

個体ベースの Gillespie アルゴリズム解説

中岡慎治 (Shinji Nakaoka)

理化学研究所統合生命医科学研究センター 免疫数理モデリング (YCI) ラボ
RIKEN Research Center for Integrative Medical Science Center (IMS-RCAI),
Laboratory for Mathematical Modeling of Immune System

ABSTRACT

異質性をもつ集団の個体群ダイナミクスに対しても適用可能な確率シミュレーション技法である, 個体ベースの Gillespie アルゴリズムについて概説する. 個体ベースの Gillespie アルゴリズムにより, 一細胞レベルで経時的に細胞分裂や死亡を定量的に計測したデータや遺伝子発現状態の変化による細胞機能の変化も考慮した細胞集団の個体群ダイナミクスを実施が可能になる.

Key words: 個体ベースモデル; 確率過程シミュレーション; Gillespie アルゴリズム;

1 はじめに

近年, 一細胞レベルで定量データを取得可能な実験技術が発展したことにより, 遺伝的に均一なクローン細胞集団においても, 遺伝子発現状態や分裂後の寿命に異質性が存在することが明らかになりつつある. たとえば [3] の一細胞計測実験によれば, 集団内で細胞死が起こる時間間隔は指数分布に従うのではなく Weibull 分布でよく近似できることが示された. また, 細胞内における分子の揺らぎが遺伝子ネットワークのダイナミクスに大きな役割を果たすことを示した報告が既にいくつも存在する [5], [4].

不均一な個体からなる集団に対する個体群ダイナミクスの研究では, これまで個体ベースモデル (IBM; individual based model) を用いたシミュレーション研究が広く行われてきた [7]. IBM は時間発展の更新ルールを記述した数理モデルであり, 柔軟かつ詳細な設定ができるため, 異質性の高い集団の個体群ダイナミクスをシミュレーションするには適している. しかしながら, IBM は本質的にアルゴリズムベースであるため, 数学的構造の分析を通じて何か情報を抽出するような解析的方法論は筆者の知る限り存在しない. 上述の一細胞計測実験から得られた細胞の振る舞いを記述するためには, IBM の柔軟性と確率過程や決定論方程式が有する解析手法の両方を活用できるような理論的枠組みが有効だと考えられる.

本稿で主題とする確率シミュレーションについて簡単に言及しておく. 確率シミュレーションとは, ある確率法則に従って生じるイベントの選択と, イベント発生に伴う系の状態変化を時間的に追跡する計算機上の仮想実験である. 個体群のダイナミクスは, 確率シミュレーションの中でも主に個体の生成と消滅を主要なイベントとし, 個体数の増減が系の状態

変化に対応する。Gillespie のアルゴリズムと呼ばれる手法は、数ある確率過程の中でもっともよく知られた確率過程のシミュレーション技法の一つであり、化学反応系やシステムズバイオロジーの分野で広く用いられている [2]。Gillespie のアルゴリズムは、個々の反応が過去に起こった反応に依存しない (無記憶性) ことを前提とした Poisson 過程のシミュレーション技法であり、どの反応が起こるかを反応の傾向 (propensity of reaction) に基づいて選択する。前提として無記憶性を仮定しているため、上述の不均一性を有する細胞集団に対して Gillespie のアルゴリズムが適用可能かどうかは検討が必要となる。Gillespie のアルゴリズムは集団を反応の単位としているものの、反応速度の異なる複数種混合系に対しても適用可能である。もし個体それぞれを反応の単位とみなせば、不均一な個体からなる集団に対しても Gillespie のアルゴリズムを適用できる可能性は残されている。

文献 [6] において、筆者らは個体ベースモデルに対して Gillespie のアルゴリズムが適用できるよう基礎理論を拡張した。開発したアルゴリズム (個体ベース Gillespie 法) を用いることで、イベント発生の待ち時間に履歴が伴うような状況であっても、Gillespie 法を適用できるようになる。本稿では、[6] で提案した手法を概説する。2 節で個体ベースモデルの数学的表現について紹介した後、3 節では従来の Gillespie アルゴリズム (the conventional Gillespie algorithm) の概要および個体ベース Gillespie 法を紹介する。

2 個体ベースモデルの数学表現

出生や死亡による個体の生成・消滅をまとめてイベントと呼ぶ。化学反応系の場合、化学反応がイベントに対応する。Gillespie のアルゴリズムとの比較から、本稿においてもイベントを反応と呼ぶことにする。

i -状態変数 (individual state variable) とは個体の年齢や位置情報、繁殖可能性といった個体の状態を表す変数の組で、個体は i -状態空間 (i -state space) における要素として定義される。以下では単純な例を考察しよう。すなわち、個体は系における存在、属する集団、年齢によって特徴付けられると仮定する。個体の状態は他個体との相互作用や環境からの影響を受けて時間と共に変化するため、一般に i -状態変数は時間 t の関数である。 S を 0 と 1 からなる二値集合 (binary set) とする。すなわち

$$S := \{0, 1\}. \quad (2.1)$$

個体 j が系に存在しているかどうかは、関数 σ_j によって定義する。すなわち、 $\sigma_j : [0, \infty) \rightarrow S$ は時刻 t において個体 j が系に存在しているとき $\sigma_j(t) = 1$ 、存在していないとき $\sigma_j(t) = 0$ となる二値関数である。

系は M 個の集団から構成されていると仮定する。たとえば生まれたばかりの個体は繁殖できないような状況を表現したい場合、繁殖の可否によって集団を定義する。 G を系に存在

する集団のインデックス集合とする.

$$G := \{1, 2, \dots, M\}. \quad (2.2)$$

個体 j が集団 $i \in G$ に属しているかどうかは, 関数 $g_j : [0, \infty) \rightarrow G$ によって定義する. すなわち, 個体 j が集団 $i \in G$ に属しているとき $g_j(t) = i$ である. 成熟して個体が繁殖可能になった場合, $g_j(t)$ の値は変化する. ある個体が同時に 2 つ以上の集団 i, \bar{i} に属していることを表現したい場合には, たとえば別の関数 $h_j(t)$ を定義して $g_j(t) = i$ かつ $h_j(t) = \bar{i}$ と表現する.

個体の年齢を表す集合 A を

$$A := \{a \in \mathbb{R} : a \geq 0\} \quad (2.3)$$

によって定義する. 時刻 t に生まれた個体 j の年齢 $a_j(t)$ を $a_j(t) = 0$ とする. 時間ステップ Δt だけ経過すると, 個体 j の年齢は $a_j(t + \Delta t) = a_j(t) + \Delta t$ とするのが自然である.

上で述べたような単純な例の場合, i -状態空間 Ω は直積集合 $\Omega := S \times G \times A$ によって表される. 時刻 t における個体 j の状態 $x_j(t)$ は, i -状態空間 Ω の要素であるベクトル $x_j(t) = (\sigma_j(t), g_j(t), a_j(t))$ によって一意に表される.

時刻 t における系の状態 $\mathbf{x}(t)$ は, 集合

$$\mathbf{x}(t) = \bigcup_{j=1}^{\infty} x_j(t) \Big|_{\sigma_j(t)=1} \quad (2.4)$$

によって定義する. ここで (2.4) は $\sigma_j(t) = 1$ である要素 $x_j(t)$ からなる和集合であり, $\mathbf{x}(t)$ は集合 Ω の部分集合であることに注意する. 個体ベースモデルでは, 系に存在する個体をカウントすることで個体数を定義する. すなわち, 時刻 t における系の個体数は, 関数 $N : [0, \infty) \rightarrow \mathbb{N} \cup \{0\}$ を用いて

$$N(t) := \sum_{j=1}^{\infty} \sigma_j(t) \quad (2.5)$$

と定義する. 定義 (2.4) からわかるように, (2.5) は系の個体数を表している. 同様に, 時刻 t における集団 i の個体数 $N_i(t)$ を

$$N_i(t) := \sum_{j=1}^{\infty} \sigma_j(t) \delta_{ig_j}(t) \quad (2.6)$$

と定義する. ここで $\delta_{ig_j}(t)$ はクロネッカーのデルタで

$$\delta_{ig_j}(t) = \begin{cases} 1 & g_j(t) = i, \\ 0 & g_j(t) \neq i. \end{cases} \quad (2.7)$$

定義により

$$N(t) = \sum_{i=1}^M N_i(t) \quad (2.8)$$

が得られる.

2.1 出生・死亡の傾向

個体の出生や死亡の傾向 (propensity) は, 個体の年齢, 他個体との相互作用や環境からの影響によって決定される. 個体 j は出生や死亡など合計 q_j 個の反応をもつと仮定し, それぞれの反応が起こる傾向を関数 $p_{jk} : [0, \infty) \times \Omega \rightarrow \mathbb{R}_+$, ($k = 1, \dots, q_j$) によって表現する. 傾向関数は非負の実数値を値にもつ連続関数とする.

個体の死亡率が年齢と相関がある例を考えよう. 時刻 t における個体 j の死亡の傾向を表す関数 p_j は, 年齢依存的な個体あたりの死亡率 (per capita death rate) $\mu(a)$ を用いて

$$p_j(t, \mathbf{x}(t)) := \mu(a_j(t)) \quad (2.9)$$

と表される. 個体の成長や生存に必須の資源などは, 環境要因として個体の状態に影響する. $E(t)$ を時刻 t における環境因子の量とすると, 傾向関数は $E(t)$ にも依存する形に拡張が必要である. もし環境因子が個体 j の死亡率に影響する場合, (2.9) は

$$p_j(t, \mathbf{x}(t)) = \mu(a_j(t), E(t)) \quad (2.10)$$

と表される. 重要な点は, 個体群ダイナミクスが傾向関数によってのみ規定されている点である. すなわち, 個体がどのような i -状態をとっていても, 個体群のダイナミクスに影響するのは, 出生・死亡の傾向のみである. また, 傾向関数は時刻 t の (合成) 関数として定義されている点も重要である. 3.2 小節で詳しく述べるように, 系内で単位時間あたりに生じるイベントの数が非斉次系 (non-homogeneous) の Poisson 過程に従うと仮定することで, 個体の異質性を含んだ集団においても Gillespie のアルゴリズムを適用できるように拡張することができる.

2.2 系における出生・死亡・成熟の表現

系の状態の時間発展は反応が 1 つ起こる毎に離散的に更新されると仮定する. t_n を第 n 番目に起こった反応とし, $\Delta t_n := t_n - t_{n-1}$ を n 番目の時間ステップとする. 3 節で紹介する確率シミュレーションでは, 時間列 $\{t_n\}_{i=0}^{\infty}$ をある確率法則に従って決定する. 断りが無い限り, 時間ステップの n に対する依存性を省略して記載する.

$\mathcal{D} := [0, \infty) \times \Omega$ とし, $C(\mathcal{D})$ を領域 \mathcal{D} で定義された連続関数全体からなる集合とする. 個体は成熟することによって機能的にも変化し得るため, それに応じて出生や死亡の傾向を規定する傾向関数も変化し得る. したがって, 系における個体の出生や死亡は, 領域 $\mathcal{D} \times C(\mathcal{D})$ に作用する写像として定義するのが自然である. 以下, 系内で生じる出生, 死亡や成熟は離散時間系, すなわち時間列 $\{t_n\}_{i=0}^{\infty}$ 上で定義された写像とする.

出生など系における個体の生成を表す写像 \mathcal{B} は, $\sigma_j(t) = 0$ である Ω の j 番目の要素を $\sigma_j(t + \Delta t) = 1$ に変換する. 死亡など系における個体の消滅を表す写像 \mathcal{L} は, $\sigma_j(t) = 1$ で

ある Ω の j 番目の要素を $\sigma_j(t + \Delta t) = 0$ に変換する. 加齢を表す写像 \mathcal{A} は, $\sigma_j(t) = 1$ である Ω の全ての要素に対して $a_j(t + \Delta t) = a_j(t) + \Delta t$ を対応させる. 個体の成熟を表す写像 \mathcal{M} は, 個体が属する集団を変化させることで成熟を表現する. 個体が成熟する年齢 τ に到達して集団 i から \bar{i} に移行する場合, 写像 \mathcal{M} は $a_j(t + \Delta t) \geq \tau$ かつ $g_j(t) = i$ である Ω の j 番目の要素に対して $g_j(t + \Delta t) = \bar{i}$ に変換する. 出生, 死亡, 加齢や成熟に対応して, 個体の傾向関数も必要に応じて変化させる必要があることに注意する. その他, 体サイズ成長や遺伝子発現レベルを i -状態変数として定義した場合, それに応じた写像を定義する必要がある.

3 個体ベースの Gillespie アルゴリズム

本節では, 2 節で紹介した個体ベースモデルに対しても Gillespie のアルゴリズムが適用できるような拡張を紹介する. まず Gillespie 直接法 [1] について簡単に説明した後, 集団・個体・個体・個体をもつ反応という 3 つの階層構造を明示化することにより, 個体ベースの Gillespie 直接法を構築する.

3.1 Gillespie 直接法 [1] の簡単な紹介

系に存在する反応の種類 (reaction channel) の総数を R とする. 時刻 t において, 微小区間 $(t + \tau, t + \tau + d\tau)$ 内で κ 番目の反応 \mathbf{R}_κ が起こる確率を $P(\tau, \kappa)d\tau$ で表す. $P(\tau, \kappa)$ は反応確率密度とも呼ばれる. $P(\tau, \kappa)$ は連続変数 τ ($0 \leq \tau < \infty$) かつ離散変数 κ ($\kappa = 1, 2, \dots, R$) からなる同時確率分布 (joint probability density) である. ここで, $d\tau$ 内に 2 つ以上の反応は起きないと仮定, すなわち与えられた微小区間で 2 つ以上の反応が起こる確率は $o(d\tau)$ とする. 子細は大幅に省略するが, $P(\tau, \kappa)$ より確率変数の組 (τ, κ) を各時間ステップ毎に数値的に計算することが目的である.

時刻 t における系の状態 $\mathbf{x}(t)$ は, 各集団の個体数 $x_i(t)$ を用いて,

$$\mathbf{x}(t) = (x_1(t), \dots, x_M(t))$$

と表される. $p_n(t, \mathbf{x})\delta t$ を微小区間 δt 内で n 番目の反応 \mathbf{R}_n が起こる確率とする.

Gillespie のアルゴリズムは

- (i) 次の反応が起こるまでの時間ステップ ΔT を決定
- (ii) R 個ある反応のうち, どの反応が起こるかを選択
- (iii) 選択された反応の種類によって, 系の個体数を更新

の 3 つの段階からなる. K 番目までの反応の傾向関数 $p_k(t, \mathbf{x})$ の累積和を $\Gamma_K(t, \mathbf{x})$ ($K \in \{1, 2, \dots, R\}$) とする. すなわち

$$\Gamma_K(t, \mathbf{x}) = \sum_{k=1}^K p_k(t, \mathbf{x}). \quad (3.1)$$

傾向関数の総和を $\Gamma_0(t, \mathbf{x})$ とすると, 定義より

$$\Gamma_0(t, \mathbf{x}) = \sum_{k=1}^R p_k(t, \mathbf{x}). \quad (3.2)$$

アルゴリズムの第 1, 第 2 段階について述べよう. Gillespie 直接法では, 2 つの一様乱数 r_1 および r_2 を用いて, 確率変数の組 (τ, κ) を決定する. τ は

$$\tau = \frac{1}{\Gamma_0(t, \mathbf{x})} \log \left(\frac{1}{r_1} \right).$$

であり, 起こる反応は

$$\frac{\Gamma_{\kappa-1}(t, \mathbf{x})}{\Gamma_0(t, \mathbf{x})} < r_2 \leq \frac{\Gamma_{\kappa}(t, \mathbf{x})}{\Gamma_0(t, \mathbf{x})} \quad (3.3)$$

を満足する κ に対応する反応 \mathbf{R}_{κ} を選択する. アルゴリズムの第 3 段階は系の状態の更新に相当する. 反応 \mathbf{R}_k それぞれに対して, 予め個体数を増加もしくは減少させるかの情報を決定しておく. たとえば, 選択された反応 \mathbf{R}_k が系の個体数を増加させる場合, 第 k 番目の要素が 1 で, その他の要素が 0 であるベクトル $\nu_k = (0, 0, \dots, 1, 0, \dots, 0)$ を用いて, 系の状態を $\mathbf{x}(t + \Delta T) = \mathbf{x}(t) + \nu_k$ とする.

3.2 個体ベースの Gillespie 直接法

系に存在する集団の個数を M とし, 集団 $I \in \{1, 2, \dots, M\}$ に属する個体 $J \in \{1, \dots, N_I(t)\}$ はそれぞれ q_J 個の反応をもつと仮定する. ここで, 個体数は時間と共に変化するため, 個体のインデックス集合の大きさも時間と共に変化することに注意する. しかしながら, アルゴリズムの各ステップでは時間を固定して考えるため, 各ステップ毎に各集団に存在する反応の総数は定数である. 系は集団, 集団内の個体, 個体をもつ反応という 3 段階の階層をもつとすると, いずれの反応も集団 I に属する個体 J のもつ K 番目の反応という形で順序付けが可能である.

時刻 t における集団 $i \in \{1, 2, \dots, M\}$ に属する個体をもつ反応数 $R_i(t)$ は, 集団 i の個体数 $N_i(t)$ を用いて

$$R_i(t) := \sum_{j=1}^{N_i(t)} q_j \quad (3.4)$$

と表される. (3.4) より, 時刻 t における系内の総反応数 $R(t)$ は

$$R(t) = \sum_{i=1}^M R_i(t) = \sum_{i=1}^M \sum_{j=1}^{N_i(t)} q_j \quad (3.5)$$

である。集団 i に属する個体 j の k 番目の反応を \mathbf{R}_{jk} とし、微小区間 δt において反応 \mathbf{R}_{jk} が起こる確率を $p_{jk}(t, \mathbf{x})\delta t$ とする。 $p_{jk}(t, \mathbf{x})$ を反応の傾向関数と呼ぶ。ここで、個体がどの集団 $i \in G$ に属しているかどうかは i -状態変数として定義されているため、傾向関数はインデックスとして i をもたないことに注意する。

$I \in \{1, 2, \dots, M\}$, $J \in \{1, \dots, N_I(t)\}$ かつ $K \in \{1, \dots, q_J\}$ に対して、傾向関数の累積和 $\Gamma_{IJK}(t, \mathbf{x})$ を

$$\begin{aligned} \Gamma_{IJK}(t, \mathbf{x}) &= \sum_{i=1}^{I-1} \sum_{j=R_{i-1}(t)+1}^{R_{i-1}(t)+N_i(t)} \sum_{k=1}^{q_j} p_{jk}(t, \mathbf{x}) + \sum_{j=R_{I-1}(t)+1}^{R_{I-1}(t)+(J-1)} \sum_{k=1}^{q_j} p_{jk}(t, \mathbf{x}) + \sum_{k=1}^K p_{JK}(t, \mathbf{x}) \\ &= \mathcal{G}(I-1) + \mathcal{I}(J-1) + \mathcal{R}(K) \end{aligned} \quad (3.6)$$

によって定義する。ここで $R_0(t) \equiv 0$, $I=1$ もしくは $J=1$ のとき, (3.6) の第 1 項もしくは第 2 項は 0 と約束する。個体ベースの Gillespie アルゴリズムでは、各階層毎の累積和を定義する。 $\mathcal{G}(I-1)$ は (3.6) の第 1 項に対応し、集団 $1, 2, \dots, I-1$ に属する個体をもつ反応に対する傾向関数の和を表す。 $\mathcal{I}(J-1)$ は (3.6) の第 2 項に対応し、集団 I に属する個体 $1, 2, \dots, J-1$ がもつ反応に対する傾向関数の和を表す。 $\mathcal{R}(K)$ は集団 I に属する個体 J がもつ K 番目までの反応に対する傾向関数の和を表す。系における傾向関数の総和 $\Gamma_0(t, \mathbf{x})$ は

$$\Gamma_0(t, \mathbf{x}) := \sum_{i=1}^M \sum_{j=R_{i-1}(t)+1}^{R_{i-1}(t)+N_i(t)} \sum_{k=1}^{q_j} p_{jk}(t, \mathbf{x}) \quad (3.7)$$

である。実際、 $I=M$, $J=N_M(t)$ かつ $K=q_{N_M(t)}$, すなわち全ての反応について傾向関数の累積和をとった場合、簡単な計算により $\Gamma_{IJK}(t, \mathbf{x}) = \Gamma_0(t, \mathbf{x})$ であることを確認できる。

アルゴリズムの第 1, 第 2 段階は、以下の通りである。

(i) 時間ステップ ΔT は、一様乱数 $r_1 \in (0, 1]$ を用いて

$$\Delta T = \frac{1}{\Gamma_0(t, \mathbf{x})} \log \left(\frac{1}{r_1} \right)$$

とする。

(ii) $I \in \{1, 2, \dots, M\}$, $J \in \{1, 2, \dots, N_I\}$, $K \in \{1, \dots, q_J\}$ かつ一様乱数 $r_2 \in (0, 1]$ に対して

$$\frac{\Gamma_{IJ(K-1)}(t, \mathbf{x})}{\Gamma_0(t, \mathbf{x})} < r_2 \leq \frac{\Gamma_{IJK}(t, \mathbf{x})}{\Gamma_0(t, \mathbf{x})} \quad (3.8)$$

を満たす組 (I, J, K) に対応する反応 \mathbf{R}_{JK} を選択する。

アルゴリズム第 2 段階の (3.8) を満たす (I, J, K) の決め方に関する詳細は省略するが、各階層における累積和を用いることで、インデックス I, J, K が順次一意に定まることを証明できる。

4 まとめと今後の課題

本稿では、異質性をもつ集団に対する個体群ダイナミクスを確率的にシミュレーションする手法である個体ベースの Gillespie アルゴリズムを紹介した。Gillespie による一連の研究では、Gillespie アルゴリズムと状態間の遷移確率に関する微分方程式系であるマスター方程式や、Langevin 方程式との関連も研究されている [2]。本研究で開発した手法は従来の Gillespie アルゴリズムの自然な拡張であるが、系に存在する反応数は各時間ステップ毎に異なる。したがって、Langevin 方程式やマスター方程式との対応は自明でない。これら方程式系との関連性について考察するのは興味深い問題である。紙面の都合上省略したが、細胞数が膨大になった場合でも現実的な計算時間でシミュレーションを実行するための近似アルゴリズムの発展と利用も重要な課題である。これらはいずれも今後の課題としたい。

参考文献

- [1] D. T. Gillespie. A general method for numerically simulation the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics*, **22** pp.403–434, (1976).
- [2] D. T. Gillespie. Stochastic simulation of chemical kinetics. *Annu Rev Phys Chem*, **58** pp.35–55, (2007).
- [3] E. D. Hawkins, J. F. Markham, L. P. McGuinness, and P. D. Hodgkin. A single-cell pedigree analysis of alternative stochastic lymphocyte fates. *Proc Natl Acad Sci U S A*, **106**, pp.13457–13462, (2009).
- [4] M. Kaern, T. C. Elston, W. J. Blake, and J. J. Collins. Stochasticity in gene expression: from theories to phenotypes. *Nat Rev Genet*, **6**, pp.451–464, (2005).
- [5] H. H. McAdams and A. Arkin. Stochastic mechanisms in gene expression. *Proc Natl Acad Sci U S A*, **94**, pp.814–819, (1997).
- [6] S. Nakaoka and K. Aihara. Stochastic simulation of structured skin cell population dynamics. *Journal of Mathematical Biology*, **66** pp.807–835, (2013).
- [7] S. F. Railsback and V. Grimm. *Agent-Based and Individual-Based Modeling: A Practical Introduction*. Princeton University Press, (2011).