ランダムネスの源を求めて

―計算の視点から

杉田　洋

プロローグ

- W先生「コルモゴロフによるランダムネスの定義というのがあるんです．x がランダムであるとは，x を記述するのに x 自身より短い記述方法がないときを言うんだそうです」(1980年頃)

“Random”のコアイメージ

「考えを挟まないこと」

||

無作為

- ラプラスのランダムネス
- 硬貨投げの記録はどれか?
- コルモゴロフの乱数
- 乱数の性質

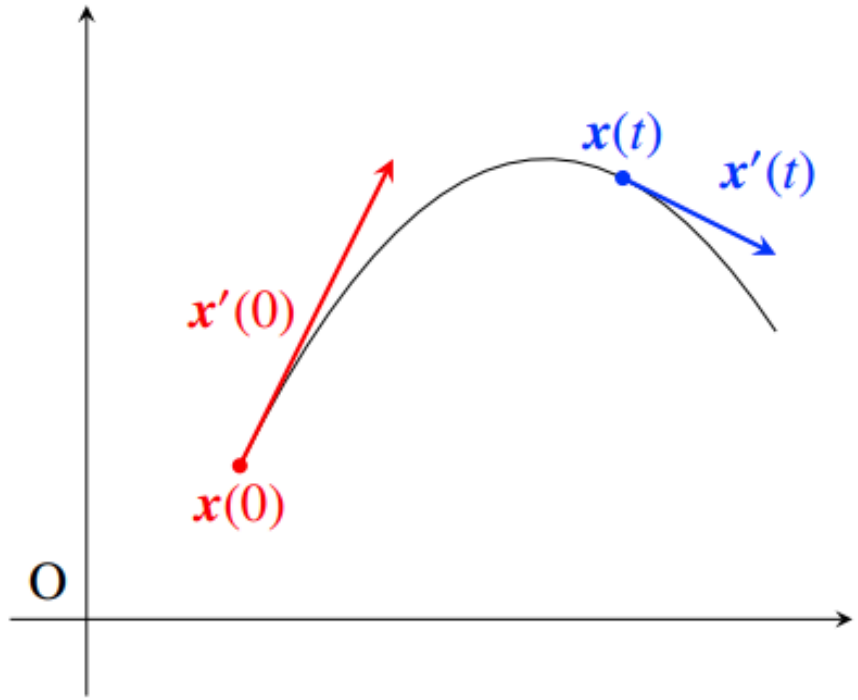
ラプラス「偶然は存在しない」

- 森羅万象は運動方程式によって
  決定論的に推移する.

初期値 + 運動方程式

⇩ 計算

未来の予測，コントロール



$\boxed{x(0),\ x'(0)} \longrightarrow \boxed{\text{運動方程式}} \longrightarrow \boxed{x(t),\ x'(t)}$

- 全能の英知（ラプラスの魔）
  - 初期値を完全に測定・設定ができる．
  - どんな運動方程式も解くことができる．

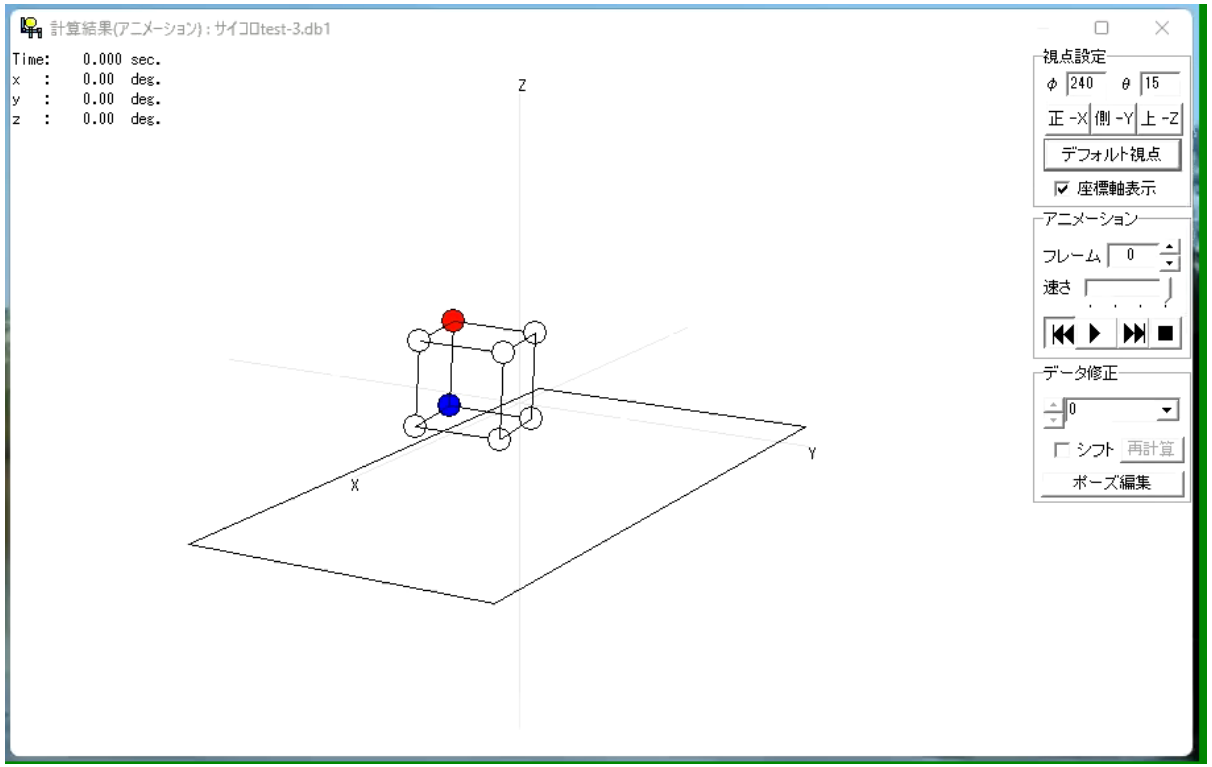
  ➡ 偶然は存在しない．

- 人間
  - 初期値の測定・設定が不完全または不可能である．
  - 複雑な運動方程式は解くことができない．

  ➡ 初期値に鋭敏に反応する系や，複雑な運動方程式を持つ系では，運動を予測したり，コントロールすることはできない．

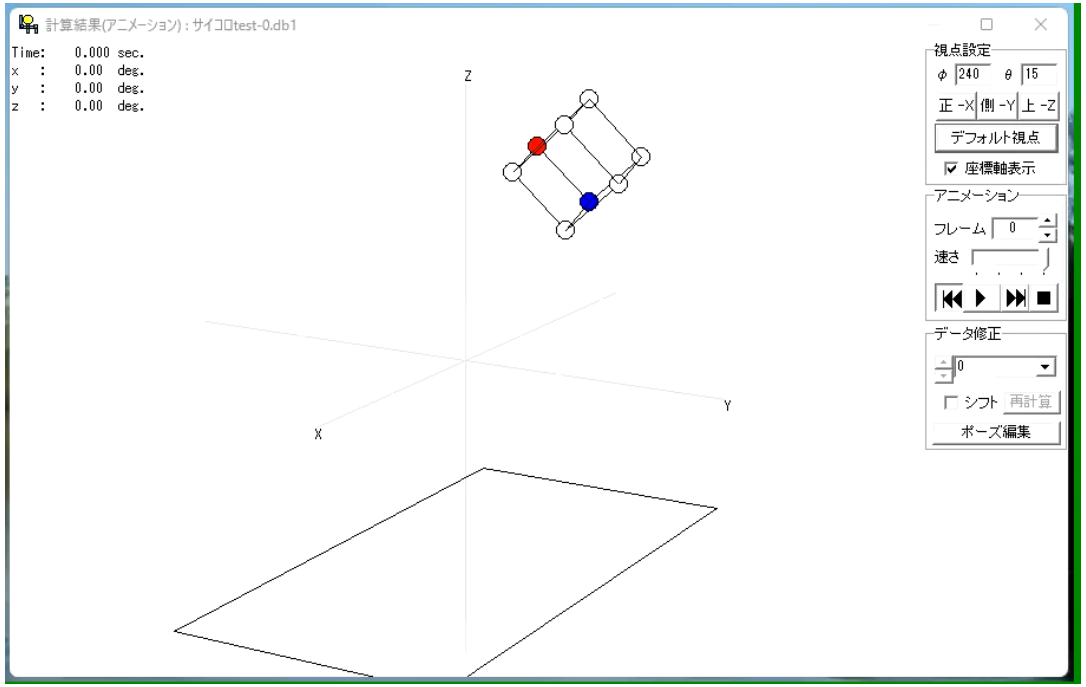
  ➡ 偶然が引き起こすランダムな現象と認識される．

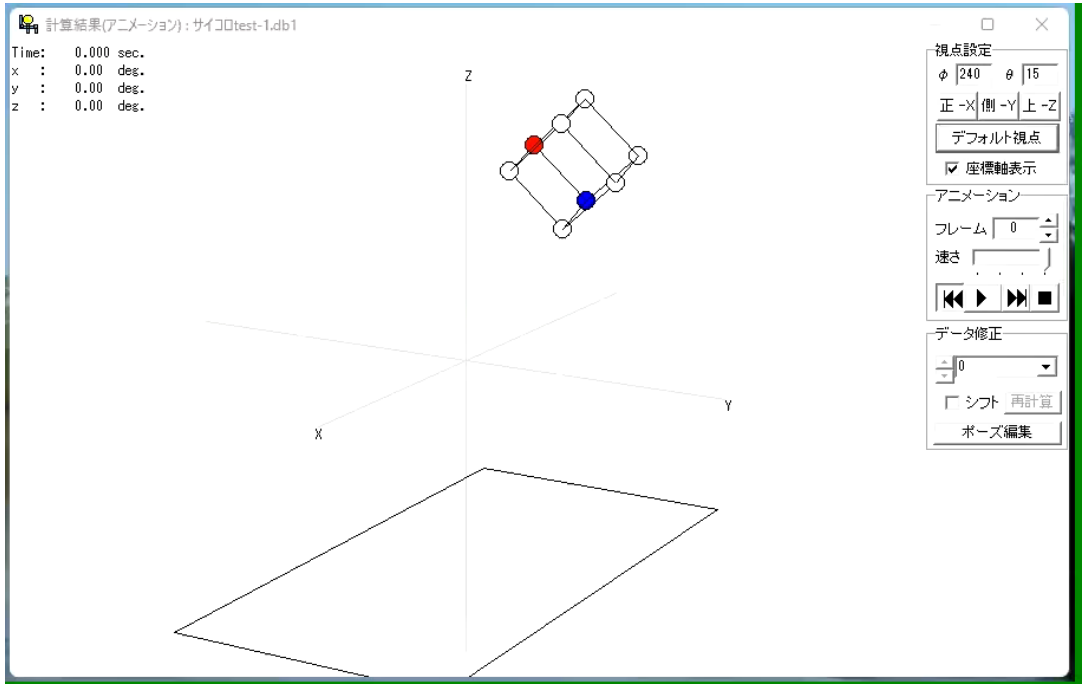
例1. サイコロ



特別な状況の下ではサイコロの運動を予測，コントロールできる．

例2. サイコロ





一般的な状況の下では，サイコロの運動は初期値に鋭敏に反応し，跳ね返りの運動は複雑で，出る目を予測・コントロールすることは不可能である．

ラプラスのランダムネス(1812年)

「考えを挟めないこと」

原因：初期値の測定・設定が不正確.
運動方程式が解けない.

注意：手間暇(設備，時間，労力，資金など)を
掛ければランダムネスが減る.

- ラプラスのランダムネス
- 硬貨投げの記録はどれか?
- コルモゴロフの乱数
- 乱数の性質

問題1.
硬貨を3回投げて表なら 1，裏なら 0 を記録した．それは次の (i) (ii) (iii) のうちどれか?

(i)　101

(ii)　111

(iii)　001

どれも普通にあり得る.

問題2.

硬貨を100回投げて表なら 1，裏なら 0 を記録した．それは次の (i) (ii) (iii) のうちどれか?

(i) 1010101010 1010101010 1010101010 1010101010 1010101010
1010101010 1010101010 1010101010 1010101010 1010101010

(ii) 1110110101 1011101101 0100000011 0110101001 0101000100
0101111101 1010000000 1010100011 0100011001 1101111101

(iii) 0010010000 1111110110 1010100010 0010000101 0100011000
0010001101 0011000100 1100011001 1000101000 1011100000

$\pi - 3$ の 2 進小数展開の小数100桁

問題2. 硬貨を100回投げて表なら 1，裏なら 0 を記録した．それは次の (i) (ii) (iii) のうちどれか?

(i) 1010101010 1010101010 1010101010 1010101010 1010101010
1010101010 1010101010 1010101010 1010101010 1010101010

(ii) 1110110101 1011101101 0100000011 0110101001 0101000100
0101111101 1010000000 1010100011 0100011001 1101111101

(iii) 0010010000 1111110110 1010100010 0010000101 0100011000
0010001101 0011000100 1100011001 1000101000 1011100000

正解

$\pi - 3$ の 2 進小数展開の小数100桁

- **ヒロミ:** (i) は規則的でランダムでない．(iii) $\pi - 3$ の2進小数展開の通りに表裏が出たとは思えない．よって答えは (ii).
- **カオル:** (i) (iii) はコンピュータで出力可能．だからランダムでない．よって答えは (ii).
- **シノブ:** どれも同じ確率 $1/2^{100}$ で起こるので判定できない.

問題2. 硬貨を100回投げて表なら 1，裏なら 0 を記録した．それは次の (i) (ii) (iii) のうちどれか?

(i) 1010101010 1010101010 1010101010 1010101010 1010101010 …

(ii) 1110110101 1011101101 0100000011 0110101001 0101000100 …

(iii) 0010010000 1111110110 1010100010 0010000101 0100011000 … $\pi - 3$ の 2 進小数展開の小数100桁

- **ヒロミ**：(i) は規則的でランダムでない．(iii) $\pi - 3$ の2進小数展開の通りに表裏が出たとは思えない．よって答えは (ii).

  - 硬貨投げによって規則的な{0,1}列は生じないと考えられる.
  - 硬貨投げによって意味のある{0,1}列は生じないと考えられる.
  - (i) (iii) はどちらも考えが挟まってる印象がある.

問題2. 硬貨を100回投げて表なら 1，裏なら 0 を記録した．それは次の (i) (ii) (iii) のうちどれか?

(i) 1010101010 1010101010 1010101010 1010101010 1010101010 …

(ii) 1110110101 1011101101 0100000011 0110101001 0101000100 …

(iii) 0010010000 1111110110 1010100010 0010000101 0100011000 … $\pi - 3$ の 2 進小数展開の小数100桁

- **カオル**: (i) (iii) はコンピュータで出力可能．だからランダムでない．よって答えは (ii).
  - 硬貨投げ：ランダムな{0,1}列を生成する.
  - コンピュータ：与えられたプログラム通りに動作し，ランダムな{0,1}列を出力できない.
- じつは(ii)もコンピュータで出力できる：

Print(“1110110101 1011101101 0100000011 0110101001 0101000100 0101111101 1010000000 1010100011 0100011001 1101111101”)

問題2. 硬貨を100回投げて表なら 1，裏なら 0 を記録した．それは次の (i) (ii) (iii) のうちどれか?

(i) 1010101010 1010101010 1010101010 1010101010 1010101010 …

(ii) 1110110101 1011101101 0100000011 0110101001 0101000100 …

(iii) 0010010000 1111110110 1010100010 0010000101 0100011000 … $\pi - 3$ の 2 進小数展開の小数100桁

- **シノブ**:どれも同じ確率 $1/2^{100}$ で起こるので判定できない．
  - 「これから硬貨を100回投げる．(i)(ii)(iii)のうちどれが最初に実現されるか」という問題なら**シノブ**の答えは正しい．
  - 問題2の場合は，統計的仮説検定を用いて判定するのが一般的．仮説「与えられた{0,1}列は硬貨投げの結果である」をどういう基準で検定するか，が問題である．**ヒロミ**と**カオル**の考えはその基準を直感的・定性的に与えているが，定量的な検定基準が欲しい．

- ラプラスのランダムネス
- 硬貨投げの記録はどれか?
- コルモゴロフの乱数
- 乱数の性質

問題3. 100回の硬貨投げを実現せよ.

- 硬貨を100回投げればよい.

問題4. 10^8回の硬貨投げを実現せよ.

- 硬貨を10^8回投げればよいが，膨大な時間が掛かり，実際には実行できない.
  - 硬貨を投げる速さを 1回/秒 とすれば，10^8回投げるのに掛かる時間は

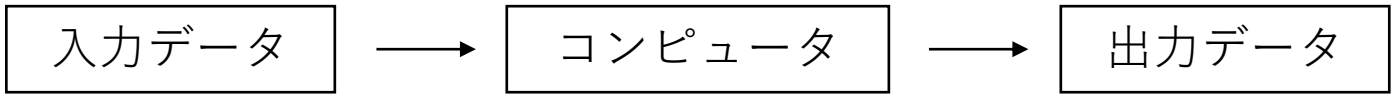
    10^8秒 $= 27{,}778$時間 $= 3$年2ヵ月(1日8時間労働で9年半)
  - その労働に見合う報酬 (1000 + 円/時) は 2800万円.
  - これだけの手間暇を掛ければ実行できる.

問題5. 10^8 回の硬貨投げをコンピュータで実現せよ.

- すべての入出力データはコンピュータでは有限$\{0,1\}$列として扱われる.
  従ってコンピュータは，有限$\{0,1\}$列(入力)を有限$\{0,1\}$列(出力)に変換する装置，数学的には関数である，と考える.

- 長さ n の$\{0,1\}$列を n ビットの$\{0,1\}$列という.

問題5. 10^8 回の硬貨投げをコンピュータで実現せよ.



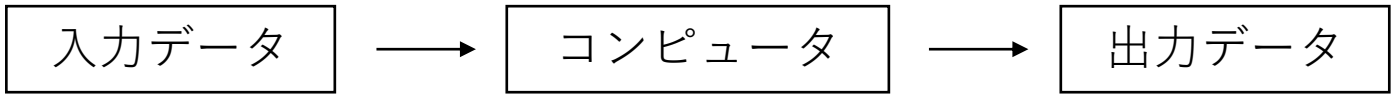
- すべての 10^8 ビットの$\{0,1\}$列 (総数 2^{10^8}) を出力できるようにしなければならない. そのためには, 10^8ビットの入力が少なくとも 1 つ必要となる.

$\because$ 10^8-1ビット以下の入力の総数は, m ビットの入力データの総数が 2^m 個だから, $2^0+2^1+\cdots+2^{10^8-1}=2^{10^8}-1$ で 1 つ足りない. 従って 10^8ビットの入力が必要となる.

- 同様に, 10^8-77ビット以下の入力の総数は $2^{10^8-76}-1<2^{10^8}/2^{76}$ となって, 全体の $1-2^{-76}$ を越える場合が出力できなくなってしまう.

$\because$ $2^0+2^1+\cdots+2^{10^8-77}=2^{10^8-76}-1$

問題5. 10^8 回の硬貨投げをコンピュータで実現せよ.



- $10^8\,(-77)$ビットのデータをキーボードから入力するには膨大な時間が掛かり，実際には実行できない.
  - $10^8(-77)$ビットの入力に要する時間を試算しよう．$0,\ 1$ を打ち込む速さを 1文字/秒 とすれば，求める時間は

$$10^8\text{秒} = 27{,}778\text{時間} = 3\text{年}2\text{ヵ月}(1\text{日}8\text{時間労働で}9\text{年半})$$

  - その労働に見合う報酬 $(1000 +$ 円/時$)$ は 2800万円.

- 入力は人間が行うので問題4と同じ困難が生じる.

問題5. 10^8 回の硬貨投げをコンピュータで実現せよ.

入力データ ⟶ コンピュータ ⟶ 出力データ

- 10^8 ビットの$\{0,1\}$列をコンピュータで出力させようとすると，ごく少数の出力可能な $\{0,1\}$ 列と大多数の出力(じつは入力)不可能な $\{0,1\}$ 列に分かれる.

乱数 (ランダムな $\{0,1\}$ 列)

コルモゴロフ複雑度 $K(x)$ と乱数

- 有限$\{0,1\}$列 xを出力するための入力(有限$\{0,1\}$列)のうち最も短いものを q_x とする． q_x の長さを$K(x)$で表す．
- x の長さが $n \gg 1$ のとき $K(x) \approx n$ を満たすx を乱数という．
  - 乱数と非乱数の間に厳格な境界線を引くことには意味がない．
  - x が硬貨投げの結果であるとすると， x はきわめて高い確率で乱数である．
- $K(x) < n$ のとき q_x は x を圧縮したデータと思うことができる．
  - 乱数 ⇄ 圧縮不可能　(非乱数 ⇄ 圧縮可能)

「x がランダムであるとは，x を記述するのに x 自身より短い記述方法がないこと」

コルモゴロフ複雑度 $K(x)$ と乱数

- $K(x)$ はすべての有限$\{0,1\}$列 x に対して定義されるが，それを計算するアルゴリズムは存在しない．
  - 与えられた x が乱数かどうかを判定するアルゴリズムは存在しない．
  - 問題2は(i)(ii)(iii)のうち「どれが乱数か」を問うている．(i)(iii)は乱数でないと言えるだろう．ただし(ii)も(そのまま Print文で書き出すという方法で)コンピュータで生成可能だから乱数とは言えない．長さが 10^8だったら(その方法は使えないので)はっきり判定できるだろう．

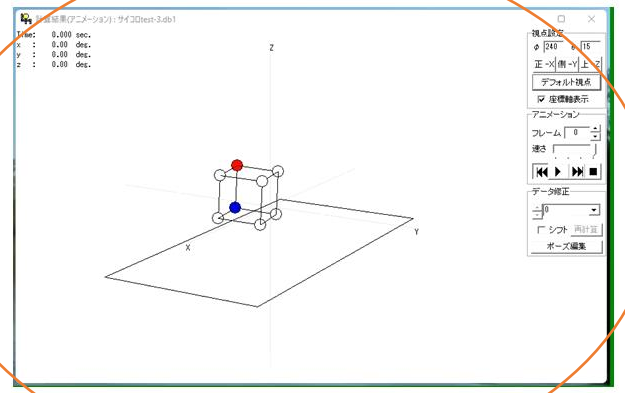
コルモゴロフの乱数(1960年代)

「考えを挟めない{0,1}列」

原因：乱数を出力するための入力データは大き過ぎて入力できない.
従って出力もできない.

注意：手間暇を掛ければ乱数は減る.

例1. サイコロ



特別な状況の下ではサイコロの運動を予測，コントロールできる．

問題1.

硬貨を3回投げて表なら 1，裏なら 0 を記録した．それは次の (i) (ii) (iii) のうちどれか?

(i) 101

(ii) 111

(iii) 001

どれも普通にあり得る．

問題2. 硬貨を100回投げて表なら 1，裏なら 0 を記録した．それは次の (i) (ii) (iii) のうちどれか?

(i) 1010101010 1010101010 1010101010 1010101010 1010101010 …

(ii) 1110110101 1011101101 0100000011 0110101001 0101000100 …

(iii) 0010010000 1111110110 1010100010 0010000101 0100011000 … $\pi - 3$ の 2 進小数展開の小数100桁

- **カオル**: (i) (iii) はコンピュータで出力可能．だからランダムでない．よって答えは (ii).
  - 硬貨投げ：ランダムな{0,1}列を生成する．
  - コンピュータ：与えられたプログラム通りに動作し，ランダムな{0,1}列を出力できない．
- じつは(ii)もコンピュータで出力できる：

Print(“1110110101 1011101101 0100000011 0110101001 0101000100 0101111101 1010000000 1010100011 0100011001 1101111101”)

問題3. 100回の硬貨投げを実現せよ．

- 硬貨を100回投げればよい．

問題4. 10^8回の硬貨投げを実現せよ．

- 硬貨を10^8回投げればよいが，膨大な時間が掛かり，実際には実行できない．
  - 硬貨を投げる速さを 1回/秒 とすれば，10^8回投げるのに掛かる時間は

    10^8秒 ＝ 27,778時間 ＝ 3年2ヵ月(1日8時間労働で9年半)
  - その労働に見合う報酬 (1000＋円/時) は 2800万円．
  - これだけの手間暇を掛ければ実行できる．

- ラプラスのランダムネス
- 硬貨投げの記録はどれか?
- コルモゴロフの乱数
- 乱数の性質

確率論の援用

- ランダムネスを語る上で「確率」は理論上必要ない.
  - 必要な「入力」の大きさ(ビット長) でランダムネスを測る.
    その際，確率は関与しない.
- 確率を語る上で「ランダムネス」は理論上必要ない.
  - 「出力」を場合に分けてその個数を数えて確率を計算する.
    その際，ランダムネスは関与しない.

- それにもかかわらず，確率論はランダムネスの分析に役に立つ.

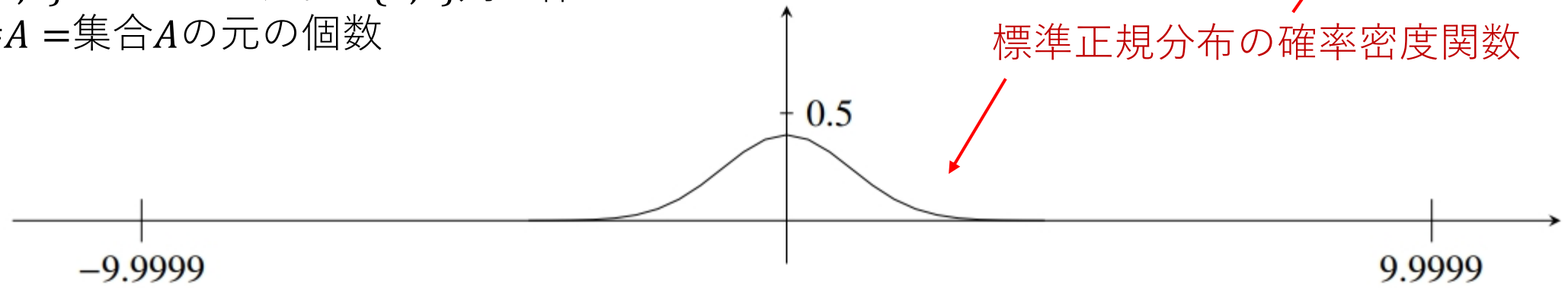
確率論の援用

中心極限定理(ラプラス-ドモアブルの定理)より

$$\frac{1}{2^{10^8}}\#\left\{x\in\{0,1\}^{10^8}\middle|\left|\frac{1}{10^8}\sum_{i=1}^{10^8}x_i-\frac{1}{2}\right|\le\frac{1}{2000}\right\}\approx\int_{-9.9999}^{9.9999}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\mathrm{d}x$$

$$=1-1.53\times10^{-23}$$

注：$\{0,1\}^{10^8}=10^8$ビットの$\{0,1\}$列全体
$\#A$ =集合Aの元の個数



標準正規分布の確率密度関数

確率論の援用

定理1. (i) $K(x) \geq 10^8 - 76$なる乱数 $x \in \{0,1\}^{10^8}$ であって

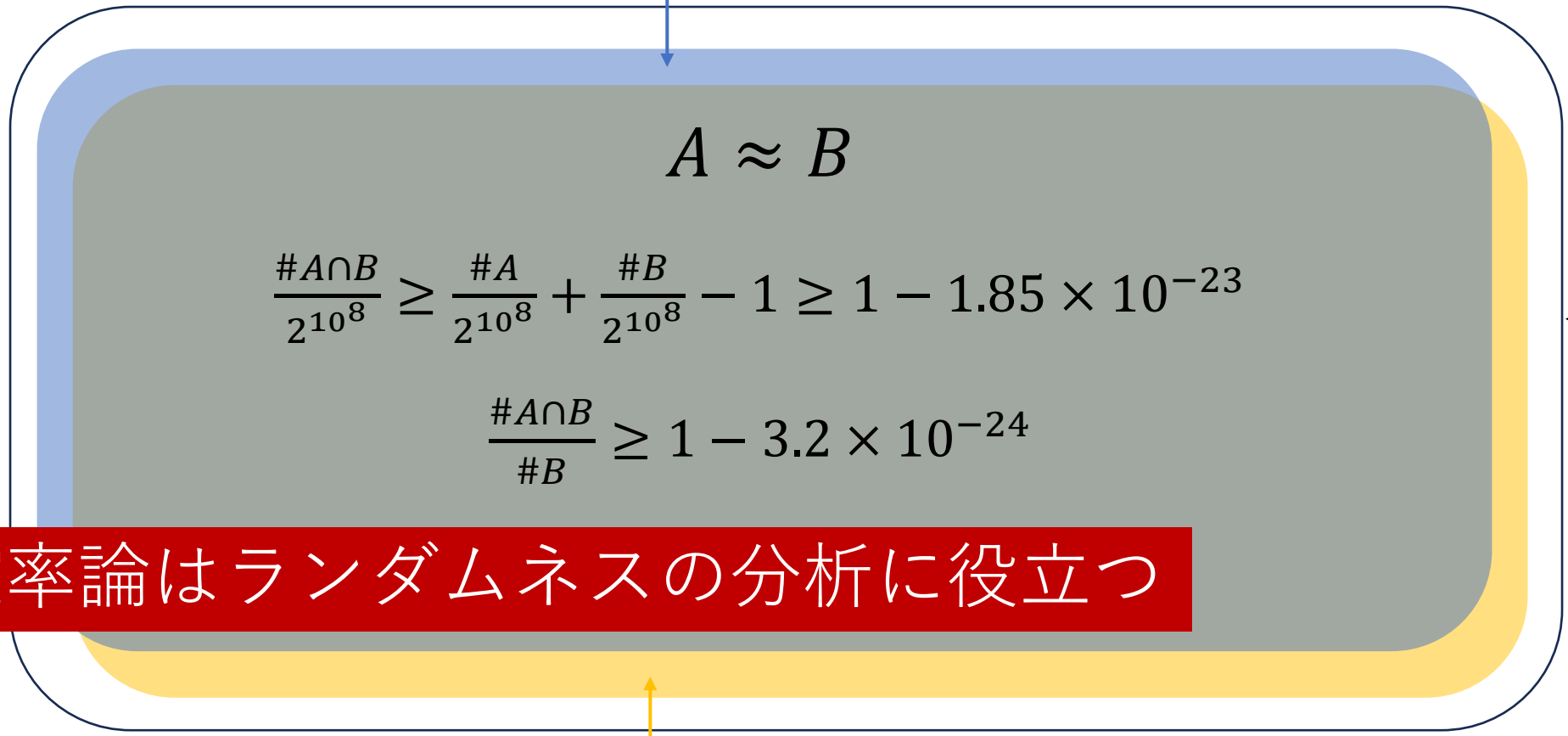
$\left|\frac{1}{10^8}\sum_{i=1}^{10^8} x_i - \frac{1}{2}\right| \leq \frac{1}{2000}$ を満たすものの割合は全体の $1 - 1.85 \times 10^{-23}$ 以上である.

(ii) $\left|\frac{1}{10^8}\sum_{i=1}^{10^8} x_i - \frac{1}{2}\right| \leq \frac{1}{2000}$ を満たす$x \in \{0,1\}^{10^8}$のうち

$K(x) \geq 10^8 - 76$なる乱数 x の割合は $1 - 3.2 \times 10^{-24}$以上である.

定理1の証明.

乱数：$A = \left\{ x \in \{0,1\}^{10^8} \middle| K(x) \geq 10^8 - 76 \right\}$ $\frac{\#A}{2^{10^8}} > 1 - 2^{-76} = 1 - 1.32 \times 10^{-23}$



$$A \approx B$$

$$\frac{\#A \cap B}{2^{10^8}} \geq \frac{\#A}{2^{10^8}} + \frac{\#B}{2^{10^8}} - 1 \geq 1 - 1.85 \times 10^{-23}$$

$$\frac{\#A \cap B}{\#B} \geq 1 - 3.2 \times 10^{-24}$$

確率論はランダムネスの分析に役立つ

$\{0,1\}^{10^8}$

中心極限定理： $B = \left\{ x \in \{0,1\}^{10^8} \middle| \left| \frac{1}{10^8} \sum_{i=1}^{10^8} x_i - \frac{1}{2} \right| \leq \frac{1}{2000} \right\}$ $\frac{\#B}{2^{10^8}} \approx 1 - 1.53 \times 10^{-23}$

ご清聴ありがとうございました.

http://www4.math.sci.Osaka-u.ac.jp/~sugita/mcm.html