数理解析研究所講究録
第 1752 巻 2011 年 77-84

On prior selection from the viewpoint of spectral density estimation

科学技術振興機構 さきがけ数学領域 田中冬彦 (Fuyuhiko Tanaka)
PRESTO, Japan Science and Technology Agency (JST)

Abstract

In the Bayesian analysis of time series model, it is inevitable to determine a prior distribution of the unknown parameter. It is known that the prior distribution could affect final result significantly and until now independently identically distributed (i.i.d.) cases have been investigated to some extent. However, for non i.i.d. cases like time series model, it remains quite unclear. Recently author has been tackling the prior selection by considering the estimation of the spectral density in the Bayesian setting. In the present paper, we focus on the alpha parallel prior, which does not always exist on the statistical model. We investigate the existence of alpha parallel prior in the autoregressive moving-average (ARMA) models, which are very fundamental and important in practical analysis of time-series data. Unfortunately, we show that there exists no such prior in the proper ARMA models.

1 Introduction

In the Bayesian analysis of the time series model, it is a considerable obstacle to select an objective prior that is based on a certain justification. In econometrics, we often see the Bayesian analysis using an *ad hoc* prior (see, e.g., Zellner [20]). Even in the most simple model like the AR(1) model, objective prior selection is very challenging as is discussed in Phillips [13].

Seeking for an objective prior based on a certain theoretical argument, Berger and Yang [4] focused on the reference prior, which was proposed first by Bernardo [5] in the i.i.d. cases. They managed to derive the explicit form of the prior in the AR(1) model, but it seems more difficult to obtain the reference prior when $p\geq 2$.

In the present paper, we focus on the Bayesian estimation of the spectral density of stationary Gaussian time series models according to Komaki [11]. Then, another criteria for the prior selection is obtained. In particular, Tanaka shows that there exist superharmonic priors, good candidates of the objective prior in the AR(p) models [19]. In numerical simulation, the estimation of the spectral density based on the superharmonic prior performs better than that based on the Jeffreys prior in the AR(2) models [18].

However, it seems very hard to validate the existence of the superharmonic prior in the given time series model because a superharmonic prior is constructed of a positive nontrivial solution of the second order differential *inequality*. As another candidate of the objective prior, we focus on the α-parallel prior. Historically speaking, it was proposed first by Hartigan [8, 9, 10]. Later, Takeuchi and Amari [16] clarified an interesting connection between the information geometrical properties of the statistical model and the existence of the α-parallel prior. Although their arguments are restricted to i.i.d. models, we expect that similar consequences hold in the asymptotic setting of time series model.

Since their existence condition is written in the information geometrical terms, we apply their argument to the ARMA(p,q) model manifold, where information geometrical quantities are explicitly calculated [14, 11] and we investigate the existence of the α-parallel prior on the ARMA(p,q) model. Although this task is not so trivial, it is still more tractable than the superharmonic prior. Since the existence of both priors is deeply related to the global properties of the statistical model manifold [12, 17], we also expect that the analysis of the ARMA(p,q) models using differential geometry yields some insights for general criterion of the prior selection for time series model.

Structure

First, an α-parallel prior on the statistical model manifold is defined and some necessary and sufficient conditions of the existence are reviewed. Then, we define the trace 2-form and another necessary and sufficient condition is given. Then, we introduce the root coordinate as a convenient coordinate in the ARMA model manifold and calculate the trace 2-form, which is in a very simple form. Finally, we mention the consequence of the main result, mainly from the statistical viewpoint.

2 α-parallel prior and trace 2-form

In the statistical model manifolds of dimension d, the affine volume element is defined by d-form (differential form of degree d). When model manifolds have good properties, such a volume element can be regarded as an extension of the invariant measure, which yields a prior distribution on the parameter space. Here we briefly review the above argument according to Takeuchi and Amari [16].

Let us consider the d-dimensional orientable smooth manifold $\mathcal{M}$ with an affine connection ∇. We shall say that an affine connection ∇ is *locally equiaffine* if around each point x of $\mathcal{M}$ there is a parallel volume element, that is, a nonvanishing d-form ω such that $\nabla\omega = 0$ on a neighborhood of each x.

Definition 2.1

By an *equiaffine connection* ∇ on $\mathcal{M}$ we mean a torsion-free affine connection that admits a parallel volume element ω on $\mathcal{M}$. If ω is a volume element on $\mathcal{M}$ such that $\nabla\omega = 0$, then we say that (∇,ω) is an *affine structure* on $\mathcal{M}$.

Now we assume that a statistical model manifold is simply connected. Then, for locally equiaffine connection ∇, there exists a volume element ω defined on $\mathcal{M}$ such that $\nabla\omega = 0$ on $\mathcal{M}$. In a statistical model manifold $\mathcal{M} := \{p(x|\theta) : \int p(x|\theta)\mathrm{d}x = 1,\ p(x|\theta) \geq 0\}$, for an arbitrary $\alpha \in \mathbf{R}$ fixed, a (symmetric) affine connection $\overset{(\alpha)}{\nabla}$ is naturally defined on $\mathcal{M}$ by

$$\overset{(\alpha)}{\Gamma}{}^{i}_{jk} := \overset{(e)}{\Gamma}{}^{i}_{jk} + \frac{1-\alpha}{2} T_{ljk} g^{il}, \quad \overset{(e)}{\Gamma}{}^{i}_{jk} := \overset{(e)}{\Gamma}{}_{l;jk}\, g^{il},$$

$$g_{ij} := \mathrm{E}[\partial_i l \partial_j l], \quad \overset{(e)}{\Gamma}{}_{l;jk} := \mathrm{E}[\partial_l l \partial_j l \partial_k l],$$

$$T_{ljk} := \mathrm{E}[\partial_l l \partial_j l \partial_k l],$$

where g^{ij} is the inverse matrix of the Fisher metric g_{ij}, $l := \log p(x|\theta)$ denotes the log likelihood function and $\mathrm{E}[\cdots]$ denotes the expectation with respect to the observation x (see, e.g., Amari and Nagaoka [3] for details). It is shown that $\overset{(\alpha)}{\nabla}$ is equiaffine for some $\alpha \neq 0$ if and only if it is equiaffine for an arbitrary α [16]. Thus, we shall say that a statistical model manifold $\mathcal{M}$ is *statistically equiaffine* if the above equivalent conditions are satisfied. In the statistically equiaffine manifolds, we may represent the α-parallel volume element ω as

$$\omega = \pi(\theta) d\theta^1 \wedge \cdots \wedge d\theta^d$$

for a certain coordinate $\theta = (\theta^1, \ldots, \theta^d) \in \Theta \subseteq \mathbf{R}^d$. Since π is positive on the whole manifold, we take this as a prior distribution on the parameter space Θ.

Definition 2.2

In a statistically equiaffine manifold, for fixed $\alpha \in \mathbf{R}$, we call the above form of π an *α-parallel prior.*

Note that it could be an improper prior. For properties of α-parallel prior, see Takeuchi and Amari [16]. When $\alpha = 1$, 1-parallel prior is so-called "MLE prior" proposed by Hartigan [10]. We also note that there always exists a $\overset{(0)}{\nabla}$-parallel volume element $\omega \propto \sqrt{g(\theta)} d\theta^1 \wedge \cdots \wedge d\theta^d$, where g is the determinant of the Fisher metric, the invariant volume element in a Riemannian manifold $(\mathcal{M}, g_{ij})$. This prior distribution $\pi \propto \sqrt{g(\theta)}$ is called the Jeffreys prior, well-known in Bayesian statistics. As Jeffreys himself pointed out, it is not necessarily reasonable to adopt the Jeffreys prior as an objective prior in a higher dimensional parametric model. (See, for example, Robert [15] and references therein.)

Now we consider the necessary and sufficient condition that there exists an α-parallel prior ($\alpha \neq 0$) on the statistical model manifold. Hartigan derived the following condition

$$\partial_j \log \pi(\theta) = q_j(\theta), \quad q_j(\theta) = g^{ik} \overset{(e)}{\Gamma}_{k;ij},$$

is necessary and sufficient condition for the existence of the MLE prior. Later, Takeuchi and Amari pointed out the above condition is invariant under reparametrization and they derived more geometrical condition, that is,

$$\partial_i T_j - \partial_j T_i = 0, \ T_i := T_{ikl} g^{kl}$$

is a necessary and sufficient condition that there exists an α-parallel prior ($\alpha \neq 0$).

In the present paper, we take another form of the above condition. Before proceeding, we introduce some notions like the connection 1-form and the curvature 2-form. For simplicity, we take more concise definitions using the coordinate vector $\{\partial_j := \frac{\partial}{\partial \theta^j}\}$. (Note that usual statistical model manifolds are covered with only one coordinate system.) The connection 1-form is defined by ${\omega^k}_j := \Gamma^k_{ij} d\theta^i$, where $\{\Gamma^k_{ij}\}$ are affine connection coefficients and $\{d\theta^i\}$ are dual basis. Note that $\{{\omega^k}_j\}$ are d^2 1-forms. Then, so-called curvature 2-form is defined by $\Omega^k_j := d{\omega^k}_j + {\omega^k}_l \wedge {\omega^l}_j$ Now we define the trace of the curvature form, the sum of the diagonal components.

Definition 2.3

We call $\Xi := \mathrm{Tr}\Omega := \Omega^j_j$ as a *trace 2-form* in the present paper.

It is easily seen that a trace 2-form is invariant under the coordinate transformation. Now we focus on the statistical models. For each α, we denote $\Xi^{(\alpha)}$ as the trace 2-form with respect to the α-connection. Then, using the coordinate representation, we see that

$$\Xi^{(\alpha)} = -\frac{\alpha}{2}\mathrm{d}T,$$

where $T := T_i \mathrm{d}\theta^i$. We obtain the following proposition.

Proposition 2.4
In the statistical model, for fixed α, there exists an α-prior distribution if and only if $\Xi^{(\alpha)} = 0$ on the model.

The above statement is one geometrical reprensentation of Hartigan's condition using differential forms, which yields a coordinate free expression.

3 Trace 2-form of the ARMA(p, q) model

3.1 Information geometry on the spectral densities

When we have time series data subject to an unknown stationary Gaussian process, the estimation of the spectral density is equivalent to that of the original stochastic process. Although we do not enter the general theory of the estimation of spectral densities, the Fisher metric on the parametric families of spectral densities is given below. The Fisher metric of a model specified by a parametric family of spectral densities $\{S(\omega|\theta)|\theta \in \Theta\}$, where θ is a finite-dimensional parameter, is defined by

$$g_{ij} := g\left(\frac{\partial}{\partial\theta^i}, \frac{\partial}{\partial\theta^j}\right) = \int_{-\pi}^{\pi} \frac{\mathrm{d}\omega}{4\pi} \partial_i \log S(\omega|\theta) \partial_j \log S(\omega|\theta) \tag{1}$$

(Amari [2]).

The above metric is defined such that it coincides with the usual Fisher information as the sample size goes to infinity. Other geometrical quantities are defined in the same manner. For our purpose, we only present the following tensor.

$$T_{ijk} := \int \frac{\mathrm{d}\omega}{2\pi} \partial_i \log S(\omega|\theta) \partial_j \log S(\omega|\theta) \partial_k \log S(\omega|\theta)$$

Note that α-connection is determined by g_{ij}, T_{ijk} like i.i.d. cases.

We expect that asymptotic theoretical arguments based on the information geometrical quantities like Fisher metric (1) in the i.i.d. cases are applied to parametric models of stationary Gaussian processes with large length of time series data. Thus, from the viewpoint of the prior selection, it is significant to investigate the existence of the α-parallel prior on the parametric models of spectral densities. In the present paper, as a typical model of stationary time series, we deal with the ARMA(p, q) model. It is already known that there exists the α-parallel prior on the AR(p) models (ARMA$(p, 0)$ model) and MA(q) models (ARMA$(0, q)$ model) because they are $e(m)$-flat (affine connection vanishes). However, as far as the author knows, it has not been investigated yet in the proper ARMA(p, q) models $(p, q > 0)$.

3.2 Geometrical quantities on the ARMA(p,q) model manifold

Here, we briefly summarize the ARMA(p,q) model. It consists of random variables $\{X_t\}$ satisfying

$$X_t = -\sum_{i=1}^{p} a_i X_{t-i} + \sum_{j=0}^{q} b_j W_{t-j},$$

where $\{W_t\}$ is a Gaussian white noise with mean 0 and variance σ^2. For basic notions and notations concerning the ARMA(p,q) model see [6].

The explicit form of the spectral density of the ARMA(p,q) model is

$$S(\omega|a_1,\ldots,a_p,b_1,\ldots,b_q,\sigma^2) = \frac{\sigma^2}{2\pi}\frac{|M_b(z)|^2}{|L_a(z)|^2}, \quad z = e^{i\omega}, \quad . \tag{2}$$

where $L_a(z)$ and $M_b(z)$ are the characteristic polynomials and satisfy

$$L_a(Z)X_t = M_b(Z)W_t,$$

where Z is the shift operator that is defined by $ZX_t = X_{t+1}$ and

$$L_a(z) := \sum_{i=0}^{p} a_i z^{-i}, \quad M_b(z) := \sum_{j=0}^{q} b_j z^{-j} \text{ with } a_0 = b_0 = 1 \ .$$

Now we adopt another coordinate system. Equation $z^p L_a(z) = z^p + a_1 z^{p-1} + \cdots + a_{p-1}z + a_p$ is a polynomial of degree p and has p complex roots, $z_1, z_2, \ldots, z_p$ (Note that $|z_i| < 1$ from the stationarity condition). Since $a_1, a_2, \ldots, a_p$ are all real, it consequently has the conjugate roots. Thus, these roots are rearranged in the order like, $z_1, \ldots, z_s, z_{s+1}, \ldots, z_{2s} \in \mathbf{C}, z_{2s+1}, \ldots, z_{2s+r} \in \mathbf{R}$, $2s + r = p$ and $z_{s+j} = \bar{z}_j (1 \leq j \leq s)$ (for simplicity, we assume that there are no multiple roots). The roots $z_1, z_2, \ldots, z_p$ correspond to the original parameter $a_1, a_2, \ldots, a_p$ one-to-one. Likewise, $w^q M_b(w) = w^q + b_1 w^{q-1} + \cdots + b_{q-1}w + b_q$ is a polynomial of degree q and has q complex roots, $w_1, w_2, \ldots, w_q$. Note that $|w_j| < 1$ from the invertibility condition. The same argument follows. Now we introduce a coordinate system $(\theta^0, \theta^1, \ldots, \theta^{p+q})$ using these roots

$$\theta^0 := \sigma^2, \quad \theta^1 := z_1, \ldots, \theta^p := z_p, \theta^{p+1} := w_1, \ldots, \theta^{p+q} := w_q.$$

The formal complex derivatives are defined by

$$\frac{\partial}{\partial\theta} := \frac{1}{2}\left(\frac{\partial}{\partial x} + i\frac{\partial}{\partial y}\right), \quad \frac{\partial}{\partial\bar{\theta}} := \frac{1}{2}\left(\frac{\partial}{\partial x} - i\frac{\partial}{\partial y}\right),$$

where x and y are both real part and imaginary part of θ. Since the conjugate complex coordinates θ^i and $\bar{\theta}^i$ correspond to x_i and y_i one-to-one, each quantity is evaluated in the original real coordinate if necessary. (See, for example, Gunning and Rossi [7].)

In the coordinate system given above, the Fisher metric on the ARMA(p,q) model g_{IJ} is written in the following way:

$$g_{IJ} = \left(\begin{array}{c|ccc} g_{00} & \cdots & g_{0i} & \cdots \\ \hline \vdots & \cdots & \cdots & \cdots \\ g_{i0} & \vdots & g_{ij} & \vdots \\ \vdots & \vdots & \cdots & \vdots \end{array}\right) \quad \text{and} \quad \left\{\begin{array}{ccc} g_{00} & = & \frac{1}{2(\theta^0)^2} = \frac{1}{2\sigma^4} \\ g_{0i} & = & g_{i0} = 0 \\ g_{ij} & = & \frac{\epsilon_i\epsilon_j}{1-\theta^i\theta^j} \end{array}\right. , \tag{3}$$

where, ϵ_i is defined by

$$\epsilon_i = \begin{cases} +1, & 1 \leq i \leq p, \\ -1, & p+1 \leq i \leq p+q. \end{cases}$$

In the above coordinate, we easily obtain

$$T_{ijk} = \epsilon_i\epsilon_j\epsilon_k \left\{ \frac{2\theta^i}{(1-\theta^i\theta^j)(1-\theta^i\theta^k)} + \frac{2\theta^j}{(1-\theta^j\theta^k)(1-\theta^j\theta^i)} + \frac{2\theta^k}{(1-\theta^k\theta^i)(1-\theta^k\theta^j)} \right\}.$$

3.3 Trace 2-form of the ARMA(p, q) model

Now we investigate the existence of the α-parallel prior for the ARMA(p, q) model by calculating its trace 2-form. We need the explicit form of the inverse matrix of the Fisher information g_{IJ}, which is given by

$$g^{00} = 2(\theta^0)^2,$$
$$g^{0i} = g^{0i} = 0,$$

and

$$g^{mh} = \epsilon_m\epsilon_h \frac{(1-\theta^m\theta^h)\prod_{l\neq h}(1-\theta^l\theta^m)\prod_{l\neq m}(1-\theta^l\theta^h)}{\prod_{l\neq h}(\theta^h-\theta^l)\prod_{l\neq m}(\theta^m-\theta^l)}.$$

By making use of the properties of symmetric polynomials, we obtain the following main result in the present paper.

Theorem 3.1
For the ARMA(p, q) model, trace 2-form is given by

$$\Xi^{(\alpha)} = -4\alpha \sum_{i=1}^{p}\sum_{j=1}^{q} \frac{\mathrm{d}w_j \wedge \mathrm{d}z_i}{(1-z_i w_j)^2}$$

where $|z_i| < 1, \ |w_j| < 1$.

4 Concluding Remarks

We introduced the trace 2-form as a geometrical quantity on the statistical model. From theorem 3.1, we see that the trace 2-form vanishes on the model manifold when $p = 0$ or $q = 0$. It implies that there exist α-parallel priors on the AR(p) model and the MA(q) model. On contrary, there is no α-parallel prior ($\alpha \neq 0$) in the proper ARMA(p, q) model (i.e., $p > 0$ and $q > 0$).

Until now, statistical applications of differential forms have not been considered so much, while other geometrical notions like metric, geodesics, curvature have been investigated considerably [1, 3]. Although our approach to the existence of the α-parallel prior is not outstanding, and equivalent to others already known, but it implies the possibility of applying the differential form to the statistical methods mainly related to the global properties of the statistical model.

Practically more and more complicated statistical models requiring numerical computation appear in various fields and the model manifolds may have nontrivial topology or other global properties. It is known that the differential form is a useful tool to analyse the global properties of the differential manifolds. Thus, further development of the analysis of statistical manifolds based on the differential form are expected to become important.

Acknowledgements

The author was supported by JST PRESTO program and Kakenhi for Young Researchers (B) (No. 20700250).

REFERENCES

[1] S. Amari: *Differential geometrical methods in statistics.* Springer-Verlag, 1985.

[2] S. Amari: Differential geometry of a parametric family of invertible linear systems - Riemannian metric, dual affine connections, and divergence. *Math. Syst. Theory*, **20** (1987), 53–82.

[3] S. Amari and H. Nagaoka: *Methods of Information Geometry.* AMS, Oxford, 2000.

[4] J. Berger and R. Y. Yang: Noninformative priors and Bayesian testing for the AR(1) model. *Econometric Theory*, **10** (1994), 461–482.

[5] J. M. Bernardo: Reference posterior distributions for Bayesian inference. *J. R. Statist. Soc. B* **41**, (1979), 113–147.

[6] P. Brockwell and R. Davis: *Time Series: Theory and Methods.* Springer-Verlag, New York, 1991.

[7] C. Gunning and H. Rossi: *Analytic Functions of Several Complex Variables.* Englewood Cliffs, NJ, Prentice-Hall, 1965.

[8] J. A. Hartigan: Invariant prior densities. *Ann. Math. Statist.*, **35** (1964), 836–845.

[9] J. A. Hartigan: The asymptotically unbiased density. *Ann. Math. Statist.*, **36** (1965), 1137–1152.

[10] J. A. Hartigan: The Maximum Likelihood Prior. *Annals of Statistics*, **26**, no.6, (1998), 2083–2103.

[11] F. Komaki: Estimating method for parametric spectral densities. *J. Time Ser. Anal.*, **20** (1999), 31–50.

[12] F. Komaki: Shrinkage priors for Bayesian prediction. *Ann. Statist.*, **34** (2006), 808–819.

[13] P. C. B. Phillips: To criticize the critics: an objective Bayesian analysis of stochastic trends. *J. Appl. Econ.*, **6** (1991), 333–364.

[14] N. Ravishanker, E. L. Melnick, and C. L. Tsai: Differential geometry of ARMA models. *J. Time Ser. Anal.* **11** (1990), 259–274.

[15] C. P. Robert: *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation.* Springer, New York, 2001.

[16] J. Takeuchi and S. Amari: α-parallel prior and its properties. *IEEE. Trans. Info. Theory,* **51**, no.3 (2005), 1011–1023.

[17] F. Tanaka and F. Komaki: The sectional curvature of AR model manifolds. *Tensor* **64**, (2003), 131–143.

[18] F. Tanaka and F. Komaki: A superharmonic prior for the autoregressive process of the second order. *J. Time Ser. Anal.*, **29** (2008), 444–452.

[19] F. Tanaka: Superharmonic priors for autoregressive models. *Mathematical Engineering Technical Reports* **2009-18**, (2009) 1–20.

[20] A. Zellner: *An Introduction to Bayesian Inference in Econometrics.* J. Wiley, New York, 1971.