

# CSŐDVALÓSZÍNŰSÉGEK INAR KÁRFOLYAMAT ESETÉN

DIPLOMAMUNKA

Írta: Árendás Ákos Tuzson

Biztosítási és pénzügyi matematika MSc  
Aktuárius szakirány

Témavezető:

Próhle Tamás

egyetemi tanársegéd

Valószínűségelméleti és Statisztika Tanszék



Eötvös Loránd Tudományegyetem

Természettudományi Kar

2015

# Köszönetnyilvánítás

Ezúton szeretnék köszönetet mondani témavezetőmnek Pröhle Tamásnak, akinek a legvégső tisztázásig terjedő rendszeres útmutatásai, tanácsai, értelmezést segítő példái segítettek szakdolgozatom elkészülését. Szeretnék köszönetet mondani mindazoknak is akik segítettek kiküszöbölni a fogalmazásbeli összeférhetlenségeket és a helyesírás ingoványos vizein is segítettek átevíckélni.

Budapest, 2015. december 30.

Árendás Ákos Tuzson

# Tartalomjegyzék

<b>Tartalomjegyzék</b>	<b>2</b>
<b>1. Bevezető</b>	<b>3</b>
1.1. Biztosítási környezet . . . . .	4
<b>2. Kockázati modellek</b>	<b>6</b>
2.1. Klasszikus kockázati modellek . . . . .	6
2.1.1. Egyedi kockázat modellje . . . . .	7
2.1.2. Összetett kockázat modellje . . . . .	8
2.2. Kockázati folyamatok . . . . .	11
2.2.1. Klasszikus rizikó folyamat . . . . .	12
2.2.2. Csődvalószínűségek a klasszikus modellekben . . . . .	13
<b>3. INAR folyamatok</b>	<b>16</b>
3.1. $A \circ$ ritkítő operátor . . . . .	17
3.2. INAR modellek . . . . .	19
3.2.1. INAR(1) folyamatok . . . . .	19
3.2.2. INAR(p) folyamatok . . . . .	22
<b>4. Poisson INAR folyamatok</b>	<b>28</b>
4.1. Poisson INAR(1) . . . . .	28
4.2. Csődvalószínűség . . . . .	32
<b>5. Adatelemzés</b>	<b>37</b>
<b>6. Kitekintés</b>	<b>45</b>
<b>Függelék</b>	<b>46</b>
<b>Irodalomjegyzék</b>	<b>57</b>

# 1. fejezet

## Bevezető

Az INAR (Integer-valued Autoregressive), azaz egészértékű autoregresszív folyamatok a '80-as évek végén kerültek először előtérbe és azóta számos cikk foglalkozott ezen folyamatokból származó modellek tulajdonságaival.

Az idősorokkal történő modellezést, azon belül is a különböző ARIMA modelleket, nagyon sok területen használták és használják is. Azonban nem feltétlenül alkalmasak nemnegatív, egészértékű folyamatok modellezésére, legfőképp a hiba tag normalitási feltételezése miatt. Módosítva a korábbi modelleket, először Al-Osh és Alzaid 1987-es [1] cikkében vezette be az INAR(1) folyamatot. Az elmúlt majdnem három évtizedben sok cikk született a témában és különféle (nem csak biztosítási, például [21] cikk) területeken alkalmazzák ezeket a modelleket. Biztosítási vonatkozásban a csődvalószínűségek minél pontosabb meghatározásai nagy szerepet játszhatnak akár egy-egy ágazat mögött lévő tőke nagyságának meghatározásánál, így a Szolvencia 2 alapú tőkeszükségleti modelleknél is.

Célunk, hogy a károk, és különös tekintetben az összkár alap modelljei ismeretése után eljussunk a kockázati folyamatok témaköréhez, ami után bevezetést nyerjünk az INAR modellek körébe. Mint ahogy később látni fogjuk, az INAR folyamatok természetes kiterjesztései a klasszikus kockázati folyamatoknak, mivel a kárszámok folyamatába egy autoregressziós dinamikát hoz

be, ezáltal biztosítva egy függőségi struktúrát. Megvizsgáljuk általánosan az INAR modellek tulajdonságait, majd speciális esetekben explicit formulákat is adunk csődvalószínűségekre. Az INAR(1) folyamatok általánosításaként megjelenő INAR( $p$ ) modellekhez, mint ahogy látni fogjuk, többféle módon juthatunk el. Definiálunk három különböző eshetőséget és ezen modelleket vizsgáljuk meg részletesebben az Csődvalószínűség fejezetben. Egy napjainkban is aktívan tevékenykedő biztosító adatait is megvizsgáljuk. A biztosítónak a kötelező gépjármű felelősségbiztosítás kárait vizsgáltuk. Alapfeltételezésünk, hogy valahogyan modellezni tudjuk a kárfolyamatot INAR folyamattal. Azt fogjuk belátni, hogy abban a megközelítésben, ahogy az adatokat megvizsgáltuk, nem modellezhetőek INAR folyamatokkal.

A klasszikus kockázati modellek és kockázati folyamatok témakörének feldolgozásakor nagyban támaszkodtam egyrészt [3] könyv első fejezetére, továbbá a [17] valamint a [24] jegyzetekre, ahonnan az alapdefiníciókat és a főbb tételeket hivatkozom, összhangba hozva és kiegészítve [6] jelöléseivel, amely alapján felépítettem a dolgozatot. Ezekben a témákban az áttekintés nem teljes körű, csak a legfontosabb definíciókat és tételeket igyekeztem kiemelni, amelyek egy egyértelmű felvezetést biztosítanak az INAR folyamatok előtt. Az INAR folyamatok kapcsán a már említett [1] mellett az INAR( $p$ ) modelleknél alapul a [2] cikket vettem. A speciális esetekre vonatkozó csődvalószínűségek meghatározásánál [6] cikkből indultam ki. További hivatkozások megtalálhatók az egyes fejezetekben.

## 1.1. Biztosítási környezet

Alapvetően a dolgozat a kár, kárszámok és az ezekből származtatott összkár fogalmak köré összpontosít. A gyakorlatban, azaz egy biztosító társaság működése során, nagyon sok ezekhez kapcsolódó fogalmat különböztetünk meg egymástól. Elkülönülnek a károk tényleges bekövekezésének időpontjai, a károk bejelentéseinek időpontjaitól, valamint a károk esetleges kifizetéseinek időpontjaitól is. Ehhez kapcsolódóan a kárigények különböznek a károk tényleges nagyságaitól, valamint ezektől különbözhet a kárral kapcsolatos ki-

fizetés tényleges nagysága. Mi most azon egyszerűsítésekkel fogunk élni, hogy kárnak fogjuk nevezni az egyes károkkal kapcsolatos tényleges kifizetéseket, valamint a bejelentett károk darabszámát kárszámnak fogjuk nevezni.

## 2. fejezet

# Kockázati modellek

### 2.1. Klasszikus kockázati modellek

Egy biztosító számára fontos, hogy ismerje különböző vonatkozásokban (szerződésenként, módozatonként, ágazatonként) a jövőbeni kárkifizetések tulajdonságait (elsősorban a tartalékképzés miatt, de a díjak meghatározásában is nagy szerepet játszhatnak). Ezen károk összességét kárösszegnek nevezük, a különböző kockázati modellek ennek eloszlását határozzák meg, rögzített időintervallumon vizsgálódva. A modellek között vannak hasonlóságok, azonban gyökeres különbséggént jelenik meg az egyes biztosítási szerződésekhez kapcsolódó vagy kapcsolódható károk száma. A két alapvető modell az *egyéni (vagy egyedi) kockázat modellje*, valamint az *összetett kockázat modellje*.

A biztosítási matematikában alapvető feltevések közé sorolhatók, hogy a károk nagyságai (kárnagyságok) függetlenek egymástól és gyakran azonos eloszlásúak, valamint, hogy a káresemények száma független a károk nagyságától. Ez a fent említett két modellben is érvényes. A fő különbség a károk darabszámának a modellezéséből adódik: míg az egyedi kockázat modelljében a károk darabszámára determinisztikusan, addig az összetett kockázat modelljében valószínűségi változóként tekintünk rá.

A következőkben jelölje  $B_i$ ,  $i = 1, 2, \dots$  a biztosítóhoz befolyt kárigényekkel

kapcsolatos kifizetések nagyságát (kárnagyság). Általános jelölésként bevezetjük, hogy egy tetszőleges  $X$  valószínűségi változó elszlásfüggvényét  $F_X(x)$ , generátorfüggvényét  $G_X(r) = E(r^X)$ , karakterisztikus függvényét  $\varphi_X(r) = E(e^{irX})$ , momentumgeneráló függvényét  $M_X(r) = E(e^{rX})$  jelöli, továbbá Laplace-transzformáltja:  $L_X(r) = E(e^{-rX})$ .

### 2.1.1. Egyedi kockázat modellje

Álljon a veszélyközösség  $n$  (determinisztikus) egyedből (kötvény, biztosítási szerződés, de leegyszerűsítve gondolhatunk rá úgy, mint a biztosítottak száma). Jelölje  $B_i$  az  $i$ -edik szerződés kárának nagyságát, azaz az  $i$ -edik szerződés kockázatát, ahol  $B_i > 0$  valószínűségi változó. Továbbá tegyük fel, hogy az egyes  $B_i$  valószínűségi változók függetlenek egymástól, valamint, hogy eloszlásuk ismert. Fontos megjegyeznünk, hogy az iménti "egyéni" (azaz egy szerződéshez kapcsolódó) kockázatok esetében egyetlen kárnagyságot vizsgálunk, azonban ezek lehetnek több kárból származó összegkárak. Legyen a biztosító által kifizetett összkár, azaz a portfólió teljes kockázata  $W$ . Ezekkel a jelölésekkel az egyedi kockázat modelljében a összkárt (azaz a teljes veszteséget vagy teljes kifizetést) a következő formulával írhatjuk le:

$$W = B_1 + \dots + B_n = \sum_{i=1}^n B_i. \quad (2.1)$$

A modellt általában akkor alkalmazzák, amikor a veszélyközösség nem homogén, azaz a  $B_i$ -k nem azonos eloszlásúak. Ezért leggyakrabban balesetbiztosítások és kockázati életbiztosítások esetében használják. Fontos még egyszer megjegyezni, hogy a modell egy adott időszak (például egy negyedév vagy egy év) vonatkozásában vizsgálja az összes kifizetést. Az összkár eloszlásának meghatározására rekurziós algoritmusokat alkalmaznak, mint például a [20]-ben bemutatott De Pril algoritmus.

#### Állítás 2.1 (De Pril algoritmus)

*Tegyük fel, hogy a veszélyközösségben minden biztosítottnak egyfajta, előre rögzített kára lehet. A szerződéseket a biztosítási összeg nagysága ( $i$ , azaz*



a kifizetendő összeg), valamint a kár bekövetkezésének valószínűsége ( $0 < q_j < 1$ ) szerint csoportosítjuk úgy, hogy az egyes csoportokbeli szerződések számát jelölje  $n_{i,j}$ . Ekkor a kárkifizetésekre valószínűségeire vezessük be a következőket:  $P(B_{i,j,l} = 0) = 1 - q_j$ , valamint  $P(B_{i,j,l} = i) = q_j$ , ahol  $i = 1, \dots, r$ ,  $j = 1, \dots, m$ ,  $l = 1, \dots, n_{i,j}$ . Ezekkel  $W$ -t írjuk fel a következő módon:

$$W = \sum_{i,j,l} B_{i,j,l} = \sum_{i=1}^r \sum_{j=1}^m \sum_{l=1}^{n_{i,j}} B_{i,j,l}.$$

Ezzel  $W$  eloszlására teljesül a következő rekurzió:

$$P(W = 0) = \prod_{i=1}^r \prod_{j=1}^m (1 - q_j)^{n_{i,j}}$$

$$P(W = k) = \frac{1}{k} \sum_{i \cdot l \leq k} h(i, l) \cdot P(W = k - i \cdot l)$$

ahol  $k > 0$  és  $h(i, l) = i \cdot (-1)^{l-1} \cdot \sum_{j=1}^m n_{i,j} \cdot \left(\frac{q_j}{q - q_j}\right)^l$ .

### 2.1.2. Összetett kockázat modellje

Mielőtt az összetett kockázat modelljével foglalkozunk mondjunk az *összetett eloszlásra* vonatkozó alapdefiníciót, majd rá vonatkozó tulajdonságokat.

#### Definíció 2.1 (Összetett eloszlás)

Az  $W$  valószínűségi változó összetett eloszlású, ha léteznek  $B, B_1, B_2, \dots$  független, azonos eloszlású val. változók, valamint egy tőlük független nemnegatív, egészértékű  $N$  val. változó, melyekre  $S_N = B_1 + \dots + B_N$  úgy, hogy  $X$  eloszlásával megegyezik az összeg eloszlása, azaz  $W \sim S_N$ .

**Tulajdonságok 2.2**

A  $W$  ( $W \stackrel{d}{=} B_1 + \dots + B_N$ ) összetett eloszlású valószínűségi változóra:

$$\varphi_W(r) = G_N(\varphi_B(r)),$$

$$M_W(r) = G_N(M_B(r)),$$

$$L_W(r) = E(e^{-rW}) = E((L_B(r))^N) = G_N(L_B(r)).$$

Ha  $E(N), E(B) < \infty$  (Wald első azonossága):

$$E(W) = E(N)E(B),$$

ha  $N$  és  $B$  szórásnégyzete véges, (Wald második azonossága):

$$D^2W = E(N)D^2(B) + D^2(N)E(B).$$

**Speciális eset****Definíció 2.2 (Összetett Poisson-eloszlás)**

Legyen  $W$  összetett eloszlású,  $N \sim \text{Poisson}(\lambda)$ ,  $B_i$ -k független, azonos eloszlásúak, melyek függetlenek  $N$ -től is. Ekkor a  $W$  összetett Poisson-eloszlású, jelölése:  $W \sim \text{Poisson}(\lambda, Q)$ , ahol  $Q = Q_B$ , a  $B_i$ -k közös eloszlását jelöli.

Most rátérhetünk az összetett kockázati modell tárgyalására. Abban az esetben, ha a portfólió minden egyedéhez több káresemény is tartozhat, akkor egy adott időszakra (pl. egy évre) az összkár a következő módon áll elő:

$$W = B_1 + \dots + B_N = \sum_{i=1}^N B_i. \quad (2.2)$$

Itt szintén  $B_i$  valószínűségi változó az  $i$ -edik kár nagyságát jelöli, azonban a károk száma  $N$  sztochasztikus. A modellben  $B_i > 0$  ( $i = 1, 2, \dots, N$ ),  $N$  és  $B_i$   $i = 1, 2, \dots, N$  függetlenek egymástól,  $N$ -nek ismert az eloszlása és a  $B_i$  változók független, azonos eloszlásúak, ismert eloszlással (a továbbiakban a közös eloszlásuk legyen  $B$ ). A veszélyközösségben a modell keretei között minden egyedhez több kár is tartozhat, mert a károk külön-külön vannak mérve, nem pedig egy szerződéshez kötve egyfajta összegkárként (mint ahogy azt az egyedi kockázat modellje esetében láttuk).

Az összkár összetett eloszlású, így igazak rá a (2.2) tulajdonságok.

Az összkár eloszlásának meghatározására az összetett kockázat modelljében szintén több lehetőség kínálkozik<sup>1</sup>, mi most a *Panjer-rekurziót* mutatjuk be a [18] cikk alapján, mivel a gyakorlatban főleg ez terjedt el. Ez előtt azonban szükséges lesz definiálni egy, a kárszámokra vonatkozó fogalmat, valamint kimondani egy hozzájuk kapcsolódó állítást, mert a Panjer-rekurzió csak ezen eloszlások körében működik.

### Definíció 2.3 (( $a, b, 0$ ) eloszlások)

Az  $N$  nemnegatív, egészértékű valószínűségi változó (amely most nevezetesen a kárszámot jelöli) ( $a, b, 0$ ) eloszlású, ha léteznek olyan  $a, b \in \mathbb{R}$  számok, melyekre a következő rekurzió teljesül:

$$P(N = n) = \left(a + \frac{b}{n}\right)P(N = n - 1) \quad n = 1, 2, \dots$$

### Állítás 2.3 (( $a, b, 0$ ) eloszlások karakterizációja)

Az  $N$  nemnegatív, egészértékű valószínűségi változó akkor és csak akkor ( $a, b, 0$ ) eloszlású, ha  $N$  Poisson, binomiális vagy negatív binomiális eloszlású.

**Megjegyzések** Az egyes paraméterekre a következő értékek adódnak: ha  $N \sim \text{Poisson}(\lambda)$ , akkor  $a = 0$  és  $b = \lambda$ ; ha  $N \sim \text{Bin}(n, p)$ , akkor  $a = -\frac{p}{1-p}$  és  $b = \frac{(n+1)p}{p}$ ; ha  $N \sim \text{NegBin}(r, q)$ , akkor  $a = q$  és  $b = (r - 1)q$ .

Természetesen nem feltétlenül jó választás mindig a kárszámokra valamelyik ( $a, b, 0$ ) eloszlást választani, a gyakorlatban előfordulnak nyilván olyan adatsorok, amelyekre más eloszlások jobb illeszkedést mutatnak.

A következőkben legyen egy tetszőleges egészértékű valószínűségi változóra érvényes az alábbi jelölés:  $p_B(n) := P(B = n)$ .

### Tétel 2.4 (Panjer-rekurzió)

Legyen  $N \sim (a, b, 0)$  eloszlású és  $B > 0$  egészértékű. Ekkor a  $W$  összkár

<sup>1</sup>A különböző módszerek és azok továbbfejlesztéseivel kapcsolatos cikkek szedete megtalálható [8] cikkben.

eloszlására teljesül a következő rekurzív azonosság:

$$p_W(0) = p_N(a)$$

$$p_W(n) = \sum_{y=1}^n \left(a + b \frac{y}{n}\right) p_B(y) p_W(n-y) \quad n = 1, 2, \dots$$

Az összkár eloszlását nem csak Panjer-rekurzióval lehetséges meghatározni, valamint az itt említett algoritmusnak is léteznek különböző általánosításai, azonban ezekről most nem ejtünk szót.

## 2.2. Kockázati folyamatok

Míg az imént tárgyalt kockázati modellek időben statikusak, azaz egy rögzített időintervallum vizsgálatára alkalmazhatók, addig a kockázati folyamatok időben dinamikusan vizsgálják a károkat. A kockázati folyamatok központi kérdése, hogy mi a valószínűsége egy biztosító társaság csődjének<sup>2</sup>. Az ilyen jellegű kérdések vizsgálatához az idő függvényében kell a károkra tekintenünk. Az összetett kockázat modelljét vesszük kiindulási pontnak, amely egy rögzített időintervallumon meghatározta a összkárt úgy, hogy a károk darabszáma is sztochasztikus volt. A 2.2 egyenlet módosításaként definiáljuk az időtől függő károk számát egy  $(N_t)_{t \geq 0}$  sztochasztikus folyamatként, melyet *kárszámfolyamatnak* neveznek. Ebből kapjuk *kárfolyamatot*, azaz az összkár folyamatát az idő függvényében, amely lényegében egy stilizált nem-élet biztosító társaság tőkéjét<sup>3</sup> mutatja  $t$  szerint.

### Definíció 2.4 (Kárfolyamat)

A *kárfolyamat* olyan  $(W_t)_{t \geq 0}$  sztochasztikus folyamat, melyre:

$$W_t = B_1 + \dots + B_{N_t} = \sum_{i=1}^{N_t} B_i \quad (2.3)$$

<sup>2</sup>Sok egyéb kérdés is fontos szerepet játszik a kockázati folyamatok témakörében, például  $u$  kezdőtőke növelésével hogyan lehet csökkenteni a csőd valószínűségét, de lehet vizsgálni a tönkre menés mértékét is („mennyire nagy a csőd”).

<sup>3</sup>Angolul „surplus process”-ként terjedt el.

ahol  $N_t$  az imént említett kárszámfolyamat, valamint a  $B_i$  ( $i = 1, \dots, N_t$ ) valószínűségi változók a korábban bevezetett károk értékei.

### Tulajdonságok 2.5

A  $W_t$  kárfolyamat összetett eloszlású minden  $t$ -re, így igazak rá a (2.2) tulajdonságok.

A kockázati folyamat (vagy más néven *rizikó folyamat*) ekkor a következő.

### Definíció 2.5 (Kockázati folyamat)

A kockázati folyamat egy olyan  $(U_t)_{t \geq 0}$  sztochasztikus folyamat, amelyre:

$$U_t = u + P_t - W_t \quad (2.4)$$

ahol:

$u$ -a biztosító társaság kezdeti tőkéje (determinisztikus),

$P_t$ -a biztosítóhoz beérkezett összbefizetés (díjakkól) a  $[0, t]$  intervallumon,

$W_t$ -az imént definiált kárfolyamat.

Mint ahogy említettük, a kockázati folyamatok egyik alapvető kérdésköre, hogy (például egy biztosító társaság) milyen valószínűséggel következik be a csőd esemény, azaz mekkora valószínűséggel lesz az  $(U_t)_{t \geq 0}$  sztochasztikus folyamat egy tetszőleges  $t$  időpillanatban negatív. A következő jelöléseket szokták alkalmazni a csőd, illetve „nem-csőd” valószínűségekre:

$$\Psi(u) = P(\exists t \geq 0 : U_t \leq 0)$$

$$\Phi(u) = P(\forall t : U_t \geq 0)$$

### 2.2.1. Klasszikus rizikó folyamat

Definiáljuk a klasszikus rizikó folyamatot, amely esetében a kimondjuk majd a csőd valószínűségére vonatkozó állításokat is.

**Definíció 2.6**

A (2.4) egyeletben definiált kockázati folyamat esetében:

$$\begin{aligned}
 P_t &= \pi \cdot t && \text{ahol } \pi > 0 \text{ állandó,} \\
 N_t &&& \lambda(> 0)\text{-paraméterű homogén Poisson-folyamat}^4, \\
 B_i &&& B_i \geq 0, i = 1, 2, \dots \text{ független, azonos eloszlású val.változók,} \\
 &&& \text{amik függetlenek } N_t\text{-től is, minden } t\text{-re.}
 \end{aligned}$$

Ezzel a klasszikus rizikó folyamat:

$$U_t = u + \pi \cdot t - W_t = u + \pi \cdot t - \sum_{i=1}^{N_t} B_i \quad (2.5)$$

**Tétel 2.6**

Klasszikus rizikó folyamat esetén teljesülnek az alábbiak:

1. A  $W_t$ ,  $t \geq 0$ , kárfolyamat és az  $U_t$ ,  $t \geq 0$ , rizikó folyamat független és stacionárius növekményű.
2. A  $\Psi(u)$  függvény monoton csökkenő, a  $\Phi(u)$  függvény monoton növekvő, és mindkét függvény càdlàg a pozitív félegyenesen.
3. A  $\Phi(u)$   $u \geq 0$ , függvény abszolút folytonos, tehát majdnem mindenhol deriválható, és a  $\Phi(u)$ ,  $u \geq 0$ , deriváltjára

$$\Phi(u) = \Phi(0) + \int_0^u \Phi'(z) dz, \quad u \geq 0.$$

**2.2.2. Csődvalószínűségek a klasszikus modellekben**

A klasszikus rizikó folyamat esetében a csődvalószínűségek meghatározásánál különböző esetekre szedhetjük szét a folyamatot a  $\pi$  – időegységre jutó díjbefizetés, a  $\lambda$  – időegységre eső káresemények száma (azaz az  $N_t$  Poisson-folyamat paramétere), valamint a  $\beta$  – átlagos kárkifizetések nagysága (azaz a  $B_i$ -k közös várható értéke) egymáshoz való viszonya szerint. Számunkra az az eset lesz a nem érdektelen eset, amikor:  $\pi > \lambda\beta$ , (megj.  $E(W) = E(N)E(B) = \lambda\beta$ )

<sup>4</sup> $N_0 = 0$ ; növekményekre igaz, hogy, ha  $s \leq t$ , akkor stacionárius növekményű, azaz  $N_t - N_s \sim Poisson(\lambda(t-s))$ ; valamint független növekményű, azaz a diszjunkt időintervallumhoz tartozó növekmények egymástól függetlenek.

azaz teljesít egyfajta fizetőképességi feltételt. Szemléletesen is ezt várjuk, hiszen az  $E(W)$  az időegységre jutó kárkifizetéseket jelenti, és ha az kisebb lenne, mint a befizetések  $\pi$  összege, akkor a biztosító nem tudná teljesíteni a kötelezettségeit. Rendezzük át a  $\pi > \lambda\beta$  összefüggést úgy, hogy  $\frac{\pi}{\lambda\beta} > 1 \Leftrightarrow \frac{\pi}{\lambda\beta} - 1 > 0$ . Ekkor a  $\eta := \frac{\pi}{\lambda\beta} - 1$  hányadost *relatív kockázati felárnak* nevezzük. Ezt foglalja össze az alábbi tétel.

### Tétel 2.7

*Klasszikus rizikófolyamat esetén igazak a következők:*

1. Ha  $\pi < \lambda\beta$ , akkor  $\Phi(u) = 0$ ,  $u \geq 0$ , azaz akármekkora  $u$  kezdődőtőkével indulva a biztosító 1 valószínűséggel csődbe fog menni (végtelen időhorizonton vizsgálódva).
2. Ha  $\pi > \lambda\beta$ , akkor a  $\Phi(\infty) := \lim_{u \rightarrow \infty} \Phi(u)$  határértékre  $\Phi(\infty) = 1$  majdnem mindenütt (azaz  $\forall u > 0 \exists \varepsilon > 0 : \Phi(u) > 1 - \varepsilon$ ).

A klasszikus elméletben különböző integrálegyenletek megoldásaként juthatunk el a „nem csőd” valószínűségekhez. A részletes kidolgozás megtalálható [17] vagy [24] jegyzetekben, itt csak a legfontosabb eredményeket közöljük (klasszikus rizikó folyamatot alapul véve).

A nem csőd valószínűségre a Cramer-féle (lásd az imént említett jegyzetekben) integrálegyenlet megoldásával az alábbi tétel igaz.

### Tétel 2.8

*Klasszikus rizikófolyamat esetén, ahol  $\pi > \lambda\beta$   $\Phi(0) = 1 - \frac{\lambda\beta}{\pi}$ .*

Mivel az integrálegyenletek általában nehezen megoldható, ezért a csődvalószínűségek aszimptotikáját szokták vizsgálni a csőd becsléséhez. Ehhez a következő konvex függvényt szokták bevezetni:

$$c(r) := M_B(r) - 1.$$

Mivel a kárnagyságok nemnegatívak, ezért a  $c$  függvény létezik a negatív félegyenesen. Ahol ez a függvény létezik, ott konvex. Vizsgáljuk a  $c(r) = \pi r / \lambda$  egyenlet megoldásait. Mivel  $c$  konvex, ezért legfeljebb két megoldása lehet,

továbbá  $c(0) = 0$ , ami miatt az egyenletnek maximum egy pozitív gyöke lehet. Ekkor az alábbi tétel igaz a csődvalószínűségekre.

**Tétel 2.9 (Cramér–Lundberg approximáció)**

Tegyük fel, hogy a  $\frac{\pi}{\lambda} = \frac{c(r)}{r}$  egyenletnek létezik pozitív  $\rho$  megoldása és tegyük fel, hogy a  $c$  függvény véges a  $\rho$  valamely környezetében. Ekkor

$$\lim_{u \rightarrow \infty} e^{\rho u} \Psi(u) = \frac{\pi - \lambda \beta}{\lambda c'(\rho) - \pi},$$

ahol a  $\rho$  érték a Lundberg-kitevő.

A tétel lényegében azt állítja, hogy a csőd valószínűsége az  $u$  kezdőtőke függvényében asszimptotikusan exponenciális. A gyakorlatban legtöbbször nem ismerjük sem az  $\lambda$  értéket (azaz az  $N_t$  Poisson-folyamat intenzitását), valamint a kárkifizetések nagyságát leíró  $B_i$  változók eloszlását sem. Emiatt a Lundberg-kitevő értékét sem tudjuk pontosan meghatározni, azonban különböző becslésekre vonatkozó tételek garantálják, hogy a csődvalószínűség jól becsülhető a Lundberg-kitevővel.

Fennálnak tehát a következő összefüggések:

$$\lim_{u \rightarrow \infty} -\frac{\ln(\Psi(u))}{u} = \rho,$$

$$\Psi(u) \simeq e^{-\rho u}.$$

Valamint a konvex függvényt definiáljuk úgy, hogy:

$$c_n(r) = \frac{1}{n} \ln \left( E \left( e^{r(S_n - n\pi)} \text{big} \right) \right).$$

Ekkor a

$$c(r) = \lim_{n \rightarrow \infty} c_n(r) = 0$$

egyenlet  $\rho$  megoldása a Lundberg-kitevő. A Lundberg-kitevő szemléletesen a biztosítási portfóliónk veszélyességének ad egyfajta mértéket.



## 3. fejezet

# INAR folyamatok

Az alábbi fejezetben tárgyaljuk azokat a folyamatokat, melyek a dolgozat fő témáját alkotják. A korábban definiált (2.6) klasszikus rizikófolyamatot szeretnénk továbbfejleszteni. A kárfolyamat esetében az egyes periódusok közötti függetlenség igen erős kritérium. A függőségi dinamika életszerű feltételezés lehet, főleg ha biztosítási környezetben vizsgáljuk. Elegendő belegondolnunk abba, hogy egy személynek akár több szerződése is lehet, vagy abba az esetben, ha a biztosítás fedezete egy családra szól, vagy egy vállalat munkavállalóira. Ezekben az esetekben azonnal észlelhető a függőségi viszony, de érdemes arra az esetre is gondolni példaként, amikor egy biztosítóról elterjed, hogy "minden kárt kifizet". Elképzelhető, hogy ebben az esetben megnövekszik az kárbejelentések száma, akár amiatt, mert normál esetben nem biztos, hogy minden biztosítási eseményt bejelentenek a kötvényesek (gondolhatunk akár a kötelező gépjármű felelősségbiztosítási ügyekre), vagy akár felmerülhet egyéb biztosítási csalások kockázata is.

Olyan modellekre van szükségünk, amik megbírkóznak a nemnegatív diszkrét értékekkel, valamint az idősorok esetében fennálló autokorrelációval.

Az alap autoregressziós  $AR(p)$  modellek nem megfelelőek, hiszen ott egyrészt az egyes  $Y_t$  értékek nem egészértékűek, de a zaj folyamat normalitása, így a negatív értékek jelenléte sem alkalmas a kárszámok folyamatának modellezésére.

Megoldásként számos modell keletkezett. Mi most az INAR (Integer Valued Autoregressive), azaz az egészértékű autoregresszív folyamatokat vizsgáljuk. Ezeknek a modelleknek is számos megközelítése lehetséges. Egyrészt elkülönülnek a szorzást felváltó új műveletek definiálásában, másrészt az egyes változók eloszlásainak megválasztásában (ami kihatással van a folyamat határeloszlására), de a modellezési környezet is többféle lehet.

### 3.1. $A \circ$ ritkító operátor

Az AR modellekben a szorzás műveletét valamilyen alternatív művelettel fel kell váltani ahhoz, hogy az egészértékűséget biztosítani tudjuk.  $A \circ$  ritkító operátort Steutel és van Harn vezette be 1979-es [23] cikkében. Az alapdefiníció és a cikk is a folytonos valószínűségi változók önfelbontható tulajdonságának a diszkretizálásáról szól, amely folytonos függvények esetében a karakterisztikus függvények, diszkrét esetben pedig a generátorfüggvények között fennálló összefüggést feltételez.<sup>1</sup> Folytonos esetben egy stacionárius AR(1) folyamat létezésére egy szükséges feltétel, hogy a határeloszlás önfelbontható tulajdonságú legyen (lásd az előbbi lábjegyzetben). Ez a tulajdonság öröklődik a diszkrét esetre is, ezért fontos egy megfelelő művelet bevezetése, ami ezt a tulajdonságot teljesíti. Erre a definícióra és az ebből következő állításra később visszatérünk.

#### Definíció 3.1 (Binomiális ritkítás)

Legyen  $X$  egy nem-negatív egészértékű valószínűségi változó, és legyen  $\alpha \in [0, 1]$  szám. Ekkor a  $\circ$  operátor definiálja a következő műveletet:

$$\alpha \circ X := \begin{cases} \sum_{i=1}^X Y_i & X > 0 \\ 0 & X = 0 \end{cases}$$

<sup>1</sup>Egy valós értékű valószínűségi változó önfelbontható (vagy  $L$  osztálybeli, angolul self-decomposable), ha a karakterisztikusfüggvénye kielégíti a  $\varphi(t) = \varphi(\alpha t)\varphi_\alpha(t)$   $t \in \mathbb{R}; \alpha \in (0, 1)$  összefüggést, ahol  $\varphi_\alpha$  karakterisztikus függvény. Megfelelő valószínűségi változókkal ez a következőt jelenti (eloszlásban):  $X = \alpha X' + X_\alpha$   $\alpha \in (0, 1)$ , ahol  $X'$  és  $X_\alpha$  függetlenek, valamint  $X'$  eloszlása megegyezik  $X$  eloszlásával.

ahol  $Y_i$  minden  $i$ -re független azonos eloszlású valószínűségi változó, amelyek mindegyike független  $X$ -től úgy, hogy

$$Y_i \sim \text{Bernoulli}(\alpha) \quad \text{azaz} \quad P(Y_i = 1) = 1 - P(Y_i = 0) = \alpha.$$

A következő állítás a definícióból azonnal következő tulajdonságait sorolja fel az operátornak. Az állítások megtalálhatók [23] és [25] cikkekben.

### Állítás 3.1 ( $A \circ$ operátor alaptulajdonságai)

1.  $\alpha \circ X \in \mathbb{N}$
2.  $0 \circ X = 0$
3.  $1 \circ X = X$
4.  $E(\alpha \circ X) = \alpha E(X)$
5.  $D^2(\alpha \circ X) = \alpha^2 D^2(X) + \alpha(1 - \alpha)E(X)$
6.  $\text{Cov}(\alpha \circ X, X) = \alpha D^2(X)$
7.  $\forall \beta \in [0, 1]: \beta \circ (\alpha \circ X) \stackrel{d}{=} (\beta\alpha) \circ X$
8.  $\alpha \circ (X + Y) \stackrel{d}{=} \alpha \circ X + \alpha \circ Y$

A binomiális szóhasználat onnan ered, hogy ha  $n \in \mathbb{N}$  darab  $\text{Bernoulli}(\alpha)$  eloszlású valószínűségi változókat összegzünk, akkor az összeg eloszlása  $\text{Bin}(n, \alpha)$  eloszlás lesz. A  $\alpha \circ X$  valószínűségi változóra igaz, hogy:  $\alpha \circ X \sim \text{Bin}(X, \alpha)$  az  $X$  feltétel mellett (tehát, ha ismerjük  $X$  értékét). Szemléletesen a  $\circ$  oprátor esetében valóban egy ritkítés történik, hiszen ha egy egész számra (jelen esetben  $X$ ) úgy gondolunk, mint 1-esek összegére, akkor minden egyes tagnál (egyed, egyén stb.) eldöntjük, hogy bent hagyjuk-e, vagy nem ( $\alpha$  valószínűséggel). Így a már megritkített  $\alpha \circ X$  valószínűségi változó a ritkítés utáni „túlélőket” számolja meg.

### Állítás 3.2

Az  $\alpha \circ X$  valószínűségi változó generátorfüggvénye:

$$G_{\alpha \circ X} = G_X(1 - \alpha + \alpha s)$$

### Definíció 3.2 (Diszkrét önfelbonthatóság tulajdonság)

Egy nemnegatív egészértékű  $X$  valószínűségi változó rendelkezik a diszkrét önfelbontható tulajdonsággal, ha minden  $\alpha \in (0, 1)$ -ra  $X$  felírható a következő

alakban:  $X \stackrel{d}{=} \alpha \circ X + X_\alpha$ , ahol az  $\alpha \circ X$  és  $X_\alpha$  függetlenek. Ugyanez generátorfüggvényekkel megfogalmazva:  $G_X(s) = G_X(1 - \alpha + \alpha s)G_{X_\alpha}(s)$ .

Megjegyezzük, hogy rögzített  $\alpha \in [0, 1]$ -re a ritkító operátorral kapott  $\alpha \circ X$  valószínűségi változó egészértékű, így az előző eloszlásbeli egyenlőségben  $X_\alpha$  érték mindenképpen egészértékű. Továbbá az is igaz, hogy  $X_\alpha$  eloszlása egyértelműen meghatározza  $X$  eloszlását is.

## 3.2. INAR modellek

Az imént definiált ritkító operátorral lehetőség nyílik arra, hogy az egyes vizsgált időszakok kárszámaiban összefüggő modellt alakítsunk ki. Először felírjuk az elsőrendű modelleket (az alapmodell Al-Osh és Alzaid [1] valamint McKenzie [14] cikkei alapján), majd az általánosabb  $p$ -rendű modelleket is. A magasabb rendű, általános modelleket is többen vizsgálták, többféle megközelítésben. Mi három különböző modellt fogunk vizsgálni (Alzaid and Al-Osh (1990) [2], Du and Li (1991) [7], [4]) de ezeken kívül nagyon sok alternatíva megjelent.<sup>2</sup>

### 3.2.1. INAR(1) folyamatok

#### Definíció 3.3 (INAR(1) modell)

Legyen  $(N_t)_{t>1}$  egy nemnegatív egészértékű valószínűségi változókból álló sorozat és  $\alpha \in [0, 1]$  pedig egy valós konstans. Ekkor az INAR(1) folyamatot a következőképpen definiáljuk:

$$N_t = \alpha \circ N_{t-1} + \varepsilon_t \quad (3.1)$$

ahol  $\varepsilon_t$  független, azonos eloszlású, diszkrét, nemnegatív értékű valószínűségi változó minden  $t$ -re, aminek létezik várható értéke és szórásnégyzete, melyeket jelöljenek rendre:  $\mu_\varepsilon$ , és  $\sigma_\varepsilon^2$ .

Egyszerűen adódik a modellre vonatkozó várható ésrétk és szórásnégyzet értékek ([1] alapján):

<sup>2</sup>Lásd bővebben Latour (1998)-as [13] cikke, Franke és Subba Rao (1995)-ös [9] cikkében definiált és Silva és Silva (2006)-os [22] cikkében taglalt módok.

**Tulajdonságok 3.3 (Várható érték és variancia)**

$$\begin{aligned}
E(N_t) &= \alpha E(N_{t-1}) + \mu_\varepsilon = \alpha^t E(N_1) + \mu_\varepsilon \sum_{j=0}^{t-1} \alpha^j \\
D^2(N_t) &= \alpha^2 D^2(N_{t-1}) + \alpha(1-\alpha)E(N_{t-1}) + \sigma_\varepsilon^2 = \\
&= \alpha^{2t} D^2(N_1) + (1-\alpha) \sum_{j=1}^t \alpha^{2j-1} E(N_{t-j}) + \sigma_\varepsilon^2 \sum_{j=1}^t \alpha^{2j-1}.
\end{aligned}$$

A várható értékre vonatkozó első egyenletnél felhasználtuk a  $\circ$  ritkító operátorra vonatkozó (3.1) tulajdonságainak a várható értékre vonatkozó állítását.

Fontos észrevétel, hogy az alapmodellben semmilyen feltételezés nincsen sem az  $N_1$ , sem pedig az  $\varepsilon_t$  hibák eloszlására nézve.

A 3.1 összefüggésben definiált rekurziós modell egy szemléletes interpretálása a következő lehet. Ha a kárszámokra úgy gondolunk, mint ha azok egyes károkozókhoz<sup>3</sup> kapcsolódnának<sup>4</sup>. Ekkor a modellt egyfajta populációs modellnek is felfoghatjuk, az  $N_t$  reprezentálja a károkozók „populációjának” méretét a  $t$  időpontban, a  $\alpha \circ N_{t-1}$  azon károkozókat, akik korábban kárt okoztak és most is (egyfajta „túlélők” – bent maradtak a rendszerben) és  $\varepsilon_t$  pedig az új károkozók száma (azaz a károkozók populációjába „bevándorlók”).<sup>5</sup>

$$\underbrace{N_t}_{\text{károkozók } t\text{-ben}} = \underbrace{\alpha \circ N_{t-1}}_{t-1 \text{ túlélői}} + \underbrace{\varepsilon_t}_{\text{bevándorlók}} \quad (3.2)$$

A korábban 3.2-ben definiált tulajdonság fontosságát mutatja az alábbi állítás.

<sup>3</sup>Persze gondolhatunk rájuk úgy is, mint károkat elszenvedők, de most nem azt a megközelítést alkalmazzuk (később az INAR(p) modellek szemléletes bevezetésénél is hasznos lesz ez a megközelítés).

<sup>4</sup>Az összetett kockázati modellekben persze azt feltételezzük, hogy egy egyénhez több kár is kapcsolódhat, azaz több kárszámban is „játszhat szerepet”, de a szemléletességhez most ettől tekintünk el.

<sup>5</sup>Többféle alternatív interpretációja is lehet a folyamatnak. Felfogható bevándorlással kiegészített szubkritikus Galton–Watson folyamatként (lásd [10]), vagy mint  $M/M/\infty$  sorbanállási rendszerként (lásd [16])

**Állítás 3.4**

Rögzített  $\alpha \in [0, 1]$  valós konstansra a 3.1 folyamat akkor és csak akkor stacionárius, ha a folyamat határeloszlása rendelkezik a diszkrét „self-decompatbility” tulajdonsággal, azaz jelen esetben a generátorfüggvénye felírható a következőképpen:

$$G_X(r) = G_X(1 - \alpha + \alpha r)G_\varepsilon(r). \quad (3.3)$$

Az állításban szereplő (3.3) egyenletet, ha átrendezzük a  $G_\varepsilon(r) = \frac{G_X(r)}{G_X(1-\alpha+\alpha r)}$  alakra, akkor egy olyan alakot kapunk, amely jó lehet az alkalmas eloszlást meghatározni az  $\varepsilon_t$ -hez, amikor a megfigyelhető  $(X_t)$  folyamat határeloszlása ismert.

A folyamat egyéb tulajdonságait fogalmazza meg a következő állítás (melyeket a megtalálhatók [1], [16] és [25] cikkekben is).

**Állítás 3.5 (INAR(1) folyamat tulajdonságai)**

1.  $N_t$  stacionaritásához  $E(N_1) = \mu_\varepsilon / (1 - \alpha)$  szükséges.
2.  $D^2(N) = (\alpha\mu_\varepsilon + \sigma_\varepsilon^2) / (1 - \alpha^2)$ .
3. Az  $N_t$  kovariancia struktúrája: minden nemnegatív egész  $k$  értékre, a  $k$ -val késleltetett  $\gamma(k)$  kovarianciára:

$$\begin{aligned} \gamma(k) &= \text{Cov}(N_{t-k}, N_t) = \\ &= \text{Cov}(N_{t-k}, \alpha^k \circ N_{t-k}) + \text{Cov}\left(N_{t-k}, \sum_{j=0}^{k-1} \alpha^j \varepsilon_{t-j}\right) \\ &= \alpha^k D^2(N_{t-k}) + \sum_{j=0}^{k-1} \alpha^j \text{Cov}(N_{t-k}, \varepsilon_{t-j}) = \alpha^k \gamma(0). \end{aligned}$$

4. Az előző pont szerint az autokorrelációs függvény  $\rho(k) = \frac{\gamma(k)}{\gamma(0)} = \alpha^k$  exponenciális sebességgel csökken (hiszen  $\alpha \in [0, 1]$ ).

5. A feltételes valószínűségekre teljesül a következő:

$$\begin{aligned} p(n_k | n_{k-1}) &= P(N_k = n_k | N_{k-1} = n_{k-1}) = \\ &= x! e^{-(1-\alpha)\lambda} \sum_{j=0}^m \frac{\alpha^k (1-\alpha)^{n_{k-1} + n_k - 2j} \lambda^{n_k - j}}{j! (n_{k-1} - j)! (n_k - j)!} \end{aligned}$$

$y = 0, 1, \dots$ , ahol  $m = \min(n_k, n_{k-1})$ .

6. Az együttes valószínűségekre

$$P(N_k = n_k, N_{k-1} = n_{k-1}) = e^{-(2-\alpha)\lambda} \sum_{j=0}^m \frac{\alpha^k (1-\alpha)^{n_{k-1}+n_k-2j} \lambda^{n_{k-1}+n_k-j}}{j!(n_{k-1}-j)!(n_k-j)!}.$$

7. A feltételes várható értékre:  $E(N_t | N_{t-1}) = \alpha \cdot N_{t-1} + \mu\varepsilon$ .

Ezen a ponton érdemes megjegyezni, hogy többfajta módon lehetséges az imént kapott modell további specifikálása. A határeloszlásokat vizsgálva a Poisson, a negatív binomiális és a geometriai eloszlások azok, amelyek szóba jönnek lehetőségekként, ugyanis ezek mindegyike rendelkezik a diszkrét önfelbonthatóság tulajdonságával. Mi elsősorban a Poisson határeloszlású eseteket vizsgáljuk majd tovább.

### 3.2.2. INAR(p) folyamatok

Az INAR( $p$ ) folyamatok esetében nem csak az egyes határeloszlások meghatározásakor van többféle lehetőségünk modellek felírására, azért mert a  $p = 1$  eset nincs (nem volt) annyira specifikálva, hogy abból egyenesen következzen egy általános eset. Míg az INAR(1) modellek esetében egy időszakra tekintettünk vissza, így - az ottani szemléletet és motivációt követve - a  $t$ -beli károkozók száma egyrészt azokból az egy időszakkal korábbi károkozókból tevődik ki, akik ismét kárt okoznak, valamint az új károkozókból. A korábbi időszakból megmaradó károkozókat ritkítással kapjuk. Ez hasonlóképpen van az INAR( $p$ ) modellek esetében is, azonban a modellek általában abban különböznek egymástól, hogy milyen valószínűségekkel ritkítjuk a korábbi időszakokat azaz, hogy az egyes kárt okozókat meddig tartjuk gyanú alatt, vagy másképpen, hogy meddig adunk annak pozitív valószínűséget, hogy az adott egyén még okozhat kárt.

Három különböző modellt fogunk vizsgálni. Az első az Alzaid, Al-Osh szerzőpárosához [2] cikkén alapszik, vizsgáljuk továbbá Du és Li [7] cikkében vázolt módszert, valamint Biswas és Song [4] úgynevezett *kevert* modelljét is. Az első két modellben az alap rekurzív dinamika megegyezik, a legfőbb különbség a

különböző ritkítási mechanizmuson alapszik. Míg a [2] bizonyos vonatkozásban direkt kiterjesztése az INAR(1) modellnek, addig [7] cikkbeli kiterjesztés inkább a lineáris magasabb rendű Gauss AR-folyamatok kiterjesztésének az analógiájára hasonlít.

### Multinomiális INAR modell

#### Definíció 3.4 (INAR(p) modell)

Egy diszkrét idejű nemnegatív egészértékű  $\{N_t\}$  sztochasztikus folyamat INAR(p) folyamat, ha kielégíti a következő egyenletet

$$\begin{aligned} N_t &= \alpha_1 \circ N_{t-1} + \alpha_2 \circ N_{t-2} + \dots + \alpha_p \circ N_{t-p} + \varepsilon_t = \\ &= \sum_{i=1}^p \alpha_i \circ N_{t-i} + \varepsilon_t \end{aligned} \quad (3.4)$$

ahol

- $\{\varepsilon_t\}$  független, azonos eloszlású, nemnegatív egészértékű valószínűségi változó sorozata, melyre a következő jelöléseket vezetjük be:  $E(\varepsilon_t) = \mu_\varepsilon$ ,  $D^2(\varepsilon_t) = \sigma_\varepsilon^2$ ,  $E(\varepsilon_t^3) = \gamma_\varepsilon$  és  $E(\varepsilon_t^4) = \kappa_\varepsilon$ ;
- $A \circ$  operátor a 3.1-ben definiált Steutel–van Harn binomiális ritkító operátor minden  $\alpha_i$ -re, azaz  $\alpha_i \circ N_{t-i} = \sum_{j=1}^{N_{t-i}} Y_{i,j}$   $i = 1, \dots, p$ , ahol  $Y_{i,j}$ , az úgynevezett számláló folyamat, független, azonos eloszlású, nemnegatív egészértékű valószínűségi változó sorozata, melyre a következő jelöléseket vezetjük be:  $E(Y_{i,j}) = \alpha_i$ ,  $D^2(Y_{i,j}) = \sigma_i^2$ ,  $E(Y_{i,j}^3) = \gamma_i$  és  $E(Y_{i,j}^4) = \kappa_i$ ;
- $0 \leq \alpha_i < 1$ ,  $i = 1, \dots, p$ ,  $0 < \alpha_p < 1$  úgy, hogy  $\sum_{i=1}^p \alpha_i < 1$ .
- A feltételes eloszlását az  $(\alpha_1 \circ N_t, \alpha_2 \circ N_t, \dots, \alpha_p \circ N_t)$  az  $N_t = n_t$  feltétel mellett. Ez a feltételes eloszlás multinomiális  $(\alpha_1, \alpha_2, \dots, \alpha_p, n_t)$  paraméterekkel és független a folyamat múltjától. Azaz azt mondhatjuk, hogy adott  $N_t = n_t$  értékre az  $\alpha_i \circ N_t$  valószínűségi változó független  $N_{t-k}$ -től és annak „túlélőitől”  $\alpha_j \circ N_{t-k}$ , ahol  $i, j = 1, 2, \dots, p$  és  $k > 0$ .

Az utolsó deklaráció adja meg a modellben a legfőbb különbséget a többitől. Személetesen a következőképpen képzelhetjük el a korábbi nézőpontba helyezve. Ha valaki kárt okozóvá válik, akkor utána azt sorsolja ki az adott  $\alpha$



paraméter, hogy 1,2 vagy  $\dots p$  ciklus múlva válik ismét károkozóvá az illető, vagy esetleg már nem lesz az soha többé. Azaz maximum kétszer okozhat kárt egy adott illető az egész folymat során. Annak a valószínűsége, hogy egy egyén ismét kárt okoz az  $i$ -ik rákövetkező periódusában  $\alpha_i$ .<sup>6</sup> Tehát mindenki esetében eldöntjük, hogy hány ciklus múlva okoz ismét kárt (ebbe beleértve azt az esetet is, amikor többé már nem hibázik), ezért nevezhetjük multinomiális modellnek. Az  $\varepsilon_t$  pedig itt is a migránsokat, azaz az új károkozókat jelentik, akik a  $(t-1, t]$  időintervallumban kárt okoztak. Ekkor a  $t$ -ik generáció leírható  $N_t$  folyamattal. Megjegyezzük, hogy ha  $\sum_{i=1}^p \alpha_i \geq 1$ , akkor  $N_t$  kirobbanó folyamat.

### Binomiális INAR modell

A definíció hasonló, mint az előző esetben, azonban itt az utolsó feltételes eloszlásra vonatkozó kitétel nincs meghagyva. Ekkor a modellre szemléletesen úgy tekinthetünk, hogy  $t$ -ben tudom azt, hogy kik okoztak kárt az előző  $p$  időszakban és minden korábbi időszak hibázóinál újra eldöntöm, hogy most  $t$ -ben fognak-e hibázni, vagy sem (innen a binomiális elnevezés - mindenki esetében vagy megint okoz kárt, vagy nem). Elképzelhető az az eset is, hogy valaki mindig kárt okoz.

Az utolsó modell tárgyalása előtt kimondunk egy tételt, amit Latour bizonyított [13] cikkében és az INAR( $p$ ) folyamatok létezéséről szól.

#### Tétel 3.6

*Legyen  $p$  olyan, hogy  $\sum_{i=1}^p \alpha_i < 1$ , és legyen továbbá adva  $(\varepsilon_t)$  független, azonos eloszlású nemnegatív egész értékű véletlen változók. Ekkor létezik egy egészértékű stacionárius  $(N_t)$  folyamat, melyre  $N_t = \sum_{i=1}^p \alpha_i \circ N_{t-i} + \varepsilon_t$  és  $Cov(N_s, \varepsilon_t) = 0$ ,  $s < t$ .*

Megjegyezzük, hogy a  $\sum_{i=1}^p \alpha_i < 1$  feltétel ekvivalens [7] cikkben felhasznált karakterisztikus polinomra és annak gyökeire vonatkozó feltétellel, miszerint

<sup>6</sup>A folyamatra ha populációs folyamatként gondolunk, akkor szemléletesen úgy képzelhetjük el, hogy egy adott emberi populációban csak a női egyedek képesek reprodukcióra és ők életük során összesen egyszer tudnak női utódot létrehozni. Az INAR(1) folyamatok esetében ez úgy mondható, hogy csak az utolsó generáció képes a reprodukcióra.

akkor stacionárius egy ilyen idősor, ha a  $\sum_{i=1}^p \alpha_i z^{p-i}$  polinom gyökei az egységkörön belül esnek<sup>7</sup>.

### Kevert INAR modell

Pegram [19] cikkében bevezetett egészértékű AR( $p$ ) folyamatoknak egy osztályát, mely folyamatok felírásához definiált egy új operátort. Az alábbi definíciót és modellt Biswas és Song [4] cikke alapján tárgyaljuk.

#### Definíció 3.5 (A \* operátor)

Legyen  $U$  és  $V$  két független egész értékű valószínűségi változó és legyen adott egy véletlen  $\phi \in (0, 1)$  együttható. A Pegram-féle \* keverő operátor a következő  $Z$  véletlen változót eredményezi:

$$Z = (U, \phi) * (V, 1 - \phi)$$

melynek eloszlása

$$P(Z = j) = \phi P(U = j) + (1 - \phi)P(V = j) \quad j = 0, 1, \dots$$

Ez az operátor minden olyan  $D(p_0, p_1, \dots)$  diszkrét eloszlásra működik, ahol  $X \sim D(p_0, p_1, \dots)$  jelölés azt jelenti, hogy  $X$  értelmezési tartománya  $\{0, 1, \dots\}$  és  $P(X = i) = p_i$ ,  $i = 0, 1, \dots$ . Ez a \* operátor lényegében két diszkrét eloszlás keverékét eredményezi, az adott  $\phi$  és  $(1 - \phi)$  keverő súlyokkal.

Tegyük fel, hogy az  $N_t$  egy egészértékű idősor, melyre  $N_t \sim D(p_0, p_1, \dots)$ . Tegyük fel továbbá, hogy  $\varepsilon_t$  változókra is igaz, hogy független, azonos eloszlásúak úgy, hogy  $\varepsilon_t \sim D(p_0, p_1, \dots)$ . Jelölje most ebben a részben  $\mu = E(N_t)$  és  $\sigma^2 = D^2(\varepsilon_t)$ .

#### Definíció 3.6 (Kevert INAR( $p$ ))

Legyen  $(N_t)$  olyan egészértékű sztochasztikus folyamat, melyre:

$$N_t = (I(N_{t-1}), \phi_1) \dots * (I(N_{t-p}), \phi_p) * (\varepsilon_t, 1 - \phi_1 - \phi_2 - \dots - \phi_p) \quad (3.5)$$

ami  $p + 1$  darab diszkrét eloszlás keveréke az adott keverő súlyokkal, ahol  $I(\bullet)$  jelöli az indikátor eloszlást. A keverő súlyokra fennáll, hogy  $\phi_j \in (0, 1)$ ,

<sup>7</sup>Ezzel ekvivalens, hogy a  $\sum_{i=1}^p \alpha_i z^i = 0$  egyenlet megoldásai az egységkörön kívülre esnek.

$j = 0, 1, \dots, p$  és  $\sum_{j=1}^p \phi_j \in (0, 1)$ . Ez azt eredményezi, hogy a feltételes valószínűségekre fennáll a következő összefüggés:

$$\begin{aligned} P(N_t = j | N_{t-1}, N_{t-2}, \dots) &= P(N_t = j | N_{t-1}, N_{t-2}, \dots, N_{t-p}) = \\ &= (1 - \phi_1 - \phi_2 - \dots - \phi_p)p_j + \phi_1 I(N_{t-1} = j) + \dots + \phi_p I(N_{t-p} = j), \end{aligned}$$

ahol a  $\phi_j$ ,  $j = 1, 2, \dots$  értékeket úgy választjuk meg, hogy az  $1 - \phi_1 z - \phi_2 z^2 - \dots - \phi_p z^p$  gyökei az egységkörön kívül esenek.

A modellre szemléletesen kicsit bonyolultabb rá gondolni, de hasonlótt mondni, mint a korábbiak, csak az eloszlások tekintetében. Ha ismerem az előző értékeket, akkor mi lesz a mostani értékem eloszlása. E szerint az eloszlás szerint fogunk sorsol egy értéket az új tagra. A múlt szerinti feltételes eloszlást veszi (azok egy keverék eloszlását) úgy, hogy az eloszlás  $1 - \sum_j \phi_j$  valószínűséggel a zaj eloszlása lesz,  $\phi_j$  valószínűséggel a korábbi  $N_{t-j}$  eloszlása. A következő példán jól lehet érzékeltetni, hogy mi is történik ebben a modellben. Legyen  $p = 3$ . Ekkor, ha  $N_t$  értékét akarjuk meghatározni, akkor ahhoz először kell lenni fog az eloszlás, ami szerint ezt az értéket meg fogjuk határozni. Legyen  $N_{t-3}, N_{t-2}, N_{t-1}$  értékei rendre  $j_1, j_2, j_3$ . Ekkor  $N_t$  eloszlását a következőképpen kapjuk:

$$N_t \sim \phi_0 \cdot p_j + \begin{cases} 0 & j \notin \{j_1, j_2, j_3\} \\ \phi_1 & j = j_1 \\ \phi_2 & j = j_2 \\ \phi_3 & j = j_3 \end{cases}$$

Így egy eloszláshoz jutunk, mert  $q_j$ -vel jelöljük az  $N_t$  eloszlását, akkor  $\sum q_j = \phi_0 \cdot \sum p_j + \phi_1 + \phi_2 + \phi_3 = 1$ .

Legfontosabb tulajdonságait foglalja össze a kevert INAR folyamatoknak a következő állítás.

### Állítás 3.7 (Keverett INAR( $p$ ) folyamatok tulajdonságai)

1.  $N_t$  stacionárius.

2. Az  $N_t$  feltételes várható értékekre ismert  $N_{t-p}$  értékek mellett:

$$E(N_t | N_{t-1}, \dots, N_{t-p}) = (1 - \phi_1 - \phi_2 - \dots - \phi_p)\mu + \phi_1 N_{t-1} + \dots + \phi_p N_{t-p}.$$

3. Autokovariancia függvénye (ACVF):

$$\gamma(k) = \text{Cov}(N_t, N_{t-k}) = \phi_1 \gamma(k-1) + \dots + \phi_p \gamma(k-p).$$

4. Autokorreláció függvénye (ACF):

$$\rho(k) = \phi_1 \rho(k-1) + \dots + \phi_p \rho(k-p).$$

Az ACF függvény megegyezik a Box-Jenkins-féle  $AR(p)$  modell ACF-jével, így  $\rho(k) = \text{Corr}(N_t, N_{t-k}) = \rho^{|k|}$ .

A három különböző modellt felhasználjuk a Csődvalószínűség fejezetben.

## 4. fejezet

# Poisson INAR folyamatok

Többször említettük, hogy a 3.1-ben definiált folyamat határeloszlását tekintve megkülönböztetjük az INAR folyamatokat. Most bővebben azt az esetet vizsgáljuk, amikor a határeloszlás Poisson. Megjegyezzük azonban, hogy negatív binomiális és geometriai határeloszlásokat is vizsgáltak. Ezekhez kapcsolódóan az egyik alapcikk McKenzie 1986-os [15] írása, továbbá a [25] és [12] cikkekben is találunk negatív binomiális és geometriai eloszlásokkal foglalkozó részeket.

### 4.1. Poisson INAR(1)

#### **Definíció 4.1 (Poisson INAR(1) folyamat)**

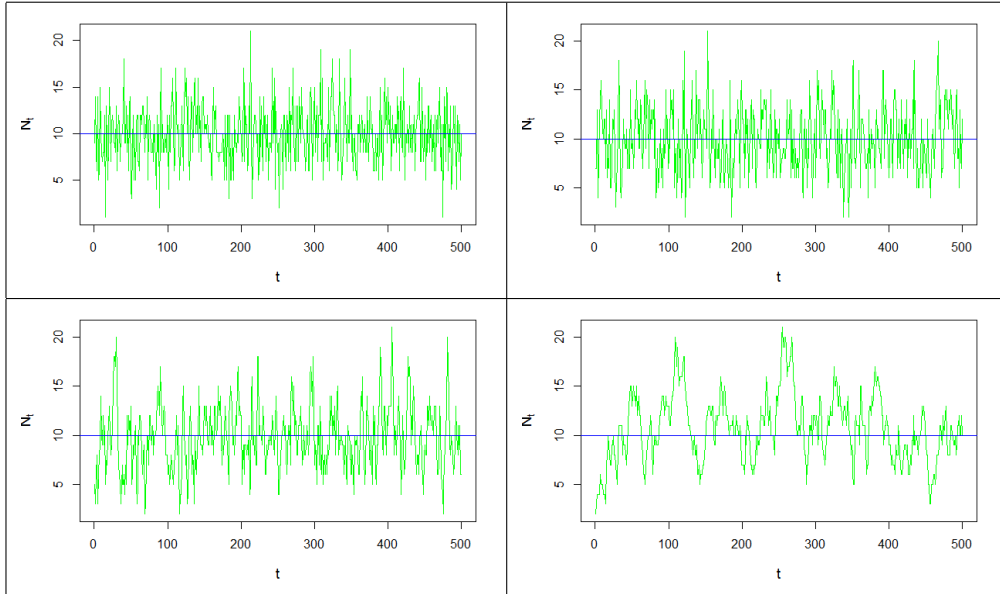
*Legyen  $(N_t)_{t>1}$  egy nemnegatív egész értékű valószínűségi változókból álló sorozat és  $\alpha \in [0, 1]$  pedig egy valós konstans. Ekkor a Poisson INAR(1) folyamatot a következőképpen definiáljuk:*

$$N_t = \alpha \circ N_{t-1} + \varepsilon_t \quad (4.1)$$

*ahol  $N_1 \sim \text{Poisson}(\lambda)$ ,  $\varepsilon_t$  független, azonos eloszlású, nemnegatív egészértékű Poisson eloszlású valószínűségi változó  $((1 - \alpha)\lambda)$  paraméterrel minden  $t$ -re.*

Nézzük meg, hogy hogyan néznek ki a Poisson INAR(1) folyamatok különböző trajektóriái különböző  $\alpha$  értékekre. Az ábrákon rendre az  $\alpha = 0$ ,  $\alpha = 0,3$ ,  $\alpha = 0,6$ ,  $\alpha = 0,9$  esetek láthatók. A folyamatot mindegyik esetben az 500.

tagig generáltuk úgy, hogy kielégítsék a fent definiált eloszlási feltételeket. A  $\lambda$  paramétert 10-nek választottuk.



Az  $\alpha = 0$  esetben látszik a függetlenség, míg ahogy növeltük a ritkítás valószínűségét úgy növekedett a korábbi tagtól való függés. Egyfajta ritkulás látszik a folyamatban.

Az autoregresszív dinamikának a függőségi struktúrájára INAR(1) folyamatok esetében explicit zárt képlet adható. A függőségi struktúra [6] cikkből származik, de a bizonyítását pontosabban levezetjük.

#### Állítás 4.1

A (4.1) autoregresszív dinamikával rendelkező Poisson INAR(1) folyamat függőségi struktúrája a következőképpen reprezentálható:

$$\begin{aligned}
N_2 &= \sum_{i=1}^{N_1} Y_{2,1,i} + \varepsilon_2, \\
N_3 &= \sum_{i=1}^{N_1} Y_{2,1,i} Y_{3,1,i} + \sum_{i=1}^{\varepsilon_2} Y_{3,2,i} + \varepsilon_3, \\
&\vdots \\
N_k &= \sum_{i=1}^{N_1} Y_{2,1,i} Y_{3,1,i} \cdots Y_{k,1,i} + \sum_{j=2}^{k-1} \sum_{i=1}^{\varepsilon_j} \prod_{l=j+1}^k Y_{l,j,i} + \varepsilon_k \quad (k = 3, 4, \dots).
\end{aligned} \tag{4.2}$$

**Bizonyítás** A cél, hogy az általános függőségi struktúrát levezessük  $N_k$ -ra Poisson INAR(1) folyamat esetén. A binomiális ritkítás (3.1) definícióját és a rá vonatkozó (3.1) alapösszefüggéseket használjuk fel. A folyamat definíciója miatt a ritkítás különböző korábbi időszakokat is érinteni fog, emiatt szükséges több index egyidejű futtatása. Kezdjük el felírni az egyes  $N_k$  értékeket. Feltesszük, hogy  $N_1 \sim \text{Poisson}(\lambda)$ .

$$\begin{aligned}
N_2 &= \alpha \circ N_1 + \varepsilon_2 = \sum_{i=1}^{N_1} Y_{2,1,i} + \varepsilon_2 \\
N_3 &= \alpha \circ [\alpha \circ N_1 + \varepsilon_2] + \varepsilon_3 = \alpha \circ (\alpha \circ N_1) + \alpha \circ \varepsilon_2 + \varepsilon_3
\end{aligned}$$

Vizsgáljuk külön az első tagot.

$$\begin{aligned}
\alpha \circ (\alpha \circ N_1) &= \alpha \circ \left( \sum_{i=1}^{N_1} Y_{2,1,i} \right) = \alpha \circ (Y_{2,1,1} + Y_{2,1,2} + \dots + Y_{2,1,N_1}) = \\
&= \alpha \circ Y_{2,1,1} + \alpha \circ Y_{2,1,2} + \dots + \alpha \circ Y_{2,1,N_1} = \\
&= \sum_{j=1}^{Y_{2,1,1}} Y_{3,1,1,j} + \sum_{j=1}^{Y_{2,1,2}} Y_{3,1,2,j} + \dots + \sum_{j=1}^{Y_{2,1,N_1}} Y_{3,1,N_1,j} = \\
&= Y_{2,1,1} \cdot Y_{3,1,1} + Y_{2,1,2} \cdot Y_{3,1,2} + \dots + Y_{2,1,N_1} \cdot Y_{3,1,N_1} = \\
&= \sum_{i=1}^{N_1} Y_{2,1,i} \cdot Y_{3,1,i}
\end{aligned}$$

Ezzel  $N_3$  karakterisztikája a következő:

$$N_3 = \sum_{i=1}^{N_1} Y_{2,1,i} \cdot Y_{3,1,i} + \sum_{i=1}^{\varepsilon_2} Y_{3,2,i} + \varepsilon_3$$

Írjuk fel hasonlóképpen  $N_4$ -et is:

$$\begin{aligned}
N_4 &= \alpha \circ N_3 + \varepsilon_4 = \alpha \circ [\alpha \circ N_2 + \varepsilon_3] + \varepsilon_4 = \\
&= \alpha \circ [\alpha \circ (\alpha \circ N_1 + \varepsilon_2) + \varepsilon_3] + \varepsilon_4 = \\
&= \alpha \circ [\alpha \circ (\alpha \circ N_1) + \alpha \circ \varepsilon_2 + \varepsilon_3] + \varepsilon_4 = \\
&= \alpha \circ [\alpha \circ (\alpha \circ N_1)] + \alpha \circ [\alpha \circ \varepsilon_2] + \alpha \circ \varepsilon_3 + \varepsilon_4
\end{aligned}$$

Vizsgáljuk ismét tagonként az összeget:

$$\begin{aligned}
\alpha \circ [\alpha \circ (\alpha \circ N_1)] &= \dots = \sum_{i=1}^{N_1} Y_{2,1,i} \cdot Y_{3,1,i} \cdot Y_{4,1,i} \\
\alpha \circ [\alpha \circ \varepsilon_2] &= \alpha \circ \left( \sum_{i=1}^{\varepsilon_2} Y_{3,2,i} \right) = \alpha \circ (Y_{3,2,1} + Y_{3,2,2} + \dots + Y_{3,2,\varepsilon_2}) = \\
&= \alpha \circ Y_{3,2,1} + \alpha \circ Y_{3,2,2} + \dots + \alpha \circ Y_{3,2,\varepsilon_2} = \\
&= \sum_{i=1}^{Y_{3,2,1}} Y_{4,2,1,i} + \sum_{i=1}^{Y_{3,2,2}} Y_{4,2,2,i} + \dots + \sum_{i=1}^{Y_{3,2,\varepsilon_2}} Y_{4,2,\varepsilon_2,i} = \\
&= \sum_{i=1}^{\varepsilon_2} Y_{3,2,i} \cdot Y_{4,2,i} \\
\alpha \circ \varepsilon_3 &= \sum_{i=1}^{\varepsilon_3} Y_{4,3,i}
\end{aligned}$$

Ezzel  $N_4$  karakterisztikája a következő:

$$N_4 = \sum_{i=1}^{N_1} Y_{2,1,i} \cdot Y_{3,1,i} + \sum_{i=1}^{\varepsilon_2} Y_{3,2,i} Y_{4,2,i} + \sum_{i=1}^{\varepsilon_3} Y_{4,3,i} + \varepsilon_4$$

Ezt az eljárást iterálva kapjuk, hogy:

$$N_k = \sum_{i=1}^{N_1} Y_{2,1,i} Y_{3,1,i} \dots Y_{k,1,i} + \sum_{j=2}^{k-1} \sum_{i=1}^{\varepsilon_j} \prod_{l=j+1}^k Y_{l,j,i} + \varepsilon_k \quad (k = 3, 4, \dots)$$

Az indexekre szemléletesen a következőféleképpen gondolhatunk: az  $Y_{k,1,i}$  esetében a  $k$  index mindig azt mondja meg, hogy az adott  $Y$  érték melyik  $N_k$  iterációnál került be, a második index pedig 1, ami azt jelzi, hogy ha elvégezzük a visszafejtést, akkor az a kiinduló  $N_1$  tagra vonatkozik (az  $i$  pedig nyilván csak egy futó index). Az  $Y_{l,j}$  esetében (ahol  $j$  2-től,  $l$  pedig  $(j+1)$ -től indul) pedig hasonlóan azt mutatja a 2., azaz a  $j$  index, hogy az adott változó az  $\varepsilon_j$ -összegzésben van benne, arra vonatkozik, valamint az  $l$  pedig az mutatja, hogy az adott változó melyik iterációnál került be.  $\square$



Az eloszlásokra vonatkozó összefüggéseket korábban az INAR(1) folyamatoknál, általánosabban láttuk.

## 4.2. Csődvalószínűség

Az alábbi tétel azt mutatja be, hogy hogyan módosul a csődvalószínűsége adott képlet a klasszikus rizikófolyamatokhoz képest, ahol nem volt összefüggőség feltételezve a kárszámok folyamatában. Most feltesszük, hogy a kárszámok Poisson INAR(1) folyamatot követnek.<sup>1</sup>

### Tétel 4.2

Tegyük fel, hogy  $\alpha M_B(r) < 1$  és legyen  $\gamma = 1 - \alpha$ . Ekkor a  $c(r)$ -re adott kifejezés:

$$c(r) = \frac{(1 - \alpha)^2 \lambda M_B(r)}{1 - \alpha M_B(r)} - (1 - \alpha)\lambda - r\pi = \frac{\gamma^2 \lambda M_B(r)}{1 - \alpha M_B(r)} - \gamma\lambda - r\pi \quad (4.3)$$

**Bizonyítás** Vezessük be a következő jelöléseket az  $n$  időpontig bekövetkezett összkárra:

$$S_n = W_1 + \dots + W_n = \sum_{j=1}^{L_n} C_j \quad \text{ahol } L_n = \sum_{j=1}^{N_n} N_j$$

és  $C_1, C_2, \dots$  független, azonos eloszlású valószínűségi változók, amik ugyanolyan eloszlásúak, mint  $B$ .

$$\begin{aligned} M_{S_n}(r) &= G_{N_1, \dots, N_n}(M_B(r), \dots, M_B(r)) = \\ &= E[M_B(r)^{N_1} \dots M_B(r)^{N_n}] = \\ &= E[M_B(r)^{N_1 \dots N_n}] = \\ &= E[M_B(r)^{L_n}] = \\ &= G_{L_n}(M_B(r)) \end{aligned}$$

Tehát ki kell fejeznünk  $G_{L_n}(M_B(r))$ -t. Ehhez vizsgáljuk meg a viselkedését az első négy perióduson ( $n = 1, 2, 3, 4$ ) keresztül. A következőkben legyen  $\gamma = 1 - \alpha$ .  $L_n$  generátorfüggvénye:  $G_{L_n}(t) = E[t^{N_1 + \dots + N_n}] = E[t^{N_1} + \dots + t^{N_n}]$ .

<sup>1</sup>A tétel és annak bizonyítása a [6] cikkben leírtak szerint.

$n = 1$  eset:

$$G_{L_n}(t) = E[t^{N_1}] = e^{\lambda(t-1)}$$

$n = 2$  eset:

$$\begin{aligned} G_{L_n}(t) &= E[t^{N_1}t^{N_2}] = E[t^{N_1}t^{(\sum_{i=1}^{N_1} Y_{2,1,i})}t^{\varepsilon_2}] = \\ &= e^{\lambda((1-\alpha)t+\alpha t^2-1)}e^{(1-\alpha)\lambda(t-1)} \end{aligned}$$

$n = 3$  eset:

$$\begin{aligned} G_{L_n}(t) &= E[t^{N_1}t^{N_2}t^{N_3}] = E[t^{N_1}t^{(\sum_{i=1}^{N_1} Y_{2,1,i}+\sum_{i=1}^{N_1} Y_{2,1,i}Y_{3,1,i})}t^{\varepsilon_2}t^{(\sum_{i=1}^{\varepsilon_2} Y_{3,2,i})}t^{\varepsilon_3}] = \\ &= E[t^{N_1}t^{(\sum_{i=1}^{N_1} Y_{2,1,i}+\sum_{i=1}^{N_1} Y_{2,1,i}Y_{3,1,i})}]E[t^{\varepsilon_2}t^{(\sum_{i=1}^{\varepsilon_2} Y_{3,2,i})}]E[t^{\varepsilon_3}] = \\ &= e^{\lambda((\gamma^2+\alpha\gamma)t+\alpha\gamma t^2+\alpha^2 t^3-1)}e^{\gamma\lambda(\gamma t+\alpha t^2-1)}e^{\gamma\lambda(t-1)} \end{aligned}$$

$n = 4$  eset:

$$\begin{aligned} G_{L_n}(t) &= E[t^{N_1}t^{N_2}t^{N_3}t^{N_4}] = \\ &= E\left[t^{N_1}t^{(\sum_{i=1}^{N_1} Y_{2,1,i}+\sum_{i=1}^{N_1} Y_{2,1,i}Y_{3,1,i}+\sum_{i=1}^{N_1} Y_{2,1,i}Y_{3,1,i}Y_{4,1,i})}\right] \cdot \\ &\quad \cdot E\left[t^{(\varepsilon_2+\sum_{i=1}^{\varepsilon_2} Y_{3,2,i}+\sum_{i=1}^{\varepsilon_2} Y_{3,2,i}Y_{4,2,i})}\right]E\left[t^{\varepsilon_3+\sum_{i=1}^{\varepsilon_3} Y_{4,3,i}}\right]E\left[t^{\varepsilon_4}\right] = \\ &= E\left[t^{N_1}t^{(\sum_{i=1}^{N_1} Y_{2,1,i}+\sum_{i=1}^{N_1} Y_{2,1,i}Y_{3,1,i}+\sum_{i=1}^{N_1} Y_{2,1,i}Y_{3,1,i}Y_{4,1,i})}\right] \cdot \\ &\quad \cdot e^{\gamma\lambda((\gamma^2+\alpha\gamma)t+\alpha\gamma t^2+\alpha^2 t^3-1)}e^{\gamma\lambda(\gamma t+\alpha t^2-1)}e^{\gamma\lambda(t-1)} \end{aligned}$$

Az alábbi kifejezést kell tehát kiszámolnunk:

$$E\left[t^{N_1}t^{(\sum_{i=1}^{N_1} Y_{2,1,i}+\sum_{i=1}^{N_1} Y_{2,1,i}Y_{3,1,i}+\sum_{i=1}^{N_1} Y_{2,1,i}Y_{3,1,i}Y_{4,1,i})}\right] =$$

Amely kibontva:

$$\begin{aligned} &= E\left[t^{N_1}E\left[t^{\sum_{i=1}^{N_1} Y_{2,1,i}}E\left[t^{\sum_{i=1}^{N_1} Y_{2,1,i}Y_{3,1,i}}\right] \cdot \right. \right. \\ &\quad \left. \left. \cdot E\left[t^{\sum_{i=1}^{N_1} Y_{2,1,i}Y_{3,1,i}Y_{4,1,i}}\middle|N_1, Y_{2,1,i}, Y_{3,1,i}\right]\middle|N_1, Y_{2,1,i}\right]\middle|N_1\right] = \end{aligned}$$

$$\begin{aligned}
&= E \left[ t^{N_1} E \left[ t^{\sum_{i=1}^{N_1} Y_{2,1,i}} E \left[ t^{\sum_{i=1}^{N_1} Y_{2,1,i} Y_{3,1,i}} \cdot \right. \right. \right. \\
&\quad \cdot \left. \prod_{i=1}^{N_1} \left( \gamma + \alpha t^{Y_{2,1,i} Y_{3,1,i}} \right) \middle| N_1, Y_{2,1,i} \right] \middle| N_1 \right] = \\
&= E \left[ t^{N_1} E \left[ t^{\sum_{i=1}^{N_1} Y_{2,1,i}} E \left[ \prod_{i=1}^{N_1} \left( \gamma t^{Y_{2,1,i} Y_{3,1,i}} + \alpha t^{2 \cdot Y_{2,1,i} Y_{3,1,i}} \right) \middle| N_1, Y_{2,1,i} \right] \middle| N_1 \right] \right] = \\
&= E \left[ t^{N_1} E \left[ t^{\sum_{i=1}^{N_1} Y_{2,1,i}} \prod_{i=1}^{N_1} \left( \gamma \left( \gamma + \alpha t^{Y_{2,1,i}} \right) + \alpha \left( \gamma + \alpha t^{2 \cdot Y_{2,1,i}} \right) \right) \middle| N_1 \right] \right] = \\
&= E \left[ t^{N_1} E \left[ \prod_{i=1}^{N_1} \left( \gamma t^{Y_{2,1,i}} \left( \gamma + \alpha t^{Y_{2,1,i}} \right) + \alpha t^{Y_{2,1,i}} \left( \gamma + \alpha t^{2 \cdot Y_{2,1,i}} \right) \right) \middle| N_1 \right] \right] = \\
&= E \left[ t^{N_1} E \left[ \prod_{i=1}^{N_1} \left( \gamma^2 t^{Y_{2,1,i}} + \alpha \gamma t^{2 \cdot Y_{2,1,i}} + \gamma \alpha t^{Y_{2,1,i}} + \alpha^2 t^{3 \cdot Y_{2,1,i}} \right) \middle| N_1 \right] \right] = \\
&= E \left[ t^{N_1} \prod_{i=1}^{N_1} \left( \gamma^2 (\gamma + \alpha t) + \alpha \gamma (\gamma + \alpha t^2) + \gamma \alpha (\gamma + \alpha t) + \alpha^2 (\gamma + \alpha t^3) \right) \right] = \\
&= E \left[ t^{N_1} \left( \gamma^2 (\gamma + \alpha t) + \alpha \gamma (\gamma + \alpha t^2) + \gamma \alpha (\gamma + \alpha t) + \alpha^2 (\gamma + \alpha t^3) \right)^{N_1} \right] = \\
&= e^{\lambda \left( (\gamma^3 + 2\alpha\gamma^2 + \alpha^2\gamma)t + (\alpha\gamma^2 + \alpha^2\gamma)t^2 + \alpha^2\gamma t^3 + \alpha^3 t^4 - 1 \right)}.
\end{aligned}$$

Tehát

$$\begin{aligned}
G_{L_n}(t) &= E[t^{N_1} t^{N_2} t^{N_3} t^{N_4}] = \\
&= e^{\lambda \left( (\gamma^3 + 2\alpha\gamma^2 + \alpha^2\gamma)t + (\alpha\gamma^2 + \alpha^2\gamma)t^2 + \alpha^2\gamma t^3 + \alpha^3 t^4 - 1 \right)} \cdot \\
&\quad \cdot e^{\gamma\lambda \left( (\gamma^2 + \alpha\gamma)t + \alpha\gamma t^2 + \alpha^2 t^3 - 1 \right)} e^{\gamma\lambda(\gamma t + \alpha t^2 - 1)} e^{\gamma\lambda(t-1)}
\end{aligned}$$

Ezzel ki tudjuk számolni  $S_4$  momentumgeneráló függvényét az  $r$  helyen:

$$\begin{aligned}
M_{S_4}(r) &= G_{L_n}(M_B(r)) = \\
&= e^{\lambda \left( (\gamma^3 + 2\alpha\gamma^2 + \alpha^2\gamma)M_B(r) + (\alpha\gamma^2 + \alpha^2\gamma)M_B(r)^2 + \alpha^2\gamma M_B(r)^3 + \alpha^3 M_B(r)^4 - 1 \right)} \cdot \\
&\quad \cdot e^{\gamma\lambda \left( (\gamma^2 + \alpha\gamma)M_B(r) + \alpha\gamma M_B(r)^2 + \alpha^2 M_B(r)^3 - 1 \right)} e^{\gamma\lambda(\gamma M_B(r) + \alpha M_B(r)^2 - 1)} \cdot \\
&\quad \cdot e^{\gamma\lambda(M_B(r)-1)}
\end{aligned}$$

Vegyük észre, hogy

$$\begin{aligned}(\gamma^2 + \alpha\gamma) &= \gamma(\gamma + \alpha) = \gamma \\(\alpha\gamma^2 + \alpha^2\gamma) &= \gamma\alpha(\gamma + \alpha) = \alpha\gamma \\(\gamma^3 + 2\alpha\gamma^2 + \alpha^2\gamma) &= \gamma(\gamma + \alpha)^2 = \gamma \\&\text{stb.}\end{aligned}$$

Következésképpen a következő általános formulát kapjuk  $M_{S_n}$ -re  $n = 2, 3, \dots$  esetén:

$$\begin{aligned}M_{S_n}(r) &= G_{L_n}(M_B(r)) = \\&= e^{\alpha\lambda(\gamma M_B(r) \sum_{k=0}^n (\alpha M_B(r))^k + \alpha^n (M_B(r))^{n-n})} \cdot \\&\cdot e^{\gamma\lambda(\gamma^n M_B(r) \sum_{k=0}^n (\alpha M_B(r))^k + M_B(r) \sum_{k=0}^n (\alpha M_B(r))^k - M_B(r) \sum_{k=0}^n (k+1)(\alpha M_B(r))^k - n)} =\end{aligned}$$

Legyen a

$$Z_\ell = \sum_{k=0}^n (\alpha M_B(r))^k = \frac{1 - (\alpha M_B(r))^{n+1}}{1 - \alpha M_B(r)}$$

és legyen

$$Z = \frac{-n(\alpha M_B(r))^{n-1}}{1 - \alpha M_B(r)}$$

Ezzel a jelöléssel:

$$= e^{\alpha\lambda(\gamma M_B(r) Z_1 + \alpha^n (M_B(r))^{n-n})} \cdot e^{\gamma\lambda(\gamma^n M_B(r) Z_1 + M_B(r) Z_1 - M_B(r) (-Z + Z_\ell) - n)}$$

Ebből meg tudjuk határozni  $c_n(r)$  értékét.

$$c_n(r) = \frac{1}{n} \ln \left( E(e^{r(S_n - \pi)}) \right)$$

Így behelyettesítve az imént  $S_n$ -re kapott összefüggést kapjuk, hogy a  $c_n(r)$  értéke:

$$\frac{\alpha\lambda(\gamma M_B(r) Z_1 - Z_0) + \gamma\lambda M_B(r)(Z_1(\gamma n + 1) - Z + Z_\ell) - n(\gamma\lambda - \pi r)}{n}$$

Ekkor, ha feltesszük, hogy  $\alpha M_B(r) < 1$  és vesszük a határértékét az előző

kifejezésnek, akkor megkapjuk a várt eredményt, miszerint:

$$c(r) = \lim_{n \rightarrow \infty} c_n(r) = \frac{\gamma^2 \lambda M_B(r)}{1 - \alpha M_B(r)} - \gamma \lambda - r \pi.$$

□

Feltettük, hogy a kárszámok Poisson INAR(1) folyamatot követnek, ebből következik, hogy a bizonyításban szereplő  $S_n$  összkár, összetett Poisson eloszlású. Ha feltételezzük, hogy a károk exponenciális eloszlásúak, akkor a Lundberg-kitevőre a következő zárt összefüggés adható. Megjegyezzük, hogy az  $\eta = \frac{\pi}{E(B)} - 1$  a relatív kockázati felárat jelentette.

#### Állítás 4.3 (Lundberg-kitevő speciális esetben)

Tegyük fel, hogy  $B \sim \text{Exp}(\beta)$ . Ekkor ha  $\alpha \in [0; 1)$ , akkor:

$$\rho = \frac{\gamma \beta \eta}{1 + \eta} = \frac{(1 - \alpha) \beta \eta}{1 + \eta}. \quad (4.4)$$

**Bizonyítás.** Tudjuk, hogy ha  $B \sim \text{Exp}(\beta)$ , akkor  $E(B) = \frac{1}{\beta}$  és  $M_B(r) = E(e^{rB}) = \frac{\beta}{\beta - r}$ . Ekkor ha behelyettesítünk 4.3 képletbe, akkor kapjuk, hogy:

$$\frac{\gamma^2 \lambda \frac{\beta}{\beta - r}}{1 - \alpha \frac{\beta}{\beta - r}} - \gamma \lambda - r \pi = 0,$$

ami átalakítva:

$$\begin{aligned} \frac{\lambda \gamma^2 \beta}{\gamma \beta - r} - r \lambda - r \pi &= \frac{\gamma^2 \beta}{\gamma \beta - r} - \gamma - r(1 + \eta)E(B) = 0 \\ \gamma^2 \beta - \gamma^2 \beta - \gamma(1 + \eta)r + r\gamma + r^2(1 + \eta)E(B) &= 0. \end{aligned}$$

Átrendezés után kapjuk a kívánt egyenletet. □

## 5. fejezet

# Adatelemzés

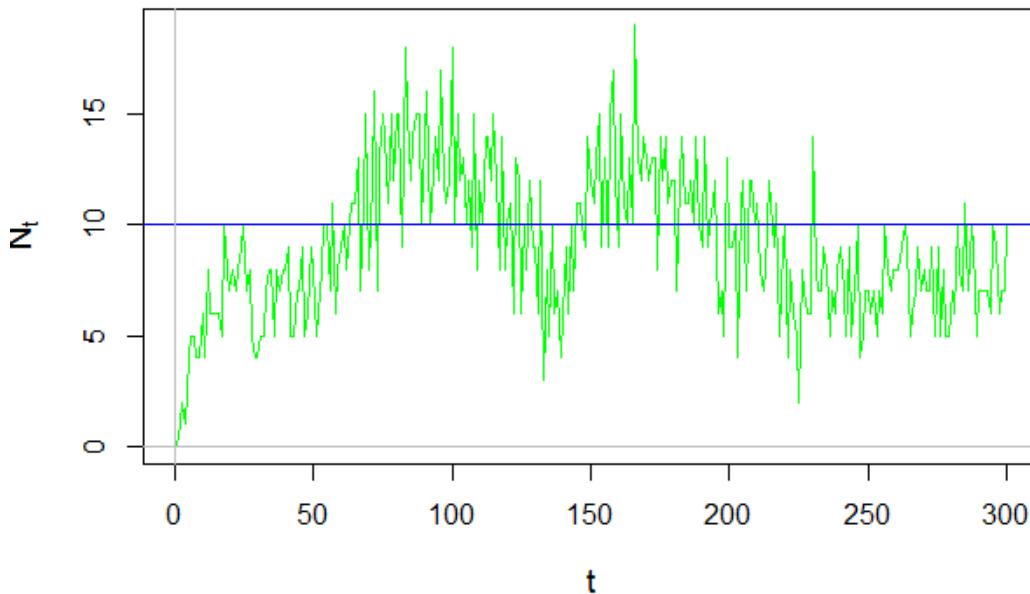
A INAR( $p$ ) folyamatok fejezetben foglalkoztunk az általános INAR( $p$ ) modellekkel, pontosabban három különböző folyamattal. Először ezekben a modellekben generálunk adatokat, majd a generált adatokon ellenőrizzük, hogy a Yule-Walker egyenletek megfelelően adják-e vissza az egyes paramétereket. Mindegyik esetben harmadrendű INAR folyamatokat generálunk. Az egyes programok **R** szoftverben készültek, mely kódok megtalálhatók a Függelék részben.

### Multinomiális modell

Elsőnek a Multinomiális INAR modell alfejezetben tárgyalt modellt vizsgáljuk. A modell értelmezése szerint a ritkítő operátorok azt mondják meg, hogy az adott  $N_{t-1}$ -ből hány ciklus múlva jelenik meg ismét egy kárt okozó. Először generáltunk INAR(3) folyamatot, amely tejesíti ezeket a feltételeket, továbbá feltételeztük, hogy a folyamat határeloszlása Poisson. Attól függően, hogy milyennek választjuk meg a  $\varepsilon_t$  paraméterét, attól függően fog változni a folyamat eloszlásának paramétere is. Korábban láttuk a folyamatok stacionaritására vonatkozó feltételeket, melyek a diszkrét önfelbonthatóággal voltak kapcsolatban (3.3). A különböző cikkekben másképp szokták megválasztani az egyes határeloszlásokat, van ahol  $N_t$  határeloszlását választják meg úgy, hogy  $Poisson(\lambda)$  legyen (mint ahogy az például a [6]-ben található, ekkor  $\alpha \circ N_{t-1} \sim Poisson(\alpha\lambda)$  a (3.1) állítás miatt, vala-

mint  $\varepsilon_t \sim \text{Poisson}((1-\alpha)\lambda)$ , vagy hogy  $\varepsilon_t$  eloszlását (például a [2] cikkben, ekkor  $\alpha \circ N_{t-1} \sim \text{Poisson}(\alpha \frac{\lambda}{1-\alpha})$  és  $N_t \sim \text{Poisson}(\frac{\lambda}{1-\alpha})$ ). Azonban az ön-felbonthatóság tulajdonságából következik, hogy lényegében egyik a másiktól számolható. Mi azt követjük most, hogy  $N_t$  eloszlását határozzuk meg,  $N_t \sim \text{Poisson}(\lambda)$ .

Először definiálunk egy függvényt (`eloszt`), amely azt határozza meg, hogy melyik hibázó mikor fog ismételten hibázni, majd generálunk az eloszlásnak megfelelő paraméterekkel rendelkező harmadrendű folyamatot. Ezután Yule-Walker egyenletekkel megpróbáljuk megbecsülni a paramétereket, hogy visszaadják-e az általunk beállított generált folyamat paramétereit. A folyamat egy realizációját mutatja az 5.1 ábra. Már a 300-as elemszám mellett is látszódik a folyamat stationaritása (a  $\lambda = 10$  paraméter mellett  $E(N_t) = 10$  nagy  $t$  értékekre, hiszen a határeloszlás  $\lambda$  paraméterű Poisson).



5.1. ábra. Multinomiális Poisson INAR(3) realizáció.  
 $t = 300$ ,  $\lambda = 10$ ,  $\alpha_1 = 0,4$ ,  $\alpha_2 = 0,3$ ,  $\alpha_3 = 0,2$

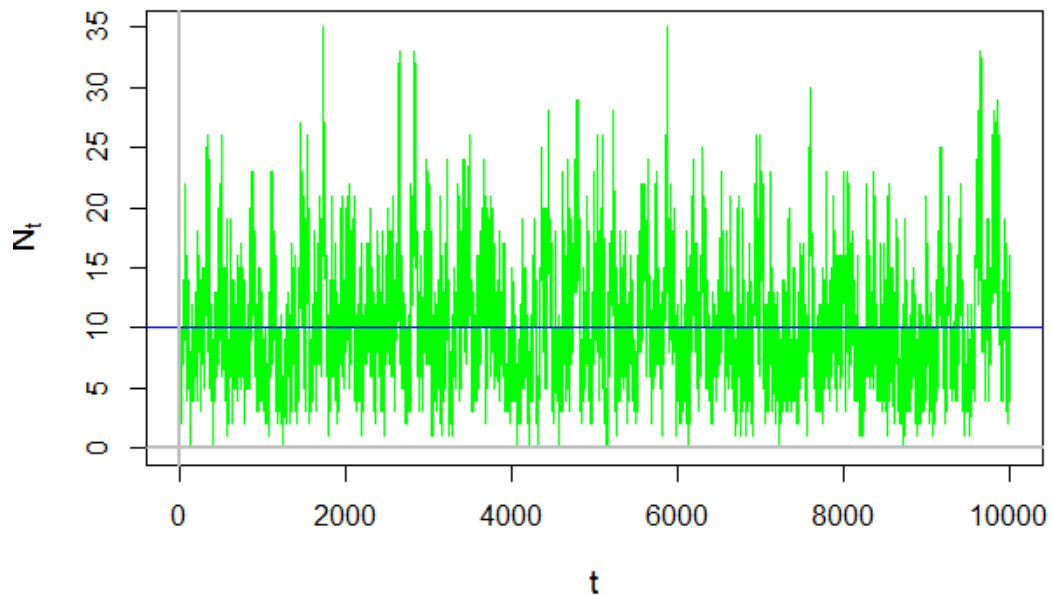
A Yule-Walker egyenleteket megoldva azt kapjuk, hogy nem adják jól vissza a paramétereket. Elvégezve 100 különböző generálást a folyamatra az eredmények azt mutatják, hogy nem megfelelő a becslés a paraméterekre. Az eredményeket a kö-

vetkező táblázat foglalja össze:

$\alpha_i$	0,4	0,3	0,2
átlag( $\hat{\alpha}_i$ )	0,2703986	0,2627153	0,2580167
szórás( $\hat{\alpha}_i$ )	0,0827736	0,1125946	0,0902973

## Binomiális modell

A szakasz 3.2.2 (3.2.2) részben bemutatott modellel is elvégezzük az imént bemutatott generálást és vizsgálatokat. A folyamat egy realizációja a következő:



5.2. ábra. Binomiális Poisson INAR(3) realizáció.  
 $t = 10000$ ,  $\lambda = 10$ ,  $\alpha_1 = 0,4$ ,  $\alpha_2 = 0,3$ ,  $\alpha_3 = 0,2$

A Yule-Walker egyenleteket megoldva nagyon jó becslést kapunk a paraméterekre. Az előzőhöz hasonlóan 100 futtatás eredményét a következő táblázatban foglaljuk össze:

$\alpha_i$	0,4	0,3	0,2
átlag( $\hat{\alpha}_i$ )	0,3987388	0,3021305	0,1978436
szórás( $\hat{\alpha}_i$ )	0,0099037	0,0102347	0,0088413



Ennek a modellnek a paramétereit már nagyon jól becslik a Yule-Walker egyenletek megoldásai. Ez után már érdemes lehet valamilyen módszerrel becslést adni a fokszámmra. A foksám becsléséhez [11] cikket használjuk fel, ami lényegben egy módosított Akaike kritérium minimumának a keresésével határozza meg a foksámot. Az eredeti Akaike információs kritérium (AIC) és annak módosított (AICc) változata:

$$\begin{aligned} AIC &= n(\log \widehat{\sigma}^2 + 1) + 2(m + 1) \\ AICc &= n \log \widehat{\sigma}^2 + n \frac{1 + m/n}{1 - (m + 2)/n} = \\ &= AIC + \frac{2(m + 1)(m + 2)}{n - m - 2} \end{aligned}$$

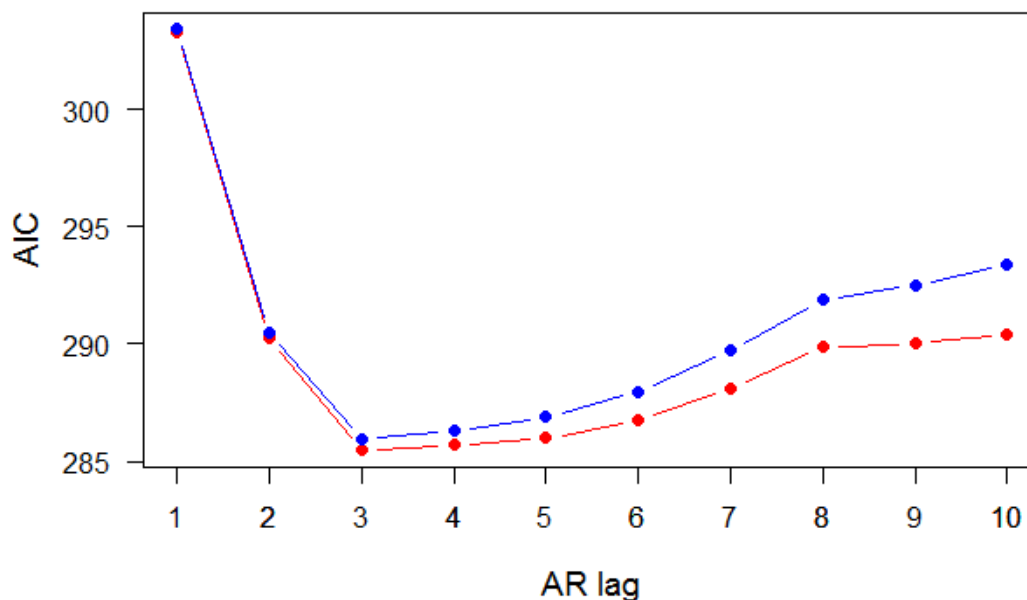
Az AIC kritérium egy relatív mérőszáma annak, hogy a választott modellünk mennyire jó. Ennek a mérésére a Kullback-Leibler távolságból szoktak kiindulni, ami két valószínűségi eloszlás különbözőségét méri. Diszkrét esetben szemléletesen ez azt jelenti, hogy az egyik a megfigyelésekből származó  $p_k$  értékek, a másik eloszlás azt mutatja, ahogy a világra gondolunk, azaz az elméleti (modell-beli) eloszlást adja meg ( $q_k$ ). Ezekből számolva a diszkrét Kullback-Leibler távolság:  $D_{KL}(P||Q) = \sum_k \log p_k \frac{p_k}{q_k}$ .<sup>1</sup> Míg az AIC torzított becslés erre a távolságra, addig az AICc már gyakori esetekben torzítatlanná válik erre. Egy olyan taggal módosítja a kritériumot, ami ne függ a konkrét adatoktól.

Tehát a módosított Akaike információs kritériummal megbecsüljük a (generált) adatsor rendjét úgy, hogy a AICc függvény minimumát vesszük, de megjelenítjük az AIC függvényt is. Az 5.3 ábrán látható, hogy a minimum visszaadja a rendet:

## Kevert modell

A bemutatott modell szerint az új rekurziós tagra, egy eloszlást választunk, ami szerint az új értéket ki fogjuk sorsolni. A modell trajektóriájának egy kimenetelét mutatja az 5.4 ábra. Hasonlóan az előzőekhez Itt is megvizsgáljuk a Yule-Walker egyenleteket, valamint több futtatásból megvizsgáljuk, hogy milyen pontossággal adja vissza a paramétereiket. Azt kapjuk az egyenletek jó közelítéssel visszaadják

<sup>1</sup>Valóban a várt és a tapasztalt értékek közötti különbséget adja meg, egy esélyhányados függvényének vesszük a várható értékét a megfigyelések szerint. Megj.: ha  $p_k/q_k=1$ , azaz ugyanaz következik be, mint amit vártunk, akkor a távolság 0.



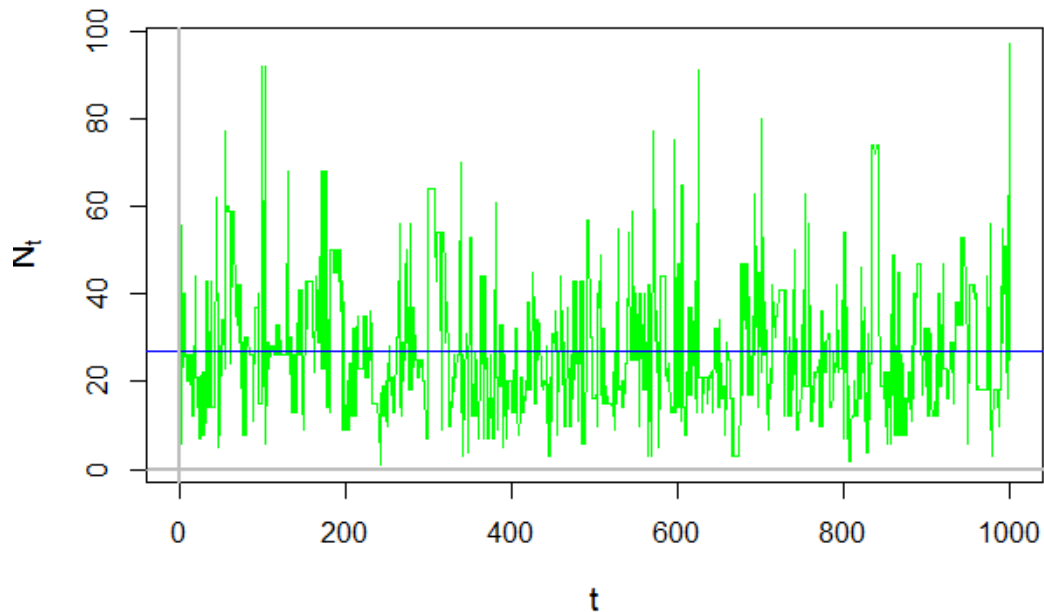
5.3. ábra. AIC (piros) és AICc(kék) függvények.  
 $t = 100$ ,  $\lambda = 1$ ,  $\alpha_1 = 0,4$ ,  $\alpha_2 = 0,3$ ,  $\alpha_3 = 0,2$

a paramétereket:

$\phi_i$	0,2	0,3	0,1
átlag( $\hat{\phi}_i$ )	0,2089384	0,3013597	0,0928193
szórás( $\hat{\phi}_i$ )	0,0425605	0,0472138	0,0444915

## Valós adatok elemzése

Egy ma is működő biztosító társaság kötelező gépjármű felelősségbiztosításhoz kapcsolódó károkat vizsgáltam. Az adatok egyfelől igen részletesek, másfelől kellőképpen megszűrték voltak. A kár bekövetkezésének időpontjának a tényleges káridőpontot vettem, így nem törődtem a kár bejelentésének és az esetleges kárkifizetés időpontjával. Azon károkat vizsgáltam, amik végül kifizetéssel zárultak. Ezen kifizetett károk nagyságát is megkaptam (de figyelmen kívül hagytam az adott kárra korábban megképzett függőkárokat). Az adatok közül a járadékos károkat is kiszűrtük, mert azok gyakran nagyon el tudnák torzítani a kifizetéseket egy-egy nagy kárral. Az adatok havi megbontásúak voltak, így viszonylag nagy időtávra nagy frekvenciájú adat állt rendelkezésre. Ezek mellett rendelkezésemre bocsáj-



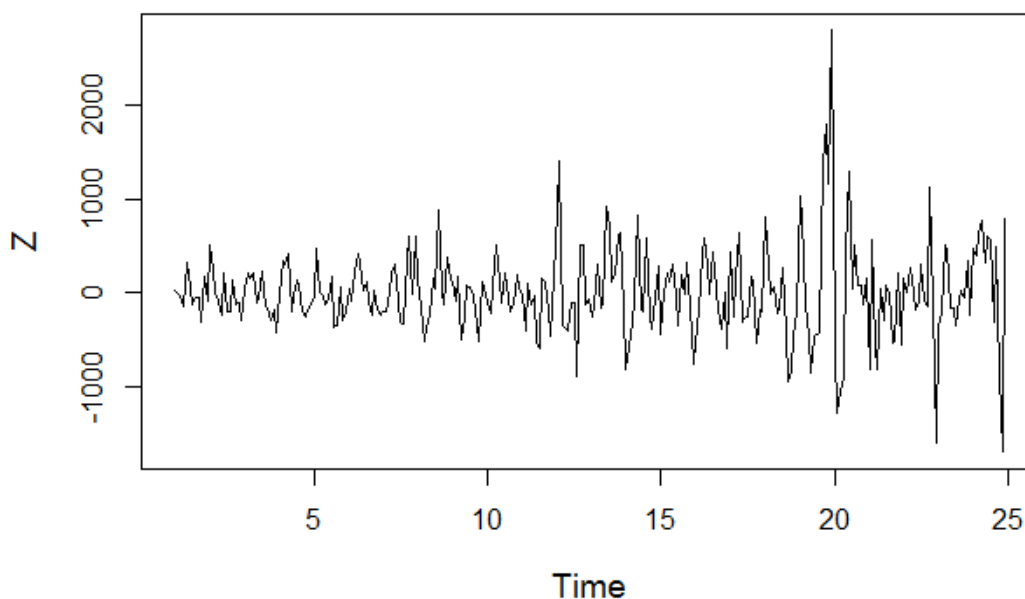
5.4. ábra. Kevert INAR(3) realizáció.  
 $t = 1000$ ,  $\phi_1 = 0,2$ ,  $\phi_2 = 0,3$ ,  $\phi_3 = 0,1$

tották a biztosító adott havi portfóliójának a nagyságát, valamint az adott havi átlagdíjt. Ezen adatokból különböző kárfolyamatok becsülhetők lennének, azonban most csak arra koncentrálnunk, hogy megpróbáljunk egy INAR modellt illeszteni az adatokra. Célunk nem a konkrét adatok bemutatása, hanem inkább egy eljárás, modell kreálása, melyet adatok birtokában tudunk hasznosítani INAR folyamatok tekintetében.

A kapott adatokat először kicsit „kozmetikázzuk”, azaz a hibás részeket esetlegesen kitöröljük, vagy adatokkal töltjük fel. Ezután idősort kreálunk belőle, melynek a frekvenciáját a havi adatok miatt 12-re állítjuk. Mivel a károk darabszámát szeretnénk modellezni, ezért azzal foglalkozunk tovább. Mivel szezonális trend figyelhető meg az adatokban, ezért egy ezt kiszűrő módszerrel kiszűrjük azt. A választott módszer a Loess-féle  $STL^2$ . A módszer az idősort három különböző komponensre bontja: trendre, szezonálisra és maradékra (bővebben a [5] cikkben olvashatunk). Az adatok elfedése érdekében csak a maradékot mutatjuk be. Több különböző

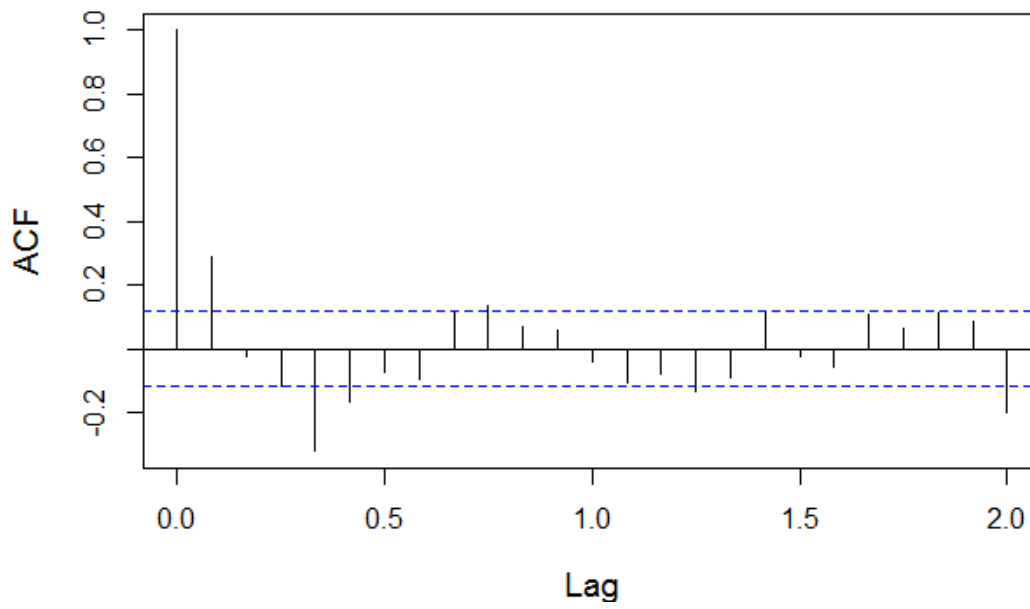
<sup>2</sup>Seasonal-Trend Decomposition

szazonalitási ablakkal is vizsgálhatjuk, de lényegi különbséget nem tapasztaltunk a 12, 18 és 24-es frekvencia paraméterbeállítások között. A frekvenciának 12-őt választva az 5.5 ábrán látható maradék folyamathoz jutunk. Vizsgáljuk meg a



5.5. ábra. Kárszám STL felbontása - maradék folyamat.

hiba folyamat autokovariancia függvényét. Azt kapjuk, hogy ACF második tagja negatív (5.6 ábra). Ebből arra tudunk következtetni, hogy a folyamat nem lehet  $INAR(p)$ -ből származó, mert annak az autokorrelációs függvénye (és így autokovariancia függvénye is) exponenciálisan lecseng és nem tartalmaz negatív értéket. Vizsgáljuk azért meg, hogy milyen értékeket adnak vissza a Yule-Walker egyenletek. Megvizsgáltuk másod, harmad és negyedrendű esetben is, de mindegyikben szerepelt negatív paraméter. Emiatt azt mondhatjuk, hogy a kapott adatsor nem modellezhető  $INAR$  folyamatokkal (vagy legalábbis nem ilyen formában). Megvizsgáltuk diszkrétizált értékeken is, de úgy sem változtak az együtthatók paraméterbecslései számottevően.



5.6. ábra. Maradék folyamat autokovarianciája.

## 6. fejezet

### Kitekintés

Megvizsgáltuk az alap kockázati modelleket, amelyek mindegyike valahogy az összkár tulajdonságait igyekszik megvizsgálni és arra becsléseket adni. Ismertettük a klasszikus kockázati folyamatok és rizikó folyamatok tekintetben a már régóta fennálló, csődvalószínűségekre vonatkozó eredményeket. Az INAR folyamatokat, mint az összefüggő kárszámokból adódó modelleket, több szempontból megvizsgáltuk és azon belül is elemeztük a csődvalószínűségekre gyakorolt hatását. Három különböző modellnek a viselkedését vizsgáltuk, melyeknél láttuk, hogy két esetben jól tudtuk becsülni a paramétereket és az egyik esetében a rendre is jó becslést adtunk. A valós adatok nem mutattak jó illeszkedést, így arra jutottunk, hogy a választott módszerrel azok nem modellezhetőek. Azonban ez nem jelenti azt, hogy semmilyen körülmények között ne lehetne ezeket az adatokat megfelelően modellezni. A modellek további vizsgálatával feltevésem szerint olyan eszözöket lehet adni, amik jól kezelik a különböző adatokat és felhasználásukkal megfelelő paraméter becslő és ez által előrejelző eljárások lennének. Egyre több cikk születik a témában, amelyekben már megvannak különböző eljárások, amelyek alapos vizsgálatával a pontosabb modellek kivitelezése elkezdhető. A pontos modellek használata nagyon hasznos lehet a Biztosítóknak a minimális tőkeszükséglet meghatározásához, így mindenképp hasznos lehet különböző Szolvencia 2 szabályozásban használható belső modellek kialakításakor.

# Függelék

## Multinomiális INAR modell

```
rm(list=ls())
eloszt<-function(k,Y,q) { p1<-length(q);x<-rep(0,p1-1)
if(Y[k]>0) {
  s<-table(sample(1:p1-1,Y[k],rep=TRUE,prob=q))
  if(names(s)[1]=="0") s<-s[-1]
  if(length(s)) x[as.numeric(names(s))]<-s };
return(x) }
```

```
N<-10000
lambda=10
```

```
a<-c(.4,.3,.2)
a0<-1-sum(a)
e<-rpois(N,lambda*a0)
p<-length(a)
X<-matrix(0,N,p)
Y<-rep(NA,N)
Y[1:3]<-e[1:3]
```

```
X[1,]<-eloszt(1,Y,c(a0,a));
X[2,]<-eloszt(2,Y,c(a0,a));
X[3,]<-eloszt(3,Y,c(a0,a))
```

```
for(k in 4:N){
  Y[k]<-X[k-1,1] + X[k-2,2] +X [k-3,3] + e[k];
  X[k,] <- eloszt(k,Y,c(a0,a))
}
plot(Y,ylab=expression(N[ t ]),xlab=expression(t),
      type="l", lwd=1.2, col="green")
abline(h=10, col="blue")
abline(h=0,v=0, col="grey", lwd=2)
```

```

#YW
ac<-acf(Y,type="covariance")
A<-matrix(NA,3,3)
c0<-ac$acf[1];
c1<-ac$acf[2];
c2<-ac$acf[3];
c3<-ac$acf[4]
A[1,1]<-A[2,2]<-A[3,3]<-c0
A[1,2]<-A[2,3]<-A[2,1]<-A[3,2]<-c1
A[1,3]<-A[3,1]<-c2
r<-c(c1,c2,c3)

solve(A,r)
a

#tesztfv
teszt<-function(a,N=123)
{
  p<-length(a)
  a0<-1-sum(a)
  Y<-rep(NA,N)
  e<-rpois(N,lambda*a0)
  Y[1:p]<-e[1:p]
  X<-matrix(0,N,p)
  for(k in 1:p) X[k,]<-eloszt(k,Y,c(a0,a))
  for(k in (p+1):N)
  { Y[k]<-sum(diag(X[k-(1:p)],1:p))+e[k];
    X[k,]<-eloszt(k,Y,c(a0,a)) }

  ac<-acf(Y,type="covariance",plot=FALSE)$acf
  A<-matrix(NA,p,p)
  diag(A)<-ac[1]
  r<-ac[-1]
  for(j in 1:(p-1))
    for(i in 1:(p-j))
      A[i,i+j]<-A[i+j,i]<-r[j]
  return(rbind(solve(A,r[1:p]),a))
}
teszt(c(.3,.2,.2,.1,.1),N=10000)
teszt(c(.3,.4,.2),N=10000)

```



```

#modosotott teszt fv
teszt2<-function(a,N=123){
  p<-length(a)
  a0<-1-sum(a)
  Y<-rep(NA,N)
  e<-rpois(N,lambda*a0)
  Y[1:p]<-e[1:p]
  X<-matrix(0,N,p)
  for(k in 1:p) X[k,]<-eloszt(k,Y,c(a0,a))
  for(k in (p+1):N){
    Y[k]<-sum(diag(X[k-(1:p),1:p]))+e[k];
    X[k,]<-eloszt(k,Y,c(a0,a))
  }
  ac<-acf(Y,type="covariance",plot=FALSE)\$acf
  A<-matrix(NA,p,p)
  diag(A)<-ac[1]
  r<-ac[-1]
  for(j in 1:(p-1)) for(i in 1:(p-j))
  A[i,i+j]<-A[i+j,i]<-r[j]
  return(solve(A,r[1:p]))
}

```

```

#modositott teszt fv 100 futtatasa
H<-matrix(NA,100,3)
for(j in 1:100){L<-teszt2(c(.8,.1,.05),N=100)
  H[j,1]<-L[1]
  H[j,2]<-L[2]
  H[j,3]<-L[3]
}

```

```

a<-c(.8,.1,.05)
colMeans(H)
sqrt(colVars(H))
a

```

## Binomiális INAR modell

```

N<-10000
lambda<-10
a<-c(.4,.3,.2)
a0<-1-sum(a)
e<-rpois(N,lambda*a0)

```

```

Y<-rep(NA,N)
Y[1:3]<-e[1:3]
for(k in 4:N)
  Y[k]<-rbinom(1,Y[k-1],a[1])+
  rbinom(1,Y[k-2],a[2])+
  rbinom(1,Y[k-3],a[3])+
  e[k]

# —
# modell illesztés Yule-Walker módszerrel

ac<-acf(Y,type="covariance")
A<-matrix(NA,3,3)
c0<-ac$acf[1]; c1<-ac$acf[2]; c2<-ac$acf[3]; c3<-ac$acf[4]
A[1,1]<-A[2,2]<-A[3,3]<-c0
A[1,2]<-A[2,3]<-A[2,1]<-A[3,2]<-c1
A[1,3]<-A[3,1] <-c2
r<-c(c1,c2,c3)

solve(A,r)
a
plot(Y,ylab=expression(N[ t ]),xlab=expression(t),
type="l",lwd=1.2,col="green")
abline(h=10,col="blue")
abline(h=0,v=0,col="grey",lwd=2)

# —
# egy általános teszt függvény

rbinom0<-function(n,p) if(n==0) 0 else rbinom(1,n,p)

teszt<-function(a,N=123)
{ # a<-c(.3,.2,.2,.1,.1)
  p<-length(a)
  a0<-1-sum(a)
  lambda=10
  Y<-rep(NA,N)
  e<-rpois(N,lambda*a0)
  Y[1:p]<-e[1:p]
  for(k in (p+1):N)
    { Y[k]<-e[k]

```

```

        for(j in 1:p) Y[k]<-Y[k]+rbinom0(Y[k-j],a[j])}
ac<-acf(Y,type="covariance",plot=FALSE)\$acf
A<-matrix(NA,p,p)
diag(A)<-ac[1]
r<-ac[-1]
for(j in 1:(p-1)) for(i in 1:(p-j))
  A[i,i+j]<-A[i+j,i]<-r[j]
return(rbind(solve(A,r[1:p]),a))
      }

teszt(c(.3,.2,.2,.1,.1),N=100000)
teszt(c(.3,.4,.2),N=100000)

#—
#modositott teszt fv
teszt2<-function(a,N=100)
{ #  $\alpha < c(.3,.2,.2,.1,.1)$ 
  p<-length(a)
  a0<-1-sum(a)
  lambda=10
  Y<-rep(NA,N)
  e<-rpois(N,lambda*a0)
  Y[1:p]<-e[1:p]
  for(k in (p+1):N)
    { Y[k]<-e[k]
      for(j in 1:p) Y[k]<-Y[k]+rbinom0(Y[k-j],a[j])}
  ac<-acf(Y,type="covariance",plot=FALSE)\$acf
  A<-matrix(NA,p,p)
  diag(A)<-ac[1]
  r<-ac[-1]
  for(j in 1:(p-1)) for(i in 1:(p-j))
    A[i,i+j]<-A[i+j,i]<-r[j]
  return(solve(A,r[1:p]))
}

#modositott teszt fv 100 futtatasa
H<-matrix(NA,100,3)
for(j in 1:100){L<-teszt2(c(.4,.3,.1),N=10000)
H[j,1]<-L[1]
H[j,2]<-L[2]
H[j,3]<-L[3]
}

```

```

a<-c(.4,.3,.1)
colMeans(H)
sqrt(colVars(H))
a

```

## Módosított AIC kritérium

```

fokszam<-function(a,N=100){
  p <- length(a)
  Y <- rep(NA,N)
  mu <- 1
  e <- rpois(N,mu)
  Y[1:p] <- e[1:p]
  for(k in (p+1):N){
    Y[k] <- e[k]
    for(j in 1:p)
      Y[k]<-Y[k]+rbinom0(Y[k-j],a[j])
  }
  ac <- acf(Y,type="covariance",plot=FALSE)\$acf

  # az AIC kulonbozo fokszamok mellett
  origAIC <- corrAIC <- rep(NA,10)
  for(q in 1:10){
    A <- matrix(NA,q,q)
    diag(A) <- ac[1]
    r <- ac[-1]
    if(q>1)
      for(j in 1:(q-1))
        for(i in 1:(q-j))
          A[i,i+j] <- A[i+j,i] <- r[j]

    akalap <- solve(A,r[1:q])
    Ykalap <- rep(NA,N)
    Ykalap[1:q] <- Y[1:q]
    for(k in (q+1):N)
      Ykalap[k] <- sum(akalap*Y[k-1:q])+mu
    s2 <- var(Y-Ykalap)
    origAIC[q] <- N*(log(s2)+1)+2*(q+1)
    corrAIC[q] <- N*log(s2)+N*(1+q/N)/(1-(q+2)/N) }
  return(rbind(1:10,origAIC,corrAIC))
}

```

```

set.seed(100)
w<-fokszam(c(.4,.3,.2),100)
plot(1,mean(w[2:3,]),xlim=c(1,10),ylim=c(min(w[2:3,]),
      max(w[2:3,])),t='n',las=1,xlab="AR_lag",ylab="AIC")
points(w[2,],t='b',col="red",pch=19)
points(w[3,],t='b',col="blue",pch=19)
axis(1,at=1:10)

```

## Kevert INAR modell

```
N<-100000
```

```
k<-3
```

```
r<-.1
```

```
e<-rbinom(N,k,r)
```

```
a<-c(.2,.3,.1);a0<-1-sum(a)
```

```
p<-length(a)
```

```
Y<-rep(NA,N)
```

```
Y[1:3]<-e[1:3]
```

```
for(j in (p+1):N)
```

```
{ # p db vget kell megvaltoztatni
```

```
  y<-sample(c(-1,Y[(j-1):(j-p)]),1,p=c(a0,a))
```

```
  Y[j]<-if(y==-1)e[j]else y
```

```
}
```

```
ac<-acf(Y,type="covariance")
```

```
A<-matrix(NA,3,3)
```

```
c0<-ac$acf[1];
```

```
c1<-ac$acf[2];
```

```
c2<-ac$acf[3];
```

```
c3<-ac$acf[4]
```

```
A[1,1]<-A[2,2]<-A[3,3]<-c0
```

```
A[1,2]<-A[2,3]<-A[2,1]<-A[3,2]<-c1
```

```
A[1,3]<-A[3,1] <-c2
```

```
r<-c(c1,c2,c3)
```

```
solve(A,r)
```

```
a
```

```
plot(Y,ylab=expression(N[ t ]),xlab=expression(t),
```

```
  type="l",lwd=1.2,col="green")
```

```
abline(h=mean(Y),col="blue")
```

```

abline(h=0,v=0, col="grey", lwd=2)

# ---
# egy altalános teszt függvény

teszt<-function(a,N=123)
{ m<-3
  r<-0.1
  e<-rnbinom(N,m,r)
  a<-c(.2,.3,.1); a0<-1-sum(a)
  p<-length(a)

  Y<-rep(NA,N)
  Y[1:p]<-e[1:p]
  for(j in (p+1):N)
    { # p db vget kell megváltoztatni
      y<-sample(c(-1,Y[(j-1):(j-p)]),1,p=c(a0,a))
      Y[j]<- if(y==-1) e[j] else y
    }
  ac<-acf(Y,type="covariance",plot=FALSE)\$acf
  A<-matrix(NA,p,p)
  diag(A)<-ac[1]
  r<-ac[-1]
  for(j in 1:(p-1)) for(i in 1:(p-j))
    A[i,i+j]<-A[i+j,i]<-r[j]
  return(rbind(solve(A,r[1:p]),a))
}

teszt(c(.3,.2,.2,.1,.1),N=1000)
teszt(c(.3,.4,.2),N=1000)

#modosított teszt fv
teszt2<-function(a,N=123)
{ m<-3
  r<-0.1
  e<-rnbinom(N,m,r)
  a<-c(.2,.3,.1); a0<-1-sum(a)
  p<-length(a)

  Y<-rep(NA,N)
  Y[1:p]<-e[1:p]
  for(j in (p+1):N)

```

```

    { # p db vget kell megvaltoztatni
      y<-sample(c(-1,Y[(j-1):(j-p)]),1,p=c(a0,a))
      Y[j]<- if(y==-1) e[j] else y
    }
ac<-acf(Y,type="covariance",plot=FALSE)\$acf
A<-matrix(NA,p,p)
diag(A)<-ac[1]
r<-ac[-1]
for(j in 1:(p-1)) for(i in 1:(p-j))
  A[i,i+j]<-A[i+j,i]<-r[j]
return(solve(A,r[1:p]))
}

#modositott teszt fv 100 futtatasa
H<-matrix(NA,100,3)
for(j in 1:100){L<-teszt2(c(.2,.3,.1),N=1000)
  H[j,1]<-L[1]
  H[j,2]<-L[2]
  H[j,3]<-L[3]
}
a<-c(.2,.3,.1)
colMeans(H)
sqrt(colVars(H))
a

Valós adatok elemzése

D<-read.csv("CountVal.csv",head=TRUE,sep=";",row.names=1)
plot(D$C,t="l",col="red",lwd=2)

which(is.na(D$C)) # 37
D[30:40,] # 199413 toroljuk
D<-D[-37,]
head(D,50)

plot(D$C,t="l",col="red",lwd=2)

ts(row.names(D),freq=12) # hibatlanok az idopontok
Y<-ts(D$C,freq=12) # ez az esetszam, ezt modellezzuk

plot(stl(Y,12)); plot(stl(Y,18)); plot(stl(Y,24)) # hasonloak

```

```

M<-stl(Y,18)
Z<-round(M$st[,3]) # ez a hiba folyamat
save(Z, file="CountVal.Rdata") # elementjuk

acf(Z) # van negativ acf, pl. a 2.

#YW egyenletek

YW<-function(Y,p){
  ac<-acf(Y,type="covariance",plot=FALSE)$acf
  A<-matrix(NA,p,p)
  diag(A)<-ac[1]
  r<-ac[-1]
  for(j in 1:(p-1)) for(i in 1:(p-j))
    A[i,i+j]<-A[i+j,i]<-r[j]
  return(solve(A,r[1:p]))
}

YW(Z,2); YW(Z,3); YW(Z,4); YW(Z,5);

#diszkretizaljuk
Zd<-round(Z/500)
table(Zd)
plot(Zd,t='b')

YW(Zd,2); YW(Zd,3)

```



# Irodalomjegyzék

- [1] AL-OSH, M. A., AND ALZAID, A. A. First-order integer-valued autoregressive (INAR(1)) process. *Journal of Time Series Analysis Vol.8* (1987), pp. 261–275.
- [2] AL-OSH, M. A., AND ALZAID, A. A. An integer-valued  $p$ th-order autoregressive structure (INAR( $p$ )) process. *Journal of Applied Probability Trust Vol.27*, No.2 (1990), pp. 314–324.
- [3] ARATÓ MIKLÓS. *Nem-élet biztosításmatematika*, 2. ed. Jan. 2001.
- [4] BISWAS, A., AND SONG, P. X.-K. Discrete-valued ARMA processes. *Statistics and Probability Letters 79*, 9 (2009), pp. 1884–1889.
- [5] CLEVELAND, R. B., CLEVELAND, W. S., MCRAE, J. E., AND TERPENNING, I. STL: A Seasonal-Trend Decomposition Procedure Based on Loess. *Journal of Official Statistics Vol.6*, No.1 (1990), pp. 3–33.
- [6] COSSETTE, H., MARCEAU, E., AND MAUME-DESCHAMPES, V. Discrete-time risk models based on time series for count random variables. *Astin Bulletin Vol.40*, No.123 (2010).
- [7] DU, J.-G., AND LI, Y. The integer-valued autoregressive (INAR( $p$ )) model. *Journal of Time Series Analysis Vol.12*, No.2 (1991), pp. 129–142.
- [8] EMBRECHTS, P., AND FREI, M. Panjer recursion versus fft for compound distributions. *Mathematical Methods of Operations Research Vol. 69*, No. 3. (2009), pp. 497–508.
- [9] FRANKE, J., AND RAO, T. S. Multivariate first order integer valued autoregressions. Tech. rep., UMIST., Math Dep., 1995.
- [10] HEYDE, C. C., AND SENETA, E. Estimation theory for growth and immigration rates in a multiplicative process. *Journal of Applied Probability Vol. 9* (1972), pp. 235–256.
- [11] HURVICH, C. M., AND TSAI, C.-L. Regression and Time Series Model Selection in Small Samples. *Biometrika*, (1989).

- [12] JUNG, R. C., AND TREMAYNE, A. R. Binomial thinning models for integer time series. *Statistical Modelling* 6 (2006), pp. 81–96.
- [13] LATOUR, A. Existence and stochastic structure of a non-negative integer-valued autoregressive process. *Journal of Time Series Analysis Vol.19*, No.4 (1998), pp. 439–455.
- [14] MCKENZIE, E. Some simple models for discrete variate time series. *Water Resources Bulletin Vol.21*, No.4 (1985), pp. 645–650.
- [15] MCKENZIE, E. Autoregressive Moving-Average Processes with Negative-Binomial and Geometric Marginal Distributions. *Advnces in Applied Probability Vol.18*, No.3 (Sep. 1986), pp. 679–705.
- [16] MCKENZIE, E. Some ARMA Models for Dependent Sequences of Poisson Counts. *Advnces in Applied Probability Vol.20*, No.4 (1988), pp. 822–835.
- [17] MICHALETZKY, G. Kockázati folyamatok. Note TEMPUS AC-JEP-13358-98, Eötvös Loránd Tudományegyetem.
- [18] PANJER, H. H. Recursive valuation of a family of compound distributions. *Astin Bulletin Vol.12* (1981), pp. 22–26.
- [19] PEGRAM, G. G. S. An Autoregressive Model for Multilag Markov Chains. *Journal of Applied Probability Vol.17*, No.2 (1980), pp. 350–362.
- [20] PRIL, N. D. On the exact computation of the aggregate claims distribution. *Astin Bulletin Vol.16*, No.2 (1986), pp. 190–112.
- [21] QUDDUS, M. A. Time series count data models: An empirical application to traffic accidents. *Accident Analysis and Prevention Vol. 40* (2008), pp. 1732–1741.
- [22] SILVA, I., AND SILVA, M. E. Asymptotic distribution of the yule-walker estimator for INAR( $p$ )processes. *Statistics and Probability Letters Vol.76*, No.15 (2006), pp. 1655–1663.
- [23] STEUTEL, F. W., AND VAN HARN, K. Discrete analogues of self-decomposability and stability. *The Annals of Probability Vol.7*, No.5 (1979), pp. 893–899.
- [24] SZŰCS, G. Kockázati folyamatok. Lecture note, Szegedi Tudományegyetem, 2015.
- [25] WEISS, C. H. Thinning operations for modelling time series of counts – a survey. *Advances in Statistical Analysis Vol.93*, No.3 (2008), pp. 319–341.