

# ČASOVÉ ŘADY

VÍTĚZSLAV VESELÝ

## PODPORA VÝUKY NA POČÍTAČOVÉ SÍTI

<http://www.math.muni.cz/~vesely>

**Algoritmy pro časové řady v MATLABu:**

`d = vesely('cas_rady')` ... připojení a výpis obsahu knihovny

`d` ... výstup=úplná cesta do adresáře knihovny.

### 1. ÚVOD

#### 1.1. Ukázky časových řad.

**Časová řada** = soubor pozorování nějaké veličiny  $\{x_t, t \in T\}$ , kde  $t$  je zpravidla čas a  $T \subseteq \mathbb{R}$ .

**Obr. 1.1:**

Měření proudu procházejícího odporem  $r$  v obvodu se střídavým napětím,  
 $v(t) = a \cos(\omega t + \Theta)$ , tj.  $x_t = \frac{a}{r} \cos(\omega t + \Theta)$ .

**Obr. 1.2 (MATLAB: `getdata('brockwell.dat/uspop.dat')`):**

Růst populace v USA v letech 1790–1980 sledovaný v desetiletých intervalech.

**Obr. 1.3 (MATLAB: `getdata('brockwell.dat/strikes.dat')`):**

Počet stávek v USA v letech 1951–1980.

**Obr. 1.4:** Výsledky Národní a Americké ligy v baseballu v letech 1933–1980.

**Obr. 1.5 (MATLAB: `getdata('brockwell.dat/sunspots.dat')`):**

Počty slunečních skvrn v letech 1770–1869.

**Obr. 1.6 (MATLAB: `getdata('brockwell.dat/deaths.dat')`):**

Počet úmrtí při nehodách v USA v letech 1973–1978.

#### 1.2. Oblasti uplatnění metod analýzy časových řad.

**fyzika, technika:**

- seismický záznam v geofyzice.
- řada nejvyšších denních teplot v meteorologii.
- průběh výstupního signálu určitého elektrického přístroje.
- tenzometrické měření povrchového napětí v provozu namáhané strojní součástky.

**biologie, ekologie:** sledování různých parametrů znečištění ovzduší.

**medicína:** záznam EKG nebo EEG.

**společenské vědy:** změny v počtu a složení obyvatelstva.

**sociologie:** vývoj rozvodovosti.

**EKONOMIE:** Teorie časových řad = jedna z nejdůležitějších kvantitativních metod pro analýzu ekonomických dat, např.:

- analýza poptávky po určitém výrobku
- analýza objemu zemědělské produkce
- analýza počtu cestujících v letecké dopravě
- analýza vývoje kurzu akcií na burze

### 1.3. Cíl analýzy.

#### Porozumění mechanismu, jímž se generují sledované údaje.

Znalost modelu tohoto mechanismu  $\Rightarrow$  znalost algoritmu, jímž můžeme chování tohoto mechanismu simulovat na počítači  $\Rightarrow$  schopnost popsat s jistou přesností jeho chování:

- mezi časovými okamžiky měření (**interpolace**)
- v budoucnosti (**extrapolace, prognóza**)
- s cílem řídit a optimalizovat činnost určitého systému vhodnou volbou vstupních a počátečních podmínek (**regulace**), např. regulace složitých technologických procesů.

### 1.4. Některé specifické problémy analýzy časových řad.

#### a) Volba okamžiků pozorování:

- jsou přímo **diskrétní svou povahou**, např. úroda obilí za jednotlivé roky.
- vznikají **diskretizací spojité časové řady**, např. teplota v danou denní dobu na daném místě.
- vznikají **akumulací (agregací)** hodnot za určité období, např. denní množství srážek, roční výroba závodu. Místo akumulace se někdy provádí průměrování.

Je-li dána možnost volby, je třeba jí věnovat pozornost:

- **málo bodů**  $\Rightarrow$  unikne charakteristický rys řady.
- **mnoho bodů**  $\Rightarrow$  zvýší se výpočetní náročnost.
- **ekvidistantní diskretizace** zpravidla usnadní numerické zpracování, ale neumožňuje adaptivně měnit hustotu diskretizace v závislosti na lokálním charakteru řady.
- **při agregaci** se mohou porušit vlastnosti původní řady.

#### b) Problémy s kalendářem:

- různá délka kalendářních měsíců.
- 4 nebo 5 víkendů v měsíci.
- různý počet pracovních dnů v měsíci.
- pohyblivé svátky: např. svátek na začátku měsíce sníží prodej potravin za tento měsíc, ale zvýší jej za předchozí v důsledku efektu předzásobení.

Příklad 'očistění' časové řady např. od proměnlivé délky měsíce:

zavedeme 'standardní' měsíc o délce 30 dnů a pak údaj třebas o produkci za leden přenásobíme korekčním faktorem  $\frac{30}{31}$ .

#### c) Problémy s nekompatibilitou jednotlivých měření:

Příklad: hodnota nějakého ukazatele se jeden rok týká např. 85 podniků, další rok jen 82 apod.

#### d) Problémy s délkou časových řad:

- zvětšení počtu měření (např. půlením časových intervalů mezi body) nemusí vždy znamenat zvětšení množství informace.
- někdy se mohou objevit protichůdné tendence: metoda zpracování vyžaduje delší řadu, ale na druhé straně řada vzniklá dlouhodobým sledováním může měnit charakteristiky svého modelu v čase  $\Rightarrow$  obtíže s konstrukcí modelu.

### 1.5. Označení a základy matematické statistiky.

$s := v$ , resp.  $v := s \dots$  označení výrazu  $v$  symbolem  $s$ .

#### Číselné obory:

$\mathbb{N} \dots$  množina všech přirozených čísel

$\mathbb{N}_0 = \mathbb{N} \cup \{0\} \dots$  množina všech nezáporných celých čísel

$\mathbb{Z} \dots$  množina všech celých čísel

$\mathbb{R} \dots$  množina všech reálných čísel

$\mathbb{R}^+ \dots$  množina všech nezáporných reálných čísel

$\mathbb{C} \dots$  množina všech komplexních čísel

$(\cdot)^+ : \mathbb{R} \rightarrow \mathbb{R}^+ \dots$  zobrazení definované předpisem  $(x)^+ = \max(0, x)$

$(a, b) \dots$  otevřený interval

$[a, b]$  ... uzavřený interval

$$\mathcal{J}(a, b) = \{x \mid \min(a, b) < x < \max(a, b)\}$$

$$\mathcal{J}[a, b] = \{x \mid \min(a, b) \leq x \leq \max(a, b)\}.$$

### Vektory:

$\mathbf{x} = (x_1, \dots, x_n)^T \in \mathbb{C}^n$  ... zpravidla sloupcové

$$\mathbf{x} + \mathbf{h} = (x_1 + h_1, \dots, x_n + h_n)^T, \quad \mathbf{h} \in \mathbb{C}$$

$$\mathbf{t} = (t_1, \dots, t_k)^T \in \mathbb{N}^k, \quad t_i \in \{1, \dots, n\} \text{ pro } i = 1, \dots, k, \quad k \leq n \quad \Rightarrow \quad \mathbf{x}_{\mathbf{t}} := (x_{t_1}, \dots, x_{t_k})^T \in \mathbb{C}^k$$

$$1 \leq i \leq n \quad \Rightarrow \quad \mathbf{x}(i) = (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n)^T$$

$$f(\mathbf{x}) = f(x_1, \dots, x_n), \quad d\mathbf{x} = dx_1 \dots dx_n$$

$\mathbf{0}, \mathbf{0}_{n \times 1}$  ... nulový sloupcový vektor délky  $n$ .

### Matice:

$A, A_{m \times n} = [a_{ij}] = [A(i, j)]$  ... matice rozměru  $m \times n$

$\mathcal{R}(A) = \{\mathbf{y} \mid \mathbf{y} = A\mathbf{x}\}$  ... obor hodnot operátoru  $A$

$\mathcal{N}(A) = \{\mathbf{x} \mid A\mathbf{x} = \mathbf{0}\}$  ... jádro operátoru  $A$

$A^T = [a_{ji}]$  ... transponovaná matice

$A^* = [\bar{a}_{ji}]$  ... hermitovsky sdružená matice

$I, I_n = I_{n \times n} = [\delta_{ij}]$  ... jednotková matice řádu  $n$

$|A|$  ... determinant čtvercové matice  $A$

$\mathbf{0}, \mathbf{0}_{m \times n} = [\mathbf{0}]$  ... nulová matice rozměru  $m \times n$

$$\text{diag}(\mathbf{x}) = \begin{bmatrix} x_1 & 0 & \dots & 0 \\ 0 & x_2 & \dots & 0 \\ & & \ddots & \\ 0 & 0 & \dots & x_n \end{bmatrix} \dots \text{diagonální matice}$$

$A(i, :) = (a_{i1}, \dots, a_{in}) =: \mathbf{r}_i$  ...  $i$ -tý řádek matice  $A$  ve stylu MATLAB

$A(:, j) = (a_{1j}, \dots, a_{mj})^T =: \mathbf{s}_j$  ...  $j$ -tý sloupec matice  $A$  ve stylu MATLAB

$A = [r_1; \dots; r_m] = [s_1, \dots, s_n]$  ... blokový zápis matice  $A$  ve stylu MATLAB

$A > 0$  (resp.  $A \geq 0$ ) ... pozitivně (semi)definitní matice.

### Norma a skalární součin vektorů:

$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^n x_i \bar{y}_i = \mathbf{y}^* \mathbf{x}$ , speciálně:

$\mathbf{y} = A^* \mathbf{x} \Leftrightarrow y_i = \langle \mathbf{x}, A(:, i) \rangle$  pro  $i = 1, 2, \dots, n$

$\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle} = \sqrt{\sum_{i=1}^n |x_i|^2}$  ... Euklidovská norma.

### Schwarzova nerovnost:

$|\langle \mathbf{x}, \mathbf{y} \rangle| \leq \|\mathbf{x}\| \|\mathbf{y}\|$ , kde rovnost nastane právě když vektory  $\mathbf{x}$  a  $\mathbf{y}$  jsou lineárně závislé.

### Pravděpodobnostní prostor $(\Omega, \mathcal{A}, P)$ :

$\Omega$  ... základní prostor elementárních jevů

$\mathcal{A} \subseteq 2^\Omega$  ...  $\sigma$ -algebra náhodných jevů

$P$  ... pravděpodobnostní míra na  $\mathcal{A}$

Všechny náhodné veličiny budeme vždy uvažovat nad tímž pravděpodobnostním prostorem. Komplexní náhodnou veličinou budeme rozumět veličinu  $X = X_r + iX_i$ , kde  $X_r$  a  $X_i$  jsou reálné náhodné veličiny představující po řadě reálnou a imaginární část  $X$

$\mathbb{X} = (X_1, \dots, X_n)^T$  ... (komplexní) náhodný vektor tvořený (komplexními) náhod. veličinami  $X_i$ .

### Střední hodnota:

$\mu = \mu_X = EX$  ... střední hodnota náhodné veličiny  $X$

$\boldsymbol{\mu} = \boldsymbol{\mu}_{\mathbb{X}} = E\mathbb{X} = (EX_1, \dots, EX_n)^T$  ... střední hodnota náhodného vektoru  $\mathbb{X}$ .

### Rozptyl a kovariance náhodných veličin:

$\sigma^2 = \sigma_X^2 = \text{var}X = E|X - EX|^2 = E|X|^2 - |EX|^2 \geq 0$  ... rozptyl  $X$

$\sigma_{XY} = \text{cov}(X, Y) = E(X - EX)(\overline{Y - EY}) = EX\bar{Y} - (EX)(\overline{EY})$  ... kovariance  $X$  a  $Y$

$\text{cov}(X, X) = \text{var}X$ ,  $\text{cov}(Y, X) = \text{cov}(X, Y)$

$\text{cov}(\sum_r X_r, \sum_s Y_s) = \sum_r \sum_s \text{cov}(X_r, Y_s)$  a odtud speciálně:

$\text{var}(X + Y) = \text{var}X + \text{cov}(X, Y) + \text{cov}(Y, X) + \text{var}Y = \text{var}X + 2\text{Re cov}(X, Y) + \text{var}Y$ .

### Varianční a kovarianční matice náhodných vektorů:

$\Sigma_{\mathbb{X}} = \text{var}\mathbb{X} = [\text{cov}(X_i, X_j)] = E(\mathbb{X} - E\mathbb{X})(\mathbb{X} - E\mathbb{X})^* = E\mathbb{X}\mathbb{X}^* - (E\mathbb{X})(E\mathbb{X})^*$  ... varianční matice  $\mathbb{X}$

$\Sigma_{\mathbb{X}\mathbb{Y}} = \text{cov}(\mathbb{X}, \mathbb{Y}) = [\text{cov}(X_i, Y_j)] = E(\mathbb{X} - E\mathbb{X})(\mathbb{Y} - E\mathbb{Y})^* = E\mathbb{X}\mathbb{Y}^* - (E\mathbb{X})(E\mathbb{Y})^*$  ...

... kovarianční matice  $\mathbb{X}$  a  $\mathbb{Y}$

$\text{cov}(\mathbb{X}, \mathbb{X}) = \text{var}\mathbb{X}$ ,  $\text{cov}(\mathbb{Y}, \mathbb{X}) = \text{cov}(\mathbb{X}, \mathbb{Y})^* \Rightarrow \text{var}\mathbb{X} = (\text{var}\mathbb{X})^* \dots$   
 ... **varianční matice  $\mathbb{X}$  je hermitovská.**

Pro konstantní komplexní vektory  $\mathbf{a}$  a  $\mathbf{c}$  a matice  $B$  a  $D$  odpovídajících rozměrů platí:

$$\text{cov}(\mathbf{a} + B\mathbb{X}, \mathbf{c} + D\mathbb{Y}) = \text{cov}(B\mathbb{X}, D\mathbb{Y}) = B \text{cov}(\mathbb{X}, \mathbb{Y}) D^*.$$

$$\Downarrow \mathbb{X} = \mathbb{Y}$$

$$\text{var}(\mathbf{a} + B\mathbb{X}) = \text{cov}(\mathbf{a} + B\mathbb{X}, \mathbf{a} + B\mathbb{X}) = \text{cov}(B\mathbb{X}, B\mathbb{X}) = B \text{var}(\mathbb{X}) B^*.$$

$$\Downarrow \mathbf{b}^* = B$$

$0 \leq \text{var}(\mathbf{b}^* \mathbb{X}) = \mathbf{b}^* \text{var}\mathbb{X} \mathbf{b} \Rightarrow \text{var}\mathbb{X} \geq 0 \dots$  **varianční matice je pozitivně semidefinitní.**

$\Sigma_{\mathbb{X}}$  je tedy celkem hermitovská a pozitivně semidefinitní a má proto reálná nezáporná vlastní čísla  $\lambda_i$ .

Zřejmě existuje matice  $\Sigma_{\mathbb{X}}^{\frac{1}{2}}$ , jejíž vlastní čísla jsou  $\lambda_i^{\frac{1}{2}}$  a přitom platí:  $\Sigma_{\mathbb{X}} = \Sigma_{\mathbb{X}}^{\frac{1}{2}} \Sigma_{\mathbb{X}}^{\frac{1}{2}}$ .

Dále platí

$$\text{cov}(\sum_r \mathbb{X}_r, \sum_s \mathbb{Y}_s) = \sum_r \sum_s \text{cov}(\mathbb{X}_r, \mathbb{Y}_s) \text{ a odtud speciálně:}$$

$$\text{var}(\mathbb{X} + \mathbb{Y}) = \text{var}\mathbb{X} + \text{cov}(\mathbb{X}, \mathbb{Y}) + \text{cov}(\mathbb{Y}, \mathbb{X}) + \text{var}\mathbb{Y} = \text{var}\mathbb{X} + 2\text{Re cov}(\mathbb{X}, \mathbb{Y}) + \text{var}\mathbb{Y}.$$

## 1.6. Normální rozdělení a rozdělení z něj odvozená.

### 1. Normální rozdělení:

$$X \sim N(\mu, \sigma^2), \mu = EX, \sigma^2 = \text{var}X \dots$$

... reálná náhodná veličina s **normálním (gaussovským) rozdělením;**

$$f(x) = (\sqrt{2\pi}\sigma)^{-1} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \dots \text{ **hustota}** náhodné veličiny  $X$ ;$$

$$\Phi(t) := Ee^{itX} = e^{it\mu - \frac{1}{2}\sigma^2 t^2} \dots \text{ **charakteristická funkce}** náhodné veličiny  $X$ ;$$

$U \sim N(0, 1) \dots$  **standardizované normální rozdělení** náhodné veličiny  $U$ ;

$u_\alpha = F^{-1}(\alpha) \dots$   $\alpha$ -kvantil pro  $U$ , kde  $F(x) = \int_{-\infty}^x f(t) dt$  značí distribuční funkci  $U$ ;

$$\mathbb{X} \sim N_n(\boldsymbol{\mu}, V), \boldsymbol{\mu} = E\mathbb{X}, V = \text{var}\mathbb{X} \dots$$

... reálný náhodný vektor s  **$n$ -rozměrným normálním rozdělením;**

$$f(\mathbf{x}) = (\sqrt{(2\pi)^n |V|})^{-1} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T V^{-1}(\mathbf{x}-\boldsymbol{\mu})} \dots \text{  **$n$ -rozměrná hustota}** náhodného vektoru  $\mathbb{X}$ ;$$

$$\Phi_n(\mathbf{t}) := Ee^{i\mathbf{t}^T \mathbb{X}} = e^{i\mathbf{t}^T \boldsymbol{\mu} - \frac{1}{2}\mathbf{t}^T V \mathbf{t}} \dots \text{ **charakteristická funkce}** náhodného vektoru  $\mathbb{X}$ ;$$

$\mathbb{U} \sim N_n(\mathbf{0}, I_n) \dots$  **standardizované normální rozdělení** náhodného vektoru  $\mathbb{U}$ .

Platí

$$X \sim N(\mu, \sigma^2) \Rightarrow a + bX \sim N(a + b\mu, b^2\sigma^2) \text{ pro } a, b \in \mathbb{R};$$

$$\mathbb{X} \sim N_n(\boldsymbol{\mu}, V) \Rightarrow \mathbf{a} + B\mathbb{X} \sim N_m(\mathbf{a} + B\boldsymbol{\mu}, BV B^T) \text{ pro } \mathbf{a} \in \mathbb{R}^m \text{ a matici } B = B_{m \times n} \text{ nad } \mathbb{R}.$$

### 2. Rozdělení $\chi^2(n)$ :

Nechť  $U_i \sim N(0, 1)$  pro  $i = 1, \dots, n$  jsou stochasticky nezávislé, pak náhodná veličina

$$C = \sum_{i=1}^n U_i^2 \sim \chi^2(n) \text{ má **Pearsonovo "chí kvadrát" rozdělení o } n \text{ stupních volnosti;}**$$

$\chi_\alpha(n) \dots$   $\alpha$ -kvantil pro  $C$ .

Platí

$$C_i \sim \chi^2(n_i) \text{ pro } i = 1, \dots, m \text{ stochasticky nezávislé} \Rightarrow C = \sum_{i=1}^m C_i \sim \chi^2(n_1 + n_2 + \dots + n_m).$$

### 3. Studentovo $t$ rozdělení:

Nechť  $U \sim N(0, 1)$  a  $C \sim \chi^2(k)$  jsou stochasticky nezávislé, pak náhodná veličina

$$T = \frac{U}{\sqrt{C/k}} \sim t(k) \text{ má **Studentovo } t \text{ rozdělení o } k \text{ stupních volnosti;}**$$

$t_\alpha(k) \dots$   $\alpha$ -kvantil pro  $T$ .

### 4. Fisher-Snedecorovo $F$ rozdělení:

Nechť  $C_1 \sim \chi^2(n_1)$  a  $C_2 \sim \chi^2(n_2)$  jsou stochasticky nezávislé, pak náhodná veličina

$$F = \frac{C_1/n_1}{C_2/n_2} \sim F(n_1, n_2) \text{ má **Fisher-Snedecorovo } F \text{ rozdělení s } n_1 \text{ a } n_2 \text{ stupni volnosti;}**$$

$F_\alpha(n_1, n_2) \dots$   $\alpha$ -kvantil pro  $F$ .

### 1.7. Prostor $L_2(\Omega, \mathcal{A}, P)$ .

$L_2(\Omega, \mathcal{A}, P)$  definujeme jako množinu všech (komplexních) náhodných veličin nad tímž pravděpodobnostním prostorem  $(\Omega, \mathcal{A}, P)$ , které mají konečné druhé momenty (resp. rozptyly - viz dále 1.11), tj.  $L_2(\Omega, \mathcal{A}, P) := \{X \mid X \text{ náhodná veličina nad } (\Omega, \mathcal{A}, P), E|X|^2 < \infty\}$ .

Poznamenejme, že do tohoto prostoru zahrnujeme také všechny konstanty z  $\mathbb{C}$ , které považujeme za náhodné veličiny s nulovým rozptylem.

**Věta 1.8.**  $L_2(\Omega, \mathcal{A}, P)$  je **Hilbertův prostor** se skalárním součinem  $\langle X, Y \rangle = EX\bar{Y}$  a normou  $\|X\|_2 = \sqrt{\langle X, X \rangle} = \sqrt{E|X|^2}$ .

*Důkaz.*  $L_2(\Omega, \mathcal{A}, P)$  je obdobou funkcionálního prostoru  $L_2(\mathcal{J})$  tvořeného funkcemi absolutně integrovatelnými v kvadrátu na intervalu  $\mathcal{J} \subseteq \mathbb{R}$ . Totiž  $E|X|^2 = \int_{\Omega} |X(\omega)|^2 dP(\omega)$ , takže namísto s Lebesgueovým integrálem pracujeme s obecnějším pojetím integrálu, kde Lebesgueova míra je nahrazena pravděpodobnostní mírou  $P$ :

- **Skalární součin  $\langle X, Y \rangle$  existuje a je konečný** pro každé  $X, Y \in L_2(\Omega, \mathcal{A}, P)$ , jak snadno nahlédneme z nerovnosti

$$4|X\bar{Y}| = (|X| + |\bar{Y}|)^2 - (|X| - |\bar{Y}|)^2 \leq (|X| + |\bar{Y}|)^2 + (|X| - |\bar{Y}|)^2 = 2(|X| + |\bar{Y}|)^2,$$

odkud užitím  $|\bar{Y}| = |Y|$  dostáváme

$$|X\bar{Y}| \leq \frac{1}{2}(|X|^2 + |Y|^2),$$

takže

$$|EX\bar{Y}| \leq \int_{\Omega} |X(\omega)\bar{Y}(\omega)| dP(\omega) \leq \frac{1}{2} \left[ \int_{\Omega} |X(\omega)|^2 dP(\omega) + \int_{\Omega} |Y(\omega)|^2 dP(\omega) \right] < \infty.$$

- $L_2(\Omega, \mathcal{A}, P)$  je **vektorovým prostorem**. Je uzavřený na násobení skaláry  $c \in \mathbb{C}$ , neboť  $E|cX|^2 = |c|^2 E|X|^2 < \infty$ . Uzavřenost vzhledem ke sčítání plyne z:

$$|X + Y|^2 \leq (|X| + |Y|)^2 = |X|^2 + 2|XY| + |Y|^2 \Rightarrow E|X + Y|^2 \leq E|X|^2 + 2E|XY| + E|Y|^2 < \infty.$$

- **Ověření, že  $L_2(\Omega, \mathcal{A}, P)$  je úplný, neboli Hilbertův prostor**, je složitější, ale provádí se opět zcela analogicky jako v případě funkcionálního prostoru  $L_2(\mathcal{J})$ . Podrobnosti lze nalézt například v monografii [1, §2.10].

□

**Důsledek 1.9** (Schwarzova nerovnost).

$$|EXY|, |EX\bar{Y}| \leq \|X\|_2 \|Y\|_2 = \sqrt{E|X|^2} \sqrt{E|Y|^2}, \quad X, Y \in L_2(\Omega, \mathcal{A}, P).$$

**Důsledek 1.10.**  $X \in L_2(\Omega, \mathcal{A}, P) \Rightarrow EX$  existuje.

*Důkaz.*

$$|EX| = |E(1 \cdot X)| \leq \underbrace{\sqrt{E|1|^2}}_1 \sqrt{E|X|^2} = \sqrt{E|X|^2} < \infty.$$

□

**Důsledek 1.11.**

$$X, Y \in L_2(\Omega, \mathcal{A}, P) \Rightarrow X - EX, Y - EY \in L_2(\Omega, \mathcal{A}, P)$$

a

$$\langle X - EX, Y - EY \rangle = E(X - EX)\overline{(Y - EY)} = \text{cov}(X, Y)$$

existuje a splňuje Schwarzovu nerovnost

$$|\text{cov}(X, Y)| \leq \sqrt{E|X - EX|^2} \sqrt{E|Y - EY|^2} = \sigma_X \sigma_Y.$$

**Důsledek 1.12.**

$$\rho(X, Y) = \begin{cases} \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} & \text{pro } \sigma_X \sigma_Y \neq 0 \\ 0 & \text{pro } \sigma_X \sigma_Y = 0 \end{cases}$$

je tzv. **korelační koeficient náhodných veličin  $X$  a  $Y$** , pro nějž platí  $|\rho(X, Y)| \leq 1$  a speciálně  $-1 \leq \rho(X, Y) \leq 1$  v případě reálných náhodných veličin  $X$  a  $Y$ .

*Poznámka* . Náhodné veličiny  $X, Y \in L_2(\Omega, \mathcal{A}, P)$  se nazývají **nekorelované**, jestliže  $\rho(X, Y) = 0$ . Vzhledem k 1.11 je nekorelovanost ekvivalentní s  $\text{cov}(X, Y) = 0$ , tj. s **ortogonalitou** centrovaných veličin  $X - EX$  a  $Y - EY$  v  $L_2(\Omega, \mathcal{A}, P)$ .

$\rho(\mathbb{X}, \mathbb{Y}) = [\rho(X_i, Y_j)]_{i,j}$  je tzv. **vzájemná korelační matice náhodných vektorů  $\mathbb{X}$  a  $\mathbb{Y}$** .

$\rho(\mathbb{X}, \mathbb{X}) = [\rho(X_i, X_j)]_{i,j}$  je tzv. **korelační matice náhodného vektoru  $\mathbb{X}$** .

**POZOR!** Nekorelovanost indikuje neexistenci **stochastické závislosti pouze lineárního typu**.

Tedy platí

$X, Y$  stochasticky nezávislé  $\Rightarrow X, Y$  nekorelované, avšak nikoliv naopak:

$X, Y$  nekorelované  $\nRightarrow X, Y$  stochasticky nezávislé.

## 2.4. IDENTIFIKACE PERIODICKÝCH KOMPONENT. PERIODOGRAM.

### 2.4.1. Fourierova řada (FŘ).

Komplexní harmonické kmity:

$$E = \{e_k(t)\}_{k \in \mathbb{Z}}, \quad e_k(t) = \frac{1}{\sqrt{T}} e^{i \frac{2\pi k t}{T}} = \frac{1}{\sqrt{T}} e^{i\omega_k t} = \frac{1}{\sqrt{T}} (\cos \omega_k t + i \sin \omega_k t),$$

kde pro  $k \neq 0$  značí

$$f_k = \frac{|k|}{T} \quad \dots \quad \text{frekvenci } |k|\text{-ho harmonického kmity s periodou } \frac{T}{|k|},$$

$$\omega_k = 2\pi f_k \quad \dots \quad \text{jeho úhlovou rychlost.}$$

$E$  je ortonormální systém v  $L_2([0, T])$ :

$$\langle e_j, e_k \rangle = \int_0^T e_j(t) \overline{e_k(t)} dt = \frac{1}{T} \int_0^T e^{i \frac{2\pi(j-k)t}{T}} dt = \delta_{j,k} = \begin{cases} 1 & \text{pro } j = k \\ 0 & \text{pro } j \neq k \end{cases}$$

$E$  je dokonce úplný (báze) v  $L_2([0, T])$  (bez důkazu):

$$x(t) \in L_2([0, T]) \text{ libovolná a } T\text{-periodická} \Rightarrow x(t) = \sum_{k=-\infty}^{\infty} x_k e_k(t),$$

kde řada konverguje (bez ohledu na pořadí) dle normy  $\|\cdot\|_2$  v prostoru  $L_2([0, T])$  a souřadnice  $x_k$  jsou jednoznačně určeny vztahem

$$\langle x, e_k \rangle = \left\langle \sum_{j=-\infty}^{\infty} x_j e_j, e_k \right\rangle = \sum_{j=-\infty}^{\infty} x_j \underbrace{\langle e_j, e_k \rangle}_{\delta_{j,k}} = x_k.$$

Komplexní tvar rozvoje do Fourierovy řady:

$$\underline{x(t)} = \sum_{k=-\infty}^{\infty} \underbrace{\langle x, e_k \rangle \frac{1}{\sqrt{T}}}_{c_k(x)} e^{i\omega_k t} = \sum_{k=-\infty}^{\infty} \underline{c_k(x)} e^{i\omega_k t},$$

kde

(2.4.1)

$$\underline{c_k := c_k(x)} = \frac{1}{\sqrt{T}} \langle x, e_k \rangle = \frac{1}{\sqrt{T}} \int_0^T x(t) \frac{1}{\sqrt{T}} e^{-i\omega_k t} dt = \underline{\frac{1}{T} \int_0^T x(t) e^{-i\omega_k t} dt}.$$

- $c_k$       ...     $k$ -tý komplexní Fourierův koeficient funkce  $x$ ,
- $c_0$       ...    střední hodnota (stejnoseměrná složka) funkce  $x$ ,
- $\{c_k\}_{k \in \mathbb{Z}}$  ...    komplexní fourierovské spektrum funkce  $x$ ,
- $x_k(t)$     :=     $c_k e^{i\omega_k t} + c_{-k} e^{-i\omega_k t}$ ,  $k \geq 1$
- ...     $k$ -tá harmonická komponenta funkce  $x$  o frekvenci  $f_k$ .

**Tvrzení 2.4.1.1** (Dirichletova věta). *Jestliže  $T$ -periodická funkce  $x(t)$  má konečnou variaci na  $[0, T]$ , pak její FŘ konverguje bodově k  $x(t)$  v každém bodě  $t$ , kde je  $x(t)$  spojitá a k  $\frac{1}{2}(x(t+0) + x(t-0))$  v každém bodě nespojitosti (u funkce s konečnou variací je to vždy pouze konečný skok).*

**Je-li  $x(t)$  reálná funkce, lze (2.4.1) upravit na další dva ekvivalentní tvary:**

Zřejmě  $c_{-k} = \overline{c_k}$ . Označíme-li

$$c_k = \frac{1}{2}(a_k - ib_k); \quad a_k, b_k \in \mathbb{R},$$

pak  $c_0 = \frac{a_0}{2} \in \mathbb{R}$ ,  $b_0 = 0$  a dostáváme

Goniometrický tvar Fourierovy řady:

$$\begin{aligned} \underline{x(t)} &= c_0 + \sum_{k=1}^{\infty} (c_k e^{i\omega_k t} + \underbrace{c_{-k} e^{-i\omega_k t}}_{c_k e^{i\omega_k t}}) = \frac{a_0}{2} + \sum_{k=1}^{\infty} 2\operatorname{Re}(c_k e^{i\omega_k t}) = \\ &= \frac{a_0}{2} + \sum_{k=1}^{\infty} (a_k \cos \omega_k t + b_k \sin \omega_k t), \end{aligned}$$

kde (2.4.2)

$$\underline{a_k} = 2\operatorname{Re}(c_k) = \frac{2}{T} \int_0^T x(t) \cos \omega_k t dt,$$

$$\underline{b_k} = -2\operatorname{Im}(c_k) = \frac{2}{T} \int_0^T x(t) \sin \omega_k t dt.$$

Položíme-li  $a_k = A_k \cos \varphi_k$  a  $b_k = A_k \sin \varphi_k$ , dostaneme

Amplitudově-fázový tvar Fourierovy řady:

$$\underline{x(t)} = \frac{a_0}{2} + \sum_{k=1}^{\infty} A_k (\cos \varphi_k \cos \omega_k t + \sin \varphi_k \sin \omega_k t) = \frac{a_0}{2} + \sum_{k=1}^{\infty} \underbrace{A_k \cos(\omega_k t - \varphi_k)}_{x_k(t)}.$$

kde (2.4.3)

$$A_k = 2|c_k| = \sqrt{a_k^2 + b_k^2},$$

$$\varphi_k = \arctan \frac{b_k}{a_k}.$$

- $A_k$  ... amplituda  $k$ -té harmonické složky funkce  $x$ ,
- $\varphi_k$  ... fázový posuv  $k$ -té harmonické složky funkce  $x$  ( $-\pi < \varphi_k \leq \pi$ ),
- $c_0 = \frac{a_0}{2}$  ... střední hodnota (stejnoseměrná složka) funkce  $x$ ,
- $\{A_k\}_{k \in \mathbb{N}}$  ... amplitudové spektrum funkce  $x$ ,
- $\{\varphi_k\}_{k \in \mathbb{N}}$  ... fázové spektrum funkce  $x$ .

### 2.4.2. Parsevalova identita (PI).

Nechť  $x \in L_2([0, T])$  (tj. s konečnou energií na intervalu  $[0, T]$ ) je  $T$ -periodická, potom

$$\begin{aligned} \|x\|_2^2 &= \int_0^T |x(t)|^2 dt = \langle x, x \rangle = \left\langle \sum_{j=-\infty}^{\infty} c_j \sqrt{T} \underbrace{\frac{e^{i\omega_j t}}{\sqrt{T}}}_{e_j(t)}, \sum_{k=-\infty}^{\infty} c_k \sqrt{T} \underbrace{\frac{e^{i\omega_k t}}{\sqrt{T}}}_{e_k(t)} \right\rangle \\ &= T \sum_{j,k=-\infty}^{\infty} c_j \overline{c_k} \underbrace{\langle e_j, e_k \rangle}_{\delta_{j,k}} = T \sum_{k=-\infty}^{\infty} |c_k|^2. \end{aligned}$$

Odtud pak dostáváme:

Komplezní tvar Parsevalovy identity:

$$\sum_{k=-\infty}^{\infty} |c_k|^2 = \frac{1}{T} \underbrace{\int_0^T |x(t)|^2 dt}_{\text{střední výkon na } [0, T]} < \infty \Rightarrow \{c_k\}_{k \in \mathbb{Z}} \in \ell_2. \quad (2.4.4)$$

### Speciálně:

- $|c_0|^2$  ... střední výkon stejnosměrné složky,
- $\frac{1}{T} \int_0^T |x_k(t)|^2 dt = |c_{-k}|^2 + |c_k|^2$  ... střední výkon  $k$ -té harmonické složky.



Amplitudový tvar Parsevalovy identity ( $x$  reálná):

$$\sum_{k=-\infty}^{\infty} |c_k|^2 = |c_0|^2 + \sum_{k=1}^{\infty} \underbrace{(|c_{-k}|^2 + |c_k|^2)}_{2|c_k|^2 = A_k^2/2} = \frac{a_0^2}{4} + \frac{1}{2} \sum_{k=1}^{\infty} A_k^2 = \frac{1}{T} \int_0^T |x(t)|^2 dt. \quad (2.4.5)$$

**Speciálně:**

$$\frac{a_0^2}{4} \dots \text{střední výkon stejnosměrné složky,}$$

$$\frac{A_k^2}{2} \dots \text{střední výkon } k\text{-té harmonické složky,}$$

$$\{2|c_k|^2\}_{k \in \mathbb{N}} = \left\{ \frac{A_k^2}{2} \right\}_{k=1}^{\infty} \dots \text{výkonová spektrální hustota,} \quad (2.4.6)$$

$$\{2T|c_k|^2\}_{k \in \mathbb{N}} = \left\{ \frac{TA_k^2}{2} \right\}_{k=1}^{\infty} \dots \text{energetická spektrální hustota.}$$

### 2.4.3. Diskretizace Fourierovy řady.

Vypočteme hodnoty rozvoje  $x(t)$  do její komplexní FŘ (2.4.1) pouze na rovnoměrné diskrétní síti  $N$  subintervalů:  $T = N\Delta t$ ,  $x_n = x(n\Delta t)$ ,  $n \in \mathbb{Z}$ .

Obdržíme

$$\underline{x}_n \stackrel{(2.4.1)}{=} \sum_{k=-\infty}^{\infty} c_k e^{i \frac{2\pi k}{N\Delta t} n \Delta t} = \sum_{k=-\lfloor \frac{N-1}{2} \rfloor}^{\lfloor \frac{N}{2} \rfloor} \overbrace{\left( \sum_{m=-\infty}^{\infty} c_{k+mN} \right)}^{\hat{c}_k} e^{i \frac{2\pi(k+mN)n}{N}} = \sum_{k=-\lfloor \frac{N-1}{2} \rfloor}^{\lfloor \frac{N}{2} \rfloor} \hat{c}_k e^{i \frac{2\pi k n}{N}}. \quad (2.4.7)$$

kde  $\lfloor \cdot \rfloor$  značí celou část hodnoty.

Platí

$$\hat{c}_k \approx c_k, \quad \text{neboť } |c_k| \rightarrow 0 \text{ pro } k \rightarrow \pm\infty,$$

$$\hat{c}_k = c_k, \quad \text{pokud } c_k = 0 \text{ pro } |k| \geq \frac{N}{2} \text{ (FŘ je konečná: } f_{max} \leq \frac{N}{2} \frac{1}{T} \text{ udává konečný frekvenční obsah } x(t)),$$

$$\{\hat{c}_k\}_{k \in \mathbb{Z}} \text{ je } N\text{-periodická posloupnost } (\hat{c}_k = \hat{c}_{k+mN}, m \in \mathbb{Z}).$$

Po úpravě:

$$\underline{x}_n = \sum_{k=-\lfloor \frac{N-1}{2} \rfloor}^{-1} \hat{c}_k e^{i \frac{2\pi k n}{N}} + \sum_{k=0}^{\lfloor \frac{N}{2} \rfloor} \hat{c}_k e^{i \frac{2\pi k n}{N}} = \sum_{k=0}^{N-1} \hat{c}_k e^{i \frac{2\pi k n}{N}}, \quad (2.4.8)$$

kde první ze sum jsme upravili záměnou sčítacího indexu  $r = k + N$  na tvar:

$$\sum_{k=-\lfloor \frac{N-1}{2} \rfloor}^{-1} \hat{c}_k e^{i \frac{2\pi k n}{N}} = \sum_{r=-\lfloor \frac{N-1}{2} \rfloor + N}^{N-1} \hat{c}_{r-N} e^{i \frac{2\pi(r-N)n}{N}} = \sum_{r=\lfloor \frac{N}{2} \rfloor + 1}^{N-1} \hat{c}_r e^{i \frac{2\pi r n}{N}}$$

s využitím vztahů

$$-\lfloor \frac{N-1}{2} \rfloor + N = \lfloor \frac{N}{2} \rfloor + 1 \text{ a } \hat{c}_r = \hat{c}_{r-N} \text{ pro } r = \lfloor \frac{N}{2} \rfloor + 1, \dots, N-1 \text{ (} N\text{-periodicita).}$$

Označíme-li  $\mathbf{x} = (x_0, x_1, \dots, x_{N-1})^T$  vzorky  $x(t)$  na její jedné periodě a  $\hat{\mathbf{c}} = (\hat{c}_0, \hat{c}_1, \dots, \hat{c}_{N-1})^T$  odhady Fourierových koeficientů, pak (2.4.8) určuje tzv. **operátor diskrétní Fourierovy transformace** ( $\mathbf{x} = \text{DFTT}_N^+(\hat{\mathbf{c}})$ ) dle následující definice.

**Definice 2.4.3.1** (Diskrétní Fourierova transformace (DFT)).

$\text{DFT}_N^\pm : \mathbb{C}^N \rightarrow \mathbb{C}^N$  je lineární operátor  $\mathbf{X} = \mathbb{W}_N^\pm \mathbf{x}$  určený maticí  $N \times N$ :

$\mathbb{W}_N^\pm = [W_N^{kn}]_{0 \leq k, n \leq N-1}$ , kde  $W_N = e^{\pm i \frac{2\pi}{N}}$  je  $N$ -tá primitivní odmocnina z 1, tj.

$W_N^N = 1$ , ale  $W_N^k \neq 1$  pro  $k = 1, \dots, N-1$ . Tedy

$$\mathbf{X} = (X_0, X_1, \dots, X_{N-1})^T, \quad \text{kde } X_k = \sum_{n=0}^{N-1} e^{\pm i \frac{2\pi kn}{N}} x_n \text{ pro } k = 0, 1, \dots, N-1.$$

Zřejmě  $\mathbb{W}_N^\pm$  je symetrická matice a  $(\mathbb{W}_N^\pm)^* = \mathbb{W}_N^\mp$ .

**Věta 2.4.3.2** (Věta o inverzi).

Platí  $\mathbb{W}_N^\pm \mathbb{W}_N^\mp = \mathbb{W}_N^\pm (\mathbb{W}_N^\pm)^* = NI_N$  a tedy  $(\mathbb{W}_N^\pm)^{-1} = \frac{1}{N} \mathbb{W}_N^\mp$  a  $\frac{1}{\sqrt{N}} \mathbb{W}_N^\pm$  je unitární matice.

*Důkaz.* Označme  $A := [a_{r,s}] = \mathbb{W}_N^\pm \mathbb{W}_N^\mp$ , pak

$$a_{r,s} = \sum_{n=0}^{N-1} W_N^{rn} W_N^{-ns} = \sum_{n=0}^{N-1} W_N^{n(r-s)} = \sum_{n=0}^{N-1} q^n, \quad q = W_N^{r-s}.$$

Odtud

$$a_{r,s} = \begin{cases} N & \text{pro } r = s \\ \frac{q^N - 1}{q - 1} = 0 & \text{pro } r \neq s \end{cases},$$

neboť  $q^N = W_N^{N(r-s)} = 1$  a  $q \neq 1$  pro  $r \neq s$  v důsledku  $0 < |r-s| \leq N-1$ . □

*Poznámka 2.4.3.3.* V systému MATLAB jsou operátory  $\text{DFT}_N^\pm$  realizovány pomocí algoritmu tzv. **rychlé Fourierovy transformace** implementovaných v procedurách **fft** a **ifft** takto:

$$\underline{\mathbf{X}} = \mathbb{W}_N^- \mathbf{x} = \text{DFT}_N^-(\mathbf{x}) = \underline{\text{fft}(\mathbf{x})}$$

a

$$\underline{\mathbf{x}} = (\mathbb{W}_N^-)^{-1} \underline{\mathbf{X}} = \frac{1}{N} \mathbb{W}_N^+ \underline{\mathbf{X}} = \frac{1}{N} \text{DFT}_N^+(\underline{\mathbf{X}}) = \underline{\text{ifft}(\underline{\mathbf{X}})}.$$

**Důsledek 2.4.3.4.** Podle (2.4.8), věty 2.4.3.2 a poznámky 2.4.3.3 platí

$$\underline{\hat{\mathbf{c}}} = \mathbb{W}_N^+ \hat{\mathbf{c}} = \text{DFT}_N^+(\hat{\mathbf{c}}) = \underline{N \text{ifft}(\hat{\mathbf{c}})},$$

$$\hat{\underline{\mathbf{c}}} = (\mathbb{W}_N^+)^{-1} \underline{\hat{\mathbf{c}}} = \frac{1}{N} \mathbb{W}_N^- \underline{\hat{\mathbf{c}}} = \frac{1}{N} \text{DFT}_N^-(\underline{\hat{\mathbf{c}}}) = \underline{\frac{1}{N} \text{fft}(\underline{\hat{\mathbf{c}}})}.$$

#### 2.4.4. Periodogram.

*Periodogram* = odhad energetické spektrální hustoty (2.4.6) reálné  $T$ -periodické funkce  $x(t)$ , kde za  $c_k$ ,  $k = 1, \dots, m$ ,  $m := \lfloor \frac{N-1}{2} \rfloor$ , dosadíme odhad  $\hat{c}_k$  z (2.4.7) spočtený pomocí  $\text{DFT}_N^-$  dle 2.4.3.4 po ekvidistantní diskretizaci jedné periody  $x(t)$ :

$T = N\Delta t$ ,  $x_n = x(n\Delta t)$ ,  $n = 0, 1, \dots, N$ .

**Periodogram** je tedy posloupnost hodnot  $\{I_k\}_{k=1}^m$ , kde

$$\underline{I}_k = I(\omega_k) = 2T |\hat{c}_k|^2 = 2N\Delta t \left| \frac{1}{N} X_k \right|^2 = \underline{\frac{2}{N} \Delta t |X_k|^2}, \quad k = 1, \dots, m \quad (2.4.9)$$

a

$$X_k = \sum_{t=0}^{N-1} x_t e^{-i \frac{2\pi kt}{N}} = \sum_{t=1}^N x_t e^{-i \frac{2\pi kt}{N}}.$$

Druhá rovnost je zde důsledkem  $N$ -periodicity  $x_0 = x_N$ .

Z hlediska kvalitativních závěrů nezáleží na multiplikační konstantě, proto bez újmy na obecnosti můžeme předpokládat  $\Delta t = 1$  a dostáváme tak **výsledný vztah pro hodnoty  $I_k$  periodogramu** ve tvaru

$$I_k = I(\omega_k) = \frac{2}{N} |X_k|^2 = \frac{2}{N} \left\{ \left( \sum_{t=1}^N x_t \cos \frac{2\pi kt}{N} \right)^2 + \left( \sum_{t=1}^N x_t \sin \frac{2\pi kt}{N} \right)^2 \right\}, \quad k = 1, \dots, m. \quad (2.4.10)$$

Velká hodnota  $I_k$  indikuje velkou energii  $k$ -té harmonické komponenty, tj. silné zastoupení složky  $x_k(t) = c_k e^{i\omega_k t} + c_{-k} e^{-i\omega_k t} = A_k \cos(\omega_k t - \varphi_k)$  v rozvoji  $x(t)$  do Fourierovy řady.

**Věta 2.4.4.1.**

$$I_k = 2 \sum_{|h| < N} \hat{\gamma}(h) e^{-i \frac{2\pi hk}{N}} \quad \text{pro } k = 1, 2, \dots, \left[ \frac{N-1}{2} \right],$$

kde  $\hat{\gamma}(h)$  je odhad autokovarianční funkce vektoru  $\mathbf{x} = (x_1, \dots, x_N)^T$  spočtený dle 1.46.

*Důkaz.*

$$I_k = \frac{2}{N} |X_k|^2 = \frac{2}{N} X_k \overline{X_k} = \frac{2}{N} \left( \sum_{s=1}^N x_s e^{-i \frac{2\pi sk}{N}} \right) \left( \sum_{t=1}^N \overline{x_t} e^{i \frac{2\pi tk}{N}} \right).$$

Protože dle důkazu věty 2.4.3.2 platí

$$\sum_{s=1}^N e^{-i \frac{2\pi sk}{N}} = \sum_{t=1}^N e^{i \frac{2\pi tk}{N}} = 0 \quad \text{pro } k \neq 0 \pmod{N},$$

lze psát

$$\begin{aligned} I_k &= \frac{2}{N} \left( \sum_{s=1}^N (x_s - \hat{x}) e^{-i \frac{2\pi sk}{N}} \right) \left( \sum_{t=1}^N (\overline{x_t - \hat{x}}) e^{i \frac{2\pi tk}{N}} \right) = \\ &= \frac{2}{N} \sum_{s=1}^N \sum_{t=1}^N (x_s - \hat{x}) (\overline{x_t - \hat{x}}) e^{-i \frac{2\pi(s-t)k}{N}} \quad h = s-t \\ &= 2 \sum_{|h| < N} \frac{1}{N} \underbrace{\sum_{t=1}^{N-h} (x_{t+h} - \hat{x}) (\overline{x_t - \hat{x}})}_{\hat{\gamma}(h)} e^{-i \frac{2\pi hk}{N}}. \end{aligned}$$

□

#### 2.4.5. Fisherův test periodicity.

Položme  $m = \left[ \frac{N-1}{2} \right]$  a definujme statistiku

$$W = \max_{k=1, \dots, m} Y_k, \quad \text{kde } Y_k = \frac{I_k}{\sum_{j=1}^m I_j}.$$

Protože  $I_j \geq 0$  pro  $j = 1, 2, \dots, m$ , je  $0 \leq W \leq 1$ .

Nechť  $X_t = P_t + E_t$ ,  $E_t \sim WN(0, \sigma^2)$  gaussovský.

Platí-li nulová hypotéza  $H_0 : X_t = E_t$ , pak testová statistika  $W$  má na intervalu  $[0, 1]$  rozdělení, pro něž platí

$$P(W > x) = \sum_{\substack{j=1 \\ 1-jx > 0}}^m (-1)^{j-1} \binom{m}{j} (1-jx)^{m-1}, \quad 0 < x < 1,$$

přičemž  $P(W > x) \approx m(1-x)^{m-1}$  je dobrá aproximace pro  $m \leq 50$ .

Nechť  $g_F(1-\alpha)$  značí  $(1-\alpha)$ -kvantil rozdělení statistiky  $W$  (tabelováno), tj.  $P(W \leq g_F(1-\alpha)) = 1-\alpha$ . Pak  $H_0$  zamítáme s rizikem  $\alpha$ , pokud  $W > g_F(1-\alpha)$ . Je-li v takovém případě  $I_{k_0} = \max_{k=1, \dots, m} I_k$ , pak  $k_0$ -tou harmonickou komponentu považujeme za významnou. Další významnou harmonickou komponentu zjistíme z periodogramu délky  $m-1$ , kde vynecháme  $I_{k_0}$ , a tak pokračujeme dále, dokud není hypotéza  $H_0$  přijata.

#### 2.4.6. Siegelův test periodicity.

Tento test je vhodnější než Fisherův v případě většího množství harmonických komponent. Siegelova statistika je tvaru

$$T_\lambda = \sum_{k=1}^m (Y_k - \lambda g_F(1 - \alpha))^+, \quad 0 < \lambda < 1 \text{ (doporučená hodnota je } \lambda = 0,6 \text{)}.$$

$H_0$  zamítáme s rizikem  $\alpha$ , jestliže  $T_\lambda > t_\lambda(1 - \alpha)$ , kde  $t_\lambda(1 - \alpha)$  je  $(1 - \alpha)$ -kvantil rozdělení  $T_\lambda$  (tabelováno).

## 2.5. SEPARACE JEDNOTLIVÝCH SLOŽEK.

### Označení.

$\mathbf{t} = (t_1, \dots, t_n)^T$ ,  $t_i \neq t_j$  pro  $i \neq j$  ... vektor "času",  
 $\mathbb{X} := \mathbb{X}_{\mathbf{t}} = (X_{t_1}, \dots, X_{t_n})^T$  ... náhodný vektor populace vybrané  
z časové řady,  
 $\mathbf{x} := \mathbf{x}_{\mathbf{t}} = (x_1, \dots, x_n)^T$ ,  $x_i = X_{t_i}(\omega)$  ... pozorování vektoru populace  $\mathbb{X}$ .

### A. Parametrické metody (P).

Tyto metody používáme v případě, že je znám analytický tvar  $D_t = D(t; \beta_1, \dots, \beta_p)$  v aditivním modelu  $X_t = D_t + E_t$  (multiplikativní model převedeme na aditivní logaritmováním) s neznámými parametry  $\beta = (\beta_1, \dots, \beta_p)^T$ , které odhadneme minimalizací výrazu  $N(\beta_1, \dots, \beta_p) = \|D(\mathbf{t}; \beta_1, \dots, \beta_p) - \mathbf{x}\|^2$  ve vhodné normě  $\|\cdot\|$ . Zpravidla se používá euklidovská norma a klasická metoda nejmenších čtverců:

$$N(\beta_1, \dots, \beta_p) = \sum_{i=1}^n |D(t_i, \beta_1, \dots, \beta_p) - x_i|^2.$$

Lokální extrém hledáme řešením systému rovnic

$$\frac{\partial N(\beta_1, \dots, \beta_p)}{\partial \beta_j} = 0 \text{ pro } j = 1, \dots, p$$

s ověřením podmínky pro lokální minimum (jakobián je pozitivně definitní matice):

$$J = \left[ \frac{\partial^2 N(\beta_1, \dots, \beta_p)}{\partial \beta_i \partial \beta_j} \right]_{i,j} > 0.$$

Pokud  $D$  závisí lineárně na  $\beta_1, \dots, \beta_p$ , pak tento systém je systémem  $p$  lineárních rovnic pro  $p$  neznámých  $\beta_1, \dots, \beta_p$  a dostáváme klasický **lineární regresní model**.

V případě nelineární regrese použijeme pro minimalizaci  $N(\beta_1, \dots, \beta_p)$  **vhodnou optimalizační metodu**, například *Optimization Toolbox* v prostředí systému MATLAB.

Lineární regresní model:  $D(t; \beta_1, \dots, \beta_p) = \varphi_1(t)\beta_1 + \dots + \varphi_p(t)\beta_p$ , kde  $\varphi_i$  jsou dané.

$$\mathbb{X} = D\beta + \mathbb{E}, \quad \mathbb{E} = (E_{t_1}, \dots, E_{t_n})^T \sim \text{IID}(0, \sigma^2),$$

kde

$$D = [\varphi_1(\mathbf{t}), \dots, \varphi_p(\mathbf{t})], \quad \varphi_j(\mathbf{t}) = (\varphi_j(t_1), \dots, \varphi_j(t_n))^T$$

je matice s plnou sloupcovou hodností  $p$ .

Pro vektor pozorování přejde výše uvedený model v

$$\mathbf{x} = D\beta + \mathbf{e}, \quad \mathbf{e} = (e_1, \dots, e_n)^T, \text{ kde } e_i = E_{t_i}(\omega).$$

### B. Neparametrické metody (NP).

Za vhodných předpokladů konstruujeme lineární nebo nelineární operátor  $S$  (smoother = vyhlazovač) takový, že

$$S(\{X_t\}) =: \{\widehat{D}_t\} \approx \{D_t\}$$

je dobrý odhad  $D_t$  v tom smyslu, že  $\{\widehat{D}_t\} \rightarrow \{D_t\}$  ve vhodné konvergenci při  $n \rightarrow \infty$ , tj. odhad je **asymptoticky přesný**.

### Značení metod:

AM-P	...	parametrická	} metoda v <b>aditivním modelu</b> .
AM-NP	...	neparametrická	
MM-P	...	parametrická	} metoda v <b>multiplikativním modelu</b>
MM-NP	...	neparametrická	

## 2.5.2. Metody pro odhad sezónní složky.

$\mathbf{x} = (x_1, \dots, x_n)^T$  ... ekvidistantní pozorování náhodného procesu  $X = \{X_t | t \in \mathbb{Z}\}$   
 $\mathbf{e} = (e_1, \dots, e_n)^T$  ... pozorované chyby  
 $t \in \{1, 2, \dots, n\}$

Předpokládáme model:

- aditivní:  $X_t = Tr_t + Sz_t + C_t + E_t$  (AM)  
 nebo  
 - multiplikativní:  $X_t = Tr_t Sz_t C_t E_t$  (MM).

Nechť  $d$  je perioda  $Sz_t$  a  $n = md$ , pak zavedeme matici  $m \times d$

$$X = \begin{bmatrix} x_1 & \dots & x_k & \dots & x_d \\ x_{d+1} & \dots & x_{d+k} & \dots & x_{2d} \\ \vdots & \dots & \vdots & \dots & \vdots \end{bmatrix} = [x_{j,k}],$$

kde  $x_{j,k} = x_{(j-1)d+k}$  (např.  $j$  =rok,  $k$  =měsíc).

Označíme-li vec operátor, který skládá sloupce matice pod sebe do jednoho sloupcového vektoru, pak můžeme psát  $\mathbf{x} = \text{vec}(X^T)$ .

V MATLABu:  $X = \text{reshape}(\mathbf{x}, d, m)$ .

Řádek  $X(j, :)$  představuje pozorování hodnot  $j$ -té periody  $Sz_t$  spolu s  $Tr_t, C_t$  a  $E_t$ .

Sloupec  $X(:, k)$  představuje pozorování  $k$ -té hodnoty v každé periodě  $Sz_t$  spolu s  $Tr_t, C_t$  a  $E_t$ .

Nechť  $E = [e_{j,k}]$  je analogicky zavedená matice chyb.

Hledaná složka  $Sz_t$  je určena hodnotami  $\mathbf{s} := (s_1, \dots, s_d) := (Sz_1, \dots, Sz_d)$  jedné své periody, přičemž platí pro  $\bar{s} := \frac{s_1 + \dots + s_d}{d}$ :

$$\bar{s} = \begin{cases} 0 & \text{v AM} \\ 1 & \text{v MM} \end{cases} \quad (2.5.4)$$

Budeme hledat odhady  $\hat{s}_k$  splňující (2.5.4).

### 2.5.2.1. Metoda malého trendu pro odhad $Sz_t$ (AM-NP, MM-NP).

Předpoklad:

$Tr_t + C_t$ , resp.  $Tr_t C_t$  je přibližně konstantní v každé periodě  $Sz_t$ :

$$\left. \begin{array}{l} \text{v AM: } x_{j,k} - s_k - e_{j,k} = Tr_t + C_t \approx m_j \\ \text{v MM: } x_{j,k} / (s_k e_{j,k}) = Tr_t C_t \approx m_j \end{array} \right\} \text{ v } j\text{-té periodě, } t = (j-1)d + k.$$

Algoritmus:

(1)  $\hat{m}_j = \frac{1}{d} \sum_{k=1}^d x_{j,k}$  ... odhad  $m_j$  pro  $j = 1, \dots, m$ .

(2) pro  $k = 1, \dots, d$  spočteme:

$$\begin{array}{l} \text{v AM: } \hat{s}_k = \frac{1}{m} \sum_{j=1}^m \underbrace{(x_{j,k} - \hat{m}_j)}_{\approx s_k + e_{j,k}} \\ \text{v MM: } \hat{s}_k = \frac{1}{m} \sum_{j=1}^m \underbrace{x_{j,k} / \hat{m}_j}_{\approx s_k e_{j,k}} \end{array}$$

Odhady zřejmě splňují (2.5.4):

v AM:

$$\widehat{s} = \frac{1}{d} \sum_{k=1}^d \left( \frac{1}{m} \sum_{j=1}^m (x_{j,k} - \widehat{m}_j) \right) = \frac{1}{m} \sum_{j=1}^m \left( \underbrace{\frac{1}{d} \sum_{k=1}^d x_{j,k}}_{\widehat{m}_j} - \underbrace{\frac{1}{d} \sum_{k=1}^d \widehat{m}_j}_{\widehat{m}_j} \right) = 0$$

v MM:

$$\widehat{s} = \frac{1}{d} \sum_{k=1}^d \left( \frac{1}{m} \sum_{j=1}^m \frac{x_{j,k}}{\widehat{m}_j} \right) = \frac{1}{m} \sum_{j=1}^m \underbrace{\frac{1}{\widehat{m}_j} \frac{1}{d} \sum_{k=1}^d x_{j,k}}_1 = \frac{1}{m} m = 1.$$

Dále pokračujeme kroky v 2.5.2.3.

### 2.5.2.2. Metoda klouzavého průměru pro odhad $Sz_t$ (AM-NP, MM-NP).

Zavedeme vektor vah  $\mathbf{w} = (w_1, \dots, w_{2q+1})^T$  pro výpočet obyčejného průměru takto:

$$\mathbf{w} = \frac{1}{d} \underbrace{(1, 1, \dots, 1)}_{(2q+1) \times}^T \quad \text{pro periodu liché délky } d = 2q + 1,$$

$$\mathbf{w} = \frac{1}{d} \underbrace{(1/2, 1, \dots, 1, 1/2)}_{(2q+1) \times}^T \quad \text{pro periodu sudé délky } d = 2q.$$

Spočteme klouzavý průměr

$$\widehat{m}_t = \begin{cases} \sum_{\tau=-q}^q w_\tau x_{t+\tau} & \text{pro } q+1 \leq t \leq n-q, \\ x_t & \text{pro } 1 \leq t \leq q \text{ nebo } n-q+1 \leq t \leq n. \end{cases}$$

Jelikož délka klouzavého průměru je rovna délce periody v sezónní složce, potlačíme tímto průměrováním nejen chybovou komponentu  $e_t$ , ale eliminujeme současně i vlastní sezónní složku  $Sz_t$ , neboť tato má v AM nulovou, resp. v MM jednotkovou, průměrnou hodnotu na každé své periodě.

V dalším tedy můžeme vycházet z platnosti následujícího předpokladu:

Předpoklad:

$$\left. \begin{array}{l} \text{v AM: } x_t - Sz_t - e_t = Tr_t + C_t \approx \widehat{m}_t \\ \text{v MM: } x_t / (Sz_t e_t) = Tr_t C_t \approx \widehat{m}_t \end{array} \right\} \text{ pro } q+1 \leq t \leq n-q.$$

Algoritmus:

- (1) Spočteme vektor klouzavých průměrů a přerovnáme jej do matice  $m \times d$  po periodách stejně jako  $\mathbf{x}$ :

$$\widehat{\mathbf{m}} = (\widehat{m}_1, \dots, \widehat{m}_n)^T =: \text{vec}(\widehat{M}^T), \text{ kde } \widehat{M} =: [\widehat{m}_{j,k}].$$

Pak  $X - \widehat{M}$  v AM, resp.  $X./\widehat{M} = [x_{j,k}/\widehat{m}_{j,k}]$  v MM, má následující strukturu s  $q$  nulami, resp. jedničkami na začátku první periody a na konci poslední periody.

Pro  $d$  liché:

$$X - \widehat{M} = \begin{bmatrix} 0 & \dots & 0 & \bullet & \bullet & \dots & \bullet \\ & & \vdots & & & \vdots & \\ \bullet & \dots & \bullet & \bullet & 0 & \dots & 0 \end{bmatrix}, \text{ resp. } X./\widehat{M} = \begin{bmatrix} 1 & \dots & 1 & \bullet & \bullet & \dots & \bullet \\ & & \vdots & & & \vdots & \\ \bullet & \dots & \bullet & \bullet & 1 & \dots & 1 \end{bmatrix}.$$

Pro  $d$  sudé:

$$X - \widehat{M} = \begin{bmatrix} 0 & \dots & 0 & \bullet & \dots & \bullet \\ & & \vdots & & & \vdots & \\ \bullet & \dots & \bullet & 0 & \dots & 0 \end{bmatrix}, \text{ resp. } X./\widehat{M} = \begin{bmatrix} 1 & \dots & 1 & \bullet & \dots & \bullet \\ & & \vdots & & & \vdots & \\ \bullet & \dots & \bullet & 1 & \dots & 1 \end{bmatrix}.$$

- (2) Pro  $k = 1, \dots, d$  spočteme:

$$\text{v AM: } \widetilde{s}_k = \frac{1}{m-\delta_k} \sum_{j=1}^m \underbrace{(x_{j,k} - \widehat{m}_{j,k})}_{\approx s_k + e_{j,k}},$$

$$\text{v MM: } \widetilde{s}_k = \frac{1}{m-\delta_k} \left[ \underbrace{\left( \sum_{j=1}^m \frac{x_{j,k}}{\widehat{m}_{j,k}} \right)}_{\approx s_k e_{j,k}} - \delta_k \right].$$

kde  $\delta_k = \begin{cases} 0 & \text{pro } k = \frac{d+1}{2} \text{ (můž e nastat jen pro } d \text{ liché, kdy } k = q + 1) \\ 1 & \text{jinak} \end{cases}$

(3)  $\bar{s} = \frac{1}{d} \sum_{k=1}^d \tilde{s}_k.$

(4) Provedeme vycentrování, aby platil vztah (2.5.4):

$\hat{s}_k = \tilde{s}_k - \bar{s}$  v AM, resp.  $\hat{s}_k = \tilde{s}_k / \bar{s}$  v MM pro  $k = 1, \dots, d.$

Dále pokračujeme opět kroky v 2.5.2.3.

### 2.5.2.3. Separace $Tr_t$ a $Sz_t$ (AM, MM).

(5) Zperiodičnění:  $\widehat{Sz}_{k+jd} = \hat{s}_k$  pro  $k = 1, \dots, d$  a  $j = 0, 1, \dots, m - 1.$

(6) Separace sezónní složky  $Sz_t$ :

$y_t := x_t - \widehat{Sz}_t \approx Tr_t + C_t + e_t$  v AM, resp.  $y_t := x_t / \widehat{Sz}_t \approx Tr_t C_t e_t$  v MM pro  $t = 1, \dots, n.$

(7) V  $y_t$  odhadujeme  $Tr_t + C_t$  (resp.  $Tr_t C_t$  v MM po zlogaritmování) vhodnou metodou (např. z odst. 2.5.1). Můžeme také odhadnout jen trend například položením  $q = 0$  v metodě 2.5.1.1 nebo užitím metod dále popsaných v odst. 2.5.3 a pak separovat cyklickou složku  $C_t$ :

(8) Separace cyklické složky  $C_t$ :

$C_t + e_t \approx y_t - \widehat{Tr}_t$  v AM, resp.  $C_t e_t \approx y_t / \widehat{Tr}_t$  v MM pro  $t = 1, \dots, n.$

(9) Analyzujeme periodické komponenty v  $C_t + e_t$  (resp. zlogaritmované  $C_t e_t$  v MM) pomocí periodogramu a testů periodicity z odst. 2.4. Ty budou nyní citlivěji detekovat harmonické komponenty cyklické složky, která již není rušena přítomností dalších deterministických složek.

### 2.5.2.4. Eliminace $Sz_t$ diferencováním na vzdálenost $d$ (AM).

Předpoklad:

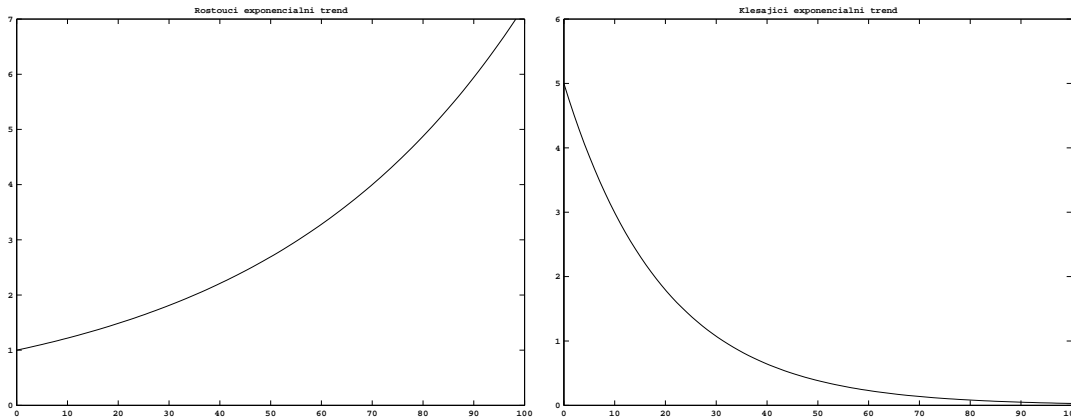
$x_t = Sz_t + m_t + e_t$ , kde  $m_t = Tr_t + C_t$  a perioda  $C_t$  je podstatně delší než perioda  $Sz_t.$

Algoritmus:

(1)  $y_t := x_t - x_{t-d} = \underbrace{Sz_t - Sz_{t-d}}_{\approx 0} + \underbrace{m_t - m_{t-d}}_{\tilde{m}_t} + \underbrace{e_t - e_{t-d}}_{\tilde{e}_t}, t = d + 1, \dots, n.$

(2) Odhadujeme  $\tilde{m}_t$  jako v 2.5.2.3(7)-(9), kde ovšem  $\tilde{m}_t = m_t - m_{t-d}$  interpretujeme jako změny oproti minulé sezóně.





a) Rostoucí:  $\alpha = 1, \beta = 1,02$

b) Klesající:  $\alpha = 5, \beta = 0,95$

OBRÁZEK 2.5.1. Exponenciální trend

### 2.5.3. Metody pro odhad trendu.

$\mathbf{x} = (x_1, \dots, x_n)^T$  ... ekvidistantní pozorování náhodného procesu  $X = \{X_t | t \in \mathbb{Z}\}$   
 $\mathbf{e} = (e_1, \dots, e_n)^T$  ... pozorované chyby  
 $t \in \{1, 2, \dots, n\}$

Následují metody pro odhad růstového trendu v ekonomických časových řadách [2, str. 34-41].

#### 2.5.3.7. Exponenciální trend (MM-P).

Model:

$$Tr_t = \alpha\beta^t, \quad \alpha, \beta \in \mathbb{R}, \beta > 0, t = 1, 2, \dots, n.$$

Tento model je vhodný pro popis trendu s **konstantním koeficientem růstu**

$$\frac{Tr_{t+2} - Tr_{t+1}}{Tr_{t+1} - Tr_t} = \frac{\alpha\beta^{t+2} - \alpha\beta^{t+1}}{\alpha\beta^{t+1} - \alpha\beta^t} = \beta,$$

kde při  $\alpha > 0$  dostáváme pro  $\beta > 1$  růst a pro  $0 < \beta < 1$  pokles (obr. 2.5.1). V případě  $\alpha < 0$  je tomu naopak.

Algoritmus:

Zlogaritmováním provedeme transformaci na aditivní model s lineárním trendem:

$$\ln Tr_t = \ln \alpha + t \ln \beta, \quad \beta = (\ln \beta, \ln \alpha)^T.$$

Odhad  $\beta$  spočteme metodou lineární regrese dle odst. 2.5.1.1. Odlogaritmováním hodnot získaných podle vztahů (2.5.1)-(2.5.3) pak dostaneme výsledky pro původní model.

Kontrolní kritérium adekvátnosti tohoto modelu:

$$\frac{x_{t+1}}{x_t} \approx \beta, \text{ resp. } \Delta \ln x_t = \ln x_{t+1} - \ln x_t \approx \ln \beta.$$

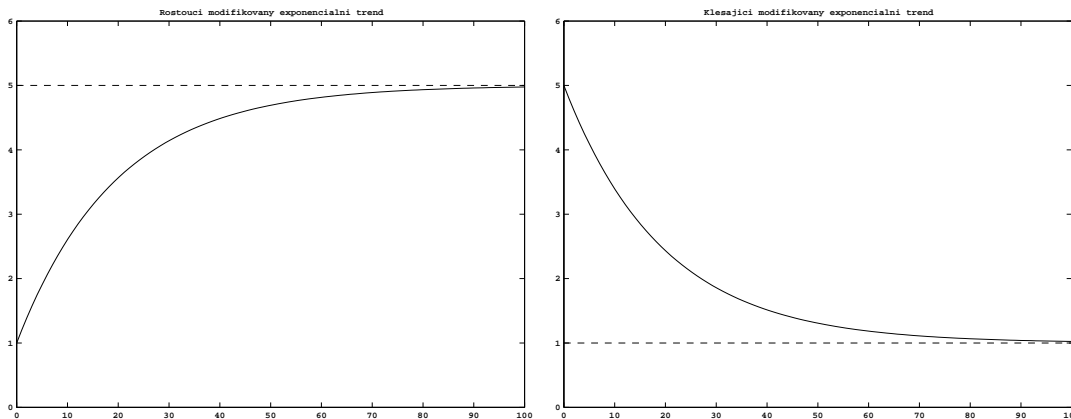
Kontrolujeme tedy, zda tato veličina osciluje náhodně kolem konstantní úrovně.

#### 2.5.3.8. Modifikovaný exponenciální trend (AM-P).

Model:

$$Tr_t = \gamma + \alpha\beta^t, \quad \alpha, \beta, \gamma \in \mathbb{R}, \beta > 0, t = 1, 2, \dots, n.$$

Jedná se o tříparametrické zobecnění předchozího modelu. Používá se především v případě, kdy trend má konstantní koeficient růstu, resp. poklesu a přitom je shora, resp. zdola asymptoticky omezen. Tehdy volíme  $0 < \beta < 1$ , přičemž  $\gamma$  pak odpovídá asymptotické úrovni, ke které trend roste ( $\alpha < 0$ ), resp. klesá ( $\alpha > 0$ ). Viz obr. 2.5.2.



a) Rostoucí:  $\alpha = -4$ ,  $\beta = 0,95$ ,  $\gamma = 5$

b) Klesající:  $\alpha = 4$ ,  $\beta = 0,95$ ,  $\gamma = 1$

OBRÁZEK 2.5.2. Modifikovaný exponenciální trend

### Algoritmus:

- Soubor pozorování  $\mathbf{x}$  rozdělíme na stejně velké třetiny délky  $m$  (pokud  $n \neq 3m$ , vynecháme ze začátku řady  $n \bmod 3$  pozorovaných hodnot).
- Pozorování v jednotlivých třetinách sečteme. Užitím vzorce pro součet  $m$  členů geometrické řady  $S := \sum_{t=1}^m \beta^t = \beta(\beta^m - 1)/(\beta - 1)$  pak obdržíme tři rovnice o třech neznámých  $\alpha, \beta, \gamma$ :

$$\begin{aligned} \Sigma_1 &:= \sum_{t=1}^m x_t && \approx \sum_{t=1}^m Tr_t && = m\gamma + \alpha S \\ \Sigma_2 &:= \sum_{t=m+1}^{2m} x_t && \approx \sum_{t=m+1}^{2m} Tr_t && = m\gamma + \alpha\beta^m S \\ \Sigma_3 &:= \sum_{t=2m+1}^{3m} x_t && \approx \sum_{t=2m+1}^{3m} Tr_t && = m\gamma + \alpha\beta^{2m} S. \end{aligned}$$

- Systém rovnic vyřešíme postupnou eliminací jednotlivých neznámých:

$$\left. \begin{aligned} \Sigma_3 - \Sigma_2 &= \alpha\beta^m S(\beta^m - 1) \\ \Sigma_2 - \Sigma_1 &= \alpha S(\beta^m - 1) \end{aligned} \right\} \Rightarrow \beta^m = \frac{\Sigma_3 - \Sigma_2}{\Sigma_2 - \Sigma_1}.$$

Pak

$$\Sigma_2 - \Sigma_1 = \alpha S(\beta^m - 1) \Rightarrow \alpha = \frac{\Sigma_2 - \Sigma_1}{S(\beta^m - 1)} \quad \text{a} \quad \Sigma_1 = m\gamma + \alpha S \Rightarrow \gamma = \frac{\Sigma_1 - \alpha S}{m}.$$

- Obdrželi jsme tak přibližné vztahy pro výpočet neznámých parametrů  $\alpha, \beta, \gamma$ :

$$\begin{aligned} \hat{\beta} &= \left( \frac{\Sigma_3 - \Sigma_2}{\Sigma_2 - \Sigma_1} \right)^{\frac{1}{m}} \\ \hat{S} &= \hat{\beta} \frac{\hat{\beta}^m - 1}{\hat{\beta} - 1} \\ \hat{\alpha} &= \frac{\Sigma_2 - \Sigma_1}{\hat{S}(\hat{\beta}^m - 1)} \\ \hat{\gamma} &= \frac{\Sigma_1 - \hat{\alpha}\hat{S}}{m}. \end{aligned}$$

### Kontrolní kritérium adekvátnosti tohoto modelu:

$$\frac{\Delta Tr_{t+1}}{\Delta Tr_t} = \frac{\gamma + \alpha\beta^{t+2} - (\gamma + \alpha\beta^{t+1})}{\gamma + \alpha\beta^{t+1} - (\gamma + \alpha\beta^t)} = \frac{\alpha\beta^{t+1}(\beta - 1)}{\alpha\beta^t(\beta - 1)} = \beta.$$

Kontrolujeme tedy, zda veličina  $\Delta x_{t+1}/\Delta x_t = (x_{t+2} - x_{t+1})/(x_{t+1} - x_t)$  osciluje náhodně kolem konstantní úrovně.

### 2.5.3.9. Logistický trend (AM-P).

Model:

$$Tr_t = \frac{\gamma}{1 + \alpha\beta^t}, \quad \alpha, \beta, \gamma \in \mathbb{R}, \beta > 0, \gamma > 0, t = 1, 2, \dots, n.$$

Tento model dále rozšiřuje možnosti modifikovaného exponenciálního trendu, neboť má inflexní bod v  $t = -\ln \alpha / \ln \beta$ , který zachycuje změnu akcelerace růstu (obr. 2.5.3).

Toto tvrzení nyní ověříme:

$$Tr'_t = -\frac{\gamma\alpha(\ln \beta)\beta^t}{(1 + \alpha\beta^t)^2} = -(\ln \beta)Tr_t \frac{\alpha\beta^t}{1 + \alpha\beta^t} = -\frac{\ln \beta}{\gamma}Tr_t(\gamma - Tr_t),$$

kde jsme použili  $\gamma - Tr_t = \gamma - \frac{\gamma}{1 + \alpha\beta^t} = \frac{\alpha\beta^t\gamma}{1 + \alpha\beta^t}$ . Oba členy  $Tr_t$  a  $\gamma - Tr_t$  působí proti sobě a příslušný lokální extrém odpovídající inflexnímu bodu nalezneme řešením rovnice  $Tr''_t = 0$ :

$$\begin{aligned} 0 &= Tr''_t = -\frac{\ln \beta}{\gamma}[Tr'_t(\gamma - Tr_t) - Tr_t Tr'_t] = -\frac{\ln \beta}{\gamma}Tr'_t(\gamma - 2Tr_t) \\ &\Leftrightarrow \gamma = 2Tr_t = \frac{2\gamma}{1 + \alpha\beta^t} \Leftrightarrow 1 + \alpha\beta^t = 2 \Leftrightarrow \alpha\beta^t = 1 \Leftrightarrow \ln \alpha + t \ln \beta = 0 \Leftrightarrow t = -\frac{\ln \alpha}{\ln \beta}. \end{aligned}$$

Algoritmus:

- Transformace časové řady:  $z_t = \frac{\Delta x_t}{x_t}$ , kde  $\Delta x_t = x_{t+1} - x_t$ .

Označíme-li  $y_t := Tr_t$ , pak  $\Delta y_t \approx Tr'_t = -\frac{\ln \beta}{\gamma}y_t(\gamma - y_t)$  a tedy  $\frac{\Delta y_t}{y_t} \approx -\ln \beta + \frac{\ln \beta}{\gamma}y_t$ .

Protože  $x_t = y_t + e_t \Rightarrow \frac{1}{x_t} = \frac{1}{y_t} + e_t^{(1)}$  a  $\Delta x_t = \Delta y_t + \Delta e_t$ , dostáváme

$$\begin{aligned} z_t &= \frac{\Delta x_t}{x_t} = \frac{\Delta y_t}{y_t} + \underbrace{\Delta y_t e_t^{(1)} + \Delta e_t \frac{1}{y_t} + \Delta e_t e_t^{(1)}}_{e_t^{(2)}} \\ &= -\ln \beta + \frac{\ln \beta}{\gamma}(x_t - e_t) + e_t^{(2)} \\ &= -\ln \beta + \frac{\ln \beta}{\gamma}x_t + e_t^{(3)}, \text{ kde } e_t^{(3)} = e_t^{(2)} - \frac{\ln \beta}{\gamma}e_t. \end{aligned}$$

- Odhad parametrů  $\beta$  a  $\gamma$ :

Parametry odhadneme v lineárním modelu

$$z_t = \begin{bmatrix} -1 & x_1 \\ \vdots & \vdots \\ -1 & x_n \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + e_t^{(3)}, \text{ kde } \beta_1 := \ln \beta, \beta_2 = \frac{\ln \beta}{\gamma} = \frac{\beta_1}{\gamma}.$$

Pak  $\hat{\beta} = e^{\hat{\beta}_1}$  a  $\hat{\gamma} = \frac{\hat{\beta}_1}{\hat{\beta}_2}$ .

- Odhad parametru  $\alpha$ :

Odhad  $\hat{\alpha}$  nalezneme z  $x_t \approx \gamma/(1 + \alpha\beta^t)$ , odkud

$$\alpha\beta^t \approx \frac{\gamma}{x_t} - 1 \Leftrightarrow \ln \alpha + t \ln \beta \approx \ln \left( \frac{\gamma}{x_t} - 1 \right).$$

Chybu této aproximace potlačíme zprůměrováním pro  $t = 1, 2, \dots, n$ :

$$\frac{1}{n} \sum_{t=1}^n (\ln \alpha + t \ln \beta) = \ln \alpha + \frac{1}{n}(\ln \beta) \frac{n(n+1)}{2} \doteq \frac{1}{n} \sum_{t=1}^n \ln \left( \frac{\gamma}{x_t} - 1 \right).$$

Odtud po dosazení odhadů  $\hat{\beta}_1 = \ln \hat{\beta}$  a  $\hat{\gamma} = \frac{\hat{\beta}_1}{\hat{\beta}_2}$  dostáváme výsledný odhad pro  $\ln \hat{\alpha}$ :

$$\ln \hat{\alpha} = -\frac{(n+1)\hat{\beta}_1}{2} + \frac{1}{n} \sum_{t=1}^n \ln \left( \frac{\hat{\beta}_1}{\hat{\beta}_2 x_t} - 1 \right) \dots \text{ tzv. Rhodesův vztah.}$$

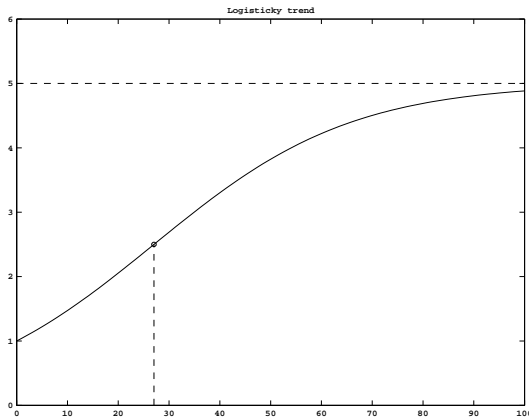
Tedy pak  $\hat{\alpha} = e^{\ln \hat{\alpha}}$ .

Kontrolní kritérium adekvátnosti tohoto modelu:

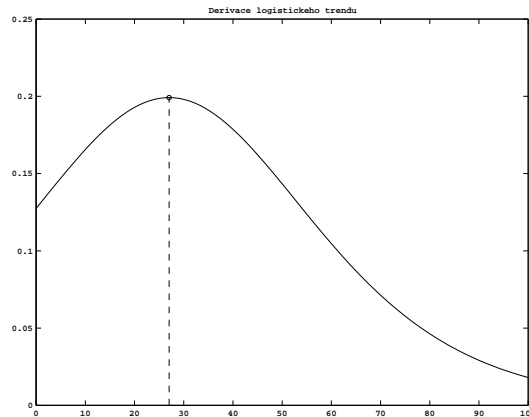
Křivka 1. diferencí  $x_{t+1} - x_t \approx Tr'_t$  osciluje kolem křivky podobné normální hustotě (obr. 2.5.3b)). Navíc  $1/Tr_t$  se zřejmě chová jako modifikovaný exponenciální trend, takže kontrolujeme, zda

$$\left( \frac{1}{x_{t+2}} - \frac{1}{x_{t+1}} \right) / \left( \frac{1}{x_{t+1}} - \frac{1}{x_t} \right)$$

osciluje kolem konstantní hodnoty.

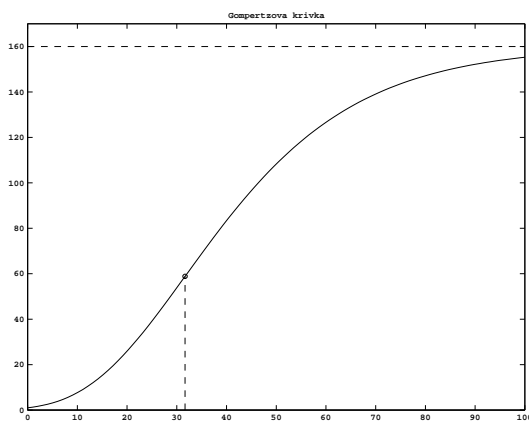


a) Logistická křivka

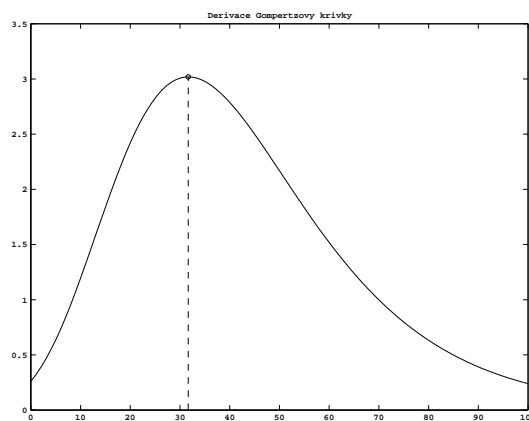


b) Derivace logistické křivky

OBRÁZEK 2.5.3. Logistický trend:  $\alpha = 4$ ,  $\beta = 0,95$ ,  $\gamma = 5$



a) Gompertzova křivka



b) Derivace Gompertzovy křivky

OBRÁZEK 2.5.4. Logistický trend:  $\alpha = -\ln 160$ ,  $\beta = 0,95$ ,  $\gamma = \ln 160$

### 2.5.3.10. Gompertzova křivka (MM-P).

Model:

$$Tr_t = e^{\gamma + \alpha \beta^t}, \quad \alpha, \beta, \gamma \in \mathbb{R}, \quad \beta > 0, \quad t = 1, 2, \dots, n.$$

Má podobné vlastnosti jako logistický trend, tj. opět je možno modelovat inflexi — tentokrát pro  $\alpha < -1$  v bodě  $t = -\ln(-\alpha)/\ln \beta$ . V tomto případě však 1. derivace  $Tr'_t$  není symetrická kolem bodu inflexe, ale je zešíkmena doleva (obr. 2.5.4).

Algoritmus:

Zlogaritmováním převedeme úlohu na modifikovaný exponenciální trend.

Kontrolní kritérium adekvátnosti tohoto modelu:

$\ln Tr_t$  se chová jako modifikovaný exponenciální trend, takže stačí kontrolovat, zda

$$(\ln x_{t+2} - \ln x_{t+1}) / (\ln x_{t+1} - \ln x_t)$$

osciluje kolem konstantní hodnoty.

**Věta 4.18** (Durbin-Levinsonův algoritmus pro rekurentní výpočet  $\Phi_n$ ).

Nechť  $X = \{X_t | t \in \mathbb{Z}\}$  je stacionární náhodný proces s nulovou střední hodnotou  $\mu_X = 0$  a autokovarianční funkcí  $\gamma_X(h) \rightarrow 0$  pro  $h \rightarrow \infty$ . Jestliže  $\hat{X}_{n+1} = \Phi_{n,1}X_n + \dots + \Phi_{n,n}X_1$  je nejlepší lineární predikce jako ve 4.10, pak pro koeficienty  $\Phi_{n,j}$  a střední kvadratickou chybu  $v_n = \mathbb{E}[X_{n+1} - \hat{X}_{n+1}]^2$  platí následující rekurentní vztahy.

**Počáteční krok pro  $n = 0$ :**

$$v_0 = \gamma_X(0). \quad (4.4a)$$

**Rekurentní krok pro  $n > 0$ :**

$$\Phi_{n,n} = [\gamma_X(n) - \overbrace{\sum_{j=1}^{n-1} \Phi_{n-1,j} \gamma_X(n-j)}{=0 \text{ pro } n=1}] v_{n-1}^{-1}, \quad (4.4b)$$

$$v_n = v_{n-1} (1 - |\Phi_{n,n}|^2), \quad (4.4c)$$

$$\begin{bmatrix} \Phi_{n,1} \\ \Phi_{n,2} \\ \vdots \\ \Phi_{n,n-1} \end{bmatrix} = \begin{bmatrix} \Phi_{n-1,1} \\ \Phi_{n-1,2} \\ \vdots \\ \Phi_{n-1,n-1} \end{bmatrix} - \Phi_{n,n} \begin{bmatrix} \bar{\Phi}_{n-1,n-1} \\ \bar{\Phi}_{n-1,n-2} \\ \vdots \\ \bar{\Phi}_{n-1,1} \end{bmatrix} \quad \text{pro } n > 1. \quad (4.4d)$$

*Důkaz.* Viz [1, str. 162]. □

**Definice 4.19** (ARMA proces).

Stochastický proces  $X = \{X_t | t \in \mathbb{Z}\}$  se nazývá **ARMA procesem řádu  $p, q$**  ( $0 \leq p, q < \infty$ ), píšeme  $X \sim ARMA(p, q)$ , jestliže

$$X_t = \underbrace{\Phi_1 X_{t-1} + \Phi_2 X_{t-2} + \dots + \Phi_p X_{t-p}}_{\text{Autoregresní část (AR)}} + \underbrace{Z_t + \Theta_1 Z_{t-1} + \Theta_2 Z_{t-2} + \dots + \Theta_q Z_{t-q}}_{\text{Část klouzavých součtů (MA=Moving Average)}}, \quad (4.5a)$$

kde  $Z := \{Z_t | t \in \mathbb{Z}\} \sim WN(0, \sigma^2)$  a  $\Phi_p \neq 0$ ,  $\Theta_q \neq 0$ ,  $\sigma \neq 0$ . Přepíšeme-li (4.5a) do ekvivalentního tvaru

$$X_t - \Phi_1 X_{t-1} - \Phi_2 X_{t-2} - \dots - \Phi_p X_{t-p} = Z_t + \Theta_1 Z_{t-1} + \Theta_2 Z_{t-2} + \dots + \Theta_q Z_{t-q}, \quad (4.5b)$$

lze použít též zkrácený zápis

$$\Phi(B)X_t = \Theta(B)Z_t \quad \text{nebo} \quad \Phi(z)X(z) = \Theta(z)Z(z), \quad (4.5c)$$

kde při  $\Phi_0 = \Theta_0 = 1$

$$\Phi(z) = 1 - \Phi_1 z - \dots - \Phi_p z^p, \quad \Theta(z) = 1 + \Theta_1 z + \dots + \Theta_q z^q, \quad z \in \mathbb{C}.$$

*Poznámka 4.20.* V předchozí definici bez újmy na obecnosti předpokládáme  $\Theta_0 = 1$ , neboť v opačném případě stačí nahradit bílý šum  $\{Z_t\} \sim WN(0, \sigma^2)$  šumem  $\{\Theta_0 Z_t\} \sim WN(0, (\Theta_0 \sigma)^2)$ .

*Poznámka 4.21* (Speciální případy).

**a) Autoregresní proces (AR proces):**

$$X \sim AR(p) = ARMA(p, 0) : \quad \Phi(B)X_t = Z_t, \text{ neboť } \Theta(z) \equiv 1.$$

Tedy

$$X_t = \Phi_1 X_{t-1} + \Phi_2 X_{t-2} + \dots + \Phi_p X_{t-p} + Z_t. \quad (4.5d)$$

V tomto případě připouštíme i  $p = \infty$  za předpokladu, že  $\Phi := \{\Phi_j\}_{j=1}^{\infty} \in \ell_1$ .

**b) Proces klouzavých součtů (MA proces):**

$$X \sim MA(q) = ARMA(0, q) : \quad X_t = \Theta(B)Z_t, \text{ neboť } \Phi(z) \equiv 1.$$

Tedy

$$X_t = Z_t + \Theta_1 Z_{t-1} + \Theta_2 Z_{t-2} + \dots + \Theta_q Z_{t-q}. \quad (4.5e)$$

V tomto případě připouštíme i  $q = \infty$  za předpokladu, že  $\Theta := \{\Theta_j\}_{j=1}^{\infty} \in \ell_1$ .

c) **Bílý šum (jediný proces, který je současně AR i MA procesem):**

$$X \sim ARMA(0, 0) = AR(0) = MA(0) = WN(0, \sigma^2) : X_t = Z_t.$$

d) **Smíšený ARMA proces:**

$$X \sim ARMA(p, q), 0 < p, q < \infty: \text{Netriviální směs autoregrese a klouzavých součtů.}$$

**Definice 4.22.** Buď  $X = \{X_t | t \in \mathbb{Z}\}$ ,  $X \sim ARMA(p, q)$ .  $X$  se nazývá **kauzální ARMA proces**, jestliže existuje  $\psi = \{\psi_j\}_{j=0}^{\infty}$ ,  $\sum_{j=0}^{\infty} |\psi_j| < \infty$  (tj.  $\psi \in \ell_1$ ) tak, že

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j} \quad (\text{zkráceně } X_t = \psi(B)Z_t), t \in \mathbb{Z}. \quad (4.6a)$$

$X$  se nazývá **invertibilní ARMA proces**, jestliže existuje  $\pi = \{\pi_j\}_{j=0}^{\infty}$ ,  $\sum_{j=0}^{\infty} |\pi_j| < \infty$  (tj.  $\pi \in \ell_1$ ) tak, že

$$\sum_{j=0}^{\infty} \pi_j X_{t-j} = Z_t \quad (\text{zkráceně } \pi(B)X_t = Z_t), t \in \mathbb{Z}. \quad (4.6b)$$

Poznámka 4.23.

Tedy **kauzalita** znamená, že ARMA proces  $X$  je kauzálním lineárním procesem, který lze generovat bílým šumem  $\{Z_t\}$ , neboli  $X \sim MA(\infty)$ .

Podobně **invertibilita** znamená, že bílý šum  $\{Z_t\}$  je kauzálním lineárním procesem, který je generován ARMA procesem  $X$ , což je ekvivalentní s  $X \sim AR(\infty)$ .

Výše jsme použili také skutečnost, že  $\psi_0 = \pi_0 = 1$  (viz dále odst. 4.32, který se týká výpočtu kauzální, resp. invertibilní reprezentace).

**Věta 4.24** (Autokovarianční funkce MA procesu).

$\{X_t\} \sim MA(q)$ ,  $q \leq \infty$  je *stacionární náhodný proces s nulovou střední hodnotou  $\mu_X = 0$  a autokovarianční funkcí*

$$\gamma_X(h) = \sigma^2 \sum_{k=0}^q \Theta_{h+k} \bar{\Theta}_k \quad \text{pro } h \geq 0. \quad (4.7a)$$

Speciálně pro  $q < \infty$  je

$$\gamma_X(h) = \begin{cases} \sigma^2 \sum_{k=0}^{q-h} \Theta_{h+k} \bar{\Theta}_k & \text{pro } 0 \leq h \leq q \\ 0 & \text{pro } h > q \end{cases} \quad (4.7b)$$

a zejména  $\gamma_X(q) = \sigma^2 \Theta_q \neq 0$ , neboť  $\Theta_0 = 1$ .

*Důkaz.* Podle 4.4 je  $\{X_t\}$  stacionární a platí

$$\mu_X = \mu_Z \sum_{j=0}^q \Theta_j = 0, \text{ neboť } \mu_Z = 0.$$

$$\gamma_X(h) = \sum_{j,k=0}^q \Theta_j \bar{\Theta}_k \gamma_Z(h-j+k), \text{ kde } \gamma_Z(h) = \begin{cases} \sigma^2 & \text{pro } h = 0 \\ 0 & \text{pro } h \neq 0 \end{cases}.$$

V sumě zůstanou nenulové členy jen pro  $h-j+k = 0$ , tj. pro  $j = h+k$ , kdy tedy dostáváme (4.7a)

$$\gamma_X(h) = \sum_{k=0}^q \Theta_{h+k} \bar{\Theta}_k \underbrace{\gamma_Z(0)}_{\sigma^2}.$$

Pokud je  $q < \infty$ , pak  $\Theta_{h+k} = 0$  pro  $h+k > q$ , tj. pro  $k > q-h$ , takže (4.7a) lze psát ve tvaru (4.7b).  $\square$

**Důsledek 4.25.**

$$\sigma_X^2 = \gamma_X(0) = \sigma^2 \sum_{k=0}^q |\Theta_k|^2. \quad (4.8a)$$

$$\sigma_X^2 = \sigma^2(1 + |\Theta_1|^2 + |\Theta_2|^2 + \dots + |\Theta_q|^2) \text{ pro } q < \infty.$$

$$\rho_X(h) = \frac{\sum_{k=0}^{q-h} \Theta_{h+k} \bar{\Theta}_k}{\sum_{k=0}^q |\Theta_k|^2} \text{ pro } h \geq 0. \quad (4.8b)$$

**Věta 4.26** (Parciální autokorelační funkce kauzálního AR procesu).

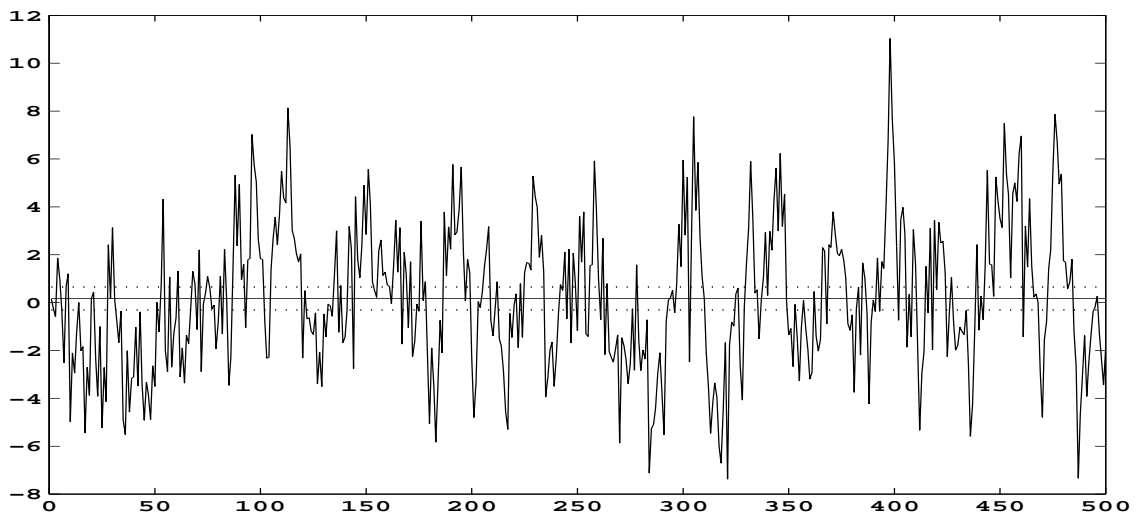
*Nechť  $\{X_t\} \sim AR(p)$ ,  $p < \infty$  je kauzální AR proces. Pak  $\{X_t\}$  je stacionární s nulovou střední hodnotou  $\mu_X = 0$  a parciální autokorelační funkcí  $\alpha_X$ , pro niž  $\alpha_X(p) = \Phi_p \neq 0$  a  $\alpha_X(h) = 0$  pro  $h > p$ .*

*Důkaz.* S uvažáním 4.23 je v důsledku kauzality  $\{X_t\} \sim MA(\infty)$  a tedy stacionární s nulovou střední hodnotou podle 4.24. Podle (4.6a) je  $X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$  a tedy  $X_t \in \mathcal{L}(Z_t, Z_{t-1}, \dots) =: \mathcal{L}_t$  (uzávěr lineárního podprostoru v  $L_2(\Omega, \mathcal{A}, P)$  generovaného náhodnými veličinami  $Z_u$ ,  $u \leq t$ ). Položíme-li  $\Phi_j = 0$  pro  $j > p$ , můžeme dle (4.5d) pro každé  $n \geq p$  psát

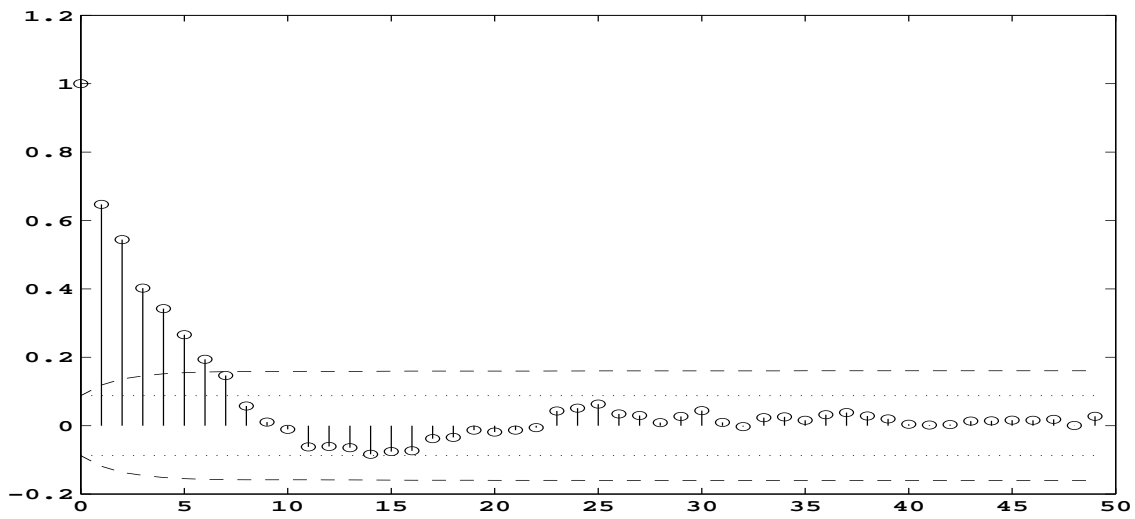
$$X_t = \underbrace{\sum_{j=1}^n \Phi_j X_{t-j}}_{\in \mathcal{L}_{t-1}} + Z_t.$$

Veličiny  $Z_t$  jsou nekorelované, takže  $Z_t \perp Z_u$  pro  $t \neq u$ . Zejména  $Z_t \perp Z_u$  pro  $u < t$  a tedy  $Z_t \perp \mathcal{L}_{t-1}$  s využitím spojitosti skalárního součinu v  $L_2(\Omega, \mathcal{A}, P)$ . Podle věty o ortogonální projekci je  $\hat{X}_t = \sum_{j=1}^n \Phi_j X_{t-j}$  jednoznačně určená nejlepší lineární predikce  $X_t$  pomocí veličin  $X_{t-1}, \dots, X_{t-n}$  pro každé  $n \geq p$ . Podle věty 4.13 pak  $\alpha(p) = \Phi_p$  a  $\alpha(n) = \Phi_n = 0$  pro  $n > p$ .  $\square$

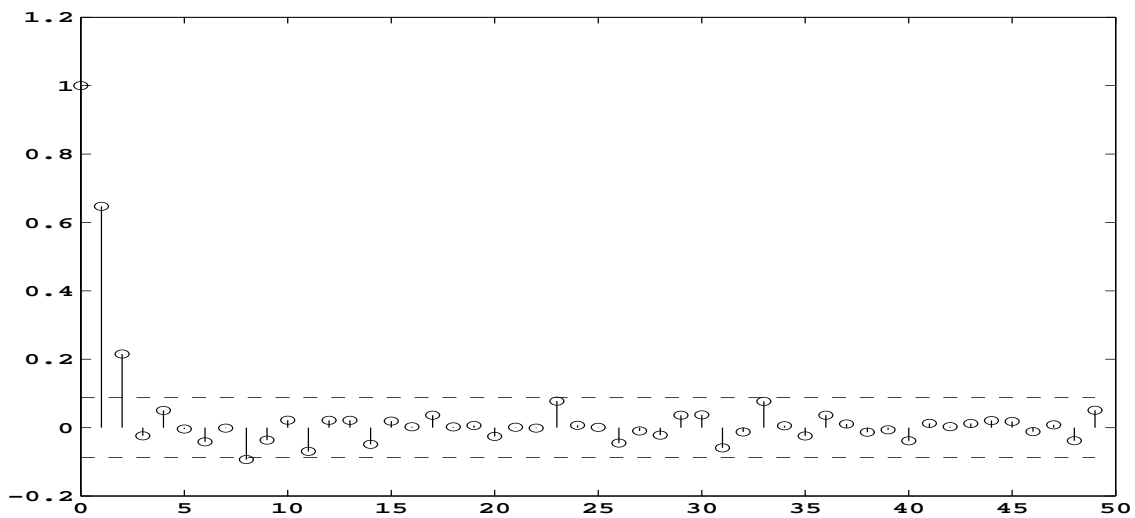
Na obrázcích 4.1, 4.2 a 4.3 jsou typické průběhy odhadnuté autokorelační a parciální autokorelační funkce simulovaných procesů po řadě  $AR(2)$ ,  $MA(2)$  a  $ARMA(2, 2)$ . Čárkovaný pás udává interval spolehlivosti 95% pro nulovou hodnotu. Vidíme, že v případě procesu  $AR(2)$  na obr. 4.1, resp.  $MA(2)$  na obr. 4.2 skutečně  $\alpha_X(h) \approx 0$ , resp.  $\rho_X(h) \approx 0$  pro  $h > 2$  v souladu s větami 4.26, resp. 4.24. V ostatních případech obálka  $\rho_X(h)$ , resp.  $\alpha_X(h)$  (v případě  $ARMA(2, 2)$  na obr. 4.3 dokonce obou) vykazuje exponenciálně klesající, případně navíc i oscilatorický, průběh. Dá se totiž ukázat, že  $\rho_X$  i  $\alpha_X$  jsou v těchto případech lineární kombinací klesajících geometrických posloupností a kosinusovek s geometricky klesající amplitudou.



Autokorelační funkce



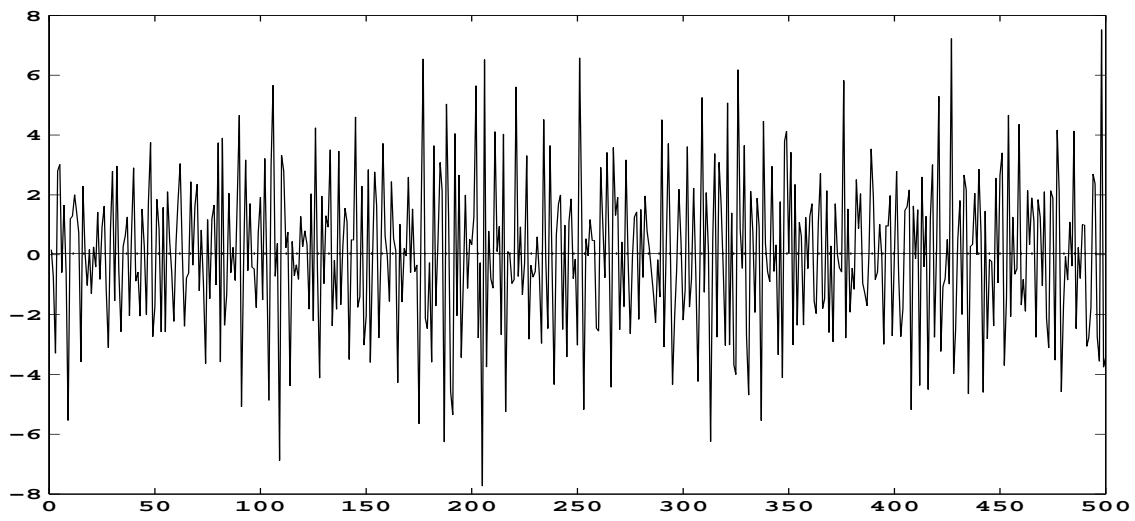
Parciální autokorelační funkce



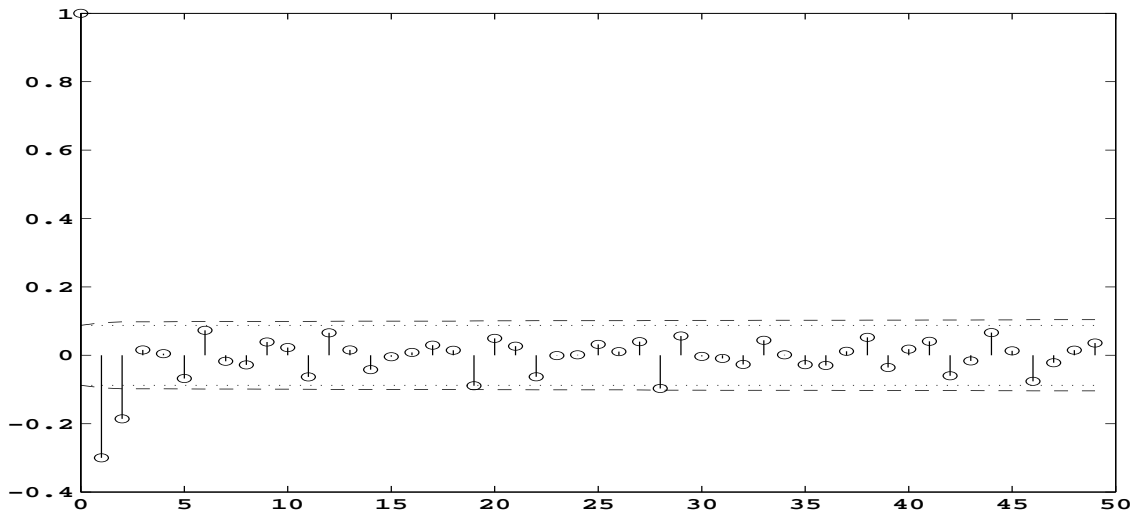
OBRÁZEK 4.1. Proces  $AR(2)$  :  $\Phi = [0.5, 0.2]$ ,  $\sigma^2 = 2.25$ .



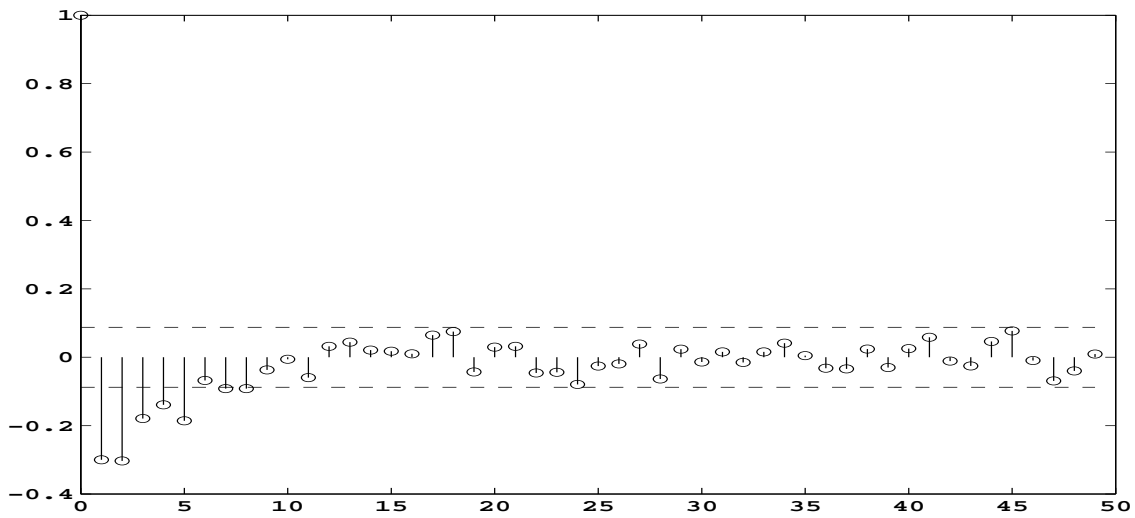
Pozorování procesu,  $n = 500$



Autokorelační funkce

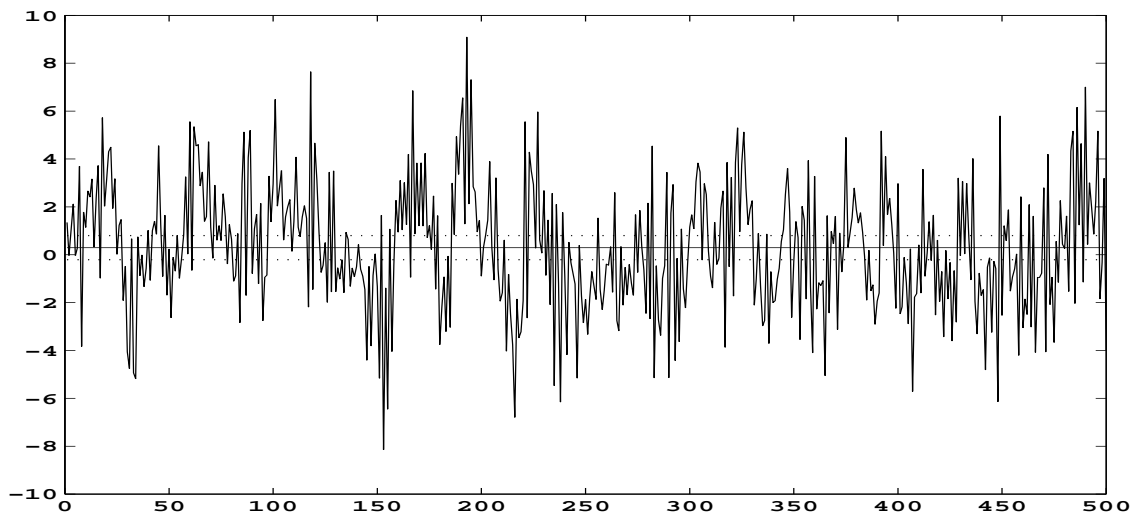


Parciální autokorelační funkce

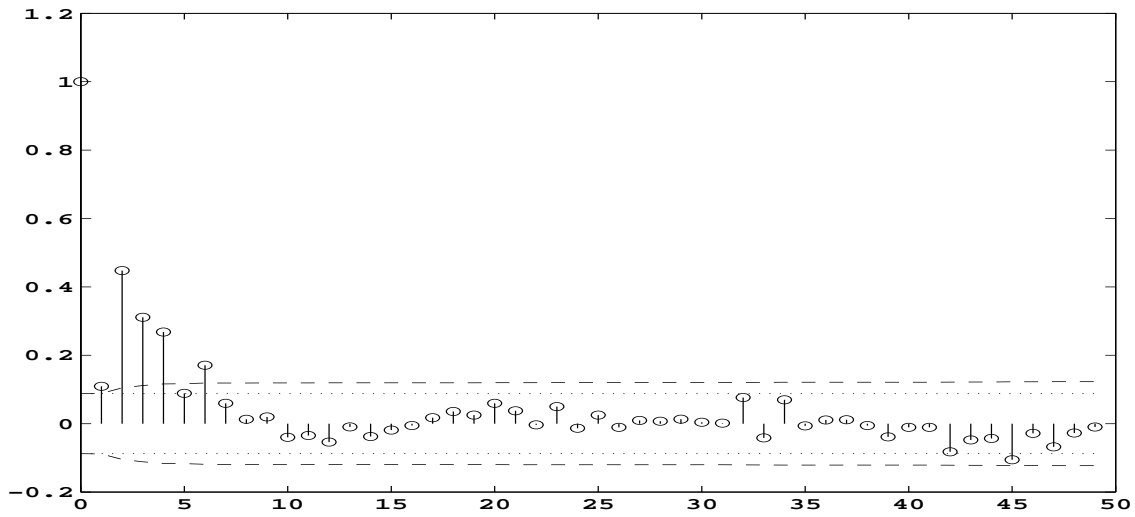


OBRÁZEK 4.2. Proces  $MA(2)$  :  $\Theta = [-0.5, -0.2]$ ,  $\sigma^2 = 2.25$ .

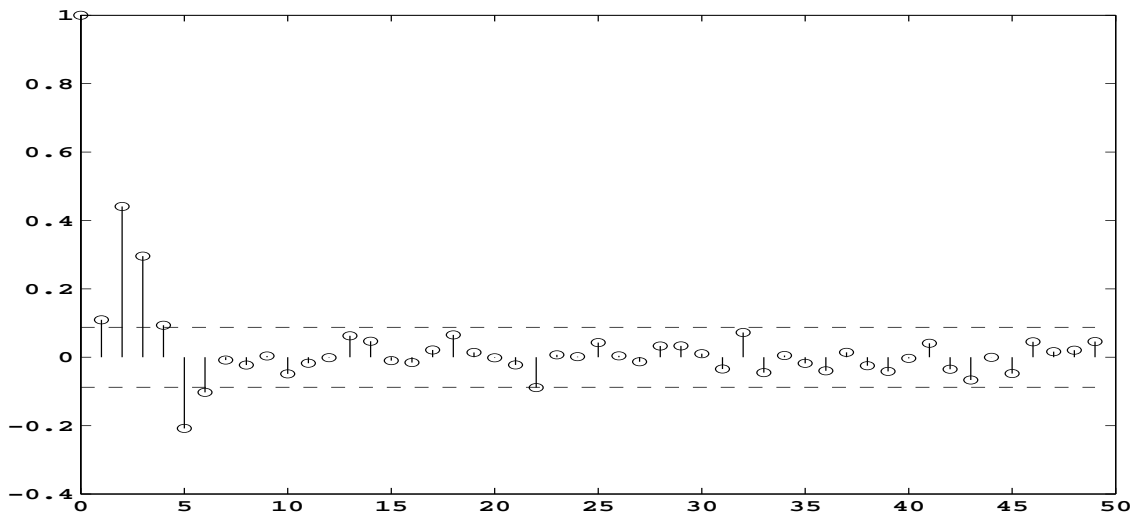
Pozorování procesu,  $n = 500$



Autokorelační funkce



Parciální autokorelační funkce



OBRÁZEK 4.3. Proces  $ARMA(2, 2)$  :  $\Phi = [0.5, 0.2]$ ,  $\Theta = [-0.6, 0.3]$ ,  $\sigma^2 = 2.25$ .

### 5.2.3. Postup hledání SARIMA modelu pro pozorovanou časovou řadu.

#### 1. Odstranění případné heteroskedasticity

Viz 5.1.3/1.

#### 2. Určení sezónní periody $s$

Za  $s$  zvolíme společnou periodu nejvíce dominantních periodických složek, neboli nejmenší společný násobek jejich period. Dominantní periodické složky nejsnáze identifikujeme z periodogramu. Pokud společná perioda vychází příliš dlouhá, volíme za  $s$  periodu dominantní složky s největším počtem jejích harmonických komponent indikovaných v periodogramu jako významných, tj. jejich frekvence jsou (alespoň přibližně) celočíselnými násobky frekvence této dominantní složky. Je přitom třeba dávat pozor, aby byl periodogram spočten z vhodně zkrácené časové řady, tak aby v ní byl obsažen celý počet period délky  $s$ .

Periodu  $s$  výrazně dominantní periodické složky lze zpravidla dosti přesně odečíst i jako periodu oscilací v odhadnuté autokorelační funkci  $\hat{\rho}_X(\cdot)$ .

#### 3. Určení řádů diferencování $d$ a $D$

- Analogicky jako v 5.1.3/2(a) se snažíme zvolit nejmenší  $d$  a  $D$  tak, aby diferencovaná řada  $\Delta^d \Delta_s^D X_t$  se jevila jako stacionární. V praxi je zpravidla  $0 \leq d, D \leq 1$ . Zejména případ  $D > 1$  je velmi málo pravděpodobný, neboť odpovídá proměnlivému kolísání sezónních amplitud, které však bývá většinou spojeno s obdobně proměnným rozptylem. Jeho odstranění vhodnou transformací ad 1 pak obvykle odstraní i kolísání v amplitudách (viz soubor `airpass.m`).
- Zkoumáme chování odhadnuté autokorelační funkce  $\hat{\rho}_W(\cdot)$  diferencované řady. Pokud ještě vykazuje oscilace s periodou  $s$  (lze samozřejmě citlivěji ověřit opět z periodogramu diferencované řady), pak je třeba řád  $D$  ještě zvýšit (oscilace s jinou periodou svědčí pro použití zobecněného SARIMA modelu popsaného v 5.2.5). Pokud  $\hat{\rho}_W(h)$  klesá pro  $h = 1, 2, \dots, \frac{s}{2}$  pomalu (lineárně) a nikoliv exponenciálně, pak je třeba ještě zvětšit řád  $d$ .
- Analogicky jako v 5.1.3(c) hledáme diferencovanou řadu s nejmenším rozptylem.

#### 4. Určení řádů $P, Q, p, q$

- rozborem autokorelační a parciální autokorelační funkce  $\hat{\rho}_W(\cdot)$  a  $\hat{\alpha}_W(\cdot)$   
Dle 4.34 musí pro  $h = 1, 2, 3, \dots$  průběh  $\hat{\rho}_W(hs)$  a  $\hat{\alpha}_W(hs)$  korespondovat s modelem  $ARMA(P, Q)$  a pro  $h = 1, 2, \dots, s-1$  s modelem  $ARMA(p, q)$ .

Zpravidla se používají modely typu  $SARIMA(0, d, q, 0, D, Q, s)$  nebo  $SARIMA(p, d, 0, P, D, 0, s)$ . Naopak modely  $SARIMA(0, d, q, P, D, 0, s)$  nebo  $SARIMA(p, d, 0, 0, D, Q, s)$  se nedoporučují, neboť obvykle vedou k vysokým řádům a tedy k velkému počtu parametrů. Identifikace obecných modelů  $SARIMA(p, d, q, P, D, Q, s)$  bývá většinou dosti obtížná.

- užitím ztrátové funkce (viz 4.36/4)

1° Nejprve určíme  $P, Q$  :

(5.2b)  $\Rightarrow W \sim SARIMA(p, 0, q, P, 0, Q, s) \stackrel{(5.3a)}{\Rightarrow} W^{(k)} \sim ARMA(P, Q), \tilde{\Phi}(B)W_t^{(k)} = \tilde{\Theta}(B)E_t^{(k)}$   
pro  $k = 1, 2, \dots, s$ . Pak  $P, Q$  nalezneme dle minima ztrátové funkce pro  $\frac{1}{s} \sum_{k=1}^s W_t^{(k)}$ , případně můžeme spočítat optimální  $P_k, Q_k$  pro každý parciální proces  $W^{(k)}$  a za  $P, Q$  vybrat dvojici, která se nejčastěji vyskytuje mezi všemi dvojicemi  $P_k, Q_k$ .

2° Pak určíme  $p, q$  :

Minimalizujeme ztrátovou funkci pro  $W \sim SARIMA(p, 0, q, P, 0, Q, s)$ , kde  $P, Q$  jsou již pevné a  $p, q$  vybíráme z dvojic ve vhodném rozmezí  $p = 0, 1, \dots, p_{\max}$  a  $q = 0, 1, \dots, q_{\max}$ . Pokud je  $P = 0$  (resp.  $Q = 0$ ), volíme většinou nejprve  $0 = p = p_{\max}$  (resp.  $0 = q = q_{\max}$ ) pevně.

#### 5. Odhad parametrů modelu a jeho verifikace

V zásadě postupujeme jako ve 4.36/5-7.

#### 5.2.4. Konstrukce predikcí v SARIMA modelu.

Přímý postup zpětného integrování analogický k 5.1.4 je algoritmicky komplikovaný vzhledem k sezónnímu diferencování  $\Delta_s$ .

##### Užijeme proto jiný přístup:

V zásadě lze totiž postupovat jako ve 4.36/7(4), neboť z hlediska predikcí se  $\{X_t\}$  řídí ARMA modelem  $\hat{\Phi}^*(B)X_t = \hat{\Theta}^*(B)Z_t$  (viz 5.2.2(1)), přičemž koeficienty  $\hat{\Phi}^*$  a  $\hat{\Theta}^*$  určíme ze součinů polynomů

$$\hat{\Phi}^*(z) = \hat{\Phi}(z)\hat{\tilde{\Phi}}(z^s)(1-z)^d(1-z^s)^D \text{ a } \hat{\Theta}^*(z) = \hat{\Theta}(z)\hat{\tilde{\Theta}}(z^s),$$

kde  $\hat{\Phi}(z)$ ,  $\hat{\tilde{\Phi}}(z)$ ,  $\hat{\Theta}(z)$  a  $\hat{\tilde{\Theta}}(z)$  jsou známé polynomy, jejichž koeficienty jsou dány odhadnutými parametry v SARIMA modelu:  $\hat{\Phi}$ ,  $\hat{\tilde{\Phi}}$ ,  $\hat{\Theta}$  a  $\hat{\tilde{\Theta}}$ .

V praxi diferencovaná řada  $\{W_t\}$  nemusí mít nulovou střední hodnotu  $\mu_W$  (viz 5.1.2(2)). Pak v (5.2b) je nutno dosadit  $W_t - \mu_W$  místo  $W_t$ :

$$\begin{aligned} \tilde{\Phi}^*(B)(W_t - \mu_W) = \Theta^*(B)Z_t &\Leftrightarrow \tilde{\Phi}^*(B)W_t - \underbrace{(1 - \tilde{\Phi}_1^* - \dots - \tilde{\Phi}_{p+Ps}^*)}_{c}\mu_W = \Theta^*(B)Z_t \Leftrightarrow \\ &\Leftrightarrow \tilde{\Phi}^*(B)W_t = c + \Theta^*(B)Z_t \Leftrightarrow \Phi^*(B)X_t = c + \Theta^*(B)Z_t. \end{aligned}$$

Pak

$$\hat{X}_t^{(1)} = \hat{\Phi}_1^* X_{t-1} + \dots + \hat{\Phi}_{p^*}^* X_{t-p^*} + \hat{\Theta}_1^* Z_{t-1} + \dots + \hat{\Theta}_{q^*}^* Z_{t-q^*} + c$$

je příslušný vztah pro 1-krokovou predikci analogický k 4.36/7(4), avšak navíc s aditivním korekčním členem  $c$ . Pro predikce  $\hat{W}_t^{(1)}$  v diferencované časové řadě stačí pouze místo  $\hat{\Phi}^*$  užít  $\tilde{\Phi}^*$ .

Závěrem poznamenejme, že pro výpočet  $\Phi^*$ ,  $\tilde{\Phi}^*$ ,  $\Theta^*$  a  $c$  lze využít procedury `ARMApar`:  
 $[\Phi^*, \Theta^*, c, \tilde{\Phi}^*] = \text{ARMApar}([p, d, q, P, D, Q, s], \{\Phi, \tilde{\Phi}, \Theta, \tilde{\Theta}\}, \mu_W)$ .

Vlastní predikce  $\hat{X}_t^{(1)}$  pak spočteme jako `c+armasim(x,  $\hat{\Phi}^*$ ,  $\hat{\Theta}^*$ , x0, z0)`.

## LITERATURA

- [1] Peter J. Brockwell and Richard A. Davis, *Time series: Theory and methods*, 2-nd ed., Springer-Verlag, New York, 1991 (corrected 2-nd printing 1993).
- [2] Tomáš Cípra, *Analýza časových řad s aplikacemi v ekonomii*, SNTL, Praha, 1986.

KATEDRA APLIKOVANÉ MATEMATIKY, MASARYKOVA UNIVERZITA V BRNĚ, JANÁČKOVO NÁM. 2A, 662 95 BRNO  
E-mail address: [vesely@math.muni.cz](mailto:vesely@math.muni.cz)