 78-79)

$$\frac{\det T_n(f)}{\Theta(f)^n} = \exp\Big(\psi(f)\Big). \tag{A.3}$$

This formula was first extended by Grenander and Szegö ([23] formula (4) p. 76) to the celebrated strong Szegö theorem which states that if f is smooth enough then (A.3) remains true, up to $o(1/n)$. Later, Golinskii and Ibragimov [22] showed that the formula (A.3) is always true up to $o(1/n)$, as soon as $\log f \in E(\mathbb{T})$. Finally, other extensions in many different directions were developed (see for example the probabilistic approach of Johansson [28] and the references therein). In this paper, we mainly use the strong Szegö theorem for complex valued functions established by Hartwing and Fisher [25].

Define the complex valued function f on $\mathbb{T}$ by $f(x) = F(e^{ix})$ where F is the generating function on $\mathbb{C}$ associated with the Fourier coefficients of f. Denote by ind (f) the winding number of f. Actually, ind (f) is the increment of the argument of f on $\mathbb{T}$ (see [9], p. 70).

Theorem A.1. *Assume that f does not vanish on $\mathbb{T}$ and ind $(f) = 0$. In addition, assume that*

$$\sum_{n=1}^{+\infty} \tilde{f}_n < +\infty \qquad \text{where} \qquad \tilde{f}_n = \max_{k \geq |n|} \Big(|\hat{f}_k|, |\hat{f}_{-k}| \Big) . \tag{A.4}$$

Then, we have

$$\frac{\det T_n(f)}{\Theta(f)^n} = \exp\Big(\psi(f) + R_n(f) \Big) \tag{A.5}$$

where the rate of convergence to zero of the remainder $R_n(f)$ can be explicitly evaluated from the behavior of the Fourier coefficients of $\log f$. For instance, if the analytical extension F of f possesses an annulus of analyticity $\rho \leq |z| \leq \rho^{-1}$ with $0 < \rho < 1$, then $R_n(f) = \mathcal{O}(\rho^{2n})$ and (A.4) is always satisfied since $\tilde{f}_n = \mathcal{O}(\rho^n)$.

Appendix B. Extension to products of Toeplitz forms

The goal of this appendix is to provide a large class of functions for which the strong Szegö theorem remains valid for products of Toeplitz forms. First of all, we recall some standard useful results concerning Toeplitz and Hankel operators (see [34] and the references therein). For $f \in L^\infty(\mathbb{T})$, we can define the $l^2(\mathbb{N})$ Toeplitz and Hankel operators respectively as

$$T(f) = \Big(\hat{f}_{i-j} \Big)_{i,j \geq 0} \qquad \text{and} \qquad H(f) = \Big(\hat{f}_{i+j+1} \Big)_{i,j \geq 0} . \tag{B.1}$$

A well-known identity relating Toeplitz and Hankel operators is

$$T(fg) - T(f)T(g) = H(f)H(\tilde{g}) \tag{B.2}$$

where $\tilde{g}(x) = g(-x)$. The analogue of (B.2) for finite Toeplitz matrices is

$$T_n(fg) - T_n(f)T_n(g) = P_n H(f) H(\tilde{g}) P_n + Q_n H(\tilde{f}) H(g) Q_n \tag{B.3}$$

where P_n and Q_n are defined by

$$\begin{aligned} P_n(x_0, x_1, \cdots) &= (x_0, x_1, \cdots, x_n, 0, \cdots) \\ Q_n(x_0, x_1, \cdots) &= (x_n, x_{n-1}, \cdots, x_0, 0, \cdots) \end{aligned}$$

and $T_n(f)$ is identified with $P_n T(f) P_n$. If $f \in E(\mathbb{T})$ then $H(f)$ is an Hilbert-Schmidt operator since

$$\mathrm{Tr}(H(f)H(\tilde{f})) = \sum_{n=1}^{+\infty} n |\hat{f}_n|^2 < +\infty.$$

It was proven in [26] that $H(f)$ is of trace class if

$$\sum_{n=1}^{+\infty} \sqrt{n} \left(\sum_{k=n}^{+\infty} |\hat{f}_{n+k} - \hat{f}_{n+k+1}|^2 \right)^{\frac{1}{2}} < +\infty. \tag{B.4}$$

In particular, $H(f)$ is of trace class if $\hat{f}_{n+1} - \hat{f}_n = \mathcal{O}(n^{-\alpha})$ with $\alpha > 2$. Thus, this is true if $\hat{f}_n = \mathcal{O}(n^{-\beta})$ with $\beta > 1$. Finally, the product of two Hilbert-Schmidt operators is of trace class and the product of a trace class operator with a compact operator is of trace class.

The following result due to Basor [5,6] is the generalization of the strong Szegö theorem for the matrix product $T_n(f)T_n(g)$. It is essentially based on identity (B.3) together with usual properties of trace class operators.

Theorem B.1. *Assume that $f, g \in L^\infty(\mathbb{T})$ and that f and g do not vanish on $\mathbb{T}$. In addition, assume that $H(f)H(\tilde{g})$ and $H(\tilde{f})H(g)$ are both of trace class. Then, we have*

$$\lim_{n\to+\infty} \frac{\det(T_n(fg))}{\det(T_n(f))\det(T_n(g))} = \Lambda(f,g) \tag{B.5}$$

where $\Lambda(f,g) = \det\left[T(f)^{-1}T(fg)T(g)^{-1}T(g)^{-1}T(fg)T(f)^{-1}\right]$.

Theorem B.2. *Assume that $f, g \in L^\infty[\mathbb{T})$ with $H(f)H(\tilde{g})$ and $H(\tilde{f})H(g)$ both of trace class. Choose $a \in \mathbb{R}$ such that the function $1 + afg$ does not vanish on $\mathbb{T}$ and $\log(1 + afg) \in E(\mathbb{T})$. Then, we have*

$$\lim_{n\to+\infty} \frac{\det(I_n + aT_n(f)T_n(g))}{\Theta(1+afg)^n} = \Gamma_a(f,g)\det\left[T(1+afg)T((1+afg)^{-1})\right]$$

where

$$\Gamma_a(f,g) = \det\left[T(1+afg)^{-1}\Big(I + aT(f)T(g)\Big)\Big(I + aT(g)T(f)\Big)T(1+afg)^{-1}\right]$$

and I_n denotes the identity on $\mathbb{C}^n$.

Remark B.3. Widom [33] proved that for $f, \log f \in E(\mathbb{T})$, the nice identity $\exp\Big(\psi(f)\Big) = \det\left[T(f)T(f^{-1})\right]$.

We now give the main lines of the proof of the last theorem. This proof follows from the same arguments as that of Theorem B.1 (see Basor [5], p. 976). For any a satisfying the assumptions of Theorem B.2, we may write by equation (B.3)

$$I_n + aT_n(f)T_n(g) = T_n(1+afg) - aP_nH(f)H(\tilde{g})P_n - aQ_nH(\tilde{f})H(g)Q_n.$$

Hence, we have

$$T_n^{-1}(1+afg)(I_n + aT_n(f)T_n(g)) = I_n + A_n + B_n$$

where

$$A_n = -aT_n^{-1}(1+afg)P_nH(f)H(\tilde{g})P_n, \tag{B.6}$$

$$B_n = -aT_n^{-1}(1+afg)Q_nH(\tilde{f})H(g)Q_n. \tag{B.7}$$

From the assumption, we may deduce that the product A_nB_n converges to 0 in the nuclear operator norm. This implies that

$$\frac{\det(I_n + aT_n(f)T_n(g))}{D_n(1+afg)} = \det(I_n + A_n)\det(I_n + B_n) + o(1).$$

On one hand, we have

$$\lim_{n\to\infty} \det(I_n + A_n) = \det(I - aT^{-1}(1+afg)H(f)H(\tilde{g})).$$

On the other hand, we also obtain that

$$\lim_{n\to\infty} \det(I_n + B_n) = \det(I - aT^{-1}(\widetilde{1+afg})H(\tilde{f})H(g)).$$

Finally, we find that

$$\begin{aligned}\lim_{n\to\infty} \frac{\det(I_n + aT_n(f)T_n(g))}{D_n(1+afg)} &= \det[T^{-1}(1+afg)(I+aT(f)T(g))]\det[T^{-1}(\widetilde{1+afg})(I+aT(\tilde{f})T(\tilde{g}))] \\ &= \det[T^{-1}(1+afg)(I+aT(f)T(g))(I+aT(g)T(f))T^{-1}(1+afg)]. \quad \square\end{aligned}$$

The authors want to thank F. Castell, D. Dacunha-Castelle, C. Léonard and A. Rouault for helpful discussions.

References

[1] Azencott R. and Dacunha-Castelle D., *Séries d'observations irrégulières*. Masson (1984).
[2] Bahadur R. and Ranga Rao R., On deviations of the sample mean. *Ann. Math. Statist.* **31** (1960) 1015–1027.
[3] Barndoff-Nielsen O.E. and Cox D.R., *Asymptotic techniques for uses in statistics*. Chapman and Hall, Londres (1989).
[4] Barone P., Gigli A. and Piccioni M., Optimal importance sampling for some quadratic forms of A.R.M.A. processes. *IEEE Trans. Inform. Theory* **41** (1995) 1834–1844.
[5] Basor E., A localization theorem for Toeplitz determinants. *Indiana Univ. Math. J.* **28** (1979) 975–983.
[6] Basor E., Asymptotic formulas for Toeplitz and Wiener-Hopf operators. *Integral Equations Operator Theory* **5** (1982) 659–665.
[7] Bercu B., Gamboa F. and Rouault A., Large deviations for quadratic forms of stationary Gaussian processes. *Stochastic Process. Appl.* **71** (1997) 75–90.
[8] Book S.A., Large deviation probabilities for weighted sums. *Ann. Math. Statist.* **43** (1972) 1221–1234.
[9] Bottcher A. and Silbermann. *Analysis of Toeplitz operators*. Springer, Berlin (1990).
[10] Bouaziz M., Testing Gaussian sequences and asymptotic inversion of Toeplitz operators. *Probab. Math. Statist.* **14** (1993) 207–222.
[11] Bryc W. and Dembo A., Large deviations for quadratic functionals of Gaussian processes. *J. Theoret. Probab.* **10** (1997) 307–332.
[12] Bryc W. and Smolenski W., On large deviation principle for a quadratic functional of the autoregressive process. *Statist. Probab. Lett.* **17** (1993) 281–285.
[13] Bucklew J.A., *Large deviations techniques in decision, simulation, and estimation*. Wiley (1990).
[14] Bucklew J. and Sadowsky J., A contribution to the theory of Chernoff bounds. *IEEE Trans. Inform. Theory* **39** (1993) 249–254.
[15] Coursol J. and Dacunha-Castelle D., Sur la formule de Chernoff pour deux processus gaussiens stationnaires. *C. R. Acad. Sci. Sér. I Math.* **288** (1979) 769–770.
[16] Cramér H., *Random variables and probability distributions*. Cambridge University Press (1970).
[17] Dacunha-Castelle D., Remarque sur l'étude asymptotique du rapport de vraisemblance de deux processus gaussiens. *C. R. Acad. Sci. Sér. I Math.* **288** (1979) 225–228.
[18] Dembo A. and Zeitouni O., *Large deviations techniques and applications*. Jones and Barblett Pub. Boston (1993).
[19] Esseen C., Fourier analysis of distribution functions. *Acta Math.* **77** (1945) 1–25.
[20] Gamboa F. and Gassiat E., Sets of superresolution and the maximum entropy method on the mean. *SIAM J. Math. Anal.* **27** (1996) 1129–1152.
[21] Gamboa F. and Gassiat E., Bayesian methods for ill posed problems. *Ann. Statist.* **25** (1997) 328–350.
[22] Golinskii B. and Ibragimov I., On Szegös limit theorem. *Math. USSR- Izv.* **5** (1971) 421–444.
[23] Grenander V. and Szegö G., *Toeplitz forms and their applications*. University of California Press (1958).
[24] Guyon X., *Random fields on a network/ modeling, statistics and applications*. Springer (1995).
[25] Hartwig R.E. and Fisher M.E., Asymptotic behavior of Toeplitz matrices and determinants. *Arch. Rational Mech. Anal.* **32** (1969) 190–225.

[26] Howland J., Trace class Hankel operators. *Quart. J. Math. Oxford Ser. (2)* **22** (1971) 147–159.
[27] Jensen J.L., Saddlepoint Approximations. *Oxford Statist. Sci. Ser.* **16** (1995).
[28] Johansson K., On Szegös asymptotic formula for Toeplitz determinants and generalizations. *Bull. Sci. Math.* **112** (1988) 257–304.
[29] Lavielle M., Detection of changes in the spectrum of a multidimensional process. *IEEE Trans. Signal Process.* **42** (1993) 742–749.
[30] Lehmann E.L., *Testing statistical hypotheses*. John Wiley and Sons, New-York (1959).
[31] Rudin W., *Real and complex analysis*. McGraw Hill International Editions (1987).
[32] Taniguchi M., *Higher order asymptotic theory for time series analysis*. Springer, Berlin (1991).
[33] Widom H., On the limit block Toeplitz determinants. *Proc. Amer. Math. Soc.* **50** (1975) 167–173.
[34] Widom H., Asymptotic behavior of block Toeplitz matrices and determinants II. *Adv. Math.* **21** (1976).