Chương 2: Biến ngẫu nhiên và luật phân phối xác suất

Lê Xuân Lý [1]

Viện Toán ứng dụng và Tin học, ĐHBK Hà Nội

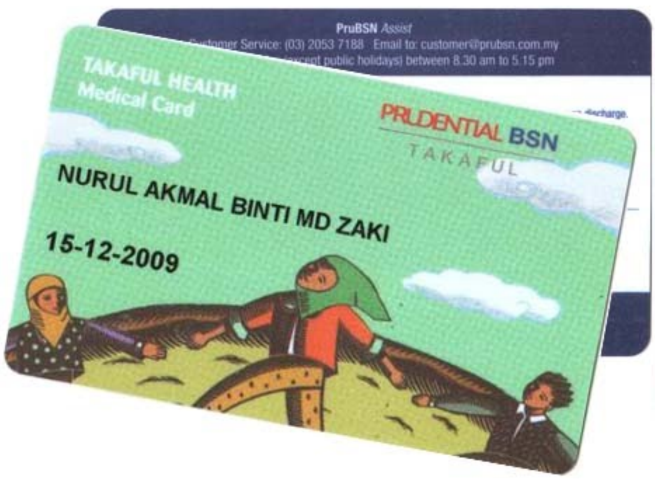
Hà Nội, tháng 12 năm 2014

[1] Email: lexuanly@gmail.com

Ví dụ

Một công ty bảo hiểm bán thẻ bảo hiểm với giá 100 ngàn đồng/1 người/1 năm. Nếu người bảo hiểm gặp rủi ro trong năm đó thì nhận được số tiền bồi thường là 1 triệu đồng. Theo thống kê biết rằng tỷ lệ người tham gia bảo hiểm bị rủi ro trong năm là 0.05.

- Hãy tính tiền lãi trung bình khi bán mỗi thẻ bảo hiểm
- Nếu bán bảo hiểm được cho 10000 khách hàng thì số tiền lãi trung bình thu về được là bao nhiêu?



Định nghĩa 1.1

Biến ngẫu nhiên (đại lượng ngẫu nhiên) là một đại lượng mà giá trị của nó là ngẫu nhiên, phụ thuộc vào kết quả phép thử. Ta thường dùng các chữ in hoa để kí hiệu biến ngẫu nhiên: $X, Y, Z, X_1, X_2, \ldots$. *Còn các giá trị mà biến ngẫu nhiên nhận thường được kí hiệu là chữ thường:* $a, b, c, \ldots, x, y, z, x_1, x_2, \ldots$.

Ví dụ 1

Biến ngẫu nhiên

- Gieo một con xúc xắc. Ta quan tâm đến số chấm xuất hiện. Gọi X là số chấm xuất hiện trên mặt con xúc xắc, ta có X là một biến ngẫu nhiên và tập giá trị có thể nhận là $\{1, 2, 3, 4, 5, 6\}$.
- Chọn ngẫu nhiên 3 đứa trẻ từ một nhóm gồm 6 bé trai và 4 bé gái. Ta quan tâm có bao nhiêu bé gái. Gọi X là số bé gái trong nhóm. Khi đó X là một biến ngẫu nhiên và tập giá trị có thể nhận là $\{0, 1, 2, 3\}$.
- Khoảng thời gian giữa 2 ca cấp cứu ở một bệnh viện nào đó là một biến ngẫu nhiên. Nó có thể nhận giá trị bất kỳ trong khoảng $[0; +\infty)$.
- Nhiệt độ của Hà Nội lúc 6h sáng hàng ngày
- Số iphone phải đi bảo hành
- ...

Phân loại biến ngẫu nhiên

Ta chỉ xét biến ngẫu nhiên ở hai dạng cơ bản sau:

- *Biến ngẫu nhiên được gọi là rời rạc, nếu tập giá trị của nó là một tập hữu hạn hoặc vô hạn đếm được các phần tử.* Nói một cách khác đối với biến ngẫu nhiên rời rạc ta có thể liệt kê tất cả các giá trị nó có thể nhận bằng một dãy hữu hạn hoặc vô hạn. Ví dụ: số điểm thi của học sinh, số cuộc gọi điện thoại của một tổng đài trong một đơn vị thời gian, số tai nạn giao thông trong một ngày, . . .
- *Biến ngẫu nhiên được gọi là liên tục, nếu tập giá trị của nó lấp kín một khoảng hoặc một số khoảng của trục số hoặc cũng có thể là cả trục số.* Ví dụ: huyết áp của một bệnh nhân, độ dài của một chi tiết máy, tuổi thọ của một loại bóng đèn điện tử,. . . Miền giá trị của một biến ngẫu nhiên liên tục sẽ gồm một số miền dạng $(a; b), [a; b), (a; b], [a; b]$ hoặc cả $\mathbb{R}$.

Biến ngẫu nhiên rời rạc

Biến ngẫu nhiên liên tục

Hàm phân phối xác suất

Định nghĩa 1.2

Hàm phân phối xác suất của biến ngẫu nhiên X, kí hiệu là $F(x)$ và được xác định như sau:

$$\boxed{F(x) = P(X < x), \quad x \in \mathbb{R}.} \tag{1.1}$$

Hàm phân phối xác suất $F(x)$ phản ánh độ tập trung xác suất ở bên trái của điểm x.

Các tính chất

- $0 \le F(x) \le 1$
- $\lim\limits_{x\to-\infty} F(x) = 0$, $\lim\limits_{x\to+\infty} F(x) = 1$
- $F(x)$ là hàm không giảm: $\forall a < b, F(a) \le F(b)$
- $P(a \le X < b) = F(b) - F(a)$

Bảng phân phối xác suất

Định nghĩa 2.1

Phân bố xác suất của một biến ngẫu nhiên rời rạc X là một bảng trên đó ta ghi cả giá trị mà X có thể nhận kèm theo xác suất để nó nhận các giá trị đó

$X = x$	x_1	x_2	$\ldots$	x_n	$\ldots$
$P(X = x)$	p_1	p_2	$\ldots$	p_n	$\ldots$

Trong đó tập các giá trị của X là $\{x_1, x_2, \ldots, x_n\}$ được sắp xếp theo thứ tự tăng dần. Các xác suất p_i thỏa mãn

- $p_i = P(X = x_i) > 0 \quad \forall i = 1, 2, \ldots;$
- $\sum\limits_i p_i = 1.$
- *Hàm phân phối xác suất của biến ngẫu nhiên rời rạc X :*
  $F(x) = P(X < x) = \sum\limits_{i:x_i<x} P(X = x_i) = \sum\limits_{i:x_i<x} p_i$

Bảng phân phối xác suất

Câu hỏi: Để lập được bảng phân phối xác suất ta cần làm gì ?
Trả lời:

- Xác định các giá trị x_i mà X có thể nhận
- Tìm các xác suất p_i tương ứng với các giá trị x_i

Bảng phân phối xác suất

Ví dụ 1

Tung một đồng tiền cân đối và đồng chất. Gọi X là biến ngẫu nhiên chỉ số lần xuất hiện mặt sấp. Ta có bảng phân phối xác suất sau:

$X = x$	0	1
$P(X = x)$	$1/2$	$1/2$

Bảng phân phối xác suất

Ví dụ 1

Tung đồng xu cân đối và đồng chất 2 lần. Gọi X là biến ngẫu nhiên chỉ số lần xuất hiện mặt sấp. Ta có bảng phân phối xác suất sau:

$X = x$	0	1	2
$P(X = x)$	$1/4$	$1/2$	$1/4$

Bảng phân phối xác suất

Ví dụ 2

Một người đem 10 nghìn đồng đi đánh một số đề. Nếu trúng thì thu được 700 nghìn đồng, nếu trượt thì không được gì. Gọi X (nghìn đồng) là số tiền thu được. Ta có bảng phân phối xác suất của X

$X = x$	0	700
$P(X = x)$	$^{99}/_{100}$	$^{1}/_{100}$

Các tham số đặc trưng

Kỳ vọng

Kỳ vọng

- *Kỳ vọng*: là đại lượng đặc trưng cho giá trị trung bình.
  (Đôi khi người ta có thể gọi nó là giá trị trung bình bởi công thức tính của nó chính là tính giá trị trung bình cho trường hợp thu được vô hạn số liệu)
- *Ký hiệu*: $E(X)$ hoặc EX
- *Công thức tính*: với X rời rạc ta có: $EX = \sum\limits_i x_i.p_i$

Các tham số đặc trưng

Kỳ vọng

Ví dụ 1

Tung một đồng tiền cân đối và đồng chất. Gọi X là biến ngẫu nhiên chỉ số lần xuất hiện mặt sấp. Ta có bảng phân phối xác suất sau:

$X = x$	0	1
$P(X = x)$	$1/2$	$1/2$

Kỳ vọng của X : $EX = 0.1/2 + 1.1/2 = 1/2$

Các tham số đặc trưng

Kỳ vọng

Ví dụ 2

Tung đồng xu cân đối và đồng chất 2 lần. Gọi X là biến ngẫu nhiên chỉ số lần xuất hiện mặt sấp. Ta có bảng phân phối xác suất sau:

$X = x$	0	1	2
$P(X = x)$	$1/4$	$1/2$	$1/4$

Kỳ vọng của X : $EX = 0.1/4 + 1.1/2 + 2.1/4 = 1$
Như vậy trong 2 lần tung đồng xu thì trung bình có một lần ra mặt sấp.

Các tham số đặc trưng
Kỳ vọng

Ví dụ 3

Một người đem 10 nghìn đồng đi đánh một số đề. Nếu trúng thì thu được 700 nghìn đồng, nếu trượt thì không được gì. Gọi X (nghìn đồng) là số tiền thu được. Ta có bảng phân phối xác suất của X

$X = x$	0	700
$P(X = x)$	$^{99}/_{100}$	$^{1}/_{100}$

Kỳ vọng của X : $EX = 0.^{99}/_{100} + 700.^{1}/_{100} = 7$
Như vậy bỏ ra 10 nghìn đồng, trung bình thu được 7 nghìn đồng, người chơi về lâu dài sẽ lỗ 30% tổng số tiền chơi.

Các tham số đặc trưng
Kỳ vọng

Các tính chất của kỳ vọng

- $Ec = c$ với c là hằng số
- $E(aX) = a.EX$
- $E(X + b) = EX + b$
  Ta suy ra kết quả: $E(aX + b) = aEX + b$
- Tổng quát với X là biến ngẫu nhiên rời rạc: $Eg(X) = \sum_i g(x_i).p_i$
  Ví dụ: $E(X^2) = \sum_i x_i^2.p_i$
- $E(X + Y) = EX + EY$

Các tham số đặc trưng

Phương sai

Phương sai

- *Phương sai*: trung bình của bình **phương sai** số.
- *Ký hiệu*: $V(X)$ hoặc VX
- *Công thức tính*: $VX = E(X - EX)^2$
  Với $(X - EX)$ là sai số, hoặc là độ lệch khỏi giá trị trung bình
  Người ta biến đổi để đưa công thức tính phương sai về dễ tính hơn:

$$\boxed{VX = E(X - EX)^2 = E(X^2) - (EX)^2}$$

Với X là biến ngẫu nhiên rời rạc:

$EX = \sum\limits_{i=1}^{n} x_i.p_i$

$E(X^2) = \sum\limits_{i=1}^{n} x_i^2.p_i$

Ý nghĩa của phương sai

- Phương sai thể hiện mức độ phân tán dữ liệu xung quanh giá trị trung bình EX, phương sai càng lớn thì độ phân tán dữ liệu càng cao và ngược lại.
- Trong công nghiệp, X thường là kích cỡ của các sản phẩm. VX lúc này biểu thị độ chính xác của các sản phẩm.
- Trong chăn nuôi, X thường là chiều cao hay cân nặng của gia súc gia cầm. VX lúc này biểu thị độ tăng trưởng đồng đều của các gia súc gia cầm.
- Trong trồng trọt, X thường là năng suất của giống cây trồng. VX lúc này biểu thị mức độ ổn định của năng suất giống cây trồng.
- Trong kinh tế, X thường là lãi suất thu được của khoản đầu tư. VX lúc này sẽ biểu thị cho mức độ rủi ro của đầu tư.

Các tham số đặc trưng

Phương sai

Ví dụ 1

Tung một đồng tiền cân đối và đồng chất. Gọi X là biến ngẫu nhiên chỉ số lần xuất hiện mặt sấp. Ta có bảng phân phối xác suất sau:

$X = x$	0	1
$P(X = x)$	$1/2$	$1/2$

$EX = 0.1/2 + 1.1/2 = 1/2$
$E(X^2) = 0^2.1/2 + 1^2.1/2 = 1/2$
Phương sai $VX = E(X^2) - (EX)^2 = 1/2 - 1/4 = 1/4$

Biến ngẫu nhiên rời rạc | Các tham số đặc trưng

Các tham số đặc trưng

Phương sai

Ví dụ 2

Tung đồng xu cân đối và đồng chất 2 lần. Gọi X là biến ngẫu nhiên chỉ số lần xuất hiện mặt sấp. Ta có bảng phân phối xác suất sau:

$X = x$	0	1	2
$P(X = x)$	$1/4$	$1/2$	$1/4$

$EX = 0.1/4 + 1.1/2 + 2.1/4 = 1$
$E(X^2) = 0^2.1/4 + 1^2.1/2 + 2^2.1/4 = 3/2$
Phương sai $VX = E(X^2) - (EX)^2 = 3/2 - 1^2 = 1/2$
Nhận xét: Phương sai của VD2 lớn hơn phương sai của VD1 cho ta kết luận rằng biên độ dao động của X xung quanh giá trị trung bình ở VD2 lớn hơn VD1.

Các tham số đặc trưng

Phương sai

Các tính chất của phương sai

- $Vc = 0$ với c là hằng số
- $V(aX) = a^2.VX$
- $V(X + b) = VX$
  Ta suy ra kết quả: $V(aX + b) = a^2 VX$

Các tham số đặc trưng

Độ lệch chuẩn

Đơn vị đo của phương sai bằng bình phương đơn vị đo của biến ngẫu nhiên. Để dễ đánh giá mức độ phân tán hơn, người ta đưa ra khái niệm độ lệch chuẩn.

Độ lệch chuẩn

- *Ý nghĩa*: dùng để đo độ phân tán dữ liệu xung quanh giá trị trung bình EX.
- *Ký hiệu*: $\sigma(X)$ hoặc σ
- *Công thức tính*: $\sigma = \sqrt{VX}$

Ví dụ: Phân tích kỹ thuật giá chứng khoán: SMA(n) và Bollinger Band(n).

Các tham số đặc trưng

Mode

Mode

- *Khái niệm*: Mode của biến ngẫu nhiên X, kí hiệu là $mod(X)$, là giá trị của biến ngẫu nhiên X có khả năng xuất hiện lớn nhất trong một lân cận nào đó của nó. Đối với biến ngẫu nhiên rời rạc, mod(X) là giá trị của X ứng với xác suất lớn nhất. Như vậy một biến ngẫu nhiên có thể có một mode hoặc nhiều mode.
- *Ký hiệu*: $mod(X)$

Các tham số đặc trưng

Phân vị mức p

- *Khái niệm*: Phân vị mức p của biến ngẫu nhiên X là giá trị z_p sao cho.

$$F(z_p) = P(X < z_p) = p$$

Một số phân vị đặc biệt:
+ Phân vị mức 25% được gọi là tứ phân vị thứ nhất
+ Phân vị mức 50% được gọi là tứ phân vị thứ hai hay trung vị.
+ Phân vị mức 75% được gọi là tứ phân vị thứ ba

- *Trung vị*: Trung vị của biến ngẫu nhiên X là giá trị của X chia phân phối xác suất thành hai phần có xác suất bằng nhau. Kí hiệu là $med(X)$:

$$P(X < med(X) = P(X \geq med(X)) = 0,5$$

Ta có thể tìm trung vị bằng cách giải phương trình: $F(x) = 0,5$.
Trong ứng dụng, trung vị là đặc trưng vị trí tốt nhất, nhiều khi tốt hơn cả kỳ vọng, nhất là trong những trường hợp số liệu có nhiều sai sót hoặc sai sót thái quá.

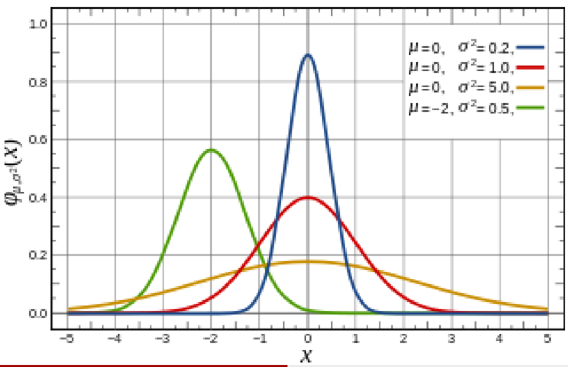
Hàm mật độ xác suất

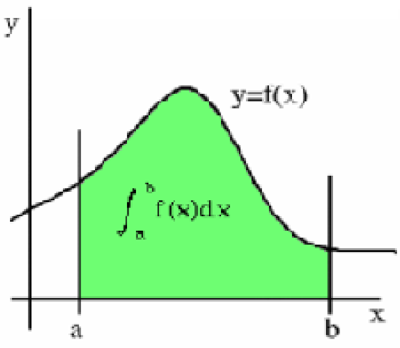
Đối với biến ngẫu nhiên liên tục, không thể dùng bảng phân phối xác suất do xác suất nó nhận tại mỗi điểm luôn bằng "0". Do đó người ta thay thế bằng hàm mật độ xác suất.

Định nghĩa 3.1

Hàm mật độ xác suất của biến ngẫu nhiên liên tục X là hàm $f(x)$ xác định trên $\mathbb{R}$ thỏa mãn:

- $f(x) \geq 0 \ \forall x \in \mathbb{R}$;
- $P(X \in B) = \displaystyle\int_B f(x)dx \quad \forall B \subset \mathbb{R}$.





Hàm mật độ xác suất

Chú ý 3.1

Hàm mật độ xác suất $f(x)$ của biến ngẫu nhiên liên tục X thể hiện mức độ tập trung xác suất của X xung quanh điểm x. Tức là với Δ_x đủ nhỏ cho trước ta có thể tính xấp xỉ:

$$P\left(x \leq X \leq x + \Delta_x\right) \approx f(x).\Delta_x.$$

Do đó ta thấy xác suất để X nhận giá trị thuộc lân cận khá bé $(x, x + \Delta_x)$ gần như tỉ lệ thuận với $f(x)$.

Hàm mật độ xác suất

Tính chất

- $\int\limits_{-\infty}^{+\infty} f(x)dx = 1;$
- $P(a \le X < b) = P(a < X < b) = P(a < X \le b) = P(a \le X \le b) = \int\limits_{a}^{b} f(x)dx$
- Hàm phân phối xác suất: $F(x) = P(X < x) = \int\limits_{-\infty}^{x} f(t)dt$

  Từ đó suy ra $f(x) = F'(x)$

Hàm mật độ xác suất

Ví dụ 3

Cho hàm số $f(x) = a.\sin 2x$. Tìm a để hàm này trở thành hàm mật độ xác suất của một biến ngẫu nhiên nhận giá trị trong $[0, \pi/2]$.

Lời giải

Để hàm này trở thành hàm mật độ xác suất của một biến ngẫu nhiên nhận giá trị trong $[0, \pi/2]$ thì:

$$f(x) = \begin{cases} a\sin 2x, & x \in [0, \pi/2] \\ 0, & x \notin [0, \pi/2] \end{cases}.$$

Do $\sin 2x \ge 0$ với mọi $x \in [0, \pi/2]$ nên $a \ge 0$. Ta có:

$1 = \int_{-\infty}^{+\infty} f(x)dx = \int_{0}^{\pi/2} a\sin 2x dx = a$. Vậy $a = 1$.

Hàm mật độ xác suất

Ví dụ 4

Tuổi thọ của một loài côn trùng là biến ngẫu nhiên X (tháng tuổi) có hàm mật độ xác suất $f(x) = \begin{cases} ax^2(4-x^2), & x \in [0,2] \\ 0, & x \notin [0,2] \end{cases}.$

a. *Xác định* a

b. *Tính* $P(0 \le X \le 1), P(X > 1)$

c. *Xác định hàm phân phối xác suất* $F(x)$

Lời giải

a. Do $ax^2(4-x^2) \ge 0$ với $\forall x \in [0,2]$ nên $a \ge 0$

Ta có $1 = \int\limits_{-\infty}^{+\infty} f(x)dx = \int\limits_0^2 ax^2(4-x^2)dx = a.\frac{64}{15} \Rightarrow a = \frac{15}{64}$

b. $P(0 \le X \le 1) = \int\limits_0^1 f(x)dx = \int\limits_0^1 ax^2(4-x^2)dx = a.\frac{17}{15} = \frac{17}{64} = 0,266$

Hàm mật độ xác suất

Lời giải

b. $P(X > 1) = \int\limits_1^{+\infty} f(x)dx = \int\limits_1^2 ax^2(4-x^2)dx = \frac{47}{64} = 0,734$

c. Hàm phân phối $F(x) = \int\limits_{-\infty}^{x} f(t)dt$

$x < 0$ suy ra $F(x) = \int\limits_{-\infty}^{x} f(t)dt = \int\limits_{-\infty}^{x} 0dt = 0$

$0 \le x \le 2$ suy ra $F(x) = \int\limits_{-\infty}^{x} f(t)dt = \int\limits_0^x at^2(4-t^2)dt = \frac{15}{64}(\frac{4x^3}{3} - \frac{x^5}{5})$

$x > 2$ suy ra $F(x) = \int\limits_{-\infty}^{x} f(t)dt = \int\limits_0^2 at^2(4-t^2)dt = 1$

Hàm mật độ xác suất

Nhận xét

Qua tính toán trên ta thấy 26.6% côn trùng sống không quá một tháng tuổi, và 73,4% côn trùng sống hơn một tháng tuổi. Do đó ta có thể nhận xét rằng tuổi thọ trung bình của loài này sẽ lớn hơn một tháng tuổi. Tuy nhiên tuổi thọ trung bình của loài côn trùng này chính xác là bao nhiêu?

Các tham số đặc trưng

Kỳ vọng

Kỳ vọng của biến ngẫu nhiên liên tục X

- *Ý nghĩa*: nó đặc trưng cho giá trị trung bình của X
- *Ký hiệu*: $E(X)$ hoặc EX
- *Công thức tính*: $EX = \int\limits_{-\infty}^{+\infty} x.f(x)dx$
- *Tính chất*:
  $+\ E(aX+b) = a.EX + b$
  $+\ Eg(X) = \int\limits_{-\infty}^{+\infty} g(x).f(x)dx$

  Ví dụ: $g(X) = X^2$ ta có $E(X^2) = \int\limits_{-\infty}^{+\infty} x^2.f(x)dx$

Các tham số đặc trưng
Phương sai

Phương sai của biến ngẫu nhiên liên tục X

- *Ý nghĩa*: nó đặc trưng cho độ phân tán dữ liệu xung quanh EX
- *Ký hiệu*: $V(X)$ hoặc VX
- *Công thức tính*: $\boxed{VX = E(X - EX)^2 = E(X^2) - (EX)^2}$

  với: $EX = \int\limits_{-\infty}^{+\infty} x.f(x)dx$ và $E(X^2) = \int\limits_{-\infty}^{+\infty} x^2.f(x)dx$
- *Tính chất*: $V(aX + b) = a^2 VX$

Các tham số đặc trưng
Độ lệch chuẩn

Độ lệch chuẩn

- *Ý nghĩa*: dùng để đo độ phân tán dữ liệu xung quanh giá trị trung bình EX.
- *Ký hiệu*: $\sigma(X)$ hoặc σ
- *Công thức tính*: $\sigma = \sqrt{VX} = \sqrt{E(X^2) - (EX)^2}$

  với X liên tục: $EX = \int\limits_{-\infty}^{+\infty} xf(x)dx$

  $E(X^2) = \int\limits_{-\infty}^{+\infty} x^2 f(x)dx$

Các tham số đặc trưng

Mode - phân vị mức p

Mode

- *Khái niệm*: Mode của biến ngẫu nhiên X, kí hiệu là $mod(X)$, là giá trị của biến ngẫu nhiên X có khả năng xuất hiện lớn nhất trong một lân cận nào đó của nó. Đối với biến ngẫu nhiên liên tục, mod(X) là giá trị của X ứng với $f(x)$ đạt cực đại địa phương.
- *Ký hiệu*: $mod(X)$

Phân vị mức p

- *Khái niệm*: Phân vị mức p của biến ngẫu nhiên X là giá trị z_p sao cho.

$$F(z_p) = P(X < z_p) = p$$

- *Trung vị*: Trung vị của biến ngẫu nhiên X là giá trị của X chia phân phối xác suất thành hai phần có xác suất bằng nhau. Kí hiệu là $med(X)$:
  $P(X < med(X)) = P(X \geq med(X)) = 0,5$

Một số phân phối xác suất thông dụng

Các quy luật thông dụng sẽ học:

Biến ngẫu nhiên rời rạc

- Luật phân phối nhị thức
- Luật phân phối Poisson

Biến ngẫu nhiên liên tục

- Phân phối đều liên tục
- Phân phối chuẩn
- Phân phối mũ
- Phân phối Khi bình phương
- Phân phối Student

Phân phối nhị thức (Binomial Distribution)

Định nghĩa 4.1

Biến ngẫu nhiên X nhận giá trị trong tập $\{0; 1; 2; ...; n\}$ với xác suất được tính theo công thức Bernoulli:
$P(X = k) = C_n^k.p^k.(1-p)^{n-k}$ *với* $k = 0, 1, \dots, n; 0 \leq p \leq 1$
gọi là tuân theo phân phối nhị thức với các tham số n và p.
Ký hiệu: $X \sim B(n; p)$

Các tham số đặc trưng

Với $X \sim B(n; p)$ ta có:

- $EX = np$
- $VX = np(1-p) = npq$ với $q = 1 - p$
- $(n+1)p - 1 \leq mod(X) \leq (n+1)p$

Phân phối nhị thức

Ứng dụng

Ta thực hiện n phép thử độc lập cùng điều kiện. Trong mỗi phép thử xác suất xảy ra sự kiện A luôn là p. Gọi X là số phép thử xảy ra A. Ta có kết quả: $X \sim B(n; p)$

Ví dụ 1

Gieo một con xúc xắc 3 lần. Gọi X là số lần ra mặt lục trong 3 lần gieo. Lập bảng phân phối xác suất của X, biết rằng khả năng ra mặt lục ở mỗi lần gieo là $1/6$.
Gợi ý:
$X \sim B(n; p)$ *với* $n = 3; p = 1/6$, $P(X = k) = C_n^k.p^k.(1-p)^{n-k}$

$X = x$	0	1	2	3
$P(X = x)$	$125/216$	$75/216$	$15/216$	$1/216$

Phân phối nhị thức

Ví dụ 2

Một người chơi đề trong 10 ngày, mỗi ngày người đó chơi 5 số. Tính xác suất trong 10 ngày chơi:
+) Người đó trúng được đúng 2 ngày.
+) Người đó trúng được ít nhất 2 ngày
+) Xác định số ngày trúng có khả năng xảy ra cao nhất?

Phân phối nhị thức

Biến nào sau đây là tuân theo phân phối nhị thức:

- Tung một đồng xu 3 lần. Gọi X là số lần được mặt ngửa.
- Hộp có 4 bi trắng và 3 bi xanh. Lấy ngẫu nhiên 3 bi. Gọi X là số bi xanh lấy được theo 2 cách:
  +) Lấy lần lượt 3 bi
  +) Lấy có hoàn lại 3 bi
- Một máy sản xuất ra sản phẩm có tỷ lệ phế phẩm là 2%. Cho máy sản xuất ra 10 sản phẩm. Gọi X là số phế phẩm có được.
- Một xạ thủ bắn 3 phát đạn vào bia. Ở lần bắn sau do rút được kinh nghiệm các lần bắn trước nên xác suất bắn trúng của 3 phát lần lượt là $0,7; 0,8; 0,9$. Gọi X là số phát bắn trúng bia.

Phân phối Poisson

Định nghĩa 4.2

Biến ngẫu nhiên X nhận giá trị trong tập $\{0; 1; 2; \ldots; n; \ldots\}$ với xác suất :

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

gọi là tuân theo phân phối Poisson với tham số λ
Ký hiệu: $X \sim P(\lambda)$

Các tham số đặc trưng

Với $X \sim P(\lambda)$ ta có:

- $EX = \lambda$
- $VX = \lambda$
- $\lambda - 1 \leq mod(X) \leq \lambda$

Phân phối Poisson

- Quá trình Poisson còn có thể gọi là quá trình đếm.
- Trong tình huống nào ta gặp phân phối Poisson?
  Xét một sự kiện E xuất hiện ở những thời điểm ngẫu nhiên. Giả sử số lần xuất hiện E trong một khoảng thời gian không ảnh hưởng tới xác suất xuất hiện của E trong các khoảng thời gian kế tiếp. Hơn nữa cường độ xuất hiện của E là không thay đổi, nghĩa là số lần trung bình xuất hiện E trong khoảng thời gian tỉ lệ với độ dài khoảng thời gian đó.
  Gọi X là số lần xuất hiện E trong khoảng thời gian (t_1, t_2). Ta có $X \sim P(\lambda)$ với $\lambda = c(t_2 - t_1)$, trong đó c là hằng số được gọi là cường độ xuất hiện của E.
- Phân phối này có nhiều ứng dụng đối với nhiều quá trình có liên quan đến số quan sát đối với một đơn vị thời gian hoặc không gian. Ví dụ: Số cuộc điện thoại nhận được ở một trạm điện thoại trong một phút, số khách hàng đến nhà băng đối với mỗi một chu kỳ 30 phút, số lỗi in sai trong một trang, ... Nói chung dòng vào của một hệ phục vụ (quán bia, hiệu cắt tóc, hiệu sửa xe, trạm điện thoại, một cửa hàng nào đó, . . .) là các biến ngẫu nhiên tuân theo luật phân phối Poisson.

Ví dụ 3

Ở một tổng đài bưu điện, các cuộc điện thoại gọi đến xuất hiện ngẫu nhiên, độc lập với nhau với tốc độ trung bình 2 cuộc gọi trong một phút. Tìm xác suất để:
a) Có đúng 5 cuộc điện thoại trong vòng 2 phút
b) Không có cuộc điện thoại nào trong khoảng thời gian 30 giây
c) Có ít nhất 1 cuộc điện thoại trong khoảng thời gian 10 giây.

Lời giải

a. Gọi X là số cuộc điện thoại xuất hiện trong vòng 2 phút. $X \sim P(\lambda)$
λ chính là số cuộc điện thoại trung bình đến trong vòng 2 phút. $\lambda = 4$
$P(X = 5) = e^{-\lambda}\frac{\lambda^5}{5!} = e^{-4}\frac{4^5}{5!} = 0,156$
b. Gọi X là số cuộc điện thoại xuất hiện trong vòng 30 giây. $X \sim P(\lambda)$ với $\lambda = 1$. Ta có
$P(X = 0) = e^{-\lambda}\frac{\lambda^0}{0!} = e^{-1} = 0,3679$
c. Gọi X là số cuộc điện thoại xuất hiện trong vòng 10 giây. $X \sim P(\lambda)$ với $\lambda = 1/3$. Ta có $P(X \geq 1) = 1 - P(X = 0) = 1 - e^{-1/3} = 0,2835$

Chú ý 4.1

Khi n lớn và p nhỏ ($n > 50; p < 0,1$) thì $X \sim B(n;p)$ có thể chuyển thành $X \sim P(\lambda)$ với $\lambda = np$

Ví dụ 4

Trong một lô thuốc, tỷ lệ ống thuốc hỏng là $p = 0,003$. Kiểm nghiệm 1000 ống. Tính xác suất để gặp 3 ống bị hỏng.
Lời giải:
Gọi X là số ống thuốc hỏng trong 1000 ống. Ta có $X \sim B(n;p)$ với $n = 1000; p - 0,003$
Do n lớn và p bé nên ta xấp xỉ $X \sim P(\lambda)$ với $\lambda = np = 3$

$$P(X = 3) = e^{-\lambda}\frac{\lambda^3}{3!} = e^{-3}\frac{3^3}{3!} = 0,224$$

Phân phối chuẩn

Định nghĩa 4.3

Biến ngẫu nhiên X được gọi là tuân theo phân phối chuẩn với hai tham số μ và σ^2 (với $\sigma > 0$) nếu hàm mật độ của X có dạng:

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

Ký hiệu: $X \sim N(\mu, \sigma^2)$

Các tham số đặc trưng

$EX = \mu$
$VX = \sigma^2$
$mod(X) = med(X) = \mu$

Mục tiêu là ta tính xác suất dạng $P(a < X < b)$

Phân phối chuẩn tắc

Đặc biệt: $X \sim N(0;1)$ với $(\mu = 0, \sigma = 1)$, X được gọi là tuân theo phân phối chuẩn tắc (hay chuẩn hoá).
Hàm mật độ xác suất hay còn gọi là hàm mật độ Gauss:

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

Để tính xác suất ta dùng hàm Laplace: $\phi(x) = \int_0^x \varphi(t)dt$

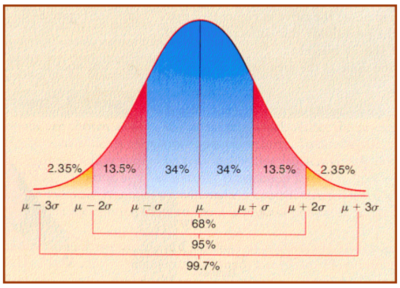
Tính chất:

- $\phi(x)$ là hàm lẻ, tăng thực sự.
- $\phi(+\infty) = 0,5$
- $X \sim N(0;1)$ ta có: $P(a < X < b) = \phi(b) - \phi(a)$
- Giá trị của hàm Laplace được tính sẵn thành bảng số liệu.

Phân phối chuẩn tổng quát

Kết quả: Nếu $X \sim N(\mu;\sigma^2)$ ta có $Z = \frac{X-\mu}{\sigma} \sim N(0;1)$
Từ đó ta xây dựng được công thức tính:

- $P(X < a) = 0,5 + \phi(\dfrac{a-\mu}{\sigma})$
- $P(X > a) = 0,5 - \phi(\dfrac{a-\mu}{\sigma})$
- $P(a \leq X < b) = \phi(\dfrac{b-\mu}{\sigma}) - \phi(\dfrac{a-\mu}{\sigma})$
- $P(|X-\mu| < \varepsilon) = 2\phi(\dfrac{\varepsilon}{\sigma})$



Phân phối chuẩn - Ví dụ

Ví dụ 5

Độ dài một chi tiết máy giả sử tuân theo luật phân phối chuẩn với giá trị trung bình là 20 cm và độ lệch chuẩn là 0,5 cm. Tính xác suất khi chọn ngẫu nhiên ra một chi tiết thì độ dài của nó:
a) lớn hơn 20 cm
b) bé hơn 19,5 cm
c) nằm trong khoảng 19 cm – 21 cm
Lời giải:
Gọi X(cm) là độ dài chi tiết máy đã chọn. $X \sim N(\mu,\sigma^2)$, $\mu = 20$, $\sigma = 0,5$.

- $P(X > 20) = 0,5 - \phi(\dfrac{20-\mu}{\sigma}) = 0,5 - \phi(0) = 0,5$
- $P(X < 19,5) = 0,5 + \phi(\dfrac{19,5-\mu}{\sigma}) = 0,5 + \phi(-1) = 0,5 - \phi(1) = 0,5 - 0,3413 = 0,1587$
- $P(19 < X < 21) = \phi(\dfrac{21-\mu}{\sigma}) - \phi(\dfrac{19-\mu}{\sigma}) = \phi(2) - \phi(-2) = 2\phi(2) = 2.0,4772 = 0,9544$

Xấp xỉ phân phối nhị thức bằng phân phối chuẩn

Với $X \sim B(n;p)$ thoả mãn $np(1-p) > 20$.
Khi đó ta xấp xỉ $X \sim N(\mu, \sigma^2)$ với $\mu = np, \sigma^2 = np(1-p)$
Tuy nhiên vì chúng ta xấp xỉ một phân phối rời rạc bằng một phân phối liên tục, nên cần một sự hiệu chỉnh để giảm sai số. Cụ thể với k, k_1, k_2 là số tự nhiên ta có:

$$P(X = k) = \phi(\frac{k + 0,5 - \mu}{\sigma}) - \phi(\frac{k - 0,5 - \mu}{\sigma})$$
$$P(k_1 \leq X \leq k_2) = \phi(\frac{k_2 + 0,5 - \mu}{\sigma}) - \phi(\frac{k_1 - 0,5 - \mu}{\sigma})$$

Xấp xỉ phân phối nhị thức bằng phân phối chuẩn

Ví dụ 6

Kiểm tra chất lượng 1000 sản phẩm với tỷ lệ chính phẩm 0,95. Tìm xác suất để số chính phẩm trong lô kiểm tra từ 940 đến 960.

Lời giải

: Gọi X là biến ngẫu nhiên chỉ số chính phẩm trong lô sản phẩm kiểm tra, ta có $X \sim B(1000; 0,95)$
Với $n = 1000, p = 0,95$, ta có $np = 950$ và $npq = 47,5$ đủ lớn nên ta xấp xỉ $X \sim N(950; 47,5)$:
$P(940 \leq X \leq 960) = \phi(\frac{960 + 0,5 - 950}{\sqrt{47,5}}) - \phi(\frac{940 + 0,5 - 950}{\sqrt{47,5}})$
$= \phi(1,52) - \phi(-1,52) = 2\phi(1,52) = 0,8716$

Phân phối chuẩn - Ý nghĩa

Phân phối chuẩn được Gauss phát minh năm 1809 nên cũng có khi nó được mang tên là phân phối Gauss.
Ta thấy biến ngẫu nhiên tuân theo phân phối chuẩn nhận giá trị trên cả trục số, tuy nhiên có thể xấp xỉ một số biến ngẫu nhiên không nhận tất cả các giá trị trên $\mathbb{R}$ theo phân phối chuẩn, đó là do qui tắc $3-\sigma$, tức là nếu ta có xác suất X rơi vào miền có xác suất bằng 0,9974 rất gần 1, nên hầu hết người ta chỉ cần quan tâm đến các giá trị trong lân cận $3-\sigma$ của kỳ vọng.
Phân phối chuẩn chiếm vị trí quan trọng trong lý thuyết xác suất, là vị trí trung tâm trong các kết luận thống kê sau này. Trong thực tế, ví dụ trong lĩnh vực kinh tế, khoa học xã hội, ... nhiều phân phối không giống phân phối chuẩn, nhưng phân phối của trung bình cộng đối với mỗi trường hợp lại có thể xem là phân phối chuẩn miễn là cỡ mẫu n đủ lớn.

Phân phối mũ

Định nghĩa 4.4

Biến ngẫu nhiên X được gọi là tuân theo phân phối mũ với tham số $\lambda > 0$ nếu nó có hàm mật độ xác suất có dạng:

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & x \le 0 \end{cases}$$

Ký hiệu: $X \sim \mathcal{E}(\lambda)$

Các tham số đặc trưng

$EX = \dfrac{1}{\lambda}$

$VX = \dfrac{1}{\lambda^2}$

Phân phối mũ

Ta có $P(X > x) = e^{\lambda x}$
Phân phối mũ có tính chất không nhớ:

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

Ý nghĩa: Phân phối mũ có nhiều nhiều ứng dụng trong thực tiễn. Nói chung với một giả thiết nào đó, khoảng thời gian giữa hai lần xuất hiện của một sự kiện E nào đó sẽ có phân phối mũ. Vì lý do này phân phối mũ còn có tên gọi là phân phối của thời gian chờ đợi ("Waiting time distribution"). Ví dụ khoảng thời gian giữa 2 ca cấp cứu ở một bệnh viện, khoảng thời gian giữa 2 lần hỏng hóc của một chiếc máy, khoảng thời gian giữa 2 trận lụt hay động đất, ...

Phân phối mũ

Ví dụ 7

Giả sử tuổi thọ (tính bằng năm) của một mạch điện tử trong máy tính là một biến ngẫu nhiên có phân phối mũ với kỳ vọng là 6,25. Thời gian bảo hành của mạch điện tử này là 5 năm. Hỏi có bao nhiêu phần trăm mạch điện tử bán ra phải thay thế trong thời gian bảo hành.

Lời giải

Gọi X là tuổi thọ của mạch. X tuân theo phân phối mũ với tham số $\lambda = \dfrac{1}{EX} = \dfrac{1}{6,25}$
$P(X \leq 5) = 1 - e^{-5\lambda} = 1 - e^{-0,8} = 0,5506$
Vậy có khoảng 55% mạch điện tử bán ra phải thay thế trong thời gian bảo hành

Phân phối Khi bình phương

Định nghĩa 4.5

Giả sử $X_i, (i = 1, 2, \ldots, n)$ *là các biến ngẫu nhiên độc lập cùng phân phối chuẩn tắc. Biến ngẫu nhiên* $Y = \sum_{i=1}^{n} X_i^2$ *được gọi là tuân theo phân phối Khi bình phương với* n *bậc tự do.*
Ký hiệu: $Y \sim \chi^2(n)$

Các tham số đặc trưng

- $EY = n$
- $VY = 2n$

Phân phối Student

Định nghĩa 4.6

Giả sử $X \sim N(0; 1)$ *và* $Y \sim \chi^2(n)$ *là hai biến ngẫu nhiên độc lập. Khi đó:*

$$T = \frac{X}{\sqrt{\frac{Y}{n}}}$$

được gọi là tuân theo phân phối Student với n *bậc tự do.*
Ký hiệu: $T \sim T(n)$

Các tham số đặc trưng

- $ET = 0$
- $VT = \frac{n}{n-2}$

Chú ý

- Phân phối Student có cùng dạng và tính đối xứng như phân phối chuẩn nhưng nó phản ánh tính biến đổi của phân phối sâu sắc hơn. Phân phối chuẩn không thể dùng để xấp xỉ phân phối khi mẫu có kích thước nhỏ. Trong trường hợp này ta dùng phân phối Student.
- Khi bậc tự do n tăng lên ($n > 30$) thì phân phối Student tiến nhanh về phân phối chuẩn. Do đó khi $n > 30$ ta có thể dùng phân phối chuẩn thay thế cho phân phối Student.