

## 第10回：擬似乱数の生成法

確率的な事象を、計算機上でシミュレーションするには、どうしたら良いか？

ある確率分布に従って生起する事象を、計算機上で実現するには、どうしたら良いか？

⇒ 何らかの方法で、「サイコロ振り」をプログラムによって実現する必要がある。

⇒ **乱数** を生成する。

## 乱数

以下の2つを満たす**数の系列**を乱数という。

- (1) **不規則性**: 数字の並び方がでたらめ(不規則)
- (2) **一様性**: 各数字が同じ割合で出現する

乱数表:

乱数サイ: 正二十面体のサイコロ

## 一様乱数

区間  $0 \leq x < 1$  の**一様乱数**

⇔

連続な確率変数  $x$  が、以下の確率密度関数をもつ

$$f(x) = \begin{cases} 1 & (0 \leq x < 1) \\ 0 & (x < 0, x \geq 1) \end{cases}$$

一様乱数の生成法

- ・乱数表から
- ・**計算機の利用**

## 生成法: 合同式法 (congruence method)

区間  $0 \leq x < 1$  の**一様乱数の生成法**

**手続き**:

$a, c, m \in N$  (自然数)。このとき、 $r_0 \in N$  を出発点として、以下の漸化式により、自然数の数列  $\{r_0, r_1, r_2, \dots\}$  を作る。

$$r_{n+1} \equiv a \cdot r_n + c \pmod{m}$$

$0 \leq r_n < m$  より、 $x_n = r_n/m$  は、 $0 \leq x_n < 1$

mod付きの等式を「合同式」と呼ぶ

例えば  $a = 1229$ ,  $c = 351750$ ,  $m = 1664501$  を使うと

$$r_0 = 1, (x_0 = 0.000001504)$$

$$r_1 = (1229 \times 1 + 351750) \% 1664501 = 352979 \% 1664501 = 352979 \\ x_1 = 352979 / 1664501 = 0.212062954$$

$$r_1 = 352979, x_1 = 0.212063$$

$$r_2 = (1229 \times 352979 + 351750) \% 1664501 = 434162941 \% 1664501 = 1392681 \\ x_2 = 1392681 / 1664501 = 0.836695802$$

$$r_2 = 1392681, x_2 = 0.836695802$$

$$r_3 = (1229 \times 1392681 + 351750) \% 1664501 = 1711956699 \% 1664501 = 849671 \\ x_3 = 849671 / 1664501 = 0.510465899$$

$$r_3 = 849671, x_3 = 0.510465899$$



## 計算機で生成できるのは擬似乱数

× 見かけは一見「でたらめ」な数列だが、先の**合同式法**の規則に基づいて(決定論的に)生成されているので、決して「不規則」とは言えない。

○ 一様性は、おおむね成立している(ことが知られている)。

⇒ 「真の」乱数ではなく、**擬似乱数**

しかも、合同式法で生成すると、周期的になる宿命。ただ、その周期が非常に長いことを期待。

→ 周期が長いような、 $a, c, m$  を選ぶ。

## 正規乱数の生成法

**正規乱数**: 連続な確率変数  $x$  が、標準正規分布  $N(0,1)$  に従うような乱数

区間  $0 \leq x < 1$  の (擬似) **一様乱数** から、  
(擬似) 正規乱数を生成する方法:

**生成法1: 中心極限定理**を用いる簡便法

手続き:

**一様乱数を、12個加えて6を引く**

つまり、 $0 \leq x < 1$  の一様乱数として12個の  $x_1, x_2, \dots, x_{12}$  を生成し、  
それから、 $y = x_1 + x_2 + \dots + x_{12} - 6$  を求めて、1つの正規乱数とする。

復習

## 中心極限定理

とはどんな定理か...

● **中心極限定理 (central limit theorem)**

$n$  個の確率変数  $x_1, x_2, \dots, x_n$  が、

互いに**独立に同じ確率分布**に従うとする。

その分布の平均を  $\mu$ 、分散を  $\sigma^2$  とする;

$$E(x_i) = \mu, \quad E((x_i - \mu)^2) = \sigma^2 \quad (i=1, 2, \dots, n)$$

この  $n$  個の変数の平均  $x = \frac{1}{n} \cdot (x_1 + x_2 + \dots + x_n)$

(これも確率変数) は、 $n$  が大きいときには、

**正規分布  $N(\mu, \sigma^2/n)$**  に従う。

元に戻って

$y = x_1 + x_2 + x_3 + \dots + x_{12} - 6$  が  $N(0,1)$  に従うこと

● **中心極限定理**が保障していること: どんな確率分布でも(よって、 $0 \leq x < 1$  の**一様分布**でも)、  
 $n$  個の確率変数の平均は、 $n$  が十分大きいとき、  
**正規分布  $N(\mu, \sigma^2/n)$**  に従う。

● いま、 $0 \leq x < 1$  の一様分布では、  
平均  $\mu = 1/2$ 、分散  $\sigma^2 = 1/12$  (計算!) であり、

$n=12$  で考えると、 $\bar{x} = \frac{1}{12} \sum_{i=1}^{12} x_i$  は  **$N(1/2, 1/12^2)$**  に従う。つまり、

$$y = 12 \cdot \bar{x} - 6 = \frac{\bar{x} - 1/2}{1/12} \text{ は、 } N(0, 1) \text{ に従う。}$$

## 生成法2: ボックス・ミュラー法

(Box-Muller's method)

手続き:

$0 \leq x < 1$  の一様乱数として2つの  $x_1, x_2$  を生成し、  
それから、

$$y_1 = \sqrt{-2 \log x_1} \cdot \cos(2\pi x_2)$$

$$y_2 = \sqrt{-2 \log x_1} \cdot \sin(2\pi x_2)$$

を求めて、2つの正規乱数  $y_1, y_2$  とする。

## $y_1, y_2$ が $N(0,1)$ に従うこと

$$dx_1 dx_2 = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{vmatrix} dy_1 dy_2 = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y_1^2}{2}\right) dy_1 \cdot \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y_2^2}{2}\right) dy_2$$

**ヤコビアン (Jacobian)**

計算は次で...

$$\underbrace{P(x_1) dx_1}_{=1} \cdot \underbrace{P(x_2) dx_2}_{=1} = \underbrace{\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y_1^2}{2}\right) dy_1}_{=f(y_1)} \cdot \underbrace{\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y_2^2}{2}\right) dy_2}_{=f(y_2)}$$

$= N(0,1) \quad \quad \quad = N(0,1)$

$$x_1, x_2 : 0 \rightarrow 1 \quad \Rightarrow \quad y_1, y_2 : -\infty \rightarrow +\infty$$

## ヤコビアン

$$\begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{vmatrix}$$

の計算 (1/2ページ目)

$$\begin{cases} y_1 = \sqrt{-2 \log x_1} \cdot \cos(2\pi x_2) \\ y_2 = \sqrt{-2 \log x_1} \cdot \sin(2\pi x_2) \end{cases}$$

$$\therefore y_1^2 + y_2^2 = -2 \log x_1, \quad \frac{y_1}{y_2} = \tan(2\pi x_2)$$

$$\therefore x_1 = \exp\left(-\frac{y_1^2 + y_2^2}{2}\right), \quad x_2 = \frac{1}{2\pi} \tan^{-1}\left(\frac{y_1}{y_2}\right)$$

$$\therefore \frac{\partial x_1}{\partial y_1} = -y_1 x_1, \quad \frac{\partial x_1}{\partial y_2} = -y_2 x_1$$

$$\frac{\partial x_2}{\partial y_1} = \frac{1}{2\pi} \cdot \frac{1}{1 + \left(\frac{y_1}{y_2}\right)^2} \cdot \frac{1}{y_2}, \quad \frac{\partial x_2}{\partial y_2} = \frac{1}{2\pi} \cdot \frac{1}{1 + \left(\frac{y_1}{y_2}\right)^2} \cdot \left(-\frac{y_1}{y_2^2}\right) \frac{d}{dx} (\tan^{-1} x) = \frac{1}{1+x^2}$$

## ヤコビアン

$$\begin{matrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{matrix}$$

## の計算 (2/2ページ目)

$$\begin{aligned} \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{vmatrix} &= \frac{1}{2\pi} \cdot \frac{1}{1 + \left(\frac{y_1}{y_2}\right)^2} \cdot x_1 \cdot \begin{vmatrix} -y_1 & -y_2 \\ y_2 & -y_2^2 \end{vmatrix} \\ &= \frac{1}{2\pi} \cdot \frac{1}{1 + \left(\frac{y_1}{y_2}\right)^2} \cdot x_1 \cdot \left(1 + \frac{y_1^2}{y_2^2}\right) = \frac{x_1}{2\pi} = \frac{1}{2\pi} \cdot \exp\left(-\frac{y_1^2 + y_2^2}{2}\right) \end{aligned}$$

$$\underbrace{\int_{-\infty}^{\infty} P(x_1) dx_1}_{=1} \cdot \underbrace{\int_{-\infty}^{\infty} P(x_2) dx_2}_{=1} = \underbrace{\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y_1^2}{2}\right)}_{=f(y_1)} dy_1 \cdot \underbrace{\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y_2^2}{2}\right)}_{=f(y_2)} dy_2 = \underbrace{N(0,1)}_{=N(0,1)}$$

## レポート課題(配布)

一様乱数から、2つの方法で正規乱数を生成するプログラムを作成し、実際に10万個の乱数を発生させて、その性質を調べなさい。

- 課題1: 「中心極限定理を用いる簡便法」
- 課題2: 「ボックス・ミュラー法」
- 課題3: 課題1, 2の結果から、2つの方法で生成された正規乱数の質の違いを考察しなさい。

テキスト138ページの図8. 7は、実行結果の1例

## 課題を実行する上での注意

- rand(), srand(), srandom()など一様乱数の発生に既存の関数を用いないこと: 一様乱数作成も課題。
- 10万次元の配列などを絶対に用いないこと: 不要でありメモリの無駄。大規模データを扱う際の基本。
- ヒストグラム(頻度分布グラフ)への頻度値の格納の際に、100回ループしたif文やelse if文を使わないこと: 計算量の無駄
- 同じ10万個のデータに対して、平均、分散、ヒストグラムを作成する: 当然ですが、個別に10万個ずつ発生させて計算したレポートがあり、それは誤りです。

以上の注意事項に反するレポートは、大きな減点です。

## プログラムを実行する上での注意

log()やsqrt(), sin(), cos()といった数学の関数を用いる場合は、それを含んでいるヘッダファイル math.h をincludeする必要があります。

つまり、プログラムの冒頭で

```
#include <math.h>
```

と書き、コンパイル時には

```
% gcc -lm program.c
```

という(マイナス エル エム)オプションが必要です。

## ポアソン乱数の生成法

ポアソン乱数: 離散的な確率変数  $n$  が、ポアソン分布  $P(\lambda)$  に従うような乱数

$$p(n) = \frac{\lambda^n}{n!} \cdot e^{-\lambda}, \quad n = 0, 1, 2, \dots$$

区間  $0 \leq x < 1$  の(擬似)一様乱数から、(擬似)ポアソン乱数を生成する方法:

一様乱数法

手続き:

$0 \leq x < 1$  の一様乱数の系列  $x_1, x_2, x_3, \dots$  から、

⇒続く

## ポアソン乱数の生成法(続き)

$$\begin{aligned} y_0 &= e^{\lambda} \cdot x_0, \quad y_1 = e^{\lambda} \cdot x_0 x_1, \quad y_2 = e^{\lambda} \cdot x_0 x_1 x_2, \\ y_3 &= e^{\lambda} \cdot x_0 x_1 x_2 x_3, \dots \end{aligned}$$

によって

- 系列  $x_0, x_1, x_2, \dots$  から系列  $y_0, y_1, y_2, \dots$  を作る。
- $x_i < 1$  より、 $y_0 > y_1 > y_2 > y_3 > \dots$  である。
- そこで、初めて  $y_n \leq 1$  となる  $n$  を求める。
- この  $n$  を、ポアソン乱数として得る。

(このようにして得た  $n$  がポアソン分布に従うことの証明は略)