Basics on Statistical Simulation

Chanseok Park (박찬석)

Applied Statistics Laboratory
Department of Industrial Engineering
Pusan National University

September 16, 2021

Overview

1. Definition of Statistical Simulation

2. Problems can be solved by statistical simulations
    - Estimation of the probability when tossing a die or dice
    - Ratio of the circumference of a circle

Overview

1. Definition of Statistical Simulation

2. Problems can be solved by statistical simulations
    - Estimation of the probability when tossing a die or dice
    - Ratio of the circumference of a circle

Definition of Statistical Simulation

Definition (wikipedia)

Monte Carlo methods (or Monte Carlo experiments) are a broad class of computational algorithms that rely on **repeated random sampling** to obtain numerical results. Their essential idea is using randomness to solve problems that might be **deterministic in principle**. They are often used in physical and mathematical problems and are most useful when it is difficult or impossible to use other approaches. *Monte Carlo methods are mainly used in three(~~actually much more~~?) problem classes*: optimization, numerical integration, and generating draws from a probability distribution.

- Monte Carlo methods $\Longrightarrow$ statistical or stochastic simulation.
- It can solve **stochastic** and **deterministic** problems.
- More applications in statistics: power, distribution, efficiency, bias-correction, etc.

Tips for Simulation

From my teaching experience, many students are confused with:

- sample size n and
- the iteration (replication) number I.

- Sample size n is in general the number of random variable(s).
- The iteration number I is the number of repeated experiments.

Problems can be solved by statistical simulations

In this class, we will solve probability and mathematics problems by statistical simulations.

1. Probability: estimation of the probability when tossing a die or dice.
2. Mathematics: the ratio of the circumference of a circle, $\pi = 3.14\ldots$

(1) Estimation of the probability

We are interested in the probability when tossing a single die.

R Code for tossing a fair die

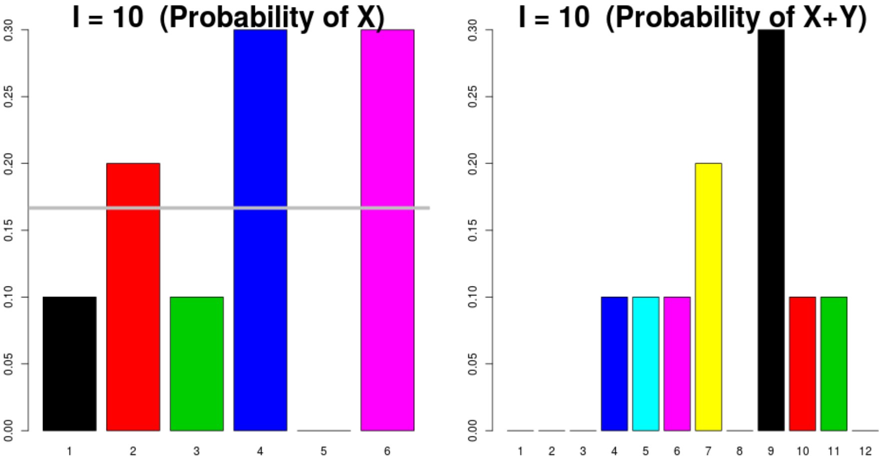
```
set.seed(1)
ITER = 10000
x = sample(1:6, size=ITER, replace=TRUE)
table(x) / ITER
table( factor(x, levels=1:6)) / ITER  # cosmetic
```

What is the prob. of sum of two outcomes when tossing two dice.

R Code for tossing two dice (sum of two)

```
set.seed(1)
ITER = 10000
x = sample(1:6, size=ITER, replace=TRUE)
y = sample(1:6, size=ITER, replace=TRUE)
table(x+y) / ITER
table( factor(x+y, levels=1:12)) / ITER  # cosmetic
```

(1) Estimation of the probability



~~Homework~~: 1. Find the probability of the difference of two dice.

2. What if a die is not fair?

(2) Ratio of the circumference of a circle, $\pi = 3.14\ldots$

The idea is very simple — consider a dart game.
The probability that a pin drops in a circle when a pin was thrown at random is given by:

$$\frac{\pi}{4} = \frac{\text{Area of unit circle}}{\text{Area of square}} \approx \frac{\#\text{ of pins in a circle}}{\#\text{ of all the pins}} = \hat{p}.$$

Thus, we have $\pi \approx 4 \times \hat{p}$.
$\Diamond$ NB: Avoid using for() loop in R.

R Code for finding $\pi = 3.14\ldots$

```
set.seed(1)
ITER = 100000
x = runif(ITER, min=-1, max=1)
y = runif(ITER, min=-1, max=1)
4 * sum (x^2 + y^2 < 1) / ITER
```

(2) Ratio of the circumference of a circle, $\pi = 3.14\ldots$