

7. 乱数とヒストグラム

2016/11/21

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。

<http://aoba.cc.saga-u.ac.jp/lecture/ModelingAndSimulation/javasrc/Random/src.zip>

をダウンロードします。この中には

```
src\histogram\Histogram.java
src\samples\Uniform.java
src\samples\Transform.java
src\samples\Rejection.java
```

の四つがあります。プロジェクト Random の下に置きましょう。また、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。

2 乱数

一様乱数とはある範囲のなかで、一様な頻度ででたらめな値を生成するものです。Java の場合、`Math.random()` によって、区間 $[0, 1)$ の乱数を生成することができます。

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()` では、`Math.random()` を用いて、区間 $[a, b)$ の一様乱数を生成しています。

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。

課題2 クラス Uniform の main() メソッドでは、区間 $[-1, 1)$ の一様乱数を 100,000 個生成するようになっている。該当する場所を見つけなさい。

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。あるいは、その現象が非常に広い区間 $[a, b)$ 、例えば $[0, 100,000)$ の間の整数値として出現する場合でも、各値の出現頻度を調べることは、そのサンプル数が 100,000 に比べて桁違いに多い場合でなければ、意味がありません。

そこで、区間を小区間に分けて、各小区間内の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が重要です。小区間を bin と呼びます。

ヒストグラムのクラス Histogram 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整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 を生成しています。

課題3 クラス Histogram のなかで、与えられた値 x がはいる bin の番号を k として、その bin の頻度を一つ増やすメソッド 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 個生成するようになっています。その結果は、Uniform-output.txt ファイルに出力されます。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が、全区間の頻度と区間幅の積の和が 1 になるように記述されます。

課題4 これを Program 3.1 を用いて図示しなさい。xrange と yrange は適当に設定すること。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。

```
set terminal png enhanced
set xlabel "x"
set ylabel "p"
set xrange [-2:2]
set yrange [0:1.5]
set ytic 0.2
set output "uniform.png"
set title "Uniform Distribution"
set style fill solid
plot "Uniform-output.txt" with boxes notitle
```

Program 3.1: histogram.plt

3.3 (発展) 平均と分散

課題5 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。また、講義で説明したように、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、乱数の総数を1000から2倍ずつ、512000個まで変え、その時の $\langle h \rangle / \sigma$ の値と厳密な値を比較しなさい。