2.1 NHẮC LẠI MỘT SỐ KIẾN THỨC VỀ XÁC SUẤT VÀ QUY TẮC TÍNH XÁC SUẤT

2.1.1 Phép thử và biến cố

		Tung đồng xu	Gieo xúc xắc
Phép thử	Biến cố sơ cấp	Sấp, ngửa	Mặt có 1 chấm đến 6 chấm
	Không gian mẫu	$\Omega=\{S,N\}$	$\Omega=\{1,2,3,4,5,6\}$
	Biến cố	$\{S\}$ $\{N\}$	Mặt có số chấm chẵn …

1

Biến cố (Event)

Với phép thử $\mathscr{C}$ ta thường xét các biến cố (còn gọi là sự kiện) mà việc xảy ra hay không xảy ra hoàn toàn được xác định bởi kết quả của $\mathscr{C}$

Mỗi kết quả ω của $\mathscr{C}$ được gọi là kết quả thuận lợi cho biến cố A nếu A xảy ra khi kết quả của $\mathscr{C}$ là ω

Ví dụ: Nếu gọi A là biến cố số chấm xuất hiện là chẵn trong phép thử tung xúc xắc thì A có các kết quả thuận lợi là 2, 4, 6

- Biến cố chắc chắn là biến cố luôn luôn xảy ra khi thực hiện phép thử. Không gian mẫu Ω biến cố một biến cố chắc chắn
- Biến cố không thể là biến cố nhất định không xảy ra khi thực hiện phép thử. Biến cố không thể được ký hiệu $\varnothing$

2

2.1.2. Quan hệ của các biến cố

a. Quan hệ biến cố đối

Biến cố đối của A ký hiệu $\bar{A}$, A xảy ra khi và chỉ khi $\bar{A}$ không xảy ra

b. Tổng của hai biến cố

Tổng của hai biến cố A, B là biến cố được ký hiệu $A\cup B$

Tổng của một dãy các biến cố $\{A_1, A_2, ..., A_n\}$ là biến cố $\bigcup_{i=1}^{n} A_i$

Biến cố tổng xảy ra khi ít nhất một biến cố xảy ra

c. Tích của hai biến cố

Tích của hai biến cố A, B là biến cố được ký hiệu AB

Tích của một dãy các biến cố $\{A_1, A_2, ..., A_n\}$ là biến cố $\prod_{i=1}^{n} A_i$

Biến cố tích xảy ra khi tất cả các biến cố cùng xảy ra

3

d. Biến cố xung khắc

Hai biến số A, B gọi là xung khắc nếu hai biến cố này không thể đồng thời cùng xảy ra

e. Hệ đầy đủ các biến cố

Dãy các biến cố $A_1, A_2, ..., A_n$ được gọi là một hệ đầy đủ các biến cố nếu:

- Xung khắc từng đôi một, nghĩa là $A_iA_j=\varnothing$ với mọi $i\neq j=1,...,n$,
- Tổng của chúng là biến cố chắc chắc, nghĩa là $\bigcup_{i=1}^{n} A_i=\Omega$

4

Ví dụ 1.7: Một nhà máy có ba phân xưởng sản xuất ra cùng một loại sản phẩm. Giả sử rằng mỗi sản phẩm của nhà máy chỉ do một trong ba phân xưởng này sản xuất.

Chọn ngẫu nhiên một sản phẩm.

Gọi A_1, A_2, A_3 lần lượt là biến cố sản phẩm được chọn do phân xưởng thứ nhất, thứ hai, thứ ba sản xuất

Khi đó hệ ba biến cố A_1, A_2, A_3 là hệ đầy đủ

f. Tính độc lập của các biến cố

Hai biến cố được gọi là độc lập với nhau nếu việc xảy ra hay không xảy ra biến cố này không ảnh hưởng tới việc xảy ra hay không xảy ra biến cố kia

5

Ví dụ 1.8: Ba xạ thủ $\mathscr{A}$, $\mathscr{B}$, $\mathscr{C}$ mỗi người bắn một viên đạn vào mục tiêu. Gọi A, B, C lần lượt là biến cố $\mathscr{A}$, $\mathscr{B}$, $\mathscr{C}$ bắn trúng mục tiêu.

ABC là biến cố cả 3 đều bắn trúng

$\bar{A}\bar{B}\bar{C}$ là biến cố cả 3 đều bắn trượt

$A\cup B\cup C$ là biến cố có ít nhất 1 người bắn trúng

Biến cố có ít nhất 2 xạ thủ bắn trúng $AB\cup BC\cup CA$

Biến cố có nhiều nhất 1 xạ thủ bắn trúng $\bar{A}\bar{B}\cup\bar{B}\bar{C}\cup\bar{C}\bar{A}$

Biến cố chỉ có xạ thủ $\mathscr{C}$ bắn trúng $\bar{A}\bar{B}C$

Biến cố chỉ có 1 xạ thủ bắn trúng $A\bar{B}\bar{C}\cup\bar{A}B\bar{C}\cup\bar{A}\bar{B}C$

Ba biến cố A, B, C độc lập? xung khắc?

6

2.1.3 Định nghĩa xác suất

Xác suất của một biến cố là một con số đặc trưng khả năng khách quan xuất hiện biến cố đó khi thực hiện phép thử

Xác suất của biến cố A ký hiệu $P(A)$. Trường hợp biến cố chỉ gồm một biến cố sơ cấp $\{a\}$ ta ký hiệu $P(a)$ thay cho $P(\{a\})$.

1 Định nghĩa cổ điển về xác suất

Xác suất của biến cố A

$$P(A) = \frac{\text{số trường hợp thuận lợi đối với } A}{\text{số trường hợp có thể}}$$

Ví dụ 1: Biến cố A: xuất hiện mặt chẵn trong phép thử gieo con xúc xắc có 3 trường hợp thuận lợi và 6 trường hợp có thể

Vậy $$P(A) = \frac{3}{6} = \frac{1}{2}$$

Ví dụ 2: Phép thử tung đồng thời 2 đồng xu

Biến cố B: xuất hiện một mặt sấp một mặt ngửa

có 2 trường hợp thuận lợi và 4 trường hợp có thể

Vậy $$P(B) = \frac{2}{4} = \frac{1}{2}$$

Để tính xác suất cổ điển ta sử dụng phương pháp đếm của giải tích tổ hợp

2. Các qui tắc đếm

a. Qui tắc cộng

b. Qui tắc nhân

c. Hoán vị Có $n!$ hoán vị n phần tử

d. Chỉnh hợp Chọn **lần lượt** k phần tử không hoàn lại trong tập n phần tử ta được một chỉnh hợp chập k của n phần tử

Số CH chập k của n phần tử $$A_n^k = \frac{n!}{(n-k)!} = n(n-1)...(n-k+1)$$

e. Tổ hợp Chọn **đồng thời** k phần tử trong tập n phần tử ta được một tổ hợp chập k của n phần tử

Số TH chập k của n phần tử $$C_n^k = \frac{A_n^k}{k!} = \frac{n!}{k!(n-k)!}$$

Ví dụ 1.10: Tung một con xúc xắc hai lần. Tìm xác suất để trong đó có 1 lần ra 6 chấm

Ví dụ 1.11: Một người gọi điện thoại quên mất hai số cuối của số điện thoại và chỉ nhớ được rằng chúng khác nhau. Tìm xác suất để quay ngẫu nhiên một lần được đúng số cần gọi

Ví dụ 1.12: Một công ty cần tuyển 2 nhân viên. Có 6 người nộp đơn trong đó có 4 nữ và 2 nam. Giả sử khả năng trúng tuyển của cả 6 người là như nhau.

Tính xác suất biến cố:

a. Hai người trúng tuyển là nam
b. Hai người trúng tuyển là nữ
c. Có ít nhất 1 nữ trúng tuyển

3. Định nghĩa thống kê về xác suất

Giả sử phép thử $\mathscr{C}$ có thể được thực hiện lặp lại nhiều lần độc lập trong những điều kiện giống hệt nhau

Nếu trong n lần thực hiện phép thử $\mathscr{C}$ biến cố A xuất hiện $k_n(A)$ lần thì tỉ số

$$f_n(A) = \frac{k_n(A)}{n}$$

gọi là tần suất xuất hiện của biến cố A trong n phép thử

Có thể chứng minh được (định lý luật số lớn) khi n tăng lên vô hạn thì $f_n(A)$ tiến đến một giới hạn xác định. Ta định nghĩa giới hạn này là xác suất của biến cố A, ký hiệu $P(A)$

Trên thực tế các tần suất $f_n(A)$ xấp xỉ nhau khi n đủ lớn và xác suất của A được chọn bằng giá trị xấp xỉ này

$$P(A) \approx f_n(A)$$

Ví dụ 1.14: Một công ty bảo hiểm muốn xác định xác suất để một người Mỹ 25 tuổi sẽ bị chết trong năm tới, người ta theo dõi 100.000 thanh niên và thấy rằng có 798 người bị chết trong vòng 1 năm sau đó. Vậy xác suất cần tìm xấp xỉ bằng 0,008.

5. Các tính chất và định lý xác suất

Các tính chất của xác suất

1. Với mọi biến cố A: $0 \le P(A) \le 1$
2. Xác suất của biến cố không thể, biến cố chắc chắn

$$P(\varnothing) = 0,\ P(\Omega) = 1$$

Qui tắc cộng xác suất

a. Trường hợp xung khắc

$$P(A \cup B) = P(A) + P(B) \qquad P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i)$$

b. Trường hợp tổng quát

$$P(A \cup B) = P(A) + P(B) - P(AB)$$

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) - \sum_{i<j} P(A_i A_j) + \cdots + (-1)^{n-1} P(A_1 A_2 \ldots A_n)$$

Quy tắc xác suất của biến cố đối

$$P(\overline{A}) = 1 - P(A) \qquad P(A) = 1 - P(\overline{A})$$

Ví dụ 1.23: Trong phòng có n người ($n < 365$).
Tính xác suất có ít nhất hai người có cùng ngày sinh?
Tính xác suất này khi $n = 10$.

Giải Gọi A là biến cố có ít nhất hai người trong phòng có cùng ngày sinh

$$P(\overline{A}) = \frac{A_{365}^{n}}{365^n} = \frac{(365)(364)\ldots(365-n+1)}{365^n},\quad P(A) = 1 - P(\overline{A})$$

Khi $n = 10$ thì

$$P(\overline{A}) = \frac{A_{365}^{10}}{365^{10}} = 0{,}883\ ;\quad P(A) = 1 - 0{,}883 = 0{,}117$$

Ví dụ: Giả sử phép thử $\mathscr{C}$ có không gian mẫu $\Omega = \{a, b, c, d\}$

với xác suất $P(a) = 0{,}2$, $P(b) = 0{,}3$, $P(c) = 0{,}4$, $P(d) = 0{,}1$.

Xét hai biến cố $A = \{a, b\}$ và $B = \{b, c, d\}$.

Do đó ta có xác suất của các biến cố sau

$$P(A) = P(a) + P(b) = 0{,}2 + 0{,}3 = 0{,}5$$

$$P(B) = P(b) + P(c) + P(d) = 0{,}3 + 0{,}4 + 0{,}1 = 0{,}8$$

$$P(\overline{A}) = P(c) + P(d) = 0{,}4 + 0{,}1 = 0{,}5$$

hoặc $P(\overline{A}) = 1 - P(A) = 1 - 0{,}5 = 0{,}5$

$$P(A \cup B) = P(\Omega) = 1$$

$$P(AB) = P(b) = 0{,}3$$

hoặc $P(AB) = P(A) + P(B) - P(A \cup B) = 0{,}5 + 0{,}8 - 1 = 0{,}3$

6. Nguyên lý xác suất lớn, xác suất nhỏ

Nguyên lý xác suất nhỏ

"Nếu một biến cố có xác suất rất nhỏ thì thực tế có thể cho rằng trong một phép thử biến cố đó sẽ không xảy ra"

Khi tung đồng xu, ngoài khả năng mặt sấp hay mặt ngửa xuất hiện còn có khả năng đồng xu ở trạng thái đứng. Tuy nhiên khả năng thứ ba rất khó xảy ra,vì vậy thực tế ta luôn công nhận chỉ có hai khả năng mặt sấp và mặt ngửa xuất hiện.

Nguyên lý xác suất lớn

"Nếu biến cố có xác suất gần bằng 1 thì trên thực tế có thể cho rằng biến cố đó sẽ xảy ra trong một phép thử"

2.1.4. Xác suất có điều kiện

1 Định nghĩa và các tính chất của xác suất có điều kiện

Xác suất của biến cố B được tính trong điều kiện biết rằng biến cố A đã xảy ra được gọi là xác suất của B với điều kiện A, ký hiệu $P(B|A)$.

Tính chất

❖ Nếu $P(A) \neq 0$ thì $P(B|A) = \dfrac{P(AB)}{P(A)}$

❖ Khi cố định A với $P(A) \neq 0$ thì xác suất có điều kiện $P(B|A)$ có tất cả các tính chất của xác suất thông thường đối với biến cố B $\quad P(\overline{B}|A) = 1 - P(B|A)$

$$P(B_1 \cup B_2 | A) = P(B_1|A) + P(B_2|A) - P(B_1 B_2 | A)$$

Ví dụ: Gieo đồng thời hai con xúc xắc cân đối. Tính xác suất để tổng số chấm xuất hiện trên hai con xúc xắc ≥10 biết rằng ít nhất một con đã ra mặt có 5 chấm

Gọi A là biến cố " ít nhất một con ra mặt có 5 chấm "

$$P(A) = 1 - P(\overline{A}) = 1 - \left(\frac{5}{6}\right)^2 = \frac{11}{36}$$

Gọi B là biến cố "tổng số chấm trên hai con ≥10"

Biến cố AB có 3 kết quả thuận lợi là (5,6), (6,5), (5,5)

Vậy $P(AB) = \dfrac{3}{36}$

$$P(B|A) = \frac{3}{36} \Big/ \frac{11}{36} = \frac{3}{11}$$

CHƯƠNG 2: BỔ SUNG KIẾN THỨC VỀ XÁC SUẤT THỐNG KÊ

2. Quy tắc nhân xác suất

Trường hợp độc lập

- Nếu A, B là hai biến cố độc lập thì

$$P(AB) = P(A)P(B)$$

- Nếu $\{A_1, A_2, ..., A_n\}$ là các biến cố độc lập thì

$$P(A_1A_2...A_n) = P(A_1)P(A_2)...P(A_n)$$

Trường hợp tổng quát

$$P(AB) = P(A)P(B|A)$$

$$P(A_1A_2...A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1A_2)...P(A_n|A_1A_2...A_{n-1})$$

19

CHƯƠNG 2: BỔ SUNG KIẾN THỨC VỀ XÁC SUẤT THỐNG KÊ

Ví dụ 1.31: Túi I chứa 3 bi trắng, 7 bi đỏ, 15 bi xanh.
Túi II chứa 10 bi trắng, 6 bi đỏ, 9 bi xanh.

Từ mỗi túi lấy ngẫu nhiên 1 bi. Tìm xác suất để 2 bi được rút từ 2 túi là cùng màu.

Gọi A_t, $A_đ$, A_x lần lượt là biến cố bi rút được từ túi I là trắng, đỏ, xanh.

B_t, $B_đ$, B_x lần lượt là biến cố bi được rút từ túi II là trắng, đỏ, xanh.

Các biến cố A_t, $A_đ$, A_x xung khắc, B_t, $B_đ$, B_x xung khắc

Các biến cố A_t, $A_đ$, A_x độc lập với các biến cố B_t, $B_đ$, B_x

Biến cố 2 bi được rút cùng mầu là $A_tB_t \cup A_đB_đ \cup A_xB_x$

$$P(A_tB_t \cup A_đB_đ \cup A_xB_x) = P(A_tB_t) + P(A_đB_đ) + P(A_xB_x)$$
$$= P(A_t)P(B_t) + P(A_đ)P(B_đ) + P(A_x)P(B_x)$$
$$= \frac{3}{25}\frac{10}{25} + \frac{7}{25}\frac{6}{25} + \frac{15}{25}\frac{9}{25} = \frac{207}{625} \approx 0,331$$

20

CHƯƠNG 2: BỔ SUNG KIẾN THỨC VỀ XÁC SUẤT THỐNG KÊ

3. Công thức xác suất đầy đủ

Giả sử $\{A_1, A_2, ..., A_n\}$ là một hệ đầy đủ các biến cố, khi đó với mọi biến cố B của cùng một phép thử, ta có

$$P(B) = \sum_{i=1}^{n} P(A_iB) = \sum_{i=1}^{n} P(A_i)P(B|A_i)$$

Ví dụ 1.37: Một túi đựng 4 bi trắng và 6 bi đen. Người thứ nhất lấy ngẫu nhiên từ túi 3 bi (không hoàn lại), người thứ hai lấy tiếp 2 bi.

Tính xác suất để người thứ hai lấy được 1 bi trắng.

Gọi lần lượt A_0, A_1, A_2, A_3 là biến cố người thứ nhất lấy được 0, 1, 2, 3 bi trắng. A_0, A_1, A_2, A_3 là một hệ đầy đủ

Gọi B là biến cố người thứ hai lấy được 1 bi trắng

21

CHƯƠNG 2: BỔ SUNG KIẾN THỨC VỀ XÁC SUẤT THỐNG KÊ

Túi đựng 10 bi = **4** bi trắng + **6** bi đen

$$P(A_0) = \frac{C_6^3}{C_{10}^3} = \frac{1}{6}, \quad P(A_1) = \frac{C_4^1C_6^2}{C_{10}^3} = \frac{1}{2}, \quad P(A_2) = \frac{C_4^2C_6^1}{C_{10}^3} = \frac{3}{10},$$

$$P(A_3) = \frac{C_4^3}{C_{10}^3} = \frac{1}{30}. \quad P(B|A_0) = \frac{C_4^1C_3^1}{C_7^2} = \frac{12}{21}, \quad P(B|A_1) = \frac{C_3^1C_4^1}{C_7^2} = \frac{12}{21},$$

$$P(B|A_2) = \frac{C_2^1C_5^1}{C_7^2} = \frac{10}{21}, \quad P(B|A_3) = \frac{C_1^1C_6^1}{C_7^2} = \frac{6}{21}.$$

Do đó $P(B) = \frac{1}{6}\cdot\frac{12}{21} + \frac{1}{2}\cdot\frac{12}{21} + \frac{3}{10}\cdot\frac{10}{21} + \frac{1}{30}\cdot\frac{6}{21} = \frac{56}{105}$

22

CHƯƠNG 2: BỔ SUNG KIẾN THỨC VỀ XÁC SUẤT THỐNG KÊ

4. Công thức Bayes

Giả sử $\{A_1, A_2, ..., A_n\}$ là một hệ đầy đủ các biến cố

B là một biến cố của cùng một phép thử có xác suất khác 0,

khi đó với mọi $k = 1, 2, \ldots, n$ ta có

$$P(A_k|B) = \frac{P(A_kB)}{P(B)} = \frac{P(A_k)P(B|A_k)}{P(B)} = \frac{P(A_k)P(B|A_k)}{\sum_{i=1}^{n} P(A_i)P(B|A_i)}$$

23

CHƯƠNG 2: BỔ SUNG KIẾN THỨC VỀ XÁC SUẤT THỐNG KÊ

Ví dụ: Một nhà máy có ba phân xưởng I, II, III cùng sản xuất ra một loại sản phẩm. Phân xưởng I, II, III sản xuất tương ứng 36%, 34%, 30% sản lượng của nhà máy, với tỷ lệ phế phẩm tương ứng là 0,12; 0,1; 0,08.

Tìm tỷ lệ phế phẩm chung của nhà máy

Lấy ngẫu nhiên một sản phẩm kiểm tra và đó là phế phẩm (biến cố B). Tính xác suất phế phẩm này là do phân xưởng I sản xuất

Gọi A_1, A_2, A_3 lần lượt là biến cố sản phẩm lấy ra kiểm tra do phân xưởng I, II, III sản xuất. $\{A_1, A_2, A_3\}$ là một hệ đầy đủ

$$P(A_1) = 0,36; \; P(A_2) = 0,34; \; P(A_3) = 0,30$$

$$P(B|A_1) = 0,12; \; P(B|A_2) = 0,10; \; P(B|A_3) = 0,08$$

$$P(B) = P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + P(A_3)P(B|A_3) = 0,1012$$

$$P(A_1|B) = \frac{P(A_1)P(B|A_1)}{P(B)} = \frac{0,36\cdot 0,12}{0,1012} = 0,427$$

24

2.1.5 Dãy phép thử Bernoulli

Dãy các phép thử lặp lại, độc lập, trong mỗi phép thử xác suất xuất hiện của biến cố A không đổi $P(A) = p, (0 < p < 1)$ được gọi là dãy phép thử Bernoulli.

p là xác suất thành công trong mỗi lần thử.

Xác suất của biến cố "A xuất hiện ra đúng k lần trong n phép thử"

$$P_n(k;p) = C_n^k p^k (1-p)^{n-k}; k = 0,1,\dots,n$$

Khi k tăng từ 0 đến n thì $P_n(k;p)$ mới đầu tăng sau đó giảm và đạt giá trị lớn nhất tại $k = m$ thoả mãn

$$(n+1)p - 1 \le m \le (n+1)p \quad P_{\max} = P_n(m;p)$$

m được gọi là số lần xuất hiện có khả năng nhất

25

Ví dụ 1.43: Tín hiệu thông tin được phát đi 3 lần độc lập nhau. Xác suất thu được mỗi lần là 0.4.

a) Xác suất để nguồn thu nhận được thông tin đúng 2 lần.

$$P_3(2;0,4) = C_3^2 (0,4)^2 (0,6) = 0,288$$

b) Xác suất để nguồn thu nhận được thông tin đó.

$$P = 1 - P_3(0;0,4) = 1 - (0,6)^3 = 0,7848$$

c) Nếu muốn xác suất thu được tin $\ge 0,9$ thì phải phát đi ít nhất bao nhiêu lần.

$$1-(0,6)^n \ge 0,9 \Leftrightarrow (0,6)^n \le 0,1 \Leftrightarrow n \ge \frac{\lg(0,1)}{\lg(0,6)} = \frac{-1}{-1+0,778} = 4,504$$

Chọn $n = 5$

26

2.2. ĐỊNH NGHĨA VÀ PHÂN LOẠI BIẾN NGẪU NHIÊN

2.2.1 Định nghĩa biến ngẫu nhiên

Biến ngẫu nhiên X là đại lượng nhận các giá trị nào đó phụ thuộc vào các yếu tố ngẫu nhiên, nghĩa là với mọi giá trị thực $x \in \mathbb{R}$ thì "X nhận giá trị nhỏ hơn bằng x", ký hiệu $\{X \le x\}$, là một biến cố ngẫu nhiên.

Tập hợp tất cả các giá trị của X được gọi là *miền giá trị* của X, ký hiệu R_X

Ví dụ 2.1: Nếu gọi X là tổng số chấm xuất hiện khi gieo hai con xúc xắc thì X là một biến ngẫu nhiên có miền giá trị

$$R_X = \{2,3,\dots,12\}$$

và $\{X = k\} = A_k; k = 2,3,\dots,12$

27

Định nghĩa 2.2: Hai biến ngẫu nhiên X, Y là độc lập nếu X nhận các giá trị nào đó không phụ thuộc Y và ngược lại. Nói cách khác với mọi số thực x, y hai biến cố $\{X \le x\}$, $\{Y \le y\}$ là độc lập.

2.2.2 Hàm phân bố xác suất

Các biến ngẫu nhiên được xét trong các phép thử khác nhau (tương ứng với các không gian xác suất khác nhau) nhưng các phân bố xác suất của chúng có thể như nhau.

Phân bố xác suất được nghiên cứu thông qua hàm phân bố xác suất

28

Hàm phân bố xác suất (cumulative distribution function, viết tắt CDF) của biến ngẫu nhiên X là hàm số $F_X(x)$ xác định với mọi $x \in \mathbb{R}$ bởi công thức:

$$F_X(x) = P\{X \le x\}; \ -\infty < x < \infty$$

trong đó $\{X \le x\}$ là ký hiệu biến cố "biến ngẫu nhiên X nhận giá trị nhỏ hơn hay bằng x".

29

Các tính chất của hàm phân bố

1. $0 \le F_X(x) \le 1$ với mọi $x \in \mathbb{R}$
2. $F_X(x)$ là hàm không giảm, liên tục bên phải

   Nếu X là biến ngẫu nhiên liên tục thì $F_X(x)$ là hàm liên tục.
3. $F_X(-\infty) = \lim_{x\to-\infty} F_X(x) = 0; \ F_X(+\infty) = \lim_{x\to+\infty} F_X(x) = 1$
4. $P\{a < X \le b\} = F_X(b) - F_X(a) \quad P\{X > a\} = 1 - F_X(a)$

   $P\{X < a\} = F_X(a^-) = \lim_{x<a, x\to a} F_X(x)$

30

Ví dụ: Một nguồn thông tin sinh ra các ký hiệu ngẫu nhiên từ bốn ký tự {*a*, *b*, *c*, *d*} với xác suất *P*(*a*)=1/2, *P*(*b*)=1/4 và *P*(*c*)=*P*(*d*)=1/8.

Mã hóa các ký hiệu này theo các mã nhị phân sau

a	0
b	10
c	110
d	111

Đặt X là biến ngẫu nhiên ký hiệu độ dài của mã, đó là số các bit

1. Tìm miền giá trị của X
2. Giả sử các ký hiệu được sinh độc lập. Tính các xác suất $P\{X=1\}$, $P\{X=2\}$ và $P\{X=3\}$
3. Tìm hàm phân bố $F_X(x)$ và vẽ đồ thị

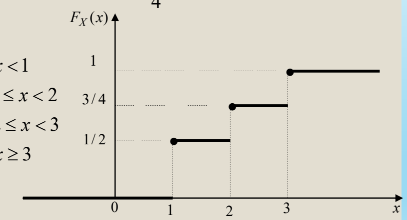
31

1. Miền giá trị $R_X = \{1,2,3\}$

2. $P\{X=1\} = P(a) = \frac{1}{2}$; $P\{X=2\} = P(b) = \frac{1}{4}$;

$P\{X=3\} = P\{c,d\} = P(c)+P(d) = \frac{1}{4}$

3.
$$F_X(x) = \begin{cases} 0 & x<1 \\ 1/2 & 1 \le x < 2 \\ 3/4 & 2 \le x < 3 \\ 1 & x \ge 3 \end{cases}$$



32

2.2.3 Phân loại

❖ *Biến ngẫu nhiên rời rạc*

Biến ngẫu nhiên X là rời rạc nếu miền giá trị gồm một số hữu hạn hoặc vô hạn đếm được các giá trị, nghĩa là có thể liệt kê các giá trị của miền giá trị thành một dãy. Do đó hàm phân bố có đồ thị dạng hình thang.

❖ *Biến ngẫu nhiên liên tục*

X là biến ngẫu nhiên liên tục nếu miền giá trị của nó có thể lấp đầy một hoặc một số các khoảng hữu hạn hoặc vô hạn và xác suất biến ngẫu nhiên nhận giá trị tại từng điểm đều bằng 0 (nghĩa là $P\{X = a\}=0$ với mọi a). Do đó hàm phân bố là hàm số liên tục.

33

Ví dụ

- Gọi X là số chấm xuất hiện khi gieo một con xúc xắc thì X là biến ngẫu nhiên rời rạc nhận các giá trị 1, 2, 3, 4, 5, 6.
- Gọi T là tuổi thọ của một thiết bị đang hoạt động thì T là biến ngẫu nhiên liên tục nhận giá trị trong một khoảng.
- Gọi Z là số khách hàng vào một điểm phục vụ trong 1 đơn vị thời gian, Z là biến ngẫu nhiên rời rạc nhận các giá trị 0, 1, 2, …
- Số cuộc gọi đến một tổng đài là biến ngẫu nhiên rời rạc nhận các giá trị 0, 1, 2, …
- Sai số Y khi đo lường một đại lượng vật lý nào đó là biến ngẫu nhiên liên tục nhận giá trị trong một khoảng.

34

2.2.4. Biến ngẫu nhiên rời rạc

1 Hàm khối lượng xác suất và bảng phân bố xác suất của biến ngẫu nhiên rời rạc

Hàm $p_X(x) = P\{X = x\}$

được gọi là hàm khối lượng xác suất (probability mass function) của biến ngẫu nhiên rời rạc X

Hàm phân bố của X

$$F_X(x) = P\{X \le x\} = \sum_{x_k \le x;\, x_k \in R_X} p_X(x_k)$$

35

Tính chất của hàm khối lượng xác suất

1. $\sum_{x_k \in R_X} p_X(x_k) = 1$
2. $p_X(x_k) > 0$, với mọi $x_k \in R_X$
3. $p_X(x) = 0$, với mọi $x \notin R_X$

Nếu $R_X = \{x_1, x_2, ...\}$ thì hàm phân bố xác suất có dạng:

$$F_X(x) = \begin{cases} 0 & \text{nếu} \quad x < x_1 \\ p_X(x_1) + \cdots + p_X(x_{k-1}) & \text{nếu} \quad x_{k-1} \le x < x_k \,,\ \forall k > 1 \end{cases}$$

36

Bảng phân bố xác suất

Bảng phân bố xác suất có hai hàng, hàng trên ghi các giá trị mà biến ngẫu nhiên nhận được, hàng dưới là giá trị của hàm khối lượng xác suất tương ứng

Bảng phân bố xác suất của biến ngẫu nhiên X có dạng

X	x_1	x_2	$\cdots$
P	$p_X(x_1)$	$p_X(x_2)$	$\cdots$

Ví dụ: Chọn ngẫu nhiên 3 bi từ một túi có 6 bi đen, 4 bi trắng. Gọi X là số bi trắng trong 3 bi vừa chọn thì X là một biến ngẫu nhiên rời rạc. Tìm bảng phân bố xác suất và hàm phân bố xác suất.

$$P\{X=0\}=\frac{C_6^3}{C_{10}^3}=\frac{5}{30},\ P\{X=1\}=\frac{C_6^2C_4^1}{C_{10}^3}=\frac{15}{30}$$

$$P\{X=2\}=\frac{C_6^1C_4^2}{C_{10}^3}=\frac{9}{30},\ P\{X=3\}=\frac{C_4^3}{C_{10}^3}=\frac{1}{30}$$

Hàm khối lượng xác suất

$$p_X(0)=\frac{5}{30},\ p_X(1)=\frac{15}{30},\ p_X(2)=\frac{9}{30},\ p_X(3)=\frac{1}{30}$$

$p_X(x)=0$ với mọi x khác 0, 1, 2, 3

Bảng phân bố xác suất

X	0	1	2	3
P	5/30	15/30	9/30	1/30

Hàm phân bố xác suất

$$F_X(x)=\begin{cases} 0 & \text{nếu } x<0 \\ 5/30 & \text{nếu } 0\le x<1 \\ 20/30 & \text{nếu } 1\le x<2 \\ 29/30 & \text{nếu } 2\le x<3 \\ 1 & \text{nếu } x\ge 3 \end{cases}$$

2. Các phân bố rời rạc thường gặp

1. Phân bố Bernoulli

Biến ngẫu nhiên rời rạc X nhận hai giá trị 0, 1 với hàm khối lượng xác suất

$$p_X(k)=P\{X=k\}=p^kq^{1-k};\ k=0,1$$

trong đó $0<p<1$, $q=1-p$,

được gọi là có phân bố Bernoulli tham số p.

Xét phép thử Bernoulli với sự thành công của phép thử là sự xuất hiện của biến cố A với xác suất xuất hiện là p. Gọi X là số lần thành công trong một lần thử thì X là biến ngẫu nhiên rời rạc có phân bố Bernoulli tham số p.

2. Phân bố nhị thức $\mathscr{B}(n;p)$

Biến ngẫu nhiên rời rạc X nhận các giá trị 0, 1, …, n với xác suất tương ứng

$$p_X(k)=P\{X=k\}=C_n^kp^kq^{n-k};\ k=0,1,\dots,n$$

trong đó n là số tự nhiên và $0<p<1$, $q=1-p$, được gọi là có phân bố nhị thức tham số n, p, ký hiệu $X\sim\mathscr{B}(n;p)$

Hàm phân bố

$$F_X(x)=\begin{cases} 0 & \text{nếu } x<0 \\ \sum_{k=0}^{m}C_n^kp^kq^{n-k} & \text{nếu } m\le x<m+1,\ 0\le m\le n-1 \\ 1 & \text{nếu } x\ge n \end{cases}$$

Thực hiện n phép thử Bernoulli với xác suất thành công của biến cố A trong mỗi lần thử là p

Với mỗi i =1, 2, …, n; nếu ở lần thử thứ i biến cố A xuất hiện ta cho X_i nhận giá trị 1, nếu biến cố A không xuất hiện ta cho X_i nhận giá trị 0

Như vậy X_i là biến ngẫu nhiên rời rạc có phân bố Bernoulli tham số p

Gọi X là số thành công trong n phép thử Bernoulli này thì

$$X=X_1+X_2+\cdots+X_n\sim\mathscr{B}(n;p)$$

Ví dụ: Một nguồn nhị phân phát ra hai ký số (digit) 1 và 0 một cách ngẫu nhiên với xác suất tương ứng 0,6 và 0,4.

1. Tính xác suất có đúng hai ký số 1 và ba ký số 0 trong dãy có năm ký số
2. Tính xác suất có ít nhất ba ký số 1 trong dãy có năm ký số

Giải: Gọi X là số các ký số 1 trong dãy có năm ký số

$$X \sim \mathscr{B}(5;0,6)$$

1. $P\{X=2\}=C_5^2(0,6)^2(0,4)^3=0,23$
2. $P\{X\le 2\}=\sum_{k=0}^{2}C_5^k(0,6)^k(0,4)^{5-k}=0,317$

$P\{X\ge 3\}=1-P\{X\le 2\}=1-0,317=0,683$

43

3. Phân bố Poisson

Biến ngẫu nhiên X được gọi là có phân bố Poisson tham số $\lambda>0$, ký hiệu $X \sim \mathscr{P}(\lambda)$, nếu X nhận các giá trị k = 0, 1, 2, … với hàm khối lượng xác suất

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

Hàm phân bố

$$F_X(x)=e^{-\lambda}\sum_{k=0}^{n}\frac{\lambda^k}{k!},\quad n\le x<n+1$$

44

Trong thực tế với một số giả thiết thích hợp thì các biến ngẫu nhiên là các quá trình đếm sau:

1. Số cuộc gọi đến một tổng đài,
2. Số khách hàng đến 1 điểm phục vụ,
3. Số xe cộ qua 1 ngã tư,
4. Số tai nạn (xe cộ); số các sự cố xảy ra ở một địa điểm …

trong một khoảng thời gian xác định nào đó sẽ có phân bố Poisson với tham số λ, trong đó λ là tốc độ trung bình diễn ra trong khoảng thời gian này.

45

Ví dụ: Ở một tổng đài điện thoại các cuộc gọi đến một cách ngẫu nhiên, độc lập. Ký hiệu $X(t)$ là số cuộc gọi đến tổng đài trong khoảng thời gian t phút.

Có thể chứng minh được $X(t)$ có phân bố Poisson tham số λt, trong đó λ là số cuộc gọi trung bình trong 1 phút.

Giả sử trung bình có 2 cuộc gọi trong 1 phút. Tìm xác suất:

1. Có đúng 5 cuộc gọi đến trong 2 phút (biến cố A).
2. Không có một cuộc gọi nào trong 30 giây (biến cố B).
3. Có ít nhất 1 cuộc gọi trong 10 giây (biến cố C).

46

Giải: Theo giả thiết $\lambda=2$, vậy ta có

1. $X(2)\sim\mathscr{P}(4)$, do đó

$$P(A)=P\{X(2)=5\}=e^{-4}\frac{4^5}{5!}\approx 0{,}156$$

2. $X(1/2)\sim\mathscr{P}(1)$, do đó

$$P(B)=P\{X(1/2)=0\}=e^{-1}\approx 0{,}3679$$

3. $X(1/6)\sim\mathscr{P}(1/3)$, do đó

$$P(C)=P\{X(1/6)\ge 1\}=1-P\{X(1/6)=0\}=1-e^{-1/3}\approx 0{,}2835$$

47

2.2.5. Biến ngẫu nhiên liên tục

1 Hàm mật độ xác suất của biến ngẫu nhiên liên tục

Giả sử X là một biến ngẫu nhiên liên tục có hàm phân bố xác suất $F_X(x)$

Hàm $f_X(x)$ thỏa mãn

$$F_X(x)=\int_{-\infty}^{x}f_X(t)dt,\ \text{với mọi}\ x\in\mathbb{R}$$

được gọi là hàm mật độ xác suất của biến ngẫu nhiên X (probability density function, viết tắt PDF)

48

Tính chất của hàm mật độ xác suất

1. $F_X'(x) = f_X(x)$ tại các điểm x mà $f_X(x)$ liên tục

   Vậy hàm phân bố của biến ngẫu nhiên liên tục là một nguyên hàm của hàm mật độ, và hàm mật độ là đạo hàm của hàm phân bố

2. $f_X(x) \geq 0$ với mọi $x \in \mathbb{R}$

3. $\int_{-\infty}^{\infty} f_X(x)dx = 1$

4. $P\{a < X < b\} = P\{a \leq X \leq b\} = P\{a \leq X < b\} = P\{a < X \leq b\} = \int_a^b f_X(x)dx$

49

Ví dụ: Hàm phân bố xác suất của biến ngẫu nhiên liên tục X có dạng

$$F_X(x) = \begin{cases} 0 & \text{với} \quad x < 0 \\ kx^2 & \text{với} \quad 0 \leq x < 1 \\ 1 & \text{với} \quad x \geq 1 \end{cases}$$

Xác định hệ số k và tìm hàm mật độ xác suất

Giải: Từ tính chất liên tục của hàm phân bố xác suất của biến ngẫu nhiên liên tục, ta có k=1

Hàm mật độ xác suất

$$f_X(x) = \begin{cases} 0 & \text{với} \quad x \leq 0 \\ 2x & \text{với} \quad 0 < x < 1 \\ 0 & \text{với} \quad x \geq 1 \end{cases}$$

50

2 Các phân bố liên tục thường gặp

1. Phân bố đều U(a; b)

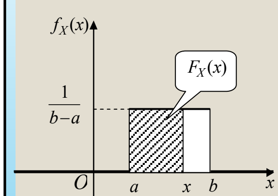
Biến ngẫu nhiên X được gọi là có phân bố đều trong khoảng (a; b) nếu hàm mật độ xác suất của nó xác định bởi

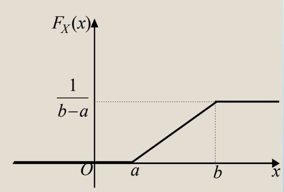
$$f_X(x) = \begin{cases} \dfrac{1}{b-a} & \text{nếu} \quad a < x < b \\ 0 & \text{nếu ngược lại} \end{cases}$$

Hàm phân bố xác suất

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt = \begin{cases} 0 & \text{nếu} \quad x \leq a \\ \dfrac{x-a}{b-a} & \text{nếu} \quad a < x < b \\ 1 & \text{nếu} \quad x \geq b \end{cases}$$

51





52

2. Phân bố mũ

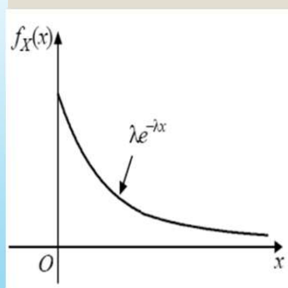
Biến ngẫu nhiên liên tục X được gọi là có phân bố mũ tham số $\lambda > 0$ nếu hàm mật độ xác suất xác định như sau

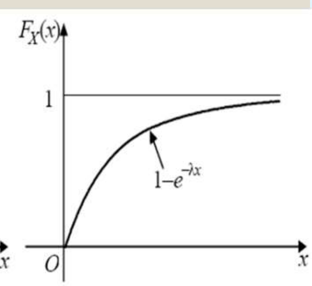
$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{nếu} \quad x > 0 \\ 0 & \text{nếu} \quad x \leq 0 \end{cases}$$

Hàm phân bố xác suất

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt = \begin{cases} 1 - e^{-\lambda x} & \text{nếu} \quad x > 0 \\ 0 & \text{nếu} \quad x \leq 0 \end{cases}$$

53

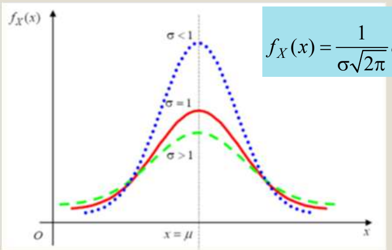




54

3. Phân bố chuẩn $\mathbf{N}(\mu;\sigma^2)$

Biến ngẫu nhiên liên tục X có phân bố chuẩn $\mathbf{N}(\mu;\sigma^2)$, ký hiệu $X \sim \mathbf{N}(\mu;\sigma^2)$, nếu hàm mật độ xác suất có dạng



$$f_X(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}\ ;\ \forall x\in\mathbb{R}$$

55

Phân bố chuẩn tắc $\mathbf{N}(0;1)$

Hàm mật độ xác suất

$$\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}},\quad \forall x\in\mathbb{R}$$

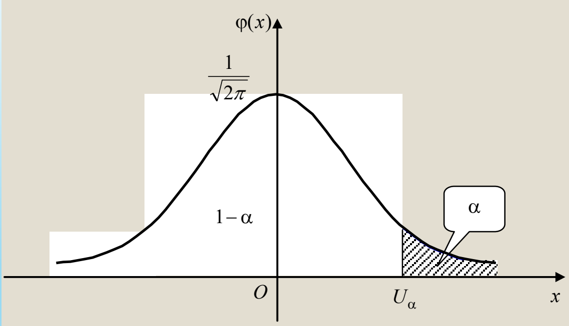
Hàm phân bố xác suất

$$\Phi(x)=\int_{-\infty}^{x}\varphi(t)dt=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{t^2}{2}}dt,\quad \forall x\in\mathbb{R}$$

Giá trị U_α gọi là giá trị tới hạn mức α của phân bố chuẩn tắc nếu

$$\Phi^{-1}(1-\alpha)=U_\alpha$$

56



57

Các tính chất của hàm phân bố xác suất $\Phi(x)$

1. $\Phi(x)+\Phi(-x)=1$, $\Phi(-x)=1-\Phi(x)$.
2. Nếu $X \sim N(0;1)$ thì

$$\forall a>0,\ P\{|X|<a\}=2\Phi(a)-1,\ P\{|X|>a\}=2\left(1-\Phi(a)\right)$$

$$P\{X>U_\alpha\}=\alpha\ ;\ P\left\{|X|>U_{\frac{\alpha}{2}}\right\}=\alpha\ ;\ P\left\{|X|<U_{\frac{\alpha}{2}}\right\}=1-\alpha$$

3. Nếu $X\sim\mathbf{N}(\mu;\sigma^2)$ thì $\frac{X-\mu}{\sigma}\sim\mathbf{N}(0;1)$

$$F_X(x)=P\{X\le x\}=P\left\{\frac{X-\mu}{\sigma}\le\frac{x-\mu}{\sigma}\right\}=\Phi\left(\frac{x-\mu}{\sigma}\right)$$

$$P\{a<X<b\}=P\{a<X\le b\}=\Phi\left(\frac{b-\mu}{\sigma}\right)-\Phi\left(\frac{a-\mu}{\sigma}\right)$$

58

Ví dụ: Giả sử $X\sim\mathbf{N}(\mu;\sigma^2);\mu=2100,\ \sigma=200$. Hãy tìm:

1. $P\{X<2400\}$, $P\{1700<X<2200\}$
2. Xác định a để $P\{X>a\}=0{,}03$

Giải:

$$P\{X<2400\}=P\{X\le 2400\}=\Phi\left(\frac{2400-2100}{200}\right)=\Phi(1{,}5)=0{,}9332$$

$$P\{1700<X<2200\}=\Phi\left(\frac{2200-2100}{200}\right)-\Phi\left(\frac{1700-2100}{200}\right)$$

$$=\Phi(0{,}5)-\Phi(-2)=0{,}6688$$

$$P\{X>a\}=1-\Phi\left(\frac{a-2100}{200}\right)=0{,}03\ \Rightarrow\ \Phi\left(\frac{a-2100}{200}\right)=0{,}97$$

59

4. Phân bố "khi bình phương"

Biến ngẫu nhiên liên tục X có phân bố "khi bình phương" n bậc tự do, ký hiệu $X\sim\chi^2_n$ nếu hàm mật độ xác suất có dạng

$$f_X(x)=\begin{cases}\dfrac{(x/2)^{n/2-1}}{\Gamma(n/2)}\dfrac{1}{2}e^{-(x/2)} & \text{nếu } x>0\\ 0 & \text{nếu } x\le 0\end{cases}$$

trong đó, $\Gamma(x)=\int_0^{+\infty}t^{x-1}e^{-t}dt,\ x>0$ là hàm Gamma.

Nếu là các biến ngẫu nhiên độc lập có cùng phân bố chuẩn tắc N(0;1) thì $X_1,X_2,\dots,X_n$

$$\sum_{i=1}^{n}X_i^2=X_1^2+X_2^2+\dots+X_n^2\sim\chi^2_n$$

60

5. Phân bố Student $\mathbf{T}(n)$

Biến ngẫu nhiên liên tục T có phân bố Student n bậc tự do, ký hiệu $T \sim \mathbf{T}(n)$, nếu hàm mật độ xác suất có dạng

$$f_T(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma(n/2)}\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}}, \quad -\infty < x < \infty$$

Có thể chứng minh được rằng nếu $Z \sim \mathbf{N}(0;1)$, $V \sim \chi^2_n$; Z và V độc lập thì

$$T = \frac{Z}{\sqrt{V/n}} \sim \mathbf{T}(n)$$

61

Giá trị tới hạn mức α của phân bố "khi bình phương" n bậc tự do, ký hiệu $\chi^2_\alpha(n)$, được xác định như sau

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

Giá trị tới hạn mức α của phân bố Student n bậc tự do, ký hiệu $t_\alpha(n)$, được xác định như sau

$$P\left\{T > t_\alpha(n)\right\} = \alpha$$

Bảng tính các giá trị tới hạn $t_\alpha(n)$ cho trong Phụ lục III và $\chi^2_\alpha(n)$ trong Phụ lục IV

62

2.2.6. CÁC THAM SỐ ĐẶC TRƯNG CỦA BIẾN NGẪU NHIÊN

1 Kỳ vọng toán

Kỳ vọng của biến ngẫu nhiên X ký hiệu EX và xác định như sau

$$\mathrm{E}X = \sum_{x_i \in R_X} x_i p_X(x_i) \quad \text{nếu } X \text{ rời rạc}$$

$$\mathrm{E}X = \int_{-\infty}^{\infty} x f_X(x)dx \quad \text{nếu } X \text{ liên tục}$$

63

Ví dụ: Theo thống kê việc một người Mỹ 25 tuổi sẽ sống thêm trên một năm có xác suất là 0,992, còn xác suất để người đó chết trong vòng một năm tới là 0,008. Một chương trình bảo hiểm đề nghị người đó bảo hiểm sinh mạng cho 1 năm với số tiền chi trả 1000 đô la, còn tiền đóng là 10 đô la. Hỏi lợi nhuận của công ty bảo hiểm nhận được là bao nhiêu?

Giải:

Lợi nhuận là biến ngẫu nhiên X với 2 giá trị là +10 đô la (nếu người bảo hiểm không chết) và –990 đô la (nếu người đó chết). Bảng phân bố xác suất tương ứng

X	–990	+10
P	0,008	0,992

$$\mathrm{E}X = (-990)\cdot 0{,}008 + 10\cdot 0{,}992 = 2$$

64

Ví dụ: Tuổi thọ của một loại côn trùng nào đó là một biến ngẫu nhiên X (đơn vị là tháng) với hàm mật độ xác suất như sau

$$f_X(x) = \begin{cases} kx^2(4-x) & \text{nếu } 0 \le x \le 4 \\ 0 & \text{nếu ngược lại} \end{cases}$$

Tìm tuổi thọ trung bình của loài côn trùng trên

Giải: $$\int_0^4 x^2(4-x)dx = \frac{64}{3} \Rightarrow k = \frac{3}{64}$$

Tuổi thọ trung bình

$$\mathrm{E}X = \frac{3}{64}\int_0^4 x^3(4-x)dx = \frac{3}{64}\left(x^4 - \frac{x^5}{5}\right)\Bigg|_0^4 = \frac{12}{5} \text{ (tháng)}$$

65

Tính chất kỳ vọng

1. E(C)=C với mọi hằng số C
2. E(CX)=C E(X) với mọi hằng số C
3. $\mathrm{E}(X_1+\cdots+X_n) = \mathrm{E}(X_1)+\cdots+\mathrm{E}(X_n)$
4. Nếu $X_1, ..., X_n$ độc lập thì $\mathrm{E}(X_1\cdots X_n) = \mathrm{E}(X_1)\cdots\mathrm{E}(X_n)$
5. Cho hàm số $g(x)$, kỳ vọng của biến ngẫu nhiên $Y=g(X)$ được tính theo công thức

$$\mathrm{E}Y = \begin{cases} \sum_{x_i \in R_X} g(x_i)p_X(x_i) & \text{nếu } X \text{ rời rạc với } p_X(x_i) = P\{X = x_i\} \\ \int_{-\infty}^{\infty} g(x) f_X(x)dx & \text{nếu } X \text{ liên tục có hàm mật độ } f_X(x) \end{cases}$$

66

2 Phương sai

Phương sai (variance) hay độ lệch (deviation) bình phương trung bình của biến ngẫu nhiên X là đại lượng đo sự phân tán bình phương trung bình của X xung quanh giá trị trung bình EX. Nói cách khác phương sai của X là kỳ vọng của $(X-EX)^2$

Phương sai của X được ký hiệu DX hoặc Var$X$

$$DX = E\left(X - EX\right)^2$$

Độ lệch chuẩn của X: $\sigma_X = \sqrt{DX}$

Khai triển vế phải công thức trên và áp dụng các tính chất của kỳ vọng ta có thể tính phương sai theo công thức sau

$$DX = EX^2 - \left(EX\right)^2$$

67

$EX^2 = \sum_i x_i^2 p_X(x_i)$ nếu X rời rạc

$EX^2 = \int_{-\infty}^{\infty} x^2 f_X(x)dx$ nếu X liên tục

Ví dụ 2.24: Tính phương sai của biến ngẫu nhiên xét trong ví dụ 2.19.

$$EX^2 = (-990)^2 \cdot 0,008 + 10^2 \cdot 0,992 = 7940$$

$$\Rightarrow DX = EX^2 - (EX)^2 = 7940 - 4 = 7936$$

$$\Rightarrow \sigma_X = \sqrt{DX} = \sqrt{7936} \approx 89,08$$

Điều này nói lên rằng mặc dù kinh doanh bảo hiểm có lãi nhưng rủi ro khá lớn

68

Tính chất của phương sai

1. D(a)=0 với mọi hằng số a
2. $D(aX+b) = a^2D(X)$ với mọi hằng số a, b
3. Nếu $X_1, ..., X_n$ độc lập và có các phương sai hữu hạn thì

$$D\left(a_1X_1 + \cdots + a_nX_n\right) = a_1^2D\left(X_1\right) + \cdots + a_n^2D\left(X_n\right)$$

Nói riêng: Nếu X, Y độc lập và DX, DY hữu hạn thì

$$D\left(X \pm Y\right) = DX + DY$$

69

Phân vị

Phân vị mức α của biến ngẫu nhiên X, ký hiệu v_α, là giá trị phân chia miền giá trị của X thỏa mãn

$$P\left\{X < v_\alpha\right\} \le \alpha \le P\left\{X \le v_\alpha\right\}$$

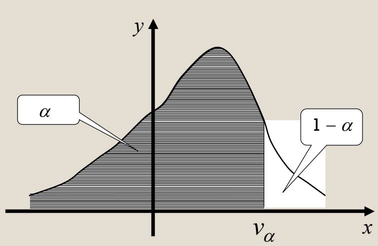
Nghĩa là $\quad F_X(v_\alpha -) \le \alpha \le F_X(v_\alpha)$

Trung vị

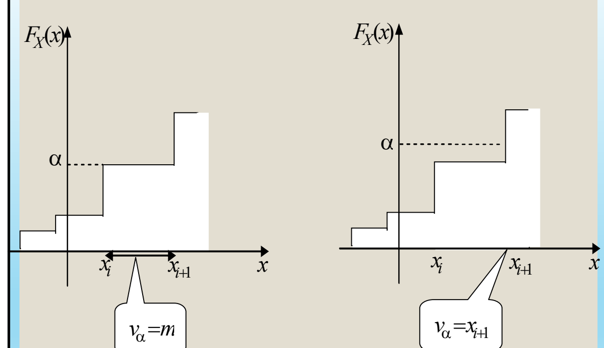
Phân vị mức 1/2 được gọi là median hay trung vị của X, ký hiệu Med X

Như vậy trung vị là điểm phân chia phân bố xác suất thành hai phần bằng nhau

70



71



72

Mốt

Mốt (Mode) của biến ngẫu nhiên X, ký hiệu Mod X, là giá trị mà biến ngẫu nhiên X nhận với xác suất lớn nhất

Một biến ngẫu nhiên có thể có nhiều Mốt

- Mốt của biến ngẫu nhiên rời rạc X

$$x_{i_0} = \text{Mod}\, X \Leftrightarrow p_X(x_{i_0}) = \max\{p_X(x_1), p_X(x_2), \ldots\}$$

- Mốt của biến ngẫu nhiên liên tục X

$$c = \text{Mod}\, X \Leftrightarrow f_X(c) = \max\{f_X(x),\ x \in \mathbb{R}\}$$

Moment, hệ số bất đối xứng, hệ số nhọn

1. Moment cấp k $\quad m_k = \mathrm{E}X^k;\ k = 1, 2, \ldots$
2. Moment quy tâm cấp k $\quad \mu_k = \mathrm{E}(X - \mathrm{E}X)^k;\ k = 1, 2, \ldots$
3. Hệ số bất đối xứng $\quad \alpha_3 = \dfrac{\mu_3}{\sigma^3};\ \sigma = \sqrt{\mathrm{D}X}$
4. Hệ số nhọn $\quad \alpha_4 = \dfrac{\mu_4}{\sigma^4}$

α_3 đo mức độ bất đối xứng của luật phân bố

α_4 đặc trưng cho độ nhọn của đồ thị hàm mật độ xác suất

2.3 KHÁI NIỆM VÉC TƠ NGẪU NHIÊN

2.3.1 Định nghĩa và phân loại

Một véc tơ ngẫu nhiên *n* chiều là một bộ có thứ tự $(X_1, X_2, \ldots, X_n)$ với các thành phần $X_1, X_2, \ldots, X_n$ là các biến ngẫu nhiên xác định trong cùng một phép thử.

Ta ký hiệu véc tơ ngẫu nhiên hai chiều là (*X*, *Y*), trong đó *X* là biến ngẫu nhiên thành phần thứ nhất và *Y* là biến ngẫu nhiên thành phần thứ hai

Véc tơ ngẫu nhiên *n* chiều là rời rạc hoặc liên tục nếu tất cả các biến ngẫu nhiên thành phần là rời rạc hoặc liên tục

2.3.2 Hàm phân bố xác suất

Hàm phân bố xác suất của véc tơ ngẫu nhiên $(X_1, X_2, \ldots, X_n)$ hay còn được gọi là hàm phân bố xác suất đồng thời của các biến ngẫu nhiên $X_1, X_2, \ldots, X_n$

$$F_{X_1 \ldots X_n}(x_1, x_2, \ldots, x_n) = P\{X_1 \le x_1, X_2 \le x_2, \ldots, X_n \le x_n\}$$

Các tính chất của hàm phân bố xác suất đồng thời

1. $0 \le F_{X_1 \ldots X_n}(x_1, \ldots, x_n) \le 1$
2. $\lim\limits_{x_k \to -\infty} F_{X_1 \ldots X_n}(x_1, \ldots, x_n) = 0$, với k nào đó thuộc $\{1, \ldots, n\}$
3. $\lim\limits_{(x_1, \ldots, x_n) \to (\infty, \ldots, \infty)} F_{X_1 \ldots X_n}(x_1, \ldots, x_n) = 1$
4. $F_{X_1 \ldots X_n}(x_1, \ldots, x_n)$ không giảm theo từng biến

5. $\lim\limits_{x_1 \to \infty} F_{X_1 \ldots X_n}(x_1, \ldots, x_n) = F_{X_2 \ldots X_n}(x_2, \ldots, x_n)$

Tương tự nếu lấy giới hạn của hàm phân bố xác suất đồng thời của $X_1, X_2, \ldots, X_n$ khi biến x_k tiến đến vô cùng, với *k* nào đó thuộc {1, … , *n*}, thì được hàm phân bố xác suất đồng thời của *n*–1 biến ngẫu nhiên còn lại $X_1, \ldots, X_{k-1}, X_{k+1}, \ldots, X_n$

Đặc biệt nếu $F_{XY}(x, y)$ là hàm phân bố xác suất của véc tơ ngẫu nhiên hai chiều (*X*,*Y*) thì

$$\lim_{y \to \infty} F_{XY}(x, y) = P\{X \le x\} = F_X(x)$$

$$\lim_{x \to \infty} F_{XY}(x, y) = P\{Y \le y\} = F_Y(y)$$

$$P\{x_1 < X \le x_2, y_1 < Y \le y_2\} = F_{XY}(x_2, y_2) - F_{XY}(x_1, y_2) - F_{XY}(x_2, y_1) + F_{XY}(x_1, y_1)$$

2.3.3 Bảng phân bố xác suất của véc tơ ngẫu nhiên rời rạc hai chiều

Tương tự trường hợp biến ngẫu nhiên rời rạc, quy luật phân bố xác suất của véc tơ ngẫu nhiên rời rạc 2 chiều có thể được xác định thông qua hàm khối lượng xác suất đồng thời hoặc bảng phân bố xác suất đồng thời

1 Hàm khối lượng xác suất đồng thời

$$p_{XY}(x_i, y_j) = P\{X = x_i, Y = y_j\}$$

thỏa mãn điều kiện $\begin{cases} p_{X,Y}(x_i, y_j) \ge 0,\ \forall i = 1, \ldots, n,\ j = 1, \ldots, m \\ \sum\limits_{i=1}^{n} \sum\limits_{j=1}^{m} p_{X,Y}(x_i, y_j) = 1 \end{cases}$

Hàm phân bố xác suất đồng thời $F_{X,Y}(x, y) = \sum\limits_{x_i \le x} \sum\limits_{y_j \le y} p_{X,Y}(x_i, y_j)$

$X \backslash Y$	y_1	y_2	…	y_j	…	y_m
x_1	$p_{XY}(x_1,y_1)$	$p_{XY}(x_1,y_2)$	…	$p_{XY}(x_1,y_j)$	…	$p_{XY}(x_1,y_m)$
x_2	$p_{XY}(x_2,y_1)$	$p_{XY}(x_2,y_2)$	…	$p_{XY}(x_2,y_j)$	…	$p_{XY}(x_2,y_m)$
⋮	⋮	⋮	…	⋮	…	⋮
x_i	$p_{XY}(x_i,y_1)$	$p_{XY}(x_i,y_2)$	…	$p_{XY}(x_i,y_j)$	…	$p_{XY}(x_i,y_m)$
⋮	⋮	⋮	…	⋮	…	⋮
x_n	$p_{XY}(x_n,y_1)$	$p_{XY}(x_n,y_2)$	…	$p_{XY}(x_n,y_j)$	…	$p_{XY}(x_n,y_m)$

79

2 Hàm khối lượng xác suất biên

$$p_Y(y_j)=P\{Y=y_j\}=\sum_{i=1}^{n}P\{X=x_i,Y=y_j\}=\sum_{i=1}^{n}p_{XY}(x_i,y_j);\ j=\overline{1,m}$$

$$p_X(x_i)=P\{X=x_i\}=\sum_{j=1}^{m}P\{X=x_i,Y=y_j\}=\sum_{j=1}^{m}p_{XY}(x_i,y_j);i=\overline{1,n}$$

Bảng phân bố xác suất biên

Từ bảng phân bố xác suất đồng thời của (*X*, *Y*), nếu ta cộng các xác suất theo cột thì ta được các xác suất tương ứng với các giá trị của *Y*.

Nếu ta cộng các xác suất theo hàng ta được các xác suất tương ứng với giá trị của *X* .

80

Ví dụ: Gieo 3 đồng tiền cân đối **A**, **B**, **C**. Gọi *X* là số mặt ngửa xuất hiện của 2 đồng tiền **A**, **B** và *Y* là số mặt ngửa xuất hiện của cả 3 đồng tiền **A**, **B**, **C**. Hãy lập bảng phân bố xác suất đồng thời của *X*, *Y*.

Các kết quả đồng khả năng

$\mathscr{A}$	N	N	N	N	S	S	S	S
$\mathscr{B}$	N	N	S	S	N	N	S	S
$\mathscr{C}$	N	S	N	S	N	S	N	S
X	2	2	1	1	1	1	0	0
Y	3	2	2	1	2	1	1	0

81

Bảng phân bố xác suất đồng thời của *X* và *Y*

$X \backslash Y$	0	1	2	3
0	1/8	1/8	0	0
1	0	2/8	2/8	0
2	0	0	1/8	1/8

Bảng phân bố xác suất của hai biến ngẫu nhiên thành phần

X	0	1	2
P	2/8	4/8	2/8

Y	0	1	2	3
P	1/8	3/8	3/8	1/8

82

2.3.4 Hàm mật độ xác suất của véc tơ ngẫu nhiên liên tục

Hàm *n* biến $f_{X_1X_2...X_n}(x_1,x_2,...,x_n)\ge 0$ thoả mãn

$$F_{X_1X_2...X_n}(x_1,x_2,...,x_n)=\int_{-\infty}^{x_1}\cdots\int_{-\infty}^{x_n}f_{X_1X_2...X_n}(t_1,t_2,...,t_n)dt_1dt_2...dt_n$$

là hàm mật độ xác suất của véc tơ ngẫu nhiên liên tục $(X_1,X_2,...,X_n)$

hoặc hàm mật độ xác suất đồng thời của các biến ngẫu nhiên $X_1,X_2,...,X_n$

83

Tính chất của hàm mật độ xác suất

1. $f_{XY}(x,y)\ge 0$ với mọi (x,y) và $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f_{XY}(x,y)dxdy=1$
2. $P\{(X,Y)\in A\}=\iint_{(x,y)\in A\cap R_{XY}}f_{XY}(x,y)dxdy$ với mọi $A\subset\mathbb{R}^2$,

   R_{XY} là miền giá trị của (X,Y)
3. $f_{XY}(x,y)=\begin{cases}\dfrac{\partial^2}{\partial x\partial y}F_{XY}(x,y) & \text{nếu tồn tại đạo hàm tại } (x,y)\\ 0 & \text{nếu ngược lại}\end{cases}$

84

CHƯƠNG 2: BỔ SUNG KIẾN THỨC VỀ XÁC SUẤT THỐNG KÊ

Hàm mật độ xác suất biên

Hàm mật độ xác suất của biến ngẫu nhiên thành phần X

$$\int_{-\infty}^{\infty} f_{XY}(x,y)\,dy = f_X(x)$$

Hàm mật độ xác suất của biến ngẫu nhiên thành phần Y

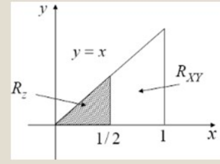
$$\int_{-\infty}^{\infty} f_{XY}(x,y)\,dx = f_Y(y)$$

85

CHƯƠNG 2: BỔ SUNG KIẾN THỨC VỀ XÁC SUẤT THỐNG KÊ

Ví dụ 3.7: Cho véc tơ ngẫu nhiên (X,Y) có hàm mật độ xác suất xác định như sau

$$f_{XY}(x,y)=\begin{cases} k & \text{nếu } 0<y\le x<1 \\ 0 & \text{nếu ngược lại} \end{cases}$$



Miền giá trị của X, Y là tam giác R_{XY}

$$1=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)\,dxdy = k\frac{1}{2} \Rightarrow k=2$$

$$f_X(x)=\begin{cases} \int_0^x 2dy = 2x & \text{nếu } 0<x<1 \\ 0 & \text{nếu ngược lại} \end{cases} \qquad f_Y(y)=\begin{cases} \int_y^1 2dx = 2(1-y) & \text{nếu } 0<y<1 \\ 0 & \text{nếu ngược lại} \end{cases}$$

$$P\{0<X\le 1/2; 0<Y\le 1/2\} = \iint_{R_z} f_{XY}(x,y)\,dxdy = 2\left(\frac{1}{8}\right)=\frac{1}{4}$$

86

CHƯƠNG 2: BỔ SUNG KIẾN THỨC VỀ XÁC SUẤT THỐNG KÊ

2.3.5 Các tham số đặc trưng của véc tơ ngẫu nhiên

1 Kỳ vọng và phương sai của các biến ngẫu nhiên thành phần

a. Trường hợp X, Y rời rạc

$$\mathrm{E}X=\sum_{i=1}^{n} x_i p_X(x_i)=\sum_{i=1}^{n}\sum_{j=1}^{m} x_i p_{XY}(x_i,y_j)$$

$$\mathrm{E}Y=\sum_{j=1}^{m} y_j p_Y(y_j)=\sum_{j=1}^{m}\sum_{i=1}^{n} y_j p_{XY}(x_i,y_j)$$

$$\mathrm{D}X=\mathrm{E}X^2-(\mathrm{E}X)^2;\ \mathrm{E}X^2=\sum_{i=1}^{n}\sum_{j=1}^{m} x_i^2 p_{XY}(x_i,y_j)$$

$$\mathrm{D}Y=\mathrm{E}Y^2-(\mathrm{E}Y)^2;\ \mathrm{E}Y^2=\sum_{j=1}^{m}\sum_{i=1}^{n} y_j^2 p_{XY}(x_i,y_j)$$

87

CHƯƠNG 2: BỔ SUNG KIẾN THỨC VỀ XÁC SUẤT THỐNG KÊ

b. Trường hợp X, Y liên tục

$$\mathrm{E}X=\int_{-\infty}^{\infty} x f_X(x)dx=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f_{X,Y}(x,y)dxdy$$

$$\mathrm{E}Y=\int_{-\infty}^{\infty} y f_Y(y)dy=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y f_{X,Y}(x,y)dxdy$$

$$\mathrm{D}X=\mathrm{E}X^2-(\mathrm{E}X)^2;\ \mathrm{E}X^2=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^2 f_{X,Y}(x,y)dxdy$$

$$\mathrm{D}Y=\mathrm{E}Y^2-(\mathrm{E}Y)^2;\ \mathrm{E}Y^2=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y^2 f_{X,Y}(x,y)dxdy$$

88

CHƯƠNG 2: BỔ SUNG KIẾN THỨC VỀ XÁC SUẤT THỐNG KÊ

2 Hiệp phương sai

Hiệp phương sai (hay còn gọi là Covariance) của hai biến ngẫu nhiên X, Y, ký hiệu $\mathrm{cov}(X,Y)$, là kỳ vọng toán của tích các sai lệch của hai biến ngẫu nhiên đó với kỳ vọng toán của chúng

$$\mathrm{cov}(X,Y)=\mathrm{E}(X-\mathrm{E}X)(Y-\mathrm{E}Y)$$

Khai triển vế phải và áp dụng tính chất của kỳ vọng ta được

$$\mathrm{cov}(X,Y)=\mathrm{E}XY-(\mathrm{E}X)(\mathrm{E}Y)$$

Nếu X, Y rời rạc thì $\mathrm{E}XY=\sum_{j=1}^{m}\sum_{i=1}^{n} x_i y_j p_{XY}(x_i,y_j)$

Nếu X, Y liên tục thì $\mathrm{E}XY=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f_{X,Y}(x,y)dxdy$

89

CHƯƠNG 2: BỔ SUNG KIẾN THỨC VỀ XÁC SUẤT THỐNG KÊ

Tính chất của hiệp phương sai

1. $\mathrm{cov}(X,Y)=\mathrm{cov}(Y,X)$
2. $\mathrm{cov}(X,X)=\mathrm{D}X$
3. $\mathrm{cov}(aX+c,bY+d)=ab\,\mathrm{cov}(Y,X)$ với mọi hằng số a,b,c,d
4. Nếu X,Y độc lập thì $\mathrm{cov}(X,Y)=0$

   nhưng ngược lại chưa chắc đúng

90

3 Ma trận hiệp phương sai

Ma trận $M = \left[C_{ij}\right]_{n\times n}$

với $C_{ij} = \operatorname{cov}(X_i, X_j);\ i,j = 1,\dots,n$

được gọi là ma trận hiệp phương sai (ma trận covariance) của véc tơ ngẫu nhiên $X=(X_1, X_2, \dots, X_n)$

Tính chất của ma trận hiệp phương sai

1. Ma trận hiệp phương sai là ma trận đối xứng
2. Với mọi $t_1, t_2, \dots, t_n \in \mathbb{R}^n$ luôn có $\sum_j \sum_i C_{ij} t_i t_j \ge 0$
3. Các định thức con chính của M không âm

91

4 Hệ số tương quan

Hệ số tương quan của hai biến ngẫu nhiên X, Y ký hiệu và định nghĩa bởi công thức

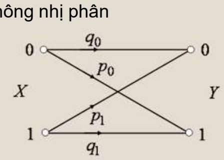
$$\rho_{X,Y} = \frac{\operatorname{cov}(X,Y)}{\sqrt{\mathrm{D}X}\sqrt{\mathrm{D}Y}}$$

Tính chất của hệ số tương quan

1. $-1 \le \rho_{X,Y} \le 1$ với mọi X, Y
2. Nếu X, Y độc lập thì $\rho_{X,Y} = 0$ điều ngược lại chưa chắc đúng
3. Với mọi hằng số a,b,c,d: $\rho_{aX+c,bY+d} = \begin{cases} \rho_{X,Y} & \text{nếu } ab>0 \\ -\rho_{X,Y} & \text{nếu } ab<0 \end{cases}$
4. $Y = aX + b$, $a \ne 0$ khi và chỉ khi $\rho_{X,Y} = \begin{cases} 1 & \text{nếu } a>0 \\ -1 & \text{nếu } a<0 \end{cases}$

92

Xét kênh viễn thông nhị phân



Ký hiệu X là đầu vào và Y là đầu ra của kênh

$$P\{X=0\} = 0{,}5$$

$$p_0 = P\{Y=1 | X=0\} = 0{,}1$$

$$p_1 = P\{Y=0 | X=1\} = 0{,}2$$

93

(X,Y) là véc tơ ngẫu nhiên có bảng phân bố xác suất đồng thời

X \ Y	0	1
0	0,45	0,05
1	0,10	0,40

Bảng phân bố xác suất thành phần X và Y

X	0	1
P	0,5	0,5

Y	0	1
P	0,55	0,45

X và Y không độc lập

94

$$\mathrm{E}X = 0{,}5,\ \mathrm{E}X^2 = 0{,}5 \Rightarrow \mathrm{D}X = 0{,}25$$

$$\mathrm{E}Y = 0{,}45,\ \mathrm{E}Y^2 = 0{,}45 \Rightarrow \mathrm{D}Y = 0{,}2475$$

Hiệp phương sai

$$\mathrm{E}XY = 0{,}4 \Rightarrow \operatorname{cov}(X,Y) = 0{,}4 - 0{,}5\times 0{,}45 = 0{,}175$$

Hệ số tương quan

$$\rho_{X,Y} = \frac{\operatorname{cov}(X,Y)}{\sqrt{\mathrm{D}X}\sqrt{\mathrm{D}Y}} = \frac{0{,}175}{\sqrt{0{,}25\times 0{,}2475}} = 0{,}704$$

Ma trận hiệp phương sai

$$M = \begin{bmatrix} 0{,}25 & 0{,}175 \\ 0{,}175 & 0{,}2475 \end{bmatrix}$$

Ta thấy giá trị $\rho_{X,Y} = 0{,}704$ khá xa 1, do đó Y không phụ thuộc tuyến tính đối với X

95