確率概念について

釜江哲朗（大阪市立大学）

＝１０月２６日津田塾大学，数学史シンポジュムでの講演＝

今日，コンピュータによって，確率微分方程式を数値的に解くといった複雑な確率現象のシミュレーションが行こなわれるようになっている．そこで用いられる乱数の多くは，線形合同法，M 系列，加算生成法等に代表される数値乱数で，これらのシミュレーションで要求される高度の独立性を満たすかどうか疑問視される．例えば，高嶋恵三氏が行っている Arcsin Law, First Passage Time 等にもとずく検定では，従来良い乱数とされてきた数値乱数の多くのものが棄却される [3].

Kolmogorov によって確立された今日の確率論は，確率を全測度 1 のルベーク測度としてとらえるもので，標本点の集合が確率 1 をもつという論拠によって，確率現象が標本点にどう反映されるかといった議論が可能となる．例えば，ベルヌイ試行において，1 回の試行の確率が標本点である数列の頻度として実現される．すなわち，頻度条件を満たす標本点の全体は確率 1 をもつのである．しかしながら，個々の標本点は確率 0 をもち，個々の標本点を集合に加減しても，集合の測度が 1 であるという性質は変わらない．すなわち，今日の確率論は，個々の数列がランダムかどうかといった議論には無力である．

他方，乱数の概念として知られているものは，von Mises による Kollektiv で，数列およびその部分列の頻度の安定性を表現したものである．すなわち，0 と 1 の列 $x_1x_2\cdots x_n\cdots$ が<u>2 進一様 Kollektiv</u>であるとは，任意の計算可能 (recursive) な関数 $f:\bigcup_{n=0}^{\infty}\{0,1\}^n\to\{0,1\}$ に対して，もし

$$\{k_1<k_2<\cdots\}:=\{n+1;f(x_1x_2\cdots x_n)=1\}$$

が正の upper density をもつならば

$$\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}x_{k_i}=\frac{1}{2}$$

が成立することをいう．

しかしながら，この定義では，一定のアルゴリスムによって生成される計算可能な数列は乱数とは認められない．生成速度と品質管理上の理由から，今日用いられる乱数はほとんどが数値乱数であり，上記の定義から排除されるものである．いかに理想化された概念とはいえ，現実の対応物のない乱数概念は空虚である．**このような観点から，現実に即しながら理論展開も十分可能な程度に理想化された乱数概念が求められる．**

E.Borel は，任意の長さのブロックに対して大数の強法則を実現する列として正規数列なる概念を導入した．例えば，<u>2 進一様正規数列</u>とは，0 と 1 の列 $x_1x_2\cdots x_n\cdots$ であって，任意の $k=1,2,\cdots$ と長さ k の 0 と 1 の列 $\xi_1\xi_2\cdots\xi_k$ に対して

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}1_{x_{n+1}\cdots x_{n+k}=\xi_1\cdots\xi_k}=2^{-k}$$

が成立するものをいう．このような列で計算可能なものは存在する．実際，Champernowne Number として知られている以下の 0 と 1 の列は 2 進一様正規数列である．

$$\underbrace{\underbrace{0}\,\underbrace{1}}_{\text{長さ1の列}}\;\underbrace{\underbrace{00}\,\underbrace{01}\,\underbrace{10}\,\underbrace{11}}_{\text{長さ2の列}}\;\underbrace{\underbrace{000}\,\underbrace{001}\,\underbrace{010}\,\underbrace{011}}_{\text{長さ3の列}}\cdots\cdots$$

どのような部分列の取り方が2進一様正規数列の頻度を保つであろうか．これについて以下の事実が知られている [1]．すなわち，正の upper density をもつ自然数の部分集合 $K=\{k_1<k_2<\cdots\}$ が，任意の2進一様正規数列 $x_1x_2\cdots$ に対して

$$\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}x_{k_i}=\frac{1}{2}$$

を満たすための必要十分条件は，Kが以下の意味で決定可能なときである．すなわち，各 $n=1,2,\cdots$ に対して，関数 $f_n:\{0,1\}^n\to\{0,1\}$ が存在して，$n\to\infty$ のとき

$$\liminf_{N\to\infty}\frac{1}{N}\sum_{i=0}^{N-1}1_{f_n(1_K(i+1),\cdots,1_K(i+n))=1_K(i+n+1)}$$

が1に収束することである．

乱数の概念を拡げるにしても，正規数列の範囲内には留まらなければならない．なぜなら，それは大数の強法則からの自然な要求だからである．

Kollektiv の概念よりやや強い乱数概念が Kolmogorov-Chaitin によって導入された Complexity を用いて与えられる．

計算可能な関数 $A:\bigcup_{n=0}^{\infty}\{0,1\}^n\to\bigcup_{n=0}^{\infty}\{0,1\}^n$ と $\xi\in\bigcup_{n=0}^{\infty}\{0,1\}^n$ が与えられたとき，A によるξのプログラムの長さ$K_A(\xi)$ を以下のように定義する．

$$K_A(\xi):=\begin{cases}\min\{n;A(\eta_1\eta_2\cdots\eta_n)=\xi\} & A(\eta)=\xi\text{となる}\\ & \eta\text{が存在するとき}\\ \infty & \text{そうでないとき}\end{cases}$$

上記の A が漸近的に最良であるとは，任意の計算可能な関数 $B:\bigcup_{n=0}^{\infty}\{0,1\}^n\to\bigcup_{n=0}^{\infty}\{0,1\}^n$ に対して，定数 Cが存在して，

$$K_A(\xi)\le K_B(\xi)+C$$

が任意の$\xi\in\bigcup_{n=0}^{\infty}\{0,1\}^n$ に対して成立することをいう．漸近的に最良な A の存在は知られている．また，そのような A,Bに対しては，定数 Cが存在して

$$|K_A(\xi)-K_B(\xi)|\le C$$

が任意の$\xi\in\bigcup_{n=0}^{\infty}\{0,1\}^n$ に対して成立する．この意味で最良の A に対する K_Aは定数差を除いて一意に定まる．これを単に Kと記し，Kolmogorov-Chaitin の Complexity Functionと呼ぶ．

0と1の列 $x_1x_2\cdots x_n\cdots$ が2進最大 Complexity 列であるとは，

$$\lim_{N\to\infty}\frac{1}{n}K(x_1x_2\cdots x_n)=1$$

が成立することをいう．2進最大 Complexity 列が2進一様 Kollektiv であることは容易にわかる．それ故，2進最大 Complexity 列を乱数列の定義として採用することは，上にのべた意味で不適当である．さらに，この Complexity Function Kは計算可能でないことも知られている．Kのもつ確率論的に重要な性質を保ちつつ，これを一般化した Complexity Function を用いて，より広い2進最大 Complexity 列を定義することが可能であろう．Complexity Function のこのような拡張は村松純氏と金谷 文夫氏によって，Universal Coding の観点からなされている [2]．

Complexity Function をどのように一般化し，一般化されたもののうちどれを新たに Kとして採用し，乱数概念としてこの Kにもとずく2進最大 Complexity 列を定義するのが適切かは今後の課題である．

参考文献

[1] Teturo Kamae, Subsequences of normal sequences, Israel J. Math. 16-2 (1973) pp.121-149.

[2] 村松 純, 金谷 文夫, "歪みを許した時の複雑度とレート歪み関数," 京都大学数理解析研究講究録, vol.975, pp.28–42, 1996.

[3] Keizo Takashima, Sojourn time test for maximum-length linearly recurring sequences with characteristic primitive trinomials, J. Jap.Soc.Comp.Stat. vol.7 (1994) 77 -87.