Aplikácia trojstavového modelu s využitím Markovových reťazcov a simulácií

Application of the three state model using Markov chains and simulations

Martin Kosztolányi[1]

Abstrakt
Cieľom článku je prezentovať využitie simulácií nehomogénnych Markovových reťazcov v diskrétnom čase pre trojstavový model. Predmetom skúmania je modelovanie rozdelenia klientov do jednotlivých stavov počas určených časových krokov, následné odhadovanie priemerného času zotrvania poistenej osoby v danom stave. V rámci použitia metódy simulácií je dôraz kladený na poskytnutie prístupov, ktoré zabezpečia presnejšie výsledky v závislosti od počtu realizovaných simulácií. Na základe štatistického spracovania údajov získaných zo simulácií je možné získať ďalšie informácie potrebné na analytické účely, na kalkuláciu peňažných prostriedkov potrebných na zabezpečenie zdravotnej starostlivosti, prípadne na cenotvorbu poistných produktov. Na implementáciu uvedených techník a ich grafické znázornenie boli použité dostupné balíky, ako napríklad *markovchain*, *lifecontingencies*, ako aj vlastný kód vytvorený v jazyku R.

Kľúčové slová
Markovove reťazce, simulácie, trojstavový model, civilizačné ochorenie, dlhodobá starostlivosť

Abstract
The objective of the article is to demonstrate the application of simulations of non-homogeneous Markov chains in discrete time for a three-state model. The analysis focuses on modelling the distribution of clients across individual states over defined time steps, followed by estimating the expected duration an insured person remains in a given state. Within the simulation framework, particular emphasis is placed on approaches that enhance the accuracy of results as the number of simulation runs increases. Statistical processing of the data generated through simulations enables to obtain additional information needed for analytical purposes, for assessing the volume of the funds necessary to provide healthcare services, or for pricing of the insurance products. The implementation of these techniques, including their graphical representation, uses available packages such as *markovchain* and *lifecontingencies*, supplemented by own custom R code.

Key words
Markov chains, simulations, three-state model, lifestyle diseases, long term health care

JEL classification
G22, C63

[1] Ekonomická univerzita v Bratislave, Fakulta hospodárskej informatiky, Katedra matematiky a aktuárstva, Dolnozemská cesta 1, 852 35 Bratislava, martin.kosztolanyi.2@euba.sk.

1 Úvod

V súčasnosti je pozorovaný jav nárastu priemernej dĺžky života a môžeme predpokladať, že tento trend bude pokračovať aj v budúcnosti (predovšetkým v rozvinutých krajinách). Práve kategória ľudí vo vyššom veku trpí rôznymi chronickými ochoreniami alebo fyzickými obmedzeniami a zároveň táto skupina najviac využíva rôzne služby zdravotnej a opatrovateľskej starostlivosti. Z tohto dôvodu venujú poskytovatelia takýchto služieb veľkú pozornosť odhadom počtu klientov, ktorí budú v budúcnosti potrebovať zdravotnú starostlivosť. Osobitne sa zameriavajú na problematiku dlhodobej starostlivosti poskytovanej ľuďom, ktorí dosiahli stav, keď sa nevedia sami o seba plnohodnotne postarať.

Zvýšený počet ľudí, ktorí v dôsledku svojho zdravotného stavu ostanú odkázaní na pomoc, zároveň vytvára tlak na nárast nákladov potrebných na primeranú zdravotnú starostlivosť. Z pohľadu jednotlivca je dôležité zamýšľať sa aj nad tým, aké úspory, prípadne aký kapitál budú mať k dispozícii v prípade potreby dlhodobej starostlivosti. Ak osoba nedisponuje vhodným poistením pre prípad takejto dlhodobej starostlivosti, náklady vynakladané na takýto typ služby sa môžu veľmi rýchlo vyčerpať a ohroziť úspory jednotlivca alebo jeho príbuzných.

Markovove reťazce patria medzi ľahko dostupné nástroje používané pri odhadoch výskytu rôznych ochorení aj pri modelovaní poskytovania (dlhodobej) zdravotnej starostlivosti. Problematike využívania takýchto Markovových náhodných procesov na modelovanie priebehu rôznych ochorení v kontexte viacstavových modelov sa venuje viacero autorov. Na ich implementáciu využívajú rôzne softvérové prostriedky, medzi ktorými je jednou z vhodných možností napríklad aj prostredie jazyka R s dostupnými balíkmi k spracovaniu Markovových reťazcov.

Napríklad balík *markovchain* (Spedicato, Kang, Yalamanchi, Yadav a Cordón, 2016) prípadne aktualizovaný materiál k balíku (Spedicato, 2026) umožňuje vytvárať objekty Markovových reťazcov, realizovať rôzne výpočty a operácie, vykonávať štatistické analýzy a simulácie homogénnych aj nehomogénnych Markovových reťazcov v diskrétnom čase. Funkcie na implementáciu Markov Chain Monte Carlo (MCMC) simulácií pomocou Metropolis algoritmu obsahuje balík *mcmc* (Geyer a Johnson, 2013). Okrem prostredia R je samozrejme možné siahnuť po akomkoľvek inom prostredí na programovanie či matematické analýzy. V prípade potreby je možné vytvoriť aj vlastný kód v uvedených programovacích jazykoch na simuláciu Markovových reťazcov v diskrétnom aj spojitom čase podľa algoritmov dostupných v rôznych publikáciách.

Ako už bolo uvedené, Markovove reťazce sa používajú aj v kontexte viacstavových modelov pri aktuárskom modelovaní v životnom aj neživotnom poistení napríklad na odhad nákladov spojených s poskytovaním zdravotnej starostlivosti, na výpočet poistného, na predpovedanie vývoja rôznych chorôb, ako aj na modelovanie očakávaného vývoja počtu poistených osôb v bonus-malus systéme v rámci povinného zmluvného poistenia motorových vozidiel.

Ďalšou oblasťou využitia Markovových reťazcov pri modelovaní kritických, prípadne civilizačných ochorení v kontexte poskytovania zdravotnej starostlivosti, je poistenie kritických chorôb. Poistenému sa v prípade diagnózy špecifického ochorenia vyplatí (zvyčajne jednorazové) poistné plnenie. Mnohí autori používajú Markovove reťazce na predpovedanie vývoja infekčných typov chorôb (Li, Dushoff a Bolker, 2018; Twumasi, Asiedu a Nortey, 2019, Mucha, Faybíková a Krčová, 2022).

V neživotnom poistení sa homogénne Markovove reťazce v diskrétnom čase používajú na modelovanie rozdelenia počtu poistníkov v bonus-malus systéme (Fernandez-Morales, 2015).

V tomto článku sa zameriame na simuláciu trajektórií nehomogénnych Markovových reťazcov v diskrétnom čase pomocou jazyka R. Na základe spracovania vygenerovaných

údajov budeme analyzovať modelovanie rozdelenia počtu klientov v jednotlivých stavoch počas určených rokov a odhad priemerného času zotrvania klienta v danom stave. Na prezentáciu uvedených techník používame údaje z balíka markovchain, ktoré sú odvodené na základe mužskej populácie v Taliansku a spracujeme aj porovnanie s použitím populačných dát z roku 2019 pre mužov zo Slovenskej republiky. V trojstavovom modeli predstavuje „chorý stav“ stav nespôsobilosti, do ktorého sa klient mohol dostať z dôvodu civilizačného ochorenia, napríklad v dôsledku Alzheimerovej choroby.

2 Prehľad problematiky Markovových reťazcov a viacstavových modelov

Náhodná premenná je nástroj na matematický opis náhodného javu. Je to funkcia v určitej množine Ω v pravdepodobnostnom priestore a s určitým pravdepodobnostným rozdelením, ktoré je charakteristické pre túto náhodnú premennú. Rozšírením diskusie o náhodnej premennej je prirodzene náhodný proces. Náhodný proces namiesto jednej náhodnej premennej opisuje sled (veľmi často časový sled) náhodných premenných. Pojem náhodného procesu v pravdepodobnostnom priestore (Ω,S,P), kde S je σ-algebra podmnožín množiny Ω a P je pravdepodobnostná miera definovaná na priestore S uvádza nasledujúca definícia.

Definícia 1. Náhodný proces je systém náhodných premenných $\{X_t\}_{t\in T}$ definovaných na pravdepodobnostnom priestore S, kde pre každé $t \in T$ je $X_t: \Omega \to R$ je náhodná premenná.

Realizácia $X_t$sa nazýva “sample path”.

Pre fixné $\omega \in \Omega$ dostávame jednu realizáciu $x(t,\omega)$náhodného procesu ako funkciu premennej $t \in T$. Množinu T je prirodzené predstaviť si ako čas. Ak T je množina prirodzených alebo celých čísel, hovoríme o náhodnom procese s diskrétnym časom. Ak T je interval na reálnej osi, ide o náhodný proces so spojitým časom. Pri riešení úloh so stochastickými procesmi je vhodné uvažovať zvlášť prípad spojitý a zvlášť diskrétny.

Náhodné premenné sú charakterizované rozdelením pravdepodobnosti daným príslušnou distribučnou funkciou. V prípade potreby charakterizovania náhodného procesu $\{X_t\}_{t\in T}$je toto zložitejšie. Pre účely tohto článku sa sústredíme najmä na rozdelenie procesov. Hodnoty, ktoré nadobúdajú náhodné premenné X_t, $t \in T$ nazývame stavmi a ich množinu označíme S. Zvyčajne ňou býva konečná, alebo spočítateľná množina nezáporných celých čísel N_0 alebo jej konečná podmnožina. Takýto náhodný proces nazývame náhodný reťazec.

Definícia 2. Markovova vlastnosť. Hovoríme, že reťazec $\{X_t\}_{t\in T}$, má Markovovu vlastnosť alebo je Markovov, ak pre každé $n = 0, 1, 2, \ldots$ a pre všetky časy $0 \le t_0 \le \cdots \le t_{n+1}$; $t_0, \ldots, t_{n+1} \in T$ a pre všetky stavy $i_0, i_1, \ldots, i_{n+1} \in S$ platí:

$$P\left(X_{t_{n+1}} = i_{n+1} \mid X_{t_n} = i_n, \ldots, X_{t_o} = i_0\right) = P\left(X_{t_{n+1}} = i_{n+1} \mid X_{t_n} = i_n\right) \qquad (1)$$

Teda za podmienky, že $X_{t_n} = i_n$, je $X_{t_{n+1}}$ nezávislé od $X_{t_o}, \ldots, X_{t_{n-1}}$.

Ak podmienená pravdepodobnosť $P\left(X_{t_{n+1}} = j \mid X_{t_n} = i\right)$ závisí len od dĺžky intervalu $t_{n+1} - t_n$, teda presnejšie, ak pre každé $t \in T$, $s > 0$ a všetky stavy $i, j \in S$ platí $P(X_{t+s} = j \mid X_t = i) = p_{ij}(s)$ hovoríme, že reťazec je homogénny.

Ak $T \in \langle 0,\infty)$, potom môžeme reťazcu priradiť systém matíc (vo všeobecnosti sú matice nekonečno rozmerné) $\{p(t); t \ge 0\}$, kde (i, j)-ty prvok matice je $p_{ij}(t) = P(X_t = j|X_0 = i)$, pre $i, j \in S$ a $t > 0$; $p_{ij}(0) = 0$ pre $i \ne j$; $p_{ij}(0) = 1$. Pre pevné $t \ge 0$ sa $P(t)$ nazýva maticou prechodu za čas $t > 0$. Ak $P(t)$ je matica prechodu za čas $t > 0$ Markovovho reťazca s množinou stavov S, tak $\sum_{j \in S} p_{ij}(t) = 1$ pre každé $i \in S$. Každý riadok matice prechodu $P(t)$ je pravdepodobnostné rozdelenie.

Markovov reťazec je jednoznačne určený počiatočným rozdelením a systémom matíc prechodu za čas $t, t \ge 0$.

Chapman-Kolmogorovova rovnosť je základná vlastnosť Markovových procesov, ktorá vyjadruje, že pravdepodobnosť prechodu medzi dvoma stavmi za určitý (dlhší) čas sa dá získať „skladaním“ prechodových pravdepodobností za kratšie časové úseky.

Nech $\{X_t\}_{t\geq 0}$ je Markovov proces so stavovým priestorom *S*. Chapman–Kolmogorovova rovnosť hovorí, že pre všetky stavy $i, j \in S$ a pre časy $0 \leq s, 0 \leq t$ platí:

$$p_{ij}(t+s) = \sum_{k \in S} p_{ik}(t) p_{kj}(s) \tag{2}$$

Inak povedané, prechod zo stavu *i* do stavu *j* za čas (*t* + *s*) sa dá rozložiť na *súčet* všetkých ciest cez medzistav *k* v čase *t*.

Veta 1. Matice prechodu $P(t)$, $t > 0$ a počiatočné rozdelenie α jednoznačne určujú konečnorozmerné rozdelenia náhodného procesu $\{X_t\}_{t\geq 0}$.

V modeloch s diskrétnym časovým krokom nás zaujímajú pravdepodobnosti prechodu za jednu časovú jednotku. Pre pravdepodobnosť prechodu zo stavu *i* do stavu *j* budeme používať označenie $p_{ij} \equiv p_{ij}(1)$. Pre pravdepodobnosti prechodov zo stavu *i* do stavu *j* za *n* diskrétnych období použijeme označenie ${p_{ij}}^n \equiv p_{ij}(n)$ a budeme ich nazývať pravdepodobnosťami prechodu *n*-tého rádu. Maticu prechodu prvého rádu označujeme $P \equiv P(1) = \left(p_{ij}\right)_{i,j\in S}$ maticu pravdepodobností *n*-tého rádu $P(n) = \left(p_{ij}(n)\right)_{i,j\in S}$. Všeobecne matica prechodu vyzerá nasledovne:

$$P = \begin{bmatrix} p_{11} & \cdots & p_{1j} \\ \vdots & \ddots & \vdots \\ p_{i1} & \cdots & p_{ij} \end{bmatrix} \tag{3}$$

Pre Markovov reťazec s tromi stavmi bude matica nasledovná:

$$P = \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{bmatrix} \tag{4}$$

Homogénny Markovov reťazec s diskrétnym časovým krokom má tieto pravdepodobnosti a tým aj matice prechodu rovnaké pre každý časový úsek. V prípade nehomogénneho Markovovho reťazca je potrebné rozšírenie – musíme uvažovať pravdepodobnosti aj matice s parametrom času, teda napr. $P(t, t+1)$.

Pravdepodobnosť prechodu z počiatočného stavu *i* v čase 0 do stavu $j, j \in S$ v *h* krokoch od času 0, teda od začiatku Markovovho reťazca sa nazýva absolútna pravdepodobnosť stavu Markovovho reťazca a označujeme ju nasledovne:

$$P_{ij}(0,h) = p_j^{(i)}(h) \tag{5}$$

Potom vektor $\boldsymbol{p}^{(i)}(h) = \left({p_j}^{(i)}(h)\right)_{j\in S}$ nazývame vektor absolútnych pravdepodobností. S použitím Chapman-Kolmogorovovej rovnosti môžeme vektor absolútnych pravdepodobností $\boldsymbol{p}^{(i)}(h)$ zapísať

$$p^{(i)}(h) = p^{(i)}(h-1) * P(h-1,h) = p^{(i)}(h-2) * P(h-2,h-1) * P(h-1,h) = p(0) * \ldots * P(h-1,h) \tag{6}$$

Komplexný prehľad matematického aparátu k Markovovým reťazcom uvádza napríklad (Janková, Kilianová, Brunovský a Bokes, 2014).

3 Aplikovaný postup

V článku je uvedený príklad trojstavového nehomogénneho Markovovho reťazca s diskrétnym časom. V prvej časti je analýza vykonaná analyticky, pomocou počiatočného rozdelenia a matíc prechodu v package „markovchain“ v programovom balíku R. Na druhej strane pre ubezpečenie a kontrolu správnosti je pripravené aj stochastické generovanie trajektórií pre rovnaký Markovov reťazec.
Generovanie trajektórií nehomogénneho Markovovho reťazca s diskrétnym časom v R:
Pre nehomogénny Markovov reťazec s maticami prechodov $P(t,t+1) = \left(p_{ij}(t,t+1)\right)_{i,j\in S}$ pre všetky hodnoty času T definujeme náhodnú premennú Z_r. Hodnoty pravdepodobnostnej funkcie $P(Z_r = j)$ predstavujú v príslušnej matici $P(t,t+1)$ hodnoty pravdepodobností $p_{i_r j}(t,t+1),\ j \in S = \{i_1, i_2, \dots, i_n\}$ konkrétne nachádzajúce sa v riadku *r*. Diskrétne rozdelenie zapíšeme $p_{Z_r}(j) = \left\{p_{i_r j}(t,t+1)\right\}_{j\in S}$. Potom algoritmus generovania hodnôt náhodnej premennej *Z* pomocou inverznej transformácie pozostáva z nasledovných krokov:

1. Vygenerovanie hodnoty *u* náhodnej premennej *U* ~ Unif(0, 1).
2. Transformovanie hodnôt *u* na hodnoty náhodnej premennej *Z*

$$Z_r = i_1 \text{ ak } u \le p_Z(i_1) \text{ alebo } Z_r = j \text{ ak } \sum_{l=i_1}^{j-1} p_{Z_r}(l) < u \le \sum_{l=i_1}^{j} p_{Z_r}(l) \qquad (7)$$

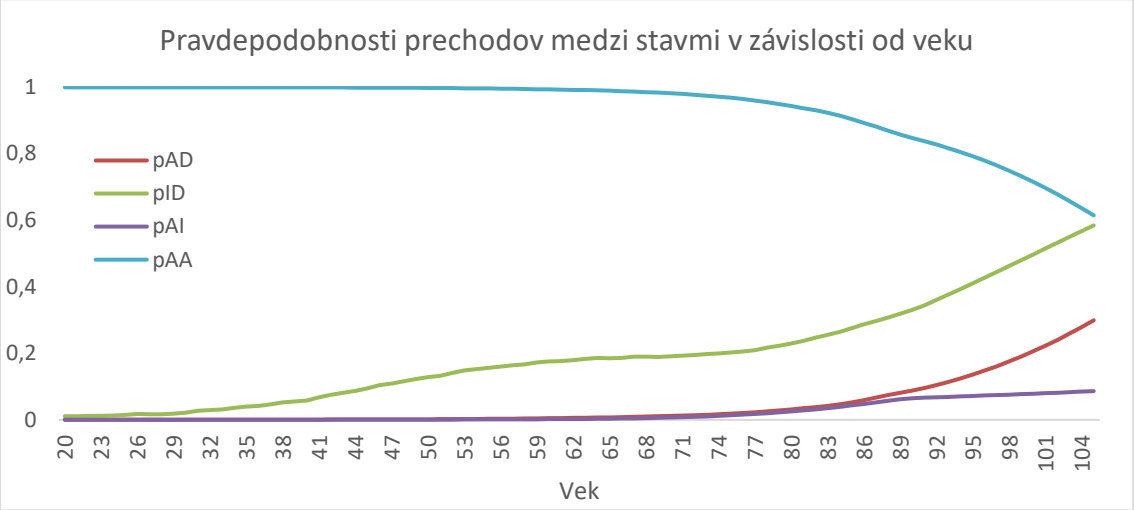
3. Budeme simulovať nehomogénny Markovov reťazec s maticami prechodov $P(t,t+1)$ a s počiatočným rozdelením $\propto = \{\propto_i\}_{i\in S}$ a stavmi $\{i_1, i_2, \dots, i_m\}$ v diskrétnom čase, teda zostrojíme trajektóriu tohto reťazca:
   a. Z diskrétneho počiatočného rozdelenia $\{\propto_i\}_{i\in S}$ vygenerujeme hodnotu $i_{t_0} = i_i$ náhodnej premennej X_{t_0} v čase nula (na začiatku reťazca).
   b. Z diskrétneho rozdelenia $\left\{p_{i_r j}(t_0,t_1)\right\}_{j\in S}$ teda z *r*-tého riadku matice prechodov $P(t_0,t_1)$ vygenerujeme hodnotu i_{t_1}, ktorá predstavuje hodnotu náhodnej premennej X_{t_1}.
   c. Ak $t_c < t_h$ a vygenerovali sme hodnotu náhodnej premennej X_{t_c} potom z rozdelenia $\left\{p_{i_r j}(t_c,t_{c+1})\right\}_{j\in S}$,teda v riadku so stavom i_{t_c} v matici prechodov $P(t_c,t_{c+1})$ vygenerujeme $i_{t_{c+1}}$, čo predstavuje hodnotu náhodnej premennej $X_{t_{c+1}}$.
4. Ak dosiahneme požadovanú hodnotu $t = t_h$ ukončíme generovanie trajektórie. Výsledkom bude realizácia *h* stavov $i_{t_0}, i_{t_1}, \dots, i_{t_h}$, ktoré získame po *h* krokoch (*h* časových úsekoch). Opakovaním tohto algoritmu *n*-krát dostávame *n* trajektórií nehomogénneho Markovovho reťazca v diskrétnom čase.

4 Popis modelu a dát

Na riešenie použijeme viacstavový model, pričom sa zameriame na model s tromi stavmi: zdravý/active **A**, chorý/ill **I** a mŕtvy/dead **D**. Zo stavu zdravý je možné sa presunúť do stavu chorý a aj do stavu mŕtvy. Ak reťazec opustí stav zdravý, už sa do tohto stavu nikdy nevráti. Zo stavu chorý je možný len prechod do stavu mŕtvy, ktorý je absorbčný. Takýto model je použiteľný na modelovanie choroby, ktorá nemá zavedenú liečbu, resp. ide o nevyliečiteľnú chorobu či trvalý následok „navždy“. Preto nie je možný prechod zo stavu chorý do zdravý.

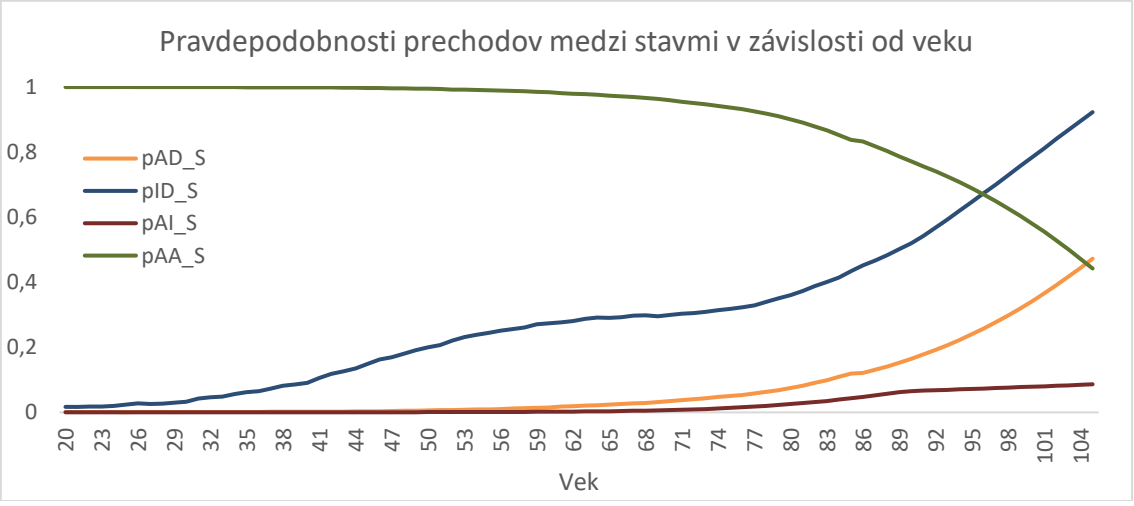
Na demonštráciu opísaných postupov pre nehomogénny Markovov reťazec s diskrétnym časom boli použité dáta zo zdroja *Assicurazioni Sulla Salute: Caratteristiche, modelli attuariali e basi tecniche* od autorov (De Angelis a Di Falco, 2016). Tieto dáta sú obsiahnuté v R package „lifecontingencies“ v dátovom sete "de_angelis_di_falco" a popisujú populáciu Talianska. Dátový set obsahuje viaceré tabuľky, resp. rôzne typy dát. Pre náš účel boli použité pravdepodobnosti prechodov medzi stavmi zdravý → chorý, zdravý → mŕtvy, chorý → mŕtvy vždy pre jednotlivé veky. Pre simuláciu boli použité dáta populácie Talianska, rok 2016 – muži. Druhá sada dát je z populačných dát VDC SAV pre Slovenskú republiku, a to úmrtnosť mužov z roku 2019. Prehľad použitých dát uvádza graf na Obr. 1 a Obr. 2.

Obr. 1: Zobrazenie vstupných dát – muži, populácia Talianska



Zdroj: Vlastné spracovanie v MS Excel

Obr. 2: Zobrazenie vstupných dát – muži, populácia Slovenska



Zdroj: Vlastné spracovanie v MS Excel

Uvažujeme model ako nástroj na generovanie trajektórii nehomogénneho Markovovho reťazca s diskrétnym časom. Keďže výpočty a simulácie Markovových reťazcov pracujú s maticami prechodov, je potrebné takéto matice definovať a vytvoriť. Postupujeme podľa pravidiel pre trojstavový model v package „markovchain“. Všeobecne matica pravdepodobností prechodov pre trojstavový Markovov reťazec a osobu vo veku *t* vyzerá nasledovne:

$$P(t,t+1) = \begin{bmatrix} p_{AA}(t,t+1) & p_{AI}(t,t+1) & p_{AD}(t,t+1) \\ 0 & p_{II}(t,t+1) & p_{ID}(t,t+1) \\ 0 & 0 & 1 \end{bmatrix} \quad (8)$$

Hodnoty pravdepodobností prechodov p_{AI}, p_{AD} a p_{ID} boli použité priamo zo zdroja dát a pravdepodobnosti prechodov p_{AA} a p_{II} boli dopočítané, aby bolo možné vytvoriť matice $P(t,t+1)$.

$$p_{AA} = 1 - p_{AI} - p_{AD} \qquad p_{II} = 1 - p_{ID} \quad (9)$$

Simulovaním nehomogénneho Markovovho reťazca vytvoríme generované trajektórie, ktoré predstavujú zobrazenie vývoja stavov modelu v čase, resp. vývoj stavu poistenej osoby. Výsledky uložíme do matice ${}^{(z)}M$ s *n* riadkami (počet simulácii) a *h* stĺpcami (pozorovaný časový úsek vyjadrený ich počtom). Prvky takejto matice budeme analyzovať zohľadňujúc vlastnosti trojstavového modelu a obsahu použitých dát.

Na základe analýzy výstupu stochastickej simulácie (za použitia dostatočného počtu simulácii) je možné z *h*-teho stĺpca odhadnúť percentuálne rozdelenie poistených do jednotlivých stavov$\{A, I, D\}$ po *h* časových úsekoch.

5 Výsledky a diskusia

V tejto časti uvedieme výsledky ako analytického postupu, tak aj simulácií, resp. stochastického prístupu s použitím programu R a kde to bude možné, tieto výsledky navzájom porovnáme. Praktická časť bola realizovaná pre osobu so vstupným vekom 55 rokov a analyzované obdobie uvažujeme 15 rokov. Využívame dáta o populácii mužov v Taliansku a druhú sadu dát s populáciou v SR, pričom sa uplatňuje nehomogénny Markovov reťazec s tromi stavmi (živý (**A**), chorý (**I**), mŕtvy (**D**)) na modelovanie správania (poistenej) osoby vo veku 55 rokov. Nehomogénny Markovov reťazec používa pravdepodobnosti prechodu, alebo maticu prechodov odlišnú pre jednotlivé veky, teda hovoríme, že ide o vekovo závislé pravdepodobnosti. Takýto prístup zodpovedá aj realite, kde pravdepodobnosti vzniku choroby, alebo pravdepodobnosti úmrtia sú obvykle závislé od veku sledovanej osoby.

Prvý pohľad na sledovanú populáciu realizujeme analytickým prístupom pomocou „markovchain“ balíka ako aj výpočtom pomocou zoznamu matíc a ich násobením pri zohľadnení vektora prvotného stavu. Postup aplikujeme na prvotný stav Zdravý. Po spracovaní v R pre projekciu na 15 rokov získame vektor absolútnych pravdepodobností pre stavy **A, I, D**.

IT [20] 0,8832272 0,01701 0,09976284
SR [20] 0,7348103 0,0110994 0,2540903

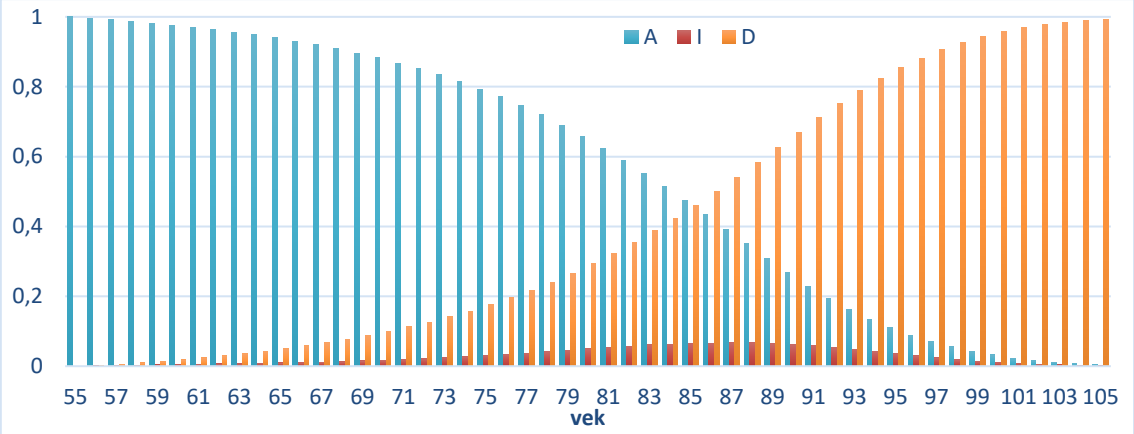
Výsledok pre talianske dáta interpretujeme ako pravdepodobnosť, že zdravá osoba, muž vo veku 55 rokov bude po 15 rokoch zdravá s pravdepodobnosťou 0,8832272, chorá s pravdepodobnosťou 0,01701 a v stave mŕtva s pravdepodobnosťou 0,09976284. Uvedené je možné vyjadriť aj v percentách, kde následne interpretujeme podiel zdravých/chorých/mŕtvych mužov po 15 rokoch projekcie s počiatočným vekom 55 rokov a počiatočným stavom zdravý nasledovným vektorom 88,32 %, 1,70 % a 9,98 %.

Analogicky, pre slovenské dáta, interpretujeme výsledky ako pravdepodobnosť, že zdravá osoba, muž vo veku 55 rokov bude po 15 rokoch zdravá s pravdepodobnosťou 0,7348103, chorá s pravdepodobnosťou 0,0110994 a v stave mŕtva s pravdepodobnosťou 0,2540903. Uvedené je možné vyjadriť aj v percentách, ktoré následne interpretujeme ako podiel zdravých/chorých/mŕtvych mužov po 15 rokoch projekcie s počiatočným vekom 55 rokov a počiatočným stavom zdravý nasledovným vektorom 73,48 %, 1,11 % a 25,41 %.

Druhý možný spôsob spracovania nehomogénneho Markovovho reťazca s tromi stavmi je použitie simulácií. Aby bola zachovaná možnosť porovnania, simuláciu vykonáme na rovnakej množine dát populácie Talianska a druhej množine dát populácie Slovenska, teda opäť pre 55 ročného muža na dobu projekcie 15 rokov s počiatočným stavom zdravý. Počiatočné rozdelenie Markovovho reťazca je [1, 0, 0]. Počas simulácií sa generuje postupne zo stavu vo veku x a s použitím pravdepodobností prechodu pre daný vek x stav pre vek $(x+1)$. Generovanie využíva funkciu *sample* so zohľadnením pravdepodobností pre príslušný vek x. Toto je potrebné zopakovať na celú dobu projekcie, teda v našom prípade pre veky od 55 do 70 rokov. Výsledkom je simulácia, resp. predikcia možnej trajektórie Markovovho reťazca. Takýchto trajektórií nagenerujeme veľký počet n, napríklad n = 100 000 simulácií. Výstupy sú ukladané do matice M a následne je možné tieto výsledky analyzovať.

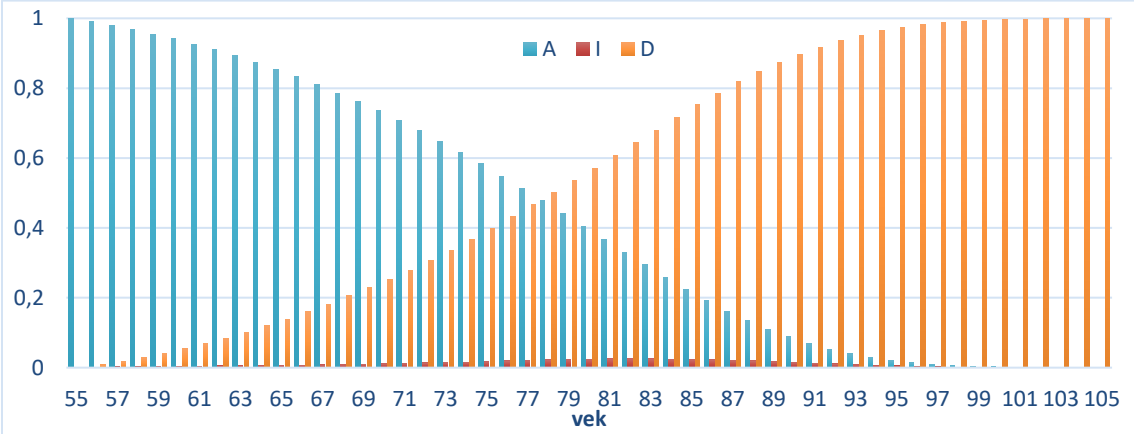
Jedným zo základných pohľadov je odsledovať a porovnať vývoj rozdelenia poistených osôb zdravých na začiatku simulácie. Takýto vývoj pre vstupný vek 55 rokov zobrazujú grafy na Obr. 3 a Obr. 4.

Obr. 3: Projekcia rozdelenia osôb do stavov – muži, populácia Talianska



Zdroj: Vlastné spracovanie v R a MS Excel

Obr. 4: Projekcia rozdelenia osôb do stavov – muži, populácia Slovenska



Zdroj: Vlastné spracovanie v R a MS Excel

Počet simulácii v opísanej analýze je n = 100 000. Následne na konci uvažovanej doby projekcie, teda po 15 rokoch, zistíme počty osôb v jednotlivých stavoch A, I, D a tieto prepočítame na percentuálne podiely. Pri jednom zo simulačných výpočtov boli zistené nasledovné percentuálne podiely jednotlivých stavov po 15 rokoch: 88,39 %, 1,63 %, 9,98 % pre dáta talianskej populácie a nasledovné podiely: 73,38 %, 1,13 %, 25,49 % pre dáta slovenskej populácie.

Porovnanie percentuálnych podielov konečných stavov Markovovho reťazca po 15 rokoch použitím analytického výpočtu a výpočtu pomocou simulácií pre dáta talianskej a slovenskej populácie je uvedené v Tab. 1.

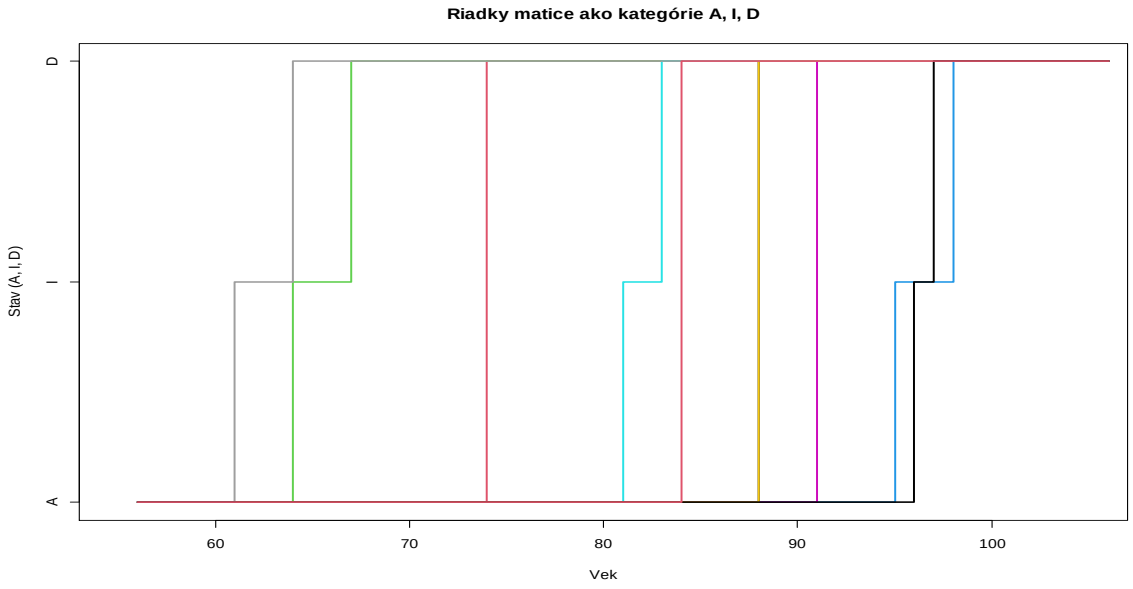
Tab. 1: Porovnanie analytického výpočtu a vybraných simulácií pre populačné dáta IT a SK

	data IT			data SK		
	A	I	D	A	I	D
analytický výpočet	88,32 %	1,70 %	9,98 %	73,48 %	1,11 %	25,41 %
simulácia 100 000 #1	88,39 %	1,63 %	9,98 %	73,38 %	1,13 %	25,49 %
simulácia 100 000 #2	88,26 %	1,74 %	10,00 %	73,50 %	1,16 %	25,34 %
simulácia 1 000 000 #1	88,34 %	1,72 %	9,95 %	73,46 %	1,11 %	25,43 %
simulácia 1 000 000 #2	88,33 %	1,71 %	9,97 %	73,48 %	1,10 %	25,42 %

Zdroj: Vlastné spracovanie v R a MS Excel

Výsledky ukazujú, že simulačné výpočty poskytujú výsledky veľmi blízke výstupom analytického výpočtu, pričom sa môžeme domnievať, že nepresnosť simulačného postupu sa zmenšuje s rastúcim počtom simulácií *n*. Simulačné metódy vo všeobecnej rovine považujeme sa flexibilnejšie a dokážu poskytnúť viaceré podkladové dáta a výstupy, ktoré v analytickom postupe nie sú priamo uvedené. Jedným so zaujímavých pohľadov na výstup simulácie je zobrazenie jednotlivých trajektórií reťazca, kde je možné sledovať presuny medzi stavmi reťazca v jednotlivých krokoch. Z výsledku spracovania simulácií bolo vybraných prvých 10 trajektórií, ktoré sú zobrazené v Obr. 5.

Obr. 5: Vykreslenie trajektórií prvých 10 simulácií



Zdroj: Vlastné spracovanie v R

Keďže výstup spracovania v R obsahuje kompletnú maticu trajektórii, je možné vykonať viacero analýz a pozorovaní. V praxi je určite dôležité v uvažovanom prípade trojstavového modelu zaoberať sa aj analýzou času zotrvania v niektorom z projektovaných stavov. Priemerná doba zotrvania napr. v stave chorý vychádza z počtu období, kedy sa reťazec nachádzal v stave „I“. Tieto počty máme vygenerované pre každú zo simulácií, teda je možné stanoviť počet

stavov „I“ v každej simulácii a následne aj priemernú dobu zotrvania v stave „I“. Podobne je možné stanoviť dobu zotrvania aj v inom stave reťazca, napríklad zdravý „A“. Z výstupu simulácií v R bolo odvodené priemerné trvanie daného stavu v rokoch. Porovnanie doby zotrvania v jednotlivých stavoch pre dáta IT a SK je uvedené v tabuľke Tab. 2.

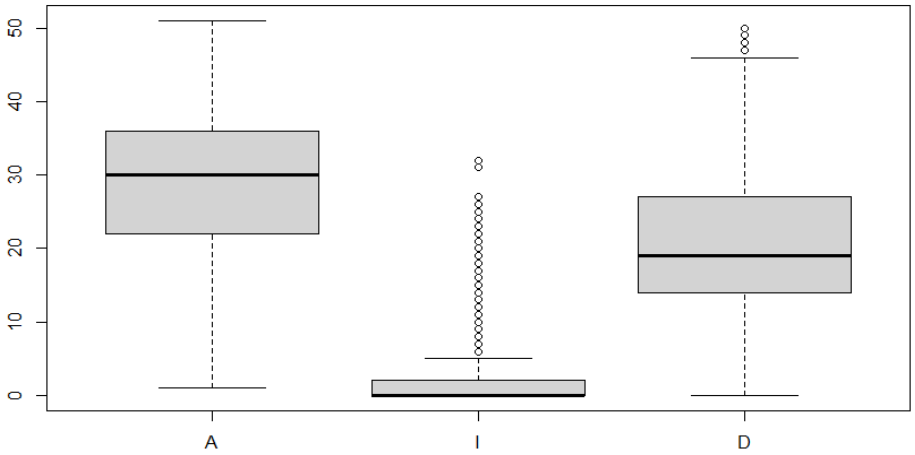
Tab. 2: Porovnanie priemerného simulovaného zotrvania v stave pre dáta IT a SK

	A	I
IT	28,6959	1,43415
SK	22,2554	0,54653

Zdroj: Vlastné spracovanie v R a MS Excel

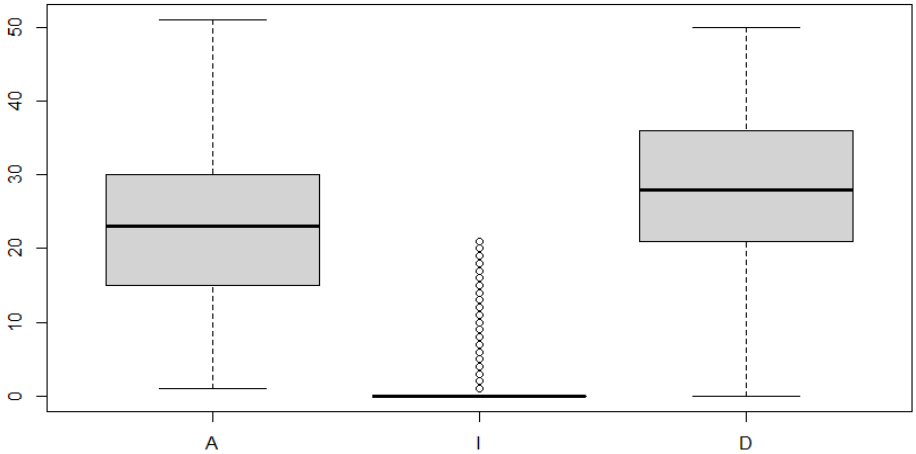
Zo simulačného behu máme k dispozícii 100 000 trajektórií, takže má zmysel urobiť nad týmito dátami štatistický rozbor, napríklad použitím BoxPlotu, ako je zobrazené na Obr. 6 a Obr. 7. Typ grafického zobrazenia BoxPlot dáva viaceré informácie o rozdelení, resp. o rozložení dát. Boxplot zobrazuje medián a pomocou kvantilov aj rozptyl alebo „šírku“ rozdelenia dát. Takisto dokáže zobraziť minimálne a maximálne hodnoty a prípadne extrémne hodnoty (outliers).

Obr. 6: BoxPloty pre simulácie zotrvania v stave talianskej populácie



Zdroj: Vlastné spracovanie v R

Obr. 7: BoxPloty pre simulácie zotrvania v stave slovenskej populácie



Zdroj: Vlastné spracovanie v R

Prezentované grafy - BoxPloty dokumentujú základné štatistické vlastnosti dát. Pre talianske dáta pri stave živý (A) vidíme, že najčastejšia hodnota, teda medián je okolo 30 rokov, a prvý kvartil sa nachádza na úrovni približne 23, čo znamená, že 75 % trajektórií obsahuje status A v dĺžke 23 rokov alebo viac. V prípade stavu chorý (I) je medián veľmi nízko, len mierne nad 1 a ¾ pozorovaní sa nachádza v oblasti do 3 rokov.

Pre slovenské dáta pri stave živý (A) vidíme, že najčastejšia hodnota, teda medián je okolo 22 rokov a prvý kvartil sa nachádza na úrovni približne 15. Stav chorý (I) má medián veľmi nízko, len mierne nad nulou. Avšak pozorujeme, že stav (I) obsahuje v oboch grafoch aj veľké množstvo extrémnych hodnôt (outliers), ktoré sú zobrazené krúžkami.

6 Záver

Použitie Markovových reťazcov pre viacstavové modely predstavuje často aplikovaný prístup na modelovanie vývoja stavov v súvislosti s výskytom chorôb a dlhodobým poskytovaním zdravotnej starostlivosti. Použitím simulácií pre nehomogénne Markovove reťaze prostredníctvom generovania ich trajektórií sa vytvára model, ktorý verne napodobňuje reálny vývoj stavov poistených osôb v čase. V rámci metódy Monte Carlo sme sa zároveň venovali vplyvu počtu simulácií na presnosť získaných výsledkov. Vzhľadom na povahu dostupných údajov sme výpočty realizovali v diskrétnom čase, a to na ročnej báze.

Analyzované údaje sa vzťahujú na mužskú populáciu v Taliansku. Stav chorý (I) podľa autora balíka markovchain predstavuje invaliditu charakterizovanú tzv. nesebestačnosťou poistenej osoby, podobnú invalidite súvisiacej s Alzheimerovou chorobou. Pre účely porovnania je spracovaná aj druhá sada údajov, ktorá popisuje mužov slovenskej populácie. Pomocou uvedeného modelovania implementovaného v programovacom jazyku R sme získali rozdelenie poistených osôb do jednotlivých stavov v priebehu niekoľkých rokov, a to na základe štatistického spracovania výsledkov simulácií. Interpretácia výstupov je doplnená grafickými a vektorovými znázorneniami.

Ďalším využitím použitého simulačného modelu je odhad očakávaného trvania zdravého stavu (A), ako aj odhad trvania obdobia stráveného v stave choroby (I). Táto analýza bola realizovaná pre vybrané veky a výsledky boli graficky prezentované prostredníctvom stĺpcových grafov a BoxPlotov. Pri prezentácii výsledkov je nutné poukázať aj na rozdiely medzi dvomi použitými sadami dát. Už pri priamom porovnaní úmrtností je viditeľný rozdiel medzi úmrtnosťou mužskej populácie Talianska a Slovenska. Následne pri výsledkoch pozorujeme výrazný rozdiel či už pri vektore absolútnych pravdepodobností ako aj pri analyzovaní doby zotrvania v konkrétnom stave.

Modelovanie vývoja stavov v čase v uvedenej oblasti pomocou simulácií Markovových reťazcov predstavuje vhodný a efektívny nástroj na riešenie týchto typov úloh.

Literatúra

1. Di Falco, L., De Angelis, P. (2016). Assicurazioni sulla salute: caratteristiche, modelli attuariali e basi tecniche. In *Assicurazioni sulla salute: caratteristiche, modelli attuariali e basi tecniche* (pp. 1-212). Società editrice il Mulino spa.
2. Fernandez-Morales, A. (2015). Application of a discrete-time Markov chain simulation in insurance. *International Journal of Recent Contributions from Engineering, Science & IT (iJES)*, *3*(3), 27-32.
3. Geyer, C. J., Johnson, L. T. (2013). Mcmc: Markov chain monte carlo. *R package version 0.9-2*. Dostupné na http://CRAN.R-project.org/package=mcmc.
4. Janková, K., Kilianová, S., Brunovský, P., Bokes, P. (2014). Markovove reťazce a ich aplikácie. Bratislava: Epos.

5. Li, M., Dushoff, J., Bolker, B. M. (2018). Fitting mechanistic epidemic models to data: A comparison of simple Markov chain Monte Carlo approaches. *Statistical methods in medical research*, *27*(7), 1956-1967.
6. Mucha, V., Faybíková, I., Krčová, I. (2022). Use of Markov Chain Simulation in Long Term Care Insurance. *Statistika: Statistics & Economy Journal*, *102*(4).
7. Spedicato, G. A., Kang, T. S., Yalamanchi, S. B., Yadav, D., Cordón, I. (2016). The markovchain package: a package for easily handling Discrete Markov Chains in R. Dostupné na https://cran.rproject.org/web/packages/markovchain/vignettes/an_introduction_to_markovchain_package.pdf citované január 2026
8. Spedicato, G. A. (2026). Easy Handling Discrete Time Markov Chains. Dostupné na https://ftp2.uib.no/cran/web/packages/markovchain/markovchain.pdf citované január 2026
9. Twumasi, C., Asiedu, L., Nortey, E. N. (2019). Markov chain modeling of HIV, tuberculosis, and Hepatitis B Transmission in Ghana. *Interdisciplinary perspectives on infectious diseases*, *2019*(1), 9362492.