

Algoritmo de Gillespie: Stochastic Simulation Algorithm (SSA)

Máster en Bioinformática y Biología Computacional - UAM

Saúl Ares

April 24, 2025

1. Motivación

- Sistemas moleculares con pocas moléculas presentan ruido intrínseco.
- Modelos deterministas (Ecuaciones Diferenciales) promedian este ruido.
- Necesidad de un método para simular dinámicas discretas y aleatorias.

2. Marco de Modelado Discreto

- Estado: vector de cuentas $\mathbf{N}(t) = (N_1, \dots, N_m)$.
- Reacción R_k caracterizada por:
 - Vector estequiométrico ν_k (cambios en \mathbf{N}).
 - Propensión $a_k(\mathbf{N})$: tasa de ocurrencia.
- Eventos discretos: $\mathbf{N} \rightarrow \mathbf{N} + \nu_k$ con probabilidad $a_k dt$.

3. Ejemplo de Red de Reacciones

Consideremos dos reacciones:



Propensiones:

$$a_1(N) = k_1 N_A N_B, \quad a_2(N) = k_2 N_C.$$

4. Probabilidad de reacción

Construimos el marco de modelado estocástico bajo las siguientes suposiciones para incrementos de tiempo muy pequeños dt :

- En cualquier intervalo de longitud dt puede ocurrir a lo sumo un único evento de reacción.
- La probabilidad de que la reacción R_k ocurra en el intervalo $[t, t + dt]$ es

$$\text{probabilidad}(R_k \text{ en } [t, t + dt]) = a_k(N(t)) dt.$$

Bajo estas hipótesis, la probabilidad de que no ocurra ninguna reacción en $[t, t + dt]$ es

$$1 - \sum_k a_k(N(t)) dt,$$

donde la suma recorre todas las reacciones del sistema.

5. Ecuación Maestra Química

Para la distribución $P(\mathbf{N}, t)$ de que en el instante t el sistema esté en el estado \mathbf{N} :

$$\frac{d}{dt}P(\mathbf{N}, t) = \sum_k a_k(\mathbf{N} - \nu_k) P(\mathbf{N} - \nu_k, t) - \sum_k a_k(\mathbf{N}) P(\mathbf{N}, t).$$

- Primer término: flujo de probabilidad hacia \mathbf{N} .
- Segundo término: flujo de salida desde \mathbf{N} .
- Sistema infinito de E.D.; generalmente intratable analíticamente.

6. Ejemplo: Producción y Degradación

Reacciones:



Propensiones: $a_1 = k_1$, $a_2 = k_2 N_A$. Ecuación maestra para $P(n, t) = P(N_A = n, t)$:

$$P(0, t + dt) = P(0, t)(1 - k_1 dt) + P(1, t)k_2 dt,$$

$$P(n, t + dt) = P(n, t)(1 - (k_1 + nk_2)dt) + \\ P(n - 1, t)k_1 dt + P(n + 1, t)(n + 1)k_2 dt.$$

7. Gillespie: Determinación de la siguiente reacción

La probabilidad de que la reacción R_i sea la siguiente es proporcional a su propensión a_i . Ejemplo: sea $a_0 = a_1 + a_2 + a_3$. Entonces:

$$P(R = R_i) = \frac{a_i}{a_0}, \quad \sum_{i=1}^3 P(R = R_i) = 1.$$

Para muestrear R :

- 1 Genere $u \sim U(0, 1)$.
- 2 Compare u con los puntos de corte

$$\frac{a_1}{a_0}, \quad \frac{a_1 + a_2}{a_0}, \quad 1.$$

- 3 Asigne

$$R = \begin{cases} R_1, & 0 \leq u \leq \frac{a_1}{a_0}, \\ R_2, & \frac{a_1}{a_0} < u \leq \frac{a_1 + a_2}{a_0}, \\ R_3, & (a_1 + a_2)/a_0 < u \leq 1. \end{cases}$$

8. Gillespie: Determinación del tiempo hasta la siguiente reacción

La variable aleatoria T representa el tiempo transcurrido entre dos eventos sucesivos.

- T puede tomar cualquier valor real ≥ 0 , por lo que la probabilidad puntual $P(T = t)$ es infinitesimal.
- En lugar de usar probabilidades puntuales, muestreamos T directamente desde su función de distribución acumulada:

$$F_T(t) = P(0 \leq T \leq t) = 1 - e^{-at},$$

donde

$$a = \sum_k a_k(\mathbf{N}(t))$$

es la suma de todas las propensiones del sistema.

- Por tanto, T sigue una distribución exponencial:

$$T \sim \text{Exp}(a).$$

9. Variables Aleatorias en SSA

- Tiempo al siguiente evento τ :

$$P(\tau > t) = e^{-a_0 t}, \quad a_0 = \sum_k a_k.$$

- Selección de reacción R_j :

$$P(R_j) = \frac{a_j}{a_0}, \quad \sum_j P(R_j) = 1.$$

10. Algoritmo de Gillespie (SSA)

- 1 Inicializar $t \leftarrow 0$, $\mathbf{N} \leftarrow \mathbf{N}_0$.
- 2 Calcular $a_k(\mathbf{N})$, $a_0 = \sum_k a_k$.
- 3 Muestrear $\tau \sim \text{Exp}(a_0)$.
- 4 Elegir reacción j con $P(R_j) = a_j/a_0$.
- 5 Actualizar: $t \leftarrow t + \tau$, $\mathbf{N} \leftarrow \mathbf{N} + \boldsymbol{\nu}_j$.
- 6 Registrar y repetir hasta $t \geq t_{\max}$.

11. Conclusiones y Lecturas

- SSA captura fluctuaciones intrínsecas en sistemas con pocas moléculas.
- Complementa métodos deterministas y ecuaciones maestras.
- Extensiones: método de primera reacción, aproximaciones de ruido-lineal.