Buổi 02 – Ngày 26-06-2025 – môn Xác suất thống kê – lớp MA005.E32.LT.CNTT

3/ Xác suất (probability):

Cho bài toán: từ một hộp bi có: 8 bi đỏ + 12 bi xanh + 6 bi vàng + 14 bi đen; ta lấy ra ngẫu nhiên cùng lúc 5 viên bi. Biết rằng các viên bi có cùng kích cỡ, hình dáng và trọng lượng.

Tính khả năng để trong 5 viên bi lấy ra,

a/ Có đúng 4 bi đỏ;

b/ Có ít nhất 3 bi đỏ, ít nhất 1 bi xanh;

c/ Có ít nhất 2 bi đỏ và ít nhất 2 bi vàng;

d/ Có đủ 4 màu bi.

Ta có xác suất của sự kiện A là khả năng xảy ra của A khi ta thực hiện phép thử, và ký hiệu là

$$P(A) = \frac{|A|}{|\Omega|} \quad (\%)$$

với $|A|$ = số cách xảy ra sự kiện A

$|\Omega|$ = số cách xảy ra trường hợp tổng quát.

Tính chất: (**Kolmogorov**)

i/ Ta có: $0 \le P(A) \le 1$;

ii/ $P(\Omega) = 1$;

iii/ $P(\phi) = 0$;

iv/ $P(\bar{A}) = 1 - P(A)$, với $\bar{A}$ là sự kiện đối lập của sự kiện A.

Gợi ý:

Một cách chọn 5 bi từ hộp có 40 bi là 1 mẫu cỡ 5:

$\left[\begin{array}{l}\text{+ không thứ tự;}\\ \text{+ không lặp (mỗi viên bi đều khác nhau)}\end{array}\right.$

⇨ Đây là tổ hợp.

a/ Gọi A = “Có đúng 4 bi đỏ”

Ta có xác suất của A là:

$$P(A)=\frac{|A|}{|\Omega|}=\frac{C_8^4.C_{32}^1}{C_{40}^5}\times 100=\frac{2240}{658008}\times 100\approx 0,34\%$$

b/ Gọi B = {có ít nhất 3 bi đỏ, ít nhất 1 bi xanh}

Các trường hợp có thể xảy ra là:

Trường hợp	Bi đỏ (có tất cả 8 bi)	Bi xanh (có tất cả 12 bi)	Bi khác đỏ và khác xanh (có tất cả 20 bi)	Kết quả
1	3	1	1	$C_8^3\times C_{12}^1\times C_{20}^1$
2	3	2	0	$C_8^3\times C_{12}^2$
3	4	1	0	$C_8^4\times C_{12}^1$

Đáp số: Xác suất cần tìm bằng

$$\frac{C_8^3\times C_{12}^1\times C_{20}^1+C_8^3\times C_{12}^2+C_8^4\times C_{12}^1}{C_{40}^5}=\frac{13440+3696+840}{658008}\approx 0,0273\approx 2,73\%$$

(**TH sai lầm**: chọn 3 bi đỏ; 1 bi xanh; và chọn tiếp 1 bi còn thiếu từ 36 bi

→ đáp số bằng $\frac{C_8^3\times C_{12}^1\times C_{36}^1}{C_{40}^5}\times 100\approx 3,6766\%$).

c/ Gọi C = {có ít nhất 2 bi đỏ và ít nhất 2 bi vàng}

Các trường hợp có thể xảy ra là:

Trường hợp	Bi đỏ (có tất cả 8 bi)	Bi vàng (có tất cả 6 bi)	Bi khác đỏ và khác vàng (có tất cả 26 bi)	Kết quả
1	2	2	1	$C_8^2\times C_6^2\times C_{26}^1$
2	2	3	0	$C_8^2\times C_6^3$
3	3	2	0	$C_8^3\times C_6^2$

Đáp số bằng

$$\frac{C_8^2\times C_6^2\times C_{26}^1+C_8^2\times C_6^3+C_8^3\times C_6^2}{C_{40}^5}=\frac{10920+560+840}{658008}\approx 0,0187\approx 1,87\%$$

d/ Gọi D = {có đủ 4 màu bi}

Các trường hợp có thể xảy ra là:

Trường hợp	Bi đỏ (có tất cả 8 bi)	Bi xanh (có tất cả 12 bi)	Bi vàng (có tất cả 6 bi)	Bi đen (có tất cả 14 bi)	Kết quả
1	1	1	1	2	$C_8^1\times C_{12}^1\times C_6^1\times C_{14}^2$
2	1	1	2	1	$C_8^1\times C_{12}^1\times C_6^2\times C_{14}^1$
3	1	2	1	1	$C_8^1\times C_{12}^2\times C_6^1\times C_{14}^1$
4	2	1	1	1	$C_8^2\times C_{12}^1\times C_6^1\times C_{14}^1$

Đáp số xác suất cần tìm:

$$=\frac{C_8^1\times C_{12}^1\times C_6^1\times C_{14}^2+C_8^1\times C_{12}^1\times C_6^2\times C_{14}^1+C_8^1\times C_{12}^2\times C_6^1\times C_{14}^1+C_8^2\times C_{12}^1\times C_6^1\times C_{14}^1}{C_{40}^5}$$

$$=\frac{52416+20160+44352+28224}{658008}\approx 0,2206\approx 22,06\%$$

4/ <u>Các công thức tính xác suất</u>:

a/ *<u>Công thức cộng xác suất</u>*:

Ta có: P(A + B) = P(A) + P(B) – P(AB);

Nếu A, B xung khắc nhau, nghĩa là AB = ϕ. Khi đó, ta có P(AB) = P(ϕ) = 0, nên

P(A + B) = P(A) + P(B).

Suy ra, khả năng để cả hai A, B không xảy ra là: P($\bar{A}.\bar{B}$) = 1 – P(A + B).

P(A + B + C) = P(A) + P(B) + P(C) – P(AB) – P(BC) – P(AC) + P(ABC) = xác suất để có ít nhất 1 trong 3 sự kiện A, B, C xảy ra.

Suy ra, khả năng để cả 3 sự kiện không xảy ra là: P($\bar{A}.\bar{B}.\bar{C}$) = 1 – P(A + B + C).

b/ *Sự độc lập theo xác suất*:

Hai sự kiện A, B được gọi là độc lập nhau theo xác suất, nếu:

P(AB) = P(A).P(B)

Ba sự kiện A, B, C được gọi là độc lập nhau theo xác suất, nếu:

P(AB) = P(A).P(B);

P(AC) = P(A).P(C);

P(BC) = P(B).P(C).

P(ABC) = P(A).P(B).P(C)

c/ *Công thức xác suất có điều kiện*: (các sự kiện diễn ra theo trình tự thời gian)

Xác suất xảy ra của A với điều kiện sự kiện B đã xảy ra trước đó được ký hiệu là:

$$P(A \mid B) = \frac{P(A.B)}{P(B)}$$

(B đã xảy ra trước A)

d/ *Công thức nhân xác suất*: (các sự kiện diễn ra theo trình tự thời gian)

P(AB) = P(A).P(B|A), nếu A xảy ra trước B;

= P(B).P(A|B), nếu B xảy ra trước A;

Và P(ABC) = P(A).P(B|A).P(C|AB), nếu A xảy ra trước, rồi đến B, đến C;

= P(A).P(C|A).P(B|AC), nếu A xảy ra trước, rồi đến C, đến B;

= P(B).P(A|B).P(C|BA), nếu B xảy ra trước, rồi đến A, đến C;

= P(B).P(C|B).P(A|BC), nếu B xảy ra trước, rồi đến C, đến A;

= P(C).P(A|C).P(B|CA), nếu C xảy ra trước, rồi đến A, đến B;

= P(C).P(B|C).P(A|CB), nếu C xảy ra trước, rồi đến B, đến A.

e/ *Công thức xác suất toàn phần* (công thức xác suất đầy đủ):

Gọi $A_1, A_2, ..., A_n$ là một nhóm đầy đủ, và

A = sự kiện cần quan tâm.

Ta có: $P(A) = P(A_1).P(A \mid A_1) + P(A_2).P(A \mid A_2) + \cdots + P(A_n).P(A \mid A_n)$.

Ngoài ra, ta có:

$$P(A_j \mid A) = \frac{P(A_j.A)}{P(A)} = \frac{P(A_j).P(A|A_j)}{P(A_1).P(A|A_1) + P(A_2).P(A|A_2) + \cdots + P(A_n).P(A|A_n)}.$$

Ta gọi đây là công thức BAYES.

Ví dụ 1: bài 3/ trang 5

Một hộp có 14 lá thăm, trong đó có 4 lá trúng thưởng và 10 lá không trúng thưởng.

Cho SV A lên bắt thăm đầu tiên.

SV B lên bắt thăm tiếp theo.

Hỏi trò chơi này có công bằng hay không? Vì sao?

Giải:

Gọi A = {SV A bắt thăm trúng thưởng}

B = {SV B bắt thăm trúng thưởng}

Suy ra $\bar{A}$ = {SV A bắt thăm không trúng thưởng}

Ta có: $P(A) = \frac{C_4^1}{C_{14}^1} = \frac{4}{14} = \frac{2}{7}$, suy ra $P(\bar{A}) = 1 - P(A) = 1 - \frac{2}{7} = \frac{5}{7}$.

Theo công thức xác suất đầy đủ (xác suất toàn phần), ta có:

$$P(B) = P(A).P(B|A) + P(\bar{A}).P(B|\bar{A})$$

$$= \frac{2}{7} \times \frac{C_3^1}{C_{13}^1} + \frac{5}{7} \times \frac{C_4^1}{C_{13}^1} = \frac{2}{7} \times \frac{3}{13} + \frac{5}{7} \times \frac{4}{13} = \frac{26}{7 \times 13} = \frac{2}{7}$$

Cho nên ta có P(A) = P(B) = 2/7, nên trò chơi là công bằng.

Ví dụ 2: bài 8/ trang 5

Gọi C = {người bệnh bị chết}, suy ra $\bar{C}$ = {người bệnh không chết}

LNT = {người bệnh bị liệt nửa thân}

L2C = {người bệnh bị liệt 2 chân}

KB = {người bệnh khỏi bệnh hoàn toàn}.

a/ Tính xác suất để người bệnh không chết.

Cách 1: P($\bar{C}$) = 1 – P(C) = 1 – 0,1 = 0,9 = 90%

Cách 2: P($\bar{C}$) = P(LNT) + P(L2C) + P(KB) = 0,3 + 0,2 + 0,4 = 0,9 = 90%.

b/ Nếu biết rằng người bệnh không chết, tính xác suất người đó bị tật.

Theo công thức Bayes, ta có:

P(bị tật| $\bar{C}$) = $\frac{P(LNT)+P(L2C)}{P(\bar{C})} = \frac{0,3+0,2}{0,9} = \frac{5}{9} \approx 55,56\%$.

<u>Ví dụ 3</u>: bài 11/ trang 6

Gọi Nam = {người đi khám bệnh là nam}

Nữ = {người đi khám bệnh là nữ}

LS = {người đi khám bệnh mắc bệnh loạn sắc}

[Ta có Nam.Nữ = ϕ; và

Nam + Nữ = toàn bộ dân cư trong vùng = Ω.

Suy ra: Nam, Nữ tạo thành một nhóm đầy đủ].

a/ Theo công thức xác suất đầy đủ, ta có:

P(LS) = P(Nam).P(LS|Nam) + P(Nữ).P(LS|Nữ)

= 0,4*0,1 + 0,6*0,15 = 0,04+0,09 = 0,13 = 13%.

b/ Nếu người này bị loạn sắc, tính xác suất người này là nam.

Ta có công thức Bayes:

P(Nam|LS) = $\frac{P(Nam).P(LS \mid Nam)}{P(LS)} = \frac{0,4 \times 0,1}{0,13} = \frac{4}{13} \approx 30,77\%$.

<u>Ví dụ 4</u>: bài 27/ trang 8

Gọi TQ = {sữa Trung Quốc},

TL = {sữa Thái Lan},

NZ = {sữa New Zealand}.

M = {sữa bị nhiễm Melamine}.

[Ta có: TQ.TL = ϕ;

TQ.NZ = ϕ;

TL.NZ = ϕ;

TQ + TL + NZ = toàn bộ sữa trong thùng = Ω.

Suy ra, TQ, TL, NZ tạo thành 1 nhóm đầy đủ].

a/ Theo công thức xác suất đầy đủ, ta có:

P(M) = P(TQ).P(M|TQ) + P(TL).P(M|TL) + P(NZ).P(M|NZ)

= 0,35*0,2 + 0,2*0,15 + 0,45*0,4 = 0,28 = 28%

b/ Theo công thức xác suất Bayes, ta có:

P(NZ|M) = $\frac{0,45\times 0,4}{0,28} = 64,29\%$.

c/ Giả sử hộp sữa kiểm tra không bị nhiễm Melamine, tính khả năng để hộp sữa này không phải là sữa Trung Quốc.

Cách 1: Ta có

$$P(\bar{M}) = 1 - P(M) = 1 - 0,28 = 0,72 = 72\%$$

Xác suất sữa có nguồn gốc từ Trung Quốc trong số sữa không nhiễm Melamine là:

$$P(TQ\,|\,\bar{M}) = \frac{P(TQ).P(\bar{M}\,|\,TQ)}{P(\bar{M})} = \frac{0,35\times 0,8}{0,72} = \frac{7}{18}.$$

Xác suất cần tìm bằng $1 - P(TQ\,|\,\bar{M}) = 1 - \frac{7}{18} = \frac{11}{18} \approx 61,11\%$.

Cách 2: xác suất sữa không bị nhiễm Melamine là:

$$P(\bar{M}) = 1 - P(M) = 1 - 0,28 = 0,72 = 72\%.$$

Theo công thức Bayes, ta có:

P(không phải sữa TQ|không nhiễm Melamine) là

$$P(\overline{TQ}\,|\,\bar{M}) = \frac{P(TL).P(\bar{M}\,|\,TL) + P(NZ).P(\bar{M}\,|\,NZ)}{0,72}$$

$$= \frac{0,2\times 0,85 + 0,45\times 0,6}{0,72} = \frac{11}{18} \approx 61,11\%$$

Bài: 15/ trang 6,

Gọi *n* là số lần cần bắn của tên lửa (*n* > 0).

Ta có P(tên lửa bắn trúng mục tiêu trong *n* lần) >= 90%.

⇔ P(có ít nhất 1 lần trúng trong *n* lần bắn) >= 0,9.

⇔ P(số lần bắn trúng là >=1 trong *n* lần bắn) >= 0,9.

⇔ 1 – P(số lần bắn trúng < 1 trong *n* lần bắn) >= 0,9.

⇔ 1 – P(số lần bắn trúng = 0 trong *n* lần bắn) >= 0,9.

$\Leftrightarrow$ 1 – P(cả *n* lần bắn đều sai) >= 0,9.

$\Leftrightarrow$ 1 – $(0,4)^n$ >= 0,9.

$\Leftrightarrow$ $(0,4)^n \le 0,1$ $\Rightarrow \log_{0,4}(0,4)^n \ge \log_{0,4}(0,1) \Rightarrow n \ge 2,5129$

Nên *n* = 3.

Vậy cần bắn ít nhất là 3 tên lửa.

Bài: 20/ trang 7

Gọi A = "xạ thủ thứ nhất bắn trúng"

B = "xạ thủ thứ hai bắn trúng"

C = "bia chỉ bị bắn trúng 1 viên".

Ta có khả năng xảy ra của C là:

P(C) = $P(A.\overline{B}) + P(\overline{A}.B) = 0,7 \times 0,4 + 0,3 \times 0,6 = 46\%$.

Theo công thức Bayes, ta có:

$$P(B \mid C) = \frac{0,3 \times 0,6}{0,46} \approx 39,13\%$$.

Bài 1: bài 67/ trang 12

Bài 2: bài 69/ trang 12-13.

Bài 3: bài 70/ trang 13.

Bài 4: bài 114/ trang 19.

Bài 5: bài 143/ trang 23

Bài 6: bài 152 trang 24

Bài 7: bài 160 trang 25

Bài 8: bài 167 trang 26.

Chương 2: BIẾN NGẪU NHIÊN

1/ *Một số khái niệm*:

Biến ngẫu nhiên là một ánh xạ:

X: $\Omega \to \mathbb{R}$.

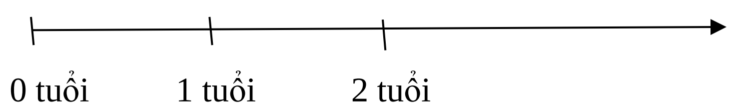
nghĩa là X là một loại biến nhận giá trị thực của mình một cách ngẫu nhiên, tùy thuộc vào kết quả xảy ra khi ta thực hiện phép thử.

* *<u>Phân loại biến ngẫu nhiên (BNN)</u>*: BNN được chia thành 2 loại

a/ **BNN rời rạc (discrete random variable)**: là loại biến có tập hợp các giá trị là những đại lượng rời rạc, đếm được, có hữu hạn phần tử.

b/ **BNN liên tục (continuous random variable)**: là loại biến có tập hợp các giá trị là một đoạn nào đó trên trục số, không đếm được, có vô hạn phần tử.

<u>Ví dụ 1</u>: Tuổi thọ của con người (đơn vị tính là "năm"), là BNN liên tục, được tính từ lúc con người sinh ra đến lúc qua đời.



<u>Ví dụ 2</u>: (Bài toán tung đồng xu ăn tiền). Một người chơi khi tham gia trò chơi này sẽ thảy 2 đồng xu. (Biết rằng khả năng xuất hiện mặt sấp của mỗi đồng xu là 30%). Nếu đồng xu xuất hiện mặt sấp thì người chơi thắng 5000 đồng, còn nếu đồng xu ra mặt ngửa thì thua 3000 đồng. Gọi X là số tiền (thắng hoặc thua) của người chơi sau khi thảy 2 đồng xu. → X là một BNN rời rạc.

Các giá trị mà X nhận được là phụ thuộc vào kết quả xảy ra khi ta thảy 2 đồng xu.

TH1: đồng xu xuất hiện 2 mặt sấp (SS)

⇨ X = 2*5000 đồng = 10000 đồng, với xác suất là:

P(X = 10000 đồng) = P(SS) = 0,3*0,3 = 0,09 = 9%.

TH2: đồng xu xuất hiện 1 sấp, 1 ngửa (SN + NS)

⇨ X = 5000 đồng – 3000 đồng = 2000 đồng, với xác suất là:

P(X = 2000 đồng) = P(SN + NS) = P(SN) + P(NS) = 0,3*0,7+0,7*0,3

= 0,42 = 42%.

TH3: đồng xu xuất hiện 2 mặt ngửa (NN)

⇨ X = 2*(– 3000 đồng) = – 6000 đồng, với xác suất là:

P(X = – 6000 đồng) = P(NN) = 0,7*0,7 = 0,49 = 49%.

Như vậy:

$$X = \begin{cases} 10000 & voi \quad xs \quad 0,09 \\ 2000 & voi \quad xs \quad 0,42 \\ -6000 & voi \quad xs \quad 0,49 \end{cases}.$$

2/ <u>Bảng phân phối xác suất của BNN rời rạc (probability distribution table of discrete random variable)</u>:

là bảng có cấu trúc sau:

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

Trong đó:

$x_1 \le x_2 \le \ldots \le x_n$;

$x_1, x_2, ..., x_n$ là các giá trị thực mà X có thể nhận được khi ta thực hiện phép thử;

$p_i = P(X = x_i) =$ khả năng mà X nhận giá trị là x_i khi ta thực hiện phép thử.

Ta có: $\boxed{p_1 + p_2 + \cdots + p_n = 1.}$

<u>Ví dụ mẫu 1</u>: lập bảng phân phối xác suất cho X là BNN thể hiện cho số tiền thu được của người chơi trong trò chơi tung 2 đồng xu ăn tiền nêu trên.

<u>Giải</u>:

Từ các trường hợp phân tích giá trị của X nêu trên, ta có bảng phân phân phối (xs):

X	-6000	2000	10000
P	0,49	0,42	0,09

Ta có mod(X) = -6000 đồng do xác suất lúc này là 0,49 là cao nhất trong tất cả các xác suất nhận được, nghĩa là khả năng cao nhất là người chơi thua 6000 đồng.

Mod(X) là giá trị mà X có nhiều khả năng nhận được nhất khi ta thực hiện phép thử, nghĩa là xác suất của X khi đó là lớn nhất.

Ta có số tiền trung bình mà người chơi nhận được (ở 1 ván) khi tham gia trò chơi này là:

E(X) = $\overline{X} = -6000 \times 0,49 + 2000 \times 0,42 + 10000 \times 0,09 = -1200$ đồng.

(nghĩa là trung bình 1 ván người chơi thua lỗ 1200 đồng).

Gợi ý: bài 22/ trang 35:

Gọi T là số tiền người chơi nhận được khi đặt 10000 đồng vào mặt "bầu" trong trò chơi lắc bầu cua.

TH1: có 0 mặt bầu => T = -10000 đồng, với xác suất là:

$$P(T=-10000) = \frac{5}{6}\times\frac{5}{6}\times\frac{5}{6}=\frac{125}{216}.$$

TH2: có 1 mặt bầu => T = 10000 đồng, với xác suất là:

$$P(T = 10000 \text{ đồng}) = \frac{1}{6}\times\frac{5}{6}\times\frac{5}{6}+\frac{5}{6}\times\frac{1}{6}\times\frac{5}{6}+\frac{5}{6}\times\frac{5}{6}\times\frac{1}{6}=\frac{75}{216}.$$

TH3: có 2 mặt bầu => T = 20000 đồng, với xác suất là:

P(T = 20000 đồng) =…..

TH4: có 3 mặt bầu => T = 30000 đồng, với xác suất là:

P(T = 30000 đồng) = ….

Nên ta có bảng phân phối xác suất của T là

T	-10000	10000	20000	30000
P	$\frac{125}{216}$	$\frac{75}{216}$		

b/ Tiền trung bình (kỳ vọng) mà người chơi thu được trong 1 ván là:

$$E(T) = \overline{T} = -10000\times\frac{125}{216}+10000\times\frac{75}{216}+20000\times\frac{...}{...}+30000\times\frac{1}{216}=\cdots$$

(ta có Mod(X) = -10000 đồng).

<u>Ví dụ mẫu 2</u>: bài 27/ trang 36

Gọi X là số lượng máy bị hỏng trong khoảng thời gian *t*.

Các giá trị mà X có thể nhận được là:

TH1: có 0 máy bị hỏng => X = 0, với xác suất là:

P(X = 0) = 0,8*0,9*0,7 = 0,504 = 50,4%

TH2: có 1 máy bị hỏng => X = 1, với xác suất là:

P(X = 1) = 0,2*0,9*0,7 + 0,8*0,1*0,7 + 0,8*0,9*0,3 = 0,398=39,8%

TH3: có 2 máy bị hỏng => X = 2, với xác suất là:

P(X = 2) = 0,2*0,1*0,7 + 0,2*0,9*0,3 + 0,8*0,1*0,3 = 0,092 = 9,2%

TH4: có 3 máy bị hỏng => X = 3, với xác suất là:

P(X = 3) = 0,2*0,1*0,3 = 0,006

Nên ta có bảng phân phối xác suất của X là

X	0	1	2	3
P	0,504	0,398	0,092	0,006

Suy ra, Mod(X) = 0, nghĩa là khả năng nhiều nhất là không có máy hỏng trong khoảng thời gian *t*.

b/ Kỳ vọng của X là:

E(X) = [0*0,504+1*0,398+2*0,092+3*0,006] = [0,6] = 1 máy

(với [x] = phép toán lấy phần nguyên của số thực x,

= số nguyên nhỏ nhất, có giá trị >= số thực x,

ví dụ: [1,25] = 2; [10,004] = 11; [12,8] = 13; [9] = 9,…).

Phương sai của X là:

Var(X) = D(X) = $(0 – 0,6)^2$*0,504 + $(1 – 0,6)^2$*0,398 + $(2 – 0,6)^2$*0,092 + $(3 – 0,6)^2$*0,006 = 0,46.

Var = Variance = độ sai lệch

D(X) = Dispersion = độ phân tán.

<u>Cách 2</u>: Var(X) = D(X) = $E(X^2) – [E(X)]^2$ = $\overline{X^2} - \left(\bar{X}\right)^2$

Với

$$E(X^2) = \overline{X^2} = 0^2 * 0,504 + 1^2 * 0,398 + 2^2 * 0,092 + 3^2 * 0,006 = 0,82$$

và

$$E(X) = \bar{X} = 0,6$$

Suy ra, Var(X) = D(X) = 0,82 – 0,6*0,6 = 0,46.

Buổi 03 – Ngày 03-07-2025 – môn Xác suất thống kê – lớp MA005.E32.LT.CNTT

Ví dụ mẫu 2: bài 27/ trang 36

Gọi X là số lượng máy bị hỏng trong khoảng thời gian *t*.

Các giá trị mà X có thể nhận được là:

TH1: có 0 máy bị hỏng => X = 0, với xác suất là:

$P(X = 0) = 0,8*0,9*0,7 = 0,504 = 50,4\%$

TH2: có 1 máy bị hỏng => X = 1, với xác suất là:

$P(X = 1) = 0,2*0,9*0,7 + 0,8*0,1*0,7 + 0,8*0,9*0,3 = 0,398=39,8\%$

TH3: có 2 máy bị hỏng => X = 2, với xác suất là:

$P(X = 2) = 0,2*0,1*0,7 + 0,2*0,9*0,3 + 0,8*0,1*0,3 = 0,092 = 9,2\%$

TH4: có 3 máy bị hỏng => X = 3, với xác suất là:

$P(X = 3) = 0,2*0,1*0,3 = 0,006$

Nên ta có bảng phân phối xác suất của X là

X	0	1	2	3
P	0,504	0,398	0,092	0,006

Suy ra, Mod(X) = 0, nghĩa là khả năng nhiều nhất là không có máy hỏng trong khoảng thời gian *t*.

b/ Kỳ vọng của X là:

E(X) = [0*0,504+1*0,398+2*0,092+3*0,006] = [0,6] = 1 máy

(với [x] = phép toán lấy phần nguyên của số thực x,

= số nguyên nhỏ nhất, có giá trị >= số thực x,

ví dụ: [1,25] = 2; [10,004] = 11; [12,8] = 13; [9] = 9,…).

Phương sai của X là:

$Var(X) = D(X) = (0 – 0,6)^2*0,504 + (1 – 0,6)^2*0,398 + (2 – 0,6)^2*0,092 + (3 – 0,6)^2*0,006 = 0,46$.

Var = Variance = độ sai lệch

D(X) = Dispersion = độ phân tán.

Cách 2: Var(X) = D(X) = E(X^2) – [E(X)]^2 = $\overline{X^2} - \left(\bar{X}\right)^2$

Với

$$E(X^2) = \overline{X^2} = 0^2 * 0,504 + 1^2 * 0,398 + 2^2 * 0,092 + 3^2 * 0,006 = 0,82$$

và

$$E(X) = \bar{X} = 0,6$$

Suy ra, Var(X) = D(X) = 0,82 – 0,6*0,6 = 0,46.

3/ Kỳ vọng và phương sai:

a/ *Kỳ vọng (trung bình) của BNN*:

Là giá trị trung bình lý tưởng khi số lần thực hiện phép thử tăng lên vô hạn.

Khi X là BNN rời rạc, có bảng phân phối xác suất

X	x_1	x_2	…	x_n
P	p_1	p_2	…	p_n

Thì kỳ vọng (Expectation) (trung bình) (mean) là:

E(X) = $\bar{X} = x_1.p_1 + x_2.p_2 + \cdots + x_n.p_n$

Tính chất:

i/ E(C) = C, với C = constant = hằng số;

ii/ E(C.X) = C.E(X), với C = constant = hằng số;

iii/ E(X + Y) = E(X) + E(Y);

iv/ E(X.Y) = E(X).E(Y); nếu X và Y là độc lập nhau

b/ *Phương sai*

Là giá trị trung bình của bình phương sai lệch các giá trị của X so với kỳ vọng;

Người ta thường dùng phương sai để đo lường độ phân tán (độ sai lệch) của dữ liệu xung quanh giá trị trung bình, và ta ký hiệu là: Var(X) = D(X) = σ^2

Var(X) = Variance = độ sai lệch;

D(X) = Dispersion = độ phân tán = σ^2;

$\sigma = \sqrt{Var(X)} = \sqrt{D(X)}$ = sig-ma (độ lệch tiêu chuẩn – standard deviation – SD, là độ lệch cho phép của dữ liệu xung quanh giá trị trung bình).

Ví dụ: tuổi thọ trung bình của bóng đèn dây tóc là X = 2000 (giờ) $\pm \underbrace{10}_{\sigma}$ (giờ)

Cách tính:

Khi X là BNN rời rạc, có bảng phân phối xác suất

X	x_1	x_2	$\cdots$	x_n
P	p_1	p_2	$\cdots$	p_n

Thì phương sai là:

Var(X) = D(X) = $(x_1 - \bar{X})^2.p_1 + (x_2 - \bar{X})^2.p_2 + \cdots + (x_n - \bar{X})^2.p_n$.

Với E(X) = $\bar{X} = x_1.p_1 + x_2.p_2 + \cdots + x_n.p_n$.

Hoặc là: Var(X) = D(X) = E(X^2) – [E(X)]^2 = $\overline{X^2} - \left(\bar{X}\right)^2$.

Với $E(X^2) = \overline{X^2} = x_1^2.p_1 + x_2^2.p_2 + \cdots + x_n^2.p_n$.

Tính chất:

i/ Var(C) = 0, với C = constant = hằng số;

ii/ Var(C.X) = C^2.Var(X), với C = constant = hằng số;

iii/ Var(X + Y) = Var(X) + Var(Y), nếu X, Y là độc lập nhau.

Làm bài: 163a,b/ trang 55; bài 180 trang 58.

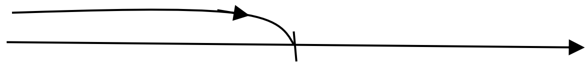
BTVN: bài 21, bài 22/ trang 35

Bài 81/ trang 43

4/ HÀM PHÂN PHỐI XÁC SUẤT CỦA BIẾN NGẪU NHIÊN: (dành cho cả BNN rời rạc và BNN liên tục)

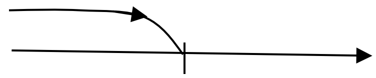
Hàm phân phối xác suất (Probability Distribution Function – PDF) của BNN X là:

$$F(x) = P(X \le x) = P(-\infty < X \le x), \forall x \in \mathbb{R}$$

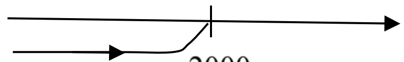


Đây chính khả năng (xác suất) để X nhận giá trị trên khoảng $(-\infty;x]$ khi ta thực hiện phép thử.

Ta có $F(10)=P(X\le 10)$



và $F(-2000)=P(X\le -2000)$



* ***<u>Tính chất</u>***:

i/ $0\le F(x)\le 1$ (do đây cũng là xác suất).

ii/ $F(x)$ là hàm không giảm, nghĩa là

$$F(a)\le F(b),\forall a\le b.$$

iii/ Ta có $P(a<X\le b)=P(-\infty<X\le b)-P(-\infty<X\le a)=F(b)-F(a)$.

iv/ $\lim\limits_{x\to+\infty}F(x)=F(+\infty)=1$.

v/ $\lim\limits_{x\to-\infty}F(x)=F(-\infty)=0$.

* ***<u>Cách tìm hàm phân phối xác suất của BNN rời rạc</u>***:

Cho X là BNN rời rạc, có bảng phân phối xác suất

X	x_1	x_2	x_3	$\ldots$	x_n
P	p_1	p_2	p_3	$\ldots$	p_n

Ta có hàm phân phối xác suất của X là

$$F(x)=\begin{cases}0 \quad khi \quad x<x_1\\ p_1 \quad khi \quad x_1\le x<x_2\\ p_1+p_2 \quad khi \quad x_2\le x<x_3\\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\ p_1+p_2+\cdots+p_{n-1} \quad khi \quad x_{n-1}\le x<x_n\\ 1 \quad khi \quad x\ge x_n\end{cases}$$

<u>Ví dụ mẫu 1</u>: Lập hàm phân phối xác suất cho X là BNN thể hiện cho số tiền mà người chơi nhận được khi tham gia trò chơi tung 02 đồng xu ăn tiền (nêu trên).

<u>Giải</u>:

Ta có bảng phân phối xác suất của X là:

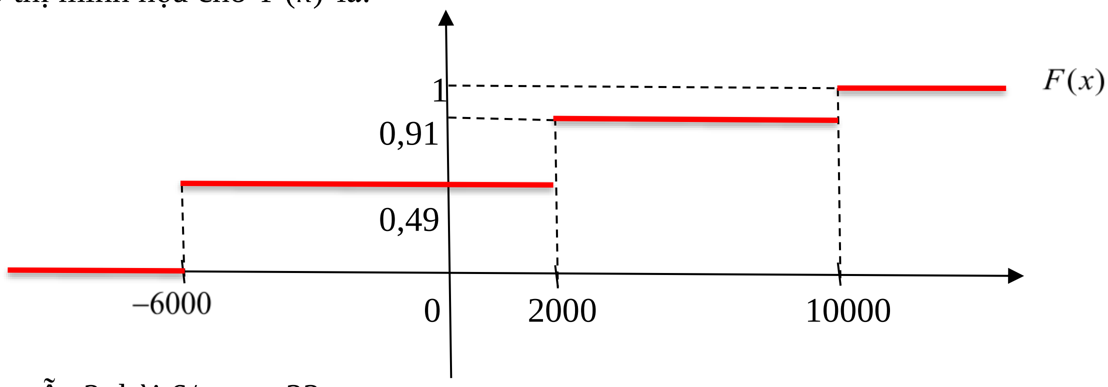
X	-6000	2000	10000
P	0,49	0,42	0,09

Gọi $F(x)=P(X\le x)=P(-\infty<X\le x),\forall x\in\mathbb{R}$ là hàm phân phối xác suất cần tìm.

Ta có:

$$F(x)=\begin{cases}0 \quad khi \quad x<-6000\\ 0,49 \quad khi \quad -6000\le x<2000\\ 0,91 \quad khi \quad 2000\le x<10000\\ 1 \quad khi \quad x\ge 10000\end{cases}$$

và đồ thị minh họa cho $F(x)$ là:



Ví dụ mẫu 2: bài 6/ trang 33

Giải: Các trường hợp có thể xảy ra là:

TH1: $X=1$ (bắn lần 1 và trúng luôn), có xác suất là:

$$P(X=1)=P(T_1)=0,75$$

TH2: $X=2$ (lần 1 bắn sai, lần 2 bắn trúng), có xác suất là:

$$P(X=2)=P(S_1T_2)=0,25\times 0,75=0,1875$$

TH3: $X=3$ (lần 1 sai, lần 2 sai, lần 3 trúng hay là lần 1 sai, lần 2 sai, lần 3 sai), có xác suất là:

$$P(X=3)=P(S_1S_2T_3)+P(S_1S_2S_3)=0,25\times 0,25\times 0,75+0,25\times 0,25\times 0,25=0,25\times 0,25=0,0625$$

a/ Ta có bảng phân phối xác suất của X là:

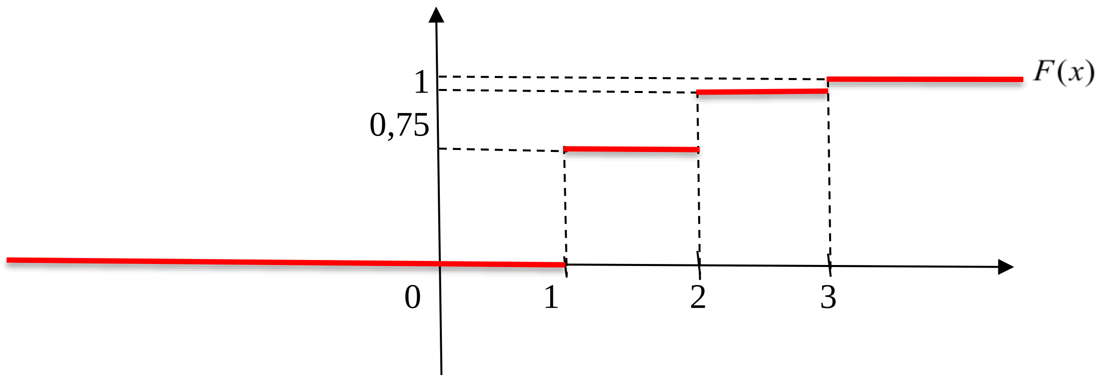
X	1	2	3
P	0,75	0,1875	0,0625

b/ Gọi $F(x)=P(X\le x)=P(-\infty<X\le x),\forall x\in\mathbb{R}$ là hàm phân phối xác suất cần tìm.

Ta có:

$$F(x)=\begin{cases}0 \quad khi \quad x<1\\ 0,75 \quad khi \quad 1\le x<2\\ 0,9375 \quad khi \quad 2\le x<3\\ 1 \quad khi \quad x\ge 3\end{cases}$$

và đồ thị minh họa cho $F(x)$ là:



c/ Tính số lần bắn trung bình là: $E(X) = \overline{X} = [1\times 0,75 + 2\times 0,1875 + 3\times 0,0625] = [1,3125] = 2$ lần.

và ta có: $\text{mod}(X) = 1$ do xác suất $P(X = 1) = 0,75$ là lớn nhất trong các xác suất nhận được.

<u>Ví dụ mẫu 3</u>: bài 27 trang 36

<u>Giải</u>:

Các trường hợp có thể xảy ra là:

TH1: $X = 0$, nghĩa là không có máy nào bị hỏng, có xác suất là:

$$P(X = 0) = 0,8\times 0,9\times 0,7 = 0,504$$

TH2: $X = 1$, nghĩa là có 1 máy hỏng, 2 máy tốt, có xác suất là:

$$P(X = 1) = 0,2\times 0,9\times 0,7 + 0,8\times 0,1\times 0,7 + 0,8\times 0,9\times 0,3 = 0,398$$

TH3: $X = 2$, nghĩa là có 2 máy hỏng, 1 máy tốt, có xác suất là:

$$P(X = 2) = 0,2\times 0,1\times 0,7 + 0,8\times 0,1\times 0,3 + 0,2\times 0,9\times 0,3 = 0,092$$

TH4: $X = 3$, nghĩa là cả 3 máy đều hỏng, có xác suất là:

$$P(X = 3) = 0,2\times 0,1\times 0,3 = 0,006$$

a/ Ta có bảng phân phối xác suất của X là:

X	0	1	2	3
P	0,504	0,398	0,092	0,006

Gọi $F(x) = P(X \le x) = P(-\infty < X \le x), \forall x \in \mathbb{R}$ là hàm phân phối xác suất cần tìm.

Ta có:

$$F(x)=\begin{cases} 0 \quad khi \quad x<0 \\ 0,504 \quad khi \quad 0\le x<1 \\ 0,902 \quad khi \quad 1\le x<2 \\ 0,994 \quad khi \quad 2\le x<3 \\ 1 \quad khi \quad x\ge 3 \end{cases}$$

Từ đây ta có thể tính xác suất: có từ 1 đến 2 máy bị hỏng trong thời gian làm việc t là

$$P(1\le X\le 2)=P(0<X\le 2)=P(-\infty<X\le 2)-P(-\infty<X\le 0)$$
$$=F(2)-F(0)=0,994-0,504=0,49=49\%$$

b/ Tìm kỳ vọng và phương sai:

Ta có kỳ vọng $E(X)=\bar{X}=[0\times 0,504+1\times 0,398+2\times 0,092+3\times 0,006]=[0,6]=1$ máy hỏng.

Phương sai của X là:

$$Var(X)=D(X)=E(X^2)-[E(X)]^2=\overline{X^2}-(\bar{X})^2$$

với

$$E(X^2)=\overline{X^2}=0^2\times 0,504+1^2\times 0,398+2^2\times 0,092+3^2\times 0,006=0,82$$

Suy ra $Var(X)=D(X)=0,82-(0,6)^2=0,46$.

5/ <u>**HÀM MẬT ĐỘ (PHÂN PHỐI) XÁC SUẤT (probability density function – pdf) của BNN liên tục**</u>

Cho X là BNN liên tục, với $F(x)$ là hàm phân phối xác suất $(\forall x\in\mathbb{R})$.

Ta có hàm mật độ xác suất của X là:

$$f(x)=F'(x),\forall x\in\mathbb{R}.$$

Nghĩa là hàm mật độ là đạo hàm của hàm phân phối xác suất.

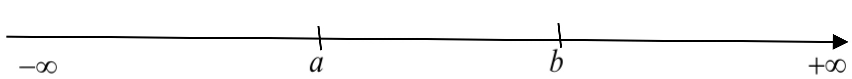
* <u>***Tính chất***</u>:

i/ Ta có $f(x)\ge 0$, do $f(x)$ là đạo hàm của hàm không giảm $0\le F(x)\le 1$.

ii/ $F(x)=P(X\le x)=P(-\infty<X\le x)=\int_{-\infty}^{x} f(t)dt,\quad \forall x\in\mathbb{R}$.

iii/ $P(a<X\le b)=P(-\infty<X\le b)-P(-\infty<X\le a)=\int_{-\infty}^{b} f(x)dx-\int_{-\infty}^{a} f(x)dx$

$$=\left[\int_{-\infty}^{a} f(x)dx+\int_{a}^{b} f(x)dx\right]-\int_{-\infty}^{a} f(x)dx$$
$$=\int_{a}^{b} f(x)dx$$



iv/ Ta có $\int_{-\infty}^{+\infty} f(x)dx = 1$ (*)

v/ Ta có $P(X = a) = P(a < X \le a) = \int_{a}^{a} f(x)dx = 0, \quad \forall a \in \mathbb{R}$, nghĩa là xác suất tại 1 điểm luôn bằng 0 khi X là BNN liên tục.

Khi X là BNN liên tục có hàm mật độ $f(x)$ thì ta có:
+ Kỳ vọng (trung bình) của X là:

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

Suy ra

$$E(8X^2 + 1) = \int_{-\infty}^{+\infty} (8x^2 + 1)f(x)dx$$

+ Phương sai là:

$$Var(X) = D(X) = \int_{-\infty}^{+\infty} (x - \bar{X})^2 f(x)dx$$

Hay là:

$$Var(X) = D(X) = E(X^2) - [E(X)]^2 = \overline{X^2} - [\bar{X}]^2$$

với

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

và

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

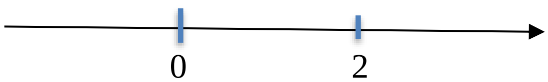
Ví dụ mẫu 1: Cho X là BNN liên tục, có hàm mật độ xác suất

$$f(x) = \begin{cases} A(2x + 3x^2) & khi \quad x \in [0;2] \\ 0 & khi \quad x \notin [0;2] \end{cases}$$

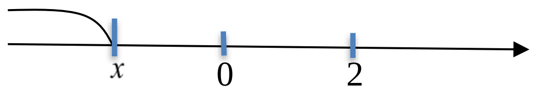
a/ Tìm hằng số A.
b/ Tìm hàm phân phối xác suất $F(x)$ cho X.
c/ Tính xác suất $P\left(-3 \le X \le \frac{3}{2}\right) = ?$
d/ Tìm kỳ vọng và phương sai cho X.
Giải:
a/ Ta có

$$\int_{-\infty}^{+\infty} f(x)dx = 1 \Leftrightarrow \int_0^2 A(2x+3x^2)dx = 1 \Leftrightarrow A[x^2+x^3]_0^2 = 1 \Leftrightarrow 12A = 1 \Leftrightarrow A = \frac{1}{12}$$

b/ Gọi $F(x) = P(X \le x) = P(-\infty < X \le x) = \int_{-\infty}^{x} f(t)dt,\ \ \forall x \in \mathbb{R}$ là hàm phân phối xác suất cần tìm.



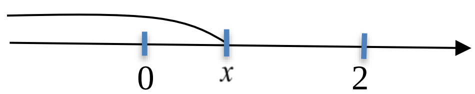
TH1: $x < 0$



Ta có

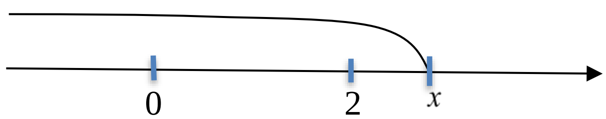
$$F(x) = \int_{-\infty}^{x} f(t)dt = \int_{-\infty}^{x} 0dt = 0$$

TH2: $0 \le x < 2$



$$\Rightarrow F(x) = \int_{-\infty}^{x} f(t)dt = \int_{-\infty}^{0} f(t)dt + \int_0^x f(t)dt = 0 + \int_0^x A(2t+3t^2)dt = \frac{1}{12}\left[t^2+t^3\right]_0^x = \frac{1}{12}(x^2+x^3)$$

TH3: $x \ge 2$



$$\Rightarrow F(x) = \int_{-\infty}^{x} f(t)dt = \int_{-\infty}^{0} f(t)dt + \int_0^2 f(t)dt + \int_2^x f(t)dt = 0 + \int_0^2 A(2t+3t^2)dt + 0 = \frac{1}{12}\left[t^2+t^3\right]_0^2 = \frac{1}{12}(2^2+2^3) = 1$$

Vậy ta có:

$$F(x) = \begin{cases} 0 \quad khi \quad x < 0 \\ \dfrac{1}{12}(x^2+x^3) \qquad khi \qquad 0 \le x < 2 \\ 1 \quad khi \quad x \ge 2 \end{cases}$$

c/ Tính xác suất

$$P\left(-3 \le X \le \frac{3}{2}\right) = \int_{-3}^{3/2} f(x)dx = \int_{-3}^{0} f(x)dx + \int_0^{3/2} f(x)dx = 0 + \int_0^{3/2} A(2x+3x^2)dx$$

$$= \frac{1}{12}\left[x^2+x^3\right]_0^{3/2} = \frac{1}{12}\left(\frac{9}{4}+\frac{27}{8}\right) \approx 0,4688 \approx 46,88\%$$

d/ Tìm kỳ vọng và phương sai cho X

Ta có:

$$E(X)=\overline{X}=\int_{-\infty}^{+\infty}xf(x)dx=\frac{1}{12}\int_{0}^{2}x(2x+3x^2)dx=\frac{13}{9}$$

Phương sai là:

Cách 1:

$$Var(X)=D(X)=\int_{-\infty}^{+\infty}(x-\overline{X})^2f(x)dx=\frac{1}{12}\int_{0}^{2}\left(x-\frac{13}{9}\right)^2(2x+3x^2)dx=\frac{73}{405}$$

Cách 2:

$$Var(X)=D(X)=E(X^2)-[E(X)]^2=\overline{X^2}-(\overline{X})^2$$

với

$$E(X^2)=\overline{X^2}=\frac{1}{12}\int_{0}^{2}x^2(2x+3x^2)dx=\frac{34}{15}$$

$$\Rightarrow Var(X)=D(X)=\frac{34}{15}-\left(\frac{13}{9}\right)^2=\frac{73}{405}$$

Ví dụ mẫu 2: Cho X là BNN liên tục, có hàm mật độ xác suất

$$f(x)=\begin{cases}C\cos(2x) & khi \quad x\in\left[-\frac{\pi}{4};\frac{\pi}{4}\right]\\ 0 & khi \quad x\notin\left[-\frac{\pi}{4};\frac{\pi}{4}\right]\end{cases}$$

a/ Tìm hằng số C.

b/ Tìm hàm phân phối xác suất $F(x)$ cho X

c/ Tính xác suất $P\left(0\le X\le\frac{\pi}{6}\right)=?$

d/ Tìm kỳ vọng và phương sai cho X.

e/ Tìm mod(X), med(X).

Ví dụ mẫu 3: Cho X là BNN liên tục, có hàm mật độ xác suất

$$f(x)=\frac{C}{1+x^2},\quad \forall x\in\mathbb{R}$$

a/ Tìm hằng số C.

b/ Tìm hàm phân phối xác suất $F(x)$ cho X.

c/ Tính xác suất $P(0\le X\le 1)=?$

d/ Tìm kỳ vọng và phương sai cho X.

e/ Tìm mod(X), med(X).

Ví dụ mẫu 4: Cho X là BNN liên tục, có hàm mật độ xác suất

$$f(x)=\begin{cases} C(4x+3x^2+1) & khi \quad x\in[0;4] \\ 0 & khi \quad x\notin[0;4] \end{cases}$$

a/ Tìm hằng số C.

b/ Tìm hàm phân phối xác suất $F(x)$ cho X.

c/ Tính xác suất $P(3\le X\le 5)=?$

d/ Tìm kỳ vọng và phương sai cho X.

e/ Tìm mod(X), med(X).

Buổi 04 – Ngày 10-07-2025 – môn Xác suất thống kê – lớp MA005.E32.LT.CNTT

6/ MỘT SỐ ĐẶC TRƯNG CỦA BIẾN NGẪU NHIÊN

a/ Mốt của BNN:

Mode của BNN X là giá trị có nhiều khả năng xảy ra nhất khi ta thực hiện phép thử, và ta kí hiệu là $\mathrm{mod}(X) = x_0$

TH1: X là BNN rời rạc, có bảng phân phối xác suất

X	x_1	x_2	x_3	$\dots$	x_n
P	p_1	p_2	p_3	$\dots$	p_n

Ta có $\mathrm{mod}(X) = x_0 \Leftrightarrow P(X = x_0) = \max\{p_i\} = \max\limits_{i\in[1;n]}\{P(X = x_i)\}$

TH2: X là BNN liên tục, có hàm mật độ $f(x)$ thì $\mathrm{mod}(X) = x_0$, với x_0 là giá trị cực đại của hàm $f(x)$. Nghĩa là ta tìm x sao cho x là nghiệm của pt: $f'(x) = 0$.

b/ Trung vị (median):

Trung vị của BNN X là giá trị tách đôi hàm phân phối xác suất ra làm 2 miền đều nhau, và ta kí hiệu là $med(X) = x_0$

TH1: X là BNN rời rạc, có bảng phân phối xác suất

X	x_1	x_2	x_3	$\dots$	x_n
P	p_1	p_2	p_3	$\dots$	p_n

Ta đi tìm hàm phân phối xác suất

$$F(x) = \begin{cases} 0 \quad khi \quad x < x_1 \\ p_1 \quad khi \quad x_1 \le x < x_2 \\ p_1 + p_2 \quad khi \quad x_2 \le x < x_3 \\ \dots\dots\dots\dots\dots\dots \\ p_1 + p_2 + \dots + p_{n-1} \quad khi \quad x_{n-1} \le x < x_n \\ 1 \quad khi \quad x \ge x_n \end{cases}$$

Ta có $med(X) = x_0 \Leftrightarrow x_0 = x_i,\ \ i = 1,2,...,n$ sao cho $F(x_i) < \frac{1}{2} \le F(x_{i+1})$.

TH2: X là BNN liên tục, có hàm mật độ $f(x)$ thì $med(X) = x_0$, với x_0 là giá trị làm cho $F(x) = \frac{1}{2}$, nghĩa là $F(x_0) = \int\limits_{-\infty}^{x_0} f(x)dx = \frac{1}{2}$.

Ví dụ mẫu: Cho X là BNN liên tục, có hàm mật độ xác suất

$$f(x) = \begin{cases} A(2x + 3x^2) \quad khi \quad x \in [0;2] \\ 0 \quad khi \quad x \notin [0;2] \end{cases}$$

Tìm mod(X), med(X).

Ta có

$$\int_{-\infty}^{+\infty} f(x)dx = 1 \Leftrightarrow \int_0^2 A(2x+3x^2)dx = 1 \Leftrightarrow A[x^2+x^3]_0^2 = 1 \Leftrightarrow 12A = 1 \Leftrightarrow A = \frac{1}{12}$$

* <u>Tìm mod(X)</u>:

Ta có khi $0 \le x \le 2$ thì

$$f(x) = A(2x+3x^2) = \frac{1}{12}(2x+3x^2) \Rightarrow f'(x) = \frac{1}{12}(2+6x)$$

Giải pt

$$f'(x) = 0 \Leftrightarrow \frac{1}{12}(2+6x) = 0 \Leftrightarrow 2+6x = 0 \Leftrightarrow x = -\frac{1}{3}(l)$$

Ta có:

$$f(0) = \frac{1}{12}(2.0+3.0^2) = 0$$

và

$$f(2) = \frac{1}{12}(2.2+3.2^2) = \frac{16}{12} = \frac{4}{3}$$

Suy ra mod(X) = 2

* <u>Tìm med(X)</u>:

Ta có $med(X) = x_0$

$$\Leftrightarrow F(x_0) = \frac{1}{2} \Leftrightarrow \int_{-\infty}^{x_0} f(x)dx = \frac{1}{2} \Leftrightarrow \int_0^{x_0} \frac{1}{12}(2x+3x^2)dx = \frac{1}{2}$$

$$\Leftrightarrow \int_0^{x_0} (2x+3x^2)dx = 6 \Leftrightarrow \left[x^2+x^3\right]_0^{x_0} = 6 \Leftrightarrow x_0^2 + x_0^3 = 6$$

$$\Leftrightarrow x_0^3 + x_0^2 - 6 = 0$$

$$\Leftrightarrow x_0 \approx 1,5377$$

(nhận)

Vậy $med(X) \approx 1,5377$.

<u>Ví dụ mẫu 2</u>: Cho X là BNN rời rạc có bảng phân phối xác suất

X	–3	0	2	5	8

P	3a	a	2a	1,5a	2,5a

a/ Tìm a.

b/ Tìm hàm phân phối xác suất cho X .

c/ Tìm mod(X), med(X).

Giải:

a/ Ta có: 3a + a + 2a + 1,5a + 2,5a = 1 $\Leftrightarrow$ 10a = 1 $\Leftrightarrow a = \frac{1}{10} = 0,1$.

b/ Gọi $F(x) = P(X \le x) = P(-\infty < X \le x), \quad \forall x \in \mathbb{R}$ là hàm phân phối xác suất cần tìm.

Ta có:

$$F(x) = \begin{cases} 0 \quad khi \quad x < -3 \\ 0,3 \quad khi \quad -3 \le x < 0 \\ 0,4 \quad khi \quad 0 \le x < 2 \\ 0,6 \quad khi \quad 2 \le x < 5 \\ 0,75 \quad khi \quad 5 \le x < 8 \\ 1 \quad khi \quad x \ge 8 \end{cases}$$

c/ Tìm mod(X):

Từ bảng phân phối xác suất của X ta có $\text{mod}(X) = -3$

do $P(X = -3) = 3a = 0,3$ là xác suất lớn nhất trong bảng phân phối.

Tìm med(X):

Ta có

$$med(X) = x_i \Leftrightarrow F(x_i) < \frac{1}{2} \le F(x_{i+1})$$

với $x_i \in \{-3, 0, 2, 5, 8\}$

Từ câu b/ ta có

$$F(0) = 0,4 < \frac{1}{2} \le 0,6 = F(2)$$

Suy ra med(X) = 0

7/ MỘT SỐ LUẬT PHÂN PHỐI XÁC SUẤT THƯỜNG GẶP:

a/ *Luật phân phối Bernoulli*:

Ta gọi X là BNN có luật phân phối Bernoulli, và kí hiệu là $X \sim B(1,p)$ nếu X là BNN rời rạc, có bảng phân phối xác suất dạng:

X	0	1
P	q	p

Khi đó ta có

$$E(X) = \overline{X} = 0.q + 1.p = p$$

$$Var(X) = D(X) = (0-p)^2 q + (1-p)^2 .p = p^2(1-p) + (1-p)^2 p$$
$$= (1-p)p(p+1-p) = (1-p).p = pq$$

với $q = 1 - p$.

* Để kiểm chứng xem X có luật phân phối Bernoulli hay không, ta làm như sau:

$$\begin{cases} \text{+ Ta có phép thử:.....} \\ \text{+ Ta có sự kiện } A \text{ cần quan tâm:......., với xác suất xảy ra } P(A) = p = \cdots \\ \text{+ Ta thực hiện phép thử 1 lần} \\ \text{+ Gọi } X \text{ là số lần xuất hiện sự kiện } A \text{ khi ta thực hiện phép thử 1 lần.} \end{cases}$$

$\Rightarrow X \sim B(1,p)$.

Ví dụ mẫu 3: Tung 1 cục xí ngầu (cục xúc xắc – a dice) 1 lần. Gọi X là số lần xí ngầu xuất hiện mặt có số chấm là số chính phương. Tìm luật phân phối xác suất của X.

Giải:

Ta có mô hình Bernoulli như sau:

+ Phép thử: tung 1 cục xí ngầu.

+ Sự kiện $A = \{$xí ngầu xuất hiện mặt có số chấm là số chính phương$\}$,

với xác suất xảy ra là $P(A) = p = \dfrac{2}{6} = \dfrac{1}{3}$.

+ Ta thực hiện phép thử 1 lần.

+ Gọi X là số lần xí ngầu xuất hiện mặt có số chấm là số chính phương khi ta thực hiện phép thử 1 lần.

$\Rightarrow X \sim B\left(1, \dfrac{1}{3}\right)$.

b/ *Luật phân phối nhị thức (Binomial)*:

Ta gọi X là BNN có luật phân phối nhị thức, và kí hiệu là $X \sim B(n,p)$, nếu X là BNN rời rạc có bảng phân phối xác suất dạng:

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

với $P(X=k) = C_n^k p^k (1-p)^{n-k}$, và $k = 0,1,2,...,n$.

Khi $X \sim B(n,p)$ thì ta có kỳ vọng: $E(X) = \overline{X} = np$ và

phương sai $Var(X) = D(X) = np(1-p)$, cùng với

$\text{mod}(X) = x_0 \Leftrightarrow x_0$ là số nguyên dương thỏa: $np - q \le x_0 \le np + p$.

* *Để kiểm chứng xem X có luật phân phối nhị thức hay không, ta làm như sau*:

+ Ta có phép thử:….
+ Ta có sự kiện $A =$ sự kiện cần quan tâm, với xác suất xảy ra $P(A) = p = \cdots$
+ Ta thực hiện phép thử n lần.
+ Gọi X là số lần xuất hiện sự kiện A trong n phép thử.

$\Rightarrow X \sim B(n,p)$.

Ví dụ mẫu 4: Một cái máy sản xuất ra sản phẩm có xác suất tạo ra sản phẩm lỗi là 5%. Máy sản xuất ra 24 sản phẩm. Gọi X là số sản phẩm lỗi trong 24 sản phẩm làm ra. Tính xác suất sao cho trong 24 sản phẩm làm ra:

a/ Có đúng 3 sản phẩm lỗi.
b/ Có từ 3 đến 5 sản phẩm lỗi.
c/ Có ít nhất 2 sản phẩm lỗi.
d/ Số sản phẩm lỗi trung bình trong 24 sản phẩm.
e/ Số sản phẩm lỗi có nhiều khả năng nhất.

Giải:

Ta có mô hình nhị thức sau:
+ Phép thử: một cái máy sản xuất ra 1 sản phẩm.
+ Sự kiện $A = \{$sản phẩm làm ra bị lỗi$\}$, với xác suất là $P(A) = p = 0,05$
+ Ta lặp lại phép thử $n = 24$ lần.
+ Gọi X là số sản phẩm bị lỗi trong 24 sản phẩm làm ra.

$\Rightarrow X \sim B(24; 0,05)$

a/ Ta có

$$P(X=3) = C_{24}^3 (0,05)^3 (1-0,05)^{24-3} \approx 0,0862 \approx 8,62\%$$

b/ Ta có

$$P(3 \le X \le 5) = P(X=3) + P(X=4) + P(X=5)$$

$$= C_{24}^3 (0,05)^3 (1-0,05)^{24-3} + C_{24}^4 (0,05)^4 (1-0,05)^{24-4} + C_{24}^5 (0,05)^5 (1-0,05)^{24-5}$$

$$\approx 0,1150 \approx 11,50\%$$

c/ Ta có:

$$P(X \geq 2) = 1 - P(X < 2) = 1 - [P(X = 0) + P(X = 1)]$$
$$= 1 - [C_{24}^{0}(0,05)^{0}(1-0,05)^{24-0} + C_{24}^{1}(0,05)^{1}(1-0,05)^{24-1}]$$
$$\approx 0,3392 \approx 33,92\%$$

d/ Số sản phẩm lỗi trung bình là:

$$E(X) = \overline{X} = [np] = [24 \times 0,05] = [1,2] = 2$$

(sản phẩm).

e/ Số sản phẩm lỗi có nhiều khả năng nhất là:

$$\operatorname{mod}(X) = x_0$$

với

$$np - q \leq x_0 \leq np + p$$
$$\Leftrightarrow 24 \times 0,05 - 0,95 \leq x_0 \leq 24 \times 0,05 + 0,05$$
$$\Leftrightarrow 0,25 \leq x_0 \leq 1,25 \Leftrightarrow x_0 = 1$$

(sản phẩm)

Ví dụ mẫu 5: Một người bán hàng đi bán hàng (mỗi ngày đi bán) ở 5 địa điểm khác nhau. Biết rằng khả năng bán được hàng ở mỗi địa điểm là 30%.

a/ Tính xác suất người này bán được hàng trong ngày đi bán.

b/ Tính số ngày bán được hàng nhiều khả năng nhất trong 1 năm, biết rằng người này đi bán 300 ngày trong 1 năm.

Giải:

a/ Ta có mô hình nhị thức như sau:

+ Phép thử: người bán hàng đi bán hàng ở 1 địa điểm trong 1 ngày đi bán.

+ Sự kiện $A =$ {người bán hàng bán được hàng ở 1 địa điểm trong 1 ngày đi bán}, với xác suất là $P(A) = p = 0,3$.

+ Ta lặp lại phép thử $n = 5$ lần.

+ Gọi X là số địa điểm bán được hàng của người bán hàng trong 5 địa điểm mà người này đến bán trong 1 ngày đi bán.

$\Rightarrow X \sim B(5;0,3)$

Ta có P(bán được hàng trong 1 ngày đi bán)=

$$P(X \geq 1) = 1 - P(X < 1) = 1 - P(X = 0)$$
$$= 1 - C_5^0(0,3)^0(1-0,3)^{5-0} = 0,83193 = 83,193\%.$$

b/ Ta có mô hình nhị thức như sau:

+ Phép thử: người bán hàng đi bán hàng 1 ngày trong năm.

+ Sự kiện $B =$ {người bán hàng bán được hàng trong 1 ngày đi bán}, có xác suất là $P(B) = p = 0,83193$.

+ Ta lặp lại phép thử $n = 300$ lần.

+ Gọi Y là số ngày bán được hàng trong 300 ngày đi bán của người bán hàng trong 1 năm.

$\Rightarrow Y \sim B(300; 0,83193)$.

Ta có số ngày bán được hàng có nhiều khả năng nhất trong 300 ngày đi bán là: $\text{mod}(Y) = y_0$, với

$$np - q \le y_0 \le np + p \Leftrightarrow 300 \times 0,83193 - 0,16807 \le y_0 \le 300 \times 0,83193 + 0,83193$$
$$\Leftrightarrow 249,41093 \le y_0 \le 250,41093$$
$$\Leftrightarrow y_0 = 250$$
(ngày).

Ví dụ mẫu 6: bài 88/ trang 44

Giải:

Ta có mô hình nhị thức như sau:

+ Phép thử: là một phép thử độc lập (nào đó) có liên quan đến X.

+ Sự kiện $A =$ {X nhận giá trị trong khoảng (1;3) khi ta thực hiện phép thử độc lập}, với xác suất là:

$$P(A) = p = P(1 \le X \le 3) = \int_1^3 f(x)dx = \int_1^3 \frac{x^2}{9}dx = \left[\frac{x^3}{27}\right]_1^3 = \frac{1}{27}(27-1) = \frac{26}{27}$$

+ Ta lặp lại phép thử $n = 3$ lần.

+ Gọi Y là số lần X nhận giá trị trong khoảng (1;3) khi ta thực hiện phép thử 3 lần.

$\Rightarrow Y \sim B\left(3; \frac{26}{27}\right)$

Ta có P(cần tìm) =

$$P(Y=2) = C_3^2\left(\frac{26}{27}\right)^2\left(1 - \frac{26}{27}\right)^1 = \frac{676}{6561} \approx 10,30\%$$

c/ *Luật phân phối Poisson*:

Ta gọi X là BNN có luật phân phối Poisson, và kí hiệu là $X \sim P(\lambda)$, nếu ta có:

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}, \text{ với } k = 0,1,2,3,...$$

Khi $X \sim P(\lambda)$ thì ta có kỳ vọng $E(X) = \overline{X} = \lambda$ và

phương sai $Var(X) = D(X) = \sigma^2 = \lambda$

* Để kiểm chứng xem X có luật phân phối Poisson hay không, ta làm như sau:

+ Ta có phép thử:….

+ Có sự kiện $A =$ sự kiện cần quan tâm, với xác suất $P(A) = p$ (p nhỏ, nghĩa là $p \le 5\%$)

+ Ta lặp lại phép thử n lần (với n lớn, nghĩa là $n > 30$).

+ Gọi X là số lần xuất hiện A trong khi lặp lại phép thử n lần.

$\Rightarrow X \sim P(\lambda)$, với $\lambda = np$.

Ví dụ mẫu 7: bài 68/ trang 41

Giải:

Ta có mô hình Poisson như sau:

+ Phép thử: tiêm vaccine cho 1 người.

+ Sự kiện $A =$ {người được tiêm vaccine bị phản ứng với vaccine}, với xác suất là

$$P(A) = p = \frac{1}{2000} = 0,0005$$

(p nhỏ)

+ Ta lặp lại phép thử $n = 5000$ lần (n lớn).

+ Gọi X là số trường hợp bị phản ứng thuốc trong 5000 trường hợp tiêm vaccine.

$\Rightarrow X \sim P(\lambda)$, với $\lambda = np = 5000 \times 0,0005 = 2,5$.

a/ P(có 3 trường hợp bị phản ứng) =

$$P(X = 3) = \frac{2,5^3 e^{-2,5}}{3!} \approx 21,38\%$$

b/ P(có nhiều nhất 3 trường hợp bị phản ứng)

$= P(0 \le X \le 3) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)$

$$= \frac{2,5^0 e^{-2,5}}{0!} + \frac{2,5^1 e^{-2,5}}{1!} + \frac{2,5^2 e^{-2,5}}{2!} + \frac{2,5^3 e^{-2,5}}{3!} \approx 75,76\%$$

c/ P(có hơn 3 trường hợp bị phản ứng) = 1 – P(có nhiều nhất 3 trường hợp bị phản ứng)

$= 1 - P(0 \le X \le 3) \approx 24,24\%$

d/ Số trường hợp bị phản ứng vaccine trung bình trong 5000 trường hợp tiêm là:

$$E(X) = \overline{X} = \lambda = [2,5] = 3$$

(trường hợp).

d/ <u>Luật phân phối siêu bội (Hyper Geometry)</u>:

https: hyper text transfer protocol – security/safely

Ta nói X là BNN có luật phân phối siêu bội, và ký hiệu là $X \sim H(N; N_A; n)$, nếu ta có

$$P(X = k) = \frac{C_{N_A}^k \times C_{N-N_A}^{n-k}}{C_N^n}, \text{ với } k = 0,1,2,...,n$$

Nghĩa là trong N phần tử ban đầu, có N_A phần tử có đặc tính A ta cần quan tâm, ta chọn ra n phần tử để khảo sát. Gọi X là số phần tử có đặc tính A cần quan tâm trong n phần tử lấy ra khảo sát. Suy ra $X \sim H(N; N_A; n)$.

Khi đó, ta có kỳ vọng là $E(X) = \overline{X} = np$, với $p = \dfrac{N_A}{N}$

và phương sai là

$$Var(X) = D(X) = \sigma^2 = np(1-p)\left(\frac{N-n}{N-1}\right)$$

với $\dfrac{N-n}{N-1}$ = hệ số hiệu chỉnh = justified factor.

Ví dụ mẫu 8: Một lô hàng có 2000 sản phẩm, trong đó có 400 sản phẩm loại I. Ta lấy ra ngẫu nhiên 25 sản phẩm để kiểm tra. Tính xác suất sao cho trong 25 sản phẩm lấy ra:

a/ Có đúng 4 sản phẩm loại I.

b/ Có ít nhất 2 sản phẩm loại I.

c/ Số sản phẩm loại I trung bình trong 25 sản phẩm.

Giải: Một lô hàng có 2000 sản phẩm, trong đó có 400 sản phẩm loại I. Ta lấy ra ngẫu nhiên 25 sản phẩm để kiểm tra. Gọi X là số sản phẩm loại I trong 25 sản phẩm lấy ra.

Ta có: $N = 2000; N_A = 400; n = 25$. Suy ra $X \sim H(2000;400;25)$

a/ Ta có:

$$P(X=4) = \frac{C_{400}^{4} C_{2000-400}^{25-4}}{C_{2000}^{25}} = \frac{C_{400}^{4} C_{1600}^{21}}{C_{2000}^{25}} \approx 18,74\%$$

b/ Ta có:

$$P(X \ge 2) = 1 - P(X<2) = 1 - P(X=0) - P(X=1)$$

$$= 1 - \frac{C_{400}^{0} C_{1600}^{25}}{C_{2000}^{25}} - \frac{C_{400}^{1} C_{1600}^{24}}{C_{2000}^{25}} \approx 97,33\%$$

c/ Số sản phẩm loại I trung bình trong 25 sản phẩm là:

$$E(X) = \overline{X} = [np] = \left[25 \times \frac{400}{2000}\right] = [5] = 5$$

(sản phẩm).

Ví dụ mẫu 9: bài 265 trang 69

Giải:

Trong 120 chậu lan có 20 chậu lan có hoa màu đỏ, khách hàng chọn ra 15 chậu hoa lan để mua. Gọi X là số chậu hoa lan màu đỏ trong 15 chậu hoa lan mà khách hàng mua.

$\Rightarrow X \sim H(120;20;15)$

a/ Ta có:

$$P(5 \le X \le 6) = P(X=5) + P(X=6) = \frac{C_{20}^{5} C_{100}^{10}}{C_{120}^{15}} + \frac{C_{20}^{6} C_{100}^{9}}{C_{120}^{15}} \approx 7,23\%$$

b/ Ta có

$$E(X) = \overline{X} = np = 15 \times \frac{20}{120} = 2,5$$

$$Var(X) = D(X) = np(1-p)\left(\frac{N-n}{N-1}\right) = 15 \times \frac{1}{6} \times \frac{5}{6} \times \left(\frac{120-15}{120-1}\right) = \frac{125}{68} \approx 1,8382$$

e/ *Luật phân phối đều (Uniform)*:

Ta gọi X là BNN liên tục, có luật phân phối đều trên khoảng $[a;b]$, và kí hiệu là $X \sim U[a;b]$ nếu X có hàm mật độ xác suất:

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

Từ đây ta có thể tìm hàm phân phối xác suất và tìm kỳ vọng, phương sai cho X .

<u>Ví dụ mẫu 10</u>: Cho X là BNN liên tục có luật phân phối đều trên đoạn [1;5].

a/ Tìm hàm mật độ cho X .

b/ Tìm hàm phân phối xác suất cho X .

c/ Tìm kỳ vọng cho X .

d/ Tính xác suất $P(-1\le X\le 2)$.

<u>Giải</u>:

Do $X\sim U[1;5]$ nên ta có

a/ Hàm mật độ xác suất là:

$$f(x)=\begin{cases}\dfrac{1}{5-1} & khi \quad x\in[1;5]\\ 0 & khi \quad x\notin[1;5]\end{cases}$$

nghĩa là

$$f(x)=\begin{cases}\dfrac{1}{4} & khi \quad x\in[1;5]\\ 0 & khi \quad x\notin[1;5]\end{cases}$$

b/ Gọi $F(x)=P(X\le x),\ \ \forall x\in\mathbb{R}$ là hàm phân phối xác suất cần tìm.

TH1: $x<1$, ta có

$$F(x)=\int_{-\infty}^{x} f(t)dt=\int_{-\infty}^{x} 0dt=0$$

TH2: $1\le x<5$, ta có:

$$F(x)=\int_{-\infty}^{x} f(t)dt=\int_{-\infty}^{1} f(t)dt+\int_{1}^{x} f(t)dt=0+\int_{1}^{x}\frac{1}{4}dt=\left[\frac{t}{4}\right]_1^x=\frac{x-1}{4}$$

TH3: $x\ge 5$, ta có

$$F(x)=\int_{-\infty}^{x} f(t)dt=\int_{-\infty}^{1} f(t)dt+\int_{1}^{5} f(t)dt+\int_{5}^{x} f(t)dt=0+\int_{1}^{5}\frac{1}{4}dt+0=\left[\frac{t}{4}\right]_1^5=\frac{5-1}{4}=1$$

Vậy

$$F(x)=\begin{cases}0 & khi \quad x<1\\ \dfrac{x-1}{4} & khi \quad 1\le x<5\\ 1 & khi \quad x\ge 5\end{cases}$$

c/ Kỳ vọng là

$$E(X)=\overline{X}=\int_{-\infty}^{+\infty} xf(x)dx=\int_{1}^{5}\frac{x}{4}dx=3$$

d/ Tính xác suất

$$P(-1\le X\le 2)=\int_{-1}^{2} f(x)dx=\int_{-1}^{1} f(x)dx+\int_{1}^{2} f(x)dx=0+\int_{1}^{2}\frac{1}{4}dx=\frac{1}{4}=25\%$$

f/ *__Luật phân phối chuẩn (Normal)__*:

Ta nói X là BNN liên tục có luật phân phối chuẩn, và kí hiệu là $X\sim N(\mu;\sigma^2)$ nếu X có hàm mật độ xác suất

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\left[\frac{(x-\mu)^2}{2\sigma^2}\right]},\forall x\in\mathbb{R}$$

đọc là "muy"

Khi $X\sim N(\mu;\sigma^2)$ thì ta có kỳ vọng $E(X)=\overline{X}=\mu$ và

phương sai là $Var(X)=D(X)=\sigma^2$

Nếu $\begin{cases}\mu=0\\ \sigma=1\end{cases}$ thì ta nói X là BNN có luật phân phối chuẩn tắc (Standardized Normal) và ta kí hiệu là $X\sim N(0;1)$. Khi đó hàm mật độ của X là

$\varphi(x)=f(x)=\frac{1}{\sqrt{2\pi}}e^{-\left(\frac{x^2}{2}\right)},\forall x\in\mathbb{R}$, và hàm phân phối lúc này là:

$$\phi(x)=F(x)=P(X\le x)=P(-\infty<X\le x)=\int_{-\infty}^{x}\varphi(t)dt=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\left(\frac{t^2}{2}\right)}dt,\quad \forall x\in\mathbb{R}.$$

Buổi 05 – Ngày 17-07-2025 – môn Xác suất thống kê – lớp MA005.E32.LT.CNTT

f/ *Luật phân phối chuẩn (Normal)*:

Ta nói X là BNN liên tục có luật phân phối chuẩn, và kí hiệu là $X \sim N(\mu;\sigma^2)$ nếu X có hàm mật độ xác suất:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\left[\frac{(x-\mu)^2}{2\sigma^2}\right]}, \quad \forall x \in \mathbb{R}.$$

đọc là "muy"

Khi $X \sim N(\mu;\sigma^2)$ thì ta có kỳ vọng $E(X) = \overline{X} = \mu$ và

phương sai là $Var(X) = D(X) = \sigma^2$.

Nếu $\begin{cases} \mu = 0 \\ \sigma = 1 \end{cases}$ thì ta nói X là BNN có luật phân phối chuẩn tắc (Standardized Normal Distribution) và ta kí hiệu là $X \sim N(0;1)$. Khi đó hàm mật độ của X là

$$\varphi(x) = f(x) = \frac{1}{\sqrt{2\pi}} e^{-\left(\frac{x^2}{2}\right)}, \quad \forall x \in \mathbb{R},$$ và hàm phân phối lúc này là:

$$\phi(x) = F(x) = P(X \le x) = P(-\infty < X \le x) = \int_{-\infty}^{x} \varphi(t)dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\left(\frac{t^2}{2}\right)} dt, \quad \forall x \in \mathbb{R}.$$

Ta thường gọi $\phi(x)$ là giá trị của hàm phân phối tại điểm x, và ta thường tra từ bảng phụ lục của BNN có luật phân phối chuẩn tắc (mà ta thường gọi là bảng phân phối Laplace, hay là bảng phân vị chuẩn).

Ví dụ:

$$\phi(1,960) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{1,960} e^{-\left(\frac{t^2}{2}\right)} dt \approx 0,975$$

$$\phi(2,576) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{2,576} e^{-\left(\frac{t^2}{2}\right)} dt \approx 0,995$$

$$\phi(1,645) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{1,645} e^{-\left(\frac{t^2}{2}\right)} dt \approx 0,950$$

Tra từ bảng phụ lục, tiếp theo ta có:

$\phi(1,717) \approx 0,957$

$\phi(2,054) \approx 0,980$

$\phi(3,090) \approx 0,999$

$\phi(0) \approx 0,5$

Suy ra nếu ta có $\phi(c) < 0,5$ thì $c < 0$

$\phi(10) \approx 1$

$\phi(1000) \approx 1$
$\phi(+\infty) \approx 1$
$\phi(-1,881) = ?$
$\phi(-1,555) = ?$

Nhớ quy tắc: $\boxed{\phi(-c) = 1 - \phi(c)}$

Nên ta có:

$$\phi(-1,881) = 1 - \phi(1,881) \approx 1 - 0,970 \approx 0,03 \approx 3\%$$
$$\phi(-1,555) = 1 - \phi(1,555) \approx 1 - 0,940 \approx 0,06 \approx 6\%$$
$$\phi(-\infty) = 1 - \phi(+\infty) \approx 1 - 1 = 0$$

Tra ngược, ta có:

Cho $\phi(c) \approx 0,90 \Rightarrow c \approx 1,282$

Cho $\phi(c) \approx 0,75 \Rightarrow c \approx 0,674$

Cho $\phi(c) \approx 0,980 \Rightarrow c \approx 2,054$

Hướng dẫn tra bảng phụ lục ABC (dùng bảng B: tích phân Laplace)

Ví dụ: tra giá trị $\phi(1,960) = ?$

bằng cách xét: $\begin{cases} \text{+ dòng = 1,9} \\ \text{+ cột = 0,06} \end{cases}$

Ta thấy kết quả là:

$$0,4750 \Rightarrow \varphi(1,960) \approx 0,4750 \Rightarrow \phi(1,960) = \varphi(1,960) + 0,5 \approx 0,975$$

Lưu ý: ta có:

$$\varphi(0) = 0$$
$$\varphi(+\infty) = 0,5$$
$$\varphi(-c) = -\varphi(c)$$
$$\varphi(-\infty) = -\varphi(+\infty) = -0,5$$

Hướng dẫn bấm MT bỏ túi để tra $\phi(c)$ khi biết $c = \cdots$

Gợi ý bằng dòng máy CASIO – fx570ES Plus

MODE → 3: STAT → 1: 1-VAR → AC → SHIFT + 1 → 5: Distr → 1: P(→ nhập số:

Ví dụ: P(1,960), ta nhập: P(1,960) → bấm dấu "=" → cho ra kết quả = 0,975

Khi X là BNN có luật phân phối chuẩn $X \sim N(\mu;\sigma^2)$ thì ta có:

$$P(a \le X \le b) \approx \phi\left(\frac{b-\mu}{\sigma}\right) - \phi\left(\frac{a-\mu}{\sigma}\right)$$

Ví dụ mẫu 1: bài 134 trang 51

Giải: Ta có

$$X \sim N(18;16) \Rightarrow \begin{cases} \mu = 18 \\ \sigma^2 = 16 \end{cases} \Rightarrow \begin{cases} \mu = 18 \\ \sigma = 4 \end{cases}$$

a/ P(cần tìm) =

$$P(12 \le X \le 18) \approx \phi\left(\frac{18-18}{4}\right) - \phi\left(\frac{12-18}{4}\right) \approx \phi(0) - \phi(-1,5) \approx 0,4332 \approx 43,32\%$$

b/ P(cần tìm) =

$$P(0 \le X < 8) = P(0 \le X \le 8) \approx \phi\left(\frac{8-18}{4}\right) - \phi\left(\frac{0-18}{4}\right) \approx \phi(-2,5) - \phi(-4,5) \approx 0,62\%$$

c/ P(cần tìm) =

$$P(X \ge 12) = P(12 \le X < +\infty) \approx \phi\left(\frac{+\infty-18}{4}\right) - \phi\left(\frac{12-18}{4}\right) \approx \phi(+\infty) - \phi(-1.5) \approx 93,32\%$$

Cách 2: P(cần tìm) =

$$1 - P(X < 12) = 1 - P(0 \le X \le 12) \approx 1 - \left[\phi\left(\frac{12-18}{4}\right) - \phi\left(\frac{0-18}{4}\right)\right] \approx$$

d/ Gọi T là khoảng thời gian cần tìm.

Ta có

$$P(T \le X < +\infty) = 0,995$$

$$\Leftrightarrow \phi\left(\frac{+\infty-18}{4}\right) - \phi\left(\frac{T-18}{4}\right) = 0,995$$

$$\Leftrightarrow \phi(+\infty) - \phi\left(\frac{T-18}{4}\right) = 0,995$$

$$\Leftrightarrow 1 - \phi\left(\frac{T-18}{4}\right) = 0,995$$

$$\Leftrightarrow \phi\left(\frac{T-18}{4}\right) = 0,005 < 0,5 \quad (*)$$

Cho nên $\frac{T-18}{4} < 0$ vì vậy, ta có

Biểu thức (*)

$$\Leftrightarrow 1 - \phi\left(\frac{18-T}{4}\right) = 0,005$$

$$\Leftrightarrow \phi\left(\frac{18-T}{4}\right) = 0,995$$

$$\Leftrightarrow \frac{18-T}{4} \approx 2,576 \Leftrightarrow T \approx 7,696$$

Đáp số: T $\approx 7,696$ (tháng)

<u>Ví dụ mẫu 2</u>: bài 136 trang 51

<u>Giải</u>: Ta có

$$X \sim N(4,2;2,25) \Rightarrow \begin{cases} \mu = 4,2 \\ \sigma^2 = 2,25 \end{cases} \Rightarrow \begin{cases} \mu = 4,2 \\ \sigma = 1,5 \end{cases}$$

Gọi Y số tiền thu được khi bán 1 bóng đèn.

$p =$ tỷ lệ bán được bóng đèn có lãi (có lời).

Ta có bảng phân phối xác suất của Y là:

Y	-300000	100000
P	$1-p$	p

Ngoài ra, ta có: kỳ vọng là

$$E(Y) = \overline{Y} = -300000 \times (1-p) + 100000 \times p = 30000$$

$$\Rightarrow p = 0,825$$

Gọi T là thời gian cần phải bảo hành cho khách hàng.

Ta có

$$P(T \le X < +\infty) = p \Leftrightarrow P(T \le X < +\infty) = 0,825$$

$$\Leftrightarrow \phi\left(\frac{+\infty - 4,2}{1,5}\right) - \phi\left(\frac{T-4,2}{1,5}\right) = 0,825$$

$$\Leftrightarrow \phi(+\infty) - \phi\left(\frac{T-4,2}{1,5}\right) = 0,825$$

$$\Leftrightarrow 1 - \phi\left(\frac{T-4,2}{1,5}\right) = 0,825$$

$$\Leftrightarrow \phi\left(\frac{T-4,2}{1,5}\right) = 0,175 < 0,5 \quad (*)$$

Cho nên $\frac{T-4,2}{1,5} < 0$, vì vậy, ta có (*)

$$\Leftrightarrow 1 - \phi\left(\frac{4,2-T}{1,5}\right) = 0,175$$

$$\Leftrightarrow \phi\left(\frac{4,2-T}{1,5}\right) = 0,825$$

Suy ra

$$\frac{4,2-T}{1,5} \approx \frac{0,915+0,954}{2} \approx 0,9345 \Rightarrow T \approx 2,79825$$ (năm).

<u>Lưu ý</u>: Cho $X \sim B(n,p)$, với $p = P(A) = \cdots$ là xác suất xảy ra sự kiện A cần quan tâm khi ta thực hiện phép thử n lần.

Nếu $n > 30$ và $n*p > 5$, ta có $X \xrightarrow{a.s.} N(np; np(1-p))$, nghĩa là X có hội tụ hầu chắc chắn (almost sure – a.s.) về phân phối chuẩn $N(np; np(1-p))$.

Khi đó:

$$P(a \le X \le b) \approx \phi\left(\frac{b-np}{\sqrt{np(1-p)}}\right) - \phi\left(\frac{a-np}{\sqrt{np(1-p)}}\right)$$

(định lý giới hạn trung tâm - Central Limit Theorem)

và ta có công thức xấp xỉ:

$$P(X = k) \approx \frac{1}{\sqrt{np(1-p)}.\sqrt{2\pi}} e^{-\left[\frac{(k-np)^2}{2np(1-p)}\right]}$$

với k là số nguyên, ≥ 0

(định lý giới hạn địa phương – Local Limit Theorem)

Ví dụ mẫu 3: bài 145 trang 52

Giải:

Ta có:

$$X \sim N(100;1) \Rightarrow \begin{cases} \mu = 100 \\ \sigma^2 = 1 \end{cases} \Rightarrow \begin{cases} \mu = 100 \\ \sigma = 1 \end{cases}$$

a/ Tỷ lệ sản phẩm đạt tiêu chuẩn là

$$P(98 \le X \le 102) \approx \phi\left(\frac{102-100}{1}\right) - \phi\left(\frac{98-100}{1}\right) \approx \phi(2) - \phi(-2) \approx 0,9545 \approx 95,45\%$$

b/ Tỷ lệ phế phẩm = 1 – Tỷ lệ sản phẩm đạt tiêu chuẩn = $1 - 0,9545 = 4,55\%$.

c/ Ta có mô hình nhị thức sau:

+ Phép thử: nhà máy sản xuất ra 1 sản phẩm

+ Sự kiện $A = \{$sản phẩm làm ra là đạt tiêu chuẩn$\}$, với xác suất là $P(A) = p = 0,9545$

+ Ta lặp lại phép thử $n = 400$ lần.

+ Gọi Y là số sản phẩm đạt tiêu chuẩn trong 400 sản phẩm làm ra.

$\Rightarrow Y \sim B(400; 0,9545)$

Do $n = 400 > 30$ và $n*p > 5$ nên

$$Y \xrightarrow{a.s.} N(400 \times 0,9545; 400 \times 0,9545 \times 0,0455)$$

hay là $Y \sim N(381,8; 17,3719)$

nên ta có

$$P(Y > 85\% \times 400) = P(Y > 340) = P(341 \le Y \le 400)$$

$$\approx \phi\left(\frac{400-381,8}{\sqrt{17,3719}}\right) - \phi\left(\frac{341-381,8}{\sqrt{17,3719}}\right) \approx 99,999\%$$

d/ Ta có mô hình nhị thức sau:

+ Phép thử: nhà máy sản xuất ra 1 sản phẩm

+ Sự kiện $B = \{$sản phẩm làm ra là đạt tiêu chuẩn$\}$, với xác suất là $P(B) = p = 0,9545$

+ Ta lặp lại phép thử $n = 500$ lần.
+ Gọi Z là số sản phẩm đạt tiêu chuẩn trong 500 sản phẩm làm ra.
$\Rightarrow Z \sim B(500; 0,9545)$

Do $n = 500 > 30$ và $n*p$ >5 nên

$$Z \xrightarrow{a.s.} N(500 \times 0,9545; 500 \times 0,9545 \times 0,0455)$$

hay là $Z \sim N(477,25; 21,714875)$

nên ta có

$$P(60\% \times 500 \le Z \le 90\% \times 500) = P(300 \le Z \le 450)$$

$$\approx \phi\left(\frac{450 - 477,25}{\sqrt{21,714875}}\right) - \phi\left(\frac{300 - 477,25}{\sqrt{21,714875}}\right) \approx 2,4995 \times 10^{-7}\%$$

Ví dụ mẫu 4: bài 160 trang 55

Giải: Gọi X là lãi suất đầu tư vào 1 dự án năm 2012 (đơn vị tính là %).

Ta có: $X \sim N(\mu; \sigma^2)$, với $E(X) = \overline{X} = \mu$ và phương sai $Var(X) = D(X) = \sigma^2$ cần tìm.

Ta có:

$$P(20 < X < +\infty) = P(20 \le X < +\infty) = 0,1587$$

$$\Leftrightarrow \phi\left(\frac{+\infty - \mu}{\sigma}\right) - \phi\left(\frac{20 - \mu}{\sigma}\right) = 0,1587$$

$$\Leftrightarrow 1 - \phi\left(\frac{20 - \mu}{\sigma}\right) = 0,1587$$

$$\Leftrightarrow \phi\left(\frac{20 - \mu}{\sigma}\right) = 0,8413 \quad (1)$$

Ta có:

$$P(25 < X < +\infty) = P(25 \le X < +\infty) = 0,0228$$

$$\Leftrightarrow \phi\left(\frac{+\infty - \mu}{\sigma}\right) - \phi\left(\frac{25 - \mu}{\sigma}\right) = 0,0228$$

$$\Leftrightarrow 1 - \phi\left(\frac{25 - \mu}{\sigma}\right) = 0,0228$$

$$\Leftrightarrow \phi\left(\frac{25 - \mu}{\sigma}\right) = 0,9772 \quad (2)$$

Từ (1), (2), ta có hệ pt

$$\begin{cases} \dfrac{20 - \mu}{\sigma} \approx 0,994 \approx 1 \\ \dfrac{25 - \mu}{\sigma} \approx 1,996 \approx 2 \end{cases} \Rightarrow \begin{cases} \mu \approx 15 \\ \sigma \approx 5 \end{cases}$$

Vậy khả năng để đầu tư không bị thua lỗ là:

$$P(X \geq 0) = P(0 \leq X < +\infty) \approx \phi\left(\frac{+\infty - 15}{5}\right) - \phi\left(\frac{0-15}{5}\right) \approx \phi(+\infty) - \phi(-3) \approx 99,865\%$$

<u>Ví dụ mẫu 5</u>: bài 262 trang 69

<u>Giải</u>: Gọi X là BNN thể hiện cho điểm TOEIC của sinh viên sắp tốt nghiệp ở 1 trường đại học.

Ta có

$$X \sim N(560; 78^2) \Rightarrow \begin{cases} \mu = 560 \\ \sigma = 78 \end{cases}$$

a/ P(cần tìm) =

$$P(600 \leq X \leq 700) \approx \phi\left(\frac{700-560}{78}\right) - \phi\left(\frac{600-560}{78}\right) \approx 26,77\%$$

b/ P(cần tìm) =

$$P(500 < X < +\infty) = P(501 \leq X < +\infty) \approx \phi\left(\frac{+\infty - 560}{78}\right) - \phi\left(\frac{501-560}{78}\right)$$

$$\approx \phi(+\infty) - \phi\left(\frac{501-560}{78}\right) \approx 77,53\%$$

c/ Gọi T là điểm TOEIC tối thiểu cần tìm để sinh viên có thể tốt nghiệp ra trường.

Ta có:

$$P(T \leq X < +\infty) = 0,8 \Leftrightarrow \phi\left(\frac{+\infty - 560}{78}\right) - \phi\left(\frac{T-560}{78}\right) = 0,8$$

$$\Leftrightarrow \phi(+\infty) - \phi\left(\frac{T-560}{78}\right) = 0,8$$

$$\Leftrightarrow 1 - \phi\left(\frac{T-560}{78}\right) = 0,8$$

$$\Leftrightarrow \phi\left(\frac{T-560}{78}\right) = 0,2 < 0,5 \quad (*)$$

Nên $\frac{T-560}{78} < 0$ vì vậy biểu thức (*)

$$\Leftrightarrow 1 - \phi\left(\frac{560-T}{78}\right) = 0,2$$

$$\Leftrightarrow \phi\left(\frac{560-T}{78}\right) = 0,8$$

$$\Leftrightarrow \frac{560-T}{78} \approx 0,842 \Leftrightarrow T \approx [494,324] = 495$$

Đáp số: điểm TOEIC tối thiểu là: 495 điểm.

<u>BTVN</u>:

Bài 1: bài 263 trang 69;

Bài 2: bài 264 trang 69;
Bài 3: bài 237 trang 65;
Bài 4: bài 235 trang 65;
Bài 5: bài 213 trang 62.

CHƯƠNG 3: ƯỚC LƯỢNG THAM SỐ

1/ <u>MỘT SỐ KHÁI NIỆM</u>:

Giả sử ta cần nghiên cứu về đặc tính X trên một tổng thể (Total) T, có số lượng phần tử là N = rất lớn.

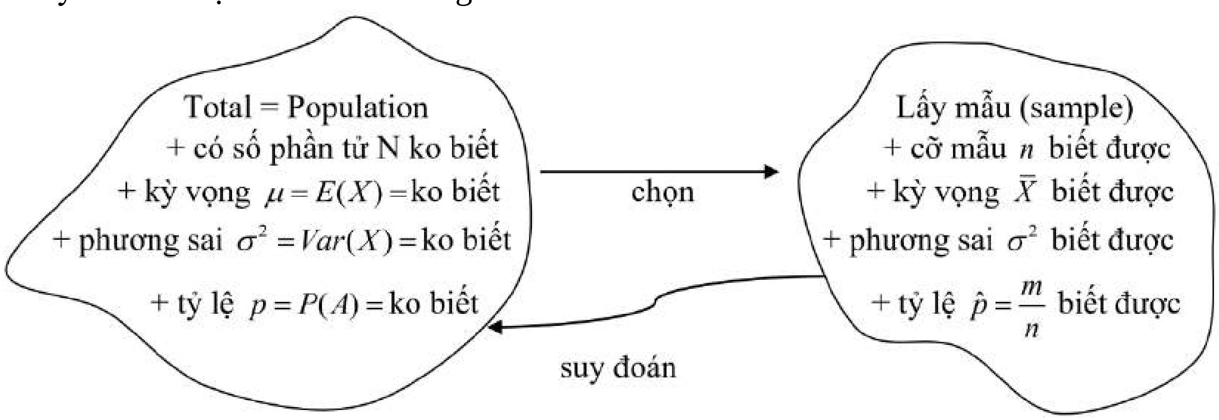
Ta thường không thể khảo sát hết các phần tử của tổng thể T vì những lý do sau:

+ N quá lớn.
+ Thời gian và kinh phí không cho phép.
+ Có thể làm hư hại các phần tử của tổng thể T.

Ví dụ: X = chiều cao (cm)
T = tất cả các SV Việt Nam thời điểm tháng 03/2025.

Ví dụ 2: X = cân nặng (kg)
T = tất cả cá basa vùng Đồng bằng sông Cửu Long, Việt Nam vào thời điểm tháng 03/2025.

Cho nên ta thường chọn ra một số lượng hữu hạn phần tử n từ tổng thể ban đầu, mà ta thường gọi là quá trình lập mẫu cỡ n, để khảo sát đặc tính X, sau đó ghi nhận lại giá trị, rồi suy đoán ra đặc tính X trên tổng thể T.



Để sự suy đoán được chính xác và mang tính khách quan, thì quá trình lấy mẫu phải đảm bảo các nguyên tắc:

+ Các phần tử của mẫu phải được chọn ngẫu nhiên từ tổng thể T.
+ Các phần tử của tổng thể T phải có cùng khả năng được chọn ra làm mẫu như nhau.

+ Quá trình lấy mẫu phải tiến hành khách quan, độc lập nhau. Kết quả của những lần đo không phụ thuộc nhau.

2/ <u>CÁC ĐẶC TRƯNG CỦA MẪU</u>:

a/ *<u>Kỳ vọng mẫu (trung bình mẫu)</u>*:

<u>TH1: Mẫu có lặp</u>:

Ta xét bảng sau:

X	Tần số n_i
x_1	n_1
x_2	n_2
$\ldots$	$\ldots$
x_k	n_k
Tổng	n

Trong đó: $x_1, x_2, ..., x_k$ là các giá trị mà X có thể nhận được khi ta khảo sát;

$n_1, n_2, ..., n_k$ là tần số (frequency) (hay số lần xuất hiện) của các giá trị x_i trong mẫu cỡ *n*.

$n_1 + n_2 + \cdots + n_k = n$;

Khi đó: kỳ vọng mẫu là: E(X) = $\bar{X} = \dfrac{1}{n}\left(n_1x_1 + n_2x_2 + \cdots + n_kx_k\right) = \dfrac{1}{n}\sum_{i=1}^{k}(n_ix_i)$.

<u>Ví dụ</u>: Đo chiều cao X (đơn vị tính là cm) của 25 SV UIT vào tháng 03 năm 2025 ta được số liệu sau:

158	162	170	165	164	158	172	180	184	172	168	160
156	168	167	165	162	160	155,5	164	172	180	164	166
168											

Tìm chiều cao trung bình của 25 SV này.

<u>Giải</u>:

Ta có bảng sau:

X (cm)	n_i
155,5	1
156	1
158	2
160	2
162	2
164	3
165	
166	
168	
…	

<u>TH2: Mẫu chia khoảng</u>:

Ta xét bảng sau:

X	n_i	θ_i
$(a_1, a_2]$	n_1	θ_1
$(a_2, a_3]$	n_2	θ_2
…	…	…
$(a_k, a_{k+1}]$	n_k	θ_k
Tổng	n	

Trong đó: $\theta_i = \frac{a_i + a_{i+1}}{2}$ = điểm giữa của khoảng $(a_i, a_{i+1}]$

Ta có trung bình mẫu là: E(X) = $\bar{X} = \frac{1}{n}\left(n_1\theta_1 + n_2\theta_2 + \cdots + n_k\theta_k\right) = \frac{1}{n}\sum_{i=1}^{k}(n_i\theta_i)$.

<u>Ví dụ</u>: Đo mức tiêu hao nhiên liệu trung bình của một loại động cơ ô tô (đơn vị tính là "lít/100 km"), ta được bộ số liệu sau:

X (lít/ 100 km)	n_i	θ_i
< 4,0	3	3,5
4,0 – 5,0	6	4,5
5,0 – 6,0	12	5,5
6,0 – 7,0	20	6,5
7,0 – 8,0	8	7,5
8,0 – 9,0	5	8,5
> 9,0	2	9,5

Hãy tính mức tiêu hao nhiên liệu trung bình của loại động cơ ô tô này.

Giải: ….

b/ *Phương sai mẫu*:

TH1: mẫu có lặp

Ta xét bảng sau:

X	n_i
x_1	n_1
x_2	n_2
…	…
x_k	n_k
Tổng	n

Ta có:

Phương sai mẫu thực nghiệm (experiment sample variance):

$$s^2 = \frac{1}{n}\left[(x_1 - \bar{X})^2 n_1 + (x_2 - \bar{X})^2 n_2 + \cdots + (x_k - \bar{X})^2 n_k\right] = \frac{1}{n}\sum_{i=1}^{k}\left[n_i(x_i - \bar{X})^2\right]$$

Phương sai mẫu có hiệu chỉnh (justified sample variance):

$$S^2 = \frac{1}{n-1}\left[(x_1-\bar{X})^2 n_1 + (x_2-\bar{X})^2 n_2 + \cdots + (x_k-\bar{X})^2 n_k\right] = \frac{1}{n-1}\sum_{i=1}^{k}\left[n_i(x_i-\bar{X})^2\right]$$

Suy ra, $S^2 = \left(\frac{n}{n-1}\right)s^2$.

Ta gọi: $s =$ độ lệch chuẩn thực nghiệm (experiment standard deviation);

$S =$ độ lệch chuẩn có hiệu chỉnh.

Ngoài ra ta có thể tính:

$$s^2 = E(X^2) - [E(X)]^2 = \overline{X^2} - (\bar{X})^2$$

với $E(X^2) = \overline{X^2} = \frac{1}{n}\left(x_1^2 n_1 + x_2^2 n_2 + \cdots + x_k^2 n_k\right) = \frac{1}{n}\sum_{i=1}^{k}(x_i^2 n_i)$.

và $S^2 = \left(\frac{n}{n-1}\right)s^2$.

TH2: mẫu chia khoảng

Ta xét bảng sau:

X	n_i	θ_i
$(a_1, a_2]$	n_1	θ_1
$(a_2, a_3]$	n_2	θ_2
$\cdots$	$\cdots$	$\cdots$
$(a_k, a_{k+1}]$	n_k	θ_k
Tổng	n	

Ta có:

Phương sai mẫu thực nghiệm:

$$s^2 = \frac{1}{n}\left[(\theta_1-\bar{X})^2 n_1 + (\theta_2-\bar{X})^2 n_2 + \cdots + (\theta_k-\bar{X})^2 n_k\right] = \frac{1}{n}\sum_{i=1}^{k}\left[n_i(\theta_i-\bar{X})^2\right]$$

Phương sai mẫu có hiệu chỉnh (justified sample variance):

$$S^2 = \frac{1}{n-1}\left[(\theta_1-\bar{X})^2 n_1 + (\theta_2-\bar{X})^2 n_2 + \cdots + (\theta_k-\bar{X})^2 n_k\right] = \frac{1}{n-1}\sum_{i=1}^{k}\left[n_i(\theta_i-\bar{X})^2\right]$$

Ngoài ra ta có thể tính:

$$s^2 = E(X^2) - [E(X)]^2 = \overline{X^2} - (\overline{X})^2$$

với $E(X^2) = \overline{X^2} = \dfrac{1}{n}\left(\theta_1^2 n_1 + \theta_2^2 n_2 + \cdots + \theta_k^2 n_k\right) = \dfrac{1}{n}\sum_{i=1}^{k}(\theta_i^2 n_i)$.

và $S^2 = \left(\dfrac{n}{n-1}\right)s^2$.

*** Dùng MT bỏ túi để tìm**: $n, \overline{X}, s, S$.

Minh họa bằng dòng máy CASIO-fx-570ES Plus (VINACAL tương tự)

SHIFT + MODE → Mũi tên xuống ⇩ → 4: STAT → Màn hình xuất hiện:

Frequency?	
1: ON	2: OFF

→ Bấm 1.

→ MODE → 3: STAT → 1: 1-VAR → Màn hình xuất hiện:

	X	FREQ (n_i)
1	155,5 =	1 =
2	156 =	1 =
3	158 =	2 =
4	160 =	2 =
5	162 =	3 =

Nhập xong bảng, bấm AC.

→ Bấm SHIFT + 1 → 4: Var → Màn hình xuất hiện

1: n	2: $\overline{x}$
3: σx (CASIO) hoặc 3: $x\sigma n$ (VINACAL)	4: sx (CASIO) hoặc 4: $x\sigma n-1$ (VINACAL)

Trong đó: nếu ta chọn

1 → rồi bấm dấu = sẽ cho ra kết quả là cỡ mẫu n;

2 → rồi bấm dấu = sẽ cho ra kết quả là kỳ vọng mẫu (trung bình mẫu) $E(X) = \bar{X}$;

3 → rồi bấm dấu = sẽ cho ra kết quả là độ lệch chuẩn thực nghiệm s;

4 → rồi bấm dấu = sẽ cho ra kết quả là độ lệch chuẩn có hiệu chỉnh S.

<u>Ví dụ</u>:

Bài 20, 21, 22 (trang 73), 23, 24 (trang 74);

3/ <u>Ước lượng điểm và ước lượng khoảng:</u>

* <u>*Ước lượng điểm (point estimator)*</u>:

Gọi θ (the-ta) là tham số của đặc tính X mà ta cần quan tâm (trên Population). Ở đây θ có thể là kỳ vọng ($\mu = E(X)$) (đọc là "muy"), có thể là phương sai ($\sigma^2 = Var(X) = D(X)$), có thể là tỷ lệ ($p = P(A)$), và θ = không biết.

Từ số liệu thực nghiệm $x_1, x_2, ..., x_n$ ta cần tìm giá trị gần đúng (xấp xỉ) dành cho θ, và ký hiệu là $\hat{\theta}$ (the-ta hat) (the-ta mũ).

Ta gọi $\hat{\theta}$ là ước lượng điểm dành cho θ.

Ta có:

$\bar{X}$ là ước lượng điểm dành cho kỳ vọng $\mu = E(X)$;

s^2 là ước lượng điểm dành cho phương sai $\sigma^2 = Var(X) = D(X)$;

S^2 là ước lượng điểm dành cho phương sai $\sigma^2 = Var(X) = D(X)$;

s là ước lượng điểm dành cho độ lệch tiêu chuẩn $\sigma = \sqrt{Var(X)} = \sqrt{D(X)}$;

S là ước lượng điểm dành cho độ lệch tiêu chuẩn $\sigma = \sqrt{Var(X)} = \sqrt{D(X)}$;

$\hat{p} = \hat{f}_n$ là ước lượng điểm dành cho tỷ lệ $p = P(A)$ (p = proportion).

Khi $\hat{\theta}$ là ước lượng điểm dành cho θ thì sai số của ước lượng là $\hat{\theta}-\theta$.

(Maximum Likelihood Estimator – MLE = pp ước lượng hợp lý nhất) là cách khác để tìm $\hat{\theta}$ cho θ.

***Ước lượng không chệch* (unbiased estimator)**: Ta gọi $\hat{\theta}$ là ước lượng không chệch cho θ nếu $E(\hat{\theta})=\theta$.

Khi $\hat{\theta}$ là ước lượng không chệch cho θ thì trung bình của sai số ước lượng là:

$$E(\hat{\theta}-\theta)=E(\hat{\theta})-E(\theta)=\theta-\theta=0.$$

Ta có:

$\overline{X}$ là ước lượng không chệch dành cho kỳ vọng $\mu=E(X)$;

S^2 là ước lượng không chệch dành cho phương sai $\sigma^2=Var(X)=D(X)$;

S là ước lượng không chệch cho độ lệch tiêu chuẩn $\sigma=\sqrt{Var(X)}=\sqrt{D(X)}$;

$\hat{p}=\hat{f}_n$ là ước lượng không chệch dành cho tỷ lệ $p=P(A)$.

Và đồng thời ta có các ước lượng chệch (biased estimator) như sau:

$$\begin{cases} s^2 \text{ là ước lượng chệch dành cho phương sai } \sigma^2=Var(X)=D(X); \\ s \text{ là ước lượng chệch dành cho độ lệch tiêu chuẩn } \sigma=\sqrt{Var(X)}=\sqrt{D(X)}. \end{cases}$$

* ***Ước lượng khoảng (interval/credible estimator)***:

Ước lượng khoảng dành cho kỳ vọng $\mu=E(X)$

Gọi $\mu=E(X)$ là kỳ vọng (trung bình) của đặc tính X trên tổng thể T

($\mu=$không biết)

Từ số liệu $x_1,x_2,...,x_n$ thực nghiệm, ta cần tìm μ_1 và μ_2 sao cho

$P(\mu_1\le\mu\le\mu_2)=\gamma$, với $\gamma=$gam-ma – độ tin cậy (confident level) cho trước.

(Mức chuẩn $\gamma=95\%$)

Ta gọi $[\mu_1,\mu_2]=$khoảng tin cậy (KTC) hay là ƯL khoảng cần tìm dành cho $\mu=E(X)$.

Trước tiên ta tìm kỳ vọng mẫu (trung bình mẫu): $\overline{X}$.

Sau đó ta tìm sai số biên (sai số lề – sai số thành phần – margin of error) như sau:

$m = \varepsilon = c.\frac{\sigma_0}{\sqrt{n}}$ nếu $\sigma_0 = \sqrt{Var(X)} = \sqrt{D(X)}$ đã biết; hoặc

$m = \varepsilon = c.\frac{S}{\sqrt{n}}$ nếu $\sigma = \sqrt{Var(X)} = \sqrt{D(X)}$ chưa biết; và $n > 30$; hoặc

$m = \varepsilon = t_{n-1}.\frac{S}{\sqrt{n}}$ nếu $\sigma = \sqrt{Var(X)} = \sqrt{D(X)}$ chưa biết; và $n \leq 30$.

Trong đó: $c =$ giá trị tra từ bảng phụ lục số 1 (bảng phân phối Laplace, bảng phân vị chuẩn) như sau:

Ta có $\gamma = 2\phi(c) - 1 \Rightarrow \phi(c) = \frac{1+\gamma}{2} = \cdots \Rightarrow c = \cdots$

Ví dụ:

$$\gamma = 95\% = 0,95 \Rightarrow \phi(c) = \frac{1+0,95}{2} = 0,975 \Rightarrow c = 1,960$$

$$\gamma = 99\% = 0,99 \Rightarrow \phi(c) = \frac{1+0,99}{2} = 0,995 \Rightarrow c = 2,576$$

$$\gamma = 90\% = 0,90 \Rightarrow \phi(c) = \frac{1+0,90}{2} = 0,95 \Rightarrow c = 1,645$$

Còn số t_{n-1} được tra từ bảng phụ lục số 3 (luật phân phối Student), như sau:

Ta xét dòng = n – 1;

Cột = $\frac{1+\gamma}{2} = \cdots$, từ đó suy ra $t_{n-1} = \cdots$

Ví dụ:

$\gamma = 95\% = 0,95$ và cho n = 20, suy ra dòng = 19 và cột = 0,975

$t_{n-1} = 2,093$

$\gamma = 99\% = 0,99$ và cho n = 14, suy ra dòng = 13 và cột = 0,995

$t_{n-1} = 3,012$

$\gamma = 90\% = 0,9$ và cho n = 10, suy ra dòng = 9 và cột = 0,95

$t_{n-1} = 1,833$

Ta có $\begin{cases} \mu_1 = \overline{X} - m \\ \mu_2 = \overline{X} + m \end{cases}$.

Ta gọi $\mu_2 - \mu_1 = 2m = 2\varepsilon =$ độ dài = đường kính (diameter) = độ dài của ƯL và

$m = \varepsilon =$ bán kính (radius) (độ chính xác – ĐCX) của ƯL.

<u>Ví dụ mẫu 1</u>:

Bài 15 trang 72

<u>Giải</u>:

Gọi $\mu = E(X)$ là mức xăng hao phí trung bình của loại động cơ ô tô khảo sát (khi đi từ A đến B) (ở đây $\mu =$ không biết).

Ta cần tìm μ_1 và μ_2 sao cho $P(\mu_1 \le \mu \le \mu_2) = \gamma$

Ta có kỳ vọng mẫu là $\overline{X} = \frac{1}{25}\sum_{i=1}^{4}(n_i\theta_i) = 20{,}12$ và $S \approx 0{,}5358$

= ước lượng điểm dành cho $\mu = E(X)$

Trường hợp phương sai chưa biết

Do *n* = 25 < 30, nên ta dùng bảng phụ lục 3 (phân phối Student)

Tiếp theo, ta tìm

$$m = \varepsilon = t_{n-1}.\frac{S}{\sqrt{n}}$$

TH1: $\gamma = 95\% = 0{,}95$. Ta có dòng = *n* – 1 = 24

Cột = $\frac{1+0{,}95}{2} = 0{,}975$

Suy ra $t_{n-1} = 2{,}064 \Rightarrow m = \varepsilon = t_{n-1}.\frac{S}{\sqrt{n}} = 2{,}064.\frac{0{,}5358}{\sqrt{25}} \approx 0{,}2212$.

$$\Rightarrow \begin{cases} \mu_1 = \overline{X} - m = 19{,}8988 \\ \mu_2 = \overline{X} + m = 20{,}3412 \end{cases}$$

Vậy ƯL khoảng dành cho μ là [19,8988; 20,3412]

TH2: $\gamma = 99\% = 0{,}99$. Ta có dòng = *n* – 1 = 24

$$\text{Cột} = \frac{1+0,99}{2} = 0,995$$

$$\text{Suy ra } t_{n-1} = 2,797 \Rightarrow m = \varepsilon = t_{n-1}.\frac{S}{\sqrt{n}} = 2,797.\frac{0,5358}{\sqrt{25}} \approx 0,2997$$

$$\Rightarrow \begin{cases} \mu_1 = \overline{X} - m = ... \\ \mu_2 = \overline{X} + m = ... \end{cases}$$

Trường hợp phương sai đã biết (bằng 4) $(\sigma_0^2 = 4 \Rightarrow \sigma_0 = 2)$

Nên ta dùng bảng phụ lục 1 (phân phối Laplace)

Tiếp theo, ta tìm

$$m = \varepsilon = c.\frac{\sigma_0}{\sqrt{n}}$$

TH1: $\gamma = 95\% = 0,95 \Rightarrow \phi(c) = \dfrac{1+0,95}{2} = 0,975 \Rightarrow c = 1,960$

$$\Rightarrow m = \varepsilon = c.\frac{\sigma_0}{\sqrt{n}} = 1,960.\frac{2}{\sqrt{25}} = ...$$

$$\Rightarrow \begin{cases} \mu_1 = \overline{X} - m = ... \\ \mu_2 = \overline{X} + m = ... \end{cases}$$

TH2: $\gamma = 99\% = 0,99 \Rightarrow \phi(c) = \dfrac{1+0,99}{2} = 0,995 \Rightarrow c = 2,576$

$$\Rightarrow m = \varepsilon = c.\frac{\sigma_0}{\sqrt{n}} = 2,576.\frac{2}{\sqrt{25}} = ...$$

$$\Rightarrow \begin{cases} \mu_1 = \overline{X} - m = ... \\ \mu_2 = \overline{X} + m = ... \end{cases}$$

Ví dụ mẫu 2: bài 16/ trang 72

Gọi $\mu = E(X)$ là năng suất lúa trung bình trên toàn địa phương A.

(ở đây $\mu =$ không biết).

Ta cần tìm μ_1 và μ_2 sao cho $P(\mu_1 \le \mu \le \mu_2) = \gamma$.

Từ số liệu đề bài, ta có:

$\overline{X} = \dfrac{1}{100}\sum_{i=1}^{7}(n_i x_i) = 46$ (tạ/ha) = là ƯL điểm (ước lượng không chệch) dành cho μ.

$S=\sqrt{\frac{1}{99}\sum_{i=1}^{7}n_i(x_i-\bar{X})^2}\approx 3,3029$ (tạ/ha).

Ta có $m=\varepsilon=c.\frac{S}{\sqrt{n}}$ (do mẫu cỡ lớn, nghĩa là có $n>30$, và phương sai chưa biết).

TH1: ta có Độ tin cậy (ĐTC) $\gamma=95\%=0,95$.

Suy ra: $\phi(c)=\frac{1+\gamma}{2}=\frac{1+0,95}{2}=0,975\Rightarrow c=1,960$

$\Rightarrow m=\varepsilon=c.\frac{S}{\sqrt{n}}=1,960.\frac{3,3029}{\sqrt{100}}\approx 0,6474$

$\Rightarrow\begin{cases}\mu_1=\bar{X}-m=46-0,6474=45,3526\\ \mu_2=\bar{X}+m=46+0,6474=46,6474\end{cases}$

(tạ/ha)

Vậy, ƯL khoảng dành cho μ là $[45,3526;46,6474]$ (tạ/ha).

TH2: ta có Độ tin cậy (ĐTC) $\gamma=99\%=0,99$.

Suy ra: $\phi(c)=\frac{1+0,99}{2}=\frac{1+0,99}{2}=0,995\Rightarrow c=2,576$

$\Rightarrow m=\varepsilon=c.\frac{S}{\sqrt{n}}=2,576.\frac{3,3029}{\sqrt{100}}\approx 0,8508$

$\Rightarrow\begin{cases}\mu_1=\bar{X}-m=46-0,8508=45,1492\\ \mu_2=\bar{X}+m=46+0,8508=46,8508\end{cases}$

(tạ/ha)

Vậy, ƯL khoảng dành cho μ là $[45,1492;46,8508]$ (tạ/ha).

<u>Ví dụ mẫu 3</u>: bài 39/ trang 76

Gọi $\mu=E(X)$ là số khuyết tật trung bình trên mỗi sản phẩm

(ở đây $\mu=$không biết).

a/ Ta cần tìm μ_1 và μ_2 sao cho $P(\mu_1\le\mu\le\mu_2)=\gamma$.

Từ số liệu đề bài, ta có:

$\bar{X}=\frac{1}{180}\sum_{i=1}^{9}(n_ix_i)\approx 3,6778$ (khuyết tật) = là ƯL điểm dành cho μ.

$$S=\sqrt{\frac{1}{179}\sum_{i=1}^{9}n_i(x_i-\bar{X})^2}\approx 1,7358 \text{ (khuyết tật).}$$

Ta có $m=\varepsilon=c.\frac{S}{\sqrt{n}}$ (do mẫu cỡ lớn, nghĩa là có $n>30$, và phương sai chưa biết).

TH1: ta có Độ tin cậy (ĐTC) $\gamma=95\%=0,95$.

Suy ra: $\phi(c)=\frac{1+\gamma}{2}=\frac{1+0,95}{2}=0,975\Rightarrow c=1,960$

$$\Rightarrow m=\varepsilon=c.\frac{S}{\sqrt{n}}=1,960.\frac{1,7358}{\sqrt{180}}\approx 0,2536$$

$$\Rightarrow\begin{cases}\mu_1=\bar{X}-m=3,6778-0,2536=3,4242\\ \mu_2=\bar{X}+m=3,6778+0,2536=3,9314\end{cases}$$

(khuyết tật)

Vậy, ƯL khoảng dành cho μ là $[3,4242;\ \ 3,9314]$ (khuyết tật).

TH2: ta có Độ tin cậy (ĐTC) $\gamma=99\%=0,99$.

Suy ra: $\phi(c)=\frac{1+0,99}{2}=\frac{1+0,99}{2}=0,995\Rightarrow c=2,576$

$$\Rightarrow m=\varepsilon=c.\frac{S}{\sqrt{n}}=2,576.\frac{1,7358}{\sqrt{180}}\approx 0,3333$$

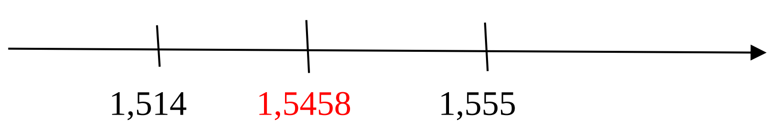
$$\Rightarrow\begin{cases}\mu_1=\bar{X}-m=3,6778-0,3333=3,3445\\ \mu_2=\bar{X}+m=3,6778+0,3333=4,0111\end{cases}$$

(khuyết tật)

Vậy, ƯL khoảng dành cho μ là $[3,3445;\ \ 4,0111]$ (khuyết tật).

b/ Ta có $m=\varepsilon=c.\frac{S}{\sqrt{n}}=0,2\Rightarrow c=\frac{0,2\times\sqrt{n}}{S}=\frac{0,2\times\sqrt{180}}{1,7358}\approx 1,5458$.

Ta dùng bảng phụ lục số 1 (bảng phân phối Laplace, hay là bảng phân vị chuẩn), ta suy ra:



Dùng chức năng nội suy của MT bỏ túi như sau

SHIFT + MODE → bấm ⇩ → bấm 4: STAT.

Màn hình xuất hiện

Frequency?
1: ON 2: OFF

Chọn 1 → MODE → 3: STAT → 2: A+BX.

Màn hình xuất hiện

	X	Y	FREQ
Nhập:	0,935 =	1,514 =	1
	0,940 =	1,555 =	1

Xong bấm AC.

SHIFT + 1 → 5: Reg → 4: $\hat{x}$ → di chuyển chuột ra trước $\hat{x}$, ta bấm "1.5458 $\hat{x}$" → "=" ta có kết quả là: 0,9399

Như vậy, ta có $\phi(c) = 0,9399 \Rightarrow \gamma = 2\phi(c) - 1 = 0,8798 = 87,98\%$ là độ tin cậy cần tìm.

c/ Muốn ĐCX là 0,1, và có độ tin cậy 99% thì cần khảo sát thêm bao nhiêu phần tử?

Ta có $m = \varepsilon = c.\dfrac{S}{\sqrt{n}} = 0,1 \Rightarrow \sqrt{n} = \dfrac{c \times S}{0,1}$, với

Độ tin cậy (ĐTC) $\gamma = 99\% = 0,99$.

Suy ra: $\phi(c) = \dfrac{1+0,99}{2} = \dfrac{1+0,99}{2} = 0,995 \Rightarrow c = 2,576$

Nên $\sqrt{n} = \dfrac{c \times S}{0,1} = \dfrac{2,576 \times 1,7358}{0,1} \approx 44,7142 \Rightarrow n \approx [1999,3596] = 2000$ (sản phẩm), nghĩa là cần khảo sát thêm 2000 – 180 = 1820 (sản phẩm nữa).

<u>Bài tập tương tự</u>:

1/ bài 13 trang 72;
2/ bài 30 trang 75;
3/ bài 32 trang 76;
4/ bài 40 trang 76-77;
5/ bài 43 trang 77.

Ước lượng khoảng dành cho tỷ lệ $p = P(A)$

Gọi $p = P(A)$ là tỷ lệ các phần tử thỏa tính chất A của đặc tính X trên tổng thể T

($p =$ không biết)

Từ số liệu $x_1, x_2, \ldots, x_n$ thực nghiệm, ta cần tìm p_1 và p_2 sao cho

$P(p_1 \le p \le p_2) = \gamma$, với $\gamma =$ gam-ma – độ tin cậy (confident level) cho trước.

(Mức chuẩn $\gamma = 95\%$)

Ta gọi $[p_1, p_2] =$ khoảng tin cậy (KTC) hay là ƯL khoảng cần tìm dành cho $p = P(A)$.

Trước tiên ta tìm ước lượng điểm dành cho tỷ lệ p là:

$$\hat{p} = \hat{f}_n = \frac{so \quad luong \quad phan \quad tu \quad thoa \quad tinh \quad chat \quad A}{n}$$

Sau đó ta tìm sai số biên (sai số lề – sai số thành phần – margin of error) như sau:

$$m = \varepsilon = c.\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Trong đó: $c =$ giá trị tra từ bảng phụ lục số 1 (bảng phân phối Laplace, bảng phân vị chuẩn) như sau:

Ta có

$$\gamma = 2\phi(c) - 1 \Rightarrow \phi(c) = \frac{1+\gamma}{2} = \ldots \Rightarrow c = \ldots$$

Suy ra

$$\begin{cases} p_1 = \hat{p} - m \\ p_2 = \hat{p} + m \end{cases}$$

Ta gọi $p_2 - p_1 = 2m = 2\varepsilon =$ độ dài = đường kính = độ dài của ƯL và

$m = \varepsilon =$ bán kính (radius) (độ chính xác – ĐCX) của ƯL.

Ví dụ mẫu 4: bài 14 trang 72

Giải: Gọi $p =$ tỷ lệ người mắc bệnh B trong tổng thể T.

($p =$ không biết).

a/ Ta có $\hat{p} = \hat{f}_n = 0,25$.

Ta cần tìm p_1 và p_2 sao cho $P(p_1 \le p \le p_2) = \gamma$

Ta có sai số biên là $m = \varepsilon = c.\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$.

TH1: ta có $\gamma = 95\% = 0,95 = 2\phi(c) - 1 \Rightarrow \phi(c) = 0,975 \Rightarrow c = 1,960$.

Suy ra $m = \varepsilon = 1,960 \times \sqrt{\frac{0,25(1-0,25)}{289}} \approx 0,0499$.

Cho nên

$$\begin{cases} p_1 = \hat{p} - m \approx 0,2001 \\ p_2 = \hat{p} + m \approx 0,2999 \end{cases}$$

Vậy ƯL khoảng dành cho tỷ lệ p là: [0,2001; 0,2999] hay [20,01; 29,99] (%)

TH2: ta có $\gamma = 99\% = 0,99 = 2\phi(c) - 1 \Rightarrow \phi(c) = 0,995 \Rightarrow c = 2,576$

Suy ra

$$m = \varepsilon = 2,576 \times \sqrt{\frac{0,25(1-0,25)}{289}} \approx 0,0656$$

Cho nên

$$\begin{cases} p_1 = \hat{p} - m \approx 0,1814 \\ p_2 = \hat{p} + m \approx 0,3156 \end{cases}$$

Vậy ƯL khoảng dành cho tỷ lệ p là: [0,1814; 0,3156] hay [18,14; 31,56] (%)

b/ Ta có ĐCX là: $m = \varepsilon = 0,02 = c.\sqrt{\frac{0,25(1-0,25)}{n}} \Rightarrow n \approx \left[\frac{0,25 \times 0,75 \times c^2}{(0,02)^2}\right]$

Với $\gamma = 95\% = 0,95 = 2\phi(c) - 1 \Rightarrow \phi(c) = 0,975 \Rightarrow c = 1,960$

$$n \approx \left[\frac{0,25 \times 0,75 \times (1,960)^2}{(0,02)^2}\right] = [1800,75] = 1801 \text{ (phần tử).}$$

<u>Ví dụ mẫu 5</u>: bài 38 trang 76.

<u>Giải</u>: Gọi N là số cá thực sự trong hồ.

Gọi $p =$ tỷ lệ cá bắt được trong hồ (ở đây $p =$ không biết)

Ta cần tìm p_1 và p_2 sao cho $P(p_1 \le p \le p_2) = \gamma$

Ta có

$$\hat{p} = \hat{f}_n = \frac{80}{400} = 0,2$$

Và sai số biên là:

$$m = \varepsilon = c.\sqrt{\frac{\hat{p}(1-\hat{p})}{n}},$$

Với $\gamma = 95\% = 0,95 = 2\phi(c) - 1 \Rightarrow \phi(c) = 0,975 \Rightarrow c = 1,960$

$$\Rightarrow m = \varepsilon = 1,960\sqrt{\frac{0,2(1-0,2)}{400}} = 0,0392$$

Suy ra

$$\begin{cases} p_1 = \hat{p} - m = 0,1608 \\ p_2 = \hat{p} + m = 0,2392 \end{cases}$$

Ta có: $p_1 = \dfrac{2000}{N} \Rightarrow N \approx [12437,81095] = 12438$ (con cá)

$p_2 = \dfrac{2000}{N} \Rightarrow N \approx [8361,204013] = 8362$ (con cá).

Vậy, số cá trong hồ có từ 8362 đến 12438 con cá, với ĐTC 95%.