1. 目的

本稿では、本大学の理工学部計算センターのワークステーションを用いて UNIX 系 OS の操作に慣れ、プログラムの作成、コンパイル、実行を行うことを目的とする。ここでは、C 言語を用いてプログラムの作成法学びながら、モンテカルロ法のシミュレーションを行う。

2. 実験結果

課題(1)

課題(1)では一様乱数を発生させて、その平均と分散、度数分布を求めた。まず、実行したプログラムの実行結果をプログラム 1 に示す。ここで、このプログラムでは、乱数を発生させる回数を$30,000$回としている。

プログラム 1: 課題(1): 実行結果

```
dosu[0] = 6040
dosu[1] = 6068
dosu[2] = 5981
dosu[3] = 5920
dosu[4] = 5991

average = 0.498628
variance = 0.083291
```

これをグラフにプロットし、度数分布を求めた結果を図 1 に示す。プログラム 1、図 1 から、それぞれの区間での度数が、理論値（6000）との誤差が80以下になっており、発生させた乱数が一様乱数になっていることが確認できる。

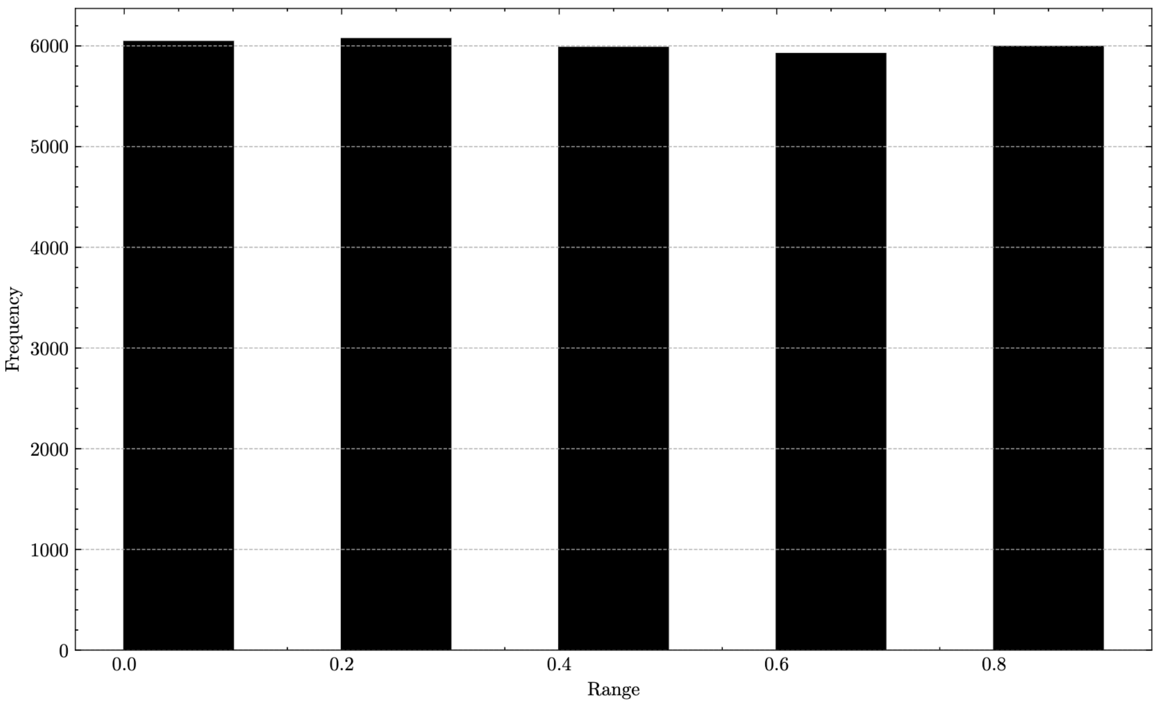

図 1: 課題(1): 度数分布のグラフ

次に、同じプログラム乱数生成法のプログラムを用いて、乱数の発生回数を1,000回から100万回まで、1,000回ずつ増やして、それぞれの平均と分散を求め、その時の平均と分散の値をグラフにプロットした。その結果を図 2 と図 3 に示す。

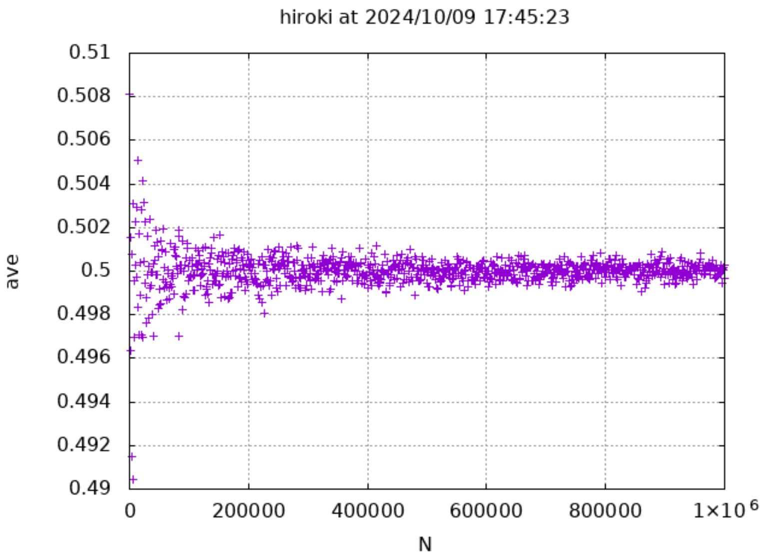

図 2: 課題(1): 平均のグラフ

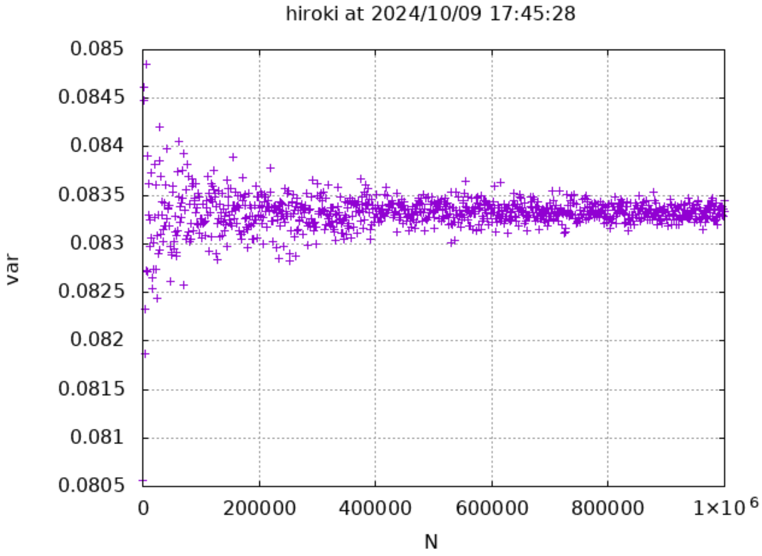

図 3: 課題(1): 分散のグラフ

図 2 から、乱数の発生回数が増えるにつれて、平均の値が0.5に収束していることがわかる。また、図 3 から、乱数の発生回数が増えるにつれて、分散の値が0.083付近に収束していることがわかる。

課題(2)

課題(2)では、buffon の針問題をシミュレーションするプログラムを実行した。このプログラムでは乱数を3万回発生させ、buffon の針問題の理論を用いて、円周率πの値を計算した。その結果をプログラム 2 に示す。

プログラム 2: 課題(2) π の計算結果

```
N= 30000 PAI = 3.125104
```

また、このプログラムを用いて、乱数の発生回数を$1,000$回から100万回まで、$1,000$回ずつ増やして、それぞれのπの値を計算し、その時のπの値をグラフにプロットした。その結果を図 4 に示す。 プログラム 2 から、乱数の発生回数が増えるにつれて、πの値がある一定の値に収束していることがわかる。

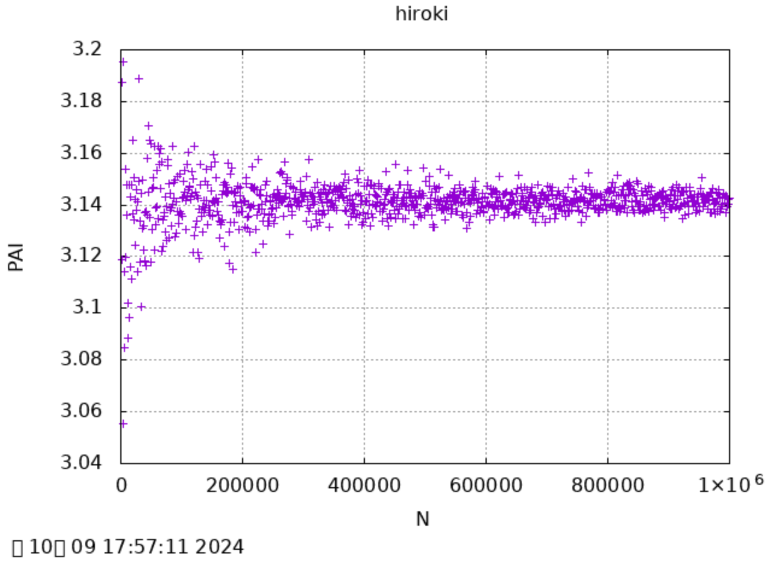

図 4: 課題(2): π への収束の様子

課題(3) 求積問題による π の値

(3)-1: 入門的モンテカルロ法により π の値を求める

入門的モンテカルロ法を用いたプログラムを実行して、π の値を計算した。この時、乱数発生の試行回数を30,000、100,000、1,000,000、10,000,000、100,000,000、300,000,000回で実行した。その結果をプログラム 3 に示す。

プログラム 3: 課題(3)-1: π の値の計算結果

```
N= 30000 PI= 3.144584
N= 100000 PI= 3.141603
N= 1000000 PI= 3.141620
N= 10000000 PI= 3.141603
N= 100000000 PI= 3.141609
N= 300000000 PI= 3.141649
```

(3)-2: 当たり外れのモンテカルロ法により π の値を求める

次に、当たり外れのモンテカルロ法を用いたプログラムを実行して、π の値を計算した。この時も乱数発生の試行回数を30,000、100,000、1,000,000、10,000,000、100,000,000、300,000,000回で実行した。その結果をプログラム 4 に示す。

プログラム 4: 課題(3)-2: π の値の計算結果

```
N= 30000 PI= 3.158400
N= 100000 PI= 3.141520
N= 1000000 PI= 3.141664
N= 10000000 PI= 3.141130
N= 100000000 PI= 3.141698
N= 300000000 PI= 3.141570
```

そして、プログラム 3 とプログラム 4 の結果からそれぞれの絶対誤差を計算し、グラフにプロットすることで、それぞれのモンテカルロ法の精度を比較した。その結果を図 5 に示す。

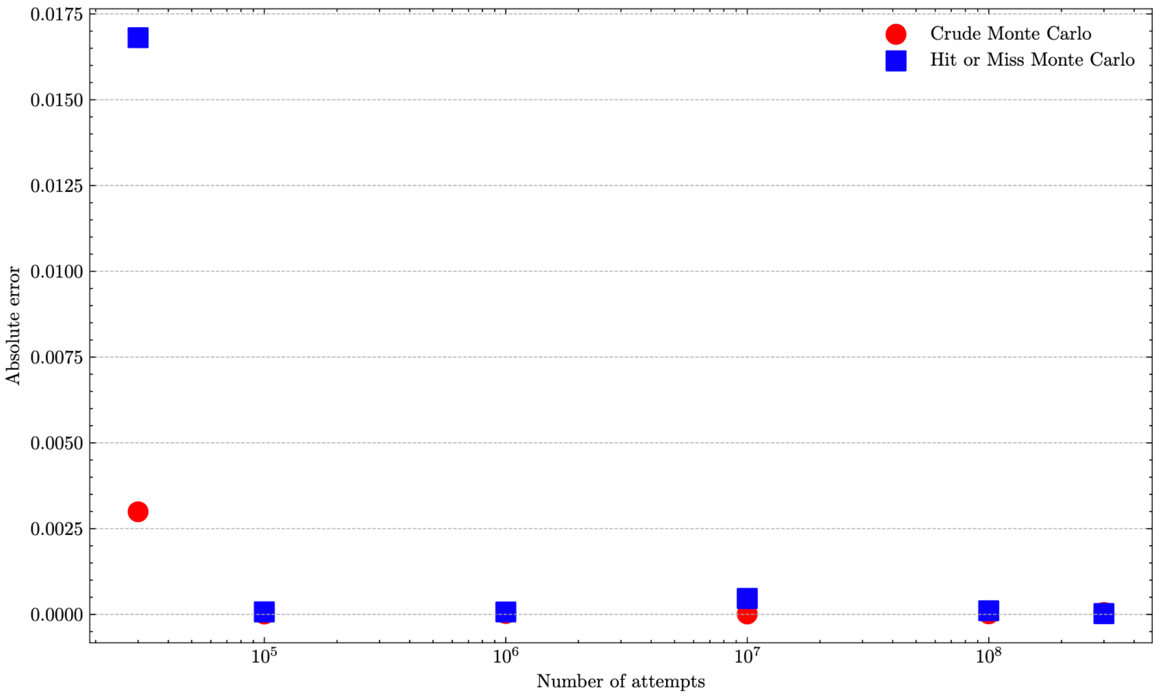

図 5: 課題(3): モンテカルロ法の比較

図 5 から、どちらの手法も初めは回数が増えるにつれて精度が向上しているが、試行回数が10^5を超えてからは、精度の向上がほとんど見られず、誤差が一定になっている。また、10^9の時以外は、入門的モンテカルロ法の方が当たり外れのモンテカルロ法よりも精度が高いことがわかる。

課題(4) モンテカルロシミュレーション

(4)-1: ランダムウォーク

ここでは、ランダムウォークをシミュレーションするプログラムを実行した。まず、今回のランダムウォークでは、1ステップの移動距離を1とし、1回のシミュレーションで10回の移動を行った。初

めに、表計算ソフトを用いて、今回の状況の確率分布の理論式から位置の確率分布を求めた。その結果を図 6 に示す。

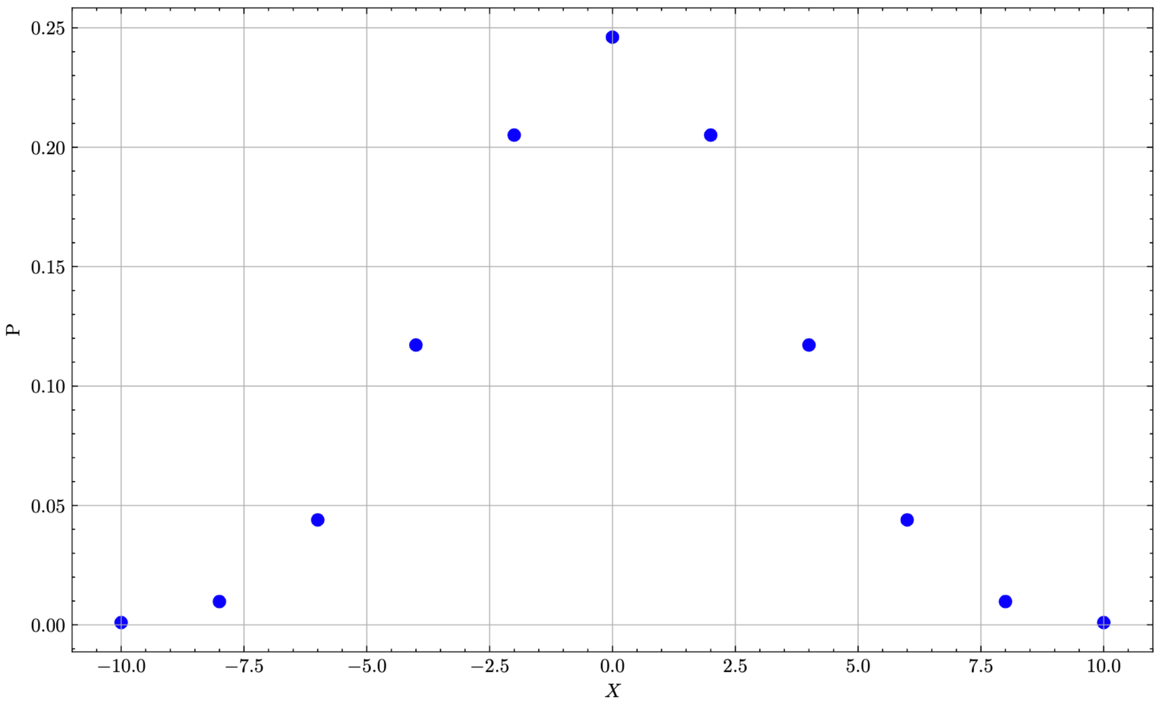

図 6: 課題(4)-1: ランダムウォークの理論値

次に、ランダムウォークのシミュレーションをするプログラムを試行回数100万回で実行し、その時の位置の確率分布をグラフとして可視化した。その結果を図 7 に示す。そして、理論値と実験値の比較を図 8 に示す。図 8 から、理論値と実験値のプロットがほとんど重なっており、プログラムの精度の良さが確認できる。

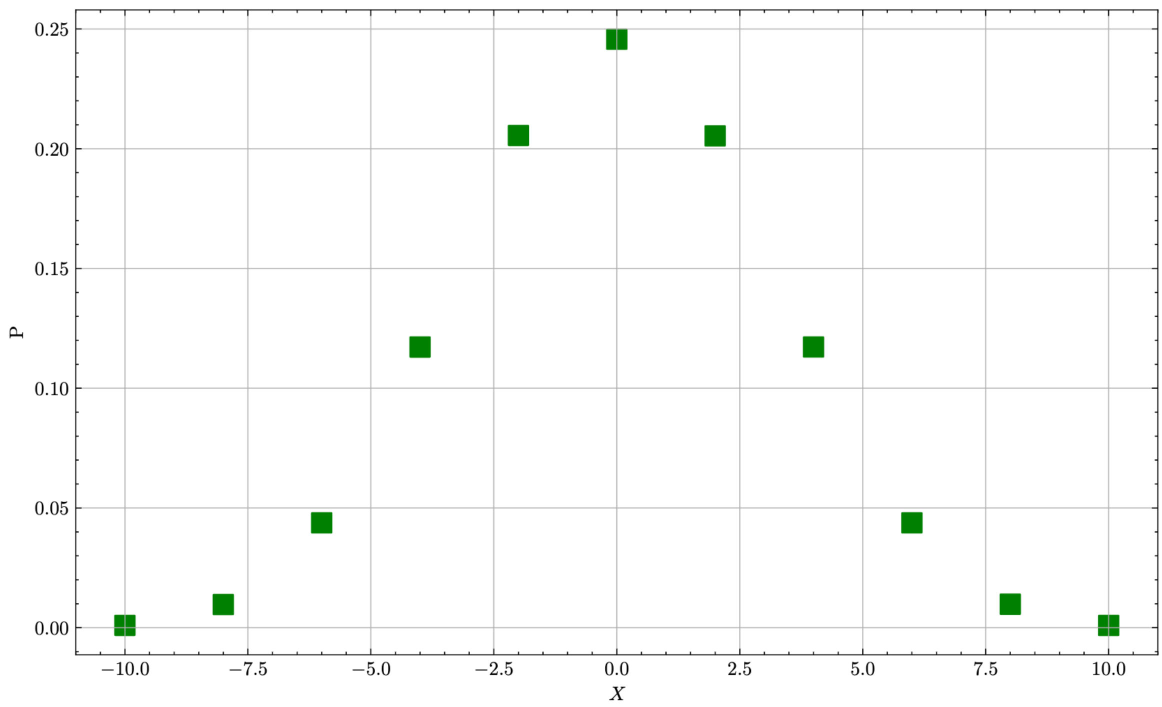

図 7: 課題(4)-1: ランダムウォークの実験値

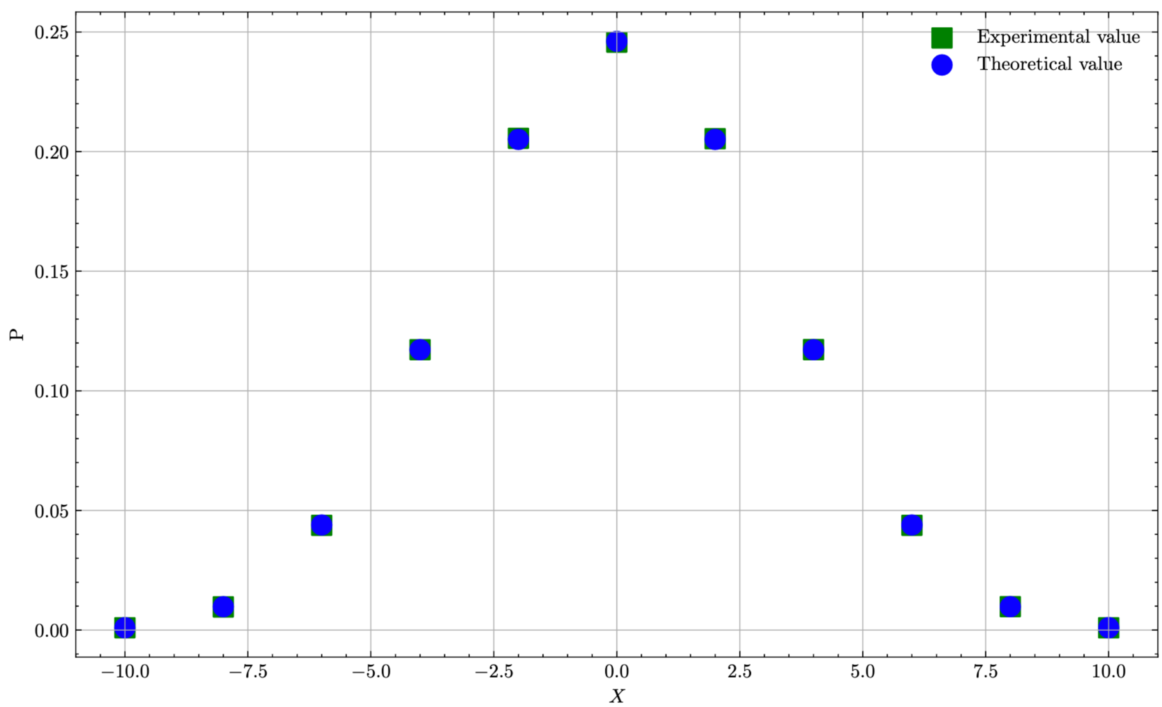

図 8: 課題(4)-1: ランダムウォークの理論値と実験値の比較

(4)-2: 二項分布に従う確率分布

ここでは、「全住民の 5% がある感染症に罹患しており、その中から無作為に 500 人を抽出する。ただし住民は 500 人よりずっと多いとする。このとき、抽出された集団の中に罹患者が 30 人以上いる確率はどれくらいか。」という問題を考え、その確率を求めるプログラムを実行した。まず、(4)-1 と同様に、表計算ソフトを用いて、この問題の確率分布の理論式から位置の確率分布を求めた。ここで、求めた値は患者数が0~29人の時までである。その結果を図 9 に示す。

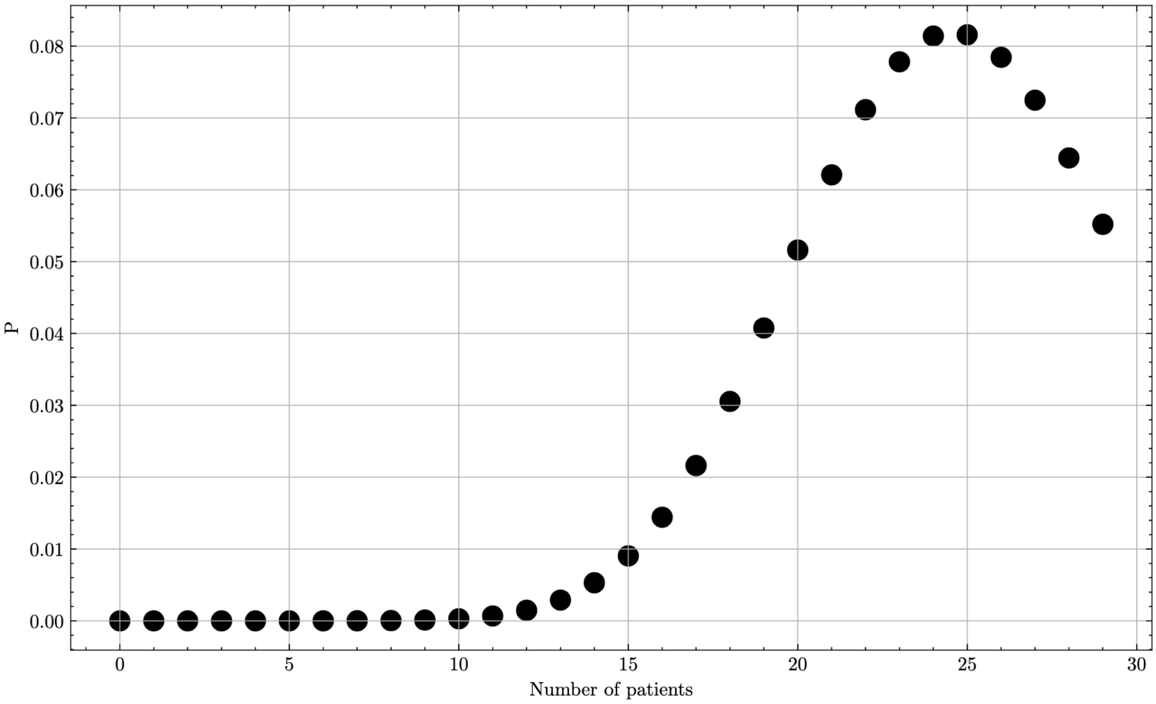

図 9: 課題(4)-2: 二項分布の理論値

次に、試行回数を100、1,000、1万、10万、100万、1,000万、1億回としてプログラムを実行し、その時の確率を計算し、その確率の値をグラフにプロットした。ここでも求めた値は患者数が0~29人の時までである。また、プログラム上で、患者数の期待値を計算させた。まず、期待値の計算結果を表 1 に示す。この表から、試行回数が増えるにつれて、期待値が理論値である25(500×0.05)に収束していることがわかる。 次に、確率の値をグラフにプロットした結果を図 10 に示す。この時、試行回数が1億回の時の確率の値を用いた。そして、理論値と実験値の比較を図 11 に示す。図 11 から、理論値と実験値のプロットが、確率が小さいところ(患者数が15以下)ではほぼ完全に重なっており、確率が大きいところでは、理論値と実験値のプロットが少しのズレがあるが、ほぼ一致していることがわかる。

表 1: 課題(4)-2: 二項分布の期待値の計算結果

試行回数	期待値	期待値の絶対誤差
100	24.400	0.600
1,000	24.880	0.120
1万	24.989	0.011
10万	24.979	0.021
100万	24.993	0.007
1,000万	24.998	0.002
1億	25.000	0.000

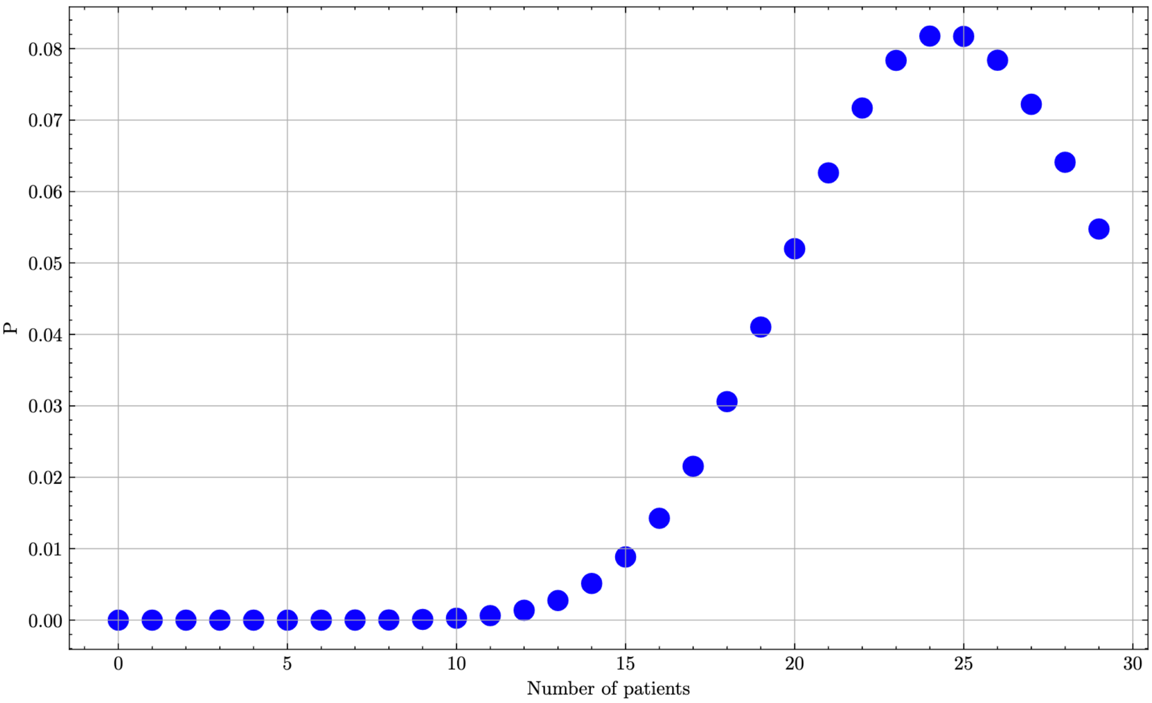

図 10: 課題(4)-2: 二項分布の実験値

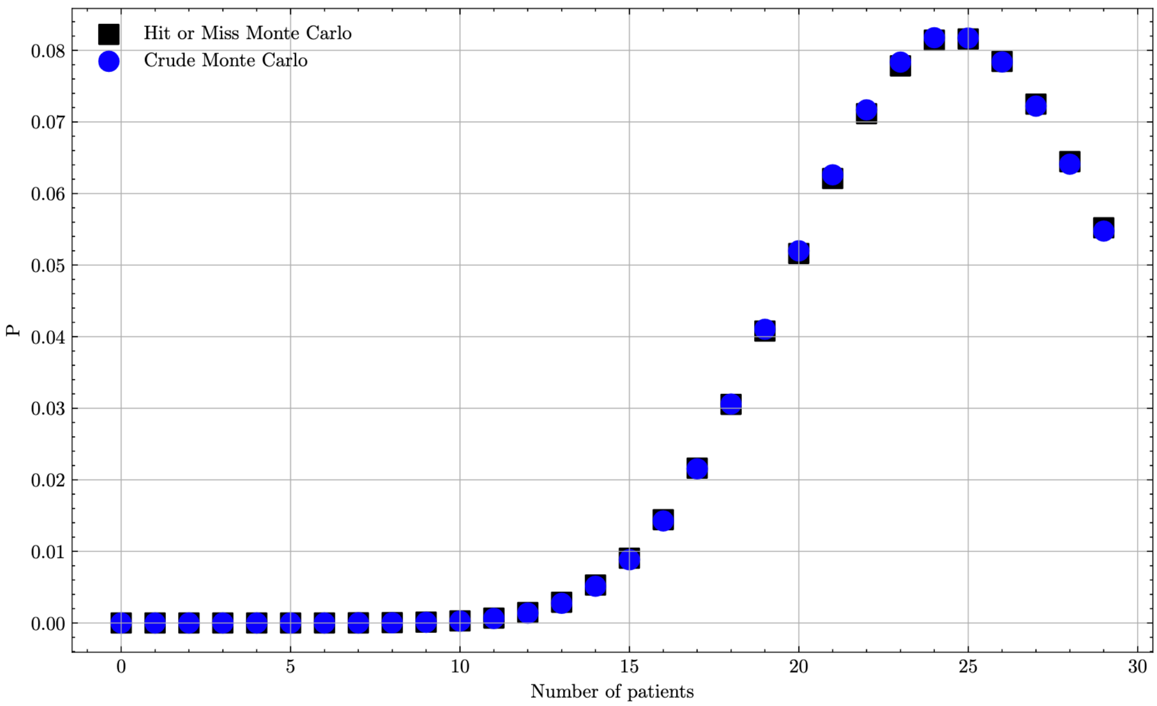

図 11: 課題(4)-2: 二項分布の理論値と実験値の比較

3. 考察

課題(1)

乱数の平均値が 0.5 になる理由、分散の値が 1/12 になる理由

まず、一様乱数の確率密度関数を考える。一般に、一様乱数の確率密度関数は次のように表される。

$$f(x) = \begin{cases} \frac{1}{b-a} \ (a \leq x \leq b) \\ 0 \ (\text{otherwise}) \end{cases} \tag{1}$$

これを用いて、平均値$E(X)$と分散$V(X)$を求めると、次のようになる。

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} xf(x)dx = \int_a^b xf(x)dx \\ &= \int_a^b \frac{x}{b-a}dx \\ &= \frac{1}{b-a}\left[\frac{x^2}{2}\right]_a^b \\ &= \frac{b+a}{2} \end{aligned} \tag{2}$$

$$
\begin{aligned}
E(X^2) &= \int_{-\infty}^{\infty} x^2 f(x)dx = \int_a^b x^2 f(x)dx \\
&= \int_a^b \frac{x^2}{b-a}dx \\
&= \frac{1}{b-a}\left[\frac{x^3}{3}\right]_a^b \\
&= \frac{b^2+ab+a^2}{3}
\end{aligned} \tag{3}
$$

$$
\begin{aligned}
V(X) &= E(X^2) - E(X)^2 \\
&= \frac{b^2+ab+a^2}{3} - \frac{(b+a)^2}{4} \\
&= \frac{b^2-2ab+a^2}{12} = \frac{(b-a)^2}{12}
\end{aligned} \tag{4}
$$

よって、今回の場合、$a=0$、$b=1$ であるため、平均値$E(X)$、分散$V(X)$はそれぞれ0.5、$\frac{1}{12}$になることがわかる。

課題(2)

buffon の針で、何故πの値が求められるのか

まず、buffon の針問題とは、床の上に等間隔に引かれた平行線群があり、その距離を$2a$として，この床に長さ$2l(l<a)$の針をデタラメに落とすと，その針が平行線と交わる確率はいくらかという問題である。 buffon の針で何故πの値が求められるかについて以下の図を基に考える。

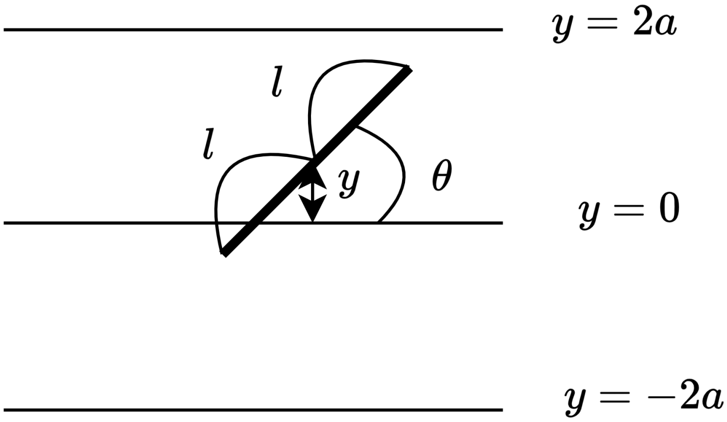

図 12: buffon の針問題

針の中点をy、針と平行線のなす角をθ、針の長さを$2l$とする。対称性より、yの区間を$[0, a]$、θの区間を$\left[0, \frac{\pi}{2}\right]$とする。この時、針が平行線と交差する条件は、

$$y \leq l \sin\theta \tag{5}$$

と表される。ここで針をデタラメに落とすということは、yとθは一様分布に従うと解釈できる。(y, θ)平面を考えると、針が平行線と交差する確率は、$y = l\sin\theta$をθの領域で積分した面積を、全体の面積で割ったものになる。全体の面積は、yの区間が$[0, a]$、θの区間が$\left[0, \frac{\pi}{2}\right]$であるため、$\frac{a\pi}{2}$になる。また、針が平行線と交差する領域は、

$$\int_0^{\frac{\pi}{2}} l \sin\theta d\theta = l \tag{6}$$

となる。よって、針が平行線と交差する確率Pは、

$$P = \frac{l}{\frac{a\pi}{2}} = \frac{2l}{a\pi} \tag{7}$$

と求めることができる。 今回の実験では、yの区間を$[a, 3a]$、θの区間を$[0, \pi]$とし、針の下端y_1、上端y_2を

$$\begin{cases} y_1 = y - l\sin\theta \\ y_2 = y + l\sin\theta \end{cases} \tag{8}$$

として、y_1とy_2の積が負になるとき、針が平行線と交差すると判定している。交わった回数をn、全試行回数をNとすると、(7) より、

$$\pi = \frac{2l}{aP} = \frac{2lN}{an} \tag{9}$$

とπを求めることができる。[参考：[1]]

フローチャート

今回 buffon の針問題を解くためのプログラムのフローチャートを以下に示す。ここで、$a = 1.0$、$l = 0.8$に設定した。そして、(7)、(8) を用いて、πの計算を行った。

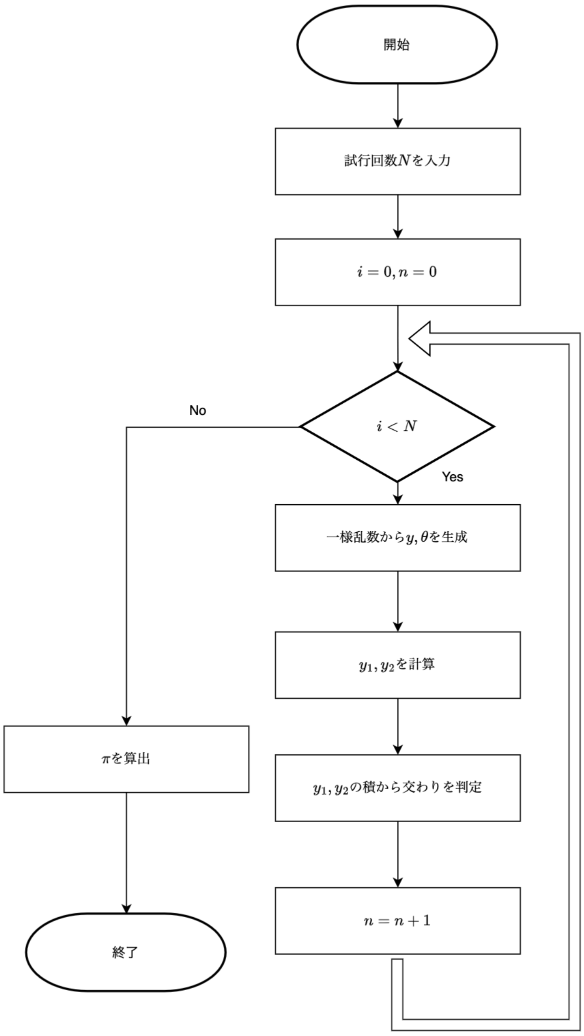

図 13: buffon の針問題のフローチャート

課題(3)

乱数からπを求める方法とその精度

πを求めるのに用いられる方法として、入門的モンテカルロ法と当たり外れのモンテカルロ法がある。これら 2 つの方法について述べる前に、大数の法則について説明する。大数の法則とは、試行回数（一般的には標本すう）を十分大きくすると、確率変数の平均値が基の分布（母集団）の平均値に収束するという法則である。ここで、$[a, b]$で定義された関数$f(x)$と$[a, b]$に一様分布する確率変数rを考える。この時、rの確率密度関数$\rho(r)$は、(1) と同様の式で表される。この時、$f(x)$の平均値$E(f(x))$は、

$$E(f(x)) = \int_a^b f(x)\rho(r)dr = \frac{I}{b-a} \tag{10}$$

となる。ここで、Iは$f(x)$の区間$[a,b]$での積分値である。また、大数の法則より、十分大きいNに対して、N個の一様乱数r_iを発生させたとすると、$f(x)$の平均値は、

$$E(f(r)) \simeq \frac{1}{N}\sum_{i=1}^{N} f(r_i) \tag{11}$$

となる。そして、(10) と(11) より、Iは、

$$I \simeq \frac{b-a}{N}\sum_{i=1}^{N} f(r_i) \tag{12}$$

と近似できる。これが入門的モンテカルロ法による求積である。

次に、$0 \leq f(x) \leq c$として、$0 \leq y \leq c$、$a \leq x \leq b$の領域を考える。この領域に、N個の一様乱数(x_i, y_i)を発生させ、$f(x_i) \geq y_i$の時の数をnとする。この時、領域と$f(x)$より下の領域の面積比は、

$$\frac{1}{c(b-a)}\int_a^b f(x)dx \simeq \frac{n}{N} \tag{13}$$

となる。この式より、Iは、

$$I \simeq c(b-a)\frac{n}{N} \tag{14}$$

となる。このように、$f(x_i) \geq y_i$に当てはまるかどうかで積分値を求める方法を当たり外れのモンテカルロ法という。

ここで、$f(x) = \sqrt{1-x^2}$、$a = 0$、$b = 1$とすることで、

$$\pi = 4I \tag{15}$$

と求めることができる。

次に、これら 2 つの方法の制度について考える。図 5 で見たように、どちらの手法も初めは回数が増えるにつれて精度が向上しているが、試行回数が10^5を超えてからは、精度の向上が0付近で緩やかになっている。大数の法則を考えると、初めは、試行回数が少ないため、平均値が母集団の平均値に収束していく様子がはっきりと見られるが、そこから試行回数が増えていくと、乱数の分布が理想的なものに近づくため誤差が0付近で一定になると考えられ、シミュレーション結果もそれを裏付けている。また、プログラム 2、プログラム 3、プログラム 4 を比べると、30000回の時に、πの値の誤差が、入門的モンテカルロ法、当たり外れのモンテカルロ法の方がbuffonの針よりも小さく、精度が良いことがわかる。よって、小さいN(10^5以下)なら、入門的モンテカルロ法を使い、大きいN(10^5以上)なら、入門的モンテカルロ法、当たり外れのモンテカルロ法のどちらかを使うことで、精度良くπの値を求めることができると考えられる。

課題(4)

ランダムウォークの理論値と実験値の比較

図 8 を見ると、ランダムウォークの理論値と実験値は酷似していることがわかる。さらに、各Xの値における理論値と実験値の相対誤差を計算した結果を表 2 に示す。

表 2: ランダムウォークの理論値と実験値の比較

X	相対誤差 “/%”
−10	3.02
−8	0.362
−6	0.363
−4	0.0192
−2	0.145
0	0.143
2	0.226
4	0.00469
6	0.367
8	1.01
10	2.41

この表から、Xの絶対値が大きくなるにつれ、相対誤差が大きくなるという傾向があることがわかる。これは、大数の法則を考えると、試行回数が多くなるにつれて平均値が理論値に収束していくため、平均値付近の値は理論値に近づくが、平均値から離れるにつれて、誤差が大きくなると考えられる。また、実際に平均値付近のXでの誤差を確認してみると、Xの絶対値が4の場所が誤差が最も小さくなっている。これは、発生させた乱数が、確率$\frac{1}{2}$で右、左に移動する様子を完全に再現できていないためであると考えられる。何故なら、確率が$\frac{1}{2}$からズレると、平均値も同時にずれてしまう。そのため、Xの絶対値が4の場所付近が平均値となり、大数の法則により、誤差が最も小さくなっていると考えられる。

二項分布に従う確率分布の理論値と実験値の比較

図 11 だけを見ると、理論値と実験値は酷似しているように見える。より精度を確認するために、各患者数における理論値と実験値の相対誤差を計算した結果を図 10 に重ねて表示させた。その結果を図 14 に示す。この図から横軸の値が$0, 1, 2$の時は、相対誤差が1になっていることがわかる。しかし、この結果は妥当であると言える。なぜなら、患者数が$0, 1, 2$の時は、理論値が0に近く、オーダーが10^{-9}よりも小さいため、試行回数が10^8の時に実験値が1を超えることがないため、実験値が0として観測され、相対誤差が1になっていると考えられる。これを踏まえて、患者数が$0, 1, 2$の時の相対誤差を除いた結果を図 15 に示す。この図から、期待値に近い患者数の時は、相対誤差が小さく、患者数が多くなるにつれて、相対誤差が大きくなるという傾向があることがわかる。しかし、患者数が3の時は、誤差が小さくなっている様子が見られる。これは、患者数が$0, 1, 2$の時と同様にオーダーの問題で、理論値が10^{-8}なため、試行回数が10^8の時に実験値が 1 の位になるため、ずれによる変動を受けにくいと考えられ、それに対し、患者数が$4, 5$などの場所では、オーダーが10^{-7}なため、桁数が 2 桁になるため、1 の位がずれの影響を受けやすいと考えられる。その結果、患者数が3の時の相対誤差が小さくなっていると考えられる。

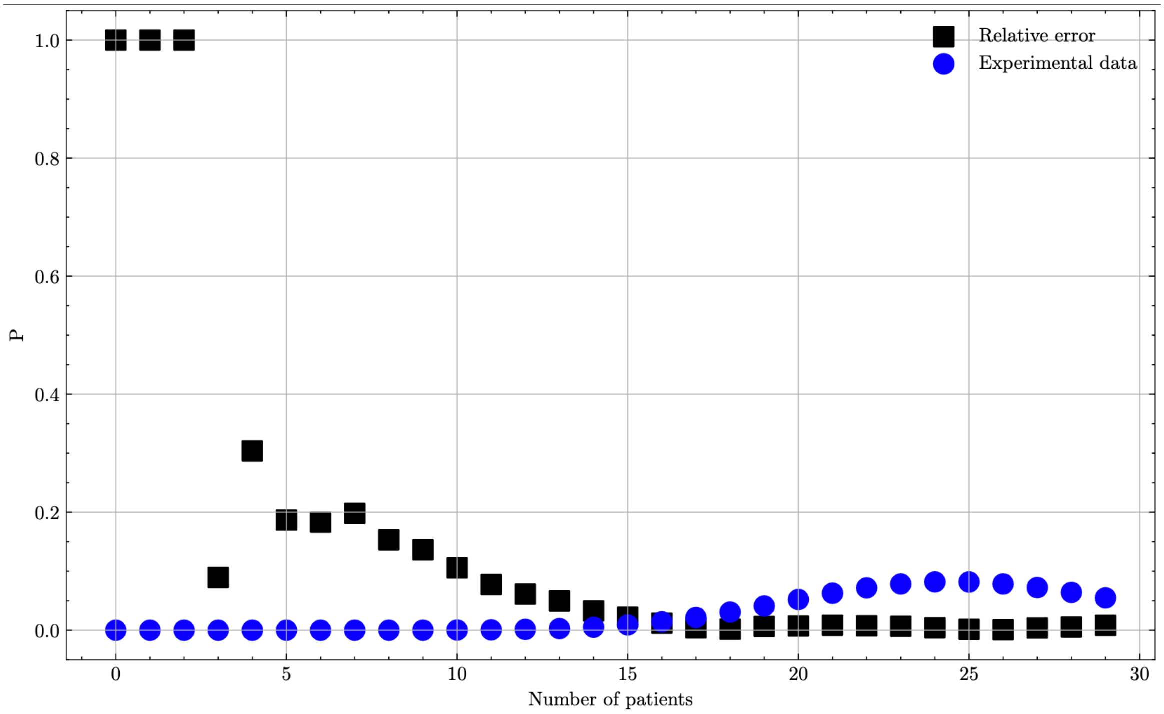

図 14: 二項分布の理論値と相対誤差

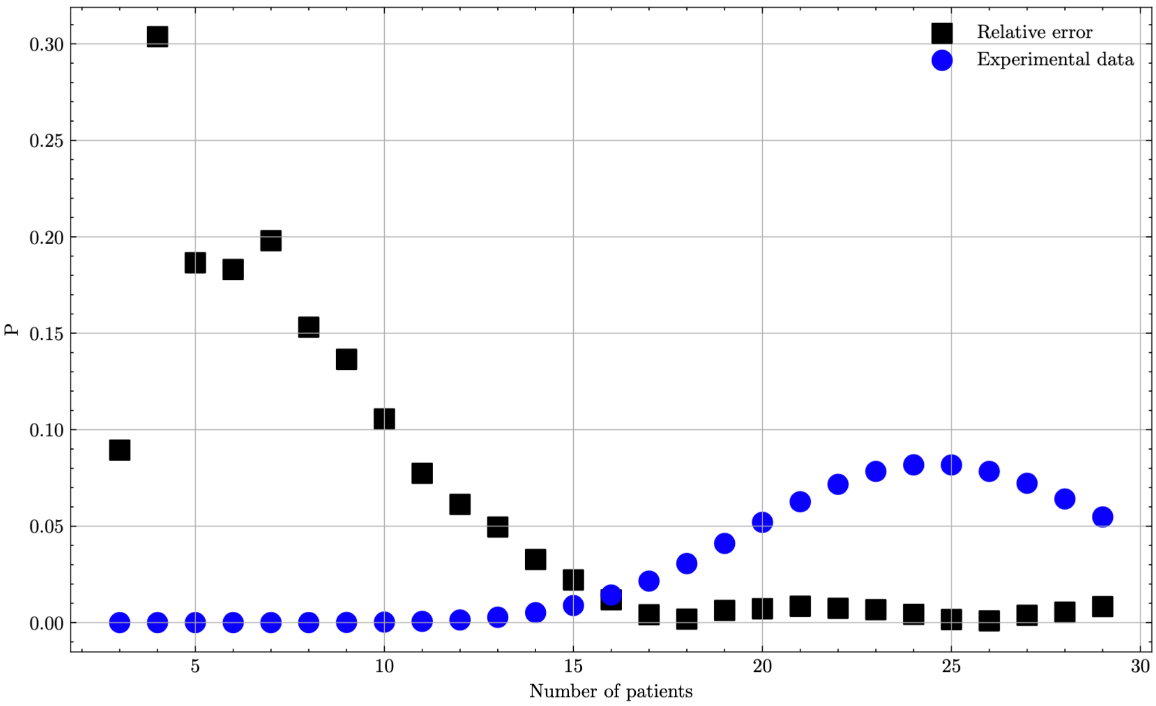

図 15: 二項分布の理論値と相対誤差(患者数が 0, 1, 2 の時を除く)

結論

今回の実験では、乱数を用いてπを求める方法、buffon の針問題、ランダムウォーク、二項分布に従う確率分布をシミュレーションするプログラムを実行し、その結果を比較した。その結果、πを求める方法として、入門的モンテカルロ法、当たり外れのモンテカルロ法があることがわかった。また、これら２つの方法の精度について考察した結果、試行回数が10^5を超えると、精度の向上が緩やかになることがわかった。また、30000回の時に、πの値の誤差が、入門的モンテカルロ法、当たり外れのモンテカルロ法の方が buffon の針よりも小さく、精度が良いことがわかった。そして、ランダムウォーク、二項分布に従う確率分布の理論値と実験値の比較を行った結果、理論値と実験値は酷似していることがわかった。また、理論値と実験値の相対誤差を計算した結果、平均値から離れるにつれて、誤差が大きくなるという傾向があることがわかった。しかし、平均値付近の値は理論値に近づくため、誤差が小さくなることがわかった。

参考文献

[1] “E5 教科書”