

TÍNH KÌ VỌNG VÀ PHÂN PHỐI BẰNG MÔ PHỎNG ĐƠN GIẢN TRONG R THỐNG KÊ MÁY TÍNH & ỨNG DỤNG

(Vũ Quốc Hoàng – 04/11/2018)

Đề: Giả sử chọn ngẫu nhiên một điểm trên một thanh có chiều dài 1 đơn vị, cắt tại điểm đó thành hai đoạn và giữ lại đoạn dài hơn. Tính kì vọng của chiều dài đoạn giữ lại theo 2 cách:

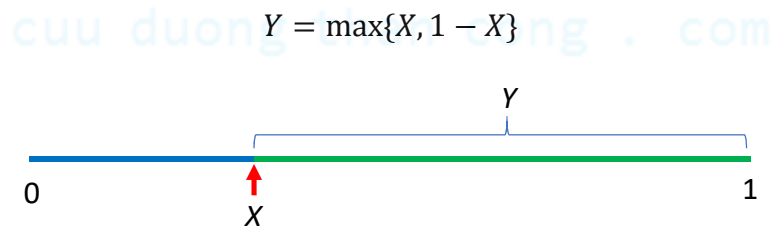
- Dùng công thức tính kì vọng của biến ngẫu nhiên phái sinh.
- Tìm phân phối của chiều dài đoạn giữ lại rồi từ đó tính kì vọng theo công thức thông thường.

Giải:

Gọi X là vị trí ngẫu nhiên chọn trên thanh thì $X \sim U(0, 1)$. Do đó X là b.n.n liên tục với hàm mật độ xác suất

$$f_X(x) = \begin{cases} 1 & \text{khi } 0 \leq x \leq 1 \\ 0 & \text{khác} \end{cases}$$

Gọi Y là chiều dài của đoạn được giữ lại (tức là đoạn dài hơn) thì



Dùng công thức tính kì vọng của biến ngẫu nhiên phái sinh:

Ta có Y là biến ngẫu nhiên phái sinh từ biến ngẫu nhiên X qua hàm số

$$r(x) = \max\{x, 1 - x\}, 0 \leq x \leq 1$$

Theo công thức tính kì vọng của biến ngẫu nhiên phái sinh liên tục

$$E(Y) = E(r(X)) = \int_{-\infty}^{\infty} r(x)f_X(x)dx$$

ta có

$$\begin{aligned}
E(Y) &= E(\max\{X, 1 - X\}) = \int_{-\infty}^{\infty} \max\{x, 1 - x\} f_X(x) dx = \int_0^1 \max\{x, 1 - x\} dx \\
&= \int_0^{1/2} \max\{x, 1 - x\} dx + \int_{1/2}^1 \max\{x, 1 - x\} dx = \int_0^{1/2} (1 - x) dx + \int_{1/2}^1 x dx \\
&= x \Big|_0^{1/2} - \frac{x^2}{2} \Big|_0^{1/2} + \frac{x^2}{2} \Big|_{1/2}^1 = \frac{1}{2} - \frac{1}{8} + \frac{1}{2} - \frac{1}{8} = \frac{3}{4}
\end{aligned}$$

Tìm phân phối của chiều dài đoạn giữ lại rồi từ đó tính kì vọng theo công thức thông thường:

Trước hết ta tìm hàm phân phối tích lũy của b.n.n $Y = \max\{X, 1 - X\}$

$$F_Y(y) = P(Y \leq y) = P(\max\{X, 1 - X\} \leq y), y \in \mathbb{R}$$

Xét các trường hợp của y :

- $y < 1/2$: $(\max\{X, 1 - X\} \leq y) = \emptyset$ vì $0 \leq X \leq 1$ nên $1/2 \leq \max\{X, 1 - X\}$. Khi đó
 $P(\max\{X, 1 - X\} \leq y) = P(\emptyset) = 0$
- $1/2 \leq y \leq 1$: $(\max\{X, 1 - X\} \leq y) = (1 - y \leq X \leq y)$.¹ Khi đó
 $P(\max\{X, 1 - X\} \leq y) = P(1 - y \leq X \leq y) = \int_{1-y}^y f_X(x) dx = \int_{1-y}^y dx = 2y$
- $y > 1$: $(\max\{X, 1 - X\} \leq y) = \Omega$ vì $0 \leq X \leq 1$ nên $\max\{X, 1 - X\} \leq 1$. Khi đó
 $P(\max\{X, 1 - X\} \leq y) = P(\Omega) = 1$

Từ đó ta có

$$F_Y(y) = \begin{cases} 0 & \text{khi } y < 1/2 \\ 2y & \text{khi } 1/2 \leq y \leq 1 \\ 1 & \text{khi } y > 1 \end{cases}$$

Bằng cách lấy đạo hàm của hàm phân phối tích lũy, ta có hàm mật độ xác suất của Y là

$$f_Y(y) = F'_Y(y) = \begin{cases} 2 & \text{khi } 1/2 \leq y \leq 1 \\ 0 & \text{khác} \end{cases}$$

Từ đó ta tính được kì vọng của Y theo công thức thông thường

$$E(Y) = \int_{-\infty}^{\infty} y f_Y(y) dy = \int_{1/2}^1 2y dy = y^2 \Big|_{1/2}^1 = 1 - \frac{1}{4} = \frac{3}{4}$$

Lưu ý: nhận xét thêm là $Y \sim U(1/2, 1)$.

Mô phỏng bằng R:

Hàm thực hiện thí nghiệm cắt và giữ lại thanh dài

```
thinghiem <- function()
```

¹ Vẽ hình ra để thấy:)

```
{
  X <- runif(1)
  Y <- max(X, 1 - X)
  return(Y)
}
```

Hàm thực hiện N lần thí nghiệm và tính trung bình chiều dài thanh giữ lại (b.n.n Y)

```
trungbinh <- function(N)
{
  ketqua <- replicate(N, thinghiem())
  return(mean(ketqua))
}
```

Chạy với N đủ lớn thì trung bình mẫu sẽ xấp xỉ kì vọng 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

```
> trungbinh(500000)
[1] 0.7501527
```

Hàm thực hiện N lần thí nghiệm và vẽ histogram cho chiều dài thanh giữ lại (b.n.n Y)

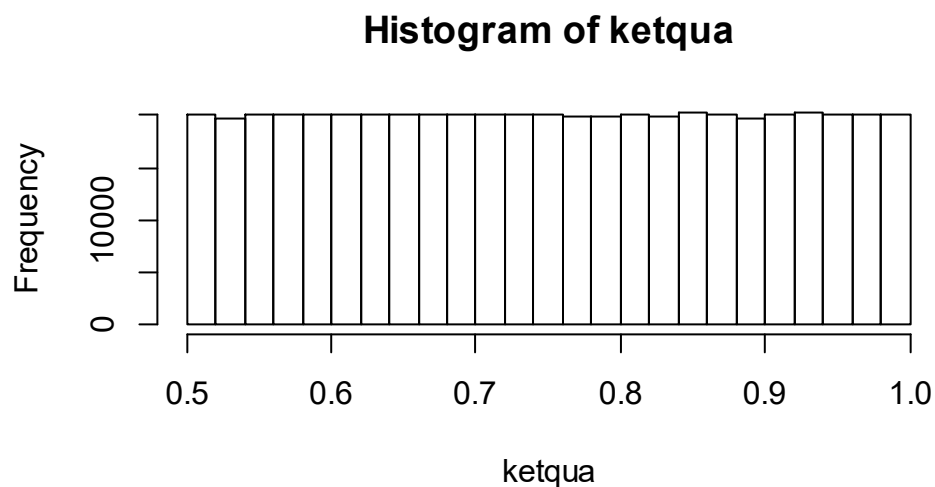
```
phanphoi <- function(N)
{
  ketqua <- replicate(N, thinghiem())
  hist(ketqua)
}
```

Chạy với N đủ lớn thì histogram của mẫu sẽ “xấp xỉ” hàm mật độ xác suất cần tìm. Dấu hiệu nhận biết N đủ lớn là khi N nhỏ thì kết quả dao động mạnh (lúc thế này lúc thế khác) còn khi N đủ lớn thì kết quả dao động ít hơn. 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

```
> phanphoi(500000)
```

cho kết quả là



Từ histogram ta kiểm chứng là $Y \sim U(1/2, 1)$.

Bài tập luyện tập: Giả sử chọn ngẫu nhiên một số thực r trong khoảng $[0, 1]$. Tính kì vọng của diện tích hình tròn có bán kính r theo 2 cách:

- Dùng công thức tính kì vọng của biến ngẫu nhiên phái sinh.
- Tìm phân phối của chiều dài đoạn giữ lại rồi từ đó tính kì vọng theo công thức thông thường.

Kiểm tra lại bằng cách mô phỏng trên R.