KYBERNETIKA — VOLUME 34 (1998), NUMBER 3, PAGES 289–308

# MAXIMUM LIKELIHOOD PRINCIPLE AND $I$-DIVERGENCE: CONTINUOUS TIME OBSERVATIONS[1]

JIŘÍ MICHÁLEK

The paper investigates the relation between maximum likelihood and minimum $I$-divergence estimates of unknown parameters and studies the asymptotic behaviour of the likelihood ratio maximum. Observations are assumed to be done in the continuous time.

## INTRODUCTION

This is a continuation of the previous paper by Michálek [6], dealing with the same problem, but in the case of discrete time observations. Here we consider Gaussian random processes that differ in the mean value. We also investigate the autoregressive Gaussian processes. It is shown that under the stationarity assumption the role of $I$-divergence is substituted by the corresponding asymptotic $I$-divergence rate.

## 1. A SIMPLE REGRESSION MODEL FOR GAUSSIAN PROCESSES

Let on the interval $\langle 0,T\rangle$ be given a real Gaussian process $\{x(t)\}$ which satisfies the following conditions

$$\mathsf{E}_\alpha\{x(t)\} = \alpha\,\varphi(t), \quad t \in \langle 0,T\rangle$$
$$\operatorname{cov}_\alpha\{x(s),\, x(t)\} = \operatorname{cov}_0\{x(s),\, x(t)\},$$

where $\varphi(\cdot)$ is a known function and $\alpha$ is a real parameter which should be estimated. The coincidence of covariance functions means that

$$\mathsf{E}_\alpha\{x(s)\,x(t)\} = \mathsf{E}_0\{x(s)\,x(t)\} + \alpha^2\,\varphi(s)\,\varphi(t).$$

For more details about this statistical model we refer to Hájek [3]. We assume, of course, that $\varphi(\cdot)$ is not the identical zero on $\langle 0,T\rangle$.

On the basis of results presented in Hájek [3] we know that in the case of the existence of a random variable $v \in \mathcal{W}$, where $\mathcal{W}$ is the closed linear hull over the

[1] This work was supported by the Grant Agency of the Czech Republic under Grant 201/96/415.

values $x(t)$, $t \in \langle 0, T \rangle$ with respect to the probability measure $P_1$, i.e. for $\alpha = 1$, the regular case occurs, i.e. $P_\alpha \sim P_0$ for each $\alpha \in R_1$ and there exists the best linear unbiased estimate of $\alpha$. Under such a situation the variable $v$ satisfies the relation

$$\mathsf{E}_0\{x(t)\, v\} = \varphi(t), \quad t \in \langle 0, T \rangle,$$

and the unbiased estimate with minimal dispersion is given by

$$u = \frac{v}{\mathsf{E}_0\{v^2\}}.$$

The variable $u$ is also a sufficient statistic for the system $\{P_\alpha\}$.

Using the existence of $v$, the corresponding Radon–Nikodym derivative has the form

$$\frac{\mathrm{d}P_\alpha}{\mathrm{d}P_0} = \exp\left\{\alpha\, v - \frac{1}{2}\alpha^2\, \mathsf{E}_0\{v^2\}\right\}.$$

We will investigate how the relation between the MLE of $\alpha$ and the $I$-divergence looks in this case. It is easy to verify that the MLE of $\alpha$ is given by

$$\hat{\alpha} = \frac{v}{\|v\|_0^2} = u.$$

Further, we can easily calculate the $I$-divergence between measures $P_\alpha$ and $P_0$, namely

$$\begin{aligned} I(P_\alpha : P_0) &= \mathsf{E}_\alpha\left\{\ln \frac{\mathrm{d}P_\alpha}{\mathrm{d}P_0}\right\} = \mathsf{E}_\alpha\left\{\alpha\, v - \frac{1}{2}\alpha^2\, \|v\|_0^2\right\} \\ &= \alpha^2\|v\|_0^2 - \frac{1}{2}\alpha^2\|v\|_0^2 = \frac{1}{2}\alpha^2\|v\|_0^2, \end{aligned}$$

because $\hat{\alpha}$ is an unbiased estimate. Then

$$\begin{aligned} \max_\alpha \ln \frac{\mathrm{d}P_\alpha}{\mathrm{d}P_0} &= \max_\alpha \left\{\alpha\, v - \frac{1}{2}\alpha^2\|v\|_0^2\right\} \\ &= \frac{1}{2}(\hat{\alpha})^2\, \|v\|_0^2 = I(P_{\hat{\alpha}} : P_0). \end{aligned}$$

In this way we have proved that in this model the relation studied in Michálek [6] and connecting the likelihood ratio maximum and the $I$-divergence is valid, too.

We can immediately utilize this relation in constructing a test based on the likelihood ratio.

Let us consider a real Gaussian process $\{x(t), t \in \langle 0, T \rangle\}$ satisfying the above regression model and we want to test the hypothesis $H_0 : \alpha \in \langle a, b \rangle$ against the alternative hypothesis $H_1 : \alpha \notin \langle a, b \rangle$. The test will be based on the statistic $T(\cdot)$ given as the ratio

$$T(x(\cdot)) = \frac{\sup_{\alpha \in \langle a,b \rangle} \frac{\mathrm{d}P_\alpha}{\mathrm{d}P_0}\{x(\cdot)\}}{\sup_{\alpha \in R_1} \frac{\mathrm{d}P_\alpha}{\mathrm{d}P_0}\{x(\cdot)\}}.$$

From the previous results about the likelihood ratio we can expect the relation

$$\ln T(x(\cdot)) = I(P_{\hat{\alpha}_0} : P_0) - I(P_{\hat{\alpha}} : P_0),$$

where $\hat{\alpha}_0$ is a local MLE obtained by maximizing over the interval $\langle a, b\rangle$ only while $\hat{\alpha}$ is a global MLE. We must verify whether really

$$\sup_{\alpha\in\langle a,b\rangle} \ln \frac{\mathrm{d}P_\alpha}{\mathrm{d}P_0}\{x(\cdot)\} = I(P_{\hat{\alpha}_0} : P_0).$$

One can easily find out that

$$\arg\sup_{\alpha\in\langle a,b\rangle} \left\{\alpha\, v - \frac{1}{2}\alpha^2 \|v\|_0^2\right\} \begin{array}{lll} = \hat{\alpha} & \text{for} & \hat{\alpha}\in\langle a,b\rangle \\ = a & \text{for} & \hat{\alpha} < a \\ = b & \text{for} & \hat{\alpha} > b. \end{array}$$

Let us define a local MLE, namely

$$\begin{array}{lll} \hat{\alpha}_0 = \hat{\alpha} & \text{for} & \hat{\alpha}\in\langle a,b\rangle \\ \hat{\alpha}_0 = a & \text{for} & \hat{\alpha} < a \\ \hat{\alpha}_0 = b & \text{for} & \hat{\alpha} > b \end{array}$$

then

$$\begin{aligned} &\sup_{\alpha\in\langle a,b\rangle} \ln \frac{\mathrm{d}P_\alpha}{\mathrm{d}P_0}\{x(\cdot)\} = \sup_{\alpha\in\langle a,b\rangle}\left\{\alpha\, v - \frac{1}{2}\alpha^2\|v\|_0^2\right\} \\ = \ &\max_{\alpha\in\{a,b,\hat{\alpha}\}}\left\{\alpha\, v - \frac{1}{2}\alpha^2\|v\|_0^2\right\} = \hat{\alpha}_0\, v - \frac{1}{2}\hat{\alpha}_0\|v\|_0^2. \end{aligned}$$

Using this relation we can write together

$$T(x(\cdot)) = e^{-\frac{\|v\|_0^2}{2}(\hat{\alpha}_0-\hat{\alpha})^2}.$$

As $\|v\|_0^2 = \frac{1}{\|\hat{\alpha}\|_0^2}$, see Hájek [3], finally

$$T(x(\cdot)) = e^{-\frac{1}{2}\frac{(\hat{\alpha}-\hat{\alpha}_0)^2}{\|\hat{\alpha}\|_0^2}}.$$

Then the critical domain of the proposed test has the form

$$\{x(\cdot) : T(x(\cdot)) < K\},$$

where $K \leq 1$ as follows from the character of the test because the hypothesis $H_0$ is not rejected if $T(x(\cdot))$ is close to one. The value of $K$ will be determined from the behaviour of the first kind error. We demand

$$\sup_{\alpha\in\langle a,b\rangle} \mathsf{P}\{x(\cdot) : T(x(\cdot)) < K_\gamma\} = \gamma$$

where $\gamma \in (0,1)$ is given beforehand. The previous inequality is equivalent to

$$\sup_{\alpha\in\langle a,b\rangle} \mathsf{P}\{x(\cdot) : \ln T(x(\cdot)) < \ln K_\gamma\} = \gamma.$$

To find the right value of $K_\gamma$, it is necessary to know the distribution of $\ln T(x(\cdot))$ under $H_0$. Surely, for each $s < 0$

$$\begin{aligned}\mathsf{P}\{\ln T(x(\cdot)) < s\} &= \mathsf{P}\{\ln T(x(\cdot)) < s \,|\, \hat\alpha \in \langle a,b\rangle)\, \mathsf{P}(\hat\alpha \in \langle a,b\rangle\} \\ &+\mathsf{P}\{\ln T(x(\cdot)) < s \,|\, \hat\alpha > b\}\, \mathsf{P}\{\hat\alpha > b\} + \mathsf{P}\{\ln T(x(\cdot)) < s \,|\, \hat\alpha < a\}\, \mathsf{P}\{\hat\alpha < a\}\end{aligned}$$

and for $s = 0$

$$\mathsf{P}\{\ln T(x(\cdot)) = 0\} = \mathsf{P}\{\alpha \in \langle a,b\rangle\}.$$

If we put $\|\hat\alpha\|_0^2 = 1$ for simplicity, then

$$\begin{aligned}F_\alpha(s) &= \mathsf{P}\{\ln T(x(\cdot)) < s\} = \Phi(a-\alpha)\left[\Phi(a-\alpha-\sqrt{2|s|}) + \Phi(\alpha-a-\sqrt{2|s|})\right] \\ &+ \Phi(\alpha-b)\left[\Phi(\alpha-b-\sqrt{2|s|}) + \Phi(b-\alpha-\sqrt{2|s|})\right],\end{aligned}$$

for each $s \le 0$ and

$$\mathsf{P}\{\ln T(x(\cdot)) = 0\} = \Phi(b-\alpha) - \Phi(a-\alpha),$$

where $\Phi(\cdot)$ is the distribution function of $N(0,1)$. The distribution function $F_\alpha(s)$ is derived from the fact that $\hat\alpha$ is Gaussian with mean $\alpha$ and unit dispersion. To find the value $K_\gamma$ defining the critical region of the proposed test it is necessary to study the behaviour of $F_\alpha(s)$ under the hypothesis $H_0$ $\alpha \in \langle a,b\rangle$. It is easy to check that the function $F_\alpha(s)$ is symmetric at the point $\frac{a+b}{2}$, i.e. if $\alpha_1 = \frac{a+b}{2} + x$, $\alpha_2 = \frac{a+b}{2} - x$ then

$$F_{\alpha_1}(s) = F_{\alpha_2}(s)$$

for each $s \le 0$ and each $x > 0$. After calculating the derivative $\frac{\partial}{\partial\alpha}F_\alpha(s)$ we can find out that it is vanishing at the point $\frac{a+b}{2}$ and positive for each $\alpha_1 = \frac{a+b}{2} + x$, $x > 0$, $s < -\frac{\delta^2}{2}$, and similarly negative for each $\alpha_2 = \frac{a+b}{2} - x$, $x > 0$, $s < -\frac{\delta^2}{2}$, where $\delta = \frac{b-a}{2}$. It means the function $F_\alpha(s)$ attains its minimum at the point $\frac{a+b}{2}$ for each $s < -\frac{\delta^2}{2}$ and its maximum is at the points $a$, $b$ satisfying the condition

$$F_a(s) = F_b(s)$$

for each $s \le 0$.

We immediately see that

$$F_a(s) = \left[\Phi\left(-\sqrt{2|s|}\right)\right] + \Phi(a-b)\left[\Phi\left(a-b-\sqrt{2|s|}\right) + \Phi(b-a) - \sqrt{2|s|}\right].$$

For $s = 0$ we have

$$F_a(0) = \frac{1}{2} + \Phi(a-b) > \frac{1}{2}. \tag{1}$$

As we demand $\sup_{\alpha\in\langle a,b\rangle} \mathsf{P}\{\ln T(x(\cdot)) < \ln K_\gamma\} = \gamma$, where $\gamma$ is close to zero, we see that
$$F_a(\ln K_\gamma) = F_b(\ln K_\gamma) \le \gamma$$
must be valid, too. But with respect to the inequality (1) this cannot be satisfied for $K_\gamma$ close to one. Hence, the restriction $s < -\frac{\delta^2}{2}$ in the behaviour of the derivative $\frac{\partial F_\alpha(s)}{\partial\alpha}$ is irrelevant because for $s = -\frac{\delta^2}{2}$ we have
$$F_a\left(-\frac{\delta^2}{2}\right) = F_b\left(-\frac{\delta^2}{2}\right) = \Phi(-2\delta)\left[\Phi(-3\delta) + \Phi(-\delta)\right] + \frac{1}{2} > \frac{1}{2}.$$
It follows from here that the critical value $K_\gamma$ for $\gamma < \frac{1}{2}$ must satisfy
$$K_\gamma < e^{-\frac{\delta^2}{2}}.$$
On the basis of the properties of $\frac{\partial F_\alpha(s)}{\partial\alpha}$ for $s < -\frac{\delta^2}{2}$ we can assert that
$$\sup_{\alpha\in\langle a,b\rangle} \mathsf{P}\left\{\ln T(x(\cdot)) < s\right\} = F_a(s) = F_b(s).$$
This fact ensures the existence of a suitable critical value $K_\gamma$ and its uniqueness.

As for the second kind error we can say that in the alternative set $\alpha \notin \langle a,b\rangle$
$$\inf_{\alpha\notin\langle a,b\rangle} F_\alpha(\ln K_\gamma) = F_a(\ln K_\gamma) = \gamma.$$

## 2. GAUSSIAN STATIONARY AUTOREGRESSIVE PROCESSES

In this part we will deal with Gaussian stationary autoregressive processes, which are defined on $\langle 0,T\rangle$ with the zero mean value for simplicity. These stochastic processes belong to the most important cases frequently applied in practice. In order to calculate the maximum value of a likelihood ratio we need to know the corresponding Radon–Nikodym derivatives. Here we mainly use the results given in Hájek [4].

We will start with a general form of a Gaussian autoregressive process, which is given by solving a linear differential equation with constant coefficients with Gaussian white noise having dispersion $\sigma^2$ on the right hand side. The equation coefficients must satisfy the condition of stability, i.e. all zeros of the characteristic polynomial are located in the left half-plane. The corresponding spectral density function has the form
$$\varphi(\lambda) = \frac{1}{2\pi}\frac{\sigma^2}{\left|\sum_{k=0}^{n} a_{n-k}(i\lambda)^k\right|^2}$$
where $a_0 = 1$. A dominating measure $Q$ defined on $\langle 0,T\rangle$ is given by a Gaussian process $\{y(t),\ t\in\langle 0,T\rangle\}$ with a zero mean and with independent increments of its $(n-1)$st derivative satisfying
$$\mathsf{E}\left\{|\mathrm{d}\,y^{(n-1)}(t)|^2\right\} = \sigma^2\,\mathrm{d}t$$

with the initial condition $(y(0),\ y'(0),\dots,y^{(n-1)}(0))$ to be independent on $y^{(n-1)}(t)-y^{(n-1)}(0)$. For more details see Hájek [4]. Then the appropriate Radon–Nikodym derivative is given by the following formula

$$\begin{aligned}\frac{\mathrm{d}P_a}{\mathrm{d}Q}(x(\cdot)) &= |D_{jk}|^{1/2}\exp\left\{\frac{1}{2}\,a_1\,T-\frac{1}{2}\sum_{k=0}^{n-1}A_{n-k}\int_0^T|x^{(k)}(t)|^2\,\mathrm{d}t\right.\\ &\left.-\frac{1}{4}\sum_{\substack{j+k\\ \text{even}}}^{n-1}\sum^{n-1}\left[x^{(j)}(t)\,x^{(k)}(t)+x^{(j)}(0)\,x^{(k)}(0)\right]D_{jk}\right\},\end{aligned}$$

where $x^{(j)}(s)$ is the $j$th derivative of $x(\cdot)$ at $s$ and $D_{jk},\ j,\ k=0,1,\dots,n-1$ are defined as

$$\begin{aligned}D_{jk} &= 2\sum_{i=\max(0,j+k+1-n)}^{\min(j,k)}(-1)^{j-i}\,a_{n-i}\,a_{n+i-j-k-1}\quad\text{for } j+k \text{ even}\\ D_{jk} &= 0\qquad\text{for } j+k \text{ odd}.\end{aligned}$$

We have put $\sigma^2=1$ for simplicity. In other words the matrix $\{D_{jk}\}_{j,k=0}^{n-1}$ is the inverse matrix to the covariance matrix of $x(0),\ x'(0),\dots,x^{(n-1)}(0)$. The coefficients $A_{n-k}$ are defined in the unique way by

$$\left|\sum_{k=0}^{n}a_{n-k}(i\lambda)^k\right|^2=\sum_{k=0}^{n}A_{n-k}\,\lambda^{2k},$$

which gives

$$A_{n-k}=\sum_{j=0}^{\min(k,n-k)}a_{n-k-j}\,a_{n-k+j}\,(-1)^j.$$

Thanks to the stability condition the polynomial $\sum_{k=0}^{n}a_{n-k}\,z^k$ has all the roots situated in the left half-plane. This condition together with $A_0=1$ determines a one-to-one correspondence between $\{a_{n-k}\}_{k=0}^{n}$ and $\{A_{n-k}\}_{k=0}^{n}$. At the first sight we see that any explicit formula for the MLE is almost impossible in a general case because of the presence of the determinant $|D_{jk}|$. For a better orientation we will now concentrate ourselves to the simpliest cases $n=1$ and $n=2$.

First, we will investigate a Gaussian autoregressive stationary process of the first order. Such a process possesses the covariance function of the form

$$R(t)=C\,e^{-\alpha|t|},$$

where $C=R(0)>0,\ \alpha>0$. Its spectral density function is then

$$\varphi(\lambda)=\frac{2C\,\alpha}{2\pi(\lambda^2+\alpha^2)},$$

for detail see e. g. Anděl [1]. To express the likelihood ratio we must find out the Radon–Nikodym derivative

$$\frac{\mathrm{d}P(x(\cdot),\beta,D)}{\mathrm{d}P(x(\cdot),\alpha,C)}$$

on $\langle 0,T\rangle$. In order to keep the measures $P(x(\cdot),\beta,D)$ and $P(x(\cdot),\alpha,C)$ equivalent, these parameters must satisfy some conditions. Here we can refer to Pisarenko [8] and Hájek [3].

Let us rewrite the spectral density function into the following form

$$\varphi(\lambda)=\frac{1}{2\pi}\,\frac{1}{|a_0(i\lambda)^2+a_1|}=\frac{1}{2\pi(a_0^2\lambda+a_1^2)}.$$

Then we obtain

$$C=\frac{1}{2a_0\,a_1},\qquad \alpha=\frac{a_1}{a_0}$$

and, vice versa,

$$a_0^2=\frac{1}{2\alpha C},\qquad a_1^2=\frac{\alpha}{2C}.$$

A necessary and sufficient condition for equivalence of the corresponding probability measures is given by

$$a_0=b_0\iff a_0^2=b_0^2\iff \alpha C=\beta D.$$

As we can put $C=1$, we obtain $D=\frac{\alpha}{\beta}$. Using the coefficients $(a_0,a_1)$ and $(b_0,b_1)$ we can express the Radon–Nikodym derivative as follows $(a_0=b_0)$

$$\begin{aligned}\frac{\mathrm{d}P(x(\cdot),b_0,b_1)}{\mathrm{d}P(x(\cdot),a_0,a_9)}&=\left(\frac{b_1}{a_1}\right)^{\frac12}\exp\left\{\frac12\,\frac{b_1T}{b_0}-\frac12\,\frac{a_1T}{a_0}\right.\\&\qquad\left.-\frac12 b_1^2\int_0^T x^2(t)\,\mathrm{d}t+\frac12 a_1^2\int_0^T x^2(t)\,\mathrm{d}t\right\}\\&=\left(\frac{\beta}{\alpha}\right)^{\frac12}\exp\left\{\frac{T\beta}{2}-\frac{T\alpha}{2}-\frac12\,\frac{\beta^2}{\alpha}\int_0^T x^2(t)\,\mathrm{d}t+\frac{\alpha}{4}\int_0^T x^2(t)\,\mathrm{d}t\right\},\end{aligned}$$

when we used the parameters $\alpha$, $\beta$ with $C=1$. Then, easily

$$\ln\frac{\mathrm{d}P(x(\cdot),\beta,D)}{\mathrm{d}P(x(\cdot),\alpha,1)}=\frac12\ln\frac{\beta}{2}+\frac{T}{2}(\beta-\alpha)-\frac{\beta^2}{4\alpha}\int_0^T x^2(t)\,\mathrm{d}t+\frac{\alpha}{4}\int_0^T x^2(t)\,\mathrm{d}t.$$

Now,

$$\frac{\partial}{\partial\beta}=\frac{1}{2\beta}+\frac{T}{2}-\frac{\beta}{2\alpha}\int_0^T x^2(t)\,\mathrm{d}t.$$

This fact gives us the MLE of the parameter $\beta$, namely

$$\frac{1}{\hat\beta}+T-\frac{\hat\beta}{\alpha}\int_0^T x^2(t)\,\mathrm{d}t=0. \tag{2}$$

Let us substitute this relation into the logarithm of the likelihood ratio. In this way we come to

$$\max_{\beta>0} \ln \frac{\mathrm{d}P(x(\cdot),\beta,D)}{\mathrm{d}P(x(\cdot),\alpha,1)} = \frac{1}{2}\ln\frac{\hat{\beta}}{\alpha} + \frac{T}{2}(\hat{\beta}-\alpha) + \frac{1}{4\alpha}(\alpha^2-(\hat{\beta})^2)\int_0^T x^2(t)\,\mathrm{d}t.$$

If we express

$$\int_0^T x^2(t)\,\mathrm{d}t = \frac{\alpha}{(\hat{\beta})^2} + \frac{\alpha T}{\hat{\beta}}$$

using (2) then

$$\max_{\beta>0} \ln \frac{\mathrm{d}P(x(\cdot),\beta,D)}{\mathrm{d}P(x(\cdot),\alpha,1)} = \frac{1}{2}\ln\frac{\hat{\beta}}{\alpha} + \frac{T}{2}(\hat{\beta}-\alpha) + \frac{\alpha^2}{4(\hat{\beta})^2} - \frac{1}{4} + \frac{T}{4}\frac{(\alpha^2-(\hat{\beta})^2)}{\hat{\beta}}.$$

This result can be easily rewritten into the form

$$\max_{\beta>0} \ln \frac{\mathrm{d}P(x(\cdot),\beta,D)}{\mathrm{d}P(x(\cdot),\alpha,1)} = \frac{1}{4}\left(\frac{\alpha^2}{(\hat{\beta})^2} - \ln\frac{\alpha^2}{(\hat{\beta})^2} - 1\right) + \frac{T}{4}\frac{(\hat{\beta}-\alpha)^2}{\hat{\beta}}.$$

At the first sight we see that this maximum is formed by two parts, namely

$$\frac{\alpha^2}{(\hat{\beta})^2} - \ln\frac{\alpha^2}{(\hat{\beta})^2} - 1,$$

which is the expression of $I$-divergence between two Gaussian random variables with dispersion $\alpha^2$, $(\hat{\beta})^2$, respectively. The other part can be expressed via an integral because

$$\frac{1}{4\pi}\int_{-\infty}^{\infty}\left(\frac{\varphi_\beta}{\varphi_\alpha} - \ln\frac{\varphi_\beta}{\varphi_\alpha} - 1\right)\mathrm{d}\lambda = \frac{(\beta-\alpha)^2}{\beta}, \tag{3}$$

when

$$\varphi_\beta(\lambda) = \frac{D\beta}{\pi(\lambda^2+\beta^2)}, \qquad \varphi_\alpha(\lambda) = \frac{C\alpha}{\pi(\lambda^2+\alpha^2)}.$$

As we have $D\beta = C\alpha$ for the equivalence of the corresponding probability measure, then the left hand side of (3) is nothing else but the asymptotic $I$-divergence rate between two Gaussian stationary processes of the first order, for more details see Michálek [5]. In this way we have just proved a close connection between the MLE and the $I$-divergence again. The contribution

$$\frac{\alpha^2}{\beta^2} - \ln\frac{\alpha^2}{\beta^2} - 1$$

can be also understood as the asymptotic $I$-divergence rate of two Gaussian white noises having dispersions $\alpha^2$, $\beta^2$ respectively, which form inputs into a linear filter whose output is created by a stationary autoregressive process of the first order. We can immediately state the following

**Theorem 1.** Let on $\langle 0,T\rangle$ be given a stationary Gaussian autoregressive process of the first order and let $\hat{\beta}_T$ be the corresponding MLE. Then

$$\lim_{T\to\infty} \frac{1}{T} \sup_{\beta>0} \ln \frac{\mathrm{d}P(x(\cdot),\beta,D)}{\mathrm{d}P(x(\cdot),\alpha,1)} = \frac{1}{4\pi}\int_{-\infty}^{\infty} \left(\frac{\varphi_{\beta_0}}{\varphi_\alpha} - \ln\frac{\varphi_{\beta_0}}{\varphi_\alpha} - 1\right) \mathrm{d}\lambda,$$

where $\beta_0$ is a true parameter, i.e. $\hat{\beta}_T \xrightarrow[n\to\infty]{} \beta_0$ a.s.

P r o o f. It immediately follows from the previous text and the properties of stationary processes. □

In the next part we will concentrate to the case of autoregressive process of the second order. Its spectral density function is of the form

$$\begin{aligned}\varphi(\lambda) &= \frac{1}{2\pi}\frac{\sigma^2}{|(i\lambda)^2+(\alpha_1+\alpha_2)(i\lambda)+\alpha_1\alpha_2|^2} \\ &= \frac{1}{2\pi}\frac{\sigma^2}{(\lambda^2+\alpha_1^2)(\lambda^2+\alpha_2^2)}.\end{aligned}$$

Here we have three parameters $\alpha_1$, $\alpha_2$, $\sigma^2$ that in the unique way determine this density function. As we want to consider a class of mutually equivalent Gaussian probability measures we are obliged to fix $\sigma^2$ because for different values of this parameter the corresponding measures are orthogonal, see Hájek [4]. Without loss of generality we put $\sigma^2 = 1$. For future purposes it is reasonable to introduce new parameters, namely

$$a_1 = \alpha_1 + \alpha_2, \qquad a_2 = \alpha_1\,\alpha_2.$$

Let us start with the Radon–Nikodym derivative of the measure $P(x(\cdot),a_1,a_2)$ on $\langle 0,T\rangle$ with respect to the measure $Q(x(\cdot))$, which was described at the beginning of this part (for more details see Hájek [4]). Then the corresponding derivative has the form

$$\begin{aligned}&\frac{\mathrm{d}P(x(\cdot),a_1,a_2)}{\mathrm{d}Q(x(\cdot))} = 2a_1\,a_2^{\frac{1}{2}}\exp\left\{\frac{a_1}{2}T\right\} \\ &\times\exp\left\{-\frac{1}{2}(a_1^2-2a_2)\int_0^T |\dot{x}(t)|^2\,\mathrm{d}t - \frac{1}{2}a_2^2\int_0^T |x(t)|^2\,\mathrm{d}t\right. \\ &\left.-\frac{1}{2}a_1\left((\dot{x}(0))^2+(\dot{x}(T))^2\right) - \frac{1}{2}a_1\,a_2(x^2(0)+x^2(T))\right\}.\end{aligned}$$

From this we can derive the set of equations for the MLE of the parameters $a_1$, $a_2$, namely

$$\begin{aligned}&\frac{1}{a_1}+\frac{T}{2}-a_1\,\Phi_1-\frac{1}{2}\Phi_3-\frac{a_2}{2}\Phi_4=0 \\ &\frac{1}{2a_1}+\Phi_1-a_2\,\Phi_2-\frac{a_1}{2}\Phi_4=0,\end{aligned}$$

where

$$\Phi_1 = \int_0^T |\dot{x}(t)|^2 \, \mathrm{d}t, \qquad \Phi_2 = \int_0^T |x(t)|^2 \, \mathrm{d}t,$$
$$\Phi_3 = (\dot{x}(0))^2 + (\dot{x}(T))^2, \qquad \Phi_4 = x^2(0) + x^2(T).$$

In next we will utilize these equations to eliminate the terms $\Phi_3$, $\Phi_4$ from the expression for the Radon–Nikodym derivative.

Multiplying the first equation by $a_1$ we obtain

$$1 + \frac{a_1}{2} T - a_1^2 T - a_1^2 \Phi_1 - \frac{1}{2} a_1 \Phi_3 - \frac{a_1 a_2}{2} \Phi_4 = 0. \tag{a}$$

Similarly, the other equation can be multiplied by $a_2$ and we get to

$$\frac{1}{2} + a_2 \Phi_1 - a_2^2 \Phi_2 - \frac{a_1 a_2}{2} \Phi_4 = 0. \tag{b}$$

Now, the likelihood ratio maximum is given by

$$\max_{(b_1, b_2)} \ln \frac{\mathrm{d}P(x(\cdot), b_1, b_2)}{\mathrm{d}P(x(\cdot), a_1, a_2)} = \ln \frac{\hat{b}_1 \hat{b}_2^{\frac{1}{2}}}{a_1 a_2^{\frac{1}{2}}} + \frac{(\hat{b}_1 - a_1)^T}{2}$$
$$- \frac{(\hat{b}_1^2 - 2\hat{b}_2 - a_2^2 + 2a_2)}{2} \Phi_1 - \frac{\hat{b}_2^2 - a_2^2}{2} \Phi_2 - \frac{\hat{b}_1 - a_1}{2} \Phi_3 - \frac{\hat{b}_1 \hat{b}_2 - a_1 a_2}{2} \Phi_4,$$

where $\hat{b}_1$, $\hat{b}_2$ are the MLE's of unknown parameters $b_1$, $b_2$.

At this moment we can use the equations (a), (b), which the MLE's must satisfy. After simple but tedious calculations we obtain the formula

$$\max_{(b_1, b_2)} \ln \frac{\mathrm{d}P(x(\cdot), b_1, b_2)}{\mathrm{d}P(x(\cdot), a_1, a_2)} = \frac{1}{2} \left( \frac{a_1}{\hat{b}_1} - \ln \frac{a_1}{\hat{b}_1} - 1 \right)$$
$$+ \frac{1}{2} \left( \frac{a_1 a_2}{\hat{b}_1 \hat{b}_2} - \ln \frac{a_1 a_2}{\hat{b}_1 \hat{b}_2} - 1 \right) + \Phi_1 \left( \frac{(\hat{b}_1 - a_1)^2}{2} + \frac{(\hat{b}_2 - a_2)(\hat{b}_1 - a_1)}{\hat{b}_1} \right)$$
$$+ \Phi_2 (\hat{b}_2 - a_2) \left( \frac{a_1 \hat{b}_2}{\hat{b}_1} + \frac{a_2 + \hat{b}_2}{2} \right).$$

In order to compare this result with the $I$-divergence of corresponding Gaussian measures, it is necessary to calculate this $I$-divergence, i.e., the integral

$$\int \ln \frac{\mathrm{d}P(x(\cdot), b_1, b_2)}{\mathrm{d}P(x(\cdot), a_1, a_2)} \, \mathrm{d}P(x(\cdot), b_1, b_2).$$

It is easy to find that

$$\mathsf{E}_{(b_1 b_2)}\{\Phi_1\} = \int_0^T \mathsf{E}\{|\dot{x}(t)|^2\} \, \mathrm{d}t = \frac{T}{2\pi} \int_{-\infty}^{\infty} \frac{\lambda^2 \, \mathrm{d}\lambda}{(\lambda^2 + \beta_1^2)(\lambda^2 + \beta_2^2)}$$

$$
\begin{aligned}
&= \frac{T}{2(\beta_1+\beta_2)} = \frac{T}{2b_1},\\
\mathsf{E}_{(b_1 b_2)}\{\Phi_2\} &= \frac{T}{2\pi}\int_{-\infty}^{\infty}\frac{\mathrm{d}\lambda}{(\lambda^2+\beta_1^2)(\lambda^2+\beta_2^2)} = \frac{T}{2(\beta_1+\beta_2)\beta_1\beta_2} = \frac{T}{2b_1 b_2},\\
\mathsf{E}_{(b_1 b_2)}\{\Phi_3\} &= \mathsf{E}_{(b_1 b_2)}\{|\dot{x}(0)|^2+|\dot{x}(T|^2\} = \frac{1}{b_1},\\
\mathsf{E}_{(b_1 b_2)}\{\Phi_4\} &= \mathsf{E}_{(b_1 b_2)}\{|x(0)|^2+|x(T)|^2\} = \frac{1}{b_1 b_2}.
\end{aligned}
$$

Then

$$
\begin{aligned}
I(P_{(b_1 b_2)} : P_{(a_1 a_2)}) &= \ln\frac{b_1 b_1^{\frac{1}{2}}}{a_1 a_2^{\frac{1}{2}}} + \frac{T(b_1-a_1)}{2} - \frac{T(b_1^2-2b_2-a_1^2+2a_2)}{4b_1}\\
&-\frac{b_1^2-a_2^2}{4b_1 b_2}T - \frac{b_1-a_1}{2b_1} - \frac{b_1 b_2 - a_1 a_2}{2b_1 b_2}\\
&= \frac{1}{2}\left(\frac{a_1}{b_1}-\ln\frac{a_1}{b_1}-1\right)+\frac{1}{2}\left(\frac{a_1 a_2}{b_1 b_2}-\ln\frac{a_1 a_1}{b_1 b_1}-1\right)+\frac{T}{4}\left(\frac{(b_1-a_1)^2}{b_1}+\frac{(b_2-a_2)^2}{b_1 b_2}\right).
\end{aligned}
$$

From this result, as a byproduct, we obtain the existence of the limit

$$
\lim_{T\to\infty}\frac{1}{T}I(P_{(b_1 b_2)} : P_{(a_1 a_2)}) = \frac{1}{4}\left[\frac{(b_1-a_1)^2}{b_1 b_1}+\frac{(b_2-a_2)^2}{b_1 b_2}\right],
$$

which is precisely the integral

$$
\frac{1}{4\pi}\int_{-\infty}^{\infty}\left(\frac{\varphi_{(b_1 b_2)}}{\varphi_{(a_1 a_2)}} - \ln\frac{\varphi_{(b_1 b_2)}}{\varphi_{(a_1 a_2)}} - 1\right)\mathrm{d}\lambda = \overline{I}(P_{(b_1 b_2)} : P_{(a_1 a_2)}),
$$

the asymptotic $I$-divergence rate. When we return to the likelihood ratio maximum and the $I$-divergence, we see that their expressions are equal in two terms

$$
\frac{1}{2}\left(\frac{a_1}{b_1}-\ln\frac{a_1}{b_1}-1\right) \quad\text{and}\quad \frac{1}{2}\left(\frac{a_1 a_2}{b_1 b_1}-\ln\frac{a_1 a_2}{b_1 b_2}-1\right)
$$

respectively. In order to achieve the equality between these expressions, it suffices to establish the relation

$$
\begin{aligned}
&\left[\frac{(\hat{b}_1-a_1)^2}{2}+\frac{(\hat{b}_2-a_2)(\hat{b}_1-a_1)}{\hat{b}_1}\right]\Phi_1 + (\hat{b}_2-a_2)\left[\frac{a_1\hat{b}_2}{\hat{b}_1}+\frac{a_2+\hat{b}_2}{2}\right]\Phi_2\\
&= \frac{T}{4}\left[\frac{(\hat{b}_1-a_1)^2}{\hat{b}_1}+\frac{(\hat{b}_2-a_2)^2}{\hat{b}_1\hat{b}_2}\right].
\end{aligned}
$$

The left hand side of this equality can be transformed into a more acceptable form, namely

$$
\frac{(\hat{b}_1-a_1)^2}{2}\Phi_1+\frac{a_1 a_2}{\hat{b}_1}(\Phi_1-\hat{b}_2\Phi_2)+\frac{a_1\hat{b}_2}{\hat{b}_1}(\hat{b}_2\Phi_2-\Phi-1)+(\hat{b}_2-a_2)\Phi_1+\frac{a_2^2-\hat{b}_2^2}{2}\Phi_2.
$$

Comparing this with the right hand side of the previous equality, we see that the equality holds if

$$\hat{b}_2\,\Phi_2 = \Phi_1,$$

in which case the left hand side equals

$$\frac{(\hat{b}_1 - a_1)^2}{2}\,\Phi_1 + \frac{(\hat{b}_2 - a_2^2)^2}{2}\,\Phi_2,$$

i.e. coincides with the right hand side.

Look in more detail when the equality $\Phi_1 = T/2\,\hat{b}_1$ takes place. It is evident that if the MLE's $\hat{b}_1$, $\hat{b}_2$ are used then this equality cannot be satisfied because of the conditions (a), (b) imposed on the MLE's. But, there exists a very interesting possibility to modify the equations (a), (b) slightly, in such a way that their solution will meet the demands. Thanks to stationarity of the investigated process we know that

$$\begin{aligned}\lim_{T\to\infty}\frac{\Phi_1}{T} &= \mathsf{E}\{|\dot{x}(0)|^2\} \quad \text{a.s.}\\ \lim_{T\to\infty}\frac{\Phi_2}{T} &= \mathsf{E}\{|x(0)|^2\} \quad \text{a.s.}\\ \lim_{T\to\infty}\frac{\Phi_3}{T} &= \lim_{T\to\infty}\frac{\Phi_4}{T} = 0 \text{ a.s.}\end{aligned}$$

If we neglect in the equations (a), (b) the terms of order $o(T)$ for $T \to \infty$, then we obtain "approximate" equations, which are linear, namely

$$\frac{T}{2} - b_1\,\Phi_1 = 0, \qquad \phi_1 - b_1\,\Phi_2 = 0.$$

Their solution is very simple and we see that the new estimates

$$b_1^0 = \frac{T}{2\Phi_1}, \qquad b_2^0 = \frac{\Phi_1}{\Phi_2}$$

are precisely those satisfying the basic relation between the likelihood ratio maximum and $I$-divergence discussed earlier.

Look in more detail how we could obtain these estimates from the corresponding Radon–Nikodym derivative.

Let us consider the logarithm of Radon–Nikodym derivative and neglect all the terms behaving $o(T)$ when $T \to \infty$. The rest of the derivative will be called the principal part of the derivative and in our case equals

$$\frac{(b_1 - a_1)T}{2} - \frac{b_1^2 - 2b_2 - a_1^2 + 2a_2}{2}\,\Phi_1 - \frac{b_2^2 - a_2^2}{2}\,\Phi_2.$$

For a better orientation, this principal part will be denoted by

$$P\,P\frac{\mathrm{d}P(x(\cdot), b_1, b_2)}{\mathrm{d}P(x(\cdot), a_1, a_2)}.$$

At this situation we can easily show that the estimates $b_1^0$, $b_2^0$, which are obtained by maximization of the principal part play a similar role as the Yule–Walker estimates in the case of a discrete time, see e.g., Dzhaparidze [2]. Simultaneously, we can express the principal part maximum via the asymptotic $I$-divergence rate

$$\sup_{b_1,b_2} \ln PP \frac{\mathrm{d}P(x(\cdot),b_1,b_2)}{\mathrm{d}P(x(\cdot),a_1,a_2)} = T\bar{I}(P_{b_1^0,b_2^0} : P_{a_1,a_2}),$$

where

$$\bar{I}(P_{b_1\,b_2} : P_{a_1\,a_2}) = \frac{1}{4\pi}\int_{-\infty}^{\infty}\left(\frac{\varphi_b}{\varphi_a} - \ln\frac{\varphi_b}{\varphi_a} - 1\right)\mathrm{d}\lambda.$$

On the other hand, it is also very interesting that the estimates $b_1^0$, $b_2^0$ can be obtained by the method of moments. Thanks to stationarity we have

$$R_0(0) = \mathsf{E}\{|x(0)|\} = \int_{-\infty}^{\infty}\varphi_b(\lambda)\,\mathrm{d}\lambda$$
$$R_1(0) = \mathsf{E}\{|\dot{x}(0)|^2\} = \int_{-\infty}^{\infty}\lambda^2\,\varphi_b(\lambda)\,\mathrm{d}\lambda.$$

Now, if we substitute the precise values $R_0(0)$, $R_1(0)$ by their estimates, namely

$$\hat{R}_0(0) = \frac{1}{T}\Phi_1, \qquad \hat{R}_1(0) = \frac{1}{T}\Phi_2$$

and we take into account these facts

$$\int_{-\infty}^{\infty}\varphi_b(\lambda)\,\mathrm{d}\lambda = \frac{1}{2b_1\,b_2}, \qquad \int_{-\infty}^{\infty}\lambda^2\,\varphi_b(\lambda)\,\mathrm{d}\lambda = \frac{1}{2b_1},$$

then solving the following equations

$$\hat{R}_0(0) = \int_{-\infty}^{\infty}\varphi_b(\lambda)\,\mathrm{d}\lambda, \qquad \hat{R}_1(0) = \int_{-\infty}^{\infty}\lambda^2\,\varphi_b(\lambda)\,\mathrm{d}\lambda$$

for unknown $b_1$, $b_2$ we immediately get the estimates $b_1^0$, $b_2^0$. There is no problem to prove the consistency of the proposed estimates because

$$\hat{R}_0(0) \longrightarrow R_0(0) = \frac{1}{2b_1^*},$$
$$\hat{R}_1(0) \longrightarrow R_1(0) = \frac{1}{2b_1^*\,b_2^*},$$

where $b_1^*$, $b_2^*$ are "true" parameters. The above obtained results we can summarize into the following theorems.

**Theorem 2.** Let $b_1^0$, $b_2^0$ be estimates obtained by maximization of the principal part of the Radon–Nikodym derivative. Then

$$\lim_{T\to\infty}\frac{1}{T}\sup_{(b_1\,b_2)}\ln PP\frac{\mathrm{d}P(x(\cdot),b_1,b_2)}{\mathrm{d}P(x(\cdot),a_1,a_2)} = \bar{I}(P_{(b_1^*\,b_2^*)} : P_{a_1\,a_2})$$

where $b_1^*$, $b_2^*$ are true values of unknown parameters.

A similar assertion we can state about the MLE's of $b_1$, $b_2$.

**Theorem 3.** Let $\hat{b}_1$, $\hat{b}_2$ be the MLE's of unknown parameter $b_1$, $b_2$. Then

$$\lim_{T\to\infty} \frac{1}{T} \sup_{(b_1\, b_2)} \ln \frac{\mathrm{d}P(x(\cdot), b_1, b_2)}{\mathrm{d}P(x(\cdot), a_1, a_2)} = \overline{I}(P_{(b_1^*\, b_2^*)} : P_{a_1\, a_2}).$$

Proof. It is well known, e.g. see Rozanov [9], that the MLE's in autoregressive cases are strongly consistent and asymptotically normal, i.e.

$$\lim_{T\to\infty} \hat{b}_1 = b_1^*, \quad \lim_{T\to\infty} \hat{b}_2 = b_2^*, \quad \text{a.s.}$$

The rest of the proof follows from the continuity of the Radon–Nikodym derivative in unknown parameters. □

The above described situation for $n = 2$ gives a hint about the general case. If we drop in the Radon–Nikodym derivative all terms of order $o(T)$ for $T \to \infty$ then the principal part satisfies the relation

$$\ln P\, P \frac{\mathrm{d}P_a}{\mathrm{d}Q}(x(\cdot)) = \frac{1}{2}\, a_1\, T - \frac{1}{2} \sum_{k=0}^{n-1} A_{n-k} \int_0^T |x^{(k)}(t)|^2 \,\mathrm{d}t.$$

This is a quadratic form in the coefficients $(a_1, a_2, \dots, a_n)$ and we can find its minimum. Applying partial derivatives to the logarithm of the principal part we get a system of linear equations for unknown parameters $a_1, a_2, \dots, a_n$. Estimates obtained in this manner were first described in Rozanov [9] or Pisarenko [8]. One can easily check that the system of linear equations has a unique solution with probability tending to one if $T \to \infty$. It follows from the fact that

$$\frac{1}{T} \int_0^T |x^{(k)(t)}|^2 \,\mathrm{d}t \underset{n\to\infty}{\longrightarrow} R_k(0) = \mathsf{E}\{|x^{(k)}(0)|^2\} \quad \text{a.s.}$$

These estimates are asymptotically normal and efficient. It is worth noting that these estimates can be obtained also by a different approach. First, we will show that these estimates of autoregressive coefficients can also be obtained by the method of moments. As seen from the form of the principal part of Radon–Nikodym derivative there is an “asymptotic” sufficient statistic given by sample characteristics

$$\frac{1}{T} \int_0^T |x(t)|^2 \,\mathrm{d}t,\ \frac{1}{T} \int_0^T |x'(t)|^2 \,\mathrm{d}t, \dots \frac{1}{T} \int_0^T |x^{n-1}(t)|^2 \,\mathrm{d}t$$

derived from the underlying process $\{x(t),\ t \in \langle 0, T \rangle\}$. Let us investigate the mapping

$$T_k(a_1, a_2, \dots, a_n) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{\lambda^{2k}}{\left|\sum_{k=0}^n a_{n-k} (i\lambda)^k\right|^2} \mathrm{d}\lambda$$

$k = 0, 1, \dots, n-1$ with $a_0 = 1$, which is defined on the stability domain of the polynomial $\sum_{k=0}^n a_{n-k}\, z^k$. There exists a one-to-one mapping between $a_1, a_2, \dots, a_n$ and the roots $\alpha_1, \alpha_2, \dots, \alpha_n$. The spectral density function $\varphi(\lambda)$ can be then expressed as

$$\varphi(\lambda) = \frac{1}{2\pi} \frac{1}{|\sum_{k=0}^{n} a_{n-k}(i\lambda)^k|^2} = \frac{1}{2\pi} \prod_{j=1}^{m} \left\{(\lambda - \operatorname{Im}\alpha_j)^2 + (\operatorname{Re}\alpha_j)^2\right\}^{-1}$$

where $\operatorname{Re}\alpha_i < 0,\ i = 1, 2, \dots, n$. This immediately implies that the stability domain is open. Let us prove that the mapping $\boldsymbol{T} = \{T_k(a_1, a_2, \dots, a_n)\}$ is one-to-one.

Let us assume there are $(a_1, a_2, \dots, a_n) \neq (b_1, b_2, \dots, b_n)$ such that

$$T_k(a_1, a_2, \dots, a_n) = T_k(b_1, b_2, \dots, b_n)$$

for each $k = 0, 1, \dots, n$. It means

$$\int_{-\infty}^{\infty} \lambda^{2k}\, \varphi_a(\lambda)\, \mathrm{d}\lambda = \int_{-\infty}^{\infty} \lambda^{2k}\, \varphi_b(\lambda)\, \mathrm{d}\lambda.$$

Let us introduce the function

$$h_{ab}(\lambda) = \frac{\varphi_a(\lambda)}{\varphi_b(\lambda)} - \ln \frac{\varphi_a(\lambda)}{\varphi_b(\lambda)} - 1.$$

From the inequality $x - \ln x - 1 \geq 0$ and properties of spectral density functions of autoregressive processes we can state the existence of

$$\int_{-\infty}^{\infty} h_{ab}(\lambda)\, \mathrm{d}\lambda \geq 0.$$

Analogously we can show that

$$\int_{-\infty}^{\infty} h_{ba}(\lambda)\, \mathrm{d}\lambda \geq 0.$$

Then the sum of both the integral exists and equals

$$\begin{aligned}
&\int_{-\infty}^{\infty} h_{ab}(\lambda)\, \mathrm{d}\lambda + \int_{-\infty}^{\infty} h_{ba}(\lambda)\, \mathrm{d}\lambda \\
=\ & \int_{-\infty}^{\infty} \left(\frac{\varphi_a(\lambda)}{\varphi_b(\lambda)} + \frac{\varphi_b(\lambda)}{\varphi_a(\lambda)} - 2\right) \mathrm{d}\lambda \\
=\ & \int_{-\infty}^{\infty} \left(\frac{1}{\varphi_b(\lambda)} - \frac{1}{\varphi_a(\lambda)}\right) (\varphi_a(\lambda) - \varphi_b(\lambda))\, \mathrm{d}\lambda \\
=\ & \int_{-\infty}^{\infty} \left(\sum_{k=0}^{n} B_{n-k}\lambda^{2k} - \sum_{k=0}^{n} A_{n-k}\lambda^{2k}\right) (\varphi_a(\lambda) - \varphi_b(\lambda))\, \mathrm{d}\lambda \\
=\ & \int_{-\infty}^{\infty} \left(\sum_{k=0}^{n-1} B_{n-k}\lambda^{2k} - \sum_{k=0}^{n-1} A_{n-k}\lambda^{2k}\right) (\varphi_a(\lambda) - \varphi_b(\lambda))\, \mathrm{d}\lambda \\
=\ & 0
\end{aligned}$$

because $B_0 = A_0$ and $T_k(a_1, \dots, a_n) = T_k(b_1, \dots, b_n)$ as we assumed. But it gives

$$\int_{-\infty}^{\infty} h_{ab}(\lambda)\,\mathrm{d}\lambda = 0,$$

which implies $h_{ab}(\lambda) = 0$ a.s. [Leb]. Hence we have obtained the fact

$$\varphi_a(\lambda) = \varphi_b(\lambda) \quad \text{a.s. [Leb]},$$

it means nothing else but $a_1 = b_1$, $a_2 = b_2, \dots, a_n = b_n$. The mapping $\boldsymbol{T}$ is one-to-one and we can define estimates

$$(\tilde{a}_1, \tilde{a}_2, \dots, \tilde{a}_n) = \boldsymbol{T}^{-1}(\hat{R}_0, \hat{R}_1, \dots, \hat{R}_{n-1})$$

if $(\hat{R}_0, \hat{R}_1, \dots, \hat{R}_{n-1}) \in \operatorname{Range} \boldsymbol{T}$ and $(\tilde{a}_1, \tilde{a}_2, \dots, \tilde{a}_n) = (0, 0, \dots, 0)$ otherwise. The statistics $\hat{R}_k$, $k = 0, 1, \dots, n-1$ are defined as

$$\hat{R}_k = \frac{1}{T} \int_0^T |x^{(k)}(t)|^2\,\mathrm{d}t.$$

As with $T \to \infty$, $\hat{R}_k \to \int_{-\infty}^{\infty} \lambda^{2k} \varphi_{a_0}(\lambda)\,\mathrm{d}\lambda$ where $a_0 = (a_1^0, a_2^0, \dots, a_n^0)$ are true autoregressive coefficients, we can assert that for large $T$ with probability close to one $(\hat{R}_0, \hat{R}_1, \dots, \hat{R}_{n-1}) \in \operatorname{Range} \boldsymbol{T}$. It remains to prove that these estimates obtained by the method of moments are identical with those given by minimizing the principal part of the Radon–Nikodym derivative. Let $\tilde{a}_1, \tilde{a}_2, \dots, \tilde{a}_n$ be the estimates obtained by minimizing the quadratic form of the principal part then this minimum equals

$$\frac{1}{2} T \tilde{a}_1 - \frac{1}{2} \sum_{k=0}^{n-1} \tilde{A}_{n-k} \int_0^T |x^{(k)}(t)|^2\,\mathrm{d}t.$$

Let $a_1^*, a_2^*, \dots, a_n^*$ be the estimates given by the method of moments, i.e.

$$\frac{1}{T} \int_0^T |x^{(k)}(t)|^2\,\mathrm{d}t = \int_{-\infty}^{\infty} \lambda^{2k} \varphi_{a^*}(\lambda)\,\mathrm{d}\lambda.$$

Then we can write

$$\begin{aligned}
\frac{1}{T} \ln P P \frac{\mathrm{d}P_a}{\mathrm{d}Q} &= \frac{1}{2}(a_1 - a_1^*) - \frac{1}{2} \sum_{k=0}^{n-1} A_{n-k} \hat{R}_k + \frac{1}{2} a_1^* \\
&= \frac{1}{2}(a_1 - a_1^*) - \frac{1}{2} \sum_{k=0}^{n-1} A_{n-k} \int_{-\infty}^{\infty} \varphi_{a^*}(\lambda)\,\mathrm{d}\lambda + \frac{1}{2} a_1^* \\
&= -\frac{1}{2} \int_{-\infty}^{\infty} \left( \frac{\varphi_{a^*}(\lambda)}{\varphi_a(\lambda)} - \frac{\varphi_{a^*}(\lambda)}{\varphi_a(\lambda)} - 1 \right) \mathrm{d}\lambda + \frac{1}{2} a_1^* \\
&= -\frac{1}{2} \overline{I}(\varphi_{a^*} : \varphi_a) + \frac{1}{2} a_1^*,
\end{aligned}$$

where $\overline{I}(\varphi_{a^*} : \varphi_a)$ is the asymptotic rate of $I$-divergence. For details see Michálek [5]. We immediately see that

$$\begin{aligned}\max_a \ln P\, P\, \frac{\mathrm{d}P_a}{\mathrm{d}Q}(x(\cdot)) &= -\frac{1}{2}\min_a \overline{I}(\varphi_{a^*} : \varphi_a) + \frac{1}{2}\, a_1^* \\ &= \frac{1}{2}\, a_1^*\end{aligned}$$

because $\overline{I}(\varphi_{a^*} : \varphi_a) \geq 0$ and equals zero if and only if $a^* = a$. This fact proves the coincidence of both the estimates.

There is another possibility how to construct these estimates as minimal distance estimates. Let $\varphi_a(\cdot)$ be a spectral density function belonging to an autoregressive random process of the $n$th order again, $a = (a_1, a_2, \ldots, a_m)$, let $\hat{f}(\cdot)$ be another spectral density function, for which the functional

$$F(\varphi_a(\cdot)) = \frac{1}{4\pi}\int_{-\infty}^{\infty}\left(\frac{\hat{f}(\lambda)}{\varphi_a(\lambda)} - \ln\frac{\hat{f}(\lambda)}{\varphi_a(\lambda)} - 1\right)\mathrm{d}\lambda$$

is finite. We wish to find the minimum of $F(\varphi_a(\cdot))$, if exists, over all spectral density functions $\varphi_a(\cdot)$. This functional can be expressed as

$$F(\varphi_a(\cdot)) = \frac{1}{2}\int_{-\infty}^{\infty}\left(\sum_{k=0}^{n}\lambda^{2k}A_{n-k}\,\hat{f}(\lambda) - \ln\left(2\pi\,\hat{f}(\lambda)\sum_{k=0}^{n}A_{n-k}\,\lambda^{2k}\right) - 1\right)\mathrm{d}\lambda.$$

Further, let us assume the existence of

$$\int_{-\infty}^{\infty}\lambda^{2n}\,\hat{f}(\lambda)\,\mathrm{d}\lambda.$$

Then we can calculate the partial derivatives $\frac{\partial}{\partial a_j}F(\varphi_a(\cdot))$, $j = 1, 2, \ldots, n$ by changing the ordering between the integral and derivatives. After simple calculation we get to the following system of non-linear equations, namely

$$\sum_{k=0}^{n-1}\int_{-\infty}^{\infty}\frac{\partial A_{n-k}}{\partial a_j}\lambda^{2k}\,f(\lambda)\,\mathrm{d}\lambda = \sum_{k=0}^{n-1}\int_{-\infty}^{\infty}\frac{\partial A_{n-k}}{\partial a_j}\lambda^{2k}\,f_a(\lambda)\,\mathrm{d}\lambda$$

$j = 1, 2, \ldots, n$. This system can be rewritten as follows

$$\sum_{k=0}^{n-1}\frac{\partial A_{n-k}}{\partial a_j}(\hat{R}_k - R_k(\varphi(a))) = 0,$$

where $\hat{R}_k$, $k = 0, \ldots, n-1$ are sample characteristics

$$\hat{R}_k = \frac{1}{T}\int_0^T |h^{(k)}(t)|^2\,\mathrm{d}t$$

derived from the observed process $\{x(t),\ t \in \langle 0,T\rangle\}$ and $R_k(\varphi(a))$, $k = 0,1,\dots,n-1$ are exact values for dispersions of derivatives of an autoregressive process with the spectral density function $\varphi_a(\cdot)$. Although the above described system of equations is very difficult to solve in general there exists an elegant solution if we solve the system

$$\hat{R}_k = R_a(\varphi(a))$$

$k = 0,1,\dots,n-1$, which is precisely the method of moments described earlier. We see that at least one solution of the underlying system exists. Let us prove that moment-method estimates determine the minimum value of $F(\varphi_a(\cdot))$. For this calculate the difference

$$\begin{aligned}
4\pi\left(F(\varphi_b(\cdot)) - F(\varphi_{a^*}(\cdot))\right) &= \int_{-\infty}^{\infty}\left(\frac{\hat{f}(\lambda)}{\varphi_b(\lambda)} - \ln\frac{\hat{f}(\lambda)}{\varphi_b(\lambda)} - 1\right)\mathrm{d}\lambda \\
&\quad - \int_{-\infty}^{\infty}\left(\frac{\hat{f}(\lambda)}{\varphi_{a^*}(\lambda)} - \ln\frac{\hat{f}(\lambda)}{\varphi_{a^*}(\lambda)} - 1\right)\mathrm{d}\lambda \\
&= \int_{-\infty}^{\infty}\left(\left(\frac{1}{\varphi_b(\lambda)} - \frac{1}{\varphi_{a^*}(\lambda)}\right)\hat{f}(\lambda) - \frac{\varphi_{a^*}(\lambda)}{\varphi_b(\lambda)}\right)\mathrm{d}\lambda \\
&= \sum_{k=0}^{n}(B_{n-k} - A^*_{n-k})\,\hat{R}_k(0) + b_1 - a^*_1 \\
&= \sum_{k=0}^{n}(B_{n-k} - A^*_{n-k})\int_{-\infty}^{\infty}\lambda^{2k}\,\varphi_{a^*}(\lambda)\,\mathrm{d}\lambda + b_1 - a^*_1 \\
&= \int_{-\infty}^{\infty}\left(\frac{\varphi_{a^*}}{\varphi_b} - 1 - \ln\frac{\varphi_{a^*}}{\varphi_b}\right)\mathrm{d}\lambda \geq 0.
\end{aligned}$$

This evidently shows that $F(\varphi_b(\cdot)) \geq F(\varphi_{a^*(\lambda)})$.

In this way we have proved that the estimates determined by the method of moments are also estimates minimizing the asymptotic $I$-divergence rate. We see that in the case of a continuous time there is an analogy with the discrete time case where the Yule–Walker estimates of autoregressive coefficients minimize also the asymptotic $I$-divergence rate. The property of the moment-method estimates to minimize the asymptotic $I$-divergence rate can be described as follows, too. Let us imagine we observe a quite arbitrary stationary Gaussian process with a zero mean and with at least the $n$th derivative on the interval $\langle 0,T\rangle$. Now, we are looking for the most similar autoregressive Gaussian process of the $n$th order to our underlying process $\{x(t),\ t \in \langle 0,T\rangle\}$. The similarity is measured via the asymptotic rate of $I$-divergence, i.e. via the functional $F(\varphi_a(\cdot))$ where the spectral density function $\hat{f}(\cdot)$ based on observations $x(t),\ t \in \langle 0,T\rangle$ is determined by

$$\hat{R}_k(0) = \frac{1}{T}\int_0^T |x^{(k)}(t)|^2\,\mathrm{d}t = \int_{-\infty}^{\infty}\hat{f}(\lambda)\,\lambda^{2k}\,\mathrm{d}\lambda,$$

$k = 0,1,\dots,n$. In other words speaking the underlying process $x(t),\ t \in \langle 0,T\rangle$ and the approximating autoregressive process must have the same dispersions of their derivatives up to the order $n-1$.

This part will be closed by two theorems whose proofs can be dropped and which deals with autoregressive processes.

**Theorem 4.** Let $\tilde{a} = (\tilde{a}_1, \tilde{a}_2, \dots, \tilde{a}_n)$ be estimates of autoregressive coefficients obtained by minimizing the principal part of the corresponding Radon–Nikodym derivative. Then

$$\lim_{T\to\infty} \frac{1}{T} \max_{a} \ln P\, P\, \frac{\mathrm{d}P_a}{\mathrm{d}P_{a_0}}(x(\cdot)) = \lim_{T\to\infty} \ln P\, P\, \frac{\mathrm{d}P_{\tilde{a}}}{\mathrm{d}P_{a_0}}(x(\cdot)) = \overline{I}(P_{a_1} : P_{a_0}) \quad \text{a.s.},$$

where $a_1 = (a_1^1, a_2^1, \dots, a_n^1)$ are true autoregressive parameters.

The same theorem is valid for the MLE's.

**Theorem 5.** Let $\hat{a} = (\hat{a}_1, \hat{a}_2, \dots, \hat{a}_n)$ be the MLE's of autoregressive coefficients then

$$\lim_{T\to\infty} \frac{1}{T} \max_{a} \ln \frac{\mathrm{d}P_a}{\mathrm{d}P_{a_0}}(x(\cdot)) = \lim_{T\to\infty} \ln \frac{\mathrm{d}P_{\hat{a}}}{\mathrm{d}P_{a_0}}(x(\cdot)) = \overline{I}(P_{a_1} : P_{a_0}) \quad \text{a.s.},$$

where $a_1 = (a_1^1, a_2^1, \dots, a_n^1)$ are true autoregressive parameters again.

The quantity $\overline{I}(P_{a_1} : P_{a_0})$ is asymptotic rate of $I$-divergence between two Gaussian autoregressive stationary measures and equals

$$\overline{I}(P_{a_1} : P_{a_0}) = \frac{1}{4\pi} \int_{-\infty}^{\infty} \left( \frac{\varphi_{a_1}(\lambda)}{\varphi_{a_0}(\lambda)} - \ln \frac{\varphi_{a_1}(\lambda)}{\varphi_{a_0}(\lambda)} - 1 \right) \mathrm{d}\lambda.$$

For more details about $\overline{I}(P_{a_1} : P_{a_0})$ see Michálek [5].

(Received November 7, 1996.)

## REFERENCES


[1] J. Anděl: Statistical Analysis of Time Series (in Czech). SNTL, Prague 1976.

[2] K. Dzhaparidze: Parameter Estimation and Hypothesis Testing in Spectral Analysis of Stationary Time Series. (Springer Series in Statistics.) Springer Verlag, Berlin 1986

[3] J. Hájek: On the simple linear model for Gaussian processes. In: Trans. of the 2nd Prague Conference, Academia, Prague 1959, pp. 185–197.

[4] J. Hájek: On linear statistical problems in stochastic processes. Czechoslovak Math. J. *12 (87)* (1962), 404–444.

[5] J. Michálek: Asymptotic Rényi's rate of Gaussian processes. Problems Control Inform. Theory *19* (1990), 3, 209–227.

[6] J. Michálek: Maximum likelihood principle and $I$-divergence: observations in discrete time. Kybernetika *34* (1998), 265–288.

[7] V. F. Pisarenko: On absolute continuity of the measures corresponding to a rational spectral density function (in Russian). Teor. Veroyatnost. i Primenen. *IV* (1959), 481–481.

[8] V. F. Pisarenko: On parameter estimations of a Gaussian stationary processes with a spectral density function (in Russian). Lithuanian Math. J. (1962).
[9] J. A. Rozanov: On application a central limit theorem. In: Proc. Fourth Berkeley Symp. Math. Stat. Prob., Berkeley 1961, Vol. 2.

*RNDr. Jiří Michálek, CSc., Institute of Information Theory and Automation – Academy of Sciences of the Czech Republic, Pod vodárenskou věží 4, 182 08 Praha 8. Czech Republic.*
*e-mail: michalek@utia.cas.cz*