

TEMA III

Modelos estocásticos. Algoritmo de Gillespie. Extensiones al marco de los sistemas P

Mario de J. Pérez Jiménez
Grupo de investigación en Computación Natural
Dpto. Ciencias de la Computación e Inteligencia Artificial
Universidad de Sevilla

Simulación y análisis computacional en Biología de Sistemas
Máster Universitario en Lógica, Computación e Inteligencia Artificial
Curso 2012-13



Modelos estocásticos clásicos de sistemas celulares (I)

El carácter **no determinista/aleatorio** de los fenómenos biológicos se captura mediante el uso de estrategias **estocásticas** o **probabilísticas**.

- ▶ A nivel microscópico, las interacciones entre las moléculas siguen la ley \sqrt{n} :
 - ▶ Los niveles de fluctuación del sistema son inversamente proporcionales a la raíz cuadrada del número de partículas.
- ▶ Los sistemas con un bajo número de moléculas muestran altas fluctuaciones.
 - ▶ Las aproximaciones clásicas deterministas y/o continuas para la modelización de sistemas celulares son cuestionables.

Modelos estocásticos clásicos de sistemas celulares (II)

Supuesto inicial: se tiene un volumen fijo homogéneo (equilibrio).

- ▶ Tipos de moléculas $\{s_1, \dots, s_n\}$.
- ▶ Tipos de interacciones o 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 i (concentración).

Estudiar la evolución del vector $\mathbf{X}(t)$ a partir de un estado de partida $\mathbf{X}(t_0)$.

Modelos estocásticos clásicos de sistemas celulares (III)

Cada reacción química r_j está caracterizada por:

- ▶ Un *vector de cambio de estado* $\mathbf{v}_j = (v_{1j}, \dots, v_{nj})$:
 - ▶ v_{ij} : cambio en el número de elementos s_i que produce aplicar r_j .
- ▶ Una *propensidad* $p_j(\mathbf{X}(t))$.
 - ▶ Se define de forma que $p_j(\mathbf{X}(t))dt$ sea la probabilidad de que r_j se ejecute en el intervalo de tiempo $[t, t + dt)$

Modelos estocásticos clásicos de sistemas celulares (IV)

Cálculo de la **propensidad** de r_j en un estado $\mathbf{X}(t)$:

- ▶ Se le asocia una **constante cinética** k_j : depende de las propiedades físicas de las moléculas involucradas y de otros parámetros físicos (se obtiene experimentalmente).

- ▶ A partir de k_j se halla una **constante estocástica** c_j así:

- ▶ Si $r_j \equiv s_i \rightarrow$ *producto*, entonces $c_j = k_j$.

- ▶ Si $r_j \equiv s_i + s_{i'} \rightarrow$ *producto*, entonces

$$\begin{cases} c_j = \frac{k_j}{V} & \text{si } i \neq i' \\ c_j = \frac{2 \cdot k_j}{V} & \text{si } i = i' \end{cases}$$

siendo V el volumen constante del medio. Si k_j viene dada en unidades de concentración, entonces hay que dividir también por el número de Avogadro.

- ▶ Si $r_j \equiv s_i \rightarrow$ *producto*, entonces su propensidad es $c_j \cdot X_i(t)$.

- ▶ Si $r_j \equiv s_i + s_{i'} \rightarrow$ *producto*, entonces su propensidad es

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

Modelos estocásticos clásicos de sistemas celulares (V)

Si el sistema está en un estado $\mathbf{X}(t)$ y se ejecuta una reacción química r_j , entonces el nuevo estado es $\mathbf{X}(t) + \mathbf{v}_j$.

$P(\mathbf{X}(t) | \mathbf{X}(t_0))$: probabilidad de que el sistema esté en el estado $\mathbf{X}(t)$ sabiendo que el estado inicial es $\mathbf{X}(t_0)$.

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

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

Sólo puede ser resuelta analíticamente en pocos casos.

Algoritmo clásico de Gillespie (I)

Algoritmo de Gillespie¹: diseñado para explorar el espacio de estados asociado a la EQM.

Es un algoritmo de Monte Carlo que genera trayectorias de cadenas de Markov que representan interacciones moleculares que tienen lugar en un **volumen fijo** (ver ² para algunas mejoras del algoritmo).

- ▶ **Entrada:** estado de un medio 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 biológicos que tienen lugar en un determinado volumen fijo.

¹T.D. Gillespie, D.T. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions, *J. Comp. Phys.*, **22** (1976), 403-434; y Exact stochastic simulation of coupled chemical reactions, *J. Phys. Chem.*, **81** (1977), 2340-2361.

²M.A. Gibson, J. Bruck. Efficient Exact Stochastic Simulation of Chemical Systems with Many Species and Many Channels. *Journal of Physical Chemistry*, **104**, 25 (2000), 1876–1889.

Algoritmo clásico de Gillespie (II)

Entrada: estado de un medio m en el que hay sustancias sometidas a posibles 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 $(0, 1)$.
3. Calcular el tiempo de espera para la siguiente reacción: $\tau = \frac{1}{p_0} \ln\left(\frac{1}{a_1}\right)$.
4. Elegir el índice j_0 , de la siguiente reacción química, que verifica:
$$\sum_{k=1}^{j_0-1} p_k < a_2 \cdot p_0 \leq \sum_{k=1}^{j_0} p_k.$$
5. Devolver (τ, j_0) .

Entonces se aplicará *una vez* la regla r_{j_0} y se actualizarán las moléculas de m

Modelos estocásticos de sistemas celulares basados en sistemas P (I)

Membrane Computing: Explora la **naturaleza computacional** de una serie de hechos relativos a las membranas biológicas.

- ▶ Juegan un papel relevante en el funcionamiento de las células.
- ▶ Están involucradas en muchas reacciones químicas que tienen lugar dentro de los compartimentos.
- ▶ Actúan como canales selectivos de comunicación entre las células y su entorno.
- ▶ Los procesos que tienen lugar en la estructura compartimentalizada de una célula viva pueden ser interpretados como **procedimientos de cálculo**.

Modelos estocásticos de sistemas celulares basados en sistemas P (II)

Sistemas P que trabajan a modo de células:

- ▶ Una *estructura de membranas* (jerarquizada) que delimita una serie de regiones o compartimentos.
 - ▶ Árbol enraizado: la raíz se denomina *piel* y las hojas se denominan *membranas elementales*.
- ▶ Unos *multiconjuntos de objetos* sobre un alfabeto colocados en los compartimentos delimitados por las membranas.
- ▶ Unas *reglas de reescritura* asociadas a cada compartimento que describen las reglas de evolución de los objetos.

Modelos estocásticos de sistemas celulares basados en sistemas P (III)

Un *sistema P multientorno de orden* (q, m, n) , con T unidades de tiempo $(q, m, n, T \geq 1)$, es una tupla $ME = (G, \Gamma, \Sigma, \Pi_q, \{f_r : r \in \mathcal{R}_{\Pi_q}\}, \{E_j : 1 \leq j \leq m\}, \mathcal{R}_E)$, en donde:

- ▶ $G = (V, S)$ es un grafo dirigido. Los elementos de $V = \{e_1, \dots, e_m\}$ se denominan **entornos**.
- ▶ Γ es el alfabeto de trabajo y $\Sigma \subsetneq \Gamma$ (los elementos de Σ son los objetos que pueden estar presente en los entornos).
- ▶ $\Pi_q = (\Gamma, H, \mu, M_1, \dots, M_q, \mathcal{R}_{\Pi_q})$ es un sistema P de grado q , cuyas reglas son del tipo $r \equiv u[v]_h^\alpha \xrightarrow{f_r} u'[v']_h^{\alpha'}$, siendo $u, v, u', v' \in \Gamma^*$, $h \in H$ y $\alpha, \alpha' \in \{0, +, -\}$.
- ▶ $f_r, r \in \mathcal{R}_{\Pi_q}$, es una función computable cuyo dominio es $\{1, 2, \dots, T\}$.
- ▶ \mathcal{R}_E es un conjunto finito de reglas de comunicación de la forma

$$(x \xrightarrow{g(j,j')} y)_{e_j} \quad (x)_{e_j} \xrightarrow{h(j,j')} (y)_{e_{j'}} \quad \text{y} \quad (\Pi_q)_{e_j} \xrightarrow{k(j,j')} (\Pi_q)_{e_{j'}}$$

donde $x \in \Sigma$, $y \in \Sigma \cup \{\lambda\}$, $(e_j, e_{j'}) \in S$, $1 \leq j, j' \leq m$, y g, h, k son funciones computables cuyo dominio es $\{1, 2, \dots, T\}$.

- ▶ $E_j, 1 \leq j \leq m$, es un multiconjunto finito de objetos de Σ .
- ▶ n es el número de copias de Π_q distribuidas aleatoriamente en los m entornos de la configuración inicial.

Modelos estocásticos de sistemas celulares basados en sistemas P (IV)

Un **sistema P multientorno** de grado (q, m, n) y con T unidades de tiempo:

- ▶ Es un conjunto de m entornos conectados entre sí por los arcos de G .
- ▶ En total existen n sistemas P que trabajan a modo de células, Π_q , todos ellos son idénticos e, inicialmente, están distribuidos aleatoriamente en los m entornos.
- ▶ Cuando se aplica una regla del tipo $(x \xrightarrow{g_{(j,j')}} y)_{e_j}$, el objeto x del entorno e_j produce el objeto y .
- ▶ Cuando se aplica una regla del tipo $(x)_{e_j} \xrightarrow{h_{(j,j')}} (y)_{e_{j'}}$, el objeto x pasa del entorno e_j al entorno $e_{j'}$ transformándose en y .
- ▶ Cuando se aplica una regla del tipo $(\Pi_q)_{e_j} \xrightarrow{k_{(j,j')}} (\Pi_q)_{e_{j'}}$, el sistema Π_q pasa del entorno e_j al entorno $e_{j'}$.
- ▶ Los valores $f(t), g(t), h(t), k(t)$ son indicadores de la predisposición de las reglas a ser aplicadas en el instante t .



Usaremos el término **compartimento** para designar a membranas o entornos, indistintamente.



Algoritmo multicompartimental de Gillespie (I)

Los sistemas P multientornos constan de:

- ▶ Una serie de entornos jerarquizados por un grafo dirigido.
- ▶ Una serie de sistemas P que trabajan a modo de células contenidos en los entornos. Dichos sistemas delimitan regiones (volúmenes), cada una con su propio conjunto de reglas y multiconjunto de objetos.

Supongamos que tenemos un sistema P multientorno en donde las funciones que intervienen formalizan **propensidades**.

El **algoritmo** (clásico) **de Gillespie** proporciona una adecuada simulación estocástica de procesos biológicos que tienen lugar en un determinado volumen.

Para capturar la semántica de los sistemas P multientornos:

- ▶ Vamos a extender dicho algoritmo permitiendo que cada membrana evolucione de acuerdo con el algoritmo clásico.

Algoritmo multicompartimental de Gillespie (II)

Entrada: Un sistema P multientorno del que se conocen las propensidades iniciales de todas las reglas.

- **Inicialización:**

- Poner el tiempo de simulación a $t = 0$.
- Para cada compartimento c en la estructura μ hacer:
 - Ejecutar el algoritmo de Gillespie clásico en c , devolviendo (τ_c, j_c) .
- Ordenar la lista de ternas (τ_c, j_c, c) con relación a τ_c .

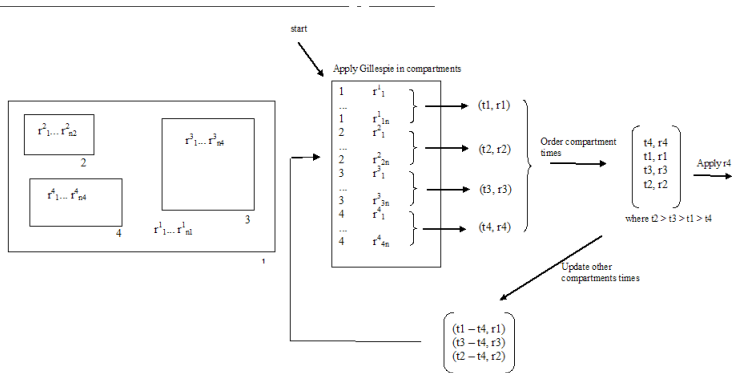
- **Iteración:**

- Elegir la primera terna de la lista, $(\tau_{c_0}, j_{c_0}, c_0)$.
- Poner el tiempo de simulación a $t = t + \tau_{c_0}$.
- Actualizar el tiempo de espera para el resto de las ternas, restando τ_{c_0} .
- Aplicar una vez la regla $r_{j_{c_0}}$ modificando los compartimentos afectados.
- Para cada compartimento c' afectado, eliminar la terna $(\tau_{c'}, j_{c'}, c')$ de la lista, actualizar las propensidades y ejecutar el algoritmo de Gillespie clásico en c' , obteniendo una nueva terna $(\tau_{c'}^*, j_{c'}^*, c')$.
- Añadir las ternas $(\tau_{c'}^*, j_{c'}^*, c')$ a la lista y ordenarlas de nuevo.
- Iterar el proceso.

- **Finalización:**

- Si el tiempo de simulación t excede al tiempo máximo prefijado, T , finalizar el proceso.

Algoritmo multicompartimental de Gillespie (III)



Algoritmo determinista con tiempo de espera (I)

El algoritmo multicompartimental de Gillespie:

- ▶ Permite definir la semántica de modelos computacionales de procesos biológicos basados en sistemas P.
- ▶ Puede ser usado para el estudio de sistemas celulares en donde las aproximaciones deterministas y/o continuas **no** son adecuadas.
- ▶ Suele ser muy costoso desde el punto de vista computacional.

En el caso de procesos en donde el número de individuos es *elevado*, las aproximaciones deterministas y/o continuas pueden ser aceptables.

- ▶ Se propone un **algoritmo determinista con tiempo de espera**.
 - ▶ Versión determinista del algoritmo multicompartimental de Gillespie.
 - ▶ Más eficiente que el algoritmo multicompartimental de Gillespie.

Algoritmo determinista con tiempo de espera (II)

Entrada: Un sistema P del que se conocen las propensiones de todas las reglas.

- **Inicialización:**

- Poner el tiempo de simulación a $t = 0$.
- Para cada regla r_i asociada a un compartimento c del sistema, calcular la terna (τ_i, i, c) , en donde $\tau_i = \frac{1}{v_i}$, y v_i se calcula aplicando la ley de acción de masas.
- Ordenar la lista de ternas (τ_i, i, c) de acuerdo a los tiempos de espera.

- **Iteración:**

- Elegir la primera terna, (τ_{i_0}, i_0, c_0) de la lista ordenada.
- Poner el tiempo de simulación a $t = t + \tau_{i_0}$.
- Actualizar los tiempos de espera de las restantes ternas restando τ_{i_0} .
- Aplicar la regla r_{i_0} una sólo vez actualizando los compartimentos afectados por dicha aplicación.
- Para cada compartimento c' afectado por r_{i_0} eliminar la correspondiente terna de la lista y volver a calcular los tiempos de espera de las reglas de c' .
- Añadir las nuevas ternas a la lista y ordenarlas de acuerdo a los tiempos de espera.

- **Finalización:**

- Si el tiempo de la simulación t alcanza o excede a un tiempo maximal prefijado, finalizar el proceso

Algoritmo determinista con tiempo de espera vs Algoritmo multicompartimental de Gillespie

- ▶ En general, el algoritmo determinista con tiempo de espera es más eficiente que el algoritmo multicompartimental de Gillespie.
- ▶ En el **algoritmo determinista con tiempo de espera**:
 - ▶ Se calculan los pares (*tiempo de espera, regla*) para **todas** las reglas del sistema (**ley de acción de masas**).
 - ▶ La lista (*tiempo de espera, regla*) consta de s pares, siendo s el número total de reglas.
- ▶ En el **algoritmo multicompartimental de Gillespie**:
 - ▶ De cada compartimento se selecciona un único par (*tiempo de espera, regla*) (**algoritmo clásico de Gillespie**).
 - ▶ La lista (*tiempo de espera, regla*) consta de s' pares, siendo s' el número total de compartimentos.