



# 乱数とヒストグラム

## モデリングとシミュレーション

1

2017年度

# 乱数(Random Numbers)

- ▶ でたらめに生成される数の列
- ▶ 例：0から $N - 1$ までの整数乱数
  - ▶ 出る順番がでたらめ
    - ▶ 次が予測できない
  - ▶ 出る頻度
    - ▶ 一様乱数：ほぼ均等
    - ▶ 指定された分布に従う場合

# 乱数とシミュレーション 確率過程

- ▶ ある確率 $p$ で事象 $E$ が起こる
- ▶ 動作が決定的でない場合
  - ▶ 対象の外からの予測不能な影響
  - ▶ 内部に不確かな要素
    - ▶ 構成要素の予測不能な挙動
  - ▶ 本質的に確率的現象
    - ▶ 量子力学

# 確率の意味

## ある確率 $p$ で事象 $E$ が起こる

- ➡ 試行を十分大きな回数 $N$ 回行う
- ➡ 事象 $E$ が起こった回数 $N_E$
- ➡  $\frac{N_E}{E} \sim p$
- ➡ 例：さいころの1の目が出る確率 $1/6$ 
  - ➡ 十分大きな回数 $N$ 回、さいころを投げる
  - ➡ 1の目がでる回数は、 $N/6$ に近づく

# 乱数とシミュレーション

## Monte Carlo 法

- ▶ 乱数を積極的に活用して、近似を行う
- ▶ 積分値の近似
  - ▶ 代表値を使って近似
- ▶ 状態のサンプリング
  - ▶ 全状態の数が大きすぎる場合
  - ▶ 代表的状態だけを使って近似

# 疑似乱数 (Pseudo Random Numbers)

## ■ 本物の乱数

- さいころ
- 熱雑音
- 放射性原子の崩壊

## ■ 疑似乱数

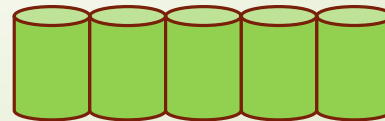
- コンピュータで生成
- アルゴリズムがあるので「偽物」

# 乱数の分布を調べる ヒストグラム (histogram)

- ▶ ヒストグラム：頻度分布
  - ▶ ある事象が何回起こったか、その頻度あるいは相対頻度
- ▶ 0から $N - 1$ までの乱数
  - ▶  $N = 2^{31} \cong 2 \times 10^9$ の場合
  - ▶ それぞれの整数は何回発生する？
  - ▶ そこから分布は分かる？

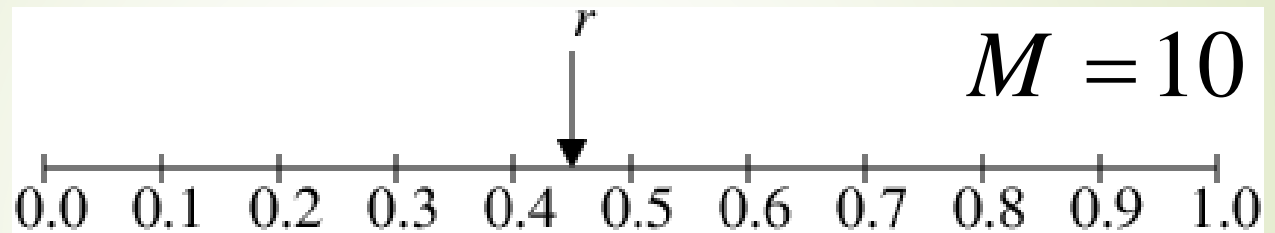
# binを使う

- ▶  $[0,1)$ の範囲の乱数
  - ▶ 各値の発生回数を数えるのは無意味
  - ▶ 区間 $[0,1)$ を $M$ 個に区切る。
    - ▶ 各区間をbinと呼ぶ
  - ▶ 各binに何個入ったかの頻度または相対頻度を求める
  - ▶ binに入る標本数が大きくなると無意味



bin





- ➡  $r$  はどの区間に入っている？
- ➡ 区間幅を  $w$  とすると

$$k = \lfloor r / w \rfloor$$

$$w = 1 / M$$

## ■ 課題：一般化

- 区間 $[a, b)$ を $M$ 個に分割した場合
  - ある数 $r$ が入る区間を判定する方法は？
- その方法で、 $[-0.1, 0.9)$ を5個に等分した区間に対して、 $\{-0.02, 0.04, 0.22, 0.57\}$ はそれぞれどの区間に入ると判定されるか？

# 理想的一様乱数

- ➡  $[0,1)$ の区間に一様乱数が発生したとする
- ➡ 乱数の総数は $N$ 、区間数は $M$ とする
- ➡ ある区間に
  - ➡ 一つの乱数が入る確率  $p = 1/M$
  - ➡  $k$ 個の乱数が入る確率

$$P(k) = \binom{N}{k} p^k (1-p)^{N-k}$$

二項係数がでてくる理由は？

# 確率母関数

## Probability generating function

$$G(z) = \sum_{k=0}^N P(k) z^k$$

和の範囲に注意

$$G(1) = \sum_{k=0}^N P(k) = 1$$

$$G'(z) = \sum_{k=1}^N kP(k) z^{k-1}$$

母関数の具体的な形が分かると、平均や二乗平均が容易に求められる

$$G'(1) = \sum_{k=1}^N kP(k) = \sum_{k=0}^N kP(k) = \langle k \rangle$$

$$G''(z) = \sum_{k=2}^N k(k-1)P(k) z^{k-2}$$

$$G''(1) = \sum_{k=2}^N k(k-1)P(k) = \sum_{k=0}^N k(k-1)P(k) = \langle k^2 \rangle - \langle k \rangle$$

# 二項分布に対する確率母関数

$$\begin{aligned} G(z) &= \sum_{k=0}^N P(k) z^k = \sum_{k=0}^N \binom{N}{k} p^k (1-p)^{N-k} z^k \\ &= \sum_{k=0}^N \binom{N}{k} (zp)^k (1-p)^{N-k} = (zp + 1 - p)^N \end{aligned}$$

具体的関数形が分かる

$$G'(z) = Np(zp + 1 - p)^{N-1}$$

$$G''(z) = N(N-1)p^2(zp + 1 - p)^{N-2}$$

# 二項分布に対する平均と標準偏差

$$G'(1) = Np = \langle k \rangle$$

$$G''(1) = N(N-1)p^2 = \langle k^2 \rangle - \langle k \rangle$$

$$\langle k^2 \rangle = N(N-1)p^2 + Np$$

$$\sigma^2 = \langle k^2 \rangle - \langle k \rangle^2 = N(N-1)p^2 + Np - N^2p^2 = Np(1-p)$$

$$\frac{\sigma}{\langle k \rangle} = \left( \frac{1-p}{p} \frac{1}{N} \right)^{1/2}$$

重要 :  $\sigma / \langle k \rangle \propto N^{-1/2}$

# 相対頻度と確率密度

- ➡ 区間 $[a, b)$ を $N$ 個のbinに分割
- ➡ 各区間の頻度 $F_i$ 、相対頻度 $f_i = F_i/N$
- ➡ 相対頻度は確率分布として理解できる？
  - ➡ 離散的確率と考えると、規格化は

$$\sum_{i=0}^{N-1} f_i = \frac{1}{N} \sum_{i=0}^{N-1} F_i = 1$$

- ➡ 連続的確率密度と考えると

$$\int_a^b p(x) dx = 1$$

- ➡ 積分は面積→小さな台形の和

- ➡ 対応する和は、binの幅  $w = (b - a)/N$  として

$$\sum_{i=0}^{N-1} w \frac{f_i}{w} = 1$$

$$\frac{f_i}{w} \rightarrow p(x_i), \quad x_i = a + \left(i + \frac{1}{2}\right)w$$