

# ミクロデータサイエンス

## Problemset1

2125178  
廣江友哉

May 22, 2024

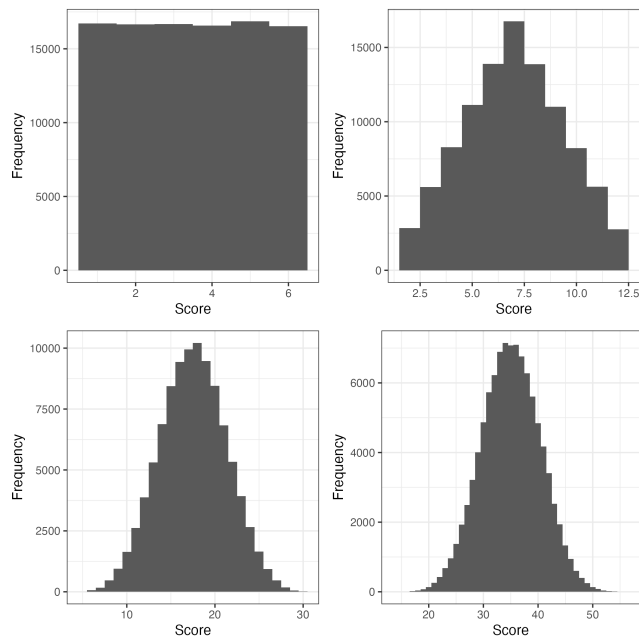
### 1 サイコロと一様分布

#### 1.1 分布の形

どの実現値においても同じ確率  $1/6$  を取る長方形のような形の分布になる。

#### 1.2 コードの実行結果の考察

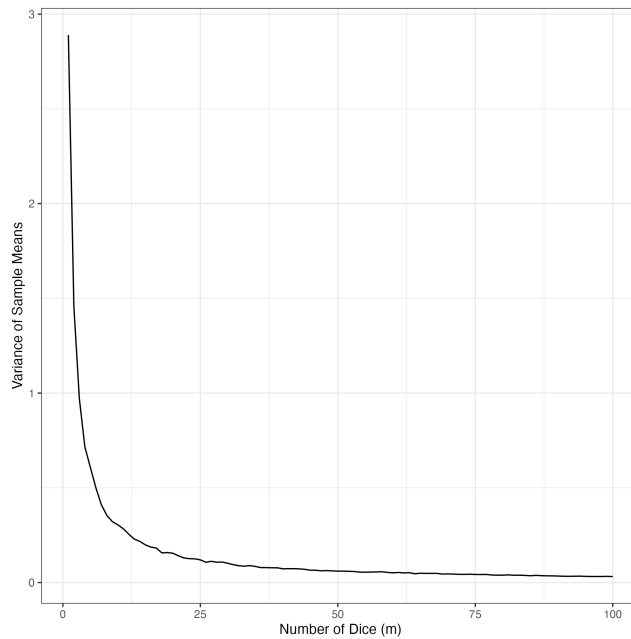
```
1 calculate_scores <- function(m, n, seed_num) {  
2     set.seed(seed_num)  
3     score <- replicate(n, sum(roll_dice(m)))  
4     df_score <- dplyr::tibble(score)  
5     plot_scores(df_score)  
6 }
```



上図は、左上から順に、 $m = 1, 2, 5, 10$  の時の分布である。 $m = 1$  の時は 1-1 で回答したように一様分布に従っている。 $m = 2$  から  $m = 10$  にかけて少しずつ正規分布に近似していくのが見て取れる。ここから、一様分布に従う確率変数の和は繰り返すと正規分布に近似することが考察できる。

### 1.3 確率分布の和の分布のシミュレーション adv

#### 1.4 標本平均の分散の変化



サイコロの数が増える、つまり、和をとる確率変数の数が増えるほど標本平均の分散は指数関数的に減少していく傾向になることがわかる。

## 2 指数分布と故障

### 2.1 タブレットの導入 (学校側)

#### 2.1.1 指数分布のパラメータ $\lambda$

「サーバの故障回数は、1 年間に平均 2.5 回の指数分布に従う」ので、以下のような式を導ける。

$$E[t] = \frac{1}{\lambda} = 2.5$$

したがって、パラメータ  $\lambda$  は 2.5 となる。

#### 2.1.2 1 年間で発生する期待コスト

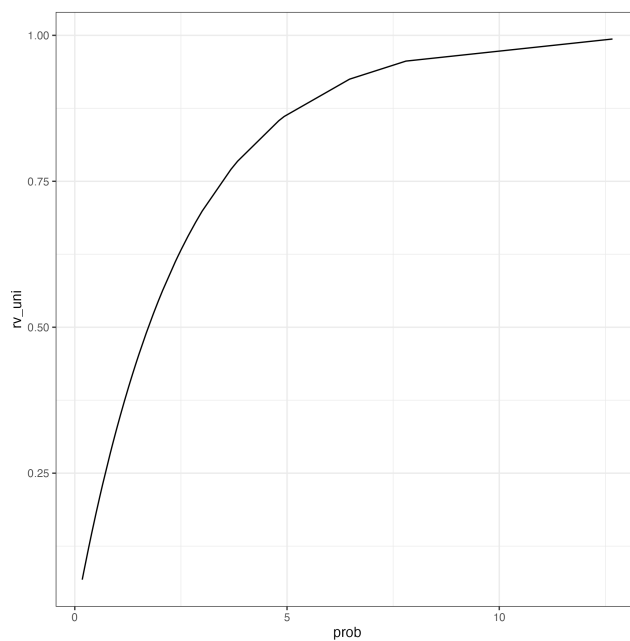
サーバは 1 年間で平均して 2.5 回故障するので、期待コストは以下の式で導る。

$$10000 + 90000 \times 2.5 = 235000$$

したがって、期待コストは 235,000 円。

### 2.1.3 サーバ故障確率の累積分布関数

```
1      simulate_exponential <- function(lambda, sample_size,  
2        seed_num) {  
3        set.seed(seed_num)  
4        df_cdf <- dplyr::tibble(  
5          rv_uni = runif(sample_size)  
6        ) |>  
7          dplyr::mutate(  
8            prob = -(log(1 - rv_uni) / (  
9              lambda))  
10         )  
11         plot_cdf(df_cdf)  
12       }
```



サーバ故障確率の累積分布関数のプロットは、X 軸方向が時間、Y 軸方向が確率となっている。  
X 時点においてサーバがすでに故障している確率が確認できる。したがって、5 年経つと約 0.875 の確率で故障していることなどが見て取れる。