

TÍNH XÁC SUẤT BẰNG MÔ PHỎNG ĐƠN GIẢN TRONG R THỐNG KÊ MÁY TÍNH & ỨNG DỤNG

(Vũ Quốc Hoàng – 29/09/2018)

Đề: Bốc ngẫu nhiên 10 viên bi trong hộp gồm 10 bi Đỏ và 20 bi Đen. Tính xác suất bốc được đúng 5 bi Đỏ trong các trường hợp:

- a) Bốc có hoàn lại.
- b) Bốc không hoàn lại.

Giải:

Bốc có hoàn lại:

Xác suất bốc được bi Đỏ trong hộp gồm 10 bi Đỏ và 20 bi Đen là:

$$p = \frac{10}{10 + 20} = \frac{1}{3}$$

Đặt X là số bi Đỏ bốc được trong $n = 10$ lần bốc thì $X \sim \text{Binomial}(n, p) = \text{Binomial}(10, \frac{1}{3})$. Xác suất để bốc được đúng 5 bi Đỏ là:

$$P(X = k = 5) = C_n^k p^k (1 - p)^{n-k} = C_{10}^5 \left(\frac{1}{3}\right)^5 \left(\frac{2}{3}\right)^5 = 0.1365$$

Chạy R

Tính xác suất trên bằng R:

```
choose(10, 5)*(1/3)^5*(2/3)^5  
[1] 0.1365645
```

Mô phỏng (simulate) bằng R:

Hàm thực hiện thí nghiệm bốc bi và kiểm tra biến cố “bốc được đúng 5 bi Đỏ” có xảy ra hay không

```
thinghiem <- function()  
{  
  hopbi <- 1:30  
  bido <- 1:10  
  muoivien <- sample(hopbi, 10, replace=TRUE)  
 kiemtra <- muoivien %in% bido
```

```

        sum(kiemtra) == 5
    }

```

Hàm thực hiện N lần thí nghiệm bốc bi và tính tần suất của biến cố cần tính

```

tansuat <- function(N)
{
    ketqua <- replicate(N, thinghiem())
    sum(ketqua)/N
}

```

Chạy với N đủ lớn thì tần suất sẽ xấp xỉ xác suất cần tính. Dấu hiệu nhận biết N đủ lớn là khi N nhỏ thì kết quả dao động mạnh (không ổn định, lúc lớn lúc nhỏ) còn khi N đủ lớn thì kết quả dao động ít hơn (ổn định, xem xem cùng bậc). Nhưng N lớn quá thì tốn thời gian chạy. Cũng lưu ý là mỗi lần chạy sẽ có kết quả khác nhau do ngẫu nhiên.

Chẳng hạn thử với N = 500000

```

tansuat(500000)
[1] 0.136996

```

Bốc không hoàn lại:

Đặt X là số bi Đỏ bốc được trong $n = 10$ lần bốc thì $X \sim \text{Hypergeometric}(n, A, B) = N(10, 10, 20)$. Xác suất để bốc được đúng 5 bi Đỏ là:

$$P(X = k = 5) = \frac{C_A^k C_B^{n-k}}{C_{A+B}^n} = \frac{C_{10}^5 C_{20}^5}{C_{30}^{10}} = 0.13$$

Chạy R

Tính xác suất trên bằng R:

```

choose(10, 5)*choose(20, 5)/choose(30, 10)
[1] 0.1300385

```

Mô phỏng bằng R:

Hàm thực hiện thí nghiệm bốc bi và kiểm tra biến cố “bốc được đúng 5 bi Đỏ” có xảy ra hay không

```

thinghiem <- function()
{
    hopbi <- 1:30

```

```

bido <- 1:10
muoivien <- sample(hopbi, 10, replace=FALSE)
kiemtra <- muoivien %in% bido
sum(kiemtra) == 5
}

```

Hàm thực hiện N lần thí nghiệm bốc bi và tính tần suất của biến cố cần tính

```

tansuat <- function(N)
{
    ketqua <- replicate(N, thinghiem())
    sum(ketqua)/N
}

```

Chạy thử với N = 500000

```

tansuat(500000)
[1] 0.130068

```

Lưu ý: kết quả 2 câu trên gần như nhau. Khi A, B rất lớn so với n thì phân phối siêu hình rất gần với phân phối nhị thức.

Bài tập luyện tập: Bỏ ngẫu nhiên 2 lá bài trong bộ bài 52 lá. Chia đều ngẫu nhiên 50 lá còn lại cho 5 người. Tính xác suất 1 người nhận được 4 lá Ách. Kiểm tra lại bằng cách mô phỏng trên R.