

APLICACIONES REALES DE MODELOS BIOINSPIRADOS

Tema 3: Estudio de sistemas biomoleculares: Aproximación clásica

David Orellana Martín
Mario de J. Pérez Jiménez

Grupo de investigación en Computación Natural
Dpto. Ciencias de la Computación e Inteligencia Artificial
Universidad de Sevilla

Máster Universitario en Lógica, Computación e Inteligencia Artificial

Curso 2023-2024



Objetivos

- Modelos estocásticos de sistemas biomoleculares basados en EDOs.
 - ★ Limitaciones.
 - ★ Aproximación continua mediante sistemas de EDOs.
- Algoritmo de Gillespie.
 - ★ Ventajas y desventajas.

Modelización de sistemas biomoleculares

Requisitos para un **buen modelo** de sistemas biomoleculares ¹ :

- ♣ **Relevancia.**
- ♣ **Comprensibilidad.**
- ♣ **Extensibilidad.**
- ♣ **Tratabilidad/Computabilidad.**

¹A. Regev, E. Shapiro. The π -calculus as an abstraction for biomolecular systems. In G. Ciobanu and Gr. Rozenberg, editors, *Modelling in Molecular Biology*, Springer Berlin, 2004, pp. 219–266.

Modelos estocásticos

- Para cada estado existe un conjunto de posibles “estados siguientes”: a ellos se pueden acceder a través de una **distribución de probabilidad**.
- A partir de un estado inicial que describe un cierto escenario, existen posibles evoluciones diferentes.
 - * La información que proporciona el modelo ante esos escenarios, requiere la realización un **número elevado de ejecuciones/simulaciones**.
- Los “resultados” de los modelos estocásticos serán los valores **promedios** de los obtenidos en las diferentes simulaciones. Su validez se basa en:
 - * El **teorema de los grandes números**: si el número de simulaciones es “grande”, el valor promedio de una variable tenderá a su “valor real”.
 - * El **teorema central del límite**: si el número de variables aleatorias es “grande”, entonces tenderá a una distribución normal.

Modelos estocásticos clásicos de sistemas biomoleculares

El carácter **aleatorio** de los procesos biomoleculares se captura mediante el uso de estrategias **estocásticas** o **probabilísticas**.

- * A nivel microscópico, las interacciones entre moléculas siguen la **ley \sqrt{n}** :
 - Los niveles de fluctuación del sistema son inversamente proporcionales a la raíz cuadrada del número de moléculas.
- * Los sistemas con un bajo número de moléculas muestran altas fluctuaciones.
 - Las aproximaciones clásicas deterministas y/o continuas para la modelización de sistemas biomoleculares, sólo son adecuadas cuando el número de moléculas que intervienen es “elevado”.

Estudiaremos procesos biomoleculares en los que cada **reacción química** involucra, a lo sumo, a **dos moléculas**.

- Las reacciones con tres reactantes son raras y se pueden describir como una sucesión de reacciones químicas con dos reactantes.

Modelos estocásticos clásicos de sistemas biomoleculares

Primer paso para la descripción estocástica de procesos biomoleculares:

- * Conjunto de **variables de estado**: la evolución del sistema dependerá sólo de los valores del estado actual.

Supuesto inicial: un **medio/volumen fijo homogéneo** en equilibrio térmico (a temperatura constante).

- * Sistema de moléculas de n especies químicas: $\{s_1, \dots, s_n\}$.
- * Las moléculas interactúan a través de q reacciones químicas: $\{r_1, \dots, r_q\}$.

Estado del sistema en un instante t : $\mathbf{X}(t) = (X_1(t), \dots, X_n(t))$.

- * $X_i(t)$: número de moléculas del tipo s_i en el instante t (**concentración**).

Objetivo: Evolución del vector de estado $\mathbf{X}(t)$ a partir de un estado inicial $\mathbf{X}(t_0)$.

Reacciones químicas en un sistema biomolecular

Cada reacción química r_j , $1 \leq j \leq q$, está caracterizada por:

- * Un **vector de cambio de estado** $\mathbf{v}_j = (v_{1j}, \dots, v_{nj})$:
 - v_{ij} : cambio en la población molecular de la especie s_i que produce la ejecución de la reacción r_j .
 - La matriz (v_{ij}) recibe el nombre de **matriz estequiométrica**.
- * Una **propensidad** $p_j(\mathbf{X}(t))$:
 - $p_j(\mathbf{X}(t))dt$: **probabilidad** de que r_j se ejecute al estado $\mathbf{X}(t)$ en el intervalo de tiempo $[t, t + dt)$.

Cálculo de las propensidades asociadas a las reacciones químicas

- * Cada reacción química r_j tiene asociada una **constante cinética** k_j :
 - Depende de las propiedades físicas de las moléculas involucradas y de otros parámetros físicos (se obtiene experimentalmente y **NO depende del tiempo**).
- * A partir de la constante cinética k_j se halla una **constante estocástica** c_j así:

* Si $r_j \equiv s_i \rightarrow \text{producto}$, entonces $c_j = k_j$.

$$* \text{ Si } r_j \equiv s_{i_1} + s_{i_2} \rightarrow \text{producto, entonces } \begin{cases} c_j = \frac{k_j}{V} & \text{si } i_1 \neq i_2 \\ c_j = \frac{2 \cdot k_j}{V} & \text{si } i_1 = i_2 \end{cases}$$

siendo V el volumen constante del medio.

- * A partir de la constante estocástica c_j se halla la **propensidad** $p_j(\mathbf{X}(t))$ así:

* Si $r_j \equiv s_i \rightarrow \text{producto}$, entonces $p_j(\mathbf{X}(t))$ es $c_j \cdot X_i(t)$.

* Si $r_j \equiv s_i + s_{i'} \rightarrow \text{producto}$, entonces $p_j(\mathbf{X}(t))$ es:

$$\begin{cases} c_j \cdot X_i(t) \cdot X_{i'}(t) & \text{si } i \neq i' \\ c_j \cdot \frac{1}{2} \cdot X_i(t) \cdot (X_{i'}(t) - 1) & \text{si } i = i' \end{cases}$$

Semántica asociada a los sistemas biomoleculares

Transición del sistema:

- ★ $\mathbf{X}(t) = (X_1(t), \dots, X_n(t)) \longrightarrow \mathbf{X}(t+1) = (X_1(t+1), \dots, X_n(t+1))$.
- ★ Ejecutando la reacción química r_j cuyo vector de cambio de estado $\mathbf{v}_j = (v_{1j}, \dots, v_{nj})$.

Se verifica que:

- ★ $\mathbf{X}(t+1) = \mathbf{X}(t) + \mathbf{v}_j = (X_1(t) + v_{1j}, \dots, X_n(t) + v_{nj})$.

Objetivo: determinar $P(\mathbf{X}(t) | \mathbf{X}(t_0))$:

- ★ Probabilidad de que el sistema alcance el estado $\mathbf{X}(t)$ sabiendo que parte del estado inicial $\mathbf{X}(t_0)$.

Sistemas biomoleculares: the chemical master equation

La evolución en el tiempo de $P(\mathbf{X}(t) | \mathbf{X}(t_0))$ viene dada por la siguiente ecuación:

$$P(\mathbf{X}(t + dt) | \mathbf{X}(t_0)) = \underbrace{\sum_{j=1}^q p_j(\mathbf{X}(t) - \mathbf{v}_j) \cdot P((\mathbf{X}(t) - \mathbf{v}_j) | \mathbf{X}(t_0)) \cdot dt}_{\text{probabilidad de que se ejecute alguna reacción en el intervalo } [t, t + dt]} + \underbrace{[1 - \sum_{j=1}^q p_j(\mathbf{X}(t)) \cdot dt] \cdot P(\mathbf{X}(t) | \mathbf{X}(t_0))}_{\text{probabilidad de no se ejecute ninguna reacción en el intervalo } [t, t + dt]}$$

Luego: $P(\mathbf{X}(t + dt) | \mathbf{X}(t_0)) - P(\mathbf{X}(t) | \mathbf{X}(t_0)) =$

$$\left[\sum_{j=1}^q p_j(\mathbf{X}(t) - \mathbf{v}_j) \cdot P((\mathbf{X}(t) - \mathbf{v}_j) | \mathbf{X}(t_0)) - \sum_{j=1}^q p_j(\mathbf{X}(t)) \cdot P(\mathbf{X}(t) | \mathbf{X}(t_0)) \right] \cdot dt$$

La siguiente ecuación (**chemical master equation, CME**) proporciona $P(\mathbf{X}(t) | \mathbf{X}(t_0))$:

$$\frac{d P(\mathbf{X}(t) | \mathbf{X}(t_0))}{dt} = \sum_{j=1}^q [p_j(\mathbf{X}(t) - \mathbf{v}_j) \cdot P((\mathbf{X}(t) - \mathbf{v}_j) | \mathbf{X}(t_0)) - p_j(\mathbf{X}(t)) \cdot P(\mathbf{X}(t) | \mathbf{X}(t_0))]$$

Se trata de un sistema de EDOs con tantas ecuaciones como posibles estados.

- ★ En la práctica, sólo se puede resolver analíticamente en casos muy simples.
- ★ Por ello, se generarán trayectorias numéricas para el vector $\mathbf{X}(t)$, usando algoritmos de tipo Monte Carlo.

Algoritmo de Gillespie (SSA, Stochastic Simulation Algorithm)

Precursor: Un trabajo de J.L. Doob sobre cadenas de Markov ²

Algoritmo de Gillespie³: explora el espacio de estados asociado a la **CME**.

- * Es un algoritmo de Monte Carlo (Ulam, von Neumann, 1944-1946): genera trayectorias de cadenas de Markov que representan interacciones moleculares que tienen lugar en un **medio/volumen fijo** (ver ⁴ para algunas mejoras).

Núcleo del algoritmo de Gillespie:

- * **Entrada**: estado inicial de **un** medio/volumen m en el que hay sustancias sometidas a reacciones químicas r_1, \dots, r_q , de las que se conocen sus *propensidades*.
- * **Salida**: la reacción química a ejecutar y su correspondiente *tiempo de espera*, para hallar el siguiente estado.

Ha sido utilizado con éxito para la simulación estocástica de procesos biomoleculares.

²J.L. Doob. Markoff chains – Denumerable case. *Transactions of the American Mathematical Society*, **58**, 3 (1945), 455–473

³D.T. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions, *Journal of Computational Physics*, **22** (1976), 403-434; y Exact stochastic simulation of coupled chemical reactions, *Journal of Physical Chemistry*, **81**, 25 (1977), 2340–2361.

⁴M.A. Gibson, J. Bruck. Efficient Exact Stochastic Simulation of Chemical Systems with Many Species and Many Channels. *Journal of Physical Chemistry*, **104**, 9 (2000), 1876–1889.

Núcleo del algoritmo de Gillespie

Entrada: estado inicial de un **medio/volumen fijo** m en el que hay sustancias sometidas a reacciones químicas r_1, \dots, r_q , de las que se conocen sus *propensidades* p_j .

1. Calcular $p_0 = \sum_{j=1}^q p_j$, siendo p_j las propensidades de las reglas.
2. Generar dos números aleatorios a_1 y a_2 uniformemente distribuidos sobre el intervalo unidad $(0, 1)$: **paso de Monte Carlo**.
3. Elegir **el** índice j_0 que verifica la siguiente condición:

$$\sum_{k=1}^{j_0-1} p_k < a_2 \cdot p_0 \leq \sum_{k=1}^{j_0} p_k.$$

4. Calcular el valor: $\tau = \frac{1}{p_0} \ln\left(\frac{1}{a_1}\right)$ **tiempo de espera**.
5. Devolver (j_0, τ) .

La **reacción química seleccionada** es r_{j_0} y el **tiempo de espera** para ejecutar la siguiente reacción es τ .

Algoritmo de Gillespie

Entrada: (a) El estado de un medio/volumen fijo m en el que hay sustancias sometidas a reacciones químicas r_j , de las que se conocen sus *propensidades* p_j ; y (b) un tiempo máximo de simulación, $T > 0$.

- **Inicialización:**

- ★ Inicializar el tiempo de simulación: $t = 0$.
- ★ Inicializar el estado del sistema: $\mathbf{X}(t) = \mathbf{X}(0)$.

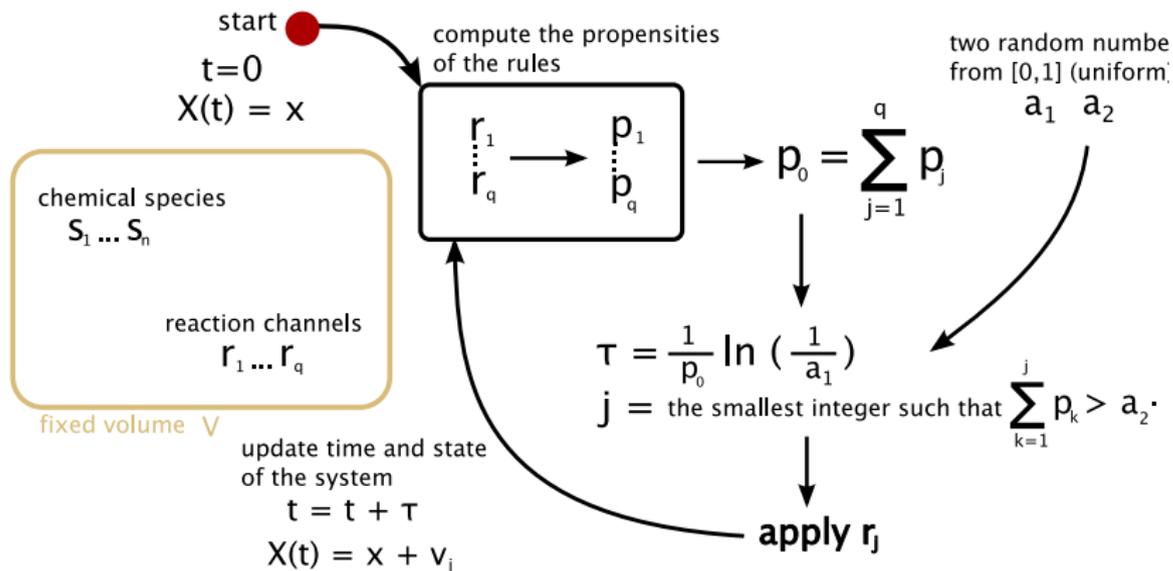
- **Iteración:**

- ★ Calcular las propensidades $p_j(\mathbf{X}(t))$.
- ★ Ejecutar el núcleo del algoritmo de Gillespie a partir de $p_j(\mathbf{X}(t))$: devuelve (j_0, τ) .
- ★ Ejecutar **una vez** la reacción química r_{j_0} .
- ★ Actualizar el tiempo de simulación: $t \leftarrow t + \tau$.
- ★ Hallar el nuevo estado del sistema $\mathbf{X}(t + \tau)$, tras la ejecución de r_{j_0} .
- ★ Iterar el proceso mientras no se verifique la condición de parada.

- **Finalización** (condición de parada):

- ★ Si el número de moléculas es cero o bien $t > T$, entonces finalizar el proceso.

Algoritmo de Gillespie



Algoritmo de Gillespie: observaciones

A la hora de resolver la CME, se consideran intervalos infinitesimales de tiempo:

- * En modelos basados en ODEs, dichos intervalos son **aproximados** por dt .
- * En el algoritmo de Gillespie, dichos intervalos son representados de manera **exacta** por **tiempos de espera**.

El algoritmo de Gillespie no resuelve numéricamente la CME ya que no proporciona la función de distribución de probabilidad de $\mathbf{X}(t)$.

- * Se puede obtener información acerca de esa función a través de los resultados obtenidos en reiteradas ejecuciones del mismo.

Algoritmo de Gillespie: ventajas y desventajas

Ventajas:

- ★ Captura los efectos del ruido (aleatoriedad inherente) en los procesos biomoleculares.
- ★ Maneja el tiempo de manera continua pero describe el sistema de forma discreta.
- ★ La implementación es relativamente simple.

Desventajas:

- ★ Alto coste computacional (si existen muchas moléculas, el valor de p_0 puede ser muy elevado, en cuyo caso los tiempos de espera serán muy pequeños).
- ★ Sólo una reacción química (y una sólo vez) se ejecutará en cada paso de computación.
- ★ Hay que realizar muchas simulaciones para obtener **resultados fiables**.
- ★ Es necesario disponer de un buen **generador de números aleatorios**.

Mejoras del algoritmo de Gillespie

En el artículo:

- D.T. Gillespie. Approximate accelerated stochastic simulation of chemically reacting systems. *Journal of Chemical Physics*, **115**, 4 (2001), 1716–1733.

se han propuesto algunas mejoras del algoritmo de Gillespie, a fin de conseguir más eficiencia computacional:

- * Perdiendo algo de “exactitud” en la selección del tiempo de espera.
 - * Permitiendo la ejecución de varias reacciones en el mismo paso de computación asociado a un mismo tiempo de espera (**tau-leaping**).
- **Condición de salto**: los cambios en el estado inducido por las reacciones que se ejecutan deben ser tales que las propensidades permanezcan aproximadamente constantes en el intervalo $[t, t + \tau)$.