

個体および集団ベース Gillespie アルゴリズムの相互検証

中岡慎治 (Shinji Nakaoka)

東京大学大学院医学系研究科 国際社会医学講座

Graduate School of Medicine, The University of Tokyo

ABSTRACT

文献 [4] において、筆者らは異質性をもつ集団の個体群ダイナミクスに対して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」を提案した。本稿では、自然な帰結として、もしある同一集団に属する個体間で出生・死亡率が等しければ、個体ベースのアルゴリズムは従来型集団ベースの Gillespie アルゴリズムに帰着することを示す。集団ベースのアルゴリズムを適用したとき、どの個体が反応に関わったを指定する方法 (個体選択法) にはいくつか異なる方法が考えうる。本稿では、Nicholson blowfly モデルを対象に、個体選択法の計算時間および数値計算結果を比較・検証した。結論として、個体抽出方法としては無作為抽出に近い方法を選択する方が望ましいという結論が得られた。

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

同一の個体から構成されると考えられるクローンのような細胞集団においても、年齢や体サイズの違いなど個体間には異質性が存在する。先進国のヒトにおいて年齢と死亡率は関連していることから、個体群ダイナミクスを考慮する上で個体の異質性を組み込んだ定式化を組み込むことが望ましい。年齢構造を考慮した個体群ダイナミクスは、一般に偏微分方程式や再生方程式、もしくは生活史におけるステージ遷移を表現した行列個体群モデルによって記述される。年齢以外にも個体の状態を表す変数として、たとえば細胞集団を考えた場合にはある代表的な遺伝子の発現量などが考えられる。

異質性を有する複雑な個体の集団の状態を追跡する手段として、柔軟に細かくルールを設定することのできる個体ベースシミュレーションが幅広く用いられている。しかしながら、どこまで細かくルールを設定すればよいのかどうかに応える一般的な指針や妥当性の指標は不在である。このため、過度の恣意的な仮定を入れて現象を再現できるという原理的な問題が存在する。一細胞計測技術の発展に代表されるように、網羅的かつ定量的な計測技術によって、集団内の異質性にはある確率的な法則が見いだされることが明らかになってきた。たとえば、細胞集団を対象にした一細胞粒度の計測において、細胞分裂を終了するまでの時間が右に裾野の長い log-normal 分布に従うという結果が報告されている [3]。実験による観測結果を適切に組み込んだ上で異質性を有する複雑な個体の集団の状態を追跡するためには、確率過程に基づいた個体ベースモデルの構築が不可欠であると考えられる。

文献 [4] において、筆者らは異質性をもつ集団の個体群ダイナミクスに対して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」を提案した。開

発した個体ベースモデルは、従来型の集団ベースアルゴリズムの自然な拡張となっているが、それ自体の証明および実際に計算機に実装させた場合に計算速度はどうなるのかという考察は行われていない。そこで本稿では、集団内の個体が均一な場合に、個体ベースのアルゴリズムが集団ベースの Gillespie アルゴリズムに帰着することを示す。続けて、Nicholson blowfly モデルを対象に、個体ベース・集団ベースアルゴリズムの計算時間および数値計算結果を比較・検証する。

2 均一な集団に対する個体ベースのアルゴリズム

Gillespie の直接法は、時刻 t において個体の状態 (年齢など) が $\mathbf{x}$ であるとき、出生・死亡率を表す傾向関数 $p_{jk}(t, \mathbf{x})$ の相対的な大きさに応じて起こる反応 (出生・死亡など個体数増減に関与するイベント) とその待ち時間を決定するアルゴリズムである。個体ベースのアルゴリズムでは、個体 J が属する集団 I と個体がもつ反応 K それぞれに傾向関数を決定する。個体がもつ反応数を q_j 、傾向関数の累積和を $\Gamma_{IJK}(t, \mathbf{x})$ とする (詳しくは [4] 参照)。

個体の均一性 (homogeneity) を定義しよう。ある集団 i に属する個体の状態が均一 (homogeneous) であるとは、任意の t かつ集団 i に属する全ての個体 j に対して、 $x_j(t) = x_{\bar{j}}(t)$ が成立する場合をいう。同様に、部分的に均一な状態というものも定義できる。ある集団 i に属する個体の状態が部分的に均一 (partially homogeneous) であるとは、 i -状態空間 Ω の部分集合 $P \subset \Omega$ に対して $x_j(t)|_P = x_{\bar{j}}(t)|_P$ が成立する場合をいう。ここで $|_P$ は P への制限とする。系に存在する全ての集団それぞれが均一な個体から構成される場合、状態が均一な系と呼ぶ。

傾向関数は i -状態とは独立に定義されているため、たとえある集団内で個体の状態が均一であったとしても、その集団内で個体の傾向関数が等しいとは限らないことに注意する。一方、ある集団内で個体が均一でなかったとしても、集団内で個体の傾向関数が等しいと仮定することもできる。したがって、集団内の個体の傾向関数が等しいような状況を独立に定義する必要が生じる。ある集団 i に属する個体が機能的に同等 (functionally equivalent) であるとは、任意の t かつ全ての個体 $j, \bar{j}$ に対して、 $p_{jk}(t, \mathbf{x}) = p_{\bar{j}k}(t, \mathbf{x})$ が全ての k に対して成立する場合をいう。ここで 個体 j および $\bar{j}$ それぞれがもつ反応の個数 q_j および $q_{\bar{j}}$ は等しい、すなわち $q_j = q_{\bar{j}}$ でなければならないことに注意する。状態が均一な系において、機能的に同等でない集団が存在すると考えるのは不自然である。したがって以下では、系の状態が均一ならば、各集団で個体は機能的に同等であると仮定する。

系の状態が均一であると仮定しよう。このとき、どの集団においても個体 j の傾向関数 p_{jk} および反応の個数 q_j は全て等しい。したがって、傾向関数および反応の個数は、集団のインデックス i を用いてそれぞれ $p_{ik}(t, \mathbf{x})$ かつ $q_j \equiv q_i$ とかける。このとき、集団 I に属する個体 J の K 番目の反応に対する傾向関数の累積和 $\Gamma_{IJK}(t, \mathbf{x})$ は、集団 $i \in \{1, 2, \dots, I-1\}$ に属する個体数 $N_i(t)$ を用いて

$$\Gamma_{IJK}(t, \mathbf{x}) = \sum_{i=1}^{I-1} \sum_{k=1}^{q_i} p_{ik}(t, \mathbf{x}) N_i(t) + (J-1) \sum_{k=1}^{q_I} p_{Ik}(t, \mathbf{x}) + \sum_{k=1}^K p_{Ik}(t, \mathbf{x}) \quad (2.1)$$

とかける [4]. 特に $I = M$, $J = N_M(t)$ かつ $K = q_M$ のとき, 定義より

$$\Gamma_0(t, \mathbf{x}) = \sum_{i=1}^M \sum_{k=1}^{q_i} p_{ik}(t, \mathbf{x}) N_i(t). \quad (2.2)$$

具体例として, 単純出生死亡過程を考えよう. 簡単のため, 個体は全て 1 つの集団に属すると仮定する. すなわち, (2.1) および (2.2) において $M = 1$ である. 対応する決定論方程式は

$$\frac{dN(t)}{dt} = bN(t) - dN(t). \quad (2.3)$$

(2.3) は [2] において最も基本的な例として考察された系で, 集団の個体数の分布は指数分布を定数倍したものに一致する.

考える集団は 1 つなので, 以下では添え字を省略して記載する. 各個体は反応を 2 つもつので, (2.1) および (2.2) において $q = 2$ である. 傾向関数 $p_k(t, \mathbf{x})$ ($k = 1, 2$) は, 個体あたりの出生率 (*per capita birth rate*) b および死亡率 (*per capita death rate*) d を用いて, $p_1(t, \mathbf{x}) = b$ かつ $p_2(t, \mathbf{x}) = d$ とする. 一方, 集団ベースの Gillespie 直接法では, 傾向関数 $\tilde{p}_k(t, \mathbf{x})$ ($k = 1, 2$) を個体群出生率 (population birth rate) $bN(t)$ および個体群死亡率 (population death rate) $dN(t)$ を用いて $\tilde{p}_1(t, \mathbf{x}) = bN(t)$ かつ $\tilde{p}_2(t, \mathbf{x}) = dN(t)$ と定めるのが自然である. ここで

$$\tilde{p}_k(t, \mathbf{x}) = p_k(t, \mathbf{x}) N(t)$$

とおくと, 個体ベース・集団ベースの Gillespie 直接法における $\Gamma_0(t, \mathbf{x})$ は一致するため, アルゴリズムの第 1 段階における操作は等しい.

アルゴリズム第 2 段階の同値性について考えよう. (2.1) に $p_k(t, \mathbf{x})$ ($k = 1, 2$) を代入すると

$$\frac{\Gamma_{IJK}(t, \mathbf{x})}{\Gamma_0(t, \mathbf{x})} = \begin{cases} \frac{J-1}{N(t)} + \frac{1}{N(t)} \times \frac{b}{b+d}, & (K=1), \\ \frac{J-1}{N(t)} + \frac{1}{N(t)} \times 1, & (K=2) \end{cases}$$

が得られる. $r_2 \in (0, 1]$ に対して

$$\frac{J-1}{N(t)} < r_2 \leq \frac{J}{N(t)}$$

を満たす $J = \hat{J} \in \{1, 2, \dots, N(t)\}$ がただ一つ存在する. r_2^J を

$$r_2^J := r_2 - \frac{\hat{J}-1}{N(t)}$$

によって定義すると, r_2^J のとり得る範囲は $0 < r_2^J \leq 1/N(t)$ である. したがって, もし $0 < r_2^J N(t) \leq b/(b+d)$ ならば $K = 1$ であり, $b/(b+d) < r_2^J N(t) \leq 1$ ならば $K = 2$ となる. 一方, 集団ベースのアルゴリズムの場合は $0 < r_2 \leq b/(b+d)$ ならば $K = 1$, $b/(b+d) < r_2 \leq 1$ ならば $K = 2$ を選択する. したがって, 個体ベースの直接法のアルゴリズムは従来の直接法の自然な拡張であることがわかる (図 1 参照).

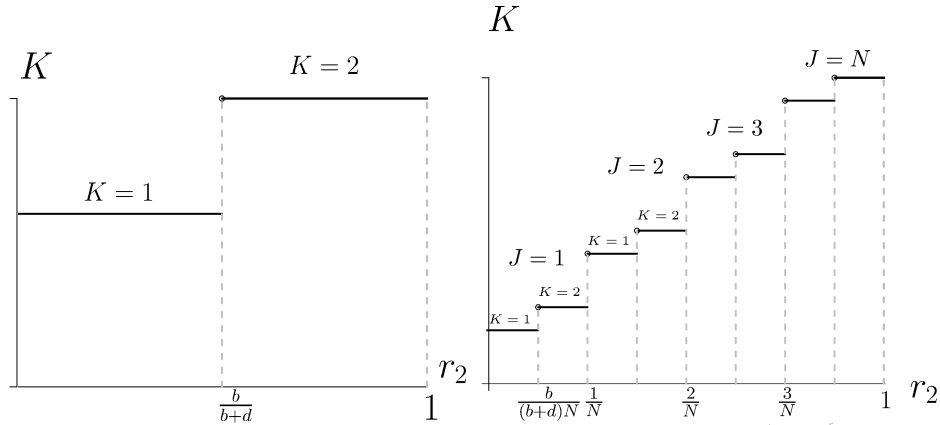


Figure 1: (2.3) に対応するアルゴリズム第 2 段階の可視化. 左パネル: 集団ベースの Gillespie 直接法. 右パネル: 個体ベースの Gillespie 直接法 (例として $N = 4$ とした場合).

3 個体・集団ベースのアルゴリズム比較

3.1 有限集団の Nicholson blowfly モデル

i -状態空間が個体の存在, 属する集団, 年齢によって特徴付けられる系を考えよう. Nisbet and Gurney によって提案された Nicholson Blowfly モデルは, 成熟にかかる時間遅れが個体群振動 (population cycle) を引き起こす古典的な例として知られている. Nicholson Blowfly モデルは, 繁殖が可能か否かで集団を成体 (adult) もしくは幼生 (juvenile) の 2 つのステージに分割し, 系の状態は完全に均一であると仮定している. 個体は, ある年齢 τ に到達すると成熟して繁殖可能になると仮定する. 決定論的 Nicholson Blowfly モデルは, 以下の時間遅れをもつ方程式によって与えられる.

$$\begin{aligned} \frac{dx_A}{dt} &= \beta e^{-d_J \tau} e^{-x_A(t-\tau)/c} x_A(t-\tau) - d_A x_A(t), \\ \frac{dx_J}{dt} &= \beta e^{-x_A(t)/c} x_A(t) - \beta e^{-d_J \tau} e^{-x_A(t-\tau)/c} x_A(t-\tau) - d_J x_J(t) \end{aligned} \quad (3.1)$$

ここで $x_A(t)$ および $x_J(t)$ はそれぞれ時刻 t における成体および幼生の個体数, β は 1 個体あたりの繁殖率, d_A および d_J はそれぞれ成体および幼生の 1 個体あたりの死亡率, c は非線型効果に関わる定数, τ は成熟にかかる時間遅れ (maturation delay) を表す [1].

個体ベースの Nicholson Blowfly モデルを構築しよう. 系は 2 つの集団 $G = \{1, 2\}$ から構成されると仮定し, インデックス 1 を幼生, インデックス 2 を成体とする. 個体がもつ反応の種類は出生・死亡とする ($q_j = 2$). 幼生個体は出生しないので出生の傾向は $p_{j1}(t, \mathbf{x}) \equiv 0$, 死亡の傾向は $p_{j2}(t, \mathbf{x}) \equiv d_J$ とする. 一方, 成体個体は傾向 $p_{j1}(t, \mathbf{x}) = \beta \exp(-\frac{x_A}{c})$ によって出生し, $p_{j2}(t, \mathbf{x}) \equiv d_A$ によって死亡すると仮定する. 生まれてくる個体は全て幼生で, 幼生はある年齢 τ に達すると成熟して成体になる. 加齢により, 時間ステップ Δt 経過毎に全ての個体の年齢は Δt 増加し, $a_j(t + \Delta t) \geq \tau$ となった個体は幼生から成体になる. 同様に, 集団ベースの Nicholson Blowfly モデルも構築できる. 成体, 幼生ステージの個体数をそれぞれ $x_J(t)$, $x_A(t)$ とする. 傾向関数 $\tilde{p}_{ik}(t, \mathbf{x})$ は, 幼生ステージ, 成体ステージそれ

それぞれにおける個体群出生率および死亡率を用いて $\tilde{p}_{11}(t, \mathbf{x}) \equiv 0$, $\tilde{p}_{12}(t, \mathbf{x}) \equiv d_J x_J$ (幼生), $\tilde{p}_{21}(t, \mathbf{x}) \equiv \beta \exp(-\frac{x_A}{c}) x_A$, $\tilde{p}_{22}(t, \mathbf{x}) \equiv d_A x_A$ (成体) とする (表 1 参照).

幼生・成体各ステージ内の個体は異なる年齢をもつため, 本小節で考察する系は部分的に均一な集団からなる系である. ここで, 部分的に均一な集団の個体群ダイナミクスに対しても, 個体が集団内で機能的に同等であれば集団ベースの Gillespie 直接法を適用できることに注意する. 3.1 小節で述べたように, 個体ベースおよび集団ベースのアルゴリズムの傾向関数は表 1 の通りである.

個体ベース	幼生	成体	集団ベース	幼生	成体
出生	0	$\beta \exp(-\frac{x_A}{c})$	出生	0	$\beta \exp(-\frac{x_A}{c}) x_A$
死亡	d_J	d_A	死亡	$d_J x_J$	$d_A x_A$

Table 1: 個体ベースおよび集団ベースのアルゴリズムの傾向関数

3.2 個体および集団ベースモデルの計算比較

図 2 はそれぞれ個体ベース (左) および集団ベース (右) のアルゴリズムを適用した結果である*. なお, 結果を比較するため決定論的 Nicholson blowfly モデルの解軌道図も併せて表示した. 末尾の (i) は個体ベース, (p) は集団ベース, direct は Gillespie 直接法, first reaction は最速反応選択法を表す†. いずれの場合も, (3.1) の解と比べて定性的に大きな差は見られない結果が得られた. 予想通り, 集団ベースのアルゴリズムは個体ベースのアルゴリズムよりも速く, 2 倍以上の差が見受けられた (表 2 参照).

手法	direct(i)	first reaction(i)	direct(p)	first reaction(p)
時間 (秒)	85	141	43	43

Table 2: 個体・集団ベースの Nicholson Blowfly モデルに対する確率シミュレーション. アルゴリズム間の計算時間の比較.

3.3 個体抽出法の比較

個体抽出は反応が起こる集団から 1 個体をサンプリングする方法であるため, 目的に応じていくつかの手法が考えられる. 決定論方程式と異なり, 集団ベースの Nicholson Blowfly モデルの各ステージ内の個体は, 出生時期が異なるため年齢も異なる. 幼生個体が死亡する際

*パラメーターの値は $\beta = 8.5$, $d_J = 0.0060455567$, $d_A = 0.27$, $c = 600$, $\tau = 15.6$, 初期個体数を $x_A(0) = 5000$ とした (以下同様). 同じ seed を用いて疑似乱数の生成を行ったため, 得られた結果は純粋にスキームの違いを反映している. CPU は XEON CPU 3.33 GHz, インテルコンパイラを利用した.

†最速反応選択法 (the first reaction method) とは, 仮に全ての反応が起こった場合, その中で最も待ち時間が最短の反応を選択するアルゴリズムである [2]. 最速反応選択法を用いて計算した反応が起きる確率は, 直接法を利用した場合と数学的に等しいことが証明されている [2]. 最速反応選択法を利用したアルゴリズムは直接法に比べて非効率であるため, 実用上で利用されることはない. ただし, 最速反応選択法は導出が直感的にわかりやすいため, その効率いかに依らず生物学の分野における確率シミュレーションにおいて採用されている場合がある. 直接法との計算時間比較のため, 本稿では最速反応選択法を用いた計算結果も載せている.

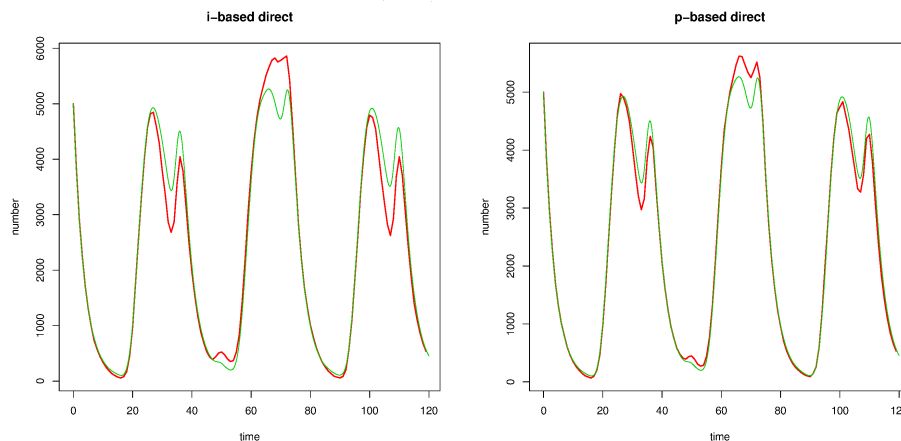


Figure 2: 個体ベースおよび集団ベースの Gillespie 直接法 (右). 赤実線: 確率シミュレーション, 緑点線: 決定論方程式 (3.1) の近似計算

に年齢の高い幼生個体を選択的に抽出するような方法では, 集団レベルで個体の成熟が遅れて, ダイナミクスの定性的な性質に影響する可能性が考えられる. 以下ではこの可能性について, 実際に異なる個体抽出法を適用して行った確率シミュレーションの結果を比較する. ただし, 以下で用いる方法のいくつかは人為的であるため, 生物学的な正当性に欠けるものも存在する. したがって, 得られた結果を生物学的に解釈することは行わない.

無作為選択法 (random-selection) は, 乱数に従って無作為に個体を選択する方法である. 意図的選択法 (intentional-selection) では予め定めておいた番号の個体を選択するが, その番号が個体数よりも多い場合は, 乱数に従って無作為選択し直す方法である. 個体数が多い場合, ある程度小さい番号を与えておくと無作為抽出に近い形で個体を選択できる上, 計算時間も短くなる. 最年少選択法 (select-by-juniority) は年齢に依存した選択方法で, 集団の中でもっとも年齢の低い個体を選択する. 一方, 最年長選択法 (select-by-seniority) は集団の中でもっとも年齢の高い個体を選択する.

図 3 では, これら 4 つの方法を用いて個体抽出を行った結果を示す. 図 3 左上パネルは無作為選択法, 右上パネルは意図的選択法, 左下パネルは最年少選択法, 右下パネルは最年長選択法を用いてシミュレーションを行った結果である. 意図的選択法では選択する個体の番号を 1000 に固定した. 計算時間は表 3 の通りである.

手法	無作為選択	意図的選択	最年少選択	最年長選択
時間 (秒)	43	33	17	22

Table 3: 集団ベースの Nicholson Blowfly モデルに対する確率シミュレーション.

最年長選択法では (3.1) の解と比べて 1 つめと 3 つめのピークに差が見られるが, その他 3 つの方法では, (3.1) の解と比べて大きな差は見られない. 最年長選択法では, 幼生が死亡するときも年齢の最も高い個体を選択されるため, 成熟が遅れて個体数のピークの位置がずれてしまうと考えられる. この結果, 時刻 $t \simeq 50$ あたりに遅れて小さなピークが現れる

など, (3.1) の解と比べて不規則な動きをすることになる. 一方, 作為的選択法や最年少選択法では, 集団レベルで成熟の遅れを引き起こすような影響はないと考えられる. 以上の結論は単に技術的な問題, もしくは対象としている問題特異的に生じた可能性もあるが, 部分的に均一な集団に対して集団ベースのアルゴリズムを用いて確率シミュレーションを行う場合は, 個体抽出スキームの違いを常に考慮に入れておく必要があることを示唆している[‡].

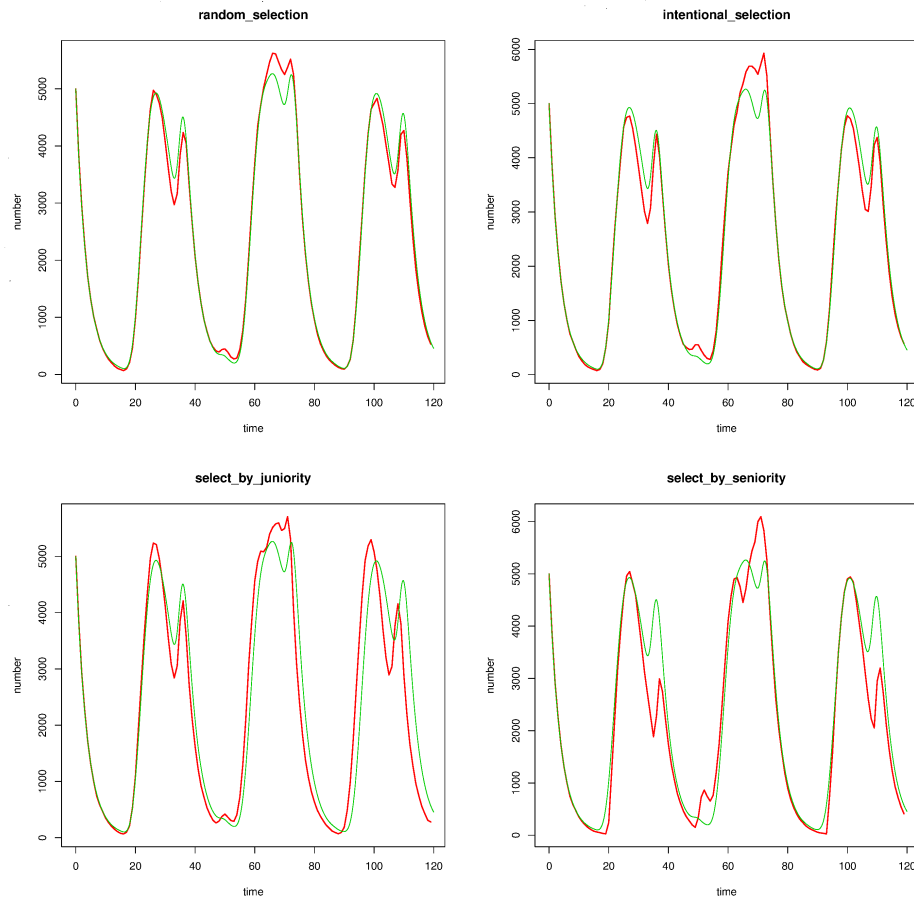


Figure 3: 無作為選択法 (左上), 作為的選択法 (右上), 最年少選択法 (左下) および最年長選択法 (右下). 赤実線: 確率シミュレーション, 緑点線: 決定論方程式 (3.1) の近似計算

[‡]プログラム上で個体は線型片方向連結リストの要素として実装した. このため, 反応に関わる個体の抽出は, リストの先頭から検索する必要がある. 検索優先法 (detected-the-first) は, 条件にマッチする個体が見つかった時点でその個体を選択する, 計算量が最も少ない方法である. Nicholson Blowfly モデルの確率シミュレーションにかかった時間は 12 秒で, 他の手法と比べて最も速い方法である. 検索優先法は, 今回考えている問題特異的に最年長選択法 (図 3 右下パネル) と同じような結果を示す. 実際, シミュレーションを行った初期条件では初めに成体の死亡が立て続けに起こり, 幼生個体は連結リストの先頭から並ぶ形となる. もし幼生の死亡が反応として選択された場合, 検索優先法は年齢の最も高い幼生個体を抽出する. したがって, 検索優先法と最年長選択法は同じ操作を行っていると考えられる. このような検索優先法が最年少選択法と一致してしまう現象は今回考えている問題特異的に生じた可能性が高いが, 計算速度の観点から最良の個体抽出法を利用する場合には, 注意が必要であることを示している.

4 まとめ

Nicholson blowfly モデルに対して, 複数の個体選択法を適用して個体ベースおよび集団ベースアルゴリズム間でどのような違いが生じるかを比較検証した. 計算時間については, どのような個体選択法であっても集団ベースアルゴリズムは個体ベースアルゴリズムよりも高速であることがわかった. 今回実施した確率シミュレーション計算と決定論方程式の数値計算の結果と比較する限り, 個体抽出法は計算結果に大きな違いを生じる原因となり得ることがわかった. 計算速度を上昇させるため意図的な個体抽出を行うと, 不適切な確率シミュレーション結果が得られる可能性が示唆された. 一方, 無作為抽出に相当する方法だと決定論方程式の結果とほぼ同様の確率シミュレーション結果が得られたことから, 個体抽出方法としては無作為抽出に近い方法を選択する方が望ましいと考えられる.

References

- [1] S. P. Ellner and J. Guckenheimer. *Dynamic Models in Biology*. Princeton University Press, (2006).
- [2] 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).
- [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] S. Nakaoka and K. Aihara. Stochastic simulation of structured skin cell population dynamics. *Journal of Mathematical Biology*, **66** pp.807–835, (2013).