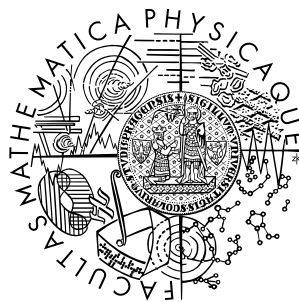


Univerzita Karlova v Praze  
Matematicko-fyzikální fakulta

## BAKALÁŘSKÁ PRÁCE



Eliška Krausová

### **Modely pro šíření infekčních chorob v malých skupinách**

Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: Mgr. Michal Kulich, PhD

Studijní program: Matematika, Obecná matematika

2007

Na tomto místě bych chtěla poděkovat vedoucímu své práce panu magistru Michalu Kulichovi, PhD za jeho trpělivost a za rady, které mi poskytl.

Prohlašuji, že jsem svou bakalářskou práci napsala samostatně a výhradně s použitím citovaných pramenů. Souhlasím se zapůjčováním práce a jejím zveřejňováním.

V Praze dne 20.5.2007

Eliška Krausová

# Obsah

<b>1</b>	<b>Úvod</b>	<b>5</b>
1.1	Motivace . . . . .	5
1.2	Základní popis infekčního procesu . . . . .	6
<b>2</b>	<b>Řetězcově-binomické modely</b>	<b>7</b>
2.1	Porovnání počtu infikovaných . . . . .	7
2.2	Epidemické řetězce . . . . .	8
2.3	Řetězcově-binomický model . . . . .	9
2.4	Analýza dat podle počtu infikovaných . . . . .	10
2.5	Analýza dat podle epidemických řetězců . . . . .	12
<b>3</b>	<b>Řetězcové modely s náhodnými vlivy</b>	<b>13</b>
3.1	Pravděpodobnost vyhnutí se infekci . . . . .	13
3.2	Náhodná infekčnost: modely . . . . .	15
3.3	Náhodná infekčnost: aplikace . . . . .	17
3.4	Náhodné vlivy domácností: modely . . . . .	17
<b>4</b>	<b>Latentní a infekční období</b>	<b>20</b>
4.1	Pozorovatelné infekční období . . . . .	20
4.2	Izolace nemocného při projevení příznaků . . . . .	23
4.3	Konstantní délka infekčního období . . . . .	24
<b>5</b>	<b>Závěr</b>	<b>25</b>
	<b>Literatura</b>	<b>26</b>

Název práce: Modely pro šíření infekčních chorob v malých skupinách  
Autor: Eliška Krausová  
Katedra (ústav): Katedra pravděpodobnosti a matematické statistiky  
Vedoucí bakalářské práce: Mgr. Michal Kulich, PhD  
e-mail vedoucího: kulich@karlin.mff.cuni.cz

Abstrakt: V předložené práci studujeme modely pro šíření infekčních chorob. V úvodu probereme základní pojmy a průběh infekčního procesu. Podrobněji se budeme zabývat řetězcově-binomickými modely, které se nejvíce hodí pro popis šíření chorob v malých skupinách, jako jsou domácnosti. Vysvětlíme použití těchto modelů v analýze dat podle počtu infikovaných a podle epidemických řetězců. Též ukážeme výpočet odhadů parametrů a důležitost testování hypotéz o vztazích mezi těmito parametry. Dále se budeme zabývat modely odvozenými z řetězcově-binomických modelů, ve kterých jsou parametry nahrazeny náhodnými veličinami. Také zmíníme analýzu chorob z hlediska délky latentního a infekčního období.

Klíčová slova: řetězcově-binomické modely, epidemické řetězce, infekční choroby

Title: Models for infectious disease spread in small populations  
Author: Eliška Krausová  
Department: Department of Probability and Mathematical Statistics  
Supervisor: Mgr. Michal Kulich, PhD  
Supervisor's e-mail address: kulich@karlin.mff.cuni.cz

Abstract: In the present work we study models for infectious disease spread. In the beginning we introduce some essential terminology and describe basic characteristics of the infection process. We discuss in more detail chain binomial models which are most appropriate for data on the spread of disease in small groups such as households. We illustrate the use of the model on analysis of size of outbreak data and analysis of chain data. We also demonstrate estimating the parameters and importance to test hypotheses in terms of relationships between the model parameters. Next we modify the chain binomial models by taking the parameters to be random variables. We also mention analyses based on durations of latent and infectious periods.

Keywords: chain binomial model, epidemic chains, infectious disease

# Kapitola 1

## Úvod

V této práci se budeme zabývat statistickými metodami pro analýzu šíření infekčních chorob. Tyto postupy se liší od klasických statistických metod z několika důvodů. Za prvé, data o šíření infekčních chorob nejsou výsledky předem plánovaných pokusů, ale pochází z přirozeně se vyskytujících epidemií. Za druhé, data jsou vysoce závislá, protože infikovaní jedinci jsou příčinou dalších infikovaných. Za třetí, průběh epidemií je většinou pozorovatelný jen částečně. Proto se používají matematické modely s vestavěnými náhodnými činiteli, ovlivňujícími šíření infekčních chorob.

Za infekční chorobu považujeme chorobu která je infekční v tom smyslu, že nakažený jedinec prochází tzv. infekčním obdobím, během kterého je schopen přenést chorobu na vnímavého jedince buď přímým dostatečně těsným kontaktem jeden na jednoho nebo infikováním prostředí (ovzduší nebo vybavení domácnosti), se kterým má potom vnímavý jedinec dostatečně blízký kontakt.

### 1.1 Motivace

Hlavním cílem analýzy infekčních chorob je získání informací, které nám umožní kontrolovat šíření těchto chorob. Pomocí statistických metod se snažíme určit způsob šíření choroby, odhadnout průměrnou délku latentního a infekčního období, jejich variabilitu a určit podíl populace, který je nutné očkovat, aby se předešlo závažnější epidemii. Můžeme zjistit, jak závisí rychlost šíření nemoci na věku, pohlaví, počtu ohrožených jedinců, počtu infekčních jedinců a hustotě populace.

## 1.2 Základní popis infekčního procesu

Infekční choroba je přenesena na vnímavého jedince, zkráceně **vnímavého**, pokud tento jedinec přijme určité množství infekčního materiálu, které je postačující k probuzení infekce. Říkáme, že proběhl **infekční kontakt**. První infekční kontakt, kterému je vnímavý vystaven, vede k jeho nakažení. O dalších infekčních kontaktech předpokládáme, že na něj nemají žádný vliv, dokud se později nestane znovu vnímavým. Nakažený jedinec nejdříve prochází **latentním obdobím**, během kterého se choroba vyvíjí pouze vnitřně bez jakéhokoliv šíření infekčního materiálu do okolí. Latentní období končí, jakmile nakažený začne být schopen šířit chorobu dál a po dobu jeho infekčního období o něm hovoříme jako o **infekčním**. Infekční období končí, když nakažený přestane být infekční a buď se znovu stane vnímavým nebo se na nějaký čas stane vyloučeným. **Vyloučený** je takový jedinec, který nehraje žádnou roli v šíření choroby. Jedinec se stane vyloučeným, pokud je izolovaný od ostatních, mrtvý nebo pokud získal imunitu. Stav izolace nebo imunity může být buď dočasný nebo trvalý. O každém nakaženém jedinci někdy také hovoříme jako o **případu**.

# Kapitola 2

## Řetězcově-binomické modely

Řetězcově-binomické modely se nejvíce hodí pro popis šíření choroby v malých skupinách jako jsou např. domácnosti. Domácností rozumíme skupinu lidí, která sdílí stejné obytné prostory a příslušenství. Předpokládáme, že poté, co je choroba zanesena do domácnosti, je šance k nakažení mimo domácnost zanedbatelná v porovnání s šancí k nakažení v domácnosti. To znamená předpoklad, že choroba se uvnitř domácnosti šíří nezávisle na ostatních domácnostech.

### 2.1 Porovnání počtu infikovaných

Uvažujme  $n$  domácností, které jsou zasaženy určitou chorobou. Pro analýzu dat z těchto domácností se obvykle předpokládá, že domácnosti jsou homogenní vzhledem k šíření choroby. Tento předpoklad ale nemusí být automaticky splněn a je nutné ho ověřit. Řekněme, že domácnosti se dají rozdělit do několika skupin, kde domácnosti z jedné skupiny jsou považovány za stejné, zatímco domácnosti z různých skupin se mohou lišit. Domácnosti mohou být například rozděleny podle míry zaplnění nebo podle počtu očkovaných členů. Cílem je testovat, zda se rozšíření choroby v jednotlivých skupinách významně liší.

Předpokládejme, že domácnosti jsou co do počtu členů stejně velké a obsahují před začátkem epidemie  $s$  vnímavých jedinců. Uvažujme rozdělení počtu infikovaných za platnosti nulové hypotézy, že všechny domácnosti jsou stejné. Označme  $\theta_j$  pravděpodobnost, že se  $j$  z  $s$  vnímavých jedné domácnosti stane infikovanými před koncem epidemie a  $N_j$  počet domácností s přesně  $j$  případy,  $j = 1, 2, \dots, s$ . Za platnosti nulové hypotézy má  $N_j$

binomické rozdělení dané

$$P(N_j = x) = \binom{n}{x} \theta_j^x (1 - \theta_j)^{n-x}, \quad x = 0, 1, \dots, n.$$

Sdružené rozdělení  $(N_1, N_2, \dots, N_s)$  je multinomické a je určeno

$$P(N_1 = x_1, \dots, N_s = x_s) = \frac{n!}{x_1! x_2! \dots x_s!} \theta_1^{x_1} \theta_2^{x_2} \dots \theta_s^{x_s},$$

kde  $\sum x_j = n$  a  $\sum \theta_j = 1$ .

Za platnosti alternativní hypotézy máme  $k$  různých typů domácností. Předpokládejme, že máme  $n_i$  domácností typu  $i$  mezi  $n$  postiženými domácnostmi. Z  $n_i$  domácností typu  $i$  nechť je  $N_{ij}$  domácností s celkovým počtem případů rovným  $j$ . Potom máme pro každý typ zvlášť multinomické rozdělení. Toto rozdělení  $(N_{i1}, N_{i2}, \dots, N_{is})$  je dáno

$$P(N_{i1} = x_{i1}, \dots, N_{is} = x_{is}) = \frac{n_i!}{x_{i1}! x_{i2}! \dots x_{is}!} \theta_{i1}^{x_{i1}} \theta_{i2}^{x_{i2}} \dots \theta_{is}^{x_{is}},$$

kde  $\theta_{ij}$  je pravděpodobnost, že postižená domácnost typu  $i$  bude mít celkem  $j$  případů.

Nechť  $n_{ij}$  jsou pozorované hodnoty náhodných veličin  $N_{ij}$ . Označme  $n_{.j} = n_{1j} + n_{2j} + \dots + n_{kj}$ . Testování nulové hypotézy, že rozdělení počtu infikovaných je pro všechny typy domácností stejné, znamená testovat nezávislost dvou znaků (typ domácnosti, počet případů). Spočítejme testovou statistiku

$$X^2 = \sum_{i=1}^k \sum_{j=1}^s (n_{ij} - e_{ij})^2 / e_{ij},$$

kde  $e_{ij} = n_i \cdot n_{.j} / n$  je odhadnutá hodnota  $E(N_{ij})$ . Tento odhad platí za platnosti nulové hypotézy, že počet infikovaných má pro každý typ domácnosti stejné rozdělení. Výsledek dostaneme porovnáním testové statistiky  $X^2$  s kritickou hodnotou  $\chi_{(k-1)(s-1)}^2$ -rozdělení. Toto porovnání můžeme použít jen pro hodnoty  $e_{ij}$  větší nebo rovny 5.

## 2.2 Epidemické řetězce

Pomocí epidemických řetězců popisujeme postup šíření choroby mezi vnímavými jedinci jedné domácnosti. Pro bližší popis vysvětlíme význam **generací**. Uvažujme vniknutí infekční choroby do skupiny. První generace případů se skládá z vnímavých, kteří se dostali do infekčního kontaktu s jedním



z primárních případů, zatímco ti, kteří se dostali do infekčního kontaktu s někým z první generace infikovaných, tvoří druhou generaci, atd. Pro přehlednost o primárních případech mluvíme jako o nulté generaci. Epidemický řetězec pro postiženou domácnost je seznam počtu případů v každé generaci včetně nulté. Například píšeme  $1 \rightarrow 1 \rightarrow 2$  pro řetězec, který se skládá z jednoho primárního případu, jednoho případu první generace, dvou případů druhé generace a žádného případu třetí (nebo vyšší) generace.

Většinou nejsou k dispozici dostatečně podrobná data, která by umožňovala rozdělení postižených domácností podle jednotlivých epidemických řetězců, ale jejich aplikací získáme rozdělení počtu infikovaných, vyjádřené pomocí parametrů s jasným epidemiologickým významem.

## 2.3 Řetězcově-binomický model

Abychom mohli najít výrazy pro pravděpodobnosti jednotlivých řetězců, musíme učinit několik předpokladů o způsobu šíření choroby. Předpokládejme, že každý vnímavý se může nakazit nezávisle na ostatních a že všichni zbývající vnímaví mají stejnou šanci se nakazit od infekčních jedinců dané generace.

Označme pravděpodobnost, že daný vnímavý unikne infekci, když je vystaven  $i$  infekčním určité generace, jako  $q_i$ ,  $i = 1, 2, \dots$ . Zatím předpokládáme, že  $q_i$  se z generace na generaci nemění. Označme  $S_t$  počet vnímavých v domácnosti vystavených  $I_t$  infekčním generace  $t$ . Pro  $t = 0, 1, 2, \dots$  pak platí

$$P(I_{t+1} = x | S_t = s, I_t = i) = \frac{s!}{x!(s-x)!} p_i^x q_i^{s-x}, \quad x = 0, 1, \dots, s,$$

kde  $p_i = 1 - q_i$ . K použití tohoto modelu je nutné znát počet infekčních a počet vnímavých každé generace. K tomu potřebujeme předpoklady, že všichni infekční jsou rozpoznatelní a že na konci infekčního období se z infekčního stane vyloučený (např. získáním imunity) po zbytek trvání epidemie. Nyní můžeme vyjádřit počet vnímavých na konci každé generace vzorcem

$$S_{t+1} = S_t - I_{t+1}, \quad t = 0, 1, 2, \dots,$$

s počátečními hodnotami  $I_0 = i_0$  and  $S_0 = s_0$ , které jsou známé. Tímto je formulace modelu dokončena.

Nyní ukážeme výpočet pravděpodobností jednotlivých řetězců. Mějme domácnost, která se skládá ze čtyř vnímavých a jednoho primárně infekčního,

takže  $s_0 = 4$  a  $i_0 = 1$ . Pro tuto domácnost nám model dává pravděpodobnost řetězce  $1 \longrightarrow 1 \longrightarrow 2$  jako

$$P \{1 \longrightarrow 1 \longrightarrow 2\} = P \{I_1 = 1 | S_0 = 4, I_0 = 1\} \cdot P \{I_2 = 2 | S_1 = 3, I_1 = 1\} \cdot$$

$$P \{I_3 = 0 | S_2 = 1, I_2 = 2\} = 4p_1q_1^3 \cdot 3p_1^2q_1 \cdot q_2 = 12p_1^3q_1^4q_2.$$

Obecně můžeme pravděpodobnost řetězce  $i_0 \longrightarrow i_1 \longrightarrow \dots \longrightarrow i_r$  vyjádřit jako

$$P \{i_0 \longrightarrow i_1 \longrightarrow \dots \longrightarrow i_r\} = \frac{s_0!}{i_1!i_2!\dots i_r!s_r!} \prod_{t=0}^r p_{i_t}^{i_{t+1}} q_{i_t}^{s_{t+1}},$$

kde  $i_{r+1} = 0$ , a tedy  $s_{r+1} = s_r$ . Všimněme si, že tato pravděpodobnost je podmíněná počtem primárních případů.

Tento obecný model v sobě obsahuje dva speciální modely, které získáme vyjádřením závislosti parametru  $q_i$  na  $i$ . Jeden z těchto modelů je Reed-Frostův model, který předpokládá, že  $q_i = q^i, i = 1, 2, \dots$ . Tento model nám říká, že pravděpodobnost úniku před infekcí při vystavení  $i$  infekčním najednou je stejná jako pravděpodobnost úniku před infekcí při vystavení jednomu infekčnímu v každé z  $i$  generací. Reed-Frostův předpoklad se používá pro nemoci, které jsou přenášeny zejména osobním kontaktem jeden na jednoho.

Druhým modelem je tzv. Greenwoodův model, který předpokládá, že  $q_i = q, i = 1, 2, \dots$ , a  $q_0 = 1$ . Tento model nám říká, že pravděpodobnost nakažení je stejná při vystavení jednomu infekčnímu jako při vystavení více infekčním najednou. Greenwoodův předpoklad se používá, pokud je prostředí domácnosti zasyceno infekčním materiálem, i když je v domácnosti přítomen jen jeden infekční, a šance nakažení vnímavého závisí spíše na jeho chování v domácím prostředí. V obou těchto modelech máme jen jeden parametr.

## 2.4 Analýza dat podle počtu infikovaných

V oddílu 2.1 jsme označili  $\theta_j$  pravděpodobnost, že počet nakažených v domácnosti je roven  $j$ . Pro domácnosti s  $s$  vnímavými máme parametry  $\theta_1, \theta_2, \dots, \theta_s$ , kde  $\sum \theta_j = 1$ . Pomocí předpokladů řetězcově-binomicého modelu můžeme snížit počet parametrů, ale hlavně můžeme parametry  $\theta_1, \theta_2, \dots, \theta_s$  vyjádřit pomocí parametrů  $q_1, q_2, \dots, q_{s-2}$ , které mají důležitější epidemiologický význam.

Nyní předpokládejme, že máme data dostatečně podrobná, abychom mohli určit počet primárních případů a původní počet vnímavých. Budeme se věnovat domácnostem se čtyřmi členy, konkrétně s jedním primárním případem a třemi vnímavými. Rozdělení počtu případů dostaneme sečtením pravděpodobností odpovídajících řetězců:

$$\theta_1 = P \{1\} = q_1^3$$

$$\theta_2 = P \{1 \longrightarrow 1\} = 3q_1^4 p_1$$

$$\theta_3 = P \{1 \longrightarrow 1 \longrightarrow 1\} + P \{1 \longrightarrow 2\} = 3q_1 p_1^2 (2q_1^3 + q_2)$$

$$\theta_4 = 1 - \theta_1 - \theta_2 - \theta_3.$$

Uvažujme  $n$  postižených domácností a mezi nimi vždy  $n_j$  domácností s právě  $j$  případy. Věrohodnostní funkce vyjádřená pomocí  $\theta$  vypadá takto:

$$l(\theta_1, \theta_2, \theta_3) = \text{konstanta} \times \theta_1^{n_1} \theta_2^{n_2} \theta_3^{n_3} (1 - \theta_1 - \theta_2 - \theta_3)^{n_4}.$$

Maximálně věrohodný odhad  $\theta_j$  je  $\hat{\theta}_j = n_j/n, j = 1, 2, 3$ . Výchozí odhady parametrů  $q_1$  a  $q_2$  dostaneme dosazením odhadů  $\hat{\theta}_1$  a  $\hat{\theta}_3$  za  $\theta_1$  a  $\theta_3$  do rovnic výše a jejich vyřešením. Výsledkem je

$$\tilde{q}_1 = \hat{\theta}_1^{1/3}, \quad \tilde{q}_2 = \{\hat{\theta}_3 / 3\tilde{q}_1 \tilde{p}_1^2\} - 2\hat{\theta}_1.$$

Tyto hodnoty jsou vhodné počáteční odhady pro výpočet maximálně věrohodného odhadu pomocí počítače. Předtím je vhodné zkontrolovat, zda se tento model shoduje s našimi daty. Je jasné, že hodnoty  $n_1$  a  $n_3$  se budou s modelem shodovat přesně, protože pomocí nich jsme vyjádřili odhady  $\tilde{q}_1$  a  $\tilde{q}_2$ . Zbývá porovnat  $n_2$  a  $n_4$  s příslušnými hodnotami. Pracujeme s podmíněným rozdělením  $(N_2, N_4)$ , určeným  $N_2 + N_4 = n_2 + n_4$ . Porovnáme tedy  $n_2$  a  $n_4$  s  $(n_2 + n_4)\tilde{a}$  a  $(n_2 + n_4)(1 - \tilde{a})$ , kde  $\tilde{a} = 3\tilde{q}_1^4 \tilde{p}_1 / (1 - \hat{\theta}_1 - \hat{\theta}_3)$ , pomocí  $\chi^2$  testu dobré shody.

Hlavním účelem použití řetězcově-binomických modelů je získat představu o způsobu šíření choroby. Proto testujeme epidemiologicky významné hypotézy o vztazích mezi parametry. Předpokládejme, že se model dostatečně shoduje s daty a chceme testovat hypotézu  $q_2 = f(q_1)$ . Pro testování Greenwoodova předpokladu volíme  $f(q_1) = q_1$ , zatímco pro Reed-Frostův předpoklad volíme  $f(q_1) = q_1^2$ . Při testování hypotézy  $q_2 = f(q_1)$  využijeme toho, že za platnosti této hypotézy můžeme  $\hat{q}_2 - f(\hat{q}_1)$  chápat jako veličinu s normálním rozdělením s nulovou střední hodnotou a rozptylem

$$\widehat{Var}(\hat{q}_2) + \{f'(\hat{q}_1)\}^2 \widehat{Var}(\hat{q}_1) - 2f'(\hat{q}_1)\widehat{Cov}(\hat{q}_1, \hat{q}_2).$$

## 2.5 Analýza dat podle epidemických řetězců

Jak bylo zmíněno výše, není snadné získat dostatečně podrobná data, která rozlišují jednotlivé řetězce. Některé vlastnosti nemocí nám k tomu ale mohou pomoci. Jednotlivé řetězce se dají rozlišit pokud (1) latentní období je relativně dlouhé oproti infekčnímu období a (2) všichni nakažení začínají ukazovat jasně rozpoznatelné symptomy přibližně ve stejnou dobu, kdy se nemoc začne rozvíjet. Mezi nemocí s těmito vlastnostmi můžeme zařadit například spalničky, příušnice nebo neštovice.

Postup analýzy ukážeme na domácnostech se čtyřmi členy. Opět uvažujeme  $n$  postižených domácností, které se skládají z jednoho primárně infekčního a tří vnímavých. Předpokládejme, že v  $n_c$  z těchto domácností odpovídá průběh nemoci epidemickému řetězci typu  $c$ , kde  $c \in C = \{1, 11, 111, 12, 1111, 112, 121, 13\}$ . Pro jednoduchost budeme při zapisování řetězců v indexech vynechávat šipky. Například  $n_{121}$  značí počet domácností, ve kterých průběh nemoci odpovídá řetězci  $1 \rightarrow 2 \rightarrow 1$ . Samozřejmě platí vztah  $\sum n_c = n$ , kde sčítáme přes všechna  $c \in C$ .

Model přizpůsobíme našim datům odhadem jeho parametrů, k čemuž použijeme metodu maximální věrohodnosti. Věrohodnostní funkce je

$$l(q_1, q_2) = \prod_{c \in C} [P\{c\}]^{n_c} = \text{konstanta} \times q_1^{x_1} p_1^{m_1 - x_1} q_2^{x_2} p_2^{m_2 - x_2},$$

kde  $m_i$  značí celkový počet vystavení vnímavého  $i$  infekčním a  $x_i$  značí počet vystavení, při kterých vnímavý infekci unikl. Přesněji

$$x_1 = 3n_1 + 4n_{11} + 4n_{111} + n_{12} + 3n_{1111} + 2n_{1112} + n_{121},$$

$$m_1 = 3n_1 + 5n_{11} + 6n_{111} + 3n_{12} + 6n_{1111} + 5n_{1112} + 3n_{121} + 3n_{13},$$

zatímco  $x_2 = n_{12}$  a  $m_2 = n_{12} + n_{121}$ . Jako maximálně věrohodné odhady parametrů dostaneme

$$\hat{q}_1 = x_1/m_1 \quad \hat{q}_2 = x_2/m_2.$$

Nyní se ujistíme že model dostatečně dobře popisuje data pomocí  $\chi^2$  testu. Spočítáme statistiku

$$\sum_{c \in C} \frac{(n_c - n\hat{p}_c)^2}{n\hat{p}_c},$$

kteřou srovnáme s kritickou hodnotou  $\chi_5^2$ -rozdělení. Velké hodnoty znamenají, že model není vhodný pro popis dat.

## Kapitola 3

# Řetězcové modely s náhodnými vlivy

O některých nemocích je známo, že nakažení jedinci dosahují různé míry infekčnosti a liší se v délce infekčního období. Proto se v této kapitole budeme věnovat řetězcovým modelům, ve kterých je různorodost jedinců nebo domácností vysvětlena nahrazením parametrů řetězcově-binomického modelu náhodnými veličinami.

### 3.1 Pravděpodobnost vyhnutí se infekci

S každým nakaženým jedincem spojujeme určitou funkci  $\Lambda$  nazývanou **infekčnostní funkce**. Tato funkce  $\Lambda(u)$  udává míru infekčnosti nakaženého jedince  $u$  časových jednotek od jeho nakažení. Tedy, jestliže  $A$  je jedinec s infekčnostní funkcí  $\Lambda_A$ , který byl nakažen v čase  $T_A$ , a  $B$  je jedinec dosud vnímavý v čase  $t$ , pak pravděpodobnost infekčního kontaktu mezi  $A$  a  $B$  v časovém intervalu  $(t, t + \delta)$  je

$$\alpha_{AB}\Lambda_A(t - T_A)\delta + o(\delta),$$

kde  $\alpha_{AB}\delta + o(\delta)$  značí pravděpodobnost blízkého kontaktu mezi jedinci  $A$  a  $B$  v čase  $(t, t + \delta)$  a  $\Lambda_A(t - T_A)$  značí pravděpodobnost, že nemoc je přenesena na jedince  $B$  během tohoto kontaktu.

Typicky platí, že infekčnostní funkce  $\Lambda$  zůstává nulová během latentního období, potom roste k vrcholu a nakonec klesá k nule, které dosáhne na konci infekčního období. To nám dovoluje popsat latentní a infekční období nakaženého jedince pomocí jeho infekčnostní funkce. Opět uvažujme jedince  $A$ ,

který je nakažen v čase  $T_A$ , s infekčností funkcí  $\Lambda_A$ . Pro úplnost definujeme  $\Lambda_A(u) = 0$ , kdykoliv  $u < 0$ . Definujeme

$$U = \inf_u \{u : \Lambda_A(u - T_A) > 0\}$$

jako první okamžik, kdy  $A$  je infekční, a

$$W = \sup_u \{u : \Lambda_A(u - T_A) > 0\}$$

poslední bod v čase, kdy  $A$  je infekční. Není problém připustit vracející se infekční období, ale my budeme předpokládat, že každý nakažený jedinec prochází jediným latentním obdobím následovaným jediným infekčním obdobím, což je nejčastější případ. Latentní období jedince  $A$  je dáno intervalem  $(T_A, U)$ , zatímco infekční období je dáno  $(U, W)$ .

Nyní odvodíme výraz pro pravděpodobnost, že  $B$  se vyhne infekčnímu kontaktu s  $A$ . Rozdělíme interval  $(U, W)$  na velké množství malých intervalů  $(U, t_1), (t_1, t_2), \dots, (t_n, W)$ , kde  $U = t_0 < t_1 < t_2 < \dots < t_n < t_{n+1} = W$ . Potom pravděpodobnost, že se  $B$  vyhne infekčnímu kontaktu s  $A$  je dána

$$\prod_{j=1}^n \{1 - \alpha_{AB} \Lambda_A(t_j - T_A) \cdot (t_{j+1} - t_j) + o(t_{j+1} - t_j)\},$$

za předpokladu, že šance blízkého kontaktu mezi  $A$  a  $B$  je  $\alpha_{AB} \times$  (délka intervalu)  $+ o$ (délka intervalu), nezávisle pro každý časový interval. Limitně, jak zvyšujeme počet intervalů a tedy délky všech intervalů se blíží k nule, dostáváme výraz

$$Q_{AB} = \exp \left\{ - \int_0^\infty \alpha_{AB} \Lambda_A(u) du \right\}$$

pro pravděpodobnost, že se  $B$  vyhne infekčnímu kontaktu s  $A$  během infekčního období druhého z nich.

V kapitole 2 jsme předpokládali, že  $Q_{AB}$  je konstantní. Uvažujme okolnosti, za kterých to bude pravda. Nechť

$$\alpha_{AB} \Lambda_A(u) = \begin{cases} \beta_{AB} & (U \leq u \leq W), \\ 0 & \text{jinak,} \end{cases} \quad (3.1)$$

takže  $Q_{AB} = \exp \{-\beta_{AB}(W - U)\}$ . Jestliže  $\beta_{AB}$ , koeficient infekčního kontaktu v infekčním období jedince  $A$ , a  $(W - U)$ , infekční období jedince  $A$ ,

jsou oba konstantní, pak  $Q_{AB}$  bude také konstantní. V následující podkapitole se budeme zabývat modely, ve kterých  $Q_{AB}$  bude náhodná veličina.

Nyní popíšeme jeden speciální případ. Nechť platí vztah (3.1). Dále předpokládejme, že  $\beta_{AB} = \beta$  je konstantní pro každý pár jedinců  $A$  a  $B$ , což je předpoklad rovnoměrného mísení. Příslušná pravděpodobnost vyhnutí se infekci je  $Q = \exp\{-\beta Y\}$ , kde  $Y = W - U$  je doba trvání infekčního období nakaženého. Nakonec předpokládejme, že  $Y$  má exponenciální rozdělení s hustotou

$$f_Y(y) = \gamma e^{-\gamma y}, \quad \gamma > 0.$$

Potom  $Q$  má degenerované beta rozdělení s hustotou

$$f_Q(x) = \rho x^{\rho-1}, \quad 0 < x < 1,$$

kde  $\rho = \gamma/\beta$ . Tento model se nazývá Kermack-McKendrickův model.

### 3.2 Náhodná infekčnost: modely

Uvažujme domácnost, ve které se členové rovnoměrně mísí, vnímavost je pro všechny vnímavé stejná a nakažení jedinci se mohou lišit infekčností. Každému nakaženému, nezávisle na ostatních, přiřadíme infekčnostní funkci náhodně vybranou z množiny vhodných infekčnostních funkcí. Označme  $s_0$  původní počet vnímavých v domácnosti s  $s_0 + 1$  členy, včetně jednoho primárního případu. Předpokládejme, že každý vnímavý nezávisle unikne infekci od primárně nakaženého s pravděpodobností  $Q$ . Potom podmíněné rozdělení počtu nakažených v první generaci je dáno

$$P(\text{počet případů první generace} = x | Q) = \binom{s_0}{x} (1 - Q)^x Q^{s_0 - x},$$

$$x = 0, 1, \dots, s_0.$$

Příslušnou nepodmíněnou pravděpodobnost získáme spočítáním podmíněné střední hodnoty. Výsledný výraz bude vyjádřen pomocí prvních  $s_0$  momentů náhodné veličiny  $Q$ .

Uvažujme teď počet případů druhé generace. Podmíněné rozdělení počtu případů druhé generace, za podmínky, že je  $x$  infekčních v první generaci, je dáno

$$\begin{aligned} P(\text{počet případů druhé generace} = y | x, Q_{(x)}) \\ = \binom{s_0 - x}{y} (1 - Q_{(x)})^y Q_{(x)}^{s_0 - x - y}, \quad y = 0, 1, \dots, s_0 - x, \end{aligned}$$

kde  $Q_{(x)}$  je pravděpodobnost, že vnímavý unikne infekci, když bude vystaven  $x$  infekčním po dobu jejich infekčních období.  $Q_{(x)}$  je opět náhodná veličina.

Jednodušší model dostaneme specifikací závislosti  $Q_{(x)}$  na  $x$ . Tady použijeme Reed-Frostův předpoklad, že  $Q_{(x)}$  je součin  $x$  nezávislých náhodných veličin se stejným rozdělením jako  $Q$ .

Předpokládejme, že máme domácnost skládající se z  $s_0 = 4$  vnímavých a  $i_0 = 1$  primárního případu. Pro tuto domácnost uvažujme řetězec  $1 \longrightarrow 1 \longrightarrow 2$ . Nechť  $Q_1, Q_2, Q_3$  a  $Q_4$  značí pravděpodobnosti, že daný vnímavý unikne infekci od každého ze čtyř nakažených jedinců. Podmíněná pravděpodobnost řetězce  $1 \longrightarrow 1 \longrightarrow 2$  se spočítá stejně jako u řetězcově-binomického modelu a je dána výrazem

$$\binom{4}{1} Q_1^3 (1 - Q_1)^1 \binom{3}{1} Q_2^1 (1 - Q_2)^2 \binom{1}{1} Q_3 Q_4 (1 - Q_3 Q_4)^0.$$

Abychom dostali nepodmíněnou pravděpodobnost, spočítáme střední hodnotu tohoto výrazu a použijeme fakt že  $Q_1, Q_2, Q_3$  a  $Q_4$  jsou nezávislé, stejně rozdělené náhodné veličiny. Vyjde nám

$$E \left\{ 4Q_1^3 (1 - Q_1) \cdot 3Q_2 (1 - Q_2)^2 \cdot Q_3 Q_4 \right\} = 12\mu_1^2 (\mu_3 - \mu_4) (\mu_1 - 2\mu_2 + \mu_3),$$

kde píšeme  $\mu_r$  pro  $E(Q^r)$ ,  $r = 0, 1, 2, \dots$

Speciálním případem tohoto modelu je i Reed-Frostův řetězcově-binomický model. Dostaneme ho dosazením za  $\mu_r = q^r$ ,  $r = 1, 2, \dots$

Z praktických důvodů je vhodné mít model jen s několika málo parametry. Toho dosáhneme například určením typu rozdělení  $Q$ . Budeme předpokládat, že  $Q$  má beta rozdělení dané hustotou

$$f(x) = \{B(\rho, \sigma)\}^{-1} x^{\rho-1} (1-x)^{\sigma-1}, \quad 0 < x < 1; (\rho, \sigma > 0).$$

Potom můžeme upravit výraz pro střední hodnotu

$$E \left\{ Q^k (1 - Q)^l \right\} = B(\rho + k, \sigma + l) / B(\rho, \sigma) = \frac{\rho^{(k)} \sigma^{(l)}}{\tau^{(k+l)}},$$

kde jsme použili označení  $\tau = \rho + \sigma$  a  $k^{(r)} = k(k+1)\dots(k+r-1)$ . Speciálně

$$\mu_r = \begin{cases} 1 & r = 0, \\ \frac{\rho^{(r)}}{\tau^{(r)}} & r = 1, 2, \dots \end{cases}$$

Za tohoto předpokladu beta rozdělení  $Q$  dostaneme pravděpodobnost řetězce  $1 \longrightarrow 1 \longrightarrow 2$  ve tvaru

$$4E \left\{ Q^3 (1 - Q) \right\} \cdot 3E \left\{ Q (1 - Q)^2 \right\} \cdot E \{ Q \} \cdot E \{ Q \} = \frac{12\sigma\sigma^{(2)}\rho^3\rho^{(3)}}{\tau^2\tau^{(3)}\tau^{(4)}} \quad (3.2)$$



Tento model v sobě opět obsahuje jako speciální případy některé jednodušší modely. Konkrétně, pokud zvolíme  $\sigma = 1$ , dostaneme Kermack-McKendrickův model. Dále, pokud necháme  $\rho$  a  $\sigma$  růst do nekonečna tak, že  $\rho/(\rho + \sigma)$  zůstane konstantní, dostaneme se zpět k řetězcově-binomickému modelu, ve kterém je  $Q$  konstantní.

### 3.3 Náhodná infekčnost: aplikace

Pro ilustraci metody hledání výchozích odhadů parametrů uvažujme  $n$  domácností s původně třemi vnímavými a jedním primárním případem. Označme  $K_i$  počet těch domácností, které mají  $i$  případů v první generaci,  $i = 0, 1, 2, 3$ . Očekávané hodnoty frekvencí případů první generace jsou dány

$$\begin{aligned} E(K_0) &= nE(Q^3) = n\mu_3, \\ E(K_1) &= nE\{3Q^2(1-Q)\} = 3n(\mu_2 - \mu_3), \\ E(K_2) &= nE\{3Q(1-Q)^2\} = 3n(\mu_1 - 2\mu_2 + \mu_3). \end{aligned}$$

Je tedy jasné, že nestranné odhady pro  $\mu_3$ ,  $\mu_2$ ,  $\mu_1$  jsou dány výrazy  $K_0/n$ ,  $(3K_0 + K_1)/3n$  a  $(3K_0 + 2K_1 + K_2)/3n$ . Odhady pro domácnosti jiných velikostí se najdou obdobně.

Abychom mohli aplikovat podrobnější model, ve kterém má  $Q$  beta rozdělení, potřebujeme odhadnout parametry  $\rho$  a  $\sigma$ . K získání výchozích odhadů těchto parametrů stačí vyřešit

$$\mu_1 = \rho/\tau, \quad \mu_2 = \frac{\rho}{\tau} \frac{\rho + 1}{\tau + 1},$$

kde  $\tau = \rho + \sigma$ . Jako řešení dostaneme výrazy

$$\rho = \mu_1(\mu_1 - \mu_2)/(\mu_2 - \mu_1^2), \quad \sigma = (1 - \mu_1)(\mu_1 - \mu_2)/\mu_2 - \mu_1^2,$$

do kterých dosadíme výchozí odhady parametrů  $\mu_1$  a  $\mu_2$  uvedené výše.

Po získání odhadů  $\mu_1$  a  $\mu_2$  je vhodné ověřit, zda je  $Q$  skutečně variabilní. Odhad  $Var(Q)$  je dán výrazem  $\tilde{\mu}_2 - \tilde{\mu}_1^2$ . Pokud vyjde příliš malé číslo, variabilita  $Q$  není podstatnou charakteristikou zkoumané nemoci.

### 3.4 Náhodné vlivy domácností: modely

V oddílu 3.2 jsme pozměnili řetězcově-binomický model tak, že jsme připustili rozdíly mezi infekčními v jejich infekčnosti. Pravděpodobnost vyhnutí se

infekci  $Q$  jsme tedy určili jako náhodnou veličinu vybranou pro každého infekčního nezávisle na ostatních. Nyní uvažujme situaci, kdy je pravděpodobnost vyhnutí se infekci  $Q$  stejná pro všechny infekční ze stejné domácnosti, ale může se lišit pro infekční z různých domácností. V podstatě předpokládáme různorodost jedinců ve společnosti, která může být vysvětlena rozdíly ve způsobu chování uvnitř domácností.

Pokud se domácnosti dají rozdělit do homogenních skupin, je možné použít řetězcově-binomický model z druhé kapitoly. Teď se budeme zabývat situací, kdy toto rozdělení do skupin možné není. Připustíme tedy rozdíly mezi domácnostmi a zvolíme pravděpodobnost vyhnutí se infekci  $Q$  pro každou domácnost jako náhodnou veličinu. Podmíněné pravděpodobnosti jednotlivých řetězců jsou stejné jako u řetězcově-binomického modelu, kde nahradíme parametry  $q_1, q_2, \dots$  náhodnými veličinami  $Q_1, Q_2, \dots$ , kde  $Q_j$  je pravděpodobnost vyhnutí se infekci od  $j$  infekčních (zatím jsme ještě nepřijali Reed-Frostův ani Greenwoodův předpoklad). Nepodmíněné pravděpodobnosti opět dostaneme spočítáním střední hodnoty. Tímto způsobem získáme pravděpodobnosti řetězců pro domácnost se třemi členy jako

$$\begin{aligned} P\{1\} &= E(Q^2) = \mu_2, & P\{1 \longrightarrow 1\} &= 2E(PQ^2) = 2(\mu_2 - \mu_3), \\ P\{1 \longrightarrow 1 \longrightarrow 1\} &= 2E(P^2Q) = 2(\mu_1 - 2\mu_2 + \mu_3), \\ P\{1 \longrightarrow 2\} &= E(P^2) = 1 - 2\mu_1 + \mu_2. \end{aligned}$$

Pro výpočet těchto pravděpodobností u domácností s větším počtem členů je nutné přijmout některé další předpoklady. První krok je použít Reed-Frostův nebo Greenwoodův předpoklad. Další krok je specifikovat rozdělení náhodné veličiny. Tady opět použijeme beta rozdělení s distribuční funkcí

$$f(x) = \{B(\rho, \sigma)\}^{-1} x^{a-1}(1-x)^{b-1}, \quad 0 < x < 1; (a, b > 0),$$

Potom

$$E\{Q^k(1-Q)^l\} = \frac{a^{(k)}b^{(l)}}{c^{(k+l)}}, \quad k, l = 1, 2, \dots,$$

kde  $c = a + b$  a  $a^{(k)} = a(a+1)\dots(a+k-1)$ , atd.

Na ukázkou spočteme pravděpodobnost řetězce  $1 \longrightarrow 1 \longrightarrow 2$  v domácnosti s původně čtyřmi vnímavými a jedním primárním případem.

$$\begin{aligned} P\{1 \longrightarrow 1 \longrightarrow 2\} &= E\{4Q_1^3(1-Q_1) \cdot 3Q_1(1-Q_1)^2 \cdot Q_2\} \\ &= \begin{cases} 12E\{Q^6(1-Q)^3\} = \frac{12a^{(6)}b^{(3)}}{c^{(9)}}, & Q_r = Q^r \\ 12E\{Q^5(1-Q)^3\} = \frac{12a^{(5)}b^{(3)}}{c^{(8)}}, & Q_r = Q \end{cases} \end{aligned}$$

První výraz nám vyšel za Reed-Frostova předpokladu, druhý za Greenwoodova předpokladu. Oba výsledky můžeme porovnat s výrazem (3.2), který vyšel pro stejný řetězec v předchozím modelu.

# Kapitola 4

## Latentní a infekční období

V této kapitole se budeme snažit použít data udávající časy, ve kterých se objevují symptomy, abychom našli časové charakteristiky nemoci jako je délka latentního a infekčního období nebo rychlost šíření nemoci v čase.

### 4.1 Pozorovatelné infekční období

Uvažujme infekční nemoc, u které jsme schopni rozpoznat začátek a konec infekčního období. Například nakažený jedinec může být infekční po dobu, po kterou se u něj objevuje viditelná vyrážka. Nebo může být například známo, že nakažený jedinec je infekční už dva dny před objevením se vyrážky až do čtvrtého dne od zmizení vyrážky. Předpokládejme, že jde o mírnou chorobu, takže příchod příznaků výrazně neovlivní chování jedince v domácnosti.

U každého infekčního budeme pozorovat jeho infekční období  $(U, W)$ . Proměnné  $U$  a  $W$  jsou měřeny vzhledem k zvolenému časovému počátku. Délka infekčního období je  $Y = W - U$  a pozorováním  $Y$  u většího množství infekčních jedinců můžeme učinit předpoklady o rozdělení  $Y$ . Pokud nás zajímá jen délka infekčního období, není problém spočítat výběrový průměr a rozptyl z pozorovaných hodnot. Pokud je ale účelem srovnání infekčních období u více nemocí, je vhodné použít rozsáhlejší analýzu a pokud možno najít rozdělení  $Y$ .

Najít obdobné vlastnosti latentního období je obtížnější, protože přesný okamžik, ve kterém proběhl infekční kontakt, není pozorovatelný. Je proto nezbytné učinit předpoklady o povaze infekční funkce. Protože infekční funkce nabývá kladných hodnot jen na intervalu  $(U, W)$ , jednou z mož-

ností je funkce

$$\Lambda(x) = \begin{cases} \beta(x - U)(W - x), & U \leq x \leq W, \\ 0, & \text{jinak.} \end{cases}$$

která je spojitá a roste od nuly k svému maximu, načež klesá zpět k nule. My použijeme jednodušší funkci

$$\Lambda(x) = \begin{cases} \beta, & U \leq x \leq W, \\ 0, & \text{jinak,} \end{cases}$$

protože dostatečně popisuje infekčnost v mnoha aplikacích. Všimněme si, že to je náhodná funkce, protože  $U$  a  $W$  jsou náhodné veličiny.

Jak souvisí míra šíření infekce  $\beta$  s latentním obdobím ukážeme na příkladu domácností se dvěma členy. Předpokládáme, že pozorované průběhy nemoci můžeme rozdělit do řetězců:

- 2 – dva primární případy;
- 1 – jeden primární případ, žádný sekundární případ
- 1  $\rightarrow$  1 – jeden primární případ, jeden sekundární případ

Data z postižené domácnosti se dvěma primárními případy neobsahují žádnou informaci o  $\beta$ , protože  $\beta$  reprezentuje míru šíření infekce uvnitř domácností. Nechť  $(U_1, W_1)$  a  $(U_2, W_2)$  jsou infekční období pro dva primární případy domácnosti, kde  $(U_1, W_1)$  odpovídá například staršímu jedinci. Předpokládáme, že se infekce u obou jedinců vyvíjí nezávisle a že oba byli nakaženi najednou. Proto  $U_1 - U_2 = X_1 - X_2$ , kde  $X_1$  a  $X_2$  jsou délky latentních období. Proměnná  $U_1 - U_2$  je funkce  $(U_1, W_1, U_2, W_2)$  která obsahuje dostupnou informaci o latentním období. Je vidět, že  $E(U_1 - U_2) = 0$  a  $\text{Var}(U_1 - U_2) = 2\text{Var}(X)$ , takže  $U_1 - U_2$  nám dává informaci o rozptylu délky latentního období.

Domácnost se dvěma primárními případy a pozorováními  $(U_1, W_1) = (u_1, w_1)$   $(U_2, W_2) = (u_2, w_2)$  přispívá do věrohodnostní funkce výrazem

$$f_Y(y_1)f_Y(y_2) \int_0^\infty f_X(x + u_1 - u_2)f_X(x)dx, \quad (4.1)$$

kde  $Y$  je délka infekčního období, ke které máme pozorování  $y_1 = (w_1 - u_1)$  a  $y_2 = (w_2 - u_2)$ . Funkce  $f_X$  a  $f_Y$  značí hustoty  $X$  a  $Y$ . Integrál je hustota  $U_1 - U_2$ .

Uvažujme teď domácnost s jedním primárním případem a bez sekundárního případu. Jediná dostupná informace z takovéto domácnosti je infekční doba ( $W - U$ ) a pozorování, že zbývající vnímavý unikl infekci. Takže data neobsahují žádné informace o latentním období, ale obsahují informaci o infekčním parametru  $\beta$ . Tato domácnost s pozorováními  $(U, W) = (u, w)$  přispívá do věrohodnostní funkce výrazem

$$f_Y(w - u)e^{-\beta(w-u)}, \quad (4.2)$$

kde  $e^{-\beta(w-u)}$  je pravděpodobnost, že zbývající vnímavý unikne infekci, když je vystaven jednomu infekčnímu po dobu jeho infekčního období  $(w - u)$  časových jednotek.

A konečně domácnost s jedním primárním a jedním sekundárním případem. Nechť  $(U_1, W_1)$  značí infekční období primárního případu a  $(U_2, W_2)$  totéž u sekundárního případu. Potom  $U_2 - U_1 = X_2 + Z$ , kde  $X_2$  je délka latentního období sekundárního případu a  $Z$  značí čas od  $U_1$  do nakažení zbývajícího vnímavého. Podmíněné rozdělení  $(U_2 - U_1)$  je dáno

$$f_{U_2-U_1}(a|Y_1 = y_1, 1 \rightarrow 1) = \int_0^{y_1} \beta e^{-\beta z} f_X(a - z) dz / 1 - e^{-\beta y_1},$$

kde  $y_1$  je délka infekčního období primárního případu. Potom tato domácnost s pozorováními  $(u_1, w_1)$   $(u_2, w_2)$  přispívá do věrohodnostní funkce výrazem

$$f_Y(y_1)f_Y(y_2) \int_0^{y_1} \beta e^{-\beta z} f_X(u_2 - u_1 - z) dz. \quad (4.3)$$

Část věrohodnostní funkce vztahující se k parametru  $\beta$  nebo parametrům rozdělení  $X$  je dána

$$\left[ \prod_{(2)} \int_0^\infty f_X(x + u_1 - u_2) f_X(x) dx \right] \left[ \prod_{(1)} e^{-\beta(w-u)} \right] \\ \times \left[ \prod_{(1 \rightarrow 1)} \int_0^{w_1 - u_1} \beta e^{-\beta z} f_X(u_2 - u_1 - z) dz \right],$$

kde  $\prod_{(c)}$  znamená součin přes všechny domácnosti s průběhem choroby podle řetězce typu  $c$ .

Abychom z věrohodnostní funkce mohli odhadnout všechny parametry, je nutné učinit předpoklady o rozdělení délky latentního období  $X$ . Jednou z možností je rozdělení s hustotou

$$f_X(x; \lambda, \gamma) = \lambda e^{-\lambda(x-\gamma)}, \quad x \geq \gamma.$$

Další možností by byl předpoklad, že  $X$  má aproximativně normální rozdělení se střední hodnotou  $\mu_X$  a rozptylem  $\sigma_X^2$ . Toto rozdělení můžeme použít jen pro dostatečně malý poměr  $\sigma_X/\mu_X$ .

## 4.2 Izolace nemocného při projevení příznaků

Uvažujme nyní nemoc s vážnějšími následky, takže všichni infekční jsou izolováni od ostatních, jakmile se u nich projeví symptomy nemoci. Je jasné, že symptomy, které vedou k odhalení nemocných, se objevují až po začátku infekčního období, jinak by se nemoc šířila jen zřídka. Okamžik vyjmutí tedy budeme považovat za konec infekčního období. Data se skládají jen z koncových okamžiků infekčních období  $W$ , zatímco počáteční okamžiky  $U$  neznáme. Ostatní předpoklady zůstávají stejné jako v předchozím modelu. Opět budeme hledat věrohodnostní funkci pro domácnosti se dvěma členy.

Průběh infekce se dvěma primárními případy nedává žádné informace o parametru  $\beta$ , ale časové body  $W_1$  a  $W_2$ , ve kterých jsou symptomy u nakažených poprvé pozorovány, dávají informaci o rozdělení  $X + Y$ . Platí tedy vztah  $W_2 - W_1 = (X_2 + Y_2) - (X_1 + Y_1)$ . Pozorování  $w_2 - w_1$  přispívá do věrohodnostní funkce výrazem

$$\int_0^\infty f_{X+Y}(x + w_2 - w_1) f_{X+Y}(x) dx. \quad (4.4)$$

Můžeme buď specifikovat rozdělení  $X$  a  $Y$  a z nich odvodit hustotu  $X + Y$  nebo přímo určit hustotu  $X + Y$ , což za našich předpokladů značí inkubační dobu nemoci.

Domácnost s pozorovaným řetězcem 1 přispívá výrazem

$$E(e^{-\beta Y}) = M(-\beta), \quad (4.5)$$

kde  $M$  značí momentovou vytvořující funkci náhodné veličiny  $Y$ .

Složitější případ nastane u domácnosti s pozorovaným řetězcem 1  $\rightarrow$  1. Označme  $W_1$  okamžik, ve kterém se poprvé objeví symptomy nemoci u primárního případu, a  $W_2$  stejný okamžik u sekundárního případu. Tato domácnost s pozorováním  $w_2 - w_1$  přispívá do věrohodnostní funkce výrazem

$$\{1 - M(-\beta)\} f_{w_2-w_1}(w_2 - w_1 | 1 \rightarrow 1).$$

Ještě musíme vyjádřit podmíněnou hustotu  $W_2 - W_1$ . Všimněme si, že  $W_2 - W_1 = X_2 + Y_1 - (Y_1 - Z)$ , kde  $Z$  je čas mezi začátkem infekčního období

primárního případu a okamžikem nakažení sekundárního případu a  $Y_1 - Z$  je délka období od nakažení sekundárního případu do konce infekčního období primárního případu. Podmíněné rozdělení  $(Y_1, Z)$  je dáno hustotou

$$f_{Y_1, Z}((y, z)|Y_1 > Z) = \frac{f_Y(y)\beta e^{-\beta z}}{1 - M(-\beta)}, \quad 0 \leq z \leq y < \infty.$$

Transformací náhodných veličin dostaneme podmíněné rozdělení  $Y_1 - Z$  dané hustotou

$$f_{Y_1 - Z}(a|Y_1 > Z) = \frac{\beta e^{\beta a} \int_a^\infty e^{-\beta u} f_Y(u) \mathrm{d}u}{1 - M(-\beta)}, \quad a \geq 0$$

a rozdělení  $W_2 - W_1$  dané hustotou

$$f_{W_2 - W_1}(w|1 \rightarrow 1) = \int_0^\infty f_{X+Y}(a+w) f_{Y_1 - Z}(a|Y_1 > Z) \mathrm{d}a.$$

Nakonec dostaneme, že domácnost s pozorovaným řetězcem  $1 \rightarrow 1$  přispívá do věrohodnostní funkce výrazem

$$\int_0^\infty f_{X+Y}(a+w_2-w_1) \beta e^{\beta a} \int_a^\infty e^{-\beta u} f_Y(u) \mathrm{d}u \mathrm{d}a. \quad (4.6)$$

Kompletní věrohodnostní funkci dostaneme vynásobením příspěvků všech postižených domácností. Opět musíme upřesnit rozdělení náhodných veličin  $Y$  a  $X+Y$ . V tomto případě můžeme použít stejná rozdělení jako u předchozího modelu.

### 4.3 Konstantní délka infekčního období

Uvažujme stejnou situaci jako v předchozím případě se zjednodušením, že  $Y$  bude konstanta, označme ji  $\mu_Y$ . Příspěvky do věrohodnostní funkce z domácností s řetězcem typů  $2, 1$  a  $1 \rightarrow 1$  jsou vyjádřeny výrazy (4.4), (4.5) a (4.6). Ty se teď zjednoduší na

$$\int_0^\infty f_X(x+w_2-w_1) f_X(x) \mathrm{d}x, \quad \exp(-\beta \mu_Y)$$

a

$$\beta \int_0^{\mu_Y} \beta e^{-\beta x} f_X(w_2-w_1-x) \mathrm{d}x.$$

Všimněme si podobnosti těchto výrazů s výrazy (4.1), (4.2) a (4.3).



# Kapitola 5

## Závěr

Epidemiologové často považují matematickou teorii infekčních nemocí spíše za teoretické cvičení než za poznatky s praktickým využitím. Doufám, že tato práce alespoň naznačila výhody používání matematických modelů v praxi. Analýza infekčních chorob nám pomáhá porozumět biologickým a sociologickým mechanismům, které jsou hlavními příčinami šíření chorob. Nejen konečný model, ale i všechny nevhodné modely odmítnuté při hledání odpovídajícího modelu, poskytují informace o charakteristikách nemoci. Každé porovnání s modelem buď potvrdí nebo vyloučí nějaký význačný rys šíření nemoci. Z konečného modelu potom získáme odhady parametrů, které udávají například infekčnost nemoci nebo průměrnou délku latentního a infekčního období. Čím více porozumíme infekčnímu nemocem, tím lépe můžeme regulovat jejich šíření.

# Literatura

- [1] Becker N. G.: *Analysis of Infectious Disease Data*, Chapman and Hall, London, 1989.