ĐẠI HỌC THĂNG LONG
THANG LONG UNIVERSITY

THỐNG KÊ ỨNG DỤNG TRONG KINH TẾ XÃ HỘI

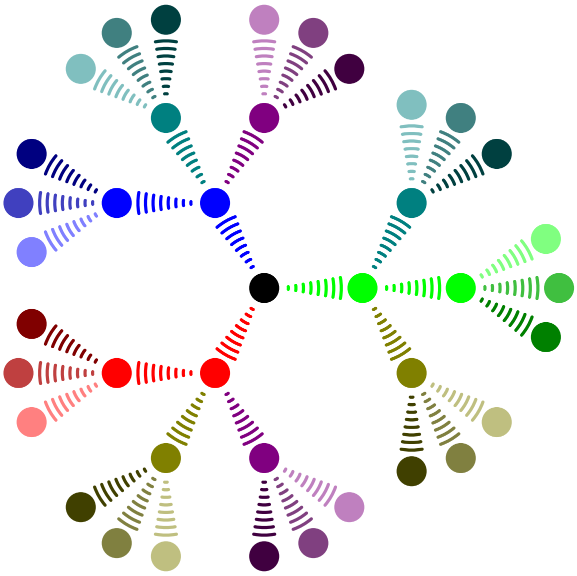
Hướng dẫn thực hành trên phần mềm R

Phan Thanh Hồng - Nguyễn Thị Nhung

NXB Thống kê, 2015

CHƯƠNG 6

BIẾN NGẪU NHIÊN



Yêu cầu đối với sinh viên

- Biết cách áp dụng những hàm trong R để tính xác suất, kì vọng và phương sai của biến ngẫu nhiên rời rạc và liên tục.
- Biết tính xác suất của những phân phối rời rạc phổ biến.
- Biết tính xác suất của những phân phối liên tục phổ biến.
- Biết tính những giá trị tới hạn phổ biến trong R.

6.1 Biến ngẫu nhiên rời rạc

Một biến ngẫu nhiên được gọi là rời rạc nếu nó chỉ nhận hữu hạn hoặc đếm được giá trị. Giả sử X là một biến ngẫu nhiên rời rạc nhận hữu hạn giá trị là $x_1, x_2, \ldots, x_n$ với xác suất tương ứng là $p_1, p_2, \ldots, p_n$. Phân phối của X được mô tả qua bảng sau:

X	x_1	x_2	$\ldots$	x_n
P_X	p_1	p_2	$\ldots$	p_n

trong đó $p_i \geqslant 0, i = 1, \ldots, n$ và $\sum_{i=1}^{n} p_i = 1$.

Hai đặc trưng cơ bản của X là kì vọng EX và phương sai VX được cho bởi biểu thức:

$$\mu = EX = \sum_{i=1}^{n} x_i p_i,$$

$$\sigma^2 = VX = E(X - EX)^2 = \sum_{i=1}^{n} (x_i - \mu)^2 p_i.$$

Độ lệch chuẩn của X là σ được tính bởi $\sigma = \sqrt{VX}$.

Chú ý: Phương sai của X cũng có thể được tính qua công thức tương đương dưới đây:

$$\sigma^2 = EX^2 - (EX)^2 = \sum_{i=1}^{n} x_i^2 p_i - \mu^2.$$

Trong R, tính xác suất cũng như các đặc trưng quan trọng của biến ngẫu nhiên có thể được thực hiện bằng những hàm tính toán thông thường hoặc một số hàm riêng trong những gói hỗ trợ thêm như gói `distrEX`. Những ví dụ sau sẽ minh họa cho ta một số cách tính toán có thể sử dụng.

Ví dụ 6.1.1 Trong một rổ có 99 quả bóng đánh số từ 1 đến 99. Nhặt ngẫu nhiên từ rổ ra 5 quả bóng và gọi X số nhỏ nhất hiện lên trên 5 quả bóng. Hãy tính kì vọng, phương sai và độ lệch chuẩn của X.

Lời giải: Ta có X là biến ngẫu nhiên rời rạc nhận các giá trị từ 1 đến 95 với xác suất tương ứng được tính như sau:

$$P(X = k) = \frac{C_{99-k}^4}{C_{99}^5}, k = 1, \ldots, 95.$$

Phân phối xác suất của X được cho qua bảng phân phối xác suất sau:

X	1	2	$\ldots$	k	$\ldots$	95
P_X	$\frac{C_{98}^4}{C_{99}^5}$	$\frac{C_{97}^4}{C_{99}^5}$	$\ldots$	$\frac{C_{99-k}^4}{C_{99}^5}$	$\ldots$	$\frac{C_4^4}{C_{99}^5}$

Kì vọng, phương sai và độ lệch chuẩn của X có thể được tính như sau:

```
> k = 1:95
> p = choose(99-k,4)/choose(99,5)
> EX = sum(k*p)
> EX
```

```
[1] 16.66667
> VX = sum(k^2*p) - EX^2
> VX
[1] 186.5079
> sigma = sqrt(VX)
> sigma
[1] 13.65679
```

Trong gói `distriEx` những đặc trưng của X có thể tính toán nhanh gọn hơn như sau:

```
> k = 1:95
> p = choose(99-k,4)/choose(99,5)
> X = DiscreteDistribution(supp = k, prob = p)
> E(X);var(X);sd(X)
[1] 16.66667
[1] 186.5079
[1] 13.65679
```

6.2 Một số phân phối xác suất rời rạc thường gặp

Trong các phân phối rời rạc có một số phân phối rời rạc thường gặp như: phân phối nhị thức, phân phối Poisson, phân phối hình học, phân phối siêu hình học, phân phối nhị thức âm,...
Mỗi phân phối có 4 loại hàm quan trọng mà ta cần biết:

- Hàm mật độ xác suất (probability density distribution);
- Hàm phân phối hay hàm xác suất tích lũy (cumulative probability distribution);
- Hàm phân vị (quantile);
- Hàm mô phỏng (simulation).

R cung cấp nhiều hàm tính toán xác suất liên quan đến những phân phối trên. Tên mỗi hàm thường bắt đầu bởi một tiếp đầu ngữ để chỉ loại hàm và tên phân phối. Tiếp đầu ngữ d (distribution) chỉ xác suất, tiếp đầu ngữ p (cumulative probability) chỉ xác suất tích lũy, tiếp đầu ngữ q (quantile) chỉ phân vị và tiếp đầu ngữ r (random) chỉ mô phỏng mẫu ngẫu nhiên. Tên các phân phối được viết tắt bởi các cụm từ tiếng anh. Phân phối nhị thức được viết tắt là `binom (binomial)`, phân phối Poisson được viết tắt là `pois (poisson)`, phân phối hình học được viết tắt là `geom (geometric)`, phân phối nhị thức âm được viết tắt là `nbinom (negative binomial)`,...
Những hàm này có thể sử dụng ngay sau khi cài R qua gói `stats` (mặc định), ngoài ra, khi cài thêm gói `distr` ta có thêm một số cách tính toán khác liên quan đến những phân phối thường gặp này.
Dưới đây, ta sẽ liệt kê những hàm liên quan đến các phân phối trên trong gói `stats` và gói `distr`. Tuy nhiên, các ví dụ được tính toán chủ yếu là trên gói `stats`, cách làm với gói `distr` có thể được tiến hành tương tự.

6.2.1 Phân phối nhị thức

Cho $X \sim B(n, p)$ là biến ngẫu nhiên tuân theo phân phối nhị thức với số lần thử nghiệm độc lập là n và xác suất thành công trong từng lần thử nghiệm là p. Xác suất có x lần thành công trong n lần thử nghiệm được tính bởi công thức sau:

$$P(X = x) = C_n^x p^x (1-p)^{n-x}, x = 0, 1, \ldots, n.$$

Bảng sau cho ta những hàm tính toán liên quan đến phân phối nhị thức trong R:

Tính	Với gói `stats`	Với gói `distr`
$P(X = x)$	`dbinom(x, size = n, prob = p)`	`d(X)(x)`
$P(X \leqslant x)$	`pbinom(x, size = n, prob = p)`	`p(X)(x)`
Tạo mẫu cỡ k	`rbinom(k, size = n, prob = p)`	`r(X)(k)`

Khi sử dụng gói `distr`, ta phải đặt `X = Binom(size = n, prob = p).`
Ta xét ví dụ minh họa sau.

Ví dụ 6.2.1 Một đề thi trắc nghiệm gồm 50 câu hỏi, mỗi câu hỏi có 4 phương án trả lời, trong đó chỉ có một phương án đúng. Một thí sinh không học bài đi thi trả lời một cách ngẫu nhiên.

a. Tính xác suất để thí sinh này trả lời đúng 10 câu hỏi.

b. Một thí sinh sẽ thi đỗ nếu trả lời đúng ít nhất một nửa số câu hỏi. Tính xác suất để thí sinh này sẽ thi đỗ.

c. Tính số câu trả lời đúng trung bình của thí sinh này.

d. Tính số câu mà thí sinh này có khả năng trả lời đúng nhiều nhất.

Lời giải: Gọi X là số câu trả lời đúng của thí sinh này. Khi đó, X tuân theo phân phối nhị thức với $n = 50$ và $p = 0.25$.
Các yêu cầu của bài toán có thể được thực hiện như sau:

a. Xác suất để thí sinh này trả lời đúng 10 câu hỏi là: $P(X = 10) = 0.0985$.

```
> dbinom(10,size=50,prob=0.25)
[1] 0.09851841
```

b. Xác suất để thí sinh này thi đỗ là: $P(X \geqslant 25) = 1 – P(X \leqslant 24) = 0.0001225$.

```
> 1 - pbinom(24,50,0.25)
[1] 0.0001225135
> #Hoặc
> pbinom(24,50,0.25, lower.tail=F)
[1] 0.0001225135
```

c. Số câu trả lời đúng trung bình là: $EX = np = 50 * 0.25 = 12.5$.

```
> x = 0:50
> p = dbinom(x,50,0.25)
> sum(x*p)
[1] 12.5
```

d. Số câu x_0 mà thí sinh này có khả năng trả lời đúng nhiều nhất ứng với xác suất $P(X = x_0)$ là cao nhất.

```
> x = 0:50
> p = dbinom(x,50,0.25)
> which(p==max(p))
[1] 13
> x[13]
[1] 12
```

Số câu có khả năng trả lời đúng nhiều nhất là 12.

6.2.2 Phân phối Poisson

Cho $X \sim P(\lambda)$ là biến ngẫu nhiên tuân theo phân phối Poisson với tham số λ. Xác suất để X nhận giá trị x được cho bởi:

$$P(X = x) = e^{-\lambda}\frac{\lambda^x}{x!}, x = 0, 1, 2, ...$$

Bảng sau cho ta những hàm tính toán liên quan đến phân phối poisson trong R:

Tính	Với gói `stats`	Với gói `distr`
$P(X = x)$	`dpois(x, lambda =` λ`)`	`d(X)(x)`
$P(X \leqslant x)$	`ppois(x, lambda =` λ`)`	`p(X)(x)`
Tạo mẫu cỡ k	`rpois(k, lambda =` λ`)`	`r(X)(k)`

Khi sử dụng gói `distr`, ta phải đặt $X = Pois(lambda = \lambda)$.
Ta xét ví dụ minh họa sau.

Ví dụ 6.2.2 Số vụ động đất trong một năm ở một khu vực tuân theo phân phối Poisson với trung bình 5 trận một năm.

a. Tính xác suất để năm tới ở khu vực này có 3 vụ động đất.

b. Tính xác suất để năm tới khu vực này có không quá 5 vụ động đất.

Lời giải: Gọi X là số vụ động đất xảy ra trong năm tới ở khu vực này. Khi đó, X tuân theo phân phối Poisson với $\lambda = 5$.
Các xác suất trong bài có thể được tính như sau:

a. Xác suất để năm tới khu vực này có 3 vụ động đất là: $P(X = 3) = 0.14$.

```
> dpois(3,lambda=5)
[1] 0.1403739
```

b. Xác suất để năm tới khu vực này có ít hơn 5 vụ động đất là: $P(X \leqslant 5) = 0.62$.

```
> ppois(5,lambda=5)
[1] 0.6159607
```

6.2.3 Phân phối hình học

Cho $X \sim Geom(p)$ là biến ngẫu nhiên tuân theo phân phối hình học với tham số $p(0 \leqslant p \leqslant 1)$. Xác suất để X nhận giá trị x được cho bởi:

$$P(X = x) = p(1-p)^{x-1}, x = 0, 1, 2, ...$$

Phân phối xác suất của phân phối hình học là phân phối của "số lần thử cho đến khi thành công". Bảng sau cho ta những hàm tính toán liên quan đến phân phối hình học trong R:

Tính	Với gói `stats`	Với gói `distr`
$P(X = x)$	`dgeom(x, prob = p)`	`d(X)(x)`
$P(X \leqslant x)$	`pgeom(x, prob = p)`	`p(X)(x)`
Tạo mẫu cỡ k	`rgeom(k, prob = p)`	`r(X)(k)`

Khi sử dụng gói `distr`, ta phải đặt `X = Geom(prob = p).`

Phân phối nhị thức âm

Cho $X \sim NB(p)$ là biến ngẫu nhiên tuân theo phân phối nhị thức âm với tham số $p(0 \leqslant p \leqslant 1)$ và $r(r > 0)$. Xác suất để X nhận giá trị x được cho bởi:

$$P(X = x) = C_{x+r-1}^{x} p^r (1-p)^x, x = 1, 2, ...$$

Bảng sau cho ta những hàm tính toán liên quan đến phân phối nhị thức âm trong R:

Tính	Với gói `stats`	Với gói `distr`
$P(X = x)$	`dnbinom(x,size = r, prob = p)`	`d(X)(x)`
$P(X \leqslant x)$	`pnbinom(x,size = r, prob = p)`	`p(X)(x)`
Tạo mẫu cỡ k	`rnbinom(k,size = r, prob = p)`	`r(X)(k)`

Khi sử dụng gói `distr`, ta phải đặt `X = Nbinom(size = r, prob = p)`.

6.3 Biến ngẫu nhiên liên tục

Cho X là biến ngẫu nhiên liên tục có hàm mật độ xác suất là $f(x)$. Khi đó, xác suất $P(a \leqslant X \leqslant b)$ được tính thông qua hàm mật độ bằng biểu thức:

$$P(a \leqslant X \leqslant b) = \int_a^b f(x)dx.$$

Hàm phân phối xác suất của X được tính qua hàm mật độ bằng công thức:

$$F_X(x) = P(X \leqslant x) = \int_{-\infty}^{x} f(x)dx.$$

Kì vọng và phương sai của X được tính theo công thức:

$$\mu = E(X) = \int_{-\infty}^{+\infty} xf(x)dx.$$

$$\sigma^2 = V(X) = \int_{-\infty}^{+\infty} (x-\mu)^2 f(x)dx.$$

Tính toán xác suất cũng như những đặc trưng quan trọng của một biến ngẫu nhiên liên tục ta phải thực hiện các phép tính tích phân. Hàm `integrate` trong R giúp tính tích phân của các hàm số. Các tham số cần thiết được cho qua hàm `integrate` như sau:

integrate(f, lower, upper)

trong đó

`f`	hàm cần tính tích phân.
`lower`	cận dưới trong tích phân (có thể hữu hạn hoặc vô hạn)
`upper`	cận trên trong tích phân (có thể hữu hạn hoặc vô hạn)

Ví dụ sau minh họa cho ta cách dùng hàm `integrate`

Ví dụ 6.3.1 Cho biến ngẫu nhiên X có hàm mật độ xác suất cho bởi

$$f(x) = \begin{cases} \frac{3}{4}(x^2 + 2x) & \text{với } x \in [0, 1], \\ 0 & \text{với } x \notin [0, 1]. \end{cases}$$

a. Chứng tỏ rằng hàm số $f(x)$ thỏa mãn là hàm mật độ xác suất của một biến ngẫu nhiên.

b. Tính xác suất $P(0.2 < X < 0.7)$.

Lời giải:

a. Nhận thấy $f(x) \geqslant 0$, để chứng minh $f(x)$ là hàm mật độ xác suất của một biến ngẫu nhiên ta chỉ cần chứng minh $\int_{-\infty}^{+\infty} f(x)dx = 1$. Do $f(x) = 0$ khi $x \notin [0, 1]$ nên:

$$\int_{-\infty}^{+\infty} f(x)dx = \frac{3}{4}\int_0^1 (x^2 + 2x)dx = \frac{3}{4}(\frac{x^3}{3} + x^2)|_0^1 = 1.$$

b. Ta có

$$P(0.2 < X < 0.7) = \frac{3}{4}\int_{0.2}^{0.7} (x^2 + 2x)dx = \frac{3}{4}(\frac{x^3}{3} + x^2)|_{0.2}^{0.7} = 0.42125.$$

Trong R, ta có thể thực hiện những tính toán trên như sau:

```
> f = function(x) 3/4*(x^2+2*x)
> integrate(f, lower=0,upper=1)
1 with absolute error < 1.1e-14
> integrate(f, lower=0.2,upper=0.7)
0.42125 with absolute error < 4.7e-15
```

Trong gói `distrEx`, ta còn có thể làm như sau:

```
> library(distrEx)
> f = function(x) 3/4*(x^2+2*x)
> X <- AbscontDistribution(d = f, low1 = 0, up1 = 1)
> p(X)(1) - p(X)(0)
[1] 1
> p(X)(0.7) - p(X)(0.2)
[1] 0.42125
```

Ngoài ra, kì vọng và phương sai của X cũng được tính bằng cách sau:

```
> library(distrEx)
> f = function(x) 3/4*(x^2+2*x)
> X <- AbscontDistribution(d = f, low1 = 0, up1 = 1)
> E(X); var(X); sd(X)
[1] 0.6872252
[1] 0.05244672
[1] 0.2290125
```

6.4 Một số phân phối liên tục thường gặp

Trong xác suất cũng như trong thống kê, ta sử dụng đến nhiều những phân phối liên tục như phân phối chuẩn, phân phối đều, phân phối mũ, phân phối student, phân phối khi - bình phương, phân phối fisher, phân phối gamma,...

Trong R các phân phối trên được viết tắt như sau: phân phối chuẩn là `norm(normal)`, phân phối đều là `unif(uniform)`, phân phối mũ là `exp(exponent)`, phân phối student là `t(student)`, phân phối khi–bình phương là `chisq(chi square)`, phân phối fisher là `f(fisher)`, phân phối gamma là `gamma(gamma)`,...

Dưới đây ta sẽ giới thiệu một số hàm trong R liên quan đến những phân phối này.

6.4.1 Phân phối đều liên tục

Biến ngẫu nhiên liên tục X được gọi là tuân theo phân phối đều trên đoạn $[a, b]$, kí hiệu là $X \sim U(a, b)$ nếu hàm mật độ xác suất của nó có dạng:

$$f(x) = \begin{cases} \dfrac{1}{b-a} & \text{nếu } x \in [a, b]. \\ 0 & \text{nếu } x \notin [a, b], \end{cases}$$

Bảng sau cho ta những hàm tính toán liên quan đến phân phối đều trong R:

Tính	Với gói `stats`	Với gói `distr`
Hàm mật độ $f(x)$	`dunif(x, min = a, max = b)`	`d(X)(x)`
$P(X \leqslant x)$	`punif(x, min = a, max = b)`	`p(X)(x)`
x để $P(X \leqslant x) = p$	`qunif(p, min = a, max = b)`	`q(X)(p)`
Tạo mẫu cỡ k	`runif(k, min = a, max = b)`	`r(X)(k)`

Khi sử dụng gói `distr` ta phải đặt `X = Unif(min = a, max = b)`.

Ta xét ví dụ minh họa sau.

Ví dụ 6.4.1 Xe buýt, bắt đầu từ 7h sáng, đỗ tại một bến cứ 15 phút một lần, tức là sẽ có các chuyến xe buýt lúc 7h15, 7h30, 7h45, ... Giả sử rằng thời điểm một hành khách đến bến xe phân phối một cách đều từ 7h đến 7h30, tìm xác suất để anh ta phải đợi xe buýt ít hơn 5 phút.

Lời giải: Gọi X là khoảng thời gian người khách đến bến tính từ 7h. Khi đó, X tuân theo phân phối đều trên [0, 30].

Muốn đợi xe ít hơn 5 phút, người khách phải đến bến lúc 7h, từ 7h10 đến 7h15, từ 7h25 đến 7h30. Vậy xác suất để người khách đợi xe ít hơn 5 phút là:

$$P(X = 0) + P(10 < X < 15) + P(25 < X < 30) = \int_0^0 \frac{dx}{30} + \int_{10}^{15} \frac{dx}{30} + \int_{25}^{30} \frac{dx}{30} = \frac{1}{30}.$$

Trong R ta có thể thực hiện những tính toán trên như sau:

```
> punif(15,0,30)-punif(10,0,30) + punif(30,0,30)-punif(25,0,30)
[1] 0.3333333
> #Hoặc
> diff(punif(c(15,10),0,30)) + diff(punif(c(30,25),0,30))
[1] 0.3333333
```

6.4.2 Phân phối chuẩn

Cho $X \sim N(\mu, \sigma^2)$ là biến ngẫu nhiên tuân theo phân phối chuẩn với trung bình là μ và phương sai là σ^2. Hàm mật độ xác suất của X được cho bởi:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Bảng sau cho ta những hàm tính toán liên quan đến phân phối chuẩn trong R:

Tính	Với gói `stats`	Với gói `distr`
Hàm mật độ $f(x)$	`dnorm(x, mean =` μ `, sd =` σ`)`	`d(X)(x)`
$P(X \leqslant x)$	`pnorm(x, mean =` μ `, sd =` σ`)`	`p(X)(x)`
x để $P(X \leqslant x) = p$	`qnorm(p, mean =` μ `, sd =` σ`)`	`q(X)(p)`
Tạo mẫu cỡ k	`rnorm(k, mean =` μ `, sd =` σ`)`	`r(X)(k)`

Khi sử dụng gói `distr`, ta phải đặt `X = Norm(mean =` μ `, sd =` σ`)`.
Ta xét ví dụ minh họa sau.

Ví dụ 6.4.2 Chỉ số IQ của con người được cho là tuân theo phân phối chuẩn với trung bình là 100 và độ lệch chuẩn là 15.

a. Một người được coi là có IQ bình thường nếu có chỉ số IQ từ 85 đến 115. Tìm tỉ lệ những người có IQ mức bình thường trên thế giới.

b. Một người được đánh giá là thiên tài nếu có chỉ số IQ cao trên 160. Tìm tỉ lệ những người được coi là thiên tài trên thế giới.

c. Tìm chỉ số IQ thấp nhất trong nhóm 10% những người có IQ cao nhất trên thế giới.

Lời giải: Gọi X là chỉ số IQ của con người. Khi đó, X tuân theo phân phối chuẩn với trung bình là 100 và độ lệch chuẩn là 15.

a. Tỉ lệ những người có mức IQ bình thường là: $P(85 \leqslant X \leqslant 115) = 0.68$.
Vậy khoảng 68% dân số thế giới có mức IQ bình thường.

Tỉ lệ này được tính toán trong R như sau:

```
> pnorm(115,100,15)-pnorm(85,100,15)
[1] 0.6826895
> #Hoặc
> diff(pnorm(c(85,115),100,15))
[1] 0.6826895
```

b. Tỉ lệ những người được coi là thiên tài trên thế giới là:

$$P(X > 160) = 1–P(X \leqslant 160) = 3 * 10^{-5}.$$

Như vậy, chỉ khoảng $3/100000$ người được đánh giá là thiên tài trên thế giới, tức là 100000 người mới có khoảng 3 người được coi là thiên tài.

Tỉ lệ này được tính trong R như sau:

```
> 1-pnorm(160,100,15)
[1] 3.167124e-05
> #Hoặc
> pnorm(160,100,15,lower.tail=F)
[1] 3.167124e-05
```

c. Gọi x là mức IQ thấp nhất trong nhóm 10% những người có IQ cao nhất thế giới. Ta có $P(X \geqslant x) = 0.1$ hay $P(X < x) = 0.9$. Ta tính được $x = 119$.
Vậy một người có IQ ít nhất là 119 sẽ thuộc nhóm 10% những người có IQ cao nhất thế giới.

Giá trị x này được tính trong R như sau:

```
> qnorm(1-0.1,100,15)
[1] 119.2233
> #Hoặc
> qnorm(0.1,100,15, lower.tail=F)
[1] 119.2233
```

6.4.3 Phân phối mũ

Cho $X \sim exp(\lambda)$ là biến ngẫu nhiên tuân theo phân phối mũ với tham số λ. Hàm mật độ xác suất của X được cho bởi:

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{nếu } x \geqslant 0, \\ 0 & \text{nếu } x < 0. \end{cases}$$

Bảng sau cho ta những hàm tính toán liên quan đến phân phối mũ trong R:

Tính	Với gói `stats`	Với gói `distr`
Hàm mật độ $f(x)$	`dexp(x, rate = `λ`)`	`d(X)(x)`
$P(X \leqslant x)$	`pexp(x, rate = `λ`)`	`p(X)(x)`
x để $P(X \leqslant x) = p$	`qexp(p, rate = `λ`)`	`q(X)(p)`
Tạo mẫu cỡ k	`rexp(k, rate = `λ`)`	`r(X)(k)`

Khi sử dụng gói `distr` ta phải đặt `X = Exp(rate = `λ`)`.
Ta xét ví dụ minh họa sau.

Ví dụ 6.4.3 Tuổi thọ (năm) của một chiếc máy giặt được cho là tuân theo phân phối mũ với tham số $\lambda = 1/10$.

a. Tính xác suất để một chiếc máy giặt có tuổi thọ ít nhất là 10 năm.

b. Một chiếc máy giặt đã dùng được 5 năm, tính xác suất để nó có thể dùng thêm được ít nhất 10 năm nữa.

Lời giải: Gọi X là tuổi thọ của một chiếc máy giặt. Khi đó, X tuân theo phân phối mũ với $\lambda = 1/10$.

a. Xác suất để một chiếc máy giặt có tuổi thọ ít nhất là 10 năm là:

$$P(X \geqslant 10) = 1 – P(X < 10) = 0.37.$$

Vậy khoảng 37% số máy giặt dùng được ít nhất bằng trung bình là 10 năm.

Xác suất này được tính trong R như sau:

```
> 1-pexp(10,1/10)
[1] 0.3678794
> #Hoặc
> pexp(10,1/10,lower.tail=F)
[1] 0.3678794
```

b. Xác suất để một chiếc máy giặt đã dùng được 5 năm, còn dùng thêm ít nhất 10 năm nữa là:

$$\begin{aligned} P(X \geqslant 15|X \geqslant 5) &= \frac{P[(X \geqslant 15)(X \geqslant 5)]}{P(X \geqslant 5)} \\ &= \frac{P(X \geqslant 15))}{P(X \geqslant 5)} = 0.37. \end{aligned}$$

Nhận thấy $P(X \geqslant 15|X \geqslant 5) = P(X \geqslant 10)$ hay tổng quát hơn ta có

$$P(X > s + t|X > t) = P(X > s).$$

Tính chất này được gọi là tính chất không có trí nhớ của phân phối mũ. Tính chất này trong ví dụ của ta cho thấy không kể chiếc máy giặt đã dùng được t là bao nhiêu năm, khả năng để nó còn dùng thêm được s năm nữa là không đổi.

Trong R, ta tính xác suất này như sau:

```
> (1-pexp(15,1/10))/(1-pexp(5,1/10))
[1] 0.3678794
> #Hoặc
> pexp(15,1/10,lower.tail=F)/pexp(5,1/10,lower.tail=F)
[1] 0.367879
```

Chú ý: Ta còn có thể sử dụng gói `distr` để làm các ví dụ (6.2.1), (6.2.2), (6.4.1), (6.4.2), (6.4.3). Chẳng hạn, các yêu cầu trong ví dụ (6.4.2) có thể được thực hiện như sau:

```
> X = Norm(100,15)
```

a. Tỉ lệ những người có mức IQ bình thường là:

```
> X = Norm(100,15)
> p(X)(115) - p(X)(85)
[1] 0.6826895
```

b. Tỉ lệ những người được coi là thiên tài trên thế giới là:

```
> 1-p(X)(160)
[1] 3.167124e-05
```

c. Chỉ số IQ thấp nhất trong nhóm 10% những người có IQ cao nhất thế giới là:

```
> q(X)(0.9)
[1] 119.2233
```

6.4.4 Một số phân phối liên tục khác

Bảng dưới đây giới thiệu những hàm tính xác suất của một số phân phối liên tục khác:

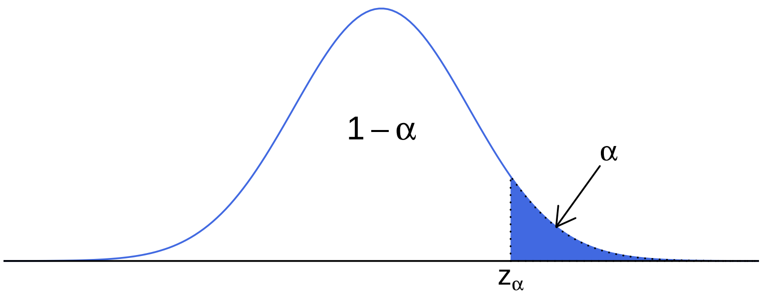
$X \sim t(n)$ tuân theo phân phối student với n bậc tự do Trong gói `distr` đặt `X = Td(df = n)`		
Tính	Gói `stats`	Gói `distr`
Hàm mật độ $f(x)$	`dt(x, df = n)`	`d(X)(x)`
$P(X \leqslant x)$	`pt(x, df = n)`	`p(X)(x)`
x để $P(X \leqslant x) = p$	`qt(p, df = n)`	`q(X)(p)`
Tạo mẫu cỡ k	`rt(k, df = n)`	`r(X)(k)`
$X \sim \chi(n)$ tuân theo phân phối khi-bình phương với n bậc tự do Trong gói `distr` đặt `X = Chisq(df = n)`		
Tính	Gói `stats`	Gói `distr`
Hàm mật độ $f(x)$	`dchisq(x, df = n)`	`d(X)(x)`
$P(X \leqslant x)$	`pchisq(x, df = n)`	`p(X)(x)`
x để $P(X \leqslant x) = p$	`qchisq(p, df = n)`	`q(X)(p)`
Tạo mẫu cỡ k	`rchisq(k, df = n)`	`r(X)(k)`
$X \sim F(m, n)$ tuân theo phân phối fisher với m, n bậc tự do Trong gói `distr` đặt `X = Fd(df1 = m, df2 = n)`		
Tính	Gói `stats`	Gói `distr`
Hàm mật độ $f(x)$	`df(x, df1 = m, df2 = n)`	`d(X)(x)`
$P(X \leqslant x)$	`pf(x, df1 = m, df2 = n)`	`p(X)(x)`
x để $P(X \leqslant x) = p$	`qf(p, df1 = m, df2 = n)`	`q(X)(p)`
Tạo mẫu cỡ k	`rf(k, df1 = m, df2 = n)`	`r(X)(k)`
$X \sim Gamma(\alpha, \lambda)$ tuân theo phân phối gamma với tham số α, λ Trong gói `distr` đặt `X = Gamma(shape = `α` , rate = `λ`)`		
Tính	Gói `stats`	Gói `distr`
Hàm mật độ $f(x)$	`dgamma(x, shape=`α`,rate=`λ`)`	`d(X)(x)`
$P(X \leqslant x)$	`pgamma(x, shape=`α`,rate=`λ`)`	`p(X)(x)`
x để $P(X \leqslant x) = p$	`qgamma(p, shape=`α`,rate=`λ`)`	`q(X)(p)`
Tạo mẫu cỡ k	`rgamma(k, shape=`α`,rate=`λ`)`	`r(X)(k)`
$X \sim Cauchy(m, \beta)$ tuân theo phân phối cauchy với tham số m, β Trong gói `distr` đặt `X = Cauchy(location=m,scale=`β`)`		
Tính	Gói `stats`	Gói `distr`
Hàm mật độ $f(x)$	`dcauchy(x, location=m,scale=`β`)`	`d(X)(x)`
$P(X \leqslant x)$	`pcauchy(x, location=m,scale=`β`)`	`p(X)(x)`
x để $P(X \leqslant x) = p$	`qcauchy(p, location=m,scale=`β`)`	`q(X)(p)`
Tạo mẫu cỡ k	`rcauchy(k, location=m,scale=`β`)`	`r(X)(k)`

6.4.5 Một số giá trị tới hạn thường gặp

Trong phần này, ta sẽ sử dụng những hàm tính toán của những phân phối liên tục được giới thiệu trên để tính một số giá trị tới hạn thường gặp. Các giá trị này được tính khi sử dụng gói cơ bản `stats`, cách tính theo gói `distr` được làm hoàn toàn tương tự.

Giá trị tới hạn z_α của phân phối chuẩn hóa

Giả sử Z là biến ngẫu nhiên tuân theo phân phối chuẩn hóa $N(0,1)$ và α là số thuộc $[0,1]$. Giá trị tới hạn z_α là số được xác định bởi $P(Z > z_\alpha) = \alpha$. Từ định nghĩa này ta có $P(Z \leqslant z_\alpha) = 1-\alpha$.



Trong R, giá trị z_α có thể được tính toán theo một số công thức sau:

- z_α `= qnorm(1 -` α`, 0, 1)` hoặc z_α `= qnorm(1 -` α`).`
- z_α `= qnorm(`α`, 0, 1, lower.tail = F)` hoặc z_α `= qnorm(`α`, lower.tail = F).`

Đoạn lệnh sau cho ta cách tính z_α với một số giá trị α thường gặp.

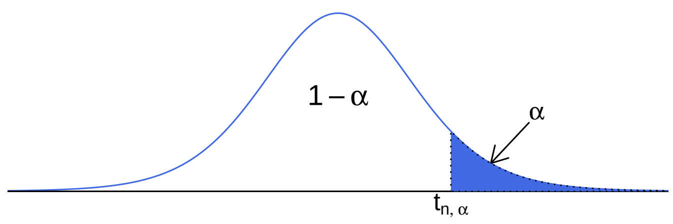
```
> alpha = c(0.2, 0.1, 0.05, 0.025, 0.01, 0.005)
> qnorm(1 - alpha)
[1] 0.8416212 1.2815516 1.6448536 1.9599640 2.3263479 2.5758293
> #Hoặc
> qnorm(alpha, lower.tail = F)
[1] 0.8416212 1.2815516 1.6448536 1.9599640 2.3263479 2.5758293
```

Bảng dưới đây ghi lại những giá trị z_α với một số giá trị α thường gặp trên.

α	0.2	0.1	0.05	0.025	0.01	0.005
z_α	0.84	1.28	1.645	1.96	2.33	2.58

Giá trị tới hạn $t_{n,\alpha}$ của phân phối Student

Giả sử t là biến ngẫu nhiên tuân theo phân phối student với n bậc tự do và α là số thuộc $[0, 1]$. Giá trị tới hạn $t_{n,\alpha}$ là số được xác định bởi $P(t > t_{n,\alpha}) = \alpha$. Từ định nghĩa này ta có $P(t \leqslant t_{n,\alpha}) = 1 - \alpha$.



Trong R, giá trị $t_{n,\alpha}$ được tính như sau:

$$\texttt{qt}(1-\alpha, \texttt{df} = \texttt{n}) \qquad \text{hoặc} \qquad \texttt{qt}(\alpha, \texttt{n}, \texttt{lower.tail} = \texttt{F})$$

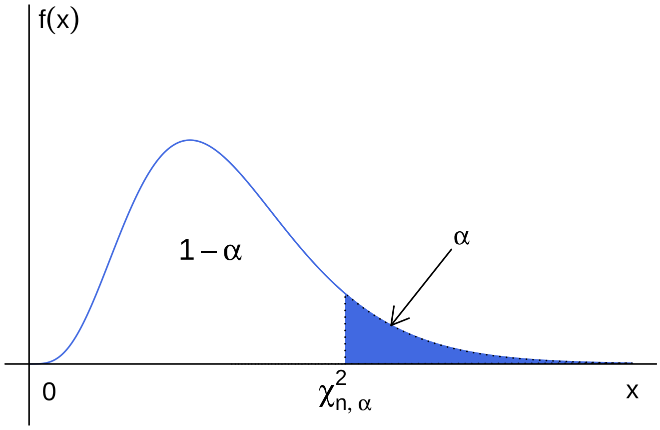
Đoạn lệnh sau cho ta cách tính một số giá trị của $t_{n,\alpha}$ với một số α thường gặp của phân phối student có bậc tự do là $n = 10$.

```
> alpha = c(0.1, 0.05, 0.01)
> qt(1-alpha, df = 10)
[1] 1.372184 1.812461 2.763769
> #Hoặc
> qt(alpha, 10, lower.tail=F)
[1] 1.372184 1.812461 2.763769
```

Từ kết quả trên ta có $t_{10,0.1} = 1.37, t_{10,0.05} = 1.81, t_{10,0.01} = 2.76$.

Giá trị tới hạn $\chi^2_{n,\alpha}$ của phân phối khi-bình phương

Giả sử χ là biến ngẫu nhiên tuân theo phân phối khi-bình phương với n bậc tự do và α là số thuộc $[0, 1]$. Giá trị tới hạn $\chi^2_{n,\alpha}$ là số được xác định bởi $P(t > \chi^2_{n,\alpha}) = \alpha$. Từ định nghĩa này ta có $P(\chi \leqslant \chi^2_{n,\alpha}) = 1 - \alpha$.



Trong R, giá trị $\chi^2_{n,\alpha}$ được tính như sau:

$$\texttt{qchisq}(1-\alpha, \texttt{df} = \texttt{n}) \qquad \text{hoặc} \qquad \texttt{qt}(\alpha, \texttt{n}, \texttt{lower.tail} = \texttt{F})$$

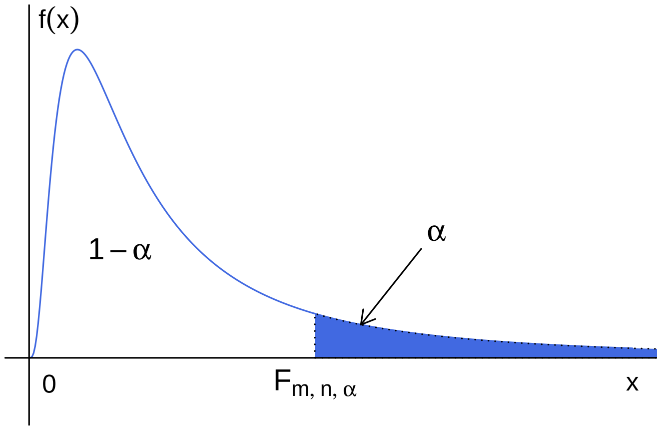
Đoạn lệnh sau cho ta cách tính một số giá trị của $\chi^2_{n,\alpha}$ với một số α thường gặp của phân phối khi-bình phương có bậc tự do là $n = 20$.

```
> alpha = c(0.1, 0.05, 0.01)
> qchisq(1-alpha, df = 20)
[1] 28.41198 31.41043 37.56623
> #Hoặc
> qchisq(alpha, 20, lower.tail=F)
[1] 28.41198 31.41043 37.56623
```

Những kết quả tính toán trên cho ta: $\chi^2_{20,0.1} = 28.41, \chi^2_{20,0.05} = 31.41, \chi^2_{20,0.01} = 37.57.$

Giá trị tới hạn $F_{m,n,\alpha}$ của phân phối Fisher

Giả sử F là biến ngẫu nhiên tuân theo phân phối Fisher với m bậc tự do ở tử, n bậc tự do ở mẫu và α là số thuộc $[0, 1]$. Giá trị tới hạn $F_{m,n,\alpha}$ là số được xác định bởi $P(F > F_{m,n,\alpha}) = \alpha$. Từ định nghĩa này ta có $P(F \leqslant F_{m,n,\alpha}) = 1 - \alpha$.



Trong R giá trị $F_{m,n,\alpha}$ được tính như sau:

$$\texttt{qf}(1-\alpha, \texttt{df1} = \texttt{m}, \texttt{df2} = \texttt{n}) \qquad \text{hoặc} \qquad \texttt{qf}(\alpha, \texttt{m}, \texttt{n}, \texttt{lower.tail} = \texttt{F})$$

Đoạn lệnh sau cho ta cách tính một số giá trị của $F_{m,n,\alpha}$ với một số α thường gặp của phân phối fisher có bậc tự do ở tử là $m = 20$ và bậc tự do ở mẫu là $n = 10$.

```
> alpha = c(0.1, 0.05, 0.01)
> qf(1-alpha, df1 = 20, df = 10)
[1] 2.200744 2.774016 4.405395
> #Hoặc
> qf(alpha, 20, 10, lower.tail=F)
[1] 2.200744 2.774016 4.405395
```

Những kết quả tính toán trên cho ta: $F_{20,10,0.1} = 2.2$, $F_{20,10,0.05} = 2.77$, $F_{20,10,0.01} = 4.41$.

BÀI TẬP

VI.45. Xác định trường hợp nào sau đây là bảng phân phối xác suất

a.

X	-1	0	1
$p(x)$	0.2	0.6	0.2

b.

X	1/2	3/4	1
$p(x)$	-1	0	2

c.

X	2	4	6
$p(x)$	0.25	0.35	0.5

d.

X	0.1	0.7	0.8
$p(x)$	2/5	1/5	2/5

VI.46. Xét phân phối xác suất như sau:

X	0	1	2	3	4	5
$p(x)$	0.15	0.35	0.2	0.1	0.15	0.05

Tìm trung bình và phương sai của X.

VI.47. Gọi X là biến ngẫu nhiên chỉ chênh lệch giữa số mặt sấp và mặt ngửa khi tung một đồng xu 4 lần.

a. Lập bảng phân phối xác suất của X.

b. Tính kì vọng và phương sai của X.

VI.48. Mười quả bóng được chọn ngẫu nhiên từ một chiếc bình có 17 quả bóng trắng và 23 quả bóng đen. Gọi X là biến ngẫu nhiên chỉ số bóng trắng được lấy ra. Tính EX.

VI.49. Trong một rổ có 99 quả bóng đánh số từ 1 đến 99. Nhặt ngẫu nhiên từ rổ ra 5 quả bóng. Gọi Y tương ứng là số lớn nhất hiện lên trên 5 quả bóng được nhặt ra.

a. Lập bảng phân phối xác suất của Y.

b. Tính kì vọng EY.

VI.50. Một công ty bảo hiểm bán một bảo hiểm nhân thọ với giá 20000 đô la và số tiền khách hàng phải đóng hàng năm là 300 đô la. Những bảng thống kê bảo hiểm cho thấy, một người mua bảo hiểm có thể chết trong một năm với xác suất 0.001. Gọi X là biến ngẫu nhiên chỉ lợi nhuận của công ty trên mỗi bảo hiểm bán ra trong một năm.

a. Hãy cho biết phân phối xác suất của X.

b. Tìm lợi nhuận kỳ vọng trên mỗi bảo hiểm của công ty.

c. Nếu không có giả thiết số tiền khách hàng phải đóng hàng năm là 300 đô la thì số tiền công ty phải thu của khách hàng mỗi năm là bao nhiêu để lợi nhuận kỳ vọng trên mỗi bảo hiểm lớn hơn 0?

VI.51. Một công ty thuê một luật sư trong một vụ kiện với hai phương án trả công như sau:

- Phương án 1: Trả 10 triệu đồng bất kể thắng hay thua kiện.
- Phương án 2: Trả 1 triệu đồng nếu thua kiện và 30 triệu đồng nếu thắng kiện.

Luật sư đánh giá khả năng thắng kiện của công ty này là 40%.

a. Lập bảng phân phối xác suất cho số tiền mà luật sư nhận được trong mỗi phương án.

b. Theo bạn luật sư nên chọn phương án nào?

VI.52. Thống kê số khách trên một xe buýt tại một tuyến giao thông thu được số liệu như sau:

Số khách trên một chuyến	20	25	30	35	40
Xác suất tương ứng	0.3	0.2	0.15	0.1	0.25

a. Tìm số khách trung bình đi trên mỗi chuyến.

b. Giả sử chi phí cho mỗi chuyến xe là 1,5 triệu đồng không phụ thuộc vào số khách đi trên xe. Nếu công ty xe buýt muốn lãi bình quân cho mỗi chuyến xe là 300 nghìn đồng thì phải qui định giá vé trên một hành khách là bao nhiêu?

VI.53. Số lượng thuyền gỗ X mà một xưởng đóng thuyền có thể làm được trong một tháng có bảng phân phối xác suất như sau:

X	2	3	4	5	6	7	8
P_X	0.2	0.2	0.3	0.1	0.1	0.05	0.05

a. Số thuyền có khả năng nhiều nhất mà xưởng đó có thể đóng được trong tháng tới là bao nhiêu?

b. Giả sử việc đóng thuyền có chi phí cố định hàng tháng là 25 triệu đồng và chi phí bổ sung cho mỗi con thuyền là 15 triệu đồng. Hãy tìm chi phí bình quân hàng tháng của xưởng đó.

VI.54. Trên một chuyến bay có 70 chỗ ngồi. Thực tế cho thấy đến giờ chót vẫn có khách bỏ chuyến bay. Để tận dụng hết chỗ bay bằng cách bán thêm vé dự phòng người ta đã thống kê 20 chuyến bay và thu được các số liệu sau:

Số khách bỏ chuyến bay	1	2	3	4	5	6	7	8	9	10
Số chuyến bay tương ứng	1	4	0	4	2	5	1	1	0	2

a. Hãy tính xác suất để trong một chuyến bay nào đó có nhiều hơn 5 hành khách bỏ chuyến bay.

b. Tìm số hành khách bỏ chuyến trung bình của mỗi chuyến bay.

VI.55. Thời gian sửa chữa một chiếc máy tính cá nhân (đơn vị: giờ) là biến ngẫu nhiên có hàm mật độ như sau:

$$f(x) = \begin{cases} 1/2, & 0 < x < 2 \\ 0, & \text{trường hợp còn lại} \end{cases}$$

a. Tìm thời gian trung bình để sửa chữa một chiếc máy.

b. Chi phí sửa chữa phụ thuộc vào thời gian theo công thức $40 + 30\sqrt{x}$ trong đó x là thời gian sửa chữa chiếc máy. Tìm chi phí kỳ vọng để sửa chữa một chiếc máy tính cá nhân.

VI.56. Một bài thi trắc nghiệm gồm 100 câu hỏi, mỗi câu có 4 đáp án trả lời nhưng chỉ có một đáp án là đúng. Nam không học bài mà chỉ trả lời một cách ngẫu nhiên.

a. Tính xác suất để Nam trả lời đúng một nửa số câu hỏi.

b. Một sinh viên thi đỗ nếu trả lời đúng ít nhất một nửa số câu hỏi. Tính xác suất để Nam thi đỗ.

c. Số câu trả lời đúng trung bình của Nam là bao nhiêu? Tính xác suất để Nam trả lời đúng số câu bằng số câu trung bình này và nhận xét.

VI.57. Một bài thi trắc nghiệm gồm 12 câu hỏi, mỗi câu hỏi có 5 phương án trả lời, trong đó chỉ có một đáp án đúng. Giả sử mỗi câu trả lời đúng được 4 điểm, mỗi câu trả lời sai bị trừ 1 điểm. Một sinh viên làm bài bằng cách trả lời ngẫu nhiên từng câu hỏi.

a. Tính số câu trả lời đúng và số câu trả lời sai để sinh viên này được 13 điểm.

b. Tính xác suất để sinh viên này được 13 điểm.

c. Tính xác suất để sinh viên này bị điểm âm.

VI.58. Cho X là biến ngẫu nhiên nhị thức với $EX = 3$, $VX = 2.1$. Tính các xác suất sau:

a. $P(X = 7)$.

b. $P(X < 7)$.

c. $P(X \leqslant 7)$.

d. $P(X > 7)$.

e. $P(X \geqslant 7)$.

VI.59. Một người tập ném bóng rổ, đứng từ một chỗ ném bóng vào rổ 6 lần. Xác suất ném trúng mỗi lần là $2/3$. Gọi X, Y tương ứng là số lần ném trúng và ném trượt, hãy tính $E(X - Y)$.

VI.60. Nếu bạn mua 50 vé xổ số và cơ hội trúng thưởng của mỗi vé số là 1/100.

a. Tính xác suất để bạn trúng ít nhất một giải, đúng một giải, ít nhất hai giải.

b. Tìm số giải mà bạn có khả năng trúng cao nhất.

VI.61. Số lần một người bị cảm lạnh trong một năm tuân theo phân phối Poisson với $\lambda = 3$.

a. Tính số lần bị cảm trung bình trong một năm của một người.

b. Tính xác suất để một người không bị cảm lạnh lần nào trong năm.

c. Tính xác suất để một người bị cảm lạnh không quá ba lần trong một năm.

VI.62. Giả sử một người có mặt tại bến xe buýt lúc 10 giờ sáng, cho biết thời điểm xe buýt đỗ tại bến tuân theo phân phối đều giữa 10h và 10h30'.

a. Tính xác suất người đó phải đợi trên 15 phút.

b. Nếu lúc 10h15' xe buýt vẫn chưa tới bến, tính xác suất để người đó phải đợi thêm ít nhất 5 phút nữa.

VI.63. Cho X là biến ngẫu nhiên phân phối chuẩn với tham số $\mu = 8, \sigma^2 = 25$. Tính các xác suất sau:

a. $P(X > 5)$.

b. $P(2 < X < 6)$.

c. $P(X < 15)$.

d. $P(X > 10)$.

e. Tìm x_0 sao cho $P(X < x_0) = 0.4$.

f. Tìm x_1 sao cho $P(X > x_1) = 0.7$.

VI.64. Giả sử tuổi thọ của một chiếc đèn hình màu trong tivi tuân theo phân phối chuẩn với trung bình 8.2 năm và độ lệch chuẩn 1.4 năm. Chọn ngẫu nhiên một chiếc đèn hình màu.

a. Tính xác suất để chiếc đèn hình màu có tuổi thọ trên 10 năm.

b. Tính xác suất để một chiếc đèn hình màu có tuổi thọ ít hơn 4 năm.

c. Tính xác suất để một chiếc đèn hình màu có tuổi thọ từ 4 đến 10 năm.

d. Nếu qui định thời gian bảo hành là 5 năm thì tỉ lệ đèn hình phải đổi lại theo bảo hành là bao nhiêu?

e. Nếu muốn tỉ lệ đèn hình phải đổi lại theo bảo hành là 5% thì phải qui định thời gian bảo hành là bao nhiêu năm?

VI.65. Chỉ số IQ của người tuân theo phân phối chuẩn với trung bình 100 và độ lệch chuẩn 14.2. Nhóm 10% những người có chỉ số IQ cao nhất có chỉ số IQ nằm trong phạm vi nào?

VI.66. Giả sử lượng mưa hàng năm (mm) của một địa phương tuân theo phân phối chuẩn với trung bình là 1800, độ lệch chuẩn là 100. Tính xác suất để có 2 năm trong 4 năm có lượng mưa không quá 1600 mm. Giả thiết rằng lượng mưa trong các năm khác nhau là độc lập.

VI.67. Giả sử quãng đường (nghìn dặm) một chiếc ôtô đi được cho đến khi không sử dụng được nữa tuân theo phân phối mũ với tham số $\lambda = 1/20$.

a. Tính quãng đường trung bình mà một chiếc ô tô đi được.

b. Một người mua một chiếc xe mới, tính xác suất để người này đi được ít nhất 20 nghìn dặm.

c. Một người mua một chiếc ôtô cũ đã đi được 10 nghìn dặm, tính xác suất để anh ta có thể sử dụng nó để đi tiếp được ít nhất 20 nghìn dặm nữa.

VI.68. Tại một trung tâm cấp cứu, số bệnh nhân đến trong một ngày tuân theo phân phối Poisson có trung bình 2 người. Tính xác suất để khoảng thời gian giữa thời điểm hai bệnh nhân đến ít hơn 8 giờ.

VI.69. Số lần động đất tại một khu vực có phân phối Poisson với trung bình 5 trận mỗi năm.

a. Xác suất có ít nhất 3 trận động đất trong nửa đầu của năm 2019 là bao nhiêu?

b. Giả sử sự kiện trên xảy ra, xác suất không có động đất ở khu vực trong năm 2020 là bao nhiêu?

c. Giả sử có một trận động đất vào tháng 5 năm 2020. Tính xác suất để ít nhất một năm sau đó không có trận động đất nào.