確率過程を用いる数理生物学

大槻 久

August 3rd, 2024

総研大・統合進化科学研究センター

概要

- 確率過程と数理生物学

具体的なトピック：

- 生物多様性と確率過程
- がんの進化と確率過程

確率過程と数理生物学

確率過程とは

確率過程 (stochastic process)

時間とともに確率変数の値が変わるようなモデル。

- 時間 t は以下のどれでも良い。
  - 離散時間：$t = 0, 1, \cdots$
  - 連続時間：$t \geq 0$
- 確率変数[*] X は以下のどれでも良い
  - 数やベクトル
  - 状態（例：{ 良い, 悪い }）やもっと複雑な対象（例：どのような樹形か）

以降、時刻 t における X の値を X_t と書く。

[*]数でない場合、「確率要素」と呼ばれる

確率過程によるモデル化

微分方程式などの決定論モデルでは、初期値を定めれば結果はいつも同じ。

確率過程を用いた確率モデルでは、初期値が同じでも異なる結果が得られる。
よって結果の統計的性質が意味を持つ。たとえば､､､

- 結果の確率分布
- 結果の期待値
- 結果の分散

数理生物学と確率モデル

不確定性を伴う様々な現象に応用されている。

例

- 遺伝子頻度の変化（例：分子進化の中立説）
- 突然変異の数（例：体細胞突然変異とがん）
- 個体群の絶滅確率（例：希少種の保全）
- 病原体の防除確率（例：感染症の水際対策）

生物多様性と確率過程

生物多様性

問：生物の多様性を決定する機構とは何だろうか？

その中でも、、、

種-個体数分布 (species-abundance distribution)

ある群集において、個体数がちょうど i である種は何種いるだろうか？

Fisher の対数級数分布

Fisher (1943)：イギリスにおける蛾やマレー半島における蝶の群集データを説明

Fisher の対数級数分布 (log-series distribution)

「i 個体からなる種の数は

$$\alpha \frac{x^i}{i} \quad (0 < x < 1)$$

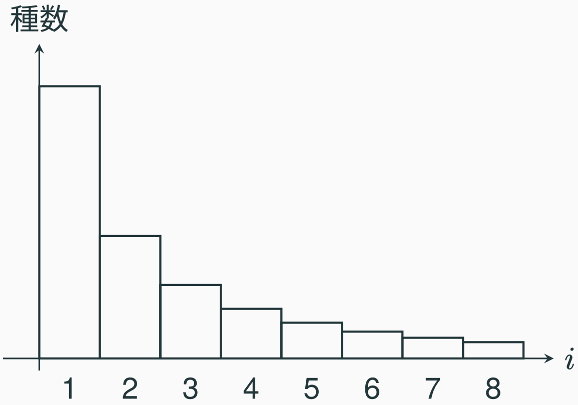
のような形で表される」

名前は対数関数の Taylor 展開

$$-\log(1-x) = x + \frac{x^2}{2} + \frac{x^3}{3} + \cdots$$

に因む。

Fisher の対数級数分布

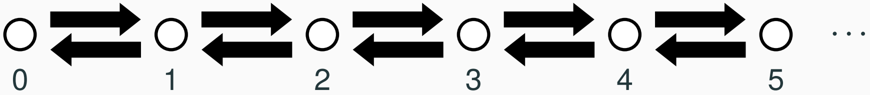


問：対数級数分布を生み出すメカニズムとは？

出生死亡過程

出生死亡過程 (birth-death process)

X が整数値で、毎回の増減が高々 ± 1 である確率過程のこと。



例）待ち行列の長さ、生物個体の数

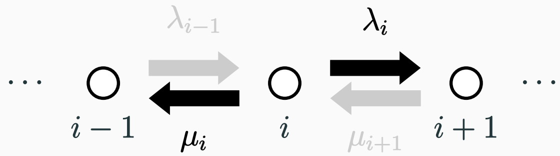
連続時間モデル

遷移率 (transition rate)

微小時間 Δt 内の遷移確率を、遷移率 λ_i, μ_i を用いて次で定める。

$$
\begin{aligned}
\mathrm{P}(X_{t+\Delta t} = i+1 \mid X_t = i) &= \lambda_i \Delta t + o(\Delta t) \\
\mathrm{P}(X_{t+\Delta t} = i-1 \mid X_t = i) &= \mu_i \Delta t + o(\Delta t) \\
\mathrm{P}(X_{t+\Delta t} = i \qquad\; \mid X_t = i) &= 1 - (\lambda_i + \mu_i)\Delta t + o(\Delta t)
\end{aligned}
$$

ただし $o(\Delta t)$ は Δt より高次の微小量。



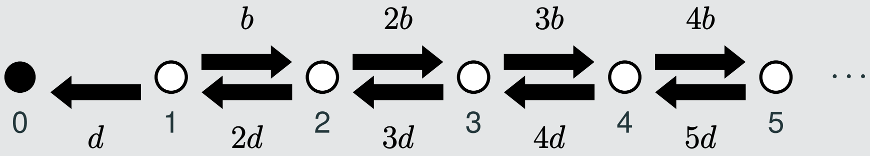
個体群動態への応用

X を個体数と考え、その増減は出生と死亡によって起こると考える。

特に以下のモデルは基本的である。

線形出生死亡過程 (linear birth-death process)

$\lambda_i = b\,i, \mu_i = d\,i$ で与えられる出生死亡過程のこと（b：出生率、d：死亡率）。



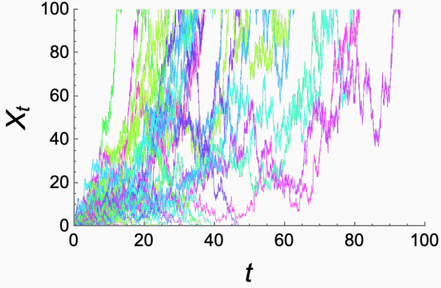
状態 $X = 0$ は吸収状態 (absorbing state) と呼ばれる。

個体群動態への応用

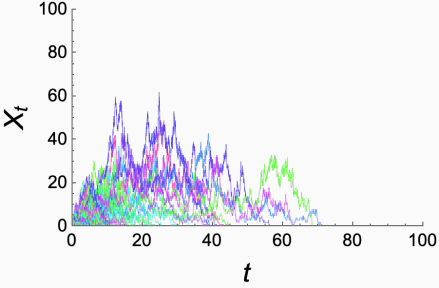
十分時間が経った後の振る舞い（$t \to \infty$）

$b > d$ **の時**　ある確率で $X = 0$ もしくは $X = \infty$ となる。

$b \leq d$ **の時**　確率 1 で $X = 0$ となる。



（$b > d$ の時）



（$b < d$ の時）

$b > d$ の時

時刻 t までに絶滅する確率 (extinction probability by time t)

$$e_{i,t} := \mathrm{P}(X_t = 0 \mid X_0 = i)$$

と定義する。

究極絶滅確率 (ultimate extinction probability)

$$e_i := \mathrm{P}(X_\infty = 0 \mid X_0 = i)$$

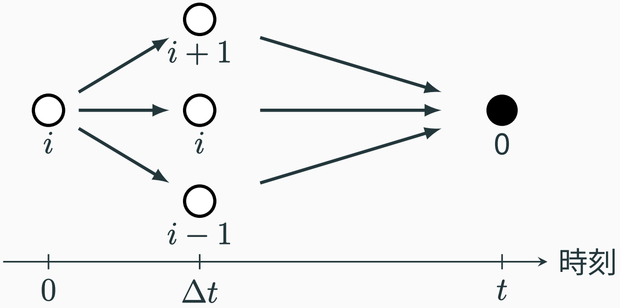
と定義する（上の $e_{i,\infty}$ に等しい)。

究極絶滅確率を求めよう。

究極絶滅確率の計算

時刻 0 で状態 i から始めて、時刻 t までに絶滅が起こるのは、以下の３つの場合。

1. 時刻 Δt で状態 $i+1$ に移っており、残り時間 $t-\Delta t$ の間に絶滅する。
2. 時刻 Δt で状態 $i-1$ に移っており、残り時間 $t-\Delta t$ の間に絶滅する。
3. 時刻 Δt で状態 i に留まっており、残り時間 $t-\Delta t$ の間に絶滅する。



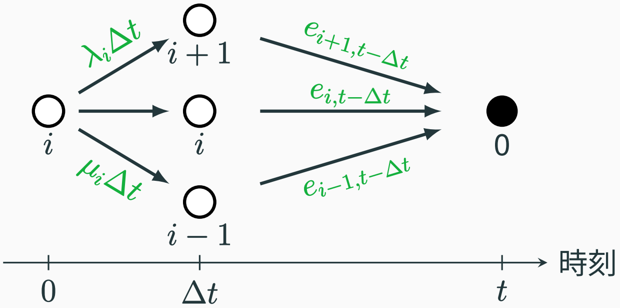
究極絶滅確率の計算

下図の３つの経路を考えることにより

$$e_{i,t} = (\lambda_i \Delta t) e_{i+1,t-\Delta t} + (\mu_i \Delta t) e_{i-1,t-\Delta t} + \{1 - (\lambda_i + \mu_i)\Delta t\} e_{i,t-\Delta t}$$

$t \to \infty$ の極限をとって

$$e_i = (\lambda_i \Delta t) e_{i+1} + (\mu_i \Delta t) e_{i-1} + \{1 - (\lambda_i + \mu_i)\Delta t\} e_i$$



究極絶滅確率の計算

$$
\begin{aligned}
e_i &= (\lambda_i \Delta t) e_{i+1} + (\mu_i \Delta t) e_{i-1} + \{1 - (\lambda_i + \mu_i)\Delta t\} e_i \\
0 &= (\lambda_i \Delta t) e_{i+1} + (\mu_i \Delta t) e_{i-1} - \{(\lambda_i + \mu_i)\Delta t\} e_i && \text{（移項）} \\
0 &= \lambda_i e_{i+1} + \mu_i e_{i-1} - (\lambda_i + \mu_i) e_i && \text{（}\Delta t\text{ で割る）}
\end{aligned}
$$

特に $\lambda_i = b\,i, \mu_i = d\,i$ を代入して、i で割って

$$0 = b e_{i+1} + d e_{i-1} - (b + d) e_i$$

境界条件は $e_0 = 1$。解は一意ではないが、（別の議論から実は）正解は

$$e_i = \left(\frac{d}{b}\right)^i$$

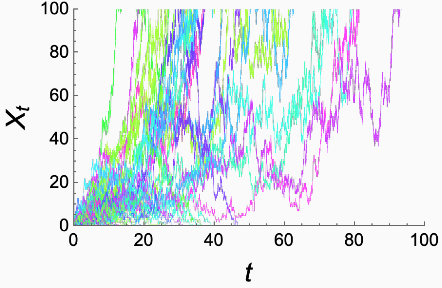
$b > d$ でも初期個体数 i が小さければ、
絶滅しうる！（人口学的確率性）

個体群動態への応用

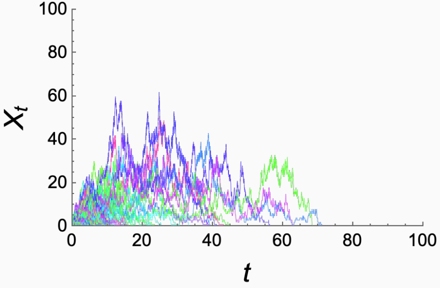
十分時間が経った後の振る舞い（$t \to \infty$）

$b > d$ **の時**　ある確率で $X = 0$ もしくは $X = \infty$ となる。

$b \le d$ **の時**　確率 1 で $X = 0$ となる。



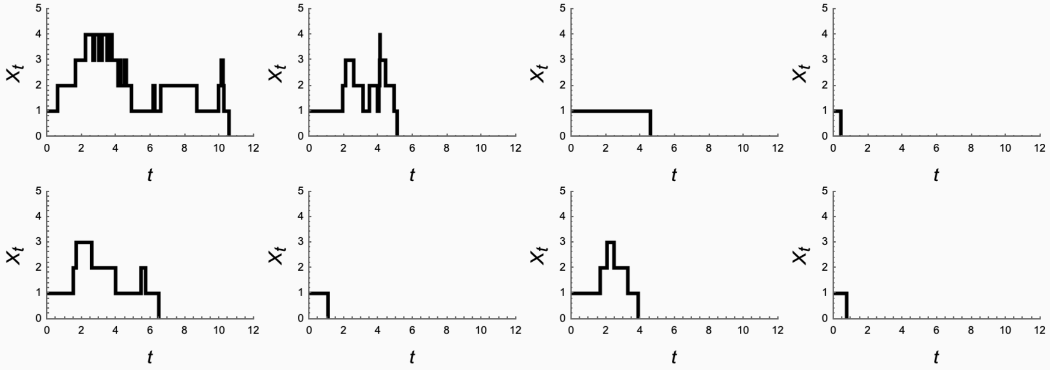
（$b > d$ の時）



（$b < d$ の時）

$b < d$ の時

$X_0 = 1$ から始めると、いずれは絶滅するが、一定時間集団は存続する。



群集モデルへの応用

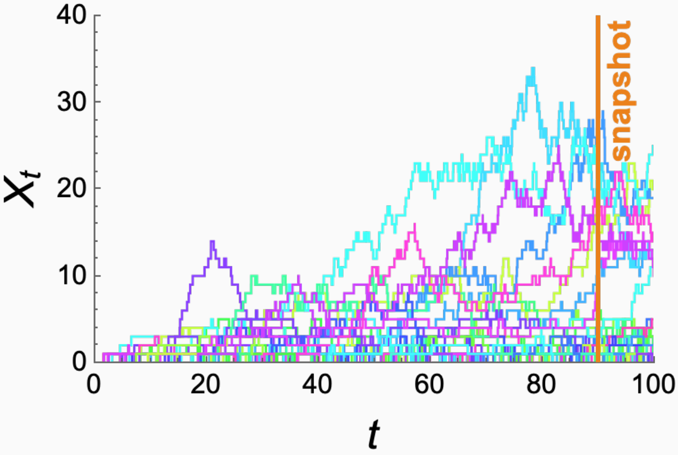
群集モデルの仮定

- 種 s は一様ランダムなタイミング t_s に起きた種分化で生まれ、$X_{t_s} = 1$ となる。
- この初期値から $b < d$ なる出生死亡過程に従い、種 s の個体数が増減。
- 上記が考えうる全ての種について起きる。
- ある瞬間 t におけるスナップショットが群集の姿。

問：この群集はどのような種-個体数分布を示すか？

群集モデルへの応用

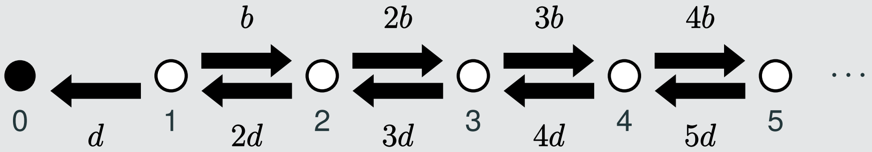
シミュレーション例：



種個体数分布

$X = 0$ なる種はそもそも観察されない。

したがって $X \geq 1$ である種に限って、その個体数 X の分布を調べれば良い。



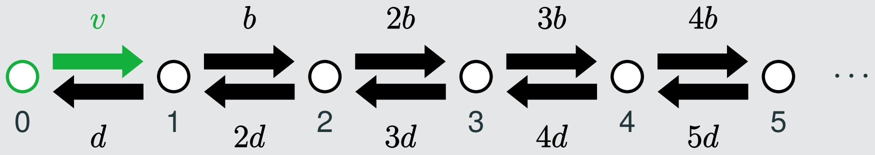
$X \geq 1$ という条件下で、確率過程がどの状態に滞在しやすいかを調べる問題。

種個体数分布の導出

数学的トリックを用いる。

- 下図の修正された確率過程を考えても問題ない。
- なぜなら、$X \geq 1$ なる各状態の滞在時間の比に影響はないから。
- $X_0 = 1$ から始まる線形出生死亡過程を、絶滅の度に繰り返すことに対応。

修正された確率過程



定常分布

定常分布 (stationary distribution)

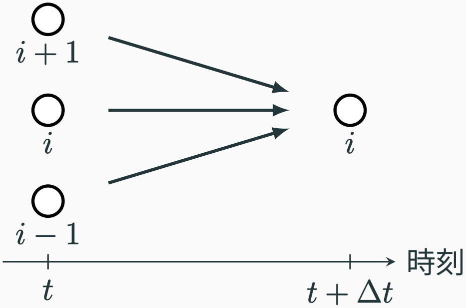
状態空間上の確率分布 $\{\pi_i\}$ で、時間が経過しても不変なもののこと。

分かりやすく言えば「どの状態に滞在しやすいか」。

定常分布の計算

時刻 $t + \Delta t$ で状態が i である確率は、以下の和。

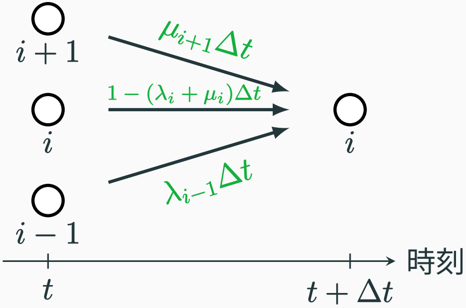
1. 時刻 t で状態 $i+1$ におり、Δt 後に状態 i にいる確率。
2. 時刻 t で状態 $i-1$ におり、Δt 後に状態 i にいる確率。
3. 時刻 t で状態 i におり、Δt 後に状態 i にいる確率。



定常分布の計算

よって下図より、定常分布 $\{\pi_i\}$ は以下の関係式を満たす。

$$\pi_i = \pi_{i+1}(\mu_{i+1}\Delta t) + \pi_{i-1}(\lambda_{i-1}\Delta t) + \pi_i\{1 - (\lambda_i + \mu_i)\Delta t\}$$



定常分布の計算

$$
\begin{aligned}
\pi_i &= \pi_{i+1}(\mu_{i+1}\Delta t) + \pi_{i-1}(\lambda_{i-1}\Delta t) + \pi_i\{1-(\lambda_i+\mu_i)\Delta t\} \\
0 &= \pi_{i+1}(\mu_{i+1}\Delta t) + \pi_{i-1}(\lambda_{i-1}\Delta t) - \pi_i(\lambda_i+\mu_i)\Delta t && \text{（移項）} \\
0 &= \mu_{i+1}\pi_{i+1} + \lambda_{i-1}\pi_{i-1} - (\lambda_i+\mu_i)\pi_i && \text{（}\Delta t\text{ で割り整理）}
\end{aligned}
$$

$\mu_0 = \lambda_{-1} = 0, \pi_{-1} = 0$ に注意すると、以下のようにまとめられる。

$$
\begin{cases}
0 = \mu_1\pi_1 - \lambda_0\pi_0 & (i=0) \\
0 = \mu_{i+1}\pi_{i+1} + \lambda_{i-1}\pi_{i-1} - (\lambda_i+\mu_i)\pi_i & (i \geq 1)
\end{cases}
$$

定常分布の計算

$$\begin{cases} 0 = \mu_1 \pi_1 - \lambda_0 \pi_0 & (i = 0) \\ 0 = \mu_{i+1}\pi_{i+1} + \lambda_{i-1}\pi_{i-1} - (\lambda_i + \mu_i)\pi_i & (i \geq 1) \end{cases}$$

数学的帰納法で、$\mu_i \pi_i = \lambda_{i-1}\pi_{i-1} \quad (i \geq 1)$ を示す。

証明

$i = 1$ は明らかに OK。i で成立すると仮定すると、

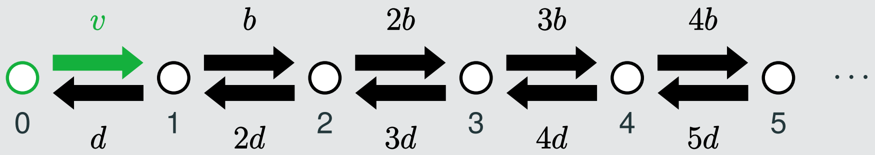
$$0 = \mu_{i+1}\pi_{i+1} + \lambda_{i-1}\pi_{i-1} - (\lambda_i + \mu_i)\pi_i$$

$$0 = \mu_{i+1}\pi_{i+1} + \lambda_{i-1}\pi_{i-1} - \lambda_i \pi_i - \mu_i \pi_i$$

$$0 = \mu_{i+1}\pi_{i+1} - \lambda_i \pi_i$$

より $i + 1$ でも OK。□

定常分布の計算



先程の結果より

$$
\begin{aligned}
\pi_1 &= \frac{\lambda_0}{\mu_1}\pi_0 = \frac{v}{d}\pi_0 &&= \frac{v}{b}\frac{\left(\frac{b}{d}\right)^1}{1}\pi_0 \\
\pi_2 &= \frac{\lambda_1}{\mu_2}\pi_1 = \frac{b}{2d}\cdot\frac{v}{b}\frac{\left(\frac{b}{d}\right)^1}{1}\pi_0 &&= \frac{v}{b}\frac{\left(\frac{b}{d}\right)^2}{2}\pi_0 \\
\pi_3 &= \frac{\lambda_2}{\mu_3}\pi_2 = \frac{2b}{3d}\cdot\frac{v}{b}\frac{\left(\frac{b}{d}\right)^2}{2}\pi_0 &&= \frac{v}{b}\frac{\left(\frac{b}{d}\right)^3}{3}\pi_0
\end{aligned}
$$

定常分布の計算

修正された出生死亡過程の定常分布

$$\pi_i = \frac{v}{b}\frac{\left(\frac{b}{d}\right)^i}{i}\pi_0 \quad (i \geq 1)$$

π_0 は $\pi_0 + \pi_1 + \cdots = 1$ より求まるが、具体的な形は必要ない。

観察される種の定常分布

$X \geq 1$ の条件下で、$X = i\,(\geq 1)$ である条件付き確率は

$$\phi_i := \mathrm{P}(X = i \mid X \geq 1) = \frac{\mathrm{P}(X = i \text{ and } X \geq 1)}{\mathrm{P}(X \geq 1)} = \frac{\pi_i}{1 - \pi_0}$$

Fisher の対数級数分布を生み出すメカニズム

つまり、次のことが分かった。

線形出生死亡過程でモデル化される群集

$X \geq 1$ の条件下で、群集における種の個体数が $X = i$ である確率は

$$\phi_i = C\frac{\left(\frac{b}{d}\right)^i}{i} \qquad (C \text{ はある定数})$$

の形をしている。

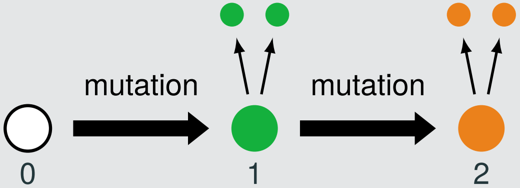
$x = b/d$ とおけば、これは Fisher の対数級数分布にほかならない！

がんの進化と確率過程

がんと進化

発がんは体細胞の進化として理解できる。

突然変異の蓄積

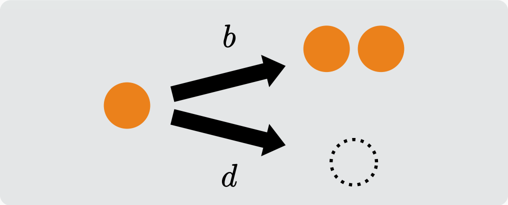


がん細胞増殖の単純モデル

各時刻 $t = 0, 1, \cdots$ で細胞ごと独立に、

- 確率 b で 2 つに増殖する。
- 確率 d で死滅する。

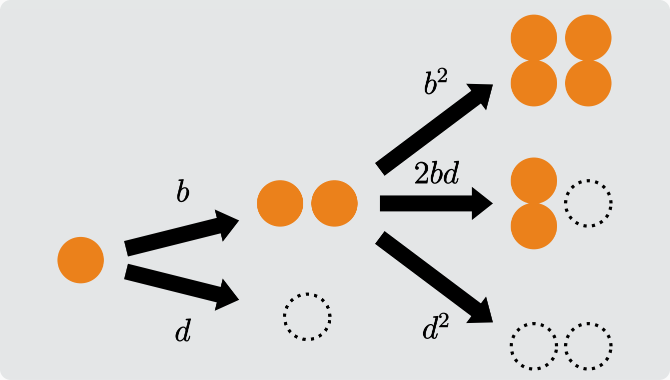
ここで $b + d = 1$ とする。



問：がん細胞 1 個から始まって、最終的にがん細胞が消失する確率は？

がん細胞増殖の単純モデル

時刻 $t = 2$ までの例



分枝過程

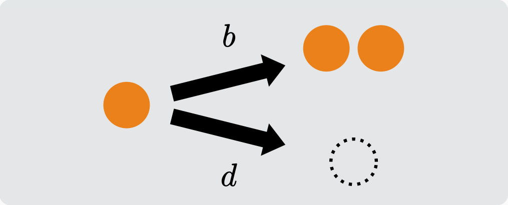
分枝過程 (branching process)

粒子数 X についてのモデルで、各粒子の増え方が同一の確率分布に従い、かつ粒子ごとに独立であるモデルを、分枝過程と呼ぶ。

消失確率の漸化式

細胞 1 個から始めて、最終的に消失する確率を P とおく。

1. $t = 0$ で増殖 (確率 b) した場合
   — 消失のためには、2 細胞の子孫がいずれも消失しなければならない。
2. $t = 0$ で死滅 (確率 d) した場合
   — 消失が達成される。



以上の考察より、次の関係式が成り立つ。

$$P = d + bP^2$$

確率母関数

分枝過程では、次の数学的道具が活躍する。

確率母関数 (probability generating function)

分枝過程において 1 粒子が次の時刻に i 粒子になる確率を p_i とおく。この時 z の多項式

$$f(z) = p_0 + p_1 z + p_2 z^2 + \cdots = \sum_{i=0}^{\infty} p_i z^i$$

をこの分枝過程の確率母関数と呼ぶ。

定義より、いつも $f(1) = 1$ が成り立つ。

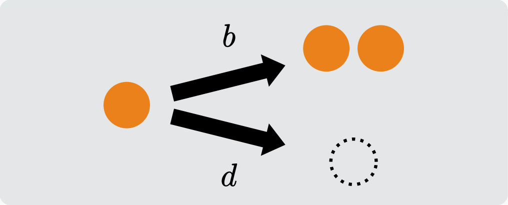
消失確率の漸化式

がん増殖モデルの確率母関数：

$$f(z) = d + bz^2$$

先程の消失確率の式は以下のように書き直せる。

$$P = f(P)$$



消失確率の計算

$$\begin{aligned} P &= f(P) \\ P &= d + bP^2 \\ (b+d)P &= d + bP^2 && (b+d=1) \\ 0 &= bP^2 - (b+d)P + d && \text{(移項)} \\ 0 &= (bP-d)(P-1) \end{aligned}$$

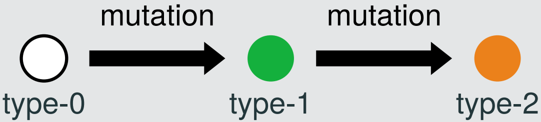
$P = (d/b), 1$ のどちらが真の消失確率だろうか？

定理

- $b > d$ ならば、消失確率は $P = d/b$ である。
- $b \leq d$ ならば、消失確率は $P = 1$ である。

発がん過程の確率モデル

突然変異の蓄積

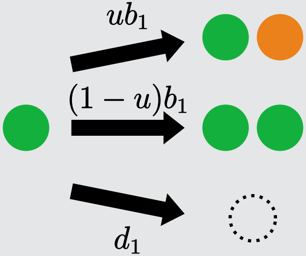


例：Tumor suppressor gene (TSG)

- Type-0 は正常細胞。
- Type-1 は TSG の 2 倍体の片方に突然変異。増殖速度は正常細胞と同じ。
- Type-2 は TSG の 2 倍体の両方に突然変異。異常増殖する。

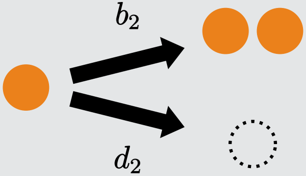
発がん過程の確率モデル

Type-1 のダイナミクス



- $b_1 = d_1 = 1/2$ を仮定。
- 突然変異率は $u\,(\ll 1)$。

Type-2 のダイナミクス



- $b_2 > d_2$ を仮定。

問：Type-1 細胞 1 つから始めたときの、がんが消失しない確率は？

マルチタイプ分枝過程 (multi-type branching process)

P_1：type-1 細胞 1 個から始めて、がんが消失する確率。

P_2：type-2 細胞 1 個から始めて、がんが消失する確率。

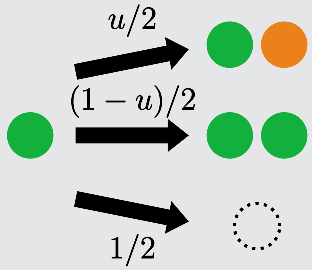
右図から

$$P_1 = \frac{1}{2} + \frac{1-u}{2}P_1^2 + \frac{u}{2}P_1P_2$$

確率母関数：

$$f_1(z_1, z_2) = \frac{1}{2} + \frac{1-u}{2}z_1^2 + \frac{u}{2}z_1z_2$$

Type-1 のダイナミクス



マルチタイプ分枝過程 (multi-type branching process)

P_1：type-1 細胞 1 個から始めて、がんが消失する確率。

P_2：type-2 細胞 1 個から始めて、がんが消失する確率。

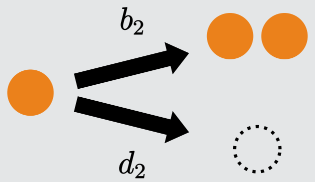
これは既に計算済み

$$P_2 = d_2 + b_2 P_2^2$$

確率母関数：

$$f_2(z_1, z_2) = d_2 + b_2 z_2^2$$

Type-2 のダイナミクス



- $b_2 > d_2$ を仮定。

マルチタイプ分枝過程 (multi-type branching process)

マルチタイプ分枝過程の確率母関数

タイプ $k\,(=1,2)$ の 1 粒子が次の時刻にタイプ 1 およびタイプ 2 の粒子をそれぞれ i_1 個と i_2 個作る確率を $p^{(k)}_{i_1,i_2}$ とおく。この時 z_1, z_2 の多項式

$$f_k(z_1, z_2) = \sum_{i_1=0}^{\infty}\sum_{i_2=0}^{\infty} p^{(k)}_{i_1,i_2} z_1^{i_1} z_2^{i_2}$$

をこの分枝過程のタイプ k の粒子の確率母関数と呼ぶ。

消失確率の計算

$$\begin{cases} P_1 = f_1(P_1, P_2) \\ P_2 = f_2(P_1, P_2) \end{cases} \quad \text{or} \quad \begin{cases} P_1 = \dfrac{1}{2} + \dfrac{1-u}{2} P_1^2 + \dfrac{u}{2} P_1 P_2 \\ P_2 = d_2 + b_2 P_2^2 \end{cases}$$

$P_2 = d_2/b_2$ が正しい消失確率であることは既に見た。これを第 1 式に代入して

$$P_1 = \frac{1}{2} + \frac{1-u}{2} P_1^2 + \frac{u}{2} P_1 \frac{d_2}{b_2}$$

この 2 次方程式の $0 \le P_1 \le 1$ の範囲にある唯一の解は

$$P_1 = 1 - \sqrt{\left(1 - \frac{d_2}{b_2}\right)} \sqrt{u} + o(\sqrt{u}) \qquad (o(\sqrt{u}) \text{ は } \sqrt{u} \text{ より高次の微小項})$$

がんのリスク評価

Type-1 細胞 1 つから始めたとき、がんが消失しない確率は

$$1 - P_1 \approx \sqrt{\left(1 - \frac{d_2}{b_2}\right)\sqrt{u}}$$

である（消失しない場合、type-2 細胞の個数はやがて ∞ に発散する）。

- リスクが $\sqrt{u}$ に比例する点が重要（u よりはるかに大きい！）。
  — type-1 細胞はやがて消失するが、消失までに時間がかかるので、それまでに type-2 細胞が生みだされる機会が十分にある。

まとめ

まとめ

- 出生死亡過程や分枝過程など、様々な確率過程モデルが数理生物学では用いられている。
- 決定論で起こり得ないことや、意外な予測が得られることがある。
- 決定論モデルと確率論モデルの違いを理解し、自己の研究にどちらを用いればよいかを判断できるようになることが大切。

参考文献

生物多様性と確率過程

- Volkov, I., Banavar, J. R., Hubbell, S. P., & Maritan, A. (2007). "Patterns of relative species abundance in rainforests and coral reefs." *Nature*, **450(7166)**, 45-49.

ガンの進化と確率過程

- Iwasa, Y., Michor, F., & Nowak, M. A. (2004). "Evolutionary dynamics of invasion and escape." *Journal of Theoretical Biology*, **226(2)**, 205-214.