Tuần 5: LT Bài 4.pdf

- Điểm danh trên gg classroom của lớp.
- Mở R, thiết lập thư mục làm việc cho R, past code TH 3.
- Mở file "Bai 4.pdf": đọc và thực hành các ví dụ.

Thực hành xác suất thống kê

Bài 4: MỘT SỐ PHÂN PHỐI THÔNG DỤNG

1. Cú pháp chung

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

- ✓ Hàm xác suất/ Hàm mật độ $f(x)$: **`dluật`**
- ✓ Hàm phân phối xác suất $F_X(x)$ của X : **`pluật`**
- ✓ Phân vị của X : **`qluật`**
- ✓ Mô phỏng giá trị của X: **`rluật`**

Tên của các phân phối phổ biến là: **`norm`** (cho phân phối chuẩn), **`binom`** (cho nhị thức), **`geom`** (cho phân phối hình học), **`pois`** (cho phân phối Poisson), **`t`** (cho phân phối Student), **`chisq`** (cho phân phối Chi bình phương), **`exp`** (cho phân phối mũ), **`f`** (cho phân phối Fisher),…

1. Cú pháp chung

Hàm phân phối	Mật độ	Tích lũy	Định bậc	Mô phỏng
Chuẩn	dnorm(x,mean,sd)	pnorm(q,mean,sd)	qnrom(p,mean,sd)	rnorm(n,mean,sd)
Nhị thức	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)
Nhị thức âm	dnbinom(x,k,p)	pnbinom(q,k,p)	qnbinom(p,k,prob)	rbinom(n,n,prob)
Beta	dbeta(x,sh1,sh2)	pbeta(q,sh1,sh2)	qbeta(p,sh1,sh2)	rbeta(n,sh1,sh2)
Gamma	dgamma(x,sh,r,s)	pgamma(q,sh,r,s)	qgamma(p,sh,r,s)	rgamma(n,sh,r,s)
Geometric	dgeom(x,p)	pgeom(q,p)	qgeom(p,prob)	rgeom(n,prob)
Hypergeometric	dhyper(x, m, n, k)	phyper(q, m, n, k)	qhyper(p, m, n, k)	rhyper(nn, m, n, k)
Exponential	dexp(x,rate)	pexp(q,rate)	qexp(p,rate)	rexp(n,rate)
Weibull	dweibull(x, shape, scale = 1)	pweibull(q, shape, scale = 1)	qweibull(p, shape, scale = 1)	rweibull(n, shape, scale = 1)
Cauchy	dcauchy(x, location, scale)	pcauchy(q,location, scale)	qcauchy(p, location, scale)	rcauchy(n,location, scale)
Fisher	df(x, df1, df2)	pdf(q, df1, df2)	qf(p, df1, df2)	rf(n, df1, df2)
Student	dt(x, df)	pt(q, df)	qt(p, df)	rt(n, df)
Chi-quared	dchisq(x,df)	pchisq(q,df)	qchisq(p,df)	rchisq(n,df)

Chú thích: Trong bảng trên, df = degrees of freedome (bậc tự do); prob = probability (xác suất); n = sample size (số lượng mẫu), s = scale (hệ số tỉ lệ), sh = shape (hệ số định dạng). 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.

2. Tính toán

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

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

```
# Tao vec to k có gia tri tu 0 den 8
k = 0:8
# Viet ham xac suat
p <- function(k) choose(8,k) * 0.3^k * 0.7^(8-k)
# Tinh gia tri xac suat P(X = k), k = 0,1,...,8
p(k)
```

```
# Dung lenh co san: X ~ B(n=8,p=0.3)
dbinom(0:8,8,0.3)
```

2. Tính toán

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

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

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))
}

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

```
# Dung lenh co san: X ~ N(mu=0,sigma^2=1)
dnorm(0,0,1)
```

Lệnh chia cửa sổ đồ thị

Khi ta muốn chia cửa sổ đồ thị thành một ma trận các cửa sổ thành n dòng và m cột, ta sử dụng

- ✓ `par(mfrow = c(n,m))`: thứ tự điền đồ thị theo dòng, từ trái qua phải và từ trên xuống.
- ✓ `par(mfcol = c(n,m))`: thứ tự điền đồ thị theo cột, từ trên xuống và từ trái qua phải.

Ví dụ:

`par(mfrow = c(2,2))`

1	2
3	4

`par(mfcol = c(2,2))`

1	3
2	4

2. Biểu diễn bằng đồ thị

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

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

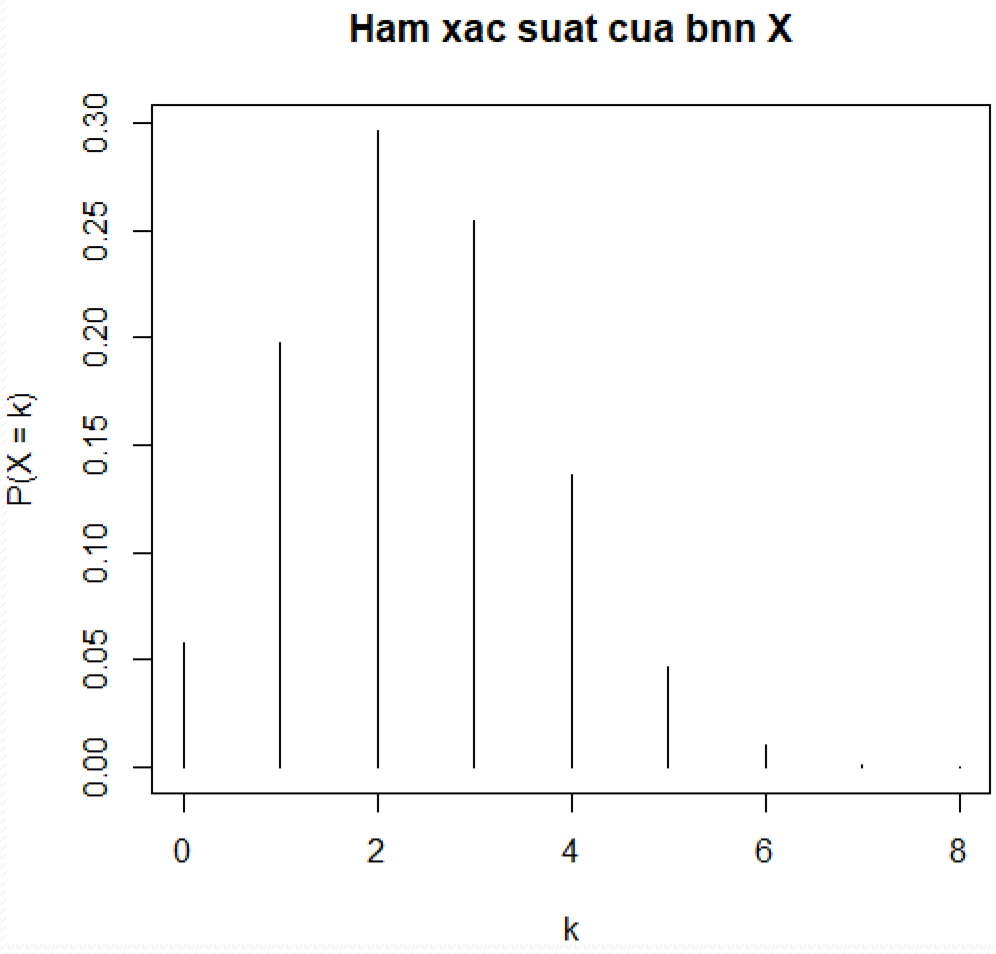
```
plot(x, f(x), type = "loại đồ thị", xlab = "nhãn trục x", ylab =
"nhãn trục y", main = "tên đồ thị")
```

```
# Ve ham xac suat o vi du 1a
plot(k, p(k), type = "h", xlab = "k", ylab = "P(X = k)", main =
"Ham xac suat cua bnn X")
```

```
# Dung lenh co san: X ~ B(n=8,p=0.3)
plot(0:8, dbinom(0:8,8, 0.3), type = "h", xlab = "k", ylab = "P(X =
k)", main = "Ham xac suat cua bnn 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)

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

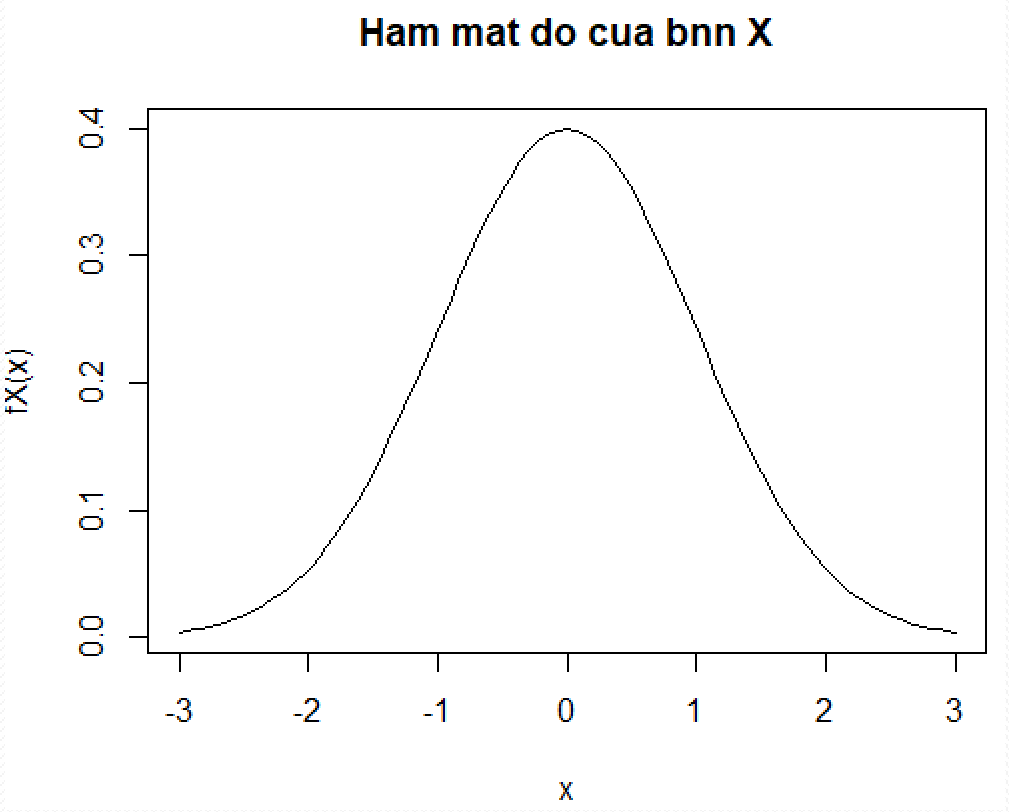
```
curve(f(x), from = cận dưới, to = cận trên, xlab = "nhãn trục x", ylab = "nhãn trục y", main = "tên đồ thị")
```

```
# Ve ham mat do xac suat o vi du 1b
curve(f(x,0,1),from=-3,to=3, xlab = "x", ylab = "fX(x)", main = "Ham mat do cua bnn X")
```

```
# Dung lenh co san: X ~ N(mu=0,sigma^2=1)
curve(dnorm(x,0,1),from=-3,to=3, xlab = "x", ylab = "fX(x)", main = "Ham mat do cua bnn 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)
```

```
# Dung lenh co san: X ~ B(n=8,p=0.3)
pbinom(4,8,0.3)

pbinom(0:8,8,0.3)
```

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)
```

```
# Dung lenh co san: X ~ N(mu=0,sigma^2=1)
pnorm(1.96,0,1)
```

3. Hàm phân phối
3.3 Vẽ đồ thị

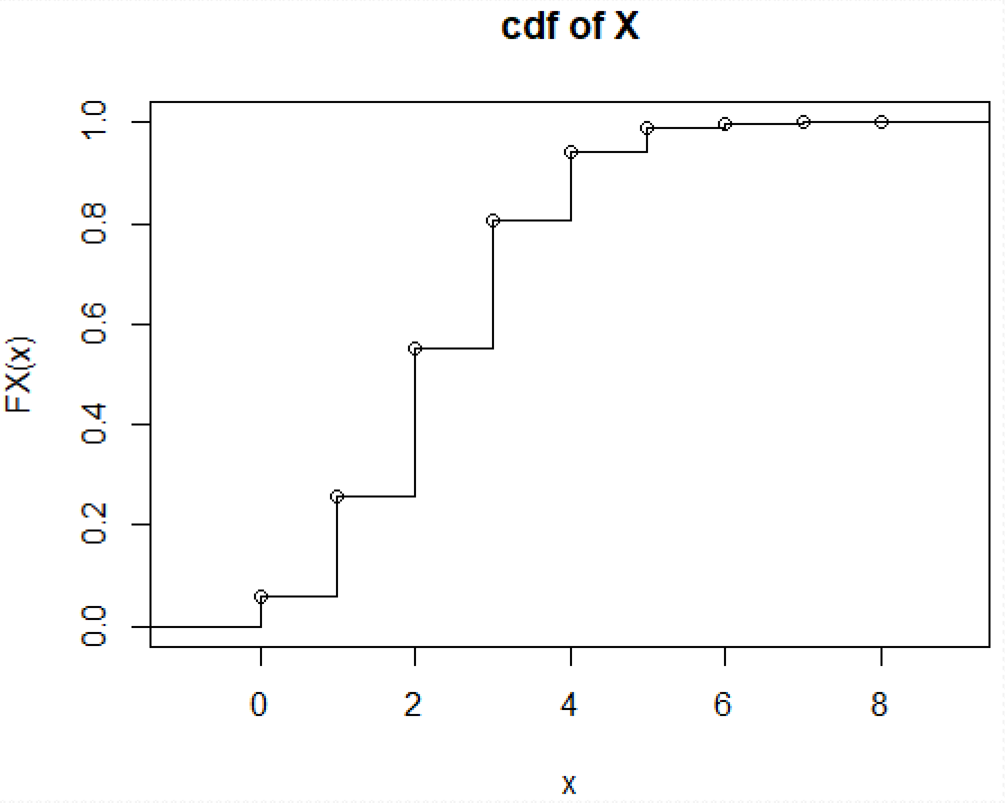
Ví dụ 4 a: Vẽ hàm phân phối xác suất của X ở ví dụ 1 a.

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

```
# Dung lenh co san: X ~ B(n=8,p=0.3)
plot(stepfun(0:8, c(0, pbinom(0:8,8,0.3))), 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

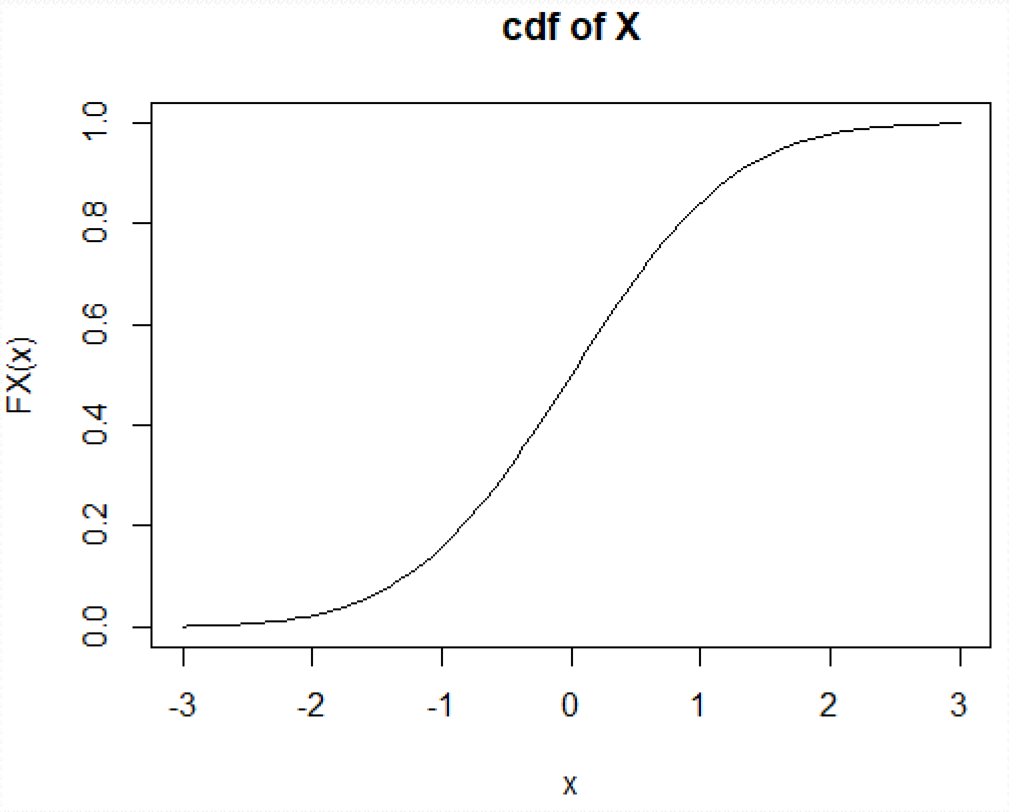
Ví dụ 4 b: Vẽ hàm phân phối xác suất của *X* ở ví dụ 1 b.

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

```
# Dung lenh co san: X ~ N(mu=0,sigma^2=1)
curve(pnorm(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
# kiem tra lai
F(0)
[1] 0.05764801
F(1)
[1] 0.2552983
```

```
# Dung lenh co san: X ~ B(n=8,p=0.3)
qbinom(0.25,8,0.3)
```

4. Phân vị
4.2 Tính toán

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

```
# Tim Nghiem cua Phuong trinh F(x) – xp = 0
uniroot(function(x) F2(x)-0.975, c(-3,3))$root
[1] 1.959992
# kiem tra lai
F2(1.96)
[1] 0.9750021
```

```
# Dung lenh co san: X ~ N(mu=0,sigma^2=1)
qnorm(0.975,0,1)
```

Bài tập

Bài 1: Vẽ một biểu đồ cột của hàm xác suất của phân phối siêu bội với $N = 100$, $M = 25$ và cỡ mẫu $n = 15$.

Lưu ý: Nếu $X \sim H(N, M, n^*)$ thì

dhyper(x, m = M, n = N - M, k = n^*)

```
# Bai tap 1

# Nhap gia tri cho cac tham so:
x <- 0:15
m = 25; n = 100-25; k = 15
dhyper(x,m,n,k)

# Ve bieu do cot ham xac suat cua X ~ H(100,25,15):

barplot(dhyper(x,m,n,k))
```

Bài tập

Bài 2: 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.

Cách 1:

$$P(5 \leq X \leq 12) = P(X = 5) + P(X = 6) + \ldots + P(X = 12)$$

```
# Bai tap 2

# Cach 1:Tinh bang tong xac suat: P(5<=X<=12) = P(X=5)+...P(X=12)
# = f(5)+...+f(12)

sum(dhyper(5:12,m,n,k))
```

Bài tập

Bài 2: 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.

Cách 2: Nhắc lại: Hàm phân phối tích lũy $F_X(x) = P(X \leq x)$.

Ta có:

$P(X \leq 12) = P(X = 0) + \ldots + P(X = 4) + P(X = 5) + \ldots + P(X = 12)$

$P(X \leq 4) = P(X = 0) + \ldots + P(X = 4)$

$\Rightarrow P(5 \leq X \leq 12) = P(X \leq 12) - P(X \leq 4)$

$= F_X(12) - F_X(4)$

```
# Bai tap 2

# Cach 2:Tinh bang FX:  P(5<=X<=12) = Fx(12) - Fx(5-1)
phyper(12,m,n,k) - phyper(4,m,n,k)
```

Bài tập

Bài 3:

a) Sử dụng lệnh **`curve`**`(dexp(x,0.6),0,10)` để vẽ hàm mật độ xác suất của phân phối mũ với tham số $\lambda = 0.6$.

b) 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`**).

c) Sử dụng hàm phân phối tích lũy để tính diện tích bên dưới của hai hàm mật độ.

```
# Bai tap 3
# Cau a: Ve ham mat do cua X ~ exp(lambda = 0.6):
curve(dexp(x,0.6),from = 0, to = 10, col = "blue")

# Cau b: Ve ham mat do cua X ~ exp(lambda = 0.3):
curve(dexp(x,0.3),from = 0, to = 10, col = "red", add = T)
```

Bài tập

Bài 3:

c) Sử dụng hàm phân phối tích lũy để tính diện tích bên dưới của hai hàm mật độ.

```
# Cau c:
# Tim nghiem cua phuong trinh: dexp(x, 0.6) - dexp(x, 0.3) = 0
x0 <- uniroot(function(x) dexp(x,0.6) - dexp(x,0.3), c(0,4))$root
x0

# Tinh dien tich ben duoi hai ham mat do:
I1 = pexp(x0,0.3)
I2 = pexp(10,0.6) - pexp(x0,0.6)
I = I1 + I2; I

# Kiem tra lai bang cach tinh tich phan:

S1 = integrate(function(x) dexp(x,0.3), lower = 0, upper = x0)$value
S2 = integrate(function(x) dexp(x,0.6), lower = x0, upper = 10)$value
S = S1 + S2; S
```

Bài tập

Bài 4: Vẽ hàm xác suất của biến $X \sim P(1)$ với $x \in \{0, \dots, 8\}$.

```
# Bai tap 4
# Nhap gia tri cho x
x <- 0:8
# Ve ham xac suat cua bien ngau nhien X ~ P(1)
plot(x,dpois(x,1),type = "h", main = "pdf of X ~ P(1)")
```

Bài tập

Bài 5: Vẽ đồ thị hàm mật độ xác suất của biến $X \sim \chi^2(3)$ với $x \in [0; 10]$.

```
# Bai tap 5
# Ve ham mat do cua bien ngau nhien X ~ X^2(3)
curve(dchisq(x,3),from = 0, to = 10, main = "pdf of X ~ X^2(3)")
```

Bài tập

Bài 6: 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 $x \in \{0, \ldots, 50\}$ 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 $P(np)$).

```
# Bai tap 6
# Chia cua so do thi thanh 2 dong, 1 cot
par(mfrow = c(2,1))

#Ve do ham xac suat cua X ~ B(50,0.08)
# Nhap gia tri cho x
x <- 0:50

plot(x,dbinom(x,50,0.08),type = "h", ylim = c(0,0.25), main = "pdf of X ~
B(50, 0.08)")

# Ve ham xac suat cua X ~ P(4)
plot(x,dpois(x,4),type = "h",ylim = c(0,0.25), main = "pdf of X ~ P(4)")
```

Bài tập

Bài 7: 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 $np(1-p)$ lớn, ta có thể xấp xỉ phân phối nhị thức $B(n,p)$ bằng phân phối chuẩn $N(np, np(1-p))$).

```
# Bai tap 7
# Ve do ham xac suat cua X ~ B(50,0.4)
# Nhap gia tri cho x
x <- 0:50
plot(x,dbinom(x,50,0.4),type = "h", col = "blue", main = "pdf of X
~ B(50, 0.4)")

curve(dnorm(x,20,sqrt(12)),from = 0, to = 50, col = "red", main =
"pdf of X ~ N(20;12)",add = T)
```