Univerzita Karlova v Praze
Matematicko-fyzikální fakulta

BAKALÁŘSKÁ PRÁCE

Petra Kochaniková

Urnové modely s náhodným vracením

Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: doc. RNDr. Zbyněk Pawlas, Ph.D.
Studijní program: Matematika
Studijní obor: Finanční matematika

Praha 2016

Rada by som vyslovila svoju vďaku za pomoc, ktorej sa mi dostalo od doc. RNDr. Zbyňka Pawlasa, Ph.D., ktorý túto prácu opakovane čítal a poskytol mi k nej cenné pripomienky. Chcela by som sa taktiež srdečne poďakovať svojej rodine a priateľom, za ich podporu pri štúdiu.

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

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

V .... dne ....

Název práce: Urnové modely s náhodným vracením

Autor: Petra Kochaniková

Katedra: Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: doc. RNDr. Zbyněk Pawlas, Ph.D., Katedra pravděpodobnosti a matematické statistiky

Abstrakt: V předložené práci čtenářům přiblížíme specifické zobecnění urnových schémat. Blíže prostudujeme urnové modely s náhodným vracením, pro které je typické, že počet kuliček v urně závisí na výsledcích z předešlých tahů. Podrobně popíšeme Ganiho model, kterým můžeme modelovat šíření epidemie v různých populacích. Nejprve budeme předpokládat homogenní populaci, pro kterou odvodíme rozdělení počtu nově infikovaných jedinců pomocí metod využívajících pravděpodobnostní vytvořující funkce. Také odvodíme střední hodnotu a rozptyl rozdělení. Potom budeme předpokládat více reálnou situaci, tedy nehomogenní populaci, pro kterou odvodíme taky pravděpodobnostní rozdělení a rovněž základní charakteristiky rozdělení. Praktickou část práce tvoří simulační studie, jejíž výsledky porovnáme s hodnotami získanými z odvozených vzorců. Na závěr se model pokusíme co nejvíce aplikovat v praxi tím, že se budeme snažit zjistit, kolik osob v populaci musíme naočkovat, aby byl počet nakažených nižší než 5 % celkové populace.

Klíčová slova: urnový model, model s náhodným vracením, pravděpodobnostní vytvořující funkce, šíření jehel

Title: Urn models with stochastic replacements

Author: Petra Kochaniková

Department: Department of Probability and Mathematical Statistics

Supervisor: doc. RNDr. Zbyněk Pawlas, Ph.D., Department of Probability and Mathematical Statistics

Abstract: The thesis purpose is to discuss urn models where the probability of success at any trial depends upon the number of previous successes. Such a scheme allows us to estimate the number of HIV cases among intravenous drug users. The coefficients in known probability generating function will be derived for the number of new infectives generated in both homogenous and inhomogenous population. The expectations and variances of the number of new infectives are also derived for both cases. These derived values will be verified for some fixed number of infectives and susceptibles by simulations. In the end of this thesis the studied model will be applied on a practical example where the effect of vaccination will be studied.

Keywords: urn model, random allocation model, probability generating function, needle sharing

Názov práce: Urnové modely s náhodným vracaním

Autor: Petra Kochaniková

Katedra: Katedra pravděpodobnosti a matematické statistiky

Vedúci bakalárskej práce: doc. RNDr. Zbyněk Pawlas, Ph.D., Katedra pravděpodobnosti a matematické statistiky

Abstrakt: V predloženej práci čitateľovi priblížime špecifické zovšeobecnenie urnových schém, konkrétne budeme študovať urnové modely s náhodným vracaním, pre ktoré je typické, že počet guľôčok v urne závisí na predchádzajúcich ťahoch. Podrobne popíšeme Ganiho model, ktorým možno modelovať šírenie epidémií v populácii. Najprv budeme predpokladať, že populácia je homogénna. Pre tento prípad odvodíme rozdelenie počtu novo infikovaných osôb pomocou pravdepodobnostných vytvárajúcich funkcií a taktiež strednú hodnotu a rozptyl tohto rozdelenia. Neskôr budeme predpokladať viac reálny model, teda nehomogénnu populáciu, pre ktorú taktiež odvodíme pravdepodobnostné rozdelenie a základné charakteristiky rozdelenia. Praktickú časť práce tvoria simulačné štúdie, ktorých výsledky budeme porovnávať s hodnotami získanými pomocou odvodených vzorcov. Na záver sa tento model pokúsime čo najviac aplikovať do praxe a budeme hľadat odpoveď na otázku, koľko osôb v populácii potrebujeme zaočkovať, aby počet nakazených neprekročil 5 % z celkovej hodnoty populácie.

Kľúčové slová: urnový model, model s náhodným vracaním, pravdepodobnostná vytvárajúca funkcia, šírenie ihiel

Obsah

Úvod

Bakalárska práca Urnové modely s náhodným vracaním, ktorá sa vám dostáva do rúk, podáva prehľad o vybraných pravdepodobnostných modeloch, ktoré môžeme stotožniť s urnovými modelmi reprezentujúcimi nejakú formu nákazy. Zároveň by mala čitateľom ukázať, že tieto vzorce sú aplikovateľné v praxi a sú stálou súčasťou ľudských životov.

Teória pravdepodobnosti rovnako ako matematická štatistika má v matematike svoje stabilné miesto. Prvé zmienky o nej nachádzame už v diele známeho gréckeho matematika Platóna „Philebos“. Taktiež grécki a čínski matematici sa zaujímali o riešenie rôznych kombinatorických úloh už v staroveku. V Európe však výraznejší rozvoj teórie pravdepodobnosti datujeme až do 16. a 17. storočia, kedy bol zaznamenaný rozvoj štúdií najstarších hazardných hier, s ktorým šla ruka v ruke snaha porozumieť problémom týkajúcich sa stávkovania. Práve v tejto oblasti sa prvýkrát objavujú urnové modely, ktoré vďaka svojej názornosti, ľahkej konštrukcii a flexibilite poskytovali dobrý prostriedok k modelovaniu rôznych situácií a zjednodušeniu úvah. Mnohé výsledky z diskrétnej pravdepodobnosti odvodené pomocou urnových modelov môžeme rovnako dobre odvodiť napr. hádzaním mince alebo hádzaním kocky, avšak urnové modely majú v porovnaní s mincami, kockami, kartami isté výhody. Vo vyššie zmienených príkladoch máme vopred obmedzenú škálu, máme 6 čísel na kocke, 52 kariet. V urnových modeloch v tomto smere limitovaní nie sme.

Urnové modely sú najčastejšie konštruované dvomi spôsobmi. Prvý je ten, že uvažujeme urnu naplnenú guľôčkami, ktoré následne žrebujeme podľa vopred stanovených pravidiel. Tieto pravidlá môžu zahrňovať buď vrátenie alebo nevrátenie vytiahnutej guľôčky späť do urny alebo napríklad výmenu vytiahnutej guľôčky za nejakú inú. V tomto prípade môžeme chcieť zisťovať rozdelenie počtu guľôčok v urne alebo čas, ktorý musíme prečkať, kým sa splní určitá vopred stanovená podmienka. Druhý spôsob je taký, že guľôčky do urny umiestňujeme a zaujíma nás, s akou pravdepodobnosťou určitý počet urien nebude obsahovať žiadne guľôčky. Umiestňovacie modely, ktoré sú mnohokrát nazývané alokačné, slúžia aj k formulácii zábavných problémov, ako je problém zberateľa kartičiek alebo problém šatniarky. Často sa stretávame s tým, že urnové modely nachádzajú využitie ako didaktická pomôcka pri popise známych pravdepodobnostných rozdelení, ako sú napríklad binomické, hypergeometrické, multinomické či mnohorozmerné negatívne hypergeometrické rozdelenie. Detailnejší popis týchto rozdelení môže čitateľ nájsť v Johnson a Kotz (1977).

Bohaté aplikácie urnových modelov vidíme aj v súčasnosti, čo súvisí najmä s rozvojom rôznych epidémií a populačných chorôb. Modelovaním šírenia týchto epidémií dostávame nástroj, ktorý nám umožní lepšie pochopiť mechanizmus priebehu choroby, a tým predikovať budúci vývoj študovaných ochorení, no hlavne pomôže vyvinúť vhodnú stratégiu, ako vypuknutú nákazu dostať pod kontrolu. Táto stratégia najčastejšie zahrňuje izoláciu, vakcináciu populácie alebo zostavenie edukačného programu. Tretia spomenutá metóda je typická práve pre HIV ochorenie, ktoré je v súčastnosti považované za celosvetovo rozšírený problém. Podľa Svetovej zdravotníckej organizácie (WHO) v roku 2014 žilo s týmto ochorením až 36,9 miliónov ľudí. Pre porovnanie v roku 2001 to bolo 29,8 miliónov, čo je nárast o 23,85 %. Práve modelom, ktorý je špecifický pre toto ochorenie, sa budeme zaoberať v tejto bakalárskej práci.

Najprv sa však vrátime do histórie a v 1. kapitole si popíšeme, ako sa tieto modely vyvíjali. V ďalších 2 kapitolách budeme podrobnejšie študovať už spomínaný urnový model s náhodným vracaním, ktorý sa používa k popisu šírenia HIV a hepatitídy. Konkrétne v 2. kapitole budeme uvažovať šírenie i infikovaných ihiel medzi fixným počtom n homogénnych osôb. Popíšeme si pravdepodobnostné rozdelenie počtu novo infikovaných osôb, odvodíme vzorec pre výpočet očakávaného počtu novo infikovaných jedincov po opakovanom používaní týchto infikovaných ihiel a ukážeme si odvodenie 2. centrálneho momentu. V závere 2. kapitoly sa pozrieme na praktické využitie a model aplikujeme na konkrétne číselné hodnoty. V poslednej kapitole si ukážeme viac realistický model a budeme uvažovať najprv všeobecne k skupín nehomogénnych drogových užívateľov, pre ktorých odvodíme rovnaké charakteristiky a simulácie ako v 2. kapitole, no kvôli zložitosti algebry si detailnejšie popíšeme tento problém iba pre dve skupiny užívateľov. V úplnom závere 3. kapitoly si ukážeme špeciálny prípad tohto modelu, keď skonkretizujeme tieto dve uvažované skupiny na vakcinované a nevakcinované osoby. Budeme predikovať, akú časť populácie musíme zaočkovať, aby sme neprekročili 5-percentný násobok nakazených ľudí v uvažovanej populácii.

Kapitola 1

História

Už v roku 1923 Eggenberger a Pólya predstavili svetu urnovú schému, dnes známu ako Pólyov urnový model, ktorá modelovala proces šírenia nákazlivej choroby. Modely tohto typu sa zvyknú taktiež označovať ako urnové modely s náhodným vracaním. Typickým znakom týchto modelov je žrebovanie s nejakým nahradením, ale nejde o jednoduché nahradenie poslednej vytiahnutej guľôčky. Tento posledný ťah je vo väčšine modelov kľúčovým, pretože určuje druh nahradenia. V už spomínanom Pólya-Eggenbergerovom modeli uvažujeme urnu s dvomi farbami guľôčok, povedzme bielymi a čiernymi. Po každom ťahu spolu s vytiahnutou guľôčku vrátime späť vopred stanovený počet guľôčok ostatných farieb. Ak sme vytiahli čiernu (bielu), do urny pridáme β_c (β_b) čiernych a ω_c (ω_b) bielych guľôčok. Ak sme napr. po n ťahoch mali v urne $B_n = b_n$ bielych a $C_n = c_n$ čiernych guľôčok, počet čiernych guľôčok odvodíme pomocou vzťahu

$$C_{n+1} = c_n + T_n\beta_c + (1 - T_n)\beta_b,$$

kde T_n je náhodná veličina, pre ktorú platí

$$\mathsf{P}[T_n = 0] = \frac{b_n}{c_n + b_n},$$

$$\mathsf{P}[T_n = 1] = \frac{c_n}{c_n + b_n}.$$

V Johnson a Kotz (1977, str. 184 – 185) je možné nájsť odvodenie momentov náhodnej veličiny C_{n+1}.

Americký matematik Max A. Woodbury tento urnový model veľmi prirodzene využil k popisu šírenia HIV a hepatitídy medzi drogovými užívateľmi zdieľajúcimi infikované ihly, uvažujúc pritom $\beta_b = \omega_b = 0$ a $\beta_c = -\omega_c > 0$. Slovne to znamená, že vytiahnutie bielej guľôčky nemení doterajšie rozloženie farieb v urne, ale vytiahnutie čiernej znamená nahradenie ω_c bielych guľôčok rovnakým počtom β_c čiernych. Woodbury (1949) popísal pravdepodobnostné rozdelenie náhodnej veličiny, ktorá popisuje, že urna obsahuje po n pokusoch x čiernych guľôčok, pomocou Newtonovho interpolačného polynómu, kým J. Gani na popis tohto pravdepodobnostného rozdelenia využil metódu pravdepodobnostných vytvárujúcich funkcií. A práve tento Ganiho prístup bude dôkladne popísaný v nasledujúcich kapitolách.

Kapitola 2

Homogénny model

2.1 Základné pojmy a vlastnosti

Na začiatku tejto kapitoly si zadefinujeme pojem pravdepodobnostnej vytvárajúcej funkcie, s ktorou budeme v nasledujúcom texte pracovať.

Definícia 1. *Nech X je diskrétna náhodná veličina nadobúdajúca celočíselné nezáporné hodnoty a označme $\mathsf{P}(X = k) = p_k$, pre $k \in \{0, 1, \dots\}$ pravdepodobnostné rozdelenie tejto náhodnej veličiny. Potom sa funkcia $f(z) = \mathsf{E}\left[z^X\right] = \sum_{k=0}^{\infty} z^k p_k$ nazýva pravdepodobnostná vytvárajúca funkcia náhodnej veličiny X.*

Pretože platí $\sum_{k=0}^{\infty} p_k = 1$, rada v definícii konverguje pre $|z| < 1$. Ďalej uvedieme základné vlastnosti pravdepodobnostnej vytvárajúcej funkcie.

Veta 1. *Nech X je diskrétna náhodná veličina nadobúdajúca celočíselné nezáporné hodnoty $k \in \{0, 1, \dots\}$ a $f(z)$ je pradepodobnostná vytvárajúca funkcia náhodnej veličiny X. Potom $\mathsf{E}[X] = \lim_{z \to 1_-} f'(z) = f'(1)$, kde $f'(z)$ označuje 1. deriváciu f podľa z.*

Dôkaz. Najprv ukážeme, ako vyzerá derivácia pravdepodobnostnej vytvárajúcej funkcie

$$f'(z) = \sum_{k=1}^{\infty} k p_k z^{k-1} = \frac{1}{z} \sum_{k=0}^{\infty} k p_k z^k,$$

pre $0 < z < 1$.

Podľa Ábelovej vety, ktorej znenie môžeme nájsť v skriptách Prášková a Lachout (2001, str. 135), aplikovanej na postupnosť kp_k dostávame

$$\lim_{z \to 1_-} f'(z) = f'(1) = \sum_{k=1}^{\infty} k p_k = \mathsf{E}[X].$$

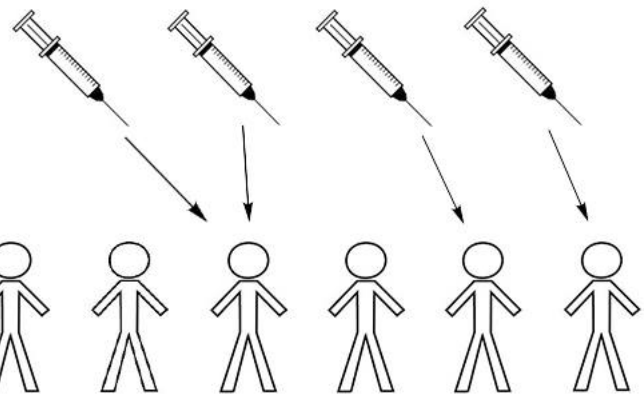
❑

Poznámka. Z dôkazu vety 1 vyplýva, že pravdepodobnostná vytvárajúca funkcia jednoznačne určuje pravdepodobnostné rozdelenie. Postupnosť pravdepodobností totiž vytvárame ďalším derivovaním $f^{(k)}(0) = k! p_k, k \in \{0, 1, \dots\}$.

2.2 Popis modelu

V tejto časti práce sa budeme venovať urnovému modelu s náhodným vracaním, ktorý je vhodný k popisu šírenia infekčných ochorení. Budeme vychádzať z Gani (2004). Je zrejmé, že vizualizácia pomocou urnového modelu je zasa ľahko predstaviteľná, no my v tejto chvíli opustíme urnovú terminológiu a viac sa zameriame na praktickú aplikáciu tohto modelu. Teda namiesto urny obsahujúcej na začiatku n bielych guľôčok, z ktorej i-krát žrebujeme, uvažujeme n zdravých osôb, ktoré medzi sebou náhodne zdieľajú i ihiel infikovaných infekčnou chorobou ako napr. HIV/AIDS alebo hepatitída. Pre lepšiu grafickú predstavu uvádzame obrázok 2.1.

V urnovom kontexte sme po každom vytiahnutí bielej guľôčky urobili výmenu za čiernu a zaujímal nás celkový počet čiernych guľôčok v urne. V epidemiologickej terminológii nás bude zaujímať pravdepodobnostné rozdelenie počtu infikovaných jedincov, ktoré popíšeme pomocou pravdepodobnostnej vytvárajúcej funkcie, a rovnako odvodíme strednú hodnotu a rozptyl tohto rozdelenia.



Obr. 2.1: Schéma náhodného šírenia infikovaných ihiel medzi osobami

Uvedomíme si, že každá ihla je použitá na daného zdravého či infikovaného jedinca s pravdepodobnosťou $1/n$ a nezávisí na ostatných ihlách. Ďalej uvažujeme, že pravdepodobnosť infikovania zdravého jedinca po zdieľaní infikovanej ihly je 1. Ihly s vírusom sa náhodne šíria medzi n zdravými jedincami, pričom infikujú S z nich, kde S je náhodná veličina, ktorá môže nadobúdať hodnoty $1, \ldots, \min(i,n)$.

Počet infikovaných ihiel môže byť ľubovoľné nezáporné číslo. V našom modeli pre jednoduchosť neberieme do úvahy, že po určitom čase sa môže infikovaný jedinec vyliečiť. Teda osoba, ktorá použila infikovanú ihlu aspoň jedenkrát, už ostane infikovaná počas celého priebehu náhodného umiesňovania ihiel.

Odvodíme rekurentnú formulu udávajúcu pravdepodobnostné rozdelenie náhodnej veličiny S, ktorá určuje počet infikovaných ľudí medzi n zdravými jedincami. Nech S_j označuje počet užívateľov, ktorí prišli do styku s práve j ihlami, kde $j = 0, \ldots, i$. Je dobré si uvedomiť, že veličina S_i, ktorá udáva četnosť osôb prichádzajúcich do styku s i ihlami, je špeciálny prípad a môže nadobúdať iba hodnoty 0 alebo 1. Nás zaujímajú iba infikovaní jedinci, preto môžme vyššie zadefinovanú veličinu S udávajúcu počet nakazených prepísať nasledujúcim spôsobom

$$S = \sum_{j=1}^{i} S_j.$$

Vďaka tomuto značeniu môžeme zapísať podmienenú pravdepodobnosť určujúcu, že ochorenie sa prejaví u s jedincov, ktorí boli v kontakte s 1 alebo viacerými ihlami, za podmienky, že máme n ľudí a i ihiel vzťahom

$$p_s(i,n) = \mathsf{P}\left(S = s \,\middle|\, \sum_{j=0}^{i} S_j = n, \sum_{j=0}^{i} jS_j = i \right).$$

Ešte uvedieme okrajové podmienky platiace pre túto pravdepodobnosť. Platí $p_s(i,n) = 0$, ak $s < 0$ alebo $s > i$. Slovne, v modeli nemôžeme mať záporný počet nakazených alebo viac nakazených osôb ako ihiel, ktoré sú k dispozícii.

Uvažujme, že pridáme jednu infikovanú ihlu. Potom pre výpočet pravdepodobností dostávame rekurzívnu rovnicu

$$p_s(i+1,n) = p_s(i,n)\frac{s}{n} + p_{s-1}(i,n)\left(1 - \frac{s-1}{n}\right), \tag{2.1}$$

kde $s = 1, 2, \ldots, \min(i+1, n)$. Teda pridanie 1 infikovanej ihly znamená, že buď sme mali s infikovaných jedincov pri počte ihiel i a pridaná ihla zasiahla nejakého z infikovaných jedincov, alebo sme mali $s-1$ infikovaných jedincov pri počte ihiel i a pridaná ihla zasiahla nejakého zo zdravých jedincov.

Zavedieme pravdepodobnostnú vytvárajúcu funkciu $f_{i,n}(u)$ náhodného umiestňovania i infikovaných ihiel

$$f_{i,n}(u) = \sum_{s=0}^{\min(i,n)} p_s(i,n)u^s,$$

kde $0 \leq u \leq 1$.

Kvôli prehľadnosti označíme $m = \min(i+1, n)$. Využitím rovnice 2.1 odvodíme pravdepodobnostnú vytvárajúcu funkciu $f_{i+1,n}(u)$ pomocou nasledujúcich úprav:

- Uvažujeme $i + 1 \leq n$. Potom sa pravdepodobnostná vytvárajúca funkcia zmení na

$$\begin{aligned}
f_{i+1,n}(u) &= \sum_{s=0}^{m} p_s(i+1,n)u^s \\
&= \sum_{s=0}^{m} p_{s-1}(i,n)u^s + \frac{1}{n}\sum_{s=0}^{m} p_s(i,n)su^s - \frac{1}{n}\sum_{s=0}^{m} p_{s-1}(i,n)(s-1)u^s \\
&= \sum_{s=0}^{i} p_s(i,n)u^{s+1} + \frac{1}{n}\sum_{s=0}^{i+1} p_s(i,n)su^s - \frac{1}{n}\sum_{s=0}^{i} sp_s(i,n)u^{s+1} \qquad (2.2) \\
&= uf_{i,n}(u) + \frac{u}{n}\sum_{s=0}^{i+1} sp_s(i,n)u^{s-1} - \frac{u^2}{n}\sum_{s=0}^{i} sp_s(i,n)u^{s-1} \\
&= uf_{i,n}(u) + \frac{u}{n}\frac{\partial f_{i,n}(u)}{\partial u} - \frac{u^2}{n}\frac{\partial f_{i,n}(u)}{\partial u}.
\end{aligned}$$

Ešte si uvedomíme, že v prvom a treťom sčítanci vo výraze 2.2 sme substituovali indexovaciu premennú v sume, takže sme zamenili jej medze. Dolná hranica je po substitúcii $s = -1$, no keď si uvedomíme okrajové podmienky, ktoré sme zadefinovali, môžeme výslednú sumu písať od indexu 0. Rovnaký argument platí aj v druhom sčítanci, kde pre index $s = i + 1$ dostávame v sume $p_{i+1}(i,n) = 0$, teda nulový člen, a preto platí

$$\frac{u}{n}\sum_{s=0}^{i+1} s p_s(i,n)u^{s-1} = \frac{u}{n}\frac{\partial f_{i,n}(u)}{\partial u}.$$

- Uvažujeme $n \leq i + 1$. V tomto prípade dostaneme

$$\begin{aligned}
f_{i+1,n}(u) &= \sum_{s=0}^{m} p_s(i+1,n)u^s \\
&= \sum_{s=0}^{m} p_{s-1}(i,n)u^s + \frac{1}{n}\sum_{s=0}^{m} p_s(i,n)su^s - \frac{1}{n}\sum_{s=0}^{m} p_{s-1}(i,n)(s-1)u^s \\
&= \sum_{s=0}^{n-1} p_s(i,n)u^{s+1} + \frac{1}{n}\sum_{s=0}^{n} p_s(i,n)su^s - \frac{1}{n}\sum_{s=0}^{n-1} p_s(i,n)su^{s+1} \qquad (2.3) \\
&= \sum_{s=0}^{n} p_s(i,n)u^{s+1} + \frac{1}{n}\sum_{s=0}^{n} p_s(i,n)su^s - \frac{1}{n}\sum_{s=0}^{n} p_s(i,n)su^{s+1} \\
&= uf_{i,n}(u) + \frac{u}{n}\frac{\partial f_{i,n}(u)}{\partial u} - \frac{u^2}{n}\frac{\partial f_{i,n}(u)}{\partial u}.
\end{aligned}$$

Všimneme si, že prvú a tretiu sumu vo výraze 2.3 môžeme písať ako sumu od 0 po n, pretože ak položíme $s = n$, dostávame pre obe sumy rovnaký výraz $p_s(i,n)u^{s+1}$, ktorý sa vďaka opačným znamienkam vykráti.

Teda sme odvodili, že pravdepodobnostná vytvárajúca funkcia $f_{i+1,n}(u)$ má nasledujúci rekurentný tvar

$$f_{i+1,n}(u) = uf_{i,n}(u) + \frac{u(1-u)}{n}\frac{\partial f_{i,n}(u)}{\partial u}, \qquad (2.4)$$

kde $f_{0,n}(u) = 1$.

Označme $a_r(i+1)$, pre $r = 1, \ldots, i+1$ koeficienty pravdepodobnostnej vytvárajúcej funkcie, kde $i+1$ je počet ihiel a r odpovedá exponentu pri argumente u^r. Potom sme podľa článku Gani (2002) schopní vyjadriť koeficienty pravdepodobnostnej vytvárajúcej funkcie napísanej v tvare

$$\begin{aligned}
f_{i+1,n}(u) = u^{i+1}\left(1-\frac{1}{n}\right)\ldots\left(1-\frac{i}{n}\right) + \cdots + \frac{u^r}{n^{i-r+1}}a_r(i+1)\left(1-\frac{1}{n}\right)\ldots \\
\times\left(1-\frac{r-1}{n}\right) + \cdots + \frac{u}{n^i}. \qquad (2.5)
\end{aligned}$$

Správnosť tohto explicitného výrazu overíme tak, že rozpíšeme výraz 2.4 a porovnáme koeficienty u rovnakých mocnín premennej u

$$
\begin{aligned}
f_{i+1,n}(u) &= \frac{u}{n}\frac{\partial f_{i,n}(u)}{\partial u} - \frac{u^2}{n}\frac{\partial f_{i,n}(u)}{\partial u} + uf_{i,n}(u) \\
&= \frac{u^i}{n} i a_i(i)\left(1-\frac{1}{n}\right)\dots\left(1-\frac{i-1}{n}\right)+\dots \\
&\quad + u^r r a_r(i)\left(1-\frac{1}{n}\right)\dots\left(1-\frac{r-1}{n}\right)/n^{i-r+1}+\dots+ua_1(i)/n^i \\
&\quad - \frac{u^{i+1}}{n} i a_i(i)\left(1-\frac{1}{n}\right)\dots\left(1-\frac{i-1}{n}\right)-\dots \\
&\quad - \frac{u^r}{n}(r-1)a_{r-1}(i)\left(1-\frac{1}{n}\right)\dots\left(1-\frac{r-2}{n}\right)/n^{i-r+1}-\dots \\
&\quad - u^2 a_1(i)/n^i \\
&\quad + u^{i+1} a_i(i)\left(1-\frac{1}{n}\right)\dots\left(1-\frac{i-1}{n}\right)+\dots \\
&\quad + u^r a_{r-1}(i)\left(1-\frac{1}{n}\right)\dots\left(1-\frac{r-2}{n}\right)/n^{i-r+1}+\dots+u^2 a_1(i)/n^{i-1}.
\end{aligned}
$$

Pri porovnávaní využijeme nasledujúce vzťahy, ktoré pre koeficienty platia

$$
\begin{aligned}
a_{i+1}(i+1) = a_1(i+1) &= 1, \\
a_r(i+1) &= r a_r(i) + a_{r-1}(i).
\end{aligned} \tag{2.6}
$$

Bez akýchkoľvek úprav vidíme, že koeficient u premennej u^1 je $1/n^i$ a je rovnaký s koeficientom v rovnici 2.5.

Jednoduché je overenie aj pre koeficient u premennej u^{i+1}. Aby sme dostali požadovaný tvar, potrebujeme odčítať a následne využiť vhodné úpravy týchto dvoch výrazov

$$
\begin{aligned}
a_i(i)\left(1-\frac{1}{n}\right)\dots\left(1-\frac{i-1}{n}\right) - \frac{i}{n}a_i(i)\left(1-\frac{1}{n}\right)\dots\left(1-\frac{i-1}{n}\right) = \\
= \left[\left(1-\frac{1}{n}\right)\dots\left(1-\frac{i-1}{n}\right)\right]\left(a_i(i)-\frac{i}{n}a_i(i)\right) = \\
= \left(1-\frac{1}{n}\right)\dots\left(1-\frac{i-1}{n}\right)\left(1-\frac{i}{n}\right).
\end{aligned}
$$

Nakoniec overíme koeficient u premennej u^r. Začneme tým, že vypíšeme všetky výrazy, ktoré sa pri tomto koeficiente vyskytujú

$$
\begin{aligned}
r\frac{a_r(i)}{n^{i-r+1}}\left(1-\frac{1}{n}\right)\dots\left(1-\frac{r-1}{n}\right) - \frac{r-1}{n}\frac{a_{r-1}(i)}{n^{i-r+1}}\left(1-\frac{1}{n}\right)\dots \\
\times\left(1-\frac{r-2}{n}\right) + \frac{a_{r-1}(i)}{n^{i-r+1}}\left(1-\frac{1}{n}\right)\dots\left(1-\frac{r-2}{n}\right).
\end{aligned}
$$

V ďalšom kroku si výraz $a_r(i)$ upravíme pomocou vzťahu 2.6 a zároveň roznásobíme jednolivé výrazy medzi sebou

$$\frac{a_r(i+1)}{n^{i-r+1}}\left(1-\frac{1}{n}\right)\dots\left(1-\frac{r-1}{n}\right)-\frac{a_{r-1}(i)}{n^{i-r+1}}\left(1-\frac{1}{n}\right)\dots\left(1-\frac{r-1}{n}\right)-$$
$$-\frac{r-1}{n}\frac{a_{r-1}(i)}{n^{i-r+1}}\left(1-\frac{1}{n}\right)\dots\left(1-\frac{r-2}{n}\right)+\frac{a_{r-1}(i)}{n^{i-r+1}}\left(1-\frac{1}{n}\right)\dots\left(1-\frac{r-2}{n}\right).$$

Vidíme, že prvý sčítanec je koeficient, ktorý chceme dostať. Ostáva nám teda ukázať, že po vhodnom vybratí pred zátvorku sa ostatné členy navzájom vyrušia. Prvý sčítanec opíšeme a po úpravach v ostatných členoch sme dostali v poslednom činiteli rovnaké výrazy s opačnými znamienkami, teda celý druhý sčítanec je rovný nule

$$\frac{a_r(i+1)}{n^{i-r+1}}\left(1-\frac{1}{n}\right)\dots\left(1-\frac{r-1}{n}\right)+\frac{a_{r-1}(i)}{n^{i-r+1}}\left(1-\frac{1}{n}\right)\dots\left(1-\frac{r-2}{n}\right)\times$$
$$\times\left(-1+\frac{r-1}{n}-\frac{r-1}{n}+1\right).$$

Ukázali sme, že koeficienty u jednotlivých premenných sú rovnaké. A navyše všetky koeficienty $a_r(i+1)$, $r=1,\dots,i+1$ môžeme priamo získať rekurentne z koeficientov $a_r(i)$, $r=1,\dots,i$ pomocou maticového zápisu

$$\begin{pmatrix} a_{i+1}(i+1) \\ a_i(i+1) \\ \vdots \\ a_2(i+1) \\ a_1(i+1) \end{pmatrix} = \begin{pmatrix} 1 & & & & & \\ i & 1 & & & & \\ & i-1 & 1 & & & \\ & & \vdots & & & \\ & & & 2 & 1 & 0 \\ & & & & 1 & 0 \end{pmatrix} \begin{pmatrix} a_i(i) \\ a_{i-1}(i) \\ \vdots \\ a_1(i) \\ 0 \end{pmatrix} = B_i A(i).$$

Všetky $A(j)$, $j=1,\dots,i+1$, sú $(i+1)$ stĺpcové vektory začínajúce koeficientom $a_j(j)$, ktorý sa postupne znižuje až ku koeficientu $a_1(j)$. Po tejto hodnote nasleduje $i+j-1$ núl. Matica B_j má v prvom riadku iba jednu hodnotu 1, ktorá je umiestnená hneď na pozícii $(1,1)$. V druhom riadku má na diagonálnej pozícii hodnotu 1 a na subdiagonálnej pozícii hodnotu j. Týmto spôsobom pokračujeme až do j-tého riadku, ktorý má na diagonálnej pozícii hodnotu 1 a na subdiagonálnej pozícii hodnotu 2. Posledný nenulový riadok je $j+1$, ktorý má iba jednu nenulovú hodnotu 1 na subdiagonálnej pozícii. Dimenzia týchto matíc je $(i+1)\times(i+1)$. Vďaka týmto známym tvarom matíc môžeme koeficienty pravdepodobnostnej vytvárajúcej funkcie v prípade, že máme v obehu $i+1$ ihiel, vyčísliť nasledujúcim násobením

$$\begin{pmatrix} a_{i+1}(i+1) \\ a_i(i+1) \\ \vdots \\ a_2(i+1) \\ a_1(i+1) \end{pmatrix} = B_i B_{i-1}\dots B_1 \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}.$$

Na konkrétnom príklade si kvôli lepšej vizuálnej predstave odvodíme, ako budú

a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}
1	511	9330	34105	42525	22827	5880	750	45	1

Tabuľka 2.1: Hodnoty koeficientov pravdepodobnostnej vytvárajúcej funkcie v homogénnom prípade pre 10 ihiel

vyzerať koeficienty získané týmto maticovým násobením, ak máme v obehu 10 ihiel, teda $i = 10$. Koeficienty uvedieme formou tabuľky 2.1.

Získané koeficienty dosadíme do pravdepodobnostnej vytvárajúcej funkcie 2.5, ktora má pre všeobecné n tvar

$$f_{10,n}(u) = u^{10}\left(1-\frac{1}{n}\right)\dots\left(1-\frac{9}{n}\right)+\dots+\frac{a_r u^r}{n^{10-r}}\left(1-\frac{1}{n}\right)\dots\left(1-\frac{r-1}{n}\right)+\dots + \frac{a_2u^2}{n^8}\left(1-\frac{1}{n}\right)+\frac{u}{n^9}.$$

Z takto zapísanej pravdepodobnostnej funkcie už vieme určiť pravdepodobnosti pre jednotlivé počty novo infikovaných ľudí. Tieto počty sú dané mocniteľom premennej u. Konkrétne pre 100 zdravých ľudí, ktorých na začiatku uvažujeme, dostávame

$$\begin{aligned} f_{10,100}(u) &= 0{,}6u^{10} + 0{,}3u^9 + 5{,}6\cdot 10^{-2}u^8 + 4{,}7\cdot 10^{-3}u^7 \\ &\quad +2\cdot 10^{-4}u^6 + 3{,}8\cdot 10^{-6}u^5 + 3{,}2\cdot 10^{-8}u^4 \\ &\quad +9{,}1\cdot 10^{-11}u^3 + 5{,}1\cdot 10^{-14}u^2 + 1\cdot 10^{-18}u. \end{aligned} \tag{2.7}$$

Je zrejmé, že pre $n < 10$ sa všetky nadbytočné členy v rovnici vykrátia. Konkrétne pre $n = 5$ dostávame

$$f_{10,5}(u) = 5{,}1\cdot 10^{-7}u + 1\cdot 10^{-3}u^2 + 5{,}7\cdot 10^{-2}u^3 + 0{,}4u^4 + 0{,}5u^5,$$

čo odpovedá našej predstave, že pri počte ihiel 5, sú pravdepodobnosti toho, že máme počet novo infikovaných ľudí väčší ako 5, nulové.

2.3 Stredná hodnota a rozptyl počtu novo infikovaných homogénnych osôb

V nasledujúcej časti si ukážeme odvodenie očakávaného počtu novo infikovaných ľudí a jeho rozptyl pomocou úvahy využívajúcej prístup klasickej kolmogorovskej pravdepodobnosti. Je namieste poznamenať, že v Gani (2002) nájdeme odvodenie prvých dvoch momentov novo infikovaných osôb generovaných šírením i ihiel medzi n zdravými jedincami pomocou zostavenia rekurentných rovníc odvodených z prvej a druhej derivácie pravdepodobnostnej vytvárajúcej funkcie 2.4.

Označme $Y_{i,n}$ náhodnú veličinu, ktorá reprezentuje počet nakazených ľudí medzi i infikovanými ihlami a celkovým počtom ľudí n. V ďalšej časti určíme strednú hodnotu a rozptyl tejto náhodnej veličiny.

Budeme postupovať tak, že zavedieme novú náhodnú veličinu $\tilde{Z}_j$, ktorá má alternatívne rozdelenie. Úspech nastáva, ak infikovaná ihla nepríde do styku s j-tým človekom, a neúspech, ak príde. Teda v reči priehradok a guľôčok by úspech znamenal, že j-tá priehradka nie je obsadená, a neúspech, ak by obsadená bola. Parameter tohto rozdelenia označíme p a vyjadruje pravdepodobnosť, že j-tého človeka neinfikuje ani jedna z i ihiel. Tento parameter odvodíme pomocou klasickej definície pravdepodobnosti. Je zrejmé, že počet všetkých možností je n^i. Uvažujme, aký je počet priaznivých možností, ktoré udávaju to, že j-tý človek nepríde do styku so žiadnou z i ihiel. Uvedomíme si, že každá infikovaná ihla môže prísť do styku so všetkými ľuďmi okrem j-tého. Teda celkom máme $(n-1)^i$ možností. A konečne môžeme napísať hľadanú pravdepodobnosť

$$p = \left(\frac{n-1}{n}\right)^i.$$

Teraz zavedieme novú náhodnú veličinu Z_j vzťahom

$$Z_j = 1 - \tilde{Z}_j, \quad j = 1, \ldots, n.$$

Táto náhodná veličina má taktiež alternatívne rozdelenie s parametrom $1 - p$. Nasledujúcim vzťahom popíšeme kedy nastane úspech (ozn. 1) a kedy neúspech (ozn. 0)

$$Z_j = \begin{cases} 1 & j\text{-tý človek prišiel do styku s ihlou,} \\ 0 & \text{inak.} \end{cases}$$

Stredná hodnota tejto náhodnej veličiny je

$$\mathsf{E}\,[Z_j] = \mathsf{P}(Z_j = 1) = 1 - \mathsf{P}(Z_j = 0) = 1 - \left(\frac{n-1}{n}\right)^i,$$

rozptyl je rovný

$$\mathrm{var}[Z_j] = \left(\frac{n-1}{n}\right)^i \left[1 - \left(\frac{n-1}{n}\right)^i\right].$$

Po tomto zadefinovaní je jednoduché uvedomiť si, že $Y_{i,n}$ vznikne ako súčet alternatívnych veličín, čo môžeme zapísať ako $Y_{i,n} = \sum_{j=1}^{n} Z_j$. Využijeme, že Z_j, $j = 1, \ldots, n$ sú rovnako rozdelené náhodne veličiny, a napíšeme vzťah, ktorý platí pre strednú hodnotu

$$\mathsf{E}\,[Y_{i,n}] = \sum_{j=1}^{n} \mathsf{E}\,[Z_j] = n\,\mathsf{E}\,[Z_1] = n\left[1 - \left(\frac{n-1}{n}\right)^i\right]. \tag{2.8}$$

Pre rozptyl platí vzťah

$$\begin{aligned} \mathrm{var}[Y_{i,n}] &= \sum_{j=1}^{n} \mathrm{var}[Z_j] + \sum_{k \neq j} \mathrm{cov}[Z_k, Z_j] \\ &= n\,\mathrm{var}[Z_1] + n(n-1)\,\mathrm{cov}[Z_1, Z_2]. \end{aligned} \tag{2.9}$$

Potrebujeme si vyjadriť kovariancie

$$\begin{aligned}\operatorname{cov}[Z_1, Z_2] &= \operatorname{cov}[1-\tilde{Z}_1, 1-\tilde{Z}_2] \\ &= \operatorname{cov}[1,1] - \operatorname{cov}[\tilde{Z}_1, 1] - \operatorname{cov}[1, \tilde{Z}_2] + \operatorname{cov}[\tilde{Z}_1, \tilde{Z}_2] \\ &= \operatorname{cov}[\tilde{Z}_1, \tilde{Z}_2].\end{aligned}$$

Ostáva nám ukázať vyjadrenie tejto kovariancie

$$\begin{aligned}\operatorname{cov}[\tilde{Z}_1, \tilde{Z}_2] &= \mathsf{E}[\tilde{Z}_1\tilde{Z}_2] - \mathsf{E}[\tilde{Z}_1]\,\mathsf{E}[\tilde{Z}_2] = P(\tilde{Z}_1 = 1, \tilde{Z}_2 = 1) - \mathsf{E}[\tilde{Z}_1]\,\mathsf{E}[\tilde{Z}_2] \\ &= \left(\frac{n-2}{n}\right)^i - \left(\frac{n-1}{n}\right)^{2i},\end{aligned}$$

kde prvý člen chápeme tak, že pri počte ihiel i vyberáme 2 jedincov, ktorí neprídu do styku so žiadnou infikovanou ihlou počas ich používania. Dosadenie tohto vzťahu do 2.9 nám konečne umožní napísať vzťah pre rozptyl a upraviť ho do jednoduchšieho tvaru

$$\begin{aligned}\operatorname{var}[Y_{i,n}] &= n\operatorname{var}[Z_1] + n(n-1)\operatorname{cov}[\tilde{Z}_1,\tilde{Z}_2] \\ &= n\left(\frac{n-1}{n}\right)^i\left[1-\left(\frac{n-1}{n}\right)^i\right] + n(n-1)\left[\left(\frac{n-2}{n}\right)^i - \left(\frac{n-1}{n}\right)^{2i}\right] \\ &= n\left(1-\frac{1}{n}\right)^i + n(n-1)\left(1-\frac{2}{n}\right)^i - n^2\left(1-\frac{1}{n}\right)^{2i}. \qquad (2.10)\end{aligned}$$

Na záver uvedieme asymptotické správanie strednej hodnoty a rozptylu. Ak v rovnici 2.8 uvažujeme zafixovaný celkový počet osôb n a veľmi vysoký počet infikovaných ihiel i, očakávame, že sa postupne nakazia všetky osoby v danej populácii. V rovnici 2.10 uvažujeme tie isté podmienky, a teda dostávame

$$\begin{aligned}\lim_{i\to\infty}\mathsf{E}[Y_{i,n}] &= \lim_{i\to\infty} n\left[1-\left(\frac{n-1}{n}\right)^i\right] = n, \\ \lim_{i\to\infty}\operatorname{var}[Y_{i,n}] &= n + n(n-1) - n^2 = 0,\end{aligned}$$

zatiaľ čo pre $i = \lambda n$, kde n je vysoké a λ je kladná fixovaná konštanta, dostávame s využitím vzťahu $\lim_{n\to\infty}(1+\frac{1}{n})^n = e$ nasledujúce asymptotické vyjadrenie strednej hodnoty

$$\lim_{n\to\infty}\mathsf{E}[Y_{i,n}]/n = 1 - e^{-\lambda},$$

k odvodeniu asymptotického rozptylu musíme navyše použiť L'Hospitalovo pravidlo a následne Maclaurinov rozvoj funkcie $\log(1-x)$ a rozvoj funkcie do geometrickej rady. Potom dostávame vzťah

$$\lim_{n\to\infty}\operatorname{var}[Y_{i,n}]/n = e^{-2\lambda}(e^{\lambda} - 1 - \lambda),$$

ktorý sa odlišuje od chybne uvedeného limitného rozptylu v článku Gani (2002).

2.4 Aplikácia modelu

Kvôli lepšej predstave a názornosti využitia tohto modelu si spočítame očakávané počty novo nakazených ľudí najprv zo vzorca 2.8, a potom generovaním náhodnej veličiny $Y_{i,n}$, keď na začiatku uvažujeme skupinu 100 osôb, ktorí medzi sebou náhodne zdieľajú rôzny počet infikovaných ihiel. Tieto hodnoty uvádzame formou tabuľky 2.2. Schválne sme ako prvú hodnotu uviedli 10 ihiel. Vidíme, že pre túto hodnotu je očakávaný počet zaokrúhlene 10 osôb, čo sa zhoduje s pravdepodobnosťou vyčíslenou v pravdepodobnostnej vytvárajúcej funkcii 2.7 , ktorá bola najvyššia práve pri koeficiente u^{10}, ktorého mocniteľ značí počet novo infikovaných jedincov. Taktiež sme ukázali správnosť asymptotického odhadu, kedže už pri používaní 675 ihiel očakávame, že približne celá populácia, ktorá ihly používa, bude nakazená.

V 3. stĺpci tabuľky odhadneme očakávané hodnoty pomocou simulácie, na základe ktorej budeme chcieť ukázať, že generované hodnoty sú blízko teoretickým, ktoré sú zaznamenané v 2. stĺpci tabuľky 2.2. V algoritme budeme postupne generovať 1000 realizácii náhodnej veličiny $Y_{i,n}$ a to nasledujúcim spôsobom: postupne vygenerujeme i náhodných čísel z rovnomerného rozdelenia na hodnotách $1, \ldots, n$ pričom budeme prihliadať na zmenu pravdepodobností. Model začína s nulovým počtom nakazených. Pre každú ihlu potom náhodne vyberáme človeka. Predstavme si situáciu, že máme 90 zdravých a 10 nakazených jedincov. Správnu pravdepodobnosť ošetríme tak, že ak vyberieme číslo medzi $11 - 100$, zvýšime počet nakazených, ktorý máme zaznamenaný do doterajšieho kroku. Ak vygenerujeme nižšie číslo, hodnota generovanej náhodnej veličiny sa meniť nebude. Takto postupujeme pre všetkých i ihiel, ktoré máme k dispozícii. Aby sme odhadli strednú hodnotu tejto náhodnej veličiny pomocou simulácií, musíme túto procedúru opakovať niekoľkokrát a následne spočítať priemer z vygenerovaných hodnôt. Dostávame výsledky, ktoré sa s niektorými hodnotami počítanými zo vzorca líšia až na 2. desatinnom mieste. Teda skutočne simulujeme študovaný urnový model s náhodnym vracaním. Celý kód simulácie je uvedený v Prílohe A na konci tejto práce. K výpočtu simulácie bol použitý program Mathematica 10.0 od Wolfram Research (2014).

Rovnakú úvahu použijeme aj pri porovnávaní smerodatných odchýliek. Teoretické hodnoty vypočítame odmocením vzťahu 2.10. A následne v 3. stĺpci tabuľky 2.3 uvedieme smerodatné odchýlky simulované pomocou vyššie popísaného algoritmu s tým rozdielom, že po niekoľko násobnom generovaní hodnôt počítame výberovú smerodatnú odchýlku. Znova môžeme zhrnúť, že simulované hodnoty sú dostatočne zhodné s tými teoretickými.

Počet ihiel i	Očakávaný počet novo infikovaných osôb	
	vzorec	simulácia
10	9,56	9,54
25	22,22	22,23
50	39,50	39,63
75	52,94	52,90
99	63,03	63,00
100	63,40	63,43
300	95,10	95,20
675	99,89	99,87

Tabuľka 2.2: Očakávané počty novo infikovaných pre vzorku 100 osôb pri zdieľaní rôzneho počtu ihiel získané pomocou vzorca a simulácie

Počet ihiel i	Smerodatné odchýlky novo infikovaných osôb	
	vzorec	simulácia
10	0,62	0,65
25	1,42	1,41
50	2,34	2,36
75	2,86	2,70
99	3,11	3,05
100	3,12	3,11
300	1,99	2,04
675	0,33	0,35

Tabuľka 2.3: Smerodatné odchýlky novo infikovaných pre vzorku 100 osôb pri zdieľaní rôzneho počtu ihiel získané pomocou vzorca a simulácie

Kapitola 3

Nehomogénny model

3.1 Popis modelu

V tejto kapitole budeme modifikovať model, ktorý sme odvodili vyššie. Naďalej budeme uvažovať i infikovaných ihiel, no n zdravých jedincov, ktorí budú ihly medzi sebou používať, rozdelíme do k skupín. Ak by sme na tento model nahliadali pomocou urnovej formulácie, dostali by sme urnu obsahujúcu k farieb, teda je dobré uvedomiť si, že dostávame mnohorozmerný model. Preň znova odvodíme, ako vyzerá pravdepodobnostná vytvárajúca funkcia, očakávané hodnoty novo nakazených osôb a rozptyl. V závere kapitoly si ukážeme numerické hodnoty týchto očakávaných hodnôt, no kvôli zložitosti algebry budeme pre populáciu obsahujúcu n osôb uvažovať iba 2 skupiny. Následne budú tieto skupiny konkretizované a pozrieme sa na niektoré ich vlastnosti.

Tento model bol podrobne študovaný v Gani (2002). Zavedieme si značenie, ktoré budeme v tejto časti používať. Symbolmi p_1 až p_k postupne označíme pravdepodobnosti, že ľubovoľná infikovaná ihla môže padnúť medzi n_1 osôb prvej skupiny až n_k osôb k-tej skupiny, pričom platí $\sum_{j=1}^{k} p_j = 1$. Ďalej uvažujeme multinomický koeficient reprezentujúci počet možností, ako môžeme z i ihiel určiť jednotlivé počty pre každú skupinu výrazom $\binom{i}{i_1,\dots,i_k}$, ktorý je možné prepísať pomocou faktoriálov vzťahom $\frac{i!}{i_1!\dots i_k!}$. Po vynásobení dostávame vzťah

$$\binom{i}{i_1,\dots,i_k} p_1^{i_1} \dots\ p_k^{i_k},$$

ktorý interpretujeme tak, že náhodný počet ihiel i_1 bude používaný užívateľmi v 1. skupine, až náhodný počet ihiel i_k bude používaný užívateľmi v k-tej skupine. Pravdepodobnostnú vytvárajúcu funkciu popisujúcu rozdelenie počtu i_j ihiel medzi n_j osobami v j-tej skupine označíme výrazom $f_{i_j,n_j}(u_j)$.

Uvedomíme si, že tento výraz má rovnakú interpretáciu ako rovnica 2.5, ktorú sme odvodili v 2. kapitole, a preto ak využijeme substitúciu $i+1$, n, u na i_j+1, n_j,u_j, kde $j = 1,\dots,k$ môžeme tento tvar ďalej využívať. Je dobré uvedomiť si, že znova máme k dispozícii vopred daný počet ihiel i, ktorý delíme medzi jednu skupinu osôb, teda môžeme pracovať so zavedenými rovnicami z 2. kapitoly a napísať finálny tvar združenej obecnej pravdepodobnostnej funkcie pre prípad nehomogénnych osôb

$$f_{i,n_1,\dots,n_k}(u_1,\dots,u_k) = \sum_{i_1,\dots,i_k} \binom{i}{i_1,\dots,i_k} p_1^{i_1}\dots\ p_k^{i_k} f_{i_1,n_1}(u_1)\dots\ f_{i_k,n_k}(u_k), \quad (3.1)$$

kde $0 < u_j \leq 1$, $j = 1,\dots,k$. Ešte upozorníme, že v nasledujúcej časti textu nebudeme kvôli rozsiahlosti rovníc písať argument výrazu $f_{i,n_1,\dots,n_k}(u_1,\dots,u_k)$.

V predošlej kapitole sme ukázali, že táto funkcia rieši diferenčnú-diferenciálnu rovnicu 2.4, ktorú uvedieme v značení pre nehomogénnych užívateľov

$$f_{i_j+1,n_j}(u_j) = u_j f_{i_j,n_j}(u_j) + \frac{u_j(1-u_j)}{n_j}\frac{\partial f_{i_j,n_j}(u_j)}{\partial u_j}, \quad (3.2)$$

kde $j = 1,\dots,k$.

Uviedli sme rozdelenie generujúce $i_j + 1$ novo nakazených osôb medzi j-tou skupinou. Ak navyše uvažujeme, že 3.2 nastane, ak budeme v práve j-tej skupine, čo sa stane s pravdepodonosťou p_j, potom sme schopní vysčítaním cez všetkých k-skupín popísať rozdelenie novo generovaných infikovaných osôb medzi všetkými skupinami

$$f_{i+1,n_1,\dots,n_k}(u_1,\dots,u_k) = \sum_{j=1}^{k}\left[p_j u_j f_{i,n_1,\dots,n_k} + p_j\frac{u_j(1-u_j)}{n_j}\frac{\partial f_{i,n_1,\dots,n_k}}{\partial u_j}\right]. \quad (3.3)$$

Aby sme boli tento vzťah schopní dokázať, zavedieme si podmienenú pravdepodobnosť

$$\begin{aligned} p_{s_1,\dots,s_k}(i,n_1,\dots,n_k) = \mathsf{P}(&s_1,\dots,s_k \text{ generovaných infekcií}| \\ &i \text{ ihiel}, n_1,\dots,n_k \text{ zdravých}), \end{aligned}$$

s pravdepodobnostnou vytvárajúcou funkciou

$$f_{i,n_1,\dots,n_k}(u_1,\dots,u_k) = \sum_{s_1,\dots,s_k} p_{s_1,\dots,s_k}(i,n_1,\dots,n_k)u_1^{s_1}\dots\ u_k^{s_k}. \quad (3.4)$$

Teraz uvažujme zvýšenie počtu infikovaných ihiel o jednotku a dostávame združenú pravdepodobnosť, ktorú interpretujeme rovnakým spôsobom ako marginálnu pravdepodobnosť

$$\begin{aligned} p_{s_1,\dots,s_k}(i+1,n_1,\dots,n_k) = {} & p_{s_1,\dots,s_k}(i,n_1,\dots,n_k)\left(p_1\frac{s_1}{n_1} + \dots + p_k\frac{s_k}{n_k}\right) \\ & + p_1 p_{s_1-1,\dots,s_k}(i,n_1,\dots,n_k)\left(1 - \frac{s_1-1}{n_1}\right) \\ & + \dots + p_k p_{s_1,\dots,s_k-1}(i,n_1,\dots,n_k)\left(1 - \frac{s_k-1}{n_k}\right), \quad (3.5) \end{aligned}$$

s okrajovou podmienkou $p_{0,\dots,0}(0,n_1,\dots,n_k) = 1$. Ukážeme si odvodenie pravdepodobnostnej vytvárajúcej funkcie 3.3. Vynásobením výrazu 3.5 premennými $u_1^{s_1},\dots,u_k^{s_k}$ a vysčítaním cez všetky možné hodnoty počtu novo nakazených jedincov $\sum_{j=1}^{k} s_j \leq \min(i, \sum_{j=1}^{k} n_j)$ odvodíme pravdepodobnostnú vytvárajúcu funkciu v tvare ako 3.4 pre $i+1$ ihiel

$$
\begin{aligned}
&f_{i+1,n_1,\dots,n_k}(u_1,\dots,u_k) = \\
&\qquad = \sum_{s_1,\dots,s_k} p_{s_1,\dots,s_k}(i+1,n_1,\dots,n_k)u_1^{s_1}\dots u_k^{s_k} \\
&\qquad = \sum_{s_1,\dots,s_k} \Bigg[p_{s_1,\dots,s_k}(i,n_1,\dots,n_k) \sum_{j=1}^{k} p_j \frac{s_j}{n_j} u_1^{s_1}\dots u_k^{s_k} \\
&\qquad\quad + \sum_{j=1}^{k} p_j p_{s_1,\dots,s_j-1,\dots,s_k}(i,n_1,\dots,n_k)\left(1-\frac{s_j-1}{n_j}\right)u_1^{s_1}\dots u_k^{s_k}\Bigg] \qquad (3.6) \\
&\qquad = \sum_{j=1}^{k} p_j \frac{u_j}{n_j} \sum_{s_1,\dots,s_k} p_{s_1,\dots,s_k}(i,n_1,\dots,n_k) s_j u_1^{s_1}\dots u_j^{s_j-1}\dots u_k^{s_k} \\
&\qquad\quad - \sum_{j=1}^{k} u_j p_j \sum_{s_1,\dots,s_k} p_{s_1,\dots,s_k}(i,n_1,\dots,n_k)\frac{s_j}{n_j} u_1^{s_1}\dots u_k^{s_k} \\
&\qquad\quad + \sum_{j=1}^{k} u_j p_j \sum_{s_1,\dots,s_k} p_{s_1,\dots,s_k}(i,n_1,\dots,n_k) u_1^{s_1}\dots u_k^{s_k} \\
&\qquad = \sum_{j=1}^{k} \left[p_j \frac{u_j}{n_j}\frac{\partial f_{i,n_1,\dots,n_k}}{\partial u_j} - p_j \frac{u_j^2}{n_j}\frac{\partial f_{i,n_1,\dots,n_k}}{\partial u_j} + p_j u_j f_{i,n_1,\dots,n_k}\right] \\
&\qquad = \sum_{j=1}^{k} \left[p_j \frac{u_j(1-u_j)}{n_j}\frac{\partial f_{i,n_1,\dots,n_k}}{\partial u_j} + p_j u_j f_{i,n_1,\dots,n_k}\right]. \qquad (3.7)
\end{aligned}
$$

Vidíme, že sme dostali presne rovnaký výraz ako 3.3. Kvôli prehľadnosti slovne okomentujeme operácie, ktoré sme používali. Druhú rovnosť sme dostali z prvej dosadením rovnice 3.5. Prvý výraz v tretej rovnosti sme dostali iba preskupením členov. Ďalší sčítanec a rozdiel roznásobením výrazu 3.6, súčasne so zmenou indexovacej premennej v sume s braním ohľadu na okrajové podmienky. Odtiaľ už bol zrejmý prepis pomocou parciálnych derivácii výrazu 3.4. Stojí za zmienku, že pre každé s_j je horný index súm, ktoré sčítavame, $\min(i, n_j)$, avšak kvôli platnosti okrajových podmienok ide iba o triviálne rozpísanie týchto súm podobné, ako sme ukázali v 2. kapitole, a preto tu tento formálny krok neuvádzame.

Ďalšia časť bude zasa zameraná viac prakticky. Bude ukázané ako pracovať s pravdepodobnostnou vytvárajúcou funkciou pre populáciu obsahujúcu 100 nehomogénnych osôb rozdelenú na 2 skupiny. Prvá obsahuje 60 a druhá 40 ľudí. Budú medzi sebou zdieľať 10 ihiel. Z rovnice 3.1 vyplýva, že dostávame funkciu

$$
f_{10,60,40}(u_1,u_2) = \sum_{i_1=0}^{10} \frac{10!}{i_1!(10-i_1)!} p^{i_1}(1-p)^{10-i_1} f_{i_1,60}(u_1) f_{10-i_1,40}(u_2). \qquad (3.8)
$$

Vďaka tomu, že budeme chcieť určiť pravdepodobnosti počtu novo nakazených osôb v oboch skupinách môžeme písať rovnicu 3.8 ako funkciu jednej premennej u a tým pádom $f_{i_1,60}(u)$ a $f_{10-i_1,40}(u)$ spočítať z 2.5. Z takto definovanej pravdepodobnostnej vytvárajúcej funkcie už vieme pomocou vhodného softvéru spočítať pravdepodobnosti počtu novo infikovaných osôb, ak uvažujeme rôzne hodnoty p, ktorá označuje pravdepodobnosť toho, že sa dostaneme do prvej skupiny. Jednotlivé pravdepodobnosti uvedieme v tabuľke 3.1.

Počet infikovaných	Pravdepodobnosti				
	0,25	0,50	0,60	0,75	1,00
1	$2{,}2 \cdot 10^{-16}$	$3{,}8 \cdot 10^{-18}$	$1{,}0 \cdot 10^{-18}$	$5{,}6 \cdot 10^{-18}$	$9{,}9 \cdot 10^{-17}$
2	$4{,}4 \cdot 10^{-12}$	$1{,}1 \cdot 10^{-13}$	$5{,}1 \cdot 10^{-14}$	$1{,}8 \cdot 10^{-13}$	$3{,}0 \cdot 10^{-12}$
3	$3{,}5 \cdot 10^{-9}$	$1{,}6 \cdot 10^{-10}$	$9{,}1 \cdot 10^{-11}$	$2{,}4 \cdot 10^{-10}$	$3{,}2 \cdot 10^{-9}$
4	$6{,}0 \cdot 10^{-7}$	$4{,}8 \cdot 10^{-8}$	$3{,}2 \cdot 10^{-8}$	$6{,}7 \cdot 10^{-8}$	$6{,}6 \cdot 10^{-7}$
5	$3{,}8 \cdot 10^{-5}$	$5{,}1 \cdot 10^{-6}$	$3{,}8 \cdot 10^{-6}$	$6{,}7 \cdot 10^{-6}$	$4{,}6 \cdot 10^{-5}$
6	$1{,}1 \cdot 10^{-3}$	$2{,}4 \cdot 10^{-4}$	$2{,}0 \cdot 10^{-4}$	$2{,}9 \cdot 10^{-4}$	$1{,}4 \cdot 10^{-3}$
7	0,02	$5{,}4 \cdot 10^{-3}$	$4{,}7 \cdot 10^{-3}$	$6{,}2 \cdot 10^{-3}$	0,02
8	0,11	0,06	$5{,}6 \cdot 10^{-2}$	$6{,}6 \cdot 10^{-2}$	0,13
9	0,38	0,32	0,31	0,33	0,40
10	0,49	0,62	0,63	0,60	0,45

Tabuľka 3.1: Pravdepodobnosti novo infikovaných nehomogénnych osôb pre rôzne hodnoty pravdepodobností vstupu do 1. skupiny

Vzhľadom k tomu, že v praxi chceme držať počet novo infikovaných osôb na minime, z tabuľky je zrejmé, že najvýhodnejšia situácia nastáva pre $p = 0{,}75$. Vtedy je totiž iba 45% pravdepodobnosť, že každá ihla, ktorá je k dispozícii, niekoho infikuje. Môžeme si skontrolovať, že pravdepodobnostná vytvárajúca funkcia bola odvodená správne. Ak totiž uvažujeme $p = 0{,}6$, teda inými slovami rovnakú pravdepodobnosť pre každú ihlu, že infikuje akúkoľvek osobu z populácie, dostávame rovnaké pravdepodobnosti ako v homogénnom prípade 2.7.

3.2 Stredná hodnota a rozptyl počtu novo infikovaných nehomogénnych osôb

Očakávané počty a rozptyl novo infikovaných osôb medzi heterogénnymi osobami používajúce i infikovaných ihiel môžeme znova odvodiť deriváciami pravdepodobnostnej vytvárajúcej funkcie. Tento postup je popísaný v Gani (2002). My sa znova prikloníme k rýchlejšiemu odvodeniu pomocou prístupu využívajúceho klasickú pravdepodobnosť. Aby sme udržali prehľadnosť a jednoduchosť používaných rovníc, budeme podobne ako v článku uvažovať rozdelenie populácie na dve skupiny. Zavedieme náhodný vektor $Y = (Y_{n_1,i}, Y_{n_2,i})^\top$ skladajúci sa z náhodnych veličín udávajúcich postupne počet infikovaných jedincov v prvej a druhej skupine. Tento vektor má multinomické rozdelenie, a teda obe jeho zložky majú binomické rozdelenie. Pre binomicky rozdelenú náhodnú veličinu platí, že vznikne ako súčet alternatívne rozdelených náhodných veličín. Teda aby sme odvodili očakávaný počet nakazených ľudí v prvej skupine používajúcich i infikovaných ihiel $\mathsf{E}[Y_{n_1,i}]$, definujeme alternatívnu náhodnu veličinu

$$\tilde{Z}_{j,n_1} = \begin{cases} 1 & j\text{-tý človek v 1. skupine neprišiel do styku s ihlou,} \\ 0 & \text{inak.} \end{cases}$$

Táto veličina označuje, že j-tá osoba v 1. skupine je neinfikovaná s pravdepodobnosťou

$$p_1 \frac{n_1 - 1}{n_1} + p_2 = 1 - \frac{p_1}{n_1},$$

pričom platí $p_1 + p_2 = 1$. Upozorňujeme, že v nasledujúcich odvodeniach budeme kvôli zjednodušeniu a lepšej čitateľnosti textu vynechávať 2. index alternatívnej veličiny $\tilde{Z}$, ktorý označuje, že sa nachádzame v 1. skupine. Vďaka nezávislosti máme pravdepodobnosť toho, že j-tá priehradka je prázdna pre i ihiel, vyjadrenú vzťahom $(1 - \frac{p_1}{n_1})^i$. Hľadanú strednú hodnotu ľahšie určíme pomocou doplnkovej alternatívnej náhodnej velčiny $Z_j = 1 - \tilde{Z}_j$ označujúcej rozdelenie toho, že j-tý človek v 1. skupine pri počte ihiel i je nakazený. Stredná hodnota tohto alternatívneho rozdelenia je

$$\mathsf{E}\,[Z_j] = \sum_{k=0}^{1} k\,\mathsf{P}(Z_j = k) = 1 - \left(1 - \frac{p_1}{n_1}\right)^i.$$

A konečne môžeme odvodiť hľadanú strednú hodnotu s okrajovou podmienkou $\mathsf{E}\,[Y_{n_1,1}] = p_1$

$$\mathsf{E}\,[Y_{n_1,i}] = n_1\,\mathsf{E}\,[Z_j] = n_1 \left[1 - \left(1 - \frac{p_1}{n_1}\right)^i\right]. \tag{3.9}$$

Strednú hodnotu pre druhú skupinu dostaneme nahradením n_1, p_1 v rovnici 3.9 za n_2, p_2.

K odvodeniu rozptylu postupne určíme rozptyl alternatívnej veličiny

$$\mathrm{var}[Z_j] = \left[1 - \left(1 - \frac{p_1}{n_1}\right)^i\right] \left(1 - \frac{p_1}{n_1}\right)^i$$

a kovarianciu, kde výraz $\mathsf{P}[\tilde{Z}_j = 1, \tilde{Z}_k = 1]$ určuje pravdepodobnosť, že j-tá a zároveň k-tá priehradka je voľná

$$\begin{aligned} \mathrm{cov}[\tilde{Z}_j, \tilde{Z}_k] &= \mathsf{E}\,[\tilde{Z}_j \tilde{Z}_k] - \mathsf{E}\,[\tilde{Z}_j]\,\mathsf{E}\,[\tilde{Z}_k] = \mathsf{P}[\tilde{Z}_j = 1, \tilde{Z}_k = 1] - \mathsf{E}\,[\tilde{Z}_j]^2 \\ &= \left(1 - \frac{2p_1}{n_1}\right)^i - \left(1 - \frac{p_1}{n_1}\right)^{2i}. \end{aligned}$$

Keď použijeme rovnakú úvahu ako v kapitole 1, môžeme napísať $\mathrm{cov}[\tilde{Z}_j, \tilde{Z}_k] = \mathrm{cov}[Z_j, Z_k]$, a teda hľadaný rozptyl je

$$\begin{aligned} \mathrm{var}[Y_{n_1,i}] &= n_1\,\mathrm{var}[Z_j] + n_1(n_1 - 1)\mathrm{cov}[Z_j, Z_k] \\ &= n_1 \left[1 - \left(1 - \frac{p_1}{n_1}\right)^i\right] \left(1 - \frac{p_1}{n_1}\right)^i \\ &+ n_1(n_1 - 1) \left[\left(1 - \frac{2p_1}{n_1}\right)^i - \left(1 - \frac{p_1}{n_1}\right)^{2i}\right]. \end{aligned} \tag{3.10}$$

Rozptyl $\mathrm{var}[Y_{n_2,i}]$ určíme nahradením n_1, p_1 za n_2, p_2.

Aby sme určili asymptotické hodnoty pre nehomogénny prípad strednej hodnoty a rozptylu očakávaného počtu novo infikovaných ľudí, uvažujeme $i = \lambda_1 n_1 + 1$, kde λ_1 je kladná konštanta, ktorá bližšie špecifikuje počet ihiel ako nejaký násobok zdravých osôb v 1. skupine, ktorý predpokladáme, že je vysoký. Teda aj počet infikovaných ihiel bude na základe tohto vzťahu vysoký. K odvodeniu asymptotických hodnôt využijeme vzťah $\lim_{n_1 \to \infty} \left(1 - \frac{p_1}{n_1}\right)^{\lambda_1 n_1} = e^{-p_1 \lambda_1}$, L'Hospitalovo pravidlo a rozvoj funkcie do geometrickej rady, dostávame

$$\lim_{n_1 \to \infty} \mathsf{E}\,[Y_{n_1,i}]/n_1 = 1 - e^{-\lambda_1 p_1}$$
$$\lim_{n_1 \to \infty} \mathrm{var}[Y_{n_1,i}]/n_1 = e^{-2\lambda_1 p_1} \left(e^{\lambda_1 p_1} - 1 - \lambda_1 p_1^2\right),$$

tak ako v homogénnom prípade zasa upozorňujeme na to, že v článku Gani (2002) je chybne uvedený asymptotický rozptyl. K odvodeniu asymptotických hodnôt pre druhú skupinu stačí zameniť n_1, p_1, λ_1 za n_2, p_2, λ_2. Je zrejmé, že pre λ_1 alebo λ_2 vysoké môžeme očakávať maximálny počet n_1 resp. n_2 nakazených osôb.

Nakoniec určíme očakávanú hodnotu počtu nakazených osôb, ak rozdeľujeme i ihiel medzi prvou a zároveň druhou skupinou

$$\begin{aligned}\mathsf{E}\,[Y_{n_1,i} Y_{n_2,i}] &= \sum_{j=1}^{n_1} \sum_{k=1}^{n_2} \mathsf{E}\,[Z_j U_k] = \sum_{j=1}^{n_1} \sum_{k=1}^{n_2} 1 - \mathsf{P}[\tilde{Z}_j = 1, \tilde{U}_k = 1] \\ &= n_1 n_2 \left[1 - \left[p_1 \left(\frac{n_1 - 1}{n_1}\right) + p_2 \left(\frac{n_2 - 1}{n_2}\right)\right]^i\right],\end{aligned}$$

kde U_k je alternatívna náhodná veličina určujúca, že k-tý jedinec v 2. skupine je nakazený.

Ešte podotkneme, že autor tento model zovšeobecnil na prípad, keď homogénni užívatelia medzi sebou zdieľajú ihly s dvomi typmi vírusov. Autor použil rovnakú metódu pomocou pravdepodobnostnej vytvárajúcej funkcie, ktorú sme už v práci popísali, a preto toto rozšírenie do práce zavedené nebolo. V prípade záujmu je možné toto rozšírenie nájsť v Gani (2001).

3.3 Aplikácia modelu

Pozrime sa na konkrétne hodnoty. Povedzme, že máme populáciu so 100 ľuďmi, z ktorých 60-ti sú v prvej a 40-ti v druhej skupine. Budeme chcieť zistiť, koľko máme očakávať novo infikovaných osôb iba v prvej skupine, ak uvažujeme, že do prvej skupiny sa dostaneme s pravdepodobnosťou $p = 0{,}75$. Logickou úvahou vyplýva, že očakávame menšiu hodnotu ako pre homogénny prípad, pretože stále rozdeľujeme napr. 10 ihiel medzi 100 osôb, no my sa pozrieme iba na nejakú ich podmnožinu a tam zisťujeme novo infikovaných.

Použitím rovnice 3.9 získavame hodnoty, ktoré zapíšeme do 2. stĺpca v tabuľke 3.2. Tieto hodnoty znovu overíme pomocou simulácie. Prvý krok v simulácii je rozhodnúť, či sme sa dostali do prvej alebo druhej skupiny. To dosiahneme tým, že generujeme číslo z rovnomerného rozdelenia z intervalu $(0, 1)$. Generovanú hodnotu porovnáme so zadanou pravdepodobnosťou p, ktorá určuje pravdepodobnosť, s akou sa dostaneme do prvej skupiny. V prípade, že generovaná hodnota je menšia alebo rovná p, vygenerujeme náhodné číslo z rovnomerného rozdelenia na hodnotách $1, \ldots, n_1$. Nového nakazeného získame, ak sme vygenerovali vyššiu hodnotu ako aktuálny počet nakazených. Tento postup opakujeme i-krát, čím získame nejakú hodnotu náhodnej veličiny $Y_{n_1,i}$. Simuláciu opakujeme 1000-krát pre každý zvolený počet ihiel. Strednú hodnotu odhadneme spočítaním priemeru, ktorý následne zapíšeme do 3. stĺpca v tabuľke 3.2.

Dostali sme približne rovnaké hodnoty ako pomocou výpočtu zo vzorca, a teda môžeme konštatovať, že skutočne simulujeme študované rozdelenie. Softvérový kód bol implementovaný v Mathematice 10.0 od Wolfram Research (2014). Celý kód je k dispozícii v Prílohe B.

Počet ihiel i	Očakávané počty novo infikovaných osôb	
	vzorec	simulácia
10	7,09	7,16
25	16,19	16,25
50	28,01	27,90
75	36,64	36,66
99	42,73	42,75
100	42,94	43,07
300	58,62	58,65
675	59,99	59,98

Tabuľka 3.2: Očakávané počty novo infikovaných osôb v prvej skupine obsahujúcej 60 osôb z celkového počtu osôb 100 pri zdieľaní rôzneho počtu ihiel získané pomocou vzorca a simulácie, pri uvažovanej pravdepodobnosti výskytu v prvej skupine rovnej 0,75

Všimnime si, že pre 10 ihiel už nie je očakávaný počet novo infikovaných ľudí približne 10, ale zhruba 7. Práve toto je smer úvah, ktorým sa J. Gani vydal, keď skúmal praktickú aplikáciu modelu. V článku Gani (2001) sa autor zaoberal špecifickým prípadom, ktorý si popíšeme. Rozdelíme n ľudí na 2 skupiny, pričom jedna bude očkovaná proti chorobe, druhá nebude a pozrieme sa, aký má vakcinácia vplyv na šírenie tejto infekčnej choroby. Môžeme uvažovať napríklad hepatitídu.

Tak ako doteraz uvažujeme, že ľudia medzi sebou zdieľajú i ihiel, pričom pravdepodobnosť, že sa ihla dostane k nevakcinovanému človeku, je n_1/n. To, že sa dostane k vakcinovanému, je určené ako doplnková pravdepodobnosť $1 - p = (n - n_1)/n$. Uvažujeme, že vakcinácia funguje na 100 %, a teda ak človek z druhej skupiny použil infikovanú ihlu, je chránený a nenakazí sa. Autor v článku ukázal, že pre očakávaný počet novo nakazených osôb platí rovnica 3.9.

Ukážeme postup ako určiť koľko percent populácie očkovať, aby sme držali očakávaný počet infikovaných ľudí pod nejakou vopred stanovenou maximálou hodnotou. Matematicky zapísané, budeme chcieť dosiahnuť nasledujúci vzťah

$$n_1 \left[1 - \left(1 - \frac{1}{n} \right)^i \right] \le an, \tag{3.11}$$

kde a označuje nejaký podiel populácie, typicky $a = 0{,}05$. Teda pre $n = 100$, $i = 10$ budeme chcieť nájsť najväčší počet nevakcinovaných ľudí n_1, tak aby platil výraz 3.11. Pre tieto konkrétne hodnoty sme stanovili $n_1 = 52{,}30$. Kedže chceme určiť počet ľudí, z tejto hodnoty vezmeme jej dolnú celú časť. Tento postup opakujeme pre rôzne počty ihiel i, zistené hodnoty zapíšeme do tabuľky 3.3. Počet ľudí, ktorých potrebujeme zaočkovať určíme ako doplnok k neočkovaným.

Počet ihiel i	10	25	50	75	99	100	300	675
Počet nevakcinovaných n_1	52	22	12	9	7	7	5	5

Tabuľka 3.3: Časť neočkovanej populácie, ktorá sa ešte môže nakaziť, aby nebola presiahnutá zvolená hodnota maximáneho počtu nakazených, ktorú pripúštame

Záver

Práca bola zameraná na urnové modely, v ktorých sa počet guľôčok menil na základe výsledkov v predchádzajúcich ťahoch. Detailne bol preštudovaný a popísaný model, ktorý bol demonštrovaný na príklade zdieľania ihiel medzi homogénnymi i nehomogénnymi drogovými užívateľmi. Pomocou pravdepodobnostných vytvárajúcich funkcií sme popísali pravdepodobnostné rozdelenia počtu novo infikovaných jedincov v oboch prípadoch. Stredné hodnoty a rozptyly boli odvodené pomocou klasickej kolmogorovskej pravdepodobnosti. Na základe odvodených vzorcov a simulácií boli ukázané očakávané počty novo infikovaných ľudí pri zdieľaní rôzneho nami zvoleného počtu ihiel v homogénnom i nehomogénnom prípade. V závere poslednej kapitoly bol ukázaný špeciálny prípad, v ktorom sme určili, akú veľkú časť populácie je potrebné zaočkovať, aby nákaza neinfikovala viac ako 5% z celkovej populácie. V dnešnom svete plnom rôznych epidemických chorôb, je tento model kvôli svojmu praktickému využitiu stále atraktívny pre väčší počet matematikov a je podrobovaný ďalším a ďalším štúdiám.

Literatúra

GANI, J. (2001). *Nonhomogeneous Susceptibles and Infectives among Needle Sharing IVDUs.* World Scientific. ISBN 978-981-238-201-6.

GANI, J. (2002). Needle sharing infections among heterogeneous ivdus. *Monatshefte für Mathematik*, **135**, pp. 25–36. ISSN 1436-5081.

GANI, J. (2004). Random-allocation and urn models. *J. Appl. Probab.*, **41**, pp. 313–320. ISSN 02109928.

JOHNSON, N. L. a KOTZ, S. (1977). *Urn Models and Their Application: An Approach to Modern Discrete Probability Theory.* Wiley, New York. ISBN 0-471-44630-0.

PRÁŠKOVÁ, Z. a LACHOUT, P. (2001). *Základy náhodnych procesů.* Karolinum, Praha. ISBN 80-7184-688-0.

WOLFRAM RESEARCH, I. (2014). *Mathematica, Version 10.0.* Wolfram Research, Inc., Champaign, Illionis.

WOODBURY, M. (1949). On a probability distribution. *Ann. Math. Statist.*, **20**, pp. 311–313. ISSN 00034851.

Zoznam obrázkov

Zoznam tabuliek

Prílohy

Príloha A: homogénny prípad

```
SimulaciaPreJednuSkupinu[i_Integer, n_Integer] :=
 Module[{PocetNakazenych = 0, j = 1, VyberCloveka},
  While[j <= i,
   VyberCloveka = RandomVariate[DiscreteUniformDistribution[{1, n}]];
   If[VyberCloveka > PocetNakazenych, PocetNakazenych++];
   j++;
   ];
  PocetNakazenych
  ]

SeedRandom[17031992];
Table[SimulaciaPreJednuSkupinu[10, 100], {1000}] // Mean // N
Table[SimulaciaPreJednuSkupinu[25, 100], {1000}] // Mean // N
Table[SimulaciaPreJednuSkupinu[50, 100], {1000}] // Mean // N
Table[SimulaciaPreJednuSkupinu[75, 100], {1000}] // Mean // N
Table[SimulaciaPreJednuSkupinu[99, 100], {1000}] // Mean // N
Table[SimulaciaPreJednuSkupinu[100, 100], {1000}] // Mean // N
Table[SimulaciaPreJednuSkupinu[300, 100], {1000}] // Mean // N
Table[SimulaciaPreJednuSkupinu[675, 100], {1000}] // Mean // N
```

Príloha B: nehomogénny prípad

```
SimulaciaPreDveSkupiny[i_Integer, n1_Integer, p_] :=
 Module[{PocetNakazenych1 = 0, j = 1, VyberCloveka, VyberSkupinu},
  While[j <= i,
   VyberSkupinu = RandomVariate[UniformDistribution[{0, 1}]];
   If[VyberSkupinu <= p,
    {VyberCloveka =
      RandomVariate[DiscreteUniformDistribution[{1, n1}]];
     If[VyberCloveka > PocetNakazenych1, PocetNakazenych1++]}
    ];
   j++;
   ];
  PocetNakazenych1
  ]

SeedRandom[17031992];
Table[SimulaciaPreDveSkupiny[10, 60, 0,75], {1000}]  // Mean // N
Table[SimulaciaPreDveSkupiny[25, 60, 0,75], {1000}]  // Mean // N
```

```
Table[SimulaciaPreDveSkupiny[50, 60, 0,75], {1000}]  // Mean // N
Table[SimulaciaPreDveSkupiny[75, 60, 0,75], {1000}]  // Mean // N
Table[SimulaciaPreDveSkupiny[99, 60, 0,75], {1000}]  // Mean // N
Table[SimulaciaPreDveSkupiny[100, 60, 0,75], {1000}] // Mean // N
Table[SimulaciaPreDveSkupiny[300, 60, 0,75], {1000}] // Mean // N
Table[SimulaciaPreDveSkupiny[675, 60, 0,75], {1000}] // Mean // N
```