ROBUST'98, 139 – 147

ODHADY S MINIMÁLNÍ VZDÁLENOSTÍ PRO GAUSSOVSKÉ PROCESY

Jiří MICHÁLEK[1]

ÚTIA ČAV

Abstract. The paper presents a close connection between maximum likelihood estimates and minimum I-divergence estimates of unknown parameters in Gaussian families and studies the asymptotic behaviour of the likelihood ratio maximum.
Резюме: Автор показывает интересное соотношение между оценками максимального правдоподобия и оценками с минимальной I-дивергенцией для гауссовских последовательностей и процессов.

1. Úvod

Problém, kterým se budeme v práci zabývat, přiblížíme na následujícím příkladu. Nechť $X_1, X_2, \ldots, X_n$ je prostý náhodný výběr z $N(\mu, \sigma^2)$, kde parametr $\theta = (\mu, \sigma^2) \in \Theta = R_1 \times R_1^+$ je neznámý. Lze snadno ukázat, že

$$\max_{\theta\in\Theta} \ln \frac{f(x_1,\ldots,x_n,\theta)}{f(x_1,\ldots,x_n,\theta_0)} = \frac{n}{2}\left(\frac{\overline{x}-\mu_0}{\sigma_0}\right)^2 + \frac{n}{2}\left(\frac{s^2}{\sigma_0^2} - \ln\frac{s^2}{\sigma_0^2} - 1\right), \tag{1}$$

kde $f(x_1,\ldots,x_n,\theta) = \prod_{i=1}^n g(x_i,\mu,\sigma^2)$, $g(\cdot,\mu,\sigma^2) \sim N(\mu,\sigma^2)$ a $\theta_0 = (\mu_0, \sigma_0^2)$. Tento výraz napravo v rovnosti (1) lze ale vyjádřit daleko zajímavějším způsobem, a to jako

$$\max_{\theta\in\Theta} \ln \frac{f(x_1,\ldots,x_n,\theta)}{f(x_1,\ldots,x_n,\theta_0)} = \frac{n}{2} I\left(N(\overline{x}, s^2) : N(\mu_0, \sigma_0^2)\right),$$

kde $\overline{x} = \frac{1}{n}\sum_{i=1}^n x_i$, $s^2 = \frac{1}{n}\sum_{i=1}^n (x_i - \overline{x})^2$ a $I(P:Q)$ je Kullback–Leiblerova I-divergence mezi dvěma pravděpodobnostními mírami P, Q daná vztahem

$$I(P:Q) = \int_{\mathcal{X}} f_P(x) \ln \frac{f_P(x)}{f_Q(x)} \,\mathrm{d}\nu(x),$$

$f_P = \frac{\mathrm{d}P}{\mathrm{d}\nu}$, $f_Q = \frac{\mathrm{d}Q}{\mathrm{d}\nu}$ a ν je libovolná σ-konečná dominující míra vůči dvojici (P,Q). Tím se ihned naskýtá otázka, zdali tento vztah mezi maximem věrohodnostního poměru a I-divergencí platí obecně či za jakých podmínek, a co z toho vztahu vyplývá.

2. Exponenciální rodina

Hustota exponenciální rodiny $f(x,\theta)$, $\theta \in \Theta$ (parametrický prostor) vůči dominantní míře $\mu(\cdot)$ má obecný tvar

$$f(x,\theta) = c(\theta)\, h(x)\, e^{\tau(\theta)\, T(x)},$$

[1]Tato práce je podporována grantem GA ČR 201/96/0415.

kde θ nechť je pro jednoduchost jednorozměrný parametr. Maximálně věrohodný odhad (dále jen MLE) $\widehat{\theta}_n$ počítaný na základě pozorování $(x_1, x_2, \dots, x_n)$ musí (za předpokladů existence patřičných derivací) splňovat vztah

$$\frac{1}{n}\sum_{j=1}^{n} T(x_j) = \frac{-1}{C(\widehat{\theta}_n)}\, C'(\widehat{\theta}_n)\, \frac{1}{\tau'(\widehat{\theta}_n)}.$$

Odhad $\widehat{\theta}_n$ existuje právě tehdy, když

$$\frac{1}{n}\sum_{j=1}^{n} T(x_j) \in \mathrm{Range}_{\theta\in\Theta}\,\{\pi(\theta)\},$$

kde

$$\pi(\theta) = -[\ln C(\theta)]'\,\frac{1}{\tau'(\theta)} \quad \text{při} \quad \tau'(\theta) \neq 0 \text{ na } \Theta.$$

Protože platí silný zákon velkých čísel, pak lze snadno vidět, že

$$\lim_{n\to\infty} P\left\{\frac{1}{n}\sum_{j=1}^{n} T(x_j) \in \mathrm{Range}_{\theta\in\Theta}\{\pi(\theta)\}\right\} = 1.$$

Důležité ale je, že lze psát

$$\frac{1}{n}\sum_{j=1}^{n} \ln f(x_j, \theta) = -I(\widehat{\theta}_n : \theta) + H(\widehat{\theta}_n),$$

kde $H(\theta) = E_\theta\{\ln f(x,\theta)\}$ (pokud ovšem existuje). Z tohoto ale ihned plyne, že

$$\max_{\theta\in\Theta}\left\{\frac{1}{n}\sum_{j=1}^{n} \ln f(x_j,\theta)\right\} = -\min_{\theta\in\Theta}\{I(\widehat{\theta}_n : \theta)\} + H(\widehat{\theta}_n) = H(\widehat{\theta}_n),$$

a tedy lze tvrdit, že MLE $\widehat{\theta}_n$ je též odhadem s minimální I-divergencí. Protože $\{\widehat{\theta}_n\}$ je superkonsistentní, tj. $\widehat{\theta}_n \to \theta_*$ s. j., pak

$$\max_{\theta\in\Theta}\frac{1}{n}\sum_{j=1}^{n} \ln f(x_j,\theta) \underset{n\to\infty}{\longrightarrow} H(\theta_*) \quad \text{s. j.}$$

Zcela obdobně lze ukázat, že

$$\max_{\theta\in\Theta}\frac{1}{n}\sum_{j=1}^{n} \ln\frac{f(x_j,\theta)}{f(x_j,\theta_0)} = I(\widehat{\theta}_n : \theta),$$

z čehož vyplývá, že

$$\lim_{n\to\infty}\max_{\theta\in\Theta}\frac{1}{n}\sum_{j=1}^{n} \ln\frac{f(x_j,\theta)}{f(x_j,\theta_0)} = I(\theta_* : \theta) \quad \text{s. j.},$$

kde θ_* je skutečná hodnota parametru θ.

3. Závislá pozorování – diskrétní čas

Zatím jsme uvažovali pouze nezávislá pozorování. Nyní se budeme zabývat závislým pozorováním, a to případem autoregrese s gaussovskými pozorováními. Máme tedy model

$$x_{i+1} + \sum_{k=1}^{p} a_k\, x_{i-k+1} = \sigma_0\, e_{i+1}, \quad e_{i+1} \sim N(0,1).$$

Označme $\boldsymbol{a} = (1, a_1, a_2, \ldots, a_p)^T$, $\boldsymbol{b} = (1, b_1, b_2, \ldots, b_p)^T$, $\sigma_0^2 > 0$, $\sigma_1^2 > 0$. Lze snadno spočítat, že

$$\max_{(\boldsymbol{b},\sigma_1^2)} \ln \frac{f(x_1,\ldots,x_n,\boldsymbol{b},\sigma_1^2)}{f(x_1,\ldots,x_n,\boldsymbol{a},\sigma_0^2)} = \frac{n}{2}\left(\frac{\boldsymbol{a}^T\widehat{\boldsymbol{R}}\boldsymbol{a}}{\sigma_0^2} - \ln\frac{(\widehat{\sigma}_1)^2}{\sigma_0^2} - 1\right), \tag{2}$$

kde $f(x_1,\ldots,x_n,\boldsymbol{b},\sigma_1^2)$ je sdružená hustota veličin $x_1, x_2, \ldots, x_n$ při parametrech $(\boldsymbol{b},\sigma_1^2)$, dále $\widehat{\sigma}_1^2 = \widehat{\boldsymbol{b}}^T\widehat{\boldsymbol{R}}\widehat{\boldsymbol{b}}$, kde

$$\widehat{\boldsymbol{R}} = \left\{\widehat{R}_{ij}\right\}_{i,j=0}^{p}, \quad \widehat{R}_{ij} = \frac{1}{n}\sum_{k=1}^{n} x_{k-i}\, x_{k-j}$$

a $\widehat{\boldsymbol{b}}$ je MLE pro parametr $\boldsymbol{b} = (1, b_1, b_2, \ldots, b_p)^T$. Ihned nás napadne, co je výraz na pravé straně rovnosti (2)? Kdybychom namísto matice $\widehat{\boldsymbol{R}}$ použili Toeplitzovy matice

$$\widehat{\boldsymbol{T}} = \{\widehat{T}_{ij}\}_{i,j=0}^{p}, \quad \widehat{T}_{ij} = \frac{1}{n}\sum_{k=1}^{n} x_k\, x_{k+|i-j|},$$

pak výraz

$$\frac{\boldsymbol{a}^T\widehat{\boldsymbol{T}}\boldsymbol{a}}{\sigma_0^2} - \ln\frac{(\widehat{s}_1)^2}{\sigma_0^2} - 1, \quad \text{kde } (\widehat{s}_1)^2 = \widetilde{\boldsymbol{b}}^T\widehat{\boldsymbol{T}}\widetilde{\boldsymbol{b}}$$

a $\widetilde{\boldsymbol{b}} = (1, \widetilde{b}_1, \widetilde{b}_2, \ldots, \widetilde{b}_p)^T$ jsou známé Yule–Walkerovy odhady autoregresních koeficientů, by nebyl nic jiného nežli

$$\frac{n}{2}\,\overline{I}\left((\widetilde{\boldsymbol{b}}, (\widehat{s}_1)^2 : (\boldsymbol{a}, \sigma_0^2))\right),$$

kde

$$\overline{I}\left((\boldsymbol{b},\sigma_1^2) : (\boldsymbol{a},\sigma_0^2)\right) = \frac{1}{4\pi}\int_{-\pi}^{\pi}\left(\frac{\varphi(\boldsymbol{b},\sigma_1^2)\,(\lambda)}{\varphi(\boldsymbol{a},\sigma_0^2)\,(\lambda)} - \ln\frac{\varphi(\boldsymbol{b},\sigma_1^2)\,(\lambda)}{\varphi(\boldsymbol{a},\sigma_0^2)\,(\lambda)} - 1\right) d\lambda,$$

kde $\varphi(\boldsymbol{b},\sigma_1^2)$ a $\varphi(\boldsymbol{a},\sigma_0^2)$ jsou příslušné spektrální hustoty, pokud náš autoregresní model bude stacionární. Tento výraz je tzv. asymptotická rychlost I-divergence mezi dvěma stacionárními autoregresními gaussovskými mírami (blíže viz [1]). Odtud snadno plyne následující

Věta 1. Za předpokladu stacionarity v gaussovském autoregresním modelu platí

$$\lim_{n\to\infty} \frac{1}{n} \max_{(\boldsymbol{b},\sigma_1^2)} \ln \frac{f(x_1,\ldots,x_n,\boldsymbol{b},\sigma_1^2)}{f(x_1,\ldots,x_n,\boldsymbol{a},\sigma_0^2)} = \frac{1}{2}\,\overline{I}\left((\boldsymbol{b}_*,\sigma_*^2) : (\boldsymbol{a},\sigma_0^2)\right), \quad \text{s.j.},$$

kde $(\boldsymbol{b}_*,\sigma_*^2)$ jsou skutečné hodnoty parametrů.

Důkaz. Je notoricky známo, že maximálně věrohodné odhady parametrů $(\boldsymbol{b},\sigma_1^2)$ jsou superkonsistentní, tj.

$$(\boldsymbol{b}_n,\widehat{\sigma}_1^2) \underset{n\to\infty}{\longrightarrow} (\boldsymbol{b}_*,\sigma_*^2) \text{ s.j.}$$

Dále je známo, že

$$\widehat{R}_{ij} \underset{n\to\infty}{\longrightarrow} R_{ij} \text{ s.j.}$$

taktéž, kde $\boldsymbol{R} = \{R_{ij}\}$ je kovariační toeplitzovská matice autoregresní posloupnosti. Použitím známých vztahů mezi $\boldsymbol{R}$ a odpovídající spektrální hustotou plynoucích z Fourierovy analýzy již snadno dokážeme tvrzení věty.

Nyní se obrátíme k obecnému případu gaussovské stacionární posloupnosti s nulovou střední hodnotou. Pak sdružená hustota pozorování $(x_1,x_2,\ldots,x_n)$ má tvar

$$f_n(x_1,\ldots,x_n) = \frac{1}{(2\pi)^{n/2}\,|\boldsymbol{R}_n|^{1/2}} \exp\left\{-\frac{1}{2}\,\boldsymbol{x}^T\,\boldsymbol{R}_n^{-1}\boldsymbol{x}\right\},$$

kde $\boldsymbol{R}_n$ je příslušná kovariační matice. což znamená, že její logaritmus má tvar

$$L_n(\boldsymbol{x}) = -\frac{1}{2}\left\{n\ln 2\pi + \ln|\boldsymbol{R}_n| - \boldsymbol{x}^T\,\boldsymbol{R}_n^{-1}\boldsymbol{x}\right\}.$$

Problém je v tom, že explicitní vyjádření pro $L_n(\boldsymbol{x})$ bývá velice těžkopádné, a tudíž případný výpočet MLE neznámého parametru θ je nesnadný a obvykle je možno získat ho pouze numericky.

V polovině padesátých let přišel Whittle, viz [2], s myšlenkou nahradit logaritmus věrohodnostního poměru vhodným výrazem $\widetilde{L}_n(\boldsymbol{x})$, pro který by platilo

$$n^{-1/2}\left(L_n(\boldsymbol{x}) - \widetilde{L}_n(\boldsymbol{x})\right) \underset{n\to\infty}{\overset{\text{slabě}}{\longrightarrow}} 0,$$

přičemž odhady parametru θ získané maximalizací $\widetilde{L}_n(\boldsymbol{x})$ by byly snadněji dostupné. Výraz $\widetilde{L}_n(\boldsymbol{x})$ se nazývá principiální částí věrohodnostního poměru a o jeho použití při konstrukcí odhadů pro gaussovské stacionární posloupnosti viz monografii [3]. Vhodný tvar pro $\widetilde{L}_n(\boldsymbol{x})$ lze totiž v případě stacionárních gaussovských posloupností najít jako

$$\widetilde{L}_n(\boldsymbol{x}) = -\frac{n}{2}\left(\ln 2\pi + \frac{1}{2\pi}\int_{-\pi}^{\pi} \ln(2\pi\varphi(\lambda))\,\mathrm{d}\lambda + \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{I_n(\lambda,\boldsymbol{x})}{\varphi(\lambda)}\,\mathrm{d}\lambda\right),$$

kde

$$I_n(\lambda, \boldsymbol{x}) = \frac{1}{2\pi n} \left| \sum_{j=1}^{n} x_j \, e^{-ij\lambda} \right|^2$$

je periodogram. Snadno lze ukázat, že

$$\widetilde{L}_n(\boldsymbol{x}) = -\frac{n}{2} \left(\text{konst} + \frac{1}{2\pi} \int_{-\pi}^{\pi} \left(\frac{\widetilde{\varphi}(\lambda, \boldsymbol{x})}{\varphi(\lambda)} \right) - 1 \right) \mathrm{d}\lambda + \frac{1}{2\pi} \int_{-\pi}^{\pi} \ln \widetilde{\varphi}(\lambda, \boldsymbol{x}) \, \mathrm{d}\lambda,$$

kde $\widetilde{\varphi}(\lambda, \boldsymbol{x})$ je spektrální hustota odvozená od výběrových momentů.

Bude-li nyní $\varphi(\lambda) = \varphi(\lambda, \theta)$, kde $\theta \in \Theta$ je konečněrozměrný parametr, pak je vidět, že odhady $\widetilde{\theta}_n$ získané maximalizací principiální částí $\widetilde{L}_n(\boldsymbol{x})$ věrohodnostního poměru jsou vlastně odhady minimalizující asymptotickou rychlost I-divergence, neboť

$$\widetilde{\theta}_n = \arg\max\nolimits_{\theta \in \Theta} \widetilde{L}_n(\boldsymbol{x}) = \arg\min\nolimits_{\theta \in \Theta} \overline{I}(\widetilde{\varphi}(\lambda, \boldsymbol{x}) : \varphi(\lambda, \theta)).$$

Ve výše uvažovaném případě autoregresního modelu, když $\varphi(\lambda, \theta)$ odpovídá autoregresnímu modelu, pak výše popsané odhady autoregresních koeficientů jsou Yule–Walkerovy odhady (viz [4]).

4. ZÁVISLÁ POZOROVÁNÍ – SPOJITÝ ČAS

Nyní se budeme zabývat stochastickými procesy, které jsou zadány na intervalu $\langle 0, T \rangle$ a jsou vesměs gaussovské.

Nejjednodušší případ je následující jednoduchý regresní model pro gaussovský proces $\{x(t),\, t \in \langle 0, T \rangle\}$ se střední hodnotou

$$E_\alpha\{x(t)\} = \alpha \, \varphi(t), \quad \varphi(\cdot) \neq 0 \quad \text{na } \langle 0, T \rangle$$

a kovariační funkcí $\operatorname{cov}_\alpha(x(s),\, x(t)) = \operatorname{cov}_0(x(s),\, x(t))$. Pak (viz [5]) buď odpovídající gaussovské míry $\{P_\alpha\}$, $\alpha \in R_1$ jsou navzájem absolutně spojité, tj. $P_\alpha \sim P_0$ a nebo jsou singulární. V regulárním případě lze psát

$$\frac{\mathrm{d}P_\alpha}{\mathrm{d}P_0}(x(\cdot)) = \exp\left\{ \alpha v - \frac{1}{2} \alpha^2 E_0\{v^2\} \right\},$$

kde $v = \widehat{\alpha} \, E_0\{v^2\}$, přičemž $\widehat{\alpha}$ je MLE parametru α. Je pak snadno vidět, že

$$\max_{\alpha \in R_1} \left\{ \ln \frac{\mathrm{d}P_\alpha}{\mathrm{d}P_0}(x(\cdot)) \right\} = I(P_\alpha : P_0),$$

kde $I(P_\alpha : P_0) = \frac{1}{2} \alpha^2 E_0\{v^2\}$. Odtud opět ihned vidíme, že MLE parametru α je rovněž odhad minimalizující $I(P_\alpha : P_0)$.

Velice širokou třídou procesu užitečných jak z teoretického, tak i praktického hlediska jsou gaussovské autoregresní procesy stacionárního typu, se střední hodnotou nula pro jednoduchost, které mají spektrální hustotu

$$\varphi(\lambda) = \frac{\sigma^2}{2\pi} \frac{1}{\left|\sum_{k=0}^{p} a_{n-k}(i\lambda)^k\right|^2},$$

kde $a_0 = 1$ a charakteristický polynom $P(z) = \sum_{k=0}^{p} a_{n-k}\, z^k$ splňuje podmínku stacionarity. Odpovídající Radon–Nikodymova hustota vůči míře Q, která je generována gaussovským procesem $\{y(t),\, t \in \langle 0, T\rangle\}$ se střední hodnotou nula a s nezávislými přírůstky, kde $(p-1)$ní derivace splňuje

$$E\left\{\mathrm{d}\, y^{(p-1)}(t)^2\right\} = \sigma^2\, \mathrm{d}t,$$

(blíže viz [6]), má pak tvar

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

kde $\{D_{jk}\}$ jsou prvky inverzní matice ke kovarianční matici náhodných veličin $(x(0), x^{(1)}(0), \dots, x^{(p-1)}(0))$.

Dále platí vztah

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

Pro $p > 1$ je nemožné explicitně spočítat MLE pro $(a_1, a_2, \dots, a_p)$. Proto se opět uvažuje o principiální části Radon–Nikodymovy derivace tím, že se zanedbají členy řádu $o(T)$ pro $T \to \infty$. Po zanedbání těchto členů se získá systém lineárních rovnic pro parametry $(a_1, a_2, \dots, a_p)$ snadno řešitelný, blíže viz ([7, 8]). Ukážeme v následujícím, že tytéž odhady lze získat i jiným přístupem. A to buď A) momentovou metodou a nebo B) minimalizací vhodné asymptotické rychlosti I-divergence.

A) Označme

$$\widehat{R}_k = \frac{1}{T} \int_0^T |x^{(k)}(t)|^2\, \mathrm{d}t, \quad k = 0, 1, \dots, p-1.$$

Uvažujme zobrazení $\boldsymbol{T} = \{T_k(\cdot)\}_{k=0}^{p-1}$, kde

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

Lze ukázat, že zobrazení $\boldsymbol{T}$ je jednojednoznačné zobrazení definované na otevřené podmnožině stacionarity parametrického prostoru, kterým je míněna oblast stacionarity. Díky konvergenci ve smyslu s. j.

$$\widehat{R}_k \underset{T\to\infty}{\longrightarrow} R_k = E\left\{|x^{(k)}(0)|^2\right\}$$

lze ukázat, že s pravděpodobností blížící se 1 při $T \to \infty$ existují odhady $(\widetilde{a}_1, \widetilde{a}_2, \dots, \widetilde{a}_p) = \boldsymbol{T}^{-1}(\widehat{R}_0, \widehat{R}_1, \dots, \widehat{R}_{p-1})$. Pak lze ukázat, že principiální část $PP \ln \frac{\mathrm{d}P_a}{\mathrm{d}Q}$ Radon–Nikodymovy derivace získaná zanedbáním členů řádu $o(T)$ při $T \to \infty$ splňuje vztah

$$\frac{1}{T} \ln PP \frac{\mathrm{d}P_a}{\mathrm{d}Q}(x(\cdot)) = -\frac{1}{2}\,\overline{I}(\widetilde{\varphi}(\cdot) : \varphi_a(\cdot)),$$

kde $\widetilde{\varphi}(\cdot)$ je spektrální hustota odvozená od výběrových charakteristik $(\widehat{R}_0, \widehat{R}_1, \dots, \widehat{R}_{p-1})$ pomocí vztahu

$$\widehat{R}_k = \int_{-\infty}^{+\infty} \lambda^{2k}\, \widetilde{\varphi}(\lambda)\, \mathrm{d}\lambda.$$

Výraz pro asymptotickou rychlost I-divergence mezi dvěma gaussovskými stacionárními mírami se střední hodnotou nula je vyjádřen jako

$$\overline{I}(P : Q) = \frac{1}{4\pi} \int_{-\infty}^{+\infty} \left(\frac{\varphi_P(\lambda)}{\varphi_Q(\lambda)} - \ln \frac{\varphi_P(\lambda)}{\varphi_Q(\lambda)} - 1 \right) \mathrm{d}\lambda,$$

kde $\varphi_P(\cdot)$, $\varphi_Q(\cdot)$ jsou odpovídající spektrální hustoty. Z výše uvedeného plyne, že odhady autoregresních koeficientů $(a_1, a_2, \dots, a_p)$ dané maximalizací principiální částí Radon–Nikodymovy derivace jsou odhady získané momentovou metodou, neboť jsme řešili systém rovnic

$$\widehat{R}_j = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{\lambda^{2j} \mathrm{d}\lambda}{\left|\sum_{k=0}^{p} a_{p-k} (i\lambda)^k\right|^2}, \quad j = 0, 1, \dots, p-1.$$

B) Nechť $\varphi_a(\cdot)$ je spektrální hustota pro AR-model s koeficienty $(a_1, a_2, \dots, a_p)$, nechť $\widehat{\varphi}(\cdot)$ je jiná hustota a uvažujme funkcionál tvaru

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

za předpokladu konečnosti $\int_{-\infty}^{+\infty} \lambda^{2p}\, \widehat{\varphi}(\lambda)\, \mathrm{d}\lambda$.

Hledejme minimum tohoto funkcionálu přes všechny možné spektrální hustoty $\varphi_a(\cdot)$ autoregresního modelu. Tuto úlohu lze tedy chápat také takto: Mějme dán gaussovský stacionární proces $\{x(t),\, t \in \langle 0, T\rangle\}$, který má derivace až do řádu $(p-1)$ a chceme najít mezi všemi autoregresními gaussovskými procesy řádu p takový, který by byl co nejvíce podoben výchozímu procesu $\{x(t),\, t \in \langle 0, T\rangle\}$ v tom smyslu, aby míra podobnosti byla maximální, tedy minimální vzdálenost měřená asymptotickou rychlostí I-divergence. Lze ukázat, že takové minimum existuje mezi autoregresními procesy p-tého řádu a je právě dáno výše popsanou metodou v bodě A). Za tímto účelem spočítejme rozdíl

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

díky nezápornosti I-divergenční rychlosti. Tím je dokázáno, že $F(\varphi_b(\cdot)) \geq F(\varphi_{\widetilde{a}}(\cdot))$ a jednoznačnost plyne z vlastností funkce $x - \ln x - 1$. ∎

Na závěr uvedeme větu, jejíž důkaz bezprostředně plyne z vlastností stacionárních autoregresních gaussovských modelů.

Věta 2. Pro stacionární AR-gaussovský proces platí

$$\lim_{T\to\infty}\frac{1}{T}\max_{a\in\Theta_p}\ln\frac{\mathrm{d}P_a(x(\cdot))}{\mathrm{d}P_{a_0}(x(\cdot))} = \frac{1}{2}\,\overline{I}(P_{a_*} : P_{a_0}),$$

kde Θ_p je je parametrický prostor a a_* je správná hodnota autoregresních parametrů.

Důkaz. Věta plyne z vlastností superkonsistence maximálně věrohodných odhadů koeficientů $(a_1, \ldots, a_p)$ autoregresního gaussovského procesu a ze vztahu mezi Radon–Nikodymovou derivací $\frac{\mathrm{d}P_a}{\mathrm{d}P_{a_0}}$ a její principiální částí, neboť platí

$$\ln\frac{\mathrm{d}P_a}{\mathrm{d}P_{a_0}} = PP\ln\frac{\mathrm{d}P_a}{\mathrm{d}P_{a_0}} + o(T) \quad \text{při } T\to\infty.$$

Dále z předchozího víme, že $\frac{1}{T}\,PP\ln\frac{\mathrm{d}P_a}{\mathrm{d}P_{a_0}} = \frac{1}{2}I(\widetilde{\varphi} : \varphi_a)$, kde spektrální hustota $\widetilde{\varphi}(\cdot)$ je určena pomocí výběrových charakteristik $(\widehat{R}_0, \widehat{R}_1, \ldots, \widehat{R}_{p-1})$, které při $T\to\infty$ konvergují ke skutečným hodnotám $R_0, R_1, \ldots, R_{p-1}$ s.j.

Literatura

[1] Michálek J. (1990), *Asymptotic Rényi's rate of Gaussian proceses*. Prob. Control and Inf. Th. Vol. 10, **3**, 209 – 227.

[2] Whittle P. (1952), *Estimation and information in time series analysis*. Skand. Aktuar. **35**, 48 – 60.

[3] Dzhaparidze K. (1986), *Parameter Estimation and Hypothesis Testing in Spectral Analysis of Stationary Time Series*. Springer-Verlag, Heidelberg.

[4] Michálek J. (1990), *Yule – Walker estimates and asymptotic I-divergence rate*. Probl. Control and Inf. Th. 5 – 6, **19**, 387 – 398.

[5] Hájek J. (1960), *On a simple regression model in Gaussian processes*. Trans. of Second Prague Conference on Inf. Theory. Publishing House of Czechoslov. Academy of Sciences, Praha, 185 – 198.

[6] Hájek J. (1962), *On linear statistical problems in stochastic processes*. Czechoslov. Math. Journal T 12 **87**, 404 – 444.

[7] Pisarenko V. F. (1962), *On parameter estimation of a Gaussian stationary processes with a spectral density function*. Lithuanian Math. Journal 27 – 43. (rusky)

[8] Rozanov I. A. (1962), *On the derivative of a Gaussian probability measure with respect to another one*. Teorija verojatnostej i prim. **7(1)**, 84 – 89. (rusky)