VÍCEROZMĚRNÝ PERIODOGRAM

Daniela Jarušková, FS ČVUT

V časových okamžicích $1,\dots,N$ pozorujeme vícerozměrnou časovou řadu $\underline{X}_1 = (X_{11},\dots,X_{1n})',\dots,\underline{X}_N = (X_{N1},\dots,X_{Nn})'$, o které předpokládáme, že je součtem deterministické periodické složky $m(t) = (m_1(t),\dots,m_n(t))'$ a vícerozměrného bílého šumu $e(t) \sim N(0,\sigma^2 I)$:

$$X_t = m(t) + e(t) .$$

O periodické složce předpokládáme, že každá její souřadnice se dá vyjádřit jako součet kosinů s frekvencemi $\Theta_1,\dots,\Theta_R$ tj.

$$m_j(t) = \sum_{r=1}^{R} \varrho_{jr} \cos(t\Theta_r + \phi_{jr}) .$$

Cílem statistického zpracování je odhadnout periodickou složku $m(t)$. Tento problém se v základě dělí na dva problémy. Za prvé najít množinu významných frekvencí $\Theta = \{\Theta_1,\dots,\Theta_R\}$, to znamená najít jejich počet a hodnoty, a za druhé najít příslušné amplitudy $\{\varrho_{jr}\}$ a fázová posunutí $\{\phi_{jr}\}$. Druhý problém je jednodušší, neboť známe-li Θ, můžeme amplitudy a fázová posunutí spočítat metodou nejmenších čtverců.

Problém první se řeší pomocí teorie Fourierovy transformace. Označme Fourierův obraz vícerozměrné posloupnosti $\underline{Z}_1,\dots,\underline{Z}_N$ $\overset{z}{\underline{v}}(\lambda) = \frac{1}{\sqrt{2\pi N}} \sum_{t=1}^{N} e^{-it\lambda} \underline{Z}_t$. Součin Fourierova obrazu $\overset{z}{\underline{v}}(\lambda)$ vícerozměrné posloupnosti $\underline{Z}_1,\dots,\underline{Z}_N$ s transponovaným a komplexně sdruženým Fourierovým obrazem $\overset{y}{\underline{v}}(\lambda)$ vícerozměrné posloupnosti $\underline{Y}_1,\dots,\underline{Y}_N$ označme $\overset{zy}{\underline{\underline{I}}}(\lambda) = \overset{z}{\underline{v}}(\lambda)\overset{y}{\underline{v}}^*(\lambda)$. Matici typu $n\times n$ $\overset{xx}{\underline{\underline{I}}}(\lambda)$ budeme říkat vícerozměrný periodogram.

Nejprve zkoumejme periodogram $\overset{mm}{\underline{\underline{I}}}(\lambda)$ odpovídající deterministické periodické složce $\underline{m}(t)$. Prvek $\overset{mm}{I}_{jk}(\lambda)$ ležící v j-tém řádku a k-tém sloupci matice $\overset{mm}{\underline{\underline{I}}}(\lambda)$ lze vyjádřit pomocí dvojné sumy $\overset{mm}{I}_{jk}(\lambda) = \sum_{r=1}^{R}\sum_{s=1}^{R} I_{jk}^{rs}(\lambda)$, kde $I_{jk}^{rs}(\lambda)$ je jednorozměrný periodogram odpovídající kosinu s frekvencí Θ_r souřadnice $m_j(t)$ a kosinu s frekvencí Θ_s souřadnice $m_k(t)$, tj.

$$I_{jk}^{rs}(\lambda) = \frac{1}{2\pi N} \sum_{t=1}^{N} e^{-it\lambda} \varrho_{jr} \cos(t\Theta_r + \phi_{jr}) \sum_{\tau=1}^{N} e^{i\tau\lambda} \varrho_{ks} \cos(\tau\Theta_s + \phi_{ks})$$

Použijeme-li toho, že $\cos x = \frac{e^{ix} + e^{-ix}}{2}$ a sečteme-li prvních N členů geometrické řady, můžeme najít vyjádření

$$I_{jk}^{rs}(\lambda) = \frac{\rho_{jr}\,\rho_{ks}}{8\pi N}\Big[\exp\Big\{ i\big\{(\phi_{ks}-\phi_{jr}) + \tfrac{1}{2}(N+1)(\Theta_s-\Theta_r)\big\}\Big\}\, S_N(\lambda-\Theta_r)\cdot S_N(\lambda-\Theta_s)$$
$$+ \exp\Big\{-i\big\{(\phi_{ks}-\phi_{jr}) - \tfrac{1}{2}(N+1)(\Theta_s-\Theta_r)\big\}\Big\}\, S_N(\lambda+\Theta_r)\cdot S_N(\lambda+\Theta_s)$$
$$+ \exp\Big\{ i\big\{(\phi_{ks}+\phi_{jr}) + \tfrac{1}{2}(N+1)(\Theta_s+\Theta_r)\big\}\Big\}\, S_N(\lambda+\Theta_r)\cdot S_N(\lambda-\Theta_s)$$
$$+ \exp\Big\{-i\big\{(\phi_{ks}+\phi_{jr}) - \tfrac{1}{2}(N+1)(\Theta_s+\Theta_r)\big\}\Big\}\, S_N(\lambda-\Theta_r)\cdot S_N(\lambda+\Theta_s)\Big],$$

kde $S_N(\lambda) = \sin(\frac{1}{2}N\lambda)\,/\,\sin(\frac{1}{2}\lambda)$.

Uvažujeme-li $\Theta_r > 0$, $\Theta_s > 0$, $\lambda > 0$, pak asymptoticky pro $N \to \infty$ pro $r=s$ a $\lambda = \Theta_r$ je $|I_{jk}^{rs}(\lambda)|$ řádu N a pro ostatní λ je řádu $1/N$. Zároveň asymptoticky platí $\overset{mm}{I}_{jk} \cong \sum_{r=1}^{a} I_{jk}^{rr}$ a $|\overset{mm}{I}_{jk}|^2 \cong \sum_{r=1}^{a} |I_{jk}^{rr}(\lambda)|^2$. Z toho vyplývá, že eukleidovská norma matice $\underline{\underline{I}}(\lambda)$ tj. odmocnina ze součtu kvadrátů absolutních hodnot jednotlivých prvků, je dobrým ukazatelem periodicit.

Vrátíme-li se k našemu modelu, pak pro střední hodnotu kvadrátu normy periodogramu $\overset{xx}{\underline{\underline{I}}}(\lambda)$, jehož prvky jsou nyní náhodné funkce, platí

$$E\,\|\overset{xx}{\underline{\underline{I}}}(\lambda)\|^2 = \sum_{j=1}^{n}\sum_{k=1}^{n}\big(|\overset{mm}{I}_{jk}(\lambda)|^2 + \sigma^2\,\overset{mm}{I}_{kk}(\lambda) + \sigma^2\,\overset{mm}{I}_{jj}(\lambda) + \sigma^4\big)$$

a tedy bude řádu N^2 pro $\lambda = \Theta_i \in \Theta$ a jinde omezená. Statistika $\|\overset{xx}{\underline{\underline{I}}}(\lambda)\|$ je tedy vhodná pro testování nulové hypotézy $\underline{m}(t) = 0$. Navíc platí

$$\|\overset{xx}{\underline{\underline{I}}}(\lambda)\|^2 = \sum_{j=1}^{n}\sum_{k=1}^{n} \overset{x}{v}_j(\lambda)\,\overset{x}{v}_k^{*}(\lambda)\,\overset{x}{v}_j^{*}(\lambda)\,\overset{x}{v}_k(\lambda) = (\operatorname{tr}\,\overset{xx}{\underline{\underline{I}}}(\lambda))^2$$

a tedy $\|\overset{xx}{\underline{\underline{I}}}(\lambda)\| = \operatorname{tr}\,\overset{xx}{\underline{\underline{I}}}(\lambda) = \sum_{j=1}^{n}\overset{xx}{I}_{jj}(\lambda)$. Předpokládáme-li, že všechna $\Theta_i \in \Theta$ jsou tvaru $\lambda_p = \frac{2\pi p}{N}$, kde $p \in \{1,\dots,N/2 - 1\}$, stačí hledat rozdělení $\|\overset{xx}{\underline{\underline{I}}}(\lambda_p)\| = \sum_{j=1}^{n}\overset{xx}{I}_{jj}(\lambda_p)$.

Za platnosti nulové hypotézy, tj. $\underline{X}_t \sim N(0,\sigma^2 I)$ budou náhodné veličiny $4\pi\,\overset{xx}{I}_{jj}(\lambda_p)/\sigma^2$, $j=1,\dots,n$, $p=1,\dots,N/2 - 1 = m$, nezávislé stejně rozdělené s rozdělením χ^2 o dvou stupních volnosti, tj. s hustotou $f(x) = \frac{1}{2}e^{-x/2}$ a tedy $2\pi\|\overset{xx}{\underline{\underline{I}}}(\lambda_p)\|/\sigma^2 \sim G(1,n)$ s hustotou $f(x) = \frac{e^{-x}\,x^{n-1}}{\Gamma(n)}$. V praxi ovšem σ^2 většinou neznáme, a tedy ho nahrazujeme konzistentním odhadem $\frac{4\pi}{nN}\sum\|\overset{xx}{\underline{\underline{I}}}(\lambda_p)\|$. Statistika používaná pro testování nulové hypotézy je pak největší pořádková statistika $g = \max\|\overset{xx}{\underline{\underline{I}}}(\lambda_p)\|/\sum\|\overset{xx}{\underline{\underline{I}}}(\lambda_p)\|$. Její přesné rozdělení lze odvodit (viz [3]) z faktu, že se jedná o největší pořádkovou statistiku výběru z Dirichletova rozdělení. Pro $m \leq 50$ lze přibližně použít vztahu

$$(1) \qquad P(g > d) \doteq m\sum_{k=0}^{n-1}\binom{mn-1}{k} d^k (1-d)^{mn-k-1} .$$

Pro větší m lze použít asymptotického testu opírajícího se o statistiku $g_m^{*} = mn\,\max\|\overset{xx}{\underline{\underline{I}}}(\lambda_p)\|\,/\sum\|\overset{xx}{\underline{\underline{I}}}(\lambda_p)\|$. Protože pro velká m má statistika

$m\,n\|\underline{\underline{I}}^{xx}(\lambda_p)\| / \sum \|\underline{\underline{I}}^{xx}(\lambda_p)\|$ přibližně $G(1,n)$ rozdělení, jedná se o problém nalezení největší, tj. m-té pořádkové statistiky výběru o $G(1,n)$ rozdělení. Platí

$$(2) \qquad P\left(g_m^* - \left\{\log m + (n-1)\log\log m\right\} < x\right) = \exp\left\{-e^{-x}/\Gamma(n)\right\} .$$

Pro nalezení množiny Θ se použije shora popsaný test tím způsobem, že pokud zamítneme nulovou hypotézu, tj. g (resp. g_m^*) > příslušný kvantil, pak frekvenci λ_i odpovídající statistice g (resp. g_m^*) zařadíme do množiny Θ . Dále pokračujeme tak, že vyloučíme frekvenci λ_i a sestrojíme novou statistiku $g = \max_{\lambda_p \neq \lambda_i} \|\underline{\underline{I}}(\lambda_p)\| / \sum_{\lambda_p \neq \lambda_i} \|\underline{\underline{I}}(\lambda_p)\|$ a testujeme, zda je větší než příslušný kvantil atd.

1% kvantily Fisherova testu pro vícerozměrný periodogram

a) získané výpočtem podle (1)

m \ n	1	2	3	4	5
10	0.53581	0.39322	0.33080	0.29452	0.27037
15	0.40661	0.28818	0.23355	0.21039	0.19189
20	0.32963	0.22868	0.18752	0.16462	0.14957
25	0.27817	0.19040	0.15521	0.13561	0.12288
30	0.24124	0.16351	0.13270	0.11562	0.10457
35	0.21338	0.14353	0.11608	0.10093	0.09114
40	0.19157	0.12808	0.10330	0.08965	0.08036
45	0.17401	0.11576	0.09313	0.08071	0.07272
50	0.15955	0.10569	0.08487	0.07346	0.06613

b) získané výpočtem podle (2)

64	0.13694	0.07960	0.05638	0.04394	0.03527
128	0.07388	0.04311	0.03105	0.02423	0.01968
256	0.03965	0.02317	0.01677	0.01318	0.01080
512	0.02118	0.01238	0.00899	0.00710	0.00585

Literatura

[1] Anděl J. : Statistická analýza časových řad, SNTL, Praha 1976

[2] Priestley M.B. : Spectral Analysis and Time Series , Academic Press, 1981

[3] Mac Neill I.B. : Tests for periodic components in multiple time series, Biometrika 1974 , 61, 1, p.57