CHAPTER I

Textík k semestrální práci (var. A)

I.1 Časová řada

Vybráno z textu [1]
Pod pojmem časová řada rozumíme data (výsledky pozorování), která jsou chronologicky uspořádána, např. seismický záznam v geofyzice, řada nejvyšších (nejnižších) denních teplot v meteorologii, vývoj koncentrace nečistot v ekologii, změny počtu jedinců nějaké populace v demografii, vývoj rozvodovosti v sociologii nebo vývoj cen v ekonomii. Máme přitom na mysli tzv. statistické (stochastické) řady, které jsou zatíženy nejistotou, nikoliv řady deterministické, jejichž chování lze jednoznačně popsat nějakým matematickým vzorcem.

Jestliže vyjdeme z teorie náhodných procesů, můžeme říci, že časová řada představuje konkrétní realizaci odpovídajícího náhodného (stochastického) procesu.

Cílem analýzy časové řady je určení modelu (mechanismu), podle něhož jsou generována sledovaná data. Znalost tohoto modelu umožňuje předpovídat budoucí vývoj systému a do jisté míry i řídit a optimalizovat chování systému vhodnou volbou vstupních parametrů a počátečních podmínek.

Volba časových okamžiků pozorování. Máme-li možnost volby, pak se doporučuje volit kompromisní řešení. Velká hustota časových bodů pozorování umožňuje dobře vystihnout charakteristické rysy časové řady, ale mohou nastat potíže při výpočtech. V každém případě se však snížíme volit ekvidistantní intervaly mezi sousedními pozorováními.

Při analýze ekonomických časových řad se mohou nepříznivě **projevit problémy spojené s kalendářem** (např. různá délka kalendářních měsíců, různý počet pracovních dnů v měsíci, pohyblivé svátky). V takových případech se zavádí tzv. „standardní měsíc“ o délce 30 dnů nebo standardní počet pracovních dnů v měsíci, nebo se pozorované údaje akumulují (např. použití kvartálních dat namísto měsíčních).

Délka časové řady se definuje jako celkový počet pozorování v časové řadě, nikoliv jako časově rozpětí mezi prvním a posledním pozorováním.

[1] https://www1.osu.cz/ bujok/files/ancas.pdf

I.2 Dekompozice časové řady

Uvažujme časovou řadu $Y_1, Y_2, \ldots, Y_n$.
Časová řada se rozkládá na čtyři základní složky, jimiž jsou

- trend (Tr),
- sezónní složka (Sz),
- cyklická složka (C) a
- náhodná (reziduální) složka (ε).

To znamená, že časovou řadu chápeme jako trend, na který se „nabalují" periodické složky (sezónní a cyklická složka) a náhodná složka (nejčastěji bílý šum).

Dekompozice (rozklad) časové řady je dvojího typu:

- a) aditivní dekompozice ve tvaru

$$Y = Tr + Sz + C + \varepsilon,$$

  kdy všechny složky se měří ve stejných jednotkách jako Yt,
- b) multiplikativní dekompozice ve tvaru

$$Y = TrSzC\varepsilon,$$

  kdy pouze trendová složka je měřena ve stejných jednotkách jako Yt a ostatní složky jsou bezrozměrné veličiny.

Trend reprezentuje dlouhodobé změny v průměrném chování řady (dlouhodobý růst nebo pokles, popř. dlouhodobá konstantní úroveň); je způsoben faktory, jež působí systematicky, tj. ve stejném směru.

Sezónní složka představuje periodické změny, které se odehrávají v průběhu roku a každý rok se opakují. Tyto změny zpravidla souvisejí se střídáním ročních období (jaro, léto, podzim a zima). Pro studium sezónních vlivů se doporučuje pracovat s řadami měsíčních, nejvýše kvartálních měření.

Cyklickou složku chápeme jako fluktuace kolem trendu, při nichž se pravidelně střídají fáze růstu s fázemi poklesu. Délka cyklu i intenzita jednotlivých fází se přitom mohou v průběhu času měnit. Příčiny vedoucí ke vzniku cyklické složky lze zpravidla jen těžko identifikovat.

Náhodná (reziduální) složka představuje náhodné fluktuace, jež nemají systematický charakter. Zahrnuje též chyby měření, chyby ve statistickém zpracování dat (např. zaokrouhlovací chyby). často se předpokládá, že má náhodná složka charakter bílého šumu, tj. že je tvořena hodnotami nezávislých náhodných veličin s nulovou střední hodnotou a nějakým konstantním rozptylem.

Dekompoziční metody pracují pouze se systematickými složkami (trend, sezónní a cyklická složka), přitom se zpravidla využívá metod **regresní analýzy**.

I.3 Předpovědi v časových řadách

Předpovědi v časových řadách mohou být dvojího druhu: bodové a intervalové. **Bodová předpověď** představuje bodový odhad hodnoty časové řady vurčitém budoucím okamžiku. **Intervalová předpověď** (předpovědní interval) je analogií intervalu spolehlivosti.

Chyba předpovědi e_t v čase t je definována vztahem

$$e_t = Y_t - \hat{Y}_t(t-1),$$

v němž Y_t značí skutečně naměřenou hodnotu v čase t a $\hat{Y}_t(t-1)$ předpověď této hodnoty pořízenou v časovém okamžiku bezprostředně předcházejícím. V praxi se nejčastěji používají následující míry kvality předpovědí:

- **Míra SSE:** součet čtvercových chyb SSE (Sum of Squared Errors)

$$SSE = \sum_{t=1}^{n} e_t^2$$

- **Míra MSE:** průměrná čtvercová chyba MSE (Mean Squared Error)

$$MSE = \frac{1}{n}\sum_{t=1}^{n} e_t^2 \quad = \quad \frac{1}{n} SSE$$

- **Míra MAD:** průměrná absolutní odchylka MAD (Mean Absolute Deviation)

$$MAD = \frac{1}{n}\sum_{t=1}^{n} |e_t|$$

- **Míra MAPE:** Průměrná absolutní procentuální chyba MAPE (Mean Absolute Percentual Error)

$$MAPE = \frac{100}{n}\sum_{t=1}^{n} \frac{|e_t|}{Y_t}$$

  Srovnáme-li všechny uvedené míry, zjistíme, že míry MSE a SSE (ve srovnání s MAD) posuzují mnohem přísněji velké chyby předpovědí než chyby malé. Všechny uvedené míry kvality předpovědí závisí na škále,v níž jsou měřeny hodnoty $\{Y_t\}$.

- **Míra SMAPE** Při porovnávání míry kvality předpovědí pro více časových řad je vhodnější použít symetrickou průměrnou absolutní chybu v procentech (SMAPE) definovanou vztahem

$$SMAPE = \frac{100}{h} \sum_{t=n+1}^{n+h} \frac{|e_t|}{\left(|Y_t| + \left|\hat{Y}_t(t-1)\right|\right)/2},$$

  v němž h značí horizont předpovědi.

I.4 Aproximace trendu matematickými funkcemi

Budeme předpokládat, že zkoumaná časová řada má tvar $Y_t = Tr_t + \varepsilon_t$.

Typ nejvhodnější matematické funkce pro danou časovou řadu se určuje na základě předběžné analýzy řady, nejčastěji pomocí grafického záznamu řady nebo teoretických znalostí o průběhu trendové složky.

I.4.1 Konstantní funkce

V tomto případě zřejmě platí $Tr_t = \beta_0, t = 1, 2, \ldots, n$. **Metodou nejmenších čtverců**, tj. minimalizací funkce $S = \sum\limits_{t=1}^{n} (Y_t - \beta_0)^2$ dostaneme pro bodový odhad b_0 parametru β_0 vztah

$$b_0 = \frac{1}{n}\sum_{t=1}^{n} y_t = \bar{y}$$

Předpověď $\hat{Y}_T$ pro $T > n$ je rovněž konstantní, totiž $\hat{Y}_T = b_0$.

I.4.2 Lineární funkce

Je-li trend lineární, tj. $Tr_t = \beta_0 + \beta_1, \quad t = 1, 2, \ldots, n$, dostaneme pro bodové odhady b_0 a b_1 pomocí **metody nejmenších čtverců soustavu dvou normálních rovnic**

$$nb_0 + b_1\sum_{t=1}^{n} t = \sum_{t=1}^{n} Y_t, b_0 + \sum_{t=1}^{n} t + b_1\sum_{t=1}^{n} t^2 = \sum_{t=1}^{n} tY_t$$

Její řešení má tvar

$$b_0 = \frac{\sum_{t=1}^{n} tY_t - \bar{t}\sum_{t=1}^{n} Y_t}{\sum_{t=1}^{n} t^2 - n\bar{t}^2}, \quad b_1 = \bar{Y} - b_0\bar{t}$$

přičemž $\bar{t} = \frac{n+1}{2}$. Pro předpovědi $\hat{Y}_T$ budoucích hodnot časové řady pak platí

$$\hat{Y}_T = b_0 + b_1 T$$

V ekonometrických časových řadách, kde body pozorování jsou zpravidla ekvidistantní, je výhodnější použít modelu $Tr_t = \gamma_0 + \gamma_1(t - \bar{t}), \quad t = 1, 2, \ldots, n$. Vzhledem k tomu, že platí $\sum_{t=1}^{n}(t - \bar{t}) = 0$, soustava normálních rovnice se zjednoduší na tvar

$$nc_0 = \sum_{t=1}^{n} Y_t, c_1\sum_{t=1}^{n}(t - \bar{t})^2 = \sum_{t=1}^{n}(t - \bar{t})Y_t$$

Odtud pro bodové odhady c_0 a c_1 snadno dostaneme

$$c_0 = \bar{Y}, \quad c_1 = \frac{\sum_{t=1}^{n} tY_t - \bar{t}\sum_{t=1}^{n} Y_t}{\sum_{t=1}^{n} t^2 - n\bar{t}^2}$$

a pro předpovědi budoucích hodnot časové řady

$$\hat{Y}_T = c_0 + c_1(T - \bar{t}).$$

I.4.3 Kvadratická funkce

V případě kvadratické funkce můžeme, stejně jako u funkce lineární, použít dvou modelů:

$$Tr_t = \beta_0 + \beta_1 t + \beta_2 t^2, \quad t = 1, 2, \dots, n$$

nebo

$$Tr_t = \gamma_0 + \gamma_1(t - \bar{t}) + \gamma_2(t - \bar{t})^2, \quad t = 1, 2, \dots, n$$

V obou těchto případech dostaneme **soustavu tří normálních rovnic** pro bodové odhady příslušných parametrů. Předpovědi budoucích hodnot časové řady se pak počítají pomocí vztahu

$$\hat{Y} = b_0 + b_1 T + b_2 T^2, \quad \hat{Y} = c_0 + c_1(T - \bar{t}) + c_2(T - \bar{t})^2$$

I.4.4 Exponenciální funkce

Uvažuje se dvouparametrická funkce tvaru

$$Tr_t = \alpha\beta^t, \quad \alpha > 0, \beta > 0, \quad t = 1, 2, \dots, n$$

který se vyznačuje tím, že podíl dvou sousedních hodnot trendu $\left(\frac{Tr_{t+1}}{Tr_t}\right)$, stejně jako podíl dvou sousedních prvních diferencí trendu, má konstantní hodnotu rovnou β. Je-li $\beta > 1$, pak uvažovaná funkce zřejmě exponenciálně roste, zatímco pro $0 < \beta < 1$ exponenciálně klesá.

Pro odhad parametrů exponenciálního trendu se nejčastěji používá „obyčejná“ metoda nejmenších čtverců. Vztah (2.1) se nejprve převede logaritmováním na tvar

$$\log Tr_t = \log\alpha + t\log\beta$$

a odhady obou parametrů se určí minimalizací výrazu

$$\sum_{t=1}^{n} (\log y_t - \log\alpha - t\log\beta)^2$$

Lepší výsledky poskytuje **metoda nejmenších vážených čtverců**, jejíž podstata spočívá v tom, že se jednotlivá pozorování opatřují statistickými vahami $w_t, t = 1, 2, \dots, n$, a minimalizuje se výraz

$$\sum_{t=1}^{n} w_t (\log y_t - \log\alpha - t\log\beta)^2$$

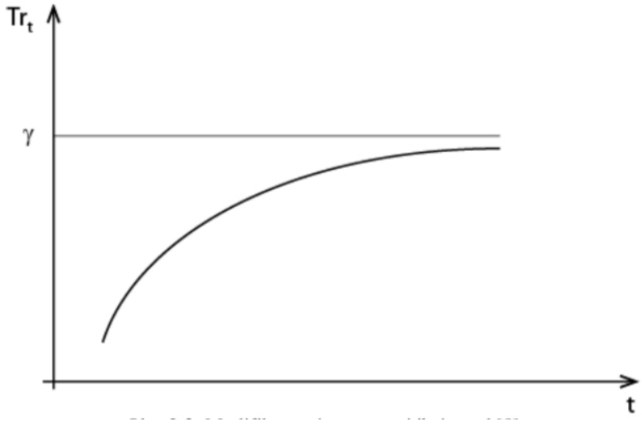
Statistické váhy je vhodné volit tak, aby platilo: $w_t = Y_t^2, t = 1, 2, \dots, n$.

I.4.5 Modifikovaná exponenciální funkce

Tato funkce ve tvaru

$$Tr_t = \gamma + \alpha\beta^t, \quad \alpha > 0,\ 0 < \beta < 1,\ \gamma > 0 \quad t = 1, 2, \dots, n$$

je vlastně zobecněním funkce exponenciální. Doporučuje se vtěch případech, kdy podíl sousedních prvních diferencí řady je konstantní a řada je omezena hodnotou parametru γ . Příklad takového trendu je na obr. 2.3.



Obr. 2.3: Modifikovaný exponenciální trend [8]

Pro odhad parametrů se často využívá následující postup. Soubor všech pozorování se (po případném vynechání jednoho nebo dvou počátečních pozorování) rozdělí na tři stejně velké části o délce m. Sečteme-li pozorování v jednotlivých částech, dostaneme

$$\begin{aligned}
\textstyle\sum_1 Tr_t \approx \ & \textstyle\sum_1 Y_t = & m\gamma + \tfrac{\alpha\beta(\beta^m-1)}{\beta-1} \\
\textstyle\sum_2 Tr_t \approx \ & \textstyle\sum_2 Y_t = & m\gamma + \tfrac{\alpha\beta^m(\beta^m-1)}{\beta-1} \\
\textstyle\sum_3 Tr_t \approx \ & \textstyle\sum_3 Y_t = & m\gamma + \tfrac{\alpha\beta^{2m+1}(\beta^m-1)}{\beta-1}
\end{aligned}$$

řešením této soustavy pak spočteme odhady všech tří parametrů

$$\begin{aligned}
b &= \left(\frac{\sum_1 Y_t - \sum_2 Y_t}{\sum_2 Y_t - \sum_1 Y_t}\right)^{1/m} \\
a &= \frac{b-1}{b(b^m-1}\left(\sum_2 Y_t - \sum_1 Y_t\right) \\
c &= \frac{\sum_1 y_t - ab(b^m-1)(b-1)}{m}
\end{aligned}$$

I.4.6 Logistická funkce

Logistická funkce má tvar

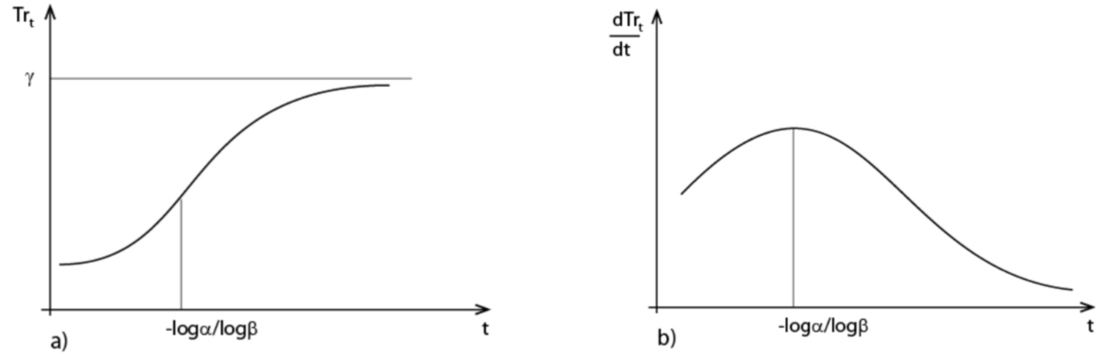
$$Tr_t = \frac{\gamma}{1+\alpha\beta^t}, \quad \alpha > 1,\ 0 < \beta < 1,\ \gamma > 0,\ t = 1, 2, \ldots, n$$

Tato funkce vykazuje inflexní bod $t_{inf} = -\log\alpha\log\beta$, je rostoucí a asymptoticky omezena hodnotou parametru γ . Její graf (viz obr. 2.4a) má průběh typický pro tzv. S-křivky. Důležitou charakteristikou logistické funkce je tzv. růstová funkce (viz obr. 2.4b), kterou dostaneme derivováním podle času

$$\frac{dTr_t}{dt} = -\frac{\ln\beta}{\gamma}Tr_t(\gamma - Tr_t)$$

nejen dosažené úrovni Tr_t , ale i vzdálenosti této úrovně od hladiny γ . Růstová funkce je navíc symetrická „kolem“ inflexního bodu. Pro odhad parametrů logistické funkce lze použít stejnou proceduru jako v případě modifikovaného exponenciálního trendu, v tomto případě se odhadovací procedura aplikuje na časovou řadu hodnot

$1/Y_t$.



Obr. 2.4: Logistický trend: a) logistická funkce, b) růstová funkce [8]

Alternativní metoda odhadu vychází z řady prvních diferencí, tj. z řady $Y_{t+1} - Y_t, t = 1, 2, \ldots, n$. Jestliže ve vztahu pro růstovou funkci nahradíme hodnoty trendové složky Tr_t hodnotami skutečných pozorování Y_t a použijeme aproximace $\dfrac{dTr_t}{dt} \approx Y_{t+1} - Y_t = \Delta_t$, kde Δ_t označuje řadu prvních diferencí původní časové řady, dostaneme (po malé úpravě)

$$\frac{\Delta_t}{y_t} = -\ln\beta + \frac{\ln\beta}{\gamma} Y_t.$$

Odtud už pomocí metody nejmenších čtverců spočteme odhady parametrů β a γ. Pro odhad zbývajícího parametru α použijeme Rhodesova vztahu

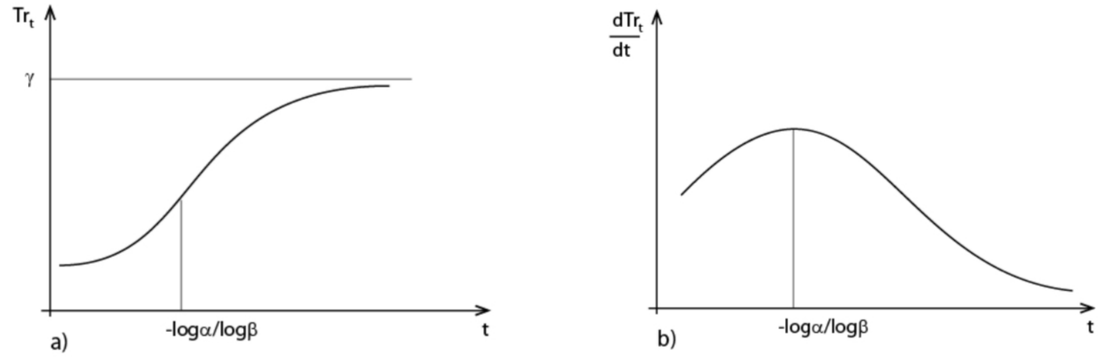
$$\ln\alpha = -\frac{(n+1)\ln\beta}{2} + \frac{1}{n}\sum_{t=1}^{n} \ln\left(\frac{\gamma}{Y_t} - 1\right)$$

I.4.7 Gompertzova funkce

Tuto funkci dostaneme jednoduše transformací modifikované exponenciální funkce na tvar

$$\ln Tr_t = \gamma + \alpha\beta^t, \quad \alpha < -1, \ 0 < \beta < 1, \ t = 1, 2, \ldots, n$$

Gompertzova funkce (viz obr. 2.5a) má inflexi v bodě $t_{inf} = -\log(-\alpha)/\log\beta$, je také rostoucí a asymptoticky omezena. Její graf má také podobu S-křivky. Příslušná růstová funkce (viz obr. 2.5b) není symetrická „kolem“ indexního bodu, ale je kladně zešikmená. Odhad parametrů se provádí stejně jako u modifikované exponenciální funkce, ovšem odhadovací procedura se aplikuje na řadu $\ln Y_t$.



Obr. 2.5: Gompertzův trend: a) Gompertzova funkce, b) růstová funkce [8]

V tabulce 2.1 uvádíme přehled testů používaných v praxi pro výběr vhodné funkce pro aproximaci trendové křivky.

Trendová funkce	Informativní test
Lineární	První diference jsou přibližně konstantní
Kvadratická	Druhé diference jsou přibližně konstantní
Exponenciální	Podíly sousedních hodnot $\dfrac{Y_{t+1}}{Y_t}$ jsou přibližně konstantní
Logistická	Podíly $\dfrac{\left(\dfrac{1}{Y_{t+2}}-\dfrac{1}{Y_{t+1}}\right)}{\left(\dfrac{1}{Y_{t+1}}-\dfrac{1}{Y_t}\right)}$ jsou přibližně konstantní
Gompertzova	Podíly $\dfrac{\ln Y_{t+2}-\ln Y_{t+1}}{\ln Y_{t+1}-\ln Y_t}$jsoupřibližně konstantní

Tab. 2.1: Testy pro výběr aproximace trendové křivky.

Úkol Vyberte libovolnou časovou řadu a znázorněte ji graficky. Navrhněte vhodnou matematickou funkci pro aproximaci trendu této řady a pokuste se určit její parametry pomocí nějakého dostupného matematického software.

Doporučená struktura:

1) stručný popis vybraných dat (včetně původu),
2) tvar vybrané funkce pro aproximaci trendu a zdůvodnění výběru,
3) interpretace hodnot vypočtených parametrů.