

Thực hành xác suất thống kê

Bài 3: BIẾN NGẪU NHIÊN

1. Tính toán

Ví dụ 1 a: Với $k = 0, 1, \dots, 8$, tính các xác suất

$$P(X = k) = C_8^k 0.3^k 0.7^{8-k}$$

```
# Tao vec to k có giá trị từ 0 đến 8
k = 0:8
# Viết hàm xác suất
p <- function(k) choose(8,k) * 0.3^k * 0.7^(8-k)
# Tính giá trị xác suất P(X = k), k = 0,1,...,8
p(k)

# Kiểm tra tổng các xác suất bằng 1
sum(p(k))
```

1. Tính toán

Ví dụ 1 b: Cho biến ngẫu nhiên X có hàm mật độ như sau

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < +\infty$$

với μ và $\sigma > 0$ là hai tham số. Viết hàm pdf của X để tính giá trị $f(x)$ tại một điểm bất kỳ.

```
# Viet ham mat do cua bien ngau nhien X ~ N(mu = 0, sigma^2 = 1)
f <- function(x, mu=0, sigma=1){
  1/sqrt(2*pi*sigma^2) * exp(-(x-mu)^2/(2*sigma^2))
}

# Kiem tra tích phan bang 1
integrate(function(x) f(x,0,1),lower=-Inf,upper=Inf)
1 with absolute error < 9.4e-05

# Tinh f(0)
f(0)
```

2. Biểu diễn bằng đồ thị

Ví dụ 2 a: Vẽ đồ thị hàm xác suất trong Ví dụ 1 a)

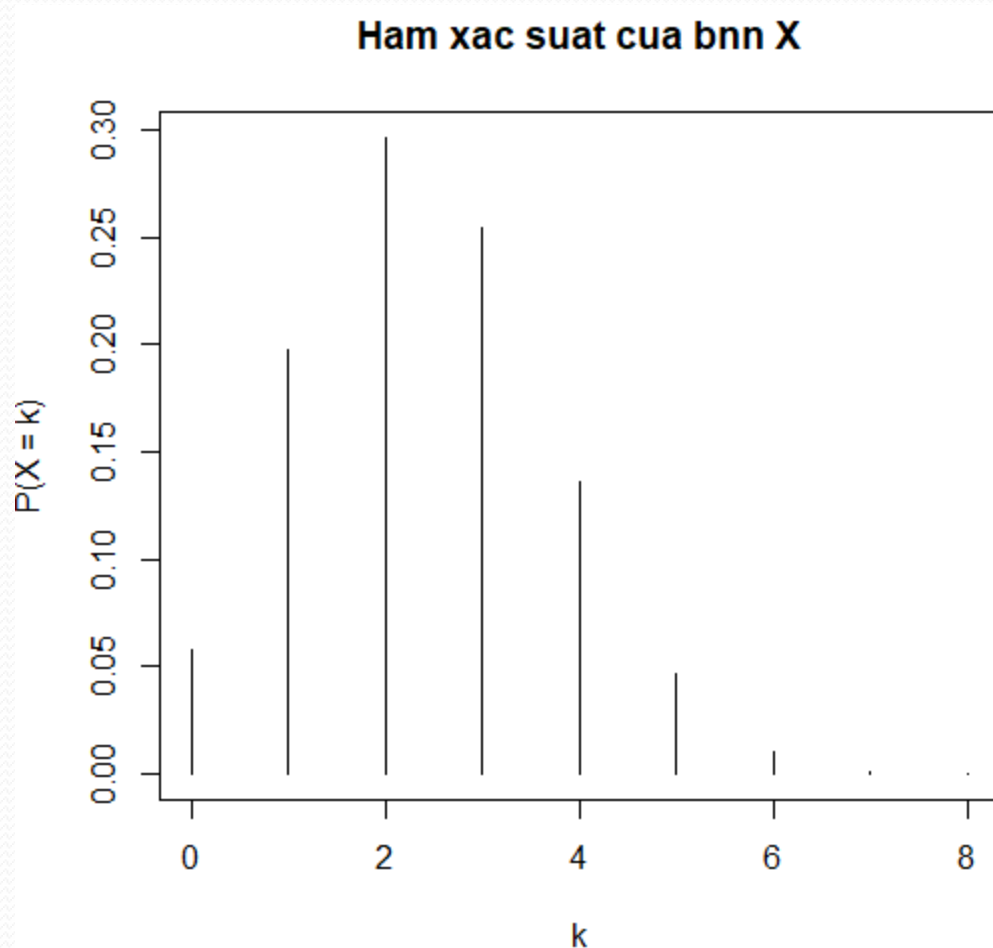
$$P(X = k) = C_8^k 0.3^k 0.7^{8-k} \text{ với } k = 0, 1, \dots, 8.$$

`plot(x, f(x), type = "loại đồ thị", xlab = "nhãn trục x", ylab = "nhãn trục y", main = "tên đồ thị")`

```
# Vẽ hàm xác suất ở ví dụ 1a
plot(k, p(k), type = "h", xlab = "k", ylab = "P(X = k)", main =
"Hàm xác suất của biến X")
```

2. Biểu diễn bằng đồ thị

Ví dụ 2 a: Vẽ đồ thị hàm xác suất trong Ví dụ 1 a)



2. Biểu diễn bằng đồ thị

Ví dụ 2 b: Vẽ đồ thị hàm mật độ xác suất trong Ví dụ 1 b)

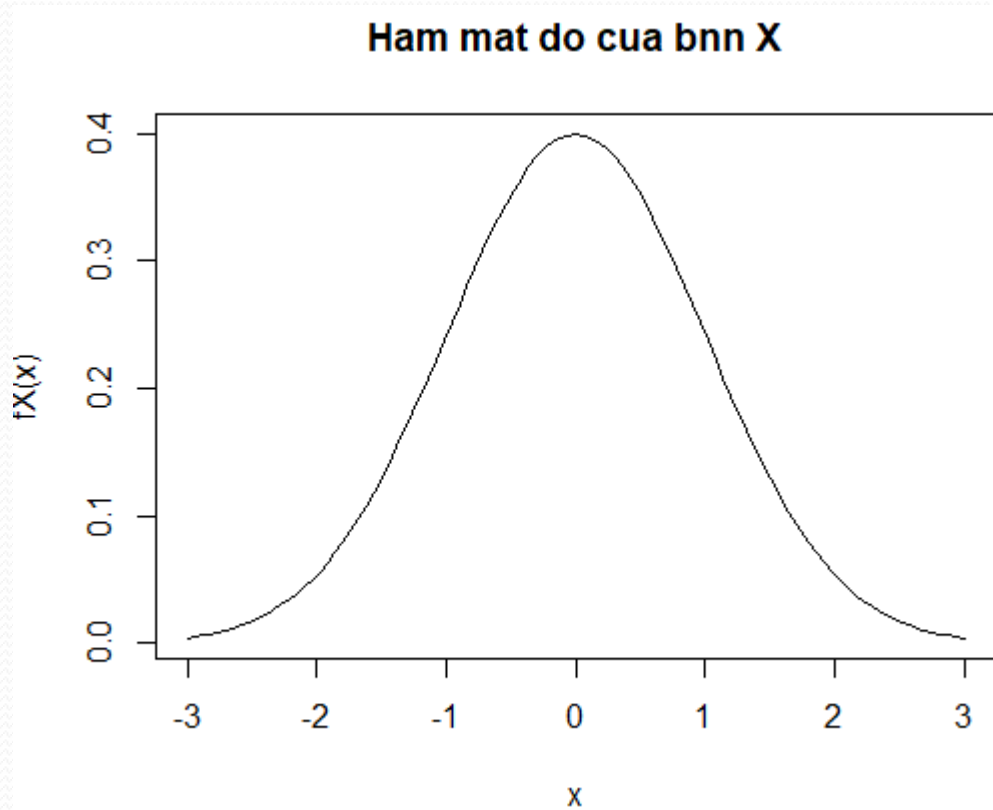
$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < +\infty$$

curve(f(x), from = cận dưới, to = cận trên, xlab = “nhãn trục x”, ylab = “nhãn trục y”, main = “tên đồ thị”)

```
# Vẽ hàm mật độ xác suất ở ví dụ 1b
curve(f(x,0,1),from=-3,to=3, xlab = "x", ylab = "fX(x)", main =
"Hàm mật độ của biến X")
```

2. Biểu diễn bằng đồ thị

Ví dụ 2 b: Vẽ đồ thị hàm mật độ xác suất trong Ví dụ 1 b)



3. Hàm phân phối

3.1 Định nghĩa

$$F_X(x) = P(X \leq x)$$

- X rời rạc

$$F_X(x) = \sum_{k \in X(\Omega) \cap (-\infty, x]} P(X = k)$$

- X liên tục

$$F_X(a) = \int_{-\infty}^a f_X(x) dx$$

3. Hàm phân phối

3.2 Tính toán

Ví dụ 3 a: Tính $F_X(4) = P(X \leq 4) = \sum_{k=0}^4 P(X = k)$

```
# Viet ham pp xac suat FX
F <- function(k) sum(p(0:k))
# Vecto hoa ham FX
F <- Vectorize(F)
# Tinh F(4)
F(4)
```

3. Hàm phân phối

3.2 Tính toán

Ví dụ 3 b: Tính $F_X(1.96) = P(X \leq 1.96) = \int_{-\infty}^{1.96} f(x)dx$

```
# Viet ham pp xac suat
F2 <- function(a,mu = 0, sigma = 1){
  integrate(function(x) f(x,mu,sigma), lower = -Inf, upper =a)$value
}
# Vecto hoa ham
F2 = Vectorize(F2)
F2(1.96)
```

3. Hàm phân phối

3.3 Vẽ đồ thị

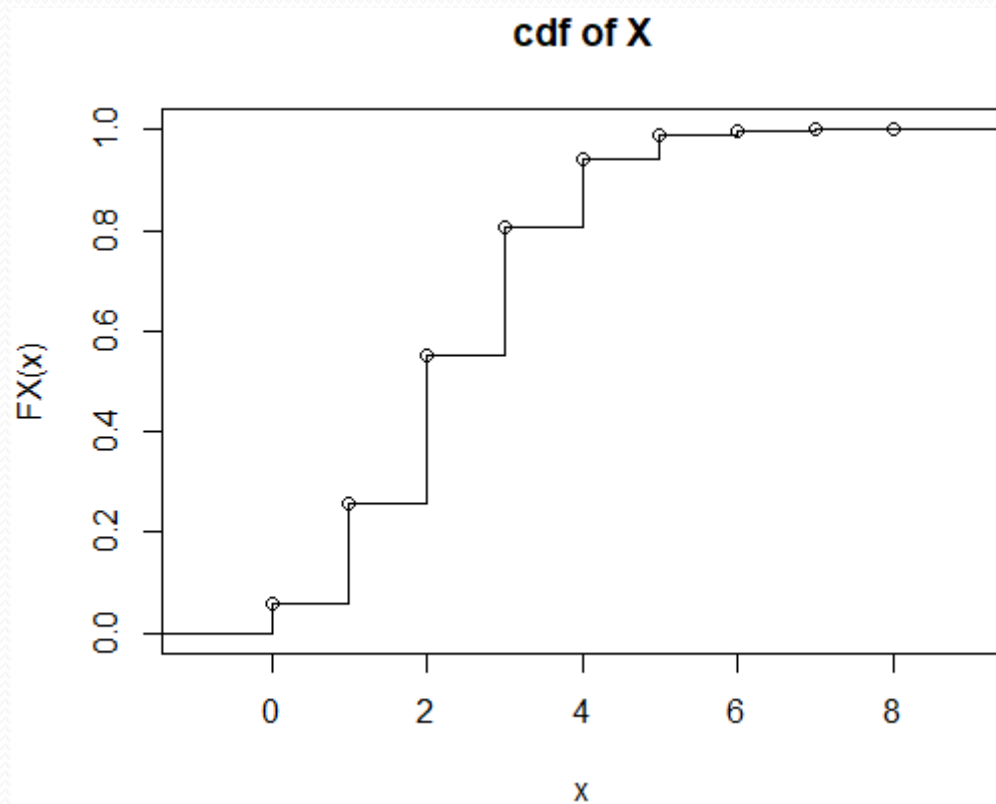
Ví dụ 4 a: Vẽ hàm phân phối xác suất của X ở ví dụ 1 a.

```
plot(stepfun(k, c(0, F(k))), ylab = "FX(x)", main = "cdf of X")
```

3. Hàm phân phối

3.3 Vẽ đồ thị

Ví dụ 4 a: Vẽ hàm phân phối xác suất của X ở ví dụ 1 a.



3. Hàm phân phối

3.3 Vẽ đồ thị

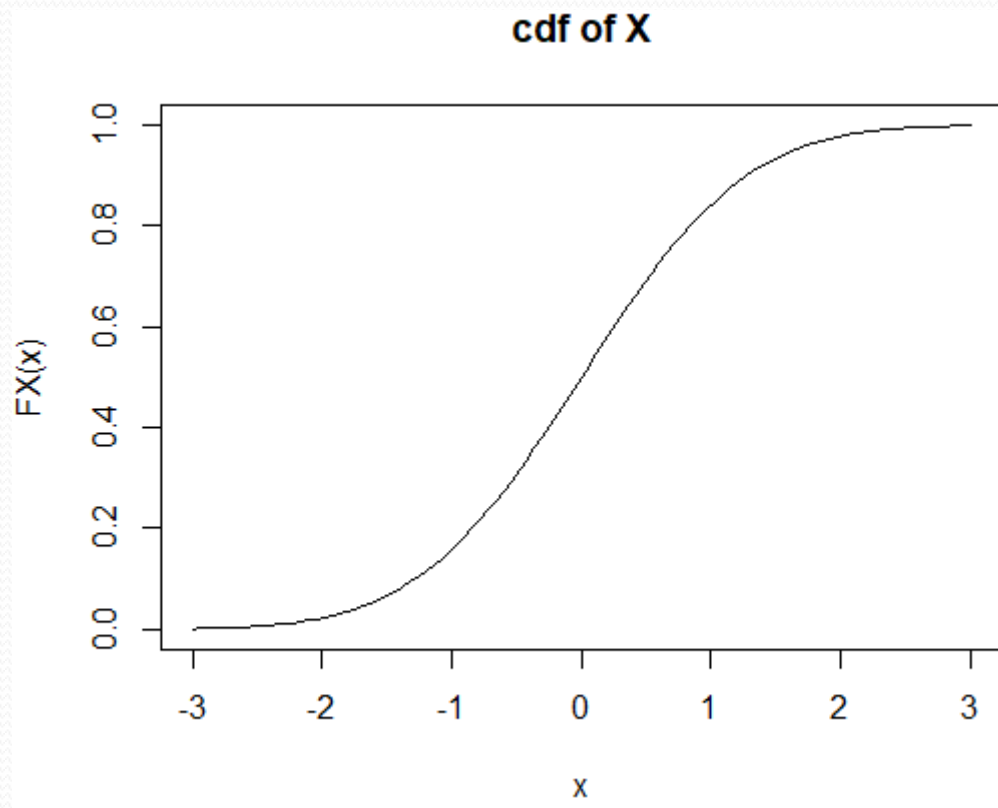
Ví dụ 4 b: Vẽ hàm phân phối xác suất của X ở ví dụ 1 b.

```
curve(F2(x), from = -3, to = 3, ylab = "FX(x)", main = "cdf of X")
```

3. Hàm phân phối

3.3 Vẽ đồ thị

Ví dụ 4 b: Vẽ hàm phân phối xác suất của X ở ví dụ 1 b.



4. Phân vị

4.1 Định nghĩa

Cho $p \in (0; 1)$ và X là một biến ngẫu nhiên.

- Nếu X rời rạc, phân vị mức p của X , ký hiệu là x_p xác định

$$x_p = \inf\{k \in \mathbb{Z}: F_X(k) \geq p\}$$

- Nếu X liên tục, phân vị mức p của X , ký hiệu x_p , là giá trị thỏa

$$F_X(x_p) = p$$

4. Phân vị

4.2 Tính toán

Ví dụ 5 a: Tính phân vị mức 0.25 của X ở ví dụ 1 a.

```
K = k[F(k) >= 0.25]
```

```
K[1]
```

```
[1] 1
```

```
# kiểm tra lại
```

```
F(0)
```

```
[1] 0.05764801
```

```
F(1)
```

```
[1] 0.2552983
```

4. Phân vị

4.2 Tính toán

Ví dụ 5 b: Tính phân vị mức 0.975 của X ở ví dụ 1 b.

```
# Tìm Nghiệm của Phương trình  $F(x) - x_p = 0$   
uniroot(function(x) F2(x)-0.975, c(-3,3))$root  
[1] 1.959992  
# kiểm tra lại  
F2(1.96)  
[1] 0.9750021
```

Bài tập

Bài 1: Biến ngẫu nhiên P nhận các giá trị giữa 0 và 1, có hàm mật độ xác suất $f_P(p) = 0.07p^{-0.93}$

- a) Tính xác suất $P(P \leq 0.2)$ bằng cách sử dụng hàm $f_P(p)$.
- b) Kiểm tra $f_P(p)$ là hàm mật độ xác suất bằng cách tích diện tích dưới đường cong từ 0 đến 1.

```
# a)
# Viet ham
f <- function(p){0.07*p^(-0.93)}
# Tich xac suat P(P <= 0.2)
integrate(function(x) f(x), lower = 0, upper = 0.2)
# b)
# Kiem tra ham mat do bang cach tinh tich phan tren toan mien
integrate(function(x) f(x), lower = 0, upper = 1)
```

Bài tập

Bài 2: Sử dụng lệnh **sample** để rút một mẫu ngẫu nhiên cỡ 100 từ phân phối với hàm xác suất

X	1	2	3	4	5
$P(X = x)$	0.1	0.2	0.4	0.2	0.1

và chứa mẫu này trong x. Tiếp theo vẽ biểu đồ cột của các xác suất thực nghiệm (mẫu) bằng lệnh **table**.

```
# khai tạo x
x <- sample(1:5,100, replace = TRUE, prob = c(0.1,0.2,0.4,0.2,0.1))
# Tạo bảng thống kê x
table(x)/100

# Vẽ biểu đồ cột của các xác suất thực nghiệm

hist(x, main = "Biểu đồ cột cho dữ liệu X")
```