

PA054: Formální modely v systémové biologii

David Šafránek

19.3.2010

Stochastická sémantika modelu

Uvažujme model $\mathcal{M} = \langle S, R, \text{reanet}, \emptyset, \text{map} \rangle$. Označme $n = |S|$.

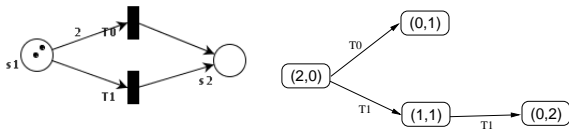
- $S\text{Val} = \mathbb{N} \dots$ množství (počet molekul)
- $R\text{Val} = \langle 0, 1 \rangle \dots$ pravděpodobnost provedení reakce v časovém intervalu $(0, t)$, $t \in \mathbb{R}^+$

Vektor $[[\vec{s}]]_{\mathcal{M}} \in S\text{Val}^n$ reprezentuje stav modelu (ohodnocení proměnných) v daném okamžiku.

Ohodnocení vektoru proměnných \vec{s} lze samplovat prostřednictvím vektorové náhodné proměnné $X_{\mathcal{M}}$, $X_{\mathcal{M}} = \langle X_1, \dots, X_n \rangle$, kde $X_i : S\text{Val}^n \rightarrow \mathbb{N}$ charakterizuje počet molekul substance $s_i \in S$ v daném okamžiku.

Stochastickou (pravděpodobnostní) sémantiku modelu \mathcal{M} definujeme jako stochastický proces $[[\mathcal{M}]] = \{X_{\mathcal{M}}(t) | t \in \mathbb{R}_0^+\}$.

Pravděpodobnostní funkce označování

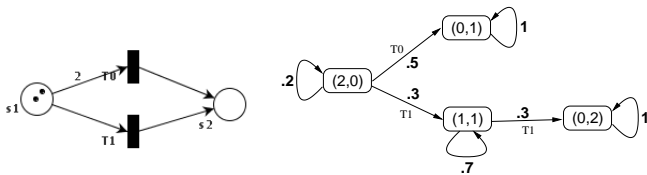


$\llbracket \bar{s} \rrbracket_{\mathcal{M}}$		$X_{\mathcal{M}}$	
s_1	s_2	X_1	X_2
2	0	2	0
1	1	1	1
0	2	0	2
0	1	0	1

- $\Pr\{X_{\mathcal{M}} = (2, 0)\} = \Pr\{X_{\mathcal{M}} = (1, 1)\} = \Pr\{X_{\mathcal{M}} = (0, 2)\} = \Pr\{X_{\mathcal{M}} = (0, 1)\} = \frac{1}{4}$
- $\Pr\{X_1 = 1\} = \Pr\{X_1 = 2\} = \Pr\{X_2 = 0\} = \Pr\{X_2 = 2\} = \frac{1}{4}$
- $\Pr\{X_1 = 0\} = \Pr\{X_2 = 1\} = \frac{2}{4} = \frac{1}{2}$
- pravděpodobnostní funkce:

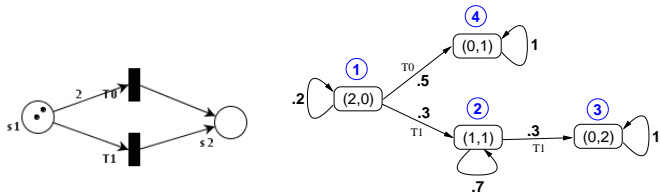
$$p((x_1, x_2)) = \Pr\{X_{\mathcal{M}} = (x_1, x_2)\} = \Pr\{X_1 = x_1 \cap X_2 = x_2\} = \Pr\{X_1 = x_1\} \cdot \Pr\{X_2 = x_2 | X_1 = x_1\}$$

Diskrétní Markovův řetězec a Markovův proces



- sledujeme časovou progresi náhodné proměnné $X_{\mathcal{M}}(t)$
- čas uvažujeme jako spočetnou veličinu
- vývoj $X_{\mathcal{M}}$ v čase lze popsat grafem $MC = \langle V, E, p \rangle$:
 - stavy V reprezentují prvky jevového pole
 - přechody E jsou ohodnoceny pravděpodobnostmi $p : E \rightarrow (0, 1)$
t.ž. $\forall v \in V. \sum_{v' \in V} p(\langle v, v' \rangle) = 1$

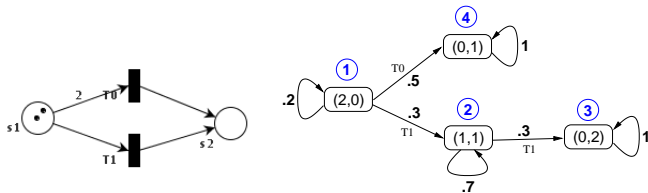
Diskrétní Markovův řetězec a Markovův proces



- sledujeme časovou progresi náhodné proměnné $X_{\mathcal{M}}(t)$
- čas uvažujeme jako spočetnou veličinu
- vývoj $X_{\mathcal{M}}$ v čase lze popsat grafem $MC = \langle V, E, p \rangle$:
 - stavy V reprezentují prvky jevového pole
 - přechody E jsou ohodnoceny pravděpodobnostmi $p : E \rightarrow (0, 1)$
t.ž. $\forall v \in V. \sum_{v' \in V} p(\langle v, v' \rangle) = 1$
 - ekvivalentně lze reprezentovat přechodovou maticí:

$$P = \begin{pmatrix} .2 & .3 & 0 & .5 \\ 0 & .7 & .3 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad P_{ij} = \Pr\{X_{\mathcal{M}}(t+1) = j | X_{\mathcal{M}}(t) = i\}$$

Diskrétní Markovův řetězec a Markovův proces



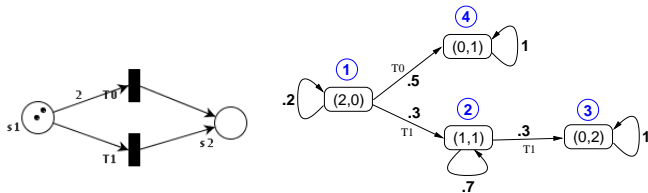
Pravděpodobnostní rozložení náhodné proměnné ve stavu i v čase t budeme značit $p_i(t) = \Pr\{X_M(t) = i\}$. Pro okamžik t máme vektorové rozložení $p(t) = \langle p_1(t), p_2(t), p_3(t), p_4(t) \rangle$.

Iniciálně předpokládejme $p_1(0) = p_2(0) = p_3(0) = p_4(0) = \frac{1}{4}$.

Vývoj pravděpodobnostního rozložení proměnné X_M v čase:

$$\begin{aligned} p(1) &= p(0)P \\ p(2) &= p(1)P \\ &\vdots \\ p(k) &= p(0)P^k \end{aligned}$$

Diskrétní Markovův řetězec a Markovův proces



Stochastický proces $\{X_M(t) | t \in \mathbb{N}\}$ rozvíjený dle Markovova řetězce z předchozího slidu se nazývá *Markovův proces*.

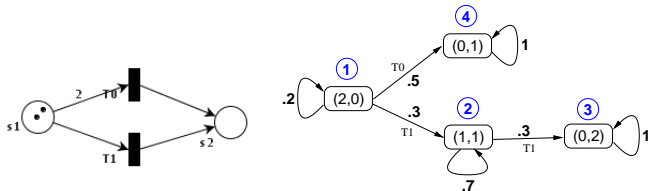
Klíčovou vlastností Markovova procesu je nezávislost na historii (tzv. “memoryless”):

Pravděpodobnostní rozložení proměnné $X_M(t + 1)$ závisí pouze na bezprostředně (v čase) předchozím rozložení $X_M(t)$.

$$p_j(t + 1) = \sum_{i=1}^k P_{ij} p_i(t)$$

Diskrétní Markovův řetězec a Markovův proces

Příklad

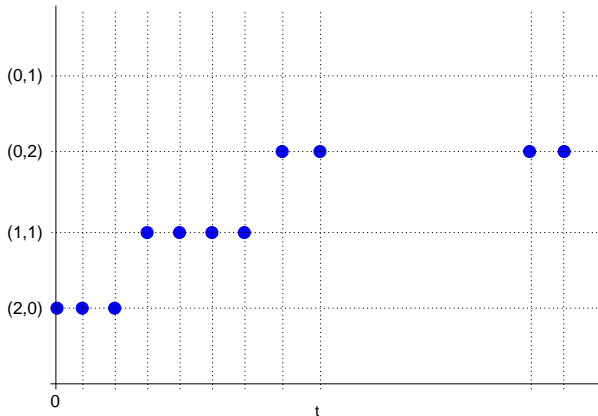


	1	2	3	4
$p(0) :$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$
$p(1) :$	0.0500	0.2500	0.3250	0.3750
$p(2) :$	0.0100	0.1900	0.4000	0.4000
$p(3) :$	0.0020	0.1360	0.4570	0.4050
$p(4) :$	0.0004	0.0958	0.4978	0.4060
$p(5) :$	0.0001	0.0672	0.5265	0.4062
$p(6) :$	0.0000	0.0470	0.5467	0.4062
$p(7) :$	0.0000	0.0329	0.5608	0.4062
$p(8) :$	0.0000	0.0231	0.5707	0.4062
		\vdots		
$p(26) :$	0.0000	0.0000	0.5937	0.4062
$p(27) :$	0.0000	0.0000	0.5937	0.4062

Diskrétní Markovův řetězec a Markovův proces

Příklad

Sémantikou Markovova procesu je (nekonečná) množina všech trajektorií samplujících stavový prostor v (diskrétním) čase. Každá trajektorie poskytuje možnou dynamiku systému (s ohledem na pravděpodobnostní rozložení).



Markovův proces a stabilita

Mějme markovův proces charakterizovaný přechodovou maticí P .
Rozložení p je *stabilní*, pokud platí

$$p = pP.$$

Stabilní rozložení je pevným bodem lineárního zobrazení určeného maticí P . Lze získat řešením homogení lineární soustavy:

$$\begin{aligned} p = pP &\Leftrightarrow p - pP = 0 \\ &\Leftrightarrow p(E - P) = 0 \end{aligned}$$

kde E je jednotková matice příslušného rozměru.

Markovův proces a biologické modely

- interpretace modelu
 - vycházíme z daného stavu (označkování)
 - v diskrétním časové kroku se projeví právě jedna (uschopněná) reakce r_i (s danou pravděpodobností) nebo stav zůstane nezměněn
 - každý přechod je vážen pravděpodobnostním rozložením závislým na konstantních pravděpodobnostech – při jeho respektování lze provádět Monte Carlo simulaci
- analyzační techniky
 - transientní analýza
 - výpočet distribuce $p(t)$ pro libovolné t
 - $p(t) = p(0)P^t$
 - stacionární analýza
 - analýza distribuce ve stabilním stavu

Stochastická sémantika modelu

- nedostatky Markovova procesu pro biologický model
 - pravděpodobnosti provedení reakcí zcela nezávislé na ohodnocení reaktantů
 - diskrétní samplování času neumožňuje dostatečně přesně vyjádřit rychlost reakce (četnost projevu reakce v čase)
- co chceme modelovat?
 - dynamika biologického systému při nízkých koncentracích
 - projev náhodných vlivů ovlivňujících provedení reakce
 - respektovat časo-prostorové jevy uvnitř buňky

Poissonův proces

Uvažujme experiment představující četnost výskytu diskrétních událostí v časovém intervalu t . Pro zachycení tohoto měření zavedeme náhodnou proměnnou X .

Pak platí:

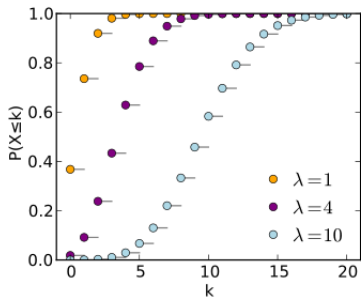
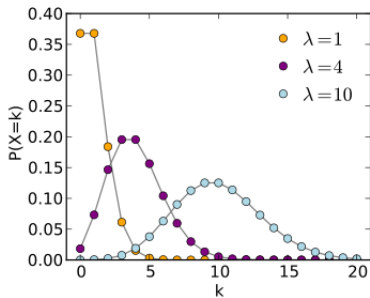
$$X \sim Po(\lambda t)$$

Jinými slovy, pravděpodobnostní funkce určující pravděpodobnost jevu, že za dobu t nastane právě k událostí, má následující tvar:

$$\Pr\{X = k\} = \frac{e^{-\lambda t}(\lambda t)^k}{k!}$$

kde λ je odpovídající parametr Poissonova rozložení.

Poissonovo rozložení



$$X \sim Po(\lambda)$$

Poissonův proces

Nyní uvažujme náhodnou proměnnou X z předchozího slidu v závislosti na čase, tzv. stochastický proces určený množinou náhodných proměnných $\{X(t) | t \in \mathbb{R}_0^+\}$. Uvažujme časové intervaly délky τ .

Předpokládejme následující podmínky:

- výskyty událostí ve dvou libovolných vzájemně disjunktních časových intervalech jsou nezávislé jevy
- pravděpodobnostní rozložení četnosti výskytu událostí v daném časovém intervalu závisí pouze na délce intervalu
- žádné dvě události nemohou nastat současně („interleaving“)

Pak pro libovolný interval $(t, t + \tau)$ platí:

$$X(t + \tau) - X(t) \sim Po(\lambda\tau)$$

Exponenciální rozložení

Problém

Uvažujme experiment, v němž budeme měřit dobu mezi bezprostředně následujícími událostmi Poissonova procesu. Měření zachytíme náhodnou proměnnou X . Jaké rozložení má X ?

- doménou X je nyní \mathbb{R}
- X je spojitá náhodná proměnná
- nutno uvažovat hustotu pravděpodobnosti:

$$f_X(x) = \lim_{\delta x \rightarrow 0} \frac{\Pr\{x \leq X \leq x + \delta x\}}{\delta x}$$

$$F_X(x) = \Pr\{X \leq x\} = \int_{-\infty}^x f_X(z) dz$$

Exponenciální rozložení

$$X \sim \text{Exp}(\lambda)$$

pokud:

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0, \\ 0, & \text{jinak.} \end{cases}$$

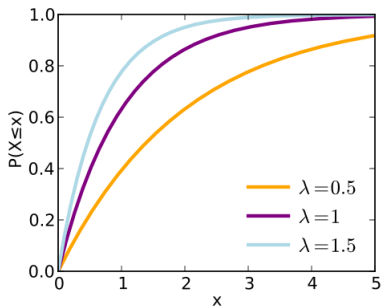
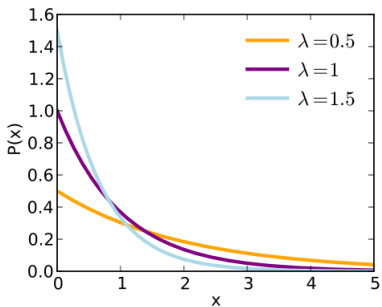
Pro distribuční funkci dostáváme:

$$F_X(x) = \begin{cases} 0, & x < 0, \\ 1 - e^{-\lambda x}, & x \geq 0. \end{cases}$$

Střední hodnota:

$$E(X) = \frac{1}{\lambda}$$

Exponenciální rozložení



Exponenciální rozložení

Vlastnosti minima exponenciálních distribucí

Uvažme X_1, \dots, X_n nezávislé náhodné proměnné t.ž.

$\forall i. X_i \sim \text{Exp}(\lambda_i)$. Pro minimální exponenciální rozložení $\min\{X_1, \dots, X_n\}$ platí:

$$\begin{aligned}\Pr\{\min\{X_1, \dots, X_n\} > x\} &= \Pr\{X_1 > x \cap X_2 > x \cap \dots \cap X_n > x\} \\ &= \prod_{i=1}^n \Pr\{X_i > x\} = \prod_{i=1}^n e^{-\lambda_i x} = e^{-x \sum_{i=1}^n \lambda_i}\end{aligned}$$

Pro parametr minimálního rozložení platí:

$$\Pr\{X_k = \min\{X_1, \dots, X_n\}\} = \frac{\lambda_k}{\lambda_1 + \dots + \lambda_n}$$

Exponenciální rozložení

Uvažujme Poissonův proces $\{X(t)|t \geq 0\}$ t.ž. $X \sim Po(\lambda)$.
Zavedeme náhodnou proměnnou T zachycující dobu do první
nejbližší události (od počátečního okamžiku).

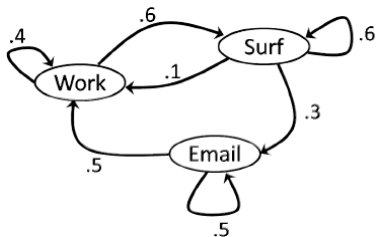
*Pro $t > 0$ uvažujme náhodnou proměnnou N_t zachycující počet událostí
v intervalu $(0, t)$. Z definice platí $N_t \sim Po(\lambda t)$.*

$$\begin{aligned}F_T(t) &= \Pr\{T \leq t\} \\&= 1 - \Pr\{T > t\} \\&= 1 - \Pr\{N_t = 0\} \\&= 1 - \frac{(e^{-\lambda t})(\lambda t)^0}{0!} \\&= 1 - e^{-\lambda t}\end{aligned}$$

Tedy platí:

$$T \sim Exp(\lambda)$$

Motivace pro spojitý Markovův řetězec

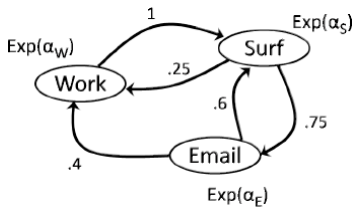
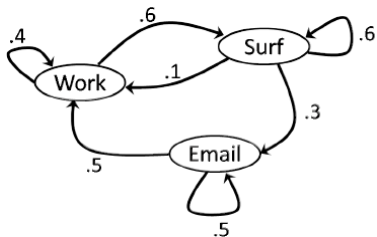


- uvažujme spojitý čas pobytu ve stavu, $t \in \mathbb{R}^+$
- lze zachytit rozložením W samplující „čekací“ dobu mezi změnami stavů
- požadujeme markovskou vlastnost nezávislosti na historii:

$$\Pr\{U > t + \tau | U > \tau\} = \Pr\{U > t\}$$

- tuto vlastnost má exponenciálně distribuovaná proměnná

Motivace pro spojitý Markovův řetězec



- uvažujme spojitý čas pobytu ve stavu, $t \in \mathbb{R}^+$
- lze zachytit rozložením W samplující „čekací“ dobu mezi změnami stavů
- požadujeme markovskou vlastnost nezávislosti na historii:

$$\Pr\{U > t + \tau | U > \tau\} = \Pr\{U > t\}$$

- tuto vlastnost má exponenciálně distribuovaná proměnná

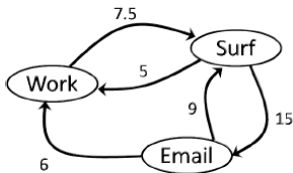
Spojité Markovův řetězec

Spojité Markovův řetězec lze definovat jako přechodový graf
 $MC = \langle V, E, p \rangle$:

- stavy V reprezentují prvky jevového pole
- přechody E jsou ohodnoceny pravděpodobnostmi $p : E \rightarrow (0, 1)$
t.ž. $\forall v \in V. \sum_{v' \in V} p(\langle v, v' \rangle) = 1$
- $\forall v \in V. \langle v, v \rangle \notin E$
- pro každý stav $v \in V$ je přiřazen parametr čekací doby
 $\alpha_v \in \mathbb{R}^+$

Spojité Markovův řetězec

Příklad



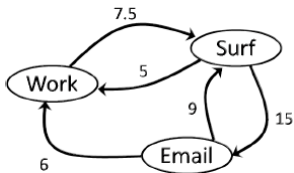
- průměrná čekací doba ve stavu *Surf* je 3 minuty, což je $\frac{3}{60} = \frac{1}{20}$ hod
 $\Rightarrow \alpha_S = 20$
- $\alpha_W = 7.5$
- $\alpha_E = 15$

Simulace spojitého Markovova řetězce

```
// inicializace počáteční distribuce X(0)
t := 0
u := X(0)
while ( true )
    wait_time := Exp( $\alpha_u$ )
    for each s, t < s < t + wait_time do
         $X_s := u$ 
    t := t + wait_time
    select v in V with probability  $Q_{uv}$ 
     $X_t := v$ 
    u := v
```

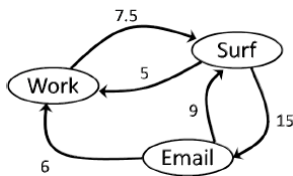
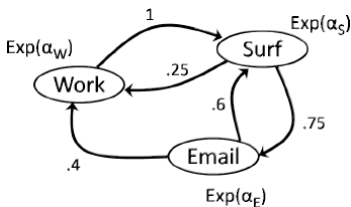
Spojité Markovův řetězec

Tradiční zápis



Spojité Markovův řetězec

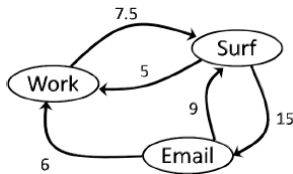
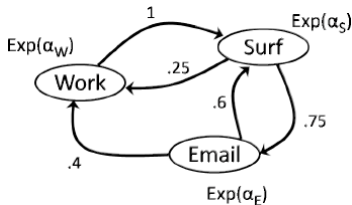
Převod na tradiční zápis



- rozhodovací procedura při opuštění stavu *Surf*
 - v okamžiku vstupu do stavu *Surf* se spustí stopky:
 A_W měřící časový úsek s distribucí $\text{Exp}(\lambda_{a_W})$
 A_E měřící časový úsek s distribucí $\text{Exp}(\lambda_{a_E})$
 - jakmile některé doběhnou, přesun do příslušného stavu
- $\Pr\{\min\{A_W, A_E\} = A_W\} = \frac{\lambda_{a_W}}{\lambda_{a_W} + \lambda_{a_E}}$
- $\Pr\{\min\{A_W, A_E\} = A_E\} = \frac{\lambda_{a_E}}{\lambda_{a_W} + \lambda_{a_E}}$

Spojité Markovův řetězec

Převod na tradiční zápis – příklad



- $A_E \sim \text{Exp}(15)$, $A_W \sim \text{Exp}(5)$
- $\Pr\{\min\{A_W, A_E\} = A_E\} = \frac{15}{15+5} = .75$
- $\Pr\{\min\{A_W, A_E\} = A_W\} = \frac{5}{15+5} = .25$

Spojité Markovův řetězec

Analýza

- CTMC definuje stochastický proces $\{X(t) | t \in \mathbb{R}^+\}$
- pravděpodobnost stavu i v čase t značena $p_i(t)$
- pro konečně stavový CTMC vektorová pravděpodobnost v čase t : $p(t) = \langle p_1(t), \dots, p_n(t) \rangle$
- přechodová matice nahrazena generující maticí Q :

$$Q_{ij} = p(i, j), i \neq j \quad Q_{ii} = - \sum_{j=1, j \neq i}^k Q_{ij}$$

- $p(t) = p(0)e^{Qt}$
- pokud je přechodový graf CTMC konečný a silně souvislý, pak existuje jednoznačné stabilní rozložení, které je řešením $pQ = 0$ a $\sum_i^n p_i = 1$

Stochastická sémantika modelu

Uvažme reakci $X + Y \rightarrow XY$.

K úspěšnému provedení je nutné, aby došlo ke kolizi molekul X a Y při Brownově pohybu. Molekuly se musí nacházet v dostatečné blízkosti.

Uvažujeme-li konstantní teplotní podmínky a dobře promíchané médium konstantního objemu (zejména konstantní uniformní rozložení molekul v prostoru nezávislé na čase), lze považovat pravděpodobnost kolize (tzv. *hazard*) za konstantní. Při těchto podmínkách lze událost “přiblížení molekul do reakční vzdálenosti” považovat za nezávislou na čase.

Stochastická sémantika modelu

Uvažme model $\mathcal{M} = \langle S, R, \text{reanet}, \emptyset, \text{map}, \text{rates} \rangle$ splňující $\forall r \in R. |\{s \in S \mid \langle s, r \rangle\}| \leq 1$ rozšířený o ohodnocení reakcí $\text{rates} : R \rightarrow \mathbb{R}^+$ *stochastickými reakčními konstantami*.

Ke každé reakci $r_i \in R$ přiřadíme *hazardní funkci* $h_i : SVaI^k \rightarrow \mathbb{R}$, $k = |\{s \in S \mid \langle s, r_i \rangle\}|$ dle následujících pravidel (pro lib. $r_i \in R$):

- $r_i : \emptyset \rightarrow *$

$$h_i(x) = \text{rates}(r_i)$$

- $r_i : s \rightarrow *, s \in S$

$$h_i(x) = \text{rates}(r_i) \cdot \llbracket s \rrbracket_{\mathcal{M}}$$

- $r_i : s_p + s_q \rightarrow *, s_p, s_q \in S, s_p \neq s_q$

$$h_i(x) = \text{rates}(r_i) \cdot \llbracket s_p \rrbracket_{\mathcal{M}} \cdot \llbracket s_q \rrbracket_{\mathcal{M}}$$

- $r_i : 2s \rightarrow *, s \in S$

$$h_i(x) = \text{rates}(r_i) \cdot \frac{\llbracket s \rrbracket_{\mathcal{M}} (\llbracket s \rrbracket_{\mathcal{M}} - 1)}{2}$$

Stochastická Petriho síť – syntax

Nechť $H = \bigcup_{t \in T} \{h_t | h_t : \mathbb{N}^{|\bullet t|} \rightarrow \mathbb{R}^+\}$ je množina hazardních funkcí.
Stochastická Petriho síť je definována jako pětice

$SPN = \langle P, T, f, \nu, m_0 \rangle$.

- P je konečná neprázdná množina *míst* (places),
- T je konečná neprázdná množina *přechodů* (transitions),
- $f : ((P \times T) \cup (T \times P)) \rightarrow \mathbb{N}$ je množina orientovaných hran vážených celými čísly,
- $\nu : T \rightarrow H$ je přiřazení hazardní funkce h_t každému přechodu $t \in T$
- $m_0 : P \rightarrow \mathbb{N}$ je *iniciální označkování* (marking).

Stochastická Petriho síť – sémantika

Hazardní funkce h_t definuje parametr $\lambda_t(m)$ exponenciálního rozložení pravděpodobnosti provedení přechodu t v čase. Toto rozložení je závislé na aktivním označování m .

Hustota pravděpodobnosti provedení přechodu t v čase $\tau \geq 0$:

$$f_{X_t}(\tau) = \lambda_t(m) \cdot e^{(-\lambda_t(m)\tau)}$$

Hazardní funkce pro biologické modely je definována v souladu s předchozím slidem:

$$h_t = \text{rates}(t) \cdot \prod_{p \in \bullet t} \binom{m(p)}{f(p, t)}$$

Stochastická Petriho síť – sémantika

- ohodnotíme graf dosažitelnosti SPN \mathcal{P} t.ž. každá hrana je ohodnocena evaluací příslušné hazardní funkce
- výsledný CTMC reprezentuje sémantiku \mathcal{P}
- CTMC je izomorfní grafu dosažitelnosti
 - je-li síť ohraničená a reversibilní, existuje (jednoznačně daný) stabilní stav nezávislý na m_0