

PB051: Výpočetní metody v bioinformatice a systémové biologii

David Šafránek

27.4.2011

Tento projekt je spolufinancován Evropským sociálním fondem a státním rozpočtem České republiky.



Obsah

Modelování dynamiky chemických reakcí

Modelování dynamiky transkripční regulace

Varianty Gillespiho algoritmu

Aproximativní metody

Deterministické metody

Hybridní metody

Obsah

Modelování dynamiky chemických reakcí

Modelování dynamiky transkripční regulace

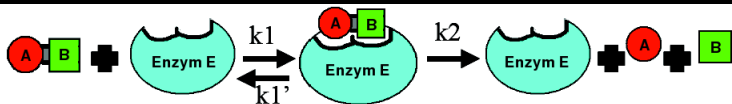
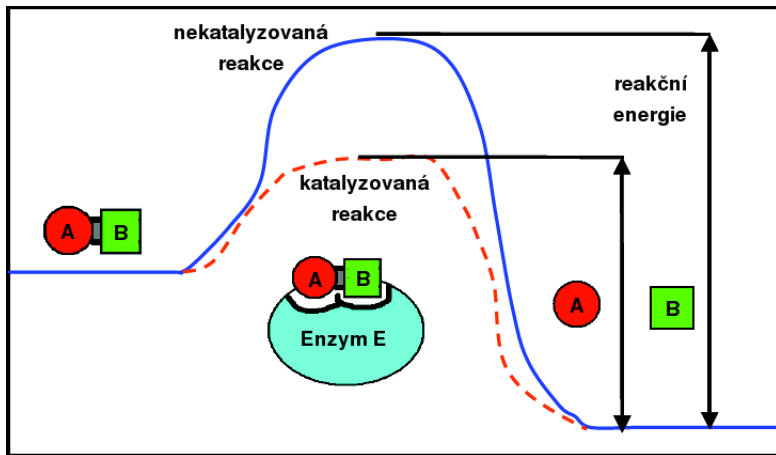
Varianty Gillespiho algoritmu

Aproximativní metody

Deterministické metody

Hybridní metody

Energetický proces chemických reakcí



Energetický proces chemických reakcí

- různé energetické stavy molekuly
 - např. komplex AB méně stabilní než individuální výskyt molekul A, B
 - při přechodu mezi energ. stavy dochází k výměně energie
 - energie požadována pro aktivaci procesu (aktivační energie)
 - energie uvolněna během procesu (volná energie)
- pro biologický systém je zdrojem většiny energie metabolismus
- absolutní teplota ovlivňuje kinetickou energii molekul
- pro reakci (úspěšnou kolizi) musí být splněno:
 - správná prostorová konfigurace (orientace) molekul
 - dostatek kinetické energie

Model dynamiky chemických reakcí

- fixujeme-li konstantní teplotu, objem a tlak, orientace a kinetická energie molekul je stochastickým jevem
 - uvažujeme dobře promíchané médium
 - lze definovat průměrnou pravděpodobnost kolize molekul v časovém okamžiku
 - určeno fyzikálními vlastnostmi reagujících molekul
- máme-li N_1 molekul látky S_1 a N_2 molekul látky S_2 , pak náhodnou proměnnou χ charakterizující pravděpodobnost kolize daného počtu molekul S_1 a S_2 v časovém okamžiku dt lze modelovat:

$$\chi = (c \cdot dt) \cdot N_1 \cdot N_2$$

- předpokládáme S_1 a S_2 různé látky
- c je stochastická konstanta charakterizující průměrnou frekvenci úspěšných kolizí molekul S_1 a S_2 za jednotku času
- dt uvažováno limitně

Model dynamiky chemických reakcí

- závislost frekvence kolizí c na absolutní teplotě T (Arheniův zákon):

$$c \propto e^{-\frac{E_A}{RT}}$$

- E_A ... aktivační energie reakce
 - R ... plynová konstanta
- rovnovážný poměr frekvence dopředné a zpětné reakce u reversibilních reakcí:

$$K_{eq} = e^{-\frac{\Delta G}{RT}}$$

- ΔG ... volná energie vyměněná při reakci
 - např. komplex kooperujících TF má vyšší stabilitu

Polach K. J., Widom J., A Model for the Cooperative Binding of Eukaryotic Regulatory Proteins to Nucleosomal Target Sites, *Journal of Molecular Biology*, Volume 258, Issue 5, 24 May 1996, Pages 800-812, ISSN 0022-2836, DOI: 10.1006/jmbi.1996.0288.

Stochastický model reakční dynamiky

- uvažujme systém n substancí $S = \{S_1, \dots, S_n\}$ provázaných m reakcemi $R = \{R_1, \dots, R_m\}$
- uvažujeme pouze reakce 1. a 2. řádu
- systém zapisujeme pomocí stoichiometrické matice M rozměru $n \times m$:

$$M_{ij} = \begin{cases} -K, \text{ je-li } K \cdot S_i \text{ reaktantem } R_j \\ K, \text{ je-li } K \cdot S_i \text{ produktem } R_j \end{cases}$$

- závislé reakce:

$$\text{dep}(R_i, R_j) \Leftrightarrow \exists k. M_{ki} \cdot M_{kj} < 0$$

Stochastický model reakční dynamiky

- počet molekul substance S_i v čase t budeme značit $N_i(t)$
- náhodnou proměnnou rozložení počtů molekul substancí v čase t charakterizujeme vektorem:

$$X(t) = \langle N_1(t), \dots, N_n(t) \rangle$$

- vývoj tohoto rozložení $X(t)$ v čase charakterizujeme jako stochastický proces:

$$\{X(t) | t \in \mathbb{R}^+\}$$

Motivace pro spojité Markovův řetězec

- stochastický proces $\{X(t) | t \in \mathbb{R}^+\}$
- spojité čas pobytu ve stavu
- lze zachytit rozložením W samplujícím “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
 - $W \sim \text{Exp}(\lambda)$
 - $\frac{1}{\lambda}$... průměrná čekací doba

Exponenciální rozložení

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

pokud:

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

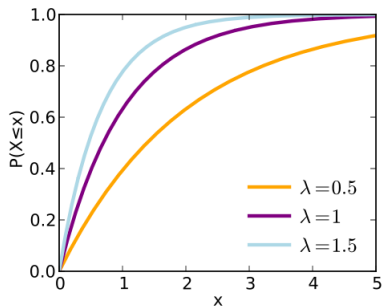
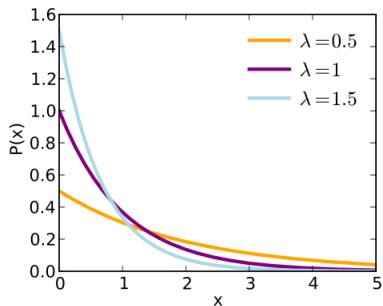
Pro distribuční funkci dostáváme:

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

Střední hodnota:

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

Exponenciální rozložení



Stochastický model reakční dynamiky

- interleaving: při přechodu $X(t) \rightarrow X(t + dt)$ je updatována právě jedna složka X
- provedení právě jedné reakce z R
- provedení reakce uvažováno jako okamžitý jev (trvá nulový čas)
- ve stavu $X(t) = \langle N_1, \dots, N_n \rangle$ je doba do provedení lib. reakce $R_i \in R$ charakterizována rozložením $\text{Exp}(\chi_i(X, c_i))$

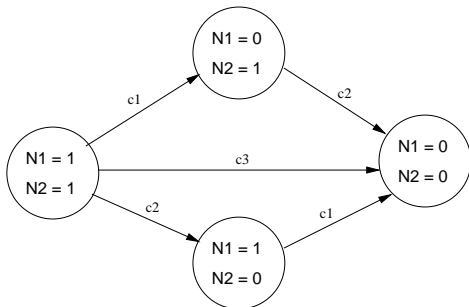
R_i	$\emptyset \rightarrow *$	$\chi_i(X, c_i) = c_i \cdot dt$
R_i	$S_j \rightarrow *$	$\chi_i(X, c_i) = (c_i \cdot dt) \cdot N_j$
R_i	$S_p + S_q \rightarrow *$	$\chi_i(X, c_i) = (c_i \cdot dt) \cdot N_p \cdot N_q$
R_i	$2S_j \rightarrow *$	$\chi_i(X, c_i) = (c_i \cdot dt) \cdot \frac{N_j \cdot (N_j - 1)}{2}$

- doba do nejbližší reakce má rozložení $\text{Exp}(\chi(X, c))$, kde

$$\chi(X, c) = \sum_{i=1}^m \chi_i(X, c_i), \quad c = \langle c_1, \dots, c_m \rangle$$

- pravděpodobnost provedení reakce R_i : $P(R_i) = \frac{\chi_i(X, c_i)}{\chi(X, c)}$

Stochastický model reakční dynamiky



R_1	$S_1 \xrightarrow{c_1}$	$\chi_1 = (c_1 \cdot dt) \cdot N_1$
R_2	$S_2 \xrightarrow{c_2}$	$\chi_2 = (c_2 \cdot dt) \cdot N_2$
R_3	$S_1 + S_2 \xrightarrow{c_3}$	$\chi_3 = (c_3 \cdot dt) \cdot N_1 \cdot N_2$

Monte Carlo simulace

Gillespiho přímá metoda

1. inicializace $X(0)$
2. výpočet $\chi_i(X, c_i) \forall i \in \{1, \dots, m\}$ v aktuálním stavu X
3. výpočet $\chi(X, c) \equiv \sum_{i=1}^m \chi_i(X, c_i)$
4. simulace doby τ do následující události – sampuj $\tau \in \text{Exp}(\chi(X, c))$
5. $t := t + \tau$
6. výběr reakce R_i s pravděpodobností $\frac{\chi_i(X, c_i)}{\chi(X, c)}$
7. $X(t) := X^T + M(j)$
8. pokud $t < T_{max}$, iteruj (2)

Obsah

Modelování dynamiky chemických reakcí

Modelování dynamiky transkripční regulace

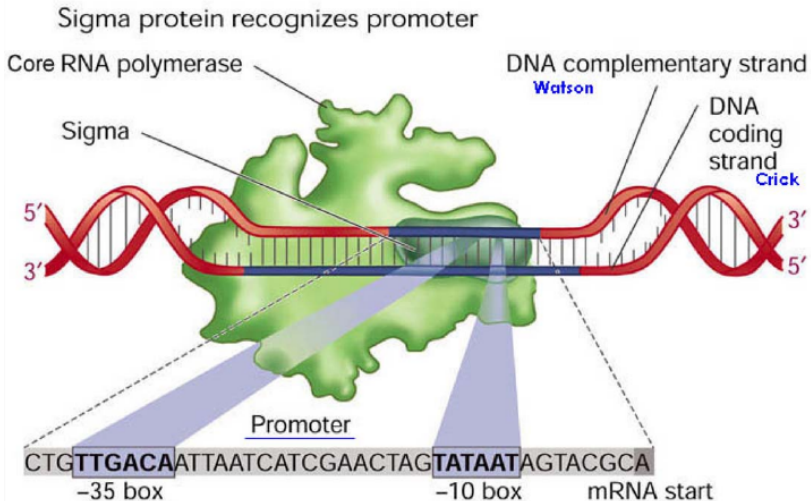
Varianty Gillespiho algoritmu

Aproximativní metody

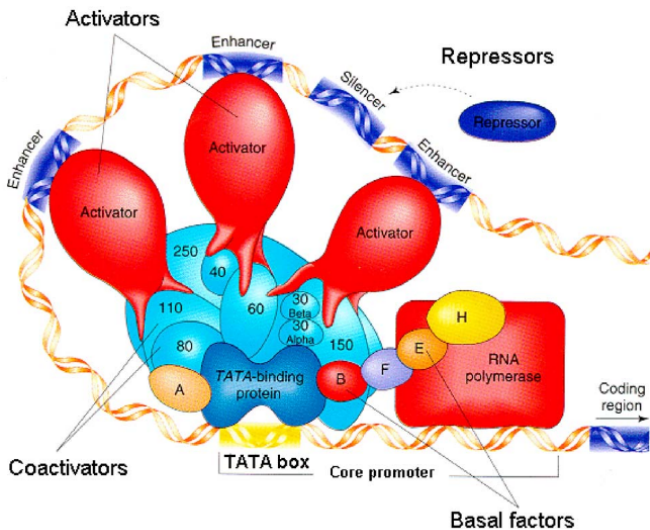
Deterministické metody

Hybridní metody

Interakce při expresi genů – prokaryota



Interakce při expresi genů – eukaryota



Interakce při expresi genů

1. vytvoření slabé vazby TF-DNA (nespecifické regiony DNA)
2. lineární pohyb TF po DNA (1D difúze)
3. disociace vazby TF-DNA
4. prostorový pohyb – “skok” (3D difúze)
5. (1-4) iterováno dokud nenalezena regulační sekvence DNA
6. afinita TFBS a kooperativní interakce transkripčního komplexu snižují vliv difúze a stabilizují funkční vazbu TF-DNA

Interakce při expresi genů

- interference při paralelním “skenování” DNA
 - velké množství molekul téhož TF
 - molekuly různých TF
- u eukaryot navíc komplikovaná lokální prostorová struktura DNA
- na úrovni buňky nutno uvažovat stochasticitu (husté prostředí v okolí DNA)

Interakce při expresi genů

- modelování na úrovni buňky
- 1D difúze uvažována v diskrétních krocích
⇒ 1 krok \sim 1 nukleotid za jednotku času
- stupně volnosti:
 - pohyb ve směru 5' – 3'
 - pohyb ve směru 3' – 5'
 - zachování pozice
- proteiny TF v buňce mají specifickou distribuci volné energie vůči vazbě s DNA
 - energie individuální molekuly fluktuuje vzhledem k náhodným kolizím s ostatními molekulami
 - fluktuace způsobuje dynamické změny afinity proteinů k DNA
 - 1D pohyb po DNA je stochastický proces

Interakce při expresi genů

- pravděpodobnost 1D pohybu lze charakterizovat exponenciálním rozložením vzhledem k rozdílu volných energií původní a cílové pozice:

$$P(m) \propto e^{-\frac{\Delta G}{RT}}$$

- $\Delta G = G' - G_0$... rozdíl volné Gibsovy energie iniciální (G_0) a cílové pozice (G')
- T ... absolutní teplota [K]
- R ... molární plynová konstanta [$J \cdot K^{-1} \cdot mol^{-1}$]

$$P(m_{3'}^{5'}) = \frac{e^{-\frac{\Delta G_{3'}^{5'}}{RT}}}{1 + e^{-\frac{\Delta G_3^{5'}}{RT}} + e^{-\frac{\Delta G_{5'}^{3'}}{RT}}}$$

$$P(m_{5'}^{3'}) = \frac{e^{-\frac{\Delta G_{5'}^{3'}}{RT}}}{1 + e^{-\frac{\Delta G_3^{5'}}{RT}} + e^{-\frac{\Delta G_{5'}^{3'}}{RT}}}$$

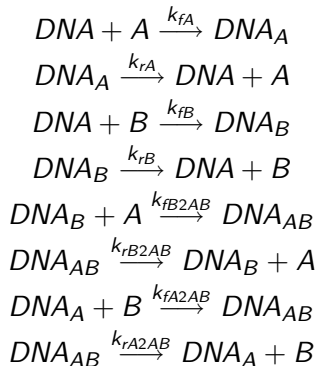
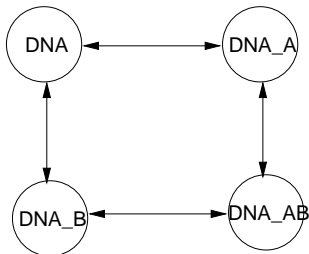
$$P(m_{zach}) = \frac{1}{1 + e^{-\frac{\Delta G_3^{5'}}{RT}} + e^{-\frac{\Delta G_{5'}^{3'}}{RT}}}$$

Modelování formace transkripčního komplexu

- uvažujeme dva TF A a B kooperativně aktivující transkripci genu
- zavedeme následující stavy regulačního regionu DNA:
 - DNA ... volný regulační region DNA
 - DNA_A ... navázán A , nikoliv B
 - DNA_B ... navázán B , nikoliv A
 - DNA_{AB} ... navázán A i B

Modelování formace transkripčního komplexu

- definujeme přechody mezi stavy:



- při kooperačním faktoru K_q lze uvažovat:
 $k_{fA2AB} \cdot k_{fA} = k_{fB2AB} \cdot k_{fB} = K_q \cdot k_{fA} \cdot k_{fB}$

Modelování formace transkripčního komplexu

- transkripce (tvorba mRNA) je proces o několi řádů pomalejší vzhledem k formaci transkripčního komplexu
- lze jej uvažovat jako nekonečně rychlý (tj. stabilní)
- okupaci promotoru pak charakterizujeme poměrem frekvence výskytu stavu DNA_{AB} vzhledem k sumě frekvencí všech možných stavů:

$$\alpha = \frac{DNA \cdot K_B \cdot K_{B2AB} \cdot B \cdot A}{DNA + DNA \cdot K_A \cdot A + DNA \cdot K_B \cdot B + DNA \cdot K_B \cdot K_{B2AB} \cdot B \cdot A}$$

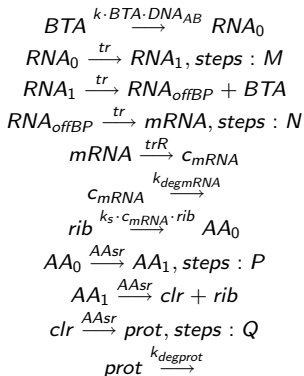
kde $K_B = \frac{k_{fB}}{k_{bB}}$, $K_A = \frac{k_{fA}}{k_{bA}}$, $K_{B2AB} = \frac{k_{fB2AB}}{k_{bB2AB}}$

- ekvivalentně lze psát:

$$\alpha = \frac{DNA \cdot K_A \cdot K_{A2AB} \cdot A \cdot B}{DNA + DNA \cdot K_A \cdot A + DNA \cdot K_B \cdot B + DNA \cdot K_A \cdot K_{A2AB} \cdot A \cdot B}$$

Modelování průběhu transkripce

- uvažujeme činnost komplexu RNA polymerázy
- předpokládáme, že všechny faktory důležité pro transkripční komplex jsou k dispozici v libovolné míře



Obsah

Modelování dynamiky chemických reakcí

Modelování dynamiky transkripční regulace

Varianty Gillespiho algoritmu

Aproximativní metody

Deterministické metody

Hybridní metody

Spojitéj Markovův řetězec

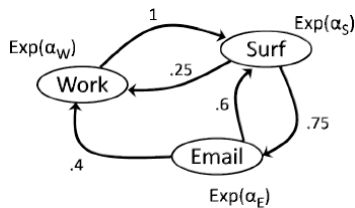
Spojitéj Markovův řetězec lze definovat jako přechodový graf $MC = \langle V, E, p \rangle$:

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

$$Q_{ij} = p(\langle v_i, v_j \rangle)$$

Spojité Markovův řetězec

Příklad



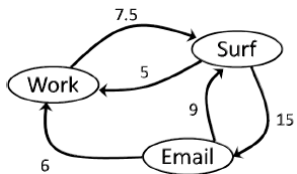
- průměrná čekací doba ve stavu *Surf* je 3 minuty, což je $\frac{3}{60} = \frac{1}{20}$ hod
 $\Rightarrow \alpha_S = 20$
- $\alpha_W = 7.5$
- $\alpha_E = 15$

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



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:

$$\Pr\{\min\{X_1, \dots, X_n\} > x\} = \Pr\{X_1 > x \cap X_2 > x \cap \dots \cap X_n > x\}$$

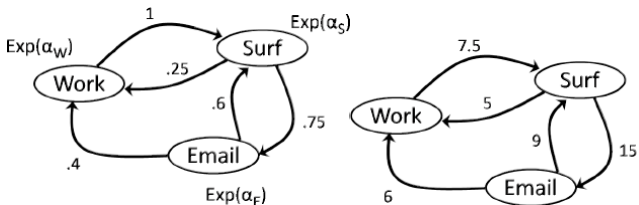
$$= \prod_{i=1}^n \Pr\{X_i > x\} = \prod_{i=1}^n e^{-\lambda_i x} = e^{-x \sum_{i=1}^n \lambda_i}$$

Pro parametr minimálního rozložení platí:

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

Spojité Markovův řetězec

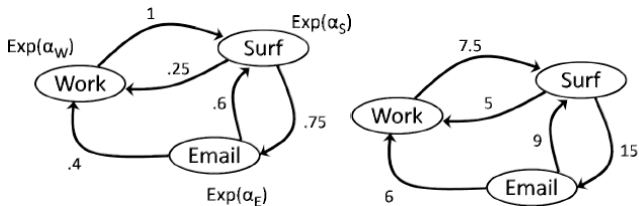
Převod na tradiční zápis



- rozhodovací procedura při opuštění stavu *Surf*
 - v okamžiku vstupu do stavu *Surf* se spustí stopky:
 - A_W měřící časový úsek s distribucí $Exp(\lambda_{a_W})$
 - A_E měřící časový úsek s distribucí $Exp(\lambda_{a_E})$
 - 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



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

Gillespiho přímá metoda (*direct*)

1. inicializace $X(0)$
2. výpočet $\chi_i(X, c_i) \forall i \in \{1, \dots, m\}$ v aktuálním stavu X
3. výpočet $\chi(X, c) \equiv \sum_{i=1}^m \chi_i(X, c_i)$
4. simulace doby τ do následující události – sampuj $\tau \in \text{Exp}(\chi(X, c))$
5. $t := t + \tau$
6. výběr reakce R_i s pravděpodobností $\frac{\chi_i(X, c_i)}{\chi(X, c)}$
7. $X(t) := X^T + M(j)$
8. pokud $t < T_{max}$, iteruj (2)

Varianta Gillespi: Metoda nejbližší reakce

1. inicializace $t := 0$, $X(t)$, $c = (c_1, \dots, c_m)$
2. inicializace $succs_times = \emptyset$
3. $\forall i \in \{1, \dots, m\}$:
 - výpočet $\chi_i(X, c_i)$ v aktuálním stavu $X(t)$
 - pro reakci R_i sampluj $t_i \in \text{Exp}(\chi_i(X, c_i))$
 - $succs_times := succs_times \cup \{t_i\}$
4. výpočet j , $t_j = \min(succs_times)$
5. $t := t + t_j$
6. $X(t) := X^T + M(j)$
7. pokud $t < T_{max}$, iteruj (2)

Přímá vs. metoda nejbližší reakce

- obě metody exaktně simulují CTMC
- přímá: v každém kroku simulace dvou náhodných čísel
- mnr: v lib. stavu i simulace $m(i)$ náhodných čísel, kde $m(i)$ je počet následníků i v CTMC
- přímá metoda efektivnější
- mnr však poskytuje filosoficky odlišný přístup, který lze dále akcelarovat

Inverzní rozložení

Uvažujme proměnnou U s rovnoměrným (uniformním) rozložením $U \sim U(0, 1)$ a nechť $F(\cdot)$ libovolná invertibilní kumulativní distribuční funkce. Pak rozložení $X = F^{-1}(U)$ je charakterizováno kumulativní distribuční funkcí $F(\cdot)$.

$$\begin{aligned} P(X \leq x) &= P(F^{-1}(U) \leq x) \\ &= P(U \leq F(x)) \\ &= F_U(F(x)) \\ &= F(x) \end{aligned}$$

Realizaci libovolného rozložení X charakterizovaného invertibilní kumulativní distribuční funkcí $F(x)$ lze simulovat pomocí uniformního rozložení U .

Souvislost uniformního a exponenciálního rozložení

Uvažujme proměnnou U s rovnoměrným (uniformním) rozložením $U \sim U(0, 1)$. Pro lib. $\lambda > 0$ má proměnná $X = -\frac{1}{\lambda} \ln(U)$ rozložení $X \sim \text{Exp}(\lambda)$.

Rozepíšeme-li distribuční a kumulativní funkci $f(x)$, $F(x)$:

$$f(x) = \lambda e^{-\lambda x} \qquad F(x) = 1 - e^{-\lambda x}$$

$F(x)$ je invertibilní:

$$F^{-1}(u) = -\frac{1}{\lambda} \ln(1 - u)$$

Triviálně platí $(1 - U) \sim U(0, 1)$ a tedy lze položit $x = -\frac{1}{\lambda} \ln(u)$.

Škálování exponenciálního rozložení

Uvažujme proměnnou X o rozložení $X \sim \text{Exp}(\lambda)$. Proměnná $Y = \alpha X$ má rozložení $Y \sim \text{Exp}(\frac{\lambda}{\alpha})$.

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(\alpha X \leq y) \\ &= P(X \leq \frac{y}{\alpha}) \\ &= F_X(\frac{y}{\alpha}) \end{aligned}$$

$$\begin{aligned} F_X(x) &= 1 - e^{-\lambda x} \\ \Rightarrow F_Y(y) &= F_X(\frac{y}{\alpha}) = 1 - e^{-\frac{\lambda}{\alpha} y} \\ \Rightarrow Y &\sim \text{Exp}(\frac{\lambda}{\alpha}) \end{aligned}$$

Algoritmus Gibson-Bruck

1. inicializace $t := 0$, $X(t)$, $c = (c_1, \dots, c_m)$
2. pro každé $i \in \{1, \dots, m\}$:
 - výpočet $\chi_i(X, c_i)$ ve stavu $X(0)$
 - pro reakci R_i sampluj $t_i \in \text{Exp}(\chi_i(X, c_i))$
3. inicializace $\text{succs_times} = \emptyset$
4. pro každé $i \in \{1, \dots, m\}$:
 - $\text{succs_times} := \text{succs_times} \cup \{t_i\}$
5. výpočet j , $t_j = \min(\text{succs_times})$
6. $t := t_j$
7. $X(t) := X^T + M(j)$
8. update $\chi_j(X, c_j)$ v novém stavu $X(t)$
9. sampluj $t_j := t + \text{Exp}(\chi_j(X, c_j))$
10. pro každé $i \in \{1, \dots, m\}$ splňující $\text{dep}(R_i, R_j)$:
 - ulož $\chi_i^{\text{last}} := \chi_i(X, c_i)$
 - update $\chi_i(X, c_i)$
 - $t_i := t + \frac{\chi_i^{\text{last}}(X, c_i)}{\chi_i(X, c_i)} (t_i - t)$
11. pokud $t < T_{\max}$, iteruj (3)

Algoritmus Gibson-Bruck

- relativní čas (do nejbližší události) nahrazen časem absolutním (čas nejbližší události)
 - není nutno generovat nové časy pro všechny reakce
 - ale pouze pro reakce závislé na provedené reakci
 - umožněno díky markovovské vlastnosti
- nové časy závislých reakcí se nesamplují, ale vypočítávají z předchozích
 - předchozí časy závislých reakcí jsou použity pro výpočet nových časů
 - škálování podmíněné požadavkem $t_i > t$

Gibson-Bruck vs. Přímá metoda

- G-B: pouze jedna simulace času v každém kroku
- klíčovou vlastností G-B je *selektivní update* $\chi_i(X, c_i)$ – pouze u relevantních (závislých) reakcí
- selektivní přímý výpočet $\chi_i(X, c_i)$ a $\chi(X, c)$ může zrychlit i přímou metodu
- rychlost obou metod výrazně závisí na implementaci
 - v G-B je obecně více operací
 - implementaci výpočtu minimálního reakčního času lze realizovat pomocí indexované prioritní fronty

Obsah

Modelování dynamiky chemických reakcí

Modelování dynamiky transkripční regulace

Varianty Gillespiho algoritmu

Aproximativní metody

Deterministické metody

Hybridní metody

Poissonův proces

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

Pak platí:

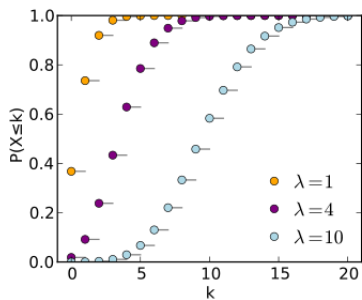
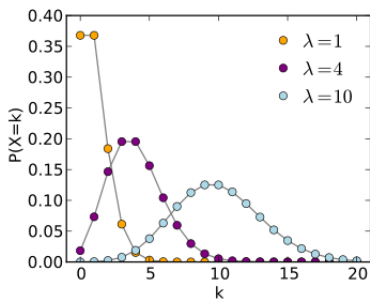
$$X \sim Po(\lambda t)$$

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:

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

Souvislost Poissonova a exponenciálního 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 aproximativní algoritmy

- exaktní metody simulují spojitý markovův řetězec
- lze aproximovat diskretizací času
- rozdělení časové osy na diskrétní intervaly
- počet událostí v lib. časovém intervalu charakterizován $Po(\lambda)$
- lze tedy samplovat počet provedení reakcí daného typu v daném časovém intervalu Δt

Obecné schema aproximativního algoritmu

1. inicializace $t := 0$, $X(t)$, $c = (c_1, \dots, c_m)$
2. pro každé $i \in \{1, \dots, m\}$:
 - výpočet $\chi_i(X, c_i)$ v aktuálním stavu $X(t)$
 - simulace počtu reakcí R_i v intervalu Δt :
 - $\#R_i \in Po(\chi_i(X, c_i)\Delta t)$
3. update $X := X + M \cdot \langle \#R_1, \#R_2, \dots, \#R_m \rangle^T$
4. update $t := t + \Delta t$
5. pokud $t < T_{max}$ iteruj (2)

Problémy aproximativního algoritmu

- volba velikosti Δt
- chceme rychlý, ale přitom dostatečně přesný výpočet
- pro jednu simulaci nemusí být konstantní Δt vyhovující
- předpokládáme konstantní frekvence reakcí po celé Δt
- možnost definovat variabilní Δt v závislosti na aktuálním stavu $X(t)$ a škále jednotlivých reakčních konstant c

Gillespiho τ -leap metoda

- aproximativní algoritmus s variabilní Δt (τ)
- τ uvžováno co největší při garanci požadované přesnosti
- přesnost určena mírou tolerance uvažování konstantní frekvence reakcí v rámci intervalu τ
- charakterizováno magnitudou změny frekvencí reakcí v průběhu τ
- τ voleno tak, aby míra změny všech reakčních frekvencí byla minimalizována

Gillespiho τ -leap metoda

- post-leap (přechod od $X(t)$ k $X' = X(t + \tau)$):
 - $\forall i \leq m$ ověření velikosti

$$|\chi_i(X', c_i) - \chi_i(X, c_i)|$$

- pokud velikosti nejsou dostatečně malé, přepočítej s menším τ
- problém: směřuje k malým změnám stavů

Gillespiho τ -leap metoda

- pre-leap (odhad přechodu $X(t)$ k $X' = X(t + \tau)$):
 - výpočet očekávaného nového stavu $E(X')$ (v $t' := t + \tau$)

$$E(X') = X + E(r)M$$

kde $E(r)$ je vektor očekávaných frekvencí reakcí:

$$E(r)_i = \chi_i(X, c_i)\tau$$

- parametr přesnosti ϵ :

$$\forall i \in \{1, \dots, m\}. |\chi_i(X', c_i) - \chi_i(X, c_i)| \leq \epsilon \cdot \chi(X, c)$$

- lze kombinovat s exaktní metodou (v kroku kde např.

$$\tau < \frac{2}{\chi(X, c)}$$

- nepřesnost vzniká při nelineární dynamice reakčních frekvencí

Obsah

Modelování dynamiky chemických reakcí

Modelování dynamiky transkripční regulace

Varianty Gillespiho algoritmu

Aproximativní metody

Deterministické metody

Hybridní metody

Deterministický model reakční dynamiky

- uvažujme systém n substancí $S = \{S_1, \dots, S_n\}$ provázaných m reakcemi $R = \{R_1, \dots, R_m\}$
- uvažujeme pouze reakce 1. a 2. řádu
- systém zapisujeme pomocí stoichiometrické matice M rozměru $n \times m$:

$$M_{ij} = \begin{cases} -K, \text{ je-li } K \cdot S_i \text{ reaktantem } R_j \\ K, \text{ je-li } K \cdot S_i \text{ produktem } R_j \end{cases}$$

Deterministický model reakční dynamiky

- uvažovány vysoké molární koncentrace látek v buňce
- koncentraci substance S_i v čase t budeme značit $[S_i](t)$
- systém v čase t charakterizujeme vektorem:

$$X(t) = \langle [S_1](t), \dots, [S_n](t) \rangle$$

- vývoj X v čase:

$$\frac{dX}{dt} = f(X)$$

- průměrné chování lze charakterizovat exponenciální funkcí

Převod počtu molekul na molární koncentraci

- molární koncentrace $[M]$:

$$m = \frac{n}{V}$$

kde n je množství látky $[mol]$, V je objem roztoku $[l]$

- vyjadřuje se pomocí Avogadrovy konstanty (počet částic v 1 molu):

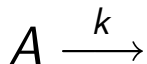
$$m = \frac{N}{N_A \cdot V}$$

kde N_A Avogadrova konstanta $[mol^{-1}]$, V objem roztoku $[l]$ a N je počet molekul.

- převodní faktor $\gamma = N_A \cdot V$:

$$N = m \cdot \gamma$$

Deterministický model reakční dynamiky



- předpokládejme nádobu jednotkového objemu obsahující v čase t látku A v molárním množství $[A]$ [mol]
- kolik množství látky A “odteče” za jednotku času?
 - hodnota přímo úměrná hodnotě $[A]$ v daném okamžiku

$$-\frac{d[A](t)}{dt} = k \cdot [A](t)$$

- koeficient úměrnosti je konstanta k [s^{-1}]
tzv. *reakční konstanta (koeficient)*
 - determinuje rychlost reakce rozpadu (“odtoku”)

Deterministický model reakční dynamiky

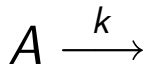
$$\frac{[A](t)}{dt} = k \cdot [A](t)$$

- jaká funkce má stejný tvar jako její derivace?
 - $f(t) = 1 + t + t^2/2! + t^3/3! + t^4/4! + \dots$

$$f(t) = e^t$$

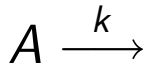
- platí

$$\frac{de^t}{dt} = e^t$$

Deterministický model reakční dynamiky

$$-\frac{d[A](t)}{dt} = k \cdot [A](t)$$

Deterministický model reakční dynamiky



$$-\frac{d[A](t)}{dt} = k \cdot [A](t) \Leftrightarrow [A](t) = [A](0) \cdot e^{-kt}$$

- lineární dif. rce 1. řádu
- jednoznačné řešení
- numericky aproximovatelné

Deterministický model reakční dynamiky

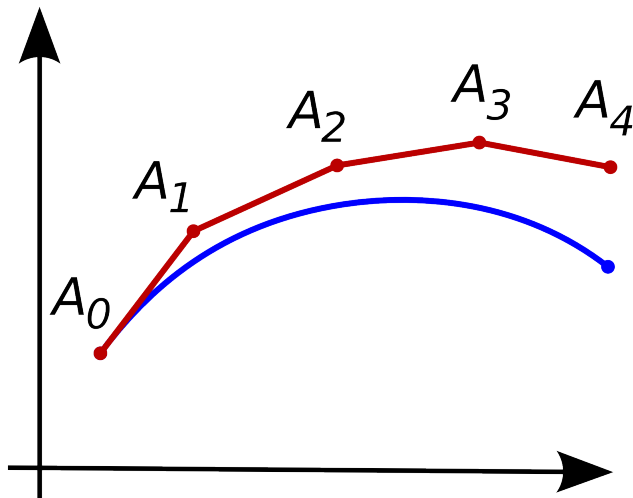
- spojité chování: při přechodu $X(t) \rightarrow X(t + dt)$ jsou updatovány všechny složky X (souběžný spojitý tok reakcí)
- časová informace o běhu reakce R_i promítnuta do okamžitého reakčního toku $\varrho_{R_i}(t)$

R_i	$\emptyset \rightarrow *$	$\varrho_{R_i}(t) = k_i \cdot dt$
R_i	$S_j \rightarrow *$	$\varrho_{R_i}(t) = (k_i \cdot dt) \cdot [S_j](t)$
R_i	$S_p + S_q \rightarrow *$	$\varrho_{R_i}(t) = (k_i \cdot dt) \cdot [S_p](t) \cdot [S_q](t)$
R_i	$2S_j \rightarrow *$	$\varrho_{R_i}(t) = (k_i \cdot dt) \cdot [S_j]^2$

Stochastický vs. deterministický model

Pro reakci $R_i \in R$ definujeme převodní vztah mezi stoch. frekvencí c_i a det. kinetickou konstantou k_i :

typ reakce R_i	$c_i \rightarrow k_i$
$S_j \rightarrow *$	$k_i = c_i$
$S_p + S_q \rightarrow *$	$k_i = c_i \cdot \gamma$
$2S_j \rightarrow *$	$k_i = \frac{c_i \cdot \gamma}{2}$

Eulerova metoda

Eulerova metoda

- aproximativní řešení $y(t)$ (Euler):

$$\begin{aligned}y'(t) &= f(t, y(t)) \\ y(0) &= y_0\end{aligned}$$

- přesné řešení $\varphi(t)$:

$$\begin{aligned}\varphi'(t) &= f(t, \varphi(t)) \\ \varphi(0) &= y_0\end{aligned}$$

- pro lib. $n \geq 0$, $t_n = n\Delta t$:

$$y_n \approx \varphi(t_n)$$

Eulerova metoda

Pro exaktní řešení $\varphi(t)$ platí:

$$\begin{aligned}\varphi(t_{n+1}) &= \varphi(t_n) + \int_{t_n}^{t_{n+1}} \varphi'(t) dt \\ &= \varphi(t_n) + \int_{t_n}^{t_{n+1}} f(t, \varphi(t)) dt\end{aligned}$$

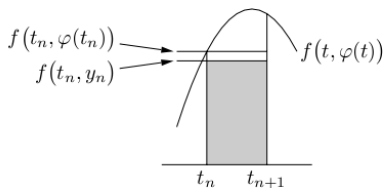
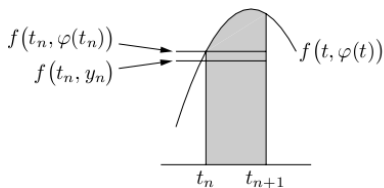
Schema numerické aproximace:

$$y_{n+1} = y_n + \sigma$$

kde

$$\sigma \approx \int_{t_n}^{t_{n+1}} f(t, \varphi(t)) dt$$

Eulerova metoda I

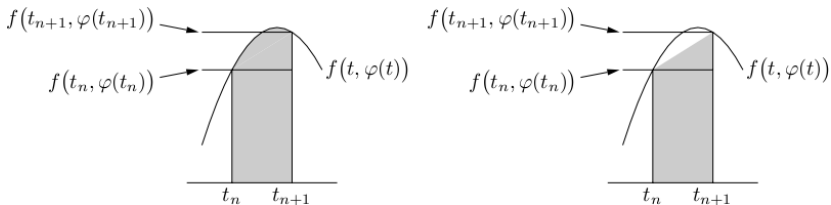


$$\frac{dy}{dt} = f(t, y)$$

$$y_{n+1} = y_n + \Delta t \cdot f(t_n, y_n)$$

1. init $t_0, y_0, \Delta t, n$;
2. for j from 1 to n do
 - 2.1 $m := f(t_0, y_0)$;
 - 2.2 $y_1 := y_0 + \Delta t m$;
 - 2.3 $t_1 := t_0 + \Delta t$;
 - 2.4 $t_0 := t_1$;
 - 2.5 $y_0 := y_1$;
3. end

Eulerova metoda II



- aproximace σ (obsah pod křivkou) lichoběžníkem o obsahu:

$$\frac{\Delta t}{2} \cdot [f(t_n, \varphi(t_n)) + f(t_{n+1}, \varphi(t_{n+1}))]$$

- pro $\varphi(t_{n+1})$ nutno předpočítat aproximaci (známe již $y_n \approx \varphi(t_n)$):

$$\varphi(t_{n+1}) \approx \varphi(t_n) + \varphi'(t_n)\Delta t \approx y_n + f(t_n, y_n)\Delta t$$

- aproximace $\sigma \approx \frac{1}{2}[f(t_n, y_n) + f(t_{n+1}, y_n + f(t_n, y_n)\Delta t)]\Delta t$
- celkem dostáváme:

$$y(t_{n+1}) \approx y_{n+1} = y_n + \frac{1}{2}[f(t_n, y_n) + f(t_{n+1}, y_n + f(t_n, y_n)\Delta t)]\Delta t$$

Runge-Kutta

- σ upřesněno Simpsonovým pravidlem pro aproximaci obsahu pod parabolou:

$$\sigma = \int_{t_n}^{t_{n+1}} \approx \frac{\Delta t}{6} [f(t_n, \varphi(t_n)) + 4f(t_n + \frac{\Delta t}{2}, \varphi(t_n + \frac{\Delta t}{2})) + f(t_n + \Delta t, \varphi(t_n + \Delta t))]$$

- nutno aproximovat $\varphi(t_n + \frac{\Delta t}{2})$ a $\varphi(t_n + \Delta t)$
- kroky algoritmu:

$$\begin{aligned} k_{n,1} &= f(t_n, y_n) \\ k_{n,2} &= f(t_n + \frac{\Delta t}{2}, y_n + \frac{h}{2}k_{n,1}) \\ k_{n,3} &= f(t_n + \frac{\Delta t}{2}, y_n + \frac{h}{2}k_{n,2}) \\ k_{n,4} &= f(t_n + \Delta t, y_n + hk_{n,3}) \\ y_{n+1} &= y_n + \frac{h}{6}[k_{n,1} + 2k_{n,2} + 2k_{n,3} + k_{n,4}] \end{aligned}$$

Porovnání metod

- diferenciální rovnice:

$$y'(t) = y - 2t$$

$$y(0) = 3$$

- exaktní řešení:

$$y(t) = 2 + 2t + e^t$$

- srovnání simulací:

steps	Euler		Improved Euler		Runge-Kutta	
	error	#evals	error	#evals	error	#evals
5	2.3×10^{-1}	5	1.6×10^{-2}	10	3.1×10^{-5}	20
50	2.7×10^{-2}	50	1.8×10^{-4}	100	3.6×10^{-9}	200
500	2.7×10^{-3}	500	1.8×10^{-6}	1000	3.6×10^{-13}	2000

Přesnost numerické simulace

Uvažujme systém:

$$y'(t) = f(t, y), y(0) = y_0$$

Označme exaktní řešení $\phi(t)$:

$$\phi'(t) = f(t, \phi(t))$$

Uvažujme Eulerovu metodu I s krokem $\Delta t = h$:

$$y_{n+1} = y_n + hf(t_n, y_n)$$

(Lokální) chyba numerické simulace:

$$\rho_{n+1} = \phi(t_{n+1}) - y_{n+1}$$

Přesnost numerické simulace

Předpokládáme-li $y_n = \phi(t_n)$, platí $y_{n+1} = \phi(t_n) + hf(t_n, \phi(t_n))$ a tedy z exaktnosti řešení $\phi(t)$ pro $\phi'(t_n) = f(t_n, \phi(t_n))$ dostáváme:

$$y_{n+1} = \phi(t_n) + h\phi'(t_n)$$

Uvažujme Taylorův rozvoj pro $\phi(t_{n+1}) = \phi(t_n + h)$:

$$\phi(t_n + h) = \phi(t_n) + \phi'(t_n)h + \frac{1}{2}\phi''(t_n)h^2 + \frac{1}{3!}\phi'''(t_n)h^3 + \dots$$

$$\begin{aligned} \rho_{n+1} &= \phi(t_{n+1}) - y_{n+1} \\ &= [\phi(t_n) + \phi'(t_n)h + \frac{1}{2}\phi''(t_n)h^2 + \frac{1}{3!}\phi'''(t_n)h^3 + \dots] - [\phi(t_n) + h\phi'(t_n)] \\ &= \frac{1}{2}\phi''(t_n)h^2 + \frac{1}{3!}\phi'''(t_n)h^3 + \dots \end{aligned}$$

Chyba aproximována (pro konstantu K a délku kroku h):

$$\rho_{n+1} \approx Kh^2 + O(h^3)$$

Adaptivní metody

- chceme, aby chyba v každém kroku nebyla větší než ϵ
- řešení pomocí prediktoru chyby
- při n -tém kroku (výpočet y_{n+1} z y_n):
 1. použití dvou různých algoritmů
(dva výsledky pro y_{n+1} — A_1 a A_2)
 2. aproximace lokální chyby $\rho_{n+1} \approx |A_1 - A_2|$
 3. pokud $\frac{\rho_{n+1}}{h} < \epsilon$, pak $y_{n+1} = A_2$
 4. jinak iteruj (1) pro $h' < h$

Adaptivní Eulerova metoda

Uvažujme systém:

$$y'(t) = f(t, y), y(0) = y_0$$

Označme $\phi(t)$ exaktní řešení $y' = f(t, y)$ t.ž. $\phi(t_n) = y_n$.

Pro A_1 uvažujme Eulerovu metodu I:

$$A_1 = y_n + hf(t_n, y_n)$$

$$A_1 = \phi(t_n + h) + Kh^2 + O(h^3)$$

Adaptivní Eulerova metoda

Pro A_2 uvažujme Eulerovu metodu 2-step (provedení dvou kroků, každý s $\Delta t = \frac{h}{2}$):

$$A_2 = y_n + \frac{h}{2}f(t_n, y_n) + \frac{h}{2}f\left(t_n + \frac{h}{2}, y_n + \frac{h}{2}f(t_n, y_n)\right)$$

Označme $y_{mid} = y_n + \frac{h}{2}f(t_n, y_n)$ a $A_2 = y_{mid} + \frac{h}{2}f\left(t_n + \frac{h}{2}, y_{mid}\right)$.

$$\rho_{mid} = K\left(\frac{h}{2}\right)^2 + O(h^3) \quad \rho_{2nd} = K\left(\frac{h}{2}\right)^2 + O(h^3)$$

Konstanta K shodná v obou případech (neplatí pro člen $O(h^3)$).

Lze tedy psát (dle Taylorovy věty):

$$A_2 = \phi(t_n + h) + \frac{1}{2}Kh^2 + O(h^3)$$

Adaptivní Eulerova metoda

$$\begin{aligned} A_1 - A_2 &= \phi(t_n + h) + Kh^2 + O(h^3) - \phi(t_n + h) - \frac{1}{2}Kh^2 - O(h^3) \\ &= \frac{1}{2}Kh^2 + O(h^3) \end{aligned}$$

Lze tedy aproximovat $\frac{1}{2}Kh^2 \approx A_1 - A_2$ a označit

$$\rho = \frac{|A_1 - A_2|}{h} \approx \frac{1}{2}Kh.$$

Pokud $\rho > \epsilon$, postup opakujeme pro h' t.ž. $\frac{1}{2}|K|h' \approx \frac{\rho}{h}h' < \epsilon$.

Např.

$$h' = .9 \frac{\epsilon}{\rho} h$$

Pokud $\rho < \epsilon$, nastavíme $y_{n+1} = 2A_2 - A_1$.

Další adaptivní metody

- Fehlbergova metoda
 - v každém kroku 3 odhady:

$$f_1 = f(t_n, y_n)$$

$$f_2 = f(t_n + h, y_n + hf_1)$$

$$f_3 = f(t_n + \frac{h}{2}, y_n + \frac{h}{4}[f_1 + f_2])$$

$$A_1 = y_n + \frac{h}{2}[f_1 + f_2]$$

$$A_2 = y_n + \frac{h}{6}[f_1 + f_2 + 4f_3]$$

- aproximace $\rho = \frac{|A_1 - A_2|}{h} \approx Kh^2$
 - update faktor $h' = .9\sqrt{\frac{\epsilon}{\rho}}h$
- Kutta-Merson, LSODE, ...

Obsah

Modelování dynamiky chemických reakcí

Modelování dynamiky transkripční regulace

Varianty Gillespiho algoritmu

Aproximativní metody

Deterministické metody

Hybridní metody

Kombinace deterministické a stochastické simulace

Předpokládáme model s kombinací spojitých a diskretních proměnných.

1. inicializuj systém, $t := 0$
2. vypočti $\chi_i(X, c_i)$ v čase t
3. na základě výsledku nastav Δt det. simulace
4. vypočti trajektorie spojitých proměnných v intervalu $[t, t + \Delta t]$
5. na základě výsledku vypočti $\chi_i(X, c_i)$ a rozhodni provedení případné diskretní události (reakce)
6. pokud žádná diskretní událost, nastav $t := t + \Delta t$ a updatuj pro tento bod spojitě proměnné
7. pokud diskretní událost
 - najdi (nejbližší) reakci R_j a čas t_j jejího provedení
 - $t := t_1$
 - updatuj spojitě proměnné k časovému bodu t_1
 - updatuj diskretní proměnné relevantní reakci R_j
8. pokud $t < T_{max}$, iteruj (2)

Multi-step metody stochastické simulace

Metoda Puchalka and Kierzek (2004) (STOCKS2). Předpokládá rozlišení “rychlých” a “pomalých” reakcí.

1. inicializuj systém, $t := 0$
2. výpočet $\chi_i(X, c_i)$ rychlých reakcí a nastav τ pro τ -leaping
3. pro pomalé reakce předpokládej konst. $\chi_i(X, c_i)$, rozhodni provedení pomalé reakce v intervalu $[t, t + \Delta t]$
4. pokud žádná pomalá reakce k provedení, vypočti τ -leap update na rychlých reakcích na $t := t + \tau$
5. pokud pomalá reakce k provedení
 - identifikuj tuto reakci R_j a čas t_j pro její provedení
 - update $t := t_j$
 - proved τ -leap update rychlých reakcí na t_j
 - proved update pomalých reakcí vzhledem k provedení R_j
6. pokud $t < T_{max}$, iteruj (2)