ランダムな現象の数理モデル

舟木 直久（東大数理）

1　ランダムとは？

- 実際の結果があらかじめきちんと予想できないとき、そのような現象をランダムという。
- 社会生活の例
  - · ある電車に乗っている人の数
  - · 銀行のキャッシュコーナーに並んでいる人の数
- 自然科学・社会科学の例
  - · 物理：物質は原子・分子でできている。気体分子はランダムな位置にあると考えられる。
  - · 生物：ランダムに子供が生まれる（男女・人数など)。あるいは遺伝子の変異。
  - · 経済：株価はランダムに変動する。
- これらは本当はランダムでなく、きちんと予想できることかもしれない。

  たとえば電車の例では、あらかじめすべての状況、すなわち誰が何時何分に家を出て、バスに乗り、バスは駅まで何分かかり...... などの情報が完全にわかっていれば、ある電車に乗る人物は特定でき、したがって人数もわかるはずである。

  しかしこれは実際には不可能だろう。そこで細かい情報は捨て去りランダムと見なす。

  つまり電車の例では、曜日・時間ごとに何度も繰り返してデータをとり、x 人乗る確率は y % であると推定する。ただし

$$y = \frac{x\ \text{人であった回数}}{\text{調べた回数}} \qquad (\text{相対頻度})$$

  である。

  銀行の例でも同様である。このようにデータから確率を推定するのは**統計学**の役割である。

数理・情報一般「数学の現在・過去・未来」、2006 年度夏学期 4 月 24 日、5 月 1 日（月）

- 自然科学では少し違う態度をとる。現象の背後にある法則を仮定して、その仮説に基づき数学的論証を行うことにより実験データに合う結果が得られれば、この法則 (= モデル) は正しいと結論付ける。たとえば、気体分子運動論 (原子の運動法則) から経験則である熱力学を導くことや、遺伝に関するメンデルの法則などがそのような例である。

- **確率論**は、定められた確率法則から何が結論付けられるかを調べる数学の分野で、その意味で統計学とは異なる理念をもつ。

2　細胞分裂のモデル

上で述べたことを具体的に実行してみるために、細胞分裂に関する簡単なモデルを取り上げる。細胞があり、それが時間の進行とともに分裂していくとする。分裂の確率法則を設定し、それを数学的に解析することにより、その性質を調べることが目標である。

このモデルでは、時間は第 0 世代、第 1 世代、第 2 世代、..... (あるいは 0 秒、1 秒後、2 秒後、.....) というように離散的に進行していくものとする。ここで考える細胞分裂の確率的な時間発展の法則は次の 2 点に集約される。

(1) 1 個の細胞○が、次の世代で 2 個の細胞○○になる確率は p、消えて 0 個になる確率は $1-p$ である。

(2) 細胞分裂はすべて独立に起きる。

(2) は、細胞分裂が互いに影響しあうのではなく、それぞれ勝手に法則 (1) にしたがって起きるという意味であり、直感的に自然に理解してもらえばよい。p は $0 \leq p \leq 1$ を満たす定数で、時間発展を特徴付けるパラメータである。$p = 1/2$ ならば 1 個の細胞が 2 個になる確率・消える確率はともに 1/2 でバランスしている。$p > 1/2$ ならば細胞数は増加する傾向にあり、$p < 1/2$ ならば減少する傾向にあると言ってよいだろう。以下では、まずモデルを数学的に正確に記述し (言い換えれば、よい記号を導入して表現し)、それを解析することによってモデルのもつ性質を解明していく道筋を簡単に見てみたい。

2.1　細胞数を表す確率変数

第 n 世代の細胞数を X_n とおく。X_0, X_1 について書けば

$$X_0 = 1, \qquad X_1 = \begin{cases} 2, & \text{確率 } p, \\ 0, & \text{確率 } 1-p, \end{cases}$$

である。X_0 は必ず 1 であるが、X_1 はランダムで 2 か 0 の値をとる。これを

$$P(X_1 = 2) = p, \quad P(X_1 = 0) = 1 - p$$

と表す。P は確率 (probability) の頭文字である。X_2 については

$$
\begin{aligned}
&P(X_2=4)=p^3,\\
&P(X_2=2)=p\cdot 2p(1-p),\\
&P(X_2=0)=(1-p)+p(1-p)^2,
\end{aligned}
$$

である。$X_2=4$ となるためには、まず第 1 世代で確率 p で 2 個に分裂し、第 2 世代で 2 個の細胞がともにまた確率 p で 2 個に分裂せねばならない。$X_2=2$ であるには、まず確率 p で 2 個に分裂し、次の世代でどちらか一方が 2 個に分裂し他方は消滅する必要がある。その場合の数が 2 通りである。当然ながら X_2 に関する上記の 3 つの確率の和は 1 になる。X_n を確率変数という。

問題　X_3 について同様に確率を求めよ。一般に X_n についてはどうか？

2.2 細胞数の期待値

$$
E[X_n]=\sum_x xP(X_n=x)
$$

を X_n の期待値 (expectation) または平均値 (mean) という。$\sum\limits_x$ は X_n がとり得る値 x 全体についての和を表す。たとえば

$$
\begin{aligned}
&E[X_0]=1,\\
&E[X_1]=2\cdot p+0\cdot(1-p)=2p,\\
&E[X_2]=4\cdot p^3+2\cdot 2p^2(1-p)=4p^2,
\end{aligned}
$$

である。一般に

$$
E[X_n]=(2p)^n
$$

が成り立つ。直観的に考えれば、1 世代ごとに 1 個の細胞は平均的に $2p$ 個に分裂し、それを n 回繰り返すと $(2p)^n$ 個になるからである。あるいはもう少し正確に言うと、第 $n-1$ 世代に X_{n-1} 個の細胞があり、そのうち k 番目 $(1\le k\le X_{n-1})$ の細胞は次の世代で $X^{(k)}$ 個 $(X^{(k)}=2$ または 0 である) に分裂するとすれば

$$
X_n=\sum_{k=1}^{X_{n-1}} X^{(k)}
$$

と表すことができる。独立性から

$$
E[X_n]=E\left[\sum_{k=1}^{X_{n-1}} E[X^{(k)}]\right]=2pE[X_{n-1}]
$$

となり、帰納法により $E[X_n]=(2p)^n$ がわかる。特に

$$
\begin{aligned}
&p>1/2 \iff \lim_{n\to\infty} E[X_n]=\infty : \text{爆発}\\
&p=1/2 \iff \qquad E[X_n]=1 : \text{均衡}\\
&p<1/2 \iff \lim_{n\to\infty} E[X_n]=0 : \text{死滅}
\end{aligned}
$$

が得られる。これは予想される範囲内の結果であるが、次に少し違う観点からモデルを見てみよう。

2.3　死滅確率と生存確率

$$\begin{aligned} D &= P(\text{ある } n \text{ について } X_n = 0), \\ S &= 1 - D \end{aligned}$$

とおく。D を死滅確率 (death probability)、S を生存確率 (survival probability) という。D は細胞がいつかは死に絶える確率で、S は永久に生き残る確率である。

$$D_n = P(X_n = 0)$$

とおけば、D_n は第 n 世代までに細胞が死滅している確率を表す。細胞はいったん死滅すると新たに生まれることはないから、D_n は n について単調増大である。D は何世代目かはわからないが、いつか死滅してしまう確率である。D_n や期待値 $E[X_n]$ は第 n 世代までの情報から計算できる量であるが、D は決まった有限の時間だけを見ていては計算できず、ある意味で無限の時間に関係して決まる量である。実際、

$$D = \lim_{n\to\infty} D_n$$

が成立する。これは数学的には確率の σ-加法性 (高校までで習うのは有限加法性) から説明できる事柄である。σ-加法性については Lebesgue 積分論で学ぶが、ここでは直感的に理解しておけばよい。確率の σ-加法性の正確な定義を知りたければ「確率論」に関する教科書を参考にしてもらいたい。

容易にわかるように

$$D \geq D_1 = P(X_1 = 0) = 1 - p$$

だから $0 \leq p < 1$ ならば $D > 0$ である。したがって $S < 1$ である。実は、D は次の定理のように具体的に求めることができる。

定理 2.1. D は p の関数として

$$D = \bar{D} := \begin{cases} 1, & 0 \leq p \leq \dfrac{1}{2}, \\ \dfrac{1}{p} - 1, & \dfrac{1}{2} \leq p \leq 1, \end{cases}$$

で与えられる。

- D は p の連続関数であるが、微分可能な関数ではない。D_n は p の多項式だから滑らかな関数であるが、その極限は微分可能でない。これは、例えば熱力学における気体・液体の相転移を記述する相図などにおいても見られる。2 相が共存する領域では、圧力は密度の滑らかな関数ではない。このような関数を数学的に実現するには、有限系からでは無理で、無限系を考える必要があることが知られている。

- $p = 1/2$ のとき定理 2.1 から $D = 1$ である。つまり細胞はいつか必ず死滅する。期待値は $E[X_n] = 1$ であり均衡しているように見えるが、実際には確率 1 で死滅する。これは矛盾するように思えるがそうではない。細胞が死滅するまでには長い時間がかかる。一方小さい確率ではあるが細胞が多数生き残ることがある。この効果がバランスして期待値は均衡することになる。

- D が微分可能でない点 $p = 1/2$ を臨界点 (critical point) という。

2.4　定理 2.1 の証明

Step 1　死滅確率 D を第 0 世代と第 1 世代からみれば、等式

$$D = (1 - p) + pD^2$$

が得られる。これは D の 2 次方程式であり、その解は

$$D = 1 \quad \text{または} \quad \frac{1}{p} - 1$$

である。

Step 2　D_n と D_{n+1} について同様に

$$D_{n+1} = (1 - p) + pD_n^2$$

が成り立つ。帰納法により $D_n \leq \bar{D}$ を示そう。ただし $\bar{D}$ は定理の中で与えた p の関数である。まず $n = 1$ のとき

$$D_1 = 1 - p \leq \bar{D}$$

だから、これは成立している。n まで成立しているとすれば

$$D_{n+1} = (1 - p) + pD_n^2 \leq (1 - p) + p\bar{D}^2 = \bar{D}$$

だから $n + 1$ でも成立する。最後の等号は $\bar{D}$ が 2 次方程式を満たすことによる。故に $D_n \leq \bar{D}$ が言えた。したがって、極限をとれば

$$D = \lim_{n\to\infty} D_n \leq \bar{D}$$

が得られる。Step 1 と合わせて $D = \bar{D}$ がわかる。 □

ここまで論じてきたモデルは分枝過程 (branching process) とよばれ、細胞分裂の確率法則 (1) はもっと一般化して考えることができる。

2.5　ガン細胞の増殖モデル

分枝過程では空間的な構造は考慮に入れていなかった。細胞が分裂するとともに空間的に広がっていくモデルもいろいろ考えられている。講義では Eden's animal とよばれるモデルについて触れる予定である。