

UNIVERZITA KOMENSKÉHO V BRATISLAVE

Fakulta matematiky, fyziky a informatiky

Katedra aplikovanej matematiky a štatistiky

**MODELOVANIE ZAMESTNANOSTI V  
PRIEMYSE V SLOVENSKEJ  
REPUBLIKE**

Diplomová práca

Peter Kudla

9.1.9 Aplikovaná matematika

Ekonomická a finančná matematika

Vedúci práce: RNDr. Katarína Boďová, PhD.

BRATISLAVA 2010

# Čestné prehlásenie

Prehlasujem, že som prácu vypracoval samostatne s využitím svojich poznatkov a s použitím uvedenej literatúry.

V Bratislave 26. apríla 2010

---

podpis študenta

# Pod'akovanie

Ďakujem svojej školiteľke RNDr. Kataríne Boďovej, PhD. za ochotu, cenné rady a usmernenie pri písaní záverečnej práce.

# Abstrakt

KUDLA, Peter: *Modelovanie zamestnanosti v priemysle v Slovenskej Republike*. [Diplomová práca] - Univerzita Komenského v Bratislave. Fakulta matematiky, fyziky a informatiky; Katedra aplikovanej matematiky a štatistiky. - Vedúca: RNDr. Katarína Boďová, PhD., Bratislava, 2010, 68 strán.

Cieľom diplomovej práce je zostaviť model zamestnanosti v priemysle v SR pomocou simulácie preskupovania agentov (zamestnancov) medzi firmami. Problém je formulovaný ako diskretný náhodný Markovov proces, motivovaný príbuznými modelmi vo fyzike materiálov, v chémii a v populačnej biológii. Odvodíme stochastický algoritmus, ktorý použijeme na simuláciu zamestnanosti. Určíme kľúčové vlastnosti zamestnancov a firiem a zostavíme modely pre tieto vlastnosti. Výsledky porovnáme s reálnymi dátami.

**Kľúčové slová:** Chapmanove-Kolmogorovove rovnice, Gillespieho algoritmus, algoritmus stochastickej simulácie, metóda Monte Carlo, adaptovaná funkcia rýchlosti, model zamestnanosti

# Abstract

KUDLA, Peter: *Modeling of employment in industry in the Slovak Republic*. [Master thesis] – Comenius University in Bratislava, Faculty of Mathematics, Physics and Informatics; Department of Applied Mathematics and Statistics. - Tutor: RNDr. Katarína Boďová, PhD., Bratislava, 2010, 68 pages.

The aim of this thesis is to establish a model of employment in industry in Slovakia by simulation regrouping agents (employees) among firms. The problem is formulated as a discrete Markov random process, motivated by related models in the physics of materials, in chemistry and population biology. We derive stochastic algorithm that we use to simulate employment. We identify key characteristics of workers and firms, and assemble models for these properties. We compare results with real data.

**Key words:** Chapman-Kolmogorov equations, Gillespie algorithm, stochastic simulation algorithm, Monte Carlo method, propensity function, model of employment

# Obsah

Úvod	12
<b>1 Konštrukcia stochastického algoritmu</b>	<b>14</b>
1.1 Chapmanove - Kolmogorovove rovnice	16
1.2 Algoritmus stochastickej simulácie	20
1.2.1 Algoritmus	22
1.2.2 Generácia náhodných čísel $\xi_1$ a $\xi_2$	23
1.2.3 Detaily simulácie	23
<b>2 Implementácia stochastického algoritmu</b>	<b>26</b>
2.1 Model starnutia	28
2.1.1 Algoritmus pre model starnutia	33
2.1.2 Rýchlostí reakcií a terminálna podmienka	33
2.1.3 Výsledky	34
2.2 Model vzdelávania	39
2.2.1 Voľba parametrov a ďalšia úprava	42
2.2.2 Systematická voľba konštant $\alpha_i$	47
2.2.3 Výsledky	48
2.3 Model zamestnanosti	52
2.3.1 Vzťahy medzi modelmi	55
2.3.2 Výsledky	61
2.4 Možnosti rozšírenia modelu zamestnanosti	63

<b>Záver</b>	<b>65</b>
<b>Literatúra</b>	<b>66</b>

# Zoznam obrázkov

1.1	Určenie typu reakcie. . . . .	24
2.1	Vekové zloženie populácie SR z roku 2008 . . . . .	28
2.2	Vekové zloženie populácie detí z roku 2008 . . . . .	29
2.3	Vekové zloženie produktívnej časti populácie z roku 2008 . . . . .	30
2.4	Schematické zobrazenie reakcií pre model starnutia . . . . .	31
2.5	Reálne čísla a výpočet našou simuláciou. . . . .	35
2.6	Celková populácia SR a percentuálna zmena výpočtu oproti skutočnosti za obdobie 2001 – 2008. . . . .	36
2.7	Ľavá tabuľka obsahuje počty vekovej kategórie <i>M3</i> , ktorej simulácia vyšla najhoršie. Pravá tabuľka popisuje vekovú kategóriu <i>Z2</i> , ktorej simulácia vyšla najlepšie. . . . .	36
2.8	Vekové kategórie <i>M3</i> a <i>Z2</i> a ich percentuálne zmeny oproti skutočnosti za obdobie 2001 – 2008. . . . .	37
2.9	Percentuálne zmeny výpočtu oproti skutočnosti všetkých vekových kategórií za obdobie 2001 – 2008. . . . .	38
2.10	Schematické zobrazenie reakcií pre model vzdelávania . . . . .	42
2.11	Schematické zobrazenie reakcií pre upravený model vzdelávania . . . . .	47
2.12	Reálne čísla a výpočet našou simuláciou. . . . .	49
2.13	Vzdelanostná štruktúra populácie SR a percentuálna zmena výpočtu oproti skutočnosti za obdobie 2001 – 2008. . . . .	49



2.14	Ľavá tabuľka obsahuje počty vzdelanostnej kategórie <i>NV</i> , ktorej simulácia vyšla najlepšie. Pravá tabuľka popisuje vzdelanostnú kategóriu <i>VSc</i> , ktorej simulácia vyšla najhoršie. . . . .	50
2.15	Vzdelanostné kategórie <i>NV</i> a <i>VSc</i> a ich percentuálne zmeny oproti skutočnosti za obdobie 2001 – 2008. . . . .	51
2.16	Percentuálne zmeny výpočtu oproti skutočnosti všetkých vzdelanostných kategórií za obdobie 2001 – 2008. . . . .	52
2.17	Histogram firiem zostavený z výberového zisťovania . . . . .	53
2.18	Histogram firiem zostavený zo súhrnných tabuliek o počtoch zamestnancov . . . . .	54
2.19	Schematické zobrazenie reakcií pre model zamestnanosti . . . . .	55
2.20	Vzťahy medzi modelmi. . . . .	55
2.21	Ilustračné zobrazenie vekovej štruktúry zamestnancov pre 4 firmy. . . . .	57
2.22	Určenie kategórie firmy, keď nastala reakcia v modeli starnutia . . . . .	60
2.23	Určenie veku zamestnanca, keď nastala reakcia v modeli zamestnanosti . . . . .	60
2.24	Vekové kategórie modelu zamestnanosti. . . . .	62
2.25	Vzdelanostné kategórie modelu zamestnanosti. . . . .	63
2.26	Kategórie firiem modelu zamestnanosti. . . . .	63

# Zoznam tabuliek

2.1	Reakcie, rýchlosti reakcií a adaptované funkcie rýchlostí pre model starnutia . . . . .	32
2.2	Reakcie, rýchlosti reakcií a adaptované funkcie rýchlostí pre model vzdelávania . . . . .	43
2.3	Reakcie, rýchlosti reakcií a adaptované funkcie rýchlostí pre upravený model vzdelávania . . . . .	48
2.4	Reakcie, rýchlosti reakcií a adaptované funkcie rýchlostí pre model zamestnanosti . . . . .	56
2.5	Rýchlosti reakcií pre model zamestnanosti. . . . .	62

# Zoznam použitých skratiek

ASS - algoritmus stochastickej simulácie

ME - master equation

ODR - obyčajná diferenciálna rovnica

ZŠ - základná škola

SŠ - stredná škola

VŠ - vysoká škola

NLP - nelineárne programovanie

SR - Slovenská Republika

# Úvod

Hospodárska kríza, ktorá dominovala svetovej ekonomike v posledných kvartáloch, mala za následok okrem iného aj výrazné poklesy v svetovom dopyte po tovaroch a službách. Firmy boli nútené kvôli zníženým objednávkam, pristúpiť k redukcii nákladov a rušiť pracovné miesta. Rastúca nezamestnanosť a nedostatok nových pracovných príležitostí nútila kompetentné osoby a odbornú verejnosť, zaoberať sa danou problematikou. Modelovanie zamestnanosti a odhady vývoja v budúcnosti sú preto v súčasnej dobe veľmi aktuálne.

Väčšina modelov zaoberajúca sa otázkou zamestnanosti, porovnáva rôzne ekonomické ukazovatele v rôznych krajinách. Náš model však neskúma túto problematiku len z ekonomického pohľadu, no sústreďujeme sa vo väčšej miere na sledovanie vlastností uchádzačov o prácu a charakteristiku jednotlivých firiem, ktoré pôsobia na trhu. Úspech zamestnancov na trhu práce je ovplyvnený viacerými činiteľmi ako vek, dosiahnuté vzdelanie, pohlavie, či nadobudnuté skúsenosti. Tento prístup je inovatívny a prináša nové prvky do modelovania zamestnanosti. Modelovanie vlastností pracujúcich a ich pohyb na trhu práce umožňuje robiť presnejšie odhady do budúcnosti a pružnejšie reagovať na zmeny na trhu práce. Rovnako dôležitá je aj identifikácia požadovaných vlastností u uchádzačov o prácu a poukázanie na nedostatky v súčasnom stave. Model je možné zostaviť dvoma spôsobmi. Pomocou sústavy lineárnych obyčajných diferenciálnych rovníc alebo pomocou stochastickej simulácie. My sme sa rozhodli model skonštruovať ako náhodný Markovov proces, teda druhým spôsobom.

Práca je rozdelená na dve časti. V prvej kapitole na ilustračnom príklade uká-

žeme, ako jednoducho je možné sledovať zložitý chemický proces. Od vysoko dimenzionálnej sústavy lineárnych diferenciálnych rovníc prejdeme k Markovovým procesom a skonštruujeme algoritmus, pomocou ktorého budeme modelovať zamestnanosť.

Druhá kapitola sa skladá zo štyroch častí, v ktorej zostrojený algoritmus aplikujeme na náš ekonomický model. Postupne vytvoríme simuláciu starnutia, vzdelávania a zamestnanosti, pričom v poslednom modeli využijeme predchádzajúce. Na konci každej časti prezentujeme výsledky a porovnáme ich s reálnymi dátami. V poslednej časti kapitoly načrtneme, ako by sa dala naša simulácia vylepšiť.

Súčasťou práce je priložené DVD, ktoré obsahuje algoritmus stochastickej simulácie, podrobnejšie výsledky ako prezentujeme v práci a dáta, ktoré sme v práci použili.

# Kapitola 1

## Konštrukcia stochastického algoritmu

Na začiatok oboznámime čitateľa s hlavným problémom práce. Načrtujeme postup celej simulácie preskupovania agentov v priemysle v Slovenskej republike na ilustračnom príklade z chémie. Postupujeme podľa článku [9].

Uvažujme proces, v ktorom  $N$  rôznych druhov chemických molekúl môže navzájom reagovať v  $M$  rôznych typoch chemických reakcií. Napríklad, molekula A a molekula B sa môžu dostať do vzájomnej blízkosti, spustiť reakciu a vytvoriť novú molekulu C. Ak poznáme počiatočnú polohu a počet molekúl, môžeme sledovať ako sa pohybujú a vzájomne reagujú. Na prvý pohľad je však zrejmé, že takéto „dynamické“ procesy už pri malom objeme chemikálie, nehovoriac o tom ak chceme tento proces pozorovať počas dlhého časového intervalu, môže byť veľmi náročné sledovať a príliš zložité numericky simulovať. Preto budeme ignorovať informáciu o polohe a sústredíme sa len na sledovanie počtu molekúl každého typu. Takéto zjednodušenie je platné, pokiaľ je naša chemikália veľmi dobre premiešaná, to znamená, budeme predpokladať, že všetky molekuly sú rovnomerne rozdelené v celej nádobe. Pre ilustráciu uvedieme jednoduchý príklad. Pri elektrolýze roztoku chloridu sodného  $NaCl$ , sa tento štiepi na kladné ióny sodíka  $Na^+$  a záporné ióny chlóru  $Cl^-$ . Kladné ióny sú priťahované k zápornej elektróde (*katóde*), z ktorej si

ión  $Na^+$  zoberie elektrón a stáva sa z neho elektricky neutrálna častica - atóm sodíka  $Na$ . Záporné ióny  $Cl^-$  sú priťahované ku kladnej elektróde (*anóde*), ktorá príma elektrón a dva atómy sa zlučujú do elektricky neutrálnej molekuly chlóru  $Cl_2$ . Predstavme si teraz, že do veľkej nádoby s prepážkou v strede, do jednej časti nalejeme čistú vodu a do druhej roztok kuchynskej soli. Prepážku odstránime. Do tej časti nádoby, v ktorej je roztok soli, vložíme elektródy. Okamžite po spojení elektrického obvodu dochádza k elektrolyze a môžeme sledovať, ako sa na kladnej elektróde uvoľňuje plynný chlór. Keď elektródy vložíme do časti nádoby, kde je naliata čistá voda a zapneme elektrický prúd, nič sa nestane. Časom sa kvapalina premieša a postupne sa chlorid sodný bude štiepiť, ale je zrejme, že to nejaký čas potrvá. A teda, aj keď sme sledovali tú istú chemickú reakciu v tej istej nádobe, rýchlosť s akou prebiehala bola rôzna.

Majme teda v počiatočnom čase  $N$  rôznych molekúl, ktorých počet poznáme. Zostrojíme *stavový vektor*,

$$\mathbf{X}(t) = (X_1(t), X_2(t), \dots, X_N(t)),$$

kde  $X_i(t)$  je nezáporné celé číslo, ktoré nám udáva počet molekúl typu  $i$  v čase  $t$ .

Stavový vektor  $\mathbf{X}(t)$  sa mení vždy, keď nastane reakcií. Keďže predpokladáme rovnomerné rozdelenie hustoty molekúl, budeme uvažovať len o pravdepodobnosti nastania reakcie, založenej na okamžitom stave uvažovaného systému v čase  $t$ . Toto prirodzene vedie k *Chapmanovým-Kolmogorovovým rovniciam* [7], čo je sústava obyčajných diferenciálnych rovníc pre rôzne stavy systému. V angličtine sa tejto sústave hovorí *master equation* [12], ďalej budeme písať len ME.  $k$ -ta rovnica v čase  $t$  určuje pravdepodobnosť, že systém sa nachádza v stave  $k$ . Je dôležité si uvedomiť, že počet rovníc neudáva počet typov molekúl  $N$ , ale počet všetkých možných stavov systému. Ten závisí nielen od počtu všetkých molekúl  $\mathbf{X}(t)$ , ale aj od množstva reakcií  $M$ , a preto môže byť dimenzia tejto sústavy diferenciálnych rovníc veľmi veľká.

Napríklad, majme  $N = 4$  molekuly  $A, B, C, D$  a  $M = 2$  reakcie:

- $A + B \rightarrow C + D$ ;
- $C + D \rightarrow A + B$  (spätná reakcia).

Ak je na začiatku  $K$  molekúl typu  $A$  a  $K$  molekúl typu  $B$ , a žiadne molekuly typu  $C$  alebo  $D$ , takže

$$\mathbf{X}(0) = (K, K, 0, 0),$$

potom stavový vektor  $\mathbf{X}(t)$  môže nadobudnúť  $K + 1$  možných stavov:

$$\begin{aligned} & (K, K, 0, 0), \\ & (K - 1, K - 1, 1, 1), \\ & (K - 2, K - 2, 2, 2), \\ & \quad \vdots \\ & (0, 0, K, K). \end{aligned}$$

Vo všeobecnosti má ME obrovskú dimenzie, že nie je možné vypočítať ju analyticky ani numericky. Namiesto riešenia takejto sústavy diferenciálnych rovníc zvolíme iný prístup, ktorý nepriamo využíva ME. Budeme ho označovať *algoritmus stochastickej simulácie* (ASS), alebo *Gillespieho algoritmus* [6]. Ten namiesto riešenia vysoko dimenzionálneho systému ODR zisťuje iba realizácie stavového vektora  $(t, X_t)$  tak, že pravdepodobnosť stavu tohto vektora odpovedá pravdepodobnosti získanej použitím ME.

## 1.1 Chapmanove - Kolmogorovove rovnice

Na ilustráciu, ako zostaviť sústavu Chapmanových - Kolmogorovových rovníc, teda ME, použijeme systém, ktorý vyžaduje štyri druhy chemických molekúl: substrát ( $S_1$ ), enzým ( $S_2$ ), komplex ( $S_3$ ), produkt ( $S_4$ ). Ako modelový systém chemických reakcií zoberieme:







Úlohu konštant - *rýchlosti nastatia reakcií*  $c_1, c_2$  a  $c_3$  vysvetlíme neskôr v tejto časti práce.

Ako už bolo spomínané v predošlej kapitole 1, pri riešení ME neberieme do úvahy informáciu o pozícii alebo objeme jednotlivých molekúl, a preto sa pokúsime popísať systém pomocou stavového vektora  $\mathbf{X}(t)$ . V systéme (1.1)-(1.3) vidíme, že molekula  $S_4$  nikdy nevystupuje v žiadnej reakcii ako reaktant, ale vždy je len produktom. Takže, ak nás nebude zaujímať množstvo vzniknutých produktov, ale len počty molekúl  $S_1, S_2, S_3$ , ktoré do nejakej reakcie vstupujú, môžeme zložku 4 vektora  $\mathbf{X}(t)$  zanedbať a tým zredukovať jeho dimenziu. Napriek tomu pre ilustráciu budeme pracovať s celým stavovým vektorom  $\mathbf{X}(t) \in \mathbb{R}^4$ .

Ak nastane reakcia (1.1), potom sa zložky  $X_1(t)$  a  $X_2(t)$  znížia o jedna a zložka  $X_3(t)$  sa zvýši o jedna. Takže  $\mathbf{X}(t)$  sa zmení na  $\mathbf{X}(t) + \mathbf{v}_1$ , kde

$$\mathbf{v}_1 = (-1, -1, 1, 0).$$

Podobne pre reakcie (1.2) a (1.3) určíme

$$\mathbf{v}_2 = (1, 1, -1, 0) \quad \text{a} \quad \mathbf{v}_3 = (0, 1, -1, 1).$$

Je zrejmé, že reakcia typu (1.1) môže nastať jedine v prípade, ak sa stretne molekula  $S_1$  s molekulou  $S_2$ . Ak je v nejakom čase  $t$  v celom objeme chemikálie veľmi málo molekúl typu  $S_1$  a  $S_2$ , je menej pravdepodobné, že nastane reakcia typu (1.1) oproti prípadu, keby bolo týchto molekúl v celej zlučeneine viac. Použitím takéhoto argumentu môžeme hovoriť o pravdepodobnosti nastatia reakcie a usudzujeme, že táto pravdepodobnosť bude úmerná koncentrácií molekúl  $S_1$  a  $S_2$  v čase  $t$ . Predpokladáme teda, že pravdepodobnosť nastatia tejto reakcie vo veľmi malom časovom intervale  $[t, t + dt)$  je daná ako  $c_1 X_1(t) X_2(t) dt$ , kde  $X_1(t) X_2(t)$

predstavuje pravdepodobnosť kolízie dvoch molekúl, a konštanta  $c_1$  predstavuje rýchlosť reakcie, ktorá okrem iného zabezpečuje aj to, že nie každá kolízia vyústi do reakcie.

Pre druhý typ reakcie, (1.2), iba molekula  $S_3$  zohráva aktívnu úlohu. Preto sa určí pravdepodobnosť tejto reakcie ako  $c_2 X_3(t)dt$ , aby proporcionálne odpovedala počtu molekúl  $S_3$ . Podobne sa určí pravdepodobnosť pre tretí typ reakcie (1.3) ako  $c_3 X_3(t)dt$ .

Takáto štruktúra platí všeobecne pre akýkoľvek súbor *unimolekulárnych* (jedna molekula na ľavej strane) alebo *bimolekulárnych* (dve molekuly na ľavej strane) reakcií. Reakcie v našom modeli budú unimolekulárne. Všeobecne, majme molekuly  $S_1, S_2, \dots, S_N$ , ktoré môžu navzájom reagovať v  $M$  rôznych typoch chemických reakcií. Pre každé  $1 \leq j \leq M$  existuje *vektor zmeny*  $\mathbf{v}_j \in \mathbb{R}^N$ , ktorého  $i$ -ta zložka vyjadruje zmenu počtu molekúl  $S_i$  spôsobenú  $j$ -tou reakciou. Takže jedna reakcia typu  $j$  zmení stavový vektor  $\mathbf{X}(t)$  na  $\mathbf{X}(t) + \mathbf{v}_j$ .

Rýchlosť reakcií závisí od koncentrácie molekúl v danom čase, preto definujeme *adaptovanú funkciu rýchlosti*  $a_j \mathbf{X}(t)$ , ktorá vyjadruje, že pravdepodobnosť nastatia tejto reakcie vo veľmi malom časovom intervale  $[t, t + dt)$  je daná

$$a_j(\mathbf{X}(t)) = \begin{cases} c_j X_m(t) X_n(t) & \text{pre reakciu typu } S_m + S_n \xrightarrow{c_j} *, \text{ kde } m \neq n, \\ c_j \frac{1}{2} X_m(t) (X_m(t) - 1) & \text{pre reakciu typu } S_m + S_m \xrightarrow{c_j} *, \\ c_j X_m(t) & \text{pre reakciu typu } S_m \xrightarrow{c_j} *. \end{cases}$$

S prvým a tretím prípadom sme sa stretli už v našom ilustračnom príklade z chémie. V druhom prípade, pri reakcii molekúl rovnakého typu, využijeme poznatky z kombinatoriky (z celkového počtu  $X_m$  vyberáme dve, pričom nezáleží na poradí).

Najprv odvodíme Chapmanove-Kolmogorovove rovnice pre časový vývoj pravdepodobností stavov. Potom v kapitole 1.2 ukážeme ako sa tento problém dá alternatívne riešiť pomocou algoritmu stochastickej simulácie.

Teraz už máme dostatok informácií, aby sme sa mohli zaoberať veličinou  $P(\mathbf{x}, t) = P[\mathbf{X}(t) = \mathbf{x} | \mathbf{X}(0) = \mathbf{x}]$ . Keďže poznáme pravdepodobnosť každého mož-

ného stavu  $\mathbf{x}$  v čase  $t$ , vypracujeme pravdepodobnosť stavu  $\mathbf{x}$  v čase  $t + dt$  za predpokladu, že  $dt$  je dostatočne malé na to, aby na intervale  $[t, t + dt)$  nastala maximálne jedna reakcia. Poznamenajme, že na to, aby sme boli v stave  $\mathbf{x}$  v čase  $t + dt$ , sú len dve možnosti pre čas  $t$ ; buď bol systém už v čase  $t$  v stave  $\mathbf{x}$  a nedošlo k žiadnej reakcii počas  $[t, t + dt)$ , alebo pre nejaké  $1 \leq j \leq M$  bol systém v stave  $\mathbf{x} - \mathbf{v}_j$  v čase  $t$  a prebehla  $j$ -ta reakcia počas  $[t, t + dt)$ , a tým preniesla systém do stavu  $\mathbf{x}$ . Pripomeňme si *vetu o úplnej pravdepodobnosti* z teórie pravdepodobnosti. Zaujímá nás, aká je pravdepodobnosť udalosti  $A$ . Majme udalosti  $H_0, H_1, H_2, \dots, H_M, H_{M+1}$  ktoré sú (a) disjunktné (najviac jedna udalosť môže nastať) a (b) úplné (jedna z nich musí nastať). Veta o úplnej pravdepodobnosti hovorí

$$P(A) = \sum_{j=0}^{M+1} P(A|H_j)P(H_j), \quad (1.4)$$

kde  $P(A|H_j)$  je podmienená pravdepodobnosť udalosti  $A$  za predpokladu, že nastala udalosť  $H_j$ . V našom prípade,  $A$  je udalosť, že systém je v stave  $\mathbf{x}$  v čase  $t + dt$ ,  $H_0$  je udalosť, že systém je v stave  $\mathbf{x}$  v čase  $t$ ,  $H_j$  pre  $1 \leq j \leq M$  je udalosť, že systém je v stave  $\mathbf{x} - \mathbf{v}_j$  v čase  $t$  a napokon  $H_{M+1}$  je udalosť, že systém je v nejakom inom stave v čase  $t$ . Pre  $1 \leq j \leq M$ ,  $P(A|H_j)$  vyjadruje pravdepodobnosť, že počas časového úseku  $[t, t + dt)$  dôjde k  $j$ -tej reakcii. Z definície adaptovanej funkcie rýchlosti dostávame

$$P(A|H_j) = a_j(\mathbf{x} - \mathbf{v}_j)dt, \quad 1 \leq j \leq M, \quad (1.5)$$

kde  $\mathbf{v}_j$  zodpovedá jednej reakcii. Podobne,  $P(A|H_0)$  je pravdepodobnosť, že nenastala žiadna reakcia po dobu  $[t, t + dt)$ . To sa musí rovnať 1 mínus pravdepodobnosť nastatia každej reakcie, takže

$$P(A|H_0) = 1 - \sum_{j=1}^M a_j(\mathbf{x})dt. \quad (1.6)$$

Nakoniec

$$P(A|H_{M+1}) = 0, \quad (1.7)$$

pretože udalosť  $H_{M+1}$  obsahuje v sebe informáciu o všetkých stavoch vzdialenejších od stavu  $\mathbf{x}$  viac ako o jednu reakciu. Pravdepodobnosť nastatia dvoch reakcií vo veľmi malom časovom intervale je zanedbateľná. Použitím vzťahov (1.5), (1.6) a (1.7) vo vzťahu (1.4) a s použitím definície pre  $P(\mathbf{x}, t)$  dostávame

$$P(\mathbf{x}, t + dt) = \left(1 - \sum_{j=1}^M a_j(\mathbf{x})dt\right) P(\mathbf{x}, t) + \sum_{j=1}^M a_j(\mathbf{x} - \mathbf{v}_j)dt P(\mathbf{x} - \mathbf{v}_j, t).$$

Túto rovnicu môžeme upraviť na

$$\frac{P(\mathbf{x}, t + dt) - P(\mathbf{x}, t)}{dt} = \sum_{j=1}^M (a_j(\mathbf{x} - \mathbf{v}_j)P(\mathbf{x} - \mathbf{v}_j, t) - a_j(\mathbf{x})P(\mathbf{x}, t)).$$

Ak prejdeme k limite  $dt \rightarrow 0$  vidíme, že ľavá strana tejto rovnice sa stáva deriváciou času, čo vedie k ME

$$\frac{dP(\mathbf{x}, t)}{dt} = \sum_{j=1}^M (a_j(\mathbf{x} - \mathbf{v}_j)P(\mathbf{x} - \mathbf{v}_j, t) - a_j(\mathbf{x})P(\mathbf{x}, t)). \quad (1.8)$$

Pripomenieme, že stavový vektor  $\mathbf{x}$  závisí od množiny diskretných hodnôt, ktorá býva obvykle veľmi veľká. ME (1.8) je lineárny systém obyčajných diferenciálnych rovníc, s jednou rovnicou pre **každý** možný stav.

## 1.2 Algoritmus stochastickej simulácie

Sústava Chapman-Kolmogorovových rovníc (ME) má spravidla takú veľkú dimenziu, že sa numericky iba ťažko dá spočítať. Namiesto počítania úplného pravdepodobnostného rozdelenia algoritmus stochastickej simulácie (ASS) sleduje jednotlivé realizácie stavového vektora, pričom jeho okamžitý stav závisí len od predchádzajúceho stavu [3, 4].

Definujme  $P_0(\tau|\mathbf{x}, t)$  ako pravdepodobnosť, že na intervale  $[t, t + \tau)$  nedošlo k žiadnej reakcii ak  $\mathbf{X}(t) = \mathbf{x}$  a uvažujme teraz časový interval  $[t, t + \tau + d\tau)$ . Predpokladajme, že udalosť ktorá nastala počas  $[t + \tau, t + \tau + d\tau)$  je nezávislá od udalosti, ktorá nastala počas  $[t, t + \tau)$ . Potom pravdepodobnosť, že počas  $[t, t + \tau + d\tau)$  nedošlo k žiadnej reakcii je

$$\begin{aligned} P_0[t, t + \tau + d\tau) &= P_0[t, t + \tau)P_0[t + \tau, t + \tau + d\tau) \\ &= P_0[t, t + \tau)(1 - \sum_j P_j[t + \tau, t + \tau + d\tau)), \end{aligned}$$

kde  $P_j[t_1, t_2)$  je pravdepodobnosť, že počas intervalu  $[t_1, t_2)$  prebehla  $j$ -ta reakcia. S využitím definície adaptovanej funkcie rýchlosti môžeme hore uvedené prepísať na

$$P_0(\tau + \delta\tau|\mathbf{x}, t) = P_0(\tau|\mathbf{x}, t) \left( 1 - \sum_{k=1}^M a_k(\mathbf{x})d\tau \right),$$

čo je ekvivalentné s

$$\frac{P_0(\tau + \delta\tau|\mathbf{x}, t) - P_0(\tau|\mathbf{x}, t)}{d\tau} = -a_{sum}(\mathbf{x})P_0(\tau|\mathbf{x}, t), \text{ kde } a_{sum}(\mathbf{x}) := \sum_{k=1}^M a_k(\mathbf{x}).$$

Poslaním  $d\tau \rightarrow 0$  dostávame obyčajnú diferenciálnu rovnicu, ktorá je lineárna a má počiatočnú podmienku  $P_0(0|\mathbf{x}, t) = 1$ . Jej riešením dostávame

$$P_0(\tau|\mathbf{x}, t) = e^{-a_{sum}(\mathbf{x})\tau}. \quad (1.9)$$

Teraz definujme  $p(\tau, j|\mathbf{x}, t)$  ako pravdepodobnosť, že najbližšie nastane  $j$ -ta reakcia a to v časovom intervale  $[t + \tau, t + \tau + d\tau)$  ak  $\mathbf{X}(t) = \mathbf{x}$ . Predpokladáme, že  $d\tau$  je dostatočne krátky čas na to, aby nastala najviac jedna reakcia. S využitím definície pre  $P_0$  a  $a_j$  dostávame

$$p(\tau, j|\mathbf{x}, t)d\tau = P_0(\tau|\mathbf{x}, t)a_j(\mathbf{x})d\tau.$$

Využitím (1.9) máme

$$p(\tau, j | \mathbf{x}, t) = a_j(\mathbf{x}) e^{-a_{sum}(\mathbf{x})\tau}.$$

Tento vzťah upravíme vynásobením pravej strany jednotkou na

$$p(\tau, j | \mathbf{x}, t) = \frac{a_j(\mathbf{x})}{a_{sum}(\mathbf{x})} a_{sum}(\mathbf{x}) e^{-a_{sum}(\mathbf{x})\tau}. \quad (1.10)$$

Funkcia hustoty (1.10) je združenou funkciou hustoty dvoch nezávislých náhodných premenných:

**Index ďalšej reakcie:** zodpovedá diskkrétnej náhodnej veličine ako  $p_j[t, t + \tau) = \frac{a_j(\mathbf{x})}{a_{sum}(\mathbf{x})}$ . Takže pravdepodobnosť  $j$ -tej reakcie je úmerná  $a_j(\mathbf{x})$ .

**Čas do ďalšej reakcie:** Zodpovedá spojitej náhodnej veličine s exponenciálnym rozdelením pravdepodobnosti, ktorej funkcia hustoty je  $a_{sum}(\mathbf{x}) e^{-a_{sum}(\mathbf{x})\tau}$ .

Táto separácia náhodných premenných nám uľahčí numerickú simuláciu. Umožní nám nezávisle simulovať index reakcie a čas, kedy k nej dôjde. Obidve rozdelenia môžeme odvodiť z rovnomerného rozdelenia  $U([0, 1])$ . Teraz uvedieme algoritmus stochastickej simulácie, ktorý sa uplatňuje v širokom spektre fyzikálnych, chemických a biologických aplikácií ako napríklad [2, 5, 10, 15, 18, 11].

### 1.2.1 Algoritmus

1. Inicializujeme stavový vektor  $\mathbf{X}(0) = \mathbf{x}_0$ .
2. Vypočítame hodnoty adaptovaných funkcií rýchlostí  $\{a_k(\mathbf{X}(t))\}_{k=1}^M$  a ich sumu  $a_{sum}(\mathbf{X}(t)) := \sum_{k=1}^M a_k(\mathbf{X}(t))$ .
3. Vygenerujeme dve náhodné čísla  $\xi_1$  a  $\xi_2$  z rovnomerného rozdelenia  $U([0, 1])$  a určíme  $j$ , teda typ reakcie k akej došlo, ako najmenšie celé číslo, ktoré vyhovuje nerovnici  $\sum_{k=1}^j a_k(\mathbf{X}(t)) > \xi_1 a_{sum}(\mathbf{X}(t))$ .
4. Určíme, koľko času prešlo, kým nastala  $j$ -ta reakcia  $\tau = \frac{\ln(1/\xi_2)}{a_{sum}(\mathbf{X}(t))}$  a obnovíme  $\mathbf{X}(t + \tau) = \mathbf{X}(t) + \mathbf{v}_j$ . Aktualizujeme čas  $t$  na  $t + \tau$ .

5. Vrátime sa na krok číslo 2.

V poslednom kroku sa testuje koncová podmienka. Napríklad, algoritmus skončí, keď  $t$  prekročí nejakú dopredu stanovenú hodnotu, keď nejaká zložka vektora  $\mathbf{X}(t)$  prekročí vrchnú alebo spodnú hranicu, alebo keď nám prebehne dostatočný počet iterácií.

### 1.2.2 Generácia náhodných čísel $\xi_1$ a $\xi_2$

Náš model zamestnanosti je naprogramovaný v MATLAB-e. Rovnomerne rozdelené náhodne čísla  $\xi_1$  a  $\xi_2$  generujeme jeho funkciou RAND.

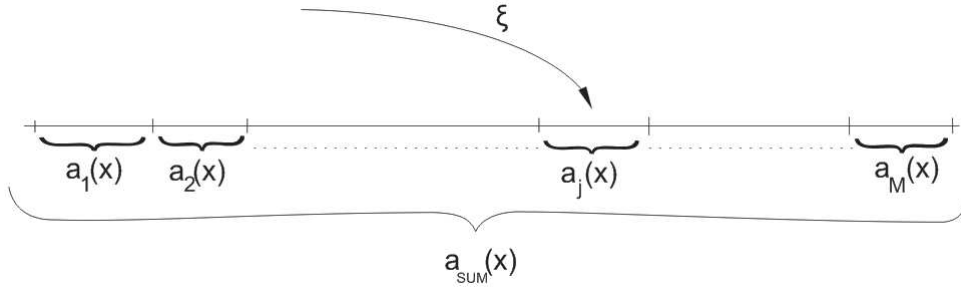
V každej iterácii potrebujeme novú dvojicu náhodných čísel. Tvorcovia programu udávajú, že MATLAB je schopný vygenerovať viac ako  $2^{1492}$  náhodných čísel kým sa niektoré z nich zopakuje. Častým volaním funkcie RAND však čísla nie sú „úplne náhodné“. Pri ich generácii vzniká určitá pravidelnosť. Je to spôsobené resetovaním počiatočného stavu generátora.

Problém jednoducho vyriešime tým, že na začiatku si vygenerujeme dostatočný počet dvojíc náhodných čísel jedným zavolaním funkcie RAND. Potom v každej iterácii zoberieme novú dvojicu, kde máme zaručené, že medzi týmito náhodnými číslami nie je žiadna korelácia.

### 1.2.3 Detaily simulácie

V tejto podkapitole veľmi stručne vysvetlíme, ako kroky 3 a 4 algoritmu náhodnej simulácie vyplývajú zo vzťahu (1.10). Našou úlohou je určiť celé číslo  $j$  z intervalu 1 až  $M$  pre identifikáciu reakcie ktorá nastala a reálne číslo  $\tau$ , ktoré nám určí, koľko času do tejto reakcie ubehlo.

Index nasledujúcej reakcie zvolíme podľa adaptovanej funkcie rýchlosti. Konkrétne, ak je jednotkový interval rozdelený na  $M$  podintervalov, kde  $j$ -ty subinterval má dĺžku  $a_j(\mathbf{x})/a_{sum}(\mathbf{x})$ , potom náhodne zvolíme bod z tohto intervalu tak,



Obr. 1.1: Určenie typu reakcie.

aby pravdepodobnosť, že tento bod leží v  $j$ -tom subintervale bol  $a_j(\mathbf{x})/a_{sum}(\mathbf{x})$ . Toto je znázornené na obrázku 1.1 a vyjadrené v 3. kroku ASS.

Zopakujme si, že exponenciálne rozdelená náhodná premenná s parametrom  $\lambda > 0$  má funkciu hustoty v tvare

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{pre } x \geq 0, \\ 0 & \text{pre } x < 0. \end{cases}$$

A teda, ak  $X$  je exponenciálne rozdelená náhodná premenná s parametrom  $\lambda$ , potom pre ľubovoľné  $0 < a < b$  platí

$$P(a < X < b) = \int_a^b \lambda e^{-\lambda x} dx = e^{-\lambda a} - e^{-\lambda b}.$$

Teraz,

$$P(a < X < b) = P(-b < -X < -a) = P(e^{-\lambda b} < e^{-\lambda X} < e^{-\lambda a}),$$

z čoho vidíme, že

$$P(e^{-\lambda b} < e^{-\lambda X} < e^{-\lambda a}) = e^{-\lambda a} - e^{-\lambda b}.$$

Toto ukazuje, že náhodná veličina  $Z := e^{-\lambda X}$  má rovnomerné rozdelenie na intervale  $[0, 1]$ . Pravdepodobnosť, že  $Z$  leží v nejakom subintervale intervalu  $[0, 1]$  je



daná dĺžkou tohto subintervalu.  $X$  môžeme teda generovať pomocou  $\ln(1/Z)/\lambda$ , kde  $Z$  je rovnomerne rozdelená náhodná premenná na intervale  $[0, 1]$ . Toto zodpovedá 4. kroku ASS pre  $\lambda = a_{sum}(\mathbf{X}(t))$ .

## Kapitola 2

# Implementácia stochastického algoritmu

V kapitole 1 sme vysvetlili, ako možno simulovať diskkrétne sa meniaci náhodný proces pomocou stochastickej numerickej simulácie. Dôležitou vlastnosťou tohto algoritmu je, že je nezávislý od voľby časovej jednotky. Hlavným cieľom práce je aplikovať tento algoritmus, v praxi používaný aj na skúmanie chemických reakcií [6, 9, 12], na modelovanie zamestnanosti v SR. Na to, aby sme mohli stochastický algoritmus použiť, potrebujeme správne identifikovať zmysluplné reakcie v našej ekonomickej aplikácii. Potom im určíme pravdepodobnosti s akými nastávajú pre rôzne relevantné scenáre a nasimulujeme počas obdobia niekoľkých rokov.

V prvom kroku určíme vlastnosti agentov (reaktantov v príklade chemických reakcií). Podľa týchto vlastností sa zamestnávateľa, teda firmy, rozhodujú, či daného uchádzača o zamestnanie príjmu alebo dajú prednosť inému uchádzačovi. V skutočnosti má zamestnávateľ dostupné množstvo údajov o uchádzačoch o zamestnanie, ako vek, dosiahnuté vzdelanie, predchádzajúce pracovné skúsenosti, znalosti a schopnosti, záľuby a iné. Z mena a adresy sa dá určiť, či sa jedná o muža alebo ženu a trvalé bydlisko. Na prípadnom ústnom pohovore môže zamestnávateľ ďalej zistiť, v akom rodinnom stave sa daný uchádzač o zamestnanie nachádza, a teda zistí, či je ochotný cestovať, prípadne, či nebude problém poslať ho pracovať na

pobočku v zahraničí. Konečné rozhodnutie, či už zamestnávateľa alebo uchádzača o zamestnanie, určite ešte ovplyvní veľa ďalších faktorov, napríklad to môže byť národnosť, zdravotný stav alebo finančné ohodnotenie.

Aby model fungoval čo najpresnejšie, všetky hore spomenuté vlastnosti by mali byť vstupom do našej simulácie. Získať však tieto štatistiky nie je vôbec jednoduché, dokonca mnohé sa nedajú ani zmerať. Jednou z kľúčových vlastností agenta pri hľadaní zamestnania sú jeho znalosti a schopnosti a predchádzajúce pracovné skúsenosti. A práve o týchto vlastnostiach agenta neexistujú žiadne tabuľky. Navyše, vlastnosti sa menia v čase. Naučiť sa ovládať nejakú aplikáciu, zlepšiť sa v cudzom jazyku na požadovanú úroveň alebo spraviť si vodičský preukaz za pár dní nie je v dnešnej dobe nič neobvyklé. Pre jednoduchosť modelu sa budeme snažiť určiť iba najdôležitejšie z týchto vlastností a iba tie zahrnieme do výsledného modelu.

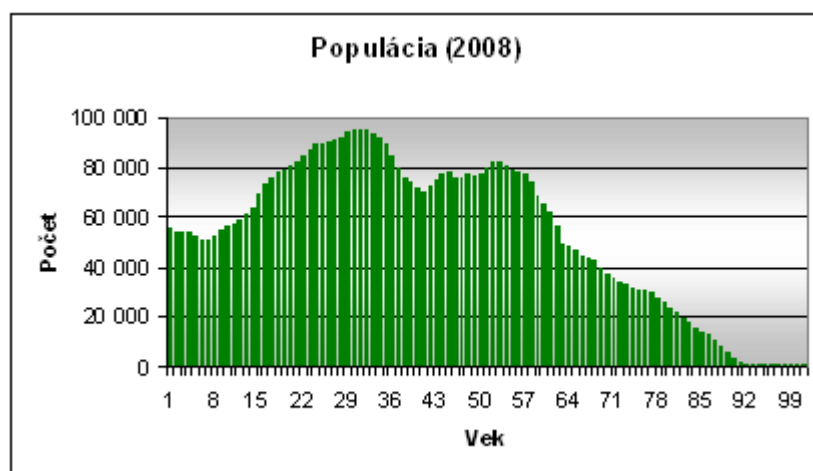
Rovnako ako agenti na pracovnom trhu aj firmy majú svoje vlastnosti, podľa ktorých hľadajú „správnych ľudí“, a tieto vlastnosti zase ovplyvňujú rozhodovanie agentov, či sa v danej firme budú uchádzať o zamestnanie alebo nie. Agentov bude zaujímať napríklad to, aká veľká je firma, kde sa nachádza (poloha), aké vypláca mzdy, akému druhu činnosti sa venuje, aká je jej kvalita. Kvalitu firmy môže určovať počet rokov strávených na trhu, priemerná dĺžka zamestnanosti jej pracovníkov, či je to domáca firma alebo pobočka nadnárodnej spoločnosti, aké získala ocenenia a podobne. Tu narazíme na rovnaký problém ako pri skúsenostiach agentov.

Pre jednoduchosť modelu a taktiež pre obmedzenosť dát budeme uvažovať dve vlastnosti agentov - vek a vzdelanie; a vlastnosti firiem - veľkosť, priemerné vzdelanie a vek ich zamestnancov. Vekovú štruktúru populácie budeme modelovať pomocou stochastického algoritmu pričom použijeme štatistické dáta úmrtnosť a pôrodnosť ako externý vstup do modelu. Podobne budeme modelovať vzdelanie. Potom rozdelíme ľudí do firiem a budeme sledovať, ako sa budú medzi nimi presúvať.

## 2.1 Model starnutia

Vek je veľmi dôležitou vlastnosťou agenta, pretože určuje, či daný občan pracuje alebo je nezamestnaný, alebo je ešte v predproduktívnom, respektíve poproduktívnom veku. Modelovať vek je jednoduché nielen preto, že o vekovom zložení obyvateľstva SR sú podrobné štatistiky, ale aj preto, že presne vieme, čo sa s každým jedincom stane, narodí sa, zostarne alebo zomrie.

Naším konečným cieľom je modelovať zamestnanosť. Preto sa sústredíme na produktívnu časť populácie. Taktiež nás (v zmysle stochastickej simulácie), nebude zaujímať osud každého jednotlivca, ale len to, v akej skupine sa nachádza, resp. aký počet ľudí sa nachádza v daných vekových kategóriách. Takéto zjednodušenie predpokladá rovnomerné rozdelenie občanov podľa veku, ako sme spomínali v kapitole 1.



Obr. 2.1: Vekové zloženie populácie SR z roku 2008

Z obrázku 2.1 vidíme, že tento predpoklad nie je splnený. Na rozdiel od chemických reakcií, v ekonomickom modeli to nepredstavuje zásadný problém. Počítačová podmienka a reakcie unimolekulárneho typu nám umožnia použiť náhodnú simuláciu, aj keď nie je splnený predpoklad rovnomerného rozdelenia, a to tak, že vhodne rozdelíme vekové kategórie. Tie musia byť dostatočne veľké aby bol mo-

del čo najjednoduchší ale zároveň dostatočne malé, aby bolo rozdelenie približne rovnomerné.

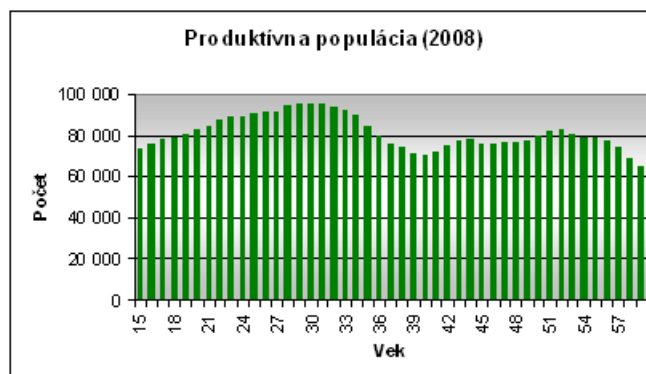
V modeli starnutia simultánne prebiehajú tieto procesy: prírastok obyvateľstva, starnutie obyvateľstva a umieranie. Sformulujeme predpoklady:

1. Populácia je po častiach rovnomerne rozdelená.
2. Ekonomicky aktívnych občanov reprezentujú ľudia vo veku 15 - 59 rokov.
3. Pravdepodobnosť, že sa narodí chlapec je rovnaká ako pravdepodobnosť, že sa narodí dievča.

Druhý predpoklad naznačuje, ako populáciu rozdelíme. Všimnime si, že početnosť detí vo veku 0 – 14 je približne rovnomerne rozdelená. Rovnako sa „vyhladili“ aj početnosti produktívnej časti populácie. Najproblematickejšiu časť grafu, ktorú predstavujú najstarší obyvatelia nemusíme brať do úvahy, pretože z tejto kategórie agenti už nikam neprestúpia, budú zo simulácie len odchádzať, teda zomrú.



Obr. 2.2: Vekové zloženie populácie detí z roku 2008



Obr. 2.3: Vekové zloženie produktívnej časti populácie z roku 2008

Tým, že nie je presne splnený predpoklad 1, očakávame, že bude v simulácii dochádzať k miernym chybám, nie však markantným. Táto chyba bude so zvyšujúcim počtom iterácií bude chyba narastať. S podrobnejším delením populácie môžeme túto chybu zmenšiť. Z obrázku 2.3 vieme určiť, ako najlepšie „rozkúskovať“ produktívnu časť populácie. Vyskúšali sme niekoľko možných delení. Pre jednoduchosť celej simulácie a s prihliadnutím na výsledky sme sa nakoniec rozhodli pre nasledujúce delenie:

**N1:** občania v neproduktívnom veku, kategória obsahuje deti vo veku 0-14 rokov.

**M1:** mladí muži, kategória obsahuje mužov vo veku 15-24 rokov.

**M2:** zrelí muži, kategória obsahuje mužov vo veku 25-49 rokov.

**M3:** starší muži, kategória obsahuje mužov vo veku 50-59 rokov.

**Z1:** mladé ženy, kategória obsahuje ženy vo veku 15-24 rokov.

**Z2:** zrelé ženy, kategória obsahuje ženy vo veku 25-49 rokov.

**Z3:** staršie ženy, kategória obsahuje ženy vo veku 50-59 rokov.

**D:** muži v dôchodkovom veku, kategória mužov vo veku 60+ rokov.

**B:** ženy v dôchodkovom veku, kategória ženy vo veku 60+ rokov.

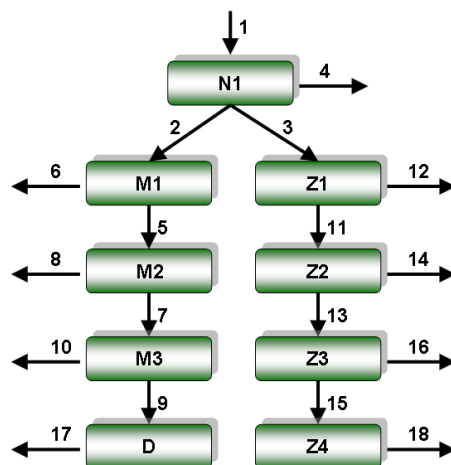
Vstupom do simulácie bude teda stavový vektor, budeme ho označovať  $\mathbf{A}(t)$ , ktorého zložky budú udávať počty ľudí v jednotlivých kategóriách meniace sa v čase. Ďalšími vstupmi budú:

$b$ : pôrodnosť.

$m_X$ : úmrtnosť vekovej kategórie  $X$ , kde  $X \in \{N1, M1, M2, M3, Z1, Z2, Z3, D, B\}$ .

Pôrodnosť je číslo, ktorého hodnota zodpovedá počtu narodených detí za rok. Úmrtnosti vekových skupín sú tiež kladné celé čísla a každé predstavuje počet zomretých v danej vekovej kategórii za rok.

Posledným krokom zostavenia modelu starnutia je určiť, k akým reakciám v ňom dochádza. Nasledujúca schéma zobrazuje všetky uvažované reakcie ako šípky, pričom voľné šípky prislúchajú úbytkom populácie v dôsledku úmrtia.



Obr. 2.4: Schematické zobrazenie reakcií pre model starnutia

Ďalej uvedieme zoznam reakcií a ich zodpovedajúce rýchlosti. K hodnotám rýchlostí reakcií sa vrátíme na konci tejto kapitoly, v sekcii 2.1.2. Teraz si popíšeme typy reakcií.

Reakcia  $\emptyset \rightarrow N1$  predstavuje rodenie detí. V prvej kapitole sme sa s podobnou reakciou nestretli. Ide o spontánnu tvorbu alebo reakciu spôsobenú nejakým vonkajším zdrojom.

$j$	Reakcia	$c_j$	$a_j(\mathbf{x})$
1	$\emptyset \rightarrow N1$	$1/12$	$1/12 * b$
2	$N1 \rightarrow M1$	$(1/(12 * 15))/2$	$(1/(12 * 15) * N1)/2$
3	$N1 \rightarrow Z1$	$(1/(12 * 15))/2$	$(1/(12 * 15) * N1)/2$
4	$N1 \rightarrow \emptyset$	$1/12$	$1/12 * N1$
5	$M1 \rightarrow M2$	$1/(12 * 10)$	$1/(12 * 10) * M1$
6	$M1 \rightarrow \emptyset$	$1/12$	$1/12 * M1$
7	$M2 \rightarrow M3$	$1/(12 * 25)$	$1/(12 * 25) * M2$
8	$M2 \rightarrow \emptyset$	$1/12$	$1/12 * M2$
9	$M3 \rightarrow D$	$1/(12 * 10)$	$1/(12 * 10) * M3$
10	$M3 \rightarrow \emptyset$	$1/12$	$1/12 * M3$
11	$Z1 \rightarrow Z2$	$1/(12 * 10)$	$1/(12 * 10) * Z1$
12	$Z1 \rightarrow \emptyset$	$1/12$	$1/12 * Z1$
13	$Z2 \rightarrow Z3$	$1/(12 * 25)$	$1/(12 * 25) * Z2$
14	$Z2 \rightarrow \emptyset$	$1/12$	$1/12 * Z2$
15	$Z3 \rightarrow B$	$1/(12 * 10)$	$1/(12 * 10) * Z3$
16	$Z3 \rightarrow \emptyset$	$1/12$	$1/12 * Z3$
17	$D \rightarrow \emptyset$	$1/12$	$1/12 * D$
18	$B \rightarrow \emptyset$	$1/12$	$1/12 * B$

Tabuľka 2.1: Reakcie, rýchlosti reakcií a adaptované funkcie rýchlostí pre model starnutia

Reakcia  $N1 \rightarrow \emptyset$  je spätnou reakciou k reakcii  $\emptyset \rightarrow N1$ . Predstavuje úmrtie jedinca z vekovej kategórie  $N1$ . Všeobecne túto reakciu môžeme zapísať v tvare

$$A_X(t) \rightarrow \emptyset \quad \text{pre} \quad X \in \{N1, M1, M2, M3, Z1, Z2, Z3, D, B\}.$$

Ostatné typy reakcií predstavujú starnutie ľudí, pričom nesledujeme starnutie po rokoch ale starnutie po vekových etapách. K týmto reakciám neexistujú reverzné reakcie.

Teraz už máme všetky potrebné informácie, aby sme spustili simuláciu starnutia agentov. Štartovacím časom bude rok 2001. Vektory zmeny  $\mathbf{v}_j$ , ktoré nám menia stavový vektor  $\mathbf{A}(t)$  vždy, keď nastane nejaká reakcia, zapíšeme do matice  $\mathbf{V}$  a budeme ju volať *matice prechodu*.



### 2.1.1 Algoritmus pre model starnutia

1. Spočítame hodnoty adaptovaných funkcií rýchlostí  $\{a_k(\mathbf{A}(t))\}_{k=1}^{18}$  a ich sumu  $a_{sum}(\mathbf{A}(t)) := \sum_{k=1}^{18} a_k(\mathbf{A}(t))$ .
2. Vygenerujeme dve náhodné čísla  $\xi_1$  a  $\xi_2$  z rovnomerného rozdelenia  $(0, 1)$  a určíme  $j$ , teda typ reakcie k akej došlo, ako najmenšie celé číslo, ktoré vyhovuje nerovnici  $\sum_{k=1}^j a_k(\mathbf{A}(t)) > \xi_1 a_{sum}(\mathbf{A}(t))$ .
3. Určíme, koľko času prešlo, kým nastala  $j$ -ta reakcia  $\tau = \frac{\ln(1/\xi_2)}{a_{sum}(\mathbf{A}(t))}$  a obnovíme  $\mathbf{A}(t + \tau) = \mathbf{A}(t) + \mathbf{V}_j$ , kde  $\mathbf{V}_j$  predstavuje  $j$ -ty riadok matice  $\mathbf{V}$ . Aktualizujeme čas  $t$  na  $t + \tau$ .
4. Vrátime sa na krok číslo 1.

### 2.1.2 Rýchlostí reakcií a terminálna podmienka

V tejto časti práce sa budeme venovať nezávislosti vekovej simulácie bez ohľadu na časové jednotky a voľbe rýchlostí reakcií.

Nezávislosť simulácie na voľbe časovej jednotky vyplýva z nasledovnej úvahy. Ak sú rýchlostí reakcií udávané v mesiacoch, čas do nasledujúcej reakcie je

$$\tau = \frac{\ln \frac{1}{\xi}}{\sum_k a_k(\mathbf{A}(t))} [\text{mesiacov}].$$

Ak sú však rýchlosti udávané v rokoch, potom

$$\begin{aligned} \tau' &= \frac{\ln \frac{1}{\xi}}{\sum_k 12a_k(\mathbf{A}(t))} = \frac{1}{12} \tau [\text{rokov}] \\ &= \tau [\text{mesiacov}]. \end{aligned}$$

Voľba časovej jednotky teda neovplyvní počet prebiehajúcich reakcií, a teda ani celkový výsledok. V našom modeli starnutia je časovou jednotkou mesiac. Simulujeme 7 rokov (2001 – 2008), terminálnou podmienkou teda bude uplynutý čas

$t = 12 * 7$  [mesiacov].

Rýchlosti reakcií sú odvodené nasledujúcou dimenzionálnou analýzou. Pre príklad uvidíme reakciu  $N1 \rightarrow M1$  (dospievanie dieťaťa na muža). Reakcia prebieha rýchlosťou

$$c_2 = \left( \frac{1}{12 * 15 * 2} \right) \left[ \frac{\text{počet}}{\text{rok} * \frac{\text{mesiac}}{\text{rok}}} \right] \iff \left[ \frac{\text{počet}}{\text{mesiac}} \right],$$

kde 12 vyjadruje, že časovou jednotkou je mesiac, 15 vyjadruje, že len pätnástina detí sa môže stať mužmi (využíva sa predpoklad 1 o po častiach rovnomernom rozdelení populácie), pretože kategória  $N1$  obsahuje deti vo veku 0 – 14 rokov, t. j. 15 ročných skupín. To, že celú rýchlosť ešte zmenšíme o polovicu vyplýva z predpokladu 3, ktorý hovorí, že len polovica týchto detí sa stane mužmi.

### 2.1.3 Výsledky

Simulácia prebieha od roku 2001 po rok 2008. Vstupmi sú:

- $A(2001)$  - počty ľudí vo vekových kategóriách (dimenzia  $1 \times 9$ ).
- $m1(2001)$  - úmrtnosti pre vekové kategórie (dimenzia  $1 \times 9$ ).
- $b(2001)$  - pôrodnosť (kladné celé číslo).
- $V$  - matica prechodu (dimenzia  $18 \times 9$ ).
- $c$  - rýchlosti reakcií (dimenzia  $1 \times 9$ ).

Terminálna podmienka  $t = 12 * 7$  (mesiacov) ukončí algoritmus. Keby sme simuláciu pustili znovu s rovnakými vstupnými dátami, dostaneme podobný, avšak iný výsledok. Je to preto, lebo náš algoritmus je stochastický. Aby sme dostali čo najlepší výsledok, nasimulujeme prechod z roku 2001 do roku 2008 veľa krát - tzv. metóda Monte Carlo [16]. Výsledkom potom bude aritmetický priemer z týchto

simulácii. Pretože výsledky pre model starnutia sú len čiastočnými výsledkami celého modelu zamestnanosti prezentujeme teraz odhady, ktoré sme získali jednou simuláciou. Pre spresnenie výpočtov po každom nasimulovanom roku aktualizujeme pôrodnosť a úmrtnosť.

Teraz uvidíme a popíšeme niektoré výstupy z nášho stochastického modelu. Porovnáme výsledky s reálnymi dátami. Pozrieme sa na grafy celej populácie a dvoch vybraných vekových kategórií, ktoré boli odhadnuté naším modelom najlepšie a najhoršie. Pribalené DVD obsahuje ďalšie výstupy a program ASS.

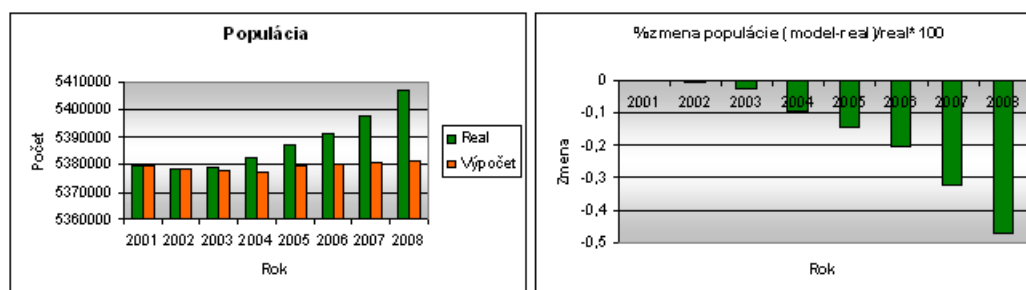
Algoritmus stochastickej simulácie je naprogramovaný v MATLAB-e.

Rok	2008			
	Úmrtnosť	Pocet	Matlab	Zmena
N1	500	843754	910885	7,956229
M1	362	417863	402468	-3,68422
M2	2941	1044961	1004560	-3,86627
M3	4881	370073	322681	-12,8061
Z1	100	400609	393304	-1,82347
Z2	976	1018958	1002081	-1,6563
Z3	1877	395883	351521	-11,2058
D	19512	361540	406941	12,55767
B	22015	553331	587070	6,097435
Spolu		5406972	5381511	
Pôrodnosť		57360		

Obr. 2.5: Reálne čísla a výpočet našou simuláciou.

V tabuľke na obrázku 2.5 sú uvedené skutočné počty ľudí (stĺpec *Pocet*) pre všetky zvolené vekové kategórie a ich suma z roku 2008. Stĺpec *Matlab* udáva počty ľudí, ktoré sme dostali z našej stochastickej simulácie po nasimulovaní siedmich rokov, pričom sme použili informáciu o počte ľudí z roku 2001 ako počiatočnú podmienku a pôrodnosti a úmrtnosti v ďalších rokoch. V poslednom stĺpci *Zmena* je spočítaná percentuálna zmena oproti skutočným počtom ( $Zmena = (Matlab - Pocet) / Pocet * 100$ ). Stochastický algoritmus veľmi dobre aproximoval skutočné hodnoty pre celú populáciu. Celková chyba je menšia ako 0.5% za 7 rokov. V jednotlivých vekových kategóriách už dochádza k podstatne

väčším výkyvom. Tieto zmeny vznikli pretože neboli splnené predpoklady nášho modelu, hlavne predpoklad o rovnomernom rozdelení populácie.



Obr. 2.6: Celková populácia SR a percentuálna zmena výpočtu oproti skutočnosti za obdobie 2001 – 2008.

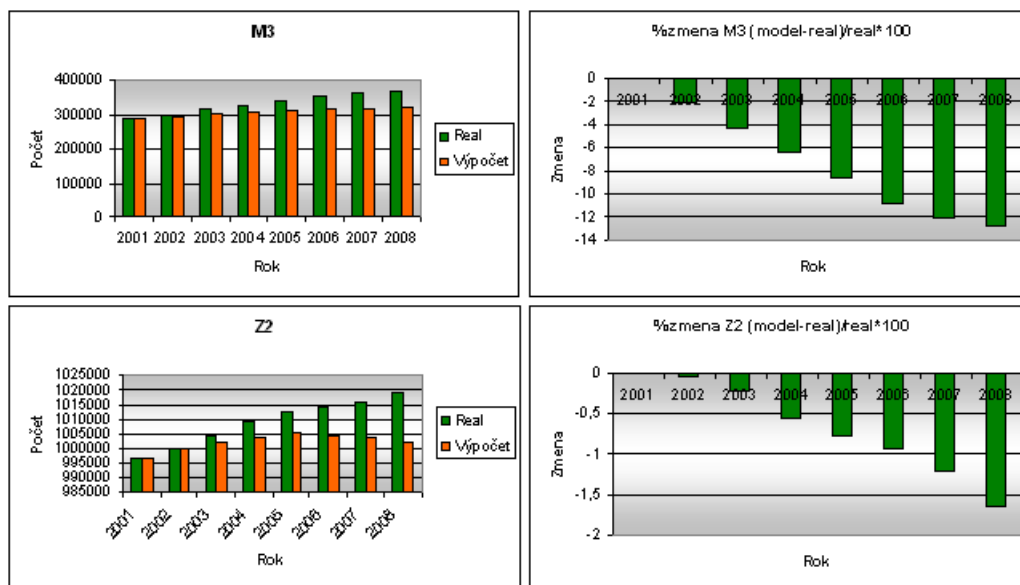
Na obrázku 2.6 vidíme, v akom vzťahu sú výsledky našej simulácie oproti reálnym číslam počas celého pozorovaného obdobia 2001 – 2008. Zelený stĺpec (*Real*) predstavuje skutočné počty obyvateľov Slovenskej Republiky. Oranžový stĺpec (*Výpočet*) predstavuje počty obyvateľov, ktoré sme získali pomocou stochastického algoritmu. Graf na pravej strane obrázku 2.6 zobrazuje, ako sa výpočet časom zhoršoval. Ide však o veľmi malú zmenu.

Teraz uvidíme výsledky pre dve vekové kategórie, *M3* a *Z2*. V kategórii *M3* vznikli oproti reálnym dátam najväčšie nepresnosti a v kategórii *Z2* naopak najmenšie.

<b>M3</b>	<b>real</b>	<b>matlab</b>	<b>Z2</b>	<b>real</b>	<b>matlab</b>
2001	286218	286218	2001	996833	996833
2002	299970	293722	2002	1000093	999795
2003	314006	300299	2003	1004264	1002080
2004	327178	306073	2004	1009374	1003817
2005	340659	311413	2005	1012725	1005033
2006	353794	315450	2006	1013997	1004478
2007	363841	319535	2007	1015899	1003650
2008	370073	322681	2008	1018958	1002081

Obr. 2.7: Ľavá tabuľka obsahuje počty vekovej kategórie *M3*, ktorej simulácia vyšla najhoršie. Pravá tabuľka popisuje vekovú kategóriu *Z2*, ktorej simulácia vyšla najlepšie.

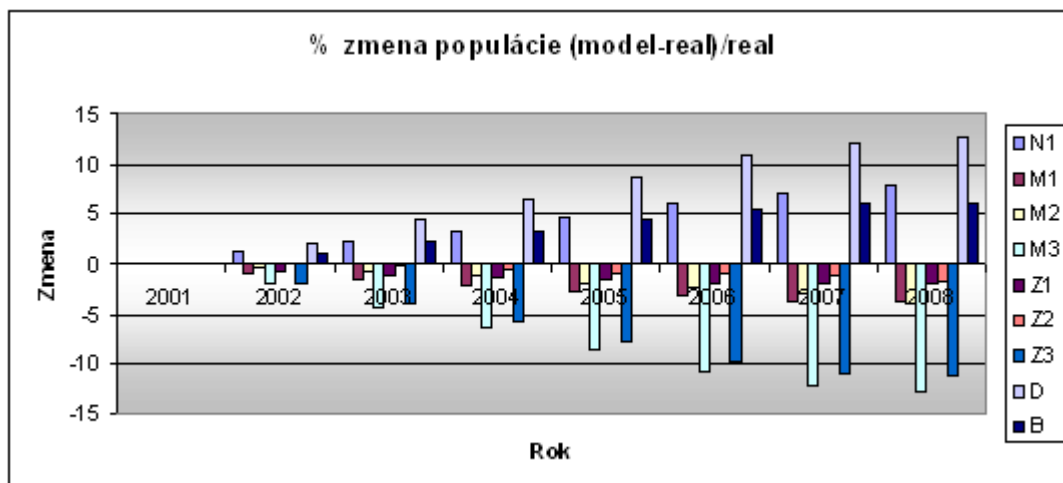
V tabuľke na obrázku 2.7 sú uvedené reálne počty ľudí (stĺpce *real*) spomínaných vekových kategórií a aj počty získané stochastickou simuláciou (stĺpce *matlab*) za celé sledované obdobie.



Obr. 2.8: Vekové kategórie *M3* a *Z2* a ich percentuálne zmeny oproti skutočnosti za obdobie 2001 – 2008.

Na obrázku 2.8 sú graficky znázornené počty ľudí v daných vekových kategóriách, ktoré boli vybrané ako najlepšie a najhoršie, a aj ich percentuálne zmeny z tabuľky na obrázku 2.7. Všimnime si škálu, na akej sú grafy znázornené. Jedným z dôvodov, prečo je veková kategória *M3* vypočítaná s takou veľkou chybou je, že aj keď naša simulácia zachytáva trend nárastu počtu ľudí v tejto kategórii, v skutočnosti rástol tento počet oveľa rýchlejšie (približne plus 10000 ľudí každý rok). Kategória *Z2* je jednou z dvoch kategórií, v ktorej nám algoritmus nezachytil trend ale napriek tomu veľmi presne spočítal počty ľudí. Je to preto, lebo v tejto kategórii sa počty ľudí časom menili minimálne. Iné delenie populácie na vekové kategórie by mohlo výsledky zlepšiť, resp. zhoršiť, avšak nikdy nebudú úplne presné. Stochastický algoritmus predpokladá rovnomerné rozdelenie, čo v reálnom svete nie je nikdy splnené.

Nakoniec uvádzame graf (obrázok 2.9), na ktorom sú zobrazené percentuálne zmeny počtov ľudí oproti skutočným dátam pre všetky vekové kategórie za celé sledované obdobie 2001 – 2008.



Obr. 2.9: Percentuálne zmeny výpočtu oproti skutočnosti všetkých vekových kategórií za obdobie 2001 – 2008.

## 2.2 Model vzdelávania

Popri veku je ďalšou kľúčovou vlastnosťou zamestnanca pri hľadaní práce nepochybne vzdelanie. Vyššie vzdelanie implikuje vyšší plat, lukratívnejšie pracovné miesto, lepší pocit zo samého seba. Nie je však neobvyklé, keď vysokoškolsky vzdelaný človek vykonáva podradné práce a naopak, niekto menej schopný zastáva významnú politickú alebo spoločenskú funkciu. Nie je jednoduché vhodne zvoliť mieru vzdelania, lebo štandardné kvantitatívne miery ako počet rokov strávených v škole, diplom, prípadne výsledok IQ testu sú často krátko navzájom neporovnateľné (je rozdiel medzi diplomami z rôznych škôl, medzi talentovaným žiakom, ktorému bolo umožnené vynechať deviaty ročník a žiakom, ktorému naopak rok na viac na základnej škole prospel, atď.). Venovať sa tejto problematike však nie je predmetom tejto diplomovej práce.

Pre jednoduchosť modelu, ako aj pre obmedzený rozsah dát zvolíme za jednotku vzdelania informáciu o najvyššom ukončenom vzdelaní.

Rozdeliť celú populáciu na kategórie podľa vzdelania už nie je také jednoznačné ako to bolo v prípade delenia podľa veku. Existujú štatistiky o počte žiakov a študentov v školách, avšak týchto je potrebné začleniť do celkovej populácie, pričom dáta s ktorými sme pracovali sú navzájom nekonzistentné.

Obyvateľov rozdelíme podľa ukončeného vzdelania do 9 skupín:

**NV:** neproduktívny vek, kategória obsahuje deti vo veku 0 – 5 rokov.

**ZS:** žiaci základných škôl, kategória obsahuje žiakov, ktorí navštevujú základnú školu.

**ZB:** ľudia, ktorí majú ukončenú len základnú školu alebo sú bez vzdelania, kategória obsahuje ľudí v produktívnom veku, ktorí majú ukončenú základnú školu alebo sú bez základného vzdelania.

**U:** uční, kategória obsahuje študentov, ktorí navštevujú stredné odborné učilište alebo združenú strednú školu.

**Uc:** ľudia, ktorí majú ukončenú učňovskú strednú školu, kategória obsahuje ľudí v produktívnom veku, ktorí majú ukončené stredné odborné učilište alebo združenú strednú školu.

**US:** stredoškoláci, kategória obsahuje študentov, ktorí navštevujú gymnázium alebo strednú odbornú školu.

**USc:** ľudia, ktorí majú ukončenú strednú školu, kategória obsahuje ľudí v produktívnom veku, ktorí majú ukončené gymnázium alebo strednú odbornú školu.

**VS:** vysokoškoláci, kategória obsahuje študentov, ktorí navštevujú vysokú školu (denná forma štúdia).

**VSc:** ľudia, ktorí majú ukončenú vysokú školu, kategória obsahuje ľudí v produktívnom veku, ktorí majú ukončené vysokú školu.

Pri modelovaní vzdelania nás už nemusia zaujímať ľudia v poproduktívnom veku (kategórie D a B z modelu starnutia v kapitole 2.1), pretože akonáhle dosiahnu dôchodkový vek z modelu odchádzajú a nie je im umožnené vrátiť sa späť. V predchádzajúcom modeli vekovej štruktúry neboli z modelu vyčlenení len preto, aby bolo možné sledovať aj zmeny počtov v celej populácii, nie len v jednotlivých kategóriách.

Ďalšími vstupmi do simulácie vzdelávania sú:

$b$ : pôrodnosť<sup>1</sup>.

$m_{NV}$ : úmrtnosť skupiny NV (vek 0 – 5 rokov, rozpätie 6 rokov).

$m_{ZS}$ : úmrtnosť skupiny ZS (vek 6 – 14 rokov, rozpätie 9 rokov).

$m_{ZB}$ : úmrtnosť skupiny ZB (vek 15 – 59 rokov, rozpätie 46 rokov).

$m_U$ : úmrtnosť skupiny U (vek 15 – 18 rokov, rozpätie 4 rokov).

---

<sup>1</sup>Rovnaká hodnota ako v modeli starnutia.



$m_{Uc}$ : úmrtnosť skupiny Uc (vek 19 – 59 rokov, rozpätie 42 rokov).

$m_{US}$ : úmrtnosť skupiny US (vek 15 – 18 rokov, rozpätie 4 rokov).

$m_{USc}$ : úmrtnosť skupiny USc (vek 19 – 59 rokov, rozpätie 42 rokov).

$m_{VS}$ : úmrtnosť skupiny VS (vek 19 – 23 rokov, rozpätie 5 rokov).

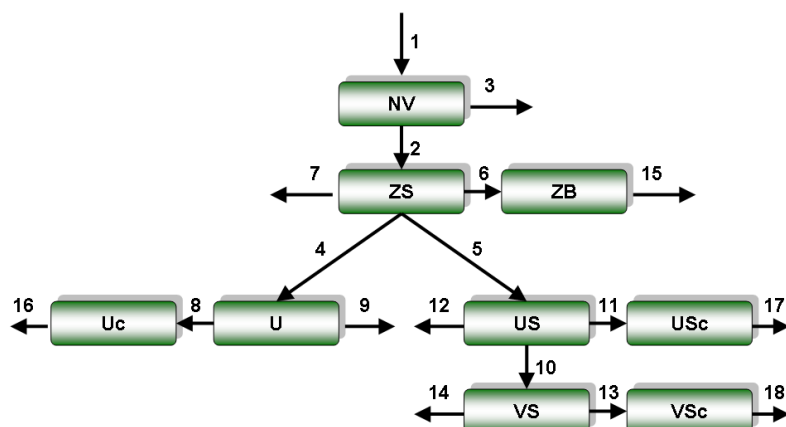
$m_{VSc}$ : úmrtnosť skupiny VSc (vek 24 – 59 rokov, rozpätie 37 rokov).

Predpokladáme, že populácia spĺňa nasledujúce predpoklady:

1. Vzdelanostné skupiny sú po častiach rovnomerne rozdelené.
2. Skupina NV obsahuje všetky deti vo veku 0 – 5 rokov, všetkým dobre vychádzajú roky, t.j. akoby sa narodili 1.9., a všetky nastúpia do školy vo veku 6 rokov.
3. V kategórii ZS strávajú všetci 9 rokov.
4. V kategóriách U a US strávajú všetci, ktorí sa do nich dostanú 4 roky.
5. V kategórii VS strávajú všetci, ktorí sa do nej dostanú 5 rokov.
6. Neexistuje cesta späť. Napríklad, nie je možné z kategórie USc „spadnúť“ do kategórie ZB.
7. Do dôchodku sa odchádza vo veku 60 rokov.
8. Spojitá dynamika - školu je možné ukončiť kedykoľvek v priebehu roka.

Za týchto predpokladov a s informáciami o počtoch ľudí v jednotlivých kategóriách zostavíme schému modelu pre vzdelanie:

Reakcie a ich rýchlosti sú zhrnuté v tabuľke 2.2, kde  $\vee$  je logický operátor ALEBO a  $Do$  predstavuje odchod do dôchodku:



Obr. 2.10: Schematické zobrazenie reakcií pre model vzdelávania

V ďalšej sekcii sa budeme venovať optimálnej voľbe parametrov  $\alpha_i$ ,  $i = \{1, 2, 3, 4, 5\}$ , ako aj dodatočným úpravám modelu.

### 2.2.1 Voľba parametrov a ďalšia úprava

Rýchlosti  $\alpha_i$  určujú, koľko žiakov/študentov prejde z aktuálneho typu školy do ďalšej kategórie. Napríklad, len časť študentov ZŠ pôjde na učňovku. Ukážeme si tri rôzne spôsoby voľby týchto parametrov. Prvým je

$$\alpha_1 = \frac{U + U_c}{ZB + U + U_c + US + US_c + VS + VS_c}$$

$$\alpha_2 = \frac{US + US_c + VS + VS_c}{ZB + U + U_c + US + US_c + VS + VS_c}$$

$$\alpha_3 = \frac{ZB}{ZB + U + U_c + US + US_c + VS + VS_c}$$

$$\alpha_4 = \frac{VS + VS_c}{US_c + VS + VS_c}$$

$$\alpha_5 = \frac{US_c}{US_c + VS + VS_c}$$

Takto vybrané rýchlosti majú vlastnosť, že rozdeľujú žiakov ZŠ, resp. SŠ v pomere, v akom sú počty ľudí v nasledujúcich vetvách modelu vzdelávania, ako je

$j$	Reakcia	$c_j$	$a_j(\mathbf{x})$
1	$\emptyset \rightarrow NV$	$1/12$	$1/12 * b$
2	$NV \rightarrow ZS$	$1/(12 * 6)$	$1/(12 * 6) * NV$
3	$NV \rightarrow \emptyset$	$1/12$	$1/12 * NV$
4	$ZS \rightarrow U$	$1/(12 * 9) * \alpha_1$	$1/(12 * 9) * \alpha_1 * ZS$
5	$ZS \rightarrow US$	$1/(12 * 9) * \alpha_2$	$1/(12 * 9) * \alpha_2 * ZS$
6	$ZS \rightarrow ZB$	$1/(12 * 9) * \alpha_3$	$1/(12 * 9) * \alpha_3 * ZS$
7	$ZS \rightarrow \emptyset$	$1/12$	$1/12 * ZS$
8	$U \rightarrow Uc$	$1/(12 * 4)$	$1/(12 * 4) * U$
9	$U \rightarrow \emptyset$	$1/12$	$1/12 * U$
10	$US \rightarrow VS$	$1/(12 * 4) * \alpha_4$	$1/(12 * 4) * \alpha_4 * US$
11	$US \rightarrow USc$	$1/(12 * 4) * \alpha_5$	$1/(12 * 4) * \alpha_5 * US$
12	$US \rightarrow \emptyset$	$1/12$	$1/12 * US$
13	$VS \rightarrow VSc$	$1/(12 * 5)$	$1/(12 * 5) * VS$
14	$VS \rightarrow \emptyset$	$1/12$	$1/12 * VS$
15	$ZB \rightarrow \emptyset \vee Do$	$1/12 + 1/(12 * 46)$	$1/12 + 1/(12 * 46) * ZB$
16	$Uc \rightarrow \emptyset \vee Do$	$1/12 + 1/(12 * 46)$	$1/12 + 1/(12 * 46) * Uc$
17	$USc \rightarrow \emptyset \vee Do$	$1/12 + 1/(12 * 46)$	$1/12 + 1/(12 * 46) * USc$
18	$VSc \rightarrow \emptyset \vee Do$	$1/12 + 1/(12 * 46)$	$1/12 + 1/(12 * 46) * VSc$

Tabuľka 2.2: Reakcie, rýchlosti reakcií a adaptované funkcie rýchlostí pre model vzdelávania

znázornené na obrázku 2.10. Inak povedané, určili sme váhu, s akou sa prestupuje do nasledujúcich kategórií. Použitím takýchto prestupových rýchlostí dostaneme z modelu výstupy, ktoré sú veľmi vzdialené skutočnosti (tieto výsledky v práci neprezentujeme). Vezmeme preto do úvahy argument, že rozhodovanie žiakov, na akej škole budú ďalej študovať je ovplyvnené voľbou ich spolužiakov (napr. „Kamarát ide na gymnázium, idem aj ja.“). Vychádzame z toho, že sa rozhodujú v prítomnosti a neuvažujú o tom, v akej kategórii nakoniec skončia. Preto zo zlomkov odstránime všetky vzdelanostné kategórie, ktoré nenasledujú bezprostredne po kategórii, z ktorej žiak/študent odchádza a tieto ešte upravíme tak, aby sa rozhodnutie žiaka/študenta, kam pôjde ďalej študovať týkalo len žiakov/študentov, ktorí z kategórie prestupujú. Dostaneme tak nové hodnoty parametra  $\alpha_i$

$$\begin{aligned}\alpha_1 &= \frac{U/4}{U/4 + US/4 + ZB/46} \\ \alpha_2 &= \frac{US/4}{U/4 + US/4 + ZB/46} \\ \alpha_3 &= \frac{ZB/46}{U/4 + US/4 + ZB/46} \\ \alpha_4 &= \frac{VS/5}{VS/5 + US/4} \\ \alpha_5 &= \frac{US/4}{VS/5 + US/4}\end{aligned}$$

Výsledky sa zmenia minimálne (v práci neuvádzame), preto vyskúšame ďalšiu možnosť. Namiesto dynamicky meniacich sa parametrov  $\alpha_i$ , použijeme konštanty. Predchádzajúce parametre  $\alpha_i$  nám približne určujú ich hodnoty. Nové parametre určíme metódou pokus-omyl. A teda

$$\alpha_1 = 0.377$$

$$\alpha_2 = 0.567$$

$$\alpha_3 = 0.057$$

$$\alpha_4 = 0.46$$

$$\alpha_5 = 0.54$$

Výsledky sa opäť zlepšili, ale stále sú oproti skutočnosti veľké rozdiely v počtoch ľudí vo vzdelanostných kategóriách. Kategórie ZS a Uc sú nadhodnotené. Celkovo však po nasimulovaní 7 rokov (2001 – 2008) v modeli chýba približne 240000 ľudí. To znamená, že oproti skutočnosti ľudia v modeli rýchlejšie starnú alebo umierajú. Je to spôsobené hlavne nekonzistentnými dátami z viacerých zdrojov a nesplnenými predpokladmi. Napríklad, neexistujú úmrtnostné tabuľky žiakov/študentov, alebo nie každé dieťa „vydrží“ 9 rokov na základnej škole. Viaceré kategórie majú spoločný prienik. To môže napríklad viesť k udalosti, že človek, ktorý by mal cho-

diť ešte na základnú školu sa už nachádza v kategórii ZB a teda v modeli zomrie dvakrát. Spôsobuje to pravdepodobnosť jeho úmrtia v kategórii ZS, aj keď sa už nachádza v kategórii ZB, kde už zomiera s inou pravdepodobnosťou. Nadhodnotenie kategórie ZS si môžeme vysvetliť tak, že na základnú školu chodí v skutočnosti menej detí. V našom modeli na základnú školu nastupujú všetky deti, teda aj tie, ktoré do školy chodiť nikdy nebudú.

Pre zlepšenie výsledkov spravíme v modeli vzdelávanie niekoľko úprav. Pozmeníme niektoré predpoklady a pridáme nové reakcie:

### 1.zmena

Zrýchlime reakciu  $NV \rightarrow ZS$ . Pozmenil sa tým predpoklad 2, ktorý okrem iného hovorí o tom, že všetkým deťom dobre vychádzajú roky, a teda v kategórii  $NV$  strávia celých 6 rokov. Teraz pripúšťame, že v modeli máme aj deti, ktoré sa nenarodili 1.9. a pôjdu do školy skôr. Ako najlepšia hodnota sa ukázalo 5.9 rokov.

### 2.zmena

Zrýchlime reakciu  $ZS \rightarrow ZB$ . Týmto sa pozmenil predpoklad o tom, že všetci strávia na základnej škole 9 rokov. Opierame sa o presvedčenie, že človek, ktorý sa nachádza v kategórii ľudí bez akéhokoľvek vzdelania alebo len s ukončenou základnou školou v nej v priemere nestrávil 9 rokov. Najlepšia sa ukázala hodnota 6 rokov.

### 3.zmena

Zrýchlime reakcie  $ZS \rightarrow U$  a  $ZS \rightarrow US$ . Rovnako sa týmto mení 3. predpoklad. Vychádzame pritom z logickej úvahy, že väčšina terajšej produktívnej populácie do deviatej triedy nechodila. Počet rokov znížime z 9 na 8 a s prihliadnutím na výsledky ešte rýchlosti upravíme o konštantu 1.23 pre reakciu  $ZS \rightarrow U$  a o 1,14 pre  $ZS \rightarrow US$ .

#### 4.zmena

Pridáme reakciu  $Uc \rightarrow USc$ , ktorá predstavuje rozhodnutie učňa spraviť si nadstavbu alebo maturitu. Rýchlosť tejto reakcie určíme nasledovne

$$c_{Uc \rightarrow USc} = \frac{1}{12 * 4} * 0.102 * 0.8014,$$

kde 4 predstavuje priemerný počet rokov, kým učeň zrealizuje svoje rozhodnutie, konštanta 0.102 určuje, aká časť učňov si vylepší vzdelanie a konštanta 0.8014 predstavuje, že si robí nadstavbu alebo maturitu.

#### 5.zmena

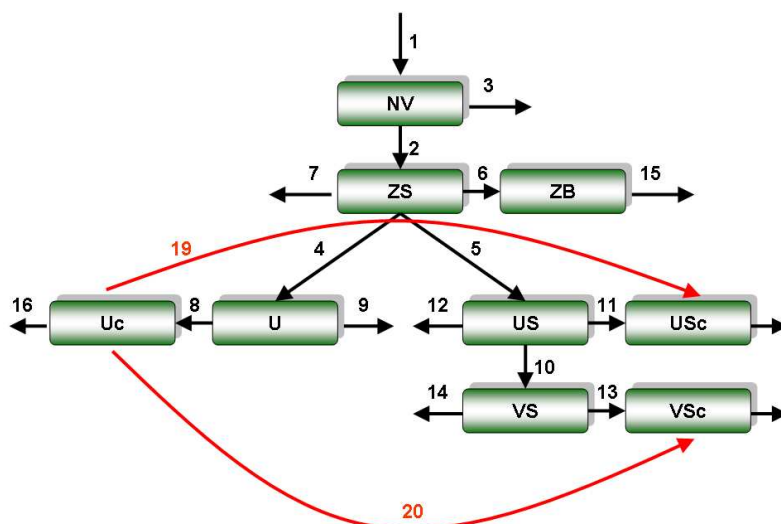
Pridáme reakciu  $Uc \rightarrow VS$ , ktorá predstavuje rozhodnutie učňa ísť študovať na vysokú školu. Budú to prevažne študenti technických smerov a hlásiť sa budú zas len na technické smery (napr. niektoré odbory na STU, FIT, FEJ). Rýchlosť reakcie je

$$c_{Uc \rightarrow VS} = \frac{1}{12 * 4} * 0.102 * 0.1986,$$

Kde 4 a 0.102 majú rovnaký význam ako v predchádzajúcom bode, konštanta 0.1986 znamená, že k takejto reakcii bude dochádzať zriedkavo.

Hodnoty konštánt v 4 a 5 bode sme získali pozorovaním štatistík vzdelanostných kategórií. Ostatné konštanty sú získané metódou pokus - omyl. K inej možnosti, ako ich čo najlepšie vyčíslíť sa ešte vrátíme.

Konečná schéma a tabuľka reakcií pre model vzdelávania budú vyzeráť nasledovne:



Obr. 2.11: Schematické zobrazenie reakcií pre upravený model vzdelávania

### 2.2.2 Systematická voľba konštánt $\alpha_i$

V modeli vzdelávania sme konštanty  $\alpha_i$  určili metódou pokus-omyl a ich hodnoty sme uhádli. Existujú metódy, ktoré tieto konštanty hľadajú systematicky.

Nájdienie správnych konštánt môžeme formulovať ako úlohu nelineárneho programovania [8]. Našou úlohou je minimalizovať zmenu výsledného vektora pre vzdelanie oproti skutočným dátam. Zvolíme si normu tejto zmeny. Napríklad, ak chceme, aby chyba bola rovnomerne rozložená v každej kategórii vzdelania použijeme *euklidovskú normu*. Alebo, ak chceme minimalizovať chybu pre najhoršie odhadnutú kategóriu, použijeme *supremovu normu*. Funkcia, závislá od piatich konštánt  $\alpha_i$ , ktorú v úlohe NLP minimalizujeme, je zakomponovaná v stochastickom algoritme.

$j$	Reakcia	$c_j$	$a_j(\mathbf{x})$
1	$NIC \rightarrow NV$	$1/12$	$1/12 * b$
2	$NV \rightarrow ZS$	$1/(12 * 5.9)$	$1/(12 * 5.9) * NV$
3	$NV \rightarrow \emptyset$	$1/12$	$1/12 * NV$
4	$ZS \rightarrow U$	$1/(12 * 8) * 0.377 * 1.23$	$1/(12 * 8) * 0.377 * 1.23 * ZS$
5	$ZS \rightarrow US$	$1/(12 * 8) * 0.567 * 1.14$	$1/(12 * 8) * 0.567 * 1.14 * ZS$
6	$ZS \rightarrow ZB$	$1/(12 * 9) * 0.057$	$1/(12 * 9) * 0.057 * ZS$
7	$ZS \rightarrow \emptyset$	$1/12$	$1/12 * ZS$
8	$U \rightarrow Uc$	$1/(12 * 4)$	$1/(12 * 4) * U$
9	$U \rightarrow \emptyset$	$1/12$	$1/12 * U$
10	$US \rightarrow VS$	$1/(12 * 4) * 0.46$	$1/(12 * 4) * 0.46 * US$
11	$US \rightarrow USc$	$1/(12 * 4) * 0.54$	$1/(12 * 4) * 0.54 * US$
12	$US \rightarrow \emptyset$	$1/12$	$1/12 * US$
13	$VS \rightarrow VSc$	$1/(12 * 5)$	$1/(12 * 5) * VS$
14	$VS \rightarrow \emptyset$	$1/12$	$1/12 * VS$
15	$ZB \rightarrow \emptyset \vee Do$	$1/12 + 1/(12 * 46)$	$1/12 + 1/(12 * 46) * ZB$
16	$Uc \rightarrow \emptyset \vee Do$	$1/12 + 1/(12 * 46)$	$1/12 + 1/(12 * 46) * Uc$
17	$USc \rightarrow \emptyset \vee Do$	$1/12 + 1/(12 * 46)$	$1/12 + 1/(12 * 46) * USc$
18	$VSc \rightarrow \emptyset \vee Do$	$1/12 + 1/(12 * 46)$	$1/12 + 1/(12 * 46) * VSc$
19	$Uc \rightarrow USc \vee Do$	$1/(12 * 4) * 0.102 * 0.8014$	$1/(12 * 4) * 0.102 * 0.8014 * Uc$
20	$Uc \rightarrow VS \vee Do$	$1/(12 * 4) * 0.102 * 0.1986$	$1/(12 * 4) * 0.102 * 0.1986 * Uc$

Tabuľka 2.3: Reakcie, rýchlosti reakcií a adaptované funkcie rýchlostí pre upravený model vzdelávania

### 2.2.3 Výsledky

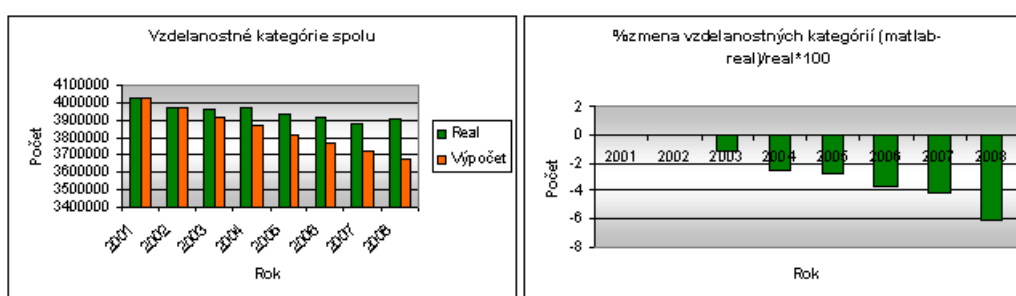
V tejto sekcii opisujeme vybrané výstupy modelu vzdelávania. Rovnako, ako v kapitole 2.1.3 sa najskôr pozrieme na výsledky celej simulácie a potom na konkrétne vzdelanostné kategórie. Pripomíname, že všetky výsledky sú na priloženom DVD ako aj program.



Rok	2008			
	Úmrtnosť	Pocet	Matlab	Zmena
NV	413	320898	320462	-0,13587
ZS	87	459173	461317	0,466926
ZB	11931	181800	159978	-12,0033
U	134	120387	116986	-2,82506
Uc	11797	928200	874867	-5,74585
US	134	173625	168146	-3,15565
Usc	11797	1179800	1139800	-3,39041
VS	270	140599	117598	-16,3593
VSc	11527	401400	310850	-22,5585
Spolu		3905882	3670004	-6,03905
Porodnosť		57360		

Obr. 2.12: Reálne čísla a výpočet našou simuláciou.

Tabuľka 2.12 je štruktúrovaná rovnako ako tabuľka 2.5 z kapitoly 2.1.3. Rovnako, ako v modeli starnutia, aj v modeli vzdelania sme použili informáciu o počte ľudí vo vzdelanostných kategóriách v roku 2001 ako počiatočnú podmienku a pôrodnosti a úmrtnosti v ďalších rokoch. V tomto prípade, keď populáciu delíme na vzdelanostné skupiny, celkové výsledky sú menej presné ako v modeli starnutia. Spôsobili to najmä nesplnené predpoklady a nekonzistentné dáta pre túto časť modelu. Ako sme sa túto chybu snažili zmierniť a prečo v modeli chýba veľké množstvo ľudí sme vysvetlili v kapitole 2.2.1.



Obr. 2.13: Vzdelanostná štruktúra populácie SR a percentuálna zmena výpočtu oproti skutočnosti za obdobie 2001 – 2008.

Na obrázku 2.13 vidíme, ako sa menili počty ľudí vo vzdelanostných kategóriách spolu oproti skutočným počtom za obdobie 2001 – 2008. Zelený stĺpec (*Real*)

predstavuje skutočné počty ľudí a oranžový (*Výpočet*) počty získané naším stochastickým algoritmom. Graf na pravej časti obrázku znázorňuje, ako sa výsledky vyvíjali v čase.

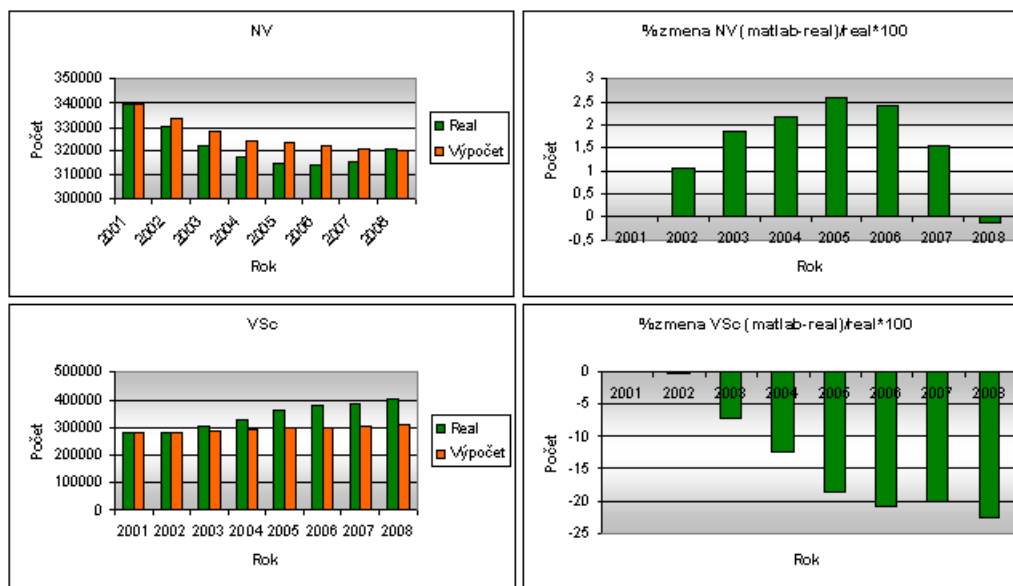
Teraz uvedieme výsledky pre dve vzdelanostné kategórie, *NV* a *VSc*. V kategórii *VSc* vznikli oproti reálnym dátam najväčšie nepresnosti a v kategórii *NV* najmenšie.

<i>NV</i>	<i>real</i>	<i>matlab</i>	<i>zmena</i>	<i>VSc</i>	<i>real</i>	<i>matlab</i>	<i>zmena</i>
2001	340079	340079	0	2001	279500	279500	0
2002	330198	333684	1,05573	2002	282600	281859	-0,26221
2003	322150	328115	1,851622	2003	307200	285189	-7,16504
2004	317440	324338	2,173009	2004	330700	289283	-12,524
2005	315176	323340	2,590299	2005	361700	294853	-18,4813
2006	314260	321861	2,418698	2006	379100	299836	-20,9085
2007	315808	320700	1,549042	2007	381900	305459	-20,016
2008	320898	320462	-0,13587	2008	401400	310850	-22,5585

Obr. 2.14: Ľavá tabuľka obsahuje počty vzdelanostnej kategórie *NV*, ktorej simulácia vyšla najlepšie. Pravá tabuľka popisuje vzdelanostnú kategóriu *VSc*, ktorej simulácia vyšla najhoršie.

V tabuľke na obrázku 2.14 sú uvedené reálne počty ľudí (stĺpce *real*) spomínaných vzdelanostných kategórií a aj počty získané stochastickou simuláciou (stĺpce *matlab*) za celé sledované obdobie.

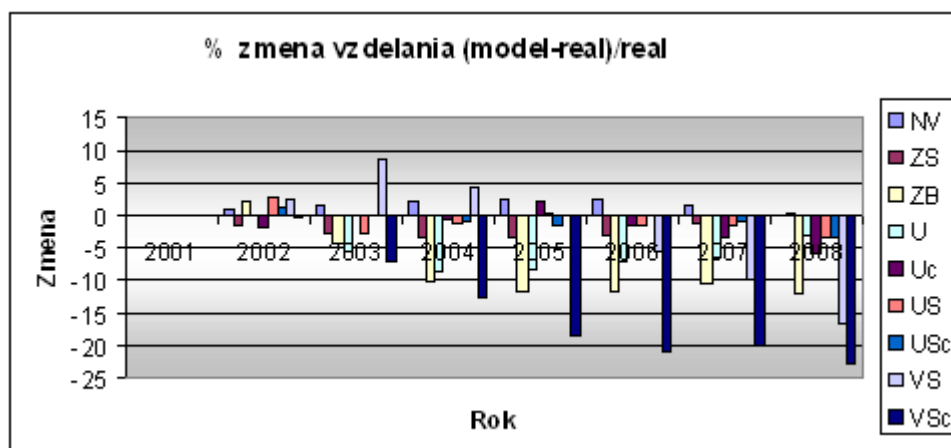
V simulácií vzdelávania algoritmus zachytil trend vývoja populácie v každej kategórii. Najlepší výsledok sme dostali pre kategóriu detí *NV*, o ktorých sme mali najpresnejšie dáta a pochádzali z jedného zdroja. O ostatných kategóriách sme museli vysloviť veľa predpokladov, ktoré neboli splnené. Navyše dáta, s ktorými sme pracovali, pochádzali z viacerých zdrojov a neprislúchali presne vzdelanostným kategóriám s akými sme pracovali v našom modeli vzdelávania. Najviac sa táto chyba prejavila v kategórii *VSc*, čo je kategória ľudí s ukončeným vysokoškolským vzdelaním. Všimnime si, ako narastal počet ľudí v tejto vekovej kategórii (cca plus 20000 za rok). Naš algoritmus presunul do kategórie *VSc* vždy len tých ľudí, ktorí predtým chodili na vysokú školu. Počet študentov VŠ však rastie pomalšie. V sku-



Obr. 2.15: Vzdelanostné kategórie *NV* a *VSc* a ich percentuálne zmeny oproti skutočnosti za obdobie 2001 – 2008.

točnom svete sa do tejto vzdelanostnej kategórie dostávajú aj ľudia, ktorí na VŠ nikdy nechodili, resp. študovali externe alebo chodili do večerných škôl. Externisti sa v kategórii *VS* nenachádzajú pretože predpokladáme, že pracujú, a teda sa už nachádzajú v niektorej kategórii s ukončeným nižším vzdelaním. Po tom, ako niekto dokončí vysokú školu (nie dennou formou štúdia), prestupuje z kategórie s nižším vzdelaním do kategórie *VSc*, ktorú náš algoritmus neprípúšťa. Pridaním ďalších reakcií vieme tento problém odstrániť. Nekonzistentné dáta však spôsobili, že v modeli chýba veľké množstvo ľudí. My sme sa snažili modelovať vzdelanie tak, aby vzniknuté chyby boli čo najmenšie pre všetky vzdelanostné kategórie.

Nakoniec uvádzame graf (obrázok 2.16), na ktorom sú zobrazené percentuálne zmeny počtov ľudí oproti skutočným dátam pre všetky vekové kategórie za celé sledované obdobie 2001 – 2008.



Obr. 2.16: Percentuálne zmeny výpočtu oproti skutočnosti všetkých vzdelanostných kategórií za obdobie 2001 – 2008.

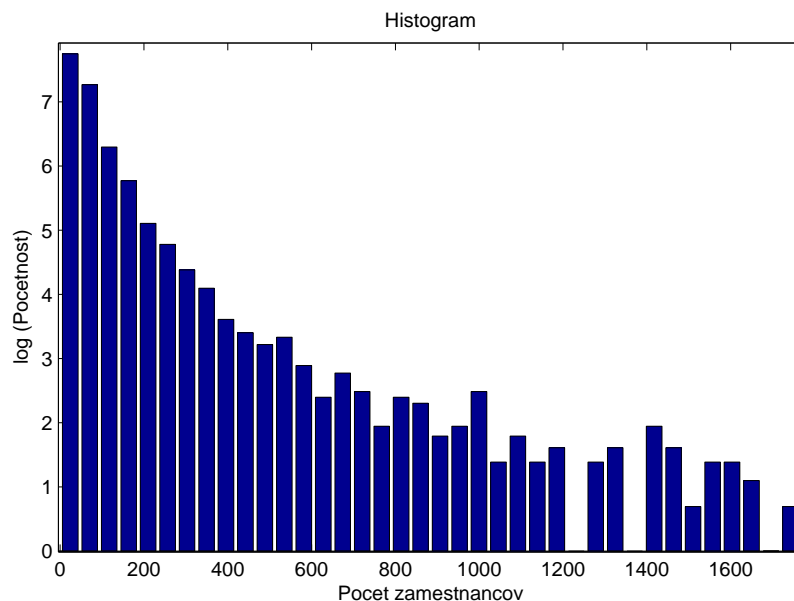
## 2.3 Model zamestnanosti

V predchádzajúcich kapitolách sme sa zaoberali najdôležitejšími vlastnosťami zamestnancov, ktorými boli vek a vzdelanie. V tejto kapitole integrujeme modely veku a vzdelania do komplexného modelu pre zamestnanosť v SR.

Zopakujeme, že pre jednoduchosť predpokladáme, že firmy majú tri základné charakteristiky: veľkosť, priemerný vek a priemerné vzdelanie zamestnancov. Tieto vlastnosti budú počas priebehu stochastickej simulácie aktualizované s prihliadnutím na nasledujúce princípy interakcie medzi modelmi:

- Zmena veku jednotlivca má vplyv na vekovú štruktúru obyvateľov SR, ale aj na priemerný vek firmy, v ktorej tento jednotlivec pracuje.
- Zmena vzdelania jednotlivca má vplyv na vzdelanostnú štruktúru obyvateľstva SR, ale aj na priemerné vzdelanie firmy, v ktorej tento jednotlivec pracuje.

Pri zostavovaní modelu zamestnanosti postupujeme podobne ako v predchádzajúcich kapitolách. Všetky firmy rozdelíme do niekoľkých kategórií podľa veľkosti.

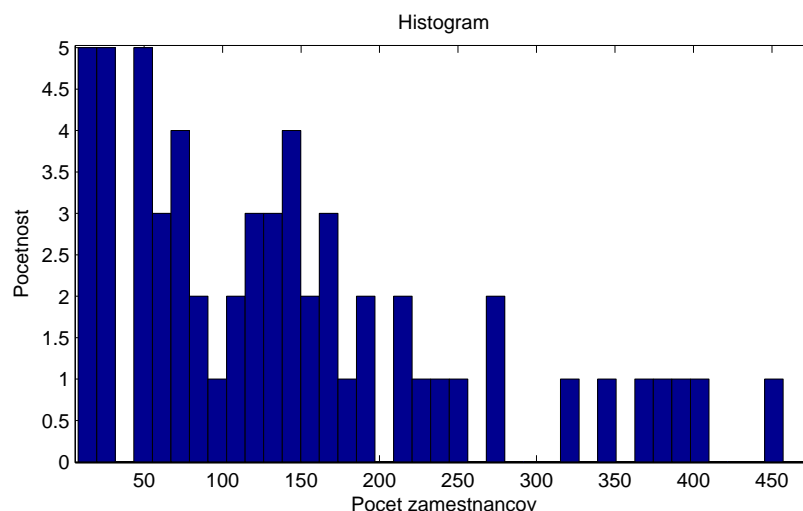


Obr. 2.17: Histogram firiem zostavený z výberového zisťovania

Na obrázku 2.17 je histogram zostavený z výberovej vzorky viac ako 5000 firiem. Na  $X$ -ovej osi sú znázornené počty zamestnancov každej firmy. Na  $Y$ -ovej osi sú logaritmy početnosti. Vidíme, že najviac je firiem s malým počtom zamestnancov a tento počet rýchlo klesá s nárastom počtu zamestnancov vo firme. Tieto dáta nám neposkytujú žiadnu informáciu o tom, ako firmy prirodzene rozdeliť podľa veľkosti. To znamená, že za malú firmu môžeme považovať samoobsluhu v obci s tromi zamestnancami, ale aj tlačiareň novín s 300 zamestnancami.

Ak použijeme hrubšie dáta zo súhrnných tabuliek štatistického úradu a počtoch zamestnancov v rôznych odvetviach hospodárstva SR, dostaneme histogram na obrázku 2.18. V nich sú zahrnuté všetky podniky a všetci evidovaní zamestnanci. Pri zostavení tohto histogramu sme predpokladali, že počet zamestnancov na firmu je približne rovný počtu zamestnancov pracujúcich v danom odvetví k počtu firiem daného odvetvia. Na základe obrázku 2.18 sme rozdelili firmy podľa veľkosti do štyroch kategórií:

**M:** malé firmy (1 – 40 zamestnancov).



Obr. 2.18: Histogram firiem zostavený zo súhrnných tabuliek o počtoch zamestnancov

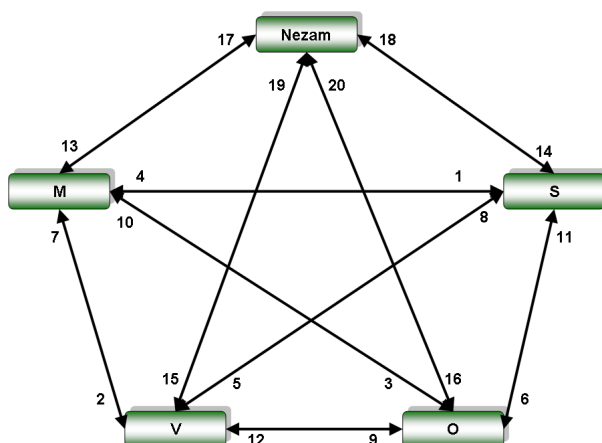
**S:** stredne veľké firmy (41 – 100 zamestnancov).

**V:** veľké firmy (101 – 200 zamestnancov).

**O:** veľmi veľké firmy (201+ zamestnancov).

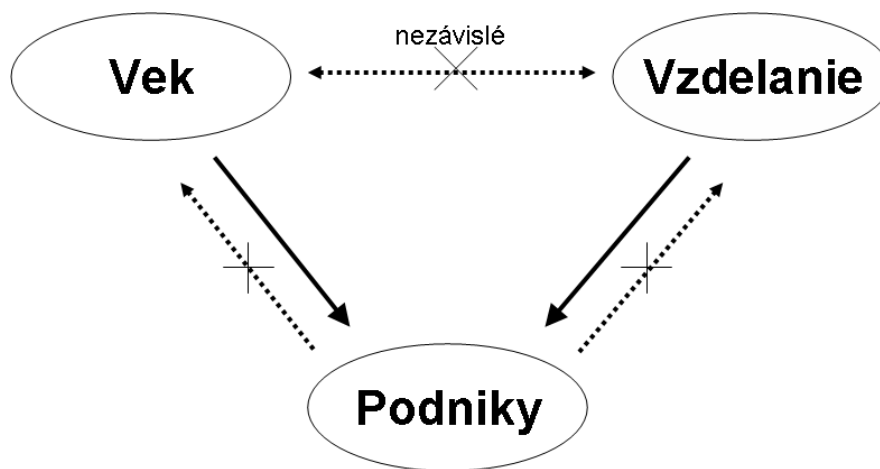
**N:** nezamestnaní.

Pripomeňme, že v algoritme stochastickej simulácie nebudeme sledovať konkrétne firmy, ale iba typy firiem. Posledná kategória samozrejme nie je kategóriou firiem. Po jej pridaní budú v modeli zahrnutí všetci ekonomicky aktívni jedinci, vrátane čerstvo skončených študentov a nezamestnaných. V prenesenom význame, všetkých nezamestnaných zamestnáme vo firme, kde pracujú nezamestnaní. V tomto modeli zamestnanosti je možné prestúpiť z akejkoľvek kategórie do akejkoľvek inej. Tieto prestupy medzi firmami predstavujú reakcie v modeli zamestnanosti. Nasledujúca schéma a tabuľka zobrazujú tieto reakcie a ich rýchlosti. V tabuľke 2.4 neuvádzame konkrétne rýchlosti reakcií, pretože práve tieto sú predmetom nášho výskumu. Konštantu  $1/12$  uvádzame preto, lebo našou časovou jednotkou je mesiac.



Obr. 2.19: Schematické zobrazenie reakcií pre model zamestnanosti

### 2.3.1 Vzťahy medzi modelmi



Obr. 2.20: Vzťahy medzi modelmi.

Model zamestnanosti sa skladá z troch submodelov, modelu starnutia, modelu vzdelávania a modelu zamestnanosti. Na obrázku 2.20 sú znázornené vzťahy medzi týmito modelmi. Reakcie, ktoré prebiehajú v modeli starnutia neovplyvňujú model vzdelávania a naopak. Avšak reakcie v modeloch starnutia a vzdelávania ovplyvňujú vlastnosti firiem, a teda aj rýchlosti reakcií modelu zamestnanosti. Napríklad,

$j$	Reakcia	$c_j$	$a_j(\mathbf{x})$
1	$M \rightarrow S$	$1/12 * c'_1$	$1/12 * c'_1 * M$
2	$M \rightarrow V$	$1/12 * c'_2$	$1/12 * c'_2 * M$
3	$M \rightarrow O$	$1/12 * c'_3$	$1/12 * c'_3 * M$
4	$S \rightarrow M$	$1/12 * c'_4$	$1/12 * c'_4 * S$
5	$S \rightarrow V$	$1/12 * c'_5$	$1/12 * c'_5 * S$
6	$S \rightarrow O$	$1/12 * c'_6$	$1/12 * c'_6 * S$
7	$V \rightarrow M$	$1/12 * c'_7$	$1/12 * c'_7 * V$
8	$V \rightarrow S$	$1/12 * c'_8$	$1/12 * c'_8 * V$
9	$V \rightarrow O$	$1/12 * c'_9$	$1/12 * c'_9 * V$
10	$O \rightarrow M$	$1/12 * c'_{10}$	$1/12 * c'_{10} * O$
11	$O \rightarrow S$	$1/12 * c'_{11}$	$1/12 * c'_{11} * O$
12	$O \rightarrow V$	$1/12 * c'_{12}$	$1/12 * c'_{12} * O$
13	$N \rightarrow M$	$1/12 * c'_{13}$	$1/12 * c'_{13} * N$
14	$N \rightarrow S$	$1/12 * c'_{14}$	$1/12 * c'_{14} * N$
15	$N \rightarrow V$	$1/12 * c'_{15}$	$1/12 * c'_{15} * N$
16	$N \rightarrow O$	$1/12 * c'_{16}$	$1/12 * c'_{16} * N$
17	$M \rightarrow N$	$1/12 * c'_{17}$	$1/12 * c'_{17} * M$
18	$S \rightarrow N$	$1/12 * c'_{18}$	$1/12 * c'_{18} * S$
19	$V \rightarrow N$	$1/12 * c'_{19}$	$1/12 * c'_{19} * V$
20	$O \rightarrow N$	$1/12 * c'_{20}$	$1/12 * c'_{20} * O$

Tabuľka 2.4: Reakcie, rýchlosti reakcií a adaptované funkcie rýchlostí pre model zamestnanosti

ak v nejakom časovom okamihu zostarne nejaký pracujúci človek, nezmenia sa len počty ľudí vo vekových kategóriách, ale zmenia sa aj vlastnosti firmy, v ktorej tento človek pracuje. Model zamestnanosti zase neovplyvňuje modely starnutia a vzdelávania. Napríklad, ak sa človek rozhodne zmeniť zamestnanie, nijako sa tým nezmení veková alebo vzdelanostná štruktúra obyvateľstva SR.

V reálnom svete pravdepodobne nie sú rovnakí zamestnanci v každom type firmy. Pravdepodobnejšie je, že napríklad v malej firme venujúcej sa počítačovým aplikáciám pracujú mladí ľudia s vyštudovanou vysokou školou informatického smeru. Naopak, vo veľkej fabrike na spracúvanie ocele pracujú robotníci s nižším vzdelaním a s vyšším vekom. To nás vedie k charakterizovaniu kategórií firiem pomocou veku a vzdelania.

Najjednoduchší spôsob, ako charakterizovať vek a vzdelanie danej firmy je

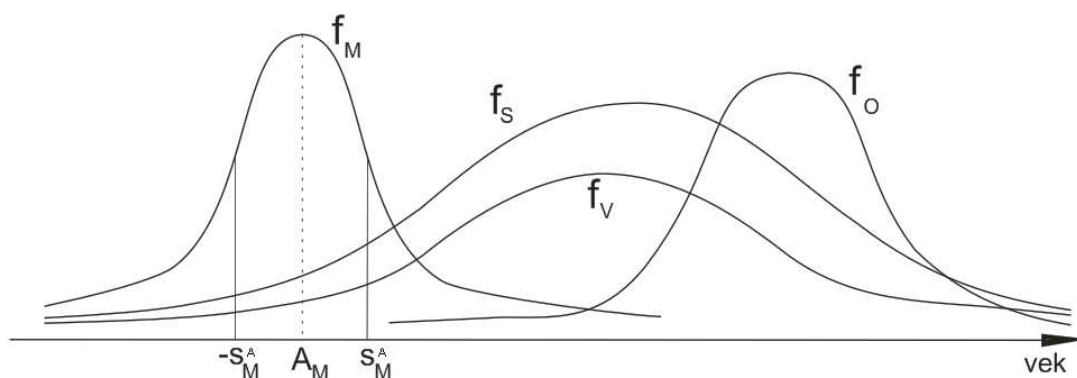


určiť, aký vek a vzdelanie majú jej súčasní zamestnanci. Budeme sledovať dve charakteristiky, strednú hodnotu a štandardnú odchýlku. Vždy, keď nastane nejaká reakcia prestupovania agenta medzi firmami, tieto štatistiky sa zmenia. Podobne, ak nastane nejaká zmena v modeli starnutia alebo vzdelávania, je potrebné tieto štatistiky správne aktualizovať.

Pri modelovaní presunu zamestnancov na trhu predpokladáme:

1. Firma je dobre charakterizovaná veľkosťou ( $N$  je počet zamestnancov), priemerným vekom a vzdelaním ( $A_j, E_j$ ), štandardnou odchýlkou veku a vzdelania ( $S_j^A, S_j^E$ ).
2. Zamestnanci vo firme  $j$  majú vek a vzdelanie rozdelené podľa normálneho rozdelenia  $\mathcal{N}(A_j, S_j^A), \mathcal{N}(E_j, S_j^E)$ .

Kvôli jednoduchosti budeme index  $j = \{M, S, V, O\}$  vynechávať, pokiaľ budeme hovoriť vo všeobecnosti. Na obrázku 2.21 je znázornený príklad, akú vekovú štruktúru môžu mať kategórie firiem.



Obr. 2.21: Ilustračné zobrazenie vekovej štruktúry zamestnancov pre 4 firmy.

Teraz ukážeme, ako aktualizujeme  $A, S^A$ , prípadne  $E, S_E$  pre dva scenáre:

- (a) Firma s  $N$  zamestnancami prijme nového zamestnanca s vekom  $a_0$  a vzdelaním  $e_0$ .

(b) Firma rozviaže pracovnú zmluvu so zamestnancom, ktorého vek a vzdelanie sú  $a_k, e_k$ .

Nech  $a_i$  je vek  $i$ -teho zamestnanca vo vybranom type firmy, v ktorom pracuje  $N$  zamestnancov. Potom priemerný vek v tomto type firmy je

$$A = \frac{1}{N} \sum_{i=1}^N a_i.$$

Ak firma prijme nového zamestnanca s vekom  $a_0$  (prípade (a)), priemerný vek sa zmení na

$$A^+ = \frac{1}{1+N} \sum_{i=0}^N a_i = \frac{1}{N+1} \sum_{i=1}^N a_i + \frac{a_0}{N+1} = \frac{N}{N+1}A + \frac{1}{N+1}a_0,$$

kde  $a_0$  predstavuje vek zamestnanca, ktorý do firmy prišiel. Naopak, ak z kategórie odíde jeden zamestnanec s vekom  $a_k$  (prípade (b)), priemerný vek sa zmení na

$$A^- = \frac{1}{N-1} \sum_{\substack{i=1 \\ i \neq k}}^N a_i = \frac{1}{N-1} \sum_{i=1}^N a_i - \frac{a_k}{N-1} = \frac{N}{N-1}A - \frac{a_k}{N-1},$$

kde  $a_k$  predstavuje vek zamestnanca, ktorý z firmy odišiel. Podobne určíme štandardnú odchýlku veku v prípade (a), keď firme danej kategórie pribudne, resp. ubudne jeden zamestnanec.

$$S^A = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (a_i - A)^2} = \sqrt{\frac{1}{N-1} \sum_{i=1}^N a_i^2 - \frac{1}{N(N-1)} \left( \sum_{i=1}^N a_i \right)^2}.$$

Príchod zamestnanca zmení štandardnú odchýlku na

$$\begin{aligned} S_+^A &= \sqrt{\frac{1}{N} \sum_{i=0}^N a_i^2 - \frac{1}{N} \sum_{i=0}^N a_i \left( \frac{N}{N+1}A + \frac{1}{N+1}a_0 \right) + \frac{1}{N} \sum_{i=0}^N \left( \frac{N}{N+1}A + \frac{1}{N+1}a_0 \right)^2} \\ &= \sqrt{\frac{1}{N+1}a_0^2 + \frac{(N+1-2a_0)A}{N+1} - \frac{NA^2}{N+1}}. \end{aligned}$$

kde sme využili, že

$$\sum_{i=1}^N a_i^2 = (N-1)(S^2 + \frac{N}{N-1}A^2).$$

Pri odchode zamestnanca pre odchýlku platí

$$S_-^A = \sqrt{\frac{2NAa_k}{N-1} - \frac{(2N-2)a_k^2}{(N-1)(N-2)} - \frac{(N-4)N^2A^2}{(N-1)(N-2)}}.$$

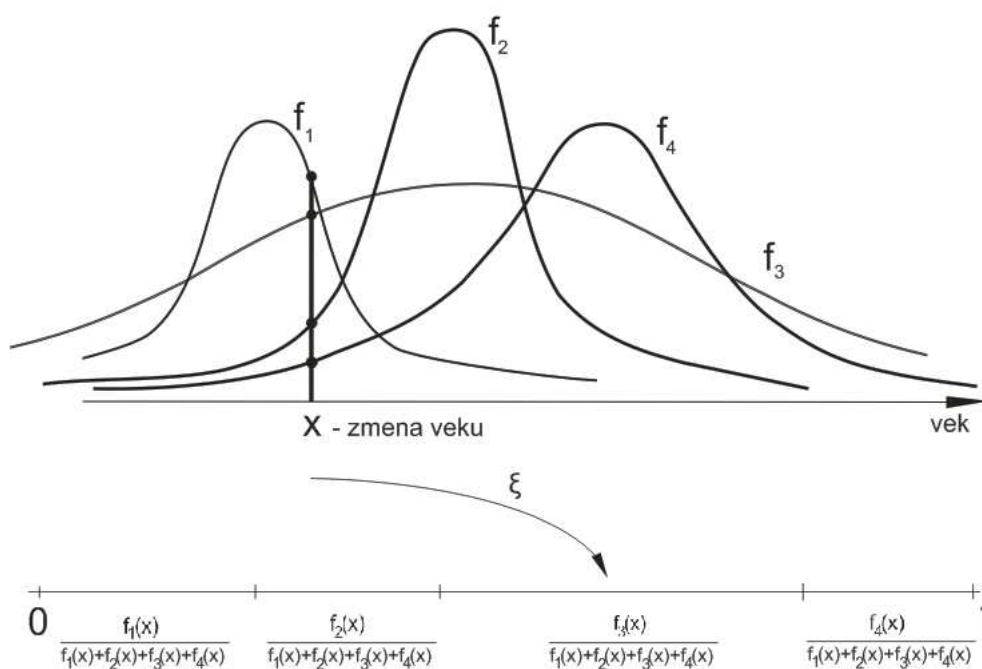
Rovnako vypočítame aj priemerné vzdelanie a štandardnú odchýlku pre vzdelanie a ich zmeny.

Aktualizovaním týchto štatistík vieme vždy, keď sa zrealizuje ľubovoľná reakcia starnutia, vzdelávania alebo zmeny pracovného pôsobiska, sledovať charakteristiky firiem.

Ako sme znázornili na obrázku 2.21, predpokladáme normálne rozdelenie veku a vzdelania zamestnancov. Teraz si ukážeme, ako vyberieme človeka alebo firmu, podľa toho, akého typu je reakcia, ktorá práve nastala. Ak nastane niektorá z reakcií modelu starnutia, potrebujeme určiť, z ktorej firmy je daný zamestnanec, prípadne ktorej firme pribudol zamestnanec v dôsledku dospievania/ukončenia školy, resp. ubudol v dôsledku úmrtia. Pri zmene vekovej kategórie agenta presne vieme, koľko má rokov. V akej firme pracuje určíme tak, že vygenerujeme náhodné číslo  $\xi$  z rovnomerného rozdelenia  $U([0, 1])$  ako je znázornené na nasledujúcom obrázku. Podľa toho, do ktorého podintervalu padne, tú variantu vyberieme.

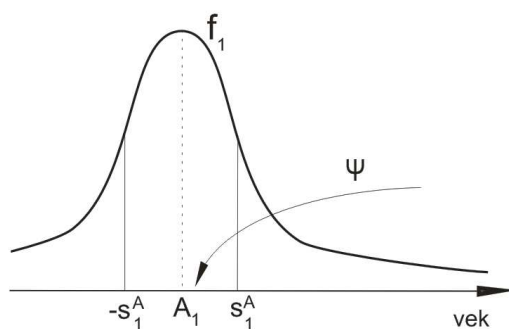
Graf na obrázku 2.22 modeluje situáciu štyroch firiem. V tomto prípade, interval  $f_3(x)/f_{SUM}(x)$  vyjadruje pravdepodobnosť, že človek vo veku  $x$  rokov je zamestnaný v tretej firme. Náhodne vygenerované číslo  $\xi \sim U(0, 1)$  určilo, že zamestnanec pracuje v tretej firme. Podobne postupujeme, ak nastane niektorá z reakcií modelu vzdelávania.

Ak nastane niektorá z reakcií modelu zamestnanosti, vieme v akej firme človek pracoval a naopak, potrebujeme určiť jeho vek a vzdelanie (aby sme vedeli zmeniť



Obr. 2.22: Určenie kategórie firmy, keď nastala reakcia v modeli starnutia

vlastnosti firiem). Keďže vo firmách predpokladáme normálne rozdelenie veku a vzdelania, s ich určením sme hotoví. Vygenerujeme náhodné číslo  $\psi \sim \mathcal{N}(A, S^A)$ . Dostaneme tak vek agenta, ktorý práve zmenil zamestnávateľa. V programe je ešte nutné zabezpečiť, aby vygenerovaný vek bol v rozmedzí 15–59 rokov (produktívny vek v našom modeli). Rovnako postupujeme aj pri určovaní vzdelania.



Obr. 2.23: Určenie veku zamestnanca, keď nastala reakcia v modeli zamestnanosti

### 2.3.2 Výsledky

Teraz uvidíme a popíšeme výsledky pre model zamestnanosti. V zmysle metódy Monte Carlo sme simuláciu spustili 1000 krát a ako výsledok uvádzame aritmetický priemer jednotlivých simulácií. Podľa predpokladov 1 a 2 pre model zamestnanosti sú agenti v jednotlivých typoch firiem normálne rozdelený a charakterizuje ich priemerný vek a vzdelanie a štandardné odchýlky týchto vlastností. Predpokladajme teda, že:

- V kategórii  $M$  pracujú mladí ľudia s vysokým vzdelaním.
- V kategórii  $S$  a  $V$  majú zamestnanci ľubovoľný vek a ľubovoľné vzdelanie.
- V kategórii  $O$  pracujú starší ľudia s nízkym vzdelaním.
- V kategórii  $N$  sa nachádzajú ľudia s ľubovoľným vekom a vzdelaním, lebo v tejto kategórii sa nachádzajú aj čerstvo skončení študenti aj dlhodobo nezamestnaní.

Keď máme informáciu o vekovom a vzdelanostnom zložení populácie, rozdeliť všetkých ľudí do firiem podľa hore uvedeného predpokladu, a určiť tým počiatočnú podmienku pre model zamestnanosti, nie je jednoduché. Existujú metódy [1], ktorými sa tento problém dá vyriešiť, v práci ich však neuvádzame. V našom modeli pre jednoduchosť rozdelíme ľudí rovnomerne.

V celom modeli zamestnanosti sa dynamicky mení veková štruktúra obyvateľov, zo študentov sa stávajú pracujúci ľudia a zamestnanci sa presúvajú medzi firmami. Teraz ukážeme, ako sme určili rýchlosť reakcií v modeli zamestnanosti, ktorá pri zmene zamestnania zohľadňuje nie len pravdepodobnosť, s akou agent zmení typ firmy v ktorej pracuje, ale aj vek a vzdelanie agentov, ktoré sa navyše časom menia. Pre  $j$ -tu reakciu sme rýchlosť určili ako

$$c'_j = \frac{1}{A(t)} E(t) \beta_j,$$

pričom vychádzame z úvahy, že čím je agent straší, tým je pravdepodobnosť, že zmení zamestnanie menšia a čím vyššie má vzdelanie, tým je pravdepodobnosť zmeny zamestnania väčšia. Konštantu beta môžeme rozpísať na

$$\beta_j = \frac{1}{k_1} \frac{1}{k_2} k_3 k_4,$$

kde  $1/k_1$  vyjadruje pravdepodobnosť, s akou agent zmení zamestnanie,  $1/k_2$  vyjadruje pravdepodobnosť, s akou zamestnanec zmení typ firmy v ktorej pracuje a  $k_3$  a  $k_4$  sú regresné koeficienty pre vek a vzdelanie. Vyskúšali sme množstvo scenárov pre rôzne konštanty. Ako tieto konštanty určiť čo najpresnejšie, môžeme opäť formulovať ako úlohu NLP. V tabuľke 2.5 sú uvedené konečné rýchlosti pre všetky typy reakcií získané metódou pokus-omyl, kde  $\Theta = \frac{1}{A(t)}E(t)$ .

$c_1$	$c_2$	$c_3$	$c_4$	$c_5$
$\frac{1}{12} * \Theta * 0.566$	$\frac{1}{12} * \Theta * 0.425$	$\frac{1}{12} * \Theta * 0.425$	$\frac{1}{12} * \Theta * 4.25$	$\frac{1}{12} * \Theta * 3.916$
$c_6$	$c_7$	$c_8$	$c_9$	$c_{10}$
$\frac{1}{12} * \Theta * 1.958$	$\frac{1}{12} * \Theta * 1.541$	$\frac{1}{12} * \Theta * 3.083$	$\frac{1}{12} * \Theta * 3.166$	$\frac{1}{12} * \Theta * 1.5$
$c_{11}$	$c_{12}$	$c_{13}$	$c_{14}$	$c_{15}$
$\frac{1}{12} * \Theta * 1.5$	$\frac{1}{12} * \Theta * 2$	$\frac{1}{12} * \Theta * 1.312$	$\frac{1}{12} * \Theta * 1.313$	$\frac{1}{12} * \Theta * 1.313$
$c_{16}$	$c_{17}$	$c_{18}$	$c_{19}$	$c_{20}$
$\frac{1}{12} * \Theta * 1.362$	$\frac{1}{12} * \Theta * 0.792$	$\frac{1}{12} * \Theta * 0.833$	$\frac{1}{12} * \Theta * 0.833$	$\frac{1}{12} * \Theta * 0.833$

Tabuľka 2.5: Rýchlosti reakcií pre model zamestnanosti.

2008	N1	M1	M2	M3	Z1	Z2	Z3	D	B	Spolu
Pocet	843754	417863	1044961	370073	400609	1018958	395883	361540	553331	5406972
Matlab	949758	425589	1010803	311880	413872	1004843	340726	373251	552609	5383330
Zmena	12,56341	1,848901	-3,26879	-15,7249	3,310616	-1,3852	-13,9328	3,23913	-0,13053	-0,43725

Obr. 2.24: Vekové kategórie modelu zamestnanosti.

V tabuľke na obrázku 2.24 sú znázornené skutočné počty ľudí vo vekových kategóriách z roku 2008 (riadok *Pocet*), počty získané po nasimulovaní siedmich rokov (2001 – 2008), pričom sme použili informáciu o počte ľudí z roku 2001 ako počiatočnú podmienku a pôrodnosti a úmrtnosti v ďalších rokoch (riadok *Matlab*); a percentuálna zmena oproti skutočnosti (riadok *Zmena*).

2008	NV	ZS	ZB	U	Uc	US	Usc	VS	VSc	Spolu
<i>Pocet</i>	320898	459173	181800	120387	928200	173625	1179800	140599	401400	3905882
<i>Matlab</i>	327081	508359	188734	124748	938718	181074	1123791	115052	295334	3802890
<i>Zmena</i>	1,926858	10,71179	3,813875	3,622069	1,133188	4,290137	-4,7473	-18,1698	-26,4241	-2,63683

Obr. 2.25: Vzdelanostné kategórie modelu zamestnanosti.

V tabuľke na obrázku 2.25 sú znázornené skutočné počty ľudí vo vzdelanostných kategóriách z roku 2008 (riadok *Pocet*), počty získané po nasimulovaní siedmich rokov (2001 – 2008), pričom sme použili informáciu o počte ľudí z roku 2001 ako počiatočnú podmienku a pôrodnosti a úmrtnosti v ďalších rokoch (riadok *Matlab*); a percentuálna zmena oproti skutočnosti (riadok *Zmena*).

2008	M	S	V	O	N	Spolu
<i>Pocet</i>	1199784	261411	352257	480613	397135	2691200
<i>Matlab</i>	1141449	273273	358613	479980	379603	2632917
<i>Zmena</i>	-4,86213	4,537778	1,804329	-0,13181	-4,41471	-2,16568

Obr. 2.26: Kategórie firiem modelu zamestnanosti.

V tabuľke na obrázku 2.26 sú znázornené skutočné počty ľudí v jednotlivých kategóriách firiem (riadok *Pocet*), počty získané po nasimulovaní siedmich rokov (riadok *Matlab*) a percentuálna zmena oproti skutočnosti (riadok *Zmena*). Tu sme ako počiatočnú podmienku použili len informáciu o počtoch ľudí v kategóriách firiem z roku 2001<sup>2</sup>.

## 2.4 Možnosti rozšírenia modelu zamestnanosti

V tejto kapitole v krátkosti naznačíme, čím by sa nami prezentovaný model zamestnanosti dal vylepšiť, resp. čo v ňom chýba.

Náš model simuluje prestupovanie zamestnancov medzi firmami v rámci celého územia Slovenskej Republiky. Zavedením geografie by model dostal nový rozmer.

<sup>2</sup>Všetky dáta, ktoré sme používali v modeli zamestnanosti sú zo stránok štatistického úradu SR.

Všetci agenti a všetky firmy by tak dostali novú vlastnosť - polohu. Potom by sa v rýchlostiach reakcií zohľadňovali aj vzdialenosti trvalého bydliska zamestnanca od sídla firmy pri hľadaní práce a vzdialenosti medzi firmami pri zmene zamestnania.

Ďalej do modelu môžeme pridať nové štatistiky, či už zamestnancov alebo firiem, ktoré sme už spomínali v úvode kapitoly 2. Vysvetlili sme, aké vlastnosti a prečo v našom modeli chýbajú, a aké problémy s nimi súvisia. Je dôležité správne odhadnúť, ktoré štatistiky sú relevantné a ktoré iba zvyšujú náročnosť algoritmu a nemajú veľký vplyv na zamestnanosť. Rozšírený model s dobre nastavenými parametrami by mohol priniesť odpovede na veľmi zaujímavé otázky. Jednoduchými úpravami vlastností, či už zamestnancov alebo firiem, vieme sledovať, aký dopad by takéto zmeny mali na zamestnanosť.

Veľmi zaujímavé môže byť pridanie procesov vzniku a zániku firiem. To znamená, že ak sa na nejakom mieste bude nachádzať dostatočný počet agentov s patričnými vlastnosťami, ktorí nepracujú v žiadnej firme, môžu spolu vytvoriť podnik a zamestnať sa v ňom. Alebo, ak sa v danej lokalite nachádza väčší počet menších firiem, venujúcich sa rovnakému odvetviu ekonomickej činnosti, spravia fúziu, prípadne môžeme uvažovať opačný proces, keď sa veľká firma rozdelí na niekoľko menších. V reálnom svete vznik a zánik firiem nie je ničím neobvyklým a bežne k takýmto udalostiam dochádza.

Aby sa zohľadnila globalizácia ekonomík, môžeme do modelu pridať príchod zahraničných firiem na náš trh. Tieto môžu so sebou priniesť nových zamestnancov, ktorí sa začlenia do populácie SR alebo vytvoria nové pracovné príležitosti pre nezamestnaných. V zmysle globalizácie môžeme uvažovať aj o imigrácii a emigrácii pracovnej sily.



# Záver

Cieľom diplomovej práce bolo odhadnúť dôležité vlastnosti zamestnancov a firmami a ich použitím zostaviť model zamestnanosti. Predstavili sme dve možnosti, ako k problematike pristupovať. Pomocou sústavy ODR alebo ako k náhodnému Markovovému procesu.

Nadstavili sme parametre modelu metódou pokus-omyl tak, aby sme pri modelovaní zamestnanosti (a populácie) s použitím historických dát dostali čo najlepší výsledok. Ukázalo sa, že model vie veľmi dobre simulovať vekovú a vzdelanostnú štruktúru obyvateľstva. Avšak presne určiť mieru, s akou na zamestnanosť a jej zmenu pôsobia vlastnosti agentov sa javí ako zložitý problém. Pri zostavovaní modelu sme vyslovili veľa predpokladov, ktoré neboli splnené. S možnosťou prístupu k lepším štatistikám o obyvateľoch Slovenskej Republiky a podrobnejšími údajmi o firmách pôsobiacich na domácom trhu by naša simulácia bola oveľa presnejšia.

Prínos práce spočíva v tom, že sme navrhli nový spôsob ako modelovať zamestnanosť. Miesto porovnávania ekonomických ukazovateľov s inými krajinami sme navrhli ako sledovať pohyb pracovnej sily, ovplyvnený vlastnosťami obyvateľov a ich meniacou sa štruktúrou.

# Literatúra

- [1] BLIEN, U. a GRAEF, F. 2006. *The Application of the Entrop Procedure*, [online], dostupné na internete <<http://www.iioa.org/pdf/Intermediate-2004/559.pdf>>
- [2] DIAZ, B. A. a FENT, T. 2003. *An Agent-based Simulation Model of Age-at-Marriage Norms*, Agent-Based Computational Modelling, Springer, 2006, ISBN 9978-3-7908-1640-2, str. 85-116.
- [3] DUPAC, V. a DUPACOVÁ, J. 1975. *Markovovy procesy I.*, MFF UK, Praha, 1975
- [4] DUPAC, V. a DUPACOVÁ, J. 1976. *Markovovy procesy II.*, MFF UK, Praha, 1976
- [5] FÁTH, F. a SARVARY, M. 2006. *Cultural Evolution in a Population of Heterogeneous Agents*, The Complex Networks of Economic Interaction, Springer, 2006, ISBN 978-3-540-28726, str. 193-205
- [6] GILLESPIE, D. T. 1976. *A General Method for Numerically Simulating the Stochastic Time Evolution of Coupled Chemical Reactions*, Journal of Computational Physics, vol. 22, 1976, str. 403-434.
- [7] HAKEN, H. a MAYER-KRESS, G. 2005. *Chapman-Kolmogorov equation and path integrals for discrete chaos in presence of noise*, Zeitschrift für Physik B Condensed Matter, vol. 43, 2005, str. 185-187.

- [8] HAMALA, M. 1976. *Nelineárne programovanie*, Alfa, 1976, 240 str.
- [9] HIGHAM, D. J. *Modeling and Simulating Chemical Reactions*, [online], dostupné na internete <<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.128.7721\&rep=rep1\&type=pdf>>
- [10] CHEN, S a CHIE, B. 2003. *A Functional Modularity Approach to Agent-based Modelling of the Evolution of Technology*, The Complex Networks of Economic Interaction, Springer, 2006, ISBN 978-3-540-28726-1, str. 165-178.
- [11] KAMPEN, N. G. 2007. *Stochastic processes in physics and chemistry*, North Holland. 2007, ISBN 0444529659, 481 str.
- [12] KRAUSE, A. *Herding Without Following the Herd: The Dynamics of Case-Based Decisions with Local Interactions*, [online], dostupné na internete <<http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=A21878D707B43152D917D724C081EFD5?doi=10.1.1.136.9718\&rep=rep1\&type=pdf>>
- [13] KURTZ, T. G. 1981. *Approximation of Population Processes*, SIAM, Philadelphia, 1981
- [14] MEYER-ORTMANN, H. 2003. *Immigration, Integration and Ghetto Formation*, International Journal of Modern Physics C, vol. 14, 2003, str. 311-320.
- [15] PYKA, A. a GREBEL, T. 2003. *Agent-Based Modeling - A Methodology for the Analysis of Qualitative Development Processes*, [online], dostupné na internete <<http://www.wiwi.uni-augsburg.de/vwl/institut/paper/251.pdf>>
- [16] RENSHAW, E. 1991. *Modelling Biological Populations in Space and Time*, Cambridge, UK, 1991

- 
- [17] TUIA, D. a kol. 2008. *Scan Statistics Analysis of Forest Fire Clusters*, Communications in Nonlinear and Numerical Simulation, vol. 13, 2008, str. 1689-1694.
- [18] VOTER, A.F. 2003. *Introduction to the Kinetic Monte Carlo Method*, [online], dostupné na internete <<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.125.3560\&rep=rep1\&type=pdf>>