Bài 4.

Một số phân phối xác suất 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:

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

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),…

Phần bên dưới sẽ trình bày chi tiết, ví dụ và minh họa cho các lệnh này.

2. Mật độ

2.1 Định nghĩa

* Với biến ngẫu nhiên X rời rạc, “mật độ” của X tại x là xác suất $P(X = x)$.
* Với biến ngẫu nhiên X có hàm mật độ f_X, “mật độ” của X tại x là $f_X(x)$.

2.2 Lệnh

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)
```

Dưới đây là một số ví dụ:

Luật	Nhị thức	Hình học	Poisson
Tham số	$n \in \mathbb{N}^*, p \in (0,1)$	$p \in (0,1)$	$\lambda > 0$
$X \sim$	$B(n, p)$	$G(p)$	$P(\lambda)$
$X(\Omega)$	$\{0,\ldots,n\}$	$\mathbb{N}^*$	$\mathbb{N}$
$P(X = x)$	$\binom{n}{x} p^x (1-p)^{n-x}$	$p(1-p)^{x-1}$	$e^{-\lambda} \frac{\lambda^k}{k!}$
Lệnh	`dninom(x, n, p)`	`dgeom(x, p)`	`dpois(x, lambda)`

Luật	Đều	Mũ	Chuẩn
Tham số	$(a,b) \in \mathbb{R}^2, a < b$	$\lambda > 0$	$\mu \in \mathbb{R}, \sigma > 0$
$X \sim$	$U([a,b])$	$E(\lambda)$	$N(\mu, \sigma^2)$
$X(\Omega)$	$[a,b]$	$[0,\infty)$	$\mathbb{R}$
$f_X(x)$	$\frac{1}{b-a}$	$\lambda e^{-\lambda x}$	$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
Lệnh	`dunif(x, a, b)`	`dexp(x, lambda)`	`dnorm(x, mu, sigma)`

Tìm hiểu thêm: `help("dgamma")`, `help("dt")`, `help("dchisq")` và `help("df")`

2.3. Tính toán

Ví dụ 1

a) Cho biến ngẫu nhiên $X \sim B(8, 0.3)$. Tính $P(X = 4) = \binom{8}{4} 0.3^4 (1-0.7)^{8-4}$.

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

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

<u>Giải:</u>

a)

Ta làm:

```
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
```

b)

```
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
```

Để tính toán mật độ tại nhiều giá trị, ta nhập `x` là vector có các giá trị này. Ta cũng có thể làm tương tự cho một tập các tham số.

Ví dụ 2

(a) Tính giá trị mật độ của $X \sim B(8, 0.3)$ tại $x \in \{4, 6\}$.

(b) Tính giá trị mật độ của $X \sim E(\lambda)$ tại $x = 2$ và $\lambda = 1, \lambda = 2$ và $\lambda = 3$.

<u>Giải :</u>

(a)

Ta làm:

```
dbinom(c(4, 6), 8, 0.3)
[1] 0.13613670  0.01000188
```

(b) Ta làm :

```
dexp(2, c(1, 2, 3))
[1] 0.135335283 0.036631278 0.007436257
```

Ta cũng có thể lưu kết quả vào một vector để sử dụng sau này.

Ta làm :

```
vec = dexp(2, c(1, 2, 3))
vec
[1] 0.135335283 0.036631278 0.007436257
```

2.4. Biểu diễn bằng đồ thị

Ta có thể biểu diễn đồ thị của hàm xác suất của biến rời rạc X với lệnh `plot` và tùy chọn `type=h`.

Ví dụ 3

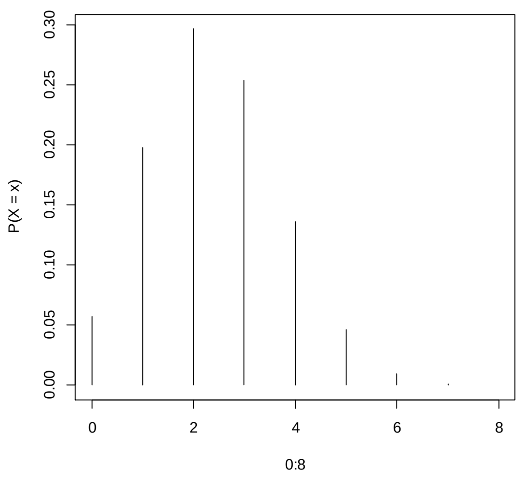
a) Vẽ đồ thị hàm xác suất trong Ví dụ 1a): $B(8, 0.3)$.

b) Vẽ hàm mật độ xác suất trong Ví dụ 1b): $N(2, 0.12^2)$.

Giải:

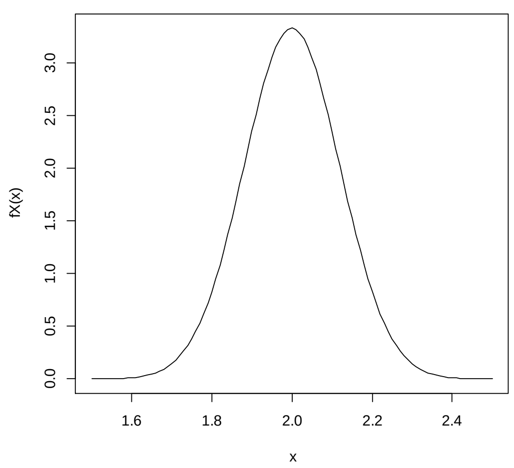
a)

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



b)

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



3. Hàm phân phối

3.1 Định nghĩa

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

* X rời rạc,

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

* X liên tục,

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt$$

3.2 Lệnh

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)
```

3.3 Tính toán

Ví dụ 4

a) Xét tiếp Ví dụ 1a). Tính $F_X(4) = P(X \le 4) = \sum_{k=0}^{4} P(X = k)$.

b) Xét tiếp Ví dụ 1b). Tính $F_X(2.1) = P(X \le 2.1) = \int_{-\infty}^{2.1} \frac{1}{\sqrt{2\pi \times 0.12^2}} e^{-\frac{(x-2)^2}{2 \times 0.12^2}} dx$.

c) Tính $P(X > 2)$, $X \sim E(3)$.

<u>Giải:</u>

a)

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

Ta có thể kiểm tra lại:

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

b)

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

c)

```
pexp(2, 3, lower.tail = FALSE)
[1] 0.002478752
```

Kiểm tra lại bằng cách tính toán:

$$P(X > 2) = \int_{2}^{\infty} f_X(x)dx = \int_{2}^{\infty} 3e^{-3x}dx = \left[-e^{-3x}\right]_2^{\infty} = e^{-3 \times 2} = e^{-6}.$$

Và tính

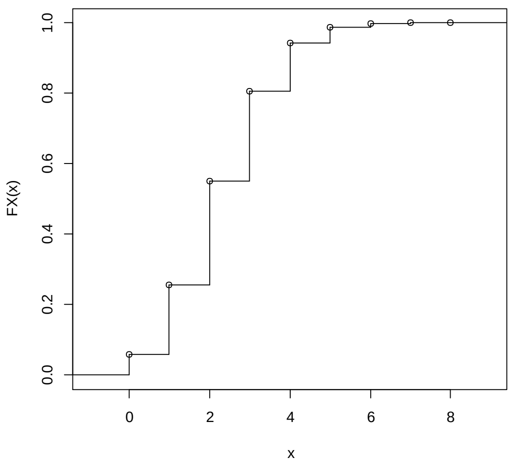
```
exp(-6)
[1] 0.002478752
```

3.4 Vẽ đồ thị

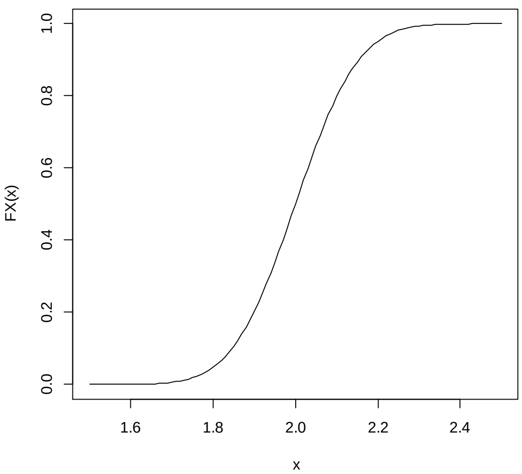
Để biểu diễn hàm phân phối của biến ngẫu nhiên rời rạc dùng lệnh `stepfun`, biến ngẫu nhiên liên tục dùng hàm `curve`.

Ví dụ 5

a) Xét tiếp Ví dụ 1a). Vẽ hàm phân phối xác suất của X .
b) Xét tiếp Ví dụ 1b). Vẽ hàm phân phối.
<u>Giải:</u>
a)
```
plot(stepfun(0:8, c(0, pbinom(0:8, 8, 0.3))), ylab = "FX(x)", main = "")
```



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



4. Phân vị

4.1 Định nghĩa

Cho $p \in (0,1)$ và X là một biến ngẫu nhiên.
* Nếu X rời rạc, thì phân vị mức p của X , ký hiệu là x_p, là

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

* Nếu X liên tục, thì phân vị mức p của X, ký hiệu là x_p, là giá trị thỏa $F_X(x_p) = p$.

4.2 Lệnh

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)
```

4.2 Tính toán

Ví dụ 6

a) Xét tiếp Ví dụ 1a). Tính phân vị mức 0.25 của X.
b) Xét tiếp Ví dụ 1b). Tính phân vị mức 0.975 của X.
Giải:
a)

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

$x_{0.25} = \inf\{k \in \mathbb{Z}; F_X(k) \geq 0.25\}$.

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

```
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
```

Khi đó ta thấy rằng $F_X(0) = 0.0576 < 0.25 \leq 0.2552 = F_X(1)$, do đó $x_{0.25} = 1$.

b)

```
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
```

5. Mô phỏng các phân phối được lập trình sẵn

5.1 Lệnh

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)
```

5.2 Tính toán

Ví dụ 7

a) Mô phỏng một mẫu ngẫu nhiên cỡ 10 của phân phối Poisson $P(2)$.

b) Mô phỏng giá trị ngẫu nhiên của biến $\sum_{i=1}^{80} X_i$, $X_1, \ldots, X_{80}$ là các biến ngẫu nhiên độc lập nhau có phân phối Bernoulli $B(1, 0.02)$.

c) Mô phỏng một mẫu ngẫu nhiên cỡ 15 của phân phối $N(2, 0.12^2)$.

Giải:
a)

```
rpois(10, 2)
[1] 0 2 2 3 3 5 1 3 1 2
```

b)

```
sum(rbinom(80, 1, 0.02))
[1] 3
```

Bởi vì $\sum_{i=1}^{80} X_i \sim B(80, 0.02)$, nên ta có thể dùng lệnh sau để mô phỏng biến $Y \sim B(80, 0.02)$:

`rbinom(1, 80, 0.02)`.

c)

```
x = rnorm(15, 2, 0.12)
x
[1] 2.021767 1.898343 2.072553 1.974686 2.044200 2.050094
[7] 2.063342 2.000698 1.702115 1.958284 1.984903 1.897726
[13] 2.011624 2.074341 1.720218
```

6. 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.

Bài 2

Nếu X có phân phối như trên, đầu tiên tính P(5 ≤ X ≤ 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 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 λ = 0.6.
b) Đối với đồ thị nhận được bạn vẽ thêm hàm mật độ xác suất của phân phối mũ với tham số λ = 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.

Bài 4

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

Bài 5

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

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 thực $B(n,p)$ bằng luật Poisson $P(np)$).

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