BÀI 2: BIẾN NGẪU NHIÊN VÀ QUY LUẬT PHÂN PHỐI XÁC SUẤT CỦA BIẾN NGẪU NHIÊN

Mục tiêu

- Phân biệt biến ngẫu nhiên, các giá trị có thể có của biến ngẫu nhiên và xác suất để biến ngẫu nhiên nhận giá trị trong miền giá trị có thể có của nó,
- Nắm vững định nghĩa, ý nghĩa, các tính chất và công thức tính giá trị các tham số đặc trưng của biến ngẫu nhiên (kỳ vọng toán, phương sai và độ lệch tiêu chuẩn)

Nội dung

I. ĐỊNH NGHĨA VÀ PHÂN LOẠI BIẾN NGẪU NHIÊN

I.1. Định nghĩa.

Biến ngẫu nhiên là đại lượng (hay biến sè) mà trong kết quả của phép thử sẽ nhận một và chỉ một trong các giá trị có thể có của nó với một xác suất tương ứng xác định .

Các biến ngẫu nhiên được ký hiệu là : X, Y, Z... hoặc $X_1, X_2, ..., X_n$.

Các giá trị có thể có của biến ngẫu nhiên X được ký hiệu là:

$$x_1, x_2, .., x_i, ..., x_n .$$

Việc X nhận 1 giá trị nào đó $(X = x_1)$, $(X = x_2)$, .. $(X = x_i)$.....,$(X = x_n)$ là hoàn toàn ngẫu nhiên nên thực chất chúng là các biến cố ngẫu nhiên.

Thí dụ: Gọi X là " số con trai trong 100 đứa trẻ sắp sinh tại 1 nhà hộ sinh".

Khi đó X là 1biến ngẫu nhiên với giá trị có thể có là 0, 1, ..., 100.

I.2. Phân loại biến ngẫu nhiên

Có 2 loại biến ngẫu nhiên là biến ngẫu nhiên rời rạc và biến ngẫu nhiên liên tục.

I.2.1. Biến ngẫu nhiên rời rạc:

Biến ngẫu nhiên gọi là rời rạc nếu các giá trị có thể có của nó lập nên 1 tập hợp hữu hạn hoặc đếm được, hay Biến ngẫu nhiên rời rạc là biến ngẫu nhiên mà ta có thể liệt kê được tất cả các giá trị có thể có của nó.

Thí dụ : (Biến ngẫu nhiên rời rạc đếm được). Gọi X là số khách hàng vào mua hàng tại 1 cửa hàng trong khoảng thời gian T nào đó. Khi đó X là 1 biến ngẫu nhiên rời rạc đếm được với các giá trị có thể có là 0, 1, 2, ...

I.2.2. Biến ngẫu nhiên liên tục:

Biến ngẫu nhiên gọi là liên tục nếu các giá trị có thể có của nó lấp đầy (lấp kín) một khoảng trên trục số.

Đối với biến ngẫu nhiên liên tục ta không thể liệt kê ra được tất cả các giá trị có thể có của nó.

Thí dụ : Bắn 1 viên đạn vào bia.

Gọi X là khoảng cách từ điểm chạm của viên đạn tới tâm bia.

X là biến ngẫu nhiên liên tục vì ta không thể liệt kê ra được tất cả các giá trị có thể có của nó.

II. QUY LUẬT PHÂN PHỐI XÁC SUẤT CỦA BIẾN NGẪU NHIÊN

II.1. Định nghĩa

Bất kỳ một hình thức thể hiện nào cho phép biểu diễn mối quan hệ giữa các giá trị có thể có của biến ngẫu nhiên và các xác suất để biến ngẫu nhiên nhận giá trị tương ứng đều được gọi là quy luật phân phối xác suất của biến ngẫu nhiên ấy.

Người ta thường dùng 3 dạng thức để thể hiện quy luật phân phối xác suất của biến ngẫu nhiên, đã là:

- Bảng phân phối xác suất
- Hàm phân phối xác suất
- Hàm mật độ xác suất .

II.2. Bảng phân phối xác suất (chỉ dùng cho biến ngẫu nhiên rời rạc)

Giả sử biến ngẫu nhiên rời rạc X có thể nhận 1 trong các giá trị có thể có là x_1, x_2,... ,x_n với các xác suất tương ứng là P_1, P_2,... ,P_n . Bảng phân phối xác suất của biến ngẫu nhiên rời rạc X có dạng:

X	x_1	x_2	...	x_n
P	P_1	P_2	...	P_n

Chú ý: Để tạo nên một bảng phân phối xác suất, các xác suất P_i phải thoả mãn 2 điều kiện sau:

1. $0 \le P_i \le 1$ với $\forall i$
2. $\sum_{i=1}^{n} P_i = 1$

Thí dụ 1: Trong 1 chiếc hộp có đựng 4 phế phẩm và 6 chính phẩm. Lấy ngẫu nhiên ra 2 sản phẩm. Xây dựng quy luật phân phối xác suất của số chính phẩm có thể lấy ra.

Giải

Gọi X là "số chính phẩm được lấy ra"

$\Rightarrow$ X là biến ngẫu nhiên rời rạc với các giá trị có thể có là 0, 1, 2 và các xác suất tương ứng là P_0, P_1, P_2 .

$\Rightarrow$ Ta tìm được P_i $\left(i = \overline{0,2}\right)$ như sau:

$$* \ P_0 = P(X = 0) = \frac{C_4^2}{C_{10}^2} = \frac{2}{15}$$

$$* \; P_1 = P(X = 1) = \frac{C_4^1 C_6^1}{C_{10}^2} = \frac{8}{15}$$

$$* \; P_2 = P(X = 2) = \frac{C_6^{02}}{C_{10}^2} = \frac{5}{15}$$

Vậy quy luật phân phối xác suất của X là:

X	0	1	2
P	$\frac{2}{15}$	$\frac{8}{15}$	$\frac{5}{15}$

II.3. Hàm phân phối xác suất

II.3.1. Định nghĩa:

Hàm phân phối xác suất của biến ngẫu nhiên X, ký hiệu F(x), là xác suất để X nhận giá trị nhỏ hơn x, với x là 1 số thực bất kỳ.

Nghĩa là: $F(x) = P(X < x)$

Nếu X là biến ngẫu nhiên rời rạc thì: $F(x) = \sum_{x_i < x} P_i$

Thí dụ: Biến ngẫu nhiên rời rạc X có bảng phân phối xác suất như sau:

X	2	4	5
P	0,2	0,5	0,3

Hãy xây dựng hàm phân phối xác suất của X.

Giải

Vậy hàm phân phối xác suất của X là:

$$F(X) = \begin{cases} 0 & \text{với } x \le 2 \\ 0{,}2 & \text{với } 2 < x \le 4 \\ 0{,}7 & \text{với } 4 < x \le 5 \\ 1 & \text{với } x > 5 \end{cases}$$

Chú ý : + Đồ thị của hàm phân phối xác suất của biến ngẫu nhiên rời rạc có dạng bậc thang với số điểm gián đoạn chính bằng số giá trị có thể có của X.

+ Nếu X là biến ngẫu nhiên liên tục thì hàm phân phối xác suất của nó liên tục và khả vi tại mọi điểm của X. Do đó đồ thị của nó là 1 đường cong liên tục.

II.3.2. Các tính chất của hàm phân phối xác suất

Tính chất 1: Hàm phân phối xác suất luôn nhận giá trị trong khoảng [0 ; 1], tức là $0 \le F(X) \le 1$

Tính chất 2: Hàm phân phối xác suất là hàm không giảm, tức là với $x_2 > x_1$ thì $F(X_2) \ge F(X_1)$

Hệ quả 1. Xác suất để biến ngẫu nhiên X nhận giá trị trong khoảng [a,b) bằng hiệu số của hàm phân phối xác suất tại 2 đầu mót cũa khoảng đó.

$$\Rightarrow P(a \le x < b) = F(b) - F(a)$$

Hệ quả 2. Xác suất để biến ngẫu nhiên liên tục X nhận 1 giá trị nào đó luôn bằng 0, tức là: $P(X = x) = 0$

Hệ quả 3. Nếu X là biến ngẫu nhiên liên tục ta có các đẳng thức sau:

$$P(a \le X < b) = P(a < X \le b) = P(a < X < b) = P(a \le X \le b)$$

Tính chất 3: Ta có biểu thức giới hạn sau: $F(-\infty) = 0; F(+\infty) = 1$

$$F(x) = 1 \text{ với } x > b$$

II.3 Ý nghĩa của hàm phân phối xác suất

Hàm phân phối xác suất phản ánh mức độ tập trung xác suất ở về phía bên trái cũa một số thực x nào đó.

II. 4. Hàm mật độ xác suất *(chỉ dùng cho biến ngẫu nhiên liên tục)*

II.4.1. Định nghĩa:

Hàm mật độ xác suất của biến ngẫu nhiên liên tục X là đạo hàm bậc nhất của hàm phân phối xác suất của biến ngẫu nhiên đó, ký hiÖu f(x)

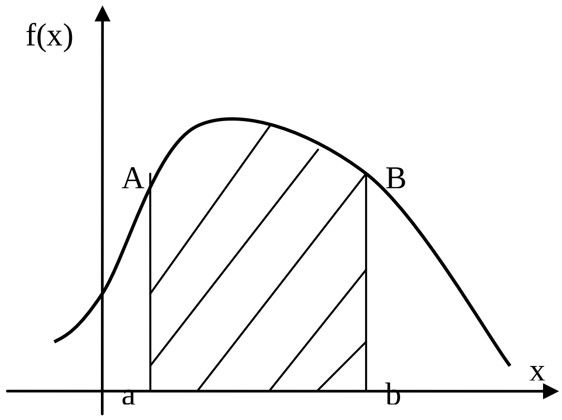
$$f(x) = F'(x)$$

II.4.2. Các tính chất của hàm mật độ xác suất

Tính chất 1: $f(x) \geq 0$ với $\forall x$.

Tính chất 2: $P(a < X, < b) = \int_a^b f(x)dx$

Về mặt hình ảnh hình học: $P(a < X < b) = S_{aABb}$



Tính chất 3: $F(x) = \int_{-\infty}^{x} f(x)dx$

Tính chất 4 : $\int_{-\infty}^{+\infty} f(x)dx = 1$

Chú ý: Để hàm f(x) có thể là hàm mật độ xác suất của 1 biến ngẫu nhiên liên tục X nào đó thì nó phải thoả mãn 2 điều kiện:

a. $f(x) \geq 0$ với $\forall x$

b. $\int_{-\infty}^{+\infty} f(x)dx = 1$

Thí dụ 1: Hàm phân phối xác suất của biến ngẫu nhiên X có dạng

$$F(x) = \begin{cases} 0 & \text{với } x \leq 0 \\ ax^2 & \text{với } 0 < x \leq 1 \\ 1 & \text{với } x > 1 \end{cases}$$

a. Tìm hệ số a,

b. Tìm hàm mật độ xác suất f(x),

c. Tìm xác suất để biến ngẫu nhiên X nhận giá trị trong (0,25 ; 0,75)

Giải

a. Tìm hệ số a

Vì F(x) là hàm liên tục nên tại x = 1 $\rightarrow ax^2 = 1 \Rightarrow a = 1$

b. Hàm mật độ xác suất của biến ngẫu nhiên liên tục:

$$f(x) = \begin{cases} 0 & \text{với } x \leq 0 \\ 2ax & \text{với } 0 < x \leq 1 \\ 0 & \text{với } x > 1 \end{cases}$$

c. $P(0{,}25 < X < 0{,}75) = F(0{,}75) - F(0{,}25) = (0{,}75)^2 - (0{,}25)^2$

$\Rightarrow P(0{,}25 < X < 0{,}75) = 0{,}5$.

Thí dụ 2: Biến ngẫu nhiên liên tục X có hàm mật độ xác suất như sau:

$$f(x) = \begin{cases} a\cos x & \text{với } x \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right] \\ 0 & \text{với } x \overline{\in} \left[-\frac{\pi}{2}, \frac{\pi}{2}\right] \end{cases}$$

a. Tìm hệ số a

b. Tìm F(x)

c. Tìm $P(0 < X < \pi/4)$

Giải

a. Theo tính chất hàm mật độ xác suất :

$$\int_{-\infty}^{+\infty} f(x)dx = 1 \qquad (*)$$

$$\Rightarrow VT* = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} a\cos x dx = a\sin x \Big|_{-\frac{\pi}{2}}^{\frac{\pi}{2}}$$

$$= a\left(\sin\frac{\pi}{2} - \sin(-\frac{\pi}{2})\right) = 2a$$

$$\Rightarrow 2a = 1 \Rightarrow a = 1/2$$

b. Để tìm hàm phân phối xác suất ta sử dụng tính chất :

$$F(x) = \int_{-\infty}^{x} f(x)dx$$

Vậy:

$$F(x) = \begin{cases} 0 & \text{với } x \le -\frac{\pi}{2} \\ \frac{1}{2}(\sin x + 1) & \text{với } -\frac{\pi}{2} \le x < \frac{\pi}{2} \\ 1 & \text{với } x > \frac{\pi}{2} \end{cases}$$

c. $P(0 < X < \frac{\pi}{4}) = F(\frac{\pi}{2}) - F(0) = \frac{\sqrt{2}}{4}$

III. CÁC THAM SỐ ĐẶC TRƯNG CỦA BIẾN NGẪU NHIÊN

III.1. Kỳ vọng toán của biến ngẫu nhiên

III.1.1. Định nghĩa:

Giả sử biến ngẫu nhiên rời rạc X nhận một trong các giá trị có thể có x_1, x_2,... ,x_n với các xác suất tương ứng

$$P_1, P_2, \ldots, P_n$$

Kỳ vọng toán của biến ngẫu nhiên rời rạc X, ký hiệu E(X) là tổng của các tích giữa các giá trị có thể có của các ngẫu nhiên với các xác suất tương ứng.

$$E(X) = \sum_{i=1}^{n} x_i P_i$$

Nếu X là biến ngẫu nhiên liên tục với hàm mật độ xác suất f(x) thì E(X) được xác định bằng biểu thức:

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

Thí dụ 1: Biến ngẫu nhiên rời rạc X có bảng phân phối xác suất :

X	2	4	5
P	0,2	0,5	0,3

Tìm kú vọng toán của nó.

Giải

Vì X là biến ngẫu nhiên rời rạc nên :

$$E(X) = \sum_{i=1}^{n} x_i P_i = 2.0,2 + 4.0,5 + 5.0,3$$

$$\Rightarrow E(X) = 3,9$$

Thí dụ 2: Biến ngẫu nhiên liên tục X có hàm mật độ xác suất như sau:

$$f(x) = \begin{cases} 0 & \text{với } x \le 1 \\ \dfrac{x^2 - 1}{2} & \text{với } 1 < x \le 2 \\ 0 & \text{với } x > 2 \end{cases}$$

Tìm E(X)

Giải

Vì X là biến ngẫu nhiên liên tục nên

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

$$= \int_{-\infty}^{1} x.f(x)dx + \int_{1}^{2} x.f(x)dx + \int_{2}^{+\infty} x.f(x)dx$$

$$= \int_{1}^{2} x.\frac{x^2 - 1}{2}dx = 1{,}1.$$

III.1.2. Các tính chất của kỳ vọng toán

Tính chất 1: E(C) = C với C = const, với C là hằng số

Tính chất 2: E(CX) = C.E(X) với C = const, với C là hằng số

Tính chất 3: E(X+Y) = E(X)+E(Y)

Hệ quả: $E(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} E(X_i)$

Tính chất 4: E(X.Y) = E(X).E(Y) nếu X và Y là 2 biến ngẫu nhiên độc lập

Hệ quả: $E(\prod_{i=1}^{n} X_i) = \prod_{i=1}^{n} E(X_i)$ nếu các X, i độc lập

III.1.3. Bản chất và ý nghĩa của kỳ vọng toán

Giả sử biến ngẫu nhiên X có các giá trị có thể có là x_1, x_2,... ,x_k

Trung bình số học của biến ngẫu nhiên X trong n phép thử này là :

$$\overline{X} = \frac{n_1x_1 + n_2x_2 + ... + n_kx_k}{n}$$

hay: $$\overline{X} = \frac{n_1}{n}x_1 + \frac{n_2}{n}x_2 + ... + \frac{n_k}{n}x_k$$

$$\Rightarrow \overline{X} = f_1x_1 + f_2x_2 + ... + f_kx_k$$

trong đó $\frac{n_i}{n} = f_1; \left(i = \overline{1,n}\right)$ chính là tần số xuất hiện của biến cố " X nhận giá trị x_i ".

$$\overline{X} = \sum_{i=1}^{k} f_i x_i \approx \sum_{i=1}^{k} P_i x_i = E(X)$$

Vậy kỳ vọng toán của biến ngẫu nhiên gần bằng trung bình số học của các giá trị quan sát của biến ngẫu nhiên ấy. Nó phản ánh giá trị trung tâm của phân phối xác suất của biến ngẫu nhiên.

III.2. Phương sai của biến ngẫu nhiên

III.2.1. Định nghĩa:

Phương sai của biến ngẫu nhiên X , ký hiệu là V(X) là kỳ vọng toán của bình phương sai lệch của biến ngẫu nhiên so với kỳ vọng toán của nó.

$$V(X) = E[X - E(X)]^2$$

hay: $$V(X) = E(X^2) - [E(X)]^2$$

* Nếu X là biến ngẫu nhiên rời rạc thì phương sai được xác định bằng công thức

$$V(X) = \sum_{i=1}^{n} [X_i - E(X)]^2 P_i \qquad \text{hay:} \qquad V(X) = \sum_{i=1}^{n} x_i^2 P_i - [E(X)]^2$$

* Nếu X là biến ngẫu nhiên liên tục thì

$$V(X) = \int_{-\infty}^{+\infty} [X - E(X)]^2 f(x)dx \qquad \text{hay:} \qquad V(X) = \int_{-\infty}^{+\infty} x^2.f(x)dx - [E(X)]^2$$

Thí dụ: Biến ngẫu nhiên rời rạc X có bảng phân phối xác suất như sau:

X	1	3	4
P	0,1	0,5	0,4

Tính V(X).

Giải

* $E(X) = 1.0{,}1 + 3.0{,}5 + 4.0{,}4 = 3.2$

* $V(X) = 1^2.0{,}1 + 3^2.0{,}4 + 4^2.0{,}4 - (3{,}2)^2 = 0{,}76$

III.2.2. Các tính chất của phương sai

Tính chất 1: V(C) = 0 với C = const

Tính chất 2: $V(CX) = C^2V(X)$ với C= const

Tính chất 3: Phương sai của tổng 2 biến ngẫu nhiên X và Y độc lập bằng tổng phương sai thành phần.

$$V(X + Y) = V(X) + V(Y)$$

Hệ quả 1: Phương sai của tổng n biến ngẫu nhiên độc lập với nhau $X_1, X_2, \ldots, X_n$ bằng tổng các phương sai thành phần.

$$V\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} V(X_i)$$

Hệ quả 2: Phương sai của tổng 1 hằng số với 1 biến biến ngẫu nhiên bằng phương sai của chính biến ngẫu nhiên đó:

$$V(C + X) = V(X)$$

Hệ quả 3. Phương sai của tổng và hiệu 2 biến ngẫu nhiên độc lập bằng tổng các phương sai thành phần

$$V(X \pm Y) = V(X) + V(Y)$$

III.2.3. ý nghĩa của phương sai

Phương sai chính là trung bình số học của bình phương các sai lệch giữa các giá trị có thể có của biến ngẫu nhiên so với giá trị trung bình của các giá trị đó. Do vËy nó phản ánh mức độ phân tán của các giá trị của biến ngẫu nhiên xung quanh giá trị trung tâm của nó là kú vọng toán.

III.3. Độ lệch tiêu chuẩn

Độ lệch tiêu chuẩn của biến ngẫu nhiên X được ký hiệu σ_X, là căn bậc hai của phương sai V(X)

$$\sigma_X = \sqrt{V(X)}$$

IV. MỘT SỐ QUY LUẬT PHÂN PHỐI XÁC SUẤT THÔNG DỤNG

IV.1. Quy luật không - một: A(P)

IV.1.1. Định nghĩa:

Biến ngẫu nhiên rời rạc X nhận 1 trong các giá trị có thể có là 0 hoặc 1 với xác suất tương ứng được xác định bởi công thức (3.1) gọi là phân phối theo quy luật Không - Một với tham sè P và được ký hiệu là A(P) (A - viÕt t¾t cña tõ Alternative)

Bảng phân phối xác suất của biến ngẫu nhiên $X \sim A(P)$ là :

X	0	1
P	(1- P)	P

IV.1.2. Các tham số đặc trưng của quy luật A(P)

* Kỳ vọng toán: $E(X) = 0.(1-P) + 1.P = P$ hay $E(X) = P$

* Phương sai : $V(X) = 0^2.(1-P) + 1^2.P - P^2 = P(1-P)$

hay $V(X) = P(1-P)$

* Độ lệch tiêu chuẩn: $\sigma(X) = \sqrt{P(1-P)}$

IV.2. Quy luật nhị thức: $B_i(n,p)$

Giả sử ta có 1 lược đồ Bernoulli:

+ Tiến hành n phép thử độc lập.

+ Trong mỗi phép thử chỉ có 2 khả năng hoặc biến cố A xuất hiện hoặc biến cố $\overline{A}$ xuất hiện .

+ Xác suất xuất hiện biến cố A trong mọi phép thử đều bằng P. Do đó xác suất xuất hiện biến cố $\overline{A}$ trong mọi phép thử đều bằng (1 – P).

Gọi X là "số lần xuất hiện của biến cố A trong n phép thử độc lập nói trên" thì X là biến ngẫu nhiên rời rạc với các giá trị có thể có là 0, 1, ... , n

Xác suất để biến ngẫu nhiên X nhận 1 trong các giá trị có thể có nói trên được tính bằng công thức Bernoulli:

$$P_x = P(X = x) = C_n^x P^x (1-P)^{n-x} \qquad (3.2)$$

với $\left(x = \overline{0, n}\right)$

IV. 2.1. Định nghĩa:

Biến ngẫu nhiên rời rạc X nhận 1 trong các giá trị có thể có 0, 1, ... , n với các xác suất tương ứng được tính bằng công thức (3.2) gọi là phân phối theo quy luật nhị thức với 2 tham số n và p, được ký hiệu là: $B_i(n, p)$

Bảng phân phối xác suất của biến ngẫu nhiên X ~ $B_i(n, p)$.

X	0	1	...	n
P	$C_n^0 P^0 (1-P)^{n-0}$	$C_n^1 P^1 (1-P)^{n-1}$	...	$C_n^n P^n (1-P)^{n-n}$

Chú ý: Đôi khi bài toán đòi hỏi phải t×m xác suất để biến ngẫu nhiên X phân phối theo quy luật nhị thức nhận giá trị trong khoảng [x, x+h] với h là 1 số nguyên dương, khi đó:

$$P(x \le X \le x+h) = P_x + P_{x+1} + ... + P_{x+h} \qquad (3.3)$$

Ở đây: P_i $(i = \overline{x, x+h})$ được xác định bởi (3.2)

IV.2.2. Các tham số đặc trưng

Người ta chứng minh được rằng: Nếu X ~ B_i(n, p) th×

* Kỳ vọng toán: $E(X) = nP$

* Phương sai : $V(X) = nP(1-P)$

* Độ lệch tiêu chuẩn: $\sigma_x = \sqrt{nP(1-P)}$

* Mốt: Mốt của biến ngẫu nhiên X, ký hiệu x_0 là giá trị tương ứng với xác suất lớn nhất trong dãy phân phối.

Thí dụ: Biến ngẫu nhiên X có bảng phân phối xác suất :

	0	3	4
	0,1	0,5	0,4

Khi đó mốt $x_0 = 3$.

Chú ý: Nếu X ~ B_i(n, p) thì $(np + p - 1 \leq x_0 \leq np + p)$ (3.4)

Thí dụ: Một phân xưởng có 5 máy hoạt động độc lập . Xác suất để trong ca mỗi máy bị hỏng đều là 0,1.

a. Tìm xác suất để trong ca có không quá 2 máy hỏng.

b. Tìm số máy hỏng trung bình .

Giải

* Nếu coi sự hoạt động của mỗi máy là một phép thử thì ta có 5 phép thử độc lập nhau (n = 5).

Do đó nếu gọi X là "số máy hỏng trong ca" thì X là biến ngẫu nhiên rời rạc, X ~ B_i(n, p) với 2 tham số n = 5 và p = 0,1.

a. Xác suất trong ca có không quá 2 máy hỏng:

$$P(0 \le X \le 2) = P_0 + P_1 + P_2$$

$$= C_5^0(0{,}1)^0(0{,}9)^5 + C_5^1(0{,}1)^1(0{,}9)^4 + C_5^2(0{,}1)^2(0{,}9)^3$$

$$\Rightarrow P(0 \le X \le 2) = 0{,}9914$$

b. Số máy hỏng trung bình trong ca chính là kỳ vọng toán E(X):

$$E(X) = np = 5.0{,}1 = 0{,}5 \text{ máy}$$

IV. 3. Quy luật Poisson – P(λ)

Quy luật Poisson có thể đượcc coi như một trường hợp riêng của quy luật nhị thức. Người ta đã chứng minh được rằng khi $n \to \infty; p \to 0$ và $np = \lambda$ không đổi. Công thức Bernoulli sẽ hội tụ về công thức Poisson sau đây:

$$P_x = e^{-\lambda}\frac{\lambda^x}{x!} \qquad (3.5)$$

với x = 0, 1, 2,... ; e = 2,71828...

IV.3.1. Định nghĩa:

Biến ngẫu nhiên rời rạc X nhận 1 trong các giá trị có thể có 0, 1, 2, ... với xác suất tương ứng được tính bằng công thức (3.5) gọi là phân phối theo quy luật Poisson với tham số λ và ký hiệu là P(λ).

Bảng phân phối xác suất của biến ngẫu nhiên X ~ P(λ) có dạng như sau :

X	0	1	...	x .....
P	$e^{-\lambda}\frac{\lambda^0}{0!}$	$e^{-\lambda}\frac{\lambda^1}{1!}$	...	$e^{-\lambda}\frac{\lambda^x}{x!}$ ......

IV.3.2. Các tham số đặc trưng của quy luật Poisson

Nếu X ~ P(λ) thì:

+ Kỳ vọng toán: E(X) = λ.

+ Phương sai: V(X) = λ

+ Độ lệch tiêu chuẩn: $\sigma_X = \sqrt{\lambda}$

+ Mốt: ($\lambda - 1 \leq x_0 \leq \lambda$)

Thí dụ: Xác suất để trong khi vận chuyển mỗi chai rượu bị vỡ là 0,001. Người ta vận chuyển 2000 chai rượu đến cửa hàng.

a. Tìm số chai vỡ trung bình khi vận chuyển.

b. Tìm số chai vỡ có khả năng nhiều nhất khi vận chuyển.

Giải

Bài toán thoả mãn lược đồ Bernoulli.

Vì n = 2000 khá lớn. p = 0,001 khá nhỏ

np = 2000.0,001 = 2 không đổi

Do đó, nếu gọi X là "số chai rượu bị vỡ khi vận chuyển" thì X là biến ngẫu nhiên rời rạc tuân theo quy luật Poisson.

a. Số chai vỡ trung bình chính là kỳ vọng toán của X. Ta có :

E(X) = λ = 2 chai.

b. Số chai vỡ có khả năng xảy ra nhiều nhất chính là mốt

$$(\lambda - 1 \leq x_0 \leq \lambda)$$

$$(1 \leq x_0 \leq 2)$$

Vậy $x_0 = 1$ và $x_0 = 2$.

IV.4. Quy luật phân phối chuẩn - $N(\mu,\sigma^2)$

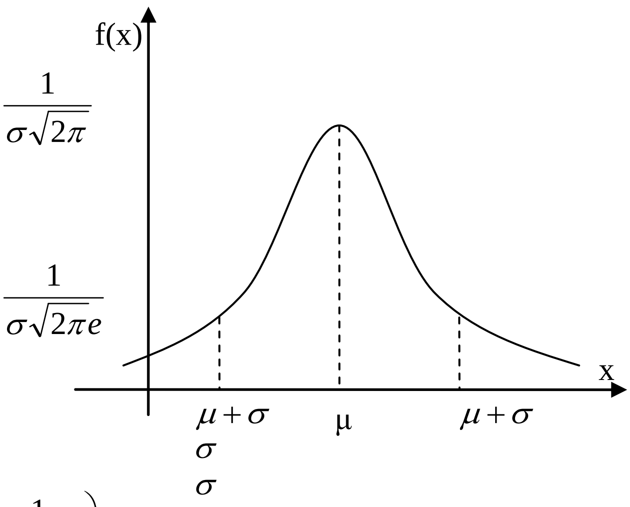
IV. 4.1. Định nghĩa:

Biến ngẫu nhiên X nhận giá trị trong khoảng $(-\infty, +\infty)$ gọi là phân phối theo quy luật chuẩn với 2 tham số $\mu \& \sigma^2$, trong đó σ > 0, nếu hàm mật độ xác suất của nó có dạng :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad (3.7)$$

và ký hiệu là $N(\mu, \sigma^2)$.

Đồ thị hàm mật độ xác suất f(x) của biến ngẫu nhiên $X \sim N(\mu, \sigma^2)$ có dạng:



Điểm cực đại: $\left(\mu, \frac{1}{\sigma\sqrt{2\pi}}\right)$

Điểm uốn : $\left(\mu-\sigma, \frac{1}{\sigma\sqrt{2\pi e}}\right)$; $\left(\mu+\sigma, \frac{1}{\sigma\sqrt{2\pi e}}\right)$

* Hàm phân phối xác suất của biến ngẫu nhiên $X \sim N(\mu, \sigma^2)$ có dạng:

$$F(x) = \int_{-\infty}^{+\infty} f(x)dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \qquad (3.8)$$

IV.4.2. Các tham số đặc trưng của quy luật chuẩn

Vì X là biến ngẫu nhiên liên tục nên:

$$E(X)=\int_{-\infty}^{+\infty}x.f(x)dx \text{ và } V(X)=\int_{-\infty}^{+\infty}x^2.f(x)dx-[E(X)]^2$$

Người ta chứng minh được rằng:

* Kỳ vọng toán: $E(X)=\mu$

* Phương sai: $V(X)=\sigma^2$

* Độ lệch tiêu chuẩn: $\sigma_X=\sqrt{V(X)}=\sigma$

Giả sử biến ngẫu nhiên $X \sim N(\mu,\sigma^2)$ với $E(X)=\mu$ và $V(X)=\sigma^2$

Xét biến ngẫu nhiên :

$$U=\frac{(X-\mu)}{\sigma}$$

IV.4. 3 Phân phối chuẩn hóa N(0,1)

Định nghĩa: Biến ngẫu nhiên U nhận các giá trị trong khoảng $(-\infty,+\infty)$ gọi là tuân theo quy luật phân phối chuẩn hoá nếu hàm mật độ xác suất của nó có dạng :

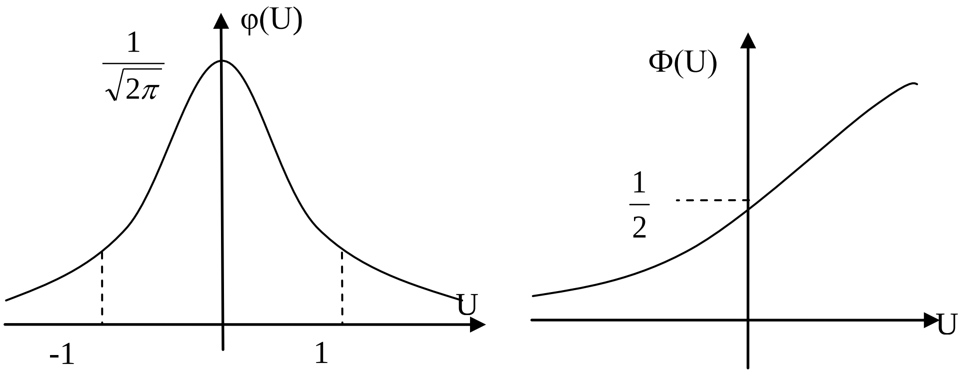
$$\varphi(U)=\frac{1}{\sqrt{2\pi}}e^{-\frac{U^2}{2}} \qquad (3.9)$$

và được ký hiệu $N(0,1)$.

* Hàm phân phối xác suất của biến ngẫu nhiên U có dạng:

$$\Phi(u)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{u}e^{-\frac{U^2}{2}}du \qquad (3.10)$$

* Đồ thị của $\varphi(u)$ và $\Phi(u)$ có dạng như sau:



Do hàm $\varphi(u)$ là đối xứng nhau qua gốc toạ độ nên :

$$P(-\infty < U < u) = P(-\infty < U < 0) + P(0 \le U < u)$$

$$\int_{-\infty}^{+\infty} \varphi(u)du = \frac{1}{2} + \int_0^u \varphi(u)du$$

hay: $$\Phi(u) = \frac{1}{2} + \Phi_0(u)$$

trong đó $\Phi_0(u) = \frac{1}{\sqrt{2\pi}} \int_0^u e^{-\frac{u^2}{2}} du$ gọi là tích phân Laplace.

* Hàm $\Phi_0(u)$ có các tính chất :

+ Là hàm lẻ : $\Phi_0(-u) = -\Phi_0(u)$

+ Giá trị của $\Phi_0(u)$ được tính sẵn trong 1 bảng. (Bảng phụ lục)

Thí dụ:

$$\Phi_0(1{,}96) = 0{,}4750$$
$$\Phi_0(0{,}32) = 0{,}1255$$
$$\Phi_0(3) = 0{,}49865$$

Người ta cũng đã chứng minh được rằng: $E(U) = 0$ và $V(U) = 1$.

IV. 4.4. Giá trị tới hạn chuẩn

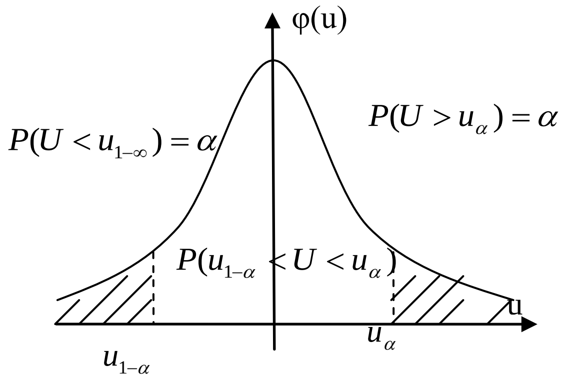
Giá trị tới hạn chuẩn mức α, ký hiệu u_α , là giá trị của biến ngẫu nhiên U có phân phối chuẩn hoá thoả mãn điều kiện :

$$P(U > u_\alpha) = \alpha$$

hay: $$P(U > u_\alpha) = \Phi(u_\alpha) = \int_{u_\alpha}^{+\infty} \varphi(u)du = \alpha$$

Điều đó có nghĩa là : cho trước α ta tính được u_α và ngược lại.

Giá trị u_α tương ứng với mức α được tính sẵn trong 1 bảng.



Tính chất của giá trị tới hạn chuẩn: $u_{1-\alpha} = -u_\alpha$

Thí dụ: $u_{0,025} = 1,96$

$u_{0,05} = 1,645$

$u_{0,975} = - u_{0,025} = - 1,96$

$u_{0,95} = - u_{0,05} = - 1,645$

IV. 4.5. Công thức tính xác suất để biến ngẫu nhiên $X \sim N(\mu,\sigma^2)$ nhận giá trị trong khoảng (a , b)

Giả sử $X \sim N(\mu,\sigma^2)$

thì : $$P(a \le X < b) = \Phi_0(\frac{b-\mu}{\sigma}) - \Phi_0(\frac{a-\mu}{\sigma}) \qquad (3.11)$$

Thí dụ: Kích thước chi tiết máy do 1 máy sản xuất ra là 1 biến ngẫu nhiên phân phối chuẩn với $\mu = 5$ cm và $\sigma = 0,81$cm. Tìm xác suất để lấy ngẫu nhiên 1 chi tiết có kích thước từ 4cm đến 7cm.

Giải

Gọi X là kích thước chi tiết $\to X \sim N(\mu, \sigma^2)$ với $\mu = 5$; $\sigma = 0{,}81$.

Vậy theo (3.11):

$$P(4 \le X \le 7) = \Phi_0(\frac{7-5}{0{,}81}) - \Phi_0(\frac{4-5}{0{,}81})$$

$$= \Phi_0(2{,}46) + \Phi_0(1{,}23) = 0{,}8838$$

IV.4.6. Xác suất của sự sai lệch giữa biến ngẫu nhiên X và kỳ vọng toán của nó

$$P(|X-\mu| < \varepsilon) = P(-\varepsilon < X - \mu < \varepsilon)$$

$$= P(\mu - \varepsilon < X < \mu + \varepsilon)$$

Theo (3.11) ta có:

$$P(|X-\mu| < \varepsilon) = 2\Phi_0\left(\frac{\varepsilon}{\sigma}\right) \qquad (3.12)$$

Thí dụ: Các vòng bi do máy tự động sản xuất ra được coi là đạt tiêu chuẩn nếu đường kính của nó sai lệch so với đường kính thiết kế không quá 0,7 mm. Biết rằng sai lệch này là biến ngẫu nhiên tuân theo quy luật phân phối chuẩn với $\mu = 0$; $\sigma = 0{,}4$ mm. Tìm tỷ lệ vòng bi đạt tiêu chuẩn của máy đó.

Giải

Tỷ lệ vòng bi đạt tiêu chuẩn chính là xác suất để lấy ra ngẫu nhiên 1 vòng bi được vòng bi đạt tiêu chuẩn.

Gọi X là sai lệch giữa đường kính của vòng bi được sản xuất so với đường kính thiết kế $\to X \sim N(\mu, \sigma^2)$ với $\mu = 0$; $\sigma = 0{,}4$.

Theo công thức (3.12) ta có:

$$P(|X-\mu| < \varepsilon) = P(|X-0| < 0{,}7)$$

$$= 2\Phi_0\left(\frac{0{,}7}{0{,}4}\right) = 2\Phi_0(1{,}75)$$

$$= 2.0{,}4599 = 0{,}9198$$

Vậy tỷ lệ vòng bi đạt tiêu chuẩn của máy đó là: 91,98%.

IV. 4.7. Quy tắc " Ba xích ma"

Nếu trong công thức (3.12) ta thay $\varepsilon = 3\sigma$ tức là bằng 3 lần độ lệch tiêu chuẩn thì:

$$P(|X-\mu| < 3\sigma) = 2\Phi_0(3) = 2.0,49865 = 0,9973$$

hay: $$P(\mu - 3\sigma < X < \mu + 3\sigma) = 0,9973$$

IV.5. Quy luật khi bình phương - $\chi^2(n)$

IV.5.1. Định nghĩa

Nếu các X_i với i = 1, 2,..., n là các biến ngẫu nhiên độc lập cùng tuân theo quy luật phân phối chuẩn hóa $N(0,1)$ thì biến ngẫu nhiên $\chi^2 = \sum_{i=1}^{n} X_i^2$ sẽ tuân theo quy luật phân phối xác suất gọi là quy luật " Khi bình phương" với n bậc tự do và ký hiệu $\chi^2(n)$.

IV.5.2. Các tham số đặc trưng:

Người ta chứng minh được rằng: Nếu $\chi^2 \sim \chi^2(n)$ thì:

* Kỳ vọng toán của X: $E(\chi^2) = n$
* Phương sai của X: $V(\chi^2) = 2n$

IV.5.3. Giá trị tới hạn Khi bình phương

Giá trị tới hạn "Khi bình phương" mức α, ký hiệu $\chi^2_\alpha(n)$ là giá trị của biến ngẫu nhiên χ^2 tuân theo quy luật " Khi bình phương với n bậc tự do" thoả mãn điều kiện :

$$P\left(\chi^2 > \chi^2_\alpha(n)\right) = \alpha$$

Nếu cho ta α ta sẽ tìm được giá trị tới hạn $\chi^2_\alpha(n)$ nhờ bảng phụ lục đã có

Chẳng hạn: $\chi^2_{0,025}(24) = 39,36$ và $\chi^2_{0,99}(20) = 8,26$

Chú ý: 1. Với cùng bậc tự do như nhau, khi α càng bé thì $\chi^2_\alpha(n)$ càng lớn. Chẳng hạn: $\chi^2_{0,95}(15) = 7,261$; $\chi^2_{0,05}(15) = 25,00$

2. Khi số bậc tự do n tăng lên và khá lớn thì $\chi^2(n) \sim N(\mu, \sigma^2)$

IV.6. Quy luật phân phối Student – $T^{(n)}$

IV.6.1. Định nghĩa:

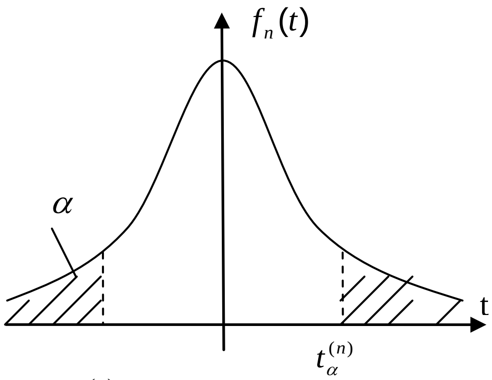
Biến ngẫu nhiên T được gọi là tuân theo quy luật phân phối Student với n bậc tự do và ký hiệu $T^{(n)}$, nếu hàm mật độ xác suất của nó có dạng:

$$f_n(t) = \frac{\Gamma\left(\frac{n}{2}\right)}{\sqrt{\pi(n-1)}.\Gamma(n-1)}.\left[1 + \frac{t^2}{n-1}\right]^{-\frac{n}{2}} \qquad (3.14)$$

với $(-\infty < t < +\infty)$

trong đó Γ (x) gọi là hàm Gama

Đồ thị của hàm $f_n(t)$



$t_{1-\alpha}^{(n)}$

$t_\alpha^{(n)}$

IV.6.2. Các tham số đặc trưng

$t_{1-\alpha}^{(n)}$

Người ta chứng minh được rằng: Nếu biến ngẫu nhiên $T \sim T^{(n)}$ thì:

* Kỳ vọng toán: $E(T) = 0$

* Phương sai: $V(T) = \dfrac{n}{n-2}$

IV.6.3. Giá trị tới hạn Student

Giá trị tới hạn Student mức α, ký hiệu $t_{\alpha}^{(n)}$, là giá trị của biến ngẫu nhiên T phân phối theo quy luật Student với n bậc tự do thoả mãn điều kiện :

$$P\left(T > t_{\alpha}^{(n)}\right) = \int_{t_\alpha}^{+\infty} f(t).dt = \alpha$$

Giá trị tới hạn Student có tính chất :

$$t_{1-\alpha}^{(n)} = -t_{\alpha}^{(n)}$$

Nếu cho α sẽ tìm được giá trị $t_{\alpha}^{(n)}$ nhờ bảng tính sẵn giá trị (xem phụ lục).

Thí dụ: $t_{0,05}^{(20)} = 1,725 \Rightarrow t_{0,95}^{(20)} = -t_{0,05}^{(20)} = -1,725$;

$t_{0,01}^{(15)} = 2,602 \Rightarrow t_{0,99}^{(15)} = -t_{0,01}^{(15)} = -2,602$.

Chú ý: Khi số bậc tự do n tăng lên thì phân phối Student hội tụ rất nhanh về phân phối chuẩn hoá. Do đó, khi n > 30 có thể dùng phân phối chuẩn hoá thay cho phân phối Student. Chẳng hạn:

$$t_{0,025}^{(100)} \approx U_{0,025} = 1,96 \qquad t_{0,025}^{(45)} \approx U_{0,025} = 1,96$$

IV.7. Quy luật Fisher-Snedecor - f(n_1, n_2)

IV.7.1. Định nghĩa

Trong thực tế quy luật Fisher- Snedecor thường được định nghĩa theo sự hình thành như sau:

Giả sử χ_1^2 và χ_2^2 là các biến ngẫu nhiên độc lập cùng phân phối theo quy luật χ^2 với số bậc tự do tương ứng là n_1 và n_2. Khi đó biến ngẫu nhiên :

$$F = \frac{\chi_1^2 / n_1}{\chi_2^2 / n_2}$$

được gọi là phân phối theo quy luật Fisher – Snedecor với n_1 và n_2 bậc tự do.

IV.7. 2. Các tham số đặc trưng

Người ta đã chứng minh được rằng: nếu F ~ F(n_1 , n_2) thì:

$$* \quad E(F) = \frac{n_2}{n_2 - 2}$$

$$* \quad V(F) = \frac{2n_2^{\,2}\left(n_1 + n_2^{\,2} + 2\right)}{n_1(n_2 - 2)^2(n_2 - 4)}$$

IV.7.3. Giá trị tới hạn Fisher – Snedecor

Giá trị tới hạn Fisher – Snedecor mức α, kí hiệu $f_\alpha(n_1, n_2)$, là giá trị của biến ngẫu nhiên F phân phối theo quy luật Fisher – Snedecor với n_1 và n_2 bậc tự do thoả mãn điều kiện :

$$P(F > f_\alpha(n_1, n_2)) = \alpha$$

Giá trị tới hạn $f_\alpha(n_1, n_1)$ có tính chất :

$$f_{1-\alpha}(n_1, n_2) = \frac{1}{f_\alpha(n_2, n_1)} \qquad (3.16)$$

Thí dụ: Với mức α = 0,05 thì $f_{0,05}(6{,}10) = 3{,}22$; $f_{0,05}(3{,}16) = 3{,}24$

Với α = 0,95, ta có: $f_{0,95}(10;6) = \dfrac{1}{f_{0,05}(6;10)} = \dfrac{1}{3{,}22} = 0{,}31$

TÓM LƯỢC

1. Có 3 hình thức thường được sử dụng để mô tả quy luật phân phối xác suất của biến ngẫu nhiên: (a) bảng phân phối xác suất của biến ngẫu nhiên rời rạc, (b) hàm phân

phối xác suất F(x) dùng cho cả 2 loại biến ngẫu nhiên và (c) hàm mật độ xác suất f(x) của biến ngẫu nhiên liên tục.

2. Các hiện tượng được đặc trưng bởi biến ngẫu nhiên rời rạc thường tuân theo các quy luật Không - Một, Nhị thức hoặc Poisson. Còn những hiện tượng được biểu thị bởi biến ngẫu nhiên liên tục thường có quy luật phân phối chuẩn, Khi bình phương, Student hoặc Fisher- Snedecor.

3. Dựa vào quy luật phân phối xác suất của biến ngẫu nhiên có thể tìm được giá trị các tham số đặc trưng của biến ngẫu nhiên đó (kỳ vọng toán, phương sai, độ lệch tiêu chuẩn), đồng thời nhờ các công thức xác suất của quy luật phân phối tính được khá thuận lợi xác suất để biến ngẫu nhiên tương ứng nhận giá trị cụ thể hoặc nhận giá trị trong một khoảng nào đó.

4. Giá trị tới hạn (định nghĩa, tính chất, bảng giá trị tính sẵn) của các quy luật phân phối liên tục là công cụ hữu ích và không thể thiếu trong việc giải quyết các bài toán suy luận thống kê (ước lượng và kiểm định giả thiết) ở các chương tiếp theo của giáo trình này.

Chúc Anh/ Chị học tập tốt!

TÀI LIỆU THAM KHẢO

1. TỐNG ĐÌNH QUỲ: Giáo trình xác suất thống kê.NXB Giáo dục, 1999

2. NGUYỄN CAO VĂN, TRẦN THÁI NINH: Giáo trình Lý thuyết Xác suất và Thống kê toán, Nhà xuất bản Giáo dục Hà Nội 2002.

3. NGUYỄN THẾ HỆ: Lý thuyết xác suất và thống kê toán, Viện Đại Học Mở Hà Nội 2011

4. NGUYỄN VĂN HỘ: Xác suất thống kê toán , Viện Đại Học Mở Hà Nội, 2001