

7. 乱数とヒストグラム

2019/11/18

1 準備

講義で説明した一様乱数のヒストグラムを作成する準備として以下の作業をします。NetBeans を開き、サンプルプログラムを取得します。また、必要に応じて、ライブラリとして、微分方程式の際に使用した MyLib を登録します。

<https://github.com/modeling-and-simulation-saga/Random>

2 乱数

一様乱数とは、ある範囲のなかで、頻度が一律な、でたらめな値の列です。Java の場合、`Math.random()` によって、値が区間 $[0, 1)$ に入る `double` 型の乱数を生成することができます。

今回は、生成する乱数の区間を変更するとともに、乱数発生分布について調べます。初めに `method` パッケージの中身を見ていきます。

様々な分布の乱数を発生させるための、インターフェース `RandomGenerator` を定義します。このインターフェースでは、`getNext()` という抽象メソッドを定義しています。

`RandomGenerator` を実装したクラス `Uniform` は、区間 $[a, b)$ の一様乱数を生成するクラスです。コンストラクタでは、区間を指定します。メソッド `getNext()` を上書きし、`Math.random()` を用いて、区間 $[a, b)$ の一様乱数を生成します。

課題 1 クラス `Uniform` では、区間 $[0, 1)$ の乱数を生成する `Math.random()` を用いて、区間 $[a, b)$ の一様乱数を生成する。その仕組みを理解し、簡単に説明しなさい。

3 ヒストグラム

3.1 プログラムの完成

ヒストグラム (histogram) とは、一般には、事象毎の頻度を表すものです。ここでは、`double` 型の乱数の出現頻度を調べます。

ある現象が区間 $[a, b)$ の間の実数値として出現する場合には、各値の出現頻度そのものを調べることは意味がありません。そこで、区間を小区間に分けて、各小区間に値が出現する頻度を調べることが有効です。その際に、各区間に十分な数のサンプルが入るように工夫することも重要です。小区間を `bin` と呼びます。

ヒストグラムのクラス `histogram.Histogram` では、コンストラクタに区間 $[a, b)$ と、それを小区間に等分するための区間数または区間の幅を与え、初期化します。コンストラクタ内では、頻度を数える配列 `hist` も生成します。

課題 2 クラス `histogram.Histogram` のなかのメソッド `put()` は、与えられた値 x が入る bin の番号を k として、その bin の頻度を一つ増やします。メソッド `put()` の内容を完成させなさい。

3.2 ヒストグラムの図示

クラス `Uniform` の `main()` メソッドでは、区間 $[-1, 1)$ の一様乱数を 100,000 個生成します。その結果を、`Uniform-output.txt` ファイルに出力します。出力ファイルには、区間の中央値とその区間に入った数値の相対頻度を、全区間の頻度と区間幅の積の和が 1 になるように記録します。

ソースコード 3.1 `histogram.plt`。"#"以降はコメントを表す

```
1 set terminal pdfcairo enhanced color size 29cm,20cm font "Times-New-Roman" fontsize 1.2
2 set xlabel "{/:Italic_x}"
3 set ylabel "{/:Italic_p}"
4 set xrange [-1.1:1.1]
5 set yrange [0:0.6]
6 set ytic 0.2 #y 軸の目盛り
7 set output "uniform.pdf"
8 set title "Uniform_Distribution"
9 set style fill solid border lc rgb "black" #ヒストグラムのスタイル
10 plot "Uniform-output.txt" with boxes notitle, 0.5 notitle
```

課題 3 これをソースコード 3.1 を用いて図示しなさい。`xrange` と `yrange` は、 x 軸及び y 軸の表示する区間を表します。各 bin に均等に乱数が入ることを示すという目的に沿って、その範囲を適切に設定しなさい。また、生成する乱数の範囲を $[-0.3, 0.7)$ と変更して、正しく乱数が生成され、ヒストグラムが図示できることを確認しなさい。

3.3 (発展) 平均と分散

課題 4 余力のある人は、ヒストグラムの高さの平均と標準偏差を数値的に求めるように、クラス `Histogram` を拡張しなさい。つまり、高さの平均を求めるメソッド `meanFrequency()` と標準偏差を求めるメソッド `frequencyStandardDeviation()` を作成しなさい。

課題 5 また、講義で説明したように、一様乱数に対しては、ヒストグラムの高さの平均 $\langle h \rangle$ と標準偏差 σ は理想の場合を厳密に計算することができる。乱数の総数を変えながら、数値的計算と比較しなさい。具体的には、乱数の総数 N を 1,000 から 2 倍ずつ、512,000 個まで変え、その時の $\sigma/\langle h \rangle$ の値と厳密な値を、横軸に乱数の総数 N 、縦軸に $\sigma/\langle h \rangle$ を表示しなさい。