

Univerzita Karlova v Praze  
Matematicko-fyzikální fakulta

## BAKALÁŘSKÁ PRÁCE



Ivana Žohová

### **Aplikace Markovových procesů při modelování průběhu choroby HIV**

Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: Mgr. Michal Kulich, Ph.D.

Studijní program: matematika, obecná matematika

2006

Na tomto místě bych chtěla poděkovat vedoucímu bakalářské práce Mgr. Michalu Kulichovi, Ph.D. za jeho pomoc při vypracovávání této práce.

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

V Praze dne 5. 8. 2006

Ivana Žohová

# Obsah

Seznam značení	5
Úvod	6
1 Choroba způsobená virem HIV	7
2 Modelování inkubačních a infekčních dob	8
2.1 Stochastické modely . . . . .	8
2.2 Vlastnosti modelů . . . . .	10
3 Markovovy procesy a jejich aplikace při modelování	14
3.1 Teorie Markovových procesů . . . . .	14
3.2 Modelování průběhu choroby HIV . . . . .	18
3.3 Vhodnost Markovových modelů . . . . .	27
Literatura	28

**Název práce:** Aplikace Markovových procesů při modelování průběhu choroby HIV

**Autor:** Ivana Žohová

**Katedra:** Katedra pravděpodobnosti a matematické statistiky

**Vedoucí bakalářské práce:** Mgr. Michal Kulich, Ph.D.

**e-mail vedoucího:** kulich@karlin.mff.cuni.cz

**Abstrakt:** V předložené práci se zabýváme stochastickým modelováním průběhu choroby HIV, především pak inkubační doby AIDS. Hlavní pozornost je věnována modelování pomocí homogenních Markovových procesů se spojitým časem a konečnou množinou stavů. V práci nejprve studujeme některá spojitá rozdělení, a to především z hlediska jejich vhodnosti k modelování inkubační doby AIDS. Další část je věnována konstrukci Markovových modelů. Průběh choroby HIV je nejprve modelován pomocí dvou konkrétních modelů a poté pomocí zobecněného modelu. V modelech jsou odvozena rozdělení inkubační doby AIDS.

**Klíčová slova:** choroba HIV, inkubační doba AIDS, homogenní Markovův proces, Markovův model

**Title:** Applications of Markov processes to modeling HIV disease progression

**Author:** Ivana Žohová

**Department:** Department of Probability and Mathematical Statistics

**Supervisor:** Mgr. Michal Kulich, Ph.D.

**Supervisor's e-mail address:** kulich@karlin.mff.cuni.cz

**Abstract:** In the present work we study stochastic modeling of HIV disease progression, and especially of the AIDS incubation period. The main attention is paid to models based on continuous time homogeneous Markov processes with a finite number of states. Firstly, in this work, some continuous distributions are considered and their suitability to model the AIDS incubation period is investigated. In the next section, we attend to the construction of Markov models. HIV disease progression is at first patterned through two concrete models, then a general model is considered. In models the AIDS incubation distributions are derived.

**Keywords:** HIV disease, AIDS incubation period, homogeneous Markov process, Markov model

# Seznam značení

$\mathbb{N}_0$	množina celých nezáporných čísel
$\mathbb{R}$	množina reálných čísel
$\mathbf{p}$	sloupcový vektor
$\mathbf{p}^T$	vektor transponovaný k vektoru $\mathbf{p}$
$\mathbf{e}_i$	jednotkový sloupcový vektor s číslem 1 na $i$ -tém místě
$\mathbf{B}$	matice
$\mathbf{B}^T$	matice transponovaná k matici $\mathbf{B}$
$\mathbf{I}$	jednotková matice
$\mathbf{I}_k$	jednotková matice řádu $k$
$\delta_{ij}$	Kroneckerovo delta
$f * g$	konvoluce funkcí $f$ a $g$

# Úvod

Onemocnění AIDS (syndrom získaného imunodeficitu) je poslední fází choroby způsobené virem HIV (dále jen choroby HIV). První zdokumentované případy AIDS se objevily na počátku 80. let 20. století mezi homosexuály ve velkých amerických městech jako New York, Los Angeles a San Francisco. Choroba HIV a zejména její poslední fáze AIDS se brzy staly celosvětovou hrozbou takových rozměrů, že hovoříme o pandemii HIV/AIDS. Podle údajů Světové zdravotnické organizace (WHO) bylo doposud virem HIV infikováno kolem 50 milionů lidí, z toho již 16 milionů zemřelo na AIDS. V Evropě žije v současné době kolem 2 milionů infikovaných jedinců.

Závažnost pandemie vedla mimo jiné k nutnosti zabývat se otázkou modelování průběhu a šíření choroby HIV. Případné modely mohou být konstruovány na základě deterministického nebo stochastického přístupu. Protože však jevy, se kterými se v těchto modelech pracuje, jsou pevně spjaty s pojmy jako variabilita, neurčitost, náhodnost apod., je při formulování modelů často vhodnější stochastický přístup.

Hlavním cílem této práce je popsat využití homogenních Markovových procesů se spojitým časem při modelování průběhu choroby HIV. Práce je rozdělena do tří kapitol. V první kapitole jsou uvedeny klinické projevy choroby HIV a způsoby jejího přenosu. Druhá a třetí kapitola jsou věnovány zejména problematice stochastického modelování inkubační doby AIDS (dále jen inkubační doby).

Ve druhé kapitole jsou popsána některá statistická rozdělení, zaměřujeme se především na posouzení jejich vhodnosti jako modelů inkubační doby.

Třetí kapitola se zabývá konstrukcí Markovových modelů. V úvodní části je shrnuta základní teorie potřebná při modelování. Průběh choroby HIV je modelován pomocí dvou konkrétních modelů, pozornost je věnována zejména odvození hustoty inkubační doby včetně jejího odhadu a odvození pravděpodobností přechodu. Dále je v práci popsán zobecněný model a v tomto modelu odvozen explicitní tvar hustoty inkubační doby. V důkazu tohoto odvození lze nalézt postup, jak k tomuto vyjádření dojít. Nicméně, aplikace těchto výsledků v praxi je numericky poměrně náročná.

# Kapitola 1

## Choroba způsobená virem HIV

Choroba HIV začíná obvykle jako lehké chřipkové onemocnění (tzv. akutní infekce). Po 2 až 3 týdnech nastává latentní fáze, která trvá několik let. V tomto období pacienti nemívají žádné vnější příznaky, dochází však u nich k postupnému oslabování imunitního systému. Důsledkem toho se objevují některé obtíže (zduřené lymfatické uzliny, únava, ztráta tělesné hmotnosti, průjem), které jsou příznakem další, tzv. ARC fáze (AIDS-Related Complex). Zhoršující se obranyschopnost organismu vede k selhávání imunity a rozvíjí se onemocnění AIDS (Acquired Immunodeficiency Syndrome). Tato fáze je charakterizována rozvojem jedné nebo více chorob v důsledku těžkého poškození imunitního systému (např. pneumonie, tuberkulóza, některé nervové poruchy, některé typy nádorů). Pacienti s projevy rozvinutého AIDS přežívají obvykle 1 až 2 roky.

Původcem nemoci je virus HIV (Human Immunodeficiency Virus) patřící mezi retroviry. V napadeném organismu tento virus přednostně infikuje  $CD4^+$  T lymfocyty (dále jen T4 buňky). V několika prvních měsících po nákaze se počet T4 buněk výrazně sníží z normálního stavu kolem 1125 T4 buněk v  $\mu l$  krve na asi 800 buněk v  $\mu l$  krve. V dalším průběhu se úbytek těchto buněk zpomalí, nicméně pokračuje dále. Množství T4 buněk v krvi proto slouží jako významný ukazatel průběhu choroby HIV.

Dosud jsou známy dvě formy viru HIV: HIV-1 a HIV-2. Virus HIV-1 byl identifikován jako první a je příčinou většiny onemocnění HIV v USA a dalších rozvinutých zemích. Virus HIV-2 byl poprvé izolován u pacientů v západní Africe. V porovnání s HIV-1 je méně patogenický a průběh nemoci je mírnější. HIV-2 je blízce příbuzný jinému viru označovanému jako SIV (Simian Immunodeficiency Virus). Tento virus byl objeven v krvi některých druhů afrických opic. Podle všeobecně přijímané domněnky mají SIV a viry HIV společného předchůdce.

Choroba HIV se šíří nechráněným pohlavním stykem, u narkomanů společným užíváním injekčních jehel, možný je i přenos z infikované ženy na její plod.

Více informací o viru HIV, chorobě HIV a nemoci AIDS lze nalézt na stránkách 3. lékařské fakulty Univerzity Karlovy [10] nebo v [7, str. 1–20].

# Kapitola 2

## Modelování inkubačních a infekčních dob

### 2.1 Stochastické modely

Ve druhé kapitole se budeme věnovat problematice stochastického modelování inkubační a infekční doby choroby HIV a inkubační doby AIDS. Jelikož se definice těchto dob mohou v odborné literatuře lišit, upřesníme nyní pojmy, se kterými budeme dále pracovat.

*Inkubační dobou choroby HIV* označíme období mezi infikací organismu virem HIV a prvními příznaky tzv. akutní infekce. Toto období trvá obvykle 1 až 3 měsíce. *Infekční dobou choroby HIV* budeme nazývat období, během něhož může infikovaný jedinec přenášet chorobu na ostatní. Toto období začíná téměř okamžitě po vniknutí viru do organismu a trvá až do konce pacientova života. *Inkubační dobou AIDS* označíme dobu mezi nákazou organismu virem HIV a rozvojem klinického stádia AIDS. (Právě toto období je někdy označováno jako inkubační doba choroby HIV.)

Délky výše definovaných dob se mohou u jednotlivých jedinců lišit. Cílem této podkapitoly je zkoumat, jak tuto variabilitu v délce dob popsat pomocí nějakého vhodného statistického rozdělení. Protože postup, který zde popíšeme, lze aplikovat na všechny tři výše definované doby (a obecně při studiu jakékoli infekční nemoci), budeme se věnovat pouze modelování inkubační doby AIDS (dále jen inkubační doby).

Inkubační dobu budeme reprezentovat spojitou náhodnou veličinou  $T$  s oborem hodnot  $R_T = [0, \infty)$  představující čas. Definujme její distribuční funkci  $F(t) = P(T \leq t)$ ,  $t \in \mathbb{R}$  a předpokládejme, že platí  $F(0) = 0$ . Hustotu veličiny  $T$  označíme  $f(t)$  a funkci přežití  $S(t)$ .

Při hledání vhodného rozdělení veličiny  $T$  je třeba mít na paměti, že inkubační doba je relativně dlouhá (průměrně kolem 10 let) a značně variabilní. Je také přirozené očekávat, že čím déle inkubační doba trvá, tím vyšší je podmíněná pravděpodobnost, že se v dalším okamžiku rozvine klinické stádium AIDS.



Nyní uvedeme některá statistická rozdělení a posoudíme, zda jsou jako modely inkubační doby vhodná. U některých rozdělení zmíníme tvar *rizikové funkce*, tedy funkce, pro kterou platí

$$\theta(t) = \lim_{h \rightarrow 0^+} \frac{P(t < T \leq t + h \mid T > t)}{h} = \lim_{h \rightarrow 0^+} \frac{F(t + h) - F(t)}{hS(t)} = \frac{f(t)}{S(t)}$$

pro všechna  $t \in \mathbb{R}$ , pro která limita existuje. Tento vztah lze ekvivalentně vyjádřit diferenciální rovnicí

$$\theta(t) = -\frac{d \ln S(t)}{dt}$$

s počáteční podmínkou  $S(0) = 1$ . Protože v našem případě je  $f(t) = 0$  pro všechna  $t \leq 0$ , budeme hustotou nadále rozumět hustotu na intervalu  $(0, \infty)$ . Obdobně pro rizikovou funkci.

### Weibullovo a exponenciální rozdělení

Nechť  $\alpha > 0, \beta > 0$ . *Weibullovo rozdělení*  $W(\alpha, \beta)$  má hustotu danou

$$f(t) = \alpha \beta t^{\alpha-1} \exp(-\beta t^\alpha), t > 0.$$

Riziková funkce má potom tvar

$$\theta(t) = \alpha \beta t^{\alpha-1}, t > 0.$$

Vlastnosti rizikové funkce určuje především hodnota parametru  $\alpha$ . Je-li  $\alpha > 1$ , je riziková funkce rostoucí. Tímto případem se budeme zabývat v další podkapitole. Naopak pro  $0 < \alpha < 1$  je riziková funkce klesající, takové rozdělení proto nemůže být pro modelování inkubační doby vhodné.

Je-li  $\alpha = 1$ , jedná se o *exponenciální rozdělení*  $\text{Exp}(\beta)$  s konstantní rizikovou funkcí. Přestože se exponenciální rozdělení často používá k modelování náhodné doby čekání na nějakou událost, jako model inkubační doby není příliš vhodné. Exponenciální rozdělení je totiž rozdělením bez paměti: Má-li veličina  $T$  exponenciální rozdělení, platí  $P(T > t + h \mid T > t) = P(T > h)$  pro každé  $t > 0, h > 0$ .

### Gama rozdělení

Nechť  $\alpha > 0, \beta > 0$ . *Gama rozdělení*  $\text{Ga}(\alpha, \beta)$  je určeno hustotou

$$f(t) = \frac{\beta^\alpha}{\Gamma(\alpha)} t^{\alpha-1} \exp(-\beta t), t > 0,$$

kde  $\Gamma(\alpha)$  je gama funkce definovaná pro  $\alpha > 0$ . Je-li  $\alpha = 1$ , získáme stejně jako u Weibullova rozdělení exponenciální rozdělení  $\text{Exp}(\beta)$ . Pro  $0 < \alpha < 1$  je riziková funkce gama rozdělení klesající a pro  $\alpha > 1$  je rostoucí. V obou případech riziková funkce konverguje k asymptotě  $\beta$  pro  $t \rightarrow \infty$ . Pro velká  $t$  má tedy tato funkce podobné vlastnosti jako riziková funkce rozdělení  $\text{Exp}(\beta)$ .

### Log-logistické rozdělení

Nechť  $\alpha \in \mathbb{R}$ ,  $\beta > 0$ . Logistické rozdělení  $L(\alpha, \beta)$  má hustotu tvaru

$$g(x) = \frac{\beta \exp[-(\alpha + \beta x)]}{\{1 + \exp[-(\alpha + \beta x)]\}^2}, \quad x \in \mathbb{R}.$$

Takové rozdělení není k modelování inkubační doby vhodné, protože jeho hustota je kladná všude na  $\mathbb{R}$ , zatímco veličina reprezentující inkubační dobu je nezáporná. Můžeme však uvažovat log-logistické rozdělení. Definujme veličinu  $T$  vztahem  $T = \exp(X)$ , kde  $X \sim L(\alpha, \beta)$ . Potom má veličina  $T$  *log-logistické rozdělení*, tedy  $\ln T \sim L(\alpha, \beta)$ , a pro hustotu  $T$  platí

$$f(t) = \frac{dF(t)}{dt} = \frac{dP(T \leq t)}{dt} = \frac{dP(\ln T \leq \ln t)}{dt} = \frac{dG(\ln t)}{dt} = \frac{1}{t}g(\ln t), \quad t > 0,$$

kde  $g$  je hustota veličiny  $X$  a  $G, F$  jsou distribuční funkce veličin  $X, T$  v tomto pořadí.

### Log-normální rozdělení

Nechť  $\mu \in \mathbb{R}$ ,  $\sigma > 0$ . Uvažujme normální rozdělení  $N(\mu, \sigma^2)$  se střední hodnotou  $\mu$  a rozptylem  $\sigma^2$ . Stejně jako logistické rozdělení má normální rozdělení hustotu kladnou na celém  $\mathbb{R}$ . Nechť veličina  $X$  má rozdělení  $N(\mu, \sigma^2)$  s hustotou  $g$ . Potom náhodná veličina  $T = \exp(X)$  má *log-normální rozdělení*. Její hustotu odvodíme analogicky jako u log-logistického rozdělení, tedy je

$$f(t) = \frac{1}{t}g(\ln t) = \frac{1}{\sqrt{2\pi} \sigma t} \exp \left[ -\frac{(\ln t - \mu)^2}{2\sigma^2} \right], \quad t > 0.$$

### Log-Cauchyovo rozdělení

Nechť  $\alpha \in \mathbb{R}$ ,  $\beta > 0$ . Cauchyovo rozdělení  $C(\alpha, \beta)$  má hustotu

$$g(x) = \frac{1}{\pi} \frac{\beta}{\beta^2 + (x - \alpha)^2}, \quad x \in \mathbb{R}.$$

Pro toto rozdělení neexistuje konečná střední hodnota. Má-li veličina  $X$  Cauchyovo rozdělení, obdobně jako v předchozích dvou případech definujeme nezápornou veličinu  $T = \exp(X)$  s *log-Cauchyovým rozdělením*.

## 2.2 Vlastnosti modelů

Při posuzování, jak jsou jednotlivá rozdělení vhodná k modelování inkubační doby, je třeba vzít v úvahu několik skutečností. Především je nutné si uvědomit, že obvykle máme k dispozici pouze malé soubory dat a že pozorování, ze kterých tato data pochází, probíhala relativně krátkou dobu v porovnání s délkou inkubační doby. Důsledkem toho mohli být z těchto pozorování vyloučeni jedinci s přirozeně dlouhou inkubační dobou. Proto

může být zjištěná průměrná délka inkubační doby podhodnocena. Je tedy dobré vědět, zda a v jaké míře jednotlivá rozdělení připouštějí odlehlá pozorování.

Dále je třeba mít na paměti, že průběh nemoci ovlivňuje celá řada faktorů. Zmiňme například rozdílný průběh chorob HIV-1 a HIV-2. Průběh choroby zřejmě závisí i na věku pacienta a pravděpodobně také na pohlaví a způsobu nákazy. Navíc nově vyvíjené léky dokáží podstatně prodloužit trvání inkubační doby.

Jednotlivá rozdělení můžeme porovnávat pomocí kvantilů. Parametry porovnávaných rozdělení lze zvolit různými způsoby. U některých rozdělení můžeme postupovat tak, že vyjádříme jejich střední hodnoty (případně i rozptyly) jako funkce neznámých parametrů. Hodnoty těchto parametrů zvolíme tak, aby střední hodnoty byly pro všechna porovnávaná rozdělení stejné. Pro některá rozdělení však nelze střední hodnotu elementárně vyjádřit a u některých rozdělení nemusí konečná střední hodnota vůbec existovat. V takovém případě můžeme hodnoty parametrů určit tak, aby se pro porovnávaná rozdělení shodovaly vybrané kvantily.

Nyní porovnáme některá rozdělení (pro zvolené parametry) z hlediska odlehlých pozorování.

**Příklad 2.1.** V tomto příkladu porovnáme Weibullovo rozdělení s gama rozdělením, rozdělením exponenciálním, log-normálním, log-logistickým a log-Cauchyovým. Parametry těchto rozdělení zvolíme tak, aby všechna rozdělení měla stejný medián a případně také stejný třetí kvartil (0,75-kvartil). U rozdělení s takto zvolenými parametry určíme některé další kvantily a pomocí nich rozdělení vzájemně porovnáme. Zdůrazněme, že takto určené parametry zde slouží pouze pro ilustrativní porovnání. Hodnoty budeme v příkladu zaokrouhlovat na 3 desetinná místa

Vyjdeme z výzkumu, který provedli Lui a kol. Ti sledovali délku inkubační doby AIDS u skupiny homosexuálních mužů. Inkubační dobu modelovali pomocí Weibullova rozdělení  $W(\alpha, \beta)$  a jeho parametry odhadli z dat metodou maximální věrohodnosti:  $\hat{\alpha} = 2,571000$  a  $\hat{\beta} = 0,003807$ .

Uvažujme tedy Weibullovo rozdělení  $W(\hat{\alpha}, \hat{\beta})$ , medián tohoto rozdělení označíme  $w_{0,50}$  a třetí kvartil  $w_{0,75}$ . U exponenciálního rozdělení  $\text{Exp}(\beta)$  zvolíme parametr  $\beta$  tak, aby byl příslušný medián roven  $w_{0,50}$ , tedy  $\beta = \ln 2/w_{0,50}$ . U gama rozdělení  $\text{Ga}(\alpha, \beta)$  budeme vzhledem k početní náročnosti požadovat pouze shodu jeho mediánu s  $w_{0,50}$ . Z teorie matematické statistiky (transformace náhodných veličin) je známo, že, má-li veličina  $X$  rozdělení  $\text{Ga}(\alpha, \beta)$ , potom veličina  $Y = kX$ , kde  $k > 0$ , má rozdělení  $\text{Ga}(\alpha, \beta/k)$ . Označme  $F, G$  distribuční funkce veličin  $X, Y$  v tomto pořadí,  $x_q$   $q$ -kvantil veličiny  $X$  a  $y_q$   $q$ -kvantil veličiny  $Y$ ,  $q \in (0, 1)$ . Jelikož platí

$$G(y_q) = P(Y \leq y_q) = P\left(X \leq \frac{y_q}{k}\right) = F\left(\frac{y_q}{k}\right) = F(x_q),$$

Tabulka 2.1: Vybrané q-quantily v rocích pro Weibullovo, gama, exponenciální, log-normální, log-logistické a log-Cauchyovo rozdělení se stejnými mediány

q	Weibull.	gama	exponen.	log-norm.	log-logis.	log-Cauchy.
0,25	5,378	5,634	3,142	5,782	5,782	5,782
0,50	7,571	7,571	7,571	7,571	7,571	7,571
0,75	9,913	9,911	15,141	9,913	9,913	9,913
0,95	13,378	14,037	32,720	14,610	15,593	41,532
0,99	15,813	17,503	50,298	19,185	23,381	$4,026 \cdot 10^4$
0,999	18,514	21,971	75,447	26,036	41,232	$1,409 \cdot 10^{38}$

jsou kvantily  $y_q$  určeny vztahem  $y_q = kx_q$ . Chceme-li pro veličinu  $Y$  určit parametry tak, aby platilo  $y_{0,50} = w_{0,50}$ , můžeme postupovat následujícím způsobem: Zvolíme nějaké  $\tilde{\alpha} > 0$ ,  $\tilde{\beta} > 0$  a pro veličinu  $X \sim \text{Ga}(\tilde{\alpha}, \tilde{\beta})$  určíme kvantil  $x_{0,50}$  (např. pomocí programu R). Veličinu  $Y$  definujeme potom vztahem  $Y = kX$ , kde konstanta  $k$  splňuje rovnici  $kx_{0,50} = w_{0,50}$ . My jsme zvolili  $\tilde{\alpha} = 6$  a  $\tilde{\beta} = 1$ , třetí kvartil  $y_{0,75}$  má v tomto případě hodnotu blízkou hodnotě  $w_{0,75}$ . Poznamenejme ještě, že jsme záměrně volili  $\tilde{\alpha} > 1$ . Pro  $\tilde{\alpha} = 1$  bychom totiž dostali exponenciální rozdělení, kterým jsme se již zabývali výše. Pro  $\tilde{\alpha} < 1$  bychom dostali gama rozdělení s klesající rizikovou funkcí, takové rozdělení není k modelování inkubační doby AIDS vhodné.

U log-normálního, log-logistického a log-Cauchyova rozdělení budeme požadovat, aby se shodovaly jejich mediány s  $w_{0,50}$  a třetí kvantily s  $w_{0,75}$ . Pro normální, logistické i Cauchyovo rozdělení platí, že lineární transformací se typ rozdělení nezmění. Toho nyní využijeme. Určíme pouze hledané log-normální rozdělení, při odvozování log-logistického a log-Cauchyova rozdělení lze postupovat analogicky. Uvažujme veličinu  $Z \sim N(0, 1)$ , její kvantily označíme  $z_q$ . Pro libovolné  $\alpha, \beta \in \mathbb{R}$  má veličina  $\alpha + \beta Z$  normální rozdělení a nezáporná veličina  $T = \exp(\alpha + \beta Z)$  odpovídající log-normální rozdělení. Kvantily  $t_q$  veličiny  $T$  splňují vztah  $t_q = \exp(\alpha + \beta z_q)$ , neboť pro  $t_q > 0$  platí

$$H(t_q) = P(T \leq t_q) = P\left(Z \leq \frac{\ln t_q - \alpha}{\beta}\right) = F\left(\frac{\ln t_q - \alpha}{\beta}\right) = F(z_q),$$

kde  $F, H$  značí distribuční funkce veličin  $Z$  a  $T$  v daném pořadí. Kvantily  $z_{0,50}$  a  $z_{0,75}$  určíme pomocí programu R. Konstanty  $\alpha, \beta$  musí splňovat rovnice  $\exp(\alpha + \beta z_{0,50}) = w_{0,50}$ ,  $\exp(\alpha + \beta z_{0,75}) = w_{0,75}$ , odkud již snadno  $\alpha$  a  $\beta$  vypočteme.

Úloha je vyřešena pomocí softwaru R. Poznamenejme, že program R pracuje s Weibullovým rozdělením  $\widetilde{W}(\alpha, \gamma)$  s parametry  $\alpha, \gamma = \beta^{-\frac{1}{\alpha}}$ , kde  $\alpha, \beta$  značí parametry námi uvažovaného rozdělení  $W(\alpha, \beta)$ . Výsledky

úlohy jsou zapsány v tabulce 2.1. Porovnáním druhého a třetího sloupce tabulky je vidět, že pro vhodně zvolené parametry mají Weibullovo a gama rozdělení podobné kvantily. Porovnáme-li s těmito rozděleními exponenciální rozdělení ve čtvrtém sloupci, vidíme, že pro daný parametr jsou jeho  $q$ -kvantily pro  $q \geq 0.95$  několikanásobně vyšší než odpovídající  $q$ -kvantily Weibullova a gama rozdělení. Ve vzorku generovaném z takto určeného rozdělení by přibližně 1% dat překročilo hodnotu 50 roků. Z výsledků dále plyne, že log-normální a především log-logistické rozdělení obsahují pro dané parametry více odlehlých pozorování než uvažované Weibullovo rozdělení. Tato rozdělení tedy v porovnání s rozdělením  $W(\hat{\alpha}, \hat{\beta})$  lépe zohledňují jedince s velmi dlouhou inkubační dobou. Log-Cauchyovo rozdělení v posledním sloupci k modelování inkubační doby AIDS vhodné není, neboť obsahuje výrazně odlehlá pozorování. Ve vzorku generovaném z tohoto rozdělení by bylo přibližně 5% dat větších než 41,5 roků a přibližně 1% dat by dokonce přesáhlo hodnotu 40260 roků, tedy dobu, která výrazně překračuje délku lidského života.  $\diamond$

Další možnosti, jak porovnávat statistická rozdělení, především z hlediska odlehlých pozorování, lze nalézt v [7, str. 37–42].

# Kapitola 3

## Markovovy procesy a jejich aplikace při modelování

### 3.1 Teorie Markovových procesů

Významnou roli mezi stochastickými modely používanými v epidemiologii hrají náhodné procesy, mezi ně patří i tzv. Markovovy procesy.

**Definice 3.1.** Necht'  $(\Omega, \mathcal{A}, P)$  je pravděpodobnostní prostor a  $T \subset \mathbb{R}$ . Systém náhodných veličin  $\{X(t), t \in T\}$  definovaných na  $(\Omega, \mathcal{A}, P)$  se nazývá *náhodný proces*.

Parametr  $t$  má význam času. Je-li  $T$  množina diskrétních hodnot, hovoříme o *procesu s diskrétním časem*. Je-li  $T$  interval, hovoříme o *procesu se spojitým časem*. Hodnoty náhodných veličin  $X(t), t \in T$  se nazývají *stavy*, jejich množinu budeme značit  $S$ . Nabývají-li veličiny  $X(t)$  pouze diskrétních hodnot, říkáme, že  $\{X(t), t \in T\}$  je *proces s diskrétními stavy*. Je-li  $S$  interval, hovoříme o *procesu se spojitými stavy*.

V další části se budeme věnovat Markovovým procesům se spojitým časem a diskrétními stavy. Připomeneme pouze základní terminologii a některé důležité vlastnosti, které využijeme v druhé podkapitole při modelování. Ucelenou teorii o těchto procesech lze nalézt např. v [9].

**Definice 3.2.** Necht' je dán pravděpodobnostní prostor  $(\Omega, \mathcal{A}, P)$ . Systém celočíselných náhodných veličin  $\{X(t), t \geq 0\}$  definovaných na  $(\Omega, \mathcal{A}, P)$  se nazývá *Markovův proces se spojitým časem a množinou stavů  $S$* , jestliže

$$\begin{aligned} P(X(t) = j \mid X(s) = i, X(t_n) = i_n, \dots, X(t_0) = i_0) = \\ = P(X(t) = j \mid X(s) = i) \end{aligned} \quad (3.1)$$

pro libovolné  $n \in \mathbb{N}_0$ , pro všechna  $i_0, \dots, i_n, i, j \in S$  a pro všechna  $0 \leq t_0 < t_1 < \dots < t_n < s < t$ , pro která je  $P(X(s) = i, X(t_n) = i_n, \dots, X(t_0) = i_0) > 0$ .

Vztah (3.1) vyjadřuje tzv. *markovskou vlastnost*, která je základní charakteristikou Markovových procesů. Podmíněné pravděpodobnosti  $P(X(t) = j \mid X(s) = i) = p_{ij}(s, t)$  (jsou-li definovány) budeme nazývat *pravděpodobnosti přechodu* ze stavu  $i$  v čase  $s$  do stavu  $j$  v čase  $t$ . Pokud pro každé  $i, j \in S$  platí

$$p_{ij}(s, t) = p_{ij}(t - s), \forall s, t : 0 \leq s < t,$$

říkáme, že Markovův proces je *homogenní*. Nadále se budeme věnovat pouze teorii homogenních procesů. Homogenní Markovův proces se spojitým časem budeme v této podkapitole dále označovat jako (CTHMP, continuous time homogeneous Markov process).

Pravděpodobnosti přechodu můžeme pro každé  $t \geq 0$  sestavit do tzv. *matice pravděpodobností přechodu*  $\mathbf{P}(t) = \{p_{ij}(t), i, j \in S\}$ . Zřejmě pro každé  $i, j \in S$  a každé  $t \geq 0$  platí

$$p_{ij}(t) \geq 0; \quad \sum_{j \in S} p_{ij}(t) = 1; \quad p_{ij}(0) = \delta_{ij},$$

kde  $\delta_{ij}$  značí Kroneckerovo delta. V důsledku markovské vlastnosti splňují pravděpodobnosti přechodu tzv. *Chapmanovu-Kolmogorovovu rovnici*.

**Věta 3.1 (Chapmanova-Kolmogorovova rovnice).** *Nechť je dán (CTHMP) s množinou diskrétních stavů  $S$ . Potom pro pravděpodobnosti přechodu platí*

$$p_{ij}(s + t) = \sum_{k \in S} p_{ik}(s)p_{kj}(t), \quad \forall i, j \in S, \quad \forall s > 0, \forall t > 0.$$

*Důkaz.* Viz [3, str. 115]. □

Protože při modelování budeme pracovat pouze s procesy s konečnou množinou stavů, v další teorii se omezíme na tento případ.

Důležitou roli v teorii Markovových procesů se spojitým časem hrají *Kolmogorovy diferenciální rovnice*. Ty lze odvodit z Chapmanovy-Kolmogorovy rovnice. K tomu je však nejprve nutné definovat tzv. intenzity přechodu.

**Věta 3.2.** *Nechť je dán (CTHMP) s konečnou množinou stavů  $S$ . Potom pro každé  $i, j \in S$  existují limity*

$$\lim_{\Delta \rightarrow 0+} \frac{1 - p_{ii}(\Delta)}{\Delta} = -q_{ii}, \quad -\infty < q_{ii} \leq 0,$$

$$\lim_{\Delta \rightarrow 0+} \frac{p_{ij}(\Delta)}{\Delta} = q_{ij}, \quad 0 \leq q_{ij} < \infty, \quad \text{kde } i \neq j,$$

a pro takto definované konstanty  $q_{ij}$  platí  $g_{ii} = -\sum_{j \neq i} q_{ij}$ .

*Důkaz.* Důkaz existence konstant  $q_{ii}$  a  $q_{ij}, j \neq i$  je uveden v [4, str. 131–133], věty 4 a 5. Důkaz vztahu  $g_{ii} = -\sum_{j \neq i} q_{ij}$  lze nalézt v [9, str. 74].  $\square$

**Definice 3.3.** Konstanty  $q_{ij}, i, j \in S$  definované v předchozí větě se nazývají *intenzity přechodu* ze stavu  $i$  do stavu  $j$ . Konstanty  $q_i, i \in S$  definované předpisem  $q_i = -q_{ii}$  se nazývají *celkové intenzity*. Matici  $\mathbf{Q} = \{q_{ij}, i, j \in S\}$  říkáme *matice intenzit přechodu*.

**Věta 3.3 (Kolmogorovy diferenciální rovnice).** *Nechť je dán (CTHMP) s konečnou množinou stavů  $S$  a intenzitami přechodu  $q_{ij}, i, j \in S$ . Potom pravděpodobnosti přechodu  $p_{ij}(t)$  splňují pro všechna  $i, j \in S$  a  $t > 0$  systémy diferenciálních rovnic*

$$p'_{ij}(t) = \sum_{k \in S} p_{ik}(t)q_{kj} \quad (\text{prospektivní rovnice}), \quad (3.2)$$

$$p'_{ij}(t) = \sum_{k \in S} q_{ik}p_{kj}(t) \quad (\text{retrospektivní rovnice}) \quad (3.3)$$

s počáteční podmínkou  $p_{ij}(0) = \delta_{ij}$ .

*Důkaz.* Viz věta 3.9 v [9, str. 82].  $\square$

Přirozeně nás zajímá, zda a jaké mají soustavy (3.2) a (3.3) řešení. Nejprve však připomeňme pojem *exponenciální maticové funkce*. Tato funkce je definována pro libovolnou čtvercovou matici  $\mathbf{B}$  vztahem

$$\exp(\mathbf{B}t) = \sum_{k=0}^{\infty} \frac{\mathbf{B}^k t^k}{k!},$$

kde  $\mathbf{B}^0$  je jednotková matice odpovídajícího řádu. Nyní již můžeme uvést větu týkající se řešení soustav (3.2) a (3.3).

**Věta 3.4.** *Nechť je dána čtvercová matice konečného řádu  $\mathbf{Q} = \{q_{ij}\}$ , pro jejíž prvky platí  $q_{ij} \geq 0, i \neq j, q_{ii} = -\sum_{j \neq i} q_{ij}$ . Pak existuje jediné řešení, společné pro obě soustavy (3.2) a (3.3), s počáteční podmínkou  $\mathbf{P}(0) = \mathbf{I}$ , kde  $\mathbf{I}$  značí jednotkovou matici příslušného řádu. Toto řešení představuje systém pravděpodobností přechodu (CTHMP) s konečnou množinou stavů. Maticově lze řešení zapsat ve tvaru  $\mathbf{P}(t) = \exp(\mathbf{Q}t)$ .*



*Důkaz.* Viz věta 3.10 v [9, str. 85]. □

Na základě intenzit přechodu můžeme klasifikovat stavy procesu. Zdefinujeme nejprve *čas prvního návratu* do stavu  $i \in S$  jako  $\tau_i = \inf\{t \geq J : X(t) = i\}$ , kde  $J$  značí časový okamžik, kdy došlo k prvnímu přechodu, tj.  $J = \inf\{t > 0 : X(t) \neq X(0)\}$ .

**Definice 3.4.** Nechť je dán (CTHMP) s konečnou množinou stavů  $S$ . Stav  $i \in S$  se nazývá *trvalý*, jestliže buď  $q_i = 0$  (absorpční stav), nebo  $q_i > 0$  a zároveň  $P(\tau_i < \infty \mid X(0) = i) = 1$ . Stav  $i \in S$  se nazývá *přechodný*, jestliže  $q_i > 0$  a zároveň  $P(\tau_i < \infty \mid X(0) = i) < 1$ .

**Věta 3.5.** Nechť je dán (CTHMP) s konečnou množinou stavů  $S$ . Je-li  $q_i > 0$ , má doba, po kterou proces setrvává ve stavu  $i \in S$ , exponenciální rozdělení  $\text{Exp}(q_i)$ . Je-li  $q_i = 0$ , potom  $p_{ii}(t) = 1$  pro všechna  $t \geq 0$ .

*Důkaz.* Viz věta 3.5 v [9, str. 77]. □

Nyní se ještě vrátíme ke Kolmogorovovým diferenciálním rovnicím. C. L. Chiang odvodil za určitých předpokladů jejich explicitní řešení.

**Lemma 3.1.** Uvažujme (CTHMP) s konečnou množinou stavů  $S$  a intenzitami  $q_{ij}$ ,  $i, j \in S$ . Nechť platí  $S = U \cup T$ , kde  $U$  je neprázdná množina přechodných stavů a  $T$  je neprázdná množina absorpčních stavů. Bez újmy na obecnosti předpokládejme, že je  $U = \{1, \dots, n\}$ . Definujme matici  $\mathbf{V} = \{q_{ij}, i, j \in U\}$ . Nechť tato matice má reálná, navzájem různá vlastní čísla  $\rho_1, \dots, \rho_n$ . Pak pravděpodobnosti přechodu mají pro  $t > 0$  tvar

$$p_{ij}(t) = \sum_{l=1}^n \frac{A_{ij}^T(l)}{\prod_{\substack{m=1 \\ m \neq l}}^n (\rho_l - \rho_m)} \exp(\rho_l t), i, j \in U, \quad (3.4)$$

$$p_{ij}(t) = \sum_{l=1}^n \sum_{k=1}^n \frac{A_{ik}^T(l)}{\rho_l \prod_{\substack{m=1 \\ m \neq l}}^n (\rho_l - \rho_m)} [\exp(\rho_l t) - 1] q_{kj}, i \in U, j \in T, \quad (3.5)$$

$A_{ij}^T(l)$  značí algebraický doplněk prvku  $a_{ij}^T(l)$  v matici  $\mathbf{A}^T(l) = \rho_l \mathbf{I} - \mathbf{V}^T$ .

*Důkaz.* Viz [3, str. 153–158] s odkazem na [3, str. 135–137]. □

## 3.2 Modelování průběhu choroby HIV

Jedinec infikovaný virem HIV prochází různými stádii choroby. Je proto přirozené modelovat průběh nemoci pomocí stavových modelů, jejichž stavy ztotožníme se stádii nemoci. V této podkapitole budeme průběh choroby HIV modelovat pomocí homogenních Markovových procesů se spojitým časem a konečnou množinou stavů.

Možností, jak definovat systém stádií choroby HIV, je samozřejmě mnoho. Zde uvedeme příklady dvou systémů, které dále využijeme při konstrukci modelů. První systém využívá fáze popsané v první kapitole, uveden je v tabulce 3.1. Druhý systém je popsán v tabulce 3.2. Stádia choroby jsou v něm definována na základě počtu T4 buněk v krvi. Jak bylo uvedeno v první kapitole, tyto buňky virus HIV infikuje ihned po nákaze a jejich počet klesá po celou dobu onemocnění.

Tabulka 3.1: Symptomatická stádia choroby HIV

Stadium	Symptomy
$E_1$	infikován virem HIV, ale séronegativní
$E_2$	séropozitivní, ale bez zřejmých příznaků
$E_3$	fáze ARC
$E_4$	klinické stádium AIDS

Tabulka 3.2: Stádia choroby HIV založená na počtu T4 buněk v krvi

Stadium	Počet T4 buněk / $mm^3$
$E_1$	více než 899
$E_2$	700-899
$E_3$	500-699
$E_4$	350-499
$E_5$	200-349
$E_6$	0-199

Nyní uvedeme příklady konkrétních modelů průběhu choroby HIV. V modelech budeme pro zjednodušení předpokládat, že se u každého jedince infikovaného virem HIV časem rozvine AIDS.

**Příklad 3.1.** Model popsáný v tomto příkladu je převzat z [6]. Uvažujme homogenní Markovův proces  $\{X(t), t \geq 0\}$  se spojitým časem a množinou stavů  $S = \{E_1, \dots, E_5\}$ , kde  $E_1, \dots, E_4$  jsou stavy definované v tabulce 3.1 a stav  $E_5$  reprezentuje úmrtí v důsledku AIDS. Jiné příčiny smrti zde

nebudeme uvažovat. Předpokládáme, že stavy  $E_1, \dots, E_4$  jsou přechodné a stav  $E_5$  absorpční. Matice intenzit  $\mathbf{Q} = \{q_{ij}\}$  nechť má tvar

$$\mathbf{Q} = \begin{pmatrix} -\alpha_1 & \alpha_1 & 0 & 0 & 0 \\ 0 & -\alpha_2 & \alpha_2 & 0 & 0 \\ 0 & 0 & -\alpha_3 & \alpha_3 & 0 \\ 0 & 0 & 0 & -\alpha_4 & \alpha_4 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \text{ kde } \alpha_i > 0 \text{ pro každé } i = 1, \dots, 4.$$

Přechody mezi stavy jsou tedy určeny následujícím diagramem:

$$E_1 \xrightarrow{\alpha_1} E_2 \xrightarrow{\alpha_2} E_3 \xrightarrow{\alpha_3} E_4 \xrightarrow{\alpha_4} E_5.$$

Náhodnou dobu, kterou proces stráví ve stavu  $E_i$ , označíme  $T_i$ , kde  $i = 1, \dots, 4$ . Jelikož uvažujeme homogenní Markovův proces se spojitým časem, jsou  $T_i$  nezávislé náhodné veličiny, pro které platí  $T_i \sim \text{Exp}(\alpha_i)$ . Veličina  $T_i$  má tedy hustotu

$$f_i(t) = \alpha_i \exp(-\alpha_i t), t > 0.$$

Príslušné distribuční funkce označíme  $F_i(t)$  a funkce přežití  $S_i(t)$ .

Nechť je proces v čase 0 ve stavu  $E_i$ ,  $i = 1, \dots, 4$ . Náhodnou veličinu, která značí dobu čekání na vstup do stavu  $E_j$ ,  $j = i + 1, \dots, 5$ , označíme  $T_{ij}$ . Jelikož v modelu připouštíme pouze jednosměrné přechody, můžeme psát  $T_{ij} = T_i + \dots + T_{j-1}$ , kde  $T_{j-1} = 0$  pro  $j = i + 1$ . Vzhledem k nezávislosti náhodných dob lze hustotu veličiny  $T_{ij}$  zapsat pomocí konvoluce

$$g_{ij}(t) = f_i * \dots * f_{j-1}(t), t > 0. \quad (3.6)$$

S odkazem na druhou kapitolu lze inkubační dobu AIDS reprezentovat náhodnou veličinou  $T_{14}$  a infekční dobu choroby HIV veličinou  $T_{15}$ . Dále se budeme zabývat pouze rozdělením inkubační doby AIDS. Hustotu veličiny  $T_{14}$  vypočteme aplikací vztahu (3.6):

$$g_{14}(t) = \alpha_1 \alpha_2 \alpha_3 \sum_{l=1}^3 \frac{\exp(-\alpha_l t)}{\prod_{\substack{m=1 \\ m \neq l}}^3 (\alpha_l - \alpha_m)}, t > 0. \quad (3.7)$$

Príslušnou distribuční funkci označíme  $G_{14}(t)$ . Platí

$$G_{14}(t) = \int_0^t g_{14}(\tau) d\tau = \alpha_1 \alpha_2 \alpha_3 \sum_{l=1}^3 \frac{1 - \exp(-\alpha_l t)}{\alpha_l \prod_{\substack{m=1 \\ m \neq l}}^3 (\alpha_l - \alpha_m)}, t > 0. \quad (3.8)$$

Jelikož je  $T_{14} = T_1 + T_2 + T_3$  a  $T_1, T_2, T_3$  jsou nezávislé, platí pro střední hodnotu a rozptyl vztah

$$E(T_{14}) = \sum_{i=1}^3 E(T_i) = \sum_{i=1}^3 \alpha_i^{-1}, \quad (3.9)$$

$$\text{var}(T_{14}) = \sum_{i=1}^3 \text{var}(T_i) = \sum_{i=1}^3 \alpha_i^{-2}.$$

Nyní určíme pravděpodobnosti přechodu. V souladu se zavedeným označením definujeme  $p_{ij}(t) = P(X(t) = E_j | X(0) = E_i)$ , kde  $i, j \in \{1, \dots, 5\}$ . Zřejmě platí  $p_{ij}(0) = \delta_{ij}$ . Budeme se tedy nadále zabývat případem  $t > 0$ . Z vlastností modelu ihned plyne, že  $p_{ij}(t) = 0$  pro  $i > j$ . Dále platí  $p_{ii}(t) = P(T_i > t) = S_i(t)$  pro každé  $i = 1, \dots, 4$  a  $p_{55}(t) = 1$ . Nyní ještě zbývá vyřešit případ  $i < j$ . Můžeme využít např. vzorce (3.4) a (3.5). Jsou-li  $\alpha_1, \dots, \alpha_4$  navzájem různá čísla, potom dle těchto vzorců dostaneme

$$p_{ij}(t) = (-1)^{j-i} \alpha_i \cdots \alpha_{j-1} \sum_{l=i}^j \frac{\exp(-\alpha_l t)}{\prod_{\substack{m=i \\ m \neq l}}^j (\alpha_l - \alpha_m)}, \quad j \neq 5, \quad (3.10)$$

$$p_{i5}(t) = (-1)^{4-i} \alpha_i \cdots \alpha_4 \sum_{l=i}^4 \frac{1 - \exp(-\alpha_l t)}{\alpha_l \prod_{\substack{m=i \\ m \neq l}}^4 (\alpha_l - \alpha_m)}. \quad (3.11)$$

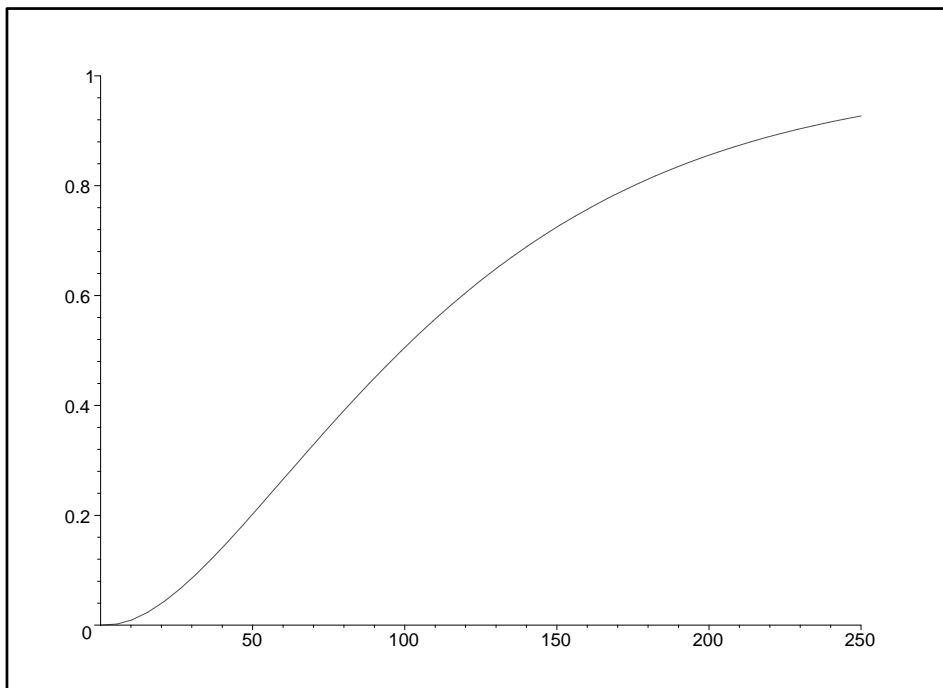
Můžeme ale postupovat i jiným způsobem, který nevyžaduje předpoklad různosti intenzit  $\alpha_i$ . Uvažujme, že proces v čase 0 vstoupil do stavu  $E_i$  a v čase  $t > 0$  se nachází ve stavu  $E_j$ . Potom proces vstoupil do stavu  $E_j$  v nějakém čase  $u \in (0, t]$  s pravděpodobností  $g_{ij}(u)du$  a zůstal v něm po dobu  $t - u$  časových jednotek s pravděpodobností  $S_j(t - u)$  pro  $j \neq 5$  a s pravděpodobností 1 pro  $j = 5$ . Platí tedy

$$p_{ij}(t) = \int_0^t g_{ij}(u) S_j(t - u) du, \quad j \neq 5,$$

$$p_{i5}(t) = \int_0^t g_{i5}(u) du.$$

Neznámé parametry  $\alpha_i, i = 1, \dots, 4$  můžeme odhadnout z dat metodou maximální věrohodnosti. Uvažujme, že máme k dispozici data pocházející od  $n$  sledovaných jedinců. Nechť u každého jedince  $k, k = 1, \dots, n$  známe stavy  $y_{km} \in S$ , ve kterých se nacházel v časech  $\tau_{km}, m = 0, 1, \dots, n_k, 0 = \tau_{k0} < \tau_{k1} \cdots < \tau_{kn_k}$ . Označme  $\alpha = (\alpha_1, \alpha_2, \alpha_3, \alpha_4)$ . Jelikož pracujeme s homogenním Markovovým procesem a osudy sledovaných jedinců považujeme za nezávislé, má věrohodnostní funkce tvar

$$L(\alpha) = \prod_{k=1}^n L_k(\alpha), \quad \text{kde } L_k(\alpha) = \prod_{m=1}^{n_k} p_{y_{k(m-1)}y_{km}}(\tau_{km} - \tau_{k(m-1)}),$$



Obrázek 3.1: Graf funkce  $\hat{G}_{14}(t)$ . Na vodorovné ose je vynesena čas v měsících.

kde  $L_k(\alpha)$  značí věrohodnostní funkci příslušející jedinci  $k$  a  $p_{ij}(t)$  jsou dány vztahy (3.10) a (3.11).

Pro úplnost uvedme hodnoty maximálně věrohodných odhadů  $\hat{\alpha}_i$  intenzit  $\alpha_i$ , které spočetli Longini a kol. (v měsíc<sup>-1</sup>):  $\hat{\alpha}_1 = 0,4571$ ,  $\hat{\alpha}_2 = 0,0190$ ,  $\hat{\alpha}_3 = 0,0159$ ,  $\hat{\alpha}_4 = 0,0424$ . Dosadíme-li  $\hat{\alpha}_i$  za  $\alpha_i$  v (3.9), dostaneme odhad střední délky inkubační doby AIDS v měsících:

$$\widehat{E}(T_{14}) = \sum_{i=1}^3 \widehat{E}(T_i) = \sum_{i=1}^3 \hat{\alpha}_i = 2,1877 + 52,6316 + 62,8931 = 117,7124,$$

což je přibližně 9,8 roku. Odhad hustoty  $\hat{g}_{14}(t)$  a odhad distribuční funkce  $\hat{G}_{14}(t)$  inkubační doby AIDS získáme dosazením  $\hat{\alpha}_i$  za  $\alpha_i$  v (3.7) a (3.8). Na obrázku 3.1 je znázorněn graf funkce  $\hat{G}_{14}(t)$ . Graf byl sestaven pomocí programu MAPLE.  $\diamond$

Stádia použitá v předešlém modelu nám umožňovala bez jakýchkoli omezení předpokládat jednosměrné přechody mezi stavy. V následujícím modelu to už platit nebude.

**Příklad 3.2.** Stejně jako v předchozím příkladu modelujeme průběh choroby HIV homogenním Markovovým procesem se spojitým časem a konečnou množinou stavů. Pro množinu stavů  $S$  nechť platí  $S = \{E_1, \dots, E_7\}$ ,

kde  $E_1, \dots, E_6$  jsou přechodné stavy definované v tabulce 3.2 a  $E_7$  absorpční stav reprezentující klinické stádium AIDS.

Při použití takto definovaných stavů je třeba mít na paměti, že zjištěný počet T4 buněk může u jedince značně kolísat, a to jak v důsledku chyb měření, tak v důsledku běžných fyziologických změn organismu. V modelu proto připustíme přechody v obou směrech. Nicméně, délka intervalů definovaných v tabulce 3.2 je dostatečně velká, aby se vliv takovéto fluktuace eliminoval. Předpokládejme tedy, že přechody mezi přechodnými stavy se řídí následujícím schématem:  $E_i \rightarrow E_j$ , kde  $j = i - 1, i + 1$  pro  $i = 2, \dots, 5$  a  $E_1 \rightarrow E_2, E_6 \rightarrow E_5$ . Dále je třeba vzít v úvahu, že seznam chorob definujících AIDS byl v průběhu času oficiálně změněn. Proto připustíme také přímé přechody typu  $E_i \rightarrow E_7, i > 1$ . Vzhledem k délce inkubační doby AIDS je však nepravděpodobné, že by se tato nemoc rozvinula přímo z časného stádia. V modelu můžeme například předpokládat, že přechody  $E_i \rightarrow E_7$  nastávají pouze pro  $i = 1, 2, 3$ .

Intenzity přechodu ze stavu  $E_i$  do stavu  $E_j$  označíme  $q_{ij}, i, j \in \{1, \dots, 7\}$ . Matice intenzit  $\mathbf{Q} = \{q_{ij}\}$  nechť má tvar

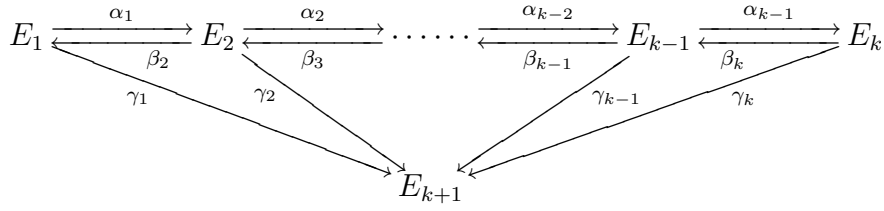
$$\mathbf{Q} = \begin{pmatrix} -q_1 & \alpha_1 & 0 & 0 & 0 & 0 & 0 \\ \beta_2 & -q_2 & \alpha_2 & 0 & 0 & 0 & 0 \\ 0 & \beta_3 & -q_3 & \alpha_3 & 0 & 0 & 0 \\ 0 & 0 & \beta_4 & -q_4 & \alpha_4 & 0 & \gamma_4 \\ 0 & 0 & 0 & \beta_5 & -q_5 & \alpha_5 & \gamma_5 \\ 0 & 0 & 0 & 0 & \beta_6 & -q_6 & \gamma_6 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

kde  $\alpha_6 = \gamma_6, \alpha_i > 0, i = 1, \dots, 6$  a  $q_i = \sum_{\substack{j=1 \\ j \neq i}}^7 q_{ij}$ .

Opět nás zajímá rozdělení inkubační doby AIDS. Je-li jedinec infikován virem HIV v čase 0, lze inkubační dobu AIDS reprezentovat veličinou  $T_{17}$ . Při určování jejího rozdělení již nelze postupovat jako v předchozím příkladu. Tomuto problému je věnována následující teorie.  $\diamond$

Model z příkladu 3.2 zobecníme. Budeme uvažovat homogenní Markovův proces  $\{X(t), t \geq 0\}$  se spojitým časem a konečnou množinou stavů  $S = \{E_1, E_2, \dots, E_k, E_{k+1}\}, k \in \mathbb{N}$ . Stavy  $E_1, \dots, E_k$  jsou přechodné, stav  $E_1$  reprezentuje počátek infekce a stavy  $E_2, \dots, E_k$  značí nějaká infekční stádia. Stav  $E_{k+1}$  je absorpční stav reprezentující klinické stádium AIDS. V modelu budeme předpokládat přechody  $E_i \rightarrow E_{i+1}, i = 1, \dots, k$  s intenzitami  $\alpha_i > 0$  a obecně připustíme i přechody  $E_i \rightarrow E_{i-1}, i = 2, \dots, k$  s intenzitami  $\beta_i \geq 0$  a přechody  $E_i \rightarrow E_{k+1}, i = 1, \dots, k$  s intenzitami  $\gamma_i$ , kde  $\gamma_k = \alpha_k$ . Model můžeme shrnout schématem na obrázku 3.2.

Předpokládejme, že proces je v čase 0 ve stavu  $E_{i_0}, i_0 = 1, \dots, k$  a označme  $T_{i_0(k+1)}$  náhodný čas, kdy je stav  $E_{i_0}$  absorbován stavem  $E_{k+1}$ . Hustotu veličiny  $T_{i_0(k+1)}$  označíme  $g_{i_0}(t)$  a distribuční funkci  $G_{i_0}(t)$ . Inkubační doba AIDS je tedy reprezentována rozdělením s hustotou  $g_1(t)$ . Rozdělení veličiny  $T_{i_0(k+1)}$  určuje následující lemma.



Obrázek 3.2: Schéma zobecněného modelu

**Lemma 3.2.** *Nechť platí výše zavedené značení. Potom pro každé  $i_0 = 1, \dots, k$  je*

$$g_{i_0}(t) = \sum_{j=0}^{k-1} \sum_{i=1}^k c_{ij}^{i_0} t^j \exp(-\rho_i t), t > 0,$$

kde  $c_{ij}^{i_0}$ ,  $\rho_i$  jsou konstanty.

Než napíšeme důkaz, uvedeme několik pomocných tvrzení. Pro následující text definujeme  $\beta_1 = 0$ .

**Lemma 3.3.** *Uvažujme čtvercovou matici  $\mathbf{A}_k$  řádu  $k$  ve tvaru*

$$\mathbf{A}_k = \begin{pmatrix} \lambda_1 & -\alpha_1 & 0 & \dots & \dots & \dots & 0 \\ -\beta_2 & \lambda_2 & -\alpha_2 & 0 & \dots & \dots & 0 \\ 0 & -\beta_3 & \lambda_3 & -\alpha_3 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & \dots & 0 & -\beta_{k-1} & \lambda_{k-1} & -\alpha_{k-1} \\ 0 & \dots & \dots & \dots & 0 & -\beta_k & \lambda_k \end{pmatrix},$$

$\alpha_i > 0$ ,  $\beta_i \geq 0$ ,  $\lambda_i \geq \alpha_i + \beta_i$  pro  $i = 1, \dots, k$ . Potom  $\det(\mathbf{A}_k) > 0$ .

*Důkaz.* Viz věta 2.2 v [8, str. 16]. □

**Lemma 3.4.** (viz [5, str. 301]) *Nechť jsou dány čtvercové matice  $\mathbf{A}, \mathbf{B}$  řádu  $k$ . Potom platí*

$$\exp(\mathbf{A} + \mathbf{B}) = \exp(\mathbf{A}) \exp(\mathbf{B}), \text{ je-li } \mathbf{AB} = \mathbf{BA}, \quad (3.12)$$

$$\frac{d}{dt} \exp(\mathbf{A}t) = \mathbf{A} \exp(\mathbf{A}t), \quad (3.13)$$

$$\mathbf{V}^{-1} \exp(\mathbf{A}) \mathbf{V} = \exp(\mathbf{V}^{-1} \mathbf{A} \mathbf{V}), \text{ kde } \mathbf{V} \text{ je libovolná regulární matice řádu } k. \quad (3.14)$$

Uvažujme matici  $\mathbf{A} = \mathbf{A}_k$ , kde  $\mathbf{A}_k$  je definována v lemmatu 3.3 a pro kterou platí  $\lambda_i = \alpha_i + \beta_i + \gamma_i$ ,  $i = 1, \dots, k-1$  a  $\lambda_k = \alpha_k + \beta_k$ . Konstanty  $\alpha_i$ ,  $\beta_i$ ,  $\gamma_i$  nechť mají význam intenzit výše definovaného procesu  $\{X(t), t \geq 0\}$  (viz obrázek 3.2). Označme  $\mathbf{0}_k$  nulový sloupcový vektor o  $k$  prvcích a  $\gamma = (\gamma_1, \dots, \gamma_k)^T$ . Potom matice

$$\mathbf{Q} = \begin{pmatrix} -\mathbf{A} & \gamma \\ \mathbf{0}_k^T & 0 \end{pmatrix}$$

je matice intenzit procesu  $\{X(t), t \geq 0\}$ .

**Lemma 3.5.** *Uvažujme matice  $\mathbf{A}, \mathbf{Q}$  definované výše. Je-li  $\mathbf{A}$  invertibilní, potom platí*

$$\mathbf{g}(t) = \exp(-\mathbf{A}t)\gamma, t > 0, \quad (3.15)$$

kde  $\mathbf{g}(t) = (g_1(t), \dots, g_k(t))^T$ .

*Důkaz.* Zřejmě platí

$$\mathbf{Q}^j = (-1)^j \begin{pmatrix} \mathbf{A}^j & -\mathbf{A}^{j-1}\gamma \\ \mathbf{0}_k^T & 0 \end{pmatrix}, j = 1, 2, \dots$$

Dle věty 3.4 je

$$\begin{aligned} \mathbf{P}(t) &= \exp(\mathbf{Q}t) = \sum_{j=0}^{\infty} \frac{1}{j!} t^j \mathbf{Q}^j = \mathbf{I}_{k+1} + \sum_{j=1}^{\infty} \frac{(-1)^j}{j!} t^j \begin{pmatrix} \mathbf{A}^j & -\mathbf{A}^{j-1}\gamma \\ \mathbf{0}_k^T & 0 \end{pmatrix} = \\ &= \begin{pmatrix} \mathbf{I}_k + \sum_{j=1}^{\infty} \frac{(-1)^j}{j!} t^j \mathbf{A}^j & -\mathbf{A}^{-1} \sum_{j=1}^{\infty} \frac{(-1)^j}{j!} t^j \mathbf{A}^j \gamma \\ \mathbf{0}_k^T & 1 \end{pmatrix} = \\ &= \begin{pmatrix} \exp(-\mathbf{A}t) & \mathbf{A}^{-1}[\mathbf{I}_k - \exp(-\mathbf{A}t)]\gamma \\ \mathbf{0}_k^T & 1 \end{pmatrix}, t > 0, \end{aligned} \quad (3.16)$$

kde  $\mathbf{I}_{k+1}$  značí jednotkovou matici řádu  $k+1$ , obdobně pro  $\mathbf{I}_k$ .

Označme  $\mathbf{p}_{k+1}(t) = (p_{1(k+1)}(t), \dots, p_{k(k+1)}(t))^T$ . Pro každé  $i = 1, \dots, k$  platí

$$\begin{aligned} p_{i(k+1)}(t) &= P(X(t) = k+1, T_{i(k+1)} > t \mid X(0) = i) + P(X(t) = k+1, \\ &+ 1, T_{i(k+1)} \leq t \mid X(0) = i) = P(X(t) = k+1, T_{i(k+1)} \leq t \mid X(0) = i) = \\ &= P(T_{i(k+1)} \leq t) = G_i(t). \end{aligned}$$

Je tedy  $\mathbf{p}_{k+1}(t) = (G_1(t), \dots, G_k(t))^T$ ,  $t > 0$ .

S využitím (3.16) a (3.13) můžeme psát

$$\mathbf{g}(t) = \frac{d}{dt} \mathbf{p}_{k+1}(t) = \frac{d}{dt} \{ \mathbf{A}^{-1}[\mathbf{I}_k - \exp(-\mathbf{A}t)]\gamma \} = \exp(-\mathbf{A}t)\gamma, t > 0.$$

□



Ještě připomeneme pojem Jordanovy matice. Problematice Jordanových matic včetně metod jejich nalezení se věnuje např. [1, str. 235-264].

**Definice 3.5.** Nechť jsou dány čtvercové matice  $\mathbf{A}, \mathbf{B}$  stejného řádu definované na témže tělese  $F$ . Řekneme, že matice  $\mathbf{A}, \mathbf{B}$  jsou podobné, jestliže existuje regulární matice  $\mathbf{V}$  nad tělesem  $F$  taková, že  $\mathbf{A} = \mathbf{V}^{-1}\mathbf{B}\mathbf{V}$ .

**Definice 3.6.** Čtvercová matice tvaru

$$\mathbf{J}(a) = \begin{pmatrix} a & 1 & 0 & \dots & \dots & \dots & 0 \\ 0 & a & 1 & 0 & \dots & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & \dots & \dots & 0 & a & 1 \\ 0 & \dots & \dots & \dots & \dots & 0 & a \end{pmatrix}$$

se nazývá *Jordanova buňka* příslušná prvku  $a$ . Diagonální bloková matice  $\mathbf{J}$ , jejíž bloky na diagonále jsou Jordanovy buňky, se nazývá *Jordanova matice*. Jordanovu matici můžeme definovat ekvivalentním způsobem jako matici

$$\begin{pmatrix} a_1 & d_1 & 0 & \dots & \dots & \dots & 0 \\ 0 & a_2 & d_2 & 0 & \dots & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & \dots & \dots & 0 & a_{k-1} & d_{k-1} \\ 0 & \dots & \dots & \dots & \dots & 0 & a_k \end{pmatrix}, \quad (3.17)$$

kde pro každé  $i = 1, \dots, k-1$  je  $d_i = 0$  nebo  $d_i = 1$  a, je-li  $d_i = 1$ , potom  $a_i = a_{i+1}$ .

**Věta 3.6.** Nechť  $\mathbf{A}$  je čtvercová matice řádu  $k$  s vlastními čísly  $\rho_1, \dots, \rho_k$ . Potom matice  $\mathbf{A}$  je podobná Jordanově matici.

*Důkaz.* Viz věta 17.8 v [2, str. 116-118]. □

Vlastní čísla  $\rho_1, \dots, \rho_k$  matice  $\mathbf{A}$  nemusí být nutně různá. Nechť  $\rho_{i_l}$ ,  $l = 1, \dots, n$  jsou všechna navzájem různá vlastní čísla, jejich násobnosti označíme  $k_{i_l}$ . Potom má Jordanova matice Jordanovy buňky  $\mathbf{J}(\rho_{i_l})$  řádu  $k_{i_l}$ , které budeme dále značit  $\mathbf{J}_l$ ,  $l = 1, \dots, n$ . Jordanova matice není ve větě 3.6 dána jednoznačně. Platí totiž, že dvě Jordanovy matice lišící se pouze pořadím Jordanových buněk jsou podobné (viz Důsledky 18.12 (i) v [1, str. 244]).

*Důkaz.* (Lemma 3.2) Nechť matice  $\mathbf{A}$  je definována jako v lemmatu 3.5. Dle lemmatu 3.3 je matice  $\mathbf{A}$  regulární a dle lemmatu 3.5 tedy platí (3.15).

Vlastní čísla matice  $\mathbf{A}$  označme  $\rho_1, \dots, \rho_k$ . Dle věty 3.6 existuje Jordanova matice  $\mathbf{J}$  a regulární matice  $\mathbf{V}$  tak, že  $\mathbf{A} = \mathbf{V}^{-1}\mathbf{J}\mathbf{V}$ . Nechť platí značení zavedené před tímto důkazem. Uvažujme matici  $\mathbf{J}$ , která se skládá z Jordanových buněk  $\mathbf{J}_1, \dots, \mathbf{J}_n$  (v tomto pořadí). Pro každé  $l = 1, \dots, n$  označme  $\mathbf{M}_l$  diagonální matici řádu  $k_{i_l}$  s diagonálním prvkem  $\rho_{i_l}$  a definujme matici  $\mathbf{N}_l = \mathbf{J}_l - \mathbf{M}_l$ . Odpovídající diagonální blokovou matici s diagonálními bloky  $\mathbf{M}_1, \dots, \mathbf{M}_n$  (v tomto pořadí) označíme  $\mathbf{M}$ , analogicky z bloků  $\mathbf{N}_1, \dots, \mathbf{N}_n$  zkonstruujeme matici  $\mathbf{N}$ . Potom platí  $\mathbf{J} = \mathbf{M} + \mathbf{N}$ . Jelikož pro každé  $l$  je  $\mathbf{M}_l\mathbf{N}_l = \mathbf{N}_l\mathbf{M}_l$ , je zřejmé také  $\mathbf{M}\mathbf{N} = \mathbf{N}\mathbf{M}$ . Označíme-li  $\overline{\mathbf{M}} = \mathbf{V}^{-1}\mathbf{M}\mathbf{V}$  a  $\overline{\mathbf{N}} = \mathbf{V}^{-1}\mathbf{N}\mathbf{V}$ , platí

$$\mathbf{A} = \mathbf{V}^{-1}\mathbf{J}\mathbf{V} = \mathbf{V}^{-1}(\mathbf{M} + \mathbf{N})\mathbf{V} = \overline{\mathbf{M}} + \overline{\mathbf{N}}, \quad (3.18)$$

$$\overline{\mathbf{M}\mathbf{N}} = \mathbf{V}^{-1}\mathbf{M}\mathbf{N}\mathbf{V} = \mathbf{V}^{-1}\mathbf{N}\mathbf{M}\mathbf{V} = \overline{\mathbf{N}\mathbf{M}}. \quad (3.19)$$

Označíme  $\mathbf{V} = \{v_{ij}\}$  a  $\mathbf{V}^{-1} = \{v_{ij}^{(-1)}\}$ . Pro  $s = 1, 2, \dots$  platí  $\mathbf{M}^s = \{m_{ij}^{(s)}\}$ , kde  $m_{ii}^{(s)} = \rho_i^s$  a  $m_{ij}^{(s)} = 0$  jinak, a tedy platí  $\exp(-\mathbf{M}t) = \{\mu_{ij}\}$ , kde  $\mu_{ii} = \exp(-\rho_i t)$  a  $\mu_{ij} = 0$  jinak. Pro  $\mathbf{N}^s = \{n_{ij}^{(s)}\}$  je

$$n_{ij}^{(s)} = \begin{cases} \prod_{m=i}^{i+s-1} d_m & \text{je-li } j = i + s \text{ a zároveň } j \leq k, \\ 0 & \text{jinak,} \end{cases}$$

kde  $d_m$  jsou prvky matice  $\mathbf{J}$  ve smyslu (3.17). Platí  $\mathbf{N}^s = \mathbf{0}$ ,  $s \geq k$  a pro  $\exp(-\mathbf{N}t) = \{\nu_{ij}\}$  je

$$\nu_{ij} = \begin{cases} \frac{(-1)^p}{p!} t^p n_{ij}^{(p)} & \text{je-li } j > i \text{ a zároveň } j = i + p, \\ 1 & \text{je-li } j = i, \\ 0 & \text{je-li } j < i. \end{cases}$$

Označme pro každé  $i_0 = 1, \dots, k$  jednotkový sloupcový vektor s číslem 1 na  $i_0$ -tém místě jako  $\mathbf{e}_{i_0}$ . Nyní využijeme (3.15) a, jelikož platí (3.18) a (3.19), také (3.12) a (3.14). Pro  $t > 0$  platí

$$\begin{aligned} g_{i_0}(t) &= \mathbf{e}_{i_0}^T \mathbf{g}(t) = \mathbf{e}_{i_0}^T \exp(-\mathbf{A}t) \boldsymbol{\gamma} = \mathbf{e}_{i_0}^T \mathbf{V}^{-1} \exp(-\mathbf{M}t) \exp(-\mathbf{N}t) \mathbf{V} (\gamma_1, \dots, \gamma_k)^T \\ &= \left( v_{i_0 1}^{(-1)} \exp(-\rho_1 t), \dots, v_{i_0 k}^{(-1)} \exp(-\rho_k t) \right) \exp(-\mathbf{N}t) \left( \sum_{i=1}^k v_{1i} \gamma_i, \dots, \sum_{i=1}^k v_{ki} \gamma_i \right)^T \\ &= \left( \exp(-\rho_1 t) v_{i_0 1}^{(-1)}, \sum_{r=1}^2 \frac{(-1)^{2-r}}{(2-r)!} t^{2-r} \exp(-\rho_r t) v_{i_0 r}^{(-1)} n_{r 2}^{(2-r)}, \dots, \sum_{r=1}^k \frac{(-1)^{k-r}}{(k-r)!} t^{k-r} \right. \\ &\quad \cdot \left. \exp(-\rho_r t) v_{i_0 r}^{(-1)} n_{rk}^{(k-r)} \right) \left( \sum_{i=1}^k v_{1i} \gamma_i, \dots, \sum_{i=1}^k v_{ki} \gamma_i \right)^T = \sum_{j=0}^{k-1} \sum_{i=1}^k c_{ij}^{i_0} t^j \exp(-\rho_i t). \end{aligned}$$

Pro konstanty  $c_{ij}^{i_0}$ , kde  $i, j, i_0 = 1, \dots, k$  platí

$$c_{ij}^{i_0} = \begin{cases} \frac{(-1)^j}{j!} v_{i_0 i}^{(-1)} \left( \prod_{m=i}^{i+j-1} d_m \right) \left( \sum_{p=1}^k \gamma_p v_{(i+j)p} \right) & \text{je-li } i + j \leq k, \\ 0 & \text{jinak,} \end{cases}$$

kde  $\prod_{m=i}^j d_m = 1$  pro  $j < i$ .

□

Důkaz nám dává návod, jak spočítat příslušné konstanty z lemmatu 3.2. Nicméně, uvažujeme-li model s obecně kladnými intenzitami  $\alpha_i$ ,  $\beta_i$  a  $\gamma_i$ , je i pro malou množinu stavů  $S$  takovýto výpočet numericky náročný.

### 3.3 Vhodnost Markovových modelů

Pro použití Markovových modelů při modelování průběhu choroby HIV hovoří několik skutečností. Při modelování je přirozené definovat jednotlivá stádia choroby, ta ztotožnit se stavy nějakého stavového modelu a problém redukovat na modelování přechodů mezi stavy tohoto modelu. Dále je nutno si uvědomit, že data, se kterými pracujeme, jsou cenzorována, a to jak intervalově, tak i zleva a zprava. Víme totiž pouze, v jakých stádiích se pacienti nacházeli ve vybraných časových okamžicích, ale neznáme přesné časy přechodů mezi stádii choroby (interval censoring). U pacientů obvykle neznáme ani přesný čas infekce (left censoring). Dále musíme počítat s tím, že vzhledem k dlouhé inkubační době se u některých pacientů během pozorování nerozvine klinické stádium AIDS (right censoring). Z těchto důvodů nelze při analýze dat použít běžné statistické metody. Přirozený charakter Markovových procesů však umožňuje dobře odhadovat neznámé parametry i z dat, která jsou silně zatížena cenzorováním.

Nicméně, jak je známo z teorie, náhodné doby mezi přechody mají v Markovových modelech exponenciální rozdělení, tedy rozdělení bez paměti. Tento fakt je zřejmě při konstrukci modelů omezující. Problém lze řešit použitím semi-Markovových procesů, jejichž doby mezi přechody jsou také nezávislé, ale mohou mít libovolné spojité rozdělení. Teorii semi-Markovových procesů se však v této práci věnovat nebudeme. Přehledný úvod do této problematiky lze nalézt např. v [7, str. 86–95].

# Literatura

- [1] Bečvář J.: *Lineární algebra*, MATFYZPRESS, Praha, 2002.
- [2] Bican L.: *Lineární algebra a geometrie*, Academia, Praha, 2002.
- [3] Chiang C. L.: *Introduction to Stochastic Processes in Biostatistics*, John Wiley & Sons, Inc., New York, 1968.
- [4] Chung K. L.: *Markov Chains with Stationary Transition Probabilities*, Springer-Verlag, New York, 1967.
- [5] Curtis C. W.: *Linear Algebra: An Introductory Approach*, Springer-Verlag, New York, 1993.
- [6] Longini I. M. a kol.: *Statistical Analysis of the Stages of HIV Infection Using a Markov Model*, *Statistics in Medicine* **8** (1989) 831–843.
- [7] Mode C. J. , Sleeman C. K.: *Stochastic Processes in Epidemiology*, World Scientific, Singapore, 2000.
- [8] Neammanee K. a kol.: *On the AIDS Incubation Distribution*, *Statistics & Probability Letters* **73** (2005) 13–23.
- [9] Prášková Z., Lachout P.: *Základy náhodných procesů*, Nakladatelství Karolinum, Praha, 2005.
- [10] HIV infekce na internetové adrese  
[http://www.lf3.cuni.cz/studium/materialy/infekce/HIV\\_infekce.doc](http://www.lf3.cuni.cz/studium/materialy/infekce/HIV_infekce.doc)