Bài giảng Xác suất Thống kê và ứng dụng

Nguyễn Thị Nhung

Bộ môn Toán - Đại học THĂNG LONG

Ngày 8 tháng 8 năm 2013

Chương V

Một số phân phối lí thuyết quan trọng

Chương V

1. Phân phối lí thuyết rời rạc
   - Phân phối nhị thức
   - Phân phối Poisson
2. Phân phối lí thuyết liên tục
   - Phân phối chuẩn
   - Phân phối đều
   - Phân phối mũ

Chương V

1. Phân phối lí thuyết rời rạc
   - Phân phối nhị thức
   - Phân phối Poisson

2. Phân phối lí thuyết liên tục
   - Phân phối chuẩn
   - Phân phối đều
   - Phân phối mũ

Câu hỏi tình huống

Ngày mai bạn phải thi một môn trắc nghiệm gồm 20 câu hỏi và mỗi câu hỏi có 4 phương án trả lời nhưng chưa ôn được gì cả. Vì là môn thi trắc nghiệm nên bạn vẫn định tham dự thi để biết đâu may mắn mình trả lời ngẫu nhiên lại đúng được nhiều câu tuy nhiên bạn vẫn băn khoăn tự hỏi:

- *Số câu mình trả lời đúng không biết là bao nhiêu nhưng chúng có tuân theo qui luật gì hay không?*
- *Khả năng mình thi đỗ là bao nhiêu? Khả năng mình được 10 điểm là bao nhiêu?*

Câu hỏi tình huống

Câu hỏi tình huống

Ngày mai bạn phải thi một môn trắc nghiệm gồm 20 câu hỏi và mỗi câu hỏi có 4 phương án trả lời nhưng chưa ôn được gì cả. Vì là môn thi trắc nghiệm nên bạn vẫn định tham dự thi để biết đâu may mắn mình trả lời ngẫu nhiên lại đúng được nhiều câu tuy nhiên bạn vẫn băn khoăn tự hỏi:

- *Số câu mình trả lời đúng không biết là bao nhiêu nhưng chúng có tuân theo qui luật gì hay không?*
- *Khả năng mình thi đỗ là bao nhiêu? Khả năng mình được 10 điểm là bao nhiêu?*

Câu hỏi tình huống

Ngày mai bạn phải thi một môn trắc nghiệm gồm 20 câu hỏi và mỗi câu hỏi có 4 phương án trả lời nhưng chưa ôn được gì cả. Vì là môn thi trắc nghiệm nên bạn vẫn định tham dự thi để biết đâu may mắn mình trả lời ngẫu nhiên lại đúng được nhiều câu tuy nhiên bạn vẫn băn khoăn tự hỏi:

- *Số câu mình trả lời đúng không biết là bao nhiêu nhưng chúng có tuân theo qui luật gì hay không?*
- *Khả năng mình thi đỗ là bao nhiêu? Khả năng mình được 10 điểm là bao nhiêu?*

Câu hỏi tình huống

Ngày mai bạn phải thi một môn trắc nghiệm gồm 20 câu hỏi và mỗi câu hỏi có 4 phương án trả lời nhưng chưa ôn được gì cả. Vì là môn thi trắc nghiệm nên bạn vẫn định tham dự thi để biết đâu may mắn mình trả lời ngẫu nhiên lại đúng được nhiều câu tuy nhiên bạn vẫn băn khoăn tự hỏi:

- *Số câu mình trả lời đúng không biết là bao nhiêu nhưng chúng có tuân theo qui luật gì hay không?*
- *Khả năng mình thi đỗ là bao nhiêu? Khả năng mình được 10 điểm là bao nhiêu?*

Câu hỏi tình huống

Câu hỏi tình huống (London nguy hiểm hay an toàn?)

Ngày 10/07/2008, có 4 vụ án mạng bằng dao xảy ra ở bốn nơi khác nhau ở London. Sự kiện này làm náo loạn dư luận đến mức thủ tướng Anh là Gordon Brown phải tuyên bố hứa sẽ tìm cách làm giảm các vụ đâm dao. London có trở nên nguy hiểm cho tính mạng hơn những năm trước không? Các nhà thống kê xã hội học đã phải đi điều tra dữ liệu để tìm hiểu. Một câu hỏi đầu tiên đặt ra đó là: Số vụ án mạng hàng năm ở London là không biết trước, tuy nhiên chúng có tuân theo qui luật nào hay không?

Câu hỏi tình huống

Câu hỏi tình huống (London nguy hiểm hay an toàn?)

Ngày 10/07/2008, có 4 vụ án mạng bằng dao xảy ra ở bốn nơi khác nhau ở London. Sự kiện này làm náo loạn dư luận đến mức thủ tướng Anh là Gordon Brown phải tuyên bố hứa sẽ tìm cách làm giảm các vụ đâm dao. London có trở nên nguy hiểm cho tính mạng hơn những năm trước không? Các nhà thống kê xã hội học đã phải đi điều tra dữ liệu để tìm hiểu. Một câu hỏi đầu tiên đặt ra đó là: Số vụ án mạng hàng năm ở London là không biết trước, tuy nhiên chúng có tuân theo qui luật nào hay không?

Câu hỏi tình huống

Câu hỏi tình huống (London nguy hiểm hay an toàn?)

Ngày 10/07/2008, có 4 vụ án mạng bằng dao xảy ra ở bốn nơi khác nhau ở London. Sự kiện này làm náo loạn dư luận đến mức thủ tướng Anh là Gordon Brown phải tuyên bố hứa sẽ tìm cách làm giảm các vụ đâm dao. London có trở nên nguy hiểm cho tính mạng hơn những năm trước không? Các nhà thống kê xã hội học đã phải đi điều tra dữ liệu để tìm hiểu. Một câu hỏi đầu tiên đặt ra đó là: Số vụ án mạng hàng năm ở London là không biết trước, tuy nhiên chúng có tuân theo qui luật nào hay không?

Câu hỏi tình huống

Câu hỏi tình huống (London nguy hiểm hay an toàn?)

Ngày 10/07/2008, có 4 vụ án mạng bằng dao xảy ra ở bốn nơi khác nhau ở London. Sự kiện này làm náo loạn dư luận đến mức thủ tướng Anh là Gordon Brown phải tuyên bố hứa sẽ tìm cách làm giảm các vụ đâm dao. London có trở nên nguy hiểm cho tính mạng hơn những năm trước không? Các nhà thống kê xã hội học đã phải đi điều tra dữ liệu để tìm hiểu. Một câu hỏi đầu tiên đặt ra đó là: Số vụ án mạng hàng năm ở London là không biết trước, tuy nhiên chúng có tuân theo qui luật nào hay không?

Câu hỏi tình huống

Câu hỏi tình huống (London nguy hiểm hay an toàn?)

Ngày 10/07/2008, có 4 vụ án mạng bằng dao xảy ra ở bốn nơi khác nhau ở London. Sự kiện này làm náo loạn dư luận đến mức thủ tướng Anh là Gordon Brown phải tuyên bố hứa sẽ tìm cách làm giảm các vụ đâm dao. London có trở nên nguy hiểm cho tính mạng hơn những năm trước không? Các nhà thống kê xã hội học đã phải đi điều tra dữ liệu để tìm hiểu. Một câu hỏi đầu tiên đặt ra đó là: Số vụ án mạng hàng năm ở London là không biết trước, tuy nhiên chúng có tuân theo qui luật nào hay không?

Câu hỏi tình huống

Câu hỏi tình huống (Chỉ số IQ của con người)

- *Trên thế giới mỗi người có một chỉ số IQ riêng, liệu chỉ số IQ của con người có tuân theo qui luật gì hay không?*
- *Tỉ lệ những người có chỉ số IQ bình thường là bao nhiêu?*
- *Tỉ lệ những người được coi là thiên tài là bao nhiêu?*

Câu hỏi tình huống

Câu hỏi tình huống (Chỉ số IQ của con người)

- *Trên thế giới mỗi người có một chỉ số IQ riêng, liệu chỉ số IQ của con người có tuân theo qui luật gì hay không?*
- *Tỉ lệ những người có chỉ số IQ bình thường là bao nhiêu?*
- *Tỉ lệ những người được coi là thiên tài là bao nhiêu?*

Câu hỏi tình huống (Chỉ số IQ của con người)

- *Trên thế giới mỗi người có một chỉ số IQ riêng, liệu chỉ số IQ của con người có tuân theo qui luật gì hay không?*
- *Tỉ lệ những người có chỉ số IQ bình thường là bao nhiêu?*
- *Tỉ lệ những người được coi là thiên tài là bao nhiêu?*

Câu hỏi tình huống

Câu hỏi tình huống

Bạn mới mở một cửa hàng bán quần áo. Bạn ước tính doanh thu của cửa hàng khoảng từ 20 đến 40 triệu và đang muốn tính xem khả năng doanh thu của hàng mình từ 35 triệu đến 40 triệu là bao nhiêu.

- *Doanh số của cửa hàng trong tình huống này có tuân theo qui luật gì hay không?*
- *Làm sao tính được khả năng doanh thu của cửa hàng từ 35 triệu đến 40 triệu?*

Câu hỏi tình huống

Bạn mới mở một cửa hàng bán quần áo. Bạn ước tính doanh thu của cửa hàng khoảng từ 20 đến 40 triệu và đang muốn tính xem khả năng doanh thu của hàng mình từ 35 triệu đến 40 triệu là bao nhiêu.

- *Doanh số của cửa hàng trong tình huống này có tuân theo qui luật gì hay không?*
- *Làm sao tính được khả năng doanh thu của cửa hàng từ 35 triệu đến 40 triệu?*

Câu hỏi tình huống

Bạn mới mở một cửa hàng bán quần áo. Bạn ước tính doanh thu của cửa hàng khoảng từ 20 *đến* 40 *triệu và đang muốn tính xem khả năng doanh thu của hàng mình từ* 35 *triệu đến* 40 *triệu là bao nhiêu.*

- *Doanh số của cửa hàng trong tình huống này có tuân theo qui luật gì hay không?*
- *Làm sao tính được khả năng doanh thu của cửa hàng từ* 35 *triệu đến* 40 *triệu?*

Câu hỏi tình huống

Bạn đang muốn mua một chiếc ô tô nhưng không đủ tiền mua ô tô mới mà định mua lại một chiếc ô tô cũ.

- *Số dặm đi được của một chiếc ô tô cho đến khi không sử dụng được nữa tuân theo qui luật gì không?*
- *Khả năng đi được bằng trung bình của một chiếc ô tô cũ có như khả năng đi được bằng trung bình của một chiếc ô tô mới không?*

Nội dung chính trong chương

- Giới thiệu về phân phối nhị thức;
- Giới thiệu về phân phối Poisson;
- Giới thiệu về phân phối chuẩn;
- Giới thiệu về phân phối đều;
- Giới thiệu về phân phối mũ;

Yêu cầu đối với sinh viên

- Nắm được phân phối nhị thức: định nghĩa, tính chất và nêu được những ví dụ áp dụng;
- Nắm được phân phối Poisson: định nghĩa, tính chất và nêu được những ví dụ áp dụng;
- Nắm được phân phối chuẩn: định nghĩa, tính chất và nêu được những ví dụ áp dụng;
- Nắm được phân phối đều: định nghĩa, tính chất và nêu được những ví dụ áp dụng;
- Nắm được phân phối mũ: định nghĩa, tính chất và nêu được những ví dụ áp dụng;

Nội dung trình bày

Phép thử Bernoulli

Định nghĩa

n phép thử được tiến hành độc lập (tức là các kết quả của phép thử này không ảnh hưởng gì đến kết quả của phép thử kia) được gọi là n phép thử Bernoulli hoặc một lược đồ Bernoulli nếu chúng thỏa mãn hai điều kiện sau:

- *Mỗi phép thử có hai kết quả* A *và* $\overline{A}$;
- $P(A) = p$; $P(A)$ *như nhau đối với mọi phép thử.*

Ví dụ:

- Gieo một con xúc xắc 100 lần, A là biến cố xuất hiện mặt lục. Đó là 100 phép thử Bernoulli;
- Một người bắn lần lượt 5 viên đạn vào mục tiêu là 5 phép thử Bernoulli. Nhưng nếu 5 người lần lượt bắn mỗi người một viên thì nói chung không phải là 5 phép thử Bernoulli.

Phép thử Bernoulli

Định nghĩa

n phép thử được tiến hành độc lập (tức là các kết quả của phép thử này không ảnh hưởng gì đến kết quả của phép thử kia) được gọi là n phép thử Bernoulli hoặc một lược đồ Bernoulli nếu chúng thỏa mãn hai điều kiện sau:

- *Mỗi phép thử có hai kết quả A và $\overline{A}$;*
- *$P(A) = p$; $P(A)$ như nhau đối với mọi phép thử.*

Ví dụ:

- Gieo một con xúc xắc 100 lần, A là biến cố xuất hiện mặt lục. Đó là 100 phép thử Bernoulli;
- Một người bắn lần lượt 5 viên đạn vào mục tiêu là 5 phép thử Bernoulli. Nhưng nếu 5 người lần lượt bắn mỗi người một viên thì nói chung không phải là 5 phép thử Bernoulli.

Tần số xuất hiện biến cố A

Mệnh đề

Tiến hành n phép thử Bernoulli. Khi đó xác suất $P_n(m,p)$ của biến cố chỉ trong n lần thử nghiệm biến cố A xuất hiện m lần được cho bởi công thức:

$$P_n(m,p) = C_n^m.p^m.(1-p)^{n-m}$$

trong đó $m = 0, 1, 2, \ldots, n$.

Tần số xuất hiện biến cố A

Chứng minh.

Thật vậy, các kết quả có thể của n phép thử Bernoulli sẽ là một dãy gồm n kết quả A hoặc $\bar{A}$. Đây có thể coi là một xâu nhị phân (gồm hai loại kí hiệu) độ dài n, có chứa đúng m kí hiệu A và $n-m$ kí hiệu $\bar{A}$.
Với mỗi một xâu như vậy, xác suất tương ứng là $p^m.(1-p)^{n-m}$.
Ta có C_n^m xâu như vậy. Xác suất cần tìm là

$$P_n(m,p) = C_n^m.p^m.(1-p)^{n-m}.$$

□

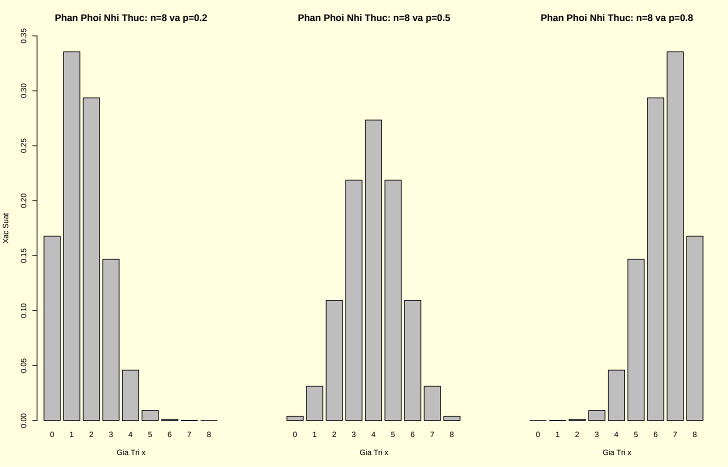
Định nghĩa phân phối nhị thức

Định nghĩa

Xét n phép thử Bernoulli với xác suất thành công $P(A) = p$. *Gọi X là số lần xuất hiện biến cố A trong n phép thử trên. Phân phối (Qui luật phân phối xác suất) của X được gọi là phân phối nhị thức và kí hiệu* $X \simeq B(n, p)$, *tức là*

$$P(X = m) = C_n^m p^m (1-p)^{n-m}, m = 0, 1, 2, \ldots n.$$

Phân phối nhị thức



Ví dụ

Bài toán

Một bài thi trắc nghiệm gồm có 20 câu hỏi, mỗi câu có 4 phương án trả lời. Mỗi câu trả lời đúng được 0.5 *điểm, trả lời sai được 0.0 điểm. Một sinh viên không học bài đi thi trả lời một cách ngẫu nhiên.*

a. *Tính xác suất để sinh viên này trả lời được* 5 *câu đúng.*

b. *Tính xác suất để sinh viên này thi đỗ (điểm ít nhất là 5).*

Lời giải

Kì vọng và phương sai của phân phối nhị thức

Mệnh đề

Cho X là biến ngẫu nhiên tuân theo phân phối nhị thức $B(n, p)$. Khi đó, kì vọng và phương sai của X được cho bởi công thức: $E(X) = np$ và $V(X) = np(1-p)$.

Chứng minh.

Gọi X_i là biến ngẫu nhiên chỉ số lần xuất hiện biến cố A ở lần thử thứ i. Ta có X_i với mọi $i = 1, 2, \ldots, n$ là các biến ngẫu nhiên độc lập, chỉ nhận hai giá trị là 0,1 với xác suất tương ứng như sau: $P(X_i = 1) = p, P(X_i = 0) = 1 - p$.
Như vậy, $E(X_i) = p$ và $V(X_i) = p(1-p)$.
Nhận thấy $X = X_1 + X_2 + \ldots + X_n$. Do đó

$$E(X) = E(X_1 + \ldots X_n) = E(X_1) + \ldots + E(X_n) = np,$$
$$V(X) = V(X_1 + \ldots X_n) = V(X_1) + \ldots + V(X_n) = np(1-p).$$

□

Ví dụ

Bài toán

Một bài thi trắc nghiệm gồm có 20 câu hỏi, mỗi câu có 4 phương án trả lời. Mỗi câu trả lời đúng được 0.5 *điểm, trả lời sai được 0.0 điểm. Một sinh viên không học bài đi thi trả lời một cách ngẫu nhiên.*

a. *Tính số câu trả lời đúng trung bình của sinh viên này.*

b. *Một nhóm gồm có* 10 *sinh viên cùng không học bài đi thi trả lời một cách ngẫu nhiên. Tính số sinh viên thi đỗ trung bình.*

Bài tập tự học: Tìm số câu mà sinh viên này có khả năng trả lời đúng nhiều nhất và cho nhận xét.

Lời giải

Tính toán liên quan đến phân phối nhị thức trong R

Giả sử $X \backsim B(n, p)$, khi đó

- `dbinom(x, n, p)` cho xác suất $P(X = x)$.
- `pbinom(x, n, p, lower.tail = TRUE (FALSE))` cho xác suất $P(X \leqslant x)$(P(X > x)).
- `qbinom(p, n, pp, lower.tail = TRUE(FALSE))` cho giá trị là x sao cho `pbinom(x, size, prob, lower.tail = TRUE(FALSE)) = p`.
- `rbinom(m, n, p)`: cho mẫu ngẫu nhiên cỡ m tuân theo phân phối nhị thức $B(n, p)$.

Tính toán phân phối nhị thức trong R

Cho $X \backsim B(n = 20, p = 0.25)$. Khi đó

- Xác suất $P(X = 5)$ là:

```
> dbinom(5, 20, 0.25)
[1] 0.2023312
```

- Xác suất $P(X \geqslant 10) = 1 - P(X \leqslant 9)$ là:

```
> 1 - pbinom(9, 20, 0.25)
[1] 0.0138644
```

Tính toán liên quan đến phân phối nhị thức trong R

- Đoạn lệnh sau cho ta cách tìm số câu trả lời đúng có khả năng nhất:

```
> x=0:20
> which.max(dbinom(x, 20, 0.25))
[1] 6
> x[6]
[1] 5
```

Nội dung trình bày

1. Phân phối lí thuyết rời rạc
   - Phân phối nhị thức
   - Phân phối Poisson
2. Phân phối lí thuyết liên tục
   - Phân phối chuẩn
   - Phân phối đều
   - Phân phối mũ

Giới thiệu về phân phối Poisson

Phân phối Poisson, được mô tả đầu tiên bởi nhà toán học Simeon Denis Poisson, là qui luật phân phối rời rạc thích hợp khi ta quan tâm đến số lần một biến cố cụ thể xảy ra trong một đơn vị thời gian hay không gian xác định, ta gọi là một phân đoạn. Chẳng hạn:

- Số lỗi trên một trang đánh máy;
- Số đơn đặt hàng gửi đến một cơ sở sản xuất trong một tháng;
- Số bệnh nhân đến một bệnh viện trong một ngày;
- Số cuộc gọi khẩn cấp nhận được mỗi 15 phút.

Các giả thiết của phân phối Poisson

Phân phối Poisson được nghiên cứu dựa trên các giả thiết sau:

- Trung bình hay kì vọng của số lần xảy ra trong một đơn vị thời gian có thể ước lượng được từ dữ liệu trong quá khứ.
- Nếu ta chia một phân đoạn thành các khoảng Δt thì các điều kiện sau được thỏa mãn:
    - Trung bình số lần biến cố xảy ra trong khoảng Δt tỉ lệ với Δt.
    - Số lần xảy ra biến cố trong khoảng thời gian Δt độc lập với thời điểm chọn Δt;
    - Số biến cố xảy ra trong những khoảng Δt rời nhau là độc lập với nhau;
    - Xác suất để có hai hay nhiều hơn lần xảy ra biến cố trong khoảng Δt là rất bé để xem như không đáng kể hay bằng 0 đối với xác suất có đúng một lần xảy ra.

Định nghĩa phân phối Poisson

Định nghĩa

Biến ngẫu nhiên X được gọi là tuân theo phân phối Poisson với tham số $\lambda(\lambda > 0)$, *kí hiệu là* P_λ, *nếu qui luật phân phối của X là:*

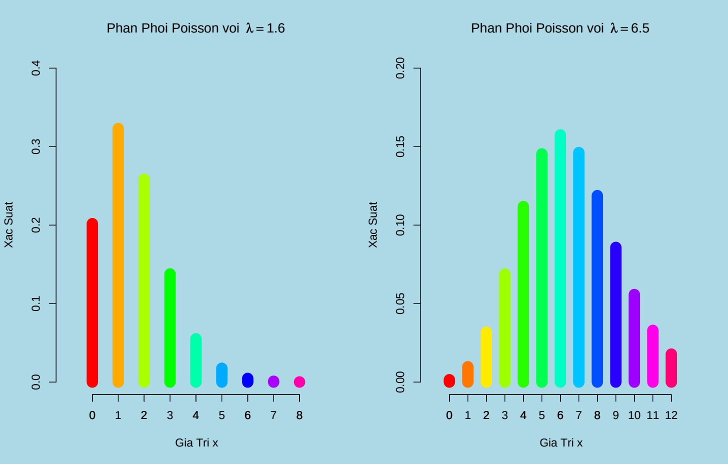
$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}, \forall k = 0, 1, 2, \dots$$

Kì vọng và phương sai của phân phối Poisson

Mệnh đề

Cho X là biến ngẫu nhiên tuân theo phân phối Poisson. Khi đó, kì vọng và phương sai của X được cho bởi công thức: $E(X) = \lambda$ *và* $V(X) = \lambda$.

Minh họa phân phối Poisson



Ví dụ

Bài toán

Theo số liệu thống kê số vụ án mạng ở London trong ngày tuân theo phân phối Poisson với trung bình là 0.44 *vụ.*

a. *Tính khả năng để trong một ngày London xảy ra 4 vụ án mạng. Từ đó hãy tính xem trung bình bao nhiêu lâu thì mới có một ngày London xảy ra 4 vụ án mạng?*

b. *Tính xem trung bình một năm London xảy ra bao nhiêu vụ án mạng.*

Lời giải

Bài tập tự học

Bài toán

Theo số liệu thống kê số vụ án mạng ở London trong ngày tuân theo phân phối Poisson với trung bình là 0.44 *vụ.*

a. *Tính khả năng để trong một ngày London xảy ra 5 vụ án mạng. Từ đó hãy tính xem trung bình bao nhiêu lâu thì mới có một ngày London xảy ra 5 vụ án mạng?*

b. *Tính xem trung bình một năm London xảy ra bao nhiêu vụ án mạng.*

Tính xác suất liên quan đến phân phối Poisson trong R

Giả sử $X \backsim P(\lambda)$, khi đó

- `dpois(x, λ)` cho xác suất $P(X = x)$.
- `ppois(x, λ), lower.tail = TRUE (FALSE))` cho xác suất $P(X \leqslant x)(P(X > x))$.
- `qpois(p, λ, lower.tail = TRUE (FALSE))` cho giá trị x sao cho `ppois(x, lower.tail = TRUE (FALSE)) = p`.
- `rpois(n, λ)` cho mẫu ngẫu nhiên cỡ n tuân theo phân phối Poisson $P(\lambda)$.

Xác suất của $P(X = 4)$ khi $X \sim P(\lambda = 0.44)$ là:

```
> dpois(4,0.44)
[1] 0.001005796
```

Nội dung trình bày

1. Phân phối lí thuyết rời rạc
   - Phân phối nhị thức
   - Phân phối Poisson

2. Phân phối lí thuyết liên tục
   - Phân phối chuẩn
   - Phân phối đều
   - Phân phối mũ

Định nghĩa

Biến ngẫu nhiên liên tục X được gọi là tuân theo phân phối chuẩn với tham số μ *và* σ^2*, kí hiệu là* $X \simeq N(\mu, \sigma^2)$ *nếu hàm mật độ xác suất của nó có dạng:*

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

Định nghĩa

Biến ngẫu nhiên Z tuân theo phân phối chuẩn với tham số $\mu = 0$ *và* $\sigma = 1$ *được gọi là biến ngẫu nhiên tuân theo phân phối chuẩn hóa, tức là hàm mật độ xác suất của nó có dạng:*

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

Định nghĩa

Biến ngẫu nhiên liên tục X được gọi là tuân theo phân phối chuẩn với tham số μ *và* σ^2, *kí hiệu là* $X \simeq N(\mu, \sigma^2)$ *nếu hàm mật độ xác suất của nó có dạng:*

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

Định nghĩa

Biến ngẫu nhiên Z tuân theo phân phối chuẩn với tham số $\mu = 0$ *và* $\sigma = 1$ *được gọi là biến ngẫu nhiên tuân theo phân phối chuẩn hóa, tức là hàm mật độ xác suất của nó có dạng:*

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

Kì vọng và phương sai của phân phối chuẩn

Mệnh đề

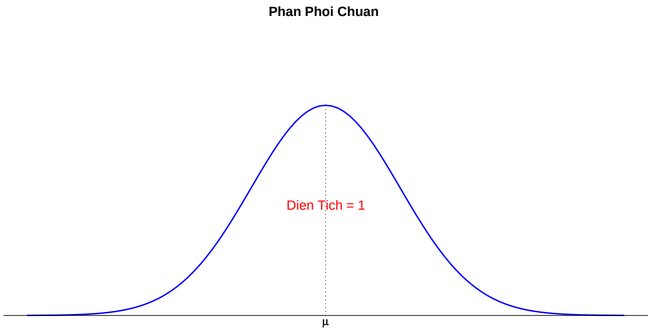
- *Nếu* $X \simeq N(\mu, \sigma^2)$ *thì* $E(X) = \mu$ *và* $V(X) = \sigma^2$.
- *Nếu* $Z \simeq N(0, 1)$ *thì* $E(Z) = 0$ *và* $V(Z) = 1$.

Một số đặc điểm của phân phối chuẩn

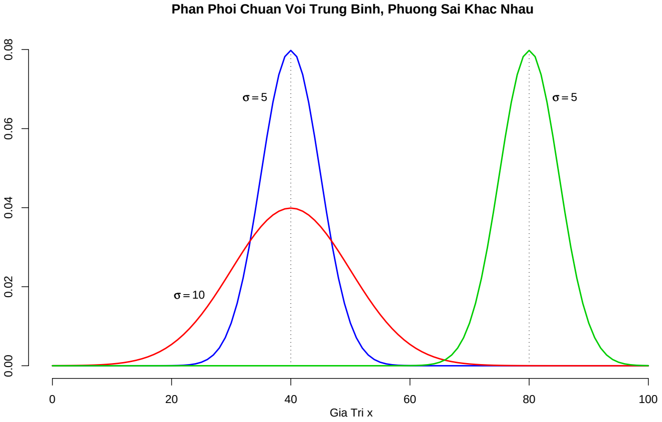
Hàm mật độ xác suất của biến ngẫu nhiên tuân theo phân phối chuẩn có một số đặc điểm quan trọng sau:

- Đường mật độ xác suất có dạng hình chuông và đối xứng xung quanh trung bình;
- Đường mật độ xác suất nhận trục ox làm tiệm cận ngang;
- Hàm số có một giá trị cực đại duy nhất bằng $\dfrac{1}{\sigma\sqrt{2\pi}}$ khi $x = \mu$;
- Diện tích của miền nằm dưới đường mật độ bằng 1.

Hình dáng phân phối chuẩn



Phân phối chuẩn với trung bình, phương sai khác nhau



Ứng dụng của qui luật chuẩn

- Phân phối chuẩn chiếm vị trí rất quan trọng trong lí thuyết xác suất, là vị trí trung tâm trong các kết luận thống kê sau này.
- Trong thực tế, nhiều biến ngẫu nhiên, nhiều qui luật tuân theo phân phối chuẩn, hoặc xấp xỉ chuẩn như trọng lượng và chiều cao của người lớn, độ thông minh của trẻ em, điểm thi của các thí sinh, khả năng chịu lực của thanh sắt, sai số đo đạc, sai số quan sát, độ bền dẻo của máy móc, trung bình cộng của một số lớn các đại lượng ngẫu nhiên độc lập...

Ứng dụng của qui luật chuẩn

- Phân phối chuẩn chiếm vị trí rất quan trọng trong lí thuyết xác suất, là vị trí trung tâm trong các kết luận thống kê sau này.
- Trong thực tế, nhiều biến ngẫu nhiên, nhiều qui luật tuân theo phân phối chuẩn, hoặc xấp xỉ chuẩn như trọng lượng và chiều cao của người lớn, độ thông minh của trẻ em, điểm thi của các thí sinh, khả năng chịu lực của thanh sắt, sai số đo đạc, sai số quan sát, độ bền dẻo của máy móc, trung bình cộng của một số lớn các đại lượng ngẫu nhiên độc lập...

Tính xác suất theo phân phối chuẩn

Mệnh đề

Nếu X là biến ngẫu nhiên tuân theo phân phối chuẩn thì $N(\mu, \sigma^2)$ thì $\dfrac{X-\mu}{\sigma}$ tuân theo phân phối chuẩn hóa $N(0,1)$.

Nhận xét: Mệnh đề trên cho phép ta chuyển cũng như tính xác suất của một phân phối chuẩn với trung bình và phương sai bất kì qua phân phối chuẩn hóa. Chẳng hạn, với $X \sim N(\mu, \sigma^2)$ và $Z \sim N(0,1)$, ta có

$$P(X < a) = P\left(\frac{X-\mu}{\sigma} < \frac{a-\mu}{\sigma}\right) = P\left(Z < \frac{a-\mu}{\sigma}\right).$$

Tính xác suất theo phân phối chuẩn

Mệnh đề

Nếu X là biến ngẫu nhiên tuân theo phân phối chuẩn thì $N(\mu, \sigma^2)$ thì $\dfrac{X-\mu}{\sigma}$ tuân theo phân phối chuẩn hóa $N(0,1)$.

Nhận xét: Mệnh đề trên cho phép ta chuyển cũng như tính xác suất của một phân phối chuẩn với trung bình và phương sai bất kì qua phân phối chuẩn hóa. Chẳng hạn, với $X \sim N(\mu, \sigma^2)$ và $Z \sim N(0,1)$, ta có

$$P(X < a) = P\left(\frac{X-\mu}{\sigma} < \frac{a-\mu}{\sigma}\right) = P\left(Z < \frac{a-\mu}{\sigma}\right).$$

Tính xác suất theo phân phối chuẩn

- Giả sử $X \sim N(\mu, \sigma^2)$ và $Z \sim N(0,1)$. Khi đó

$$P(X < a) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{a} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx.$$

$$P(X < a) = P(Z < \frac{a-\mu}{\sigma}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{a-\mu}{\sigma}} e^{-\frac{x^2}{2}} dx.$$

- Nếu gọi $F_X(x)$ và $F_Z(x)$ tương ứng là các hàm phân phối của biến ngẫu nhiên X và Z thì ta có:

$$P(X < a) = F_X(a) = F_Z(\frac{a-\mu}{\sigma}).$$

- Để tính $P(X < a)$ ta có thể dùng những hàm tính xác suất trong R hoặc chuyển qua phân phối chuẩn hóa để tra bảng phân phối xác suất của phân phối chuẩn hóa.

Tính xác suất theo phân phối chuẩn

- Giả sử $X \sim N(\mu, \sigma^2)$ và $Z \sim N(0, 1)$. Khi đó

$$P(X < a) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{a} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx.$$

$$P(X < a) = P(Z < \frac{a-\mu}{\sigma}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{a-\mu}{\sigma}} e^{-\frac{x^2}{2}} dx.$$

- Nếu gọi $F_X(x)$ và $F_Z(x)$ tương ứng là các hàm phân phối của biến ngẫu nhiên X và Z thì ta có:

$$P(X < a) = F_X(a) = F_Z(\frac{a-\mu}{\sigma}).$$

- Để tính $P(X < a)$ ta có thể dùng những hàm tính xác suất trong R hoặc chuyển qua phân phối chuẩn hóa để tra bảng phân phối xác suất của phân phối chuẩn hóa.

Tính xác suất theo phân phối chuẩn

- Giả sử $X \sim N(\mu, \sigma^2)$ và $Z \sim N(0, 1)$. Khi đó

$$P(X < a) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{a} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx.$$

$$P(X < a) = P(Z < \frac{a-\mu}{\sigma}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{a-\mu}{\sigma}} e^{-\frac{x^2}{2}} dx.$$

- Nếu gọi $F_X(x)$ và $F_Z(x)$ tương ứng là các hàm phân phối của biến ngẫu nhiên X và Z thì ta có:

$$P(X < a) = F_X(a) = F_Z(\frac{a-\mu}{\sigma}).$$

- Để tính $P(X < a)$ ta có thể dùng những hàm tính xác suất trong R hoặc chuyển qua phân phối chuẩn hóa để tra bảng phân phối xác suất của phân phối chuẩn hóa.

Tính toán liên quan đến phân phối chuẩn trong R

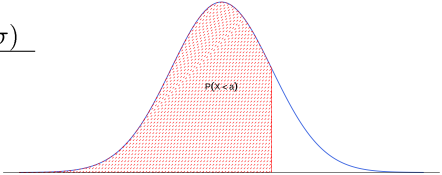
Cho biến ngẫu nhiên $X \backsim N(\mu, \sigma^2)$ có hàm mật độ xác suất là $f(x)$, khi đó

- `dnorm(x,` μ`,` σ`)` cho giá trị của hàm mật độ $f(x)$ tại x.
- `pnorm(x,` μ`,` σ`, lower.tail = TRUE (FALSE))` cho giá trị $P(X \leqslant x)$ (P(X > x)).
- `qnorm(p,` μ`,` σ`, lower.tail = TRUE (FALSE))` cho giá trị là x sao cho `pnorm(x,` μ`,` σ`, lower.tail = TRUE (FALSE)) = p`.
- `rnorm(p,` μ`,` σ`)` cho mẫu ngẫu nhiên cỡ n tuân theo phân phối chuẩn $N(\mu, \sigma^2)$.

Minh họa khoảng xác suất của phân phối chuẩn

Cho $X \sim N(\mu, \sigma^2)$. Khi đó

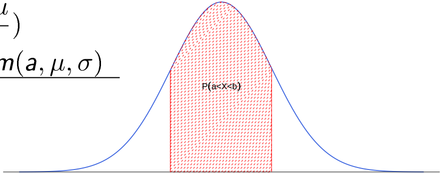
$$
\begin{aligned}
P(X \leqslant a) &= F_X(a) \\
&= F_Z(\frac{b-\mu}{\sigma}) \\
&= pnorm(a, \mu, \sigma)
\end{aligned}
$$



Minh họa khoảng xác suất của phân phối chuẩn

Cho $X \sim N(\mu, \sigma^2)$. Khi đó

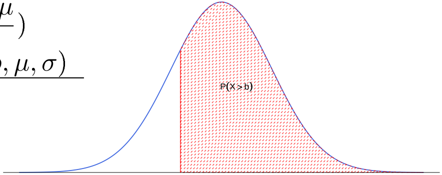
$$
\begin{aligned}
&P(a \leqslant X \leqslant b) \\
&= F_X(b) - F_X(a) \\
&= F_Z\left(\frac{b-\mu}{\sigma}\right) - F_Z\left(\frac{a-\mu}{\sigma}\right) \\
&= pnorm(b, \mu, \sigma) - pnorm(a, \mu, \sigma)
\end{aligned}
$$



Minh họa khoảng xác suất của phân phối chuẩn

Cho $X \sim N(\mu, \sigma^2)$. Khi đó

$$\begin{aligned} P(X > b) &= 1 - P(X \leqslant b) \\ &= 1 - F_X(b) \\ &= 1 - F_Z(\frac{b-\mu}{\sigma}) \\ &= 1 - pnorm(b, \mu, \sigma) \end{aligned}$$



Ví dụ

Bài toán

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

a. *Một người được cho là bình thường nếu có chỉ số IQ thuộc đoạn* $[85, 115]$. *Tìm tỉ lệ những người được coi là bình thường trên thế giới.*
b. *Một người được coi là thiên tài nếu có chỉ số IQ cao hơn 160. Tính tỉ lệ những người được coi là thiên tài trên thế giới.*
c. *Hãy tính xem chỉ số IQ thấp nhất trong nhóm 1/1 triệu có chỉ số IQ cao nhất thế giới?*

Lời giải

Tính toán phân phối chuẩn trên R

Giả sử $X \sim N(100, 15^2)$. Khi đó

a. Xác suất $P(85 \leqslant X \leqslant 115)$ được tính như sau:

```
> pnorm(115, mean = 100, sd = 15)
- pnorm(85, mean = 100, sd = 15)
[1] 0.6826895
> # Hoặc
> pnorm(115, 100, 15) - pnorm(85, 100, 15)
[1] 0.6826895
```

- Xác suất $P(X > 160)$ được tính như sau:

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

- Giá trị x_1 để $P(X > x_1) = 10^{-6}$ được tính như sau:

```
> qnorm(1-0.000001, 100, 15)
[1] 171.3014
> # Hoặc
> qnorm(0.000001, 100, 15,lower.tail=F)
[1] 171.3014
```

Bài tập tự học

Bài toán

a. *Một dân tộc có IQ tuân theo phân phối chuẩn với trung bình là 110, độ lệch chuẩn là 15. Tìm tỉ lệ những người được coi là thiên tài (IQ cao hơn 160) của dân tộc này.*

b. *Hãy tính xem chỉ số IQ thấp nhất trong nhóm 1/1 tỉ người có chỉ số IQ cao nhất thế giới? (IQ con người tuân theo phân phối chuẩn với trung bình là 100, độ lệch chuẩn là 15).*

Nội dung trình bày

1. Phân phối lí thuyết rời rạc
   - Phân phối nhị thức
   - Phân phối Poisson

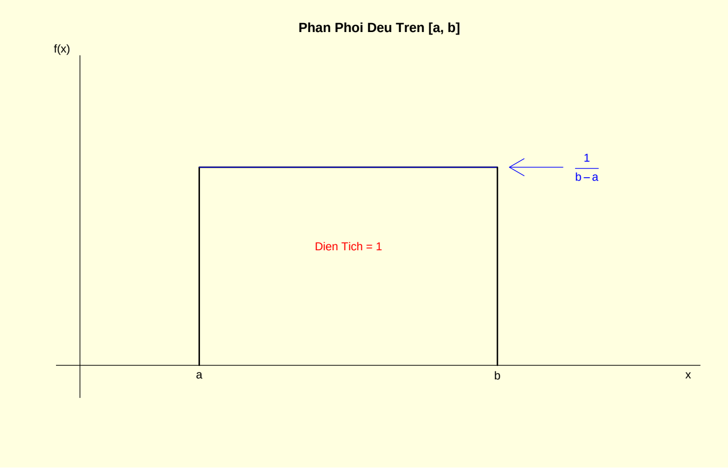
2. Phân phối lí thuyết liên tục
   - Phân phối chuẩn
   - Phân phối đều
   - Phân phối mũ

Định nghĩa

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

$$f(x) = \begin{cases} 0 & \text{nếu } x < a \text{ hoặc } x > b, \\ \dfrac{1}{b-a} & \text{nếu } a \leqslant x \leqslant b. \end{cases}$$

Hình dạng phân phối đều



Kì vọng và phương sai của phân phối đều

Mệnh đề

Giả sử X là biến ngẫu nhiên tuân theo phân phối đều trên $[a, b]$*, khi đó*
$E(X) = \dfrac{a+b}{2}, V(X) = \dfrac{(b-a)^2}{12}.$

Ví dụ

Bài toán

Khi thâm nhập vào một thị trường mới, doanh nghiệp không thể khẳng định được một cách chắc chắn doanh số hàng tháng có thể đạt được sẽ là bao nhiêu mà chỉ dự kiến được rằng doanh số tối thiểu sẽ là 20 triệu/tháng và tối đa là 40 triệu/tháng. Tìm xác suất để doanh nghiệp đạt được doanh số tối thiểu là 35 triệu/tháng.

Lời giải

Tính toán liên quan đến phân phối đều trong R

Cho biến ngẫu nhiên $X \backsim U([a, b])$. Khi đó

- `dunif(x, a, b)` cho giá trị của hàm mật độ $f(x)$ tại x.
- `punif(x, a, b, lower.tail=TRUE(FALSE))` cho giá trị $P(X \leqslant x)(P(X > x))$.
- `qunif(p, a, b, lower.tail=TRUE(FALSE))`: cho giá trị là x sao cho `punif(x, a, b, lower.tail=TRUE(FALSE))=p`.
- `runif(n, a, b)`: cho ta mẫu ngẫu nhiên cỡ n tuân theo phân phối đều $U([a, b])$.

Tính toán liên quan đến phân phối đều trong R

Giả sử $X \backsim U([20, 40])$. Ta có thể tính $P(X > 35) = 1 - P(X \leqslant 35)$ như sau:

```
> 1 - punif(35, 20, 40)
[1] 0.25
> # Hoặc
> punif(35, 20, 40, lower.tail = F)
[1] 0.25
```

Nội dung trình bày

1. Phân phối lí thuyết rời rạc
   - Phân phối nhị thức
   - Phân phối Poisson

2. Phân phối lí thuyết liên tục
   - Phân phối chuẩn
   - Phân phối đều
   - Phân phối mũ

Định nghĩa

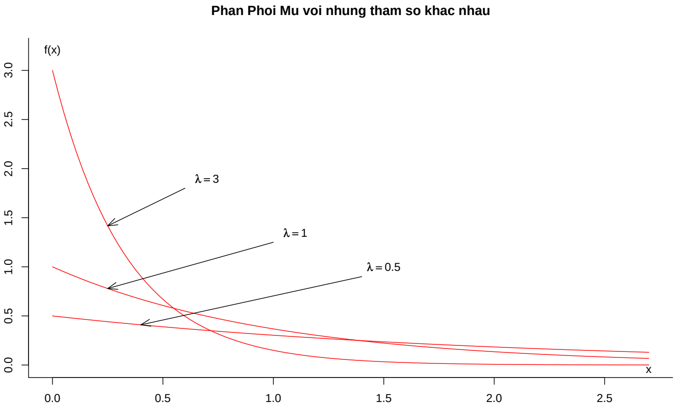
Biến ngẫu nhiên liên tục X được gọi là tuân theo phân phối mũ với tham số λ ($\lambda > 0$), kí hiệu $X \sim exp(\lambda)$ nếu hàm mật độ xác suất của nó có dạng:

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

Qui luật phân phối mũ có ứng dụng trong nhiều lĩnh vực khác nhau. Người ta có thể chứng minh rằng các khoảng thời gian sau tuân theo phân phối mũ:

- Thời gian nhân viên ngân hàng phục vụ một khách đến vay hay rút tiền; thời gian một cô bán sách phục vụ một người đến mua sách;...
- Khoảng thời gian giữa hai lần có sự cố của một cái máy; khoảng thời gian giữa hai lần nhận được điện thoại;...

Minh họa phân phối mũ



Kì vọng và phương sai của phân phối mũ

Mệnh đề

Giả sử X là biến ngẫu nhiên tuân theo phân phối mũ với tham số λ, khi đó

a. *Kì vọng của X:* $E(X) = \dfrac{1}{\lambda}$ *và phương sai của X:* $V(X) = \dfrac{1}{\lambda^2}$.

b. *Hàm phân phối* $F_X(x) = 0, x < 0$ *và* $F_X(x) = 1 - e^{-\lambda x}, x \geqslant 0$.

c. *Ta có:*
   - $P(X < a) = 1 - e^{-\lambda a}, a > 0$.
   - $P(a < X < b) = F_X(b) - F_X(a) = e^{-\lambda a} - e^{-\lambda b}$, $0 < a < b$.
   - $P(X > a) = e^{-\lambda a}$.

Tính chất không có trí nhớ của phân phối mũ

Mệnh đề

Giả sử X là biến ngẫu nhiên tuân theo phân phối mũ với tham số λ, khi đó

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

Tính chất trên được gọi là tính chất không có trí nhớ của phân phối mũ.

Nhận xét: Nếu ta coi X là thời gian sống của một vật nào đó thì tính chất không có trí nhớ chỉ ra rằng khả năng để X sống lâu hơn $t + s$ biết rằng nó đã sống lâu hơn t cũng giống như khả năng X sống lâu hơn s tại thời điểm đầu tiên. Như vậy là nếu X còn sống đến thời điểm t thì nó như mới. Do đó có một vài trường hợp tự nhiên có thể minh họa cho phân phối mũ là tuổi thọ của một sản phẩm chất lượng cao hay thời gian sống của một người lính trong chiến tranh.

Tính chất không có trí nhớ của phân phối mũ

Mệnh đề

Giả sử X là biến ngẫu nhiên tuân theo phân phối mũ với tham số λ, khi đó

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

Tính chất trên được gọi là tính chất không có trí nhớ của phân phối mũ.

Nhận xét: Nếu ta coi X là thời gian sống của một vật nào đó thì tính chất không có trí nhớ chỉ ra rằng khả năng để X sống lâu hơn $t + s$ biết rằng nó đã sống lâu hơn t cũng giống như khả năng X sống lâu hơn s tại thời điểm đầu tiên. Như vậy là nếu X còn sống đến thời điểm t thì nó như mới. Do đó có một vài trường hợp tự nhiên có thể minh họa cho phân phối mũ là tuổi thọ của một sản phẩm chất lượng cao hay thời gian sống của một người lính trong chiến tranh.

Ví dụ

Bài toán

Theo những tìm hiểu của John thì tổng số dặm (nghìn dặm) mà một chiếc ô tô đi được cho đến khi nó không sử dụng được nữa tuân theo phân phối mũ với tham số $\lambda = 1/20$. *Smith đang sử dụng một chiếc ô tô và khẳng định rằng nó mới đi được* 10000 *dặm.*

a. *Tổng số dặm đi được trung bình của một chiếc ô tô đi được cho đến khi không sử dụng được nữa là bao nhiêu?*

b. *Nếu John mua một chiếc xe mới thì khả năng chiếc xe này đi được ít nhất* 20000 *dặm là bao nhiêu?*

c. *Nếu John mua xe của Smith thì khả năng chiếc xe này đi được thêm ít nhất* 20000 *dặm là bao nhiêu?*

Lời giải

Tính toán liên quan đến phân phối mũ trong R

Cho $X \sim exp(\lambda)$, khi đó

- `dexp(x,` λ`)` cho giá trị của hàm mật độ $f(x)$ tại x.
- `pexp(x,` λ`, lower.tail=TRUE(FALSE))` cho giá trị $P(X \leqslant x)(P(X > x))$.
- `qexp(p,` λ`, lower.tail=TRUE(FALSE))` cho giá trị là x sao cho `pexp(x,` λ`, lower.tail=TRUE(FALSE))=p`.
- `rexp(n,` λ`)` cho mẫu ngẫu nhiên cỡ n tuân theo phân phối mũ.

Tính toán phân phối mũ trong R

Khi X tuân theo phân phối mũ với $\lambda = 1/20$. Để tính $P(X \geqslant 20) = 1 - P(X < 20)$, ta thực hiện như sau:

```
> 1-pexp(20,1/20)
[1] 0.3678794
```

Mối liên hệ giữa phân phối mũ và phân phối Poisson

Nếu số lần xuất hiện của một biến cố trong một khoảng thời gian cho trước tuân theo một phân phối Poisson với trung bình a thì khoảng thời gian giữa hai lần xuất hiện liên tiếp của biến cố ấy tuân theo phân phối mũ với trung bình $\frac{1}{a}$.

Bài toán

Giả sử số khách đến rút tiền tại một máy rút tiền tự động ATM tuân theo phân phối Poisson với trung bình 20 người trong một giờ. Giả sử một khách hàng vừa đến, tính xác suất để khách hàng tiếp theo sẽ đến trong vòng 6 phút.

Mối liên hệ giữa phân phối mũ và phân phối Poisson

- Gọi X là biến ngẫu nhiên chỉ thời gian rút tiền giữa hai khách hàng liên tiếp. Do số người đến rút tiền trong một giờ tuân theo phân phối Poisson với trung bình bằng 20 nên X là biến ngẫu nhiên tuân theo phân phối mũ với trung bình $\frac{1}{a} = \frac{1}{20} = 0.05$ giờ hay X tuân theo phân phối mũ với tham số $\lambda = 20$ giờ.
- Xác suất để khách hàng tiếp theo đến trong vòng 6 phút hay 0.1 giờ là
$$P(X < 0.1) = F_X(0.1) = 1 - e^{-0.1\times 20} = 0.8647.$$