

BÀI THỰC HÀNH SỐ 3

1. Biến ngẫu nhiên (Bằng cách viết hàm)

a) Biến ngẫu nhiên rời rạc

i. Với $k = 0, 1, \dots, 8$, tính các xác suất $P(X = k) = C_8^k 0.3^k 0.7^{8-k}$

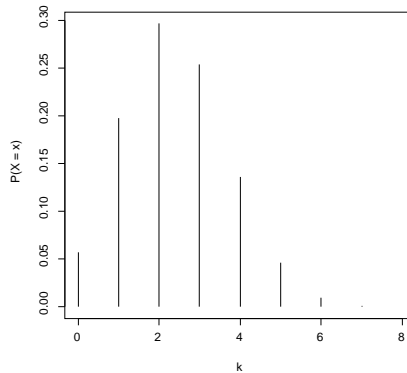
```
k = 0:8;
p = function(k) choose(8,k) * 0.3^k * 0.7^(8-k);
p(k)
[1] 0.05764801 0.19765032 0.29647548 0.25412184
0.13613670
0.04667544
[7] 0.01000188 0.00122472 0.00006561
```

* Ta có thể kiểm tra lại rằng tổng các xác suất bằng 1 với lệnh

```
sum(p(k))
[1] 1
```

ii. Vẽ đồ thị hàm mật độ xác suất của X .

```
plot(k, p(k), type = "h", ylab = "P(X = x)")
```

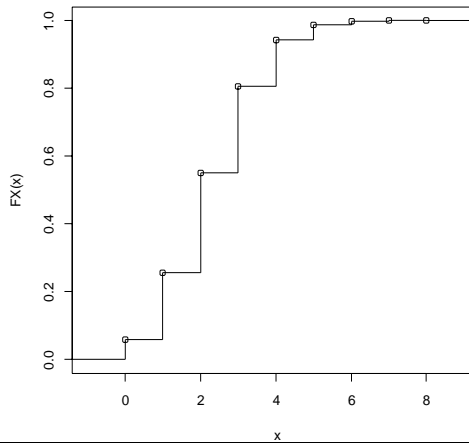


iii. Tính hàm phân phối xác suất $F_X(4) = P(X \leq 4) = \sum_{k=0}^4 P(X = k)$

```
F = function(k) sum(p(0:k))
F = Vectorize(F)
F(4)
[1] 0.9420323
```

iv. Vẽ hàm phân phối xác suất của X .

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



v. Tính phân vị mức 0.25 của X

```
K = k[F(k) >= 0.25]
K[1]
[1] 1
```

* Kiểm tra lại bằng lệnh:

```
F(0)
[1] 0.05764801
F(1)
[1] 0.2552983
```

b) Biến ngẫu nhiên liên tục

i. 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 mật độ xác suất của X để tính giá trị $f(x)$ tại một điểm bất kỳ.

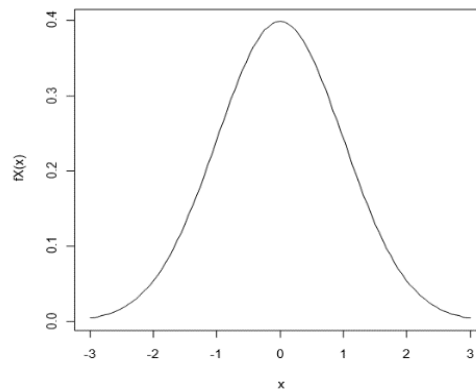
```
f = function(x, mu=0, sigma=1){
  1/sqrt(2*pi*sigma^2) * exp(-(x-mu)^2/(2*sigma^2))
}
```

* Kiểm tra tích phân có bằng 1?

```
integrate(function(x) f(x,0,1),lower=-Inf,upper=Inf)
[1] 1 with absolute error < 9.4e-05
```

ii. Vẽ hàm mật độ xác suất của X .

```
curve(f(x,0,1),from=-3,to=3, ylab = "fX(x)")
```

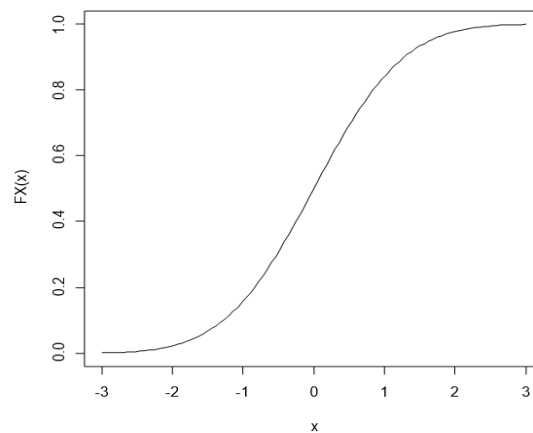


iii. Tính hàm phân phối xác suất $F_X(1.96) = P(X \leq 1.96) = \int_{-\infty}^{1.96} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}} dx$

```
F2 = function(a,mu = 0, sigma = 1){  
  integrate(function(x) f(x,mu,sigma),  
    lower = -Inf, upper = a)$value  
}  
F2 = Vectorize(F2)  
F2(1.96)  
[1] 0.9750021
```

iv. Vẽ hàm phân phối xác suất của X .

```
curve(F2(x), from = -3, to = 3, ylab = "FX(x)")
```



v. Tính phân vị mức 0.975 của X

```
uniroot(function(x) F2(x)-0.975, c(-3,3))$root  
[1] 1.959992
```

* Kiểm tra lại bằng lệnh

```
F2(1.959992)  
[1] 0.9750016
```

2. Một số phân phối xác suất thông dụng (Hàm có sẵn trong R)

a) Cú pháp chung

Với biến ngẫu nhiên X có phân phối (luật) được định nghĩa sẵn trong R, cú pháp chung là như sau:

- Để nhận mật độ của X , dùng lệnh: **d**luật; bằng cách thêm ký tự d trước luật,
- Để nhận giá trị của hàm phân phối của X , dùng lệnh: **p**luật; bằng cách thêm ký tự p trước luật,
- Để nhận giá trị phân vị của X , dùng lệnh: **q**luật; bằng cách thêm ký tự q trước luật,
- Để mô phỏng giá trị của X , dùng lệnh: **r**luật; bằng cách thêm ký tự r trước luật.

b) Hàm mật độ xác suất

Nếu phân phối của X phụ thuộc vào một hoặc nhiều tham số, tham1 và tham2, thì mật độ của X tại x được cho bởi lệnh:

```
dluật(x, tham1, tham2)
```

c) Hàm phân phối xác suất

Nếu phân phối của X phụ thuộc vào một hoặc nhiều tham số, tham1 và tham2, thì hàm phân phối của X tại x được cho bởi lệnh:

```
pluật(x, tham1, tham2)
```

Có thể tính: $P(X > x) = 1 - F_X(x)$ bằng lệnh:

```
pluật(x, tham1, tham2, lower.tail = FALSE)
```

d) Phân vị

Nếu phân phối của X phụ thuộc vào một hoặc nhiều tham số, tham1 và tham2, thì phân vị mức p của X được cho bởi lệnh:

```
qluật(x, tham1, tham2)
```

e) Mô phỏng các phân phối

Nếu phân phối của X phụ thuộc vào một hoặc vài tham số, tham1 và tham2, thì mô phỏng n biến độc lập có cùng phân phối như X bằng lệnh:

```
rluật(n, tham1, tham2)
```

f) Ví dụ

Trường hợp rời rạc

i. Cho biến ngẫu nhiên $X \sim B(8, 0.3)$. Tính giá trị hàm mật độ tại $x = 4$:

$$P(X = 4) = C_8^4 0.3^4 0.7^{8-4}$$

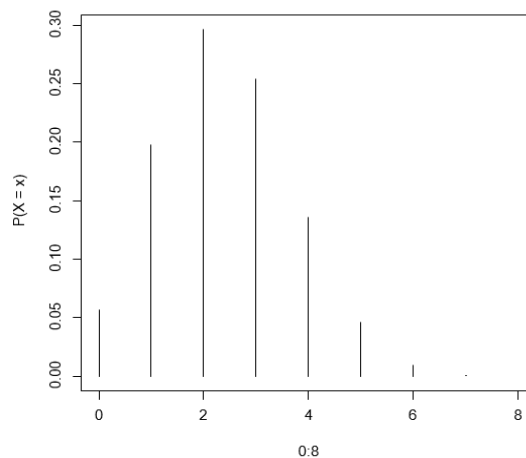
```
dbinom(4, 8, 0.3)
[1] 0.1361
```

* Kiểm tra lại bằng lệnh:

```
choose(8, 4) * 0.3^4 * (1 - 0.3)^(8 - 4)
[1] 0.1361
```

ii. Vẽ đồ thị hàm mật độ xác suất của X

```
plot(0:8, dbinom(0:8, 8, 0.3), type = "h", ylab = "P(X = x) ")
```



iii. Tính $F_X(4) = P(X \leq 4) = \sum_{k=0}^4 P(X = k)$

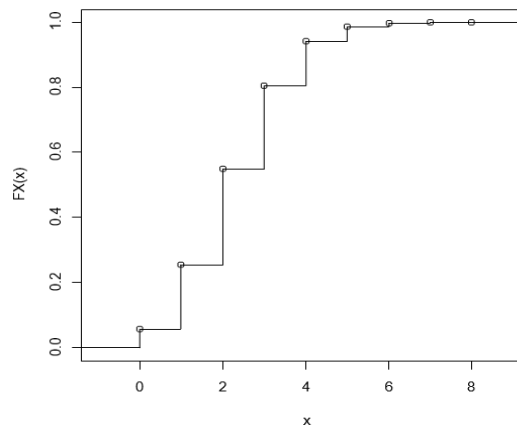
```
pbinom(4, 8, 0.3)
[1] 0.9420324
```

* Ta có thể kiểm tra lại:

```
sum(dbinom(0:4, 8, 0.3))  
[1] 0.9420324
```

iv. Vẽ hàm phân phối xác suất của X

```
plot(stepfun(0:8, c(0, pbinom(0:8, 8, 0.3))), ylab =  
"FX(x)", main = " ")
```



v. Tính phân vị mức 0.25 của X

```
qbinom(0.25, 8, 0.3)  
[1] 1
```

* Ta có thể kiểm tra lại bằng cách tính các giá trị $F_X(x)$ tại $x \in \{0, 1, \dots, 7, 8\}$:

```
pbinom(0:8, 8, 0.3)  
[1] 0.05764801 0.25529833 0.55177381 0.80589565 0.94203235  
[6] 0.98870779 0.99870967 0.99993439 1.00000000
```

Trường hợp liên tục

i. Cho biến ngẫu nhiên $X \sim N(2, 0.12^2)$. Tính giá trị hàm mật độ tại $x = 1.7$:

$$f(1.7) = \frac{1}{\sqrt{2\pi}0.12^2} e^{-\frac{(1.7-2)^2}{2 \times 0.12^2}}, \quad -\infty < x < \infty$$

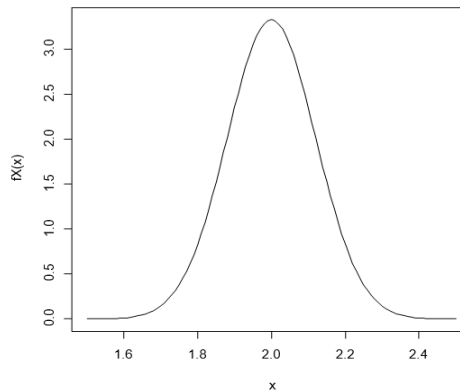
```
dnorm(1.7, 2, 0.12)  
[1] 0.1460692
```

* Kiểm tra lại bằng lệnh:

```
(1 / sqrt(2 * pi * 0.12^2)) * exp(- (1.7 - 2)^2 / (2 * 0.12^2))  
[1] 0.1460692
```

ii. Vẽ đồ thị hàm mật độ xác suất của X

```
curve(dnorm(x, 2, 0.12), from=1.5, to=2.5, ylab = "fX(x)")
```

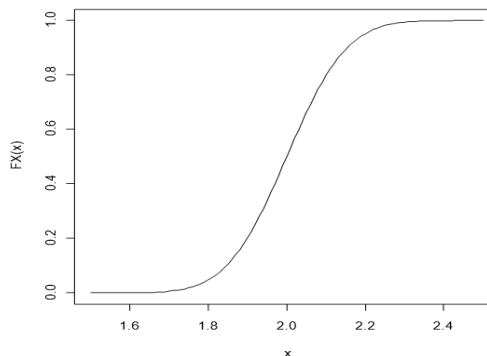


iii. Tính $F_X(2.1) = P(X \leq 2.1) = \int_{-\infty}^{2.1} \frac{1}{\sqrt{2\pi 0.12^2}} e^{-\frac{(x-2)^2}{2 \times 0.12^2}} dx$

```
pnorm(2.1, 2, 0.12)  
[1] 0.7976716
```

iv. Vẽ hàm phân phối xác suất của X

```
curve(pnorm(x, 2, 0.12), from = 1.5, to = 2.5, ylab = "FX(x)")
```



v. Tính phân vị mức 0.975 của X

```
qnorm(0.975, 2, 0.12)
[1] 2.235196
```

Như vậy ta đã tính được phân vị mức $p = 0.975$ của biến $X \sim N(2, 0.12)$: là $x_{0.975}$ sao cho $F_X(x_{0.975}) = 0.975$

* Kiểm tra lại bằng lệnh:

```
pnorm(2.235196, 2, 0.12)
[1] 0.9750002
```

Hàm phân phối	Mật độ	Tích lũy	Định bậc	Mô phỏng
Chuẩn	dnorm(x, mean, sd)	pnorm(q, mean, sd)	qnorm(p, mean, sd)	rnorm(n, mean, sd)
Nhị phân	dbinom(k, n, p)	pbinom(q, n, p)	qbinom(p, n, p)	rbinom(k, n, prob)
Poisson	dpois(k, lambda)	ppois(q, lambda)	qpois(p, lambda)	rpois(n, lambda)
Uniform	dunif(x, min, max)	punif(q, min, max)	qunif(p, min, max)	runif(n, min, max)
Negative binomial	dnbinom(x, k, p)	pnbinom(q, k, p)	qnbinom(p, k, prob)	rbinom(n, n, prob)
Beta	dbeta(x, shape1, shape2)	pbeta(q, shape1, shape2)	qbeta(p, shape1, shape2)	rbeta(n, shape1, shape2)
Gamma	dgamma(x, shape, rate, scale)	gamma(q, shape, rate, scale)	qgamma(p, shape, rate, scale)	rgamma(n, shape, rate, scale)
Geometric	dgeom(x, p)	pgeom(q, p)	qgeom(p, prob)	rgeom(n, prob)
Exponential	dexp(x, rate)	pexp(q, rate)	qexp(p, rate)	rexp(n, rate)
Weibull	dnorm(x, mean, sd)	pnorm(q, mean, sd)	qnorm(p, mean, sd)	rnorm(n, mean, sd)
Cauchy	dcauchy(x, location, scale)	pcauchy(q, location, scale)	qcauchy(p, location, scale)	rcauchy(n, location, scale)
F	df(x, df1, df2)	pf(q, df1, df2)	qf(p, df1, df2)	rf(n, df1, df2)
T	dt(x, df)	pt(q, df)	qt(p, df)	rt(n, df)
Chi-squared	dchisq(x, df)	pchi(q, df)	qchisq(p, df)	rchisq(n, df)

Chú thích: Trong bảng trên, df = degrees of freedom (bậc tự do); prob = probability (xác suất); n = sample size (số lượng mẫu). Các thông số khác có thể tham khảo thêm cho từng luật phân phối. Riêng các luật phân phối F, t, Chi-squared còn có một thông số khác nữa là non-centrality parameter (ncp) được cho số 0. Tuy nhiên người sử dụng có thể cho một thông số khác thích hợp, nếu cần.

3. Bài tập

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

- Tính xác suất $P(P \leq 0.2)$ bằng cách sử dụng hàm: 1) $f = \text{function}(p)\{.07*p**(-0.93)\}$, và 2) $\text{integrate}(f, \text{lower}=0, \text{upper}=.2)$.

- b. Kiểm tra $f_p(p)$ là hàm mật độ xác suất bằng cách tính diện tích dưới đường cong từ 0 đến 1.

Bài 2: Sử dụng lệnh `x=sample(1:5,100,TRUE,c(0.1,0.2,0.4,0.2,0.1))` để rút một mẫu ngẫu nhiên cỡ 100 từ phân phối với hàm xác suất

x	1	2	3	4	5
P(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(x)/100`

Bài 3: Vẽ một biểu đồ cột của hàm xác suất của phân phối siêu bội với $N = 100$, $M = 25$ và cỡ mẫu $n = 15$.

Bài 4: Nếu X có phân phối như trên, đầu tiên tính $P(5 \leq X \leq 12)$ bằng cách lấy tổng các xác suất được cho bởi hàm xác suất, và sau đó bằng cách sử dụng hàm phân phối tích lũy.

Bài 5:

- Sử dụng lệnh `curve(dexp(x,0.6),0,10)` để vẽ hàm mật độ xác suất của phân phối mũ với tham số $\lambda = 0.6$.
- Đối với đồ thị nhận được bạn vẽ thêm hàm mật độ xác suất của phân phối mũ với tham số $\lambda = 0.3$ (đảm bảo bạn thêm `add=T` trong lệnh `curve`).
- Sử dụng hàm phân phối tích lũy để tính diện tích bên dưới của hai hàm mật độ.

Bài 6: Vẽ hàm xác suất của biến $X \sim P(1)$ với $x \in \{0, \dots, 8\}$

Bài 7: Vẽ đồ thị hàm mật độ xác suất của biến $X \sim \chi^2(3)$ với $x \in [0, 10]$.

Bài 8: Chia cửa sổ đồ thị thành hai phần trên và dưới.

- Trong phần trên, vẽ đồ thị của hàm xác suất của biến $X \sim B(50, 0.08)$ lấy `ylim=c(0,0.25)`.
- Trong phần dưới, vẽ đồ thị của hàm xác suất của biến $X \sim P(4)$ với $x \in \{0, \dots, 50\}$ với cùng lựa chọn `ylim=c(0,0.25)`.

(Điều này minh họa kết quả là khi n đủ lớn và np đủ nhỏ ta có thể xấp xỉ phân phối nhị thức $B(n, p)$ bằng luật Poisson $P(np)$)

Bài 9: Vẽ đồ thị của hàm mật độ của biến $X \sim B(50, 0.4)$ và thêm vào đồ thị này hàm mật độ của biến $Y \sim N(20, 12)$

(Điều này minh họa kết quả rằng khi n lớn, np lớn và $n(1-p)$ lớn, ta có thể xấp xỉ phân phối nhị thức $B(n, p)$ bằng phân phối chuẩn $N(np, np(1-p))$)