



**MATEMATICKO-FYZIKÁLNÍ
FAKULTA**
Univerzita Karlova

BAKALÁŘSKÁ PRÁCE

Karla Strachoňová

Základní stochastické epidemiologické modely

Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: RNDr. Šárka Hudecová, Ph.D.

Studijní program: Matematika

Studijní obor: Obecná matematika

Praha 2019

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 má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle §60 odst. 1 autorského zákona.

V dne

Podpis autora

Chtěla bych poděkovat své vedoucí bakalářské práce RNDr. Šárce Hudecové, Ph.D. za odborné vedení, za pomoc a cenné rady při zpracování této práce.

Název práce: Základní stochastické epidemiologické modely

Autor: Karla Strachoňová

Univerzita Karlova: Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: RNDr. Šárka Hudecová, Ph.D., Katedra pravděpodobnosti a matematické statistiky

Abstrakt: V této práci je popsáno modelování šíření epidemií v uzavřených populacích pomocí dvou základních modelů: Greenwoodova a Reedova-Frostova. Nejprve uvedeme shrnutí znalostí o Markovových řetězcích a náhodných veličinách. Následně je popsán klasický postup sledující počet náchylných a počet infikovaných jedinců. Poté se práce zabývá rozdělením doby trvání epidemie a celkového počtu nakažených jedinců do jejího skončení. Oba tyto přístupy jsou aplikovány na příkladech. Dále je diskutován postup odhadování pravděpodobnosti nákazy metodou maximální věrohodnosti a také projednávány možnosti, jakými můžeme pomocí matematických modelů ovlivnit průběh epidemie. Nakonec jsou představené modely aplikovány na reálná data a je provedena diskuze o jejich přesnosti.

Klíčová slova: stochastické modely, epidemiologie, Greenwoodův model, Reedův-Frostův model

Title: Basic stochastic epidemic models

Author: Karla Strachoňová

Charles university: Department of probability and mathematical statistics

Supervisor: RNDr. Šárka Hudecová, Ph.D., Department of probability and mathematical statistics

Abstract: This thesis deals with two basic models which are used for epidemic modelling in closed populations, namely Greenwood and Reed-Frost models. At first, knowledge which a reader needs to have about Markov chains and random variables is summarized. Then the two models are described by modelling the number of susceptible and infectious individuals, as well as the duration and size of the epidemic. All of these approaches to modelling an epidemic are then illustrated on examples. Finally, the maximum likelihood method of the probability of infection is described and illustrated on real data in the last chapter, where the obtained results are discussed as well.

Keywords: stochastic models, epidemiology, Greenwood model, Reed-Frost model

Obsah

| | | |
|----------|---|-----------|
| 1 | Úvod | 2 |
| 2 | Náhodné procesy | 3 |
| 2.1 | Markovovy řetězce s diskretním časem | 3 |
| 2.2 | Některé vlastnosti náhodných veličin | 5 |
| 3 | Stochastické modely v epidemiologii | 7 |
| 3.1 | Stochastické modely v diskretním čase | 7 |
| 3.1.1 | Obecná formulace modelů | 8 |
| 3.1.2 | Greenwoodův model | 9 |
| 3.1.3 | Reedův-Frostův model | 20 |
| 4 | Odhad parametru α | 27 |
| 4.1 | Odhad podle znalosti W | 27 |
| 4.2 | Odhad podle znalosti řetězce | 31 |
| 5 | Aplikace na data | 36 |
| 5.1 | Data o spalničkách | 36 |
| 5.2 | Data o nachlazení | 38 |
| 5.3 | Aplikace a pozorování | 40 |
| 5.3.1 | Kontrola epidemie | 41 |
| | Závěr | 42 |
| | Seznam použité literatury | 43 |
| | Seznam obrázků | 44 |
| | Seznam tabulek | 45 |

1. Úvod

Šíření infekční choroby je náhodný proces. Jeden z prvních stochastických modelů vytvořil v roce 1926 Anderson Gray McKendrick. Zabýváme-li se epidemiologií, jedním z nejznámějších a nejjednodušších modelů je řetězový binomický model Reeda a Frosta, kteří ho formulovali roku 1928 na Univerzitě Johnse Hopkinse v USA. Nikdy však své dílo nepublikovali, a tak se jej později ujali další vědci, viz [1]. Nicméně některé aspekty Reed-Frostova modelu sepsal již roku 1889 Pyotr Dimitrievich Enko, který zkoumal epidemii spalniček v ruském Petrohradu. Více viz [8].

Mluvíme-li o matematických modelech reálného světa, jde vždy o zjednodušení jevů, které se kolem nás dějí. Nelze tedy ani v nejmenším předpokládat, že bychom matematickými modely dokázali naprosto přesně popsat, jak bude epidemie nějaké choroby probíhat. Jedná se o velmi zjednodušující postup, který naznačuje, jaká by mohla být situace v určitém čase. I přesto je matematické modelování v praxi užitečné.

V této práci se budeme zabývat modelováním šíření infekce v uzavřené populaci pomocí stochastických modelů, tj. modelů náhodných jevů. V práci se zaměříme na modely v diskrétním čase a na infekční nemoci, kterými může jedinec projít pouze jedenkrát za život, neboť výsledkem prodělání nemoci je smrt či doživotní imunita. Konkrétně jde o Reedův-Frostův a Greenwoodův model. V práci je popsán detailní postup vytváření těchto dvou modelů z vícero pohledů spolu s názornými příklady a nakonec aplikací na reálných datech. Je kladen důraz na pochopení chování popisovaných náhodných veličin a na jejich vlastnosti a souvislosti s teorií Markovových řetězců.

Ve druhé kapitole jsou zavedeny definice a vlastnosti týkající se náhodných procesů, především homogenních Markovových řetězců v diskrétním čase. Také jsou zde k připomenutí některé pojmy týkající se diskrétních náhodných veličin, které budeme v práci využívat. Třetí kapitola je hlavní teoretickou částí celé práce. Je zde detailně popsán postup vytváření Greenwoodova a Reedova-Frostova modelu. Nejprve se zaměříme na modelování binomických řetězců počtu náchylných a počtu infikovaných jedinců. Další částí této kapitoly je pak modelování velikosti a délky trvání epidemie, zejména je odvozeno rozdělení těchto dvou náhodných veličin. Ve čtvrté kapitole si předvedeme postup odhadování neznámého parametru modelu metodou maximální věrohodnosti pro různě početné domácnosti. To pak využijeme v kapitole páté, kde postupy aplikujeme na reálná data o epidemii spalniček a epidemii nachlazení.

Práce především vychází z knihy [7], ale v teoretické části bylo čerpáno i ze článků [10] a [16], přičemž řada matematických odvození je v práci provedena podrobněji a s vysvětlujícími komentáři. Analyzovaná data byla převzata z knihy [9] a článků [5] a [3]. Pro simulace a příklady byly využity programy R a Mathematica.

Práce by měla být využívána jako zdroj teoretických metod vytváření základních modelů v epidemiologii a jejich aplikací. Především by měla sloužit k pochopení základních a nejjednodušších přístupů k modelování epidemií.

2. Náhodné procesy

Šíření epidemie modelujeme pomocí náhodného procesu. V této kapitole jsou uvedeny některé základní definice a vlastnosti týkající se náhodných procesů, se kterými budeme pracovat v dalších kapitolách.

Definice 1 (Náhodný proces). *Nechť (Ω, \mathcal{A}, P) je pravděpodobnostní prostor a buď $T \subseteq \mathbb{R}$ neprázdná. Pak soustavu reálných náhodných veličin $\{X_t : t \in T\}$ definovaných na (Ω, \mathcal{A}, P) nazýváme **reálný náhodný (stochastický) proces**.*

Poznámka. Je-li $T \in \mathbb{N}_0$ nebo $T \in \mathbb{Z}$, jedná se o náhodný proces s diskrétním časem. Pokud $T = [a, b]$, kde $-\infty \leq a < b \leq \infty$, pak mluvíme o procesu se spojitým časem. V této práci budeme uvažovat pouze náhodné procesy s diskrétním časem.

Mezi základní parametry popisující náhodné veličiny patří jejich střední hodnota. Definujme proto střední hodnotu celého náhodného procesu.

Definice 2 (Střední hodnota náhodného procesu). *Nechť $\{X_t : t \in T\}$ je náhodný proces, kde pro každé $t \in T$ existuje $E X_t$. Pak **střední hodnota náhodného procesu** $\{X_t : t \in T\}$ je funkce $E X_t$ definovaná na T .*

Jedním z typů náhodných procesů je Galtonův-Watsonův proces větvení, který sleduje počet potomků v postupně vznikajících generacích.

Definice 3 (Větvící se proces). *Mějme populaci jedinců, kteří žijí přesně jednu jednotku času a každý z nich má náhodný počet potomků. Nechť jsou počty potomků jednotlivých jedinců na sobě nezávislé a stejně rozdělené náhodné veličiny a zároveň nezávislé na předchozím průběhu procesu. Označme náhodnou veličinu U_{nj} jako počet potomků j -tého jedince v n -té generaci. Pak **větvící se proces** (tzv. Galtonův-Watsonův proces větvení) je náhodný proces $\{X_n : n \in \mathbb{N}_0\}$, v němž náhodná veličina X_n vyjadřuje celkový počet jedinců v n -té generaci pro $n = 0, 1, \dots$, tedy*

$$X_n = \sum_{j=1}^{X_{n-1}} U_{nj}.$$

2.1 Markovovy řetězce s diskrétním časem

V následující kapitole využíváme znění definic a vět z knihy [14] uvedené v seznamu literatury. Pro důkazy uvedených vět a tvrzení se odkazujeme na tuto publikaci.

Náhodné procesy, kterými se budeme v modelování epidemií zabývat, jsou tzv. Markovovy řetězce. Jedná se o náhodné procesy definované v Definici (4).

Definice 4 (Markovův řetězec s diskrétním časem). *Náhodný proces $\{X_n : n \in \mathbb{N}_0\}$ nazveme **Markovovým řetězcem s diskrétním časem** a množinou stavů S , splňuje-li tzv. markovskou vlastnost, tj. pro všechna $n = 0, 1, \dots$ a všechna $i, j, i_0, i_1, \dots, i_{n-1} \in S$ splňující*

$$P(X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) > 0$$

platí:

$$P(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = P(X_{n+1} = j | X_n = i).$$

Pravděpodobnost na pravé straně rovnosti nazýváme **pravděpodobnost přechodu** ze stavu i v čase n do stavu j v čase $n + 1$ a značíme ji $p_{ij}(n, n + 1)$. Říká se jí též pravděpodobnost přechodu 1. řádu. Podobně zavedeme pravděpodobnost přechodu m -tého řádu ze stavu i v čase n do stavu j v čase $n + m$ jako $p_{ij}(n, n + m) = P(X_{n+m} = j | X_n = i)$ pro nějaké přirozené $m \geq 1$. Pravděpodobnost přechodu m -tého řádu ze stavu i do stavu j značíme obecně $p_{ij}^{(m)}$ pro $m \geq 1$.

Definice 5 (Homogenní Markovův řetězec). Řekneme, že Markovův řetězec $\{X_n : n \in \mathbb{N}_0\}$ je **homogenní**, pokud pravděpodobnosti přechodu $p_{ij}(n, n + m)$ nezávisí na n , ale pouze na m .

Galtonův-Watsonův proces z Definice 3 je též homogenní Markovův řetězec. Jeho formulace využíváme dále v kapitole 3.1.2.

Pro usnadnění práce s pravděpodobnostmi přechodu nám poslouží matice jimi tvořená. Její význam si ukážeme v kapitolách 3.1.3 a 3.1.2 při modelování epidemie pomocí vytvářující funkce.

Definice 6 (Matice pravděpodobností přechodu). Čtvercovou matici $P = \{p_{ij}\}$ nazýváme **maticí pravděpodobností přechodu** prvního řádu homogenního Markovova řetězce, kde p_{ij} jsou pravděpodobnosti přechodu prvního řádu ze stavu i do stavu j , kde i a j jsou z množiny stavů S . Pro matici P platí dvě následující vlastnosti:

- (i) $p_{ij} \geq 0$ pro všechna $i, j \in S$,
- (ii) $\sum_j p_{ij} = 1$ pro všechna $i \in S$.

Podobně definujeme matici $P^m = \{p_{ij}^{(m)}\}$ matici pravděpodobností přechodu m -tého řádu.

Definice 7 (Počáteční rozdělení a absolutní pravděpodobnosti). **Počátečním rozdělením** homogenního Markovova řetězce rozumíme vektor $p(0) = (p_i^{(0)})_{i \in S}$, kde $p_i^{(0)} = P(X_0 = i)$. Obdobně definujeme **absolutní pravděpodobnosti** Markovova řetězce v čase n jako $p(n) = (p_i(n))_{i \in S}$, kde $p_i(n) = P(X_n = i)$.

Tvrzení 1. Buď $p(0)$ počáteční rozdělení homogenního Markovova řetězce. Pak

$$p(n)^T = p(0)^T P^n, n \in \mathbb{N}_0.$$

Chceme-li vyjádřit pravděpodobnost, že se z konkrétního počátečního stavu dostaneme přes konkrétní mezistavy do stavu konečného, poslouží nám následující věta. Ta říká, že takovou pravděpodobnost lze vyjádřit součinem dílčích pravděpodobností přechodu mezi jednotlivými stavy a pravděpodobnosti počátečního stavu.

Věta 2. Necht $\{X_t : t \in T\}$ je náhodný proces s množinou stavů $S = \{0, 1, 2, \dots\}$, necht \mathbf{p} je vektor splňující vlastnosti v Definici 7 a \mathbf{P} je matice splňující vlastnosti Definice 6. Pak je $\{X_t : t \in T\}$ homogenní Markovův řetězec s počátečním rozdělením \mathbf{p} a maticí pravděpodobností přechodu \mathbf{P} právě tehdy, když platí

$$P(X_0 = x_0, X_1 = x_1, \dots, X_n = x_n) = p_{x_0} p_{x_0 x_1} \cdots p_{x_{n-1} x_n}$$

pro všechna $x_i \in S$ a všechna $n \in \mathbb{N}_0$.

Krom takového vyjádření bude pro výpočty užitečné definovat pravděpodobnost $p_{ij}^{(n)}$ jako pravděpodobnost přechodu ze stavu i do stavu j v celkem n krocích. Následující tvrzení popisuje, jak lze rekurentně tuto hodnotu vyjádřit jako součet součinů pravděpodobnosti, že se po $n - 1$ krocích dostaneme ze stavu i do stavu k , s pravděpodobností přechodu ze stavu k do stavu j v jednom kroku.

Tvrzení 3. Položme $p_{ij}^{(0)} = \delta_{ij}$, kde δ_{ij} je Kroneckerovo delta a $p_{ij}^{(1)} = p_{ij}$. Definujeme rekurentně pro $n \in \mathbb{N}$ pravděpodobnosti $p_{ij}^{(n)}$ tak, že

$$p_{ij}^{(n)} = \sum_{k \in S} p_{ik}^{(n-1)} p_{kj}.$$

Pak z tohoto vztahu plyne, že matice $P^{(n)}$ prvků $p_{ij}^{(n)}$ je n -tou mocninou matice P , tj.

$$P^{(n)} = P^n$$

pro všechna $n \in \mathbb{N}$.

Definice 8 (Uzavřenost množiny). Necht $C \subset S$ je neprázdná množina. Řekneme, že C je **uzavřená**, pokud neexistují stavy $j \notin C$ a $i \in C$ takové, že j je dosažitelný stav z i , tj. pokud existuje $n \in \mathbb{N}_0$ takové, že $p_{ij}^{(n)} > 0$.

Definice 9. Stav j nazýváme **absorpční**, pokud je množina $\{j\}$ uzavřená.

Poznámka. Každý homogenní Markovův řetězec je určen třemi parametry:

- (i) množinou stavů S ,
- (ii) vektorem počátečního rozdělení $p(0)$,
- (iii) maticí pravděpodobností přechodu P .

2.2 Některé vlastnosti náhodných veličin

Při popisu modelů budeme využívat i vytvořující funkce, neboť nám to v některých případech velmi usnadní počítání. Vytvořující funkce jednoznačně určuje rozdělení náhodné veličiny, tedy lze z ní spočítat např. střední hodnotu či rozptyl náhodné veličiny, a to jen pomocí derivování. V kapitole 3.1.3 je naznačeno, jak náročným způsobem bychom museli počítat, kdybychom se chtěli modelování pomocí vytvořující funkce vyhnout.

Definujme proto charakteristickou a vytvořující funkci náhodné veličiny.

Definice 10 (Vytvořující funkce diskrétní náhodné veličiny). Necht X je náhodná veličina s diskrétním rozdělením s hodnotami v \mathbb{N}_0 . **Vytvořující funkci** vlastní náhodné veličiny X definujeme jako

$$P_X(s) = E[s^X] = \sum_{n=0}^{\infty} P(X = n) s^n,$$

což je dobře definovaná řada pro $|s| < 1$.

Definice 11 (Sdružená vytvořující funkce diskrétních náhodných veličin). Necht X, Y jsou náhodné veličiny s diskrétním rozdělením s hodnotami v \mathbb{N}_0 . Pak **sdruženou vytvořující funkci** náhodného vektoru $(X, Y)^T$ je funkce

$$P_{(X,Y)}(s, t) = E[s^X t^Y] = \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} P(X = n, Y = m) s^n t^m.$$

Je zřejmé, že rozdělení náhodné veličiny, resp. náhodného vektoru, vyjadřuje vektor koeficientů vytvořující funkce.

Následující věta popisuje výpočet střední hodnoty a rozptylu náhodné veličiny pomocí její vytvořující funkce.

Věta 4 (Střední hodnota, rozptyl a vytvořující funkce). *Nechť X je nezáporná celočíselná náhodná veličina. Označme $P'_X(1-) = \lim_{s \rightarrow 1-} P'_X(s)$. Pak pro momenty veličiny X platí:*

(i) $EX = P'_X(1-)$,

(ii) $EX(X-1)\dots(X-k+1) = P_X^{(k)}(1-)$,

(iii) $\text{var} X = P''_X(1-) + P'_X(1-) - (P'_X(1-))^2$, je-li $EX < \infty$.

3. Stochastické modely v epidemiologii

V epidemiologii má matematické modelování tři základní cíle. Prvním z nich je lépe pochopit **mechanismy**, díky nimž dochází k šíření choroby. Příkladem je popis parametru ρ jakožto relativní míry odebrání (anglicky *relative removal rate*) v Kermack-McKendrickově modelu. Jedná se o číslo, které musí být nižší než počáteční počet náchylných jedinců x_0 , aby se epidemie rozrůstala.

Druhým z cílů je umět předpovědět budoucí **vývoj** epidemie. Zajímá nás, kolik jedinců bude ovlivněno epidemií, čímž dokážeme stanovit zdravotní náklady na léčbu během celé epidemie.

Třetím aspektem je porozumět, jak **kontrolovat růst** epidemie. Nejčastěji se zabýváme otázkou, jak bránit šíření choroby pomocí vzdělávání obyvatel, imunizací náchylných či izolací infikovaných. Pokud by kupříkladu bylo možné vakcinací zredukovat počet náchylných x_0 v Kermack-McKendrickově modelu pod hodnotu ρ , epidemie by nabývala mnohem nižších rozměrů.

3.1 Stochastické modely v diskrétním čase

Abychom popsali šíření choroby, jejíž infekční doba (tj. doba, kdy je jedinec infekční) je relativně krátká oproti latentnímu období (tj. době, kdy je jedinec již nakažen, ale ještě neprojevuje známky choroby a není infekční), je výhodné použít model v diskrétním čase, jehož časová jednotka je délka latentního období. Tou může být jeden či více dní.

Existují dva základní modely tohoto typu, a to Reedův-Frostův (1928) a Greenwoodův (1931). V obou modelech platí, že každý infikovaný jedinec je infekční právě jednu časovou jednotku. Jedná se o modely typu XYZ (také označovány jako SIR), které popisují vývoj epidemie pouze pro choroby, po jejichž prodělání je člověk buď navždy imunní, nebo zemře. Příkladem takové choroby jsou neštovice, spalničky, či zjednodušená reprezentace chřipky.

Popis stavů: Každý jedinec je v každé časové jednotce v právě jednom ze tří stavů. Ve stavu X (S: susceptibles) jsou tzv. náchylní jedinci, kteří nemoc nikdy předtím neprodělali, a tedy nejsou imunní. Ve stavu Y (I: infectives) jsou infikovaní jedinci, kteří jsou v časové jednotce po infikování schopni nemoc přenést na náchylného jedince. Ve stavu Z (R: recovered and immune / removed) jsou jedinci, kteří si nemoc již prodělali, a jsou buď navždy imunní, nebo zemřeli. Ze stavu X může jedinec přejít pouze do stavu Y, nebo zůstat ve stavu X. Ze stavu Y jedinec po jedné časové jednotce přejde do stavu Z, ve kterém už zůstane. Modely budeme tedy stavět na vytváření náhodných procesů popisujících počty náchylných a infikovaných jedinců v čase, přičemž v každé časové jednotce t odebereme z procesu všechny infikované a soustředíme se pouze na ty nově vzniklé v čase $t + 1$.

Jak popis již naznačuje, oba modely předpokládají, že se v průběhu šíření choroby nemění celkový počet uvažovaných osob, tj. $X_t + Y_t + Z_t$ je během epidemie pro všechna t konstantní. Proto říkáme, že modelujeme vývoj epidemie

v uzavřené populaci. Speciálně, pokud $Y_t = 0$, epidemie končí.

3.1.1 Obecná formulace modelů

Definujme náhodné veličiny:

- X_t jako počet náchylných jedinců v populaci v čase $t = 0, 1, \dots$
- Y_t jako počet infikovaných jedinců v populaci v čase $t = 0, 1, \dots$

Pro smysl modelu předpokládáme, že $Y_0 = y_0 \geq 1$. Jelikož každý z infikovaných jedinců je infekční pouze po jednu časovou jednotku, platí rovnost

$$X_t = X_{t+1} + Y_{t+1}.$$

Značení: Budeme-li v textu pracovat s vektory náhodných veličin nebo hodnot záviselých na čase, budeme používat zjednodušený zápis:

$$(X_t, Y_t)^T = (X, Y)_t^T.$$

Dále označme

- $p \in (0, 1)$ pravděpodobnost **kontaktu** mezi infikovaným a náchylným jedincem,
- $\beta \in (0, 1)$ pravděpodobnost **přenesení infekce** z infikovaného na náchylného jedince, došlo-li ke vzájemnému kontaktu.

Pak pravděpodobnost, že se daný náchylný jedinec nenakazí infekcí, je součet pravděpodobnosti, že ke kontaktu vůbec nedojde, s pravděpodobností, že nedojde k přenosu infekce při vzájemném kontaktu, tedy

$$\alpha = (1 - p) + p(1 - \beta) = 1 - p\beta.$$

Předpokládáme, že všechny zde uvedené pravděpodobnosti jsou **konstantní** v průběhu celé epidemie, tedy dokud nenastane její konec.

Protože se jedná o model v uzavřené populaci, předpokládáme, že pro celkový počet jedinců n v této populaci platí $X_0 + Y_0 = n$, a tedy

$$X_t + \sum_{i=0}^t Y_i = n.$$

Je-li $X_t = x_t$ a $Y_t = y_t$, pak s pravděpodobností $\alpha(y_t)$ (tj. funkcí proměnné y_t) bude jedinec náchylný v čase t stále náchylným i v čase $t + 1$. Náhodná veličina X_{t+1} má podmíněně binomické rozdělení $\mathbb{B}i(x_t, \alpha(y_t))$ při daném $(X, Y)_t^T$. Proto můžeme napsat vztah pro pravděpodobnost přechodu ze stavu $(X, Y)_t^T = (x, y)_t^T$ do stavu $(X, Y)_{t+1}^T = (x, y)_{t+1}^T$ jako podmíněnou pravděpodobnost:

$$P(X_{t+1} = x_{t+1} | X_t = x_t, Y_t = y_t) = \binom{x_t}{x_{t+1}} \alpha(y_t)^{x_{t+1}} (1 - \alpha(y_t))^{y_{t+1}},$$

kde $x_{t+1} = x_t - y_{t+1}$.

Stochastické modely, které jsou popsány v kapitolách 3.1.3 a 3.1.2, tj. Reedův-Frostův a Greenwoodův model, se od sebe liší ve vyjádření funkce α . V Greenwoodově modelu je $\alpha(y_t) = \alpha$, tedy konstantní funkce, která nezávisí na hodnotě y_t , zatímco v Reedově-Frostově modelu na ní závisí a je rovna $\alpha(y_t) = \alpha^{y_t}$.

Epidemie se zastaví v čase $t = T$ ve dvou možných případech, přičemž oba uvažujeme vzhledem k námi zkoumané uzavřené populaci, díky čemuž se jedná o konečný stav.

(i) V prvním případě nastane $X_t = X_{t-1}$, což znamená, že počet náchylných jedinců se v čase t nezměnil. Proto nepřibyl žádný nakažený jedinec, tj. $Y_t = 0$ a nemoc se již nemůže šířit.

(ii) Druhý případ nastane, nezbyde-li v populaci žádný náchylný jedinec, tedy $X_t = 0$.

Definujme tedy náhodnou veličinu T , kterou nazýváme *doba trvání epidemie* jako

$$T = \min\{t : Y_t = 0\} = \min\{t : X_t = X_{t-1}\}.$$

Vývoj epidemie budeme vždy sledovat pouze do času T , tedy do jejího zastavení.

Dále definujme náhodnou veličinu W značící celkový počet jedinců nakažených během epidemie, neboli *velikost epidemie*, jako

$$W = \sum_{t=1}^{T-1} Y_t.$$

Všimněme si, že velikost epidemie nezahrnuje počet počátečních infikovaných.

Zajímají-li nás vlastnosti těchto dvou náhodných veličin, musíme popsat model popisující chování náhodného vektoru $(W, T)^\top$. Takový model lze sestavit pomocí vytvářející funkce, jak si ukážeme v kapitolách 3.1.3 a 3.1.2.

Tato modifikovaná formulace matematického modelu nám může přinést mnoho výhod, především hlubší vhled do modelování epidemických procesů a ustanovení relativně snadných algoritmů pro počítání pravděpodobností pro náhodné veličiny T a W .

3.1.2 Greenwoodův model

Greenwoodův model předpokládá, že pravděpodobnost nakažení se nezávisí na počtu infikovaných, neboli vystavení se dvěma či více infikovaným ve stejném čase je ekvivalentní vystavení se jednomu infikovanému člověku. Pravděpodobnost α je v tomto případě pravděpodobnost nepřenesení infekce, neboli pravděpodobnost, že se konkrétní jedinec v jednom časovém úseku nenakazí infekcí. V tomto případě opět platí, že

$$x_t = x_{t+1} + y_{t+1},$$

a přechodovou pravděpodobnost mezi časy t a $t + 1$ vyjádříme jako

$$\begin{aligned} P_{(x,y)_t^\top, (x,y)_{t+1}^\top} &= \mathbf{P}\left(\left((X, Y)_{t+1}^\top = (x, y)_{t+1}^\top \mid (X, Y)_t^\top = (x, y)_t^\top\right)\right) \\ &= \binom{x_t}{x_{t+1}} \alpha^{x_{t+1}} (1 - \alpha)^{y_{t+1}} = \binom{x_t}{x_{t+1}} \alpha^{x_{t+1}} (1 - \alpha)^{x_t - x_{t+1}}. \end{aligned} \quad (3.1)$$

Dostáváme tak Markovův řetězec $\{X_t : t = 0, 1, \dots\}$, kde $Y_{t+1} = X_t - X_{t+1}$, $0 \leq X_{t+1} \leq X_t$.

Poznámka. Protože sledujeme model pouze do času T zastavení epidemie, promítá se ve vzorci skutečnost, že $Y_t > 0$, jinak by vzorec nebyl korektní.

Všimneme si zde návaznosti na **Galtonův-Watsonův proces větvení** popsaný v Definici 3. Můžeme totiž formulovat Greenwoodův model jako Galtonův-Watsonův proces pro posloupnost náhodných veličin $\{X_t : t \in \mathbb{N}_0\}$, a to následujícím způsobem.

Uvažujme, že každý náchylný jedinec v čase t může mít za svého „potomka“ buď sám sebe v případě, že se vyhnul nákaze, nebo žádného, pokud byl v čase t infikován. Označíme-li stejně jako v Definici 3 náhodnou veličinu U_{tj} jako počet potomků j -tého jedince v t -té generaci, pak má tato náhodná veličina pro všechna t a j alternativní rozdělení s parametrem α , neboli:

$$\begin{aligned} P(U_{tj} = 1) &= \alpha, \\ P(U_{tj} = 0) &= 1 - \alpha, \end{aligned}$$

a platí

$$X_t = \sum_{j=1}^{X_{t-1}} U_{tj}.$$

Jelikož jsou počty infikovaných v navazujících generacích dány binomickým rozdělením, můžeme některé charakteristiky řetězce snadno spočítat. V Greenwoodově modelu platí (z již známé pravděpodobnosti přechodu):

$$E[X_{t+1}|X_t] = \alpha X_t,$$

tedy

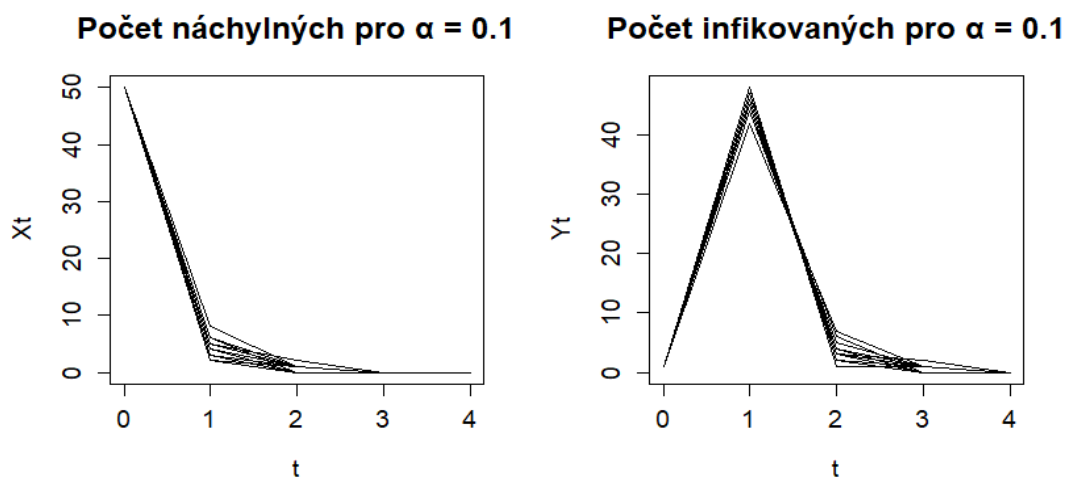
$$E[X_t|X_0 = x_0] = \alpha^t x_0,$$

a protože $Y_t = X_{t-1} - X_t$, pak

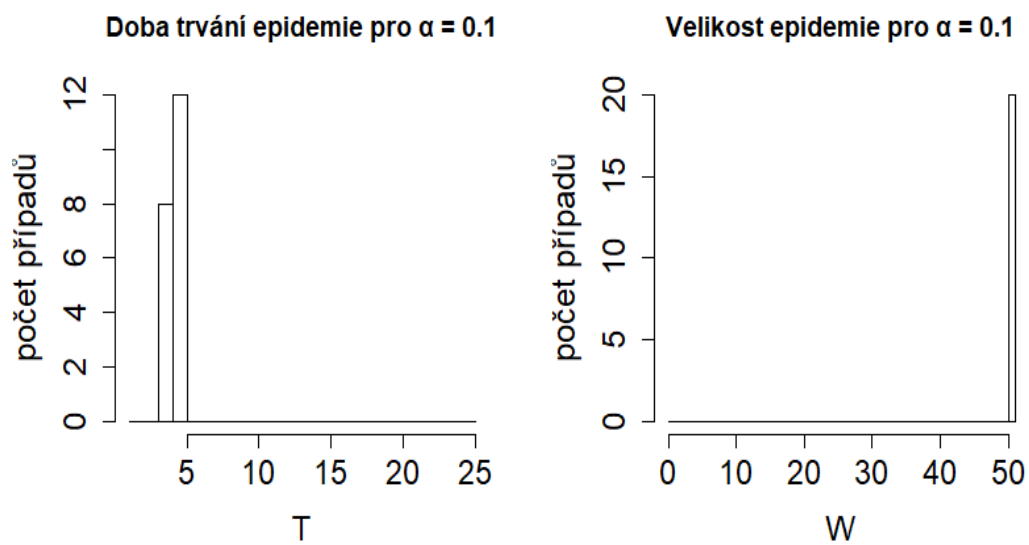
$$\begin{aligned} E[Y_t|X_0 = x_0] &= E[X_{t-1} - X_t|X_0 = x_0] \\ &= E[X_{t-1}|X_0 = x_0] - E[X_t|X_0 = x_0] \\ &= \alpha^{t-1}x_0 - \alpha^t x_0 \\ &= \alpha^{t-1}(1 - \alpha)x_0. \end{aligned}$$

Příklady: Grafy na obrázcích 3.1, 3.2, 3.3 a 3.4 vykreslují možné průběhy epidemie pro $X_0 = 50$, $Y_0 = 1$ a $\alpha = 0,1$ nebo $\alpha = 0,9$. Všechny vykreslují celkem 20 nezávislých realizací procesu.

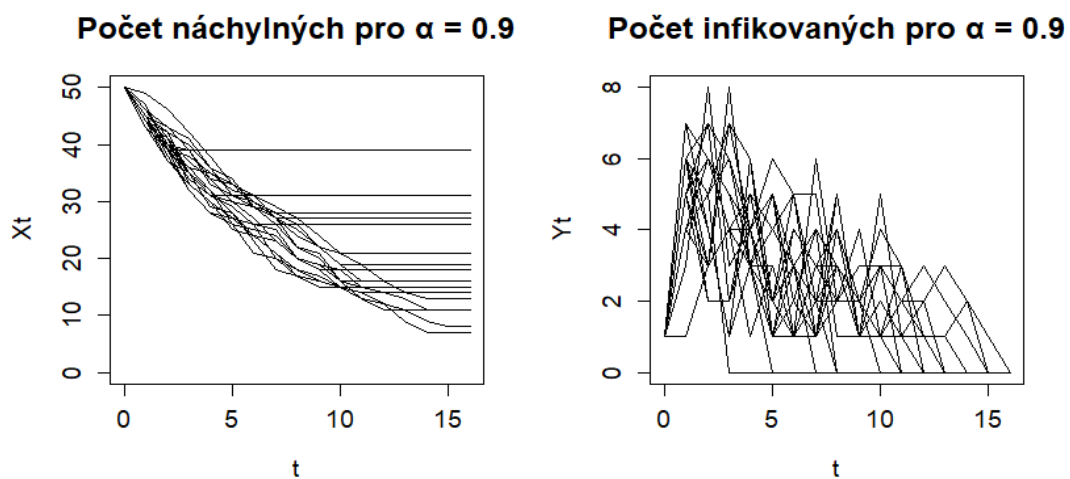
Všimneme si, že pro pravděpodobnost nepřenesení infekce $\alpha = 0,1$ skončila epidemie velmi rychle, nejčastěji po třech až čtyřech časových jednotkách. Počet infikovaných narostl nejvíce v prvním časovém úseku. Pro $\alpha = 0,9$ se epidemie šířila daleko pomaleji. V některých případech se během epidemie nakazilo z celkových 50 náchylných jedinců méně než 20 a epidemie netrvala déle než 5 časových jednotek. V jiných případech se nakazila většina populace, ale velmi pozvolna, a epidemie skončila až po 15. časovém úseku. Z toho lze usoudit, že pro velké hodnoty α je těžké předpokládat vývoj epidemie na základě Greenwoodova modelu, neboť můžou nastat příliš různé závěry, a to i co se týče předpokladu pro rozdělení náhodné veličiny W velikosti epidemie, nebo T doby trvání epidemie, jak znázorňují histogramy na obrázcích 3.2 a 3.4.



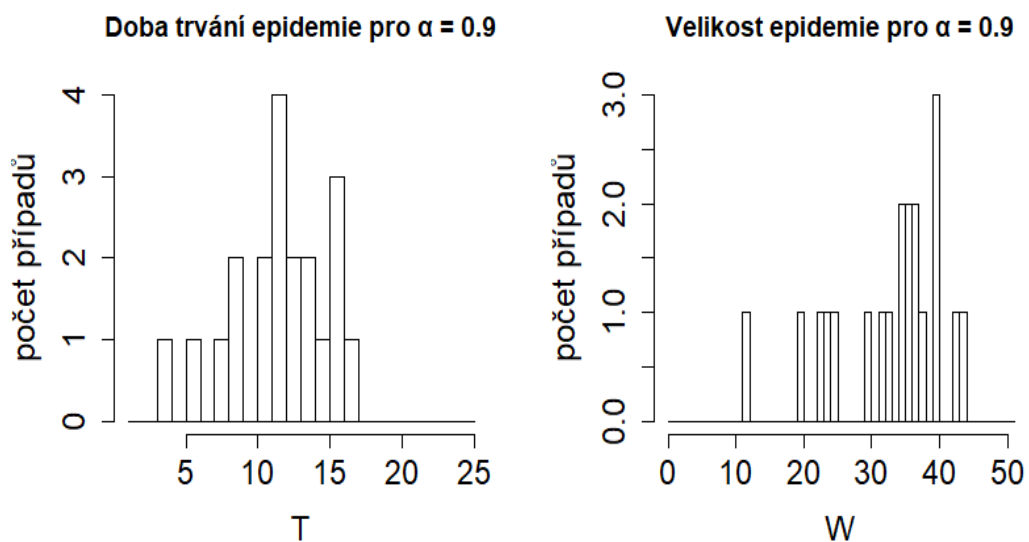
Obrázek 3.1: Simulace procesů X a Y v Greenwoodově modelu pro $\alpha = 0,1$



Obrázek 3.2: Simulace veličin T a W v Greenwoodově modelu pro $\alpha = 0,1$



Obrázek 3.3: Simulace procesů X a Y v Greenwoodově modelu pro $\alpha = 0,9$



Obrázek 3.4: Simulace veličin T a W v Greenwoodově modelu pro $\alpha = 0,9$

Popis rozdělení $(\mathbf{W}, \mathbf{T})^\top$: Pro $k, t \in \mathbb{N}$ je rozdělení náhodného vektoru $(W, T)^\top$ určeno podmíněnými pravděpodobnostmi

$$\mathbf{P}((W, T)^\top = (k, t)^\top | X_0 = i, Y_0 > 0).$$

Tuto pravděpodobnost lze snadno vyjádřit jako součin pravděpodobnosti, že v čase $(t-1)$ zbylo $(i-k)$ náchylných jedinců z původních k , a pravděpodobnosti, že v čase t bude stále $(i-k)$ náchylných, což znamená, že během přechodu mezi časy $(t-1)$ a t nepřibyl žádný nemocný, čímž jsme dosáhli celkem k jedinců, kteří se během epidemie trávající t časových jednotek nakazili. Neboli velikost epidemie je k . Matematicky formulováno platí:

$$\begin{aligned} \mathbf{P}((W, T)^\top = (k, t)^\top | X_0 = x_0, Y_0 > 0) \\ &= \mathbf{P}(X_{t-1} = x_0 - k, Y_{t-1} > 0 | X_0 = x_0) \cdot \\ &\quad \cdot \mathbf{P}(X_t = x_0 - k | X_{t-1} = x_0 - k, Y_{t-1} > 0) \\ &= \mathbf{P}(X_{t-1} = x_0 - k, Y_{t-1} > 0 | X_0 = x_0) \alpha^{x_0 - k}, \end{aligned}$$

kde α je pravděpodobnost, že se konkrétní náchylný jedinec v jedné časové jednotce nenakazí.

Označme p_{ij} pravděpodobnost přechodu ze stavu i do stavu j , dále $p_i(t)$ pravděpodobnost, že se v čase t nacházíme ve stavu i a počáteční stav označme x_0 . Potom

$$\begin{aligned} p_i(t) &= \mathbf{P}(X_t = i, Y_t > 0 | X_0 = x_0), \\ p_{ij} &= \mathbf{P}(X_{t+1} = j | X_t = i, Y_t > 0, X_0 = x_0). \end{aligned}$$

Mezi těmito pravděpodobnostmi můžeme napsat vztah

$$p_i(t) = \sum_{k=i+1}^{x_0-(t-1)} p_k(t-1) p_{ki}, \quad (3.2)$$

neboli $p_i(t)$ spočítáme jako součet pravděpodobností přes všechny stavy, které mohly nastat v předchozím čase (tj. v čase $t-1$) násobené pravděpodobnostmi přechodu z toho konkrétního stavu do stavu i . Nyní už můžeme psát, že

$$\mathbf{P}((W, T)^\top = (k, t)^\top | X_0 = x_0, Y_0 > 0) = p_{x_0-k}(t-1) \alpha^{x_0-k}.$$

Pravděpodobnost $p_{x_0-k}(t-1)$ pak spočteme rekurzivně pomocí vzorců (3.2) a (3.1).

Sdružené rozdělení vektoru $(W, T)^\top$ lze vyjádřit ještě jiným způsobem, který bude vhodný i pro vyjádření jeho rozdělení v Reedově-Frostově modelu. Nejprve vyjádříme matici pravděpodobností přechodu prvního řádu, tedy matici, která má na i -tém řádku v j -tém sloupci pro libovolné $t \in \mathbb{N}_0$ prvek $\mathbf{P}(X_{t+1} = j | X_t = i)$ s počátečním stavem x_0 . Dosazením do vzorce (3.1) dostáváme:

$$P = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 1 - \alpha & \alpha & 0 & \dots & 0 \\ (1 - \alpha)^2 & 2(1 - \alpha)\alpha & \alpha^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ (1 - \alpha)^{x_0} & x_0(1 - \alpha)^{x_0-1}\alpha & \binom{x_0}{2} (1 - \alpha)^{x_0-2}\alpha^2 & \dots & \alpha^{x_0} \end{pmatrix}. \quad (3.3)$$

Matici P rozdělíme na součet dvou matic $Q + R$, kde

$$Q = \text{diag}\{p_{jj}\}_{j=0}^{x_0}$$

je matice pravděpodobností přechodu mezi stejnými stavy a $R = P - Q$, tj.

$$R = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\ 1 - \alpha & 0 & 0 & \cdots & 0 \\ (1 - \alpha)^2 & 2(1 - \alpha)\alpha & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ (1 - \alpha)^{x_0} & x_0(1 - \alpha)^{x_0-1}\alpha & \begin{pmatrix} x_0 \\ 2 \end{pmatrix} (1 - \alpha)^{x_0-2}\alpha^2 & \cdots & 0 \end{pmatrix}$$

je striktně dolní trojúhelníková matice pravděpodobností přechodu mezi různými stavy. Tedy matice R^{t-1} je matice pravděpodobností přechodu mezi různými stavy i a j v $t - 1$ časových jednotkách.

Pravděpodobnost, že epidemie skončí v čase $T = t$ a celkový počet nakažených během epidemie je $W = x_0 - j$, pokud byl počet náchylných jedinců na počátku $X_0 = x_0$, je

$$\mathbb{P}(T = t, X_{t-1} = X_t = j | X_0 = x_0) = \{R^{t-1}\}_{x_0j} p_{jj}, \quad (3.4)$$

kde R^{t-1} je $(t - 1)$ -ní mocnina matice R , což je podle Tvzení 3 matice pravděpodobností přechodu $(t - 1)$ -ního řádu $R^{(t-1)}$. Zápis $\{R^{t-1}\}_{x_0j}$ značí prvek matice R^{t-1} v x_0 -tém řádku a j -tém sloupci, což je pravděpodobnost, že se z počátečního stavu x_0 dostaneme po $t - 1$ krocích do stavu j , aniž by někdy došlo k setrvání v některém ze stavů déle, než jednu časovou jednotku. Dále s pravděpodobností p_{jj} v čase t zůstaneme ve stavu j , což je tedy konečný stav řetězce $\{X_t\}$.

Pravděpodobnost, že epidemie skončila v čase t , pokud $X_0 = x_0$, můžeme zapsat jako

$$\begin{aligned} \mathbb{P}(T = t | X_0 = x_0) &= \sum_{j=0}^{x_0} \mathbb{P}(T = t, W = x_0 - j | X_0 = x_0) \\ &= \sum_{j=0}^{x_0} \mathbb{P}(X_t = X_{t-1} = j, Y_{t-1} > 0 | X_0 = x_0) \\ &= \sum_{j=0}^{x_0} \{R^{t-1}\}_{x_0j} p_{jj}, \end{aligned}$$

což je pravděpodobnost, že řetězec $\{X_t\}$ přecházel během prvních $t - 1$ kroků mezi různými stavy a v posledním t -tém kroku zůstal ve stejném stavu j . V posledním kroku jsme dosadili vzorec (3.4).

Tedy rozdělení náhodné veličiny T můžeme rozepsat jako

$$\mathbb{P}(T = t) = \mathbb{P}(X_0 = x_0) \sum_{j=0}^{x_0} \{R^{t-1}\}_{x_0j} \{Q\}_{jj},$$

neboť procházíme přes všechny počty náchylných jedinců j v posledním t -tém kroku. Je-li $p(0)^\top$ počáteční rozdělení X_0 (viz Definice 7), můžeme konečně vyjádřit rozdělení náhodné veličiny T jako

$$\mathbb{P}(T = t) = p(0)^\top R^{t-1} Q E,$$

pro $t = 1, 2, \dots, x_0 + 1$, kde pomocí $(x_0 + 1)$ -rozměrného vektoru $E = (1, 1, \dots, 1)^\top$ dostáváme součet všech prvků vektoru $p(0)^\top R^{t-1}Q$.

Všimneme si, že $p(0)^\top = (0, 0, \dots, 0, 1)$, neboť předpokládáme $\mathbb{P}(X_0 = x_0) = 1$. Vytvořující funkci náhodné veličiny T vyjádříme z Definice 10 jako

$$\begin{aligned}\Psi_T(\theta) &= \sum_{n=1}^{x_0+1} \mathbb{P}(T = n)\theta^n \\ &= \sum_{n=1}^{x_0+1} p(0)^\top R^{n-1}QE\theta^n \\ &= p(0)^\top \sum_{n=1}^{\infty} (R^{n-1}\theta^{n-1})\theta QE,\end{aligned}$$

tedy pomocí vzorce pro maticový součet geometrické řady dostáváme

$$\Psi_T(\theta) = p(0)^\top (I - \theta R)^{-1}\theta QE. \quad (3.5)$$

Sčítání jsme mohli rozšířit do nekonečna, neboť matice R se po své x_0 -té mocnině zcela vynuluje, tj. $R^j = 0, j \geq x_0 + 1$.

Ze vzorce vytvořující funkce můžeme spočítat některé vlastnosti rozdělení náhodné veličiny T , jako např. její střední hodnotu a rozptyl (viz Větu 4). Pro výpočet střední hodnoty potřebujeme derivaci vytvořující funkce,

$$\Psi'(\theta) = p(0)^\top (I - \theta R)^{-2}QE,$$

a dosazením $\theta = 1$ dostáváme

$$\mathbb{E} T = p(0)^\top (I - R)^{-1}E.$$

Druhá derivace vytvořující funkce je

$$\Psi''(\theta) = 2p(0)^\top R(I - \theta R)^{-3}QE,$$

a tedy

$$\begin{aligned}\text{var}(T) &= 2p(0)^\top R(I - R)^{-3}QE + p(0)^\top (I - R)^{-2}QE - (p(0)^\top (I - R)^{-1}E)^2 \\ &= p(0)^\top (2R(I - R)^{-2}E + (I - R)^{-1}E) - (p(0)^\top (I - R)^{-1}E)^2.\end{aligned}$$

Poznámka. Rozdělení náhodné veličiny T se někdy říká Markovovo geometrické rozdělení, neboť je podobné klasickému geometrickému rozdělení, jehož vytvořující funkce je tvaru $(1 - \theta p)^{-1}\theta q$, kde p je pravděpodobnost úspěchu a $q = 1 - p$ je pravděpodobnost neúspěchu. Více viz [10].

Podobným způsobem, jakým jsme vyjádřili vytvořující funkci náhodné veličiny T , spočteme vytvořující funkci velikosti epidemie W . Musíme zde brát v úvahu, že pokaždé, když se počet náchylných X_t sníží, počet infikovaných Y_{t+1} se zvýší o stejnou hodnotu.

Tentokrát

$$\mathbb{P}(W = k) = \mathbb{P}(X_0 = x_0) \sum_{t=1}^{x_0} \{R^{t-1}\}_{x_0, x_0-k} \{Q\}_{x_0-k, x_0-k}.$$

V tomto případě zahrnujeme všechny možné délky času, během nichž se z počátečního stavu x_0 náchylných dostaneme do konečného stavu $x_0 - k$ náchylných jedinců. Prvek $\{Q\}_{x_0-k, x_0-k} = p_{x_0-k, x_0-k}$ je pak pravděpodobnost, že v tomto konečném stavu zůstaneme.

Vytvořující funkce náhodné veličiny W bude tedy ve tvaru

$$\begin{aligned}\Psi_W(\varphi) &= \sum_{k=1}^{x_0} \mathbb{P}(W = k) \varphi^k \\ &= \mathbb{P}(X_0 = x_0) \sum_{k=1}^{x_0} \sum_{t=1}^{x_0} \{R^{t-1}\}_{x_0, x_0-k} \{Q\}_{x_0-k, x_0-k} \varphi^k.\end{aligned}\quad (3.6)$$

Vzorec (3.6) můžeme ještě dále rozložit na součet přes stavy, které mohou nastat, tj.

$$\mathbb{P}(X_0 = x_0) \sum_{k=1}^{x_0} \sum_{t=1}^{x_0} \sum_{m=0}^{t-2} p_{i_m, i_{m+1}} p_{x_0-k, x_0-k} \varphi^{i_m - i_{m+1}}, \quad (3.7)$$

neboť $\varphi^k = \varphi^{x_0-i_1} \varphi^{i_1-i_2} \dots \varphi^{i_{t-2}-x_0+k}$. Zde uvažujeme i_m navzájem různé stavy jednoho konkrétního řetězce, kde $i_0 = x_0, i_{t-1} = x_0 - k$ a k je stále celková velikost epidemie.

Zavedeme matici $R(\varphi)$ tvořenou prvky $p_{ij}(\varphi) = p_{ij} \varphi^{i-j}$, $i > j$, tedy

$$R(\varphi) = \begin{pmatrix} 0 & 0 & 0 & \dots & 0 \\ (1-\alpha)\varphi & 0 & 0 & \dots & 0 \\ ((1-\alpha)\varphi)^2 & 2(1-\alpha)\alpha\varphi & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ ((1-\alpha)\varphi)^{x_0} & x_0((1-\alpha)\varphi)^{x_0-1}\alpha & \begin{pmatrix} x_0 \\ 2 \end{pmatrix} ((1-\alpha)\varphi)^{x_0-2}\alpha^2 & \dots & 0 \end{pmatrix}.$$

Zřejmě platí, že $R = R(1)$.

Protože matice R je tvořena prvky $p_{i_m, i_{m+1}}$, matice $R(\varphi)$ je díky vzorci (3.7) tvořena prvky $p_{i_m, i_{m+1}} \varphi^{i_m - i_{m+1}}$. Tedy pro pevné t můžeme psát, že $\{R^{t-1}\}_{x_0, x_0-k} \varphi^k$ je totéž, jako $\{R(\varphi)^{t-1}\}_{x_0, x_0-k}$.

Dosazením do vzorce (3.7) dostáváme

$$\Psi_W(\varphi) = \mathbb{P}(X_0 = x_0) \sum_{k=1}^{x_0} \sum_{t=1}^{x_0} \{R(\varphi)^{t-1}\}_{x_0, x_0-k} \{Q\}_{x_0-k, x_0-k},$$

což je vzorec, který lze stejně jako u vzorce vytvořující funkce náhodné veličiny T snadno sečíst jako maticový součet geometrické řady, čímž dostáváme

$$\begin{aligned}\Psi_W(\varphi) &= \sum_{k=0}^{x_0} \mathbb{P}(W = k) \varphi^k \\ &= \sum_{n=1}^{x_0+1} p(0)^\top R(\varphi)^{n-1} Q E \\ &= p(0)^\top (I - R(\varphi))^{-1} Q E.\end{aligned}\quad (3.8)$$

Sdruženou vytvořující funkci náhodného vektoru $(W, T)^\top$ vyjádříme podobným způsobem a pomocí Definice 11 jako

$$\begin{aligned}
\Psi_{(W,T)}(\varphi, \theta) &= \mathbb{P}(X_0 = x_0) \sum_{k=1}^{x_0} \sum_{t=1}^{x_0} \{R^{t-1}\}_{x_0, x_0-k} \{Q\}_{x_0-k, x_0-k} \varphi^k \theta^t \\
&= \mathbb{P}(X_0 = x_0) \sum_{k=1}^{x_0} \sum_{t=1}^{x_0} \{R(\varphi)^{t-1}\}_{x_0, x_0-k} \{Q\}_{x_0-k, x_0-k} \theta^t \\
&= \sum_{t=1}^{x_0} p(0)^\top R(\varphi)^{t-1} Q E \theta^t \\
&= p(0)^\top (I - \theta R(\varphi))^{-1} \theta Q E.
\end{aligned} \tag{3.9}$$

Všimneme si, že zřejmě platí rovnost

$$\Psi_{(W,T)}(\varphi, 1) = \Psi_W(\varphi),$$

což plyne z definice sdružené vytvořující funkce, viz Definice 11.

Příklad. Aplikujeme Greenwoodův model na příklad domácnosti 4 zdravých neimunních jedinců a jednoho nakaženého člověka. Ukážeme si rovnou výpočet sdružené vytvořující funkce $\Psi_{(W,T)}(\varphi, \theta)$.

Matice $R(\varphi)$ je v tomto případě

$$R(\varphi) = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ (1-\alpha)\varphi & 0 & 0 & 0 & 0 \\ ((1-\alpha)\varphi)^2 & 2(1-\alpha)\alpha\varphi & 0 & 0 & 0 \\ ((1-\alpha)\varphi)^3 & 3((1-\alpha)\varphi)^2\alpha & 3(1-\alpha)\varphi\alpha^2 & 0 & 0 \\ ((1-\alpha)\varphi)^4 & 4((1-\alpha)\varphi)^3\alpha & 6((1-\alpha)\varphi)^2\alpha^2 & 4(1-\alpha)\varphi\alpha^3 & 0 \end{pmatrix}.$$

Pak

$$\Psi_{(W,T)}(\varphi, \theta) = (0, 0, 0, 0, 1) (I + \theta R(\varphi) + \theta^2 R(\varphi)^2 + \theta^3 R(\varphi)^3 + \theta^4 R(\varphi)^4) \theta \begin{pmatrix} 1 \\ \alpha \\ \alpha^2 \\ \alpha^3 \\ \alpha^4 \end{pmatrix}.$$

Tento výraz rozepíšeme:

$$\begin{aligned}
\Psi_{(W,T)}(\varphi, \theta) &= (0, 0, 0, 0, 1) \left\{ \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \right. \\
&+ \theta \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ (1-\alpha)\varphi & 0 & 0 & 0 & 0 \\ ((1-\alpha)\varphi)^2 & 2(1-\alpha)\alpha\varphi & 0 & 0 & 0 \\ ((1-\alpha)\varphi)^3 & 3((1-\alpha)\varphi)^2\alpha & 3(1-\alpha)\varphi\alpha^2 & 0 & 0 \\ ((1-\alpha)\varphi)^4 & 4((1-\alpha)\varphi)^3\alpha & 6((1-\alpha)\varphi)^2\alpha^2 & 4(1-\alpha)\varphi\alpha^3 & 0 \end{pmatrix} \left. \right\}
\end{aligned}$$

$$\begin{aligned}
& +\theta^2 \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 2((1-\alpha)\varphi)^2\alpha & 0 & 0 & 0 & 0 \\ 3((1-\alpha)\varphi)^3\alpha(1+\alpha) & 6((1-\alpha)\varphi)^2\alpha^3 & 0 & 0 & 0 \\ ((1-\alpha)\varphi)^4\alpha(4+6\alpha+4\alpha^2) & 12((1-\alpha)\varphi)^3\alpha^3(1+\alpha) & 12((1-\alpha)\varphi)^2\alpha^5 & 0 & 0 \end{pmatrix} \\
& +\theta^3 \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 6((1-\alpha)\varphi)^3\alpha^3 & 0 & 0 & 0 & 0 \\ 12((1-\alpha)\varphi)^4\alpha^3(1+\alpha+\alpha^2) & 24((1-\alpha)\varphi)^3\alpha^6 & 0 & 0 & 0 \end{pmatrix} \\
& +\theta^4 \left(\begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 24((1-\alpha)\varphi)^4\alpha^6 & 0 & 0 & 0 & 0 \end{pmatrix} \right) \theta \begin{pmatrix} 1 \\ \alpha \\ \alpha^2 \\ \alpha^3 \\ \alpha^4 \end{pmatrix}.
\end{aligned}$$

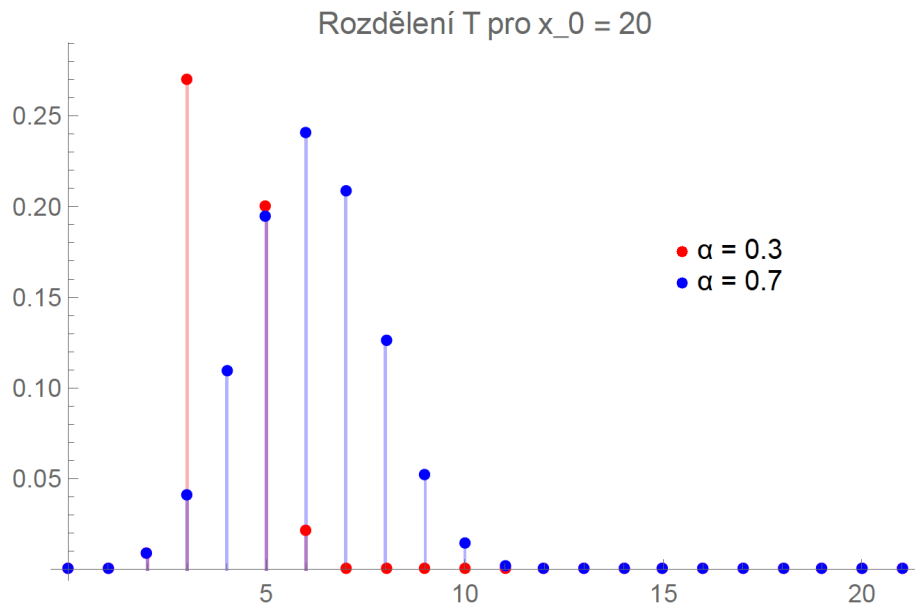
Tedy $P(W = w, T = t)$ je koeficient vytvořující funkce u proměnných φ^w a θ^t , kde $w \in \{0, 1, 2, 3, 4\}$ a $t \in \{1, 2, 3, 4, 5\}$. Jednotlivé pravděpodobnosti tedy udává Tabulka 3.1.

| $\begin{matrix} T \\ W \end{matrix}$ | 1 | 2 | 3 | 4 | 5 |
|--------------------------------------|------------|-------------------------|---|---|--------------------------|
| 0 | α^4 | 0 | 0 | 0 | 0 |
| 1 | 0 | $4(1-\alpha)\alpha^6$ | 0 | 0 | 0 |
| 2 | 0 | $6(1-\alpha)^2\alpha^4$ | $12(1-\alpha)^2\alpha^7$ | 0 | 0 |
| 3 | 0 | $4(1-\alpha)^3\alpha^2$ | $12(1-\alpha)^3\alpha^4(1+\alpha)$ | $24(1-\alpha)^3\alpha^7$ | 0 |
| 4 | 0 | $(1-\alpha)^4$ | $(1-\alpha)^4\alpha(4+6\alpha+4\alpha^2)$ | $12(1-\alpha)^4\alpha^3(1+\alpha+\alpha^2)$ | $24(1-\alpha)^4\alpha^6$ |

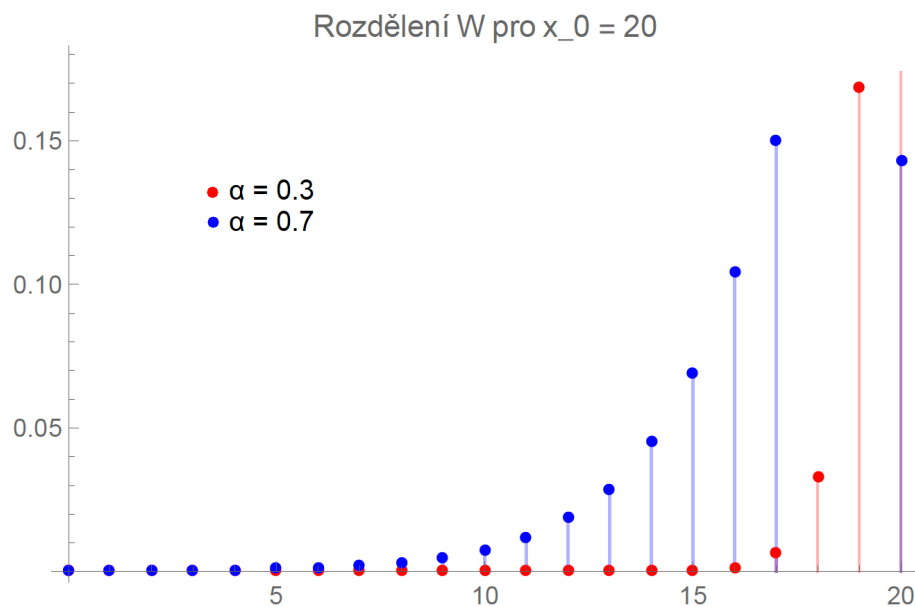
Tabulka 3.1: Rozdělení náhodného vektoru $(W, T)^\top$ pro $x_0 = 4$ a $y_0 = 1$

Příklad. Aplikujme Greenwoodův model na příklad uzavřené skupiny 20 zdravých neimunních jedinců a jednoho nakaženého člověka. Tentokrát se budeme dívat na rozdělení náhodných veličin T a W zvlášť. Vložení příslušných hodnot do vzorce (3.5) pro vytvořující funkci náhodné veličiny T vykreslíme vektor jejích koeficientů u proměnné θ , čímž z definice vytvořující funkce získáme rozdělení náhodné veličiny T . Stejný postup aplikujeme pro výpočet vytvořující funkce náhodné veličiny W dosazením do vzorce (3.8). Výsledná rozdělení zobrazují grafy na Obrázcích 3.5 a 3.6.

V Obrázku 3.6 zasahuje poslední hodnota mimo graf, neboť je její hodnota $P(W = 20) = 0,79$ příliš velká oproti ostatním.



Obrázek 3.5: Rozdělení veličiny T v Greenwoodově modelu pro $\alpha = 0,3$ a $\alpha = 0,7$



Obrázek 3.6: Rozdělení veličiny W v Greenwoodově modelu pro $\alpha = 0,3$ a $\alpha = 0,7$

3.1.3 Reedův-Frostův model

Reedův-Frostův model byl poprvé formulován americkým biostatistikem Lowellem Reedem a americkým epidemiologem Wadem Frostem roku 1928.

Hlavním předpokladem Reedova-Frostova modelu je, že vystavení se dvěma či více infikovaným lidem ve stejném čase jsou nezávislé jevy. Neboli předpokládá, že pravděpodobnost přenosu infekce na náchylného jedince závisí na počtu infikovaných jedinců. Je-li $X_t = x_t$ a $Y_t = y_t$, pak s pravděpodobností α^{y_t} bude jedinec náchylný v čase t stále náchylným i v čase $t + 1$, a proto má náhodná veličina X_{t+1} podmíněně binomické rozdělení $\mathbb{B}i(x_t, \alpha^{y_t})$ při daném $(X, Y)_t^\top$. Tudíž dostáváme vztah pro pravděpodobnost přechodu v Reedově-Frostově modelu ze stavu $(X, Y)_t^\top = (x, y)_t^\top$ do stavu $(X, Y)_{t+1}^\top = (x, y)_{t+1}^\top$, neboli podmíněnou pravděpodobnost:

$$\begin{aligned} p_{(x,y)_t, (x,y)_{t+1}} &= \mathbf{P}((X, Y)_{t+1}^\top = (x, y)_{t+1}^\top | (X, Y)_t^\top = (x, y)_t^\top) \\ &= \binom{x_t}{x_{t+1}} \alpha^{y_t x_{t+1}} (1 - \alpha^{y_t})^{x_t - x_{t+1}}, \end{aligned} \quad (3.10)$$

kde $x_t = x_{t+1} + y_{t+1}$ a $x_t > 0$. Je-li $x_t = 0$, pak epidemie končí v čase t a dále model nesledujeme. Pravděpodobnost α^{y_t} vyjadřuje, že v čase t žádný z infikovaných nenakazí konkrétního náchylného jedince.

Realizace epidemie jako dvojice $(x, y)_0, \dots, (x, y)_T$ bude tedy tzv. binomický řetězec. Tento výraz plyne z již zmíněného rozdělení náhodné veličiny X_t a tedy z výrazu (3.10) pro pravděpodobnost přechodu. Celková pravděpodobnost takovéto realizace bude součinem dílčích pravděpodobností v každém z časů $0, 1, \dots, T$, přičemž T je čas, v němž se proces zastaví. Neboli platí:

$$\mathbf{P}[(X, Y)_0^\top = (x, y)_0^\top, \dots, (X, Y)_n^\top = (x, y)_n^\top] = p_{(x,y)_0^\top} p_{(x,y)_0^\top, (x,y)_1^\top} \cdots p_{(x,y)_{n-1}^\top, (x,y)_n^\top},$$

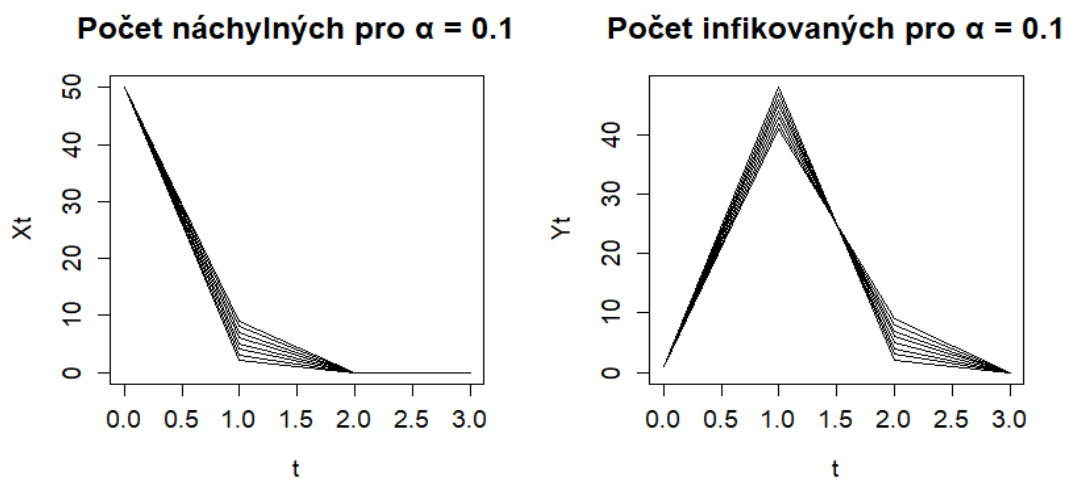
kde $p_{(x,y)_0^\top}$ je pravděpodobnost, že na počátku řetězce je $X_0 = x_0$ a $Y_0 = y_0$ a $p_{(x,y)_i^\top, (x,y)_j^\top}$ jsou pravděpodobnosti přechodu ze stavu i do stavu j (viz Definici 4 a vzorec (3.10)).

Narozdíl od Greenwoodova modelu (viz kapitolu 3.1.2) zde sledujeme v každém časovém úseku také náhodnou veličinu Y , nikoli pouze X . Zřejmě tak dostáváme dvourozměrný Markovův řetězec $\{(X, Y)_{t+1}^\top : t = 0, 1, \dots\}$ s pravděpodobnostmi přechodu $p_{(x,y)_t^\top, (x,y)_{t+1}^\top}$ danými vztahem (3.10). Ekvivalentně bychom mohli modelovat řetězec $\{(X_t, X_{t-1})^\top : t = 1, 2, \dots\}$, neboť $X_{t-1} = X_t + Y_t$, pokud dodefinujeme $X_{-1} = X_0 + Y_0$. Budeme se však držet předchozí formulace.

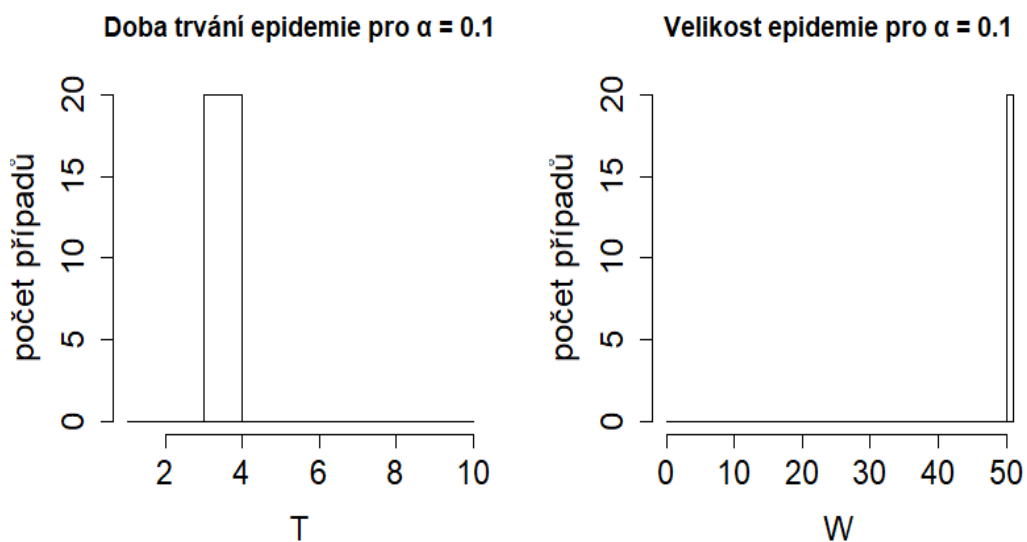
Podmíněnou střední hodnotu v Reedově-Frostově modelu vyjádříme pomocí již zmíněného vzorce pravděpodobnosti přechodu:

$$\mathbf{E} [(X, Y)_{t+1}^\top | (X, Y)_t^\top = (x, y)_t^\top] = (x_t \alpha^{y_t}, x_t (1 - \alpha^{y_t}))^\top.$$

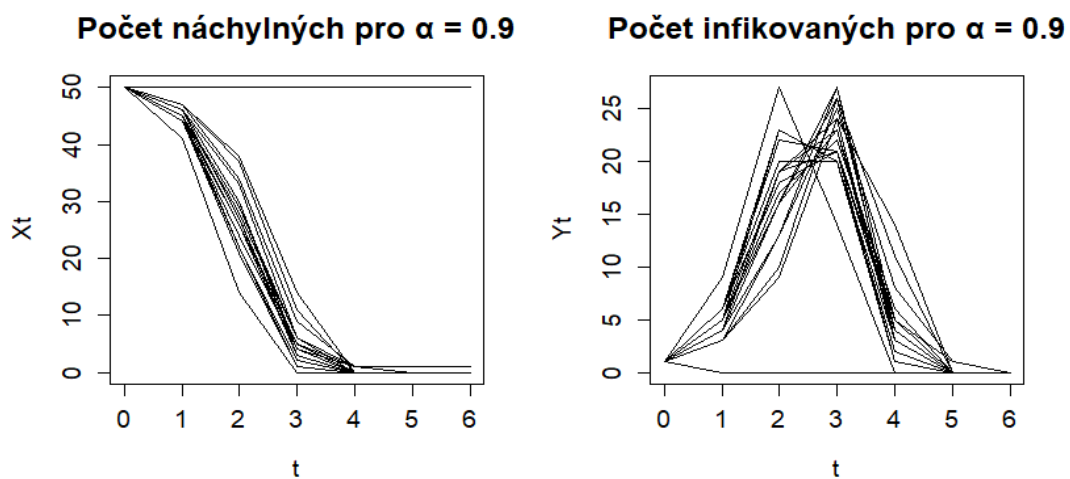
Příklad. Grafy na Obrázcích 3.7, 3.8, 3.9 a 3.10 vykreslují možné průběhy epidemie v Reedově-Frostově modelu pro $X_0 = 50$, $Y_0 = 1$ a $\alpha = 0,1$ nebo $\alpha = 0,9$. Všechny vykreslují celkem 20 nezávislých realizací procesu.



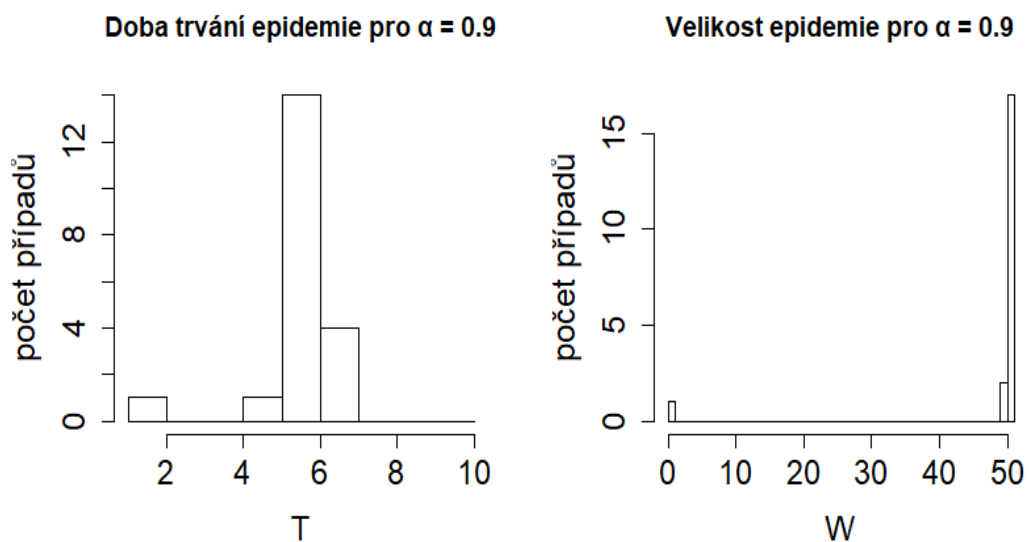
Obrázek 3.7: Simulace procesů X a Y v Reedově-Frostově modelu pro $\alpha = 0,1$



Obrázek 3.8: Simulace veličin T a W v Reedově-Frostově modelu pro $\alpha = 0,1$



Obrázek 3.9: Simulace procesů X a Y v Reedově-Frostově modelu pro $\alpha = 0,9$



Obrázek 3.10: Simulace veličin T a W v Reedově-Frostově modelu pro $\alpha = 0,9$

V příkladu pro $\alpha = 0,1$ je průběh epidemie vždy velmi podobný a rychlý. Nejčastěji končí již ve třetí časové jednotce a nakazí se většina populace (v našem případě dokonce celá populace). Realizace Reedova-Frostova modelu je pro takové α velmi podobná realizaci Greenwoodova modelu (viz Obrázek 3.1). Oproti tomu, výsledky realizace Reedova Frostova modelu pro $\alpha = 0,9$ jsou v mnoha směrech odlišné oproti Obrázku 3.3 Greenwoodova modelu. V nynějším případě nepřesáhne délka trvání epidemie sedm časových jednotek a i přes tak vysokou hodnotu α dojde ve většině případech k nakažení většiny populace. Proto není již tak těžké předpokládat velikost epidemie W ani její dobu trvání T , jak ukazuje následující histogram. I přesto se jedno pozorování úplně odprostito od chování většiny případů, kdy už v prvním kroku procesu nedošlo k přenesení infekce na žádného z náchylných, a tedy epidemie se okamžitě zastavila. Tento případ, ačkoliv je velmi zajímavý, nastane s velmi malou pravděpodobností, a to přibližně 0,52 %. Proto i nadále můžeme tvrdit, že dokážeme celkem dobře odhadnout průběh epidemie.

Popis rozdělení $(\mathbf{W}, \mathbf{T})^\top$: V Reedově-Frostově modelu není ani zdaleka jednoduché vyjádřit rozdělení náhodných veličin T a W . Rozdělení náhodného vektoru $(W, T)^\top$ se nám v Greenwoodově modelu podařilo vyjádřit pěkným vzorcem i bez využití matic R a Q . To však není případ Reedova-Frostova modelu. Příklad takového vyjádření pravděpodobnosti $\mathbb{P}(T = t)$ si čtenář může přečíst např. v článku [13].

Budeme tedy při popisu rozdělení náhodného vektoru $(W, T)^\top$ postupovat podobně, jako v Greenwoodově modelu. Matice pravděpodobností přechodu pro Reedův-Frostův model je tentokrát čtvercová bloková matice P velikosti $(x_0 + 1)^2 \times (x_0 + 1)^2$ tvaru

$$P = \begin{pmatrix} P_{00} & 0 & \cdots & 0 \\ P_{10} & P_{11} & \cdots & P_{1x_0} \\ \vdots & \vdots & \ddots & \vdots \\ P_{x_00} & P_{x_01} & \cdots & P_{x_0x_0} \end{pmatrix}, \quad (3.11)$$

kde podmatice $P_{ij} = (p_{(k,i)^\top(l,j)^\top})$ jsou čtvercové matice pravděpodobností přechodu velikosti $x_0 + 1$, tedy podle vzorce (3.10) platí

$$(p_{(k,i)^\top(l,j)^\top}) = \binom{k}{l} \alpha^{il} (1 - \alpha^i)^j,$$

pro $k, l = 0, 1, \dots, x_0$ a $j + l = k$, a jinak je rovná nule. Matice P_{ij} je nulová kromě své j -té poddiagonály, kde jsou prvky binomického rozdělení $\mathbb{B}i(k, \alpha^i)$, $j \leq k \leq x_0 + 1 - j$, neboli

$$P_{ij} = \begin{pmatrix} 0 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ (1 - \alpha^i)^j & 0 & \cdots & 0 & \cdots & 0 \\ 0 & (j + 1)\alpha^i(1 - \alpha^i)^j & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 & \cdots & 0 \\ 0 & 0 & \cdots & \begin{pmatrix} x_0 \\ x_0 - j \end{pmatrix} \alpha^{i(x_0 - j)}(1 - \alpha^i)^j & \cdots & 0 \end{pmatrix}.$$

Matici P rozdělíme na součet dvou matic $Q + R$, kde

$$Q = \begin{pmatrix} P_{00} & 0 & \cdots & 0 \\ P_{10} & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ P_{x_0 0} & 0 & \cdots & 0 \end{pmatrix},$$

je čtvercová matice velikosti $(x_0 + 1)^2 \times (x_0 + 1)^2$ tvořená pravděpodobnostmi přechodu do absorpčních stavů $\{(x, 0) : x = 0, \dots, x_0\}$ (viz Definici 9), tj. Q obsahuje prvních $x_0 + 1$ sloupců matice P a všechny ostatní prvky jsou nulové. Matice R je pak definovaná jako $P - Q$, tedy

$$R = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & P_{11} & \cdots & P_{1x_0} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & P_{x_0 1} & \cdots & P_{x_0 x_0} \end{pmatrix},$$

jejíž nenulové prvky jsou pravděpodobnosti přechodu prvního řádu do přechodných stavů (l, j) . Pak analogicky jako v Greenwoodově modelu dostáváme pravděpodobnost, že epidemie skončí v čase t , jako

$$\mathbf{P}(T = t) = p(0)^\top R^{t-1} Q E,$$

a vytvářející funkci náhodné veličiny T je opět

$$\begin{aligned} \Psi_T(\theta) &= \sum_{n=1}^{x_0+1} \mathbf{P}(T = n) \theta^n \\ &= \sum_{n=1}^{x_0+1} p(0)^\top R^{n-1} Q E \theta^n \\ &= p(0)^\top \sum_{n=1}^{\infty} (R^{n-1} \theta^{n-1}) \theta Q E \\ &= p(0)^\top (I - \theta R)^{-1} \theta Q E. \end{aligned}$$

Zde $E = (1, 1, \dots, 1)^\top$ je tentokrát vektor velikosti $(x_0 + 1)^2$ a vektor počátečního rozdělení je jednotkový vektor $p(0)^\top = (0, 0, \dots, 1, 0, \dots, 0)^\top$ s prvkem 1 na $x_0 \cdot (x_0 + 1)$ -ním místě.

Definujeme matici $R(\varphi)$ jako modifikaci matice R tak, že $R(\varphi)$ má v j -tých poddiagonálách svých blokových matic P_{ij} prvky vynásobené prvkem φ^j . Pak

stejně jako u Greenwoodova modelu můžeme psát, že

$$\Psi_W(\varphi) = p(0)^\top (I - R(\varphi))^{-1} Q E.$$

A analogickým způsobem, jakým jsme u Greenwoodova modelu popisovali sdruženou vytvořující funkci náhodného vektoru $(W, T)^\top$, píšeme

$$\Psi_{(W,T)}(\varphi, \theta) = p(0)^\top (I - \theta R(\varphi))^{-1} \theta Q E.$$

Opět platí rovnost

$$\Psi_{(W,T)}(\varphi, 1) = \Psi_W(\varphi).$$

Pozorování. Vzorce pro výpočet vytvořujících funkcí náhodných veličin T a W a náhodného vektoru $(W, T)^\top$ jsou shodné s příslušnými vzorci v Greenwoodově modelu. Liší se tedy pouze o matici pravděpodobností přechodu, a tedy o matice Q , R a $R(\varphi)$.

Příklad. Ukažme si výpočet rozdělení náhodného vektoru $(W, T)^\top$ na příkladu domácnosti o dvou náchylných neimunních jedincích a jednom infikovaném jedinci, tj. $x_0 = 2$ a $y_0 = 1$.

Matice $R(\varphi)$ je tvaru

$$R(\varphi) = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & (1 - \alpha)\varphi & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2\alpha(1 - \alpha)\varphi & 0 & (1 - \alpha)^2\varphi^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & (1 - \alpha^2)\varphi & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2\alpha^2(1 - \alpha^2)\varphi & 0 & (1 - \alpha^2)^2\varphi^2 & 0 & 0 & 0 \end{pmatrix}.$$

Pak

$$\Psi_{(W,T)}(\varphi, \theta) = (0, 0, 0, 0, 0, 1, 0, 0, 0) (I + \theta R(\varphi) + \theta^2 R(\varphi)^2) \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ \alpha \\ \alpha^2 \\ 1 \\ \alpha^2 \\ \alpha^4 \end{pmatrix}.$$

Tento výraz rozepíšeme:

$$\Psi_{(W,T)}(\varphi, \theta) = (0, 0, 0, 0, 0, 1, 0, 0, 0) \left\{ I \right.$$

$$\begin{aligned}
& +\theta \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & (1-\alpha)\varphi & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2\alpha(1-\alpha)\varphi & 0 & (1-\alpha)^2\varphi^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & (1-\alpha^2)\varphi & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2\alpha^2(1-\alpha^2)\varphi & 0 & (1-\alpha^2)^2\varphi^2 & 0 & 0 \end{pmatrix} \\
& +\theta^2 \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2\alpha(1-\alpha)^2\varphi^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2\alpha^2(1-\alpha)(1-\alpha^2)\varphi^2 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \left. \vphantom{\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2\alpha(1-\alpha)^2\varphi^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2\alpha^2(1-\alpha)(1-\alpha^2)\varphi^2 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}} \right\} \theta \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ \alpha \\ \alpha^2 \\ 1 \\ \alpha^2 \\ \alpha^4 \end{pmatrix}.
\end{aligned}$$

Z tohoto výrazu zjistíme následující čtyři pravděpodobnosti:

$$\begin{aligned}
\mathbf{P}(T = 1, W = 0) &= \alpha^2, \\
\mathbf{P}(T = 2, W = 1) &= 2\alpha^2(1 - \alpha), \\
\mathbf{P}(T = 2, W = 2) &= (1 - \alpha)^2, \\
\mathbf{P}(T = 3, W = 2) &= 2\alpha(1 - \alpha)^2.
\end{aligned}$$

Ostatní pravděpodobnosti jsou rovny 0.

4. Odhad parametru α

V praxi je důležité umět odhadnout parametry popsaných modelů a vědět, jak dobře se shodují dostupná data s matematickými modely. Tento fakt totiž pomáhá vylepšit přesnost předpokladů o chování epidemie na základě modelů. V této části se proto zaměříme na principy odhadu parametru α a probereme možnosti, jak aplikovat teoretické poznatky matematického modelování v praxi tak, abychom dokázali epidemii ovládat.

Chceme-li umět předpovědět chování začínající epidemie, potřebujeme znát pravděpodobnost α u dané nemoci. Označme jako v předchozích kapitolách α pravděpodobnost nepřenesení infekce.

Postup výpočtu maximálně věrohodného odhadu si ukážeme na dvou možných případech. Ty se liší podle toho, kolik informace o epidemii máme. V prvním případě známe pouze velikost epidemie W každé z pozorovaných domácností, tj. kolik celkem jedinců v domácnosti bylo během epidemie nakaženo. Ve druhém případě máme navíc informaci o tom, jak se nemoc šířila, neboli z každé pozorované domácnosti známe řetězec $\{X_t : t \in \{0, 1, \dots, T\}\}$.

Pro ověření dostatečné shody modelu s reálnými daty použijeme χ^2 -test dobré shody pro multinomická data, viz [2]. Testujeme hypotézu shody napozorovaných dat s Greenwoodovým, resp. Reedovým-Frostovým modelem pro nějaké α proti alternativě jejich neshody. Proti platnosti nulové hypotézy svědčí velké hodnoty testové statistiky, přesněji

$$H_0 \text{ zamítáme} \Leftrightarrow \chi^2 > \chi_{K-d-1}^2,$$

kde χ^2 je testová statistika, K je počet kategorií a d je počet odhadovaných parametrů.

4.1 Odhad podle znalosti W

Mějme data, která vyjadřují, kolik členů z pozorovaných domácností bylo během epidemie celkově nakaženo. Předpokládejme, že všechny pozorované domácnosti mají na počátku epidemie stejný počet infikovaných a stejný počet neimunizovaných jedinců. Tabulka 4.1 naznačuje, jak tato data vypadají.

| w | četnost |
|-------|-----------|
| 0 | n_0 |
| 1 | n_1 |
| 2 | n_2 |
| ... | ... |
| x_0 | n_{x_0} |

Tabulka 4.1: Obecná data o znalosti velikosti epidemie

Z kapitol 3.1.2 a 3.1.3 víme, jak spočítat pravděpodobnosti $P(W = w)$ pro $w = 0, 1, \dots, x_0$ pomocí vytvářející funkce. Díky této znalosti vyjádříme věrohodnostní funkci jako

$$L = \prod_{w=0}^{x_0} P(W = w)^{n_w},$$

z čehož zlogaritmováním a nalezením maxima vyjádříme maximálně věrohodný odhad $\hat{\alpha}$ parametru α .

Testová statistika je rovna

$$\chi^2 = \sum_{w=0}^{x_0} \frac{(n_w - n\hat{p}_w)^2}{n\hat{p}_w},$$

kde n je stále počet pozorovaných domácností, n_w vyjadřuje počet domácností, které dosáhly velikosti epidemie w a \hat{p}_w jsou rovny pravděpodobnostem $P(W = w)$ po dosažení odhadu $\hat{\alpha}$. Test budeme vždy provádět na hladině 0,05.

V následující části si ukážeme postup výpočtu pro konkrétní velikosti domácností.

Tříčlenné domácnosti: Předpokládejme, že jsme pozorovali n tříčlenných domácností, z nichž byl vždy jeden počáteční nakažený a dva náchylní jedinci. Tabulka 4.2 naznačuje obecně, jak tato data vypadají.

| | Předpokládaný počet domácností | | |
|-----|--------------------------------|--------------------------------|-------|
| w | Greenwood | Reed-Frost | n_w |
| 0 | $n\alpha^2$ | $n\alpha^2$ | a |
| 1 | $2n(1 - \alpha)\alpha^2$ | $2n(1 - \alpha)\alpha^2$ | b |
| 2 | $n(1 - \alpha)^2(1 + 2\alpha)$ | $n(1 - \alpha)^2(1 + 2\alpha)$ | c |

Tabulka 4.2: Obecná data – tříčlenná domácnost, $y_0 = 1$

První sloupec udává, kolik ze dvou jedinců domácnosti se během epidemie nakazili od jednoho počátečního infikovaného. Druhý a třetí sloupec vyjadřují předpokládaný počet domácností, které dosáhly celkem w počtu nakažených. Je vidět, že pro tak málo uvažovaných náchylných jedinců se Greenwoodův model zcela shoduje s Reedovým-Frostovým. To je zřejmé, neboť nikdy nemůže v tomto případě domácnosti nastat, že by v jednom časovém kroku byl ještě nějaký náchylný jedinec a zároveň byl více než jeden infikovaný. V posledním sloupci jsou napozorovaná data, tedy kolik domácností z celkových n dosáhlo w počtu nakažených, kde $n = a + b + c$.

Věrohodnost L z pozorovaných dat pro oba modely můžeme psát, jako

$$\begin{aligned} L &= \prod_{w=0}^2 P(W = w)^{n_w} \\ &= \alpha^{2a} [2(1 - \alpha)\alpha^2]^b [(1 - \alpha)^2(1 + 2\alpha)]^c. \end{aligned}$$

Maximalizací logaritmu věrohodnosti L získáme maximálně věrohodný odhad parametru α . Tedy

$$\begin{aligned} l = \log L &= 2a \log \alpha + b \log(2\alpha^2(1 - \alpha)) + c \log((1 - \alpha)^2(2\alpha + 1)) \\ &= [2a + 2b] \log \alpha + [b + 2c] \log(1 - \alpha) + c \log(2\alpha + 1) + b \log 2, \end{aligned}$$

což zderivujeme na

$$l' = \frac{2a + 2b}{\alpha} - \frac{b + 2c}{1 - \alpha} + \frac{2c}{1 + 2\alpha}.$$

Tento výraz položíme rovno 0, čímž dostáváme kvadratickou rovnici pro α :

$$\alpha^2(-4a - 6b - 6c) + \alpha(2a + b) + 2(a + b) = 0. \quad (4.1)$$

Dosazením napozorovaných hodnot a , b , c a vyřešením kvadratické rovnice získáváme odhad $\hat{\alpha}$.

Čtyřčlenné domácnosti: Ukažme si nyní stejný postup na trochu složitější situaci. Předpokládejme, že jsme pozorovali n čtyřčlenných domácností, z nichž byl jeden počáteční nakažený a tři náchylní jedinci. Zde už se budou lišit pravděpodobnosti pro Greenwoodův a Reedův-Frostův model.

| | Předpokládaný počet domácností | | |
|---|--|--|-------|
| w | Greenwood | Reed-Frost | n_w |
| 0 | $n\alpha^3$ | $n\alpha^3$ | a |
| 1 | $3n(1 - \alpha)\alpha^4$ | $3n(1 - \alpha)\alpha^4$ | b |
| 2 | $3n(1 - \alpha)^2\alpha^2(1 + 2\alpha^2)$ | $3n(1 - \alpha)^2\alpha^3(1 + 2\alpha)$ | c |
| 3 | $n(1 - \alpha)^3(1 + 3\alpha + 3\alpha^2 + 6\alpha^3)$ | $n(1 - \alpha)^3(1 + 3\alpha + 6\alpha^2 + 6\alpha^3)$ | d |

Tabulka 4.3: Obecná data – čtyřčlenná domácnost, $y_0 = 1$

Věrohodnost L_G z pozorovaných dat pro Greenwoodův model můžeme psát, jako

$$\begin{aligned} L_G &= \prod_{w=0}^3 \mathbb{P}(W = w)^{n_w} \\ &= \alpha^{3a} [3(1 - \alpha)\alpha^4]^b [3(1 - \alpha)^2\alpha^2(1 + 2\alpha^2)]^c [(1 - \alpha)^3(1 + 3\alpha + 3\alpha^2 + 6\alpha^3)]^d. \end{aligned}$$

Maximalizací logaritmu věrohodnosti L_G získáme maximálně věrohodný odhad parametru α . Tedy

$$\begin{aligned} l_G = \log L_G &= 3a \log \alpha + b \log(3(1 - \alpha)\alpha^4) + c \log(3(1 - \alpha)^2\alpha^2(1 + 2\alpha^2)) \\ &\quad + d \log((1 - \alpha)^3(1 + 3\alpha + 3\alpha^2 + 6\alpha^3)) \\ &= (3a + 4b + 2c) \log \alpha + (b + 2c + 3d) \log(1 - \alpha) \\ &\quad + c \log(1 + 2\alpha^2) + d \log(1 + 3\alpha + 3\alpha^2 + 6\alpha^3), \end{aligned}$$

což zderivujeme na

$$l'_G = \frac{3a + 4b + 2c}{\alpha} - \frac{b + 2c + 3d}{1 - \alpha} + \frac{4c\alpha}{1 + 2\alpha^2} + \frac{3d(1 + 2\alpha + 6\alpha^2)}{1 + 3\alpha + 3\alpha^2 + 6\alpha^3}.$$

Výraz položíme roven 0, z čehož dostáváme

$$\begin{aligned} &\alpha^6[-36d - 24c - 12B - 12A] + \alpha^5[24d + 12c - 6B + 6A] \\ &+ \alpha^4[-12d - 12B - 6A] + \alpha^3[18d + 8c - 5B + 7A] \\ &+ \alpha^2[3d + 4c - 3B + 2A] + \alpha[3d - B + 2A] + A = 0, \end{aligned} \quad (4.2)$$

kde

$$\begin{aligned} A &= 3a + 4b + 2c, \\ B &= b + 2c + 3d. \end{aligned}$$

Dosazením napozorovaných hodnot a, b, c, d a vyřešením rovnice bychom získali odhad $\hat{\alpha}_G$.

Použitím stejného postupu pro Reedův-Frostův model získáme věrohodnost L_{RF} jako

$$L_{RF} = \alpha^{3a}[3(1-\alpha)\alpha^4]^b[3(1-\alpha)^2\alpha^3(1+2\alpha)]^c[(1-\alpha)^3(1+3\alpha+6\alpha^2+6\alpha^3)]^d,$$

a tedy zlogaritmováním a zderivováním dostáváme

$$l'_{RF} = \frac{3a+4b+2c}{\alpha} - \frac{b+2c+3d}{1-\alpha} + \frac{2c\alpha}{1+2\alpha} + \frac{3d(1+4\alpha+6\alpha^2)}{1+3\alpha+3\alpha^2+6\alpha^3}.$$

Výraz položíme roven 0, z čehož dostáváme

$$\begin{aligned} & \alpha^6[-12c] + \alpha^5[-36d+6c-12\tilde{B}] \\ & + \alpha^4[-6d-12\tilde{B}] + \alpha^3[24d+4c-9\tilde{B}] \\ & + \alpha^2[15d+2c-5\tilde{B}-2\tilde{A}] + \alpha[3d-\tilde{B}+\tilde{A}] + \tilde{A} = 0, \end{aligned} \quad (4.3)$$

kde

$$\begin{aligned} \tilde{A} &= 3a+4b+2c, \\ \tilde{B} &= b+2c+3d. \end{aligned}$$

Dosazením napozorovaných hodnot a, b, c, d a vyřešením rovnice bychom získali odhad $\hat{\alpha}_{RF}$.

Pětičlenné domácnosti: Ukažme si tento postup pro pětičlenné domácnosti o čtyřech náchylných neimunních jedincích a jednom počátečním infikovaným. Tabulka 4.4 uvádí, jak tato obecná data vypadají.

| w | Předpokládaný počet domácností | | n_w |
|---|--|---|-------|
| | Greenwood | Reed-Frost | |
| 0 | $n\alpha^4$ | $n\alpha^4$ | a |
| 1 | $4n(1-\alpha)\alpha^6$ | $4n(1-\alpha)\alpha^6$ | b |
| 2 | $6n(1-\alpha)^2\alpha^4(1+2\alpha^3)$ | $6n(1-\alpha)^2\alpha^6(1+2\alpha)$ | c |
| 3 | $4n(1-\alpha)^3\alpha^2(1+3\alpha^2+3\alpha^3+6\alpha^5)$ | $4n(1-\alpha)^2\alpha^4(1+2\alpha+3\alpha^2-6\alpha^4)$ | d |
| 4 | $n(1-\alpha)^4(1+4\alpha+6\alpha^2+16\alpha^3+12\alpha^4+12\alpha^5+24\alpha^6)$ | $n(1-\alpha)^2(1+2\alpha+3\alpha^2+4\alpha^3-4\alpha^5-18\alpha^6-12\alpha^7+24\alpha^8)$ | e |

Tabulka 4.4: Obecná data – pětičlenná domácnost, $y_0 = 1$

Pomocí tabulky vyjádříme věrohodnost L_G jako

$$\begin{aligned} L_G &= \alpha^{4a}[4(1-\alpha)\alpha^6]^b[6(1-\alpha)^2\alpha^4(1+2\alpha^3)]^c \\ & \cdot [4(1-\alpha)^3\alpha^2(1+3\alpha^2+3\alpha^3+6\alpha^5)]^d \\ & \cdot [(1-\alpha)^4(1+4\alpha+6\alpha^2+16\alpha^3+12\alpha^4+12\alpha^5+24\alpha^6)]^e, \end{aligned}$$

tedy

$$l'_G = \frac{4a + 6b + 4c + 2d}{\alpha} - \frac{b + 2c + 3d + 4e}{1 - \alpha} + \frac{6c\alpha^2}{1 + 2\alpha^3} + \frac{d(6\alpha + 9\alpha^2 + 30\alpha^4)}{1 + 3\alpha^2 + 3\alpha^3 + 6\alpha^5} + \frac{e(4 + 12\alpha + 48\alpha^2 + 48\alpha^3 + 60\alpha^4 + 144\alpha^5)}{1 + 4\alpha + 6\alpha^2 + 16\alpha^3 + 12\alpha^4 + 12\alpha^5 + 24\alpha^6}. \quad (4.4)$$

Položíme rovno nule a dosazením do polynomu a vyřešením příslušné rovnice získáváme odhad $\hat{\alpha}_G$.

Po Reedův-Frostův model bychom stejným postupem dostali

$$l'_{RF} = \frac{4a + 6b + 6c + 4d}{\alpha} - \frac{b + 2c + 2d + 2e}{1 - \alpha} + \frac{2c}{1 + 2\alpha} + \frac{d(2 + 6\alpha - 24\alpha^3)}{1 + 2\alpha + 3\alpha^2 - 6\alpha^4} + \frac{e(2 + 6\alpha + 12\alpha^2 - 20\alpha^4 - 108\alpha^5 - 84\alpha^6 + 193\alpha^7)}{1 + 2\alpha + 3\alpha^2 + 4\alpha^3 - 4\alpha^5 - 18\alpha^6 - 12\alpha^7 + 24\alpha^8}. \quad (4.5)$$

Položíme rovno nule a vyřešením příslušné polynomiální rovnice získáváme odhad $\hat{\alpha}_{RF}$.

4.2 Odhad podle znalosti řetězce

Známe-li podrobnější průběhy epidemií v jednotlivých domácnostech, můžeme pravděpodobnost α odhadnout ještě přesněji. Budeme tedy sledovat, kolik rodin prošlo jednotlivými různými možnými realizacemi řetězce počtu náchylných jedinců. Neboli, jak přesně se z pohledu počtu náchylných jedinců šířila nemoc v domácnosti.

Předpokládejme, že všechny pozorované domácnosti mají na počátku epidemie stejný počet infikovaných a stejný počet neimunních jedinců. Tabulka 4.5 naznačuje, jak tato data vypadají.

| Řetězec | Četnost |
|---|---------------|
| $x_0 - x_0$ | n_1 |
| $x_0 - (x_0 - 1) - (x_0 - 1)$ | n_2 |
| $x_0 - (x_0 - 1) - (x_0 - 2) - (x_0 - 2)$ | n_3 |
| ... | ... |
| $x_0 - 0 - 0$ | $n_{2^{x_0}}$ |

Tabulka 4.5: Podrobná data o znalosti řetězce šíření epidemie

Z kapitol 3.1.2 a 3.1.3 víme, jak spočítat pravděpodobnost, že nastane konkrétní řetězec $\{X_t\}$. Díky této znalosti vyjádříme věrohodnostní funkci jako

$$L = \prod_{k=1}^{2^{x_0}} \mathbb{P}(\text{nastane } k\text{-tý řetězec})^{n_k},$$

z čehož zlogaritmováním a nalezením maxima vyjádříme maximálně věrohodný odhad $\hat{\alpha}$ parametru α .

Testová statistika je rovna

$$\chi^2 = \sum_{k=1}^{2^{x_0}} \frac{(n_k - n\hat{p}_k)^2}{n\hat{p}_k},$$

kde n je stále počet pozorovaných domácností, n_k je vyjadřuje počet domácností, u nichž epidemie probíhala řetězcem k , kde řetězců je celkem 2^{x_0} možných, tj. $k = 1, 2, \dots, 2^{x_0}$, a \hat{p}_k jsou pravděpodobnosti, že nastane právě řetězec k po dosažení odhadu $\hat{\alpha}$. Test budeme provádět vždy na hladině 0,05.

V následující části si ukážeme postup výpočtu pro konkrétní velikosti domácností.

Tříčlenné domácnosti: Předpokládejme, že jsme pozorovali n tříčlenných domácností, z nichž byl vždy jeden počáteční nakažený a dva náchylní jedinci. Tabulka 4.6 udává obecně, jak tato data vypadají. Čísla v řetězci značí počet náchylných jedinců v časové jednotce. Například řetězec $2-1-1$ značí, že v prvním časovém kroku se nakazil jeden člen domácnosti a ve druhém už nikdo. Opět se pro tak malé množství náchylných jedinců shoduje Greenwoodův a Reedův-Frostův model.

| Řetězec | Předpokládaný počet domácností | Data |
|-----------|--------------------------------|------|
| $2-2$ | $n\alpha^2$ | a |
| $2-1-1$ | $2n(1-\alpha)\alpha^2$ | b |
| $2-1-0-0$ | $2n(1-\alpha)^2\alpha$ | c |
| $2-0-0$ | $n(1-\alpha)^2$ | d |

Tabulka 4.6: Podrobná data – tříčlenná domácnost, $y_0 = 1$

Sestavme věrohodnostní rovnici stejným postupem, jako v předchozím případě.

$$L = \alpha^{2a}[2\alpha^2(1-\alpha)]^b[2\alpha(1-\alpha)^2]^c[(1-\alpha)^2]^d.$$

Dále zlogaritmuje:

$$l = \log L = (2a + 2b + c) \log \alpha + (b + 2c + 2d) \log(1 - \alpha) + (b + c) \log 2,$$

a zderivujeme:

$$l' = \frac{2a + 2b + c}{\alpha} - \frac{b + 2c + 2d}{1 - \alpha}.$$

Položíme výraz roven 0 a dostáváme

$$\hat{\alpha} = \frac{2a + 2b + c}{2a + 3b + 3c + 2d}. \quad (4.6)$$

Čtyřčlenné domácnosti: Zkusme tento postup aplikovat na data o čtyřčlenných domácnostech s jedním počátečním infikovaným. V tabulce 4.7 jsou rozepsané stejně jako v předchozí části kapitoly předpokládané počty domácností, u nichž epidemie probíhala tím konkrétním způsobem, který naznačuje první sloupec tabulky. Čísla v řetězci značí počet náchylných jedinců v časové jednotce.

Sestavme věrohodnostní rovnice pro oba modely. Začneme s Greenwoodovým:

$$L_G = \alpha^{3a}[3\alpha^4(1-\alpha)]^b[6\alpha^4(1-\alpha)^2]^c[3\alpha^2(1-\alpha)^2]^d \cdot [6\alpha^3(1-\alpha)^3]^e[3\alpha^2(1-\alpha)^3]^f[3\alpha(1-\alpha)^3]^g[1-\alpha]^{3h}.$$

| Řetězec | Předpokládaný počet domácností | | Data |
|-------------------|--------------------------------|--------------------------------------|------|
| | Greenwood | Reed-Frost | |
| 3 – 3 | $n\alpha^3$ | $n\alpha^3$ | a |
| 3 – 2 – 2 | $3n(1 - \alpha)\alpha^4$ | $3n(1 - \alpha)\alpha^4$ | b |
| 3 – 2 – 1 – 1 | $6n(1 - \alpha)^2\alpha^4$ | $6n(1 - \alpha)^2\alpha^4$ | c |
| 3 – 1 – 1 | $3n(1 - \alpha)^2\alpha^2$ | $3n(1 - \alpha)^2\alpha^3$ | d |
| 3 – 2 – 1 – 0 – 0 | $6n(1 - \alpha)^3\alpha^3$ | $6n(1 - \alpha)^3\alpha^3$ | e |
| 3 – 2 – 0 – 0 | $3n(1 - \alpha)^3\alpha^2$ | $3n(1 - \alpha)^3\alpha^2$ | f |
| 3 – 1 – 0 – 0 | $3n(1 - \alpha)^3\alpha$ | $3n(1 - \alpha)^3\alpha(1 + \alpha)$ | g |
| 3 – 0 – 0 | $n(1 - \alpha)^3$ | $n(1 - \alpha)^3$ | h |

Tabulka 4.7: Podrobná data – čtyřčlenná domácnost, $y_0 = 1$

Pak

$$l_G = \log L_G = (3a + 4b + 4c + 2d + 3e + 2f + g) \log \alpha + (b + 2c + 2d + 3e + 3f + 3g + 3h) \log(1 - \alpha).$$

Zderivováním výrazu podle α dostaneme

$$l'_G = \frac{A}{\alpha} + \frac{B}{1 - \alpha},$$

kde

$$A = 3a + 4b + 4c + 2d + 3e + 2f + g$$

a

$$B = b + 2c + 2d + 3e + 3f + 3g + 3h.$$

Položíme $l'_G = 0$ a dostaneme

$$(1 - \alpha)A - \alpha B = 0.$$

Odtud pak

$$\hat{\alpha}_G = \frac{A}{A + B} = \frac{3a + 4b + 4c + 2d + 3e + 2f + g}{3a + 5b + 6c + 4d + 6e + 5f + 4g + 3h}. \quad (4.7)$$

Podobně pro Reedův-Frostův model bychom dostali:

$$l'_{RF} = \frac{3a + 4b + 4c + 3d + 3e + 2f + g}{\alpha} + \frac{g}{1 + \alpha} - \frac{b + 2c + 2d + 3e + 3f + 3g + 3h}{1 - \alpha},$$

$$l'_{RF} = \frac{\tilde{A}}{\alpha} + \frac{g}{1 + \alpha} + \frac{\tilde{B}}{1 - \alpha},$$

kde

$$\tilde{A} = 3a + 4b + 4c + 3d + 3e + 2f + g$$

a

$$\tilde{B} = b + 2c + 2d + 3e + 3f + 3g + 3h.$$

Položíme $l'_{RF} = 0$ a vyřešením této rovnice po dosazení shromážděných dat bychom dostali výraz pro $\hat{\alpha}_{RF}$ jako

$$\hat{\alpha}_{RF} = \frac{\tilde{B} - g \pm \sqrt{(g - \tilde{B})^2 + 4\tilde{A}(\tilde{A} + \tilde{B} + g)}}{-2(\tilde{A} + \tilde{B} + g)}. \quad (4.8)$$

Pětičlenné domácnosti: Opět si ukážeme ten samý postup, tentokrát pro pětičlennou domácnost s jedním počátečním infikovaným.

| Řetězec | Předpokládaný počet domácností | | Data |
|-----------------------|--------------------------------|--|----------|
| | Greenwood | Reed-Frost | |
| 4 – 4 | $n\alpha^4$ | $n\alpha^4$ | <i>a</i> |
| 4 – 3 – 3 | $4n(1 - \alpha)\alpha^6$ | $4n(1 - \alpha)\alpha^6$ | <i>b</i> |
| 4 – 3 – 2 – 2 | $12n(1 - \alpha)^2\alpha^7$ | $12n(1 - \alpha)^2\alpha^7$ | <i>c</i> |
| 4 – 3 – 2 – 1 – 1 | $24n(1 - \alpha)^3\alpha^7$ | $24n(1 - \alpha)^3\alpha^7$ | <i>d</i> |
| 4 – 3 – 2 – 1 – 0 – 0 | $24n(1 - \alpha)^4\alpha^6$ | $24n(1 - \alpha)^4\alpha^6$ | <i>e</i> |
| 4 – 3 – 2 – 0 – 0 | $12n(1 - \alpha)^4\alpha^5$ | $12n(1 - \alpha)^4\alpha^5$ | <i>f</i> |
| 4 – 3 – 1 – 1 | $12n(1 - \alpha)^3\alpha^5$ | $12n(1 - \alpha)^3\alpha^6$ | <i>g</i> |
| 4 – 3 – 1 – 0 – 0 | $12n(1 - \alpha)^4\alpha^4$ | $12n\alpha^4(1 - \alpha)^3(1 - \alpha^2)$ | <i>h</i> |
| 4 – 3 – 0 – 0 | $4n(1 - \alpha)^4\alpha^3$ | $4n(1 - \alpha)^4\alpha^3$ | <i>i</i> |
| 4 – 2 – 2 | $6n(1 - \alpha)^2\alpha^4$ | $6n(1 - \alpha)^2\alpha^6$ | <i>j</i> |
| 4 – 2 – 1 – 1 | $12n(1 - \alpha)^3\alpha^4$ | $12n\alpha^5(1 - \alpha)^2(1 - \alpha^2)$ | <i>k</i> |
| 4 – 2 – 1 – 0 – 0 | $12n(1 - \alpha)^4\alpha^3$ | $12n\alpha^4(1 - \alpha)^3(1 - \alpha^2)$ | <i>l</i> |
| 4 – 2 – 0 – 0 | $6n(1 - \alpha)^4\alpha^2$ | $6n\alpha^2(1 - \alpha)^2(1 - \alpha^2)^2$ | <i>m</i> |
| 4 – 1 – 1 | $4n(1 - \alpha)^3\alpha^2$ | $4n\alpha^4(1 - \alpha)^3$ | <i>o</i> |
| 4 – 1 – 0 – 0 | $4n(1 - \alpha)^4\alpha$ | $4n\alpha(1 - \alpha)^3(1 - \alpha^3)$ | <i>p</i> |
| 4 – 0 – 0 | $n(1 - \alpha)^4$ | $n(1 - \alpha)^4$ | <i>q</i> |

Tabulka 4.8: Podrobná data – pětičlenná domácnost, $y_0 = 1$

Sestavíme věrohodnostní rovnice. Začneme s Greenwoodovým:

$$\begin{aligned}
L_G = & \alpha^{4a}[4\alpha^6(1 - \alpha)]^b[12\alpha^7(1 - \alpha)^2]^c[24\alpha^7(1 - \alpha)^3]^d[24\alpha^6(1 - \alpha)^4]^e \\
& \cdot [12\alpha^5(1 - \alpha)^4]^f[12\alpha^5(1 - \alpha)^3]^g[12\alpha^4(1 - \alpha)^4]^h[4\alpha^3(1 - \alpha)^4]^i \\
& \cdot [6\alpha^4(1 - \alpha)^2]^j[12\alpha^4(1 - \alpha)^3]^k[12\alpha^3(1 - \alpha)^4]^l[6\alpha^2(1 - \alpha)^4]^m \\
& \cdot [4\alpha^2(1 - \alpha)^3]^o[4\alpha(1 - \alpha)^4]^p[(1 - \alpha)^4]^q.
\end{aligned}$$

Tedy

$$\begin{aligned}
l'_G = & \frac{4a + 6b + 7c + 7d + 6e + 5f + 5g + 4h + 3i + 4j + 4k + 3l + 2m + 2o + p}{\alpha} \\
& - \frac{b + 2c + 3d + 4e + 4f + 3g + 4h + 4i + 2j + 3k + 4l + 4m + 3o + 4p + 4q}{1 - \alpha} \\
= & \frac{A}{\alpha} - \frac{B}{1 - \alpha}.
\end{aligned}$$

Výraz položíme roven 0 a získáváme tak maximálně věrohodný odhad

$$\hat{\alpha}_G = \frac{A}{A + B}, \tag{4.9}$$

kde

$$A = 4a + 6b + 7c + 7d + 6e + 5f + 5g + 4h + 3i + 4j + 4k + 3l + 2m + 2o + p$$

a

$$B = b + 2c + 3d + 4e + 4f + 3g + 4h + 4i + 2j + 3k + 4l + 4m + 3o + 4p + 4q.$$

Postupujeme analogicky pro Reedův-Frostův model. Věrohodnostní funkce je

$$L_{RF} = \alpha^{4a} [4\alpha^6(1-\alpha)]^b [12\alpha^7(1-\alpha)^2]^c [24\alpha^7(1-\alpha)^3]^d [24\alpha^6(1-\alpha)^4]^e \cdot [12\alpha^5(1-\alpha)^4]^f [12\alpha^6(1-\alpha)^3]^g [12\alpha^4(1-\alpha)^3(1-\alpha^2)]^h [4\alpha^3(1-\alpha)^4]^i \cdot [6\alpha^6(1-\alpha)^2]^j [12\alpha^5(1-\alpha)^2(1-\alpha^2)]^k [12\alpha^4(1-\alpha)^3(1-\alpha^2)]^l \cdot [6\alpha^2(1-\alpha)^2(1-\alpha^2)^2]^m [4\alpha^4(1-\alpha)^3]^o [4\alpha(1-\alpha)^3(1-\alpha^3)]^p [(1-\alpha)^4]^q,$$

z níž spočteme

$$l'_{RF} = \frac{4a + 6b + 7c + 7d + 6e + 5f + 6g + 4h + 3i + 6j + 5k + 4l + 2m + 4o + p}{\alpha} \cdot \frac{b + 2c + 3d + 4e + 4f + 3g + 3h + 4i + 2j + 2k + 3l + 2m + 3o + 3p + 4q}{1 - \alpha} - \frac{2\alpha(h + k + l + 2m)}{1 - \alpha^2} - \frac{3p\alpha^2}{1 - \alpha^3} = \frac{\tilde{A}}{\alpha} - \frac{\tilde{B}}{1 - \alpha} - \frac{2\tilde{C}\alpha}{1 - \alpha^2} - \frac{3p\alpha^2}{1 - \alpha^3}.$$

Výraz položíme roven 0 a získáváme tak rovnici

$$\alpha^5[\tilde{A} + \tilde{B} + 2\tilde{C} + 3p] + \alpha^4\tilde{B} - \alpha^3[\tilde{A} + 3p] - \alpha^2[\tilde{A} + \tilde{B} + 2\tilde{C}] - \alpha\tilde{B} + \tilde{A} = 0, \quad (4.10)$$

kde

$$\begin{aligned} \tilde{A} &= 4a + 6b + 7c + 7d + 6e + 5f + 6g + 4h + 3i + 6j + 5k + 4l + 2m + 4o + p, \\ \tilde{B} &= b + 2c + 3d + 4e + 4f + 3g + 3h + 4i + 2j + 2k + 3l + 2m + 3o + 3p + 4q, \\ \tilde{C} &= h + k + l + 2m. \end{aligned}$$

Po dosazení příslušných hodnot a po vyřešení rovnice získáme odhad $\hat{\alpha}_{RF}$.

5. Aplikace na data

5.1 Data o spalničkách

Nyní si ukážeme dvě možnosti odhadu parametru α metodou maximální věrohodnosti na skutečných datech o spalničkách z Providence RI, která jsem čerpala z knih [7] a [3]. V obou uvedených příkladech uvažujeme s jedním z domácnosti počátečním infikovaným.

Tříčlenné domácnosti: Mějme celkem $n = 334$ pozorovaných tříčlenných domácností.

| w | Předpokládaný počet domácností | n_w |
|---|--------------------------------|-------|
| 0 | 24,7 | 34 |
| 1 | 36,0 | 25 |
| 2 | 273,3 | 275 |

Tabulka 5.1: Obecná data o spalničkách pro tříčlenné domácnosti

Dosažením hodnot $a = 34$, $b = 25$, $c = 275$ do vzorce (4.1) dostáváme $\hat{\alpha} = 0,272$ odhad α pro Greenwoodův i Reedův-Frostův model. Druhý sloupec tabulky 5.1 jsou pak odhadované počty domácností po dosažení spočteného odhadu.

Pro ověření dostatečné shody modelu s reálnými daty použijeme χ^2 -test dobré shody pro multinomická data na hladině 0,05. Do vzorce pro výpočet testové statistiky dosadíme np_w , což jsou hodnoty ve druhém sloupci Tabulky 5.1 a Z_w jsou rovny získaným datům, tj. třetímu sloupci Tabulky 5.1.

Dosažením příslušných hodnot dostáváme hodnotu testové statistiky.

$$\chi^2 = \frac{(34 - 24,7)^2}{24,7} + \frac{(25 - 36,0)^2}{36,0} + \frac{(275 - 273,3)^2}{273,3} = 6,87,$$

což je rozhodně vyšší než kvantil χ^2 rozdělení o jednom stupni volnosti, tj.

$$\chi^2 > \chi_1^2(0,950) = 3,8,$$

tedy zamítáme nulovou hypotézu. To znamená, že test dobré shody vyhodnotil Reedův-Frostův i Greenwoodův model jako nedostatečně shodné s pozorovanými daty.

Nyní použijme stejná data o spalničkách z Providence RI, tentokrát využijeme podrobnější data popsaná v Tabulce 5.2.

| Řetězec | Předpokládaný počet domácností | Data |
|---------------|--------------------------------|------|
| 2 - 2 | 14,9 | 34 |
| 2 - 1 - 1 | 23,5 | 25 |
| 2 - 1 - 0 - 0 | 87,7 | 36 |
| 2 - 0 - 0 | 207,9 | 239 |

Tabulka 5.2: Podrobná data o spalničkách pro tříčlenné domácnosti

Po dosazení napozorovaných dat do vzorce (4.6) pro odhad α metodou maximální věrohodnosti dostáváme $\hat{\alpha} = 0,211$. Druhý sloupec tabulky 5.2 jsou pak odhadované počty domácností po dosazení spočteného odhadu.

Opět ověříme χ^2 -testem dobré shody pro multinomická data, jestli se naše modely s odhadem $\hat{\alpha}$ dostatečně shodují s realitou. Testová statistika je ve tvaru

$$\chi^2 = \frac{(14,9 - 34)^2}{14,9} + \frac{(23,5 - 25)^2}{23,5} + \frac{(87,7 - 36)^2}{87,7} + \frac{(207,9 - 239)^2}{207,9} = 37,79.$$

Tedy testová statistika vyšla příliš velká oproti kvantilu χ^2 -rozdělení o dvou stupních volnosti, který je roven $\chi_2^2(0,950) = 6,0$. To znamená, že i nyní zamítáme nulovou hypotézu a můžeme tvrdit, že se Greenwoodův ani Reedův-Frostův model dostatečně neshodují s napozorovanými daty.

Čtyřčlenné domácnosti: Provedeme nyní to samé na datech z Providence RI o celkem $n = 100$ pozorovaných, tentokrát čtyřčlenných domácnostech zaznamenaných v Tabulce 5.3.

| | Předpokládaný počet domácností | | |
|---|--------------------------------|------------|-------|
| w | Greenwood | Reed-Frost | n_w |
| 0 | 2,5 | 4,2 | 4 |
| 1 | 1,5 | 2,8 | 3 |
| 2 | 14,9 | 9,0 | 9 |
| 3 | 81,1 | 84,0 | 84 |

Tabulka 5.3: Obecná data o spalničkách pro čtyřčlenné domácnosti

Dosazením hodnot $a = 4$, $b = 3$, $c = 9$, $d = 84$ do vzorce (4.2) dostáváme $\hat{\alpha}_G = 0,291$ odhad α pro Greenwoodův model a dosazením do vzorce (4.3) dostáváme $\hat{\alpha}_{RF} = 0,347$ pro Reedův-Frostův model.

Pro ověření dostatečné shody modelu s reálnými daty použijeme opět χ^2 -test dobré shody pro multinomická data. Za platnosti H_0 má testová statistika asymptotické rozdělení $\chi_{K-d-1}^2 = \chi_{K-2}^2$. Protože se jedná o asymptotickou aproximaci, je potřeba, aby byl počet pozorování každé kategorie dostatečně velký. Jedním takovým orientačním pravidlem je požadovat alespoň 5 očekávaných pozorování každé kategorie. Jelikož $P(Z_w = 0)$ a $P(Z_w = 1)$ jsou příliš malé a v našem případě obě kategorie předpokládají méně než 5 pozorování, sloučíme je dohromady při výpočtu testové statistiky. Neboli sloučíme 1. a 2. řádek Tabulky 5.3.

Dosazením příslušných hodnot vypočítáme testovou statistiku jako

$$\chi_G^2 = \frac{(7 - 4)^2}{4} + \frac{(9 - 14,9)^2}{14,9} + \frac{(84 - 81,1)^2}{81,1} = 4,69,$$

tedy

$$\chi_G^2 > \chi_1^2(0,950) = 3,8.$$

To znamená, že v tomto případě zamítáme nulovou hypotézu, tedy tvrdíme, že Greenwoodův model není dostatečně shodný s našimi daty.

Pro Reedův-Frostův model je hodnota testové statistiky rovna

$$\chi_{RF}^2 = \frac{(6,9959 - 7)^2}{6,9959} + \frac{(9,0328 - 9)^2}{9,0328} + \frac{(83,9713 - 84)^2}{83,9713} = 1,3 \cdot 10^{-4},$$

tedy v tomto případě test vyhodnotil Reedův-Frostův model jako velmi dobrý model pro napozorovaná data.

Nyní použijeme stejná data o spalničkách z Providence RI, tentokrát využijeme podrobnější data, která jsou popsána v tabulce 5.4.

| | Předpokládaný počet domácností | | |
|-------------------|--------------------------------|------------|------|
| Řetězec | Greenwood | Reed-Frost | Data |
| 3 – 3 | 0,9 | 1,2 | 4 |
| 3 – 2 – 2 | 0,4 | 0,7 | 3 |
| 3 – 2 – 1 – 1 | 0,7 | 1,0 | 1 |
| 3 – 1 – 1 | 8,2 | 2,2 | 8 |
| 3 – 2 – 1 – 0 – 0 | 2,7 | 3,4 | 4 |
| 3 – 2 – 0 – 0 | 6,5 | 7,3 | 3 |
| 3 – 1 – 0 – 0 | 31,0 | 38,7 | 10 |
| 3 – 0 – 0 | 49,6 | 45,5 | 67 |

Tabulka 5.4: Podrobná data o spalničkách pro čtyřčlenné domácnosti

Po dosažení napozorovaných dat do vzorců (4.7) a (4.8) pro odhad α metodou maximální věrohodnosti dostáváme $\hat{\alpha}_G = 0,209$ a $\hat{\alpha}_{RF} = 0,231$.

Stejným způsobem jako v předchozí části kapitoly ověříme χ^2 -testem dobré shody pro multinomická data, jestli jsou naše odhady dostatečně správné. Pro Greenwoodův model je dobré sloučit řetězce v řádcích 1, 2, 3 a 5 tabulky 5.4, neboť pravděpodobnosti, že takové řetězce nastanou, jsou příliš malé oproti zbylým pozorováním. Pak

$$\chi_G^2 = \frac{(4,7 - 12)^2}{4,7} + \frac{(8,2 - 8)^2}{8,2} + \frac{(6,5 - 3)^2}{6,5} + \frac{(31,0 - 10)^2}{31,0} + \frac{(49,6 - 67)^2}{49,6} = 33,56.$$

Pro výpočet testové statistiky Reedova-Frostova modelu tentokrát sloučíme řádky 1, 2, 3, 4 a 5 tabulky 5.4 ze stejného důvodu, jako předtím. Dostáváme

$$\chi_{RF}^2 = \frac{(8,5 - 20)^2}{8,5} + \frac{(7,3 - 3)^2}{7,3} + \frac{(38,7 - 10)^2}{38,7} + \frac{(45,5 - 67)^2}{45,5} = 49,5.$$

V obou případech vyšla testová statistika příliš velká oproti kvantilu χ^2 -rozdělení, Konkrétně $\chi_G^2 > \chi_3^2(0,950) = 7,8$ a i $\chi_{RF}^2 > \chi_2^2(0,950) = 6,0$. Tedy zamítáme nulovou hypotézu shody modelů s reálnými daty a můžeme tvrdit, že pro tento případ neodpovídají naše modely dostatečně přesně.

5.2 Data o nachlazení

Pětičlenné domácnosti: Máme k dispozici data o nachlazení po pozorování 664 pětičlenných domácností, v nichž byl na začátku epidemie vždy jeden nakažený a čtyři zdraví jedinci. Data jsou čerpána z [7] a [3]. Informace o velikosti epidemie v domácnostech je zaznamenána v Tabulce 5.5.

| | Předpokládaný počet domácností | | |
|---|--------------------------------|------------|-------|
| w | Greenwood | Reed-Frost | n_w |
| 0 | 399,3 | 404,4 | 423 |
| 1 | 147,9 | 147,2 | 131 |
| 2 | 80,8 | 71,2 | 60 |
| 3 | 30,0 | 32,0 | 36 |
| 4 | 6,0 | 9,2 | 14 |

Tabulka 5.5: Obecná data o nachlazení pro pětičlennou domácnost

Dosazením do vzorců 4.4 a 4.5 a vyřešením rovnic dostáváme maximálně věrohodné odhady parametru α jako $\hat{\alpha}_G = 0,881$ a $\hat{\alpha}_{RF} = 0,883$. Dosazením těchto hodnot do Tabulky 4.4 dostáváme druhý a třetí sloupec Tabulky 5.5.

Spočtíme testovou statistiku pro Greenwoodův model pro tento případ.

$$\chi_G^2 = \frac{(399,3 - 423)^2}{399,3} + \frac{(147,9 - 131)^2}{147,9} + \frac{(80,8 - 60)^2}{80,8} + \frac{(30,0 - 36)^2}{30,0} + \frac{(6,0 - 14)^2}{6,0} = 20,56,$$

což je rozhodně vyšší hodnota než $\chi_3^2(0,950) = 7,8$, tedy zamítáme nulovou hypotézu a tvrdíme, že model není dostatečně přesný pro napozorovaná data.

Nyní spočtíme testovou statistiku pro Reedův-Frostův model této situace.

$$\chi_{RF}^2 = \frac{(404,4 - 423)^2}{404,4} + \frac{(147,2 - 131)^2}{147,2} + \frac{(71,2 - 60)^2}{71,2} + \frac{(32,0 - 36)^2}{32,0} + \frac{(9,2 - 14)^2}{9,2} = 7,40,$$

tudíž v tomto případě je testová statistika nižší než $\chi_3^2(0,950) = 7,8$. Můžeme tedy tvrdit, že Reedův-Frostův model je přijatelným modelem epidemie pro tato data.

Využijme nyní stejná, tentokrát podrobnější data o nachlazení uvedena v Tabulce 5.6 a spočítejme odhad pravděpodobnosti α analogicky jako v části 5.1.

Dosazením napozorovaných hodnot do vzorce (4.9) dostáváme odhad $\hat{\alpha}_G = 0,883$ a dosazením do vzorce (4.10) dostáváme odhad $\hat{\alpha}_{RF} = 0,884$. Tyto hodnoty dosadíme do Tabulky 4.8, čímž dostáváme druhý a třetí sloupec Tabulky 5.6.

Pro výpočet testové statistiky pro Greenwoodův model sloučíme řádky 5, 6, 8, 9, 12, 13, 14, 15 a 16, neboť ve všech těchto řádcích je očekávaná hodnota počtu pozorování menší než pět. Pak

$$\chi_G^2 = \frac{(403,3 - 423)^2}{403,3} + \frac{(147,3 - 131)^2}{147,3} + \frac{(45,7 - 36)^2}{45,7} + \frac{(10,7 - 14)^2}{10,7} + \frac{(8,4 - 17)^2}{8,4} + \frac{(5,4 - 8)^2}{5,4} + \frac{(33,2 - 24)^2}{33,2} + \frac{(7,8 - 11)^2}{7,8} = 19,76,$$

což je větší než kvantil $\chi_6^2(0,950) = 12,6$. Tedy zamítáme nulovou hypotézu a tvrdíme, že náš model není v souladu s napozorovanými daty.

| Řetězec | Předpokládaný počet domácností | | Data |
|-----------------------|--------------------------------|------------|------|
| | Greenwood | Reed-Frost | |
| 4 – 4 | 403,3 | 405,2 | 423 |
| 4 – 3 – 3 | 147,3 | 133,5 | 131 |
| 4 – 3 – 2 – 2 | 45,7 | 40,2 | 36 |
| 4 – 3 – 2 – 1 – 1 | 10,7 | 10,4 | 14 |
| 4 – 3 – 2 – 1 – 0 – 0 | 1,4 | 1,8 | 4 |
| 4 – 3 – 2 – 0 – 0 | 0,6 | 1,1 | 2 |
| 4 – 3 – 1 – 1 | 5,4 | 6,1 | 8 |
| 4 – 3 – 1 – 0 – 0 | 0,7 | 2,4 | 2 |
| 4 – 3 – 0 – 0 | 0,3 | 0,5 | 2 |
| 4 – 2 – 2 | 33,2 | 23,2 | 24 |
| 4 – 2 – 1 – 1 | 7,8 | 13,3 | 11 |
| 4 – 2 – 1 – 0 – 0 | 1,0 | 2,4 | 3 |
| 4 – 2 – 0 – 0 | 0,6 | 3,2 | 1 |
| 4 – 1 – 1 | 3,3 | 2,8 | 3 |
| 4 – 1 – 0 – 0 | 0,4 | 2,2 | 0 |
| 4 – 0 – 0 | 0,1 | 0,2 | 0 |

Tabulka 5.6: Podrobná data o nachlazení pro pětičlennou domácnost

Pro výpočet testové statistiky pro Reedův-Frostův model sloučíme opět ty stejné řádky 5, 6, 8, 9, 12, 13, 14, 15 a 16, neboť ve všech těchto řádcích je očekávaná hodnota počtu pozorování menší než pět. Pak

$$\chi_{RF}^2 = \frac{(405,2 - 423)^2}{405,2} + \frac{(133,5 - 131)^2}{133,5} + \frac{(40,2 - 36)^2}{40,2} + \frac{(10,7 - 14)^2}{10,4} + \frac{(16,6 - 17)^2}{16,6} + \frac{(6,1 - 8)^2}{6,1} + \frac{(23,2 - 24)^2}{23,2} + \frac{(13,3 - 11)^2}{13,3} = 3,26,$$

což je rozhodně nižší než kvantil $\chi_6^2(0,950) = 12,6$. Tedy v tomto případě jsme vyhodnotili Reedův-Frostův model jako dostatečně shodný s napozorovanými daty.

5.3 Aplikace a pozorování

V kapitolách 5.1 a 5.2 se nám podařilo nalézt shodu modelu s reálnými daty pouze ve třech případech, a to pro Reedův-Frostův model při znalosti velikosti epidemie pro čtyřčlenné domácnosti, pro Reedův-Frostův model pro znalost velikosti epidemie pro pětičlenné domácnosti a opět Reedův-Frostův model pro znalost řetězce pro pětičlenné domácnosti. Tedy Greenwoodův model nebyl dostatečně přesný pro žádná z uvedených dat. Výsledky testování ve většině případů vyvracely dobrou shodu modelů s daty, někdy vyšla testová statistika χ^2 -testu dobré shody dokonce šestkrát vyšší, než příslušný kvantil χ^2 -rozdělení.

Můžeme diskutovat o více důvodech, proč se někdy naše modely neshodují s reálnými daty v tak velké míře. Jedním z nich je, že se imunita jedinců může v každé domácnosti lišit. Dalším je vliv okolí, který může způsobit, že některý z náchylných jedinců bude nakažen mimo domácnost, což odporuje předpokladu uzavřenosti zkoumané populace. Pro alternativní možnosti odhadování parametru α viz [7]. Výsledkem jeho práce byl prokazatelně přesnější odhad parametru α .

5.3.1 Kontrola epidemie

Praktickým využitím matematického modelování je zjistit, jakým způsobem je vhodné upravit předpoklady, abychom dosáhli lepších výsledků. Modelujeme-li šíření infekčních chorob, snažíme se o zabránění šíření infekce mezi jedinci tak, aby se epidemie dotkla co nejnižšího počtu lidí. Můžeme se snažit i o úplný opak, chceme-li například kontrolovat nadměrnou populaci nějakého živočišného druhu a chceme tak dosáhnout co největšího počtu infikovaných, nebo přímo určitého množství.

Epidemii lze kontrolovat třemi základními způsoby:

- vzděláváním a šířením informace o infekci,
- imunizací (např. očkováním),
- izolací infekčních jedinců.

Vliv povědomí o infekčních nemocech na jejich šíření v populaci je nejnázornější pro AIDS či žloutenku. Díky kampaním, např. STOP-AIDS z roku 1987, které vzdělávají širokou veřejnost o možnostech nákazy, se prokazatelně snížil počet nových případů. Více se podobnými vlivy vzdělávání zabývali ve svých studiích [11] nebo [12].

Imunizace je dlouho používaná metoda pro kontrolu šíření infekce. Konkrétním příkladem jsou pravé neštovice, u nichž nastalo prokazatelné snížení výskytu a počtu úmrtí po vynalezení spolehlivé vakcinace na počátku 19. století. Během druhé poloviny 20. století se Světová zdravotnická organizace pokusila o úplné vymýcení pravých neštovic pomocí povinné vakcinace, čehož dosáhla roku 1977, kdy je znám poslední případ výskytu této nemoci. Proto se v dnešní době proti pravým neštovicím již neočkuje. Více o tomto projektu se může čtenář dozvědět z podrobné dokumentace Světové zdravotnické organizace, viz [9].

Metoda izolace infekčních jedinců je nejen používaná v každodenním životě, kdy od zdravé populace izolujeme infekční osoby. Je především hojně používaná ve věznicích, kde dochází k častému šíření HIV mezi vězni. Krom pohlavního styku je to kvůli používání stejných jehel pro intravenózní užívání drog. Více si čtenář může zjistit ve studii [6] a o matematických metodách izolace v knize [7].

Závěr

V této práci bylo podrobně vysvětleno vytváření Greenwoodova a Reedova-Frostova stochastického modelu pro popis průběhu epidemie některých infekčních nemocí v uzavřené populaci. Řetězec počtu náchylných a počtu infikovaných jedinců byl vytvářen přímo pomocí znalosti pravděpodobnosti přechodu mezi stavy. Jiný pohled na modelování nám umožnilo zkoumat rozdělení náhodných veličin velikosti a doby trvání epidemie. Teoretické poznatky byly aplikovány na příkladech, které čtenáři naznačily, ve kterých případech mohou být tyto modely spolehlivým popisem reálné situace a ve kterých je příliš velké riziko odlišnosti. Později byl ilustrován postup odhadování parametru pravděpodobnosti nepřenesezení infekce, který byl použit na reálných datech o spalničkách a nachlazení v různých početných domácnostech.

Vzhledem k velmi zjednodušujícím předpokladům Greenwoodova a Reedova-Frostova modelu se mnohdy data uvedená v této práci neshodovala s těmito stochastickými modely. Vylepšením odhadu parametru pravděpodobnosti nepřenesezení infekce se zabývalo více vědců, viz [4], [5] a [15]. Norman T. J. Bailey publikoval své výsledky s daty o spalničkách (viz kapitola 5.1), v nichž dokázal vylepšit odhad pravděpodobnosti nepřenosu infekce tím, že ji uvažoval jako náhodnou veličinu s beta rozdělením. Tento nápad pak uplatňovali další matematici ve svých výzkumech.

Práce by mohla být rozšířena o hlubší zkoumání vlivu různorodosti pozorovaných skupin z hlediska jejich odlišné imunity či délky infekčního období na samotné modelování, neboť v popisovaných modelech předpokládáme jejich homogenitu. Jak je zmíněno v kapitole 5.3.1, praktickým využitím matematických modelů je vědět, jak epidemii ovlivnit různými prostředky, abychom docílili stanovených výsledků. Tedy výzkum vlivu vzdělávání, imunizace a izolace na chování modelů by byl další možný způsob rozšíření této práce o zajímavé souvislosti mezi teorií a praxí modelování epidemií.

Tato práce slouží čtenáři k pochopení základních postupů při vytváření dvou SIR modelů epidemií. Tyto modely jsou základem složitějších stochastických modelů s diskrétním časem, jako jsou např. MSIR, SEIR, či MSEIR modely. Ty kromě fází S (susceptible), I (immune) a R (removed) uvažují též fázi M (maternally-derived immunity), což je fáze pasivní imunity dítěte po jeho narození, či fázi E (exposed) značící fázi latentního období nemoci. Pomocí těchto modifikací mohou tyto modely popisovat přesnější vývoj epidemie v populaci.

Seznam použité literatury

- [1] ABBEY, H. (1952) *An exmination of the Reed-Frost theory of epidemics*, *Human Biology*, 24(3), pp. 201–233.
- [2] ANĎĚL, J. (2005) *Základy matematické statistiky*. První vydání. Matfyzpress, Praha. ISBN 80-86732-40-1.
- [3] BAILEY, N. T. J. (1953) *The use of chain-binomials with a variable chance of infection for the analysis of intra-household epidemics*, *Biometrics*, 40(3), pp. 279–286.
- [4] BAILEY, N. T. J. (1975) *The mathematical theory of infectious diseases and its applications*. Druhé vydání. Charles Griffin & Co., London. ISBN 0852642318.
- [5] BECKER, Niels G. (1981) *A general chain binomial model for infectious diseases*, *Biometrics*, 37(2), pp. 251–258.
- [6] BREWER, T. F., VLAHOV, D., TAYLOR, E., HALL, D., MUNOZ, A. a POLK, F. (1988) *Transmission of HIV1 within a statewide prison system*, *AIDS 2*, pp. 363–367.
- [7] DALEY, D. J. a GANI, J. (1999) *Epidemic modelling: An introduction*. První vydání. Cambridge University Press, Cambridge. ISBN 978-0-521-64079-4.
- [8] EN'KO, P. D. (1889) *On the course of epidemics of some infectious diseases*. *International Journal of Epidemiology*, 18, pp. 749–755.
- [9] FENNER, F., HANDERSON, D. A., ARITA, I., JEŽEK, Z., LADNYI, I. D. (1988) *Smallpox and its eradication*. První vydání. Office of Publications, WHO, Geneva, Switzerland. ISBN 9241561106.
- [10] GANI, J. a JERWOOD, D. (1971) *Markov chain methods in chain binomial epidemic models*. *Biometrics*, 27(3), pp. 591–603.
- [11] KAPLAN, E. H. (1995) *Probability models of needle exchange*, *Operations Research*, 43(4), pp. 558–569.
- [12] O'NEILL, P. D. (1995) *Epidemic models featuring behaviour change*, *Advances in Applied Probability*, 27(4), pp. 960–979.
- [13] O'NEILL, P. D. (2003) *Perfect simulation for Reed-Frost epidemic models*, *Statistics and Computing*, 13(2), pp. 37–44.
- [14] PRÁŠKOVÁ, Z. a LACHOUT, P. (2012) *Základy náhodných procesů I*. Druhé vydání. MatfyzPress, Praha. ISBN 978-80-7378-210-8.
- [15] SCHENZLE, D. (1982) *Problems in drawing epidemiological inferences by fitting epidemic chain models to lumped data*, *Biometrics*, 38, pp. 843–847.
- [16] TUCKWELL, H. C. a WILLIAMS, R. J. (2007) *Some properties of a simple stochastic epidemic model of SIR type*. *Mathematical Biosciences*, 208(1), pp. 76–78.

Seznam obrázků

| | | | |
|------|--|-------|----|
| 3.1 | Simulace procesů X a Y v Greenwoodově modelu pro $\alpha = 0,1$ | . . . | 11 |
| 3.2 | Simulace veličin T a W v Greenwoodově modelu pro $\alpha = 0,1$ | . . . | 11 |
| 3.3 | Simulace procesů X a Y v Greenwoodově modelu pro $\alpha = 0,9$ | . . . | 12 |
| 3.4 | Simulace veličin T a W v Greenwoodově modelu pro $\alpha = 0,9$ | . . . | 12 |
| 3.5 | Rozdělení veličiny T v Greenwoodově modelu pro $\alpha = 0,3$ a $\alpha = 0,7$ | | 19 |
| 3.6 | Rozdělení veličiny W v Greenwoodově modelu pro $\alpha = 0,3$ a $\alpha = 0,7$ | | 19 |
| 3.7 | Simulace procesů X a Y v Reedově-Frostově modelu pro $\alpha = 0,1$ | . | 21 |
| 3.8 | Simulace veličin T a W v Reedově-Frostově modelu pro $\alpha = 0,1$ | . | 21 |
| 3.9 | Simulace procesů X a Y v Reedově-Frostově modelu pro $\alpha = 0,9$ | . | 22 |
| 3.10 | Simulace veličin T a W v Reedově-Frostově modelu pro $\alpha = 0,9$ | . | 22 |

Seznam tabulek

| | | |
|-----|---|----|
| 3.1 | Rozdělení náhodného vektoru $(W, T)^\top$ pro $x_0 = 4$ a $y_0 = 1$ | 18 |
| 4.1 | Obecná data o znalosti velikosti epidemie | 27 |
| 4.2 | Obecná data – tříčlenná domácnost, $y_0 = 1$ | 28 |
| 4.3 | Obecná data – čtyřčlenná domácnost, $y_0 = 1$ | 29 |
| 4.4 | Obecná data – pětičlenná domácnost, $y_0 = 1$ | 30 |
| 4.5 | Podrobná data o znalosti řetězce šíření epidemie | 31 |
| 4.6 | Podrobná data – tříčlenná domácnost, $y_0 = 1$ | 32 |
| 4.7 | Podrobná data – čtyřčlenná domácnost, $y_0 = 1$ | 33 |
| 4.8 | Podrobná data – pětičlenná domácnost, $y_0 = 1$ | 34 |
| 5.1 | Obecná data o spalničkách pro tříčlenné domácnosti | 36 |
| 5.2 | Podrobná data o spaničkách pro tříčlenné domácnosti | 36 |
| 5.3 | Obecná data o spalničkách pro čtyřčlenné domácnosti | 37 |
| 5.4 | Podrobná data o spalničkách pro čtyřčlenné domácnosti | 38 |
| 5.5 | Obecná data o nachlazení pro pětičlennou domácnost | 39 |
| 5.6 | Podrobná data o nachlazení pro pětičlennou domácnost | 40 |