



**MATEMATICKO-FYZIKÁLNÍ  
FAKULTA**  
Univerzita Karlova

## **BAKALÁRSKA PRÁCA**

Marek Rathouský

## **Autoregresné modely**

Katedra pravděpodobnosti a matematické statistiky

Vedúci bakalárskej práce: RNDr. Jitka Zichová, Dr.

Študijný program: Matematika

Študijný obor: Finanční matematika

Praha 2017

Prohlašuji, že jsem tuto bakalářskou práci vypracoval(a) samostatně a výhradně s použitím citovaných pramenů, literatury a dalších odborných zdrojů.

Beru na vědomí, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorského zákona v platném znění, zejména skutečnost, že Univerzita Karlova má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle §60 odst. 1 autorského zákona.

V ..... dne .....

Podpis autora

Názov práce: Autoregresné modely

Autor: Marek Rathouský

Katedra: Katedra pravděpodobnosti a matematické statistiky

Vedúci bakalárskej práce: RNDr. Jitka Zichová, Dr., Katedra pravděpodobnosti a matematické statistiky

Abstrakt: Náplňou tejto práce je porovnanie klasického a celočíselného autoregresného modelu prvého rádu. Vzhľadom k rozšírenosti klasického AR(1) modelu je v tejto práci spracovaný v menšej podrobnosti. Vo väčšom detaile je spracovaná teória celočíselného autoregresného modelu prvého rádu. V práci je definovaný operátor  $\circ$  potrebný k zavedeniu INAR(1) a jeho základné vlastnosti s dôkazmi. Pre INAR(1) sú všetky netriviálne vlastnosti v podrobnosti dokázané, odvodená je aj podmienka slabej stacionarity. Pre poissonovský INAR(1) sú popísané základné odhadové metódy. Práca obsahuje aj simulačnú štúdiu sústredenú na skúmanie konvergenzie odhadov.

Kľúčové slová: náhodný proces, časový rad, autoregresný model, INAR

Title: Autoregressive models

Author: Marek Rathouský

Department: Department of Probability and Mathematical Statistics

Supervisor: RNDr. Jitka Zichová, Dr., Department of Probability and Mathematical Statistics

Abstract: The purpose of this thesis is to compare the classic autoregressive model of order 1 to integer autoregressive model of order 1. Considering the popularity of AR(1) model, only the basics are covered within this thesis. The main focus is on the INAR(1) model. Operator  $\circ$  necessary for INAR(1) definition is introduced alongside with its properties with proof. All of the non-trivial properties of INAR(1) are followed by detailed proof, stationarity condition is also derived. Common estimation techniques are described for poisson INAR(1) model. This thesis also contains simulation study, which focuses on the rate of convergence of estimates of parameters.

Keywords: stochastic process, time series, autoregressive model, INAR

Na tomto mieste by som sa chcel poďakovať RNDr. Jitke Zichovej, Dr. za odborné vedenie, cenné rady a pripomienky, ochotu a všetok venovaný čas pri vypracovaní tejto bakalárskej práce.

# Obsah

Úvod	2
<b>1 Náhodné procesy</b>	<b>3</b>
1.1 Základné vlastnosti náhodných procesov . . . . .	3
1.2 Odhady momentových charakteristík . . . . .	4
1.3 Reprezentácie časového radu . . . . .	5
<b>2 Autoregresný model prvého rádu</b>	<b>7</b>
2.1 Definícia a vlastnosti . . . . .	7
2.2 Odhad parametrov modelu . . . . .	8
<b>3 Celočíselný autoregresný model prvého rádu</b>	<b>11</b>
3.1 Operátor $\circ$ . . . . .	11
3.2 Definícia a vlastnosti procesu INAR(1) . . . . .	13
3.3 Poissonovský INAR(1) . . . . .	16
<b>4 Simulačná štúdia</b>	<b>20</b>
4.1 Proces AR(1) . . . . .	20
4.2 Proces INAR(1) . . . . .	22
<b>Záver</b>	<b>24</b>
<b>Zoznam použitej literatúry</b>	<b>25</b>
<b>Zoznam obrázkov</b>	<b>26</b>
<b>Zoznam tabuliek</b>	<b>27</b>
<b>Prílohy</b>	<b>28</b>

# Úvod

Náhodné procesy, ktoré majú charakter časového radu, sa vyskytujú vo všetkých aspektoch bežného života. Väčšina týchto časových radov môže nadobúdať ľubovoľné hodnoty z oboru reálnych čísel, ale existujú aj také, ktorých hodnota udáva počet a teda nadobúdajú iba nezáporné celočíselné hodnoty. K analýze takýchto časových radov nie je vhodné použiť klasický autoregresný model, a preto bol formulovaný celočíselný autoregresný model.

Úlohou tejto práce je porovnanie všeobecného autoregresného modelu prvého rádu a celočíselného autoregresného modelu prvého rádu.

V prvej časti tejto práce je zadaný náhodný proces podľa Lachout a Prášková (2012) a sú uvedené jeho základné vlastnosti potrebné pri práci s autoregresnými modelmi. Hlavným zdrojom pre túto časť je kniha Wei (1990), kapitola 2.

V druhej kapitole je zadaný všeobecný autoregresný model prvého rádu a stručne zhrnuté jeho základné vlastnosti. Taktiež sú popísané základné metódy pre odhad parametrov autoregresného modelu. V tejto časti čerpáme z knihy Wei (1990), kapitola 3.

V tretej časti práce je predstavený celočíselný autoregresný model prvého rádu, uvedené sú jeho najdôležitejšie vlastnosti s odvodením. Pozornosť je venovaná aj špeciálnemu prípadu celočíselného autoregresného modelu, v ktorom má chybový člen Poissonovo rozdelenie. Pre tento proces sú uvedené aj základné odhadové metódy. V tejto časti čerpáme z článku Al-Osh a Alzaid (1987) a knihy Kedem (2002), kapitola 5.

V poslednej časti je pomocou softwaru Mathematica skúmané správanie modelov v závislosti na hodnote regresného parametra a skúmaná je aj rýchlosť konvergencie niektorých odhadových metód a jej závislosť na hodnote parametra.

# 1. Náhodné procesy

## 1.1 Základné vlastnosti náhodných procesov

**Definícia 1** (Náhodný proces). *Náhodným (stochastickým) procesom pre daný pravdepodobnostný priestor  $(\Omega, \mathcal{A}, \mathcal{P})$  nazývame množinu  $\{X_t, t \in T\}$ , kde  $X_t$  je náhodná veličina z pravdepodobnostného priestoru  $(\Omega, \mathcal{A}, \mathcal{P})$  a  $T \subset \mathbb{R}$ .*

Ak v definícii 1 zvolíme množinu  $T = \mathbb{Z} = \{0, \pm 1, \pm 2, \dots\}$  alebo  $T = \mathbb{N}_0$ , potom takýto náhodný proces nazývame procesom s diskretným časom alebo časovým radom. Náhodný proces nazveme reálnym náhodným procesom, ak nadobúda iba hodnoty reálnych čísel. Ďalej v tomto texte, ak nebude napísané inak, pod pojmom náhodný proces budeme rozumieť reálny časový rad.

Pre náhodný proces definujeme 4 základné momentové charakteristiky: stredná hodnota procesu

$$\mu_t = \mathbb{E} X_t,$$

rozptyl procesu

$$\text{var } X_t = \sigma_t^2 = \mathbb{E} (X_t - \mu_t)^2,$$

autokovariančná funkcia procesu

$$\gamma(t_1, t_2) = \mathbb{E} [(X_{t_1} - \mu_{t_1})(X_{t_2} - \mu_{t_2})],$$

autokorelačná funkcia procesu

$$\rho(t_1, t_2) = \frac{\gamma(t_1, t_2)}{\sqrt{\sigma_{t_1}^2} \sqrt{\sigma_{t_2}^2}}. \quad (1.1)$$

Uvažujme konečnú  $n$  prvkovú množinu náhodných veličín  $X_{t_1}, X_{t_2}, \dots, X_{t_n}$  z náhodného procesu  $\{X_t, t = 0, \pm 1, \pm 2, \dots\}$ . Združená distribučná funkcia týchto náhodných veličín je definovaná ako

$$F(x_{t_1}, \dots, x_{t_n}) = \mathbb{P}(X_{t_1} \leq x_{t_1}, \dots, X_{t_n} \leq x_{t_n}).$$

**Definícia 2** (Striktná stacionarita). *Náhodný proces nazývame stacionárny  $n$ -tého rádu v distribúcii, ak je združená distribučná funkcia časovo invariantná pre všetky  $n$  prvkové množiny náhodných veličín. Náhodný proces nazývame striktné stacionárny, ak je stacionárny  $n$ -tého rádu v distribúcii pre  $\forall n \in \mathbb{N}$ .*

*Poznámka.* Stacionarita  $n$ -tého rádu v distribúcii znamená, že pre ľubovoľnú  $n$ -ticu náhodných veličín z procesu platí  $F(x_{t_1}, \dots, x_{t_n}) = F(x_{t_1+k}, \dots, x_{t_n+k})$  pre všetky  $k \in \mathbb{Z}$ .

Ako dôsledok toho, že združená distribučná funkcia jednoznačne určuje marginálne rozdelenia náhodných veličín, stacionarita  $n$ -tého rádu implikuje stacionaritu  $k$ -tého rádu pre  $k \leq n$ . Striktná stacionarita nám zaručuje, že všetky náhodné veličiny z procesu majú rovnaké rozdelenie a v prípade, že  $\mathbb{E} X_t^2 < \infty$ , sú stredná hodnota procesu  $\mu_t$  a rozptyl procesu  $\sigma_t^2$  konštantné funkcie. Ak označíme

$$\mathbb{E} X_t = \mu,$$

pre autokovariančnú funkciu striktnie stacionárneho procesu platí

$$\gamma(t_1, t_2) = \mathbf{E}(X_{t_1} X_{t_2}) - \mu^2 = \mathbf{E}(X_{t_1+k} X_{t_2+k}) - \mu^2$$

pre ľubovoľné  $t_1, t_2, k \in \mathbb{Z}$ . Rovnosť  $\mathbf{E}(X_{t_1} X_{t_2}) = \mathbf{E}(X_{t_1+k} X_{t_2+k})$  plynie z faktu, že vektory  $(X_{t_1}, X_{t_2})^T$  a  $(X_{t_1+k}, X_{t_2+k})^T$  majú rovnaké rozdelenie. Pomocou substitúcie  $t_1 = t - k$ ,  $t_2 = t$  dostaneme

$$\gamma(t - k, t) = \gamma(t, t + k) = \gamma_k.$$

Autokovariančná funkcia teda nezávisí na konkrétnych časoch  $t_1$ ,  $t_2$ , ale je funkciou rozdielu  $t_2 - t_1$ . Z (1.1) priamo vyplýva rovnaká vlastnosť aj pre autokorelačnú funkciu.

Striktná stacionarita nám zabezpečuje mnoho užitočných vlastností, ale v reálnych aplikáciách je len ťažko overiteľná, pretože vyžaduje znalosť združených distribučných funkcií. Preto sa v analýze časových radov častejšie používa stacionarita do momentov 2. rádu, nazývaná aj slabá stacionarita.

**Definícia 3** (Slabá stacionarita). *Náhodný proces je slabo stacionárny, ak sú stredná hodnota a rozptyl procesu konštantné v čase a autokovariančná funkcia je závislá len na rozdieli  $t_2 - t_1$ .*

Autokovariančnú funkciu slabo stacionárneho procesu môžeme teda vyjadriť v tvare

$$\gamma_k = \text{Cov}(X_t, X_{t+k}) = \mathbf{E}[(X_t - \mu)(X_{t+k} - \mu)].$$

Dosadením do (1.1) dostaneme vyjadrenie pre autokorelačnú funkciu

$$\rho_k = \frac{\text{Cov}(X_t, X_{t+k})}{\sqrt{\text{var } X_t \text{ var } X_{t+k}}} = \frac{\gamma_k}{\gamma_0}. \quad (1.2)$$

*Poznámka.* Z rovnosti  $\text{Cov}(X_{t-k}, X_t) = \text{Cov}(X_t, X_{t-k})$  vyplýva, že autokovariančná a autokorelačná funkcia sú párne funkcie a platí teda  $\gamma_k = \gamma_{-k}$  a  $\rho_k = \rho_{-k}$ . Preto je zvykom skúmať tieto funkcie len pre  $k \in \mathbb{N}$ .

V tomto texte si vystačíme so slabou stacionaritou, ktorej výhodou je jednoduchosť overenia, k čomu nám stačí znalosť prvých dvoch momentov. Preto, keď ďalej budeme hovoriť o stacionárnom procese, budeme predpokladať slabú stacionaritu.

## 1.2 Odhady momentových charakteristík

Neznáme stredné hodnoty náhodných veličín z obecného časového radu môžeme odhadnúť skupinovým priemerom v prípade, ak máme k dispozícii dostatok realizácií. V praxi ale často máme v konkrétnom čase  $t$  k dispozícii iba jedinú realizáciu, z ktorej nemá zmysel počítať skupinový priemer. U stacionárnych časových radov je však možné nahradiť skupinový priemer časovým priemerom vzhľadom k tomu, že za predpokladu stacionarity majú všetky náhodné veličiny rovnakú strednú hodnotu.



Predpokladajme teda, že máme k dispozícii realizáciu  $X_1, \dots, X_n$  stacionárneho časového radu. Odhadom strednej hodnoty je potom výberový priemer  $\bar{X}_n$ , ktorý spočítame ako

$$\bar{X}_n = \frac{1}{n} \sum_{t=1}^n X_t.$$

Odhad autokovariančnej funkcie  $\gamma_k$  získame rovnako ako u strednej hodnoty, časovým priemerom

$$\hat{\gamma}_k = \frac{1}{n} \sum_{t=1}^{n-k} (X_t - \bar{X}_n)(X_{t+k} - \bar{X}_n).$$

Keď zvolíme  $k = 0$ , získame odhad rozptylu procesu.

Výberovú autokorelačnú funkciu definujeme v súlade s (1.2) pomocou výberovej autokovariančnej funkcie predpisom

$$\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0} = \frac{\sum_{t=1}^{n-k} (X_t - \bar{X}_n)(X_{t+k} - \bar{X}_n)}{\sum_{t=1}^n (X_t - \bar{X}_n)^2}. \quad (1.3)$$

### 1.3 Reprezentácie časového radu

**Definícia 4** (Biely šum). *Náhodný proces  $\{a_t\}$  nazývame biely šum, ak sa jedná o postupnosť nekorelovaných náhodných veličín s rovnakým rozdelením.*

Z definície bieleho šumu hneď plynie, že sa jedná o stacionárny proces a pre strednú hodnotu a rozptyl procesu platí, že sú konštantné. Väčšinou navyše predpokladáme, že stredná hodnota  $E a_t = 0$ . Pre autokovariančnú funkciu platí

$$\gamma_k = \begin{cases} \text{var } a_t = \sigma^2 & \text{ak } k = 0 \\ 0 & \text{ak } k \neq 0, \end{cases}$$

pre autokorelačnú funkciu

$$\rho_k = \begin{cases} 1 & \text{ak } k = 0 \\ 0 & \text{ak } k \neq 0. \end{cases}$$

Biely šum označujeme ako gaussovský, ak má každá náhodná veličina z postupnosti  $\{a_t\}$  normálne rozdelenie.

Dá sa dokázať, že každý stacionárny časový rad, ktorý neobsahuje žiadnu deterministickú časť, ktorá by sa dala predpovedať z jej vlastnej minulosti, je možné zapísať ako lineárnu kombináciu nekorelovaných náhodných veličín

$$X_t = \mu + a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \dots = \mu + \sum_{j=0}^{\infty} \psi_j a_{t-j}, \quad (1.4)$$

kde  $\psi_0 = 1$ ,  $\{a_t\}$  je biely šum s nulovou strednou hodnotou a  $\sum_{j=0}^{\infty} (\psi_j)^2 < \infty$ . Ak označíme

$$\dot{X}_t = X_t - \mu \quad (1.5)$$

a zavedieme operátor časového posunu  $B$ , pre ktorý platí  $B^j X_t = X_{t-j}$ , môžeme predchádzajúcu rovnosť prepísať v kompaktnej forme

$$\dot{X}_t = \psi(B)a_t,$$

kde  $\psi(B) = \sum_{j=0}^{\infty} \psi_j B^j$ . Forma zápisu (1.4) sa nazýva MA reprezentácia (z anglického moving average), niekedy sa používa aj názov lineárny proces.

Druhou užitočnou formou zápisu je autoregresná reprezentácia, v ktorej vyjadrujeme hodnotu  $X_t$  pomocou minulých hodnôt. Pri konštantnej strednej hodnote môžeme písať

$$\dot{X}_t = \pi_1 \dot{X}_{t-1} + \pi_2 \dot{X}_{t-2} + \dots + a_t, \quad (1.6)$$

alebo ekvivalentne vyjadrené pomocou operátora časového posunu

$$\pi(B)\dot{X}_t = a_t,$$

kde  $\pi(B) = 1 - \pi_1 B - \pi_2 B^2 - \dots$ ,  $\{a_t\}$  je biely šum s nulovou strednou hodnotou a  $\sum_{j=1}^{\infty} |\pi_j| < \infty$ . Náhodný proces nazývame invertibilným, ak je možné ho zapísať v tomto tvare. Platí, že nie každý invertibilný proces je stacionárny. Postačujúca podmienka na stacionaritu invertibilného procesu je, aby korene rovnice  $\pi(B) = 0$  ležali mimo jednotkového kruhu v komplexnej rovine.

## 2. Autoregresný model prvého rádu

Na konci minulej kapitoly sme zaviedli autoregresnú reprezentáciu náhodného procesu, ktorá je veľmi užitočná pri predpovedaní. Problémom ale je, že tento model obsahuje nekonečné množstvo parametrov, ktoré nie je možné odhadnúť z konečne mnoho pozorovaní časového radu, ktoré máme k dispozícii. Preto sa zameriame iba na modely s konečným počtom parametrov. Náhodný proces sa nazýva autoregresný proces rádu  $p$ , ak je vo vyjadrení (1.6) prvých  $p$  parametrov nenulových a zvyšné sa rovnajú nule.

### 2.1 Definícia a vlastnosti

Autoregresný proces prvého rádu, skrátene označovaný aj AR(1), má iba jeden nenulový parameter  $\pi_1 = \phi$  a zvyšné  $\pi_i = 0, i > 1$ .

**Definícia 5** (Proces AR(1)). Pre  $\phi \in \mathbb{R}$  a biely šum  $\{a_t\}$  s nulovou strednou hodnotou definujeme autoregresný proces prvého rádu  $X_t, t \in \mathbb{Z}$  rovnicou

$$X_t = \phi X_{t-1} + a_t, \quad (2.1)$$

alebo pomocou operátora časového posunu

$$(1 - \phi B)X_t = a_t.$$

Tento proces je vždy invertibilný a za podmienky, že korene rovnice  $(1 - \phi B) = 0$  ležia mimo jednotkového kruhu, čo nastane práve vtedy, keď  $|\phi| < 1$ , je aj stacionárny.

*Poznámka.* Aplikáciou strednej hodnoty na rovnosť (2.1) dostávame

$$0 = \mathbb{E} a_t = \mathbb{E} X_t - \phi \mathbb{E} X_{t-1} = \mu(1 - \phi).$$

Za predpokladu stacionarity musí teda byť stredná hodnota procesu AR(1) rovná 0, v opačnom prípade by nastal spor s podmienkou slabšej stacionarity. Nahradením náhodných veličín  $X_t$  a  $X_{t-1}$  v rovnosti (2.1) centrovanými náhodnými veličinami  $\dot{X}_t$  a  $\dot{X}_{t-1}$  podľa (1.5), získame definíciu pre stacionárny AR(1) proces s nenulovou strednou hodnotou  $\mu$ .

Za predpokladu stacionarity môžeme skúmať vlastnosti autokovariančnej a autokorelačnej funkcie procesu AR(1).

Autokovariančnú funkciu pre  $k \geq 1$  spočítame ako

$$\mathbb{E}(\dot{X}_{t-k}\dot{X}_t) = \mathbb{E}(\phi\dot{X}_{t-k}\dot{X}_{t-1}) + \mathbb{E}(\dot{X}_{t-k}a_t),$$

kde člen  $\mathbb{E}(\dot{X}_{t-k}a_t)$  upravíme opakovaným dosadením za  $\dot{X}_{t-k}$  podľa (2.1). Tým získame zápis  $\dot{X}_{t-k}$  v tvare lineárneho procesu (1.4) a využijeme fakt, že náhodné veličiny v postupnosti  $\{a_t\}$  sú nekorelované.

Dostaneme  $E(\dot{X}_{t-k}a_t) = E(\sum_{j=0}^{\infty} \phi^j a_{t-k-j}a_t) = 0$ . Autokovariančnú funkciu procesu AR(1) teda vieme zapísať ako

$$\gamma_k = \phi\gamma_{k-1}, \quad k \geq 1. \quad (2.2)$$

K tomu, aby sme tento rekurentný vzorec k výpočtu autokovariančnej funkcie mohli použiť, potrebujeme ešte určiť počiatočnú hodnotu  $\gamma_0$ , teda rozptyl procesu AR(1). Ten spočítame rozpisáním  $\dot{X}_t$  v tvare lineárneho procesu

$$\gamma_0 = E \dot{X}_t^2 = E \left( \sum_{j=0}^{\infty} \phi^j a_{t-j} \right)^2 = \sigma^2 \sum_{j=0}^{\infty} \phi^{2j} = \sigma^2 \frac{1}{1-\phi^2}. \quad (2.3)$$

Autokorelačnú funkciu dostaneme postupným dosadzovaním podľa (2.2) do (1.2)

$$\rho_k = \frac{\phi\gamma_{k-1}}{\gamma_0} = \phi\rho_{k-1} = \phi^k, \quad k \geq 1, \quad (2.4)$$

pričom v poslednej rovnosti sme využili fakt, že  $\rho_0 = 1$ .

## 2.2 Odhad parametrov modelu

Ďalej sa zameriame na odhad neznámych parametrov  $\mu$ ,  $\phi$  a  $\sigma^2$  na základe realizácie  $X_1, \dots, X_n$  časového radu.

### Odhad momentovou metódou

Odhadom strednej hodnoty stacionárneho autoregresného procesu prvého rádu momentovou metódou je  $\bar{X}_n = \frac{1}{n} \sum_{t=1}^n X_t$ , ako bolo uvedené v kapitole 1.2. Pre odhad parametra  $\phi$  nahradíme v rovnosti (2.4) teoretickú autokorelačnú funkciu jej odhadom  $\hat{\rho}_k$  spočítaným podľa (1.3) a dosadíme  $k = 1$ . Dostaneme

$$\hat{\phi} = \hat{\rho}_1 = \frac{\sum_{t=1}^{n-1} (X_t - \bar{X}_n)(X_{t+1} - \bar{X}_n)}{\sum_{t=1}^n (X_t - \bar{X}_n)^2}.$$

Neznámy rozptyl bieleho šumu  $\{a_t\}$  odhadneme vyjadrením z rovnosti (2.3) a nahradením teoretických momentových charakteristík ich empirickými hodnotami

$$\hat{\sigma}^2 = \hat{\gamma}_0(1 - \hat{\phi}^2) = \frac{[\sum_{t=1}^n (X_t - \bar{X}_n)^2]^2 - [\sum_{t=1}^{n-1} (X_t - \bar{X}_n)(X_{t+1} - \bar{X}_n)]^2}{n \sum_{t=1}^n (X_t - \bar{X}_n)^2}.$$

### Odhad metódou najmenších štvorcov

Odhad parametra  $\phi$  pomocou metódy najmenších štvorcov konštruujeme tak, aby platilo

$$\hat{\phi} = \arg \min_{\phi \in (-1,1)} \sum_{t=2}^n (\dot{X}_t - \phi\dot{X}_{t-1})^2.$$

Argument minima sumy nájdeme jej deriváciou, ktorú následne položíme rovnú nule

$$\begin{aligned} -\sum_{t=2}^n 2(\dot{X}_t - \hat{\phi}\dot{X}_{t-1})\dot{X}_{t-1} &= 0 \\ \sum_{t=2}^n (\dot{X}_t\dot{X}_{t-1}) - \hat{\phi}\sum_{t=2}^n \dot{X}_{t-1}^2 &= 0 \\ \hat{\phi} &= \frac{\sum_{t=2}^n \dot{X}_t\dot{X}_{t-1}}{\sum_{t=2}^n \dot{X}_{t-1}^2}. \end{aligned}$$

Keď navyiac dosadíme do predchádzajúceho vyjadrenia  $\dot{X}_t = X_t - \mu$  a neznámu strednú hodnotu nahradíme odhadom  $\bar{X}_n$  dostávame

$$\hat{\phi} = \frac{\sum_{t=1}^{n-1} (X_t - \bar{X}_n)(X_{t+1} - \bar{X}_n)}{\sum_{t=1}^{n-1} (X_t - \bar{X}_n)^2}.$$

Podľa Cipra (2008) str. 344 spĺňa stacionárny autoregresný proces prvého rádu predpoklady regresného modelu s náhodnými regresormi, z čoho vyplýva, že vyššie uvedený odhad parametra  $\phi$  je konzistentný a pre odhad rozptylu bieleho šumu  $\sigma^2$  môžeme použiť odhad odvodený z teórie regresného modelu. Ten má všeobecný tvar  $\hat{\sigma}^2 = (\mathbf{Y}^T \mathbf{Y} - \mathbf{b}^T \mathbf{X}^T \mathbf{Y}) / (n - \dim(\mathbf{b}))$ , kde v našom prípade je  $\mathbf{Y} = (\dot{X}_2, \dots, \dot{X}_n)^T$ ,  $\mathbf{X} = (\dot{X}_1, \dots, \dot{X}_{n-1})^T$  a  $\mathbf{b} = \hat{\phi}$ . Odhad rozptylu bieleho šumu je potom

$$\hat{\sigma}^2 = \frac{\sum_{t=2}^n (X_t - \bar{X}_n)^2 \sum_{t=1}^{n-1} (X_t - \bar{X}_n)^2 - [\sum_{t=1}^{n-1} (X_t - \bar{X}_n)(X_{t+1} - \bar{X}_n)]^2}{(n-1) \sum_{t=1}^{n-1} (X_t - \bar{X}_n)^2}.$$

## Odhad metódou maximálnej vierohodnosti

Maximálne vierohodný odhad je bod, v ktorom je maximalizovaná vierohodnostná funkcia pozorovaní. Na to, aby sme mohli vierohodnostnú funkciu zostrojiť, potrebujeme združenú hustotu pozorovaní  $X_1, \dots, X_n$ . Ak predpokladáme, že  $\{a_t\}$  je gaussovský biely šum, tak z toho vyplýva, že  $X_1 \sim N(\mu, \sigma^2 / (1 - \phi^2))$  a hustota  $X_1$  teda je

$$f_{X_1}(x_1; \boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi\sigma^2/(1-\phi^2)}} \exp\left\{-\frac{(x_1 - \mu)^2}{2\sigma^2/(1-\phi^2)}\right\}, \text{ kde } \boldsymbol{\theta} = (\mu, \phi, \sigma^2).$$

Ďalej uvažujme podmienené rozdelenie  $X_2$  pri známom  $X_1 = x_1$ . Z (2.1) plynie, že  $X_2 = \phi(x_1 - \mu) + \mu + a_2$ , a teda  $X_2$  ma podmienené rozdelenie  $N(\mu(1-\phi) + \phi x_1, \sigma^2)$  s hustotou

$$f_{X_2|X_1}(x_2|x_1; \boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{[x_2 - \mu(1-\phi) - \phi x_1]^2}{2\sigma^2}\right\}.$$

Združená hustota oboch pozorovaní je potom

$$f_{X_1, X_2}(x_1, x_2; \boldsymbol{\theta}) = f_{X_2|X_1}(x_2|x_1; \boldsymbol{\theta})f_{X_1}(x_1; \boldsymbol{\theta}).$$

Podobne, podmienené rozdelenie tretieho pozorovania  $X_3$  pri známom  $X_2 = x_2$  a  $X_1 = x_1$  má hustotu

$$f_{X_3|X_2, X_1}(x_3|x_2, x_1; \boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{[x_3 - \mu(1 - \phi) - \phi x_2]^2}{2\sigma^2} \right\}$$

a združená hustota prvých troch pozorovaní je

$$\begin{aligned} f_{X_1, X_2, X_3}(x_1, x_2, x_3; \boldsymbol{\theta}) &= f_{X_3|X_2, X_1}(x_3|x_2, x_1; \boldsymbol{\theta}) f_{X_1, X_2}(x_1, x_2; \boldsymbol{\theta}) \\ &= f_{X_3|X_2, X_1}(x_3|x_2, x_1; \boldsymbol{\theta}) f_{X_2|X_1}(x_2|x_1; \boldsymbol{\theta}) f_{X_1}(x_1; \boldsymbol{\theta}). \end{aligned}$$

Opakovaním analogického postupu pre každé pozorovanie dostaneme združenú hustotu všetkých pozorovaní  $X_1, \dots, X_n$

$$\begin{aligned} f_{X_1, \dots, X_n}(x_1, \dots, x_n; \boldsymbol{\theta}) &= f_{X_n|X_{n-1}, \dots, X_1}(x_n|x_{n-1}, \dots, x_1; \boldsymbol{\theta}) \\ &\quad \times f_{X_1, \dots, X_{n-1}}(x_1, \dots, x_{n-1}; \boldsymbol{\theta}) \\ &= f_{X_1}(x_1; \boldsymbol{\theta}) \prod_{t=2}^n f_{X_t|X_{t-1}, \dots, X_1}(x_t|x_{t-1}, \dots, x_1; \boldsymbol{\theta}). \end{aligned} \quad (2.5)$$

Vierohodnostná funkcia  $L(\boldsymbol{\theta})$  je náhodná funkcia (2.5) braná ako funkcia neznámeho parametru  $\boldsymbol{\theta}$ . Pre  $n$  pozorovaní má vierohodnostná funkcia tvar

$$\begin{aligned} L(\boldsymbol{\theta}) &= \frac{1}{\sqrt{2\pi\sigma^2/(1-\phi^2)}} \exp \left\{ -\frac{(X_1 - \mu)^2}{2\sigma^2/(1-\phi^2)} \right\} \\ &\quad \times \prod_{t=2}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{[X_t - \mu(1-\phi) - \phi X_{t-1}]^2}{2\sigma^2} \right\}. \end{aligned}$$

Vzhľadom k jej tvaru je jednoduchšie hľadať argument minima logaritmu vierohodnostnej funkcie  $l(\boldsymbol{\theta})$

$$\begin{aligned} l(\boldsymbol{\theta}) &= -\frac{n}{2} \log(2\pi\sigma^2) + \frac{1}{2} \log(1-\phi^2) - \frac{(X_1 - \mu)^2}{2\sigma^2/(1-\phi^2)} \\ &\quad - \sum_{t=2}^n \frac{[X_t - \mu(1-\phi) - \phi X_{t-1}]^2}{2\sigma^2}. \end{aligned}$$

Minimum vierohodnostnej funkcie je v bode, v ktorom sú všetky parciálne derivácie podľa neznámych parametrov  $\mu$ ,  $\phi$ ,  $\sigma^2$  rovné nule. Hľadáme teda riešenie sústavy rovníc

$$\begin{aligned} \frac{(X_1 - \mu)(1 - \phi^2)}{\sigma^2} + \sum_{t=2}^n \frac{[X_t - \mu(1 - \phi) - \phi X_{t-1}](1 - \phi)}{\sigma^2} &= 0 \\ -\frac{\phi}{1 - \phi^2} + \frac{(X_1 - \mu)^2 \phi}{\sigma^2} + \sum_{t=2}^n \frac{[X_t - \mu(1 - \phi) - \phi X_{t-1}](X_{t-1} - \mu)}{\sigma^2} &= 0 \\ -\frac{n}{2\sigma^2} + \frac{(X_1 - \mu)^2(1 - \phi^2)}{2(\sigma^2)^2} + \sum_{t=2}^n \frac{[X_t - \mu(1 - \phi) - \phi X_{t-1}]^2}{2(\sigma^2)^2} &= 0. \end{aligned}$$

Explicitné riešenie  $\hat{\mu}$ ,  $\hat{\phi}$ ,  $\hat{\sigma}^2$  tejto sústavy neuvádzame, pretože v praxi sa rieši numericky pomocou vhodného softwaru.

# 3. Celočíselný autoregresný model prvého rádu

V predchádzajúcej kapitole sme sa venovali autoregresnému modelu pre náhodné procesy, ktoré nadobúdali ľubovoľné hodnoty z oboru reálnych čísel. V tejto kapitole sa zameriame na procesy, ktoré nadobúdajú iba nezáporné celočíselné hodnoty.

## 3.1 Operátor $\circ$

Na to, aby sme mohli riadne definovať celočíselný autoregresný model, musíme zaviesť nový operátor  $\circ$ .

**Definícia 6** (Operátor  $\circ$ ). *Nech  $X$  je nezáporná celočíselná náhodná veličina a  $\alpha \in [0,1]$ . Potom operátor  $\circ$  definujeme ako*

$$\alpha \circ X = \sum_{i=1}^X Y_i, \quad (3.1)$$

kde  $\{Y_i\}$  je postupnosť nezávislých rovnako rozdelených náhodných veličín s alternatívnym rozdelením s parametrom  $\alpha$ , ktorá je nezávislá na  $X$ .

*Poznámka.* Niekedy sa rozdelenie náhodnej veličiny v tvare  $\sum_{i=0}^X Y_i$ , kde  $\{Y_i\}$  je postupnosť nezávislých rovnako rozdelených náhodných veličín a  $X$  je nezáporná celočíselná náhodná veličina nezávislá na  $\{Y_i\}$ , nazýva zložené rozdelenie. Uplatňuje sa napríklad v teórii neživotného poistenia pri výpočte úhrnu škôd za dané časové obdobie.

Náhodná veličina  $\alpha \circ X$  teda vyjadruje počet úspechov, ktoré nastali v náhodnom počte nezávislých pokusov, pričom pravdepodobnosť úspechu  $\alpha$  je rovnaká pre všetky pokusy. V nasledujúcom tvrdení uvedieme základné vlastnosti náhodnej veličiny  $\alpha \circ X$ .

**Tvrdenie 1.** *Pre náhodnú veličinu  $\alpha \circ X$  definovanú vzťahom (3.1) platí:*

$$(i) \quad \beta \circ (\alpha \circ X) \sim (\beta\alpha) \circ X \quad (3.2)$$

$$(ii) \quad E(\alpha \circ X | X) = \alpha X$$

$$(iii) \quad E(\alpha \circ X) = \alpha E X \quad (3.3)$$

$$(iv) \quad \text{var}(\alpha \circ X | X) = \alpha(1 - \alpha)X$$

$$(v) \quad \text{var}(\alpha \circ X) = \alpha^2 \text{var} X + \alpha(1 - \alpha) E X \quad (3.4)$$

*Dôkaz.*

- (i) Vzťah (3.2) dokážeme pomocou vytvárajúcej funkcie postupnosti pravdepodobností, ktorej definícia a vlastnosti potrebné pre tento dôkaz sú uvedené v Prílohe 1.

Výraz  $\beta \circ (\alpha \circ X)$  vieme podľa (3.1) zapísať ako  $\sum_{i=1}^{\alpha \circ X} Z_i$ . Vytvárajúcu funkciu pre túto náhodnú veličinu vieme zapísať podľa tvrdenia 3 v Prílohe 1 ako  $P_X(P_{Y_1}(P_{Z_1}(s)))$ . Napíšme teraz vytvárajúcu funkciu pre náhodnú veličinu  $Y_1$  s alternatívnym rozdelením

$$P_{Y_1}(s) = P(Y_1 = 0) + P(Y_1 = 1)s = 1 - \alpha + \alpha s.$$

Analogicky pre veličinu  $Z_1$  dostávame

$$P_{Z_1}(s) = 1 - \beta + \beta s.$$

Po dosadení vytvárajúcej funkcie náhodnej veličiny  $Z_1$  dostaneme

$$P_{Y_1}(P_{Z_1}(s)) = 1 - \alpha + \alpha(1 - \beta + \beta s) = 1 - \alpha\beta + \alpha\beta s,$$

čo je vytvárajúca funkcia postupnosti pravdepodobností náhodnej veličiny s alternatívnym rozdelením s parametrom  $\alpha\beta$ . Vďaka tomu, že vytvárajúca funkcia jednoznačne určuje pravdepodobnostné rozdelenie, je rovnosť rozdelení dokázaná.

- (ii) Pre pevné  $X$  dostávame súčet  $X$  nezávislých alternatívne rozdelených náhodných veličín s rovnakým parametrom  $\alpha$ , ktorý má binomické rozdelenie s parametrami  $X$  a  $\alpha$ . Stredná hodnota binomického rozdelenia s týmito parametrami je  $\alpha X$ .
- (iii) Pri dôkaze rovnosti (3.3) použijeme vlastnosť strednej hodnoty

$$\mathbf{E}(\alpha \circ X) = \mathbf{E}[\mathbf{E}(\alpha \circ X|X)] = \mathbf{E}(\alpha X) = \alpha \mathbf{E} X.$$

- (iv) Použitím rovnakej úvahy ako v časti dôkazu (ii) dostaneme rozptyl náhodnej veličiny s binomickým rozdelením s parametrami  $X$  a  $\alpha$ , ktorý je rovný  $\alpha(1 - \alpha)X$ .
- (v) V tejto časti dôkazu využijeme vlastnosť rozptylu, podľa ktorej môžeme písať

$$\begin{aligned} \text{var}[\alpha \circ X] &= \text{var}[\mathbf{E}(\alpha \circ X|X)] + \mathbf{E}[\text{var}(\alpha \circ X|X)] \\ &= \alpha^2 \text{var} X + \alpha(1 - \alpha) \mathbf{E} X, \end{aligned}$$

kde v druhej rovnosti sme využili výsledky z časti (ii) a (iv) tohto dôkazu.

□



## 3.2 Definícia a vlastnosti procesu INAR(1)

Štruktúra celočíselného autoregresného procesu prvého rádu sa podobá na štruktúru autoregresného modelu popísaného v kapitole 2 v tom zmysle, že namiesto násobenia medzi parametrom a náhodnou veličinou sa využíva operátor  $\circ$ . Budeme ho značiť INAR(1).

**Definícia 7** (Proces INAR(1)). *Nech  $\alpha \in [0,1]$  a  $\{\epsilon_t\}$  je postupnosť celočíselných nezáporných nezávislých rovnako rozdelených náhodných veličín, pre ktoré platí  $E \epsilon_t = \mu$  a  $\text{var} \epsilon_t = \sigma^2$ . Celočíselný autoregresný proces prvého rádu  $X_t$ ,  $t \in \mathbb{N}$  potom definujeme rovnicou*

$$X_t = \alpha \circ X_{t-1} + \epsilon_t, \quad (3.5)$$

pričom predpokladáme, že  $X_0$  je celočíselná nezáporná náhodná veličina, alternatívne rozdelené veličiny z postupnosti  $\{Y_{t,i}\}$ , z ktorých vznikajú náhodné veličiny  $\alpha \circ X_t$  pre  $t = 0,1, \dots$  sú nezávislé a postupnosti  $\{Y_{t,i}\}$ ,  $\{\epsilon_t\}$  a náhodná veličina  $X_0$  sú taktiež vzájomne nezávislé.

Hodnotu takto definovaného procesu v čase  $t$  môžeme interpretovať ako počet elementov pozostalých v procese z času  $t-1$ , pričom každý má pravdepodobnosť prežitia  $\alpha$ , a elementov, ktoré vstúpili do systému v časovom intervale  $(t-1, t]$  v podobe inovácie  $\epsilon_t$ .

Opakovaným dosadzovaním za  $X_t$  podľa (3.5) dostaneme využitím vlastnosti (3.2) MA reprezentáciu procesu INAR(1) v tvare

$$\begin{aligned} X_t &= \alpha \circ X_{t-1} + \epsilon_t \\ &= \alpha \circ (\alpha \circ X_{t-2} + \epsilon_{t-1}) + \epsilon_t \\ &= \alpha^2 \circ X_{t-2} + \alpha \circ \epsilon_{t-1} + \epsilon_t \\ &= \dots \\ &= \sum_{i=0}^{\infty} \alpha^i \circ \epsilon_{t-i} = \sum_{i=0}^{\infty} \sum_{j=1}^{\epsilon_{t-i}} Y_{ij}, \end{aligned} \quad (3.6)$$

kde  $\forall i,j$  sú  $Y_{ij}$  nezávislé alternatívne rozdelené náhodné veličiny.

### Základné momenty

Strednú hodnotu procesu INAR(1) vieme vyjadriť v tvare

$$E X_t = E(\alpha \circ X_{t-1} + \epsilon_t) = \alpha E X_{t-1} + \mu = \dots = \alpha^t E X_0 + \mu \sum_{j=0}^{t-1} \alpha^j. \quad (3.7)$$

V prvej rovnosti sme  $X_t$  rozpísali podľa (3.5), v druhej sme využili (3.3) a opakovaním sme dospeli k finálnemu tvaru.

Pre rozptyl platí

$$\begin{aligned} \text{var} X_t &= \text{var}(\alpha \circ X_{t-1} + \epsilon_t) = \alpha^2 \text{var} X_{t-1} + \alpha(1-\alpha) E X_{t-1} + \sigma^2 \\ &= \dots = \alpha^{2t} \text{var} X_0 + (1-\alpha) \sum_{j=1}^t \alpha^{2j-1} E X_{t-j} + \sigma^2 \sum_{j=1}^t \alpha^{2(j-1)}, \end{aligned} \quad (3.8)$$

kde pri úprave sme využili postupne definíciu procesu (3.5), vlastnosti (3.2), (3.3), (3.4) a v druhej rovnosti sme navyš predpokladali nulovú kovarianciu  $\epsilon_t$  a  $\alpha \circ X_{t-1}$ , ktorú teraz dokážeme. Najskôr zapíšeme  $\alpha \circ X_{t-1}$  v MA reprezentácii. Vzhľadom k tomu, že  $\alpha^0 \circ \epsilon_t = \epsilon_t$  a  $X_t = \alpha \circ X_{t-1} + \epsilon_t$ , dostávame z (3.6)  $\alpha \circ X_{t-1} = \sum_{i=1}^{\infty} \alpha^i \circ \epsilon_{t-i}$ . Ďalej máme

$$\text{Cov}(\alpha \circ X_{t-1}, \epsilon_t) = \text{Cov}\left(\sum_{i=1}^{\infty} \alpha^i \circ \epsilon_{t-i}, \epsilon_t\right) = \sum_{i=1}^{\infty} \text{Cov}(\alpha^i \circ \epsilon_{t-i}, \epsilon_t).$$

Pre každé  $i \in \mathbb{N}$  platí

$$\begin{aligned} \text{Cov}(\alpha^i \circ \epsilon_{t-i}, \epsilon_t) &= \mathbf{E} \left[ (\alpha^i \circ \epsilon_{t-i}) \epsilon_t \right] - \mathbf{E} (\alpha^i \circ \epsilon_{t-i}) \mathbf{E} \epsilon_t \\ &= \mathbf{E} \left[ \mathbf{E} \left( \epsilon_t \sum_{k=1}^{\epsilon_{t-i}} Y_k \middle| \epsilon_{t-i}, \epsilon_t \right) \right] - \mathbf{E} \left[ \mathbf{E} \left( \sum_{k=1}^{\epsilon_{t-i}} Y_k \middle| \epsilon_{t-i} \right) \right] \mathbf{E} \epsilon_t \\ &= \mathbf{E} [\alpha^i \epsilon_{t-i} \epsilon_t] - \mathbf{E} [\alpha^i \epsilon_{t-i}] \mathbf{E} \epsilon_t \\ &= \alpha^i \text{Cov}(\epsilon_{t-i}, \epsilon_t) = 0 \end{aligned}$$

vzhľadom k nezávislosti  $\epsilon_t$  a  $\epsilon_{t-i}$ .

## Stacionarita

Nutnou podmienkou slabej stacionarity je konštantnosť strednej hodnoty a rozptylu procesu. K tomu, aby bol proces INAR(1) slabo stacionárny, musí pre strednú hodnotu platiť  $\mathbf{E} X_t = m$ ,  $\forall t \in \{0, 1, \dots\}$  a úpravou (3.7) získame

$$\begin{aligned} m &= \alpha^t m + \mu \sum_{j=0}^{t-1} \alpha^j \\ m(1 - \alpha^t) &= \mu \frac{1 - \alpha^t}{1 - \alpha} \\ m &= \frac{\mu}{1 - \alpha}. \end{aligned} \tag{3.9}$$

To znamená, že každá náhodná veličina zo slabo stacionárneho procesu INAR(1) musí mať strednú hodnotu rovnú  $\frac{\mu}{1 - \alpha}$ , a to nastane práve vtedy, keď je stredná hodnota počiatocnej hodnoty procesu  $\mathbf{E} X_0 = \frac{\mu}{1 - \alpha}$ . Pre rozptyl musí rovnako platiť  $\forall t : \text{var} X_t = v$ . Dosadením konštantnej strednej hodnoty (3.9) do (3.8) a úpravou dostávame

$$\begin{aligned} v &= \alpha^{2t} v + \mu \sum_{j=1}^t \alpha^{2j-1} + \sigma^2 \sum_{j=1}^t \alpha^{2j-2} \\ v(1 - \alpha^{2t}) &= \mu \alpha \frac{1 - \alpha^{2t}}{1 - \alpha^2} + \sigma^2 \frac{1 - \alpha^{2t}}{1 - \alpha^2} \\ v &= \frac{\alpha \mu + \sigma^2}{1 - \alpha^2}. \end{aligned}$$

Každá náhodná veličina zo slabo stacionárneho procesu INAR(1) musí teda mať rozptyl rovný  $\frac{\alpha \mu + \sigma^2}{1 - \alpha^2}$ , a to nastane práve vtedy, keď je rozptyl počiatocnej hodnoty procesu  $\text{var} X_0 = \frac{\alpha \mu + \sigma^2}{1 - \alpha^2}$ .

**Tvrdenie 2.** *Pre slabo stacionárny proces INAR(1) platí*

(i)

$$E X_t = \frac{\mu}{1 - \alpha}, \quad t = 0, 1, 2, \dots$$

(ii)

$$\text{var } X_t = \frac{\alpha\mu + \sigma^2}{1 - \alpha^2}, \quad t = 0, 1, 2, \dots$$

(iii)

$$\gamma_k = \alpha^k \text{var } X_0, \quad k = 0, 1, 2, \dots$$

(iv)

$$\rho_k = \alpha^k, \quad k = 0, 1, 2, \dots$$

*Dôkaz.*

(i) a (ii) Vyplývajú priamo z definície slabej stacionarity a nutnej podmienky pre slabú stacionaritu procesu INAR(1).

(iii) Dosadením  $k$  krát za  $X_t$  podľa (3.5) dostaneme

$$\begin{aligned} \text{Cov}(X_t, X_{t-k}) &= \text{Cov}\left(\alpha^k \circ X_{t-k} + \sum_{i=0}^{k-1} \alpha^i \circ \epsilon_{t-i}, X_{t-k}\right) \\ &= \text{Cov}(\alpha^k \circ X_{t-k}, X_{t-k}) + \sum_{i=0}^{k-1} \text{Cov}(\alpha^i \circ \epsilon_{t-i}, X_{t-k}). \end{aligned}$$

Oba sčítance ďalej upravíme samostatne využitím definície 7 a (3.3)

$$\begin{aligned} \text{Cov}(\alpha \circ X_{t-k}, X_{t-k}) &= E(\alpha^k \circ X_{t-k} X_{t-k}) - E \alpha^k \circ X_{t-k} E X_{t-k} \\ &= E \left[ E(\alpha^k \circ X_{t-k} X_{t-k}) | X_{t-k} \right] - \alpha^k (E X_{t-k})^2 \\ &= E \left[ \alpha^k X_{t-k}^2 \right] - \alpha^k (E X_{t-k})^2 \\ &= \alpha^k \text{var } X_{t-k} = \alpha^k \text{var } X_0 \end{aligned}$$

vzhľadom k (ii).

V druhom sčítanci najprv zapíšeme  $X_{t-k}$  v MA reprezentácii

$$\begin{aligned} \sum_{i=0}^{k-1} \text{Cov}(\alpha^i \circ \epsilon_{t-i}, X_{t-k}) &= \sum_{i=0}^{k-1} \text{Cov}\left(\alpha^i \circ \epsilon_{t-i}, \sum_{j=0}^{\infty} \alpha^j \circ \epsilon_{t-k-j}\right) \\ &= \sum_{i=0}^{k-1} \sum_{j=0}^{\infty} \text{Cov}(\alpha^i \circ \epsilon_{t-i}, \alpha^j \circ \epsilon_{t-k-j}) \end{aligned}$$

a pre každé  $i \in \{0, \dots, k-1\}$  a  $j \in \{0, 1, \dots\}$  platí podľa definície 6 a (3.3)

$$\begin{aligned}
\text{Cov}(\alpha^i \circ \epsilon_{t-i}, \alpha^j \circ \epsilon_{t-k-j}) &= \mathbf{E} [(\alpha^i \circ \epsilon_{t-i})(\alpha^j \circ \epsilon_{t-k-j})] \\
&\quad - \mathbf{E} \alpha^i \circ \epsilon_{t-i} \mathbf{E} \alpha^j \circ \epsilon_{t-k-j} \\
&= \mathbf{E} [\mathbf{E} (\alpha^i \circ \epsilon_{t-i})(\alpha^j \circ \epsilon_{t-k-j}) | \epsilon_{t-i}, \epsilon_{t-k-j}] \\
&\quad - \mathbf{E} [\mathbf{E} \alpha^i \circ \epsilon_{t-i} | \epsilon_{t-i}] \mathbf{E} [\mathbf{E} \alpha^j \circ \epsilon_{t-k-j} | \epsilon_{t-k-j}] \\
&= \mathbf{E} \left[ \mathbf{E} \sum_{m=1}^{\epsilon_{t-i}} Y_m \sum_{n=1}^{\epsilon_{t-k-j}} Z_n \middle| \epsilon_{t-i}, \epsilon_{t-k-j} \right] \\
&\quad - \alpha^i \mathbf{E} \epsilon_{t-i} \alpha^j \mathbf{E} \epsilon_{t-k-j} \\
&= \mathbf{E} \left[ \left( \mathbf{E} \sum_{m=1}^{\epsilon_{t-i}} Y_m \middle| \epsilon_{t-i} \right) \left( \mathbf{E} \sum_{n=1}^{\epsilon_{t-k-j}} Z_n \middle| \epsilon_{t-k-j} \right) \right] \\
&\quad - \alpha^i \mathbf{E} \epsilon_{t-i} \alpha^j \mathbf{E} \epsilon_{t-k-j} \\
&= \mathbf{E} [\alpha^i \epsilon_{t-i} \alpha^j \epsilon_{t-k-j}] - \alpha^{i+j} \mathbf{E} \epsilon_{t-i} \mathbf{E} \epsilon_{t-k-j} \\
&= \alpha^{i+j} \text{Cov}(\epsilon_{t-i}, \epsilon_{t-k-j}) = 0,
\end{aligned}$$

vďaka vzájomnej nezávislosti  $\epsilon_{t-i}$ ,  $\epsilon_{t-k-j}$  a postupností  $\{Y_m\}$ ,  $\{Z_n\}$ .

- (iv) Hodnotu autokorelačnej funkcie v bode  $k$  dostaneme hneď po dosadení výsledku bodu (iii) z tohto tvrdenia do (1.2).

□

*Poznámka.* Z tvaru autokovariančnej funkcie (iii) ďalej vyplýva, že nutná podmienka slabej stacionarity procesu INAR(1),  $\mathbf{E} X_0 = \frac{\mu}{1-\alpha}$  a  $\text{var} X_0 = \frac{\alpha\mu + \sigma^2}{1-\alpha^2}$ , je zároveň aj postačujúcou podmienkou.

### 3.3 Poissonovský INAR(1)

Ďalej sa budeme venovať iba špeciálnemu prípadu procesu INAR(1), v ktorom uvažujeme  $\{\epsilon_t\}$  s Poissonovým rozdelením s parametrom  $\lambda$ . Z definície 7, ako dôsledok tejto voľby, vyplýva, že pri známom  $X_{t-1} = x$  je náhodná veličina  $X_t$  súčtom náhodnej veličiny  $\alpha \circ x$  s binomickým rozdelením s parametrami  $x$ ,  $\alpha$  a náhodnej veličiny  $\epsilon_t$  s Poissonovým rozdelením s parametrom  $\lambda$ . Podmienené rozdelenie  $X_t$  pri známom  $X_{t-1} = x$  vieme spočítať pomocou vzorca

$$\mathbf{P}(X_t = y | X_{t-1} = x) = \sum_{k=0}^{\min(y,x)} \mathbf{P}(\alpha \circ x = k, \epsilon_t = y - k).$$

Horná medza sumy je rovná  $\min(y,x)$ , pretože nosičom veličiny  $\alpha \circ x$  je množina  $\{0, \dots, x\}$  a pravdepodobnosť nadobudnutia hodnoty mimo nosiča je nulová. Ďalej môžeme vďaka nezávislosti  $\alpha \circ x$  a  $\epsilon_t$  písať

$$\mathbf{P}(X_t = y | X_{t-1} = x) = \exp(-\lambda) \sum_{k=0}^{\min(y,x)} \binom{x}{k} \frac{\alpha^k (1-\alpha)^{x-k} \lambda^{y-k}}{(y-k)!}, \quad y \in \mathbb{N}_0. \quad (3.10)$$

Z (3.7), (3.3) a (3.5) vieme vyjadriť podmienenú strednú hodnotu  $X_t$  pri pevnom  $X_{t-1}$

$$\mathbf{E}(X_t | X_{t-1}) = \alpha X_{t-1} + \lambda,$$

a taktiež podmienený rozptyl

$$\text{var}(X_t|X_{t-1}) = \alpha(1 - \alpha)X_{t-1} + \lambda,$$

pričom pre platnosť tejto rovnosti musíme ešte dokázať nulovú podmienenú kovarianciu  $\alpha \circ X_{t-1}$  a  $\epsilon_t$  pri pevnom  $X_{t-1}$

$$\begin{aligned} \text{Cov}(\alpha \circ X_{t-1}, \epsilon_t | X_{t-1}) &= \mathbf{E} [(\alpha \circ X_{t-1})\epsilon_t | X_{t-1}] - \mathbf{E}(\alpha \circ X_{t-1} | X_{t-1}) \mathbf{E}(\epsilon_t | X_{t-1}) \\ &= \mathbf{E} \left( \epsilon_t \sum_{i=1}^{X_{t-1}} Y_i \middle| X_{t-1} \right) - \alpha X_{t-1} \mathbf{E} \epsilon_t \\ &= \mathbf{E} [\epsilon_t (Y_1 + \dots + Y_{X_{t-1}})] - \alpha X_{t-1} \mathbf{E} \epsilon_t. \end{aligned}$$

V uvedených rovnostiach sme opakovane využili nezávislosť  $\epsilon_t$  a  $X_{t-1}$ , ktorá vyplýva z toho, že  $X_{t-1}$  je funkciou náhodných veličín nezávislých na  $\epsilon_t$ . Keďže  $\epsilon_t$  je taktiež nezávislé na  $Y_1, \dots, Y_{X_{t-1}}$  platí, že  $\mathbf{E} \epsilon_t Y_i = \mathbf{E} \epsilon_t \mathbf{E} Y_i$  a prvý člen z poslednej rovnosti môžeme upraviť do tvaru

$$\mathbf{E} [\epsilon_t (Y_1 + \dots + Y_{X_{t-1}})] = \mathbf{E} \epsilon_t (\mathbf{E} Y_1 + \dots + \mathbf{E} Y_{X_{t-1}}) = \alpha X_{t-1} \mathbf{E} \epsilon_t,$$

čím je dôkaz hotový.

## Odhad momentovou metódou

Jedná sa o najjednoduchší odhad neznámeho parametra  $\alpha$ , ktorý spočítame analogicky ako v procese AR(1) nahradením autokorelačnej funkcie  $\rho_k$  spočítanej v bode  $k = 1$  jej odhadom (1.3)

$$\hat{\alpha} = \frac{\sum_{t=1}^{n-1} (X_t - \bar{X}_n)(X_{t+1} - \bar{X}_n)}{\sum_{t=1}^n (X_t - \bar{X}_n)^2}.$$

Pomocou odhadu  $\hat{\alpha}$  vieme taktiež spočítať odhad parametra  $\lambda$ , ktorý charakterizuje Poissonovo rozdelenie náhodných veličín  $\{\epsilon_t\}$ . Označme  $\hat{\epsilon}_t = X_t - \hat{\alpha}X_{t-1}$  pre  $t = 2, \dots, n$ . Keďže parameter  $\lambda$  je strednou hodnotou Poissonovho rozdelenia, odhad spočítame výberovým priemerom

$$\hat{\lambda} = \frac{1}{n-1} \sum_{t=2}^n \hat{\epsilon}_t.$$

## Odhad metódou najmenších štvorcov

Odhad parametrov pre proces INAR(1) je komplikovanejší ako v prípade klasického AR(1) procesu, pretože pri známom  $X_{t-1}$  a  $\epsilon_t$  je  $X_t$  stále náhodná veličina. Označme preto podmienenú strednú hodnotu  $X_t$  pri pevnom  $X_{t-1}$

$$\mathbf{E}(X_t | X_{t-1}) = \alpha X_{t-1} + \lambda \equiv g(\alpha, \lambda, X_{t-1}).$$

Výpočet odhadu je založený na minimalizácii sumy štvorcov odchýliek pozorovaní od podmienenej strednej hodnoty. Chceme teda nájsť bod  $(\alpha, \lambda)^T$ , v ktorom je minimalizovaný výraz

$$K_n(\alpha, \lambda) = \sum_{t=2}^n [X_t - g(\alpha, \lambda, X_{t-1})]^2.$$

K nájdeniu argumentu minima potrebujeme spočítať parciálne derivácie funkcie  $K_n(\alpha, \lambda)$  podľa oboch parametrov

$$\begin{aligned}\frac{\partial K_n(\alpha, \lambda)}{\partial \alpha} &= -2 \sum_{t=2}^n (X_t - \alpha X_{t-1} - \lambda) X_{t-1} \\ \frac{\partial K_n(\alpha, \lambda)}{\partial \lambda} &= 2 \sum_{t=2}^n (X_t - \alpha X_{t-1} - \lambda).\end{aligned}$$

Obe parciálne derivácie položíme rovné nule a z druhej rovnice vyjadríme  $\lambda$ , čo môžeme dosadiť do prvej rovnice a dostaneme vyjadrenie pre odhad  $\hat{\alpha}$ ,  $\hat{\lambda}$

$$\begin{aligned}\hat{\alpha} &= \frac{\sum_{t=2}^n X_t X_{t-1} - \frac{1}{n-1} \sum_{t=2}^n X_t \sum_{t=2}^n X_{t-1}}{\sum_{t=2}^n X_{t-1}^2 - \frac{1}{n-1} (\sum_{t=2}^n X_{t-1})^2} \\ \hat{\lambda} &= \frac{1}{n-1} \left( \sum_{t=2}^n X_t - \hat{\alpha} \sum_{t=2}^n X_{t-1} \right).\end{aligned}$$

## Maximálne vierohodný odhad

Rovnako ako pri procese AR(1) potrebujeme k zostrojeniu vierohodnostnej funkcie poznať združenú hustotu pozorovaní  $X_1, \dots, X_n$ . Tá má tvar (2.5) a vyžaduje znalosť hustoty prvého pozorovania  $X_1$  a podmienené hustoty zvyšných pozorovaní. Za predpokladu, že náhodné veličiny z postupnosti  $\{\epsilon_t\}$  majú Poissonovo rozdelenie, má  $X_1$  taktiež Poissonovo rozdelenie s parametrom  $\lambda/(1-\alpha)$  a jeho hustota je

$$f_{X_1}(x_1) = \exp\left(\frac{-\lambda}{1-\alpha}\right) \frac{[\lambda/(1-\alpha)]^{x_1}}{x_1!}, \quad \text{pre } x_1 \in \mathbb{N}_0.$$

Pre náhodné veličiny z procesu INAR(1) platí, že  $f_{X_t|X_{t-1}, \dots, X_1} = f_{X_t|X_{t-1}}$  a podmienené hustoty potrebné k zostaveniu združenej hustoty majú tvar (3.10). Zo združenej hustoty dostávame vierohodnostnú funkciu v tvare

$$L(\alpha, \lambda) = \frac{[\lambda/(1-\alpha)]^{X_1}}{X_1!} \exp[-\lambda/(1-\alpha)] \prod_{t=2}^n P_t(X_t),$$

kde

$$P_t(y) = \exp(-\lambda) \sum_{k=0}^{\min(X_t, X_{t-1})} \binom{X_{t-1}}{k} \frac{\alpha^k (1-\alpha)^{X_{t-1}-k} \lambda^{y-k}}{(y-k)!} \quad \text{pre } t = 2, \dots, n.$$

K nájdeniu argumentu maxima vierohodnostnej funkcie potrebujeme spočítať parciálne derivácie logaritmickej vierohodnosti  $l(\alpha, \lambda) = \log[L(\alpha, \lambda)]$  podľa  $\alpha$  a  $\lambda$

$$\begin{aligned}\frac{\partial l(\alpha, \lambda)}{\partial \alpha} &= S_\alpha + \frac{1}{1-\alpha} \left( X_1 - \frac{\lambda}{1-\alpha} \right) \\ \frac{\partial l(\alpha, \lambda)}{\partial \lambda} &= S_\lambda + \frac{X_1}{\lambda} - \frac{1}{1-\alpha},\end{aligned}$$

kde

$$S_\alpha = \sum_{t=2}^n \frac{\partial \log P_t(X_t)}{\partial \alpha} = \sum_{t=2}^n \frac{X_t - \alpha X_{t-1} \frac{P_t(X_t - 1)}{P_t(X_t)}}{\alpha(1 - \alpha)}$$
$$S_\lambda = \sum_{t=2}^n \frac{\partial \log P_t(X_t)}{\partial \lambda} = \sum_{t=2}^n \frac{P_t(X_t - 1)}{P_t(X_t)} - n + 1.$$

Tieto parciálne derivácie položíme rovné nule, čím dostaneme dve rovnice s dvomi neznámymi parametrami. Keďže tieto rovnice nemajú explicitné riešenie, je potrebné ich v praxi riešiť s konkrétnymi pozorovaniami pomocou softwaru.

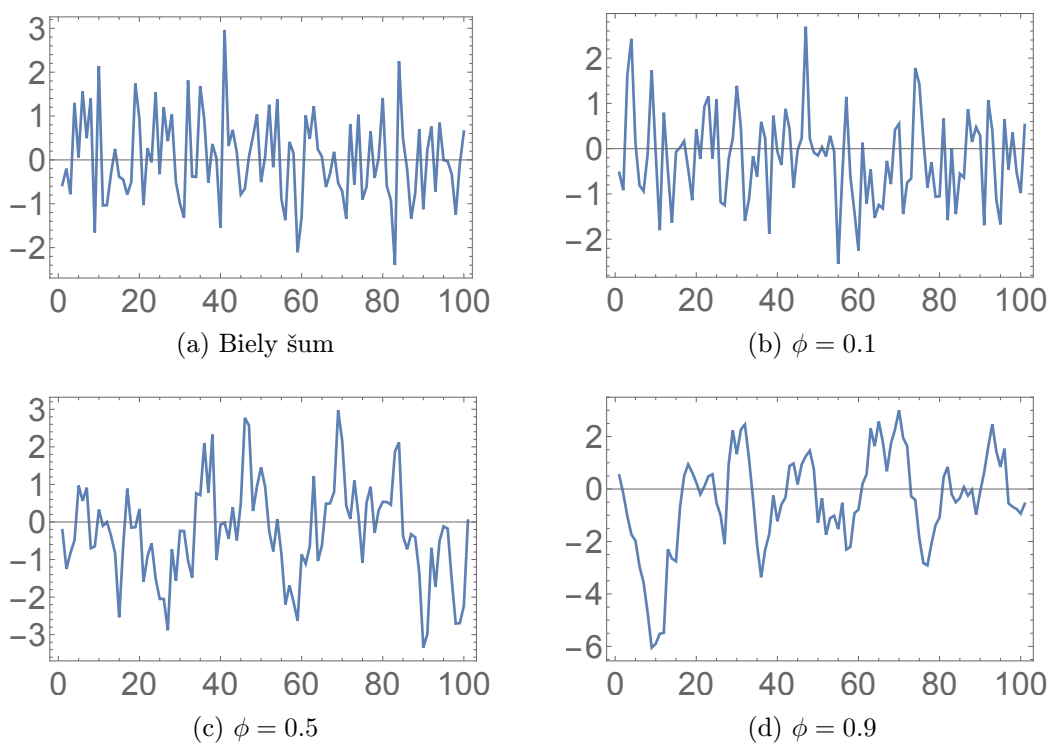
## 4. Simulačná štúdia

V tejto kapitole budeme skúmať správanie oboch modelov pri rôznych hodnotách regresného parametra a taktiež porovnáme presnosť niektorých odhadových procedúr popísaných v predchádzajúcich kapitolách.

### 4.1 Proces AR(1)

#### Správanie modelu

Skúmanie správania procesu má zmysel len pre hodnoty parametra, pre ktoré je proces slabo stacionárny, to znamená pre  $|\phi| < 1$ . Pre  $|\phi| \geq 1$  potom stredná hodnota procesu buď nemá limitu, alebo konverguje k nekonečnu. Pri generovaní všetkých nasledujúcich realizácií procesov sme zvolili pevnú strednú hodnotu  $\mu = 0$  a rozptyl gaussovského bieleho šumu  $\sigma^2 = 1$ . Za počiatočnú hodnotu procesu sme zvolili aritmetický priemer 20 hodnôt procesu s rovnakými parametrami a počiatočnou hodnotou 0. Pre ilustráciu je na obrázku 4.1 v časti (a) realizácia

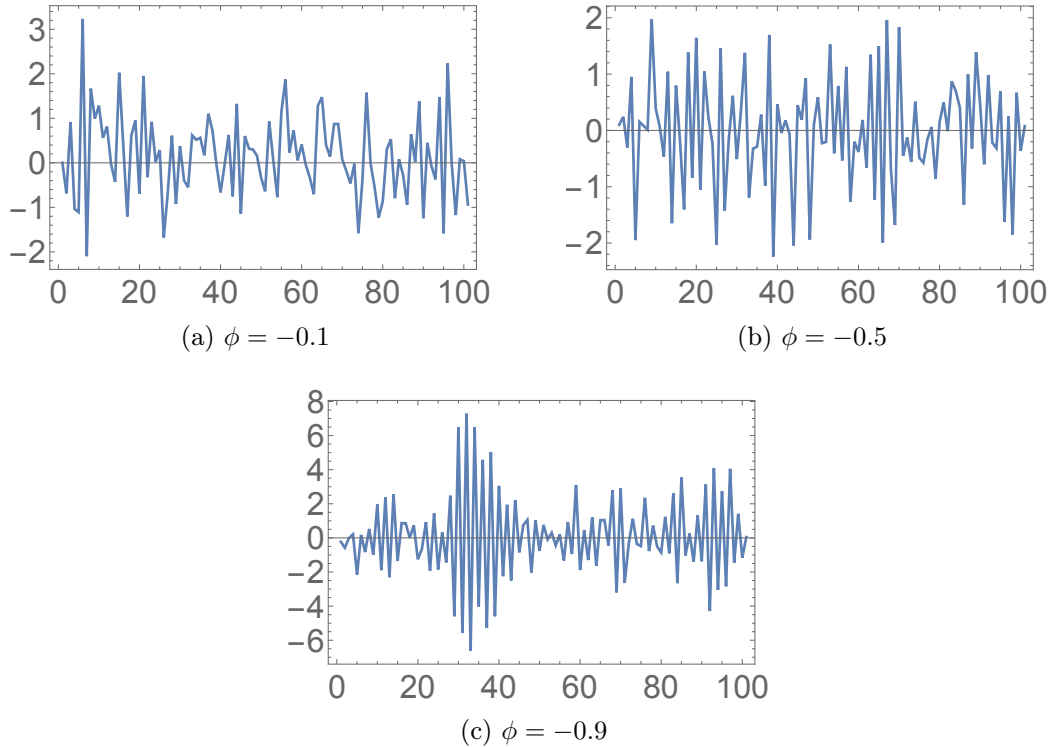


Obr. 4.1: Slabo stacionárny AR(1) proces s nezáporným parametrom

gaussovského bieleho šumu s nulovou strednou hodnotou a jednotkovým rozptylom. Môžeme si všimnúť, že proces AR(1) je pre malé hodnoty parametra  $\phi$  bielemu šumu veľmi podobný a závislosť na predchádzajúcej hodnote má zanedbateľný vplyv. Pre narastajúce hodnoty  $\phi$  je potom možné pozorovať vyššiu mieru korelácie medzi susednými hodnotami a vytrácajúci sa charakter bieleho šumu.

Pre záporné hodnoty parametra  $\phi$  je z definície zrejmé, že s rastúcou absolútnou





Obr. 4.2: Slabo stacionárny AR(1) proces so záporným parametrom

hodnotou parametra bude klesať vplyv bieleho šumu a znamienko hodnôt procesu bude často alternovať, ako je možné pozorovať na obrázku 4.2.

## Odhad parametrov

Pre proces AR(1) budeme skúmať a porovnávať rýchlosť konverencie odhadov  $\hat{\mu}$ ,  $\hat{\phi}$  a  $\hat{\sigma}$  spočítaných momentovou metódou a metódou maximálnej vierohodnosti a taktiež jej závislosť na skutočnej hodnote parametra  $\phi$ . Metóda najmenších štvorcov poskytuje odhady, ktorých štruktúra je veľmi podobná momentovým odhadom a preto ju v tejto práci neskúmame. Skutočná hodnota strednej hodnoty procesu bola vygenerovaná z intervalu  $(-100,100)$  s výsledkom 6.8877 a hodnotu smerodajnej odchýlky bieleho šumu  $\sigma$  sme vygenerovali z intervalu  $(0,10)$  s výsledkom 4.7772. Za účelom skúmania závislosti rýchlosti konverencie odhadov na  $\phi$  sme jeho hodnotu generovali z 8 disjunktných podintervalov intervalu  $(-1,1)$  dĺžky  $1/4$ . Ako počiatočnú hodnotu všetkých generovaných realizácií sme použili strednú hodnotu procesu. V tabuľkách z Prílohy 2 sú uvedené hodnoty aritmetických priemerov a smerodajných odchýliek odhadov parametrov  $\mu$ ,  $\phi$  a  $\sigma$  spočítané z  $N = 100$  realizácií pre rôzne, ale pevné dĺžky realizácií  $n$ .

Z tabuliek 4.1, 4.2 a 4.3 je vidieť, že dĺžka radu má vplyv na priblíženie priemernej hodnoty odhadu k odhadovanému parametru a na veľkosť smerodajnej odchýlky. Momentová metóda a metóda maximálnej vierohodnosti nevykazujú zásadné rozdiely v rýchlosti konverencie. V oboch prípadoch je priemer uspokojivo blízko odhadovanej hodnote už pre dĺžku radu  $n = 50$ . Smerodajná odchýlka sa výrazne znižuje v prípade stoviek pozorovaní ( $n = 500$ ).

Výrazný vplyv hodnoty  $\phi$  na presnosť odhadov  $\phi$  a  $\sigma$  sme v dátach nespozorovali.

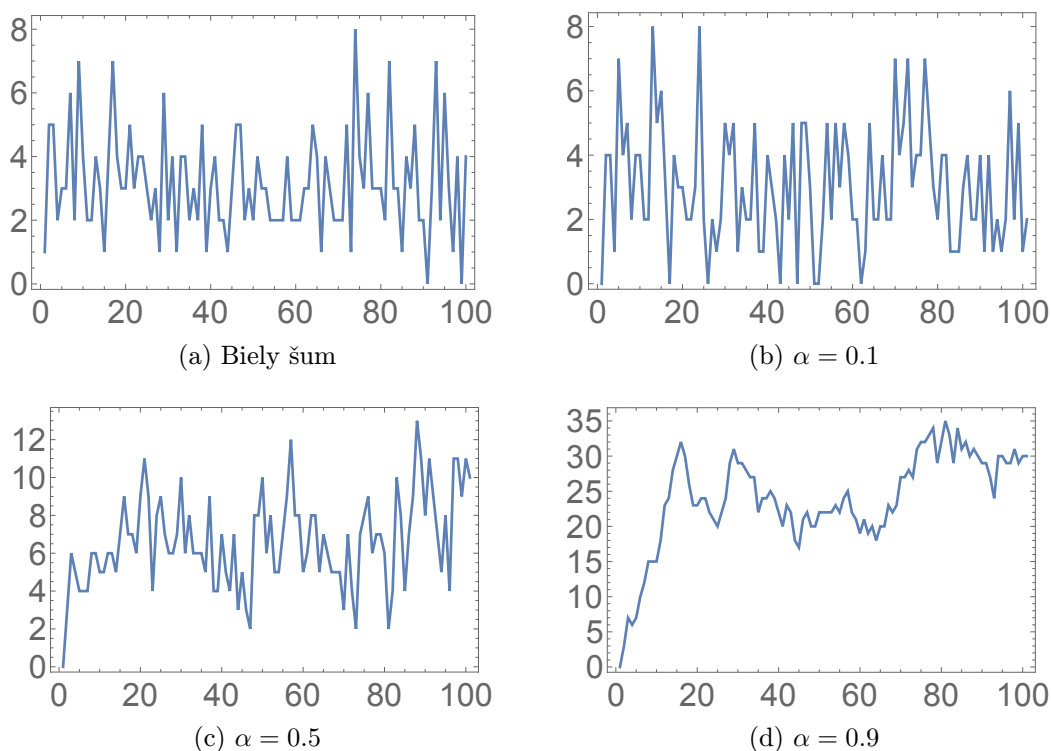
Pri odhadoch strednej hodnoty procesu je však možné pozorovať rastúci trend smerodajnej odchýlky odhadov pre rastúcu hodnotu parametra  $\phi$  u oboch odhadových techník. Je dôležité dodať, že odhad metódou maximálnej vierohodnosti bol počítaný numericky v software Mathematica a výsledok preto nemusí byť presný.

## 4.2 Proces INAR(1)

V nasledujúcich častiach budeme pracovať s INAR(1) procesom, pre ktorý má  $\{\epsilon_t\}$  Poissonovo rozdelenie s parametrom  $\lambda$ .

### Správanie modelu

Správanie procesu INAR(1) v závislosti na parametre  $\alpha \in (0,1)$  je podobné ako u procesu AR(1). Pre nízke hodnoty parametra je správanie procesu podobné s tým poissonovského bieleho šumu a s rastúcou hodnotou  $\alpha$  sa zvyšuje závislosť hodnoty procesu na minulej hodnote. Na rozdiel od procesu AR(1) ovplyvňuje hodnota parametra  $\alpha$  aj strednú hodnotu celočíselného autoregresného procesu, čo je zrejmé aj zo vzorca (3.9) strednej hodnoty INAR(1).



Obr. 4.3: Proces INAR(1)

### Odhad parametrov

Analogicky ako u procesu AR(1), budeme aj teraz skúmať závislosť rýchlosti konvergencie odhadov parametrov na hodnotu parametra  $\alpha$ . Porovnáваме

momentovú metódu a metódu najmenších štvorcov. Metóda maximálnej viero-  
hodnosti nie je do simulačnej štúdie zahrnutá z dôvodu výpočtovej náročnosti.  
Skutočnú hodnotu parametra poissonovského bieleho šumu procesu  $\lambda$  sme vyge-  
nerovali z množiny  $\{1, \dots, 10\}$  s výsledkom 8 a hodnoty parametra  $\alpha$  sme genero-  
vali z disjunktných podintervalov intervalu  $(0,1)$  s dĺžkou  $1/4$ . Za počiatočný bod  
generovaných realizácií sme zvolili strednú hodnotu procesu (3.9) zaokrúhlenú na  
celé číslo. V tabuľkách v Prílohe 3 sú uvedené výberové priemery a výberové smer-  
rodajné odchýlky odhadov parametrov  $\alpha$  a  $\lambda$  spočítaných z  $N = 100$  realizácií s  
pevnou dĺžkou  $n$ . Z dát v tabuľkách (4.4) a (4.5) je opäť zrejмый vplyv dĺžky radu  
na priblíženie priemernej hodnoty odhadu k skutočnej hodnote parametra  
a na veľkosť výberovej smerodajnej odchýlky. Pri priemeroch odhadov parametra  
 $\alpha$  je s narastajúcou hodnotou  $\alpha$  možné pozorovať zväčšujúce sa rozdiely medzi  
priermi odhadov spočítaných momentovou metódou a metódou najmenších  
štvorcov. Tento rozdiel je najvýraznejší pre malú dĺžku radu ( $n = 20, 50$ ) a väčšiu  
presnosť vykazujú priemery odhadov spočítané metódou najmenších štvorcov,  
ale pre väčšie dĺžky sa postupne tento rozdiel vytráca. Vo všetkých prípadoch je  
však priemer odhadov uspokojivo blízko skutočnej hodnote s dostatočne malou  
výberovou smerodajnou odchýlkou pre rad dĺžky  $n = 100$ .

Pre priemery odhadov parametra  $\lambda$  môžeme pozorovať rovnaký trend rozdie-  
lov medzi odhadovými metódami, aký sme popísali v predchádzajúcom odstavci.  
Naviac môžeme s narastajúcou hodnotou  $\alpha$  pozorovať spomalenie konvergencie  
priemerov odhadov a nárast výberovej smerodajnej odchýlky pre obe odhadové  
metódy. Pre  $\alpha$  blízke 1 je tento fenomén tak výrazný, že výsledky sú pre dĺžky  
 $n = 20, 50, 100$  prakticky nepoužiteľné. Zároveň je dôležité si uvedomiť, že pre  
 $\alpha = 1$  nemá vzhľadom k (3.9) proces  $X_t$  konečnú strednú hodnotu.

# Záver

V tejto práci sa nám podarilo vytvoriť porovnanie medzi dvoma diskutovanými modelmi. Pre INAR(1) sme navyše spracovali dôkazy dôležitých vlastností, ktoré boli v použitej literatúre vynechané. V simulačnej štúdii sa preukázala uspokojivá rýchlosť konvergencie vybraných odhadových procedúr pre rady dĺžky aspoň 50, čo je doporučená minimálna dĺžka pre analýzu autoregresných časových radov napríklad podľa Cipra (2008), str. 330.

# Zoznam použitej literatúry

AL-OSH, M. a ALZAID, A. A. (1987). First-order integer-valued autoregressive (INAR (1)) process. *Journal of Time Series Analysis*, **8**(3), 261–275.

CIPRA, T. (2008). *Finanční ekonometrie*. Ekopress, Praha.

KEDEM, B; FOKIANOS, K. (2002). *Regression models for time series analysis*. Wiley-Interscience, New York.

LACHOUT, P. a PRÁŠKOVÁ, Z. (2012). *Základy náhodných procesů I*. Matfyzpress, Praha.

WEI, W. W. (1990). *Time series analysis: univariate and multivariate methods*. Pearson Addison Wesley, New York.

# Zoznam obrázkov

4.1	Slabo stacionárny AR(1) proces s nezáporným parametrom . . . .	20
4.2	Slabo stacionárny AR(1) proces so záporným parametrom . . . .	21
4.3	Proces INAR(1) . . . . .	22

# Zoznam tabuliek

4.1	Odhady strednej hodnoty $\mu$ . . . . .	29
4.2	Odhady parametra $\phi$ . . . . .	30
4.3	Odhady smerodatnej odchýlky $\sigma$ . . . . .	31
4.4	Odhady parametra $\alpha$ . . . . .	32
4.5	Odhady parametra $\lambda$ . . . . .	33

# Prílohy

## Príloha 1

### Vytvárajúca funkcia postupnosti pravdepodobností

**Definícia 8.** Pre diskrétnu náhodnú veličinu  $X$  s pravdepodobnostným rozdelením  $p_0, p_1, p_2, \dots$ , kde  $p_i = P(X = i)$ , nazveme funkciu  $P_X(s) = \sum_{k=0}^{\infty} p_k s^k$  vytvárajúcou funkciou postupnosti pravdepodobností náhodnej veličiny  $X$ .

*Poznámka.* Pre  $s$  také, že suma  $\sum_{k=0}^{\infty} p_k s^k$  konverguje platí vzťah  $P_X(s) = E s^X$ .

**Tvrdenie 3.** Nech  $\{X_k\}_{k=1}^{\infty}$  je postupnosť nezávislých rovnako rozdelených diskrétnych náhodných veličín, ktorá je nezávislá s diskrétnou náhodnou veličinou  $N$ . Potom pre veličinu  $S_N = \sum_{k=1}^N X_k$  a pre  $s$  také, že  $E|s|^{S_N} < \infty$  platí

$$E s^{S_N} = E(E s^{X_1})^N,$$

alebo ekvivalentne zapísané pomocou vytvárajúcich funkcií

$$P_{S_N}(s) = P_N(P_{X_1}(s)).$$

*Dôkaz.*

$$E s^{S_N} = E s^{\sum_{k=1}^N X_k} = E \left[ E(s^{\sum_{k=1}^N X_k} | N) \right] = E (E s^{X_1})^N$$

□

Vlastnosť, ktorá hovorí, že vytvárajúca funkcia postupnosti pravdepodobností jednoznačne určuje pravdepodobnostné rozdelenie náhodnej veličiny, vyplýva z toho, že pravdepodobnosti  $p_0, p_1, p_2, \dots$  vieme vyjadriť derivovaním funkcie  $P_X(s)$  ako

$$p_n = P(X = n) = \frac{1}{n!} P_X^{(n)}(0).$$

## Príloha 2

V tejto prílohe sa nachádzajú tabuľky obsahujúce výsledky simulačnej štúdie pre proces AR(1) popísanej v kapitole 4.



$\phi$	$n$	$\hat{\mu}^{MM}$		$\hat{\mu}^{MLE}$	
		Priemer	Smer. odchýlka	Priemer	Smer. odchýlka
0.0818	20	6.8233	1.1748	6.8127	1.1729
	50	6.828	0.6484	6.8282	0.6451
	100	6.8951	0.4671	6.8952	0.4669
	500	6.9404	0.2217	6.9404	0.2216
	1000	6.8912	0.175	6.8912	0.175
0.4523	20	6.8618	1.4933	6.8588	1.4362
	50	6.7898	1.1354	6.7968	1.122
	100	6.7552	0.7809	6.758	0.7658
	500	6.9592	0.3537	6.9603	0.353
	1000	6.8889	0.272	6.8889	0.2713
0.6573	20	6.644	2.9226	6.6738	2.7306
	50	6.8136	1.7364	6.8298	1.697
	100	6.9032	0.4307	6.9036	0.4289
	500	6.9414	0.7062	6.9426	0.7022
	1000	6.8682	0.4328	6.8685	0.431
0.9116	20	6.4617	6.6873	6.6109	6.0877
	50	6.6955	6.9552	6.6768	5.9708
	100	6.2094	5.0757	6.3097	4.7276
	500	6.9686	2.4215	6.9799	2.3685
	1000	7.017	1.6365	7.0289	1.6231
-0.1075	20	6.7618	1.0009	6.7573	1.0022
	50	6.8587	0.5964	6.8591	0.5979
	100	6.8369	0.3917	6.837	0.3916
	500	6.9122	0.2048	6.9124	0.2051
	1000	6.8864	0.1473	6.8864	0.1474
-0.3442	20	6.9341	0.7384	6.9342	0.7479
	50	6.9396	0.4592	6.9410	0.4622
	100	6.8555	0.3487	6.8555	0.3478
	500	6.8832	0.1761	6.8832	0.1761
	1000	6.9015	0.1104	6.9016	0.1106
-0.6443	20	6.9101	0.5303	6.9164	0.5439
	50	6.8916	0.3931	6.8893	0.395
	100	6.8705	0.3133	6.8677	0.3154
	500	6.8954	0.1218	6.895	0.1219
	1000	6.8744	0.0974	6.8744	0.0972
-0.9296	20	6.8545	0.5621	6.8454	0.4885
	50	6.8877	0.3612	6.9005	0.335
	100	6.8673	0.2667	6.8717	0.2587
	500	6.8901	0.1171	6.8905	0.1157
	1000	6.8854	0.082	6.885	0.0819

Tabuľka 4.1: Odhady strednej hodnoty  $\mu$

$\phi$	$n$	$\hat{\phi}^{MM}$		$\hat{\phi}^{MLE}$	
		Priemer	Smer. odchýlka	Priemer	Smer. odchýlka
0.0818	20	0.0083	0.2112	0.0068	0.2166
	50	0.0619	0.1246	0.0617	0.1237
	100	0.0514	0.0904	0.0515	0.0903
	500	0.0798	0.0481	0.0798	0.0481
	1000	0.0815	0.0306	0.0816	0.0306
0.4523	20	0.3126	0.2056	0.3163	0.206
	50	0.3995	0.1219	0.399	0.122
	100	0.432	0.0756	0.4334	0.0771
	500	0.4472	0.0392	0.4473	0.039
	1000	0.4478	0.0261	0.4477	0.0261
0.6573	20	0.4367	0.1962	0.447	0.1971
	50	0.5983	0.1079	0.598	0.1075
	100	0.6553	0.0227	0.6554	0.2286
	500	0.6486	0.0337	0.6484	0.3365
	1000	0.6531	0.0223	0.6531	0.0223
0.9116	20	0.6589	0.1696	0.7059	0.1738
	50	0.7914	0.1003	0.8075	0.0995
	100	0.8596	0.0585	0.8649	0.0564
	500	0.9049	0.0193	0.905	0.0194
	1000	0.9072	0.0134	0.9071	0.0134
-0.1075	20	-0.1587	0.2014	-0.1623	0.2028
	50	-0.1177	0.1432	-0.1192	0.1463
	100	-0.144	0.0867	-0.1439	0.0867
	500	-0.1176	0.044	-0.1176	0.0439
	1000	-0.1106	0.0308	-0.1106	0.0308
-0.3442	20	-0.3512	0.2001	-0.3510	0.1964
	50	-0.3359	0.1322	-0.3374	0.133
	100	-0.3219	0.1015	-0.3226	0.1017
	500	-0.3428	0.0432	-0.3429	0.0433
	1000	-0.3429	0.0433	-0.3454	0.0311
-0.6443	20	-0.5971	0.2018	-0.6056	0.2005
	50	-0.6256	0.1208	-0.625	0.1204
	100	-0.6281	0.085	-0.6298	0.084
	500	-0.641	0.037	-0.6411	0.0366
	1000	-0.6427	0.0235	-0.6427	0.0235
-0.9296	20	-0.7764	0.1578	-0.7964	0.1623
	50	-0.883	0.0697	-0.8835	0.0728
	100	-0.9022	0.0535	-0.9055	0.0524
	500	-0.9252	0.0163	-0.9253	0.0161
	1000	-0.9266	0.0125	-0.9267	0.0127

Tabuľka 4.2: Odhady parametra  $\phi$

$\phi$	$n$	$\hat{\sigma}^{MM}$		$\hat{\sigma}^{MLE}$	
		Priemer	Smer. odchýlka	Priemer	Smer. odchýlka
0.0818	20	4.2603	0.7788	4.253	0.7791
	50	4.6363	0.4548	4.6357	0.4546
	100	4.6968	0.3128	4.6966	0.3129
	500	4.7631	0.1494	4.7630	0.1494
	1000	4.7655	0.1233	4.7655	0.1233
0.4523	20	4.3619	0.7697	4.3404	0.7658
	50	4.6228	0.4925	4.6133	0.4927
	100	4.6041	0.336	4.5963	0.337
	500	4.7496	0.1589	4.7483	0.1587
	1000	4.7866	0.1089	4.7861	0.1089
0.6573	20	4.4579	0.6305	4.4058	0.6297
	50	4.6334	0.4678	4.6058	0.4613
	100	4.7947	0.1063	4.7927	0.1064
	500	4.7518	0.1513	4.7489	0.151
	1000	4.7783	0.11	4.7766	0.1099
0.9116	20	4.6918	0.9565	4.3217	0.7102
	50	4.9043	0.6265	4.6877	0.5329
	100	4.8313	0.4039	4.7098	0.3807
	500	4.7857	0.1457	4.7613	0.1416
	1000	4.7677	0.1117	4.7573	0.109
-0.1075	20	4.4252	0.7776	4.4153	0.7778
	50	4.6374	0.4485	4.6339	0.4473
	100	4.7107	0.3995	4.7101	0.3996
	500	4.7939	0.1677	4.7938	0.1677
	1000	4.7558	0.0958	4.7557	0.0958
-0.3442	20	4.4247	0.728	4.4053	0.7297
	50	4.5752	0.538	4.5666	0.5384
	100	4.6888	0.3087	4.6851	0.3084
	500	4.7744	0.1373	4.7736	0.1373
	1000	4.7698	0.1072	4.7694	0.1071
-0.6443	20	4.4431	0.9111	4.3401	0.886
	50	4.6581	0.5012	4.6256	0.4951
	100	4.7034	0.3334	4.6828	0.3293
	500	4.7254	0.1663	4.7218	0.1654
	1000	4.7683	0.1088	4.7672	0.1088
-0.9296	20	4.958	1.0681	4.3723	0.7077
	50	4.7367	0.6365	4.4817	0.4423
	100	4.9028	0.4203	4.7242	0.3193
	500	4.7812	0.1451	4.7505	0.1387
	1000	4.79	0.1193	4.7727	0.115

Tabuľka 4.3: Odhady smerodatnej odchýlky  $\sigma$

## Príloha 3

V tejto prílohe sa nachádzajú tabuľky obsahujúce výsledky simulačnej štúdie pre proces INAR(1) popísanej v druhej časti kapitoly 4.

$\alpha$	$n$	$\hat{\alpha}^{MM}$		$\hat{\alpha}^{LS}$	
		Priemer	Smer. odchýlka	Priemer	Smer. odchýlka
0.0761	20	0.0095	0.2123	0.0116	0.2242
	50	0.0336	0.1424	0.0343	0.1445
	100	0.0546	0.0909	0.0549	0.0922
	500	0.0622	0.044	0.0623	0.044
	1000	0.0774	0.0258	0.0774	0.0258
0.3861	20	0.2551	0.2032	0.2686	0.2162
	50	0.3373	0.1252	0.3441	0.1279
	100	0.3749	0.0989	0.3787	0.1001
	500	0.3839	0.0437	0.3847	0.0437
	1000	0.3868	0.0295	0.3871	0.0295
0.6668	20	0.4407	0.2092	0.4721	0.2242
	50	0.6048	0.113	0.6202	0.115
	100	0.6288	0.0643	0.6353	0.0645
	500	0.6551	0.0351	0.6561	0.035
	1000	0.6642	0.0269	0.665	0.027
0.9859	20	0.6849	0.1407	0.7559	0.1515
	50	0.861	0.0782	0.8888	0.0814
	100	0.9156	0.0424	0.9294	0.0428
	500	0.9737	0.0118	0.9764	0.0116
	1000	0.9799	0.0063	0.9814	0.0062

Tabuľka 4.4: Odhady parametra  $\alpha$

$\alpha$	$n$	$\hat{\lambda}^{MM}$		$\hat{\lambda}^{LS}$	
		Priemer	Smer. odchýlka	Priemer	Smer. odchýlka
0.0761	20	8.5392	1.8632	8.5214	1.9632
	50	8.3256	1.2559	8.3191	1.273
	100	8.2003	0.8292	8.1981	0.8385
	500	8.1155	0.4122	8.1144	0.4127
	1000	7.9949	0.2464	7.9943	0.2466
0.3861	20	9.8985	2.8708	9.7244	3.0305
	50	8.6499	1.7186	8.5599	1.7506
	100	8.1633	1.3047	8.1134	1.3195
	500	8.0405	0.5763	8.031	0.5745
	1000	7.9996	0.3909	7.9954	0.391
0.6668	20	13.4065	5.2972	12.6695	5.6405
	50	9.5744	2.8273	9.2072	2.8803
	100	8.9446	1.6031	8.7901	1.6088
	500	8.2818	0.8439	8.2575	0.8419
	1000	8.0496	0.6563	8.0325	0.6565
0.9859	20	179.316	79.8365	138.886	85.895
	50	78.8076	44.6023	63.0139	46.3519
	100	47.7324	24.0454	39.9039	24.2381
	500	14.9399	6.7059	13.3797	6.5628
	1000	11.4062	3.6083	10.5464	3.5323

Tabuľka 4.5: Odhady parametra  $\lambda$