Stat Inference Stoch Process (2012) 15:225–239
DOI 10.1007/s11203-012-9071-9

On large deviations in testing simple hypotheses for locally stationary Gaussian processes

Inder Tecuapetla-Gómez · Michael Nussbaum

Received: 10 March 2011 / Accepted: 25 November 2011 / Published online: 25 September 2012

Abstract We derive a large deviation result for the log-likelihood ratio for testing simple hypotheses in locally stationary Gaussian processes. This result allows us to find explicitly the rates of exponential decay of the error probabilities of type I and type II for Neyman–Pearson tests. Furthermore, we obtain the analogue of classical results on asymptotic efficiency of tests such as Stein's lemma and the Chernoff bound, as well as the more general Hoeffding bound concerning best possible joint exponential rates for the two error probabilities.

Keywords Hypothesis testing · Likelihood ratio · Large deviations · Locally stationary Gaussian processes · Hoeffding bound · Stein's lemma · Chernoff bound

1 Introduction and main results

Consider a locally stationary Gaussian process $X := (X_{k,n})_{1\leq k\leq n}$ and the problem of testing

$$H_0 : X \sim \mathbf{P}_n \quad \text{vs} \quad H_1 : X \sim \mathbf{Q}_n, \tag{1.1}$$

where $\mathbf{P}_n \sim \mathcal{N}_n(0, \Sigma_n)$ and $\mathbf{Q}_n \sim \mathcal{N}_n(0, \widetilde{\Sigma}_n)$. In our setting, both hypotheses remain fixed as $n \to \infty$, so the study of error probabilities has to be based on large deviation theory. The main objective of this paper is to characterize the exponential rates of decrease for the error probabilities of the first and second kind and their interdependence. In the information theoretical literature, this characterization (for the i.i.d. case) is known as the Hoeffding bound (Hoeffding 1965; Csiszár and Longo 1971; Blahut 1974); Stein's lemma and the Chernoff bound then appear as special cases. For a concise presentation of these results in the context of large deviation theory and asymptotic statistics see Genon-Catalot and Picard (1993).

I. Tecuapetla-Gómez (✉)
Department of Statistical Science, Cornell University, Ithaca, NY 14853-3801, USA
e-mail: irt4@cornell.edu

M. Nussbaum
Department of Mathematics, Cornell University, Ithaca, NY 14853, USA
e-mail: nussbaum@math.cornell.edu

In order to treat the case of locally stationary Gaussian processes we mainly follow some work by Bouaziz (1993) and Taniguchi and Kakizawa (2000) who have obtained large deviations for stationary Gaussian sequences.

Locally stationary Gaussian processes were introduced and amply studied by Dahlhaus (1996a,b, 1997, 2009). We begin with a brief introduction to these processes and also to the hypothesis testing problem that we are concerned with in this paper.

Definition 1 A triangular array of random variables $X := (X_{k,n})_{1\le k\le n}$ is called a *locally stationary Gaussian process* with trend μ and transfer function A°, if it allows the following spectral representation

$$X_{k,n} = \mu\left(\frac{k}{n}\right) + \int_{-\pi}^{\pi} \exp(\boldsymbol{i}\lambda k)\, A^\circ_{k,n}(\lambda)\, d\epsilon(\lambda) \tag{1.2}$$

where

1. $\epsilon(\lambda) = \epsilon_1(\lambda) + \mathbf{i}\epsilon_2(\lambda)$ is a complex-valued Brownian motion on $[0,\pi]$ extended to $[-\pi,0]$ by $\epsilon(\lambda) = \overline{\epsilon(-\lambda)}$, $\lambda \in [-\pi,0]$, where $\epsilon_1(\lambda), \epsilon_2(\lambda)$ are independent standard Brownian motions on $\lambda \in [0,\pi]$.
2. There exist a constant K and a 2π-periodic function $A : [0,1]\times[-\pi,\pi] \to \mathbb{C}$ with $A(u,-\lambda) = \overline{A(u,\lambda)}$ and
$$\sup_{k,\lambda} \left|A^\circ_{k,n}(\lambda) - A(k/n,\lambda)\right| \le K\, n^{-1},$$
for all n.

Note that when $A^\circ_{k,n}(\cdot)$ does not depend on k, we obtain a spectral representation of stationary Gaussian processes. Definition 1 allows for more stochastic processes to be studied within the framework of oscillatory processes. This family of processes was introduced by Priestly (1981). Priestley provided a stochastic representation of a certain type of non-stationary processes, in addition, Dahlhaus' approach allows for an asymptotic treatment of statistical inference problems. A concise comparison between these two approaches can be found in Dahlhaus (1996b).

The function $f : [0,1]\times[-\pi,\pi] \to \mathbb{R}$ given by

$$f(u,\lambda) = \frac{1}{2\pi}\,|A(u,\lambda)|^2$$

is called the *time-varying spectral density* of the process.

Although f, as given above, may not be unique, if the function A is sufficiently smooth, then the function f is asymptotically uniquely determined by the whole triangular array $(X_{k,n})_{1\le k\le n}$. Indeed, if A is uniformly Lipschitz continuous in both components with index $\alpha > 1/2$ then for all $u \in (0,1)$, we have

$$\int_{-\pi}^{\pi} |f_n(u,\lambda) - f(u,\lambda)|^2\, d\lambda = o(1),$$

where $f_n(u,\lambda)$ is the Wigner–Ville spectrum for fixed n:

$$f_n(u,\lambda) := \frac{1}{2\pi}\sum_{s=-\infty}^{\infty} \operatorname{cov}\left(X_{[un+s/2],n}, X_{[un-s/2],n}\right)\exp(-\boldsymbol{i}\lambda s),$$

where $X_{s,n}$ obeys the representation (1.2) and $[x]$ is the largest integer less or equal than x. See Theorem 2.2 in Dahlhaus (1996a).

Along with the time-varying spectral density f we have the *time-varying covariance function* of lag k at time u

$$c(u,k) := \int_{-\pi}^{\pi} f(u,\lambda)\exp(\boldsymbol{i}\lambda k)\, d\lambda.$$

For two transfer functions A° and B°, set

$$\Sigma_n(A,B) := \left(\int_{-\pi}^{\pi} \exp\{\boldsymbol{i}\lambda(r-s)\} A^\circ_{r,n}(\lambda)\overline{B^\circ_{s,n}(\lambda)}\, d\lambda \right)_{r,s=1,\ldots,n}. \tag{1.3}$$

If the true process is locally stationary with transfer function A° and approximating function A, then $\Sigma_n = \Sigma_n(A,A)$ is the true covariance matrix of the process. Instead of working directly with the theoretical covariance matrix, we can appropriately approximate Σ_n by using the time-varying covariance function. For instance, if $A(u,\lambda)$ is uniformly differentiable in u then

$$\operatorname{cov}\left(X_{[un],n}, X_{[un]+k,n}\right) = c(u,k) + O(n^{-1}).$$

See Remark 2.8 in Dahlhaus (2009).

Throughout this paper $X := (X_{k,n})_{1\le k\le n}$ will denote observations from a locally stationary Gaussian process with zero trend function, transfer function A°, approximating function A and time-varying spectral density f. We study asymptotic statistical inference about X under the two simple hypotheses established in (1.1).

In accordance with the results above under H_0, the covariance matrix, Σ_n, is asymptotically defined through the time-varying spectral density f. Under H_1, the covariance matrix of X, denoted by $\widetilde{\Sigma}_n$, is asymptotically defined via another time-varying spectral density, say g.

Let T_n be a test for the hypotheses in (1.1). The performance of T_n is determined by the false-alarm rate (type I error probability or simply α_n) and the nondetection rate (type II error probability or just β_n) which by definition are

$$\alpha_n = \mathbb{P}\{\, T_n \text{ Rejects } H_0 \mid H_0\,\}, \quad \beta_n = \mathbb{P}\{\, T_n \text{ Rejects } H_1 \mid H_1\,\}. \tag{1.4}$$

We are interested in obtaining exponential rates of decrease of α_n and β_n for Neyman–Pearson and Bayes-type tests. In order to formulate our main results some notation is in order. The *torus* $\mathbb{T}$ will be the set $\mathbb{T} = [0,1] \times [-\pi,\pi]$. Let $B_{\mathbb{T}}$ denote the set of functions $h : \mathbb{T} \to \mathbb{C}$ such that:

1. There exists a constant $C > 0$ such that $|h(u,\lambda)| > C$, for all $(u,\lambda) \in \mathbb{T}$, and
2. the derivative $\frac{\partial}{\partial u}\frac{\partial}{\partial \lambda} h$ is uniformly bounded.

Proposition 1 *Suppose that $f, g \in B_{\mathbb{T}}$ and $f \neq g$ on $\mathbb{T}$. Let β_n^ϵ be the infimum of β_n among all tests with $\alpha_n < \epsilon$. Then for any $\epsilon < 1$*

$$\lim_{n\to\infty} \frac{1}{n} \log \beta_n^\epsilon = -K_L(f;g),$$

where

$$K_L(f;g) = \frac{1}{4\pi} \int_{\mathbb{T}} \left[\log \frac{g}{f} + \frac{f-g}{g} \right] d\lambda\, du.$$

Recall that $f := f(u, \lambda)$ *and* $g := g(u, \lambda)$.

In the scenario of having some a priori probability on H_0 we have the following result.

Proposition 2 *Suppose that the conditions of Proposition 1 hold. Let* $\mathbb{P}_n^B = \alpha_n \mathbb{P}\{H_0\} + \beta_n \mathbb{P}\{H_1\}$ *be the Bayes probability of error for testing the simple hypotheses* (1.1). *If* $0 < \mathbb{P}\{H_0\} < 1$ *then*

$$\inf_{S} \lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}_n^B = -\Psi^{\sharp}(0).$$

The infimum is taken over all tests for (1.1) *with error probabilities given by* (1.4). *Moreover,*

$$\Psi^{\sharp}(0) = -K_L(g_{\alpha_0}; f),$$

with $K_L(f; g)$ *given above and*

$$g_{\alpha_0} = \frac{fg}{\alpha_0 f + (1-\alpha_0)g}.$$

The value of α_0 *is determined in the proof of this result.*

Remark 1 Hirukawa and Taniguchi (2006) have studied non-Gaussian locally stationary processes, focusing on proving local asymptotic normality (LAN). We comment on possible relations to our results at Sect. 2.1.

This paper is organized as follows. In Sect. 2 we derive a large deviation result for the log-likelihood ratio for testing simple hypotheses for locally stationary Gaussian processes. Also in this section, we present a study about the rates of exponential decay of the type I and type II error probabilities given in (1.4), as a result we obtain a Hoeffding bound for testing the simple hypotheses (1.1). A proof of Proposition 1 is given in Sect. 3 along with an introduction that refers to the classical Stein's lemma for i.i.d. observations. In the same spirit, Sect. 4 presents the classical Chernoff bound and the corresponding proof of Proposition 2. Finally, some technical details used in our proofs are relegated to the Appendix.

2 Large deviations for locally stationary Gaussian processes

We begin this section with some definitions. Let $\mathbf{P}_n$ and $\mathbf{Q}_n$ be probability measures defined on some space $(\Omega_n, \mathcal{F}_n)$. For each n, suppose that $\mathbf{Q}_n$ is absolutely continuous w.r.t. $\mathbf{P}_n$ (in short notation $\mathbf{Q}_n \ll \mathbf{P}_n$). Let $\Lambda_n = \log \frac{d\mathbf{Q}_n}{d\mathbf{P}_n}$ be the log-likelihood ratio between $\mathbf{Q}_n$ and $\mathbf{P}_n$. Let $\boldsymbol{\mu}_n$ be a measure dominating $\mathbf{P}_n$ and $\mathbf{Q}_n$. Since $d\mathbf{P}_n = \mathbf{p}_n \, d\boldsymbol{\mu}_n$ and $d\mathbf{Q}_n = \mathbf{q}_n \, d\boldsymbol{\mu}_n$, where $\mathbf{p}_n$ and $\mathbf{q}_n$ are the Radon-Nykodim derivatives of $\mathbf{P}_n$ and $\mathbf{Q}_n$ w.r.t. $\boldsymbol{\mu}_n$, $\Lambda_n = \log \frac{\mathbf{q}_n}{\mathbf{p}_n}$.

Under $\mathbf{P}_n$, let $\Phi_n(\cdot)$ be the m.g.f. of Λ_n; $\Phi_n(\alpha) = \mathbb{E}_{\mathbf{P}_n}[\exp\{\alpha \Lambda_n\}]$. Note that for any $\alpha \in [0, 1]$

$$\Phi_n(\alpha) = \int \mathbf{p}_n^{1-\alpha}(x) \mathbf{q}_n^{\alpha}(x) \, d\boldsymbol{\mu}_n(x). \tag{2.1}$$

The RHS of (2.1) is known as the Hellinger transform between $\mathbf{P}_n$ and $\mathbf{Q}_n$. For mean zero Gaussian processes, i.e., $\mathbf{P}_n \sim \mathcal{N}_n(0, \Gamma_n)$, $\mathbf{Q}_n \sim \mathcal{N}_n(0, \widetilde{\Gamma}_n)$, Taniguchi and Kakizawa (2000) have introduced the terms *asymptotic Kullback–Leibler divergence* and *asymptotic Chernoff information with index* α, $0 < \alpha < 1$, as

$$K_L = \lim_{n\to\infty} \frac{1}{n} \mathbb{E}_{\mathbf{P}_n}\left[\log \frac{d\mathbf{P}_n}{d\mathbf{Q}_n}\right] \tag{2.2}$$

and

$$C_\alpha = \lim_{n\to\infty} -\frac{1}{n}\log \mathbb{E}_{\mathbf{P}_n}\left[\left\{\frac{d\mathbf{Q}_n}{d\mathbf{P}_n}\right\}^\alpha\right], \tag{2.3}$$

respectively. See Sect. 7.6 in Taniguchi and Kakizawa (2000).

Following these definitions we have that for any two equivalent Gaussian measures $\mathbf{P}_n$ and $\mathbf{Q}_n$

$$\lim_{n\to\infty}\frac{1}{n}\log\Phi_n(\alpha) = \lim_{n\to\infty}\frac{1}{n}\log\int \mathbf{p}_n^{1-\alpha}(x)\mathbf{q}_n^\alpha(x)\,dx = -C_\alpha. \tag{2.4}$$

That is, computing the asymptotic Chernoff divergence is equivalent to knowing the logarithmic limit of the Hellinger transform which in turn, and thanks to Bouaziz' lemma below, is equivalent to establishing a large deviation principle for the log-likelihood ratio of $\mathbf{Q}_n$ w.r.t $\mathbf{P}_n$. This is summarized by the Eq. (2.4).

Bouaziz (1993) showed the following general result on large deviations. Let S_n be a sequence of real-valued random variables defined on some probability space $(\Omega_n^*, \mathcal{F}_n^*, \mathbb{P}_n)$. Let ϕ_n be the moment generating function of S_n, i.e., $\phi_n(\alpha) = \mathbb{E}_{\mathbb{P}_n}[\exp\{\alpha S_n\}]$.

Bouaziz' Lemma (Lemma 2.1 in Bouaziz 1993) *Suppose that*

1. $\phi_n(1) < \infty$ *for all (sufficiently large) n,*
2. $(1/n)\log\phi_n(\alpha) \to \psi(\alpha)\ \forall\alpha \in [0,1]$,
3. ψ *is differentiable and strictly convex on* $(0,1)$.

Then for all $a \in \big(\psi'(0), \psi'(1)\big)$

$$\lim_{n\to\infty}\frac{1}{n}\log\mathbb{P}_n\{\,S_n > na\,\} = -\psi^\sharp(a),$$

where $\psi^\sharp$ *is the Fenchel-Legendre conjugate of* ψ *defined by*

$$\psi^\sharp(a) = \sup_{0\le\alpha\le 1}[a\alpha - \psi(\alpha)].$$

In addition, the same result holds for $\{\,S_n \ge na\,\}$.

We will apply this lemma to the log-likelihood Λ_n of two locally stationary Gaussian processes. More precisely, let $X = (X_{k,n})_{1\le k\le n}$ and $\widetilde{X} = (\widetilde{X}_{k,n})_{1\le k\le n}$ be realizations of locally stationary Gaussian processes with probability laws $\mathbf{P}_n$, $\mathbf{Q}_n$ and time-varying spectral densities f, g, respectively. From now on $\mathbf{P}_n \sim \mathcal{N}_n(0,\Sigma_n)$ and $\mathbf{Q}_n \sim \mathcal{N}_n(0,\widetilde{\Sigma}_n)$ where Σ_n and $\widetilde{\Sigma}_n$ are defined in (1.3). Let $\Lambda_n = \log\frac{d\mathbf{Q}_n}{d\mathbf{P}_n}$ be the log-likelihood ratio of $\mathbf{Q}_n$ w.r.t. $\mathbf{P}_n$ and $\Phi_n(\cdot)$ its moment generating function. From (2.1), it can be seen that $\Phi_n(1) = 1$ for all n. Let $\mathbb{T}$ and $B_\mathbb{T}$ be as in the introduction. We have the following

Lemma 1 *If* $f, g \in B_\mathbb{T}$ *then for any* $\alpha \in [0,1]$

$$\lim_{n\to\infty}\frac{1}{n}\log\Phi_n(\alpha) = \Psi(\alpha),$$

where

$$\Psi(\alpha) = \frac{1}{4\pi}\int_\mathbb{T}\left[\alpha\log\frac{f}{g} - \log\left(\frac{\alpha f + (1-\alpha)g}{g}\right)\right]d\lambda\,du, \tag{2.5}$$

where $f := f(u,\lambda)$ *and* $g := g(u,\lambda)$.

Proof Since $\Phi_n(1) = \Phi_n(0) = 1$, the result is trivial for $\alpha \in \{0, 1\}$. For $\alpha \in (0, 1)$ we show in the Appendix that for any $n \in \mathbb{N}$ we have that

$$\frac{1}{n}\log \Phi_n(\alpha) = \frac{\alpha}{2n}\left(\log\det \Sigma_n - \log\det \widetilde{\Sigma}_n\right) \\ -\frac{1}{2n}\left(\log\det\left[\alpha\Sigma_n + (1-\alpha)\widetilde{\Sigma}_n\right] - \log\det \widetilde{\Sigma}_n\right).$$

Since $f, g \in B_{\mathbb{T}}$ it follows that for any $\alpha \in (0, 1)$, $\alpha f + (1-\alpha)g \in B_{\mathbb{T}}$. An application of Theorem 3.2(ii) in Dahlhaus (1996a) yields

$$\frac{1}{n}\log\det \Sigma_n \to \frac{1}{2\pi}\int \log 2\pi f \, d\lambda\, du,$$
$$\frac{1}{n}\log\det \widetilde{\Sigma}_n \to \frac{1}{2\pi}\int \log 2\pi g \, d\lambda\, du,$$
$$\frac{1}{n}\log\det\left[\alpha\Sigma_n + (1-\alpha)\widetilde{\Sigma}_n\right] \to \frac{1}{2\pi}\int \log 2\pi(\alpha f + (1-\alpha)g)\, d\lambda\, du.$$

The result now follows by summing up all the terms above.

According to (2.3) and (2.5), the asymptotic Chernoff information between two locally stationary Gaussian processes is given by the integral

$$C_\alpha = \frac{1}{4\pi}\int\limits_{\mathbb{T}}\left[\log\left(\frac{\alpha f + (1-\alpha)g}{g}\right) - \alpha\,\log\frac{f}{g}\right] d\lambda\, du.$$

In the Appendix we show that (2.2) is given by the quantity

$$K_L = \frac{1}{4\pi}\int\limits_{\mathbb{T}}\left(\log\frac{g}{f} + \frac{f-g}{g}\right) d\lambda\, du.$$

In accordance with the terminology introduced by Taniguchi and Kakizawa, this expression is called the asymptotic Kullback–Leibler divergence for locally stationary Gaussian processes.

Observe that when $f(u, \lambda)$ and $g(u, \lambda)$ do not depend on u the two expressions above generalize well-known formulas for the asymptotic Chernoff information between stationary Gaussian sequences (Coursol and Dacunha-Castelle 1979) and for the asymptotic Kullback–Leibler divergence between stationary Gaussian sequences (obtained by Taniguchi 2001 as the approximate slope of the primitive likelihood ratio test between stationary Gaussian processes with spectral density $f(\lambda)$).

Let f, g and Ψ be as in Lemma 1. Since the function $h(x) = -\log(1+\theta x)$ is strictly convex $\forall\theta \neq 0$ it follows that $\Psi(\cdot)$ is strictly convex whenever $f \neq g$. In addition, the function $\Psi(\cdot)$ is differentiable and $\Psi'(0, 1) = (\Psi'(0), \Psi'(1))$ where

$$\Psi'(0) = \frac{1}{4\pi}\int\limits_{\mathbb{T}}\left[\log\frac{f}{g} - \frac{f-g}{g}\right] d\lambda\, du \tag{2.6}$$

$$\Psi'(1) = \frac{1}{4\pi}\int\limits_{\mathbb{T}}\left[\log\frac{f}{g} - \frac{f-g}{f}\right] d\lambda\, du. \tag{2.7}$$

Since $\log(x) < x-1$, $\forall x > 0$, $x \neq 1$ it follows that $\Psi'(0) < 0 < \Psi'(1)$ whenever $f \neq g$. So, we have found conditions under which the m.g.f., $\Phi_n(\cdot)$, of the log-likelihood ratio between two locally stationary Gaussian processes satisfies the assumptions of Bouaziz' lemma. We have the following

Theorem 1 *Let $\mathbf{P}_n$ and $\mathbf{Q}_n$ be probability measures corresponding to two locally stationary Gaussian processes with time-varying spectral densities f and g, respectively. Let $f, g \in B_{\mathbb{T}}$ and suppose that $f \neq g$ on $\mathbb{T}$. Let Λ_n be the log-likelihood ratio of $\mathbf{Q}_n$ w.r.t. $\mathbf{P}_n$. Then for all $a \in \big(\Psi'(0), \Psi'(1)\big)$*

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbf{P}_n\{\, \Lambda_n > na \,\} = -\Psi^{\sharp}(a),$$

where

$$\Psi^{\sharp}(a) = \sup_{0\le\alpha\le 1} [a\alpha - \Psi(\alpha)],$$

with $\Psi(\alpha)$ given as in (2.5). The same is true for $\{\, \Lambda_n \ge na \,\}$.

This result is a large deviation for the log-likelihood ratio between the laws of two locally stationary Gaussian processes. Large deviations for quadratic forms of locally stationary processes have been obtained by Zani (2002). We now use Theorem (1) to establish exponential rates of decrease for the false-alarm and nondetection rates for testing the hypotheses considered in (1.1). Following Bouaziz let us recall that given a false-alarm level α_n in (0, 1), the most powerful tests of $\mathbf{P}_n$ against $\mathbf{Q}_n$ at the level α_n are the Neyman–Pearson tests of size α_n, i.e., those tests τ_n which satisfy

$$\mathbb{1}_{\{\Lambda_n > a_n\}} \le \tau_n \le \mathbb{1}_{\{\Lambda_n \ge a_n\}}, \quad \mathbb{E}_{\mathbf{P}_n}[\tau_n] = \alpha_n, \quad \mathbb{E}_{\mathbf{Q}_n}[\tau_n] = 1 - \beta_n.$$

For every $n \in \mathbb{N}$ we have the following inequalities:

$$\frac{1}{n} \log \mathbf{P}_n\{\, \Lambda_n > a_n \,\} \le \frac{1}{n} \log \alpha_n \le \frac{1}{n} \log \mathbf{P}_n\{\, \Lambda_n \ge a_n \,\}.$$

Under the conditions of Theorem 1 and assuming that $a_n \sim na$, it follows that for all $a \in (\Psi'(0), \Psi'(1))$, $\frac{1}{n} \log \alpha_n(a) \to -\Psi^{\sharp}(a)$. Let $\widehat{\Phi}_n(\alpha) = \mathbb{E}_{\mathbf{Q}_n}[\exp\{\alpha\widetilde{\Lambda}_n\}]$ be the m.g.f. of $\widetilde{\Lambda}_n = \log \frac{d\mathbf{P}_n}{d\mathbf{Q}_n}$ under $\mathbf{Q}_n$. Observe that $\widetilde{\Lambda}_n = -\Lambda_n$. A straightforward calculation allows us to see that $\widehat{\Phi}_n(\alpha) = \Phi_n(1-\alpha)$, for all $\alpha \in [0, 1]$. Consequently,

$$\lim_{n\to\infty} \frac{1}{n} \log \widehat{\Phi}_n(\alpha) \to \Psi(1-\alpha).$$

The Fenchel-Legendre conjugate of this limiting function is given by

$$\widehat{\Psi}^{\sharp}(a) = \sup_{\alpha\in[0,1]} [a\alpha - \Psi(1-\alpha)] = a + \Psi^{\sharp}(-a).$$

Thus, under conditions of Theorem 1 we have that

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbf{Q}_n\{\, \widetilde{\Lambda}_n > na \,\} = -a - \Psi^{\sharp}(-a),$$

and the same holds for $\{\widetilde{\Lambda}_n \ge na\}$.

Since $\{\, \Lambda_n < nb \,\} = \{\, \widetilde{\Lambda}_n > -nb \,\}$ it follows that

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbf{Q}_n\{\, \Lambda_n < nb \,\} = b - \Psi^{\sharp}(b).$$

Thus, for Neyman–Pearson tests and for every $n \in \mathbb{N}$ we have the following inequalities:

$$\frac{1}{n} \log \mathbf{Q}_n\{\, \Lambda_n < a_n \,\} \le \frac{1}{n} \log \beta_n \le \frac{1}{n} \log \mathbf{Q}_n\{\, \Lambda_n \le a_n \,\}.$$

Hence for locally stationary Gaussian processes satisfying the conditions of Theorem 1 we conclude that for all $a \in (\Psi'(0), \Psi'(1))$,

$$\lim_{n\to\infty} \frac{1}{n} \log \beta_n(a) = a - \Psi^\sharp(a),$$

where $\Psi^\sharp(\cdot)$ is the Fenchel-Legendre conjugate of $\Psi(\cdot)$ given in (2.5). Summarizing, we have the following

Corollary 1 *Consider the assumptions of Theorem 1. Let* $\Psi'(0)$ *and* $\Psi'(1)$ *be given as in* (2.6) *and* (2.7)*, respectively. Then for all* $a \in (\Psi'(0), \Psi'(1))$

$$\lim_{n\to\infty} \frac{1}{n} \log \alpha_n(a) = -\Psi^\sharp(a) < 0, \tag{2.8}$$

$$\lim_{n\to\infty} \frac{1}{n} \log \beta_n(a) = a - \Psi^\sharp(a) < 0, \tag{2.9}$$

where

$$\Psi^\sharp(a) = \frac{1}{4\pi} \int_{\mathbb{T}} \left[\log \left(\frac{\alpha_a f + (1-\alpha_a) g}{g} \right) + \frac{\alpha_a (g-f)}{\alpha_a f + (1-\alpha_a) g} \right] d\lambda\, du,$$

with α_a *the unique solution to* $\Psi'(\alpha_a) = a$.

Proof It remains to establish the expression for $\Psi^\sharp(a)$. Define $G(a,\alpha) = a\alpha - \Psi(\alpha)$. Observe that $\frac{\partial}{\partial\alpha} G(a,\alpha) = a - \Psi'(\alpha)$, $\frac{\partial^2}{\partial^2\alpha} G(a,\alpha) = -\Psi''(\alpha)$. Since for all $\alpha \in [0,1]$, $\Psi''(\alpha) = (4\pi)^{-1} \int_{\mathbb{T}} (f-g)^2/[\alpha f + (1-\alpha) g]^2\, d\lambda\, du > 0$, the function $\frac{\partial}{\partial\alpha} G(a,\alpha)$ is strictly decreasing on $\big(\Psi'(0), \Psi'(1)\big)$. Then, the equation $a = \Psi'(\alpha)$ has a unique solution α_a, i.e., there exists a unique $\alpha_a \in [0,1]$ such that $a = \Psi'(\alpha_a)$. It follows that $\Psi^\sharp(a) = G(a, \alpha_a)$. A straightforward calculation yields

$$\begin{aligned} G(a,\alpha_a) &= \frac{1}{4\pi} \int_{\mathbb{T}} \left[\log \left(\frac{\alpha_a f + (1-\alpha_a) g}{g} \right) + \frac{\alpha_a (g-f)}{\alpha_a f + (1-\alpha_a) g} \right] d\lambda\, du \\ &= K_L(g_{\alpha_a}, f), \end{aligned} \tag{2.10}$$

where $K_L(f, g)$ is given in Proposition 1 and

$$g_{\alpha_a} = \frac{fg}{\alpha_a f + (1-\alpha_a) g}.$$

Recall that $f := f(u,\lambda)$ and $g := g(\lambda, u)$.

Results of this type can be found in the literature on large deviations for stochastic processes. For instance, Genon-Catalot and Picard (1993) (Theorem 3.4.3 §4.2) have obtained Corollary 1 for the case of a sequence of n i.i.d. random variables, Theorem 2.4 in Bouaziz (1993) and Lemma 8.2.3 in Taniguchi and Kakizawa (2000) treat the same problem for stationary Gaussian sequences. In the information theoretic literature this type of result is referred to as the Hoeffding-Blahut-Csiszár-Longo bound. Recently, Audenaert et al. (2008) have obtained a Hoeffding bound in quantum hypothesis testing, and Gapeev and Küchler (2008) have provided similar large deviation results for testing Ornstein-Uhlenbeck-type models.

An interpretation of Corollary 1 is that for every $a \in (\Psi'(0), \Psi'(1))$ there exists a test τ_a whose type I and type II error probabilities, $\alpha_n(a)$ and $\beta_n(a)$, satisfy (2.8) and (2.9), respectively. Moreover, the ratio $\frac{\alpha_n(a)}{\beta_n(a)}$ behaves roughly like $\mathbf{e}^{-na}$, and we have the following cases:

1. For $a > 0$, $\frac{\alpha_n(a)}{\beta_n(a)} \to 0$ as $n \to \infty$. That is, $\alpha_n(a) \to 0$ and $\beta_n(a) \to c_0$.
2. For $a = 0$, $\frac{\alpha_n(a)}{\beta_n(a)} \to 1$ as $n \to \infty$. That is, for n large enough $\alpha_n(a)$ and $\beta_n(a)$ share the same growth rate.
3. For $a < 0$, $\frac{\alpha_n(a)}{\beta_n(a)} \to \infty$ as $n \to \infty$. That is, $\alpha_n(a) \to c_1$ and $\beta_n(a) \to 0$.

In the next sections we study these three cases in more detail. For instance, the rate with which $\beta_n \to 0$, case $a < 0$, is given by Stein's Lemma and the rate in which both type I error and type II error grow equally, case $a = 0$, appears as the Chernoff bound for testing simple hypotheses.

2.1 A comment on large deviations for non-Gaussian locally stationary processes

To our knowledge, Hirukawa and Taniguchi (2006) have presented the only systematic study on non-Gaussian locally stationary processes. We now present an outline of a possible extension of our large deviation results to the non-Gaussian processes introduced in Hirukawa and Taniguchi (2006).

Let us begin by saying that a non-Gaussian locally stationary process satisfies the representation given in (1.2):

$$X_{k,n} = \int_{-\pi}^{\pi} \exp(\boldsymbol{i}\lambda k)\, A^{\circ}_{k,n}(\lambda)\, d\epsilon(\lambda).$$

As in Definition 1.1, ϵ is a complex-valued stochastic process defined on $[-\pi, \pi]$ but it is not necessarily a Brownian motion. Instead, it is supposed that

$$\text{cum}\,\{\, d\epsilon(\lambda_1), \ldots, d\epsilon(\lambda_k)\,\} = \eta\left(\sum_{i=1}^{k} \lambda_i\right) h_k(\lambda_1, \ldots, \lambda_{k-1})\, d\lambda_1 \cdots d\lambda_k,$$

where cum stands for the cumulant of k-th order, η is a 2π-periodic extension of the Dirac delta function and $|h_k(\lambda_1, \ldots, \lambda_{k-1})| \le C_k$. As in Definition 1.2 the trend function $A^{\circ}_{k,n}(\cdot)$ has a uniformly approximating function $A(k/n, \cdot)$ which is assumed to be uniformly bounded from above and bounded away from zero as well as continuous in the first component.

Since Hirukawa and Taniguchi (2006) were focused on locally asymptotic normality (LAN), it was natural to assume that the process had a parametric structure. Namely, realizations $(X_{k,n})_{1\le k\le n}$ of a non-Gaussian locally stationary process with trend function A°_{θ}, approximating function A_θ and time-varying function $f_\theta(u, \lambda) := |A_\theta(u, \lambda)|^2$ where $\theta = (\theta_1, \ldots, \theta_q) \in \Theta \in \mathbb{R}^q$ have been considered.

In Hirukawa and Taniguchi (2006) the following assumptions were made:

A1. $A_\theta(u, \lambda)$, the gradient $\nabla A_\theta(u, \lambda)$, and the Hessian $\nabla^2 A_\theta(u, \lambda)$ have components which are differentiable in u and λ with uniformly continuous derivative $\frac{\partial}{\partial u}\frac{\partial}{\partial \lambda}$. Also, it is assumed that the gradient and Hessian of $A^{\circ}_{\theta,k,n}(\lambda) - A_\theta(k/n, \lambda)$ are uniformly bounded in k and λ.

A2. Write

$$\xi_k = \int_{-\pi}^{\pi} \exp(\boldsymbol{i}\lambda k)\, d\epsilon(\lambda),$$

and assume that the ξ_k's are i.i.d. random variables with $\mathbb{E}[\xi_k] = 0$, $\mathbb{E}[\xi_k^2] = 1$ and $\mathbb{E}[\xi_k^4] < \infty$. Furthermore, assume that the distribution of ξ_k is absolutely continuous w.r.t. Lebesgue measure and that the corresponding probability density p satisfies regularity conditions such as being light-tailed and having smooth first two derivatives p' and p''.

A3. Assume that $\{X_{k,n}\}$ has the MA(∞) and AR(∞) representations

$$X_{k,n} = \sum_{j=0}^{\infty} a^{\circ}_{\theta,k,n}(j)\, \xi_{k-j},$$

$$a^{\circ}_{\theta,k,n}(0)\, \xi_k = \sum_{t=0}^{\infty} b^{\circ}_{\theta,k,n}(t)\, X_{k-t,n}.$$

It is also assumed that the non-random coefficients $a^{\circ}_{\theta,k,n}(\cdot)$ and $b^{\circ}_{\theta,k,n}(\cdot)$ satisfy some smoothness and summability conditions.

Let $\mathbf{P}_{\theta,n}$ be the probability distribution of $\left(\xi_s,\, s \le 0,\, X_{1,n}, \ldots, X_{n,n}\right)$. Then under assumptions A1, A2 and A3, the log-likelihood ratio between two non-Gaussian locally stationary processes defined via the points $\theta, \theta_n \in \Theta$ is given by

$$\Lambda_n(\theta, \theta_n) = \log \frac{d\mathbf{P}_{\theta_n,n}}{d\mathbf{P}_{\theta,n}} = 2 \sum_{k=1}^{n} \log \Phi_{k,n}(\theta, \theta_n),$$

where

$$\Phi^2_{k,n}(\theta, \theta_n) = \frac{g_{\theta_n,k,n}(z_{\theta,k,n} + q_{k,n})}{g_{\theta,k,n}(z_{\theta,k,n})}$$

with

$$g_{\theta,k,n}(\cdot) = \frac{1}{a^{\circ}_{\theta,k,n}(0)}\, p\left(\frac{\cdot}{a^{\circ}_{\theta,k,n}(0)}\right), \qquad z_{\theta,k,n} = a^{\circ}_{\theta,k,n}(0)\, \xi_k$$

$$a^{\circ}_{\theta,k,n}(0) \equiv \exp\left\{\frac{1}{4\pi} \int_{-\pi}^{\pi} \log f_{\theta,k,n}(\lambda)\, d\lambda\right\}, \qquad f_{\theta,k,n}(\lambda) = \left|A^{\circ}_{\theta,k,n}(\lambda)\right|^2$$

$$q_{k,n} = \sum_{t=1}^{k-1} \left[\frac{1}{\sqrt{n}} h^\top \nabla b^{\circ}_{\theta,k,n}(t) + \frac{1}{2n}\, h^\top\, \nabla^2 b^{\circ}_{\theta^*,k,n}(t)\, h\right] X_{k-t,n} + \sum_{r=0}^{\infty} \frac{1}{\sqrt{n}} h^\top\, \nabla c^{\circ}_{\theta^{**},k,n}(r)\, \xi_{-r}$$

$$c^{\circ}_{\theta,k,n}(r) = \sum_{s=0}^{r} b^{\circ}_{\theta,k,n}(k+s)\, a^{\circ}_{\theta,0,n}(r-s).$$

The points θ^* and θ^{**} belong to the segment defined by θ and $\theta + \frac{h}{\sqrt{n}}$, for some $h \in \mathbb{R}^q$.

As shown in the present paper, to establish a large deviation for the log-likelihood ratio between non-Gaussian locally stationary processes we could start by studying the behavior of

$$\frac{1}{n} \log \mathbb{E}_{\mathbf{P}_{\theta,n}} \left[\exp\{s \Lambda_n(\theta, \theta_n)\}\right]$$

as $n \to \infty$ with $\Lambda_n(\theta, \theta_n)$ as given above. This task is beyond the scope of the present article; however, as it will be illustrated in the next sections such a large deviation result might prove useful to determine the asymptotic efficiency of Neyman–Pearson and Bayes tests even in a non-Gaussian setting.

3 Proof of Proposition 1

In the i.i.d. case, Stein's lemma for testing simple hypotheses can be stated as follows. Let $X = (X_1, \ldots, X_n)$ be a sequence of i.i.d. random variables. Consider the following hypotheses

$$H_0 : X \sim p^n \quad \text{vs} \quad H_1 : X \sim q^n. \tag{3.1}$$

The Neyman–Pearson lemma dictates to use the log-likelihood ratio test between q^n and p^n to minimize the type II error probability, β_n, of this test. Stein's lemma tells us that, when the type I error probability, α_n, of this test is bounded then β_n will converge to zero at an exponential rate given by the Kullback–Leibler divergence between p and q. More precisely, and following Dembo and Zeitouni (1998), we have

Stein's Lemma *Let β_n^ϵ be the infimum of β_n among all tests whose false-alarm rate $\alpha_n < \epsilon$. Then for any $\epsilon < 1$*

$$\lim_{n\to\infty} \frac{1}{n} \log \beta_n^\epsilon = -K_L(p; q),$$

where $K_L(p; q)$ is the Kullback–Leibler information divergence between p and q.

Many authors have provided proofs of this lemma. For instance, Lemma 6.1 in Bahadur (1971), Section B, Chapter 6 in Bucklew (1990), Lemma 3.4.6 in Dembo and Zeitouni (1998) or the original statement, Theorem 2, in Chernoff (1956) all present proofs of the so-called Stein's lemma. Roughly speaking, in an i.i.d. framework when testing between two probability measures, p and q, if the false-alarm detection rate of the test statistic remains bounded when $n \to \infty$ then the nondetection rate, β_n, will decrease to zero exponentially fast.

We have stated an analogue of Stein's lemma for locally stationary Gaussian processes in the introduction of this paper. In this section we present a proof of this lemma which is based on Corollary 1. From now on we assume that $a_n \sim na$, that $\mathbf{Q}_n$ is absolutely continuous w.r.t. $\mathbf{P}_n$ and that $\Psi'(0) = -K_L(f, g)$ where $\Psi'(0)$ is given in (2.6).

We now give a proof of Proposition 1.

Proof Let $\epsilon > 0$ be given. Let β_n^ϵ be the infimum of β_n among all tests with $\alpha_n < \epsilon$. It is enough to consider Neyman–Pearson tests, i.e., those tests τ_n such that

$$\mathbb{1}_{\{\Lambda_n > a_n\}} \le \tau_n \le \mathbb{1}_{\{\Lambda_n \ge a_n\}}, \quad \mathbb{E}_{\mathbf{P}_n}[\tau_n] = \alpha_n, \quad \mathbb{E}_{\mathbf{Q}_n}[\tau_n] = 1 - \beta_n. \tag{3.2}$$

We get that

$$\frac{1}{n} \log \mathbf{Q}_n\{\Lambda_n < a_n\} \le \frac{1}{n} \log \beta_n \le \frac{1}{n} \log \mathbf{Q}_n\{\Lambda_n \le a_n\}. \tag{3.3}$$

Since $\mathbf{Q}_n \ll \mathbf{P}_n$ we have $\int \mathbb{1}_{\{\Lambda_n \le a_n\}} \mathbf{e}^{\Lambda_n} d\mathbf{P}_n = \int \mathbb{1}_{\{\Lambda_n \le a_n\}} d\mathbf{Q}_n$. This implies that

$$\frac{1}{n} \log \mathbf{Q}_n\{\Lambda_n \le a_n\} \le a. \tag{3.4}$$

For any $a \in (\Psi'(0), \Psi'(1))$ there exists a test, τ_a say, such that its error probabilities of first type, $\alpha_n(a)$, and second type, $\beta_n(a)$, satisfy (2.8) and (2.9). For n large enough,

$\alpha_n(a) < \epsilon$. Hence $\frac{1}{n}\log\beta_n^\epsilon < \frac{1}{n}\log\beta_n(a)$. Now, use the RHS of (3.3) and (3.4) to deduce that $\frac{1}{n}\log\beta_n^\epsilon \le \Psi'(0)$. Therefore $\forall\delta > 0$ we get that

$$\limsup_n \frac{1}{n}\log\beta_n^\epsilon \le \Psi'(0) + \delta.$$

To continue with the proof we need the following law of large numbers for Λ_n. For every $\theta \in (0, 1)$, define $d\mathbf{P}_{n,\theta} = [\mathbf{e}^{\theta\Lambda_n}/\Psi_n(\theta)]\,d\mathbf{P}_n$, $\Psi_n = \frac{1}{n}\log\Phi_n$. Then

$$\frac{1}{n}\Lambda_n \to \Psi'(\theta) \text{ in } \mathbf{P}_{n,\theta} \text{ probability.} \tag{3.5}$$

The argument to show that (3.5) holds is contained in the proof of Lemma 2.1 in Bouaziz (1993).

Choose $a \in \big(\Psi'(0), \Psi'(1)\big)$ whose corresponding test, τ_a, is such that for n large enough $\alpha_n(a) = \mathbf{P}_n\{\Lambda_n > a_n\} \le \epsilon$. The law of large numbers for Λ_n just presented implies that for all $\delta > 0$

$$\liminf_n \mathbf{P}_n\{\,\Lambda_n \in [n(\Psi'(0) - \delta), a_n]\,\} \ge 1 - \epsilon. \tag{3.6}$$

Use the LHS of the inequality in (3.2) to get that

$$\begin{aligned}\beta_n &\ge \int \mathbb{1}_{\{\,\Lambda_n < a_n\,\}}\,d\mathbf{Q}_n = \int \mathbb{1}_{\{\,\Lambda_n < a_n\,\}}\,\mathbf{e}^{\Lambda_n}\,d\mathbf{P}_n \\ &\ge \int \mathbb{1}_{\{\,\Lambda_n \in [n(\Psi'(0)-\delta), a_n]\,\}}\,\mathbf{e}^{\Lambda_n}\,d\mathbf{P}_n \\ &\ge \mathbf{e}^{n(\Psi'(0)-\delta)} \int \mathbb{1}_{\{\,\Lambda_n \in [n(\Psi'(0)-\delta), a_n]\,\}}\,d\mathbf{P}_n.\end{aligned}$$

Hence for all $\delta > 0$

$$\frac{1}{n}\log\beta_n(a) \ge \Psi'(0) - \delta + \frac{1}{n}\log\mathbf{P}_n\{\,\Lambda_n \in [n(\Psi'(0) - \delta), a_n]\,\}.$$

Finally, use (3.6) and recall that by assumption $\epsilon < 1$ to deduce that for all $\delta > 0$

$$\liminf_n \frac{1}{n}\log\beta_n^\epsilon \ge \Psi'(0) - \delta.$$

This completes the proof.

We have extended Stein's Lemma from the classical i.i.d. setup to nonstationary processes. In the context of model selection, Dahlhaus (1996a) found $\Psi'(0)$ and termed it as the Kullback–Leibler divergence between two locally stationary processes wih time-varying spectral densities f and g. According to the conclusion of Proposition 1 we can now adopt the notation $K_L(f, g)$ and refer to it as the asymptotic Kullback–Leibler divergence between two locally stationary Gaussian processes.

4 Proof of Proposition 2

Consider the setting given in (3.1). Because of the Neyman–Pearson lemma, to study the asymptotic efficiency of any test in (3.1) it is sufficient to have some a priori probability on H_0 and to consider the two types of error probabilities given by

$$\alpha_n = \mathbb{P}\{S_n \geq a_n \mid H_0\}, \quad \beta_n = \mathbb{P}\{S_n < a_n \mid H_1\},$$

where $S_n = \sum_{k=1}^n X_k$. A principle to find the threshold a_n is that of minimizing $\alpha_n + \lambda\,\beta_n$ for some $\lambda > 0$. Basically, this criterion is equivalent to requiring that as $n \to \infty$, α_n and β_n approach zero at the same rate. A solution to this problem was provided by Chernoff (1952). Namely, we have

Chernoff Bound

$$\lim_{n\to\infty} \left[\inf_{a_n} (\beta_n + \lambda\,\alpha_n)\right]^{1/n} = \inf_{0<t<1} \int p(x)^{1-t}\, q(x)^t\, dx.$$

Dembo and Zeitouni (1998) consider the same problem. Their solution is mainly based on elements of large deviations for the log-likelihood ratio test. Taniguchi and Kakizawa (2000) have obtained the Chernoff bound for testing simple hypotheses for stationary processes. Their solution is also mainly based on large deviations. The content of Proposition 2 is an extension of these results to the framework of locally stationary Gaussian processes. We now present a proof of such proposition.

Proof Again it is sufficient to consider Neyman–Pearson tests, τ_n. Let $\alpha_n(0)$ and $\beta_n(0)$ be the error probabilities of the Neyman–Pearson test with zero threshold. For any other Neyman–Pearson test, τ_n, either $\alpha_n \geq \alpha_n(0)$ $(a_n \leq 0)$ or $\beta_n(0) \leq \beta$ $(a_n \geq 0)$. Thus the inequalities

$$\begin{aligned}\min\{\alpha_n(0), \beta_n(0)\}\, \min\{\mathbb{P}\{H_0\}, \mathbb{P}\{H_1\}\} &\leq \mathbb{P}_n^B \\ &\leq 2\max\{\alpha_n(0), \beta_n(0)\}\max\{\mathbb{P}\{H_0\}, \mathbb{P}\{H_1\}\}\end{aligned}$$

yield

$$\inf_S \liminf_{n\to\infty} \frac{1}{n}\log \mathbb{P}_n^B \geq \liminf_{n\to\infty} \min\{\frac{1}{n}\log\alpha_n(0), \frac{1}{n}\log\beta_n(0)\}$$

and

$$\inf_S \limsup_{n\to\infty} \frac{1}{n}\log \mathbb{P}_n^B \leq \limsup_{n\to\infty} \max\{\frac{1}{n}\log\alpha_n(0), \frac{1}{n}\log\beta_n(0)\}.$$

Recall that $0 < \mathbb{P}\{H_0\} < 1$ and S denotes the set of all tests for (1.1) with probability errors given by (1.4). Given (2.8) and (2.9)

$$\lim_{n\to\infty} \frac{1}{n}\log\alpha_n(0) = \lim_{n\to\infty} \frac{1}{n}\log\beta_n(0) = -\Psi^\sharp(0).$$

Observe that by definition

$$-\Psi^\sharp(0) = -\sup_{\alpha\in[0,1]} [-\Psi(\alpha)] = \inf_{\alpha\in[0,1]} \Psi(\alpha) = f(0, t_0)$$

where $f(0, t_0)$ is given by (2.10). This completes the proof.

Acknowledgements The authors are grateful to a referee for his/her valuable comments and suggestions. This research was partially supported by CONACyT grant 230788 and NSF grant DMS-08-05632.

Appendix

In this section we use the notation introduced in Sect. 2 and the assumptions of Lemma 1. Consider the log-likelihood of $\mathbf{Q}_n$ w.r.t. $\mathbf{P}_n$, $\Lambda_n = \log \frac{d\mathbf{Q}_n}{d\mathbf{P}_n}$, and its moment generating function with respect to $\mathbf{P}_n$, i.e., the function $\Phi_n : [0,1] \to [0,1]$ given by

$$\Phi_n(\alpha) = \mathbb{E}_{\mathbf{P}_n}\left[\exp\{\alpha\Lambda_n\}\right] = \mathbb{E}_{\mathbf{P}_n}\left\{\frac{d\mathbf{Q}_n}{d\mathbf{P}_n}\right\}^{\alpha} = \int \mathbf{p}_n^{1-\alpha}(x)\,\mathbf{q}_n^{\alpha}(x)\,dx.$$

Let us calculate this integral:

$$\begin{aligned}\Phi_n(\alpha) &= \int \det(2\pi\Sigma_n)^{-(1-\alpha)/2}\mathbf{e}^{-\frac{(1-\alpha)}{2}x^\top\Sigma_n^{-1}x}\det(2\pi\widetilde{\Sigma}_n)^{-(\alpha)/2}\mathbf{e}^{-\frac{\alpha}{2}x^\top\widetilde{\Sigma}_n^{-1}x}\,dx\\ &= \det(\Sigma_n)^{-(1-\alpha)/2}\det(\widetilde{\Sigma}_n)^{-\alpha/2}\det[(1-\alpha)\Sigma_n^{-1}+\alpha\widetilde{\Sigma}_n^{-1}]^{-1/2}\\ &= \det(\Sigma_n)^{\alpha/2}\det(\widetilde{\Sigma}_n)^{(1-\alpha)/2}\det[\alpha\Sigma_n+(1-\alpha)\widetilde{\Sigma}_n]^{-1/2}.\end{aligned}$$

The latter follows after multiplying the integrand by a proper constant, using the fact that every probability density function integrates to 1 and using properties of the determinant.

Therefore

$$\log\Phi_n(\alpha) = \frac{1}{2}\left(\alpha\log\frac{\det\Sigma_n}{\det\widetilde{\Sigma}_n} - \log\frac{\det[\alpha\Sigma_n+(1-\alpha)\widetilde{\Sigma}_n]}{\det\widetilde{\Sigma}_n}\right).$$

Similar calculations show that the m.g.f. of the log-likelihood ratio of $\mathbf{P}_n$ w.r.t. $\mathbf{Q}_n$ is equal to

$$\log\frac{d\mathbf{P}_n}{d\mathbf{Q}_n}(x) = -\frac{1}{2}\log\frac{\det\Sigma_n}{\det\widetilde{\Sigma}_n} - \frac{1}{2}x^\top\left(\Sigma_n^{-1}-\widetilde{\Sigma}_n^{-1}\right)x. \tag{4.1}$$

Recall that if $Z \sim \mathcal{N}(0,S)$ then $\mathbb{E}[Z^\top AZ] = \mathrm{tr}(AS)$. Hence

$$\mathbb{E}_{\mathbf{P}_n}\left[\log\frac{\mathbf{P}_n}{\mathbf{Q}_n}\right] = -\frac{1}{2}\left[\log\frac{\det\Sigma_n}{\det\widetilde{\Sigma}_n} + n - \mathrm{tr}\left(\widetilde{\Sigma}_n^{-1}\Sigma_n\right)\right].$$

Lemma 4.8 in Dahlhaus (1996a) now yields that

$$\frac{1}{n}\mathrm{tr}\left(\widetilde{\Sigma}_n^{-1}(\widetilde{A},\widetilde{A})\Sigma_n(A,A)\right) \to \frac{1}{2\pi}\int_{\mathbb{T}}\frac{f}{g}\,d\lambda\,du.$$

Hence the asymptotic Kullback–Leibler information of two locally stationary Gaussian processes (as termed by Taniguchi and Kakizawa) with time-varying spectral densities f and g, respectively, is

$$\begin{aligned}K_L &= \lim_{n\to\infty}\frac{1}{n}\mathbb{E}_{\mathbf{P}_n}\left[\log\frac{d\mathbf{P}_n}{d\mathbf{Q}_n}\right]\\ &= \left(-\frac{1}{2}\right)\lim_{n\to\infty}\frac{1}{n}\left[\log\frac{\det\Sigma_n}{\det\widetilde{\Sigma}_n} + n - \mathrm{tr}\left(\widetilde{\Sigma}_n^{-1}\Sigma_n\right)\right]\\ &= \left(-\frac{1}{2}\right)\frac{1}{2\pi}\int_{\mathbb{T}}\left(\log\frac{f}{g}+1-\frac{f}{g}\right)d\lambda\,du\\ &= \frac{1}{4\pi}\int_{\mathbb{T}}\left(\log\frac{g}{f}+\frac{f-g}{g}\right)d\lambda\,du.\end{aligned}$$

References

Audenaert K, Nussbaum M, Szkola A, Verstraete F (2008) Asymptotic error rates in quantum hypotesis testing. Commun Math Phys 278:251–283

Bahadur RR (1971) Some limit theorems in statistics. Regional Conference Series in Applied Mathematics, vol 4. SIAM, Philadelphia

Blahut RE (1974) Hypothesis testing and information theory. IEEE Trans Inf Theory 20:405–417

Bouaziz M (1993) Testing Gaussian sequences and asymptotic inversion of Toeplitz operators. Probab Math Stat 14(2):207–222

Bucklew J (1990) Large deviations techniques in decision, simulation, and estimation. Wiley Series in Probability and Mathematical Statistics. Wiley, New York

Chernoff H (1952) A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. Ann Math Stat 23(4):493–507

Chernoff H (1956) Large-sample theory: parametric case.Ann Math Stat 27(1):1–22

Coursol J, Dacunha-Castelle D (1979) Sur la formule de Chernoff pour deux processus gaussiens stationnaires. C R Acad Sci Paris 288 A:769–770

Csiszár I, Longo G (1971) On the error exponent for source coding and fortesting simple statistical hypotheses. Studia Sci Math Hungarica 6:181–191

Dahlhaus R (1996a) On the Kullback–Leibler information divergence of locally stationary processes. Stoch Process Appl 62:139–168

Dahlhaus R (1996b) Asymptotic statistical inference for nonstationary processes with evolutionary spectra. In: Robinson PM, Rosenblatt M (eds) Athens conference on applied probability and time series. Springer, New York, pp 145–159

Dahlhaus R (1997) Fitting time series models to nonstationary processes. Ann Stat 25(1):1–37

Dahlhaus R (2009) Local inference for locally stationary time series based on the empirical spectral measure. J Econ 151:101–112

Dembo A, Zeitouni O (1998) Large deviations techniques and applications. Stochastic Modelling and Applied Probability. Springer, Heidelberg

Gapeev P, Küchler U (2008) On large deviations in testing Ornstein-Uhlenbeck-type models. Stat Inference Stoch Process 11:143–155

Genon-Catalot V, Picard D (1993) Eléments de statistique asymptotique. Mathématiques & Applications. Springer, Paris

Hirukawa J, Taniguchi M (2006) LAN theorem for non-Gaussian locally stationary processes and its applications. J Stat Plan Inference 136:640–688

Hoeffding W (1965) Asymptotically optimal tests for multinomial distributions. Ann Math Stat 36:369–401

Priestly MB (1981) Spectral analysis and time series, vol 2. Monographs and textbooks. Academic Press, London

Taniguchi M (2001) On large deviation asymptotics of some tests in time series. J Stat Plan Inference 97:191–200

Taniguchi M, Kakizawa Y (2000) Asymptotic theory of statistical inference for time series. Springer Series in Statistics. Springer, New York

Zani M (2002) Large deviations for quadratic forms of locally stationary processes. J Multivar Anal 81: 205–228