Bernoulli **8**(4), 2002, 451–473

Non-informative priors in the case of Gaussian long-memory processes

ANNE PHILIPPE[1] and JUDITH ROUSSEAU[2]

[1] *Laboratoire de Mathématiques Appliquées, CNRS-FRE 2222, Bâtiment M2 Université des Sciences et Technologies de Lille, 59655 Villeneuve d'Ascq Cedex, France.*
E-mail: Anne.Philippe@univ-lille1.fr
[2] *Laboratoire de Statistique, Université Paris V, 45 rue des St-Pères, 75 006 Paris, France.*
E-mail: rousseau@ensae.fr

In this paper, we consider an asymptotic Bayesian analysis for Gaussian processes with long memory. First, we determine the asymptotic expansion of the posterior density based on a normal approximation. This expansion leads to the construction of Bayesian confidence regions such as highest posterior density regions and to the determination of matching prior. Then, we generalize Clarke and Barron's result in the long-memory set-up. More precisely, we establish the asymptotic expansion of the Kullback–Leibler distance between the true density and the marginal density of the observations. As in the independent and identically distributed case, this result gives an asymptotic justification of Berger and Bernardo's algorithm to obtain reference priors.

Keywords: Kullback–Leibler distance; Laplace expansion; long memory; matching prior; reference prior; spectral density

1. Introduction

In this paper we consider a stationary Gaussian long-memory process $(X_n)_{n\in\mathbb{N}}$ whose spectral density has the following form:

$$f_\theta(\lambda) \sim |\lambda|^{-\alpha(\theta)} L_\theta(\lambda), \qquad \theta \in \Theta \subset \mathbb{R}^k, \qquad \textit{as } \lambda \to 0, \tag{1}$$

where $\alpha(\theta) \in (0, 1)$, and L_θ is a slowly varying function around 0, continuous on the subset $[-\pi, \pi] - \{0\}$. For instance, the well-known autoregressive fractionally integrated moving average (ARFIMA) processes are contained in this class of processes. The ARFIMA processes satisfy equations of the form

$$P(B)(1-B)^d X_n = Q(B)\epsilon_n$$

where ϵ_n is a Gaussian white noise, $d \in (-\frac{1}{2}, \frac{1}{2})$, B denotes the backward shift operator and P and Q are polynomials whose roots are outside the unit circle (see Granger and Joyeux 1980; and Hosking 1981). The spectral density of X_n is then

$$|1 - \mathrm{e}^{-\mathrm{i}\lambda}|^{-2d} \frac{\sigma^2}{2\pi} \frac{|Q(\mathrm{e}^{\mathrm{i}\lambda})|}{|P(\mathrm{e}^{\mathrm{i}\lambda})|}. \tag{2}$$

Thus, $\theta = (d, \sigma, \psi, \phi)$ where (ψ, ϕ) are the parameters of the ARMA process, $\alpha(\theta) = 2d$ and

$$L_\theta(\lambda) = \frac{\sigma^2}{2\pi} \frac{|Q(\mathrm{e}^{\mathrm{i}\lambda})|}{|P(\mathrm{e}^{\mathrm{i}\lambda})|}.$$

An extensive literature is devoted to the analysis of stationary long-memory time series. Many heuristic methods have been considered to estimate the long-memory parameter $\alpha(\theta)$ (see Beran 1994, for a review). From a frequentist point of view, many authors have studied the Gaussian maximum likelihood estimator as well as the Whittle estimator (see Dahlhaus 1989; Fox and Taqqu 1986; Giraitis and Surgalis 1990). In particular, the asymptotic properties of these estimators, such as asymptotic normality, are well established.

Bayesian analysis was introduced for Gaussian long-memory processes, by Carlin *et al*. (1985) in the special case of Gaussian ARFIMA models. More recently, Koop *et al*. (1994; 1997) and Pai and Ravishanker (1998) have proposed a different Bayesian analysis in the same set-up using the exact likelihood to obtain the posterior distribution. We recall that in the Gaussian case, the likelihood function is given by

$$p_\theta^n(X) = \frac{\mathrm{e}^{-X^{\mathrm{T}}\Sigma_n^{-1}X/2}}{\det[\sum_n]^{1/2}(2\pi)^{n/2}},$$

where Σ_n is the covariance matrix and X^{T} denotes the transpose of X, as it does throughout this paper. The difficulty here is to calculate Σ_n^{-1} explicitly. Koop *et al*. (1997) use Sowell's method to calculate Σ_n^{-1} recursively (see Sowell 1992). Pai and Ravishanker (1998) incorporate latent variables to obtain the explicit form of the likelihood function. In both studies, given the complexity of the models, Monte Carlo methods are used to evaluate the properties of the posterior distribution. None of these studies give asymptotic results, such as the convergence of their estimators. There are no results on the approximation of the posterior density, which is a common way of constructing Bayesian confidence regions.

The choice of the prior distribution in the Bayesian set-up is important. When there is no, or hardly any, prior information on the model, it is necessary to use default procedures to determine the prior. Osiewalski and Steel (1993) study some priors in terms of robustness with respect to the model considered. However, none of the papers mentioned previously consider fully non-informative priors.

Two main procedures are often considered in the non-informative set-up: one of these leads to the reference prior defined by Bernardo (1979) and Berger and Bernardo (1992a; 1992b); the other is the matching procedure, as defined by Welch and Peers (1963), for instance. There is a vast literature on both types of non-informative prior in the independent and identically distributed (i.i.d.) set-up. In particular, Clarke and Barron (1990) study very carefully the asymptotic behaviour of the Kullback–Leibler divergence between the posterior and the prior, when there are no nuisance parameters, which validates asymptotically the reference prior algorithm. In a similar way, Ghosh and Mukerjee (1992) obtain the same type of expansions in the nuisance parameter case. Other approaches, based on the same kind of ideas, are also considered; see, for instance, Clarke and Sun (1997). The frequentist validation of priors, as used in the construction of matching priors, is also a criterion for the choice of a non-informative prior. Berger and Bernardo (1989), in particular, use these matching priors to choose a single prior among a certain number of candidate reference priors; see Ghosh *et al*. (1995). There is a large literature on the validation and determination of matching priors. Welch and Peers (1963) and Peers (1965) study matching priors based on one-sided intervals in the one-dimensional and in the nuisance parameter case respectively, when the observations are i.i.d. Their results are improved by Tibshirani (1989), who gives the general form of such non-informative priors, provided there is an orthogonal parametrization. Datta and Ghosh (1995) and Datta (1996) study simultaneous and joint matching priors in the multidimensional set-up; Mukerjee and Dey (1993) and Datta and Ghosh (1995) determine second-order matching priors in the two-dimensional case and in the multidimensional case, respectively. Other types of matching priors are also considered; see, for instance, Ghosh and Mukerjee (1992; 1993). However, none of these studies consider dependent processes.

In Section 2, we determine an asymptotic expansion for the posterior density based on a normal approximation. This is of primary interest since it allows for the construction of confidence regions such as Bayesian intervals or highest posterior density regions. Another important application of such expansions is the determination of matching priors, which is done in Section 2.2.

In Section 3, we establish a result similar to that of Clarke and Barron (1990) in the long-memory set-up. In other words, we obtain a first-order asymptotic expansion of the relative entropy (or the Kullback–Leibler distance) between the true distribution and the mixtures of distributions (or the marginal distribution of the sample). The applications considered by Clarke and Barron (1990) can therefore be obtained in the long-memory case, and, in particular, the asymptotic validation of the reference priors.

The asymptotic results obtained in Sections 2 and 3 are formally the same as in the i.i.d. case, although the techniques used are different. The main tools used here are central limit theorems for quadratic forms and uniform convergence theorems for products of Toeplitz matrices, instead of the usual law of large numbers and central limit theorems for sums of i.i.d. random variables.

In this paper we confine ourselves to Gaussian processes; this condition is particularly important in a Bayesian framework. The Bayesian inference is obtained from the posterior distribution, which follows from an application of Bayes's rule. Therefore, it requires a closed form of the likelihood function. In general, a closed form of the likelihood function is not available for long-memory processes unless the Gaussian assumption holds. Note that there is work on Bayesian inference (via simulation) without an explicitly available likelihood for other types of process; see, for instance, Geyer (1999) for point processes.

Moreover, in this paper we study processes which follow the same kind of assumptions as those of Dahlhaus (1989). We assume that if $\theta \neq \theta'$ then $\{f_\theta(\lambda) \neq f_{\theta'}(\lambda)\}$ has positive Lebesgue measure. We denote the covariance function of the process by $\gamma_\theta(n)$, $n \in \mathbb{N}$, where

$$\gamma_\theta(n) = \int_{-\pi}^{\pi} e^{i\lambda n} f_\theta(\lambda) d\lambda.$$

We also need the following assumptions:

Assumption 1. *$f_\theta(\lambda)$ is $s+1$ times continuously differentiable with respect to θ, where $s \geqslant 3$, and each derivative is continuous in (λ, θ), $\lambda \neq 0$. In addition, $f_\theta(\lambda)^{-1}$ is continuous in (λ, θ) for all (λ, θ).*

Assumption 2. *The derivatives $(\partial/\partial\lambda) f_\theta(\lambda)^{-1}$ and $(\partial^2/\partial\lambda^2) f_\theta(\lambda)^{-1}$ are continuous in (λ, θ) for $\lambda \neq 0$ and*

$$\left(\frac{\partial}{\partial\lambda}\right)^k f_\theta(\lambda)^{-1} = O(|\lambda|^{\alpha(\theta)-k-\delta}),$$

for $k = 0, 1, 2$ and all $\delta > 0$.

Assumption 3. *For all $\delta > 0$, for all $(j_1, \ldots, j_m)$, $m \leqslant s+1$, and for $\lambda \in (0, \pi)$,*

$$\frac{\partial^m f_\theta(\lambda)^{-1}}{\partial\theta_{j_1} \ldots \partial\theta_{j_m}} = O\big(|\lambda|^{\alpha(\theta)-\delta}\big).$$

Assumption 4. *In differentiating $\gamma_\theta(n)$ with respect to θ, the derivatives indicated in Assumption 3 may be taken inside the integral sign.*

Assumption 5. *Over any compact subset $\Theta^* \subset \Theta$, the constants implied by the order bounds in Assumptions 2 and 4 may be chosen independently of θ.*

Assumption 6. *The function $\alpha(\theta)$ is continuous in θ.*

These assumptions are fairly classical; the classical ARFIMA and fractional exponential (FEXP) models satisfy them. Note that FEXP models are presented in Section 3.2 to illustrate the construction of the reference priors.

In the following section we prove that the posterior density is approximately Gaussian and that it allows for a Laplace expansion to the order s. We then use this result to determine matching priors.

2. Laplace expansion for long-memory processes

Laplace expansions of posterior distributions have many applications: among others, the construction of Bayesian confidence regions and the asymptotic performance of Bayes estimates. An asymptotic expansion such as the one obtained in Theorem 1 below enables us to approximate terms like $\int h(\theta)\pi(\theta|X^n)\mathrm{d}\theta$. This leads to the posterior mean, for instance, if $h(\theta) = \theta$. If $h(\theta)$ is an indicator function of a set, we obtain a posterior coverage and we can then construct a good approximation of Bayesian confidence regions. Moreover, this can also be used to calculate Edgeworth expansions of frequentist coverages of confidence sets, by using the shrinkage argument as developed by Sweeting (1995). Another application of this expansion is the determination of matching priors, which is developed in Section 2.2.

Let φ and Φ denote respectively the density and the cdf of a standard Gaussian random vector, and φ_V the density of a Gaussian random vector with zero mean and covariance matrix V. We denote

$$D^\nu g(\theta) = \frac{\partial^{|\nu|} g(\theta)}{\partial\theta_1^{\nu_1} \ldots \partial\theta_k^{\nu_k}}$$

where $\nu = (\nu_1, \ldots, \nu_k)$ and $|\nu| = \sum_{i=1}^k \nu_i$. We also write $D_{j_1 \ldots j_r} g(\theta)$ for the rth derivative of g with respect to $(\theta_{j_1}, \ldots, \theta_{j_r})$, $(j_1, \ldots, j_r) \in \{1, \ldots, k\}^r$, i.e.

$$D_{j_1 \ldots j_r} g(\theta) = \frac{\partial^r g(\theta)}{\partial\theta_{j_1} \ldots \partial\theta_{j_r}}.$$

Finally, $l_n(\theta)$ is the log-likelihood function, and $\hat{\theta}$ the maximum likelihood estimator.

2.1. Expansion of the posterior distribution

Theorem 1. *Let θ_0 be fixed in the interior of $\Theta \subset \mathbb{R}^k$, and X be a Gaussian process satisfying Assumptions 1–6. In this section, we denote $X^n = (X_1, \ldots, X_n)$. Let π be a prior on Θ with compact support, and such that $\pi(\theta_0) > 0$. Assume that π is $s-2$ times continuously differentiable. Then, for any smooth Borel subset A of Θ,*

$$\int_A \pi(\theta|X^n)\mathrm{d}\theta = \int_{\sqrt{n}(A-\hat{\theta})} \varphi_{J_n^{-1}}(u)\left[1 + \sum_{j=1}^{s-2} \frac{P_j(u, X^n)}{n^{j/2}}\right]\mathrm{d}u + o_{P_{\theta_0}^n}(n^{-(s-2)/2}), \tag{3}$$

where the $P_j(u, X^n)$ are polynomial functions of u whose coefficients depend only on X^n and J_n is the observed information matrix calculated at $\hat{\theta}$, i.e.

$$(J_n)_{r,s} = -\frac{\partial^2 l_n(\hat{\theta})}{\partial\theta_r\partial\theta_s}, \qquad r, s \leqslant k.$$

The above result is uniform over all compact subset of Θ.

Proof. The proof follows the same ideas as in Johnson (1970). We have

$$l_n(\theta) = l_n(\hat{\theta}) - \frac{n(\theta - \hat{\theta})J_n(\theta - \hat{\theta})}{2} + \ldots + \frac{1}{s!}\sum_{j_1,\ldots,j_s=1}^{k} (\theta - \hat{\theta})_{j_1} \ldots (\theta - \hat{\theta})_{j_s} D_{j_1 \ldots j_s} l_n(\tilde{\theta}),$$

where $\tilde{\theta} \in (\theta, \hat{\theta})$. Similarly, if we write $\psi(\theta) = \log(\pi(\theta))$, we obtain

$$\psi(\theta) = \psi(\hat{\theta}) + \ldots + \sum_{j_1,\ldots,j_{s-2}=1}^{k} \frac{(\theta - \hat{\theta})_{j_1} \ldots (\theta - \hat{\theta})_{j_{s-2}}}{(s-2)!} D_{j_1 \ldots j_{s-2}} \psi(\bar{\theta}),$$

where $\bar{\theta} \in (\theta, \hat{\theta})$.

Denote

$$k_n(\theta) = -\frac{n(\theta - \hat{\theta})J_n(\theta - \hat{\theta})}{2} + \ldots + \frac{1}{(s-1)!}\sum_{j_1,\ldots,j_{s-1}=1}^{k} (\theta - \hat{\theta})_{j_1} \ldots (\theta - \hat{\theta})_{j_{s-1}} D_{j_1 \ldots j_{s-1}} l_n(\hat{\theta}).$$

The idea of the proof is to integrate the above Taylor expansions. To do so we need the following lemmas, whose proofs are given in the Appendix. These lemmas are along the lines of Johnson's (1970), but the proofs are typical of Gaussian long-memory processes and differ from those of the i.i.d. case.

Lemma 1. *For all θ_0 in the interior of Θ, for any $\delta > 0$ (such that $\{|\theta - \theta_0| \leqslant \delta\} \subset \Theta$), there exist $M \geqslant 0$ and $S_1 \subset \mathbb{R}^n$ such that, for any $|\nu| \leqslant s$,*

$$\left| \frac{D^\nu l_n(\theta')}{n} \right| \leqslant M, \qquad \forall |\theta' - \theta_0| < \delta, \forall X \in S_1$$

with $P^n_{\theta_0}[S_1^c] = o(n^{-h})$, for all $h \geqslant 0$.

Lemma 2. *Let K be a compact subset of Θ, and let $\delta > 0$. For all $\alpha \in (0, 1)$ and all $\epsilon > 0$, there exists a set S_α such that, for all $X \in S_\alpha$,*

$$l_n(\theta) - l_n(\theta_0) \leqslant -n^\alpha \epsilon, \qquad \forall \theta \in K \cap N_\delta^c,$$

and

$$P^n_{\theta_0}[S_\alpha^c] = o(n^{-h}), \qquad \forall h > 0.$$

Let $S_2 = \{|\theta_0 - \hat{\theta}| \leqslant \delta/2\} \cap S_1(\delta)$. Then, according to Lieberman *et al.* (1999),

$$P^n_{\theta_0}[|\hat{\theta} - \theta_0| > \delta/2] = o(n^{-h}), \qquad \forall h > 0.$$

Therefore, Lemma 1 implies that $P^n_{\theta_0}[S_2^c] = o(n^{-h})$, for all $h > 0$, and when $|\theta - \theta_0| < \delta/2$,

$$|\tilde{l}_n(\theta) - \tilde{l}_n(\hat{\theta}) - k_n(\theta)| \leqslant Mn^{-(s-2)/2}$$

on S_2, where $\tilde{l}_n = l_n + \psi$. Let X be in S_2. We have

$$\int_{-\delta}^{\delta} |\exp\{[\tilde{l}_n(\theta) - \tilde{l}_n(\hat{\theta})]\} - \exp\{k_n(\theta)\}| \mathrm{d}\theta < M_1 n^{-(s-2)/2}. \tag{4}$$

Let $z = \sqrt{n}(\theta - \hat{\theta})$. Since

$$k_n(z) = -z^{\mathrm{T}} J_n z/2 + \sum_{i=1}^{s-3} n^{-i/2} P_i(z;\, D^\nu l_n(\hat{\theta})/n),$$

where the P_i are polynomial functions of z with coefficients depending only on terms such as $D^\nu l_n(\hat{\theta})/n$ which are bounded on S_2,

$$\exp[k_n(z)] = \mathrm{e}^{-z^{\mathrm{T}} J_n z/2}\left(1 + \sum_{j=1}^{s-3} \frac{Q_j(z;\, D^\nu l_n(\theta)/n)}{n^{j/2}} + \frac{R_n(z)}{n^{(s-2)/2}}\right),$$

where $R_n(z)$ is an infinite sum of powers of the P_j, and the Q_j are polynomial functions of z similar to the P_j. When $|\theta - \theta_0| < \delta$, each term appearing in the P_j is bounded by a constant M, thus there exists M_3 such that on S_2,

$$\left|\sum_{i=1}^{s-3} n^{-i/2} P_i(z;\, D^\nu l_n(\hat{\theta})/n)\right| \leqslant M_3, \qquad \forall |z| \leqslant n^{1/2}\delta.$$

The infinite sum $R_n(z)$ converges uniformly on $|z| \leqslant \sqrt{n}\delta$; therefore, there exists M_2 such that

$$\int_{-\delta}^{\delta}\left|\exp[k_n(\theta)] - \mathrm{e}^{-n(\theta-\hat{\theta})^{\mathrm{T}} J_n(\theta-\hat{\theta})/2}\left(1 + \sum_{j=1}^{s-3} \frac{Q_j(\sqrt{n}(\theta-\hat{\theta});\, D^\nu l_n(\hat{\theta})/n)}{n^{j/2}}\right)\right| \mathrm{d}\theta \leqslant M_2 n^{-(s-2)/2}. \tag{5}$$

Lemma 2 and equations (4) and (5) imply that there exist $b_1, \ldots, b_{\lfloor (s-3)/2 \rfloor}$ such that $b_j = O_{P^n_{\theta_0}}(1)$ and

$$\int_\Theta \mathrm{e}^{l_n(\theta)+\psi(\theta)}\,\mathrm{d}\theta = \mathrm{e}^{l_n(\hat{\theta})}(2\pi)^{k/2}\det[J_n]^{-1/2}\left\{1 + \frac{b_1}{n} + \ldots + \frac{b_{\lfloor (s-3)/2 \rfloor}}{n^{\lfloor (s-3)/2 \rfloor}} + O_{P^n_{\theta_0}}(n^{-(s-2)/2})\right\},$$

and Theorem 1 follows. □

2.2. Application to the determination of matching priors

We apply this result to the determination of priors matching the frequentist and the Bayesian coverage of one-sided intervals. Let $\theta = (\theta_1, \ldots, \theta_k)$. We consider one-sided intervals for θ_1, the nuisance parameter being $(\theta_2, \ldots, \theta_k)$:

$$P^\pi[\{\theta_1 \leqslant k_n(\alpha)\}|X] = \alpha.$$

Let $I(\theta)$ be the asymptotic Fisher information matrix:

$$I(\theta) = \lim_{n\to\infty} I_n(\theta), \qquad (I_n(\theta))_{r,s} = -\frac{1}{n}\mathrm{E}_\theta\left[\frac{\partial^2 l_n(\theta)}{\partial\theta_r \partial\theta_s}\right], \qquad r, s \leqslant k.$$

We thus obtain

$$(I(\theta))_{r,s} = \frac{1}{4\pi}\int_{-\pi}^{\pi} \frac{D_r f_\theta(\lambda) D_s f_\theta(\lambda)}{f_\theta^2(\lambda)}\,\mathrm{d}\lambda, \qquad r, s \leqslant k.$$

We have the following theorem:

Theorem 2. *Under Assumptions 1–6, the frequentist coverage of* $\{\theta_1 \leqslant k_n(\alpha)\}$ *has Edgeworth expansion*

$$P_\theta^n[\theta_1 \leqslant k_n(\alpha)] = \alpha + \frac{\varphi(\Phi^{-1}(\alpha))}{\sqrt{n}} \left\{ \frac{I^1(\theta) D \log \pi(\theta)}{\sqrt{I^{11}(\theta)}} - D^{\mathrm{T}}\left((I^1(\theta))^{\mathrm{T}} / \sqrt{I^{11}(\theta)} \right) \right\}$$

$$+ \frac{P_2(\theta, \pi, \alpha)}{n} + \ldots + \frac{P_{s-2}(\theta, \pi, \alpha)}{n^{(s-2)/2}} + o(n^{-(s-2)/2}), \tag{6}$$

where I^1 denotes the first row of the inverse of I, and $P_2, \ldots, P_{s-2}$ are continuous functions of θ and are formally the same as in the i.i.d. case.

Proof. The proof follows from Theorem 1 and the result of Lieberman *et al*. (1999) result on the existence of the Edgeworth expansion. □

Therefore, matching priors are the solutions of

$$\frac{I^1(\theta) D \log \pi(\theta)}{\sqrt{I^{11}(\theta)}} - D^{\mathrm{T}}\left((I^1(\theta))^{\mathrm{T}} / \sqrt{I^{11}(\theta)} \right) = 0. \tag{7}$$

In the one-dimensional case, we obtain Jeffrey's prior. In the multivariate case, however, the determination of the solutions of (7) can be very difficult. Consider, for instance, an ARFIMA(1, d, 1); some terms in I^1 are defined as infinite sums and the solutions, if they exist, would also be defined as infinite sums. However, in the case of ARFIMA(0, d, 0), it is possible to determine the matching priors in closed form. In this case the spectral density is equal to

$$f_\theta(\lambda) = \frac{\sigma^2}{2\pi} \mathrm{e}^{-d \log[2(1-\cos\lambda)]}.$$

The parameter is then (d, σ). Suppose that d is the parameter of interest. Then the matching priors are in the form

$$\pi(d, \sigma) = \sigma \mathrm{e}^{-c_1 d} h(\sigma^2 \mathrm{e}^{-c_1 d}),$$

where h is any real function twice continuously differentiable and the constant c_1 is given by

$$c_1 = \frac{1}{4\pi} \int_{-\pi}^{\pi} \log[2(1-\cos\lambda)]^2 \, \mathrm{d}\lambda. \tag{8}$$

For ARFIMA(p, d, q) models, Pai and Ravishanker (1998) and Koop *et al*. (1997) use the prior $\pi(\theta) = \sigma^{-1}$. In the particular case of the ARFIMA(0, d, 0) model, this prior is a matching prior by taking $h(x) = x$. However, it does not satisfy equation (7) in other ARFIMA(p, d, q) processes (p or $q > 0$) and therefore it is not a matching prior for these models.

We will see in Section 4 that these priors lead to proper posteriors, except on a subset with null Lebesgue measure, if h is integrable on $\mathbb{R}^+$ or if $h(u) \propto u^{-p}$, $p > 0$. In particular $\pi(d, \sigma) \propto \sigma^{-1}$ is obtained by considering $h(u) = u^{-1}$.

3. Information-theoretic asymptotics

Another type of asymptotics is important in the Bayesian framework, namely the information-theoretic asymptotics. This was studied in the i.i.d. case by Clarke and Barron (1990) and can be applied to many set-ups. In particular, they give an information interpretation to Berger and Bernardo's reference priors.

In this section, we state the validity of the expansion of the Kullback–Leibler divergence between the distribution $P^n_{\theta_0}$ and the marginal distribution of X, M_n, i.e.

$$K[P^n_{\theta_0}|M_n] = \mathrm{E}_{P^n_{\theta_0}}(\log(p^n_{\theta_0}(X)/m(X))),$$

where $m(X)$ and $p^n_{\theta_0}$ are the densities of M_n and $P^n_{\theta_0}$ respectively, with respect to Lebesgue measure. We prove that the expansion is formally the same as in the i.i.d. case. We then use this result to validate asymptotically Berger and Bernardo's reference priors.

3.1. Main theorem

Theorem 3. *Let θ_0 be fixed in the interior of Θ, and suppose that the Gaussian process satisfies Assumptions 1–6 with $s = 3$. Let π be a prior on Θ continuously differentiable with $\pi(\theta_0) > 0$ and compact support. Then, uniformly in θ_0 over all compact subsets of Θ,*

$$\lim_{n\to\infty}\left(K[P^n_{\theta_0}|M_n] - \frac{k}{2}\log\frac{n}{2\pi}\right) = -\log\pi(\theta_0) + \frac{1}{2}\log\det[I(\theta_0)] - \frac{k}{2}. \tag{9}$$

In $L_1(P^n_{\theta_0})$ as well as in probability,

$$\lim_{n\to\infty}\left(\log\frac{p^n_{\theta_0}(X)}{m(X)} + \frac{1}{2}S_n^{\mathrm{T}}I(\theta_0)^{-1}S_n - \frac{k}{2}\log\frac{n}{2\pi}\right) = -\log\pi(\theta_0) + \frac{1}{2}\log\det[I(\theta_0)], \tag{10}$$

where $S_n = n^{-1/2}Dl_n(\theta_0)$.

Proof. The proof is essentially the same as that of Clarke and Barron in the sense that the decomposition of R_n is the same. However, the control of $\limsup \mathrm{E}[R_n]$ differs.

Let

$$R_n = \log\frac{p^n_{\theta_0}(X)}{m(X)} - \left(\frac{k}{2}\log\frac{n}{2\pi} - \log\pi(\theta_0) + \frac{\log\det[I(\theta_0)]}{2} - \frac{S_n^{\mathrm{T}}I(\theta_0)^{-1}S_n}{2}\right).$$

The idea is to prove that $\lim \mathrm{E}|R_n| = 0$. Let

$$N_\delta = \{\theta;\ |\theta - \theta_0| \leqslant \delta\},$$

$$A_n(\delta, \epsilon) = \left\{\int_{N_\delta^c} p^n_\theta(X)\pi(\theta)\mathrm{d}\theta \leqslant \epsilon\int_{N_\delta} p^n_\theta(X)\pi(\theta)\mathrm{d}\theta\right\},$$

$$\begin{aligned}B_n(\delta, \epsilon) = \{&(1-\epsilon)(\theta-\theta_0)^{\mathrm{T}}I(\theta_0)(\theta-\theta_0) \leqslant (\theta-\theta_0)^{\mathrm{T}}J_n(\tilde\theta)(\theta-\theta_0)\\ &\leqslant (1+\epsilon)(\theta-\theta_0)^{\mathrm{T}}I(\theta_0)(\theta-\theta_0),\ \forall\theta, \tilde\theta \in N_\delta\},\end{aligned}$$

where $J_n(\theta) = -D^2 l_n(\theta)/n$ and

$$C_n(\delta) = \{n^{-1} S_n(\theta_0)^{\mathrm{T}} I(\theta_0)^{-1} S_n(\theta_0) \leqslant \delta^2\}$$

for $\delta > 0$ and $\epsilon > 0$. Using the same kind of argument as Clarke and Barron, we obtain that if $P^n_{\theta_0}[A^c_n] = o(\log n^{-1})$ and $P^n_{\theta_0}[B^c_n] = O(n^{-1})$, then

$$\liminf \mathrm{E}[R_n] \geqslant 0.$$

We also have

$$\begin{aligned} R_n \leqslant{}& \frac{\epsilon}{2(1+\epsilon)} S_n^{\mathrm{T}} I(\theta_0)^{-1} S_n + \frac{k}{2}\log(1+\epsilon) + \rho(\delta, \theta_0) - \log[1 - 2^{k/2}\,\mathrm{e}^{-\epsilon^2 n\delta^2/8}] \\ &+ \mathbb{I}_{[B_n \cap C_n]^c}\left[\left(\log \frac{p^n_{\theta_0}(X)}{m(X)}\right)^+ + \left|\frac{k}{2}\log\frac{n}{2\pi} - \log \pi(\theta_0) + \log\det[I(\theta_0)]/2\right|\right] \\ &+ \mathbb{I}_{(B_n\cap C_n)^c} S_n^{\mathrm{T}} I(\theta_0)^{-1} S_n. \end{aligned} \tag{11}$$

We need only look at the first term on the right-hand side of (11), which we will denote by A_1. Let $E_n = \{p^n_{\theta_0}(X)/m(X) > 1\}$. Then, by restricting the integral in the definition of $m(X)$ to N_δ,

$$\begin{aligned} \mathrm{E}_{\theta_0}[A_1] &\leqslant -\mathrm{E}_{\theta_0}\left[\mathbb{I}_{(B_n\cap C_n)^c\cap E_n} \log\left(\int_{N_\delta} \mathrm{e}^{l_n(\theta)-l_n(\theta_0)} \frac{\pi(\theta)}{\pi(N_\delta)}\mathrm{d}\theta\right)\right] - P^n_{\theta_0}[(B_n\cap C_n)^c \cap E_n]\log\pi(N_\delta) \\ &\leqslant \mathrm{E}_{\theta_0}[\mathbb{I}_{(B_n\cap C_n)^c\cap E_n}]^{1/2}\left(\int_{N_\delta}\mathrm{E}^n_{\theta_0}|l_n(\theta) - l_n(\theta_0)|^2 \pi(\theta)\mathrm{d}\theta\right)^{1/2} - P^n_{\theta_0}[(B_n\cap C_n)^c]\log\pi(N_\delta), \end{aligned}$$

and

$$\begin{aligned} \mathrm{E}^n_{\theta_0}|l_n(\theta) - l_n(\theta_0)|^2 = \frac{1}{4}\Bigg\{& \mathrm{E}^n_{\theta_0}(X^{\mathrm{T}}(\Sigma_n^{-1}(\theta) - \Sigma_n^{-1}(\theta_0))X)^2 \\ &- 2\mathrm{E}_{\theta_0}(X^{\mathrm{T}}(\Sigma_n^{-1}(\theta) - \Sigma_n^{-1}(\theta_0))X)\log\frac{\det[\Sigma_n(\theta)]}{\det[\Sigma_n(\theta_0)]} \\ &+ \left[\log\frac{\det[\Sigma_n(\theta)]}{\det[\Sigma_n(\theta_0)]}\right]^2\Bigg\}. \end{aligned}$$

Using the same argument as in the proof of Theorem 1, we obtain that

$$\mathrm{E}^n_{\theta_0}|l_n(\theta) - l_n(\theta_0)|^2 \leqslant nM\delta(n\delta^{3/2} + 2),$$

where M does not depend on θ (when θ varies in a compact), which implies that

$$\mathrm{E}^n_{\theta_0}[A_1] \leqslant n^{1/2}M\delta^{1/2}\big(n\delta^{3/2}+2\big)^{1/2} P^n_{\theta_0}[(B_n\cap C_n)^c\cap E_n]^{1/2} - P^n_{\theta_0}[(B_n\cap C_n)^c]\log\pi(N_\delta).$$

Hence, to obtain (9) and (10) it only remains to prove

$$P^n_{\theta_0}[A^c_n] = o(\log n^{-1}), \qquad P^n_{\theta_0}[B^c_n] = o(n^{-1}), \qquad P^n_{\theta_0}[C^c_n] = o(n^{-1}),$$

uniformly over any compact subset of Θ. First, Theorem 1 implies that

$$P^n_{\theta_0}[A^c_n] = P^n_{\theta_0}\left[\int_{N^c_\delta} \pi(\theta|X)\mathrm{d}\theta > \epsilon(1+\epsilon)^{-1}\right] \leqslant \frac{M}{\sqrt{n}},$$

with M independent of θ on a compact. Further, if $J_{ij}(\theta)$ is the (i, j)th component of $J_n(\theta)$, we have,

$$P^n_{\theta_0}[B^c_n] \leqslant \max_{i,j}\left\{ P^n_{\theta_0}\left[\sup_{\theta\in N_\delta}|J_{ij}(\theta) - I_{ij}(\theta)| > \epsilon/2k\right] + P^n_{\theta_0}\left[\sup_{\theta\in N_\delta}|I_{ij}(\theta) - I_{ij}(\theta_0)| > \epsilon/2k\right]\right\}.$$

Similarly as in the proof of Lemma 1, we obtain that

$$P^n_{\theta_0}\left[\sup_{\theta\in N_\delta}|J_{ij}(\theta) - I_{ij}(\theta)| > \epsilon/2k\right] = o(n^{-h}),$$

for all $h > 0$ and all $i, j \leqslant k$. Since $I_{ij}(\theta)$ is continuous in θ, there exists $\delta_0 > 0$ such that, for all $\delta_0 > \delta > 0$, for all $i, j \leqslant k$,

$$\sup_{\theta\in N_\delta}|I_{ij}(\theta) - I_{ij}(\theta_0)| < \epsilon/2k,$$

therefore,

$$P^n_{\theta_0}[B^c_n] = o(n^{-h}), \qquad \forall h > 0.$$

We prove in the same way that

$$P^n_{\theta_0}[C^c_n] = o(n^{-h}), \qquad \forall h > 0.$$

uniformly on compacts, and Theorem 3 is proved. □

This result can be applied to the asymptotic justification of Berger and Bernardo's algorithm. Indeed, π has compact support, thus expansion (9) can be integrated:

$$\int\pi(\theta)\left[K[P^n_\theta|M_n] - \frac{k}{2}\log\frac{n}{2\pi}\right]\mathrm{d}\theta = -\frac{k}{2} - \int\log\left(\pi(\theta)/\sqrt{\det[I(\theta)]}\right)\pi(\theta)\mathrm{d}\theta + o(n^{-1}).$$

Moreover,

$$\int\pi(\theta)K[P^n_\theta|M_n]\mathrm{d}\theta = \mathrm{E}[\log(\pi(\theta|X)/\pi(\theta))],$$

therefore this term is maximised asymptotically with

$$\pi(\theta) \propto \sqrt{\det[I(\theta)]},$$

as in the i.i.d. case.

The extension to the nuisance parameter case can thus be done exactly as in the i.i.d. case.

We now present Berger and Bernardo's algorithm. For simplicity's sake, we confine ourselves to the case of a two-bloc parameter. Consider $\theta = (\theta_1, \theta_2)$, where $\theta_1 \in \mathbb{R}^p$ is the parameter of interest and $\theta_2 \in \mathbb{R}^{k-p}$ is the nuisance parameter. According to Berger and Bernardo (1992a; 1992b), the construction of the reference prior proceeds as follows:

1. $\pi(\theta_2|\theta_1) \propto \sqrt{|I_{2,2}|}$, where $I_{2,2}$ is the lower right $(k-p, k-p)$ corner of $I(\theta_1, \theta_2)$.
2. $\pi(\theta_1) \propto \exp\{\frac{1}{2}\mathrm{E}[\log(|S_{p,p}^{-1}(\theta)|)|\theta_1]\}$, where $S_{p,p}$ is the upper left (p, p) corner of I^{-1} and $\mathrm{E}[g(\theta)|\theta_1] = \int g(\theta)\pi(\theta_2|\theta_1)\,\mathrm{d}\theta_2$.
3. $\pi(\theta) = \pi(\theta_2|\theta_1)\pi(\theta_1)$.

We use this algorithm, in the following subsection, to determine reference priors for ARFIMA and FEXP models.

3.2. Application to the determination of reference priors

3.2.1. ARFIMA models

Consider for instance the ARFIMA$(0, d, 0)$ case. d is the parameter of interest and $w = \sigma^2$ is the nuisance parameter. The Fisher information matrix is

$$I(\theta) = \begin{pmatrix} c_1 & -c_2 w^{-1} \\ -c_2 w^{-1} & w^{-2} \end{pmatrix},$$

where c_1 is given by (8) and

$$c_2 = \frac{1}{4\pi}\int_{-\pi}^{\pi} \log[2(1-\cos\lambda)]\mathrm{d}\lambda. \tag{12}$$

Note that the Fisher information matrix does not depend on d. The reference prior is of the form $\pi(d, w) = \pi(w|d)\pi(d)$, with

$$\pi(w|d) \propto \sqrt{\frac{1}{w^2}} = w^{-1},$$

$$\pi(d) \propto 1,$$

according to Bernardo and Berger's algorithm and the reference is given by $\pi(d, w) = w^{-1}$. Note that the reference prior is contained in the class of matching priors; recall that this was the prior used by Pai and Ravishanker (1998) and Koop *et al*. (1997) in ARFIMA(p, d, q) models.

In the ARFIMA$(1, d, 1)$ case, we have not obtained any explicit form for matching priors, but we derive a reference prior below. The spectral density is

$$f_\theta(\lambda) = w\mathrm{e}^{-d\log[2(1-\cos\lambda)]}\frac{|1-\phi\mathrm{e}^{\mathrm{i}\lambda}|^2}{|1-\psi\mathrm{e}^{\mathrm{i}\lambda}|^2}$$

and the Fisher information matrix is

$$I(\theta) = \begin{pmatrix} c_1 & -c_2/w & I_{13} & I_{14} \\ -c_2/w & w^{-2}/2 & 0 & 0 \\ I_{13} & 0 & I_{33} & I_{34} \\ I_{14} & 0 & I_{34} & I_{44} \end{pmatrix},$$

where I_{13}, I_{14} depend only on (ϕ, ψ) and $I_{33} = 1/(1-\phi^2)$, $I_{44} = 1/(1-\psi^2)$, $I_{34} = -1/(1-\psi\phi)$.

If d is the parameter of interest and ω, ϕ, ψ the nuisance parameters, we obtain

$$\pi(d|\omega, \phi, \psi) \propto \begin{vmatrix} w^{-2}/2 & 0 & 0 \\ 0 & I_{33} & I_{34} \\ 0 & I_{34} & I_{44} \end{vmatrix}^{1/2} \propto w^{-1}\pi^J(\phi, \psi),$$

$$\pi(d) \propto 1,$$

where

$$\pi^J(\phi, \psi) \propto \left(\frac{1}{(1-\psi^2)(1-\phi^2)} - \frac{1}{(1-\psi\phi)^2}\right)^{1/2}$$

is Jeffreys prior for an ARMA(1, 1) process when the variance of the white noise is known and equals 1. We obtain

$$\pi(d, w, \phi, \psi) \propto w^{-1}\left(\frac{1}{(1-\psi^2)(1-\phi^2)} - \frac{1}{(1-\psi\phi)^2}\right)^{1/2}.$$

Thus, for this ARFIMA(1, d, 1) model, the prior used by Pai and Ravishanker (1998) and Koop *et al*. (1997) does not coincide with the reference prior.

3.2.2. FEXP models

The FEXP process is defined in Beran (1994). Let $g : [-\pi, \pi] \to \mathbb{R}^+$ be a positive symmetric function satisfying $g(\lambda) \sim \lambda$, as λ goes to zero. Define $f_0 = 1$ and let $f_1, \ldots, f_p$ be smooth and symmetric functions on $[-\pi, \pi]$. Assume that there exists an integer n^* such that the matrix in $\mathbb{R}^{p \times n^*}$, with (i, j)th element $f_i(2\pi j/n)$ is non-singular. The spectral density of an FEXP process is given by

$$g(\lambda)^{1-2H} \exp\left(\sum_{i=0}^{p} \eta_i f_i(\lambda)\right), \tag{13}$$

where $H \in [\frac{1}{2}, 1)$ and $\eta_k \in \mathbb{R}$ for $k \in \{0, \ldots, p\}$. We assume that the functions f_i and g are known explicitly and thus the set of parameters to estimate is $\theta = (H, \eta_0, \ldots, \eta_p)$ with $\alpha(\theta) = 1 - 2H$, H being the parameter of interest. It is easy to show that the Fisher information matrix $I(H, \eta_0, \ldots, \eta_p)$ does not depend on the parameter $(H, \eta_0, \ldots, \eta_p)$. Therefore, the reference prior is the Lebesgue measure (i.e. $\pi(H, \eta_0, \ldots, \eta_p) \propto 1$).

To our knowledge there are no other Bayesian analyses for these models.

4. Posterior distributions

In this section, we study the propriety of the posterior distribution in some special cases. Generally speaking, the likelihood function is

$$p_\theta^n(X) = \frac{\mathrm{e}^{-X^{\mathrm{T}}\Sigma_n^{-1}X/2}}{\det[\Sigma_n]^{1/2}(2\pi)^{n/2}}.$$

Suppose that the spectral density can be expressed as

$$f_\theta(x) = \frac{\sigma^2}{2\pi}(1-\cos x)^{-d} L_{\theta_2}(x), \tag{14}$$

where L_{θ_2} is continuous in x on $[-\pi, \pi]$, and $\theta = (d, \sigma, \theta_2)$, $\theta_2 \in \mathbb{R}^{k-2}$.

Consider a prior of the form

$$\pi(\theta) = \sigma^{-\alpha} p(d, \theta_2)$$

or bounded by a function in this form. Integrating over σ leads to

$$\pi(d, \theta_2|X) \propto \frac{(X^{\mathrm{T}}S(d, \theta_2)^{-1}X)^{-(n+\alpha-1)/2}}{\det[S]^{-1/2}} p(d, \theta_2), \tag{15}$$

where $\Sigma_n = \sigma^2 S$ with S independent of σ. A proper prior on (d, θ_2) is not enough to obtain the propriety of the posterior, since the right-hand side of (15) is not integrable as a function of X. Therefore, it is necessary to specify the behaviour of (15), in the neighbourhood of the boundaries of the stationary domain of X : $\{(d, \theta_2); \forall\sigma > 0$, such that the process X is stationary$\}$. For example, in the case of an ARFIMA(p, d, q), defined by (2), these boundaries are defined by $d = \frac{1}{2}$ or the norm of at least one of the roots of P and Q is equal to 1. In the case of a FEXP process defined by (13) the boundaries are defined by $d = 1/2$ or $\eta_i = \infty$ for at least one of the i in $\{1, \ldots, p\}$.

Assume that $\Theta = [0, \frac{1}{2}[\times\mathbb{R}^+ \times \Theta_2$, where Θ_2 is an open subset of $\mathbb{R}^{k-2}$. Let L_{θ_2} be continuously differentiable on $[-\pi, \pi]$, for all $\theta_2 \in \Theta_2$. We then have the following result:

Lemma 3. *Uniformly over all compact subsets of* Θ_2*,*

$$S = \frac{\rho}{(\frac{1}{2}-d)} J + \Gamma + O(\tfrac{1}{2}-d), \tag{16}$$

where J is the matrix in $\mathbb{R}^n$ whose components are all equal to 1 *and Γ is a non-singular matrix with elements $\Gamma_{i,j} = g(i-j)$, where*

$$g(k) = -\frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{1-\mathrm{e}^{\mathrm{i}kx}}{\sqrt{1-\cos x}} L_{\theta_2}(x)\mathrm{d}x,$$

and $\rho = \rho_0 + t\rho_1$, with $\rho_0 > 0$.

Let $X = \sum_{i=1}^{n-1} x_i \Gamma U_i + x_n U_n$, *where the U_i are vectors whose components are null apart from the ith component which is equal to* 1 *and the* $(i+1)$*th component which is equal to*

-1 *and* $U_n = (1, 1, \ldots, 1)^{\mathrm{T}}$. Then $\Gamma U_1, \ldots, \Gamma U_{n-1}, U_n$ *are linearly independent and, as* $d \to \frac{1}{2}$,

$$\frac{(X^{\mathrm{T}} S(d, \theta_2)^{-1} X)^{-(n+\alpha-1)/2}}{\det[S]^{1/2}}$$

$$= \frac{\sqrt{\frac{1}{2} - d}\left(\sum_{i,j=1}^{n-1} x_i x_j U_i^{\mathrm{T}} \Gamma U_j + (1/2 - d) x_n^2/\rho\right)^{-(n+\alpha-1)/2}}{\sqrt{\rho \operatorname{tr}[\operatorname{Cof}(\Gamma) J]}} (1 + O(\tfrac{1}{2} - d)).$$

This implies in particular that, for ARFIMA (0, d, 0) models, priors of the form

$$\pi(d, \sigma) = \sigma^{-\alpha} \pi(d),$$

where $\pi(d) = O((\frac{1}{2} - d)^{-3/2+\epsilon})$ lead to proper posteriors, except on a subset with null Lebesgue measure.

For more complex models, one would also have to consider the behaviour of $[X^{\mathrm{T}} S^{-1} X]^{-(n+\alpha-1)/2}/\det[S]^{1/2}$ in a neighbourhood of the boundaries of Θ_2. However, the techniques would be very similar to those used in the following proof.

Proof. Let $k \geqslant 1$. Consider $\tilde{g}(k) = \gamma(k) - \gamma(0)$. Then

$$\begin{aligned}
\tilde{g}(k) &= -\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1 - \mathrm{e}^{\mathrm{i}kx}}{(1 - \cos x)^d} L_{\theta_2}(x) \mathrm{d}x, \\
&= -k r(0) - 2 \sum_{p=1}^{k-1} (k - p) r(p) + (\tfrac{1}{2} - d) R(d, k) \\
&= g(k) + (\tfrac{1}{2} - d) R(d, k),
\end{aligned}$$

where

$$r(p) = \int_{-\pi}^{\pi} \mathrm{e}^{\mathrm{i}px} \sqrt{(1 - \cos x)} L_{\theta_2}(x) \mathrm{d}x,$$

and $R(d, k)$ converges to a constant depending on k and θ_2, as d goes to $\frac{1}{2}$.

The matrix Γ given in the lemma is non-singular. Indeed, transformations of the form row $(i) =$ row $(i) -$ row $(i+1)$, for $i = 1, \ldots, n-1$, and column $(i) =$ column $(i) -$ column $(i+1)$, $i = 1, \ldots, n-1$, lead to the matrix

$$M = \begin{pmatrix} -2r(0) & -2r(1) & \dots & -2r(n-2) & r(0) + \sum_{i=1}^{n-2} r(i) \\ -2r(1) & -2r(0) & \dots & -2r(n-3) & r(0) + 2\sum_{i=1}^{n-3} r(i) \\ \dots & \dots & \dots & \dots & \dots \\ -2r(n-2) & -2r(n-3) & \dots & -2r(0) & r(0) \\ r(0) + 2\sum_{i=1}^{n-2} r(i) & r(0) + 2\sum_{i=1}^{n-3} r(i) & \dots & r(0) & 0 \end{pmatrix}.$$

Denote by V the $(n-1) \times (n-1)$ upper left part of the above matrix. Then V can be interpreted as -2 times a covariance matrix, since $-2V = T(\sqrt{(1 - \cos x)} L_{\theta_2}(x))$, where T denotes the Toeplitz operator. Then, $MX = 0$ if and only $-2V\tilde{X} + Hx_n = 0$ and $H^{\mathrm{T}}\tilde{X} = 0$, where $X = (\tilde{X}^{\mathrm{T}}, x_n)^{\mathrm{T}}$ and

$$H = (r(0) + 2(r(1) + \dots + r(n-2)),\ r(0) + 2(r(1) + \dots + r(n-3)),\ \dots,\ r(0))^{\mathrm{T}}.$$

This implies that $H^{\mathrm{T}}\tilde{X} = H^{\mathrm{T}}V^{-1}Hx_n/2 = 0$ if and only if $x_n = 0$, which also implies that $X = 0$. Moreover, U_n is linearly independent of $(\Gamma U_1, \dots, \Gamma U_{n-1})$, therefore $(\Gamma U_1, \dots, \Gamma U_{n-1}, U_n)$ is a basis, and for $j = 1, \dots, n-1$,

$$(J + t/\rho\Gamma)^{-1}\Gamma U_j = \frac{\rho}{t} U_j.$$

So let $X = \sum_{i=1}^{n-1} \Gamma U_i + x_n U_n$;

$$X^{\mathrm{T}}[J + t/\rho\Gamma]^{-1}X = \frac{\rho}{t}\sum_{i,j=1}^{n-1} x_i x_j U_i^{\mathrm{T}}\Gamma U_j + x_n^2 U_n^{\mathrm{T}}[J + t/\rho\Gamma]^{-1}U_n.$$

Let $[J + t\rho\Gamma]V_n = U_n$, with $V_n = \sum_{i=1}^{n} v_i U_i$;

$$U_n^{\mathrm{T}}[J + t/\rho\Gamma]^{-1}U_n = U_n^{\mathrm{T}}V_n = v_n U_n^{\mathrm{T}}U_n.$$

We have

$$nv_n U_n + \frac{t}{\rho}\sum_{i=1}^{n} v_i \Gamma U_i = U_n.$$

Let $\Gamma U_n = \alpha_1 \Gamma U_1 + \dots + \alpha_{n-1}\Gamma U_{n-1} + \alpha_n U_n$; then $nv_n + t/\rho v_n \alpha_n = 1$. Obviously $\alpha_n \neq 0$ and is independent of t, so

$$v_n = \left(n + \frac{t\alpha_n}{\rho}\right)^{-1} = n - \frac{t\alpha_n}{\rho} + O(t^2).$$

Finally,

$$X^{\mathrm{T}}[J + t/\rho\Gamma]^{-1}X = \frac{\rho}{t}\sum_{i,j=1}^{n-1} x_i x_j U_i^{\mathrm{T}}\Gamma U_j + \frac{x_n^2}{\rho}\left[1 - \frac{t\alpha_n}{n\rho}\right] + O(t^2).$$

Moreover, $U_i^{\mathrm{T}}\Gamma U_j = 2r(i-j)$, so the matrix $(U_i^{\mathrm{T}}\Gamma U_j)_{i,j=1,\ldots,n-1}$ is positive. We also have:

$$\begin{aligned}\det[S] &= \frac{\rho^n}{t^n}\det[\Gamma]\det[\Gamma^{-1}J + t/\rho \mathrm{Id}]\\ &= \frac{\rho^n}{t^n}\det[\Gamma]\left(\frac{t^{n-1}}{\rho^{n-1}}\mathrm{tr}[\Gamma^{-1}J] + \frac{t^n}{\rho^n}\right)\\ &= \frac{\rho\,\mathrm{tr}[\mathrm{Cof}(\Gamma)J]}{t} + \det[\Gamma].\end{aligned}$$

Γ is non-singular so $\mathrm{rk}[\mathrm{Cof}(\Gamma)J] = 1$ and its trace is positive, and Lemma 3 is proved. □

Appendix

Proof for Lemma 1. Let $S_1 = \{X; \forall 1 \leqslant |\nu| \leqslant s,\ |D^\nu l_n(\theta')|/n \leqslant M;\ |\theta' - \theta_0| < \delta\}$, where M will be specified later.

For all $1 \leqslant |\nu| \leqslant s$,

$$D^\nu l_n(\theta) = -\frac{1}{2}X^{\mathrm{T}}D^\nu\left[\Sigma_n^{-1}\right]X - \frac{1}{2}D^\nu \log\det[\Sigma_n].$$

Thus this is a sum of quadratic forms associated with products of Toeplitz matrices in the form

$$\Sigma_n^{-1}T_n(g_1)\ldots\Sigma_n^{-1}T_n(g_p)\Sigma_n^{-1}$$

and of traces of matrices such as

$$\Sigma_n^{-1}T_n(g_1)\ldots\Sigma_n^{-1}T_n(g_p),$$

where $g_1, \ldots, g_p$ are derivatives of the spectral density. Then, $X \in S_1^c$ if and only if there exists $|\theta^* - \theta_0| < \delta$ such that

$$\left|\frac{D^\nu l_n(\theta^*)}{n}\right| \geqslant M.$$

This implies that there exist $(g_1(\theta^*), \ldots, g_p(\theta^*))$ such that either

$$\left|n^{-1}\mathrm{tr}\left[\Sigma_n^{-1}T_n(g_1)\ldots\Sigma_n^{-1}T_n(g_p)(\theta^*)\right]\right| \geqslant cM$$

or

$$n^{-1}\left|X^{\mathrm{T}}\Sigma_n^{-1}T_n(g_1)\ldots\Sigma_n^{-1}T_n(g_p)(\theta^*)\Sigma_n^{-1}(\theta^*)X\right| \geqslant cM,$$

where c is some positive constant depending only on ν. Since (see Lieberman *et al.* 1999)

$$\sup_{|\theta-\theta_0|<\delta}\left|n^{-1}\mathrm{tr}\left[\Sigma_n^{-1}T_n(g_1)\ldots\Sigma_n^{-1}T_n(g_p)(\theta)\right]-\frac{1}{2\pi}\int_{-\pi}^{\pi}f_\theta(\lambda)^p\prod_{j=1}^{p}g_{j,\theta}(\lambda)\mathrm{d}\lambda\right|$$

goes to zero when n tends to infinity, if M is large enough (independently of n and θ), we only have to study

$$P_n = P_{\theta_0}^n\left[\exists|\theta^*-\theta_0|<\delta;\ \left|X^{\mathrm{T}}\Sigma_n^{-1}T_n(g_1)\ldots\Sigma_n^{-1}T_n(g_p)(\theta^*)\Sigma_n^{-1}(\theta^*)X\right|\geqslant cMn\right].$$

Let θ be in a δ-neighbourhood of θ_0. We denote by $\Gamma_n(\theta)$ the matrix

$$\Gamma_n(\theta)=\Sigma_n^{-1}(\theta)T_n(g_1(\theta))\ldots\Sigma_n^{-1}T_n(g_p)(\theta)\Sigma_n^{-1}(\theta).$$

Then, for $\frac{1}{2}>a>0$, let

$$Z_n=n^{-1/2-a}\left\{X^{\mathrm{T}}[\Gamma_n(\theta)-\Gamma_n(\theta_0)]X-\mathrm{tr}[\Sigma_n(\theta_0)(\Gamma_n(\theta)-\Gamma_n(\theta_0))]\right\}$$

We have that

$$\begin{aligned}P_{\theta_0}^n[|Z_n|\geqslant Kn^{1/2-a}|\theta-\theta_0|] &\leqslant \mathrm{e}^{-tKn^{1/2-a}}\mathrm{E}_{\theta_0}[\mathrm{e}^{tZ_n/|\theta-\theta_0|}+\mathrm{e}^{-tZ_n/|\theta-\theta_0|}]\\ &\leqslant 2\mathrm{e}^{-tKn^{1/2-a}}\sum_{k=0}^{\infty}\frac{t^k|\mathrm{E}_{\theta_0}[Z_n^k]|}{k!|\theta-\theta_0|^k}.\end{aligned}$$

We proceed as in Dahlhaus's (1989) proof of his Lemma 6.2. Let Id be the identity matrix. The cumulant generating function, $k_n(t)$, is equal to

$$\begin{aligned}k_n(t)=&-\frac{t\,\mathrm{tr}[\Sigma_n(\theta_0)(\Gamma_n(\theta)-\Gamma_n(\theta_0))]}{n^{1/2+a}}\\ &-\frac{1}{2}\log\det\left[\mathrm{Id}-\frac{2t}{n^{1/2+a}}\Sigma_n^{1/2}(\theta_0)[\Gamma_n(\theta)-\Gamma_n(\theta_0)]\Sigma_n^{1/2}(\theta_0)\right];\end{aligned}$$

unless otherwise specified, the covariance matrix Σ_n is calculated at θ_0. So

$$k_n'(t)=-\frac{\mathrm{tr}[\Sigma_n(\theta_0)(\Gamma_n(\theta)-\Gamma_n(\theta_0))]}{n^{1/2+a}}+\mathrm{tr}[(\mathrm{Id}-tH_n)^{-1}H_n],$$

where $H_n=2n^{-1/2-a}\Sigma_n(\theta_0)^{1/2}[\Gamma_n(\theta)-\Gamma_n(\theta_0)]\Sigma_n^{1/2}$ and, for all $j\geqslant 2$,

$$\frac{\mathrm{d}^j k_n(t)}{\mathrm{d}t^j}=2^{-1}(j-1)!\mathrm{tr}[((\mathrm{Id}-tH_n)^{-1}H_n))^j];$$

therefore, the jth cumulant of Z_n is equal to

$$C_j(Z_n) = 2^{-1}(j-1)!\mathrm{tr}[H_n^j].$$

As in Dahlhaus (1989), we denote

$$|A|^2 = \mathrm{tr}[A^{\mathrm{T}}A],$$

for any matrix A, and

$$\|A\|^2 = \sup_{Y^{\mathrm{T}}Y=1} Y^{\mathrm{T}}A^{\mathrm{T}}AY.$$

Then,

$$\begin{aligned}\mathrm{tr}[H_n^j] &= \frac{\mathrm{tr}\left\{[\Sigma_n(\theta_0)(\Gamma_n(\theta) - \Gamma_n(\theta_0))]^j\right\}}{n^{j/2+ja}} \\ &\leqslant \frac{\left|\Sigma_n^{1/2}(\theta_0)\right|^2 \left\|\left(\Sigma_n(\theta_0)^{1/2}\right)\right\|^{j-2} \|(\Gamma_n(\theta) - \Gamma_n(\theta_0))\|^j}{n^{j/2+ja}}\end{aligned}$$

Since

$$\left|\Sigma_n^{1/2}(\theta_0)\right|^2 = \mathrm{tr}\left[\Sigma_n(\theta_0)\right] = n\gamma(0)$$

and

$$\left\|\Sigma_n^{1/2}(\theta_0)\right\| = O(n^{\alpha(\theta_0)+\epsilon}), \qquad \forall \epsilon > 0,$$

we only have to study the behaviour of $\Gamma_n(\theta) - \Gamma_n(\theta_0)$. As in Dalhaus (1989), if each $g_j(\theta)$ is continuously differentiable in θ, then

$$\|\Gamma_n(\theta) - \Gamma_n(\theta_0)\| \leqslant |\theta - \theta_0| \|\nabla\Gamma_n(\theta')\|$$

where $\theta' \in (\theta, \theta_0)$ and

$$\|\Gamma_n(\theta) - \Gamma_n(\theta_0)\| \leqslant K|\theta - \theta_0| n^b, \qquad \forall b > 0.$$

Finally, we obtain that

$$|C_j(Z_n)| \leqslant K^j (j-1)! |\theta - \theta_0|^j n^{1+\alpha(\theta_0)(j-2)-j/2-aj+\epsilon j},$$

for all $\epsilon > 0$, where K depends only on ϵ and δ. Let $a > \alpha - \frac{1}{2}$; then there exists $c > 0$ such that

$$|C_j(Z_n)| \leqslant K^j (j-1)! n^{-cj} |\theta - \theta_0|^j,$$

for all $j \geqslant 2$. So, for all θ such that for all $|\theta - \theta_0| \leqslant \delta$,

$$\begin{aligned}P_{\theta_0}^n\left[|Z_n| \geqslant Kn^{1/2-a}|\theta - \theta_0|\right] &\leqslant e^{-tKn^{1/2-a}} \sum_{j=1}^{\infty} (2Kt)^j n^{-jc} \\ &\leqslant 2\mathrm{e}^{-tKn^{1/2-a}} \qquad (17)\end{aligned}$$

if $2Ktn^{-c} < \frac{1}{2}$. To obtain a bound for the supremum over $T = \{|\theta - \theta_0| \leqslant \delta\}$, we proceed similarly to Pollard (1984); we thus construct a countable set T^*, dense in T, such that

$$P^n_{\theta_0}\left[|Z_n(\theta)| > 5n^{1/2-a}J(|\theta-\theta_0|), \text{ for some } \theta \in T^*\right] \leqslant 2^{-1}\exp\{-tn^{1/2-a}\log C\}K, \quad (18)$$

where J is a positive function on $\mathbb{R}$ satisfying $J(\delta) < \infty$. Since $Z_n(\theta)$ is continuous in θ,

$$\sup_T |Z_n(\theta)| = \sup_{T^*}|Z_n|,$$

and (4) becomes

$$P^n_{\theta_0}\left[\sup_T |Z_n(\theta)| > Kn^{1/2-a}\right] \leqslant K'\mathrm{e}^{-C'n^{1/2-a}}$$

where K, K' and C' are constants independent of n. Finally, since

$$n^{-1}X^{\mathrm{T}}\Sigma_n^{-1}T_n(g_1)\ldots\Sigma_n^{-1}T_n(g_p)(\theta)\Sigma_n^{-1}(\theta)X = n^{-1/2-a}Z_n(\theta) + n^{-1}\mathrm{tr}[\Sigma_n(\theta_0)\Gamma_n(\theta)]$$
$$+ n^{-1/2}\frac{X^{\mathrm{T}}\Gamma_n(\theta_0)X - \mathrm{tr}[\Sigma_n(\theta_0)\Gamma_n(\theta_0)]}{n^{1/2}},$$

there exist K and C such that

$$P_n \leqslant K\mathrm{e}^{-C'n^{1/2-a}}$$

if M is large enough. □

Proof of Lemma 2. We have

$$l_n(\theta) - l_n(\theta_0) = -\frac{1}{2}X^{\mathrm{T}}\left[\Sigma_n^{-1}(\theta) - \Sigma_n^{-1}(\theta_0)\right]X + \frac{1}{2}\log\det\left[\Sigma_n(\theta_0)\Sigma_n^{-1}(\theta)\right].$$

Considering the same Taylor expansion as Dahlhaus (1989, p. 1754), we have

$$\log\det\left[\Sigma_n(\theta_0)\Sigma_n^{-1}(\theta)\right] = \mathrm{tr}\left[\Sigma_n(\theta_0)\Sigma_n^{-1}(\theta) - \mathrm{Id}\right] - \frac{1}{2}\sum_{i=1}^n \frac{(\lambda_{in}-1)^2}{[1+\tau(\lambda_{in}-1)]^2},$$

where $\tau \in (0, 1)$ and $(\lambda_{1n}, \ldots, \lambda_{nn})$ are the eigenvalues of $\Sigma_n(\theta_0)\Sigma_n^{-1}(\theta)$. Therefore,

$$l_n(\theta) - l_n(\theta_0) = -\frac{1}{2}\left\{X^{\mathrm{T}}\left[\Sigma_n^{-1}(\theta) - \Sigma_n^{-1}(\theta_0)\right]X - \mathrm{tr}\left[\Sigma_n(\theta_0)\Sigma_n^{-1}(\theta) - \mathrm{Id}\right]\right\}$$
$$-\frac{1}{4}\sum_{i=1}^n \frac{(\lambda_{in}-1)^2}{[1+\tau(\lambda_{in}-1)]^2}.$$

We first consider the second term on the right-hand side, dealing with it in a similar way to Dalhaus (1989). If $\alpha(\theta) \leqslant \alpha(\theta_0)$, then $0 < \lambda_{in} < [C|\theta - \theta_0| + 1]$ and

$$\frac{(\lambda_{in}-1)^2}{[1+\tau(\lambda_{in}-1)]^2} \geqslant [C|\theta-\theta_0|+1]^{-2}(1-\lambda_{in})^2.$$

Therefore, for all $K \subset \Theta$ compact, there exists $N > 0$ such that, for all $n \geqslant N$, for all $\theta \in K \cap N_\delta^c$,

$$\frac{1}{4}\sum_{i=1}^{n} \frac{(\lambda_{in}-1)^2}{[1+\tau(\lambda_{in}-1)]^2} \geqslant \frac{n[C|\theta-\theta_0|+1]^{-2}}{8\pi}\int_{-\pi}^{\pi}\left(\frac{f_{\theta_0}(x)}{f_\theta(x)}-1\right)^2 \mathrm{d}x.$$

If $\alpha(\theta) \geqslant \alpha(\theta_0)$, then calculations of the same type imply that

$$\frac{1}{4}\sum_{i=1}^{n} \frac{(\lambda_{in}-1)^2}{[1+\tau(\lambda_{in}-1)]^2} \geqslant \frac{n[C|\theta-\theta_0|+1]^{-2}}{8\pi}\int_{-\pi}^{\pi}\left(\frac{f_{\theta}(x)}{f_{\theta_0}(x)}-1\right)^2 \mathrm{d}x.$$

Let

$$Z_n^0 = X^{\mathrm{T}}\left[\Sigma_n^{-1}(\theta) - \Sigma_n^{-1}(\theta_0)\right]X - \mathrm{tr}\left[\Sigma_n(\theta_0)\Sigma_n^{-1}(\theta) - \mathrm{Id}\right].$$

Then

$$P_{\theta_0}^n[|Z_n| > n^{1/2+\alpha}] \leqslant \frac{\mathrm{E}_{\theta_0}[Z_n^{2k}]}{n^{k+2k\alpha}} \leqslant \frac{M_k}{n^{2k\alpha}}$$

for all $n \geqslant N$ and for all $\theta \in K$ (N depends only on k and K). Finally, for all $\alpha < 1$, for all $n \geqslant N$,

$$P_{\theta_0}^n[l_n(\theta) - l_n(\theta_0) \geqslant -n^\alpha\epsilon] \leqslant P_{\theta_0}^n[|Z_n| \geqslant n\alpha(\theta) - n^\alpha\epsilon] = o(n^{-h}),$$

for all $h > 0$. □

Acknowledgement

We thank the referee and the Editor for their comments that were very helpful in clarifying the paper. This work was partially supported by the TMR network, contract C.E. CT 96–0095.

References

Beran J. (1994) *Statistics for Long Memory Processes*. London: Chapman & Hall.

Berger, J.O. and Bernardo, J.M. (1989) Estimating a product of means – Bayesian analysis with reference priors. *J. Amer. Statist. Assoc.*, **84**, 200–207.

Berger, J. and Bernardo, J. (1992a) Reference priors in a variance components problem. In P.K. Goel and N.S. Iyengar (eds), *Bayesian Analysis in Statistics and Econometrics*. New York: Springer-Verlag.

Berger, J. and Bernardo, J. (1992b) On the development of the reference prior method. In J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith (eds), *Bayesian Statistics 4*, pp. 35–49. Oxford: Oxford University Press.

Bernardo, J.M. (1979) Reference posterior distributions for Bayesian inference (with discussion). *J. Roy. Statist. Soc. Ser. B*, **41**, 113–147.

Carlin, J., Dempster, A.P. and Jonas, A.B. (1985) On models and methods for Bayesian time series analysis. *J. Econometrics*, **30**, 67–90.

Clarke, B. and Barron, A. (1990) Information-theoretic asymptotics of Bayes methods. *IEEE Trans. Inform. Theory*, **36**, 453–471.

Clarke, B. and Sun, D. (1997) Reference priors under the chi-squared distance. *Sankhyā Ser. A*, **59**, 215–231.

Dahlhaus R. (1989) Efficient parameter estimation for self-similar processes, *Ann. Statist.*, **17**, 1749–1766.

Datta, G.S. (1996) On priors providing frequentist validity for Bayesian parametric functions. *Biometrika*, **83**, 287–298.

Datta, G.S. and Ghosh, J.K. (1995). On priors providing frequentist validity for Bayesian inference. *Biometrika*, **82**, 37–45.

Fox, R. and Taqqu, M.S. (1986) Large sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *Ann. Statist.*, **14**, 517–532.

Geyer, C. (1999) Likelihood inference for spatial point processes. In *Stochastic Geometry, Likelihood and Computation.*, O.E. Barndorff-Nielsen, W.S. Kendall and M.N.M. van Lieshout (Eds). London: Chapman & Hall.

Ghosh, J.K. and Mukerjee, R. (1992) Non-informative prior. In J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith (eds), *Bayesian Statistics 4*, pp. 195–203. Oxford: Oxford University Press.

Ghosh, J.K. and Mukerjee, R. (1993) Frequentist validity of highest posterior density regions in the multiparameter case. *Ann. Inst. Statist. Math.*, **45**, 293–302.

Ghosh, M., Carlin, B.P. and Srivastava, M.S. (1995) Probability matching priors for linear calibration. *T.E.S.T.*, **4**, 333–358.

Giraitis, L. and Surgalis, D. (1990) A central limit theorem for quadratic forms in strongly dependent linear variables and its applications to asymptotic normality of Whittle's estimate. *Probab. Theory Related Fields.*, **86**, 87–104.

Granger, C.W.J. and Joyeux, R. (1980) An introduction to long memory time series models and fractional differencing. *J. Time Ser. Anal.*, **1**, 15–29.

Hosking, J.R.M. (1981) Fractional differencing. *Biometrika*, **68**, 165–176.

Johnson, R.A. (1970) Asymptotic expansions associated with posterior distributions. *Ann. Math. Statist.*, **41**, 851–864.

Koop, G., Ley, E., Osiewalski, J. and Steel, M.F.J. (1994) Posterior properties of long-run impulse responses. *J. Business Econom. Statist.*, **12**, 489–492.

Koop, G., Ley, E., Osiewalski, J. and Steel, M.F.J. (1997) Bayesian analysis of long memory and persistence using ARFIMA models. *J. Econometrics*, **76**, 149–169.

Lieberman, O., Rousseau, J. and Zucker D.M. (1999) Valid asymptotic expansions for the maximum likelihood estimator of the parameter of a stationary, Gaussian, strongly dependent process. Preprint.

Mukerjee, R. and Dey, D.K. (1993) Frequentist validity of posterior quantiles in the presence of nuisance parameter: higher order asymptotics. *Biometrika*, **80**, 499–505.

Osiewalski, J. and Steel, M.F.J. (1993) Bayesian marginal equivalence of elliptical regression models. *J. Econometrics*, **59**, 391–403.

Pai, J.S. and Ravishanker, N. (1998) Bayesian analysis of autoregressive fractionally integrated moving-average processes. *J. Time Ser. Anal.*, **19**, 99–112.

Peers, H.W. (1965) On confidence sets and Bayesian probability points in the case of several parameters. *J. Roy. Statist. Soc. Ser. B*, **27**, 9–16.

Pollard (1984) *Convergence of Stochastic Processes*. New York: Springer-Verlag.

Sowell, F.B. (1992) Maximum likelihood estimation of stationary univariate fractionally integrated time series models. *J. Econometrics*, **53**, 165–188.

Sweeting, T.J. (1995) A framework for Bayesian and likelihood approximations in statistics. *Biometrika*, **82**, 1–23.

Tibshirani, R. (1989) Noninformative priors for one parameter of many. *Biometrika*, **76**, 604–608.

Welch, B. and Peers, H. (1963) On formulae for confidence points based on intervals of weighted likelihoods. *J. Roy. Statist. Soc. Ser. B*, **25**, 318–329.

Received February 2000 and revised January 2002