

ミクロデータサイエンス

Problemset2

2125178
廣江友哉

June 19, 2024

1 囚人と主観的確率

1.1 3つの扉

事前確率として、扉が正解である事象 X の確率は $P(X) = \frac{1}{3}$ とする。また、看守がいずれかの扉を選ぶ事象を Y とする。例えば、囚人が A の扉を選び看守が不正解の扉として C の扉を開けたとすると、A と B それぞれの扉が正解の扉である事後確率はベイズの定理を用いて以下のように表せる。

$$P(X = A|Y = C) = \frac{P(Y = C|X = A)}{P(Y = C)}P(X = A) \quad (1)$$

$$P(X = B|Y = C) = \frac{P(Y = C|X = B)}{P(Y = C)}P(X = B) \quad (2)$$

ここで、A の扉が正解だったとすると以下のように具体的な確率を求めることができる。

$$P(Y = C) = \frac{1}{2} \quad P(Y = C|X = A) = \frac{1}{2} \quad (3)$$

$$P(Y = C) = \frac{1}{2} \quad P(Y = C|X = B) = 1 \quad (4)$$

上記の式は、囚人が正解の扉を選んでいる場合看守は残りのどちらの扉も選ぶことができるのでその確率を仮に $\frac{1}{2}$ とし、囚人が不正解の扉を選んでいる場合に看守は扉を一つしか選択する余地がないので確率が 1 となることを表している。上記の確率の値を使用して、ベイズの定理の式 1 と式 2 を計算すると、

$$P(X = A|Y = C) = \frac{\frac{1}{2}}{\frac{1}{2}} \times \frac{1}{3} = \frac{1}{3} \quad (5)$$

$$P(X = B|Y = C) = \frac{1}{\frac{1}{2}} \times \frac{1}{3} = \frac{2}{3} \quad (6)$$

したがって、囚人は A から B へ扉を変更した方が正解の確率が上がるので、囚人の扉を選び直すという行為は正しいことがわかる。

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 λ

「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 λ 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 サーバー故障確率の累積分布関数

```

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

```

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。