

レポート提出票

科目名: 情報工学実験1

実験課題名: 課題4 統計学入門

実施日: 2020年 8月 3日

学籍番号: 4619055

氏名: 辰川力駆

共同実験者:

1 実験の目的

(1) 離散分布の理解

代表的な離散分布である二項分布を理解する。

(2) 標本平均の分布

中心極限定理を理解する。

(3) 連続分布の理解

代表的な連続分布である正規分布を理解する。

2 実験 1

二項分布を基にした実験

2.1 目標

離散分布の基本として二項分布の特徴を理解する。

2.2 実験手順

(1) “赤 4 個、白 6 個の合計 10 個の玉が入った箱から 1 つ取り出して色を確認後、取り出した玉を箱に戻す作業を 5 回繰り返す。” というのを 1 回の試行として、これを 30 回試行して記録する。

(2) 取り出された赤玉の回数の度数表と棒グラフを Excel で作成する。

(3) 取り出された赤玉の個数の平均値を求める。

2.3 実験結果

以下に実験結果の表とグラフを示す。また、取り出された赤玉の個数の平均値は約 2.433333333 となった。

表 1: 試行における赤玉の個数

試行	回数					赤玉の個数
	1	2	3	4	5	
1	○	○	○	×	○	4
2	×	○	×	○	×	2
3	○	×	○	○	×	3
4	×	○	×	×	○	2
5	×	×	○	○	×	2
6	○	×	○	×	×	2
7	○	×	○	×	○	3
8	○	×	○	○	○	4
9	×	×	×	○	×	1
10	×	×	○	×	×	1
11	×	○	×	×	○	2
12	×	○	○	○	×	3
13	○	○	○	○	×	4
14	○	○	×	○	×	3
15	○	○	○	×	×	3
16	○	×	○	○	○	4
17	×	×	×	×	×	0
18	×	×	×	○	○	2
19	○	×	×	○	○	3
20	○	×	○	○	×	3
21	×	○	○	×	○	3
22	○	×	×	×	×	1
23	×	×	×	×	○	1
24	×	○	○	○	○	4
25	×	×	○	○	○	3
26	○	×	×	×	○	2
27	○	×	○	×	×	2
28	○	○	×	○	○	4
29	○	×	×	×	×	1
30	×	×	○	×	×	1
平均						2.43333333

表 2: 出現回数の度数表

赤玉の個数	度数
0	1
1	6
2	8
3	9
4	6
5	0

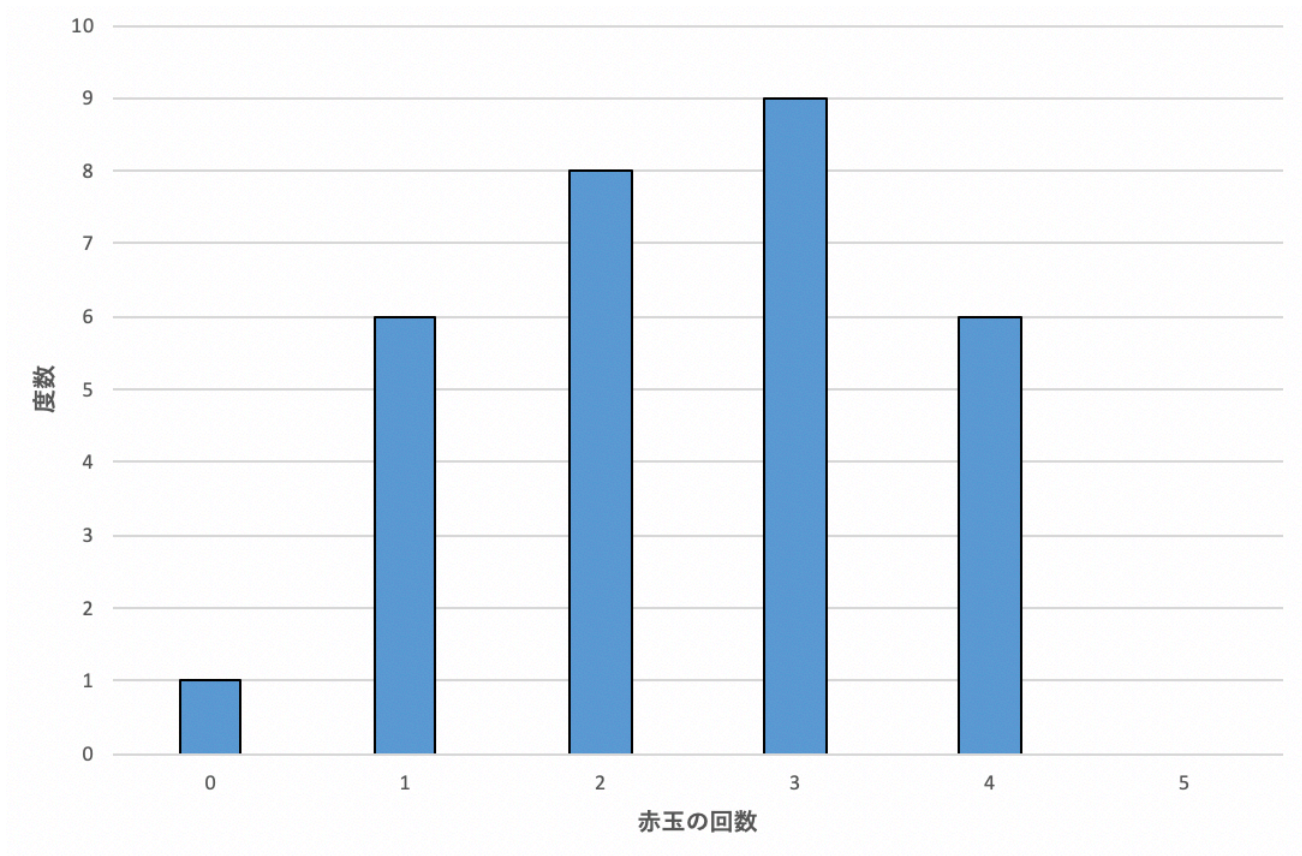


図 1: 赤玉出現回数の棒グラフ

2.4 レポート課題

レポート課題 1-1

実験 1 の結果をまとめよ。

表 1 の実験結果より、平均値は約 2.433333333 である。また、表 2 の度数分布より最頻値は 3 であり、中央値は 2.5 である。ヒストグラムを見ると明らかであるが、0 と 5 の頻度が少ない。そして 1 から 4 はほぼ頻度が同じなので、平均値と中央値がほとんど同じと考えられる。

レポート課題 1-2

実験 1 の結果は、二項分布に従っているといえるか述べて。

実験 1 の結果は、二項分布に従っていない。

二項分布の公式より

$$P(X = k) = {}_n C_k p^k (1 - p)^{n-k}$$

である。今回は、条件より $n = 5$ 、 $p = 0.4$ として $k = 0, 1, 2, 3, 4, 5$ を計算すると、

$$P(X = 0) = {}_5 C_0 0.4^0 (1 - 0.4)^{5-0} \doteq 0.0778$$

$$P(X = 1) = {}_5 C_1 0.4^1 (1 - 0.4)^{5-1} \doteq 0.2592$$

$$P(X = 2) = {}_5 C_2 0.4^2 (1 - 0.4)^{5-2} \doteq 0.3456$$

$$P(X = 3) = {}_5 C_3 0.4^3 (1 - 0.4)^{5-3} \doteq 0.2304$$

$$P(X = 4) = {}_5 C_4 0.4^4 (1 - 0.4)^{5-4} \doteq 0.0768$$

$$P(X = 5) = {}_5 C_5 0.4^5 (1 - 0.4)^{5-5} \doteq 0.0102$$

となるが、実験から出される確率は、

$$P(X = 0) = 1 \div 30 \doteq 0.3333$$

$$P(X = 1) = 6 \div 30 = 0.2$$

$$P(X = 2) = 8 \div 30 \doteq 0.2667$$

$$P(X = 3) = 9 \div 30 = 0.3$$

$$P(X = 4) = 6 \div 30 = 0.2$$

$$P(X = 5) = 0 \div 30 = 0$$

となる。これは、二項分布の公式から出した値と一致しない。

したがって、実験 1 の結果は、二項分布に従っていない。

3 実験2

中心極限定理を基にした実験

3.1 目標

標本平均の分布について理解する。

3.2 実験手順

- (1) Excel の [データ] から [データ分析] を選び、続いて [乱数発生] を選択する。
- (2) 0 から 1 までの 1000 個の乱数を発生させる。
- (3) 発生させた乱数のヒストグラムを作成する。
- (4) 区間幅を 0.01 としてヒストグラムを作成する。
- (5) 0 から 1 までの 1000 個の乱数を 2 つ発生させ、その平均値のヒストグラムを作成する。
- (6) 0 から 1 までの 1000 個の乱数を 8 つ発生させ、その平均値のヒストグラムを作成する。
- (7) 0 から 1 までの 1000 個の乱数を 100 つ発生させ、その平均値のヒストグラムを作成する。

3.3 実験結果

以下に実験結果のグラフを示す。

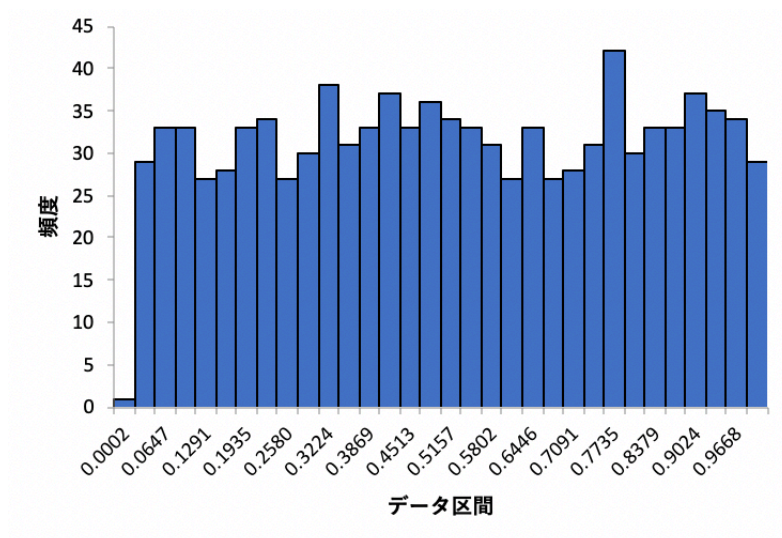


図 2: 1000 個の乱数を発生させた結果のヒストグラム (変数 1 つの区間幅指定なし)

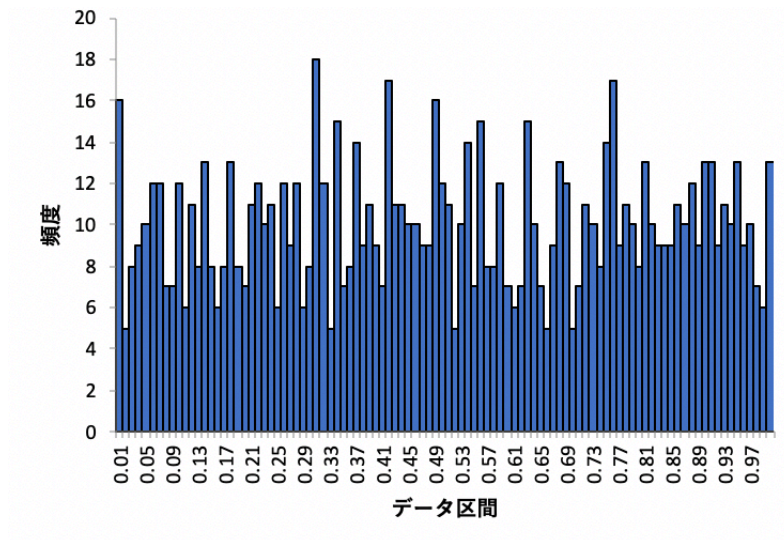


図 3: 1000 個の乱数を発生させた結果のヒストグラム (変数 1 つ)

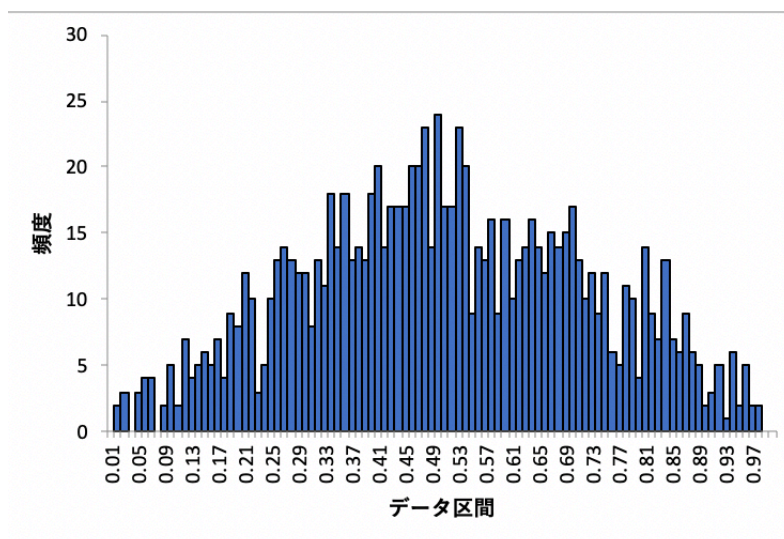


図 4: 1000 個の乱数を発生させた結果の平均値のヒストグラム (変数 2 つ)

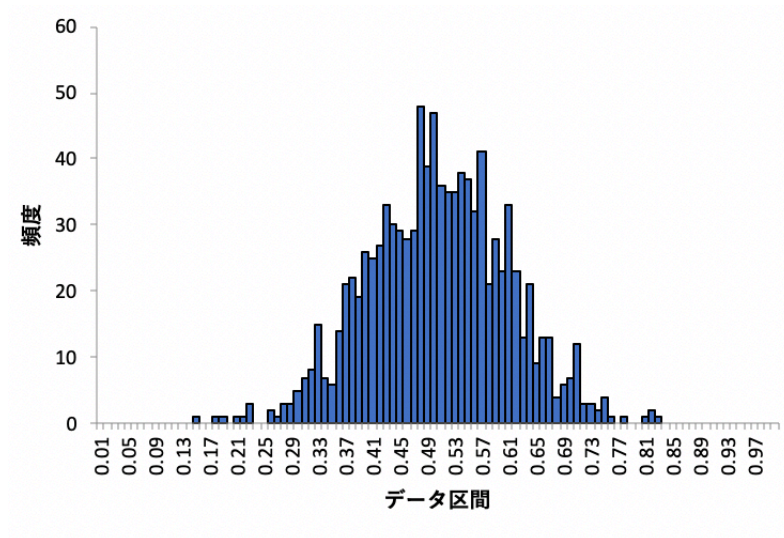


図 5: 1000 個の乱数を発生させた結果の平均値のヒストグラム (変数 8 つ)

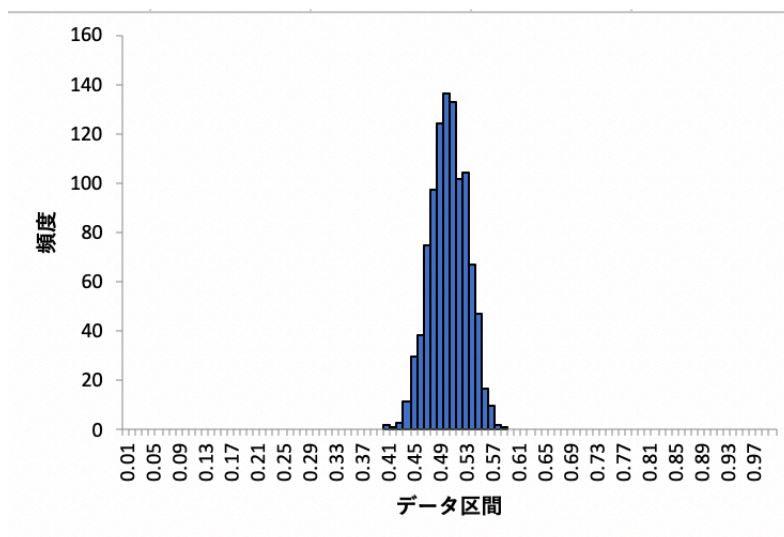


図 6: 1000 個の乱数を発生させた結果の平均値のヒストグラム (変数 100 つ)

3.4 レポート課題

レポート課題 2-1

実験 2 の結果をまとめよ。

それぞれの図より、変数の数を増加させるとデータが中央に寄った。今回の場合は 0 から 1 までの乱数を 1000 個作ったので 0.5 に寄ったことになる。また、変数の数を更に大きくすると、さらに中央に寄ると考えられる。

レポート課題 2-2

大数の法則とは何か、実験 2 の結果に基づいて説明せよ。

大数の法則とは、確率 p で起こる事象において、試行回数を増やすほど、その事象が実際に起こる確率は p に近づくという法則である。

今回は Excel 上で均一分布で乱数を発生させたので、無限個の母集団の平均値は範囲の中央の値になる。従って 0.5 に寄ることになるが、レポート課題 2-1 でも述べたように変数の数を大きくするにつれて 0.5 に寄っている。

レポート課題 2-3

中心極限定理とは何か、実験 2 の結果に基づいて説明せよ。

中心極限定理とは、標本を抽出する母集団が平均 μ 、分散 σ^2 の正規分布に従う従わないに関わらず、抽出するサンプルサイズ n が大きくなるにつれて標本平均の分布は「平均 μ 、分散 $\frac{\sigma^2}{n}$ 」の正規分布 $N(\mu, \frac{\sigma^2}{n})$ に近づくという定理である。実際、今回の実験で変数の数を大きくするにつれてグラフの概形が正規分布に近づいている。

レポート課題 2-4

区間 $[0,1]$ の一様分布から 12 個選び、その合計から 6 を引いた値を Y とする。 Y はどのような分布に従うか述べよ。

区間 $[0,1]$ の一様分布から 12 個選ぶので、分散と平均は元の一様分布に従う。したがって、 Y は正規分布に近づくと考えられる。

4 実験3

正規分布を基にした実験

4.1 目標

正規分布の概要について理解する。

4.2 実験手順

- (1) Excel の [データ] から [データ分析] を選び、続いて [乱数発生] を選択する。
- (2) 0 から 1 までの 1000 個の乱数を平均を 0、標準偏差を 1 として正規分布に従って発生させる。
- (3) [データ区間] に指定する数値 (区間幅 0.5 刻みの -10 から 10) を作る。
- (4) ヒストグラムを作成する。
- (5) 同様の手順で、平均を変更した場合に発生する乱数を用いてヒストグラムを作成する。
 - 平均 -2、標準偏差 1
 - 平均 2、標準偏差 1
- (6) 同様の手順で、標準偏差を変更した場合に発生する乱数を用いてヒストグラムを作成する。
 - 平均 0、標準偏差 2
 - 平均 0、標準偏差 4

4.3 実験結果

以下に実験結果のグラフを示す。

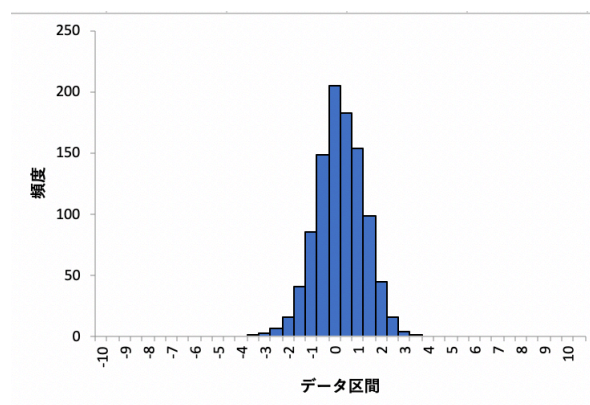


図 7: 平均が 0、標準偏差が 1 の乱数を発生させた結果のヒストグラム

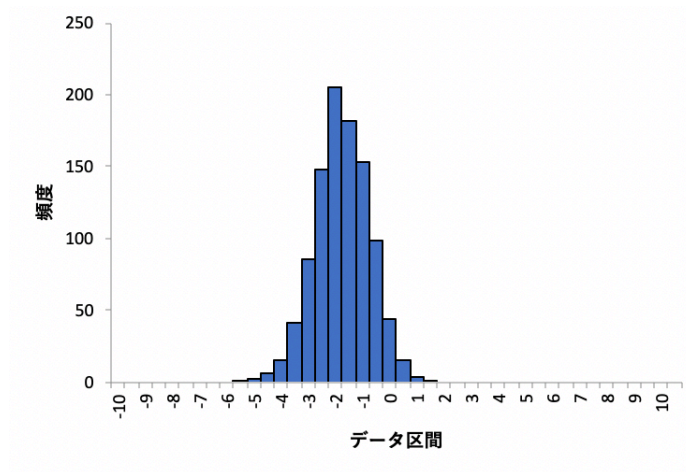


図 8: 平均が-2、標準偏差が1の乱数を発生させた結果のヒストグラム

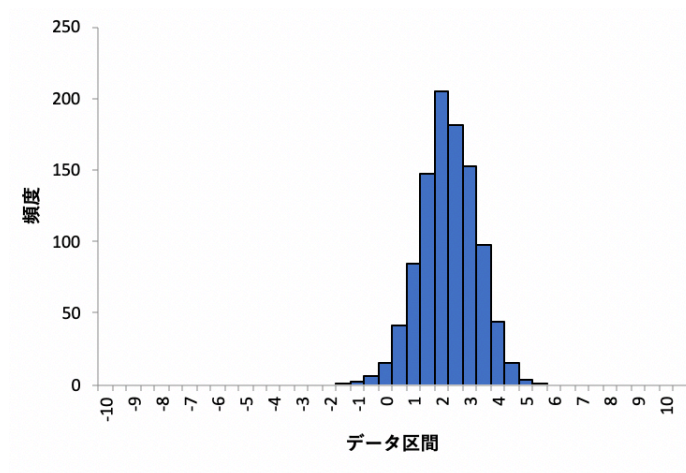


図 9: 平均が2、標準偏差が1の乱数を発生させた結果のヒストグラム

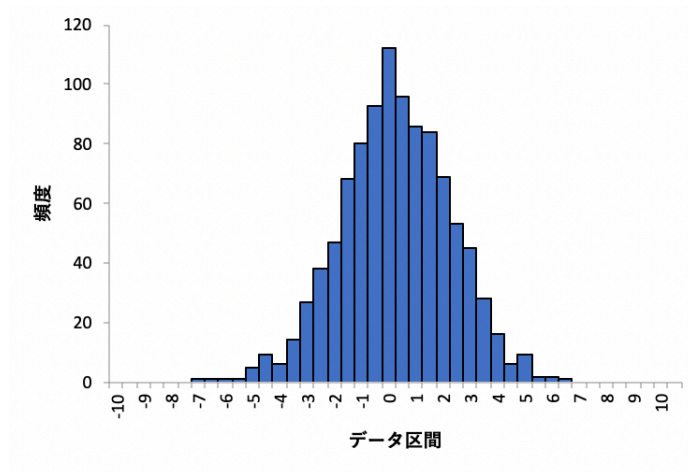


図 10: 平均が 0、標準偏差が 2 の乱数を発生させた結果のヒストグラム

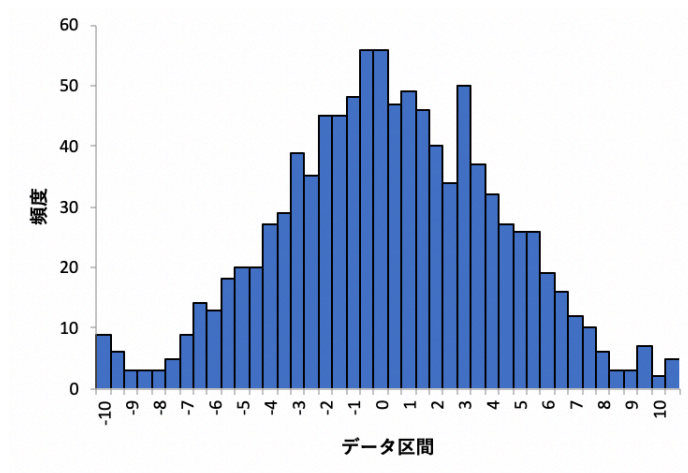


図 11: 平均が 0、標準偏差が 4 の乱数を発生させた結果のヒストグラム

4.4 レポート課題

レポート課題 3-1

実験 3-4、3-5 で作成したヒストグラムを比較して、考察せよ。

図 7、8、9 を比べてみると、グラフの概形は変わらないが山なりなグラフの場所が移動していることがわかる。また、その山なりな概形の中心は平均値に一致している。したがって、図 8 は図 7 のグラフに -2、図 9 は図 7 のグラフに +2 したグラフの概形と考えられる。よって、乱数を生成する時に標準偏差を一定にしたままで平均値を変更すると、その平均値の値だけヒストグラムは移動することが分かる。

レポート課題 3-2

実験 3-4、3-6 で作成したヒストグラムを比較して、考察せよ。

図 7、10、11 を比べてみると、山なりな概形の中心は変わっていないが、標準偏差を大きくすることによりデータのばらつきが大きくなっている。また、縦軸のラベルを見ると分かるが、標準偏差を大きくすると同じ数字があまり存在しなくなっている。よって、乱数を生成する時に平均値を一定にしたままで標準偏差を増やすと、データのばらつきが大きくなることが分かる。

レポート課題 3-3

課題 3-1、3-2 から、正規分布における期待値と標準偏差は確率分布の何を決定しているといえるか述べて。

正規分布において、期待値は確率分布における平均値を決定している。また、正規分布における標準偏差は確率分布のデータのばらつきを決定している。

レポート課題 3-4

偏差値の確率密度関数を書け。

正規分布である確率密度関数 $f(x)$ は、平均を μ 、標準偏差を σ とすると次のようになる。

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \quad (1)$$

ここで、偏差値は平均を $\mu = 50$ 、標準偏差を $\sigma = 10$ としたものであるから、式 (1) に代入して

$$f(x) = \frac{1}{10\sqrt{2\pi}} \exp\left(-\frac{(x-50)^2}{200}\right)$$

となる。

レポート課題 3-5

試験の点数や順位と比較して偏差値の短所・長所を述べよ。

偏差値は、平均値と標準偏差が決められているので、平均値や標準偏差が異なる試験においても比較ができるようになっている。

しかし、偏差値は母集団のどのくらいの位置にいるのかが分かるものなので、母集団が変化した時に偏差値だけでは試験の実力差が分からない。したがって、母集団が違うもの同士で比較はできない。

5 結論

実験を通して、代表的な離散分布である二項分布、標本平均の分布、代表的な連続分布である正規分布を理解した。また、考察で実験値は二項分布に従っているかどうかや、標本平均の分布が中心極限定理に従っているかどうかを考えた。

6 感想

今回の実験で様々なヒストグラムを比べたので、ヒストグラムについての細かい違いを理解することができた。また、確率密度関数の使い方について知ることができた。

参考文献

[1] 東京理科大学工学部情報工学科「情報工学実験1 2020 年度」(2020/4/6)

[2] 大数の法則1 — 統計学の時間 — 統計WEB

<https://bellcurve.jp/statistics/course/8539.html>

最終閲覧日:2020/8/5

[3] 中心極限定理1 — 統計学の時間 — 統計WEB

<https://bellcurve.jp/statistics/course/8543.html>

最終閲覧日:2020/8/5