

Jiný přístup k analýze časových řad - Box-Jenkinsova metodologie

Napřed některé základní pojmy:

Stacionarita - znamena, že časová řada je v nějakém smyslu dočasově ustálená

slabá stacionarita - předpokladem rozdělení náhodného vektoru $(Y_{t_1}, \dots, Y_{t_k})$ je stejné jako rozdělení vektoru $(Y_{t_1+h}, \dots, Y_{t_k+h})$ pro libovolné h

silná stacionarita - proces má konstantní střední hodnotu, konstantní rozptyl a kovarianční fce invariantní vůči posunutí v čase, tj: $\text{cov}(Y_t, Y_s) = \text{cov}(Y_{t+h}, Y_{s+h})$

Autokoranační funkce (jen pro slabě stacionární řady, což budeme nadále předpokládat nyní)

$$R(k) = \text{cov}(Y_{t+k}, Y_t) = E(Y_{t+k} - \mu)(Y_t - \mu), \quad k = \dots, -1, 0, 1, \dots$$

speciálně $R(k) = E Y_{t+k} Y_t$ když $\mu = 0$

rozptyl je roven $R(0)$

Autokorelační funkce $\rho_k = \frac{R(k)}{R(0)}$

(obdobá korelační koeficientu)

obě funkce jsou sudé, takže se dá omezit na $k \geq 0$

rovněž $\rho_0 = 1$

Odhady - počítají se z řady delší alespoň 50

odhad střední hodnoty: $\bar{Y} = \frac{1}{n} \sum_{t=1}^n Y_t$

odhad autokovarianční fce: $C_k = \sum_{t=1}^{n-k} \frac{(Y_t - \bar{Y})(Y_{t+k} - \bar{Y})}{n}$ (někdy $\approx$ dělí $n-k$)

odhad autokorelační fce: $r_k = \frac{C_k}{C_0}$

Příklady stacionárních procesů:

Bílý šum - posloupnost navzájem nezávislých náh. veličin s nulovou střední hodnotou a konstantním rozptylem

$$R(0) = \sigma^2, \quad R(k) = 0 \text{ pro } k \neq 0, \quad \rho_0 = 1, \quad \rho_k = 0 \text{ pro } k \neq 0$$

Posloupnost klouzavého průměru MA(q)

$$Y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}$$

kde ε_t je bílý šum, $\theta_1, \dots, \theta_q$ parametry, q je řád procesu (libovolný)

zřejmě $EY_t = 0$

$$\text{var } Y_t = (1 + \theta_1^2 + \dots + \theta_q^2) \sigma_\varepsilon^2 = R(0)$$

$$R(k) = 0 \text{ pro } k > q$$

$$\begin{aligned} R(k) = EY_{t+k} Y_t &= E(\varepsilon_{t+k} + \theta_1 \varepsilon_{t+k-1} + \dots + \theta_q \varepsilon_{t+k-q}) \times \\ &\quad \times (\varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}) = \end{aligned}$$

$$= (\theta_k + \theta_1 \theta_{k+1} + \dots + \theta_{q-k} \theta_q) \sigma_\varepsilon^2$$

např. MA(1) : $y_t = \varepsilon_t + \theta \varepsilon_{t-1}$

$$R(1) = E y_{t+1} y_t = E (\varepsilon_{t+1} + \theta \varepsilon_t) (\varepsilon_t + \theta \varepsilon_{t-1}) = \theta \sigma_\varepsilon^2$$

$$R(0) = (1 + \theta^2) \sigma_\varepsilon^2 = E y_t y_t = E (\varepsilon_t + \theta \varepsilon_{t-1})^2$$

$$R(k) = 0 \text{ pro } k > 1$$

$$\rho_k = \frac{R(k)}{R(0)} = \frac{\theta}{1 + \theta^2} \text{ pro } k=1$$

da' se napíše i jinak: $\varepsilon_t = y_t - \theta \varepsilon_{t-1}$, pokud $|\theta| < 1$

$$\begin{aligned} \text{pak } y_t &= \varepsilon_t + \theta (y_{t-1} - \theta \varepsilon_{t-2}) = \\ &= \varepsilon_t + \theta y_{t-1} - \theta^2 \varepsilon_{t-2} = \varepsilon_t + \theta y_{t-1} - \theta^2 (y_{t-2} - \theta \varepsilon_{t-3}) = \\ &= \dots = \varepsilon_t + \theta y_{t-1} - \theta^2 y_{t-2} + \theta^3 y_{t-3} - \dots \end{aligned}$$

To znamená, že hodnota procesu v čase t je závislá na řadě minulých hodnotách prostřednictvím parametru θ , přičemž vliv vzdálenějších členů čím dál víc slabne (plus náhodná složka ε_t).

Autoregresní proces AR(p)

$$y_t = g_1 y_{t-1} + g_2 y_{t-2} + \dots + g_p y_{t-p} + \varepsilon_t$$

kde p je řád modelu, $g_1, \dots, g_p$ parametry (*)

speciálně AR(1): $y_t = g y_{t-1} + \varepsilon_t$, $|g| < 1$

(*) musí splňovat podmínku, že řeší nějaký lineární polynom

$$f(y) = 1 - g_1 y - \dots - g_n y^n$$

tedy má jednorázové kořeny

$$\text{AR}(1): f(y) = 1 - g y = 0$$

$$y = \frac{1}{g} \quad \text{má jednorázové kořeny}$$

tedy g smí být $\neq 0$

hady napíšat se dá napísať $Y_t = \varepsilon_t + \varphi \varepsilon_{t-1} + \varphi^2 \varepsilon_{t-2} + \dots$

tj. jako lineární proces

krájme $EY_t = 0$

$$R(k) = EY_{t+k}Y_t = E(\varphi Y_{t+k-1} + \varepsilon_{t+k})Y_t = \varphi EY_{t+k-1}Y_t + E\varepsilon_{t+k}Y_t$$

čili $R(k) = \varphi R(k-1)$ pro $k \geq 1$

$$\begin{aligned} R(0) &= EY_tY_t = E(\varphi Y_{t-1} + \varepsilon_t)(\varphi Y_{t-1} + \varepsilon_t) = \\ &= \varphi^2 R(0) + \sigma_\varepsilon^2 \Rightarrow R(0) = \frac{\sigma_\varepsilon^2}{1 - \varphi^2} \end{aligned}$$

$$R(k) = \varphi R(k-1) = \varphi^2 R(k-2) = \dots = \varphi^k R(0) = \frac{\varphi^k}{1 - \varphi^2} \sigma_\varepsilon^2$$

$$\rho_k = \frac{R(k)}{R(0)} = \varphi^k \quad \%$$

Smíšený proces ARMA(p, q)

$$Y_t = \varphi_1 Y_{t-1} + \dots + \varphi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}$$

Modely pro nestacionární řady

ARIMA - od prvního řady se přijde k prvnímu, druhému, atd. diferenciám, až dostaneme řadu, která je stacionární, po které konstruuje model typu ARMA

např. řada o lineárním trendu lze cyklických složek -

stačí přijít k prvnímu diferenciám

SARIMA - pro řady se sezónními složkami

%. Pro $AR(p)$:

$$EY_t = 0$$

$$R(k) = \varphi_1 R(k-1) + \varphi_2 R(k-2) + \dots + \varphi_p R(k-p) \quad \text{pro } k > 0$$

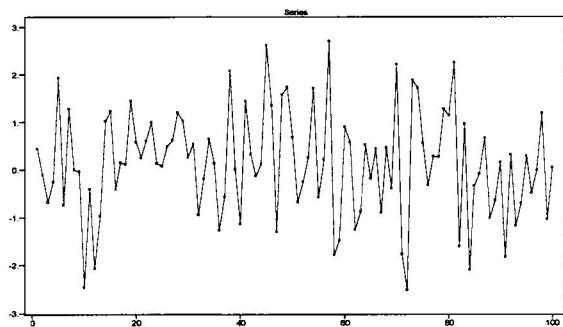
$$R(0) = \varphi_1 R(1) + \varphi_2 R(2) + \dots + \varphi_p R(p) + \sigma_\varepsilon^2$$

to je soustava tzv. Yuleových-Walkerových rovnic

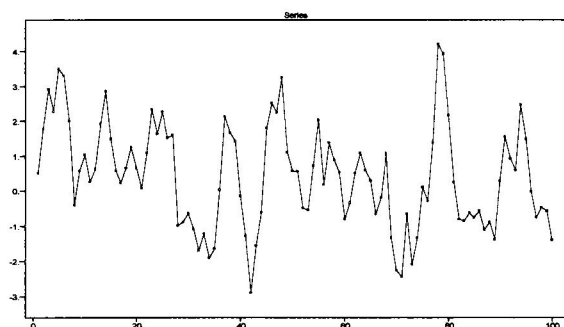
Analytické řešení je dost komplikované - jsou to lineární
diferenční rovnice.

Řeší se algoritmicky

Simulovaný bílý šum

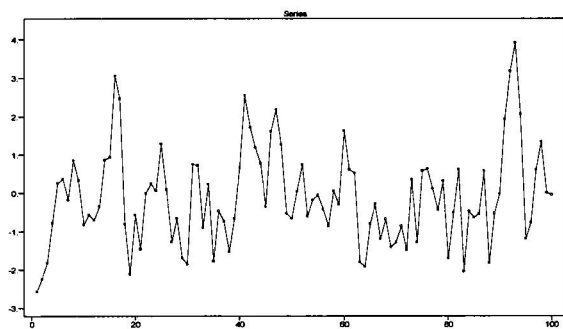


Simulovaná řada MA(2)



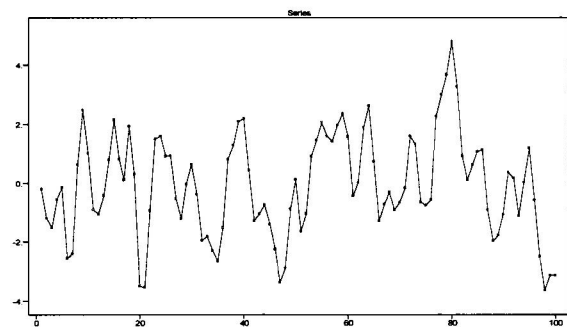
$$X_t = Z_t + Z_{t-1} + 0.6 Z_{t-2}$$

Simulovaná řada AR(1)



$$X_t = 0.7X_{t-1} + Z_t$$

Simulovaná řada ARMA(1,1)



$$X_t = 0.5 X_{t-1} + Z_t + 0.75 Z_{t-1}$$