

# TEORIE HROMADNÉ OBSLUHY

Gejza Dohnal

září 1997

Gejza Dohnal, fakulta strojní ČVUT Praha.

Doplňkový text k přednášce pro 4. ročník dopravní fakulty ČVUT Praha

## Předmluva

*Vážení přátelé,*

*setkáváme se na přednášce o tématu, které může být pro řadu z Vás zajímavé, předpokládám že i užitečné, nicméně je velmi obtížné. Ani ne tak z hlediska použitých matematických nástrojů, jako spíše z hlediska filosofie problému, Vaší představivosti a potřebné intuice. Teorie hromadné obsluhy je de facto především aplikací teorie pravděpodobnosti, speciálně náhodných procesů, optimalizačních metod a metod matematického modelování (simulace, Monte Carlo metody). Její praktické aplikace vyžadují především použití matematické statistiky.*

*Tato přednáška by Vás měla seznámit s problematikou a základními pojmy z Teorie hromadné obsluhy. Bude jakýmsi úvodem do problému s důrazem na aplikace v dopravě. Pro ty z Vás, kteří by se chtěli s touto teorií seznámit hlouběji, uvedu na konci tohoto textu seznam literatury, kde lze nalézt více a hlubších informací, než které dostanete v rámci tohoto kurzu.*

### 1. Co se předpokládá?

*Předpokládám znalosti integrálního a diferenciálního počtu v rozsahu Matematiky I - III. Především derivace a integrace elementárních funkcí, limitu funkce a posloupnosti, práce s posloupnostmi a s nekonečnými řadami, řešení obyčejných diferenciálních rovnic a jejich lineárních soustav. Dále předpokládám znalosti základních pojmů z teorie pravděpodobnosti, jako je pojem náhodné veličiny a jejích charakteristik, pravděpodobnostního rozdělení, hustoty a distribuční funkce. Některé tyto pojmy a metody si znovu zopakujeme v rámci cvičení.*

### 2. Co se nabízí?

*Přednáška a cvičení v celkovém rozsahu 3 hodiny týdně po dobu 15 týdnů. Celkem tedy 45 hodin, z nichž bych chtěl věnovat cca 26 hodin přednáškám k danému tématu (tj. 13 dvouhodinových přednášek), 13 hodin cvičení, 3 hodiny si ponechám pro opakování a doplnění partií z matematiky a pravděpodobnosti (sem patří zejména soustavy lineárních diferenčních rovnic,*

*pravděpodobnostní rozložení součtu náhodných veličin, rozdělení Erlangovo, algoritmy pro generování náhodných veličin na počítači) a 3 hodiny jako rezervu pro každý případ.*

*Obsahy přednášek budou následující:*

1. Úvod, základní pojmy, příklady.
2. Bodový proces, definice, pravděpodobnostní charakteristiky
3. Základní typy procesů, proces obnovy.
4. Markovské procesy.
5. Markovské modely, Kendallova klasifikace, model M/M/1
6. Modely M/M/n
7. Nemarkovské modely, model M/G/n.
8. Modely G/G/n. *neuskutečnila se*
9. Modely se spojitým tokem. *neuskutečnila se*
10. Obslužné sítě, úvod.
11. Obslužné sítě, příklady.
12. Petriho síť. *neuskutečnila se*
13. Počítačové simulace. *rozšířená o generátory pseudonáhodných čísel a metody generování náhodných veličin.*

*Protože nejsou k dispozici žádná skripta a literatura uvedená na konci tohoto textu je mnohem širší, než bude obsah uvedených přednášek, pokusím se průběžně připravit učební texty k jednotlivým přednáškám. (Tento text je už jejich součástí.) Budou ve formě L<sup>A</sup>T<sub>E</sub>Xových dokumentů a předpokládám, že se mi podaří je umístit někde na Síti tak, aby byly dostupné všem. Tyto texty by měly být doplňkem přednášek a Vaší pomůckou při přípravě na zkoušku.*

### **3. Co se požaduje?**

*Především pokud možno aktivní účast. Pasivní sledování přednášky (o nesledování ani nemluvě) považuji za ztrátu Vašeho i mého času. Aktivní účastí rozumím jak řešení zadávaných úloh a vyhledávání příkladů ve svém okolí, tak i vzájemnou komunikaci v průběhu přednášek a cvičení. Touto cestou*

*můžete ovlivnit do jisté míry i obsah přednášek tak, aby pro Vás byl co nejužitečnější. Tato přednáška se koná ve školním roce 1997/98 poprvé a tudíž její obsah se bude teprve dotvářet. Moje představa může být poněkud jiná, než Vaše potřeby a možnosti. Proto uvítám veškeré připomínky z Vaší strany a pokusím se na ně reagovat.*

*Podle rozdělení témat přednášek budou rozděleny i požadavky ke zkoušce. Součástí hodnocení u zkoušky (cca 30%) bude i hodnocení práce v průběhu semestru. Každý student zpracuje jednu úlohu ve formě samostatné práce (referátu), která bude též součástí hodnocení. Úlohy si vyberete sami a uvítám, když tyto práce budou v přímé souvislosti s ostatními Vašemi pracemi na této fakultě (projekty, diplomové práce apod.).*

*V Praze, září 1997*



# Obsah

<b>1</b>	<b>Úvod</b>	<b>1</b>
1.1	Model a skutečnost. . . . .	1
1.2	Základní pojmy. . . . .	2
<b>2</b>	<b>Bodový proces</b>	<b>5</b>
2.1	Definice . . . . .	5
2.2	Předpoklad A: . . . . .	6
2.3	Předpoklad B: . . . . .	7
2.4	Předpoklad C: . . . . .	8
<b>3</b>	<b>Základní typy procesů, proces obnovy</b>	<b>9</b>
3.1	Základní třídy procesů . . . . .	9
3.2	Homogenní proces . . . . .	10
3.3	Ordinární procesy . . . . .	11
3.4	Procesy obnovy . . . . .	12
3.5	Poissonův proces . . . . .	13
3.5.1	Nehomogenní Poissonovy procesy . . . . .	15
<b>4</b>	<b>Markovské procesy</b>	<b>17</b>
4.1	Markovské řetězce . . . . .	17
4.2	Markovské procesy . . . . .	20
4.3	Nemarkovské procesy . . . . .	24
<b>5</b>	<b>Kendallova klasifikace, model M/M/1</b>	<b>27</b>
5.1	Systémy M/M/1 . . . . .	28
5.2	Počet zákazníků v systému M/M/1 . . . . .	31
5.3	Rozložení doby čekání v systému M/M/1 s $\lambda < \mu$ . . . . .	31

5.4	Rozdělení periody nepřetržitého provozu v $M/M/1$ . . . . .	33
5.5	Frontové režimy . . . . .	33
<b>6</b>	<b>Modely <math>M/M/n</math></b>	<b>35</b>
6.1	Podmínka stability systému $M/M/n$ . . . . .	35
6.2	Charakteristiky systému $M/M/n$ . . . . .	38
6.3	Systém $M/M/\infty$ . . . . .	40
6.4	Systém $M/M/n/r$ (systém s omezenou frontou) . . . . .	40
<b>7</b>	<b>Nemarkovské modely, model <math>M/G/n</math></b>	<b>43</b>
7.1	Systémy s obecně rozloženou dobou obsluhy . . . . .	45
7.2	Příklad 7.3 . . . . .	46
7.3	Systém s minimální dobou obsluhy . . . . .	46
<b>8</b>	<b>Modely <math>G/G/n</math></b>	<b>51</b>
<b>9</b>	<b>Modely se spojitým tokem</b>	<b>53</b>
<b>10</b>	<b>Obslužné sítě, úvod</b>	<b>55</b>
10.1	Příklad systému se dvěma obslužnými stanicemi . . . . .	56
10.2	Obslužný systém jako markovský proces . . . . .	57
10.3	Rovnováha systému hromadné obsluhy . . . . .	59
10.4	Model otevřené obslužné sítě . . . . .	61
<b>11</b>	<b>Obslužné sítě – obecnější rozdělení požadavku na obsluhu</b>	<b>71</b>
11.1	Limitní přechod k obecnému rozdělení požadavku na obsluhu	74
11.2	Síť stanic s kladnou minimální dobou obsluhy . . . . .	78
<b>12</b>	<b>Petriho sítě</b>	<b>81</b>
<b>13</b>	<b>Počítačové simulace</b>	<b>83</b>
13.1	Metody simulace . . . . .	84
13.2	Simulace modelu obslužné sítě . . . . .	85
<b>14</b>	<b>Dodatky</b>	<b>97</b>
14.1	Některá důležitá pravděpodobnostní rozložení . . . . .	97
14.1.1	Rozložení počtu pokusů do prvního úspěchu . . . . .	97
14.1.2	Rozložení doby čekání . . . . .	98
14.1.3	Gamma rozložení . . . . .	99



14.2	Rozložení součtů nezávislých náhodných veličin . . . . .	99
14.2.1	Diskrétní případ . . . . .	99
14.2.2	Spojité případ . . . . .	100
<b>15</b>	<b>Doplňky – simulace náhodných veličin.</b>	<b>101</b>
15.1	Pseudonáhodné posloupnosti . . . . .	101
15.2	Algoritmické generátory pseudonáhodných čísel . . . . .	103
15.2.1	Kongruenční generátory . . . . .	104
15.2.2	Generátory založené na posunu v registrech . . . . .	105
15.3	Testování pseudonáhodných posloupností . . . . .	105
15.3.1	Testy shody s rovnoměrným rozložením pravděpodobnosti . . . . .	105
15.3.2	Testy nezávislosti . . . . .	105
15.3.3	Test délky cyklu . . . . .	105
15.4	Generování náhodných veličin . . . . .	105
15.4.1	Metoda inverze . . . . .	106
15.4.2	Metoda zamítací . . . . .	106
15.4.3	Některá diskrétní rozložení . . . . .	106
15.4.4	Některá spojitá rozložení . . . . .	106
15.4.5	Náhodný výběr, míchání . . . . .	106
15.4.6	Metody Monte Carlo . . . . .	106
15.5	Simulace náhodných procesů . . . . .	106



# Kapitola 1

## Úvod

### 1.1 Model a skutečnost.

Model je zjednodušený obraz skutečnosti. Přejchod od reality k modelu bývá obvykle velmi obtížný. Vždy je třeba stanovit míru podobnosti, tj. míru abstrakce, s níž model tvoříme, tak, aby v rámci této míry model přesně odpovídal realitě, resp. aby formulace problému v termínech modelu odpovídala formulaci v termínech reality. Tato míra nemůže být nikdy stoprocentní shodou, neboť potom bychom museli dospět k poznání, že „nejlepším modelem kočky je kočka“. Proto se často používá k modelování reálného světa teorie pravděpodobnosti a k určování parametrů modelu výsledky statistického pozorování modelovaného reálného objektu.

Prvotní model bývá obvykle verbální. Ten v sobě nese největší míru zobecnění. Na tomto modelu (formulaci problému) je třeba v další fázi vybudovat model pracovní – fyzikální, matematický, chemický, logický, simulační či jiný. Tyto modely se liší většinou pouze vyjadřovacími a zobrazovacími prostředky. Častou chybou při modelování je to, že v případě selhání pracovního modelu se snažíme o úpravu verbálního modelu. Tato cesta je však chybná! Spíše se musíme snažit najít jiný model, případně snížit nároky na jeho přesnost, použít simulaci apod.

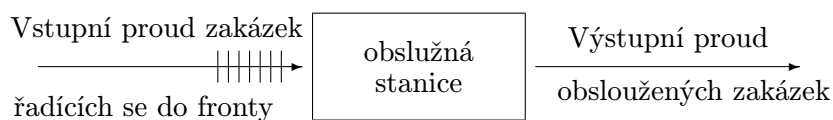
Teorie hromadné obsluhy vychází především z matematiky a z teorie pravděpodobnosti. Patří do oblasti matematického modelování a optimalizace. Jedná se zde o modelování systému *zákazník – obsluha*. V reálném světě se s tímto systémem se setkáváme téměř na každém kroku a to nejen v obchodě či u lékaře. Nejzávažnější problémy, které byly touto metodikou zkoumány, se týkaly především komunikace (přenosu a zpracování informací), dopravy a dopravních uzlů, výrobních systémů.

Počátky teorie lze umístit někam do 30. let tohoto století a jsou spojeny především s matematickým modelováním telefonních ústředen, kde „zákazníky“ jsou přicházející hovory účastníků telefonní sítě a „obsluhu“ tvořily spojovatelky. Hovory přicházejí na ústřednu v náhodných, na sobě nezávislých časových okamžicích a buď jsou ihned spojeny (obslouženy) nebo se řadí do front, ze kterých jsou spojovány postupně podle nějakého předem stanoveného pravidla (tzv. „frontového režimu“). Přitom však může dojít i k poruše v obslužném systému (spojovatelka – ústředna – volaný účastník). V uvedeném příkladu se předpokládá, že jak délka obsluhy, tak i okamžiky poruch a doba potřebná k jejich odstranění mohou být náhodné veličiny. Toto schema lze uplatnit i v řadě jiných situací.

**Cvičení 1:** Najděte tři různé příklady z oblasti dopravy a přepravy osob, na které lze použít výše popsané schéma zákazník – obsluha.

## 1.2 Základní pojmy.

Základní jednotkou systémů hromadné obsluhy je *obslužný systém*, [*obslužná systém, queueing system*]. Ten je složen z trojice zákazník – linka – obsluha.



- *Zákazník* [*zakazčik, customer*] – subjekt, který požaduje vyřízení (zpracování) svého požadavku – *zakázky* [*trebovanije, order*]. Příkladem zákazníka může být telefonní účastník, počítačový program, součástka, ale i letadlo, cestující nebo náklad. Zakázkou potom může být například požadavek na spojení, požadavek na přidělení výpočetní kapa-

city, montáž nebo obrábění, požadavek na přistání nebo požadavek na přepravu.

- *Obslužná linka* – místo a osoba nebo zařízení, kde se zakázky zpracovávají. V našich příkladech to mohou být telefonní ústředna, CPU počítače, dělník u montážní linky nebo obráběcí stroj, přistávací dráha, přepravní jednotka.
- *Obsluha* [*obsluživanije, service*] – zpracování zakázky. Činnost vedoucí k uspokojení požadavku zákazníka. Například spojení s volaným účastníkem, zpracování programu, vlastní montáž nebo obrábění, přistání nebo přeprava.

Vlastnosti obslužného systému:

- Zákazníci přicházejí se svými požadavky do systému opakovaně v čase (tedy ne nutně titíž) - to je předpoklad tzv. *hromadnosti* obsluhy. Okamžiky příchodů zákazníků do systému mohou být deterministické nebo náhodné, závislé či nezávislé, se známým rozdělením pravděpodobnosti, nebo s rozdělením, které neznáme. Navíc zakázky mohou být několika typů, lišících se jak v rozložení doby příchodů, tak i v dalších parametrech.
- Systém obsahuje jednu či několik paralelních obslužných linek. V jedné obslužné lince může být v daný okamžik obsluhováno vždy pouze určité množství zákazníků (podle kapacity linky). Ostatní přichodí se řadí spolu s dříve přišedšími (a dosud neobsloženými) zákazníky do tzv. *fronty* [*očered, queue*]. Způsob řazení do fronty a výběr zákazníků k obsluze – tzv. *frontový režim* [*disciplina, discipline*] může být opět různý (viz dále). Různé systémy se mohou lišit i v maximální možné délce fronty (kapacitě „odložených hovorů“, zásobníku či čekárny).
- Zpracování zakázky představuje čerpání určité (omezené) kapacity obslužné linky - *množství obsluhy*. Množství obsluhy může být měřeno v čase, množství spotřebované energie, finančních nákladech apod. Navíc toto množství může být – podobně jako časy příchodů zákazníků – náhodné, se známým či neznámým rozdělením pravděpodobnosti, závislé na typu zakázky, . . .

V případě, že počet zákazníků přicházejících do systému je větší, než je systém momentálně schopen obsloužit, hromadí se zákazníci v tzv. „frontě“, která může mít řadu různých konkrétních podob. Pro nás bude důležitý způsob, jak je fronta organizována, to znamená, zda existuje pravidlo pro zařazování do fronty, případně pro výběr zákazníků z fronty k obsluze. Totuto pravidlu říkáme *frontový režim*. V našich modelech budeme hovořit o několika typech frontových režimů:

- FIFO (= first in, first out)
- LIFO (= last in, first out)
- s preferencemi
- náhodný (zákazníci se řadí do fronty, případně jsou z fronty vybírání podle pravděpodobnostního rozdělení na pořadí)
- s omezenou frontou (s kapacitou  $K$ )
- s rezignací (zákazník, pro něhož je fronta příliš dlouhá, se vůbec nezařadí, případně po určitém čase odchází)

Obslužné linky se obvykle spojují do *obslužných sítí*, kde je třeba navíc uvažovat způsob, jakým zákazník touto sítí prochází, než je zcela obsloužen. Takovou obslužnou sítí může být například počítačový systém (CPU, disky, periferie), dílna ve výrobním závodě (obáběcí centra, montážní linky), systém křižovatek, přepravní systém (kombinovaná přeprava několika dopravními prostředky, přes různé uzly).

V některých případech se na systém hromadné obsluhy díváme jako na systém nejrůznějších toků : tok času, energie, peněz, případně jiného média. Pokud budeme nadále hovořit například o *čase*, lze si pod tímto termínem představit například „pouze“ provozní hodiny, nebo počet impulzů, případně množství energie atp.

## Kapitola 2

# Bodový proces

### 2.1 Definice

Představme si posloupnost nějakých událostí, které nastávají náhodně v čase. Příkladem mohou být příjezdy vozidel k výběrčímu mýtného na dálnici nebo příchody cestujících do stanice metra. Uvažujme posloupnost náhodných veličin  $\{t_n\}_{-\infty}^{\infty}$ ,  $t_n : \Omega \rightarrow R$ , takových, že  $-\infty \leq \dots \leq t_{-n} \leq \dots \leq t_{-1} \leq t_0 \leq t_1 \leq \dots \leq t_n \leq \dots \leq \infty$ . Náhodná veličina  $t_n$  reprezentuje *dobu n-té události*, přičemž  $t_n(\omega) = x_n$  je *čas, kdy n-tá událost nastala*.

Označme

$$N(t, \omega) = \sum_{-\infty}^{\infty} \chi_{(t_n \leq t)}(\omega).$$

Často budeme používat pouze zkrácené označení  $N(t)$ .  $N(t)$  budeme nazývat *bodovým procesem*. (Symbol  $\chi_A$  označuje indikátorovou funkci jevu  $A$ , to jest takovou funkci, která nabývá hodnoty 1 pokud jev  $A$  nastává, jinak je rovna 0.)

Všimněme si, že platí  $N(t) = n \iff t_n \leq t < t_{n+1}$ .

Dále označíme  $\tau_n = t_{n+1} - t_n$  posloupnost náhodných veličin, majících význam *doby mezi dvěma sousedními událostmi*.

Rozdíl  $N(s+t) - N(s) = N(s, t)$  potom označuje *počet událostí, které nastaly v časovém intervalu  $(s, s+t]$ ,  $s \in R, t \in R^+$* . Pro tento proces platí následující vztah:  $N(s, t) = n \iff \exists k \in N : s < t_k \leq \dots \leq t_{k+n-1} \leq s+t < t_{k+n}$ .

*Poznámka:* Následující výrazy představují ekvivalentní formy zápisu bodového procesu:

$$\{t_n\}, \{\tau_n\}, N(t), N(s, t).$$

## 2.2 Předpoklad A:

Nechť  $-\infty = \dots = t_{-n} = \dots = t_{-1} < t_0 < t_1 < \dots < t_n < \dots < \infty$ .

Za tohoto předpokladu

- a)  $t_0$  je čas první pozorované události (počátek pozorování procesu),
- b) v jeden časový okamžik nemohou nastat současně dvě události.

Za platnosti předpokladu A můžeme zavést následující označení:

$$N(t) = N(0, t) \text{ a platí zřejmě } N(t, 0) = 0.$$

### 2.3. Tvzení:

$$N(s, t) = N(s, n) + N(s + n, t - n)$$

pro  $0 \leq n \leq t$ .

Nadále budeme předpokládat Předpoklad A. Označme

$$v_n(t) = P(N(t) = n), \quad n = 0, 1, 2, \dots$$

Protože pro každé  $t \in R$  je  $v_n(t)$  pravděpodobnostním rozdělením, musí platit vztah  $\sum_{n=0}^{\infty} v_n(t) = 1$  a  $E(N(t)) = \sum_{n=0}^{\infty} n v_n(t)$ .

Dále označme

$$v_n(s, t) = P(N(s, t) = n), \quad n = 0, 1, 2, \dots$$

Podobně jako v předchozím případě, i zde musí pro každou dvojici  $s, t \in R^+$  platit vztah  $\sum_{n=0}^{\infty} v_n(s, t) = 1$  a  $E(N(s, t)) = \sum_{n=0}^{\infty} n v_n(s, t)$ .

*Cvičení:* Jakou hodnotu má  $E(N(s, 0))$ ?

**2.4. Tvzení:** Pro libovolné  $0 \leq u \leq t$  a  $s \in R^+$  platí vztah

$$E(N(s, t)) = E(N(s, u)) + E(N(s + u, t - u))$$

a tedy i

$$E(N(t)) = E(N(u)) + E(N(u, t - u)).$$



## 2.3 Předpoklad B:

Pro doby mezi událostmi  $\tau_n$  označme

- distribuční funkci  $A_k(x) = P(\tau_k \leq x)$ ,  $k = 0, 1, 2, \dots$ ,
- rizikovou funkci  $B_k(x) = 1 - A_k(x) = P(\tau_k > x)$ ,  $k = 0, 1, 2, \dots$ .

$\tau_n$  jsou spojité náhodné veličiny, to znamená, že existují hustoty  $a_n(x)$  tak, že

$$A_n(x) = \int_0^x a_n(u) du,$$

neboli

$$a_n(x) = A_n'(x) = -B_n'(x).$$

Z předpokladu A plyne  $A_n(0) = 0$ , neboli mezi  $n$ -tou a  $n + 1$ -ní událostí je s pravděpodobností 1 doba větší než nula (tedy nenastanou zároveň).

*Cvičení:* Co lze vyčíst ze zápisu  $A_5(0) = \frac{1}{2}$ ?

## 2.6. Tvrzení:

$$E(\tau_k) = \int_0^\infty u a_k(u) du = \int_0^\infty B_k(u) du, \quad k = 0, 1, \dots$$

*Důkaz:* Nabývá-li  $x$  pouze nezáporných hodnot a integrál  $\int_0^\infty x f(x) dx$  existuje, pak užitím metody per partes  $u = x, v' = -f(x), u' = 1, v = 1 - F(x)$  dostaneme

$$-\int_0^\infty x f(x) dx = [x(1 - F(x))]_0^\infty - \int_0^\infty (1 - F(x)) dx.$$

Nyní si stačí uvědomit, že z předpokladu existence  $\int_0^\infty x f(x) dx$  plyne, že  $\lim_{x \rightarrow \infty} x(1 - F(x)) = 0$ .

Pro okamžiky událostí  $\{t_n\}$  definujme

- distribuční funkci  $U_n(t) = P(t_n \leq t) = P(N(t) \geq n) = \sum_{j=n}^\infty P(N(t) = j) = \sum_{j=n}^\infty v_j(t)$ ,  $n = 0, 1, \dots$ ,
- rizikovou funkci  $V_n(t) = 1 - U_n(t) = P(t_n > t) = \sum_{j=0}^{n-1} v_j(t)$ ,  $n = 0, 1, \dots$ .

**2.7. Tvrzení:**

$$E(t_n) = \sum_{j=0}^{n-1} E(\tau_j), \quad n = 0, 1, \dots$$

*Důkaz:* Tvrzení plyne z rovnosti  $t_n = \sum_{j=0}^{n-1} \tau_j$ .

**2.8. Tvrzení:**

$$E(\tau_k) = \int_0^\infty B_k(t) dt = \int_0^\infty v_k(t) dt, \quad k = 0, 1, \dots$$

*Důkaz:*

$$E(t_n) = \int_0^\infty t dU_n(t) = \int_0^\infty (1 - U_n(t)) dt = \int_0^\infty \sum_{j=0}^{n-1} v_j(t) dt = \sum_{j=0}^{n-1} \int_0^\infty v_j(t) dt = \sum_{j=0}^{n-1} E(\tau_j).$$

**2.4 Předpoklad C:**

$\tau_n, n = 1, 2, \dots$  jsou nezávislé, stejně rozdělené spojité náhodné veličiny, to znamená, že  $A_n(x) = F(x), n = 1, 2, \dots$  s hustotami  $f(x)$ .

Nechť pro  $\tau_0$  je  $A_0(x) = A(x)$ . Potom

$$P(t_1 \leq t) = U_1(t) = P(\tau_0 \leq t) = A_0(t)$$

$$P(t_2 \leq t) = P(\tau_0 + \tau_1 \leq t) = \int_0^t A_0(t-u) f(u) du = U_2(t) = \int_0^t F(t-u) a(u) du = A * F(t)$$

Potom

$$U_n(t) = P(t_n \leq t) = A * F^{(n-1)}(t), \quad n = 1, 2, \dots,$$

přičemž definujeme  $A * F^{(0)}(t) = A(t)$ . (Symbol  $*$  označuje konvoluci.)

## Kapitola 3

# Základní typy procesů, proces obnovy

### 3.1 Základní třídy procesů

*Posloupností nezávislých náhodných veličin* nazýváme posloupnost náhodných veličin  $\{X_n\}_{n=-\infty}^{\infty}$  takovou, že pro libovolnou  $k$ -tici indexů  $(i_1, \dots, i_k)$ ,  $k = 1, 2, \dots$  platí

$$P(X_{i_1} \leq x_1, X_{i_2} \leq x_2, \dots, X_{i_k} \leq x_k) = \prod_{j=1}^k P(X_{i_j} \leq x_j).$$

Bodový proces  $N(t)$  nazýváme *procesem s nezávislými přírůstky* právě když pro libovolnou  $k$ -tici disjunktních intervalů  $(s_1, s_1 + t_1], (s_2, s_2 + t_2], \dots, (s_k, s_k + t_k]$ ,  $k = 2, \dots$  tvoří  $(N(s_1, s_1 + t_1), N(s_2, s_2 + t_2), \dots, N(s_k, s_k + t_k))$  posloupnost nezávislých náhodných veličin.

Bodový proces  $N(t)$  nazýváme *regenerativním* právě když posloupnost dob mezi událostmi  $\{\tau_n\}$  je posloupností nezávislých náhodných veličin.

Bodový proces  $N(t)$  nazýváme *rekurentním* právě když posloupnost dob mezi událostmi  $\{\tau_n\}$  je posloupností nezávislých a stejně rozdělených náhodných veličin, tzn. je-li  $A_k(t) = A(t)$ ,  $k = 1, 2, \dots$

*Příklad:* Mezi procesem s nezávislými přírůstky a regenerativním procesem je obecně rozdíl. Tak například průjezdy motorových vozidel určitým místem na dálnici mohou tvořit bodový proces s nezávislými přírůstky, ale

tento proces už nemusí být (a zpravidla není) regenerativní. Naproti tomu, volání na telefonní ústřednu lze považovat za regenerativní (a dokonce i rekurentní) bodový proces.

V případě rekurentního procesu označme  $A_0(t) = P(\tau_0 \leq t)$ . Tato distribuční funkce se od ostatních  $A_k(t) = A(t)$ ,  $k = 1, 2, \dots$  může lišit, nicméně nemůže být libovolná:

**Tvrzení:**

$$A_0(t) = \frac{1}{T} \int_0^t B(u) du,$$

kde  $T = \int_0^\infty B(u) du$ .

*Důkaz: za cvičení.*

### 3.2 Homogenní proces

Bodový proces  $N(t)$  nazýváme *homogenním*, jestliže pro každé  $s \in R$  má  $N(s, t)$  stejné rozdělení pravděpodobnosti jako  $N(0, t)$  a tedy jako  $N(t)$ . To znamená, že musí platit  $v_n(s, t) = v_n(0, t) = v_n(t)$  pro libovolné  $s, t \in R^+$ ,  $n \in N$ .

**Tvrzení:** Je-li  $N(t)$  homogenní bodový proces, potom platí  $E(N(t)) = tE(N(1)) = t\mu$ , kde  $\mu = E(N(1))$  je průměrný počet událostí za jednotku času (hustota procesu).

*Důkaz: za cvičení.*

**Poznámka:** Předpoklad homogenity se obvykle vztahuje na určitý (nedlouhý) časový úsek. Tímto postupem lze pracovat s „po částech homogenními procesy“.

Za předpokladu A zřejmě musí platit, že  $v_0(0) = P(N(0) = 0) = 1$ . Pravděpodobnost  $P(N(t) = 0) = v_0(t)$  je zřejmě neklesající funkcí času  $t$ .

**Tvrzení:** Pro každý homogenní bodový proces existuje

$$\lim_{t \rightarrow 0^+} \frac{1 - v_0(t)}{t} = \lambda \geq 0.$$

$\lambda$  se nazývá *parametrem (intenzitou)* homogenního procesu.

*Důkaz:*  $\lambda$  je vlastně záporně vzatou derivací  $v_0(t)$  zprava v bodě  $t = 0$ . Další plyne z předpokladů A a B.

**Tvrzení:** Pro každý homogenní bodový proces platí  $\lambda \leq \mu$ .

*Důkaz:*  $1 - v_0(t) = \sum_{j=1}^{\infty} v_j(t) \leq \sum_{j=1}^{\infty} j v_j(t) = E(N(t)) = \mu t$ . Odtud po vydělení  $t > 0$  dostaneme

$$\frac{1 - v_0(t)}{t} \leq \mu \Rightarrow \lambda = \lim_{t \rightarrow 0^+} \frac{1 - v_0(t)}{t} \leq \mu.$$

### 3.3 Ordinární procesy

Bodový proces  $N(t)$  nazýváme *ordinárním* právě když pro libovolné  $s \leq 0$  platí

$$\lim_{t \rightarrow 0} \frac{1 - v_0(s, t) - v_1(s, t)}{t} = 0.$$

To znamená, že ve velmi krátkém časovém intervalu je pravděpodobnost že nastane více než 1 událost řádově menší, než je délka tohoto intervalu.

**Tvrzení:** Nechť  $N(t)$  je homogenní bodový proces. Potom platí  $\lambda = \mu \iff$  proces je ordinární.

*Důkaz:* za cvičení.

Z výše uvedeného tvrzení plyne, že  $\lambda = \frac{v_1(t)}{t}$ . Dále pro homogenní ordinární proces platí při  $t \rightarrow 0$ :

$$\begin{aligned} v_0 &= 1 - \lambda t + o(t) \\ v_1 &= \lambda t + o(t) \\ P(N(t) \geq 2) &= U_2(t) = o(t) \end{aligned}$$

*Cvičení:* Dokažte následující tvrzení:

- Ordinární proces s nezávislými přírůstky je regenerativní.
- Ordinární homogenní proces je regenerativní právě když je rekurentní.

**Poznámka 1.** V praxi pozorujeme i procesy, v nichž nastává více událostí v jediném okamžiku – například průjezdy na víceproudé komunikaci – v takovém případě, nelze-li už zjemnit časovou škálu tak, aby časy událostí byly navzájem různé, lze namísto posloupnosti  $\{t_n\}$  registrovat posloupnost  $\{t_n, r_n\}$ , kde  $t_n$  je čas a  $r_n$  je počet událostí v okamžiku  $t_n$ . Ve většině případů už bude podmínka ordinárnosti procesu splněna.

**Poznámka 2.** Regenerativní procesy se nazývají též *procesy obnovy*. Ve chvíli, kdy nastává událost, proces začíná jakoby „nanovo“. V takových případech se zkoumají vlastnosti procesu pro  $t \rightarrow \infty$  (při dlouhé době provozu) jako „průměrné v čase“ (time averages).

### 3.4 Procesy obnovy

Nechť  $\{\tau_n\}$  je posloupnost nezávislých náhodných veličin. Označme

$$\begin{aligned} P(\tau_j \leq x) &= A_j(x) = F(x), \quad j = 1, 2, \dots, \\ P(\tau_0 \leq x) &= A_0(x). \end{aligned}$$

Proces  $N(t) = \max\{n : \sum_{k=0}^n \tau_k \leq t\}$  budeme nazývat *procesem obnovy* a časy  $t_n = \sum_{k=0}^n \tau_k$  nazveme *okamžiky obnovy*.

Funkce  $n(t) = E(N(t)) = \sum_{j=0}^{\infty} E(\chi_{\{t_j \leq t\}}) = P(t_j \leq t) = \sum_{j=0}^{\infty} A_0 \star F^{(j-1)}(t)$  se nazývá *funkce obnovy*.

Je zřejmé, že  $P(\lim_{t \rightarrow \infty} n(t) = +\infty) = 1$ . Bude nás zajímat průměrná hodnota  $n(t)$  v čase, t.j.  $\lim_{t \rightarrow \infty} \frac{n(t)}{t}$ .

Označme dále

$$\mu = \frac{1}{\int_{-\infty}^{\infty} xF(x)dx},$$

pokud je integrál konečný,  $\mu = 0$  v opačném případě.  $\mu$  se nazývá *intenzitou obnovy*.

**Tvrzení (základní věta teorie obnovy):**

Pro proces obnovy  $N(t)$  platí skoro jistě (t.j. s pravděpodobností 1)

$$\lim_{t \rightarrow \infty} \frac{N(t)}{t} = \mu$$

(verze pro náhodné veličiny) a

$$\lim_{t \rightarrow \infty} \frac{n(t)}{t} = \mu$$

(verze pro střední hodnoty).

*Důkaz:* Zákon velkých čísel  $\Rightarrow Et_n = \sum E\tau_n = n \frac{1}{\mu} \Rightarrow$

$$\begin{aligned} P(\omega : \lim_{n \rightarrow \infty} \frac{t_n(\omega)}{n} = \frac{1}{\mu}) &= 1 \\ P(\omega : \lim_{n \rightarrow \infty} N(t, \omega) = \infty) &= 1 \end{aligned}$$

Odtud plyne, že

$$P(\omega : \lim_{n \rightarrow \infty} \frac{t_{N(t)}(\omega)}{N(t, \omega)} = \frac{1}{\mu}) = 1$$

Protože je  $t_{N(t)} \leq t \leq t_{N(t)+1}$  a  $P(\lim_{t \rightarrow \infty} \frac{t_{N(t)}}{t_{N(t)+1}} = 1) = 1$ , dostáváme skoro jistě nerovnosti

$$\frac{t_{N(t)}}{t_{N(t)+1}} < \frac{t_{N(t)}}{t} \leq 1$$

a

$$\lim_{n \rightarrow \infty} \frac{t_{N(t)}}{t} = 1,$$

ze kterých dostáváme konvergenci  $\frac{N(t)}{t} = \frac{t_{N(t)}}{t} \frac{N(t)}{t_{N(t)}} \rightarrow \mu$  skoro jistě.

### 3.5 Poissonův proces

**Definice:** Homogenní proces s nezávislými přírůstky, v němž platí

$$P(N(s, t) = n) = P(N(t) = n) = v_n(s, t) = v_n(t) = e^{-\lambda t} \frac{(\lambda t)^n}{n!},$$

pro  $n = 1, 2, \dots$ ,  $\lambda > 0$ , se nazývá Poissonův proces.

**Tvrzení:** Homogenní proces s nezávislými přírůstky je ordinární právě když je Poissonovým procesem.

*Důkaz:* Nejprve ukážeme, že Poissonův proces je ordinární: je totiž

$$\lim_{t \rightarrow 0} \frac{1 - v_0(t) - v_1(t)}{t} = \lim_{t \rightarrow 0} \frac{1 - e^{-\lambda t} - \lambda t e^{-\lambda t}}{t} = \lim_{t \rightarrow 0} \frac{1 - e^{-\lambda t}}{t} - \lim_{t \rightarrow 0} \lambda e^{-\lambda t} = 0.$$

Naopak: z homogenity a nezávislosti plyne, že  $v_0(t+h) = v_0(t)v_0(h)$  pro všechna  $t, h \in \mathbb{R}^+$ , a tedy  $v_0(t) = \left(v_0\left(\frac{t}{n}\right)\right)^n$  pro každé  $n \in \mathbb{N}$  a  $t \geq 0$ . Navíc z ordinarity procesu vyplývá, že  $v_0\left(\frac{t}{n}\right) = 1 - \lambda\frac{t}{n} + o\left(\frac{1}{n}\right)$ , pro  $n \rightarrow \infty$ . Spojením obou důsledků dostaneme

$$\lim_{n \rightarrow \infty} v_0(t) = \left(v_0\left(\frac{t}{n}\right)\right)^n = \lim_{n \rightarrow \infty} \left(1 - \frac{\lambda t}{n}\right)^n = e^{-\lambda t}.$$

Tedy pro každé  $t \geq 0$  je  $v_0(t) = e^{-\lambda t}$ . Zbývá odvodit vztah pro  $v_n(t)$ .

Uvažujme interval délky  $t+h$ , v němž nastalo  $k$  událostí. Z nezávislosti plyne, že

$$v_k(t+h) = v_k(t)v_0(h) + v_{k-1}(t)v_1(h) + \dots + v_0(t)v_k(h).$$

Odtud

$$\begin{aligned} v_k(t+h) - v_k(t) &= v_k(t)(v_0(h) - 1) + \dots \\ \lim_{h \rightarrow 0} \frac{v_k(t+h) - v_k(t)}{h} &= v'_k(t) = -v_k(t) + \lambda v_{k-1}(t) \end{aligned}$$

kde poslední rovnost plyne z ordinárnosti. Tedy  $v'_k(t) = \lambda(v_{k-1}(t) - v_k(t))$ .

Položme  $v_k(t) = e^{-\lambda t} u_k(t)$ . Pak musí být  $u_0(t) = 1$  pro každé  $t \geq 0$ , pro  $k \geq 1$  je  $v_k(0) = u_k(0) = 0$ . Potom

$$v'_k(t) = -\lambda e^{-\lambda t} u_k(t) + e^{-\lambda t} u'_k(t) = \lambda e^{-\lambda t} (u_{k-1}(t) - u_k(t)).$$

Odtud postupně dostáváme

$$u_0(t) = 1 \Rightarrow u_1(t) = \lambda t \Rightarrow u_2(t) = \frac{\lambda^2 t^2}{2} \Rightarrow \dots \Rightarrow u_n(t) = \frac{(\lambda t)^n}{n!} \Rightarrow \dots$$

**Tvrzení:**  $N(t)$  je regenerativním procesem s  $\tau_n \sim \text{Exp}(\lambda)$  právě když je Poissonovým procesem.

*Důkaz:* Opět začneme implikací zprava doleva: je-li  $N(t)$  Poissonův proces, potom je ordinární s nezávislými přírůstky a tedy je regenerativní a platí

$$P(\tau_0 \leq t) = 1 - P(\tau_0 > t) = 1 - v_0(t) = 1 - e^{-\lambda t}.$$



Obráceně: nechť  $\tau_n \sim \text{Exp}(\lambda)$ . Potom je  $t_k$  součtem nezávislých náhodných veličin s exponenciálním rozdělením a má tedy Erlangovo rozdělení s parametry  $\lambda, k$ . Odtud

$$P(t_k > t) = e^{-\lambda t} \sum_{j=0}^{k-1} \frac{(\lambda t)^j}{j!} = P(N(t) < k).$$

Srovnáním plyne, že  $N(t)$  je Poissonův proces.

**Tvrzení:** Nechť  $N(t)$  je Poissonův proces a nechť  $t_n = t$ . Potom podmíněné rozdělení  $t_1, t_2, \dots, t_{n-1}$  je rovnoměrné v intervalu  $(0, t)$ , tzn. že

$$P(t_k < s | t_n = t) = \frac{s}{t}, \quad k < n, s < t.$$

### 3.5.1 Nehomogenní Poissonovy procesy

Nehomogenním Poissonovým procesem nazveme proces s nezávislými přírůstky, v němž platí

$$P(N(s, t) = n) = v_n(s, t) = e^{-(\Lambda(s+t) - \Lambda(s))} \frac{(\Lambda(s+t) - \Lambda(s))^n}{n!},$$

$n = 1, 2, \dots$ , a  $\Lambda(s)$  je nezáporná neklesající funkce pro  $s \geq 0$ .

*Poznámky:* Pokud je  $\Lambda(s) = \lambda s$ , pak je proces homogenní. Je-li  $\Lambda(s)$  absolutně spojitá, t.j. existuje nezáporná funkce  $\lambda(u)$  tak, že  $\Lambda(s) = \int_0^s \lambda(u) du$ , nazýváme tuto funkci *proměnnou intenzitou* procesu, závisující na čase. Pro homogenní proces je  $\lambda(s) = \lambda$  a tedy  $\Lambda(s) = \lambda s$ .

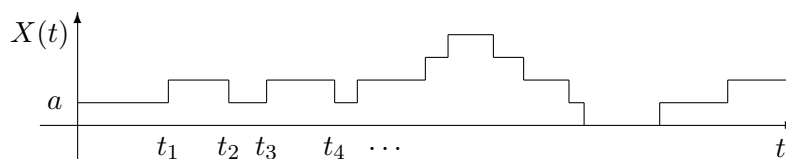


## Kapitola 4

# Markovské procesy

### 4.1 Markovské řetězce

Pozorujeme-li počet vozidel, která jsou v daném okamžiku v určitém úseku silnice (například na celnici). Změnu tohoto počtu způsobují události dvojího typu: příjezdy nových vozidel a odjezdy vozidel, která úsek projela (byla odbavena). Označme počet automobilů v úseku  $X(t)$ . Použijeme-li označení pro časy událostí  $t_n, n = 1, 2, \dots$ , potom nechť  $X_n = X(t_n)$ .



**Definice 4.1.1:** Platí-li pro libovolné  $n \in \{1, 2, \dots\}$

$$\begin{aligned} P(X_n = j | X_{n-1} = i, X_{n-2} = i_{n-2}, \dots, X_0 = i_0) &= P(X_n = j | X_{n-1} = i) \\ &= p_{ij}(n) \end{aligned}$$

a  $P(X_0 = k) = a_k(0)$ , řekneme, že posloupnost náhodných veličin  $\{X_n\}$  splňuje *Markovskou vlastnost*. Posloupnost  $\{X_n\}$  budeme nazývat (nehomogenním) *Markovským řetězcem*.

Nezávisí-li  $p_{ij}(n)$  na čase  $n$ , tedy je-li  $p_{ij}(n) = p_{ij}$ ,  $n = 1, 2, \dots$ , budeme Markovský řetězec  $\{X_n\}$  nazývat *homogenním*. V tomto případě budeme používat i označení  $a_k(0) = a_k$ .

Označme  $\Omega = \{E_1, E_2, \dots\}$  množinu (spočetnou) stavů, které mohou ve sledovaném modelu nastat. Potom  $X_n = j \iff$  systém je v čase  $t_n$  ve stavu  $E_j$ . Při tomto značení bude pro Markovský řetězec se spočetně mnoha stavy platit

$$P(E_{j_0}, E_{j_1}, \dots, E_{j_n}) = a_{j_0} p_{j_0 j_1}(1) \dots p_{j_{n-1} j_n}(n),$$

v nehomogenním případě, pro homogenní řetězec potom

$$P(E_{j_0}, E_{j_1}, \dots, E_{j_n}) = a_{j_0} p_{j_0 j_1} \dots p_{j_{n-1} j_n}.$$

Dosud jsme značili  $p_{ij}(n)$  pravděpodobnost přechodu mezi časy  $t_n$  a  $t_{n+1}$ , tedy mezi  $n$ -tým a  $n + 1$ -ním krokem (v jednom kroku). Označme nyní  $p_{ij}(n, m)$  pravděpodobnost přechodu ze stavu  $i$  do  $j$  mezi  $n$ -tým a  $m$ -tým krokem (přirozeně pro  $n < m$ ), t.j. pravděpodobnost  $P(X(m) = j | X(n) = i)$ . Uvažujme tzv. „matici přechodu“

$$P = \begin{pmatrix} p_{00} & p_{01} & \dots & p_{0n} \\ p_{10} & p_{11} & \dots & p_{1n} \\ \dots & \dots & \dots & \dots \\ p_{n0} & p_{n1} & \dots & p_{nn} \end{pmatrix}.$$

Zřejmě je (v homogenním případě)

$$p_{ij}(n-2, n) = P(X(n) = j | X(n-2) = i) = \sum_{k=0}^n p_{ik} p_{kj} = p_{ij}^{(2)},$$

což je prvek matice  $P^2 = P.P$ . Rozšířením této úvahy dojdeme k závěru, že pro  $n < m$  je  $p_{ij}(m, n) = p_{ij}^{(m-n)}$  prvek matice  $P^{m-n}$ . Toto platí i pro řetězec se spočetně (i nekonečně) mnoha stavy. Vlastní výpočet prvků  $n$ -té mocniny matice  $P$  je ovšem poměrně problematický. V případě nekonečně mnoha stavů musíme vycházet z definice  $p_{ij}^{(n)}$  jako podmíněné pravděpodobnosti; pro řetězec s konečně mnoha stavy můžeme buď postupně umocňovat matici  $P$ , nebo použít aproximaci jako je například tzv. „Peronův vzorec“ (viz Dodatek).

**Definice 4.1.2:**

- Stav  $E_j$  je *přechodný*  $\iff \sum_{n=1}^{\infty} p_{jj}^{(n)} < \infty$   
V tomto případě je  $\sum_{n=1}^{\infty} p_{ij}^{(n)} < \infty$  pro každé  $i$
- Stav  $E_j$  je *trvalý, nulový*  $\iff \sum_{n=1}^{\infty} p_{jj}^{(n)} = +\infty, \lim_{n \rightarrow \infty} p_{jj}^{(n)} = 0$   
Opět je  $\lim_{n \rightarrow \infty} p_{ij}^{(n)} = 0$  pro každé  $i$
- Stav  $E_j$  je *trvalý, nenulový, neperiodický*  $\iff$  existuje  $\mu_j \neq 0$ , tak, že  $\sum_{n=1}^{\infty} p_{jj}^{(n)} = +\infty, \lim_{n \rightarrow \infty} p_{jj}^{(n)} = 1/\mu_j$
- Stav  $E_k$  je *dosažitelný z  $E_j$*   $\iff$  existuje  $n > 0$ , tak, že  $p_{jk}^{(n)} > 0$
- Množina stavů  $C = \{E_{j_1}, \dots, E_{j_n}\}$  je *uzavřená*  $\iff p_{jk} = 0$  pro každé  $E_j \in C$  a  $E_k \notin C$ .
- Stav  $E_j$  se nazývá *absorpčním*  $\iff p_{jj} = 1$ .
- Markovův řetězec se nazývá *nerozložitelným*  $\iff$  nejmenší nerozložitelná množina je množina všech stavů.

Je-li Markovův řetězec nerozložitelný, potom je každý stav dosažitelný z jakéhokoli jiného stavu. Naopak, je-li rozložitelný, potom lze matici přechodu vyjádřit ve tvaru

$$P = \begin{pmatrix} P_1 & O \\ A & B \end{pmatrix},$$

kde  $P_1, B$  jsou čtvercová matice a  $O$  je nulová matice.

**Tvrzení 4.1.3:** Jsou-li  $E_i$  a  $E_j$  vzájemně dosažitelné stavy, potom jsou i téhož typu.

**Tvrzení 4.1.4 (Chapman–Kolmogorovova rovnost):**

$$p_{ij}^{(n)} = \sum_{k=1}^{\infty} p_{ik}^{(m)} \cdot p_{kj}^{(n-m)}, \quad 0 < m < n.$$

Markovův řetězec je plně určen maticí přechodu  $P = \{p_{ij}\}$  a počátečním rozložením  $\{a_j(0)\}$ . Označme  $a_j(n) = P(X(n) = j)$ . Podle Chapman–Kolmogorovy rovnosti je

$$a_j(1) = \sum_{k=1}^{\infty} a_k(0) \cdot p_{kj}, \quad \dots, \quad a_j(n) = \sum_{k=1}^{\infty} a_k(0) \cdot p_{kj}^{(n)}, \quad \dots,$$

nebo též

$$a_j(n) = \sum_{k=1}^{\infty} a_k(m) \cdot p_{kj}^{(n-m)}, \quad 0 < m < n.$$

**Definice 4.1.5:** Limity  $\lim_{n \rightarrow \infty} a_j(n) = \pi_j$ , pokud existují, nazýváme *stacionárním rozložením* Markovova řetězce.

Stacionární rozložení je řešením soustavy rovnic  $\pi_j = \sum_{k=1}^{\infty} \pi_k p_{kj}$ , neboli  $\vec{\pi} = \vec{\pi}P$ . V teorii Markovových řetězců se zkoumají podmínky, za kterých toto stacionární rozložení existuje. Lze například ukázat, že pokud je množina všech stavů konečná, potom k existenci stacionárního rozložení stačí, aby

1. Markovův řetězec byl nerozložitelný
2. pro alespoň jeden stav  $E_j$  platilo  $p_{jj} > 0$ .

Jinou takovou podmínkou je, že musí existovat  $n$  tak, že pro všechny dvojice  $i, j$  platí  $p_{ij}^{(n)} > 0$ .

## 4.2 Markovské procesy

Uvažujme proces  $\{X(t), t \geq 0\}$  se spočetnou množinou stavů  $\{E_0, \dots, E_n, \dots\}$ . Pravděpodobnost, že v čase  $t$  bude proces ve stavu  $E_k$  označíme  $p_k(t) = P(X(t) = E_k)$ .  $p_k(t)$  tvoří pravděpodobnostní rozdělení, tedy musí platit  $\sum_{k=1}^{\infty} p_k(t) = 1$  pro každé  $t \geq 0$ . *Markovská vlastnost* má v případě spojitého času tvar: pro každou posloupnost časů  $0 \leq s_n \leq s_{n-1} \leq \dots \leq s_1 \leq s < t$  platí

$$P(X(t) = j | X(s) = i, X(s_1) = i_1, \dots, X(s_n) = i_n) = P(X(t) = j | X(s) = i) = p_{ij}(s, t),$$

kde  $p_{ij}(s, t)$  je *pravděpodobnost přechodu*. Chapman–Kolmogorovova rovnost zde má tvar

$$p_{ij}(s, t) = \sum_{k=1}^{\infty} p_{ik}(s, l) \cdot p_{kj}(l, t), \quad 0 \leq s < l < t,$$

nebo

$$p_j(t) = \sum_{k=1}^{\infty} p_k(s) \cdot p_{kj}(s, t), \quad 0 \leq s < t.$$

Markovův proces nazveme *homogenním*, pokud  $p_{ij}(s, t) = p_{ij}(t - s)$ . Pro homogenní Markovovy procesy platí Chapman–Kolmogorovova rovnost v následujícím tvaru:

$$p_{ij}(s + t) = \sum_{k=1}^{\infty} p_{ik}(s) \cdot p_{kj}(t),$$

nebo

$$p_j(s + t) = \sum_{k=1}^{\infty} p_k(s) \cdot p_{kj}(t).$$

**Tvrzení 4.2.1:** Doba setrvání homogenního Markovova procesu  $X(t)$  ve stavu  $E_j$  má exponenciální rozdělení s parametrem  $q_j$ .

**Definice 4.2.2:** Parametr  $q_j$  z předchozího tvrzení nazýváme *intenzitou výstupu* procesu  $X(t)$  ze stavu  $E_j$ .

*Důkaz:* Označme  $\pi_j(t) = P(X(s) = j, s \in (l, l + t], l \geq 0 | X(l) = j)$ . Pro libovolné  $t_1, t_2 \geq 0$  je zřejmě  $\pi_j(t_1 + t_2) = \pi_j(t_1) \cdot \pi_j(t_2)$  a po úpravě  $\pi_j(t + h) - \pi_j(t) = \pi_j(t)(\pi_j(h) - 1)$ . Odtud můžeme odvodit diferenciální rovnici

$$\pi_j'(t) = \lim_{h \rightarrow 0} \frac{\pi_j(t + h) - \pi_j(t)}{h} = -\pi_j(t) \lim_{h \rightarrow 0} \frac{1 - \pi_j(h)}{h} = -\pi_j(t) \cdot q_j,$$

pokud existuje  $q_j = \lim_{h \rightarrow 0} \frac{1 - \pi_j(h)}{h} = -\pi_j'(0)$ . Tato limita bude kladná, neboť  $\pi_j(t)$  je klesající funkcí času a tedy její derivace je záporná. Řešením výše uvedené rovnice je  $\pi_j(t) = \exp(-q_j t)$  pro  $t \geq 0$ .

Všimněme si, že je

$$1 \geq \lim_{h \rightarrow 0} p_{jj}(h) \geq \lim_{h \rightarrow 0} \pi_j(h) = \lim_{h \rightarrow 0} \exp(-q_j h) = 1$$

Tedy  $\lim_{h \rightarrow 0} p_{jj}(h) = 1 \Rightarrow \lim_{h \rightarrow 0} p_{jk}(h) = 0$ , neboť  $\sum_{k=1}^{\infty} p_{jk}(h) = 1$ .

**Tvrzení 4.2.3:** (Homogenní) Poissonův proces je (homogenním) Markovovým procesem.

*Důkaz:*

$$P(N(t) = n | N(s) = m, N(s_1) = m_1, \dots, N(s_k) = m_k) = P(N(s, t-s) = n-m) = \exp -\lambda(t-s) \frac{[\lambda(t-s)]^{n-m}}{(n-m)!} = p_{mn}(s, t), \quad n > m$$

Odtud plyne že  $p_{jj}(t) = \pi_j(t) = \exp -\lambda t$ , neboli v Poissonově procesu jsou všechny intenzity stejné a rovné  $\lambda$ . Protože je  $N(t)$  neklesající, je  $p_{kl}(t) = 0$  pro  $k > l$ .

Zřejmě je  $\lim_{h \rightarrow 0} p_{jj}(h) = 1$ ,  $\lim_{h \rightarrow 0} p_{ij}(h) = 0$ , pro každou dvojici  $i, j > 0$ . Pro každé  $j \neq i, i > 0, j > 0$  předpokládejme dále existenci limit ( $\star$ ):

$$q_j(t) = \lim_{h \rightarrow 0} \frac{1 - p_{jj}(t+h)}{h}$$

$$q_{ij}(t) = \lim_{h \rightarrow 0} \frac{p_{ij}(t, t+h)}{h}.$$

O  $q_j$  a  $q_{ij}$  navíc předpokládáme spojitost v proměnné  $t \geq 0$  stejnoměrnost výše uvedené konvergence. V dalším budeme používat tzv. „Kroneckerovo  $\delta$ “:

$$\delta_{ij} = \begin{cases} 0, & \text{pro } i \neq j, \\ 1, & \text{pro } i = j. \end{cases}$$

**Tvrzení 4.2.4:** Za výše uvedených předpokladů ( $\star$ ) splňují pravděpodobnosti přechodu  $p_{ij}(t, \tau)$  soustavám tzv. „Kolmogorovových diferenciálních rovnic“

$$\frac{\partial p_{ij}(t, \tau)}{\partial \tau} = -q_j(\tau) p_{ij}(t, \tau) + \sum_{k \neq j} q_{kj}(\tau) p_{ik}(t, \tau)$$



s počáteční podmínkou  $p_{ij}(t, t) = \delta_{ij}, t > \tau$ ,

$$\frac{\partial p_{ij}(t, \tau)}{\partial \tau} = q_i(t)p_{ij}(t, \tau) - \sum_{k \neq i} q_{ik}(t)p_{kj}(t, \tau)$$

s počáteční podmínkou  $p_{ij}(\tau, \tau) = \delta_{ij}, t < \tau$ .

*Poznámka:* První soustava se nazývá „prospektivní“, druhá je „retrospektivní“ soustava Kolmogorovových diferenciálních rovnic.

*Důkaz* plyne přímo z Chapman–Kolmogorovovy nerovnosti. Pro  $t < \tau < \tau + h$  dostáváme

$$\begin{aligned} p_{ij}(t, \tau + h) &= \sum_k p_{ik}(t, \tau)p_{kj}(\tau, \tau + h) \\ &= p_{ij}(t, \tau)p_{jj}(\tau, \tau + h) + \sum_{k \neq j} p_{ik}(t, \tau)p_{kj}(\tau, \tau + h). \end{aligned}$$

Z předpokladů plyne, že  $p_{ij}(\tau, \tau + h) = 1 - q_j(\tau)h + o(h)$  a tedy

$$p_{ij}(t, \tau + h) - p_{ij}(t, \tau) = -q_j(\tau)hp_{ij}(t, \tau) + o(h) + \sum_{k \neq j} p_{ik}(t, \tau)p_{kj}(\tau, \tau + h),$$

Po vydělení  $h$  a limitním přechodem k nule dostaneme tvrzení.

**Důsledek 4.2.5** Nechť homogenní markovův proces splňuje předpoklady  $(\star)$ . Potom pro  $p_{ij}(t)$  platí

$$\begin{aligned} p'_{ij}(t) &= -q_j p_{ij}(t) + \sum_{k \neq j} q_{kj} p_{ik}(t), \quad t > 0 \\ p'_{ij}(t) &= -q_i p_{ij}(t) + \sum_{k \neq i} q_{ik} p_{kj}(t), \quad t > 0 \end{aligned}$$

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

*Poznámka:* Obě rovnice jsou ekvivalentní, to znamená, že řešení jedné je zároveň i řešením druhé.

**Důsledek 4.2.6** Nechť homogenní markovův proces splňuje předpoklady  $(\star)$ . Potom pro  $p_{ij}(t)$  platí

$$p'_j(t) = -q_j p_j(t) + \sum_{k \neq j} q_{kj} p_k(t), \quad t > 0$$

**Důsledek 4.2.7** Pro homogenní Poissonův proces platí

$$\begin{aligned} p'_{ij}(t) &= -\lambda p_{ij}(t) + \lambda p_{ij-1}(t), & j > 0 \\ p'_{i0}(t) &= -\lambda p_{i0}(t) \end{aligned}$$

*Důkaz:* Do rovnic z důsledku 4.2.5 dosadíme  $q_j = \lambda = q_{j-1,j}$ ,  $j = 1, 2, \dots$ ,  $q_0 = \lambda$ .

**Cvičení** Najděte řešení soustavz Kolmogorovových rovnic pro homogenní Poissonův proces.

Zpravidla nás zajímá pouze tzv. „stacionární rozložení“, to jest limity  $\lim_{t \rightarrow \infty} p_{ij}(t) = \pi_{ij}$ ,  $\lim_{t \rightarrow \infty} p_j(t) = \pi_j$ .

- Pro každý homogenní markovův proces tyto limity existují, splňují-li podmínku  $\lim_{t \rightarrow 0} p_{ij}(t) = 0$ ,  $\lim_{t \rightarrow 0} p_{jj}(t) = 1$ , obecně však tyto limity závisí na  $i$ .
- Má-li markovův proces konečnou množinu stavů, pak k nezávislosti limit na  $i$  stačí, aby existovalo  $t \geq 0$  tak, že  $p_{ij}(t) > 0$  pro všechna  $i, j = 0, 1, \dots, M$ .

### 4.3 Nemarkovské procesy

Proces, který je Markovův, musí nutně mít časové intervaly mezi událostmi (změnami stavu) s exponenciálním rozložením. To je ovšem omezení, které *nelze často splnit*.

Jak postupovat v takových případech?

1. V některých případech lze použít metodu tzv. „vnořené markovova řetězce“, pocházející od Kendalla. Není-li splněna Markovská vlastnost pro libovolnou volbu  $t > s \geq s_1 \geq \dots \geq s_n \geq 0$ ,  $k, j, j_1, \dots, j_n$ , snažíme se najít určitou podposloupnost  $0 \leq t_1^* < t_2^* < \dots < t_m^* < \dots$ ,  $\lim_{n \rightarrow \infty} t_n^* = +\infty$  takovou, aby pro čísla z této podposloupnosti Markovská vlastnost už platila. To znamená, že posloupnost  $\{X(t_n^*)\}_{n=1}^\infty$  bude tvořit Markovův řetězec. Tím však ztrácíme informaci o chování procesu mimo tuto posloupnost  $\{t_n^*\}_{n=1}^\infty$ .

*Příklad:* Do obchodu přicházejí zákazníci v Poissonovském proudu (t.j. s exponenciální dobou mezi příchody) a jsou obslouženi v době, která má rozdělení s distribuční funkcí  $F(t)$ . Označme  $X(t)$  počet zákazníků v obchodě. Tento proces není obecně Markovův. Označme dále  $t_n^*$  časy odchodů zákazníků z obchodu. Posloupnost  $\{X(t_n^*)\}_1^\infty$  již tvoří Markovův řetězec.

2. Jindy lze situaci napravit *vhodným rozšířením zápisu procesu*.

*Příklad:* Uvažujme rekurentní proces  $N(t)$ . Kdyby byl markovský, pak ze znalosti hodnoty  $N(t^*) = k$  pro jedno dané  $t^*$  by bylo možno stanovit pravděpodobnosti budoucích hodnot  $N(t + \tau)$ . To však obecně neplatí. Pravděpodobnosti změny  $N(t)$  v intervalu  $(t, t + \tau)$  závisí nejen na  $N(t)$ , ale i na tom, jak dlouho už  $N(t) = k$  platí. Namísto  $N(t)$  budeme registrovat dvojice  $[N(t), T(t)]$ , kde  $T(t)$  je doba od poslední události do času  $t$ . Tyto dvojice už tvoří Markovův proces. Pokuste se za cvičení spočítat (s pomocí  $B(t)$ ) podmíněnou pravděpodobnost

$$P(N(t + \tau) = k, T(t + \tau) = w | N(t) = j, T(t) = v)$$

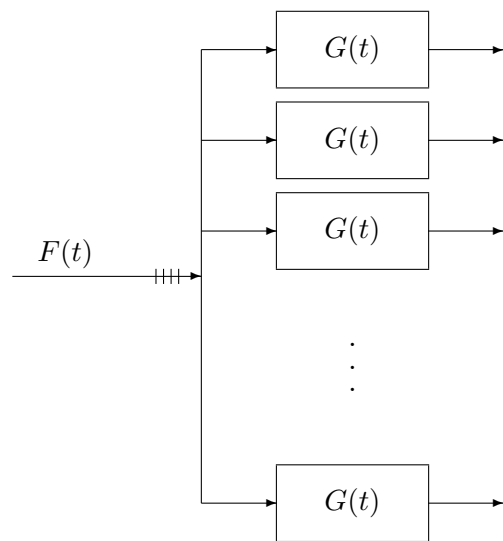
3. Studují se tzv. „semimarkovské procesy“, o nichž až někdy jindy.



## Kapitola 5

# Kendalova klasifikace, model M/M/1

Systém hromadné obsluhy lze zjednodušeně znázornit pomocí následujícího obrázku. Zákazníci přicházejí do systému v obecně náhodných okamžicích a řadí se nějakým způsobem do fronty. Intervaly mezi příchody nechť mají rozdělení pravděpodobnosti s distribuční funkcí  $F(t)$ .



Systém se skládá z  $n$  obslužných stanic. Každá z nich může najednou obsluhovat jediného zákazníka, přičemž doba obsluhy se opět obecně považuje za náhodnou veličinu s distribuční funkcí  $G(t)$ . Zákazníci čekající ve frontě se nějakým způsobem (podle frontového režimu) zařazují do obsluhy; je-li některá stanice obsluhy volná, ihned se do ní zařadí některý zákazník čekající ve frontě. Je-li obsluha ukončena, obslužený zákazník opouští stanici a tím i celý systém obsluhy. V teorii hromadné obsluhy se

používá tzv. „Kendalova klasifikace“ systémů hromadné obsluhy. V rámci

této klasifikace se systémy hromadné obsluhy dělí podle rozdělení vstupního proudu zákazníků, podle rozdělení doby obsluhy a podle počtu obslužných stanic v systému. Někdy se bere v úvahu i kapacita fronty (maximální počet zákazníků, který může čekat ve frontě). Používá se následující označení

$$X/Y/k \quad \text{nebo} \quad X/Y/k/n$$

kde

- $X$  = označení typu stochastického procesu popisujícího příchody zákazníků k obsluze
- $Y$  = označení rozdělení pravděpodobnosti délky doby obsluhy
- $k$  = počet obslužných stanic
- $n$  = kapacita fronty (neomezená fronta se označuje symbolem  $\infty$ )

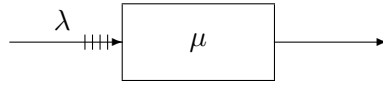
V tabulce je přehled nejčastěji používaných symbolů:

$X, Y$	proces příchodů	rozdělení doby obsluhy
$M$	Poissonův proces, t.j. exponenciální rozdělení vzájemně nezávislých intervalů mezi příchody zákazníků	exponenciální rozdělení doby obsluhy
$E_n$	Erlangovo rozdělení intervalů mezi nezávislými příchody zákazníků	Erlangovo rozdělení doby obsluhy
$K_n$	$\chi^2$ rozdělení intervalů mezi nezávislými příchody zákazníků	$\chi^2$ rozdělení doby obsluhy
$D$	pravidelné deterministické příchody zákazníků	konstantní doba obsluhy
$G$	obecný případ – žádné speciální předpoklady o příchodech zákazníků	obecné rozdělení doby obsluhy – jakékoli
$GI$	rekurentní proces	

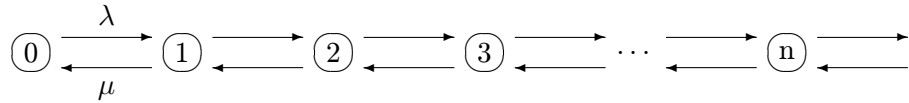
## 5.1 Systémy M/M/1

Zákazníci přicházející do systému se řadí do jediné fronty. Jejich příchody tvoří Poissonovský *ustupní proud*, to jest příchody jsou nezávislé, časové

intervaly mezi nimi jsou náhodné a mají exponenciální rozdělení pravděpodobnosti s parametrem  $\lambda$ . Ve stanici obsluhy je vždy obsluhován jeden zákazník vybraný z fronty. Doba jeho obsluhy je opět náhodná, exponenciálně rozdělená s parametrem  $\mu$ . Po ukončení obsluhy zákazník systém opouští.



Označme  $X(t)$  = počet zákazníků v systému v čase  $t$ . Potom například jev  $X(t) = 0$  znamená, že systém je v čase  $t$  prázdný a čeká na zákazníka; je-li systém ve stavu  $X(t) > 0$ , je právě jeden zákazník obsluhován a ostatních  $X(t) - 1$  zákazníků čeká ve frontě. Proces  $X(t)$  má typický průběh znázorněný grafem na začátku odstavce 4.1. Jiný způsob grafického znázornění procesu  $X(t)$  je následující:



**Tvrzení 5.1.1:**  $X(t)$  je homogenní Markovův proces.

*Důkaz:* Necht'  $X(s) = j$  pro nějaké  $s \geq 0$ . Potom

$$P(\text{v čase } (s, s+h] \text{ přijde 1 zákazník}) = 1 - e^{-\lambda h} @ >> h \rightarrow 0 > \lambda h + o(h)$$

$$P(\text{v čase } (s, s+h] \text{ nepřijde 1 zákazník}) = e^{-\lambda h} @ >> h \rightarrow 0 > 1 - \lambda h + o(h)$$

$$P(\text{v čase } (s, s+h] \text{ odejde 1 zákazník}) = 1 - e^{-\mu h} @ >> h \rightarrow 0 > \mu h + o(h)$$

$$P(\text{v čase } (s, s+h] \text{ přijdou 2 zákazníci}) = (1 - e^{-\lambda h})(1 - e^{-\lambda h}) @ >> h \rightarrow 0 > (\lambda h + o(h))^2 = o(h)$$

$$P(\text{v čase } (s, s+h] \text{ se počet zákazníků nezmění}) = P(X(s+h) = j | X(s) = j) =$$

$$= P(\text{v čase } (s, s+h] \text{ žádný nepřijde a žádný neodejde}) =$$

$$= (1 - \lambda h + o(h))(1 - \mu h + o(h)) = 1 - \lambda h - \mu h + o(h) = p_{jj}(h)$$

$$P(\text{v čase } (s, s+h] \text{ se počet zákazníků zvýší o 1}) = P(X(s+h) = j+1 | X(s) = j) =$$

$$= P(\text{v čase } (s, s+h] \text{ jeden přijde a žádný neodejde}) =$$

$$= (\lambda h + o(h))(1 - \mu h + o(h)) = \lambda h + o(h) = p_{j,j+1}(h)$$

Podobně  $p_{j,j-1}(h) = \mu h + o(h)$  a  $p_{ji}(h) = o(h)$  pro  $|j - i| > 1$ . Výjimečné postavení má stav  $j = 0$ :

$$p_{00}(h) = 1 - \lambda h + o(h) \text{ (nemá, kdo by odešel)}$$

Soustava Kolmogorovových diferenciálních rovnic dle Důsledku 4.2.6 zde má tvar:

$$\begin{aligned} p_0'(t) &= -\lambda p_0(t) + \mu p_1(t) \\ p_k'(t) &= -(\lambda + \mu)p_k(t) + \lambda p_{k-1}(t) + \mu p_{k+1}(t) \end{aligned}$$

pro  $k = 1, 2, \dots$ , neboť je  $q_0 = \lambda, q_k = \lambda + \mu, q_{kk-1} = \lambda, q_{kk+1} = \mu, q_{ik} = 0$  jinak. Navíc musí platit vztah  $\sum p_k(t) = 1$  pro každé  $t$ . Řešení je obecně známé a možné, nicméně výsledek je komplikovaný. Jednodušší situace nastane, bude-li nás zajímat pouze limitní chování systému pro  $t \rightarrow \infty$ , tedy pravděpodobnosti  $p_k = \lim_{t \rightarrow \infty} p_k(t), k = 1, 2, \dots$ . Přitom pokud tato limita existuje, platí zřejmě  $\lim_{t \rightarrow \infty} p_k'(t) = 0$  a tedy soustava přejde do tvaru

$$\begin{aligned} 0 &= -\lambda p_0 + \mu p_1 \\ 0 &= -(\lambda + \mu)p_k + \lambda p_{k-1} + \mu p_{k+1} \end{aligned}$$

s podmínkou  $\sum_{k=0}^{\infty} p_k = 1$ . Řešení této soustavy budeme hledat takto: označme  $z_k = -\lambda p_k + \mu p_{k+1}$ . Po dosazení do soustavy dostaneme

$$\begin{aligned} 0 &= z_0 \\ 0 &= -z_{k-1} + z_k, \quad k = 1, 2, \dots \end{aligned}$$

Odtud  $z_k = 0$  pro všechna  $k$ . Tedy

$$\lambda p_k = \mu p_{k+1} \Rightarrow \frac{p_{k+1}}{p_k} = \frac{\lambda}{\mu} \Rightarrow p_k = \left(\frac{\lambda}{\mu}\right)^k p_0$$

Abychom mohli splnit podmínku  $\sum_{k=0}^{\infty} p_k = 1$ , je třeba aby  $\rho = \frac{\lambda}{\mu} < 1$ . Potom

$$\sum_{k=0}^{\infty} p_0 \left(\frac{\lambda}{\mu}\right)^k = \frac{p_0}{1 - \rho} \Rightarrow p_0 = 1 - \rho$$

. Tak dostáváme řešení

$$p_k = \left(1 - \frac{\lambda}{\mu}\right) \left(\frac{\lambda}{\mu}\right)^k, \quad k = 1, 2, \dots$$

což je geometrické rozdělení pravděpodobnosti s parametrem  $\rho = \frac{\lambda}{\mu}$ . Číslo  $\rho$  se nazývá *intenzitou obsluhy*



## 5.2 Počet zákazníků v systému M/M/1

- Průměrný počet zákazníků v systému:

$$\kappa = \sum_{i=1}^{\infty} ip_k = \rho(1-\rho) \sum_{i=1}^{\infty} i\rho^{i-1} = \frac{\rho}{1-\rho}$$

$$\text{Variance počtu zákazníků v systému} = \sum_{i=1}^{\infty} i^2 p_k = \frac{\rho}{(1-\rho)^2}$$

- Průměrný počet zákazníků ve frontě:

$$\gamma = \sum_{i=1}^{\infty} ip_{k+1} = \frac{\rho^2}{1-\rho} = \frac{\lambda^2}{\mu(\mu-\lambda)}$$

$$\text{Variance tohoto počtu je rovna } \frac{\rho^2(1+\rho-\rho^2)}{(1-\rho)^2}$$

*Poznámka:*

1.  $\kappa - \gamma = \rho$  a nikoli 1, jak by se dalo očekávat!
2.  $p_k$  = pravděpodobnost, s jakou náhodně vybraný zákazník, vstupující do systému, najde v něm dalších  $k$  zákazníků. Bude tedy po vstupu do systému  $k$ -tý ve frontě ( $k > 0$ ).

$p_0$  = pravděpodobnost, že náhodně vybraný zákazník nebude čekat ve frontě.

## 5.3 Rozložení doby čekání v systému M/M/1 s

$$\lambda < \mu$$

Označme  $W_k, k = 1, 2, \dots$  dobu po kterou bude muset čekat zákazník, který vstoupil do systému s  $k$  zákazníky jako  $(k+1)$ ní, t.j. jako  $k$ -tý do fronty s pravděpodobností  $p_k$ .

$W_k$  = (součet doby obsluhy  $(k-1)$  čekajících zákazníků + zbytek doby obsluhy právě obsluhovaného zákazníka) =  $O_1 + O_2 + \dots + O_{k-1} + Z_0$ .

První část  $O_1 + O_2 + \dots + O_{k-1}$  je součtem nezávislých exponenciálně rozložených náhodných veličin s parametrem  $\mu$ , má tedy Erlangovo rozložení pravděpodobnosti s parametry  $(k, \mu)$ . Tedy

$$P(W_k > w) = e^{-\mu w} \sum_{j=0}^{k-1} \frac{(\mu w)^j}{j!}, \quad k = 1, 2, \dots$$

Zřejmě je  $P(W_0) = 1$  (zákazník, který přijde do prázdného systému, nečeká).

- *Rozložení doby čekání  $W$ :*

$$\begin{aligned} P(W > w) &= \sum_{k=0}^{\infty} p_k P(W_k > w) = \sum_{k=1}^{\infty} (1 - \rho) \rho^k e^{-\mu w} \sum_{j=0}^{k-1} \frac{(\mu w)^j}{j!} = \\ &= (1 - \rho) \rho e^{-\mu w} \sum_{j=0}^{\infty} \frac{(\lambda w)^j}{j!} \sum_{r=0}^{\infty} \rho^r = \rho e^{-\mu w} e^{\lambda w} = \rho e^{-(\mu - \lambda)w} \end{aligned}$$

a  $P(W = 0) = p_0 = 1 - \rho$  (t.j. skok o velikosti  $(1 - \rho)$  v bodě nula).

- *Střední hodnota a rozptyl doby čekání:*

$$E(W) = \frac{\rho}{\mu - \lambda}, \quad \text{Var}(W) = \frac{\rho(2 - \rho)}{(\mu - \lambda)^2}.$$

- *Celková doba strávená zákazníkem v systému:* Tato doba se skládá z doby strávené čekáním a z doby obsluhy:  $R = W + O$ .

$$P(R > w) = \sum_{k=0}^{\infty} p_k e^{-\mu w} \sum_{j=0}^k \frac{(\mu w)^j}{j!} = \sum_{j=0}^{\infty} \sum_{r=0}^{\infty} (1 - \rho) e^{-\mu w} \rho^r \rho^j \frac{(\mu w)^j}{j!} = e^{-(\mu - \lambda)w},$$

pro  $w > 0$  a  $P(R < 0) = 0$ .

Pro střední hodnotu a rozptyl platí následující vztahy:

$$E(R) = \frac{1}{\mu - \lambda}, \quad \text{Var}(R) = \frac{1}{(\mu - \lambda)^2}.$$

*Poznámka:* Je-li  $\rho = 1$ , t.j.  $\mu = \lambda$ , je systém nestabilní a soustava Kolmogorových rovnic nemá řešení, všechna  $p_k(t)$  konvergují k nule při  $t \rightarrow \infty$ .

## 5.4 Rozdělení periody nepřetržitého provozu v M/M/1

Uvažujme proces  $Y(t)$ , začínající v 1 (t.j.  $p_1(0) = 1$ ) s pohlcujícím stavem nula: pokud se systém do tohoto stavu dostane, už nikdy se z něho nedostane ven, neboli pokud bude existovat  $t^* > 0$  takové, že  $Y(t^*) = 0$ , pak už bude  $Y(t) = 0$  pro všechna  $t \geq t^*$ . Pro takovýto proces lze odvodit následující soustavu Kolmogorovových diferenciálních rovnic:

$$\begin{aligned} p_0'(t) &= \mu p_1(t), \\ p_1'(t) &= -(\lambda + \mu)p_1(t) + \mu p_2(t), \\ p_k'(t) &= \lambda p_{k-1}(t) - (\lambda + \mu)p_k(t) + \mu p_{k+1}(t), \quad k = 2, 3, \dots \end{aligned}$$

s počátečními podmínkami  $p_1(0) = 1$  a  $\sum_{k=0}^{\infty} p_k(t) = 1$ .

V této soustavě je  $p_0(t)$  pravděpodobnost jevu, že se systém v časovém intervalu  $(0, t]$  zastavil. Tuto pravděpodobnost lze zároveň chápat jako distribuční funkci periody nepřetržitého provozu. Řešení uvedené soustavy rovnic je poměrně komplikované. To, co nás většinou pouze zajímá je střední hodnota a rozptyl (variance) periody nepřetržitého provozu  $\Phi$ :

$$E(\Phi) = \frac{1}{\mu - \lambda}, \quad Var(\Phi) = \frac{\mu + \lambda}{(\mu - \lambda)^3}.$$

Mezi časovými úseky, v nichž systém pracuje bez zastavení, leží intervaly, v nichž systém čeká na příchod zákazníka a tedy nepracuje. Tyto intervaly mají exponenciální rozložení s parametrem  $\lambda$ , tedy se střední hodnotou  $\frac{1}{\lambda}$ . Na takovýto proces se můžeme dívat jako na proces obnovy, v němž okamžiky obnovy jsou právě ony okamžiky, v nichž přichází zákazník do prázdného, čekajícího systému. Střední délka cyklu obnovy je potom součtem střední doby nepřetržitého provozu a doby čekání prázdného systému na zákazníka  $\tau = \frac{1}{\mu - \lambda} + \frac{1}{\lambda} = \frac{\mu}{\lambda(\mu - \lambda)} = \frac{1}{\rho(\mu - \lambda)}$ . V dlouhém časovém intervalu délky  $T$  bude potom průměrný počet cyklů obnovy roven  $\frac{T}{\tau} = T\rho(\mu - \lambda)$ . To je také průměrný počet zastavení linky obsluhy při dlouhém běhu.

## 5.5 Frontové režimy

Frontovým režimem rozumíme soubor pravidel pro řazení příchozích zákazníků do fronty a výběr zákazníka do obsluhy. Řada charakteristik systému M/M/1 je na frontovém režimu nezávislá (počet zákazníků v systému, ve

frontě, perioda nepřetržitého provozu). Změní se pouze rozdělení doby čekání  $W$  a tedy i doba  $R$  strávená zákazníkem v systému.

- *FIFO (First In First Out)*  
Tento režim jsme popisovali na předchozích stránkách.
- *LIFO (Last In First Out)*  
Rozdíl oproti režimu FIFO je v tom, že
  - a)  $W_k = W$  – doba čekání nezávisí na počtu zákazníků ve frontě před příchodem zákazníka.
  - b)  $W$  má stejné rozdělení jako rozdělení periody nepřetržitého provozu.

$$P(W = 0) = p_0 = 1 - \rho$$

$$P(W > w) = \rho(1 - F(w)), \quad w \geq 0,$$

kde  $F(w)$  je distribuční funkce délek period nepřetržitého provozu. Odtud plyne, že

- \* střední doba čekání ve frontě LIFO je rovna  $\frac{\rho}{\mu - \lambda}$ , což je stejné jako v režimu FIFO
- \* rozptyl doby čekání ve frontě LIFO je  $\frac{\lambda(\rho^2 - \rho + 2)}{(\mu - \lambda)^2}$ , což je větší, než při režimu FIFO.

- *Náhodný výběr pro obsluhu*  
Vždy je (pro jakýkoli frontový režim)  $E(R) = E(W) + \frac{1}{\mu} = \frac{1}{\mu - \lambda}$  a  $\text{Var}(R) = \text{Var}(W) + \frac{1}{\mu^2}$ . Jest

$$\text{Var}_{FIFO}(W) \leq \text{Var}(W) \leq \text{Var}_{LIFO}(W).$$

## Kapitola 6

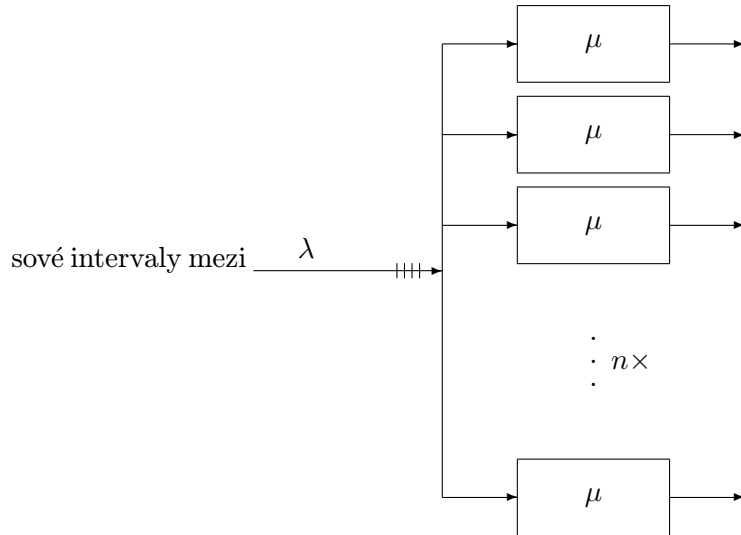
# Modely M/M/n

### 6.1 Podmínka stability systému M/M/n

Podmínka  $\lambda \leq \mu$  stability systému  $M/M/1$  je často velmi těžko realizovatelná. V takovém případě máme několik možností:

- a) zvýšit intenzitu obsluhy  $\mu$  – toho lze dosáhnout organizačními či technickými prostředky. Můžeme však narazit na fyzické meze obsluhy.
- b) omezit příliv zákazníků, t.j. snížit intenzitu  $\lambda$  – ekonomickými či administrativními opatřeními. Sníží se však výkon obslužného systému.
- c) znásobit intenzitu obsluhy  $\mu$  zvýšením počtu obslužných linek. To znamená přejít k systému  $M/M/n$ ,  $1 < n < \infty$ .

Zákazníci přicházející do systému se řadí do jediné fronty. Jejich příchody tvoří Poissonovský *ustupní proud*, to jest příchody jsou nezávislé, ča-



nimi jsou náhodné a mají exponenciální rozdělení pravděpodobnosti s parametrem  $\lambda$ . Zákazníci postupují dále do některé z  $n$  obslužných stanic, z nichž každá může najednou obsluhovat jediného zákazníka, přičemž doba obsluhy se opět považuje za náhodnou veličinu s exponenciálním rozložením s parametrem  $\mu$ . Je-li některá stanice obsluhy volná, ihned se do ní zařadí zákazník čekající ve frontě a po ukončení obsluhy obslužený zákazník opouští stanici a tím i celý systém obsluhy. V takto postaveném modelu se prakticky kapacita obsluhy násobí  $n$ -krát a tedy podmínka stability zde bude zřejmě ve tvaru  $\lambda < n\mu$ .

Stavy procesu  $X(t)$  i v tomto případě jsou počty zákazníků v systému. Vyjádříme pravděpodobnosti přechodu v takovémto systému: zde musíme rozlišit tři případy

- 1) První případ odpovídá situaci, kdy je systém prázdný –

$$\begin{aligned} p_{00}(t) &= 1 - \lambda t + o(t), \\ p_{01}(t) &= \lambda t + o(t), \\ p_{0k}(t) &= o(t), \quad k = 2, 3, \dots \end{aligned}$$

- 2) Jiná situace nastane v případě, kdy systém není zcela zaplněn, to znamená, že  $1 \leq j \leq n$  (není prázdný, ale ještě se nevytváří fronta):

$$\begin{aligned} p_{jj-1}(t) &= j\mu t + o(t), \\ p_{jj}(t) &= 1 - \lambda t - j\mu t + o(t), \\ p_{jj+1}(t) &= \lambda t + o(t), \\ p_{jk}(t) &= o(t), \quad \text{pro } |j - k| > 1 \end{aligned}$$

- 3) Bude-li systém zaplněn, t.j. kdy  $j < n$  (příchozí zákazníci se řadí do

fronty), potom dostaneme

$$\begin{aligned} p_{jj-1}(t) &= n\mu t + o(t), \\ p_{jj}(t) &= 1 - \lambda t - n\mu t + o(t), \\ p_{jj+1}(t) &= \lambda t + o(t), \\ p_{jk}(t) &= o(t), \quad \text{pro } |j - k| > 1 \end{aligned}$$

Nyní můžeme sestavit soustavu Kolmogorovových diferenciálních rovnic

$$\begin{aligned} p'_0(t) &= -\lambda p_0(t) + \mu p_1(t), \\ p'_k(t) &= \lambda p_{k-1}(t) - (\lambda + k\mu)p_k(t) + (k+1)\mu p_{k+1}(t), \quad \text{pro } 1 \leq k < n \\ p'_n(t) &= \lambda p_{n-1}(t) - (\lambda + n\mu)p_n(t) + n\mu p_{n+1}(t), \quad \text{pro } n \leq k \end{aligned}$$

a pro stabilizovaný systém

$$\begin{aligned} 0 &= -\lambda p_0 + \mu p_1, \\ 0 &= \lambda p_{k-1} - (\lambda + k\mu)p_k + (k+1)\mu p_{k+1}, \quad \text{pro } 1 \leq k < n \\ 0 &= \lambda p_{n-1} - (\lambda + n\mu)p_n + n\mu p_{n+1}, \quad \text{pro } n \leq k \end{aligned}$$

s podmínkou  $\sum_{k=0}^{\infty} p_k = 1$ .

Pro řešení této soustavy zavedme substituci  $z_k = \lambda p_{k-1} - k\mu p_k, k = 1, \dots, n$ . Potom výše uvedená soustava dostane tvar

$$\begin{aligned} 0 &= z_1, \\ 0 &= z_k, \quad \text{pro } k = 1, \dots, n-1 \end{aligned}$$

Odtud dostáváme  $z_k = 0$  pro  $k = 1, \dots, n$  a tedy – při označení  $\beta = \frac{\lambda}{\mu}$  –

$$p_k = \beta \frac{p_{k-1}}{k}, \quad \text{neboli} \quad p_k = p_0 \frac{\beta^k}{k!}.$$

Dále má být  $n\mu(p_{k+1} - p_k) = \lambda(p_k - p_{k-1}), k = n, n+1, \dots$

Sečtením  $r$  rovnic (pro  $j = n, n+1, \dots, n+r$ ) dostaneme

$$\begin{aligned} n\mu(p_{n+r+1} - p_n) &= \lambda(p_{n+r} - p_{n-1}) \\ n\mu p_{n+r+1} - z_n &= \lambda p_{n+r} \end{aligned}$$

a protože  $z_n = 0$ , bude

$$p_{n+r+1} = \frac{\beta}{n} p_{n+r}, \quad \text{neboli} \quad p_{n+r+1} = p_0 \left(\frac{\beta}{n}\right)^{r+1} \frac{\beta^n}{n!},$$

tedy pro  $k \geq n$  je

$$p_k = p_0 \frac{\beta^k}{n! n^{k-n}} = p_n \left(\frac{\beta}{n}\right)^{k-n}.$$

$$\sum_{k=0}^{\infty} p_k = 1 \Rightarrow p_0 = \left[ \sum_0^{n-1} \frac{\beta^k}{k!} + \frac{n^n}{n!} \sum_n^{\infty} \left(\frac{\beta}{n}\right)^k \right]^{-1}$$

Z posledního vztahu plyne nutná podmínka pro stabilitu systému, t.j. pro to, aby  $p_k > 0, k = 0, 1, 2, \dots$ :

$$\rho = \frac{\beta}{n} = \frac{\lambda}{n\mu} < 1 \iff \lambda < n\mu.$$

## 6.2 Charakteristiky systému M/M/n

- *Pravděpodobnost okamžité obsluhy (bez čekání)*

$$\pi = \sum_{k=0}^{n-1} p_k = p_0 \sum_{k=0}^{n-1} \frac{\beta^k}{k!}.$$

Zákazník bude čekat s pravděpodobností

$$1 - \pi = \sum_{k=n}^{\infty} p_k = p_0 \frac{\beta^n}{n!} \cdot \sum_{k=0}^{\infty} \left(\frac{\beta}{n}\right)^k = \frac{p_0 \frac{\beta^n}{n!}}{1 - \rho} = \frac{p_n}{1 - \rho}.$$

- *Průměrný počet zákazníků ve frontě*

$$\gamma = \sum_{r=0}^{\infty} r p_{n+r} = p_n \sum_{r=0}^{\infty} r \rho^r = \frac{p_n \rho}{1 - \rho} = \frac{p_0 \beta^{n+1}}{(n-1)!(n-\beta)^2}$$



- Průměrný počet obsazených obslužných linek

$$\begin{aligned} \nu &= \sum_{k=1}^n kp_k + \sum_{k=n+1}^{\infty} np_k = p_0 \sum_{k=1}^n k \frac{\beta^k}{k!} + p_0 \sum_{k=n+1}^{\infty} n \beta^k \frac{n^{n-k}}{n!} = \\ &= p_0 \beta \left[ \sum_0^{n-1} \frac{\beta^k}{k!} + \frac{n^n}{n!} \sum_n^{\infty} \left( \frac{\beta}{n} \right)^k \right] = \beta = \frac{\lambda}{n} = n\rho. \end{aligned}$$

Průměrný počet neobsazených obslužných linek je potom  $(n - \beta) = n(1 - \rho)$ , je-li  $\beta < n$ , t.j.  $\rho < 1$ .

- Průměrný počet zákazníků v systému

$$\begin{aligned} \kappa &= \sum_{k=0}^{\infty} kp_k = \sum_{k=1}^n kp_k + \sum_{k=n+1}^{\infty} kp_k = \\ &= \sum_{k=1}^n kp_k + \sum_{k=n+1}^{\infty} np_k + \sum_{k=n+1}^{\infty} (k - n)p_k = \beta + \gamma. \end{aligned}$$

- Rozložení doby čekání  $W$

$$P(W > 0) = \Pi = 1 - P(W = 0)$$

$$P(W > w) = \sum_{k=1}^{\infty} p_{n+k-1} P(W_k > w),$$

kde  $W_k$  má stejný význam jako v předchozí kapitole. Jest

$$P(W_k > w) = \sum_{j=0}^{k-1} e^{-n\mu w} \frac{(n\mu w)^j}{j!}$$

a tedy po dosazení

$$\begin{aligned} P(W > w) &= \sum_{k=0}^{\infty} \left( \frac{\beta}{n} \right)^k p_n \sum_{j=0}^k e^{-n\mu w} \frac{(n\mu w)^j}{j!} = \\ p_n e^{-n\mu w} \sum_{k=0}^{\infty} \sum_{j=0}^k \left( \frac{\beta}{n} \right)^k \frac{(n\mu w)^j}{j!} &= p_n e^{-n\mu w} \sum_{j=0}^{\infty} \frac{(n\mu w)^j}{j!} \sum_{k=j}^{\infty} \left( \frac{\beta}{n} \right)^k = \\ \frac{p_n e^{-n\mu w} \sum_{j=0}^{\infty} \frac{\left( \frac{n\mu\beta w}{n} \right)^j}{j!}}{\left( 1 - \frac{\beta}{n} \right)} &= \frac{p_n e^{-(n\mu-\lambda)w}}{1 - \frac{\beta}{n}} = \Pi e^{-(n\mu-\lambda)w} \end{aligned}$$

Všimněte si, že toto rozložení má obdobný charakter jako v případě systému M/M/1: má skok o velikosti  $(1 - \Pi)$  v bodě nula (diskrétní složka) a spojitá složka je exponenciální s parametrem  $(n\mu - \lambda)$ . Pro střední hodnotu a rozptyl platí

$$E(W) = \frac{\Pi}{n\mu - \lambda}, \quad \text{Var}(W) = \frac{\Pi(2 - \Pi)}{n\mu - \lambda}$$

*Cvičení:* Jaké je rozdělení doby setrvání v systému  $R$ ?

### 6.3 Systém M/M/ $\infty$

Tento systém je limitním případem předchozího. Soustavy diferenciálních rovnic pro stacionární rozdělení (Kolmogorovova soustava) se od předchozí liší pouze tím, že odpadají poslední řádky pro  $k \geq n$  a zbývající rovnice platí pro  $k = 1, 2, \dots$ . Řešení má potom tvar

$$p_k = \frac{p_0 \beta^k}{k!} = e^{-\beta} \frac{\beta^k}{k!}, \quad k = 0, 1, 2, \dots$$

Z toho plyne, že počet zákazníků v systému má Poissonovo rozdělení s parametrem  $\beta = \frac{\lambda}{\mu}$ . To je také průměrný počet obsazených linek.

### 6.4 Systém M/M/n/r (systém s omezenou frontou)

Funkce tohoto systému se od klasického M/M/n (neboli M/M/n/ $\infty$ ) liší v tom, že pokud má fronta délku  $r \geq 0$ , další příchozí zákazníci nezůstávají v systému a ihned odcházejí neobslouženi. Tento model si lze představit jako systém s omezenou kapacitou čekárny nebo zásobníku. Soustava Kolmogorových rovnic bude oproti minulému modelu rozšířena ještě o rovnici pro  $k = n + r$ :

$$\begin{aligned} p'_0(t) &= -\lambda p_0(t) + \mu p_1(t), \\ p'_k(t) &= \lambda p_{k-1}(t) - (\lambda + k\mu)p_k(t) + (k+1)\mu p_{k+1}(t), \quad \text{pro } 1 \leq k < n \\ p'_k(t) &= \lambda p_{k-1}(t) - (\lambda + n\mu)p_k(t) + n\mu p_{k+1}(t), \quad \text{pro } n \leq k < n+r \\ p'_{n+r}(t) &= \lambda p_{n+r-1}(t) - n\mu p_{n+r}(t). \end{aligned}$$

Jedná se tedy o konečnou soustavu diferenciálních rovnic, která má vždy řešení a také vždy zde existují limity  $\lim_{t \rightarrow \infty} p_k(t)$ , bez ohledu na velikost poměru  $\rho = \frac{\lambda}{\mu}$ . To znamená, že takovýto systém se vždy stabilizuje (proč? – otázka pro cvičení). Platí zde

$$\begin{aligned} 0 &= -\lambda p_0 + \mu p_1, \\ 0 &= \lambda p_{k-1} - (\lambda + k\mu)p_k + (k+1)\mu p_{k+1}, \quad \text{pro } 1 \leq k < n \\ 0 &= \lambda p_{k-1} - (\lambda + n\mu)p_k + n\mu p_{k+1}, \quad \text{pro } n \leq k < n+r \\ 0 &= \lambda p_{n+r-1} - n\mu p_{n+r}. \end{aligned}$$

s podmínkou  $\sum_{k=0}^{n+r} p_k = 1$ .  
Odtud dostáváme řešení

$$\begin{aligned} p_k &= p_0 \frac{\beta^k}{k!}, \quad k = 0, 1, \dots, n, \\ p_k &= p_0 \frac{\beta^k}{n! n^{k-n}} = p_n \left(\frac{\beta}{n}\right)^{k-n}, \quad k = n+1, \dots, n+r, \end{aligned}$$

kde je opět  $\beta = \frac{\lambda}{\mu}$ . Z podmínky pro  $p_0$  dostáváme

$$p_0 = \left[ \sum_0^n \frac{\beta^k}{k!} + \frac{\beta^n}{n!} \sum_1^r \left(\frac{\beta}{n}\right)^k \right]^{-1}$$

Dále si uvedeme několik charakteristik tohoto modelu:

- $p_{n+r}$  je pravděpodobnost ztráty (náhodně vybraného) zákazníka
- $\sum_{k=0}^{n-1} p_k$  je pravděpodobnost, že zákazník bude obslužen bez čekání
- $\sum_{k=n}^{n+r-1} p_k$  je pravděpodobnost, že zákazník bude muset čekat
- $\kappa = \sum_{k=1}^{n+r} k p_k$  = průměrný počet zákazníků v systému
- $\gamma = \sum_{k=1}^r k p_{n+k}$  = průměrný počet zákazníků ve frontě

$$- \nu = \sum_{k=1}^n k p_k + n \sum_{k=1}^r p_{n+k} = \beta(1 - p_{n+r}) = \text{průměrný počet obsa-} \\ \text{zených obslužných linek}$$

Dobu čekání  $W$  a její pravděpodobnostní charakteristiky je třeba odvozovat jako podmíněné podmínkou, že nedošlo ke ztrátě zákazníka.

$P(\text{zákazník se zařadí na místo } k \text{ ve frontě — neztratí se}) = \frac{p_{n+k-1}}{1-p_{n+r}}$   
 $W_k$  se řídí stejnými zákony rozdělení jako v systému M/M/ $\infty$  (předpokládáme režim FIFO), tedy

$$\begin{aligned} P(W > w) &= \sum_{k=0}^{r-1} \frac{p_{n+k}}{1-p_{n+r}} P(W_k > w) = \\ &= \frac{p_n}{1-p_{n+r}} \sum_{k=0}^{r-1} \left(\frac{\beta}{n}\right)^k \int_w^\infty e^{-n\mu y} \frac{(n\mu y)^k}{k!} n\mu dy = \\ &= \frac{p_n}{1-p_{n+r}} \int_{n\mu w}^\infty e^{-z} \sum_{k=0}^{r-1} \frac{(\lambda z)^k}{k!} dz, \\ P(W = 0) &= \frac{\sum_{k=0}^{n-1} p_k}{\sum_{k=0}^{n+r-1} p_k}, \end{aligned}$$

$$E(W_k) = \frac{k}{n\mu}, \quad (\text{jedná se o Erlangovo rozdělení})$$

$$E(W) = \sum_{k=0}^{r-1} \frac{p_{n+k}}{1-p_{n+r}} \frac{k+1}{n\mu} = \frac{\frac{\beta^n}{n!} \sum_{k=0}^{r-1} (k+1)\rho^k}{n\mu \left( \sum_{k=0}^{n-1} \frac{\beta^k}{k!} + \frac{1-\rho^n}{n!(1-\rho)} \right)},$$

kde  $\rho = \frac{\beta}{n} = \frac{\lambda}{n\mu}$ .

*Poznámka:* Pro  $r \rightarrow \infty$  dostaneme obyčejný systém M/M/n s neomezenou frontou. Potom ovšem potřebujeme opět podmínku  $\lambda < n\mu$ , neboli  $\rho < 1$ .

*Cvičení:* Spočítejte charakteristiky systému M/M/n/0, to jest v případě, že se fronta vůbec nevytváří.

## Kapitola 7

# Nemarkovské modely, model $M/G/n$

Předpoklad exponenciálního rozložení, který vyhovuje například představě rozložení doby obsluhy v telefonní ústředně nebo v obchodě, není zřejmě příliš vhodným pro řadu jiných modeů. Exponenciální rozložení totiž připouští s kladnou pravděpodobností libovolně krátkou dobu obsluhy. Rozložení doby zpracování zakázky ve výrobě nebo doby přepravy nákladu či cestujících však tuto vlastnost obvykle nemá. Zpravidla existuje jistá minimální doba pro provedení potřebných operací či překonání vzdálenosti, která však může být z různých příčin, nejčastěji způsobených náhodnými vlivy, delší. Proto lze předpokládat spíše různé typy zleva posunutých či useknutých rozložení doby obsluhy. Například v případě pravidelné (kyvadlové) dopravy, kdy je sice doba dopravy pro všechny jízdy přibližně stejná, ale může kolísat na obě strany, lze předpokládat (zleva useknuté) normální rozložení doby přepravy. V případě automatizované přepravy, například v provozu metra, můžeme rozložení doby přepravy popsat posunutým exponenciálním rozložením, kde posunutí je dáno minimální dobou potřebnou k průjezdu úseku a exponenciální rozložení zde odpovídá působení náhodných vlivů, které mohou dobu průjezdu prodloužit (překážky na trase, nával cestujících, drobné závady apod.).

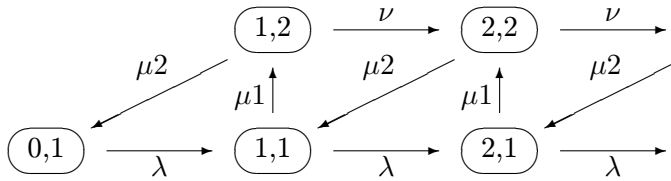
Do jisté míry lze problém obecnějšího rozložení doby obsluhy řešit předpokladem vícefázové obsluhy, kdy každá fáze (stadium) obsluhy má exponenciální rozložení. Tuto metodu, které se říká *metoda stadií* (viz [Basket & spol, 1975]) si naznačíme v následujícím příkladu.

### 7.1 Příklad: Systém s jednou obslužnou stanicí a dvoufázovou obsluhou.

Uvažujme systém, kdy máme jednu pracovní stanicí, v níž jsou obsluhováni zákazníci ve dvou fázích:

- přípravné, která bude probíhat s intenzitou  $\mu_1$ .
- vlastní obsluha s intenzitou  $\mu_2$ .

Příchody zakázek tvoří Poissonův proces s intenzitou příchodu  $\lambda$ . Stav systému lze charakterizovat dvojicí  $(N(t), I(t))$ , kde  $N(t)$  je počet zakázek v systému a  $I(t) \in \{1, 2\}$  je fáze obsluhy. Grafické znázornění tohoto systému je na následujícím obrázku.



Při dlouhém provozu budeme požadovat, aby byl systém v „rovnovážném stavu“, to znamená, aby průměrný počet zakázek přicházejících do každého uzlu byl roven průměrnému počtu zakázek, které tento uzel opustí. Uvedený požadavek je vyjádřen soustavou rovnic

$$\begin{aligned} \lambda p_{n-1,1} + \mu_2 p_{n+1,2} &= (\lambda + \mu_1) p_{n,1} & \mu_2 p_{1,2} &= \lambda p_{0,1} \\ \lambda p_{n-1,2} + \mu_1 p_{n,1} &= (\lambda + \mu_2) p_{n,2} & \mu_1 p_{1,1} &= (\lambda + \mu_2) p_{1,2} \end{aligned}$$

pokud existují limity  $\lim_{t \rightarrow \infty} P(N(t) = n, I(t) = i) = p_{n,i}$ ,  $n = 0, 1, 2, \dots$ ,  $i = 1, 2$ .  $p_{n,i}$  potom tvoří stacionární rozložení.

Výpočet lze provést rekurentně pomocí vztahů

$$\begin{aligned} p_{1,2} &= \frac{\lambda}{\mu_2} p_{0,1} & p_{1,1} &= \frac{\lambda + \mu_2}{\mu_1} p_{1,2} \\ p_{n,2} &= \left( \frac{\lambda + \mu_1}{\mu_2} p_{n,1} - \frac{\lambda}{\mu_2} p_{n-1,2} \right) & p_{n+1,2} &= \left( \frac{\lambda + \mu_1}{\mu_2} p_{n,1} - \frac{\lambda}{\mu_2} p_{n-1,1} \right) \end{aligned}$$

pro  $n = 1, 2, \dots$

## 7.1 Systémy s obecně rozloženou dobou obsluhy

Nyní uvedeme jiný přístup, který sice neposkytuje v obecném případě spočítat stacionární rozložení, ale umožňuje spočítat některé důležité charakteristiky.

Předpokládejme tedy, že doba obsluhy zakázky v systému je náhodná veličina se spojitým rozložením pravděpodobnosti s hustotou  $f(t)$ . Vstupní proud zakázek nechť tvoří opět Poissonovský proces jako v předchozích příkladech. V Kendallově klasifikaci (viz [Zítek, 1969]) je takovýto model označen jako M/G/1. Budeme sledovat střední dobu  $W$ , kterou stráví zakázka v systému od svého příchodu až do opuštění systému po ukončení obsluhy. Tato doba je součtem středních hodnot tří složek:

$W_1$  = střední doba, která zbývá do dokončení právě probíhající obsluhy zakázky,

$W_2$  = střední čas potřebný ke zpracování zakázek čekajících v systému,

$W_3$  = střední doba vlastní obsluhy zakázky.

Z předpokladu dostáváme

$$W_3 = \int_0^{\infty} t f(t) dt = s.$$

Číslo  $\rho = \lambda s$  je potom intenzitou provozu. Označíme-li  $L$  střední počet všech zakázek v systému, pak střední počet čekajících zakázek je  $(L - \rho)$  a  $W_2 = (L - \rho)s$ .

Zbývá spočítat  $W_1$ . Tento čas je roven

$$W_1 = \rho \int_0^{\infty} \frac{t}{2} g(t) dt,$$

kde  $g(t) = \frac{t f(t)}{\int_0^{\infty} u f(u) du}$ . Potom

$$W_1 = \frac{\rho}{2s} \int_0^{\infty} t^2 f(t) dt = \rho \frac{M_2(s)}{2s},$$

označíme-li  $M_2(s)$  druhý obecný moment rozdělení doby obsluhy zakázky. Tedy

$$W = \rho \frac{M_2(s)}{2s} + (L - \rho)s + s.$$

Mezi  $L$  a  $W$  platí tzv. „Littleova formule“:  $L = \lambda W$ . Dosazením této formule a jednoduchou úpravou dostáváme vztah, kterému se v teorii hromadné obsluhy říká „Pollaczekova–Chinčinova formule“:

$$W = \frac{\rho M_2(s)}{2s(1-\rho)} + s.$$

## 7.2 Příklad 7.3

Nechť zakázky přicházejí do systému s intenzitou  $\lambda$  a doba obsluhy zakázky má exponenciální rozdělení s intenzitou  $\mu$  (Příklad 1.1). Střední hodnota tohoto rozložení je  $\frac{1}{\mu}$  a druhý obecný moment  $\frac{2}{\mu^2}$ .

Potom střední celková doba, kterou stráví zakázka v systému bude

$$W = \frac{2\rho\mu^{-2}}{2\mu^{-1}(1-\rho)} + \frac{1}{\mu} = \frac{1}{\mu} \frac{\rho}{1-\rho} + \frac{1}{\mu} = \frac{1}{\mu(1-\rho)} = \frac{1}{\mu - \lambda}$$

a střední počet zakázek v systému

$$L = \frac{\rho}{1-\rho} = \frac{\lambda}{\mu - \lambda}.$$

**Příklad 7.4.** Nechť doba obsluhy zakázky má normální rozdělení se střední hodnotou  $\tau$  a směrodatnou odchylkou  $\sigma = \frac{\tau}{2}$ . Zakázky přicházejí do systému s intenzitou  $\lambda$ . Potom střední celková doba, kterou stráví zakázka v systému bude

$$W = \frac{5\rho\tau}{8(1-\rho)} + \tau$$

a střední počet zakázek v systému

$$L = \frac{5\rho^2}{8(1-\rho)} + \rho.$$

## 7.3 Systém s minimální dobou obsluhy

Nechť zakázky přicházejí do systému s intenzitou  $\lambda$ . Předpokládejme, že doba zpracování zakázky je náhodná veličina  $T_W = t_0 + T_R$ , kde  $t_0$  je minimální doba zpracování zakázky, daná výrobní technologií a  $T_R$  je náhodná



veličina s exponenciálním rozdělením s intenzitou  $\mu$ .  $T_R$  je doba, o kterou se zpracování prodlouží vlivem různých poruch, výpadků apod. Střední hodnota veličiny  $T_W$  je potom rovna  $s = t_0 + \frac{1}{\mu}$  a druhý obecný moment  $M_2(s) = t_0^2 + 2t_0\frac{1}{\mu} + \frac{2}{\mu^2} = s^2 + \frac{1}{\mu^2}$ .

Potom střední celková doba, kterou stráví zakázka v systému bude

$$W = \frac{\rho(s^2 + \mu^{-2})}{2s(1 - \rho)} + s = \frac{s^2(2 - \rho) + \rho\mu^{-2}}{2s(1 - \rho)} = \frac{s}{2(1 - \rho)} \left( 2 - \rho + \frac{\rho}{s^2\mu^2} \right)$$

a střední počet zakázek v systému

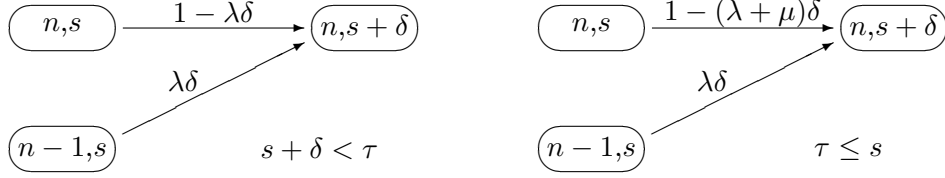
$$L = \frac{\rho}{2(1 - \rho)} \left( 2 - \rho + \frac{\rho}{s^2\mu^2} \right).$$

Zabývejme se nyní poněkud blíže systémem popsáním v tomto příkladu. Doba mezi příchody zakázek zde má exponenciální rozložení s parametrem  $\lambda$  a doba obsluhy má posunuté exponenciální rozložení s parametry  $\tau$  a  $\mu$  (jeho hustota je na obr. 1.5). Stavy systému lze popsat dvojicemi  $X(t) = (N(t), S(t))$ , kde  $N(t) \in \mathbb{N}$  je počet zakázek v systému v čase  $t \in \mathbb{R}$  a  $S(t) \geq 0$  je doba, která uplynula od počátku zpracovávání právě zpracovávané zakázky. Pro  $n = 1, 2, \dots$  označme

$$\begin{aligned} p(n, s; t) ds &= P(N(t) = n, S(t) \in (s, s + ds)) & s < \tau \\ \pi(n; t) &= P(N(t) = n, S(t) \geq \tau) & s \geq \tau. \end{aligned}$$

Dále položíme  $p(0, 0; t) = p_0$ . Zřejmě je  $p(0, s; t) = 0$  pro  $s > 0$  a  $\pi(0; t) = 0$ .

Uvažujme časový interval  $(s, s + \delta)$ ,  $s + \delta < \tau$ . Je-li systém ve stavu  $(n, s + \delta)$ , mohl se tam dostat s pravděpodobností  $\lambda\delta + o(\delta)$  ze stavu  $(n - 1, s)$  (došla nová zakázka) nebo s pravděpodobností  $1 - \lambda\delta + o(\delta)$  ze stavu  $(n, s)$  (nová zakázka nedošla). V intervalu  $(s, s + \delta)$ ,  $s \geq \tau$  je situace podobná: je-li systém ve stavu  $(n, s + \delta)$ , mohl se tam dostat s pravděpodobností  $\lambda\delta + o(\delta)$  ze stavu  $(n - 1, s)$  (došla nová zakázka) nebo s pravděpodobností  $1 - (\lambda + \mu)\delta + o(\delta)$  ze stavu  $(n, s)$  (nedošla nová zakázka a obsluha neskončilo). Po ukončení obsluhy zpracovávané zakázky přejde systém ze stavu  $(n, s)$  do stavu  $(n - 1, 0)$  s pravděpodobností  $\mu\delta$ .



Sestavíme nyní *Kolmogorouovy prospektivní rovnice* pro popisovaný systém. Z předchozího vyplývá, že pro  $n = 1, 2, \dots$  :

$$\begin{aligned}
 p(n, s; t + dt) &= p(n, s - dt; t)(1 - \lambda dt) + p(n - 1, s - dt; t)\lambda dt + o(dt) \\
 \pi(n; t + dt) &= \pi(n; t)(1 - (\lambda + \mu)dt) + \pi(n - 1; t)\lambda dt + o(dt) \\
 p(n, 0; t + dt) &= \pi(n + 1; t)\mu dt + o(dt) \\
 p(0, 0; t + dt) &= p(0, 0; t)(1 - \lambda dt) + \pi(1; t)\mu dt + o(dt)
 \end{aligned}$$

Limitním přechodem pro  $dt \rightarrow 0$  dostaneme soustavu rovnic

$$\begin{aligned}
 \frac{\partial}{\partial t} p(n, s; t) &= \frac{\partial}{\partial s} p(n, s; t) - \lambda p(n, s; t) + \lambda p(n - 1, s; t), \quad n = 1, 2, \dots \\
 \frac{\partial}{\partial t} \pi(n; t) &= -(\lambda + \mu)\pi(n; t) + \lambda\pi(n - 1; t), \quad n = 2, 3, \dots \\
 \frac{\partial}{\partial t} p(0, 0; t) &= -\lambda p(0, 0; t) + \mu\pi(1; t)
 \end{aligned}$$

s počátečními podmínkami

$$p(n, 0; t) = \mu\pi(n + 1; t) \quad \text{pro } n = 1, 2, \dots,$$

za předpokladu, že příslušné parciální derivace existují. Pokud existuje stacionární rozložení, tzn. pokud existují limity

$$p_n(s) = \lim_{t \rightarrow \infty} p(n, s; t) \quad \text{a} \quad \pi_n = \lim_{t \rightarrow \infty} \pi(n; t),$$

potom při  $t \rightarrow \infty$  dostáváme soustavu diferenciálně-diferenčních rovnic

$$\begin{aligned}
 0 &= \frac{\partial}{\partial s} p_n(s) - \lambda p_n(s) + \lambda p_{n-1}(s), \quad n = 1, 2, \dots \\
 0 &= -(\lambda + \mu)\pi_n + \lambda\pi_{n-1}, \quad n = 2, 3, \dots
 \end{aligned}$$

s počátečními podmínkami

$$p_0 = \frac{\mu}{\lambda} \pi_1, \quad p_n(0) = \mu \pi_{n+1}, \quad n = 1, 2, \dots$$

Rovnice (1.4.3) se řeší podobným způsobem jako (1.2.2). Jejich řešení dostaneme ve tvaru

$$\pi_n = \left( \frac{\lambda}{\lambda + \mu} \right)^{n-1} \pi_1, \quad n = 2, 3, \dots,$$

přičemž  $\pi_1$  lze vyjádřit z počáteční podmínky jako  $\pi_1 = \frac{\lambda}{\mu} p_0$  a tedy

$$\pi_n = \left( \frac{\lambda}{\lambda + \mu} \right)^{n-1} \frac{\lambda}{\mu} p_0 = \left( \frac{\lambda}{\lambda + \mu} \right)^n \frac{\lambda + \mu}{\mu} p_0, \quad n = 1, 2, \dots$$

Řešení rovnic (1.4.3) budeme předpokládat ve tvaru  $K(s)\rho^n e^{-s}$ . Dosazením do rovnic a jednoduchým výpočtem obdržíme řešení

$$p_n(s) = \rho^n e^{-s} \left( e^{\frac{\lambda}{\rho} s} + C \right),$$

které dosadíme do počáteční podmínky, ze které dostaneme  $C = \lambda p_0$  a  $\rho = \frac{\lambda}{\lambda + \mu}$ . Potom

$$p_n(s) = \left( \frac{\lambda}{\lambda + \mu} \right)^n e^{-s} \left( e^{(\lambda + \mu)s} + \lambda p_0 \right), \quad s \in (0, \tau), \quad n = 1, 2, \dots$$

Zbývá určit konstantu  $p_0$ , která je určena podmínkou

$$p_0 + \sum_{n=1}^{\infty} \left( \int_0^{\tau} p_n(s) ds + \pi_n \right) = 1. \quad (*)$$

Protože je

$$\begin{aligned} \int_0^{\tau} p_n(s) ds &= \left( \frac{\lambda}{\lambda + \mu} \right)^n \left[ \int_0^{\tau} e^{(\lambda + \mu - 1)s} ds + \lambda p_0 \int_0^{\tau} e^{-s} ds \right] = \\ &= \left( \frac{\lambda}{\lambda + \mu} \right)^n \left[ \frac{e^{(\lambda + \mu - 1)\tau} - 1}{\lambda + \mu - 1} + (1 - e^{-\tau}) \lambda p_0 \right], \end{aligned}$$

budeme vyšetřovat zvlášť dva případy: (a)  $\lambda + \mu \neq 1$  a (b)  $\lambda + \mu = 1$ .

(a) Necht' tedy  $\lambda + \mu \neq 1$ . Dosazením do podmínky  $(\star)$  dostaneme

$$p_0 = \frac{\mu(\lambda + \lambda\mu + \mu^2 - \mu - \lambda e^{(\lambda+\mu-1)\tau})}{(\lambda + \mu - 1)(\lambda^2 + \lambda^2\mu + \lambda\mu + \mu^2 - \lambda^2\mu e^{-\tau})}.$$

(b) Pokud  $\lambda + \mu = 1$ , pak

$$p_n(s) = \lambda^n(1 + \lambda p_0 e^{-s}), \quad \pi_n = \lambda^n \frac{p_0}{\mu}$$

a podmínka  $(\star)$  bude mít tvar

$$p_0 + \sum_{n=1}^{\infty} (\tau + \lambda p_0(1 - e^{-\tau})) + \lambda^n \frac{p_0}{\mu} = p_0 + \frac{\lambda}{1 - \lambda} \left( \tau + \lambda p_0(1 - e^{-\tau}) + \frac{1}{\mu} p_0 \right) = 1$$

odkud

$$p_0 = \frac{1 - \lambda - \tau\lambda}{1 - \lambda + \lambda^2(1 - e^{-\tau}) + \frac{\lambda}{\mu}}.$$

**Poznámka.** Uvedené výsledky neplatí pro případ, kdy doba obsluhy je nenáhodná, vždy stejná, rovnající se konstantě (například obsluha automatem s pevně nastaveným taktom). V tomto případě výsledek neovlivňuje variabilita, způsobená náhodnou dobou obsluhy a nelze odvodit soustavu rovnic, obdobných  $(\star)$ .

## Kapitola 8

## Modely $G/G/n$



## Kapitola 9

# Modely se spojitým tokem

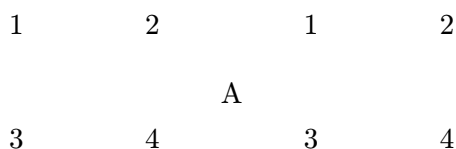




## Kapitola 10

# Obslužné sítě, úvod

Pod pojmem „sítě obslužných stanic“ nebo „obslužná síť“ budeme rozumět systém *stanic obsluhy* vzájemně propojených fyzickými i logickými vazbami. Jednotlivé *zakázky (zákazníci)* vstupují v určitém místě do systému, pohybují se dále od stanice ke stanici buď podle předem stanoveného plánu obsluhy nebo náhodně (např. je-li v systému více stanic, vykonávajících stejnou činnost), dokud jejich obsluha není ukončena. Poté systém opouštějí. Grafické schéma jednoduché obslužné sítě si lze představit jako některý z grafů na následujícím obrázku.



Obrázky znázorňují jednoduché sítě čtyř obslužných stanic. Na obrázku vlevo přicházející zakázky (do stanic 1 a 3) tvoří tzv. *vstupní proud*, obslužené zakázky opouštějí systém (ze stanic 2 a 4) v tzv. *výstupním proudu*. Takovýto systém se nazývá *otevřený*.

Pokud je počet zakázek v systému konstantní v čase (např. když každá obslužená zakázka se ihned opět zařazuje do systému nebo je nahrazena

novou), můžeme si systém představit jako *uzavřený*, v němž se zakázky cyklicky pohybují mezi stanicemi, žádné nepřibývají ani neodcházejí. Schéma uzavřeného systému je na obrázku vpravo.

Studiu takovýchto sítí se věnuje řada autorů již od 50. let. Jednou z nejstarších aplikací je model J. Taylora a R. R. P. Jacksona z roku 1954, zabývající se pohybem strojů a obslužných zařízení mezi letadly na letištní ploše a řazením obslužených strojů. Podobné systémy byly použity pro modelování výrobní dílny s několika odděleními, mezi nimiž procházejí zakázky před jejich zkompletováním (J. R. Jackson, 1957), model pohybu strojů a lidí mezi zařízeními uhelného dolu (E. Koenigsberg, 1958) a řady dalších. Od 70. let se pomocí podobných sítí modeluje činnost, probíhající v elektrických obvodech, speciálně v obvodech centrálních procesorových jednotek číslicových počítačů (Baskett, Chandy, Herzog a řada dalších autorů). Teoretické základy této problematiky lze nalézt v knize F. P. Kellyho (1979).

Jednou z nejdůležitějších otázek, spojených s analýzou obslužných sítí je otázka *rovnováhy systému*, o které se zmíníme v následujících odstavcích.

## 10.1 Příklad systému se dvěma obslužnými stanicemi

V následujícím příkladu se budeme zabývat systémem, který sestává ze dvou obslužných stanovišť, která si lze představit jako dvě obslužné stanice nebo jednu stanicí s dvoufázovou obsluhou. Všechny zakázky přicházející do systému s intenzitou příchodu  $\lambda$  musí být nejprve obslouženy na stanici 1 a poté budou obslouženy na stanici 2. Po ukončení obsluhy na stanici 2 zakázka systém opustí. U každé ze stanic se může vytvářet fronta zakázek, čekajících na obsluhu, přičemž stanice 1 pracuje s intenzitou  $\mu$  a stanice 2 s intenzitou  $\nu$ . Stav systému lze charakterizovat dvojicí  $[n_1(t), n_2(t)]$ , kde  $n_1(t)$  je počet zakázek v systému u stanici 1 (zakázky ve frontě a ve fázi obsluhy),  $n_2(t)$  je počet zakázek v systému u stanici 2. Celkový počet zakázek v systému je tedy  $n(t) = n_1(t) + n_2(t)$ . Na následujícím obrázku jsou zobrazeny intenzity přechodu v uvedeném modelu.

$$\begin{array}{cccc}
 0,0 & \nu & 0,1 & 1,3 \\
 & \mu & \lambda & \\
 1,0 & & 1,1 & 1,2 \\
 & & \text{A} & \\
 2,0 & & 2,1 & 0,2
 \end{array}$$

Požadavek rovnováhy systému je v tomto případě vyjádřen soustavou diferenčních rovnic

$$\begin{aligned}
 \mu p_{m+1,n-1} + \lambda p_{m-1,n} + \nu p_{m,n+1} &= (\lambda + \mu + \nu) p_{m,n}, & \nu p_{0,1} &= \lambda p_{0,0}, \\
 \mu p_{1,n-1} + \nu p_{0,n+1} &= (\lambda + \nu) p_{0,n}, & \lambda p_{m-1,0} + \nu p_{m,1} &= (\lambda + \mu) p_{m,0}
 \end{aligned}$$

pro  $m, n = 1, 2, \dots$ . Řešení soustavy budeme hledat ve tvaru  $p_{m,n} = \alpha^m \beta^n$ . Dosadíme-li tento předpoklad do soustavy, dostaneme soustavu

$$\begin{aligned}
 \mu \alpha^2 + \lambda \beta + \nu \alpha \beta^2 &= (\lambda + \mu + \nu) \alpha \beta, & \nu \beta^2 &= \lambda, \\
 \mu \alpha + \nu \beta^2 &= (\lambda + \nu) \beta, & \lambda + \nu \alpha \beta &= (\lambda + \mu) \alpha,
 \end{aligned}$$

jejíž řešení je  $\alpha = \frac{\lambda}{\mu}$ ,  $\beta = \frac{\lambda}{\nu}$ . Z podmínky  $\sum_{m,n=1}^{\infty} p_{m,n} = 1$  dostaneme ještě normalizační konstantu. Konečné řešení je potom

$$p_{m,n} = \left(1 - \frac{\lambda}{\mu}\right) \left(1 - \frac{\lambda}{\nu}\right) \left(\frac{\lambda}{\mu}\right)^m \left(\frac{\lambda}{\nu}\right)^n.$$

Toto řešení odpovídá představě, že obě pracovní stanice pracují nezávisle jako dva systémy M/M/1 (viz Příklad 1.1). Potom je i  $p_{m,n} = p_m p_n$ .

## 10.2 Obslužný systém jako markovský proces

Matematický model obslužné sítě je tvořen *náhodným procesem* se spočetnou množinou stavů  $\mathcal{S} = \{\mathcal{C}_\infty, \mathcal{C}_\epsilon, \dots\}$  a se spojitým časem  $t \in \mathcal{T} \subset \mathbb{R}$ . Prvky množiny  $\mathcal{S}$  obsahují všechny informace, které jednoznačně popisují stavy obslužné linky z hlediska její činnosti v čase.

Proces  $X(t)$  budeme nazývat *stacionárním*, jestliže pro libovolnou  $n$ -tici časů  $t_1, t_2, \dots, t_n \in \mathcal{T}$  a pro libovolné  $\tau \in \mathcal{T}$  jsou sdružená rozložení pravděpodobnosti náhodných vektorů  $(X(t_1), X(t_2), \dots, X(t_n))$  a  $(X(t_1 + \tau), X(t_2 + \tau), \dots, X(t_n + \tau))$  totožná.

Náhodný proces  $X(t)$  se nazývá *reverzibilním*, jestliže pro libovolnou  $n$ -tici časů  $t_1, t_2, \dots, t_n \in \mathcal{T}$  a pro libovolné  $\tau \in \mathcal{T}$  jsou sdružená rozložení pravděpodobnosti náhodných vektorů  $(X(t_1), X(t_2), \dots, X(t_n))$  a  $(X(\tau - t_1), X(\tau - t_2), \dots, X(\tau - t_n))$  totožná.

Proces  $X(t)$  budeme nazývat *markovským*, jestliže pro libovolnou  $n$ -tici časů  $t_1 < t_2 < \dots < t_n \in \mathcal{T}$  platí

$$\begin{aligned} P(X(t_n) = C_n | X(t_1) = C_1, X(t_2) = C_2, \dots, X(t_{n-1}) = C_{n-1}) = \\ = P(X(t_n) = C_n | X(t_{n-1}) = C_{n-1}), \end{aligned} \quad (10.1)$$

pokud má jev  $(X(t_1) = C_1, X(t_2) = C_2, \dots, X(t_{n-1}) = C_{n-1})$  kladnou pravděpodobnost. Budeme-li chápat čas  $t_{n-1}$  jako „současnost“, potom lze uvedenou vlastnost interpretovat jako „zapomínání procesu“. Tato vlastnost totiž říká, že pravděpodobnost budoucího stavu v čase  $t_n$  závisí pouze na současném stavu  $X(t_{n-1})$  a nezávisí na stavech minulých.

Markovský proces se nazývá *homogenní v čase*, nezávisí-li pravděpodobnost  $p(C, D, \tau) = P(X(t + \tau) = D | X(t) = C)$  na čase  $t$ .  $p(C, D, \tau)$  označuje pravděpodobnost přechodu mezi stavy  $C$  a  $D$  za dobu  $\tau$ .

Proces  $X(t)$  se nazývá *irreducibilní*, jestliže pro každou dvojici  $C, D \in \mathcal{S}$  existuje  $\tau \in \mathcal{T}$  tak, že  $p(C, D, \tau) > 0$ . Tato vlastnost znamená, že každý stav  $C \in \mathcal{S}$  je dosažitelný z kteréhokoli jiného stavu procesu  $X(t)$ .

*Intenzitu přechodu* mezi stavy  $C$  a  $D$  definujeme jako limitu

$$q(C, D) = \begin{cases} \lim_{\tau \rightarrow 0} \frac{p(C, D, \tau)}{\tau} & C \neq D, \\ 0 & C = D, \end{cases}$$

Označme  $q(C) = \sum_{D \in \mathcal{S}} q(C, D)$  *intenzitu opuštění* stavu  $C$ .

Hodnota  $q(C)$  má následující interpretaci: doba, po kterou proces  $X(t)$  setrvává ve stavu  $C$  má exponenciální rozložení pravděpodobnosti s parametrem  $q(C)$ . Tedy *střední doba setrvání procesu ve stavu  $C$*  je rovna  $\frac{1}{q(C)}$ .

Nadále budeme předpokládat, že  $q(C)$  je konečná a že proces nemůže provést nekonečně mnoho změn stavu v konečném čase.

Pomocí  $q(C, D)$  a  $q(C)$  lze definovat *pravděpodobnost přechodu* mezi stavy  $C$  a  $D$  vztahem

$$p(C, D) = \frac{q(C, D)}{q(C)}, \quad C, D \in \mathcal{S}.$$

### 10.3 Rovnováha systému hromadné obsluhy

Pojem *rovnováhy* pochází z teorie homogenních markovských procesů. Existuje-li posloupnost nezáporných čísel  $\{\pi(C)\}_{C \in \mathcal{S}}$  takových, že  $\sum_{C \in \mathcal{S}} \pi(C) = 1$ , splňujících rovnice

$$\pi(C) \sum_{D \in \mathcal{S}} q(C, D) = \sum_{D \in \mathcal{S}} \pi(D) q(D, C), \quad C \in \mathcal{S} \quad (10.2)$$

nazýváme ji *rovnovážným rozložením* pravděpodobnosti a rovnice (10.2) se nazývají rovnice rovnováhy procesu  $X(t)$ . Pokud rovnovážné rozložení  $\pi(C)$  existuje, je určeno rovnicemi (10.2) jednoznačně a je zároveň *stacionárním rozložením*.

Pokud rovnovážné rozložení neexistuje, potom

- (i) pro všechna  $C \in \mathcal{S}$  je  $\lim_{t \rightarrow \infty} P(X(t) = C) = 0$
- (ii) lze nalézt množinu nezáporných čísel  $\pi(C)$  splňujících (1) takovou, že  $\sum_{C \in \mathcal{S}} \pi(C) = \infty$

Má-li množina stavů  $\mathcal{S}$  konečný počet prvků, potom rovnovážné rozložení existuje vždy.

Existuje-li rovnovážné rozložení  $\pi(C)$ , potom pro všechny dvojice stavů  $C, D \in \mathcal{S}$  platí tzv. *rovnice detailní rovnováhy*

$$\pi(C)q(C, D) = \pi(D)q(D, C). \quad (10.3)$$

V jednoduchých případech, jako jsou příklady minulého odstavce, vystačíme při hledání rovnovážného rozložení s rovnicemi (10.2) či (10.3). V řadě komplikovanějších případů je tento postup neschůdný. Proto se zavádí pojem *reverzibility* a *opačného procesu*. Věnujme se nyní poněkud blíže vztahu mezi stacionaritou a reverzibilitou markovského procesu  $\{X(t), t \in \mathbb{R}\}$ .

**Lemma 1** Každý reverzibilní proces je stacionární.

*Důkaz* Je zřejmý, uvědomíme-li si, že pro reverzibilní proces je sdružené rozložení  $(X(t_1), X(t_2), \dots, X(t_n))$  i  $(X(t_1 + \tau), X(t_2 + \tau), \dots, X(t_n + \tau))$  stejné jako rozložení  $(X(\tau - t_1), X(\tau - t_2), \dots, X(\tau - t_n))$  pro libovolnou  $n$ -tici časů  $t_1, t_2, \dots, t_n \in \mathcal{T}$  a pro libovolné  $\tau \in \mathcal{T}$ .

Všimněme si, že z lemmatu 1 plyne, že pokud proces  $X(t)$  není stacionární, není ani reverzibilní, i kdyby byly rovnice (10.3) splněny. Naopak, stacionarita sama o sobě nestačí k tomu, aby proces byl reverzibilní. Následující lemma dává nutnou a postačující podmínku pro reverzibilitu markovského procesu.

**Lemma 2** Stacionární markovský proces je reverzibilní právě když existuje rovnovážné rozložení  $\pi(C)$ ,  $C \in \mathcal{S}$ , splňující rovnice detailní rovnováhy (10.3).

*Důkaz* tohoto Lemmatu lze nalézt v [Kelly, 1979, str. 7–8].

Obrácený proces  $X(-t)$  si lze představit jako proces  $X(t)$  sledovaný „pozpátku“. Označme dále  $q'(C)$ ,  $q'(C, D)$  intenzity přechodu obráceného procesu  $X(-t)$ .

**Lemma 3** Je-li  $X(t)$  stacionární markovský proces s intenzitami přechodu  $q(C, D)$ ,  $C, D \in \mathcal{S}$  a rovnovážným rozložením  $\pi(C)$ ,  $C \in \mathcal{S}$ , potom obrácený proces  $X(-t)$  je stacionární markovský proces s intenzitami přechodu

$$q'(D, C) = \frac{\pi(C)q(C, D)}{\pi(D)}, \quad C, D \in \mathcal{S}$$

a se stejným rovnovážným rozložením pravděpodobnosti  $\pi(C)$ ,  $C \in \mathcal{S}$ .

*Důkaz* Z markovské vlastnosti procesu  $X(t)$  plyne rovnost

$$P(X(t) = C)P(X(t+\tau) = D | X(t) = C) = P(X(t+\tau) = D)P(X(t) = C | X(t+\tau) = D),$$

neboli

$$\pi(C)p(C, D, \tau) = \pi(D)p(D, C, -\tau), \quad C, D \in \mathcal{S}.$$

Vydělením obou stran rovnice číslem  $\tau$  a limitním přechodem  $\tau \rightarrow 0$  dostaneme rovnici

$$\pi(C)q(C, D) = \pi(D)q'(D, C), \quad C, D \in \mathcal{S}. \quad (10.4)$$

Z rovnice (10.4) plyne tvrzení lemmatu 3.

Rovnice (10.4) jsou opět rovnicemi detailní rovnováhy. Navíc, jsou-li tyto rovnice splněny, platí  $q(C) = q'(C)$  pro všechna  $C \in \mathcal{S}$ . Tento důsledek je zřejmý, uvědomíme-li si, že  $\frac{1}{q(C)}$  je střední doba setrvání procesu  $X(t)$  ve stavu  $C$  a ta je stejná i pro proces  $X(-t)$ .

**Tvrzení 1** *Předpokládejme, že  $X(t)$  je stacionární markovský proces s intenzitami přechodů  $q(C, D)$ ,  $q(C)$ . Pokud lze nalézt množinu kladných čísel  $\pi(C)$ ,  $C \in \mathcal{S}$  takových, že  $\sum_{C \in \mathcal{S}} \pi(C) = 1$  a nezáporná čísla  $q'(C, D)$ ,  $q'(C) = \sum_{D \neq C} q'(C, D)$  splňující rovnice (10.4), potom  $\pi(C)$  je stacionárním rozdělením procesu  $X(t)$  a zpětný proces  $X(-t)$  má intenzity přechodu  $q'(C, D)$  a  $q'(C)$ .*

*Důkaz* Sečteme-li rovnice (10.4) při pevném  $D$  přes všechna  $C \in \mathcal{S}$ , dostaneme

$$\sum_{C \in \mathcal{S}} \pi(C)q(C, D) = \pi(D) \sum_{C \in \mathcal{S}} q'(D, C) = \pi(D)q'(D) = \pi(D)q(D).$$

Tedy  $\pi(D)$  je rovnovážným rozložením procesu  $X(t)$ . Z lemmatu 3 vyplývá, že  $q'(C, D)$  a  $q'(C)$  jsou intenzity přechodu procesu  $X(-t)$ .

## 10.4 Model otevřené obslužné sítě

Pokusme se nyní o poněkud obecnější vymezení modelu obslužné sítě. Popisovaný model zavedl F. P. Kelly, 1976. Představme si síť jako otevřený systém  $J$  stanišť v nichž jsou zpracovávány zakázky  $I$  typů. Na každém stanovišti se zakázky řadí do front a jsou zpracovávány na  $K$  stanicích obsluhy s určitou intenzitou. Vstup zakázek typu  $i$  do sítě tvoří Poissonův proces s intenzitou  $\lambda_i$ ,  $i = 1, 2, \dots, I$ . Každá zakázka typu  $i$  má předem stanovenou posloupnost  $r^i = (r(i, 1), r(i, 2), \dots, r(i, S_i))$  stanovišť, kterými musí postupně projít než opustí síť (plán obsluhy).

**Jakým způsobem lze popsat stavy systému v takovémto modelu?** Řekneme, že zakázka je v síti na pozici  $(j, l)$ , je-li zařazena na  $l$ -tém místě ve frontě u stanice  $j$ . Podívejme se blíže na  $j$ -té stanoviště. Zde je fronta  $n_j$  zakázek různých typů a v různém stadiu obsluhy. Tuto frontu lze popsat vektorem

$$\mathbf{c}^j(t) = (c_1^j(t), c_2^j(t), \dots, c_{n_j}^j(t)) \quad j = 1, 2, \dots, J,$$

kde  $c_l^j(t) = (t_l^j, s_l^j)$ ,  $t_l^j$  identifikuje typ zakázky na pozici  $(j, l)$  a  $s_l^j$  označuje fázi obsluhy, ve které se tato zakázka v čase  $t$  nalézala.

Stav celé sítě je potom jednoznačně popsán vektorem

$$\mathbf{C} = (\mathbf{c}^1, \mathbf{c}^2, \dots, \mathbf{c}^J).$$

**Jak síť funguje?** Podrobněji si nyní popíšeme funkci sítě na  $j$ -tém stanovišti. Zavedeme několik předpokladů:

(I) – Každá zakázka na  $j$ -tém stanovišti vyžaduje obsluhu, jejíž „velikost“ je náhodná veličina s exponenciálním rozložením se střední hodnotou 1 (měřeno v jednotkách množství práce, energie či nákladů, potřebných pro obsluhu zakázky).

(II) – Stanoviště  $j$  zpracovává zakázky s celkovou intenzitou  $\phi^j(n_j)$  (měřeno v jednotkách množství práce, energie či nákladů za jednotku času). Budeme předpokládat, že pokud je  $n_j > 0$ , potom i  $\phi^j(n_j) > 0$ .

(III) – Předpokládáme, že na stanovišti lze obsluhovat více zakázek najednou. Zakázka na pozici  $(j, l)$  bude obsluhována s intenzitou, která je rovna  $\gamma_l^j(n_j)\phi^j(n_j)$ ,  $l = 1, 2, \dots, n_j$ .  $\gamma_l^j(n_j)$  představuje podíl z celkové intenzity  $\phi^j(n_j)$  poskytované na stanovišti  $j$  zakázce na místě  $l$  ve frontě, obsahující celkem  $n_j$  zakázek. Předpokládáme, že je  $\sum_{l=1}^{n_j} \gamma_l^j(n_j) = 1$ .

(IV) – Jeli obsluha zakázky na místě  $l$  ukončena, zakázka opouští frontu. Zakázky na místech  $l+1, l+2, \dots, n_j$  se posunou na místa  $l, l+1, \dots, n_j-1$ .

(V) – Pokud je obsluha zakázky na stanovišti  $j$  ukončena a byla poslední v jejím plánu, zakázka opustí síť. Nebyla-li tato obsluha poslední, zakázka se zařadí do fronty u  $k$ -tého stanoviště, kde  $k = r(t_l^j, s_l^j + 1)$ . V  $k$ -té frontě se zakázka zařadí na místo  $l$  s pravděpodobností  $\delta_l^k(n_k)$ ,  $l = 1, 2, \dots, n_k$ . Přitom zakázky, které byly na místech  $l, l+1, \dots, n_k$  se posunou na místa  $l+1, l+2, \dots, n_k+1$ . Opět předpokládáme, že  $\sum_{l=1}^{n_k} \delta_l^k(n_k) = 1$ .

**Příklad 10.4.1.** V příkladu 10.1.1 jsme uvažovali síť dvou obslužných stanic se sériovým zpracováním jednoho typu zakázek. To znamená, že v tomto modelu je  $J = 2$ ,  $I = 1$ ,  $r(1, 1) = 1$ ,  $r(1, 2) = 2$ . Obě stani-ceobsluhují vždy jen jednu zakázku s intenzitami  $\mu, \nu$ , tedy je

$$\phi^1(m) = \mu, \quad \gamma_1^1(m) = 1, \quad \gamma_l^1(m) = 0, \quad l = 2, 3, \dots, m, \quad (10.5)$$

$$\phi^2(n) = \nu, \quad \gamma_1^2(m) = 1, \quad \gamma_l^2(m) = 0, \quad l = 2, 3, \dots, m, \quad (10.6)$$



pro libovolná  $m, n$ , což jsou délky front u stanice 1, resp. 2. Pravděpodobnosti zařazení zakázky na  $l$ -té místo ve frontě jsou

$$\delta_l^1(m) = 0, \quad l = 1, 2, 3, \dots, m-1, \quad \delta_m^1(m) = 1, \quad (10.7)$$

$$\delta_l^2(n) = 0, \quad l = 1, 2, 3, \dots, n-1, \quad \delta_n^2(n) = 1 \quad (10.8)$$

(zakázky se řadí vždy na konec fronty).

**Příklad 10.4.2.** Zakázky mají být zpracovávány v síti popsané v minulém příkladu, pouze s tím rozdílem, že po zpracování ve druhé stanici se musí vrátit ještě jednou na stanici první. V takovémto případě zůstane popis modelu podobný jako v příkladu 1, jen se změní „plán obsluhy“:  $r(1,1) = 1$ ,  $r(1,2) = 2$ ,  $r(1,3) = 1$ .

**Příklad 10.4.3** popisuje situaci, kdy  $j$ -té stanoviště je schopno obsluhovat až  $k$  zakázek najednou se stejnou intenzitou. Potom je

$$\phi^j(m) = \mu_j \cdot \min(k, m), \quad (10.9)$$

$$\gamma_l^j(m) = \frac{1}{m}, \quad l = 1, 2, \dots, m, \quad m \leq k \quad (10.10)$$

$$\gamma_l^j(m) = \frac{1}{k}, \quad l = 1, 2, \dots, k, \quad m > k \quad (10.11)$$

$$\gamma_l^j(m) = 0, \quad l = k+1, k+2, \dots, m, \quad m > k \quad (10.12)$$

**Jak vypadají intenzity přechodu?** Přejít od stavu  $C$  ke stavu  $D$  je v našem modelu určen plánem obsluhy  $r$ . Ten určuje, kam a kdy se která zakázka v síti přemístí. Na množině stavů sledovaného procesu zavedeme operátory  $T$ , odpovídající následujícím přechodům:

- (i) *Vstup nové zakázky do sítě.* Nová zakázka typu  $i$  vstupuje do sítě na stanoviště  $j = r(i, 1)$ . Byla-li síť před příchodem nové zakázky ve stavu  $C$  a nová zakázka se zařadí na stanoviště  $j$  do fronty na místo  $m$ , označíme tento nový stav jako  $T^{im}C$ .
- (ii) *Přechod zakázky v síti.* Předpokládejme, že síť je ve stavu  $C$  a obsluha zakázky na pozici  $(j, l)$  je ukončena. Není-li tato obsluha poslední v plánu  $r$ , přejde zakázka do dalšího stanoviště  $k = r(t_l^j, s_l^j + 1)$  na místo  $m$ . Tento nový stav označíme  $T_{jlm}C$ .

- (iii) *Odchod zakázky ze sítě.* Předpokládejme, že síť je ve stavu  $C$  a obsluha zakázky typu  $i$  na pozici  $(j, l)$  je ukončena a je poslední v seznamu  $\mathbf{r}^i$ . V tomto případě zakázka opustí síť a tato přejde do stavu který označíme  $T_{jl}C$ .

Jiné typy přechodů v síti v modelu nepředpokládáme. Odpovídající intenzity přechodu zřejmě budou

- (i) Intenzita příchodu zakázky typu  $i$  do sítě je  $\lambda_i$ . Do fronty u stanoviště  $k = r(i, 1)$  na místo  $m$  se zařadí s pravděpodobností  $\delta_m^k(n_k + 1)$ , kde  $n_k$  je délka fronty  $k$  před zařazením. Tedy je

$$q(C, T^{im}C) = \lambda_i \delta_m^k(n_k + 1).$$

- (ii) Intenzita obsluhy zakázky na pozici  $(j, l)$  je  $\gamma_l^j(n_j)\phi^j(n_j)$ . To spolu s pravděpodobností zařazení na nové stanoviště  $k = r(t_l^j, s_l^j + 1)$  na místo  $m$  dává intenzitu

$$q(C, T_{jlm}C) = \gamma_l^j(n_j)\phi^j(n_j)\delta_m^k(n_k + 1).$$

- (iii) Intenzita odchodu ze soustavy bude podobná předcházejícímu případu, pouze s tím rozdílem, že zakázka se už nikam nezařazuje. Proto

$$q(C, T_{jl}C) = \gamma_l^j(n_j)\phi^j(n_j).$$

**Jak vypadá obrácený proces?** Obrácený proces popisuje soustavu stejného typu, s obráceným pořadím průchodů zakázek stanovišti. Do této soustavy přicházejí zakázky typu  $i$  s intenzitou  $\lambda_i$ ,  $i = 1, 2, \dots, I$  a procházejí stanovišti s plánem  $\mathbf{r}_i = (r(i, S_i), r(i, S_i - 1), \dots, r(i, 1))$ . Nejpodstatnější změnou je záměna funkcí intenzit  $\gamma_i^j(n_j)$  a  $\delta_i^j(n_j)$ . Tuto záměnu lze vysvětlit následujícím způsobem: zatímco  $\delta_i^j(n_j)$  je pravděpodobností zařazení zakázky  $i$  na místo vstupu do fronty  $j$ ,  $\gamma_i^j(n_j)$  lze interpretovat jako pravděpodobnost zařazení zakázky v této frontě na místo, odkud po obslužení odejde. Při obráceném procesu se tato místa obměňují – vstupní se stane výstupním a naopak. Proto se obrací i úloha funkcí  $\gamma$  a  $\delta$ . Tedy

- (i') Intenzita příchodu zakázky typu  $i$  do obrácené sítě na pozici  $(j, l)$  je

$$q'(T_{jl}C, C) = \lambda_i \gamma_l^j(n_j).$$

- (ii') Intenzita přechodu zakázky ze stanoviště  $k = r(t_l^j, s_l^j + 1)$  z místa  $m$  zpět na pozici  $(j, l)$  je

$$q'(T_{jlm}C, C) = \delta_m^k(n_k + 1)\phi^k(n_k + 1)\gamma_l^j(n_j).$$

- (iii') Intenzita opuštění obrácené soustavy bude podobná případu (ii), pouze s tím rozdílem, že zakázka se už nikam nezařazuje. Proto

$$q'(T^{im}C, C) = \delta_m^k(n_k + 1)\phi^k(n_k + 1).$$

Nyní odvodíme rovnovážné rozložení a rovnice detailní rovnováhy. Nejprve vyjádříme střední počet  $a_j$  zakázek typu  $i$ , přicházejících na stanoviště  $j$  za jednotku času. Za rovnovážného stavu bude  $a_j$  rovno součtu středních počtů přicházejících zakázek na stanoviště  $j$  pro všechny typy  $i$ . Zavedme funkce  $\alpha_j(i, s)$

$$\alpha_j(i, s) = \begin{cases} \lambda_i & \text{když } r(i, s) = j, \\ 0 & \text{jindy.} \end{cases}$$

Funkce  $\alpha_j(i, s)$  reprezentuje intenzitu příchodu zakázky  $i$  na stanoviště  $j$  (je nenulová, pokud má zakázka  $i$  uzel  $j$  v plánu a je rovna nule, pokud zakázka  $i$  uzel  $j$  nenavštíví). Potom

$$a_j = \sum_{i=1}^I \sum_{s=1}^{S(i)} \alpha_j(i, s).$$

Hledání rovnovážného rozložení má smysl jen tehdy, nebude-li se u žádného stanoviště vytvářet „nekonečná“ fronta, tzn. bude-li mít každá zakázka kladnou pravděpodobnost, že bude obsloužena. Tato podmínka je vyjádřena nerovností

$$b_j = \sum_{n=0}^{\infty} \left( a_j^n / \prod_{l=1}^n \phi^j(l) \right) < \infty \quad \text{pro } j = 1, 2, \dots, J. \quad (10.13)$$

Tato podmínka bude splněna například pokud  $a_j < \max_{1 \leq l \leq n_j} \phi^j(l)$ .

**Tvrzení 2** *Rovnovážné rozložení v popisovaném systému je rovno*

$$\pi(C) = \prod_{j=1}^J \pi_j(c^j),$$

kde  $\pi_j(\mathbf{c}^j) = \frac{1}{b_j} \prod_{l=1}^{n_j} (\alpha_j(t_l^j, s_l^j) / \phi^j(l))$ .

*Důkaz* Podle Tvrzení 1 stačí ukázat, že  $\pi(C)$  splňuje rovnice detailní rovnováhy (10.4). Označme  $b = \left[ \prod_{j=1}^J b_j \right]^{-1}$  normalizační konstantu. Dosaďme do všech případů (i)–(iii).

$$\begin{aligned}
\text{(i)} \quad \pi(C)q(C, T^{im}C) &= b \prod_{j=1}^J \pi_j(\mathbf{c}^j) \lambda_i \delta_m^k(n_k+1) = B(C) \pi_k(\mathbf{c}^k) \lambda_i \delta_m^k(n_k+1) = \\
&= B(C) \left( \prod_{l=1}^{n_k+1} \alpha_k(t_l^k, s_l^k) / \prod_{l=1}^{n_k+1} \phi^k(l) \right) \phi^k(n_k+1) \delta_m^k(n_k+1) = \\
B(C) \pi_k(T^{im} \mathbf{c}^k) \phi^k(n_k+1) \delta_m^k(n_k+1) &= \pi(T^{im}C) q'(T^{im}C, C). \quad (10.14)
\end{aligned}$$

$$\begin{aligned}
\text{(ii)} \quad \pi(C)q(C, T_{hlm}C) &= b \prod_{j=1}^J \pi_j(\mathbf{c}^j) \gamma_l^h(n_h) \phi^h(n_h) \delta_m^k(n_k+1) = B(C) \pi_h(\mathbf{c}^h) \\
\pi_k(\mathbf{c}^k) \gamma_l^h(n_h) \phi^h(n_h) \delta_m^k(n_k+1) &= B(C) \left( \prod_{l=1}^{n_h-1} \alpha_h(t_l^h, s_l^h) / \prod_{l=1}^{n_h-1} \phi^h(l) \right) \\
\left( \prod_{l=1}^{n_k+1} \alpha_k(t_l^k, s_l^k) / \prod_{l=1}^{n_k+1} \phi^k(l) \right) &\gamma_l^h(n_h) \phi^k(n_k+1) \delta_m^k(n_k+1) = \\
= B(C) \pi_h(T_{hlm} \mathbf{c}^h) \pi_k(T_{hlm} \mathbf{c}^k) q'(T_{hlm}C, C) &= \pi(T_{hlm}C) q'(T_{hlm}C, C). \quad (10.15)
\end{aligned}$$

$$\begin{aligned}
\text{(iii)} \quad \pi(C)q(C, T_{hl}C) &= b \prod_{j=1}^J \pi_j(\mathbf{c}^j) \gamma_l^h(n_h) \phi^h(n_h) = B(C) \pi_h(\mathbf{c}^h) \gamma_l^h(n_h) \phi^h(n_h) = \\
= B(C) \left( \prod_{l=1}^{n_h-1} \alpha_h(t_l^h, s_l^h) / \prod_{l=1}^{n_h-1} \phi^h(l) \right) &\lambda_i \gamma_l^h(n_h) = \\
B(C) \pi_h(T_{hl} \mathbf{c}^h) q'(T_{hl}C, C) &= \pi(T_{hl}C) q'(T_{hl}C, C). \quad (10.16)
\end{aligned}$$

Dále je zřejmé

$$q(C) = q'(C) = \sum_{j=1}^J \phi^j(n_j) + \sum_{i=1}^I \lambda(i).$$

Speciálně platí následující důsledky Tvrzezení 1 a 2.

**Důsledek 1** Je-li  $X(t)$  stacionární otevřená síť odpovídající výše popsanému modelu, potom i obrácený proces  $X(-t)$  je stacionární otevřenou sítí stejného typu.

**Důsledek 2** Je-li síť v rovnovážném stavu, potom

- (i) Zakázky typu  $i$  opouštějí soustavu v Poissonovském proudu s intenzitou  $\lambda_i$ ,  $i = 1, 2, \dots, I$ . Tyto proudy jsou nezávislé a stav  $X(t_0)$  je nezávislý na odchodech ze soustavy před časem  $t_0$ .
- (ii) Stav  $\mathbf{c}^j$  systému ve stanici  $j$  je nezávislý na stavech v ostatních stanicích a má pravděpodobnost  $\pi_j(\mathbf{c}^j)$ . Pravděpodobnost, že délka fronty ve stanici  $j$  bude rovna  $n$  je

$$P(n_j = n) = \frac{1}{b_j} \frac{a_j^n}{\prod_{l=1}^n \phi^j(l)}. \quad (10.17)$$

- (iii) Zakázka přicházející podle svého plánu na stanici  $j$  „nachází“ tuto stanici ve stavu  $\mathbf{c}^j$  s pravděpodobností  $\pi_j(\mathbf{c}^j)$ . Pravděpodobnost, že ve frontě bude  $n$  čekajících zakázek je rovna (10.13).
- (iv) Pravděpodobnost, že zakázka ve frontě u stanice  $j$  na místě  $l$  je typu  $i$  ve stadiu obsluhy  $s$  (na  $s$ -tém místě ve svém plánu) je rovna  $\alpha_j(i, s)/a_j$ .

**Příklad 2.4.4.** Uvažujme následující model sítě dvou pracovních stanic. Zakázky, přicházející do stanice 1 s intenzitou  $\lambda$  jsou dvojího typu: první typ může ihned po obslužení na stanici 1 síť opustit, zatímco druhý typ je po obslužení ve stanici 1 obsluhován ve stanici 2 a vrácen opět do stanice 1. Teprve poté síť opustí. Zakázek druhého typu je v průměru  $\frac{1}{3}$ , stanice 1 pracuje s intenzitou  $\phi^1 = \mu$ , stanice 2 s intenzitou  $\phi^2 = \nu$ . Obsluha v obou stanicích bude probíhat podle disciplíny FIFO (*First In First Out*), což je charakterizováno pravděpodobnostmi

$$\gamma_1^1(m) = 1, \quad \gamma_l^1(m) = 0, \quad l = 2, 3, \dots, m, \quad (10.18)$$

$$\gamma_1^2(m) = 1, \quad \gamma_l^2(m) = 0, \quad l = 2, 3, \dots, m, \quad (10.19)$$

$$\delta_m^1(m) = 1, \quad \delta_l^1(m) = 0, \quad l = 1, 2, \dots, m-1, \quad (10.20)$$

$$\delta_n^2(n) = 1, \quad \delta_l^2(n) = 0, \quad l = 1, 2, \dots, n-1. \quad (10.21)$$

Zakázka 1 vstupuje do sítě s intenzitou  $\frac{2}{3}\lambda$  a její plán sestává z jedné položky  $\mathbf{r}^1 = r(1, 1) = 1$ . Zakázka typu 2 má vstupní intenzitu  $\frac{1}{3}\lambda$  a řídí se plánem

$r^2 = (r(2, 1), (r(2, 2))) = (1, 2)$ . Stav sítě lze plně charakterizovat trojicí  $(n_1^1, n_1^2, n_2)$ , kde  $n_1^1$  je počet zakázek 1. typu ve stanici 1,  $n_1^2$  je počet zakázek typu 2 ve stanici 1 a  $n_2$  je počet zakázek ve stanici 2 (kde mohou být pouze zakázky 2. typu). Označme  $n_1 = n_1^1 + n_1^2$ . Dosazením do výrazů v tvrzení 2 dostaneme

$$\pi_1(n_1^1, n_1^2) = \frac{1}{b_1} \prod_{l=1}^{n_1} \frac{\alpha_1}{\mu} = \frac{1}{b_1} \left( \frac{2\lambda}{3\mu} \right)^{n_1^1} \left( \frac{\lambda}{3\mu} \right)^{n_1^2} = \frac{1}{b_1} 2^{n_1^1} \left( \frac{\lambda}{3\mu} \right)^{n_1}$$

$$\pi_2(n_2) = \frac{1}{b_2} \prod_{l=1}^{n_2} \frac{\alpha_2}{\nu} = \frac{1}{b_2} \left( \frac{\lambda}{3\nu} \right)^{n_2}$$

Za předpokladu  $4\lambda < 3\mu$  a  $\lambda < 3\nu$  má potom normalizační konstanta tvar

$$b = \frac{1}{b_1} \frac{1}{b_2} = \left( 1 - \frac{4\lambda}{3\mu} \right) \left( 1 - \frac{\lambda}{3\nu} \right)$$

Odtud

$$\pi(n_1^1, n_1^2, n_2) = \left( 1 - \frac{4\lambda}{3\mu} \right) \left( 1 - \frac{\lambda}{3\nu} \right) 2^{n_1^1} \left( \frac{\lambda}{3\mu} \right)^{n_1} \left( \frac{\lambda}{3\nu} \right)^{n_2}.$$

Pravděpodobnosti délky fronty v jednotlivých stanicích jsou

$$P(n_1 = n) = \left( 1 - \frac{4\lambda}{3\mu} \right) \left( \frac{4\lambda}{3\mu} \right)^n$$

$$P(n_2 = n) = \left( 1 - \frac{\lambda}{3\nu} \right) \left( \frac{\lambda}{3\nu} \right)^n$$

a střední počet zakázek ve stanici

$$L_1 = \sum_{n=0}^{\infty} n P(n_1 = n) = \left( 1 - \frac{4\lambda}{3\mu} \right) \sum_{n=0}^{\infty} n \left( \frac{4\lambda}{3\mu} \right)^n = \frac{4\lambda}{3\mu - 4\lambda}$$

$$L_2 = \frac{\lambda}{3\nu - \lambda}$$

Střední dobu strávenou ve stanici dostaneme pomocí Littleovy formule

$$W_1 = \frac{1}{\lambda} L_1 = \frac{1}{3\mu - 4\lambda}, \quad W_2 = \frac{1}{\lambda} L_2 = \frac{1}{3\nu - \lambda}.$$

**Příklad 2.4.5** Spočtěme rovnovážné rozložení  $j$ -té stanice pro případ popsaný v příkladu 2.4.3. Zakázky jsou pouze jednoho typu, přicházejí do sítě v Poissonovském proudu s intenzitou  $\lambda$  a postupně procházejí stanicemi  $1, 2, \dots, J$ . Stav sítě je opět plně charakterizován vektorem  $(n_1, n_2, \dots, n_J)$ , kde  $n_j$  je počet zakázek ve stanici  $j$ ,  $j = 1, 2, \dots, J$ . Označme  $\rho_j = \frac{\lambda}{\mu_j}$ . Potom

$$\pi(n_j) = \begin{cases} \frac{1}{b_j} \frac{\rho_j^{n_j}}{n_j!}, & n_j \leq k, \\ \frac{1}{b_j} \frac{\rho_j^{n_j}}{k! k^{n_j - k}}, & n_j > k. \end{cases}$$

Konstanta  $b_j$  má tvar (samozřejmě za předpokladu  $\rho < 1$ )

$$b_j = \sum_{i=1}^k \frac{\rho_j^i}{i} + \sum_{i=k+1}^{\infty} \frac{\rho_j^i}{k^{i-k}} = \rho_j + \frac{\rho_j^2}{2} + \dots + \frac{\rho_j^k}{k} + \frac{\rho_j^{k+1}}{k - \rho}.$$

Pravděpodobnosti (2.4.1) mají v tomto případě stejný tvar jako  $\pi(n_j)$ . Střední počet zakázek ve stanici  $j$  potom bude

$$L_j = \frac{1}{b_j} \sum_{i=1}^k \frac{\rho_j^i}{(i-1)!} + \frac{k^k}{k!} \frac{\rho}{k - \rho}.$$





## Kapitola 11

# Obslužné sítě – obecnější rozdělení požadavku na obsluhu

Zakázky v každé stanici lze rozdělit do několika *tříd*. Ujijeme-li označení zavedené dříve, lze tyto třídy reprezentovat dvojicemi  $(i, s)$ , kde  $i$  je typ zakázky a  $s$  je fáze obsluhy a je  $r(i, s) = j$ . Zakázka během svého průchodu stanicí  $j$  svoji třídu nemění. Označme  $\mathcal{C}^j$  množinu všech tříd ve stanici  $j$ .

Požadovaná míra obsluhy (např. množství práce, času, energie) byla v modelu v odstavci 10.4 v předpokladu (I) zavedena jako náhodná veličina s exponenciálním rozložením se střední hodnotou 1 (tedy s parametrem 1). Ukážeme si nyní, jak lze modelovat obecné rozložení požadavku na obsluhu tzv. *metodou stadií*. Představme si, že zakázka třídy  $\varsigma = (i, s)$  ve stanici  $j$  prochází celkem  $z^j(\varsigma)$  stadii obsluhy. Množství obsluhy požadovaná v jednotlivých stadiích nechť mají nezávislá exponenciální rozložení se střední hodnotou  $d^j(\varsigma)$  (např.  $z^j(\varsigma)$  pracovních úkonů, z nichž každý představuje  $d^j(\varsigma)$  práce). Celkové rozložení požadavku na obsluhu ve stanici  $j$  je potom rozložení gamma (speciálně Erlangovo rozložení) s parametry  $z^j(\varsigma)$  a  $\mu^j = \frac{1}{d^j}$ , tedy se střední hodnotou  $z^j(\varsigma)d^j(\varsigma)$  a rozptylem  $z(\varsigma)(d^j(\varsigma))^2$ .

Stav systému ve stanici  $j$  na místě  $l$  budeme popisovat dvojicí  $c_l^j(t) = (\varsigma_l^j, x_l^j)$ , kde  $\varsigma_l^j = (t_l^j, s_l^j)$  je třída, do které patří zakázka na pozici  $(j, l)$  a  $1 \leq x_l^j \leq z^j(t_l^j, s_l^j)$  označuje stadium obsluhy této zakázky. Frontu lze

potom popsat vektorem

$$\mathbf{c}^j(t) = (c_1^j(t), c_2^j(t), \dots, c_{n_j}^j(t)) \quad j = 1, 2, \dots, J,$$

a

$$\mathbf{C}(t) = (\mathbf{c}^1(t), \mathbf{c}^2(t), \dots, \mathbf{c}^J(t))$$

tvoří markovský proces.

Model, popsaný v odstavci 10.4 pozměníme následujícím způsobem:

(I') – Zakázka třídy  $\zeta$  na  $j$ -tém stanovišti vyžaduje obsluhu, jejíž velikost je náhodná veličina s rozložením gamma se střední hodnotou  $z^j(\zeta)d^j(\zeta)$  a rozptylem  $z^j(\zeta)(d^j(\zeta))^2$ .

(II) – (IV) Tyto předpoklady zůstanou beze změny

(V') – Pokud je obsluha zakázky třídy  $c$  na stanovišti  $j$  ukončená a byla poslední v jejím plánu, zakázka opustí síť. Nebyla-li tato obsluha poslední, zakázka se zařadí do fronty u  $k$ -tého stanoviště, kde  $k = r(t_l^j, s_l^j + 1)$ . V  $k$ -té frontě se zakázka zařadí na místo  $l$  s pravděpodobností  $\gamma_l^k(n_k + 1)$ ,  $l = 1, 2, \dots, n_k$ . Přitom zakázky, které byly na místech  $l, l+1, \dots, n_k$  se posunou na místa  $l+1, l+2, \dots, n_k+1$ .

**Příklad 11.1.1.** Necht'  $\gamma_l^j(n) = 0$ ,  $l = 1, 2, \dots, n-1$ ,  $\gamma_n^j(n) = 1$ ,  $n = 1, 2, \dots$ . Tento model popisuje disciplinu LIFO (*Last In First Out*) kterou si lze představit tak, že, zakázky se řadí na jakousi „hromadu“ a ke zpracování je vždy zařazena ta zakázka, která je navrch (tedy ta, co přišla jako poslední).

**Příklad 11.1.2.**  $\gamma_l^j(n) = \frac{1}{n}$ ,  $l = 1, 2, \dots, n = 1, 2, \dots$ . Tento model popisuje stanici, jejíž činnost je rozdělena mezi všechny zakázky, které přijdou se stejnou intenzitou (příklad tzv. „sdílení procesoru“, známý z činnosti číslicových počítačů).

**Příklad 11.1.3.** Model z příkladu 10.4.3 popisuje případ stanice, která je schopna obsluhovat zároveň maximálně  $k$  zakázek najednou se stejnou intenzitou. Pokud je stanice „obsazena“ (je v ní  $k$  zakázek), další zakázky již nepřijímá.

**Jak vypadají intenzity přechodu?** Intenzity přechodu lze odvodit podobně jako v odstavci 10.4 pouze s tím rozdílem, že  $\delta_m^k(n) = \gamma_m^k(n)$  a

musíme počítat i s intenzitou zpracování v jednotlivých stádiích. Označme tuto intenzitu  $\mu_m^k = \frac{1}{d_m^k}$ , kde  $d_m^k = d^k(\varsigma)$ , pokud zakázka na místě  $m$  ve stanici  $k$  je třídy  $\varsigma$ . Dále je třeba uvažovat nové přechody mezi stavy. K přechodům (i)–(iii) z odstavce 10.4 přidáme další

- (iv) *Přechod zakázky z jednoho stadia obsluhy do dalšího* před ukončením obsluhy ve stanici. Označíme-li  $C$  stav sítě v okamžiku kdy je na pozici  $(j, l)$  zakázka třídy  $(i, s)$  ve stadiu obsluhy  $x < z^j(i, s)$ , symbolem  $T_{jlx}C$  označme stav kdy tato zakázka přejde do stadia  $x + 1$ .

Intenzity odpovídající přechodům (i)–(iv) potom budou

- (i)  $q(C, T^{im}C) = \lambda_i \gamma_m^k (n_k + 1)$ ,  
(ii)  $q(C, T_{jlm}C) = \gamma_l^j(n_j) \phi^j(n_j) \mu_l^j \gamma_m^k (n_k + 1)$ ,  
(iii)  $q(C, T_{jl}C) = \gamma_l^j(n_j) \phi^j(n_j)$ ,  
(iv)  $q(C, T_{jlx}C) = \gamma_l^j(n_j) \phi^j(n_j) \mu_l^j$ .

**Jak vypadá obrácený proces?** V tomto procesu reprezentují čísla  $x$  počet stadií, které mají být ještě vykonána před opuštěním stanice zakázkou. Jinak zůstává interpretace obráceného procesu podobná jako v odstavci 10.4. Vzhledem k symetrii, nedochází zde k záměně funkcí intenzit  $\gamma_i^j(n_j)$  a  $\delta_i^j(n_j)$ . Intenzity obráceného procesu jsou

- (i')  $q'(T_{jl}C, C) = \lambda_i \gamma_l^j(n_j)$ ,  
(ii')  $q'(T_{jlm}C, C) = \gamma_m^k(n_k + 1) \phi^k(n_k + 1) \mu_m^k \gamma_l^j(n_j)$ ,  
(iii')  $q'(T^{im}C, C) = \gamma_m^k(n_k + 1) \phi^k(n_k + 1)$ ,  
(iv')  $q'(T^{jlx}C, C) = \gamma_l^j(n_j) \phi^j(n_j) \mu_l^j$

Pro třídu zakázek  $\varsigma = (i, s)$  označme

$$\alpha_j(\varsigma) = \begin{cases} \lambda_i d^j(\varsigma) & \text{když } r(\varsigma) = j, \\ 0 & \text{jindy.} \end{cases}$$

Potom

$$a_j = \sum_{i=1}^I \sum_{s=1}^{S(i)} \alpha_j(\varsigma) z^j(\varsigma),$$

je střední požadavek na míru obsluhy ve stanici  $j$  za jednotku času, pokud je systém v rovnováze. Postačující podmínkou rovnováhy je splnění nerovnosti (10.4.1). Podobným způsobem jako v odstavci 10.4 lze dokázat, že i pro tento model platí Tvzení 1,2 a jejich Důsledky 1 a 2.

Představme si nyní, že činnost stanice  $j$  budeme dělit na stále větší počet stadií  $z_n^j(\varsigma)$  s velmi malými požadavky  $d_n^j(\varsigma)$  tak, že  $z_n^j(\varsigma)d_n^j(\varsigma) = y$ ,  $n = 1, 2, \dots$ . Tímto způsobem lze přiblížit rozdělení požadavku na obsluhu zakázky třídy  $\varsigma$  konstantě  $y$  s libovolně malým rozptylem. Současným zvyšováním počtu fází obsluhy zakázky typu  $i$  ve stanici  $j$  můžeme nyní modelovat sítě, v nichž požadavky na obsluhu zakázek ve stanicích mají rozložení, jež lze vyjádřit jako konečnou směs rozložení gamma. Přitom zůstávají v platnosti všechna tvrzení z odstavce 10.4.

## 11.1 Limitní přechod k obecnému rozdělení požadavku na obsluhu

Při důkazech a odvozování vlastností systémů hromadné obsluhy, mezi něž patří i sítě obslužných stanic, se často používá metoda, kterou nazýváme *metodou vnořené markovské řetězce*. V obecnějších případech, kdy  $\{X(t), t \in \mathcal{T}\}$  netvoří markovský proces se spojitým časem, zúžíme náš pohled na systém tak, že vybereme posloupnost časových okamžiků  $\{t_0, t_1, \dots, t_n, \dots\}$ ,  $\lim_{n \rightarrow \infty} t_n = \infty$  takovou, že posloupnost  $\{X(t_0), \dots, X(t_n), \dots\}$  tvoří markovský proces. To znamená, že pro libovolnou  $m$ -prvkovou podposloupnost  $\tau_1 < \tau_2 < \dots < \tau_m$  posloupnosti  $\{t_n\}$  platí

$$\begin{aligned} P(X(\tau_m) = C_m \mid X(\tau_1) = C_1, X(\tau_2) = C_2, \dots, X(\tau_{m-1}) = C_{m-1}) = \\ = P(X(\tau_m) = C_m \mid X(\tau_{m-1}) = C_{m-1}). \end{aligned} \quad (11.1)$$

Systém potom budeme vyšetřovat pouze v okamžicích  $\{t_0, t_1, \dots, t_n, \dots\}$  a získané výsledky přeneseme na původní proces.

V našem modelu budeme sledovat pouze stavy v okamžicích  $\tau_n$ ,  $n = 0, 1, 2, \dots$  pohybu zakázek mezi jednotlivými stanicemi. *Pohybem zakázky* budeme rozumět buď příchod nové zakázky do sítě, přechod zakázky mezi stanicemi v síti nebo odchod zakázky ze sítě. Ukážeme si, že znalost posloupnosti  $\{(\tau_n, X(\tau_n))\}_{n \geq 0}$  a počátečního stavu  $x = X(0)$  nám umožňuje rekonstrukci celého procesu  $\{X(t), t \geq 0\}$ . Stavy procesu  $X(t)$  v čase

### 11.1. LIMITNÍ PŘECHOD K OBECNĚMU ROZDĚLENÍ POŽADAVKU NA OBSLUHU

$\tau_n \leq t < \tau_{n+1}$  mezi dvěma pohyby zakázek v síti jsou určeny předpokladem (II) a (III) modelu. Předpokládejme, že celkový stav sítě v čase  $t$  je

$$X(t) = \mathbf{C}(t) = (\mathbf{c}^1(t), \mathbf{c}^2(t), \dots, \mathbf{c}^J(t)),$$

kde pro  $j = 1, 2, \dots, J$  je

$$\mathbf{c}^j(t) = ((\varsigma_1^j, x_1^j), (\varsigma_2^j, x_2^j), \dots, (\varsigma_{n_j}^j, x_{n_j}^j),) = (\boldsymbol{\varsigma}^j, \mathbf{x}^j).$$

Známe-li stav  $\mathbf{c}^j(t)$ ,  $\tau_n \leq t < \tau_{n+1}$ , potom pro  $0 \leq h < (\tau_{n+1} - t)$  bude

$$\mathbf{c}^j(t+h) = ((\varsigma_1^j, x_1^j + \gamma_1^j(n_j)\phi^j(n_j)h + o(h)), \dots, (\varsigma_{n_j}^j, x_{n_j}^j + \gamma_{n_j}^j(n_j)\phi^j(n_j)h + o(h))).$$

Tedy známe-li všechny okamžiky pohybu zakázek v síti a stavy sítě v těchto okamžicích, potom můžeme určit stav sítě v libovolném čase.

Nechť posloupnost  $\{\tau_n\}_{n \geq 0}$  časů pohybu zakázek v síti je úplná, tzn. že mezi dvěma časy  $\tau_n$  a  $\tau_{n+1}$  nedojde k pohybu žádné zakázky v síti pro všechna  $n = 0, 1, 2, \dots$ . Této posloupnosti odpovídá

- posloupnost intervalů  $I_n = [\tau_n, \tau_{n+1})$ ,
- posloupnost stavů  $\{\boldsymbol{\varsigma}_n\}_{n \geq 0}$ ,
- posloupnost přirozených čísel  $\{r_n\}_{n \geq 0}$ , kde  $r_n$  je rovno počtu zakázek v síti v čase  $\tau_n$ , který se po celý časový interval  $I_n$  (tedy až do času  $\tau_{n+1}$ ) nemění,
- posloupnost vektorů  $\{\boldsymbol{\omega}_n^1, \dots, \boldsymbol{\omega}_n^J\}_{n \geq 0}$ , kde  $\boldsymbol{\omega}_n^j = (\mathbf{q}(\varsigma_1^j, n), \mathbf{q}(\varsigma_2^j, n), \dots, \mathbf{q}(\varsigma_{n_j}^j, n))$ .  
Pro  $\varsigma_m^j = (i, s)$  je  $\mathbf{q}((i, s), n) = (q_{i,s}(n), \dots, q_{i,S(i)}(n))$  vektor momentálních požadavků na obsluhu ve zbývajících fázích  $s, s+1, \dots, S(i)$  (množství práce, času, energie,  $\dots$ , které ještě požaduje zakázka třídy  $\varsigma_m^j$  na pozici  $(j, m)$  v čase  $\tau_n$ .

Uvažujme nyní ještě obecnější model než v odstavci 2.5. Předpoklady (II) – (IV), (V') popisující model budou stejné jako v 2.5, předpoklad (I) změníme následovně:

(I\*) – Zakázka třídy  $\varsigma$  na  $j$ -tém stanovišti vyžaduje obsluhu, jejíž velikost je náhodná veličina  $Q(\varsigma)$ . Vektor požadavků  $\{Q(i, s)\}_{s=1}^{S(i)}$  zakázky typu  $i$  má sdružené rozložení  $F_i$ , nezávislé na ostatních stavech sítě.

Ostatní předpoklady nechť zůstanou beze změny. V tomto případě bude střední požadavek na míru obsluhy ve stanici  $j$  za jednotku času, pokud bude systém v rovnováze, určen vztahem

$$a_j = \sum_{i=1}^I \sum_{s=1}^{S(i)} \alpha_j(i, s) E(Q(i, s)).$$

Postačující podmínkou rovnováhy je potom opět splnění nerovnosti (2.4.1) pro všechna  $j$ . Označme  $b = \prod_{j=1}^J b_j^{-1}$ , kde  $b_j$  je definováno ve vztahu (2.4.1).

Rozšíříme nyní pojem *stav sítě*. Stav sítě v čase  $t$  vyjádříme jako

$$\mathbf{G} = (\vartheta, (\boldsymbol{\varsigma}^1, \boldsymbol{\omega}^1), \dots, (\boldsymbol{\varsigma}^J, \boldsymbol{\omega}^J)),$$

kde

- $\boldsymbol{\vartheta} = (\vartheta_1, \dots, \vartheta_I)$  jsou doby do příchodu nové zakázky typu  $i = 1, \dots, I$ . Veličiny  $\vartheta(i)$  jsou nezávislé a každá z nich má exponenciální rozložení s parametrem  $\lambda(i)$ .
- spolu s každou zakázkou v síti na místě  $(m, j)$  budeme uvažovat kromě odpovídající třídy  $\zeta_m^j = (t_m^j, s_m^j)$  uvažovat i vektor jejich zbývajících požadavků na obsluhu  $(q_{t_m^j, 1}, \dots, q_{t_m^j, S(t_m^j)})$  (momentální zbývajících požadavky ve všech fázích obsluhy  $s = 1, 2, \dots, S(t_m^j)$ ). Podle předpokladu ( $I^*$ ) se tyto hodnoty řídí sdruženým rozložením  $F_i(q_1, \dots, q_{S(i)})$ ,  $i = t_m^j$ .

Pro  $\mathbf{c}^j = (\boldsymbol{\varsigma}^j, \boldsymbol{\omega}^j)$  potom definujeme

$$\mu_j(\mathbf{c}^j) = \frac{1}{b_j} \prod_{m=1}^{n_j} \frac{\lambda_i}{\phi^j(m)} \int_{\mathbb{R}_+^{S(i)}} (t_s - q_{i,s}) \prod_{r=s}^{S(i)} I[t_r \geq q_{i,r}] dF_i(t_1, \dots, t_{S(i)}),$$

kde  $i = t_m^j$ . A. D. Barbour v práci [Barbour, 1976] dokázal následující tvrzení.

**Tvrzení 3** *Mějme síť, splňující předpoklady ( $I^*$ ), (II) – (IV), ( $V'$ ) a podmínku (2.4.1). Předpokládejme, že s pravděpodobností 1 jsou všechny požadavky  $Q(\zeta)$  jsou kladné. Potom rovnovážné rozložení v síti má tvar*

$$\mu(\mathbf{G}) = \left( \prod_{l=1}^I e^{-\lambda_l \vartheta_l} \right) \left( \prod_{j=1}^J \mu(\mathbf{c}^j) \right).$$

### 11.1. LIMITNÍ PŘECHOD K OBECNĚMU ROZDĚLENÍ POŽADAVKU NA OBSLUHU

**Důsledek 3** Je-li síť v rovnovážném stavu, potom

- (i) Zakázky typu  $i$  opouštějí soustavu v nezávislých poissonovských prouděch s intenzitami  $\lambda_i$ ,  $i = 1, 2, \dots, I$ .
- (ii) Stav  $\mathbf{c}^j$  systému ve stanici  $j$  je nezávislý na stavech v ostatních stanicích a má pravděpodobnost  $\mu_j(\mathbf{c}^j)$ . Pravděpodobnost, že délky front ve stanicích budou rovny  $\mathbf{n} = (n_1, \dots, n_J)$  je

$$P(\mathbf{n}) = \prod_{j=1}^J \left[ a_j^{n_j} / \prod_{m=1}^{n_j} \phi^j(m) \right].$$

**Příklad 11.2.1.** V kapitole 4 jsme se zabývali stanicí s minimální obsluhou. Uvažujme nyní model sítě složený ze tří stanic tohoto typu. Předpoklady modelu jsou následující:

- síť sestává ze tří obslužných stanic ( $J=3$ ) s intenzitami obsluhy  $\phi_1 = \nu_1$ ,  $\phi_2 = \nu_2$  a  $\phi_3 = \nu_3$ ,
- v síti je obsluhován pouze jeden druh zakázek ( $I=1$ ), které do sítě přicházejí s intenzitou  $\lambda$  a každou stanicí procházejí jednou v pořadí 1, 2, 3,
- požadavky na obsluhu v jednotlivých stanicích obsluhy jsou nezávislá s posunutým exponenciálním rozložením s parametry  $\tau_1, \mu_1, \tau_2, \mu_2$  a  $\tau_3, \mu_3$ .

Z třetího předpokladu plyne, že sdružená hustota požadavků na obsluhu zakázky v síti je

$$f(x, y, z) = \mu_1 \mu_2 \mu_3 e^{-\mu_1(x-\tau_1)} e^{-\mu_2(y-\tau_2)} e^{-\mu_3(z-\tau_3)}, \quad x \leq \tau_1, y \leq \tau_2, z \leq \tau_3.$$

Střední hodnota tohoto rozložení je rovna  $E(Q) = (\tau_1 + \frac{1}{\mu_1}, \tau_2 + \frac{1}{\mu_2}, \tau_3 + \frac{1}{\mu_3}) = (\alpha_1, \alpha_2, \alpha_3)$ . Ke splnění nerovnosti (2.4.1) stačí, aby

$$\frac{\lambda \alpha_1}{\nu_1} < 1 \quad \frac{\lambda \alpha_2}{\nu_2} < 1 \quad \frac{\lambda \alpha_3}{\nu_3} < 1.$$

Z téhož vztahu potom dostaneme normalizační konstantu

$$b = \frac{(\nu_1 - \lambda \alpha_1)(\nu_2 - \lambda \alpha_2)(\nu_3 - \lambda \alpha_3)}{\nu_1 \nu_2 \nu_3}.$$

Uvažujme nyní případ, kdy na stanici 1 je celkem  $k$  zakázek, na stanici 2  $l$  a na stanici 3  $m$  zakázek. Pravděpodobnost tohoto jevu je podle důsledku 3 rovna

$$P(k, l, m) = \left( \frac{\lambda\alpha_1}{\nu_1} \right)^k \left( \frac{\lambda\alpha_2}{\nu_2} \right)^l \left( \frac{\lambda\alpha_3}{\nu_3} \right)^m.$$

Představme si stav sítě  $\mathbf{G} = (w, (q, u, v), (0, u, v), (0, 0, v))$ . Stav  $\mathbf{G}$  lze interpretovat takto:

- do příchodu další zakázky zbývá čas  $w$ .
- zakázkám v síti zbývá ještě množství  $q$  obsluhy ve stanici 1, množství  $u$  obsluhy ve stanici 2 a  $v$  ve stanici 3.
- zakázky ve stanici 2 už nebudou požadovat obsluhu stanicí 1 a podobně zakázky ve stanici 3 už nebudou obsluhovány ve stanicích 1 a 2.

Předpokládejme, že např.  $q \geq \tau_1$ ,  $u \geq \tau_2$ ,  $v \geq \tau_3$ . Označme  $q' = q - \tau_1$ ,  $u' = u - \tau_2$  a  $v' = v - \tau_3$ . Dosazením těchto hodnot do výrazu v tvrzení 3 dostaneme pravděpodobnost

$$\mu(\mathbf{G}) = be^{-\lambda w} \frac{\lambda^{k+l+m}}{(\nu_1\mu_1)^k(\nu_2\mu_2)^l(\nu_3\mu_3)^m} \exp(-\mu_1 q' k - \mu_2 u' (k+l) - \mu_3 v' (k+l+m)).$$

## 11.2 Síť stanic s kladnou minimální dobou obsluhy

Předpokládejme, že síť je tvořena stanicemi popsanými v kapitole 4. Požadavek na obsluhu zde má posunuté exponenciální rozložení s parametry  $\tau, \mu$ . Tento požadavek lze interpretovat jako dvoufázovou obsluhu, přičemž první fáze vyžaduje pevné (nenáhodné) množství  $\tau$  obsluhy a ve druhé fázi má obsluha exponenciální rozložení s parametrem  $\mu$ . Pro náš model to znamená konkretizaci předpokladu ( $I^*$ ):

( $I^*$ ) – Zakázka typu  $i$  vyžaduje v  $s$ -té fázi obsluhu, jejíž velikost je náhodná veličina  $Q(i, s)$  s distribuční funkcí

$$F(i, s; x) = \begin{cases} 0, & x < \tau(i, s), \\ 1 - \exp(-\mu(i, s)(x - \tau(i, s))), & x \geq \tau(i, s), \end{cases}$$



nezávislá na ostatních stavech sítě.

Protože je  $E(Q(i, s)) = \tau(i, s) + \frac{1}{\mu(i, s)}$ , bude střední požadavek na obsluhu ve stanici  $j$  dán vztahem

$$a_j = \sum_{i=1}^I \sum_{s=1}^{S(i)} \alpha_j(i, s) \left[ \tau(i, s) + \frac{1}{\mu(i, s)} \right],$$

pomocí něhož vyjádříme podmínku rovnováhy (2.4.1) a hodnoty normujících konstant  $b_j$ ,  $j = 1, \dots, J$ .

Stav sítě v čase  $t$  vyjádříme vektorem

$$\mathbf{G} = (\boldsymbol{\vartheta}, \mathbf{c}^1, \dots, \mathbf{c}^J),$$

kde

$$\mathbf{c}^j = ((t_1^j, s_1^j, q_1^j), \dots, (t_{n_j}^j, s_{n_j}^j, q_{n_j}^j)),$$

přičemž  $q_m^j = q(t_m^j, s_m^j)$  je velikost zbývajících požadavků na obsluhu zakázky na místě  $m$  ve stanici  $j$  vzhledem k této stanici.

Podle tvrzení 3 v odstavci 2.6 má rovnovážné rozložení v síti tvar

$$\mu(\mathbf{G}) = \left( \prod_{l=1}^I e^{-\lambda_l \vartheta_l} \right) \left( \prod_{j=1}^J \mu(\mathbf{c}^j) \right),$$

přičemž

$$\mu_j(\mathbf{c}^j) = \frac{1}{b_j} \prod_{m=1}^{n_j} \frac{\lambda_{t_m^j}}{\phi^j(m)} \int_{q_m^j}^{\infty} (x - q_m^j) dF_m^j(x),$$

označíme-li  $F_m^j(x) = F(t_m^j, s_m^j, x)$ .

Pro  $x \geq \tau_m^j$  je navíc

$$dF_m^j(x) = \mu(t_m^j, s_m^j) \exp(-\mu(i, s)(x - \tau(i, s))) dt = \mu_m^j \exp(-\mu_m^j(x - \tau_m^j)) dt.$$

Tedy pro  $q_m^j \geq \tau_m^j$ ,  $m = 1, \dots, n_j$  dostaneme

$$\mu_j(\mathbf{c}^j) = \frac{1}{b_j} \prod_{m=1}^{n_j} \frac{\lambda_{t_m^j}}{\phi^j(m)} \int_{q_m^j}^{\infty} (x - q_m^j) dF_m^j(x) = \frac{1}{b_j} \prod_{m=1}^{n_j} \frac{\lambda_{t_m^j} e^{-\mu_m^j(q_m^j - \tau_m^j)}}{\phi^j(m) \mu_m^j}.$$



## Kapitola 12

# Petriho síť



## Kapitola 13

# Počítačové simulace

V některých případech nám vlastnosti modelu buď nedovolují výpočet některých charakteristik analytickým způsobem vůbec, nebo jen s velkými obtížemi. Potom zpravidla použijeme *numerickou simulaci* modelu. Výhody numerické simulace jsou nesporné a její význam prudce vzrůstá s rostoucími možnostmi a výkonností současné výpočetní techniky. Simulaci, kterou ještě před několika lety bylo možno provést pouze na velkém počítači a trvala několik hodin, lze v současné době uskutečnit na stolním počítači v mnohem kratším čase. Velkého pokroku je též dosahováno v možnostech grafické prezentace výsledků. Simulací lze modelovat sítě s různými pravděpodobnostními rozloženími dob příchodů zakázek, požadavků na obsluhu, dob do poruchy či délky oprav obslužných stanic. K simulaci lze použít i empiricky zjištěných rozložení, bez nutnosti aproximace teoretickým rozložení. Simulační modely nám umožňují modelovat i velmi složité sítě z hlediska plánů obsluhy jednotlivých zákazníků.

Proč se tedy s poměrně značným úsilím zabýváme analytickými rozbory, jako např. v kapitolách 5,6,7 ? Především numerická simulace není možná bez předchozího analytického rozboru. Má-li simulace probíhat efektivně, musí vycházet z analytických vztahů. Další důvody souvisí s obecným vztahem mezi numerickými a analytickými výpočty. Numerické výpočty a simulace mají – stejně jako empirické zjišťování – pouze „lokální“ charakter v tom smyslu, že jsou vždy vztaženy k momentální situaci či podmínkám, za kterých jsou uskutečněny. Naproti tomu analytické výpočty nám mohou poskytnout globální pohled na zkoumaný model. Například řešení soustavy diferenciálních rovnic. Numerická metoda nám sice poskytne řešení, ale

pouze jedno a v jediném bodě. Toto můžeme potom mnohokrát opakovat a získat tak záplavu čísel, kterou lze zobrazit i graficky a získat tak jakousi „globální“ představu o řešení dané soustavy. Výsledek je zpravidla názorný a postačující, nicméně je k němu třeba velmi mnoho výpočtů, náročného technického vybavení a čas. Naproti tomu analytické řešení – pokud je vůbec možné – bývá obvykle složitější a nebývá ani přesnější, neboť zvláště v nelineárních případech jsme nuceni provádět řadu aproximací. Nicméně analytický výsledek má obecnější charakter a umožňuje diskusi z hlediska parametrů. Dojde-li v modelu ke změně parametrů, analytický výsledek nám řekne, co se může stát s řešením. Numerické řešení se však musí celé opakovat. Z tohoto hlediska numerické řešení neumožňuje ani vyšetřování asymptotického (mezního) chování modelu. Numerický výpočet může být i velmi pomalý v případech, kdy je třeba vyhodnotit situaci pokud možno „ihned“. Dosazení okamžitých parametrů do jednoho vzorce, který je výsledkem analytického výpočtu provedeného dříve, bývá rychlejší, než provést výpočet či simulaci.

Výjimku v předchozí úvaze tvoří některé počítačové programy, které jsou schopny analytických výpočtů a symbolického vyjádření jejich výsledků (např. Maple, Mathematica), nicméně těchto prostředků se k simulaci zatím ve větší míře nepoužívá.

Z uvedeného vyplývá, že optimálních výsledků lze dosáhnout teprve při vhodném spojení obou metod. Proto je důležité odvozovat výsledky analyticky a pokud to model umožňuje, použít jich při numerických výpočtech.

## 13.1 Metody simulace

Metody simulace závisí do jisté míry na použitých technických a programových prostředcích. V zásadě však lze při numerické simulaci obslužných systémů postupovat dvěma způsoby:

- (a) **Simulace v diskrétním čase.** Je to tzv. „ $\Delta t$ “ metoda, která je založena na diskretizaci časové osy. Časový interval  $T$  rozdělíme na  $N$  dílů o délce  $\Delta = \frac{T}{N}$ . Model budeme vyšetřovat v časových okamžicích  $0, \Delta, 2\Delta, 3\Delta, \dots$ . V každém tomto okamžiku zjísíme, zda došlo k nějaké změně stavu – *události* – či nikoli. Pokud ano, provedeme příslušné výpočty a změny.

- (b) **Simulace diskretních událostí.** Tento přístup předpokládá, že jsme schopni postupně generovat úplný seznam okamžiků změn stavů modelu  $\{\tau_1, \tau_2, \dots, \tau_n, \dots\}$ . Simulace potom probíhá sekvenčně tak, že v  $i$ -tém kroku je simulována událost, příslušející k času  $\tau_i$ .

Každá z uvedených metod má své výhody i nevýhody. Jejich použití zpravidla závisí na konkrétním modelu. Obecně lze říci, že první metoda bývá jednodušší z hlediska algoritmu, ale na úkor rychlosti a přesnosti. Druhá metoda je – z hlediska pravděpodobnostních charakteristik – přesnější, nicméně její algoritmy mohou být komplikovanější a náročnější na počet použitých parametrů. Nejčastěji se však používá kombinace obou uvedených metod.

Oba přístupy jsou založeny na simulační metodě které se říká *metoda Monte Carlo*. Při ní numerickými prostředky generujeme číselné hodnoty, odpovídající realizacím náhodných veličin. Těmto číslům říkáme *pseudonáhodná čísla* a algoritmům pro jejich vytváření nazýváme *generátory (pseudo)náhodných čísel*. Kvalita simulace – to jest míra podobnosti mezi teoretickým modelem a výsledkem simulace – závisí především na kvalitě použitého generátoru. Problematice generování pseudonáhodných čísel a posuzování kvality generátorů se věnuje např. [Ripley, 19??] Běžné programovací jazyky poskytují generátory pseudonáhodných čísel z rovnoměrného rozložení na intervalu  $(0, 1)$ , případně z normálního rozložení  $N(0, 1)$ . Srovnání generátorů různých programovacích jazyků a statistických systémů lze nalézt v [Antoch,1990].

## 13.2 Simulace modelu obslužné sítě

V dalším odstavci se budeme věnovat simulaci otevřené sítě, popisované v předchozí kapitole. Použijeme k tomu metodu diskretních událostí. Konkrétní podoba programu pro simulaci bude záviset na programových prostředcích, které máme k dispozici. Proto se zde nebudeme zabývat detaily jednotlivých procedur, ale pouze algoritmizací úlohy. Konkrétní kód v jazyce systému PC–MATLAB, verze 3.1 je uveden v příloze této práce.

Simulační model se řídí řadou parametrů. Jejich stanovení se provádí obvykle empiricky, formou pozorování a měření reálného systému. Parametry také můžeme volit experimentálně, případně mohou být předmětem optimalizace. Ke konstrukci modelu je však musíme znát. Především jsou

to tyto **parametry modelu**:

- počet typů zakázek  $I$  a počet obslužných stanic  $J$ .
- pro každý typ zakázky musíme znát rozložení doby mezi příchody do systému  $L_i$ , plán obsluhy  $\vec{r}_i = (r(i, 1), r(i, 2), \dots, r(i, S(i)))$  a rozložení požadavků na zpracování  $F_{is}$ ,  $s = 1, \dots, S(i)$ ,  $i = 1, \dots, I$ .
- v každé obslužné stanici  $j$  musí být definován frontový režim prostřednictvím rozložení pravděpodobnosti  $\Delta_j$  a  $\Gamma_j$ , kde  $\Delta_j = \{\delta_{jn}\}_{n=0}^{\infty}$  jsou pravděpodobnosti zařazení nové zakázky na místo  $n$  ve frontě a  $\Gamma_j = \{\gamma_{jn}\}_{n=0}^{\infty}$  jsou pravděpodobnosti zařazení zakázky na místě  $n$  ke zpracování.

Obvykle stanovíme také maximální počet  $K$  zakázek, které mohou být najednou v síti a kritérium pro ukončení simulačního procesu.

K simulaci obslužné sítě budeme dále potřebovat generátory pseudonáhodných čísel ze zadaných rozložení s distribučními funkcemi  $L_i$ ,  $F_{is}$  a z rozložení  $\Delta$  a  $\Gamma$ . Jejich konstrukce s využitím generátoru pseudonáhodných čísel z rovnoměrného rozložení na intervalu  $(0, 1)$  je popsána v příloze.

V průběhu simulace budeme používat řadu dalších proměnných, z nichž některé jsou nutné pro běh programu, další zavádíme účelově pro výpočet sledovaných charakteristik. **Nutné proměnné** lze zavést například takto:

- posloupnost událostí v síti  $\mathcal{U} = \{U_n\}_{n=0}^K$ . Událostí zde budeme rozumět trojici  $U_n = (\tau_n, j_n, m_n)$ , kde  $\tau_n$  je čas kdy událost nastane,  $j_n$  je číslo stanice a  $m_n$  je místo ve frontě, kde událost nastane.
- stav  $C_j$  ve stanici  $j$  bude vektor  $((i_1, s_1), \dots, (i_K, s_K))$ , kde  $i_m$  je typ a  $s_m$  je fáze zpracování zakázky na místě  $m$  ve frontě. Pokud na místě  $m$  nebude žádná zakázka, budou obě hodnoty nulové.

V uvedeném příkladě budeme simulací zjišťovat střední dobu průchodu zakázky sítí a střední počet zakázek v každé stanici při dlouhodobém provozu. Pro tyto výpočty zavedeme další **účelové proměnné**:

- $T_1, \dots, T_I$ , do kterých budeme postupně načítat časy, strávené zakázkami odpovídajícího typu v síti,  $N_1, \dots, N_I$  budou sloužit jako čítače zakázek odpovídajícího typu a  $M_1, \dots, M_J$ , což budou množství jakýchsi „zakázkohodin“ v každé stanici.



- časové stavy  $D_j$  pro každou stanici  $j$ ,  $D_j = ((t_1, x_1), \dots, (t_K, x_K))$ , kde  $t_m$  je okamžik vstupu zakázky na místě  $m$  do sítě a  $x_m$  je okamžik vstupu této zakázky do stanice  $j$ .
- $\bar{T}_1, \dots, \bar{T}_I$  – průměrné doby strávené zakázkami v síti. Spočteme je podle vztahu  $\bar{T}_i = \frac{T_i}{N_i}$ ,  $i = 1, \dots, I$ .  $\bar{M}_1, \dots, \bar{M}_J$  – průměrné počty zakázek ve stanicích za jednotku času, spočtené podle vztahu  $\bar{M}_j = \frac{M_j}{\tau}$ , kde  $\tau$  je čas poslední události v síti.

Podobným způsobem lze doplnit proměnné pro výpočet dalších charakteristik, které nás mohou v souvislosti se sítí zajímat.

**Vlastní simulace** potom bude probíhat v následujících základních krocích:

- 0) **Nastavení počátečních hodnot.** Pro všechny třídy  $i = 1, \dots, I$  vygenerujeme posloupnost prvních událostí  $\mathcal{U} = \{U_i\}_{i=1}^I = \{(\tau_i, j_i, m_i)\}_{i=1}^I$  (první příchody do systému). Číslo  $j_i$  položíme zatím rovno nule a hodnotu  $m$  položíme rovnu  $i$ . Generovat budeme pouze čas  $\tau_i$  pomocí rozložení  $L_i$ . Události  $\mathcal{U}$  uspořádáme vzestupně podle času  $\tau_i$ . Tímto krokem jsou pouze vytvořeny okamžiky budoucích vstupů zakázek do systému. Vlastní zařazení zakázek bude probíhat v kroku 1) postupně, podle pořadí. Na začátku simulace také naplníme všechny stavy  $C_{jm}$  nulami,  $j = 1, \dots, J$ ,  $m = 1, \dots, K$ .
- 1) **Vyhodnocení aktuální události.** Z množiny  $\mathcal{U}$  vybereme událost  $U = (\tau, j, m)$ , která má minimální hodnotu  $\tau$  (která nejdříve nastane) a tuto událost z  $\mathcal{U}$  odstraníme. Přitom mohou nastat následující případy:
  - a)  $j = 0$ . To znamená, že se jedná o **vstup nové zakázky typu  $i = m$  do sítě** (její zařazení do sítě). Položíme  $j = r(i, 1)$ . Pokud ve stanici  $j$  není žádná zakázka ( $C_{j1} = (0, 0)$ ), položíme  $C_{j1} = (i, 1)$  a ihned vygenerujeme čas  $\tau$  budoucího dokončení zakázky ve stanici  $j$ . Událost  $U = (\tau, j, 1)$  doplníme do množiny  $\mathcal{U}$ . Jsou-li ve stanici  $j$  zakázky, vygenerujeme hodnotu  $m$  podle rozložení  $\Delta_j$ . Pokud je ve stavu  $C_j$  složka  $C_{jm}$  nenulová, provedeme „posun“ zakázek o jednu pozici zpět, tzn.  $C_{jm} \rightarrow C_{j(m+1), \dots}$  a  $C_{jm}$  naplníme hodnotou  $(i, 1)$ . V tomto případě žádný čas pro stanici  $j$  negenerujeme. V obou případech však nakonec vygenerujeme příští čas  $\pi$  vstupu další zakázky typu  $i$  do sítě a do množiny  $\mathcal{U}$  doplníme událost  $(\pi, 0, i)$ .

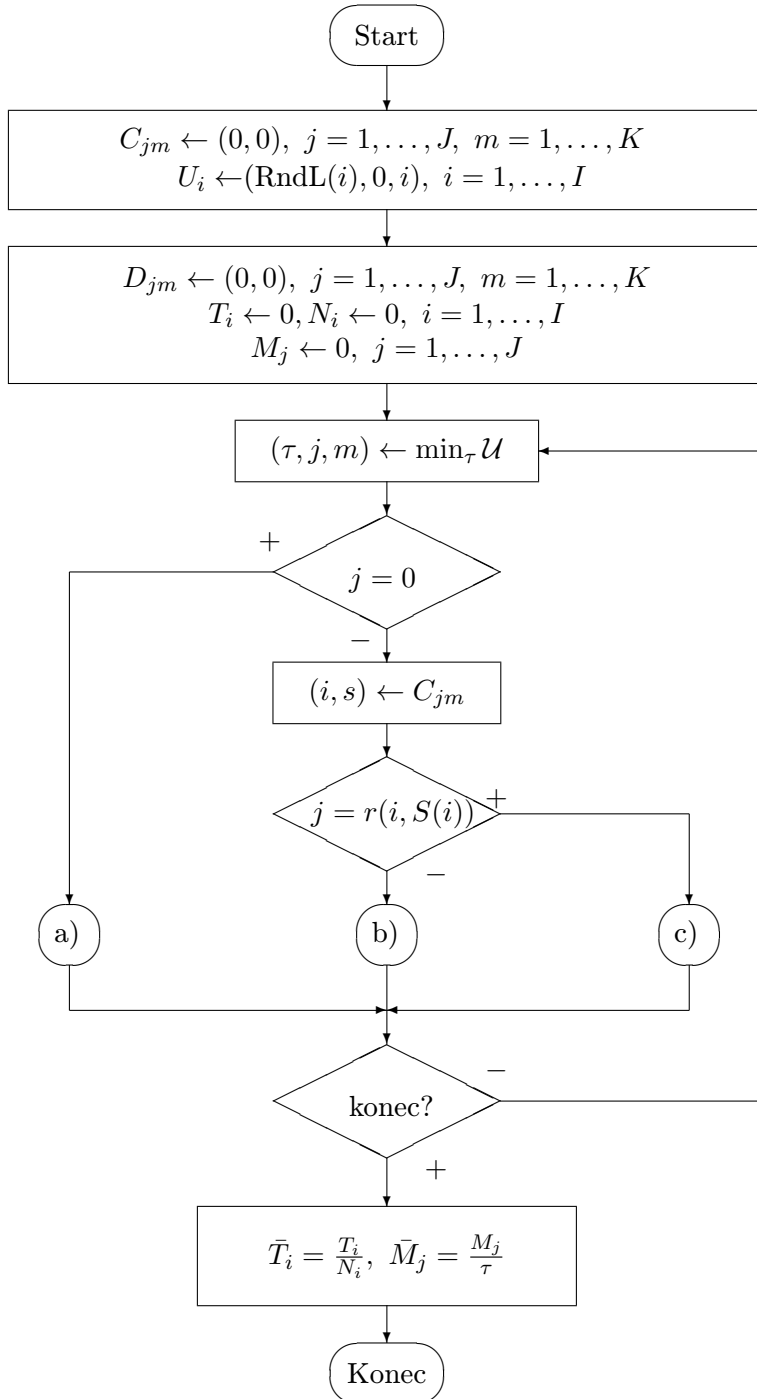
- b)  $1 \leq j < r(i, S(i))$ , kde typ  $i$  zakázky zjistíme z hodnoty  $C_{jm}$  spolu s fází zpracování  $s$ . Potom se jedná o **přechod dosud nedokončené zakázky mezi stanicemi v síti**. Přechod bude uskutečněn ze stanice  $j$  do stanice  $k = r(i, s + 1)$ . Provedeme přesun zakázek  $C_{j(m+1)} \rightarrow C_{jm}, \dots$ . Pokud jsou ve stanici  $j$  ještě nějaké zakázky (nenulové  $C_{j1}$ ), vygenerujeme pomocí rozložení  $\Gamma_j$  pozici  $m$  a čas dokončení  $\tau$  další zakázky ve stanici  $j$ . Množinu událostí  $\mathcal{U}$  doplníme trojicí  $(\tau, j, m)$ . Dále provedeme zařazení zakázky do stanice  $k$  podobně jako v případě a). Pokud ve stanici  $k$  není žádná zakázka ( $C_{k1} = 0$ ), položíme  $m = 1$  a ihned vygenerujeme čas  $\tau$  budoucího dokončení zakázky ve stanici  $k$ . Událost  $U = (\tau, k, 1)$  doplníme do množiny  $\mathcal{U}$ . Jsou-li ve stanici  $k$  zakázky, vygenerujeme hodnotu  $m$  podle rozložení  $\Delta_k$ . Pokud je ve stavu  $C_k$  složka  $C_{km}$  nenulová, provedeme „posun“ zakázek o jednu pozici zpět, tzn.  $C_{km} \rightarrow C_{k(m+1)}, \dots$  a  $C_{km}$  naplníme hodnotou  $(i, 1)$ . V tomto případě žádný čas pro stanici  $k$  negenerujeme.
- c)  $j = r(i, S(i))$ , kde typ  $i$  zakázky zjistíme z hodnoty  $C_{jm}$ . V takovém případě se jedná o **ukončení zpracování zakázky a její odchod ze sítě**. Provedeme přesun zakázek  $C_{j(m+1)} \rightarrow C_{jm}, \dots$  a vygenerujeme pozici  $m$  a čas dokončení  $\tau$  další zakázky ve stanici  $j$ . Množinu událostí doplníme trojicí  $(\tau, j, m)$ .
- 2) Je-li splněno kritérium pro ukončení simulace, ukončíme ji. Jinak pokračujeme v simulaci opět krokem 1).

Uvedené kroky jsou pouze základní, simulující vlastní proces obsluhy. Budeme-li chtít počítat zároveň průměrnou dobu strávenou zakázkou v síti a průměrný počet zakázek ve stanici za jednotku času, doplníme tyto kroky o příslušné operace (odpovídající skutečnému měření v reálném případě). Především budeme zaznamenávat okamžiky vstupu zakázky do sítě, abychom mohli při jejím výstupu vyhodnotit celkovou dobu strávenou v síti a při každém přechodu zaznamenáme čas vstupu zakázky do stanice a změříme dobu, kterou zakázka ve stanici strávila. Výše uvedené kroky doplníme o následující akce:

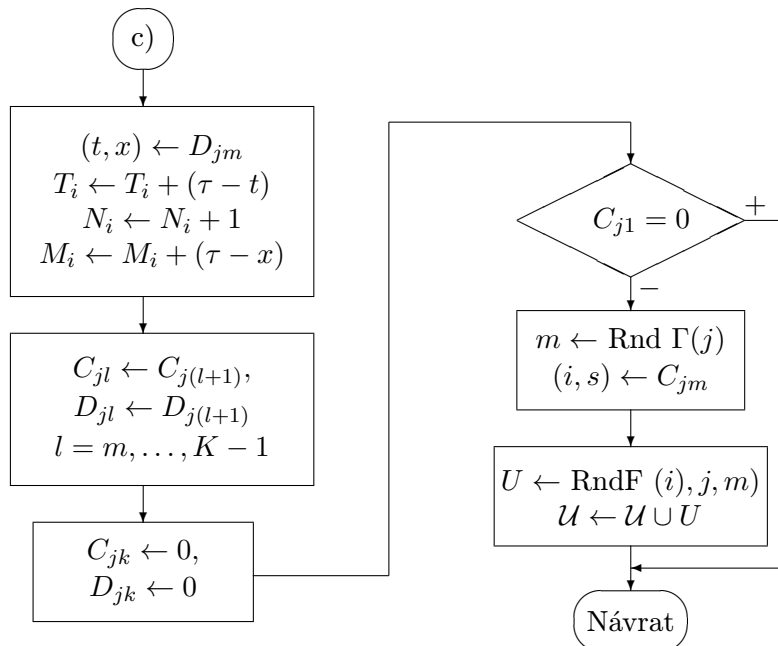
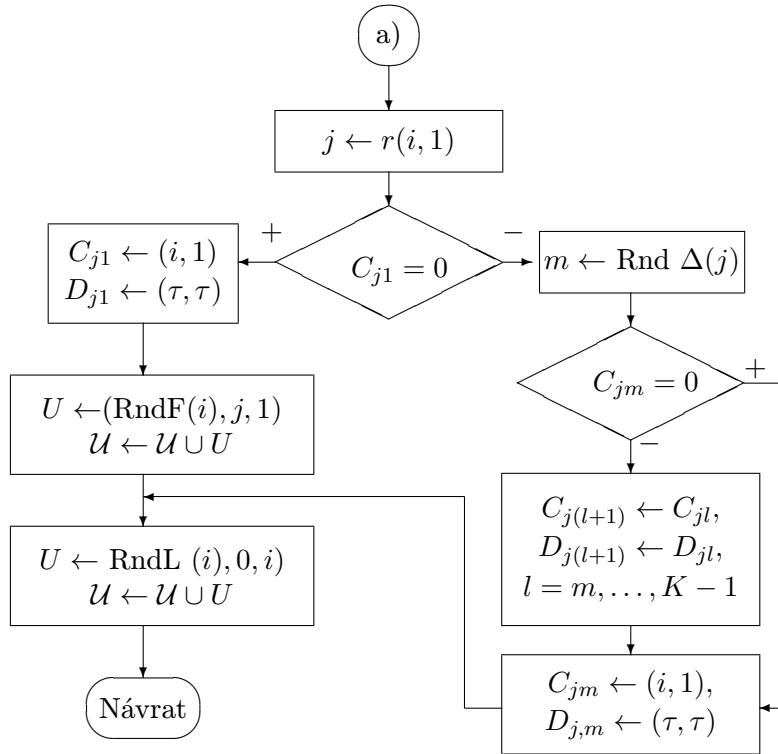
- 0) Hodnoty  $T_1, \dots, T_I, N_1, \dots, N_I, M_1, \dots, M_J$  položíme rovny nule. Nulami naplníme také hodnoty  $D_{jm}$ ,  $j = 1, \dots, J$ ,  $m = 1, \dots, K$ .
- 1) Při vyhodnocení události v čase  $\tau$  budeme zaznamenávat časové informace a aktualizovat čítače.

- a) Stav  $D_{j1}$  resp.  $D_{jm}$  naplníme hodnotou  $(\tau, \tau)$  (uložíme čas vstupu zakázky do sítě, který je zároveň okamžikem vstupu zakázky do první stanice).
- b) Při přesunu zakázky ze stanice  $j$  do stanice  $k$  nejprve zjistíme časy  $(t, x)$  ze stavu  $D_{jm}$  před přesunem. Hodnotu  $M_j$  zvětšíme o dobu  $(\tau - x)$ , po kterou zakázka setrvala ve stanici  $j$  a naplníme stav  $D_{km} = (t, \tau)$  (stav po přesunu, v němž uchováváme informaci  $t$  o čase vstupu zakázky do sítě). V tomto kroku provedeme přesuny hodnot  $D_{jm}$  a  $D_{km}$  analogicky jako s  $C_{jm}$  a  $C_{km}$ .
- c) Při výstupu zakázky ze sítě spočteme rozdíly  $(\tau - t)$ ,  $(\tau - x)$  a přičteme je k hodnotám  $T_i$  a  $M_j$ . Zároveň zvětšíme čítač zakázek prošlých sítí  $N_i$  o 1 a provedeme přesun hodnot  $D_{j(m+1)} \rightarrow D_{jm, \dots}$ .
- 2) Má-li být konec simulace, vypočteme průměrnou dobu strávenou zakázkou typu  $i$  v síti podle vztahu  $\bar{T}_i = \frac{T_i}{N_i}$  a průměrný počet zakázek  $M_j$  ve stanici  $j$  za jednotku času jako  $\bar{M}_j = \frac{M_j}{\tau}$ , kde  $\tau$  je čas poslední události v síti.

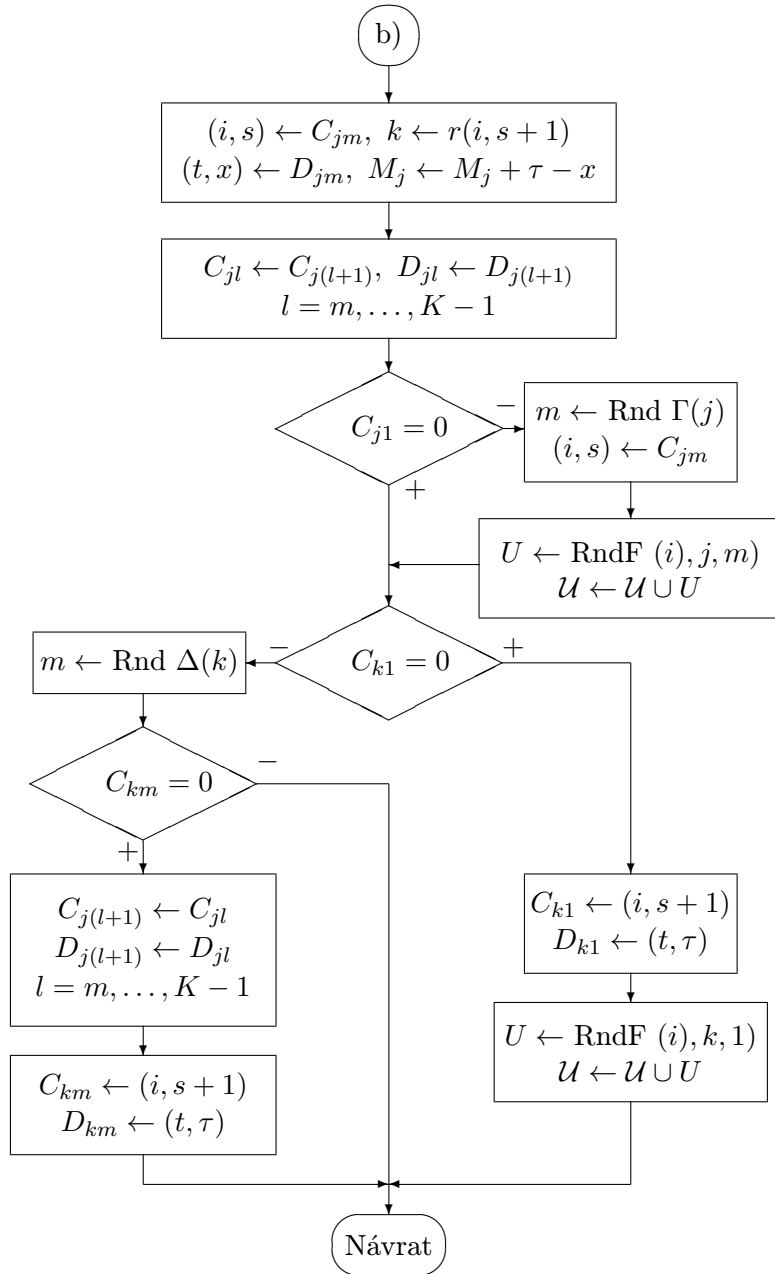
Na následujících obrázcích jsou bloková schemata programu pro simulaci uvedeného modelu. Obr. 13.1a zobrazuje hlavní schema, na obr. 13.1b a 13.1c jsou schemata pro vyhodnocení jednotlivých událostí, tak jak jsou uvedeny v 1a), 1b) a 1c).



Obr. 13.1a. Blokové schéma programu pro simulaci obslužné sítě.



Obr. 13.1b. Blokové schéma programu pro simulaci obslužné sítě, části a), c).



Obr. 13.1c. Blokové schéma programu pro simulaci obslužné sítě, část b).



Ve schemech užívám následující značení:

$\text{RndL}(i)$  – generátor pseudonáhodných čísel z rozdělení  $L(i)$

$\text{RndF}(i)$  – generátor pseudonáhodných čísel z rozdělení  $F(i)$

$\text{Rnd}\Gamma(j)$  – generátor pseudonáhodných čísel z rozdělení  $\Gamma(j)$

$\text{Rnd}\Delta(j)$  – generátor pseudonáhodných čísel z rozdělení  $\Delta(j)$

$\min_{\tau} \mathcal{U}$  – nejbližší událost (událost s minimálním  $\tau$ )



# Kapitola 14

## Dodatky

### 14.1 Některá důležitá pravděpodobnostní rozložení

V tomto doplňku si připomeneme a doplníme některá pravděpodobnostní rozdělení, která se v teorii hromadné obsluhy používají. Některá z nich jsou známa již ze základního kurzu pravděpodobnosti, o některých se dosud nemluvalo.

#### 14.1.1 Rozložení počtu pokusů do prvního úspěchu

Nechť pravděpodobnost realizace jevu  $A$  v jednom náhodném pokusu je rovna  $p$ . Budeme-li provádět sérii nezávislých pokusů tak dlouho, dokud jev  $A$  nenastane, bude nás zajímat rozdělení počtu  $N$  těchto „neúspěšných“ pokusů.  $N$  je náhodnou veličinou, která má geometrické rozdělení pravděpodobnosti, pro něž platí:

$$P(N = k) = p(1 - p)^k$$

*Cvičení:* Spočítejte základní charakteristiky náhodné veličiny s geometrickým rozdělením.

*Cvičení:* Ukažte, že geometrické rozdělení „nemá paměť“, tzn. že  $P(N \geq k + m | N \geq m) = P(N \geq k)$ .

### 14.1.2 Rozložení doby čekání

Rozložení doby  $T$  do první realizace nějakého náhodného jevu  $A$  může být velmi různé. Takovým jevem může být například příjezd soupravy metra do stanice. Potom bude doba mezi příjezdy jednotlivých souprav mít nějakou „typickou“ hodnotu, kolem níž budou skutečné doby kolísat s nějakou odchylkou. Zde můžeme předpokládat například normální rozdělení doby mezi příjezdy. V některých případech, například automatických dopravníků, lze považovat dobu mezi „příjezdy“ dvou po sobě jdoucích vozíků za deterministickou, tj. konstantní, danou rychlostí dopravníku.

Jiná situace je v případě, že o pozorovaném systému nemáme žádnou apriorní informaci, tj. předpokládáme, že realizace jevu  $A$  bude výsledkem působení mnoha nezávislých vlivů a může nastat prakticky kdykoli. O výskytech jevu  $A$  máme pouze informace statistického charakteru, například dlouhodobý průměrný počet událostí za časovou jednotku. V takovém případě obvykle předpokládáme, že náhodná veličina má Weibullovo rozložení pravděpodobnosti  $W(\alpha, \beta)$ , kde parametry  $\alpha$  a  $\beta$  jsou kladná čísla a mají význam *parametru měřítka* a *parametru tvaru*. Hustota tohoto rozdělení má tvar

$$f(x) = \frac{\beta x^{\beta-1}}{\alpha^\beta} \exp \left[ - \left( \frac{x}{\alpha} \right)^\beta \right], \quad x > 0.$$

Distribuční funkce Weibullova rozdělení má tvar

$$F(x) = 1 - \exp \left[ - \left( \frac{x}{\alpha} \right)^\beta \right], \quad x > 0.$$

Modus Weibullova rozdělení roven nule pro  $\beta \leq 1$ ; pro  $\beta > 1$  je roven  $\alpha[1 - (1/\beta)]^{1/\beta}$ . Střední hodnotu a rozptyl lze vyjádřit pouze pomocí gamma-funkce:

$$E(X) = \alpha \Gamma[1 + (1/\beta)], \quad D(X) = \alpha^2 \{ \Gamma[1 + (2/\beta)] - \{ \Gamma[1 + (1/\beta)] \}^2 \}.$$

*Poznámka:* Gamma funkce je definována vztahem

$$\Gamma(\delta) = \int_0^\infty x^{\delta-1} \exp(-x) dx, \quad x > 0$$

a platí pro ni některé známé vztahy, jako např.  $\Gamma(x+1) = x\Gamma(x)$ ,  $x > 0$ ,  $\Gamma(1/2) = \sqrt{\pi}$ ,  $\Gamma(n) = (n-1)!$  pro přirozená  $n$ .

V základním kurzu teorie pravděpodobnosti se obvykle uvádí jako rozdělení doby čekání rozdělení exponenciální, které má pouze jeden parametr  $\lambda$ . Velmi snadno nahlédneme, že se jedná o speciální případ Weibullova rozdělení.

*Cvičení:* Jak je třeba volit parametry Weibullova rozdělení, abychom dostali rozdělení exponenciální s parametrem  $\lambda$ ?

### 14.1.3 Gamma rozložení

Dalším rozdělením, které se objevuje v souvislosti s dobou mezi událostmi nebo ještě spíše jako rozdělení doby životnosti je rozdělení gamma s hustotou

$$f(x) = \left[ \frac{1}{\Gamma(\beta)} \right] \left( \frac{x^{\beta-1}}{\alpha^\beta} \right) \exp\left(\frac{-x}{\alpha}\right), \quad x > 0.$$

Distribuční funkci lze vyjádřit „jednoduše“ jako

$$F(x) = \Gamma(x/\alpha, \beta), \quad x > 0,$$

modus je opět pro  $\beta \leq 1$  roven nule, pro  $\beta > 1$  je roven  $\alpha(\beta - 1)$ , střední hodnota  $E(X) = \beta\alpha$  a rozptyl  $D(X) = \beta\alpha^2$ .

Speciálním případem tohoto rozdělení je opět rozdělení exponenciální, ale také rozdělení chí-kvadrát, známé z matematické statistiky.

## 14.2 Rozložení součtů nezávislých náhodných veličin

V mnoha aplikacích je třeba pracovat se součtem dvou i více nezávislých náhodných veličin, jejichž rozdělení pravděpodobnosti je známé.

### 14.2.1 Diskrétní případ

Budeme se zabývat rozložením náhodné veličiny  $Z = X + Y$ , kde  $X$  a  $Y$  jsou vzájemně nezávislé diskrétní náhodné veličiny s rozdělením pravděpodobnosti daným posloupnostmi  $p_i$  a  $q_j$ . Náhodná veličina  $Z$  je zřejmě také

diskrétní a pro její rozložení pravděpodobnosti zřejmě platí

$$P(Z = z_k) = \sum_{i=1}^{\infty} P(X = x_i)P(Y = z_k - x_i) = \sum_{i,j:x_i+y_j=z_k} p_i q_j.$$

*Cvičení:* Dokažte, že součet dvou Poissonovsky rozdělených náhodných veličin má opět Poissonovo rozdělení pravděpodobnosti.

### 14.2.2 Spojitý případ

Nechť  $X$  a  $Y$  jsou vzájemně nezávislé náhodné veličiny s rozdělením pravděpodobnosti spojitěho typu s hustotami  $p(x)$  a  $q(y)$ . Potom hustota spojitě náhodné veličiny  $Z$  bude vyjádřena pomocí konvoluce  $p(x)$  a  $q(y)$ :

$$P(Z = z) = \int_{-\infty}^{\infty} p(x)q(z-x)dx = \int_{-\infty}^{\infty} p(z-y)q(y)dy.$$

Konvoluce funkcí  $p$  a  $q$  se v matematice označuje symbolem  $p * q$ . Konvoluce  $p$  s  $p$  je druhá konvoluční mocnina, tedy  $p^{(2)}$ . Budeme-li tedy vyšetřovat pravděpodobnostní rozložení dvou stejně rozdělených náhodných veličin, dostaneme se ke druhé konvoluční mocnině jejich hustoty. Hustotu rozložení součtu  $n$  nezávislých, stejně rozdělených náhodných veličin potom budeme moci vyjádřit pomocí  $n$ -té konvoluční mocniny jejich společné hustoty.

*Cvičení:* Najděte hustotu rozdělení součtu  $n$  nezávislých náhodných veličin s exponenciálním rozdělením pravděpodobnosti  $\text{Exp}(\lambda)$ . Toto rozdělení se nazývá Erlangovo a má dva parametry:  $\lambda$  a  $n$ .

*Cvičení:* Najděte střední hodnotu a rozptyl náhodné veličiny s Erlangovým rozdělením pravděpodobnosti.

*Cvičení:* Najděte několik praktických příkladů náhodné veličiny s Erlangovým rozdělením pravděpodobnosti.

*Cvičení:* Ukažte, že Erlangovo rozdělení pravděpodobnosti je speciálním případem gamma rozdělení.

## Kapitola 15

# Doplňky – simulace náhodných veličin.

Význam metod pro stochastické modelování stále roste. Čím složitější úlohy se snažíme řešit, čím komplikovanější je jejich analytické řešení (je-li vůbec možné) a čím nákladnější je realizace reálných experimentů, tím častěji používáme simulační metody. Moderní výpočetní technika nám umožňuje ”na stole” realizovat poměrně rychle a snadno i takové projekty, do nichž bychom se dříve raději ani nepouštěli.

Základem všech stochastických simulací je kvalitní generátor pseudonáhodné posloupnosti. Proto první odstavec budeme věnovat právě jemu.

### 15.1 Pseudonáhodné posloupnosti

Snahy o konstrukci „stroje na generování náhody“ započaly již v minulém století. Předcházeli tomu různé pokusy s házením mincí, různých mnohostěnnů (kostek), s ruletou. Jako zdroj posloupnosti náhodných čísel byly používány i údaje získané při sčítání lidu (Kadyrov 1936). Spolu s těmito snahami se zároveň rozvíjela diskuse o tom, co to vlastně ta náhoda je, zda lze vůbec náhodnou posloupnost uměle získat a jak nahodilost ověřovat. Velká snaha byla věnována různým analogovým elektronickým zařízením pro generování takovýchto posloupností. Pomocí nich byly sestavovány tabulky čísel pro širší použití (1939 Kendall, Babbington-Smith). Generátor společnosti RAND – v roce 1955 vygeneroval jeden milión čísel z elektronického šumu. Nicméně i tyto posloupnosti nevyhovují zcela našim představám

o náhodném chování takovéto posloupnosti. Tabulky RAND lze používat pro některé metody běžné v aplikované statistice (např. znáhodňování pacientů v klinických pokusech, bootstrap) ale již nevyhovují pro jednoduché simulační úlohy ( $10^6$  je příliš málo). Cesta generování pomocí fyzikálních generátorů má z hlediska simulace ještě jeden závažný nedostatek – nemožnost opakování přesně stejné dostatečně dlouhé posloupnosti.

Posledně jmenovaný nedostatek odstraňují tzv. "algoritmické" generátory. Základ k nim položili von Neumann (1946) a Lehmer (1951). Fakt, že pomocí deterministického algoritmu lze vygenerovat posloupnost čísel, která se podle všech (tehdy) dostupných testů chová jako náhodná, byl ve své době šokující! Pro tyto posloupnosti se začal používat název "pseudonáhodné".

Příklad (J. von Neumann, "the middle square method"): vezměme čtyřmístné číslo, např.  $x_0 = 8653$ . Další číslo  $x_1$  dostaneme tak, že předchozí číslo umocníme na druhou a z takto získaného čísla vezmeme prostřední čtyři číslice:  $x_1 = 8744$  ( $8653^2 = 74874409$ ).

**Definice:** Posloupností pseudo-náhodných čísel  $U_i$  budeme nazývat deterministickou posloupnost čísel z intervalu  $[0,1]$  mající stejné odpovídající statistické vlastnosti jako posloupnost čísel náhodných.

Základní vlastnosti požadované od pseudonáhodné posloupnosti:

- jednotlivé členy posloupnosti by měly být nezávislé a to i v případě posloupnosti vektorů vyšších dimenzí.

V praxi požadujeme aby sdružené rozložení  $(U_{i+1}, \dots, U_{i+k})$  nebylo „příliš vzdálené“ od rovnoměrného rozložení na  $[0,1]^k$  pro malé hodnoty  $k$  ( $\leq 6$ ).

- požadujeme „nepředvídatelnost“, to znamená co největší délku posloupnosti bez opakování stejných sekvencí.

S touto vlastností souvisí pojem "perioda" pseudonáhodné posloupnosti.

**Příklad:** Vezmeme-li jako počáteční hodnotu v předchozím příkladu číslo  $x_0 = 2100$ , potom dostaneme posloupnost:  $\{2100, 4100, 8100, 6100, 2100, \dots\}$ . Tedy perioda je rovna 4.

Obecně lze uvést, že



- je-li proces generován nějakým algoritmem, potom není náhodný, neboť jej lze za týchž podmínek opětovně zrealizovat
- je-li proces generován nealgoritmicky (například prostřednictvím nějakého fyzikálního děje), otázka náhodnosti je obecně nerozhodnutelná.

## 15.2 Algoritmické generátory pseudonáhodných čísel

**Příklad:** (Fibonacciho generátor) Uvažujme algoritmus  $U_i = (U_{i-1} + U_{i-2}) \bmod 1$ . Tento algoritmus poskytne posloupnost 50ti čísel (po sloupcích):

0,563	0,810	0,478	0,067	0,218	0,471	0,396	0,824	0,455	0,825
0,624	0,809	0,527	0,603	0,163	0,397	0,527	0,197	0,692	0,811
0,187	0,620	0,005	0,671	0,382	0,867	0,923	0,020	0,147	0,635
0,811	0,429	0,531	0,274	0,545	0,264	0,450	0,217	0,839	0,446
0,999	0,049	0,536	0,945	0,926	0,132	0,373	0,238	0,986	0,082

Všimněte si, že nikdy není  $U_{i-2} < U_i < U_{i-1}$  !

Přestože naším cílem je získat co nejdelší posloupnost čísel z intervalu  $[0, 1]$ , jejíž vlastnosti se co nejvíce blíží posloupnosti realizací nezávislých a stejně rozložených (rovnoměrně na  $[0, 1]$ ) náhodných veličin, většina generátorů na číslicových počítačích pracuje v celočíselné aritmetice. To má hned několik důvodů:

- výpočet je rychlejší, než výpočty v pohyblivé desetinné čárce,
- máme k dispozici více bitů ve slově a tím lze dosáhnout delší periody,
- výpočet se většinou redukuje na násobení, při kterém je potlačena signalizace přetečení registru (ignoruje se).

**Příklad:**  $U_n = MU_{n-1} \bmod 1$ ,  $U_0 = 2^{-m}$ . Lépe bude použít algoritmus  $X_n = MX_{n-1} \bmod 2^m$ ,  $U_0 = 1$  a transformace  $U_n = X_n 2^{-m}$ .

(Dělení  $2^m$  představuje ve dvojkové aritmetice posun bitů slova o  $m$  pozic vpravo.)

### 15.2.1 Kongruenční generátory

Tento typ generátorů tvoří nejrozšířenější a nejlépe prostudovanou třídu algoritmických generátorů. Kongruenční generátory jsou definovány rekurentním lineárním vztahem

$$X_i = (a_0 + a_1 X_{i-1}) \pmod{m}$$

s počáteční hodnotou  $X_0$  (tzv. „seed“), kde  $a_1$  (multiplier),  $a_0$  (shift) a  $m$  (modulus) jsou předem vhodně zvolené konstanty. Pseudonáhodná posloupnost je dána vztahem  $U_i = X_i/m$ ,  $i = 1, 2, \dots$

- Je-li  $a_0 = 0$ , jedná se o tzv. *multiplikativní generátor*. Popsal jej poprvé Lehmer (1951) a doporučil konstanty  $a_1 = 23$ ,  $m = 10^8 + 1$  pro dekadický počítač.
- Obecněji lze kongruenční generátor popsat rekurentním vztahem

$$X_i = (a_0 + a_1 X_{i-1} + a_2 X_{i-2} + \dots + a_k X_{i-k}) \pmod{m}, \quad i = k, k+1, \dots$$

s počátečními hodnotami  $X_0, \dots, X_k$ .

- *Fibonacciho generátory* jsou definovány obecně rekurentním vztahem

$$V_i = V_{i-r} \otimes V_{i-s}, \quad r \geq 1, s \geq 1, r \neq s,$$

kde symbol  $\otimes$  označuje nějakou vhodnou matematickou operaci definovanou na objektech  $V_i$ . Tak například

- $V_i$  jsou reálná čísla a  $\otimes$  reprezentuje sčítání nebo odečítání modulo 1
- $V_i$  jsou  $(n-1)$ bitová čísla a  $\otimes$  je sčítání, odčítání či násobení modulo  $2^n$
- $V_i$  jsou binární vektory a  $\otimes$  je binární sčítání, odčítání či násobení, binární EOR (exclusive OR, t.j.  $0 + 0 = 1 + 1 = 0$ ,  $0 + 1 = 1 + 0 = 1$ ).

**15.2.2 Generátory založené na posunu v registrech****15.3 Testování pseudonáhodných posloupností**

Pro posloupnost nezávislých náhodných veličin  $U_i, i = 1, 2, \dots$ , rovnoměrně rozložených na intervalu  $[0, 1]$  požadujeme aby platilo

$$P(U_{i_1} \leq u_1, \dots, U_{i_k} \leq u_k) = \prod_{j=1}^k P(U_{i_j} \leq u_j) = \begin{cases} \prod_{j \in \mathcal{J}} u_j, & \mathcal{J} = \{j : 1 \leq j \leq k, 0 < u_j < 1\}, \\ 0 & \text{je-li některé } u_j \leq 0, j = 1, \dots, k \end{cases}$$

Specielně pro  $k = 1$ :

$$P(U \leq u) = \begin{cases} 0 & u \leq 0, \\ u & u \in (0, 1), \\ 1 & u \geq 1. \end{cases}$$

a pro  $k = 2$ :

$$P(U_1 \leq u_1, U_2 \leq u_2) = \begin{cases} 0 & u_1 \leq 0 \text{ nebo } u_2 \leq 0, \\ u_1 u_2 & u_1, u_2 \in (0, 1), \\ u_1 & u_2 \geq 1. \\ u_2 & u_1 \geq 1. \\ 1 & u_1, u_2 \geq 1. \end{cases}$$

pro libovolnou podmnožinu  $i_1, i_2, \dots, i_k$  množiny přirozených čísel.

**15.3.1 Testy shody s rovnoměrným rozložením pravděpodobnosti****15.3.2 Testy nezávislosti****15.3.3 Test délky cyklu****15.4 Generování náhodných veličin**

Viz [1].

15.4.1 Metoda inverze

15.4.2 Metoda zamítací

15.4.3 Některá diskrétní rozložení

15.4.4 Některá spojitá rozložení

15.4.5 Náhodný výběr, míchání

15.4.6 Metody Monte Carlo

15.5 Simulace náhodných procesů

# Literatura

- [1] Antoch, J., Vorlíčková, D.: *Vybrané metody statistické analýzy dat*, kap. 5. Programové zabezpečení pro statistickou analýzu dat. str. 209 ...
- [2] Baccelli, F., Brémaud, P.: *Elements of Queueing Theory*. Applications of Mathematics 26, Springer–Verlag, 1994.
- [3] Dohnal, G.: *Analýza výrobních systémů*. Habilitační práce, ČVUT Praha (1993).
- [4] Kleinrock, L.: *Queueing Systems*. orig. John Wiley & Sons, 1975, ruský překlad Moskva, 1979.
- [5] Machek, J., Likeš, J.: *Teorie pravděpodobnosti*. Matematika pro vysoké školy technické, SNTL Praha (1987).
- [6] Wolf, R. W.: *Stochastic Modeling and Theory of Queues*. Prentice–Hall Int., 1989.
- [7] Zítek, F.: *Ztracený čas*. ACADEMIA Praha, 1969.