LỜI NÓI ĐẦU

Lý thuyết xác suất là bộ môn toán học nghiên cứu tính qui luật của các hiện tượng ngẫu nhiên.

Các khái niệm đầu tiên của xác suất do các nhà toán học tên tuổi Pierre Fermat (1601-1665) và Blaise Pascal (1623-1662) xây dựng vào giữa thế kỷ 17, dựa trên việc nghiên cứu các qui luật trong các trò chơi may rủi. Do sự hạn chế của trình độ toán học đương thời, nên suốt một thời gian dài các trò chơi may rủi vẫn là cơ sở duy nhất cho các khái niệm và phương pháp của lí thuyết xác suất với công cụ chủ yếu là phép tính tổ hợp và số học sơ cấp. Hiện nay, tuy lí thuyết xác suất đã có nền tảng toán học đồ sộ, nhưng các phương pháp "ngây thơ" ban đầu đó vẫn còn tác dụng, đặc biệt đối với các ngành khoa học thực nghiệm.

Việc giải quyết các bài toán nảy sinh trong lí thuyết sai số và đo lường đã đem lại bước phát triển mới cho lí thuyết xác suất. Các nhà toán học Jacob Bernoulli (1654-1705), A.Moivre (1667-1754), Laplace (1749-1827), Gauss (1777-1855), Poisson (1781-1840) đã có công lao xứng đáng phát triển lí thuyết xác suất bằng phương pháp giải tích.

Từ giữa thế kỉ 19 đến đầu thế kỉ 20, sự phát triển của lí thuyết xác suất gắn liền với tên tuổi các nhà toán học Nga như Bunhiacốpxki (1804-1889), Trebưsep (1821-1894), Markov (1856-1922) và Liapunov (1857-1918).

Trong quá trình phát triển mạnh mẽ của của lí thuyết xác suất, vấn đề xây dựng một cơ sở toán học chặt chẽ trở thành cấp thiết. Sự ra đời của lí thuyết tập hợp và độ đo đã cung cấp công cụ toán học giải quyết vấn đề này, và vinh quang xây dựng lí thuyết xác suất tiên đề thuộc về nhà toán học Nga Kolmogorov (1929).

Sự ra đời của thống kê toán học bắt nguồn từ các vấn đề thực tiễn và dựa trên những thành tựu của lí thuyết xác suất. Các thí nghiệm trong các ngành khoa học khác nhau như vật lý, hóa học, sinh học, y học, ... phụ thuộc vào nhiều yếu tố ngẫu nhiên như con người, môi trường,... Do đó kết quả thực nghiệm thường là các đại lượng ngẫu nhiên.

Có thể định nghĩa thống kê toán học là ngành khoa học về các phương pháp tổng quát xử lý các kết quả thực nghiệm. Cùng với sự phát triển của lí thuyết xác suất, thống kê toán học đã có bước tiến nhanh, với sự đóng góp của các nhà toán học như Gantơn (1822-1911), Piếcxơn (1857-1936), Cramer, Fisher, Von Neuman, ... Thống kê toán học đã có các ứng dụng hiệu quả trong nhiều lĩnh vực khoa học, công nghệ, kinh tế và xã hội khác nhau như vật lí, hóa học, cơ học, sinh vật, y học, dự báo, khí tượng, thủy văn, vô tuyến, điện tử, ngôn ngữ học, xã hội học, ...

Có thể nói lí thuyết xác suất và thống kê toán học đã trở thành kiến thức cơ sở không thể thiếu của mỗi kỹ sư tương lai.

Giáo trình được biên soạn lần đầu nên chắc chắn còn nhiều khiếm khuyết. Tác giả chân thành cảm ơn những ý kiến đóng góp quý báu của độc giả để giáo trình ngày một hoàn thiện.

Xin chúc các bạn thành công!

Đà nẵng 1/2005

Tác giả.

CHƯƠNG 0
GIẢI TÍCH KẾT HỢP

I. TẬP HỢP

1. CÁC KHÁI NIỆM CƠ BẢN

• ***Định nghĩa***: Khái niệm *tập hợp* là khái niệm nền tảng cho toán học cũng như ứng dụng của nó. Tập hợp là khái niệm nguyên thuỷ không định nghĩa chính xác dựa trên các khái niệm khác. Tập hợp được coi là kết hợp các đối tượng có cùng bản chất (thuộc tính, dấu hiệu) chung nào đó.

Tập hợp thường được ký hiệu bằng các chữ cái A, B, C , ... Các phần tử của tập hợp ký hiệu bằng các chữ thường a, b, c,...

Để chỉ x là phần tử của tập hợp X ta viết :

$$x \in X \text{ (đọc : x thuộc X)}$$

Để chỉ x không phải là phần tử của X ta viết :

$$x \notin X \text{ (đọc : x không thuộc X)}$$

Tập không có phần tử gọi là tập rỗng và ký hiệu $\varnothing$.

• ***Biểu diễn tập hợp***:

Có hai cách biểu diễn tập hợp như sau

(i) Liệt kê các phần tử :

+ *Ví dụ*

$A = \{ a, b, c \}$

$X = \{ x_1, x_2, \ldots , x_n \}$

(ii) Biểu diễn tập hợp bằng cách mô tả tính chất :

+ *Ví dụ*

$C = \{ n \mid n$ là số chẵn $\}$

$Y = \{ x \mid x$ là nghiệm phương trình $x^2 + 2x - 5 = 0 \}$

• ***Lực lượng tập hợp***:

Số phần tử của tập A, ký hiệu là $|A|$, gọi là *lực lượng* của tập A.

Nếu $|A| < \infty$, ta nói A là tập hữu hạn, nếu $|A| = \infty$, ta nói A là tập vô hạn.

Trong chương trình này ta giả thiết các tập hợp là hữu hạn.

• ***Quan hệ bao hàm***: Cho hai tập A, B.

Nếu mỗi phần tử thuộc A cũng thuộc B ta nói A là tập con của B và ký hiệu

$$A \subset B$$

Nếu A không phải tập con của B ta ký hiệu

$$A \not\subset B$$

Nếu $A \subset B$ và $B \subset A$ ta nói A bằng B và ký hiệu

$$A = B$$

Nếu $A \subset B$, $A \neq \varnothing$ và $B \neq A$, thì ta nói A là *tập con thực sự* của B.

+ *Ví dụ*

(i) Tập rỗng ∅ có lực lượng bằng 0, |∅| = 0. Với mọi tập A, ∅ ⊂ A.

(ii) Cho đa thức P(x). Ký hiệu S = {x | P(x) = 0}. S là tập hữu hạn.

(iii) Ký hiệu

N là tập số tự nhiên, **N** = {0, 1, 2, … };

Q là tập số hữu tỷ; **R** là tập só thực.

Ta có **N** ⊂ **Q** ⊂ **R.**

Bây giờ ta xét tập hữu hạn A. Ký hiệu tập tất cả tập con của A là **P**(A)

• **Định lý 1**. Nếu |A| = n , thì |**P**(A)| = 2^n

Chứng minh. Quy nạp theo n.

2. CÁC PHÉP TOÁN TẬP HỢP

Cho các tập A, B, X_1, X_2, ... , X_n (n ∈ **N**) là các tập con của tập "vũ trụ" **U** nào đó. Ta định nghĩa các phép toán sau.

+ *Phép hiệu: Hiệu của A và B, ký hiệu A \ B là tập:*

$$A \setminus B = \{ x \mid x \in A \ \&\ x \notin B \}$$

+ *Phần bù*: Phần bù của A (trong **U**) là tập

$$\overline{A} = U \setminus A$$

+ *Phép hợp: Hợp của A và B, ký hiệu A ∪ B là tập*

$$A \cup B = \{ x \mid x \in A \text{ hoặc } x \in B \}$$

Tương tự, hợp của X_1, X_2, ... , X_n là tập

$$\bigcup_{i=1}^{n} X_i = \{x \mid \exists k, 1 \le k \le n, x \in X_k\}$$

+ *Phép giao: Giao của A và B, ký hiệu A ∩ B là tập*

$$A \cap B = \{ x \mid x \in A \ \&\ x \in B \}$$

Tương tự, giao của X_1, X_2, ... , X_n là tập

$$\bigcap_{i=1}^{n} X_i = \{x \mid \forall k, 1 \le k \le n, x \in X_k\}$$

+ *Tích Đề-các*

- Tích Đề-các của hai tập A, B là tập

$$A \times B = \{ (a,b) \mid a \in A \ \& \ b \in B \}$$

- Tích Đề-các của các tập $X_1, X_2, \ldots, X_n$ là tập

$$X_1 \times X_2 \times \ldots \times X_n = \{ (x_1, x_2, \ldots, x_n) \mid x_1 \in X_1 \ \& \ x_2 \in X_2 \ \& \ldots \& \ x_n \in X_n \}$$

+ *Phân hoạch*:

- Nếu $A \cap B = \varnothing$, ta nói A và B *rời nhau*.

- Nếu các tập $X_1, X_2, \ldots, X_n$ thoả

$$A = X_1 \cup X_2 \cup \ldots \cup X_n$$

và chúng rời nhau từng đôi một, ta nói $\{ X_1, X_2, \ldots, X_n \}$ là một *phân hoạch* của tập hợp A.

• ***Định lý 1***. Giả sử $\{ X_1, X_2, \ldots, X_n \}$ là một phân hoạch của tập S. Khi đó

$$|S| = |X_1| + |X_2| + \ldots + |X_n|$$

Chứng minh. Hiển nhiên.

• ***Định lý 2***. Cho các tập A, B, C trong tập vũ trụ U, khi đó ta có :

(i) *Luật kết hợp* :

$$(A \cup B) \cup C = A \cup (B \cup C)$$
$$(A \cap B) \cap C = A \cap (B \cap C)$$

(ii) *Luật giao hoán* :

$$A \cup B = B \cup A$$
$$A \cap B = B \cap A$$

(iii) *Luật phân bố* :

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$
$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

(iv) *Luật đối ngẫu De Morgan*:

$$\overline{A \cup B} = \overline{A} \cap \overline{B} \qquad \& \qquad \overline{A \cap B} = \overline{A} \cup \overline{B}$$

$$\overline{\bigcup_{i=1}^{n} X_i} = \bigcap_{i=1}^{n} \overline{X_i} \quad \& \quad \overline{\bigcap_{i=1}^{n} X_i} = \bigcup_{i=1}^{n} \overline{X_i}$$

Chứng minh. (bài tập).

• ***Định lý 3*** (về lực lượng tập hợp).

(i) Lực lượng tập con:

$$A \subset B \Rightarrow |A| \leq |B|$$

(ii) Lực lượng của hợp

$$|A \cup B| = |A| + |B| - |A \cap B|$$

(iii) Nguyên lý bù trừ Poincaré:

$$\left|\bigcup_{k=1}^{n} A_k\right| = \sum_{m=1}^{n}\left[(-1)^{m-1} \sum_{1 \leq i_1 \leq i_2 \leq \ldots \leq i_m \leq n} \left|A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_m}\right|\right]$$

(iv) Lực lượng tích Đề-các

$$|X_1 x\, X_2\, x \ldots x\, X_n| = |X_1| \,.\, |X_2| \,.\, \ldots \,.\, |X_n|$$

(v) Lực lượng tương đương:

|A| = |B| ⇔ Tồn tại song ánh từ A vào B.

Chứng minh. (bài tập).

II. GIẢI TÍCH KẾT HỢP

1. BÀI TOÁN GIẢI TÍCH KẾT HỢP

Trong thực tế ta thường gặp bài toán sau: Cho một tập hữu hạn X. Các phần tử của X được chọn và ghép theo quy luật nào đó. Hãy tính số nhóm tạo thành. Ngành toán học nghiên cứu các bài toán loại này gọi là *Giải tích kết hợp*.

• ***Ví dụ***: Công ty phát hành sách bán sách thông qua hệ thống hiệu sách. Giả sử có 12 đầu sách và các đầu sách ký hiệu là 1, 2, …, 12. Có 3 khách hàng đến hiệu sách đặt mua, mỗi người 1 quyển. Gọi x_1, x_2, x_3 lần lượt là quyển sách mà khách hàng thứ nhất, thứ hai, thứ ba đặt mua ($x_1, x_2, x_3 \in \{1, 2, \dots, 12\}$).

Hỏi có bao nhiêu bộ (x_1, x_2, x_3) ?

Kết quả bài toán đếm này phụ thuộc vào việc ai giao sách: hiệu sách hay công ty.

(i) Trường hợp 1:

Người giao sách là hiệu sách và các khách hàng đặt mua các đầu sách khác nhau.

Khi đó hiệu sách cần biết thứ tự của bộ (x_1, x_2, x_3). Số bộ (x_1, x_2, x_3) sẽ là

$$12.11.10 = 1320$$

(ii) Trường hợp 2:

Người giao sách là hiệu sách và các khách hàng có thể đặt mua các đầu sách giống nhau.

Khi đó hiệu sách cần biết thứ tự của bộ (x_1, x_2, x_3) và x_1, x_2, x_3 có thể giống nhau . Số bộ (x_1, x_2, x_3) sẽ là

$$12^3 = 1728$$

(iii) Trường hợp 3:

Người giao sách là công ty và các khách hàng đặt mua các đầu sách khác nhau.

Khi đó công ty không cần biết thứ tự của bộ (x_1, x_2, x_3). Số bộ (x_1, x_2, x_3) sẽ là

$$12.11.10 / 1.2.3 = 1320 / 6 = 220$$

(iv) Trường hợp 4:

Người giao sách là công ty và các khách hàng có thể đặt mua các đầu sách giống nhau.

Khi đó công ty không cần biết thứ tự của bộ (x_1, x_2, x_3) và x_1, x_2, x_3 có thể giống nhau. Số bộ (x_1, x_2, x_3) sẽ gồm các trường hợp sau:

+ Trường hợp 3 người cùng đặt mua 1 đầu sách:

có 12 khả năng.

+ Trường hợp 3 người cùng đặt mua 2 đầu sách:

có C(12,2). 2 = 132 khả năng (C(n, k) là số tổ hợp chập k của n phần tử).

+ Trường hợp 3 người cùng đặt mua 3 đầu sách:
có 220 khả năng.

Tổng cộng số bộ (x_1, x_2, x_3) là

$$12 + 132 + 220 = 364$$

2. CÁC KẾT HỢP CƠ BẢN

a) Nguyên lý nhân:

Xét bài toán giải tích kết hợp ở trên. Ta giả sử mỗi nhóm kết hợp các phần tử của tập X được xây dựng qua k bước:

Bước 1 có n_1 khả năng
Bước 2 có n_2 khả năng
. . .
Bước k có n_k khả năng

Khi đó số nhóm kết hợp là

$$n_1.n_2. \ldots . n_k$$

b) Chỉnh hợp

+ *Định nghĩa*: Một chỉnh hợp chập k của n phần tử là một bộ có thứ tự gồm k thành phần lấy từ n phần tử đã cho. Các thành phần không được lặp lại.

Một chỉnh hợp chập k của n có thể được xây dựng qua k bước kế tiếp như sau :

Chọn thành phần đầu : có n khả năng.
Chọn thành phần thứ hai : có n - 1 khả năng.
...
Chọn thành phần thứ k : có n - k + 1 khả năng.

Như vậy, theo nguyên lý nhân, số tất cả chỉnh hợp không lặp chập k của n phần tử là

$$A(n,k) = n(n-1)...(n-k+1) = \frac{n!}{(n-k)!}$$

+ *Ví dụ 1*:

Tính số hàm đơn ánh từ tập X có k phần tử đến tập Y có n phần tử.

Giải : Mỗi hàm đơn ánh từ X vào Y tương ứng với một chỉnh hợp không lặp chập k của n phần tử của Y. Như vậy số cần tìm là

$$A(n, k) = n.(n-1).....(n-k+1).$$

+ *Ví dụ 2*:

Quay lại ví dụ ở mục trước. Trong trường hợp 1, mỗi bộ (x_1, x_2, x_3) là một chỉnh hợp chập 3 của 12. Vậy số bộ là

$$A(12, 3) = 12.11.10 = 1320$$

c) Chỉnh hợp lặp

+ *Định nghĩa*: Một *chỉnh hợp lặp* chập k của n phần tử là một bộ có thứ tự gồm k thành phần lấy từ n phần tử đã cho. Các thành phần có thể được lặp lại.

Một chỉnh hợp lặp chập k của n có thể xem như một phần tử của tích Đề-các X^k, với X là tập n phần tử. Như vậy số tất cả các chỉnh hợp lặp chập k của n là: $\boldsymbol{n^k}$

+ *Ví dụ* 1: Tính số hàm từ tập X có k phần tử đến tập Y có n phần tử.

Mỗi hàm từ X vào Y tương ứng với một bộ có thứ tự k thành phần của n phần tử của Y, các phần tử có thể lặp lại . Như vậy số hàm từ X vào Y là n^k .

+ *Ví dụ 2*:

Quay lại ví dụ ở mục trước. Trong trường hợp 2, mỗi bộ (x_1, x_2, x_3) là một chỉnh hợp lặp chập 3 của 12. Vậy số bộ là

$$12^3 = 1728$$

d) Hoán vị

+ *Định nghĩa* : Một hoán vị của n phần tử là một cách sắp xếp thứ tự các phần tử đó.

Hoán vị có thể coi như trường hợp riêng của chỉnh hợp không lặp chập k của n trong đó k = n. Ta có số hoán vị là

$$\boldsymbol{P(n) = n!}$$

+ *Ví dụ*: Có 6 người xếp thành hàng ngang để chụp ảnh. Hỏi có thể bố trí bao nhiêu kiểu khác nhau ?

Giải: Mỗi kiểu ảnh là một hoán vị của 6 người. Vậy số kiểu ảnh là 6! = 720.

e) Tổ hợp

+ *Định nghĩa*: Một tổ hợp chập k của n phần tử là một bộ không kể thứ tự gồm k thành phần khác nhau lấy từ n phần tử đã cho. Nói cách khác ta có thể coi một tổ hợp chập k của n phần tử là một tập con có k phần tử của n phần tử đó.

Gọi số tổ hợp chập k của n phần tử là C(n,k) ta có :

$$A(n,k) = C(n,k) * k!$$

Suy ra

$$C(n,k) = \frac{n!}{k!.(n-k)!}$$

+ *Ví dụ* 1: Có n đội bóng thi đấu vòng tròn. Phải tổ chức bao nhiêu trận đấu bóng tất cả ?

Giải : Mỗi trận ứng với một tổ hợp chập 2 của n. Vậy có C(n,2) trận đấu.

+ *Ví dụ 2*:

Quay lại ví dụ ở mục trước. Trong trường hợp 3, mỗi bộ (x_1, x_2, x_3) là một tổ hợp chập 3 của 12. Vậy số bộ là

$$C(12, 3) = \frac{12!}{3!.(12-3)!} = \frac{12.11.10}{1.2.3} = 220$$

+ *Hệ quả* : Tích k số tự nhiên liên tiếp chia hết k!

Chứng minh. Vì C(n,k) = (n-k+1).(n-k+2).....n / k! là số nguyên.

3. CÁC KẾT HỢP NÂNG CAO

a) Hoán vị lặp

+ Ví dụ: Có 3 viên bi đỏ, 2 viên bi xanh và 4 viên bi trắng. Hỏi có bao nhiêu cách sắp các viên bi trên theo hàng ngang.

Ta có tất cả 9 chỗ trống để xếp các viên bi. Ta có C(9,3) khả năng xếp 3 viên bi đỏ, C(6,2) khả năng xếp 2 viên bi xanh, còn lại 1 khả năng xếp các viên bi trắng. Theo nguyên lý nhân ta có

$$C(9,3).C(6,2) = \frac{9!}{3!.6!}.\frac{6!}{2!.4!} = \frac{9!}{3!.2!.4!}$$

cách xếp.

+ *Định nghĩa*: *Hoán vị lặp* là hoán vị trong đó mỗi phần tử được ấn định một số lần lặp lại cho trước.

+ *Định lý*: Giả sử tập S có n phần tử, trong đó có n_1 phần tử kiểu 1, n_2 phần tử kiểu 2, ..., n_k phần tử kiểu k. Khi đó số các hoán vị n phần tử của S là

$$C_n(n_1, n_2, ..., n_k) = \frac{n!}{n_1!.n_2!...n_k!}$$

b) Tổ hợp lặp

+ *Ví dụ*: Giả sử ta có 3 đầu sách : Toán, Tin, Lý và mỗi đầu sách có ít nhất 6 bản photocopy. Hỏi có bao nhiêu cách chọn ra 6 quyển.

Giải: Bài toán đặt ra là chọn 6 phần tử, không kể thứ tự và cho phép lặp lại. Mỗi cách chọn được xác định duy nhất bởi số lượng của mỗi loại sách. Như vậy ta có thể biểu diễn mỗi cách chọn như sau

Toán Tin Lý
x x x | x x | x

trong đó 6 dấu x chỉ quyển sách chọn và 2 dấu | chỉ phân cách giữa các loại sách. Như vậy mỗi cách chọn tương đương với tổ hợp chập 2 (dấu |) từ 8 phần tử. Ta có số cách chọn là C(8,2) = 28.

+ *Định nghĩa*: *Tổ hợp lặp* chập k từ n phần tử là một nhóm không phân biệt thứ tự gồm k phần tử trích từ n phần tử đã cho, trong đó các phần tử có thể được lặp lại.

+ *Định lý*: Giả sử X có n phần tử. Khi đó số tổ hợp lặp chập k từ n phần tử của X là: C(k + n - 1, n - 1) = C(k + n - 1, k).

+ *Ví dụ*:

Quay lại ví dụ ở mục 1. Trong trường hợp 4, mỗi bộ (x_1, x_2, x_3) là một tổ hợp chập 3 của 12. Vậy số bộ là

C(3 + 12 − 1, 3) = C(14, 3) = 14.13.12 / 1.2.3 = 364

+ *Ví dụ*: Phương trình: $x_1 + x_2 + x_3 + x_4 = 10$

có bao nhiêu bộ nghiệm nguyên không âm ?

Giải : Mỗi bội nghiệm nguyên không âm của phương trình tương ứng 1-1 với một cách chọn 10 phần tử, trong đó phần tử kiểu i lặp lại x_i lần, i=1,…,4. Vậy số bộ nghiệm là số tổ hợp lặp chập 10 của 4. Vậy ta có số nghiệm là

C(10 + 4 -1 , 4 - 1) = C(13, 3) = 286

c) Tổ hợp lặp tổng quát

+ *Định nghĩa*: Tổ hợp lặp tổng quát chập *k* từ *n* phần tử là nhóm không phân biệt thứ tự gồm k phần tử trích từ n phần tử đã cho, trong đó phần tử thứ i lặp lại không quá k_i lần (i=1,…,n), với $k_1 + … + k_n \geq k$.

+ *Công thức*: Gọi Ω là tập hợp tất cả các tổ hợp lặp chập k từ n phần tử. Ta có

|Ω| = C(k + n − 1, k).

Ký hiệu A_i , i = 1, … n, là số tổ hợp lặp trong Ω có phần tử thứ i lặp lại hơn k_i lần.

Như vậy tập hợp tổ hợp lặp tổng quát là

$$\Omega \setminus \bigcup_{i=1}^{n} A_i$$

Suy ra số tổ hợp lặp tổng quát là

$$C_n^k(k_1,...,k_n) = C(k+n-1,k) + \sum_{m=1}^{n}\left[(-1)^m \sum_{1\le i_1\le...\le i_m\le n}\left|A_{i_1}\cap...\cap A_{i_m}\right|\right]$$

Mặt khác mỗi phần tử của $A_{i_1}\cap...\cap A_{i_m}$ sau khi loại $k_{i_1}+1$ phần tử thứ i_1, … , $k_{i_m}+1$ phần tử thứ i_m, là một tổ hợp lặp chập k− ($k_{i_1}+k_{i_2}+...+k_{i_m}$ + m).

Như vậy ta có

$$\left|A_{i_1}\cap...\cap A_{i_m}\right| = C(n-1+k-(k_{i_1}+...+k_{i_m}+m), n-1))$$

Suy ra

$$C_n^k(k_1,...,k_n) = C(k+n-1,k) + \sum_{m=1}^{n}\left[(-1)^m \sum_{1\le i_1\le ...\le i_m\le n} C(n-1+k-(k_{i_1}+...+k_{i_m}+m), n-1))\right]$$

+ *Ví dụ*: Cho 1 bi đỏ, 2 bi xanh và 3 bi vàng. Tính số tổ hợp chập 3 của các viên bi trên.

Mỗi tổ hợp là một tổ hợp lặp chập 3 của 3 phần tử bi đỏ, bi xanh và bi vàng, trong đó bi đỏ lặp không quá 1 lần, bi xanh lặp không quá 2 lần, bi vàng lặp không quá 3 lần. Vậy số tổ hợp là

$$C_3^3(1,2,3) = C(3-1+3,3-1) -$$
$$[C(3-1+(3-1-1),3-1) + C(3-1+(3-2-1),3-1) + C(3-1+(3-3-1),3-1)] + 0 =$$
$$C(5,2) - C(3,2) - C(2,2) - C(1,2) = 10-3-1 = 6$$

4. HỆ SỐ NHỊ THỨC

a) Các tính chất cơ bản

Với mọi n, k $\in$ **N**, k ≤ n.

(i) C(n,k) = C(n,n-k)
C(n,0) = C(n,n) = 1

(ii) Công thức tam giác Pascal
C(n,k) = C(n-1,k-1) + C(n-1,k)

(iii) Công thức giảm bậc
k.C(n,k) = n.C(n-1,k-1)

b) Nhị thức Newton

Với n $\in$ **N**, x, y $\in$ **C** ta có

$(x+y)^n = C(n,0).x^n + C(n,1).x^{n-1}.y + ...+ C(n,n-1).x.y^{n-1} + C(n,n).y^n$

+ Hệ quả nhị thức Newton:

(i) C(n,0) + C(n,1) + ... + C(n,n) = 2^n
(số các tập con của n phần tử là 2^n)

(ii) C(n,0) - C(n,1) + ... + $(-1)^n$C(n,n) = 0

(iii) $\sum_{j=0}^{\left[\frac{n}{2}\right]} C(n,2j) = \sum_{j=0}^{\left[\frac{n-1}{2}\right]} C(n,2j+1) = 2^{n-1}$
(số tập con chẵn bằng số tập con lẻ).

c) Công thức Vandermonde

Cho a,b,n $\in$ **N**. Ta có

$$\sum_{k=0}^{n} C(a,k).C(b,n-k) = C(a+b,n)$$

CM. Gọi E là tập có a+b phần tử, A, B $\subset$ E rời nhau, A có a phần tử và B có b phần tử. Khi đó mỗi tổ hợp chập n của các phần tử trong E là một kết hợp của một tổ hợp chập k của các phần tử trong A và tổ hợp chập n−k của các phần tử trong B. Từ đó suy ra công thức.

Áp dụng công thức cho a = b = n suy ra

• *Hệ quả*: Với n $\in$ **N** ta có

$$\sum_{k=0}^{n} C(n,k)^2 = C(2n,n)$$

CHƯƠNG I
SỰ KIỆN VÀ XÁC SUẤT

I. PHÉP THỬ VÀ SỰ KIỆN

1. Định nghĩa

• ***Phép thử*** là sự thực hiện một bộ điều kiện xác định, có thể là một thí nghiệm cụ thể, quan sát đo đạc hay thu thập dữ liệu về một hiện tượng nào đó.

• ***Sự kiện*** của phép thử là một kết cục xảy ra nào đó của phép thử. Một phép thử có thể có nhiều sự kiện.

+ *Ví dụ*

(i) Gieo một đồng tiền là phép thử. Hai sự kiện có thể xảy ra là xuất hiện mặt sấp, hoặc xuất hiện mặt ngửa.

(ii) Gieo một con xúc sắc là phép thử. Các kết cục sau là các sự kiện của phép thử:

- Xuất hiện mặt 1 chấm
- Xuất hiện mặt 2 chấm
- Xuất hiện mặt 3 chấm
- Xuất hiện mặt 4 chấm
- Xuất hiện mặt 5 chấm
- Xuất hiện mặt 6 chấm
- Xuất hiện mặt có số chấm lẻ
- Xuất hiện mặt có số chấm chẵn

(iii) Quan sát ghi nhận tuổi thọ của một chi tiết máy, hay của một loại bóng đèn, là một phép thử. Sự kiện của nó có thể là giá trị bất kỳ trong khoảng $[0,+\infty)$, hoặc một khoảng $(a,b) \subset [0,+\infty)$ nào đó mà tuổi thọ rơi vào.

• ***Sự kiện sơ cấp***

Trong các sự kiện ta thấy, có sự kiện là kết hợp của các sự kiện khác, chẳng hạn như sự kiện xuất hiện mặt có số chấm lẻ ở ví dụ (ii) là hợp của ba sự kiện xuất hiện mặt 1 chấm, mặt 3 chấm và mặt 5 chấm.

Những sự kiện không thể phân chia ra các sự kiện nhỏ hơn gọi là *sự kiện sơ cấp*, ví dụ như sự kiện xuất hiện mặt 1 chấm, 2 chấm,..., mặt 6 chấm ở ví dụ (ii).

• *Không gian các sự kiện sơ cấp* của phép thử là tập hợp tất cả các sự kiện sơ cấp của phép thử đó, thường ký hiệu là Ω.

+ *Ví dụ*. Không gian các sự kiện sơ cấp của phép thử gieo con xúc sắc là

$$\Omega = \{\omega_i \mid i = 1, 2, \ldots, 6\}$$

với ω_i , i=1,...,6, là sự kiện xuất hiện mặt có i chấm.

Bây giờ ta cho Ω là không gian sự kiện sơ cấp của phép thử α. Dễ thấy rằng, mỗi sự kiện A của phép thử α là tập con của Ω.

Các sự kiện của phép thử tạo thành không gian sự kiện, được định nghĩa chính xác như sau.

• *Không gian sự kiện*. Cho không gian sự kiện sơ cấp Ω của phép thử α. Cho **B** là σ−đại số trên Ω, tức **B** thoả

(i) $\Omega \subset$ **B**; (ii) $A \in$ **B** $\Rightarrow \overline{A} \in$ **B**; (iii) $A_i \in$ **B** $\Rightarrow \bigcup_i A_i \in$ **B**

Khi đó **B** gọi là một *không gian sự kiện* của phép thử α.

Sự kiện $\varnothing$ không bao giờ xảy ra, gọi là *sự kiện bất khả*.
Sự kiện Ω luôn xảy ra, gọi là *sự kiện tất yếu*.
Sự kiện ngẫu nhiên là sự kiện khác $\varnothing$ và Ω.

2. Quan hệ và phép tính sự kiện

Cho phép thử α với không gian sự kiện **B**, A, B ∈ **B.**

• Sự kiện A gọi là *sự kiện riêng* của sự kiện B, ký hiệu $A \subset B$, nếu sự kiện A xuất hiện kéo theo sự kiện B cũng xuất hiện.

• Sự kiện A gọi là *tương đương* sự kiện B, ký hiệu $A = B$, nếu $A \subset B$ và $B \subset A$.

• *Sự kiện tổng* của sự kiện A và sự kiện B, ký hiệu $A \cup B$, là sự kiện xảy ra khi và chỉ khi xảy ra sự kiện A hoặc sự kiện B.

• *Sự kiện tích* của sự kiện A và sự kiện B, ký hiệu $A \cap B$ hay A.B, là sự kiện xảy ra khi và chỉ khi xảy ra sự kiện A và sự kiện B.

Tương tự ta định nghĩa sự kiện tổng và sự kiện tích của nhiều sự kiện

$$\bigcup_{i=1}^{n} A_i \quad , \qquad \bigcap_{i=1}^{n} A_i$$

• *Sự kiện hiệu* của sự kiện A đối với sự kiện B, ký hiệu $A \setminus B$, là sự kiện xảy ra khi và chỉ khi xảy ra sự kiện A và không xảy ra sự kiện B.

• *Sự kiện đối lập* của sự kiện A là sự kiện $\overline{A} = \Omega \setminus A$.

• A và B gọi là *xung khắc* nếu $A \cap B = \varnothing$.

• Tập hợp các sự kiện $\{ A_1, \ldots, A_n \}$ gọi là *nhóm đầy đủ các sự kiện* nếu chúng xung khắc từng cặp một và tổng của chúng là sự kiện tất yếu Ω.

+ *Ví dụ*. Xét phép thử gieo con xúc xắc. Các sự kiện xuất hiện mặt i chấm ω_i, $i=1,\ldots,6$, tạo thành nhóm đầy đủ các sự kiện.

Nếu ta ký hiệu A là sự kiện xuất hiện mặt lẻ và B là sự kiện xuất hiện mặt chẵn thì {A, B} cũng là nhóm đầy đủ các sự kiện.

II. XÁC SUẤT

1. Khái niệm xác suất

Quan sát các sự kiện ngẫu nhiên ta thấy khả năng xuất hiện của chúng không giống nhau. Từ đó nảy sinh vấn đề đo lường khả năng xuất hiện của sự kiện ngẫu nhiên. Mỗi sự kiện A được gán một số không âm P(A) để đo khả năng xuất hiện được gọi là xác suất của sự kiện.

• *Định nghĩa*. Cho không gian sự kiện **B** của phép thử α. Ánh xạ P : **B** → [0; 1] gọi là xác suất trên **B**, nếu

(i) P(Ω) = 1

(ii) Với mọi tập sự kiện $\{A_i \mid i \in I\} \subset$ **B** , I là tập chỉ số hữu hạn hoặc vô hạn, từng đôi một xung khắc ta có

$$P\left(\bigcup_{i\in I} A_i\right) = \sum_{i\in I} P(A_i)$$

Khi đó (Ω, **B**, P) gọi là *không gian xác suất*.

• Trường hợp $\Omega = \{\omega_1, \ldots, \omega_n\}$ là không gian sự kiện rời rạc hữu hạn.

+ *Mệnh đề*.

Tập hợp tất cả tập con của Ω, ký hiệu **P**(Ω), là không gian sự kiện.

+ *Định lý*. Cho Ω là không gian sự kiện sơ cấp hữu hạn.

Cho P là xác suất trên **P**(Ω). Ký hiệu $p_i = P(\omega_i)$, ∀i=1,…,n. Khi đó ta có

(i) $p_i \geq 0, \forall i = 1, \ldots, n$

(ii) $\sum_{i=1}^{n} p_i = 1$

Ngược lại, Mọi tập $\{ p_i \mid i = 1, \ldots, n \}$ thoả (i), (ii) xác định một xác suất trên **P**(Ω) với

$$P(A) = \sum_{\omega_i \in A} p_i, \forall A \in P(\Omega)$$

+ *Ví dụ*: Cho $\Omega = \{ \omega_i \mid i = 1, \ldots, 6 \}$ là không gian sự kiện sơ cấp của phép thử gieo xúc xắc, trong đó ω_i là sự kiện xuất hiện mặt i. Khi đó tập $\{ p_i = 1/6 \mid i=1,\ldots,6\}$ xác định một xác suất trên **P**(Ω).

2. Các tính chất của xác suất

Cho không gian xác suất (Ω, **B**, P) của phép thử nào đó. Khi đó

(i) $P(\emptyset) = 0$

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

(iii) $A \subset B \Rightarrow P(B \setminus A) = P(B) - P(A)$

(iv) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

(v) $$P\left(\bigcup_{k=1}^{n} A_k\right) = \sum_{m=1}^{n} (-1)^{m-1} \sum_{1\le i_1\le i_2\le\ldots\le i_m\le n} P(A_{i_1} \cap A_{i_2} \cap\ldots\cap A_{i_m})$$

Chứng minh. Các tính chất (1)-(iv) suy ra từ định nghĩa. Tính chất (v) chứng minh bằng quy nạp.

3. Cách tính xác suất trong trường hợp đồng khả năng

a) *Trường hợp sự kiện sơ cấp hữu hạn*

• *Định nghĩa*. Cho không gian sự kiện sơ cấp $\Omega = \{\omega_1, \ldots, \omega_n\}$ của phép thử nào đó. Ta nói Ω *đồng khả năng* nếu các sự kiện sơ cấp có xác suất như nhau:

$$P(\omega_i) = P(\omega_j)\ ,\ \forall i, j = 1,\ldots, n$$

Khi đó, từ tính chất xác suất, ta có:

(i) $P(\omega_i) = 1/n$ $\quad\forall i = 1, \ldots, n$

(ii) $P(A) = |A|/n$ $\quad\forall A \subset \Omega$

trong đó |A| là lực lượng của tập A.

Nói cách khác

$$P(A) = \frac{\text{Số kết cục thuận lợi sự kiện A}}{\text{Số kết cục đồng khả năng}}$$

+ *Ví dụ*

(i) Gieo một xúc xắc hoàn toàn đối xứng. Tính xác suất

(a) xuất hiện mặt 6 chấm

(b) xuất hiện mặt bội của 3

Giải

Gọi A là sự kiện xuất hiện mặt 6 chấm, B là sự kiện xuất hiện mặt bội của 3.

Số sự kiện sơ cấp đồng khả năng là 6, số kết cục thuận lợi sự kiện A là 1, số kết cục thuận lợi sự kiện B là 2. Vậy

$$P(A) = 1/6 \text{ và } P(B) = 2/6 = 1/3$$

(ii) Trong thùng có a quả cầu trắng, b quả cầu đen giống hệt nhau. Lấy ngẫu nhiên n quả ($n \le a + b$). Tính xác suất rút được k quả cầu trắng.

Giải.

Mỗi kết cục của phép thử (rút n quả cầu) là một tổ hợp chập n của (a+b) phần tử. Như vậy số kết cục đồng khả năng là C(a+b, n).

Gọi A_k là sự kiện rút được k quả cầu trắng. Như vậy những kết cục rút được k quả cầu trắng và n − k quả cầu đen là thuận lợi cho sự kiện A_k. Số kết cục này là C(a,k)*C(b,n-k).

Vậy xác suất của sự kiện rút k quả cầu trắng là

$$P(A_k) = \frac{C(a,k).C(b,n-k)}{C(a+b,n)}$$

Từ tính chất $\sum_{k=0}^{n} P(A_k) = 1$ suy ra

↳ Hệ quả: công thức *de Vandermonde*

$$\sum_{k=0}^{n} C(a,k).C(b,n-k) = C(a+b,n)$$

(iii) Giả thiết giống ví dụ (ii). Tính xác suất rút được ít nhất 1 quả cầu trắng.

Giải.

Gọi A là sự kiện rút được ít nhất 1 quả cầu trắng. Khi đó sự kiện bù của A, tức $\overline{A}$, là sự kiện cả n quả cầu được rút đều đen. Số kết cục thuận lợi $\overline{A}$ là C(b, n), nên xác suất sự kiện $\overline{A}$ là C(b,n)/C(a+b,n).

Suy ra xác suất sự kiện A là

P(A) = 1 − P($\overline{A}$) = 1 − C(b,n)/C(a+b,n).

b) *Trường hợp sự kiện sơ cấp vô hạn*

• *Định nghĩa*. Cho không gian xác suất (Ω, **B**, P), |Ω| = ∞. Giả thiết m:**B**→R^+ là ánh xạ định nghĩa trên **B** thoả

(i) 0 < m(Ω) < ∞

(ii) Với mọi tập sự kiện {A_i | i ∈ I } ⊂ **B** , I là tập chỉ số hữu hạn hoặc vô hạn, từng đôi một xung khắc ta có

$$m\left(\bigcup_{i\in I} A_i\right) = \sum_{i\in I} m(A_i)$$

Ánh xạ m gọi là *độ đo* trên **B**.

Ta nói Ω *đồng khả năng* nếu xác suất của mỗi sự kiện trong **B** tỉ lệ với độ đo của nó.

Khi đó, từ tính chất xác suất, ta có:

P(A) = m(A)/m(Ω) ∀A ⊂ Ω

Nói cách khác

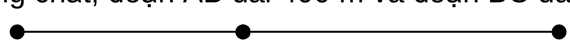
$$P(A) = \frac{\text{Độ đo miền kết cục thuận lợi sự kiện A}}{\text{Độ đo miền kết cục đồng khả năng}}$$

• *Độ đo hình học*.

- Độ đo của đoạn thẳng hay đường cong là độ dài.
- Độ đo của miền phẳng hay miền cong là diện tích.
- Độ đo của vật thể là thể tích.

+ *Ví dụ*:

(i) Đường dây điện thoại ngầm nối ba trạm A, B và C (xem hình). Bỗng nhiên liên lạc giữa A và C bị ngắt do đứt dây. Hãy tính xác suất dây đứt trong đoạn dây từ A đến B. Biết rằng dây đồng chất, đoạn AB dài 400 m và đoạn BC dài 600m.



A 400m B 600m

C

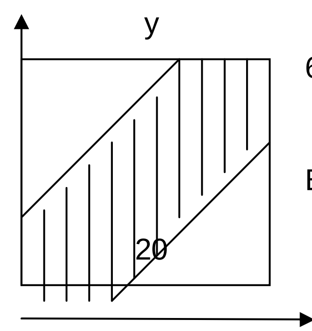
Giải.

Rõ ràng khả năng dây đứt tại mỗi điểm bất kỳ là như nhau. Như vậy xác suất dây đứt trong một đoạn tỉ lệ với độ dài của đoạn dây đó. Suy ra xác suất dây đứt trong đoạn AC là

$$400/(400 + 600) = 0.4$$

(ii) Hai người A và B hẹn gặp nhau tại một địa điểm xác định trong khoảng từ 0 đến 1 giờ. Người đến trước sẽ chờ tối đa 20 phút, nếu người kia chưa đến thì sẽ bỏ đi. Tính xác suất họ gặp nhau. Biết rằng mỗi người có thể đến chỗ hẹn vào thời điểm bất kỳ trong khoảng thời gian trên với khả năng như nhau.
Giải.

Gọi x là thời điểm đến chỗ hẹn của A và y là thời điểm đến chỗ hẹn của B (tính ra phút). Mọi kết cục đồng khả năng là một cặp (x,y), $0 \le x,y \le 60$. Tập không gian sự kiện sơ cấp Ω sẽ là hình vuông



$$\Omega = \{(x,y) \mid 0 \le x, y \le 60\}$$

Miền kết cục thuận lợi cho hai người gặp nhau là phần hình vuông chắn giữa hai đường thẳng y=x+20 và y=x−20 (xem hình bên)

$$B = \{(x,y) \mid 0 \le x, y \le 60 \text{ và } |x-y| \le 20\}$$

0 20 60 x

Suy ra xác suất hai người gặp nhau là diện tích B chia cho diện tích Ω, tức là

$$(60^2 - 40^2)/60^2 = 5/9$$

4. Tần suất

Trong thực tế có những phép thử không có số sự kiện sơ cấp đồng khả năng. Chẳng hạn như phép thử bắn một viên đạn vào bia. Khi đó sự kiện bắn trúng bia và sự kiện bắn không trúng bia không thể coi là đồng khả năng.

Trong những trường hợp như thế này người ta sử dụng khái niệm *tần suất*.

• *Định nghĩa*. Cho A là sự kiện của phép thử. Giả sử phép thử được lặp lại n lần và sự kiện A xuất hiện m lần. Tỉ số m/n gọi là tần suất xuất hiện sự kiện A trong loạt n phép thử. Người ta đã chứng minh

$$\frac{m}{n} \xrightarrow[n\to\infty]{} P(A)$$

Vì vậy trong thực tế , khi n đủ lớn, người ta coi P(A) = m/n.

+ *Ví dụ*. Nhà toán học Laplace đã thống kê tần suất sinh con trai ở các thành phố lớn châu Âu là 22/43 = 0.512.

III. XÁC SUẤT CÓ ĐIỀU KIỆN

1. Khái niệm xác suất có điều kiện

Cho không gian xác suất (Ω, **B**, P), B $\in$ **B** có P(B) > 0. Với mỗi sự kiện A $\in$ **B**, xác suất để A xuất hiện với giả thiết sự kiện B xảy ra gọi là *xác suất có điều kiện* của sự kiện A với điều kiện B.

\+ *Ví dụ*. Người ta điều tra các gia đình có hai con thì thấy tỉ lệ sinh con trai con gái bằng nhau. Vì vậy 4 khả năng sau có xác suất bằng ¼:

(T,T), (T,G), (G,T), (G,G)

trong đó T ký hiệu con trai, G ký hiệu con gái và trong các cặp trên anh hoặc chi đứng trước, em đứng sau.

Giả sử người ta gõ cửa một nhà có hai con, và có bé gái ra mở cửa (tức gia đình có bé gái). Hãy tính xác suất để đứa bé còn lại là con trai.

Không gian sự kiện sơ cấp là

$$\Omega = \{ (T,T), (T,G), (G,T), (G,G) \}$$

Gọi A là sự kiện gia đình có 1 trai và 1 gái. Ta có P(A) = ½.

Nhưng với điều kiện gia đình có bé gái thì các sự kiện sơ cấp đồng khả năng là

$$\Omega' = \{ (T,G), (G,T), (G,G) \}$$

và có 2 kết cục thuận lợi cho A là (T,G) và (G,T). Vậy với điều kiện gia đình có bé gái thì xác suất cần tìm của A là P'(A) = 2/3.

Bây giờ ta ký hiệu B là sự kiện gia đình có con gái. Ta nói P'(A) là xác suất có điều kiện của A đối với B và ký hiệu là P(A/B).

Ký hiệu
n là số kết cục đồng khả năng
n_X là số kết cục thuận lợi sự kiện X, X $\in$ **B**.

Ta có

$$P(A/B) = \frac{n_{A.B}}{n_B} = \frac{n_{A.B} : n}{n_B : n} = \frac{P(A.B)}{P(B)}$$

Từ đó ta đi đến định nghĩa sau

• *Định nghĩa*. Cho không gian xác suất (Ω, **B**, P), B $\in$ **B** có P(B) > 0. Với mỗi sự kiện A $\in$ **B**, ta định nghĩa *xác suất có điều kiện* của sự kiện A với điều kiện B là đại lượng

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

Ta dễ dàng thấy rằng định nghĩa này thoả mãn các tiên đề xác suất và ánh xạ. Vậy

$P(\bullet, B) : A \rightarrow P(A/B)$ cũng là xác suất trên không gian sự kiện **B.**

Bằng qui nạp ta dễ dàng chứng minh định lý sau

• *Định lý nhân xác suất*. Cho các sự kiện $A_1,\ldots, A_n$. Khi đó

$$P(A_1.A_2.\ldots.A_n) = P(A_1).P(A_2/A_1).P(A_3/A_1.A_2)\ldots.P(A_n/A_1\ldots.A_{n-1})$$

2. Sự kiện độc lập

Cho không gian xác suất $(\Omega$, **B**, P), A, B $\in$ **B**.

• Sự kiện A gọi là *độc lập* với sự kiện B nếu kết quả của sự kiện B không ảnh hưởng đến xác suất của sự kiện A.

Hiển nhiên là nếu P(A)=0 hoặc P(B)=0 thì sự kiện A độc lập với B và B cũng độc lập với A.

Nếu $P(A) \neq 0$ và $P(B) \neq 0$ thì, theo định nghĩa,

A độc lập với B $\Leftrightarrow$ $P(A/B) = P(A)$ $\Leftrightarrow$ $P(A.B) = P(A).P(B)$

và

B độc lập với A $\Leftrightarrow$ $P(B/A) = P(B)$ $\Leftrightarrow$ $P(A.B) = P(A).P(B)$

Kết hợp các kết quả trên ta có

Tính độc lập có tính chất tương hỗ, tức là A độc lập với B thì B cũng độc lập với A và

A và B *độc lập* với nhau $\Leftrightarrow P(A.B) = P(A).P(B)$

Bây giờ ta cho các sự kiện $A_i \in$ **B**, i=1,…,n.

• Các sự kiện $A_1,\ldots, A_n$ gọi là *độc lập từng đôi*, nếu

$$\forall\, i,j \in \{1,\ldots,n\},\ i\neq j \Rightarrow A_i, A_j \text{ độc lập}$$

• Các sự kiện $A_1,\ldots, A_n$ gọi là *độc lập tương hỗ*, nếu mỗi sự kiện A_k, k=1,…,n, độc lập với tích nhóm bất kỳ các sự kiện còn lại.

Từ định lý nhân xác suất suy ra

• Định lý. Các sự kiện $A_1,\ldots, A_n$ độc lập tương hỗ khi và chỉ khi

$$\forall\, I \subset \{1,\ldots,n\},\ P\left(\bigcap_{i\in I} A_i\right) = \prod_{i\in I} P(A_i)$$

+ Ghi chú. Với n>2, khái niệm độc lập tương hỗ mạnh hơn độc lập từng đôi.

Xét ví dụ điều tra gia đình hai con ở trên. Gọi

A là sự kiện gia đình có 1 trai, 1 gái.

B là sự kiện gia đình có con gái đầu

C là sự kiện gia đình có con trai thứ

Ta có: P(A) = P(B) = P(C) = ½

Và: P(A.B) = P(B.C) = P(C.A) = ¼

Vậy A, B, C độc lập từng đôi.

Nhưng A, B, C không độc lập tương hỗ vì

$$P(A.B.C) = 1/4 \neq P(A).P(B).P(C) = 1/8$$

• *Định lý*. Nếu A, B độc lập thì các cặp sự kiện $(A, \overline{B})$, $(\overline{A}, B)$ và $(\overline{A}, \overline{B})$ cũng độc lập.

Chứng minh. Ta có

$P(A, \overline{B}) = P(A \setminus A.B) = P(A) - P(A.B)$

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

$\Rightarrow$ A và $\overline{B}$ độc lập.

Đối với các cặp còn lại cũng chứng minh tương tự.

Từ định lý trên và bằng qui nạp suy ra

• *Định lý*. Cho $A_1, \ldots, A_n$ là các sự kiện độc lập tương hỗ. Khi đó các sự kiện $B_1, \ldots, B_n$, trong đó B_i là A_i hoặc $\overline{A_i}$ $\forall i=1,\ldots,n$, cũng độc lập tương hỗ

+ *Ví dụ*.

(i) Một thùng đựng *n* sản phẩm, trong đó có *m* phế phẩm (m<n). Rút ngẫu nhiên 1 sản phẩm, sau đó rút tiếp 1 sản phẩm nữa (sản phẩm rút lần đầu không bỏ lại vào thùng). Tính xác suất sản phẩm rút đầu là phế phẩm và sản phẩm rút sau là chính phẩm.

Giải. Gọi

A là sự kiện sản phẩm rút đầu là phế phẩm.

B là sự kiện sản phẩm rút sau là chính phẩm.

Khi đó xác suất cần tìm là

$$P(A.B) = P(A).P(B/A) = \frac{m}{n}.\frac{n-m}{n-1}$$

(ii) Một công nhân đứng 3 máy, các máy hoạt động độc lập với nhau. Xác suất để trong thời gian T máy 1, 2, 3 không bị hỏng tương ứng là 0.9 ; 0.8 ; 0.7. Tính xác suất có ít nhất 1 máy bị hỏng trong thời gian T.

Giải. Gọi

A là sự kiện máy 1 không bị hỏng trong thời gian T.

B là sự kiện máy 2 không bị hỏng trong thời gian T.

C là sự kiện máy 3 không bị hỏng trong thời gian T.

Xác suất để cả ba máy không bị hỏng trong thời gian T là

$$P(A.B.C) = P(A).P(B).P(C) = 0.9 * 0.8 * 0.7 = 0.504$$

Sự kiện có ít nhất 1 máy hỏng đối lập với sự kiện A.B.C. Vậy xác suất cần tìm là:

$$1 - P(A.B.C) = 1 - 0.504 = 0.496$$

IV. CÔNG THỨC XÁC SUẤT TOÀN PHẦN và CÔNG THỨC BAYES

1. Công thức xác suất toàn phần

Cho không gian xác suất $(\Omega, \mathbf{B}, P)$. Giả sử $A_1,\ldots, A_n$ là nhóm đầy đủ sự kiện và B là sự kiện bất kỳ trong **B**.

Bài toán đặt ra là: biết xác suất $P(A_i)$ và $P(B/A_i)$, i=1,…,n, tính xác suất P(B).

Ta có

$$B = B.(A_1 \cup A_2 \cup \ldots \cup A_n) = B.A_1 \cup B.A_2 \cup \ldots \cup B.A_n$$

Vì các sự kiện $A_1, \ldots, A_n$ xung khắc từng đôi nên các sự kiện $B.A_1, \ldots, B.A_n$ cũng xung khắc từng đôi. Suy ra

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

Mặt khác ta có

$$P(B.A_i) = P(A_i).P(B/A_i) \quad \forall\, i = 1,\ldots, n$$

Từ đó ta có

• *Công thức xác suất toàn phần*

$$P(B) = \sum_{i=1}^{n} P(A_i).P(B/A_i)$$

+ *Ví dụ.*

(i) Một nhà máy sản xuất bóng đèn gồm ba phân xưởng. Phân xưởng 1 sản xuất 50%, phân xưởng 2 sản xuất 20% và phân xưởng 3 sản xuất 30% số bóng đèn. Tỷ lệ phế phẩm của phân xưởng 1, 2 và 3 tương ứng là 2%, 3% và 4%. Hãy tính tỷ lệ phế phẩm chung của nhà máy.

Giải.

Tỷ lệ phế phẩm chung của nhà máy là xác suất một bóng đèn chọn ngẫu nhiên là phế phẩm. Ta gọi

A_i là sự kiện bóng đèn chọn ra thuộc phân xưởng i, i=1,2,3.

B là sự kiện bóng đèn chọn ra là phế phẩm.

Theo công thức xác suất toàn phần ta có

$$\begin{aligned} P(B) &= P(A_1).P(B/A_1) + P(A_2).P(B/A_2) + P(A_3).P(B/A_3) \\ &= 50\%.2\% + 20\%.3\% + 30\%.4\% \\ &= 1\% + 0.6\% + 1.2\% = 2.8\% \end{aligned}$$

(ii) Biết xác suất trong khoảng thời gian t có k cuộc gọi đến tổng đài là $p_t(k)$. Giả thiết rằng số các cuộc gọi trong các khoảng thời gian rời nhau là các sự kiện độc lập. Hãy tính xác suất $p_{2t}(n)$.

Giải.

Gọi

B là sự kiện có n cuộc gọi trong khoảng thời gian 2t

A_k, k=0, 1, …,n, là sự kiện có k cuộc gọi trong nửa đầu thời gian t.

Ta có

$$P(A_k) = p_t(k) \text{ và } \quad P(B/A_k) = p_t(n-k) \quad \forall k=0,1,\ldots, n$$

Theo công thức xác suất toàn phần suy ra

$$P(B) = p_{2t}(n) = \sum_{k=0}^{n} p_t(k).p_t(n-k)$$

2. Công thức Bayes.

Cho không gian xác suất (Ω, **B**, P). Giả sử $A_1, \ldots, A_n$ là nhóm đầy đủ sự kiện và B là sự kiện bất kỳ trong **B**. Biết xác suất $P(A_i)$ và $P(B/A_i)$ $\forall i=1,\ldots,n$.

Giả thiết phép thử được thực hiện và sự kiện B xảy ra. Hãy tính xác suất $P(A_i/B)$ $\forall i=1,\ldots,n$.

Từ công thức nhân xác suất

$$P(A_i.B) = P(A_i).P(B/A_i) = P(B).P(A_i/B) \qquad \forall i=1,\ldots,n$$

suy ra

$$P(A_i/B) = \frac{P(A_i).P(B/A_i)}{P(B)} \qquad \forall i=1,\ldots,n$$

Thế P(B) theo công thức xác suất toàn phần ta được

• *Công thức Bayes*

$$P(A_i/B) = \frac{P(A_i).P(B/A_i)}{\sum_{k=1}^{n} P(A_k).P(B/A_k)} \qquad \forall i=1,\ldots,n$$

Xác suất $P(A_i/B)$ gọi là *xác suất hậu nghiệm*, còn xác suất $P(A_i)$ gọi là *xác suất tiên nghiệm*.

+ *Ví dụ.*

(i) Một thiết bị gồm ba loại linh kiện, loại 1 chiếm 35%, loại 2 chiếm 25%, loại 3 chiếm 40% tổng số linh kiện của thiết bị. Xác suất hư hỏng sau khoảng thời gian hoạt động T của từng loại tương ứng là 15%, 25% và 5%.

Thiết bị đang hoạt động bỗng bị hỏng. Hãy tính xác suất hư hỏng của tứng loại linh kiện.

Giải.

Gọi

B là sự kiện thiết bị bị hỏng

A_k là sự kiện linh kiện hỏng thuộc loại k, k=1,2,3.

Theo công thức xác suất toàn phần, ta có

$$\begin{aligned} P(B) &= P(A_1).P(B/A_1) + P(A_2).P(B/A_2) + P(A_3).P(B/A_3) \\ &= 35\%.15\% + 25\%.25\% + 40\%.5\% \\ &= 5.25\% + 6.25\% + 2\% = 13.5\% \end{aligned}$$

$$\Rightarrow \quad \begin{aligned} P(A_1/B) &= 5.25\% / 13.5\% = 7/18 \\ P(A_2/B) &= 6.25\% / 13.5\% = 25/54 \\ P(A_3/B) &= 2\% / 13.5\% = 4/27 \end{aligned}$$

(ii) Giả sử một hệ thống truyền tin có 1 máy phát và 1 máy thu.

Tại máy phát có thể xảy ra hai sự kiện: phát tín hiệu (sự kiện B) và không phát tín hiệu (sự kiện $\overline{B}$).

Tại máy thu có thể xảy ra hai sự kiện: nhận tín hiệu (sự kiện A) và không nhận tín hiệu (sự kiện $\overline{A}$).

Vì ảnh hưởng nhiễu nên có thể xảy ra hiện tượng máy thu không nhận được tín hiệu của máy phát, hoặc ngược lại máy phát không phát tín hiệu nhưng máy thu vẫn nhận tín hiệu giả do tạp âm.

Giả sử biết xác suất P(B), P(A/B) và P(A/$\overline{B}$). Để xác định độ tin cậy của hệ thống, hãy tính xác suất P(B/A) và P($\overline{B}/\overline{A}$).

Giải.

Theo công thức Bayes ta có

$$P(B/A) = \frac{P(B).P(A/B)}{P(B).P(A/B)+P\left(\overline{B}\right).P\left(A/\overline{B}\right)}$$

$$P(\overline{B}/\overline{A}) = \frac{P\left(\overline{B}\right).P\left(\overline{A}/\overline{B}\right)}{P\left(\overline{B}\right).P\left(\overline{A}/\overline{B}\right)+P(B).P\left(\overline{A}/B\right)}$$

trong đó các xác suất P($\overline{B}$), P($\overline{A}$/B), P($\overline{A}/\overline{B}$) tính như sau

P($\overline{B}$) = 1 − P(B), P($\overline{A}$/B) = 1 − P(A/B), P($\overline{A}/\overline{B}$) = 1 − P(A/$\overline{B}$)

Chẳng hạn, cho P(B) = 5/8, P(A/B) = 3/5 và P(A/$\overline{B}$) = 1/3 ta có:

P($\overline{B}$) = 1 − P(B) = 3/8, P($\overline{A}$/B) = 1 − P(A/B) = 2/5, P($\overline{A}/\overline{B}$) = 1 − P(A/$\overline{B}$) = 2/3

Suy ra

$$P(B/A) = \frac{\frac{5}{8}\cdot\frac{3}{5}}{\frac{5}{8}\cdot\frac{3}{5}+\frac{3}{8}\cdot\frac{1}{3}} = \frac{\frac{3}{8}}{\frac{4}{8}} = \frac{3}{4} \quad \& \quad P(\overline{B}/\overline{A}) = \frac{\frac{3}{8}\cdot\frac{2}{3}}{\frac{3}{8}\cdot\frac{2}{3}+\frac{5}{8}\cdot\frac{2}{5}} = \frac{\frac{2}{8}}{\frac{4}{8}} = \frac{1}{4}$$

CHƯƠNG 2
BIẾN NGẪU NHIÊN

I. KHÁI NIỆM BIẾN NGẪU NHIÊN

1. Ví dụ

(i) Một xạ thủ bắn ba phát đạn vào bia với xác suất trúng bia mỗi phát là p. Gọi X là tần suất bắn trúng bia. X có thể nhận các giá trị 0, 1/3, 2/3, 1.

Ta thấy X có hai tính chất: biến thiên và ngẫu nhiên. Tính biến thiên là hiển nhiên vì X có thể nhận các giá trị khác nhau. Tính ngẫu nhiên thể hiện ở chỗ giá trị của X phụ thuộc vào kết cục của phép thử bắn ba phát đạn.

Phép thử có các kết cục sơ cấp sau:

$$\omega_1 = (0, 0, 0); \omega_2 = (0, 0, 1); \omega_3 = (0, 1, 0); \omega_4 = (0, 1, 1)$$

$$\omega_5 = (1, 0, 0); \omega_6 = (1, 0, 1); \omega_7 = (1, 1, 0); \omega_8 = (1, 1, 1)$$

trong đó bộ ba (b_1, b_2, b_3) với $b_i = 0$ hoặc 1 tương ứng với lần bắn thứ i trượt hoặc trúng đích. Chẳng hạn, nếu ω_2 xảy ra thì X = 1/3. Như vậy có thể coi X là hàm từ $\Omega = \{\omega_i \mid i=1,\ldots, 8\}$ vào **R**.

(ii) Tuổi thọ trung bình của chi tiết là a. Tuy nhiên thời gian làm việc của các chi tiết không giống nhau. Ký hiệu X là độ lệch của tuổi thọ chi tiết so với tuổi thọ trung bình:

$$X = | t - a |.$$

trong đó t là tuổi thọ chi tiết.

Như vậy X cũng là đại lượng ngẫu nhiên. Ở đây không gian sự kiện sơ cấp là $\Omega = [0; +\infty)$.

2. Định nghĩa.

Cho không gian xác suất (Ω, **B**, P). Ánh xạ

$$X : \Omega \rightarrow \mathbf{R}$$

gọi là ***biến ngẫu nhiên***, nếu

$$\forall x \in \mathbf{R}, \{\omega \in \Omega \mid X(\omega) \leq x\} \in \mathbf{B}.$$

• Biến ngẫu nhiên X gọi là *biến ngẫu nhiên rời rạc*, nếu $X(\Omega)$ là hữu hạn hoặc vô hạn đếm được.

Chẳng hạn, biến ngẫu nhiên ở ví dụ (i) là biến ngẫu nhiên rời rạc.

• Biến ngẫu nhiên X gọi là *biến ngẫu nhiên liên tục*, nếu $X(\Omega)$ lấp đầy một khoảng nào đấy trên trục số.

Chẳng hạn, biến ngẫu nhiên ở ví dụ (ii) là biến ngẫu nhiên liên tục.

II. BIẾN NGẪU NHIÊN RỜI RẠC

1. Luật phân phối xác suất rời rạc

• *Định nghĩa*. Cho không gian xác suất (Ω, **B**, P) và biến ngẫu nhiên rời rạc X. Giả sử tập các giá trị của X là

$$X(\Omega) = \{ x_i \mid i \in I \}$$

trong đó I là tập chỉ số hữu hạn hoặc vô hạn đếm được.

Ta gọi *luật phân phối xác suất* của X là tập

$$\{ p_k \mid k \in I \}$$

trong đó p_k , $k \in I$, là xác suất sự kiện $X = x_k$:

$$p_k = P(X = x_k) = P(\omega \in \Omega \mid X(\omega) = x_k)$$

Luật phân phối xác suất của biến ngẫu nhiên rời rạc X được biểu diễn dạng dãy

$$X = \{ (x_i, p_i) \mid i \in I \}$$

hoặc dạng bảng

X	x_1	x_2	. . .	x_k	. . .
P	p_1	p_2	. . .	p_k	. . .

Vì x_k, $k \in I$, khác nhau từng đôi nên các sự kiện $X = x_k$, $k \in I$ xung khắc từng đôi. Suy ra

$$\sum_{k\in I} p_k = P\left(\bigcup_{k\in I} \{\omega\in \Omega \mid X(\omega) = x_k\}\right) = P(\Omega) = 1$$

+ *Ví dụ*.
(i) ***Luật phân phối xác suất nhị thức B(n,p)***. Giả sử sự kiện A của phép rhử α có xác suất là p. Thực hiện phép thử α n ($n\in$ **N**) lần một cách độc lập. Gọi X là biến ngẫu nhiên có giá trị là tần số xuất hiện sự kiện A. Hãy tìm luật xác suất của X.
Giải.

X có thể nhận các giá trị 0, 1, 2, … , n. Để tính xác suất $p_k = P(X=k)$ ta làm như sau:

Ký hiệu A_i là sự kiện A ở phép thử lần thứ i , i=1, …, n.

Để X=k thì trong n lần thực hiện phép thử α, phải có đúng k lần sự kiện A xảy ra.

Như vậy , để X=k thì trong dãy sự kiện (A_1, … , A_n) có đúng k sự kiện xảy ra và n−k sự kiện không xảy ra. Mỗi khả năng như vậy có xác suất là

$$p^k.(1-p)^{n-k}$$

Mặt khác mỗi khả năng trên ứng với mọt tổ hợp chập k của n phần tử. Vậy ta nhận được công thức Bernoulli

$$p_k = C(n,k).p^k.(1-p)^{n-k}, \qquad \forall k=0, 1, \ldots, n$$

Luật phân phối xác suất

$$\{(k, C(n,k).p^k.(1-p)^{n-k} \mid k=0, 1, \ldots, n\}$$

gọi là *luật phân phối xác suất nhị thức B(n,p)*.

Xác suất để số lần xuất hiện A nằm trong khoảng từ k_1 đến k_2 là

$$P_n(k_1, k_2) = \sum_{k=k_1}^{k_2} C(n,k).p^k(1-p)^{n-k}$$

(ii) Có n xạ thủ bắn bia. Xác suất bắn trúng đích của xạ thủ k là p_k , k=1,…, n. Gọi X là số xạ thủ bắn trúng đích. Tìm luật phân phối xác suất của X.
Giải.

Lập luận tương tự như trên ta thấy có C(n,k) khả năng k xạ thủ bắn trúng đích và n−k xạ thủ bắn trượt. Tuy nhiên xác suất mỗi khả năng không giống nhau. Vì vậy xác suất P(X=k) là tổng của C(n,k) số hạng, mỗi số hạng là tích của k xác suất p_i và n−k xác suất $q_j = 1-p_j$ dạng

$$p_{i_1}.p_{i_2}.....p_{i_k}.q_{j_1}.q_{j_2}.....q_{j_{n-k}}$$

Như vậy

$$P(X=k) = \sum_{1\le i_1<i_2<\ldots<i_k\le n} \prod_{h=1}^{k} p_{i_h} \prod_{j\notin\{i_1,\ldots,i_k\}} (1-p_j)$$

trong đó mỗi bộ $\{i_1, \ldots , i_k\}$ là tổ hợp chập k của n phần tử {1, 2, … , n}.

Chẳng hạn, với

$$n = 4,\ p_1 = 0.1 ;\ p_2 = 0.2 ;\ p_3 = 0.3 ;\ p_4 = 0.4$$

ta có

P(X=2) = 0.1 x 0.2 x 0.7 x 0.6 + 0.1 x 0.8 x 0.3 x 0.6 + 0.1 x 0.8 x 0.7 x 0.4
+ 0.9 x 0.2 x 0.3 x 0.6 + 0.9 x 0.2 x 0.7 x 0.4 + 0.9 x 0.8 x 0.3 x 0.4
= 0.2144

Ta có thể tính xác suất bằng *hàm sinh*:

$$\varphi(z) = \prod_{i=1}^{n}(q_i + p_i.z) = a_n z^n + a_{n-1}z^{n-1} + \ldots + a_k z^k + \ldots + a_1 z + a_0$$

Khi đó

$$P(X=k) = a_k ,\ \forall k=0, 1, \ldots, n$$

Quay lại số liệu trên ta có

$$\varphi(z) = (0.9 + 0.1z)(0.8 + 0.2z)(0.7 + 0.3z)(0.6 + 0.4z) =$$
$$= 0.302 + 0.440z + 0.215z^2 + 0.040z^3 + 0.002z^4$$

Suy ra

$$P(X=0) = 0.302 ;\ P(X=1) = 0.440 ;\ P(X=2) = 0.215 ;$$
$$P(X=3) = 0.040 ;\ P(X=4) = 0.002$$

(iii) Giả sử một phép thử có m sự kiện $A_1, \ldots , A_m$ hợp thành nhóm đầy đủ các sự kiện. Xác suất sự kiện A_i là p_i, i=1,…,m, không phụ thuộc vào lần thử. Phép thử được thực hiện n lần. Tính xác suất A_i xuất hiện k_i lần, i=1,…,m, trong đó

$$k_1 + \ldots + k_m = n$$

Giải.

Mỗi kết cục: A_i xuất hiện k_i lần, i=1,…,m, là một hoán vị lap và có xác suất là

$$p_1^{k_1}.p_2^{k_2}.....p_m^{k_m}$$

Vậy xác suất cần tìm là

$$P_n(k_1, \ldots , k_m) = \frac{n!}{k_1!.k_2!.....k_m!} p_1^{k_1}.p_2^{k_2}.....p_m^{k_m}$$

(iv) Một hộp kín có a quả cầu trắng và b quả cầu đen. Lấy ngẫu nhiên n quả cầu ($n \le a+b$). Gọi X là số quả cầu trắng được lấy ra. Tìm luật phân phối xác suất của X.

Giải.

Theo ví dụ chương trước,

$$p_k = P(X = k) = \frac{C(a,k).C(b,n-k)}{C(a+b,n)} \qquad , \forall k=0,1,\ldots,n.$$

Đây là luật phân phối *xác suất siêu hình học* H(a+b, n, a/(a+b)).

(v) Trong một số ứng dụng kỹ thuật, ta thường gặp biến ngẫu nhiên rời rạc vô hạn có luật xác suất

$$\{(\mathrm{k}, \mathrm{e}^{-\lambda}.\frac{\lambda^k}{k!} \mid k = 0,1,2,....\}$$

trong đó $\lambda > 0$ là hằng số.

Luật này gọi là *luật phân phối xác suất Poisson* , ký hiệu là $P(\lambda)$.

• *Biến ngẫu nhiên độc lập*. Hai biến ngẫu nhiên X và Y trên cùng không gian xác suất gọi là độc lập, nếu luật phân phối của X không phụ thuộc giá trị của Y và ngược lại.

2. Hàm phân phối xác suất

• *Định nghĩa*. Cho biến ngẫu nhiên X (rời rạc hoặc liên tục) trên không gian xác suất (Ω, **B**, P). *Hàm phân phối xác suất* của X là hàm

$$F(x) = P(X \le x), \forall x \in R$$

Ta xét trường hợp biến ngẫu nhiên rời rạc.

• Cho biến ngẫu nhiên rời rạc X có luật xác suất

$$\{(x_i , p_i) \mid i=1, \dots , n\}$$

Từ định nghĩa suy ra hàm phân phối F(x) được xác định theo công thức sau

$$F(x) = P(X \le x) = \sum_{x_i \le x} p_i$$

Như vậy hàm phân phối là hàm gián đoạn bậc thang có bước nhảy p_k tại $x=x_k$, k=1,…,n, và là hàm liên tục phải.

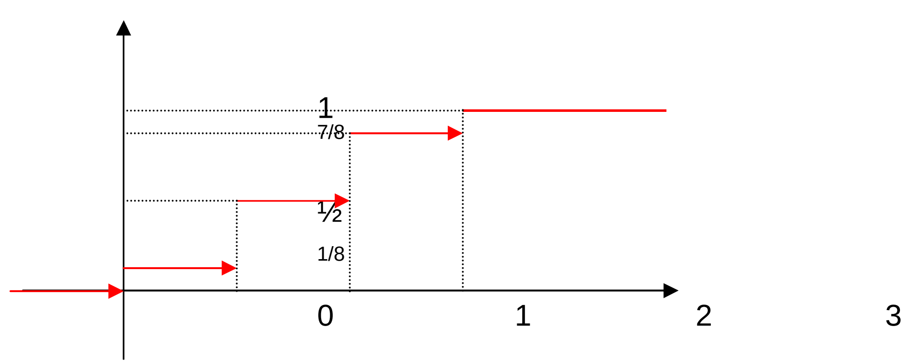
+ *Ví dụ*: Xét luật nhị thức B(3, ½). Ta có

$$p_0 = 1/8 , \ p_1 = p_2 = 3/8 , \ p_3 = 1/8$$

Vậy hàm phân phối là

$$F(x) = \begin{cases} 0 & , x < 0 \\ \frac{1}{8} & ,0 \le x < 1 \\ \frac{1}{2} & ,1 \le x < 2 \\ \frac{7}{8} & 2 \le x < 3 \\ 1 & 3 \le x \end{cases}$$

Đồ thị của hàm phân phối như sau



• *Tính chất*. Cho hàm phân phối F(x) của biến ngẫu nhiên X. Khi đó
(i) F(x) không giảm trên R
(ii) $\lim\limits_{x \to +\infty} F(x) = 1$ và $\lim\limits_{x \to -\infty} F(x) = 0$
(iii) F(x) liên tục phải trên R
(iv) Nếu X là biến ngẫu nhiên rời rạc với luật xác suất

$$\{(x_i , p_i) \mid i=1, \dots , n\}$$

thì hàm phân phối F(x) được xác định theo công thức sau

$$F(x) = \sum_{x_i \le x} p_i$$

và

$$P(X=x) = F(x) - \lim_{t \to x_{-0}} F(t)$$

↳ Ghi chú: Từ (i) và (ii) suy ra

$$0 \le F(x) \le 1 , \forall x \in R$$

3. Các tham số đặc trưng

a) Kỳ vọng.

Cho biến ngẫu nhiên rời rạc X với luật xác suất

$$\{(x_i , p_i) \mid i \in I\}$$

Kỳ vọng của X, ký hiệu E(X), là giá trị xác định bởi công thức

$$E(X) = \sum_{i \in I} x_i p_i$$

Trong trường hợp tập chỉ số I vô hạn thì chuỗi phải hội tụ tuyệt đối.

+ *Ví dụ*.

(i) *Phân phối Bernoulli.*

Cho A là sự kiện của phép thử α có xác suất P(A) = p. Gọi X là biến ngẫu nhiên chỉ số lần xuất hiện sự kiện A trong 1 lần thực hiện phép thử α. Luật phân phối của X là

$$\{(1,p), (0,1-p)\}$$

Vậy

$$E(X) = 1.p + 0.(1-p) = p$$

(ii) Cho X là biến ngẫu nhiên có luật phân phối nhị thức B(n,p)

$$\{(k, C(n,k).p^k.(1-p)^{n-k} \mid k=0, 1, \ldots, n\}$$

Sử dụng đẳng thức

$$k.C(n,k) = n.C(n-1,k-1),\ \forall n,k\in N,\ n\geq k$$

ta có

$$E(X) = \sum_{k=0}^{n} kC(n,k).p^k(1-p)^{n-k} = \sum_{k=1}^{n} kC(n,k).p^k(1-p)^{n-k} =$$

$$= \sum_{k=1}^{n} n.C(n-1,k-1).p^k(1-p)^{n-k} = n.p.\sum_{h=0}^{n-1} C(n-1,h).p^h(1-p)^{n-1-h} = n.p$$

(iii) Cho biến ngẫu nhiên X có luật phân phối xác suất Poisson P(λ)

$$\{(k, e^{-\lambda}.\frac{\lambda^k}{k!} \mid k = 0,1,2,....\}$$

Ta có

$$E(X) = \sum_{k=0}^{\infty} k.e^{-\lambda}.\frac{\lambda^k}{k!} = \sum_{k=1}^{\infty} \lambda.e^{-\lambda}.\frac{\lambda^{k-1}}{(k-1)!} = \lambda.e^{-\lambda}\sum_{h=0}^{\infty}\frac{\lambda^h}{h!} = \lambda.e^{-\lambda}.e^{\lambda} = \lambda$$

(iv) Cho X là biến ngẫu nhiên có luật phân phối xác suất rời rạc vô hạn

$$\left\{\left(k, \frac{1}{k.(k+1)}\right) \middle| k = 1,2,...\right\}$$

Ta có

$$\sum_{k=1}^{\infty}\frac{1}{k(k+1)} = \sum_{k=1}^{\infty}\left(\frac{1}{k} - \frac{1}{k+1}\right) = 1$$

nhưng

$$\sum_{k=1}^{\infty} k\frac{1}{k(k+1)} = \sum_{k=1}^{\infty}\frac{1}{k+1} = +\infty$$

Chuỗi trên phân kỳ nên X không có kỳ vọng.

• *Định lý*. Cho X là biến ngẫu nhiên rời rạc của phép thử α có luật phân phối xác suất

$$\{(x_i\ ,\ p_i) \mid i \in I\}$$

Giả sử X có kỳ vọng E(X) và ánh xạ φ:R→R . Khi đó φ(X) cũng là biến ngẫu nhiên của phép thử α. Nếu φ(X) có kỳ vọng thì

$$E[\varphi(X)] = \sum_{i\in I}\varphi(x_i).p_i$$

CM. Suy ra trực tiếp từ định nghĩa.

• *Tính chất*. Cho X, Y là các biến ngẫu nhiên, C là hằng. Khi đó
(i) E(C) = C
(ii) E(C.X) = C.E(X)
(iii) E(X + Y) = E(X) + E(Y)
(iv) Nếu X, Y là các biến ngẫu nhiên độc lập, thì

$$E(X.Y) = E(X).E(Y)$$

b) *Mode*.
• *Định nghĩa*. Cho biến ngẫu nhiên X với luật phân phối xác suất

$$\{(x_i , p_i) \mid i \in I\}$$

Mode của X là trị của X có xác suất cực đại, ký hiệu M(X).

+ *Ví dụ*. Cho biến ngẫu nhiên X có luật phân phối xác suất nhị thức B(n,p). Ký hiệu

$$P_n(k) = P(X=k) = C(n,k).p^k.q^{n-k} \quad (q=1-p)$$

Ta có

$$\frac{P_n(k+1)}{P_n(k)} = \frac{C(n,k+1).p^{k+1}.q^{n-k-1}}{C(n,k).p^k.q^{n-k}} = \frac{n!}{(k+1)!.(n-k-1)!}\,\frac{k!.(n-k)!}{n!}\,\frac{p}{q} = \frac{n-k}{k+1}\,\frac{p}{q}$$

Từ đó suy ra

$$\begin{aligned} P_n(k+1) > P_n(k) &\Leftrightarrow (n-k).p > (k+1).q \\ &\Leftrightarrow n.p - k.p > (k+1)(1-p) \\ &\Leftrightarrow n.p - k.p > -k.p - p + k + 1 \\ &\Leftrightarrow k < n.p - q \end{aligned}$$

$$P_n(k+1) = P_n(k) \Leftrightarrow k = n.p - q$$
$$P_n(k+1) < P_n(k) \Leftrightarrow k > n.p - q$$

Ta thấy khi k tăng từ 0 đến n, hàm $P_n(k)$ tăng đến cực đại rồi giảm dần. Vậy
Nếu $n.p - q \geq 0$ là số nguyên thì $P_n(k)$ đạt cực đại tại

$$k_0 = n.p - q \qquad \text{và} \qquad k_0 + 1 = np - q + 1 = n.p + p$$

Nếu $n.p - q > 0$ không nguyên thì $P_n(k)$ đạt cực đại tại

$$k_0 = \lceil n.p - q \rceil \; (\lceil x \rceil \text{ là số nguyên nhỏ nhất lớn hơn } x)$$

Nếu $n.p - q < 0$ thì $P_n(k)$ đạt cực đại tại $k_0 = 0$.

c) *Phương sai và độ lệch quân phương*.
• *Định nghĩa*. Cho biến ngẫu nhiên X với luật phân phối xác suất

$$\{(x_i , p_i) \mid i \in I\}$$

và kỳ vọng E(X) = a. Giả thiết chuỗi $\sum_{i\in I}(x_i - a)^2.p_i$ hội tụ. Khi đó tổng của chuỗi gọi là *phương sai* của X và ký hiệu là D(X), tức là

$$D(X) = \sum_{i\in I}(x_i - a)^2.p_i = E(X-a)^2$$

Nếu X có phương sai D(X) thì đại lượng

$$\sigma(X) = \sqrt{D(X)}$$

gọi là *độ lệch quân phương* của biến ngẫu nhiên X.

• *Công thức Koenig−Huyghens*.

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

CM.

$$D(X) = E(X-a)^2 = E(X^2 - 2.a.X + a^2)$$

$$= E(X^2) - 2E(X).E(X) + E(X)^2 = E(X^2) - [E(X)]^2$$

+ *Ví dụ*.

(i) Tính phương sai của biến ngẫu nhiên X có luật phân phối nhị thức B(n,p) (q=1−p).

Ta có

$$E(X^2) = \sum_{k=0}^{n} k^2.C(n,k).p^k.q^{n-k}$$

$$= \sum_{k=1}^{n} k.n.p.C(n-1,k-1).p^{k-1}.q^{(n-1)-(k-1)}$$ (áp dụng k.C(n,k) = n.C(n−1,k−1))

$$= n.p.\sum_{h=0}^{n-1}(h+1).C(n-1,h).p^h.q^{(n-1)-h}$$ (h = k − 1)

$$= n.p.\sum_{h=0}^{n-1}h.C(n-1,h).p^h.q^{(n-1)-h} + n.p.\sum_{h=0}^{n-1}C(n-1,h).p^h.q^{(n-1)-h}$$

= n.p.E(X') + n.p.1

trong đó X' có luật nhị thức B(n−1,p). Vậy

$$E(X^2) = n.p.(n-1).p + n.p = n.(n-1).p^2 + n.p$$

Theo công thức Koenig−Huyghens suy ra

$$D(X) = n.(n-1).p^2 + n.p - (n.p)^2 \quad = \quad n.p.q$$

Độ lệch quân phương cuả X là

$$\sigma(X) = \sqrt{n.p.q}$$

(ii) Tính phương sai của biến ngẫu nhiên X có luật phân phối Poisson tham số λ > 0, P(λ).

Ta có

$$E(X^2) = \sum_{k=1}^{\infty} k^2.e^{-\lambda}\frac{\lambda^k}{k!} = e^{-\lambda}.\lambda.\sum_{k=1}^{\infty} k\frac{\lambda^{k-1}}{(k-1)!} = e^{-\lambda}.\lambda.\sum_{h=0}^{\infty}(h+1)\frac{\lambda^h}{h!} \qquad (h=k-1)$$

$$= \lambda.\sum_{h=0}^{\infty} h.e^{-\lambda}\frac{\lambda^h}{h!} + \lambda.e^{-\lambda}.\sum_{h=0}^{\infty}\frac{\lambda^h}{h!} = \lambda.E(X) + \lambda.e^{-\lambda}.e^{\lambda} = \lambda.\lambda + \lambda = \lambda^2 + \lambda$$

Theo công thức Koenig−Huyghen suy ra

$$D(X) = \lambda^2 + \lambda - \lambda^2 = \lambda$$

Độ lệch quân phương cuả X là

$$\sigma(X) = \sqrt{\lambda}$$

• *Tính chất*. Cho X, Y là các biến ngẫu nhiên, C là hằng. Khi đó

(i) D(C) = 0

(ii) $D(C.X) = C^2.D(X)$

(iii) σ(C.X) = |C|.σ(X)

(iv) Nếu X, Y là các biến ngẫu nhiên độc lập, thì

$$D(X + Y) = D(X) + D(Y)$$

CM

(iv) $D(X + Y) = E(X + Y)^2 - [E(X + Y)]^2 = E(X^2 + 2X.Y + Y^2) - [E(X) + E(Y)]^2$

$= E(X^2) - [E(X)]^2 + E(Y^2) - [E(Y)]^2 + 2.E(X.Y) - 2.E(X).E(Y)$

= D(X) + D(Y) (vì E(X.Y) = E(X).E(Y))

• *Hệ quả*. Cho $X_1, \dots, X_n$ là các biến ngẫu nhiên độc lập có cùng phương sai D. Ký hiệu

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Khi đó

$$D(\overline{X}) = \frac{D}{n}$$

CM

$$D(\overline{X}) = D\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} D(X_i) = \frac{1}{n^2} n.D = \frac{D}{n}$$

d) Mômen

Cho biến ngẫu nhiên X, a $\in$ R, k $\in$ N.

• *Mômen cấp k đối với a* của X là giá trị

$$\nu_k(a) = E[(X - a)^k]$$

• *Mômen gốc cấp k* của X là giá trị

$$\nu_k = \nu_k(0) = E(X^k)$$

↳ *Mômen gốc cấp 1* của X là kỳ vọng E(X) (nếu tồn tại).

• *Mômen trung tâm cấp k* của X, với điều kiện X có kỳ vọng, là giá trị

$$\mu_k = \nu_k(E(X)) = E[(X - E(X))^k]$$

↳ *Mômen trung tâm cấp 2* của X là phương sai D(X) (nếu tồn tại).

• *Định lý*. Cho r, s $\in$ N , r ≤ s, và biến ngẫu nhiên rời rạc X. Khi đó, nếu tồn tại mômen gốc cấp s của X, thì tồn tại mômen gốc cấp r của X.

CM

Giả sử luật phân phối xác suất của X là

$$\{(x_i , p_i) \mid i \in I\}$$

Vì ν_s tồn tại nên chuỗi $\sum_{i\in I} |x_i^s| p_i$ hội tụ. Ta có

$$\left|\sum_{i\in I} x_i^r . p_i\right| \le \sum_{i\in I} |x_i^r| p_i \le \sum_{i\in I} \left(1 + |x_i^s|\right) p_i$$

(vì $|x_i^r| \le |x_i^s|$, nếu $|x_i| \ge 1$ và $|x_i^r| \le 1$, nếu $|x_i| \le 1$).

Suy ra

$$\left|\sum_{i\in I} x_i^r . p_i\right| \le \sum_{i\in I} p_i + \sum_{i\in I} |x_i^s| . p_i < \infty$$

Cuối cùng ta suy ra $\sum_{i\in I} |x_i^r| p_i$ hội tụ tuyệt đối.

Đpcm

• *Hệ quả*. Cho r, s $\in$ N , r $\le$ s, và biến ngẫu nhiên rời rạc X. Khi đó, nếu tồn tại mômen gốc cấp s của X, thì tồn tại mômen cấp r tại a $\in$ R bất kỳ của X.

CM

Sử dụng định lý và đẳng thức sau

$$\gamma_k(a) = \mathrm{E}[(\mathrm{X} - \mathrm{a})^{\mathrm{r}}] = \mathrm{E}\left[\sum_{i=0}^{\mathrm{n}} \mathrm{C}(\mathrm{n},\mathrm{k}).(-1)^{\mathrm{r}-\mathrm{i}}.a^{r-i}x^i\right] = \sum_{i=0}^{\mathrm{r}} \mathrm{C}(\mathrm{n},\mathrm{k}).(-1)^{\mathrm{r}-\mathrm{i}}.a^{r-i}.\mathrm{E}(\mathrm{X}^{\mathrm{i}})$$

• *Mômen trung tâm cấp 3*.

$$\begin{aligned}\mu_3 &= E[X - E(X)]^3 = E[X^3 - 3.X^2.E(X) + 3.X.E(X)^2 - E(X)^3] = \\ &= \nu_3 - 3.\nu_1\nu_2 + 3.\nu_1^3 - \nu_1^3 = \nu_3 - 3.\nu_1\nu_2 + 2.\nu_1^3\end{aligned}$$

Bây giờ ta giả thiết rằng X đối xứng qua E(X), tức X có luật phân phối

$$\{(x_i, p_i) \mid i = -\infty , \dots , -2, -1, 1, 2, \dots , +\infty \}$$

trong đó

$$p_i = p_{-i} \text{ và } x_i - \nu_1 = -(x_{-i} - \nu_1) , \forall i = 1, 2, \dots , \infty$$

Khi đó mọi mômen trung tâm cấp lẻ đều bằng 0.

Mômen μ_3 đặc trưng cho tính bất đối xứng của biến ngẫu nhiên đối với kỳ vọng. Nếu $\mu_3 > 0$ thì phân phối lệch về bên phải kỳ vọng, nếu $\mu_3 < 0$ thì phân phối lệch về bên trái kỳ vọng.

Đại lượng

$$s_3 = \frac{\mu_3}{\sigma^3}$$

gọi là *hệ số bất đối xứng* của biến ngẫu nhiên X.

• *Mômen trung tâm cấp 4*.

$$\begin{aligned}\mu_3 &= E[X - E(X)]^4 = E[X^4 - 4.X^3.E(X) + 6.X^2.E(X)^2 - 4.X.E(X)^3 + E(X)^4] = \\ &= \nu_4 - 4.\nu_3\nu_1 + 6.\nu_1^2.\nu_2 - 4.\nu_1^3.\nu_1 + \nu_1^4 = \nu_4 - 4.\nu_3\nu_1 + 6.\nu_1^2.\nu_2 - 3.\nu_1^4\end{aligned}$$

4. Một số luật phân phối quan trọng

a) *Luật phân phối đều rời rạc* U(1,n).

• *Định nghĩa*. Biến ngẫu nhiên X gọi là tuân theo luật phân phối đều U(1,n) nếu có phân phối

$$\{(k, 1/n) \mid k = 1, 2, \dots, n\}$$

• Kỳ vọng và phương sai:

$$E(X) = \frac{n+1}{2} \qquad \& \qquad D(X) = \frac{n^2-1}{12}$$

CM

$$E(X) = \sum_{k=1}^{n} k.\frac{1}{n} = \frac{1}{n}\sum_{k=1}^{n} k = \frac{1}{n}\frac{n(n+1)}{2} = \frac{n+1}{2}$$

$$D(X) = E(X^2) - E(X)^2 = \frac{1}{n}\sum_{k=1}^{n} k^2 - \frac{(n+1)^2}{4} = \frac{1}{n}\frac{n(n+1)(2.n+1)}{6} - \frac{(n+1)^2}{4} = \frac{n^2-1}{12}$$

+ *Ví dụ*. Trong thùng có n quả cầu, trong đó có 1 quả cầu đen. Người ta lấy ngẫu nhiên từng quả cầu ra khỏi thùng (không bỏ lại vào thùng) cho đến khi lấy được quả cầu đen. Gọi X là biến ngẫu nhiên có giá trị bằng số quả cầu lấy ra. Xác định luật phân phối của X.
Giải.

Miền giá trị của X là 1, 2, …, n.

Gọi A_i là sự kiện quả cầu đen được rút ra ở lần thứ i, i=1,2,…,n. Với mọi k=1,2,…,n ta có

P(X=k) $= P\left(\overline{A_1}.\overline{A_2}.....\overline{A_{k-1}}.A_k\right)$

$= P\left(\overline{A_1}\right)P\left(\overline{A_2}/\overline{A_1}\right)P\left(\overline{A_3}/\overline{A_1}\,\overline{A_2}\right).....P\left(\overline{A_{k-1}}/\overline{A_1}\,\overline{A_2}.....\overline{A_{k-2}}\right)P\left(A_k/\overline{A_1}\,\overline{A_2}.....\overline{A_{k-1}}\right)$

$= \frac{n-1}{n}\frac{n-2}{n-1}.....\frac{n-k+1}{n-k+2}\frac{1}{n-k+1} = \frac{1}{n}$

Vậy X có phân phối đều U(1,n).

b) *Phân phối Bernoulli* B(1,p).

• *Định nghĩa*. biến ngẫu nhiên X gọi là tuân theo luật phân phối Bernoulli B(1,p), 0<p<1, nếu X có phân phối xác suất

{(1,p), (0,1−p)}

• Kỳ vọng và phương sai:

E(X) = p & D(X) = p.(1−p)

+ *Ví dụ*. Hai đối thủ ngang sức chơi *2.n* ván cờ, với điều kiện mỗi ván phải có người thắng. Gọi X_n là biến ngẫu nhiên,

X_n = 1, nếu trận đấu hoà, ngược lại X_n = 0.

Xác định luật phân phối của X_n.
Giải.

Sự kiện X_n = 1 xảy ra nếu người thứ nhất thắng n trận. Xác suất một khả năng như vậy là

$$(½)^{n}.(½)^{2n-n} = (½)^{2n}$$

và mỗi khả năng ứng với một tổ hợp chập n của 2n. Vậy xác suất

$$p_n = P(X_n=1) = C(2n,n).(½)^{2n}$$

và X_n có phân phối B(1,p_n).
* Ghi chú. biến ngẫu nhiên

$$X_1 + X_2 + \ldots + X_n$$

là số thời điểm hai người có số ván thắng thua bằng nhau.

c) *Phân phối nhị thức* B(n,p).
• *Định nghĩa*. biến ngẫu nhiên X gọi là tuân theo luật phân phối nhị thức B(n,p), $n \in N$, $0<p<1$, nếu X có phân phối xác suất

$$\{(k, C(n,k).p^k.(1-p)^{n-k}) \mid k=0,1,2, \ldots ,n\}$$

• Kỳ vọng và phương sai:

$$E(X) = n.p \qquad \& \qquad D(X) = n.p.(1-p)$$

+ *Ví dụ*. Có n khách hàng và m quầy hàng. Mỗi khách hàng chọn ngẫu nhiên một quầy hàng với xác suất giống nhau và hoàn toàn độc lập với nhau. Gọi X là số khách hàng chọn quầy số 1. Xác định luật phân phối của biến ngẫu nhiên X.
Giải.
Mỗi khách hàng chọn quầy số 1 với xác suất *p = 1/m*. Suy ra xác suất

$$P(X=k) = C(n,k).(1/m)^k.(1-1/m)^{n-k}$$

Như vậy X có luật phân phối B(n,1/m).

d) *Phân phối hình học* G(p).
• *Định nghĩa*. biến ngẫu nhiên X gọi là tuân theo luật phân phối hình học G(p), $0<p<1$, nếu X có phân phối xác suất

$$\{(k, p.(1-p)^{k-1}) \mid k = 1, 2, \ldots ,n, \ldots , +\infty \}$$

• Kỳ vọng và phương sai: Ký hiệu $q = 1 - p$, ta có

$$E(X) = 1/p \qquad \& \qquad D(X) = q/p^2$$

CM.

(i) $E(X) = p\sum_{k=1}^{\infty} k.(1-p)^{k-1} = p.\dfrac{d\left(\sum_{k=1}^{\infty} q^k\right)}{dq} = p.\dfrac{d\left(\dfrac{1}{1-q}\right)}{dq} = \dfrac{p}{(1-q)^2} = \dfrac{1}{p}$

(ii) Ta có

$$D(X) = p.\sum_{k=1}^{\infty} k^2.q^{k-1} - \frac{1}{p^2} \qquad (1)$$

Theo (i) ta có

$$\sum_{k=1}^{\infty} k.q^{k-1} = \frac{1}{(1-q)^2} \Rightarrow \sum_{k=1}^{\infty} k.q^{k} = \frac{q}{(1-q)^2} \Rightarrow \frac{d\left(\sum_{k=1}^{\infty} k.q^k\right)}{dq} = \frac{d\left(\frac{q}{(1-q)^2}\right)}{dq}$$

==> $$\sum_{k=1}^{\infty} k^2.q^{k-1} = \frac{(1-q)^2 + 2.q.(1-q)}{(1-q)^4} = \frac{p^2 + 2.p.q}{p^4} = \frac{p+2.q}{p^3}$$

Thế vào (1) ta có

$$D(X) = p\frac{p+2.q}{p^3} - \frac{1}{p^2} = \frac{q}{p^2}$$

+ *Ví dụ*. Ba người chơi trò chơi gieo xúc sắc. Mỗi ván mỗi người gieo một lần, nếu có người gieo được mặt 6 chấm thì người đó thắng. Trò chơi kết thúc nếu có người thắng. Gọi X là số ván chơi. Xác định luật phân phối của biến ngẫu nhiên X.
Giải.
(i) Trường hợp có thể có nhiều người thắng trong một ván.
Sự kiện X = k xảy ra nếu trong k − 1 ván: 1,2,…,k−1 không có ai gieo được 6 chấm và đến ván thứ k thì có người gieo được 6 chấm.
Gọi A là sự kiện không có ai thắng trong một ván. Ta có

$$P(A) = \left(\frac{5}{6}\right)^3 \qquad \& \qquad P(\overline{A}) = 1 - \left(\frac{5}{6}\right)^3$$

Vậy

$$P(X=k) = \left[1-\left(\frac{5}{6}\right)^3\right]\left(\frac{5}{6}\right)^{3.(k-1)}, \ k=1,2,\dots,+\infty$$

tức X có luật phân phối hình học G(p), với p = $1-\left(\frac{5}{6}\right)^3 = \frac{216-125}{216} = \frac{91}{216}$

(ii) Trường hợp chỉ cho phép một người thắng trong một ván.
Sự kiện X = k xảy ra nếu ở ván thứ k có đúng 1 người gieo được 6 chấm và trong mỗi ván: 1,2,…,k−1 hoặc không có ai gieo được 6 chấm hoặc có ít nhất 2 người gieo được 6 chấm.
Gọi A là sự kiện có đúng 1 người thắng trong một ván. Ta có

$$P(A) = 3.\frac{1}{6}\left(\frac{5}{6}\right)^2 = \frac{25}{72} \qquad \& \qquad P(\overline{A}) = 1 - \frac{25}{72}$$

Vậy

$$P(X=k) = \frac{25}{72}\left(1-\frac{25}{72}\right)^{(k-1)}, \ k=1,2,\dots,+\infty$$

tức X có luật phân phối hình học G(p), với p = $\frac{25}{72}$

e) *Phân phối siêu hình học* H(N,n,p).

• *Định nghĩa*. biến ngẫu nhiên X gọi là tuân theo luật phân phối siêu hình học H(N,n,p) (N, n ∈ **N**, 0<p<1, q=1−p, N.p ∈ **N**), nếu X có phân phối xác suất

$$\{(k, \frac{C(N.p,k).C(N.q,n-k)}{C(N,n)}) \mid k = 0,1,2,\ldots,n\}$$

• Kỳ vọng và phương sai:

$$E(X) = n.p \qquad \& \qquad D(X) = n.p.q.\frac{N-n}{N-1}$$

CM.
(i) Kỳ vọng: Ta có

$$E(X) = \sum_{k=0}^{n} k.\frac{C(N.p,k).C(N.q,n-k)}{C(N,n)}$$

Sử dụng đẳng thức k.C(n,k) = n.C(n−1,k−1), 1≤ k≤ n, suy ra

$$E(X) = \sum_{k=0}^{n} N.p.\frac{C(Np-1,k-1).C(Nq,n-k)}{C(N,n)}$$

Tiếp theo, áp dụng công thức Vandermonde, ta có

$$E(X) = N.p.\frac{C(N.p+N.q-1,n-1)}{C(N,n)} = N.p\frac{C(N-1,n-1)}{C(N,n)}$$

$$\Rightarrow \qquad E(X) = N.p\frac{(N-1)!}{(n-1)!(N-n)!}\frac{n!(N-n)!}{N!} = N.p\frac{n}{N} = n.p$$

(ii) Phương sai. Áp dụng hai lần công thức giảm bậc, ta có

$$E[X.(X-1)] = \sum_{k=1}^{n} k(k-1)\frac{C(N.p,k).C(N.q,n-k)}{C(N,n)}$$

$$= N.p.(N.p-1).\sum_{k=2}^{n}\frac{C(N.p-2,k-2).C(N.q,n-k)}{C(N,n)}$$

Tiếp theo, áp dụng công thức Vandermonde, ta có

E[X.(X−1)] = N.p.(N.p − 1).C(N.p + N.q −2, n − 2) / C(N,n)
= N.p.(N.p − 1).C(N −2, n − 2) / C(N,n)

$$= N.p.(N.p-1).\frac{(N-2)!}{(n-2)!(N-n)!}\frac{n!(N-n)!}{N!}$$

$$= N.p.(N.p-1).\frac{n.(n-1)}{N.(N-1)}$$

Từ trên suy ra

$$D(X) = E[X(X-1)] + E(X) - E(X)^2$$

$$= N.p.(N.p-1).\frac{n.(n-1)}{N.(N-1)} + n.p - (n.p)^2$$

$$= n.p.\left[\frac{(N.p-1).(n-1)}{N-1} + 1 - n.p\right]$$

$$= n.p.\left[\frac{N + n.p - n - n.p}{N-1}\right] = n.p.q\frac{N-n}{N-1}$$

+ *Ví dụ*.

Trong thùng có a quả cầu trắng, b quả cầu đen. Lấy ngẫu nhiên n quả cầu ($n \le a+b$). Ký hiệu X là biến ngẫu nhiên chỉ số quả cầu trắng được lấy ra. Xác định luật phân phối của X.

Giải.

X nhận giá trị nguyên k trong khoảng

$$[\max(0, n-b), \min(n,a)]$$

và xác suất

$$P(X = k) = \frac{C(a,k).C(b,n-k)}{C(a+b,n)}$$

Như vậy X có phân phối siêu hình học H(a+b, n, a/(a+b)) với kỳ vọng

$$E(X) = n.a/(a+b)$$

Và phương sai

$$D(X) = n\frac{a}{a+b}\frac{b}{a+b}\frac{a+b-n}{a+b-1}$$

f) *Phân phối Poisson* P(λ) (λ>0).

• *Định nghĩa*. biến ngẫu nhiên X gọi là tuân theo luật phân phối Poisson P(λ), 0<λ, nếu X có phân phối xác suất

$$\{(k, e^{-\lambda}\cdot\frac{\lambda^k}{k!}) \mid k = 0, 1, 2, \ldots, n, \ldots, +\infty\}$$

• Kỳ vọng và phương sai:

$$E(X) = \lambda \quad \& \quad D(X) = \lambda$$

+ *Ví dụ*. Luật phân phối Poisson là luật giới hạn.

Cho dãy thùng $(U_n)_{n\ge 1}$, trong đó thùng U_n, $n \ge 1$, chứa các quả cầu trắng và cầu đen theo tỉ lệ p_n. Với n=1,2,… ta lần lượt lấy n quả cầu, có trả lại, từ thùng U_n.

Ký hiệu X_n là số quả cầu trắng lấy ra từ thùng U_n.

Giả thiết $n.p_n \to \lambda$ khi $n\to\infty$. Ta nghiên cứu xác suất $P(X_n = k)$ khi $n\to\infty$. Ký hiệu $1 - p_n = q_n$, ta có

$$P(X_n = k) = C(n,k).p_n^k.q_n^{n-k} = \frac{n(n-1)...(n-k+1)}{k!} p_n^k e^{(n-k)\ln(q_n)}$$

Vì $n.p_n \to \lambda$ khi $n\to\infty$, nên $p_n \to 0$. Suy ra

$$\lim_{n\to\infty}(n-k)\ln(q_n) = \lim_{n\to\infty}(n-k)\ln(1-p_n) = -\lim_{n\to\infty}(n-k)p_n = -\lambda$$

kéo theo

$$\lim_{n\to\infty} P(X_n = k) = \lim_{n\to\infty}\left(\frac{n^k}{k!} p_n^k e^{-\lambda}\right) = e^{-\lambda}\frac{\lambda^k}{k!}$$

Như vậy phân phối của X_n tiệm cận đến phân phối Poisson $P(\lambda)$.

III. BIẾN NGẪU NHIÊN LIÊN TỤC

1. Hàm mật độ phân phối

• *Định nghĩa*. Biến ngẫu nhiên X trên không gian xác suất (Ω, **B**, P) gọi là liên tục khi và chỉ khi tồn tại hàm f:R→R thoả

(i) f(x) ≥ 0 ∀x∈R

(ii) Số điểm gián đoạn của f là hữu hạn

(iii) Tại các điểm gián đoạn, f có giới hạn phải và giới hạn trái, hữu hạn hoặc vô hạn

(iv) Tích phân

$$\int_{-\infty}^{+\infty} f(t)dt = 1$$

(v) Hàm phân phối F(x) thoả

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

Hàm f(t) gọi là *hàm mật độ phân phối* của biến ngẫu nhiên X.

Ngược lại, hàm f(t) thoả (i)−(iv) là hàm mật độ của biến ngẫu nhiên nào đó.
Từ định nghĩa suy ra hệ quả

• *Hệ quả*. Cho biến ngẫu nhiên liên tục X với hàm mật độ f(t). Khi đó

(i) P(X=a) = 0 ∀a∈R

(ii) $P(a<X<b) = P(a \le X \le b) = P(a<X \le b) = P(a \le X<b) = \int_a^b f(t)dt$ ∀a,b∈R, a<b

(iii) Hàm phân phối F(x) liên tục trên R và khả vi tại các điểm liên tục của f và
F'(t) = f(t) ∀t là điểm liên tục của f

+ *Ví dụ*.

(i) Cho a,b∈R, a<b, và X là điểm ngẫu nhiên chọn trong đoạn [a,b]. Giả thiết rằng xác suất X rơi vào các khoảng bằng nhau trong [a,b] là giống nhau. Xác định hàm mật độ của X.

Giải.

Vì xác suất P(a<X≤ b) = 1, ta có

$$P(\alpha < X \le \beta) = \frac{\beta-\alpha}{b-a} \qquad \forall[\alpha,\beta] \subset [a,b]$$

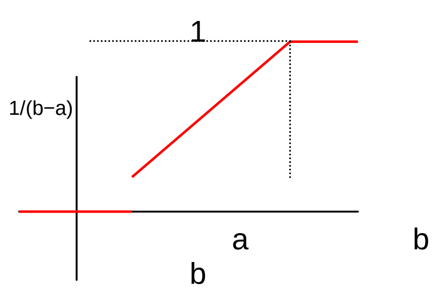
Vậy hàm phân phối của X là

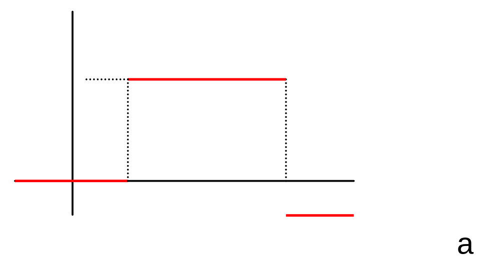
$$F(x) = \begin{cases} 0 & ,x \le a \\ \dfrac{x-a}{b-a} & ,a < x \le b \\ 1 & ,b < x \end{cases}$$

Suy ra hàm mật độ f(t) của X:

$$f(t) = F'(t) = \begin{cases} \dfrac{1}{b-a} & \forall a \le t \le b \\ 0 & \forall t < a \vee t > b \end{cases}$$

Đồ thị của hàm phân phối và hàm mật độ như sau





hàm phân phối F(x) hàm mật độ f(t)

Biến ngẫu nhiên X gọi là tuân theo luật phân phối đều trên [a,b], U([a,b]).

(ii) Một chất phóng xạ phân rã. Tuổi thọ T của chất phóng xạ tại thời điểm bất kỳ không phụ thuộc vào thời gian sống trước đó của nó. Xác định hàm mật độ của T.
Giải.

Theo giả thiết ta có

$$P(T \ge t+u \,/\, T \ge t) = P(T \ge u) \qquad \forall t,u > 0$$

Suy ra

$$\frac{P(T \ge t+u \,\&\, T \ge t)}{P(T \ge t)} = P(T \ge u)$$

$$\Rightarrow \qquad P(T \ge t+u) = P(T \ge t).P(T \ge u) \qquad \forall t,u > 0$$

Ký hiệu F là hàm phân phối của T và H = 1 − F. Theo trên ta có

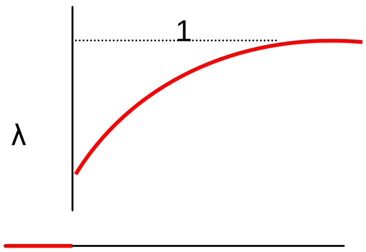
$$H(t+u) = H(t).H(u) \qquad \forall t,u > 0$$

$$\Rightarrow \qquad H(x) = e^{a.x} \quad \Rightarrow \quad F(x) = 1 - e^{a.x} \qquad \forall x > 0$$

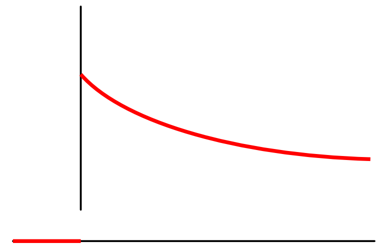
Vì $F(x) \to 1$ khi $x \to +\infty$, nên a<0. Đặt λ = −a, ta có

$$F(x) = 1 - e^{-\lambda.x} \quad \Rightarrow \quad f(x) = F'(x) = \lambda.e^{-\lambda.x} \quad \forall x \ge 0$$

Đồ thị của hàm phân phối và hàm mật độ như sau:



hàm phân phối F(x)



hàm mật độ f(t)

T gọi là tuân theo *luật phân phối mũ* tham số λ , E(λ).

(iii) Biến ngẫu nhiên X có hàm mật độ

$$f(x) = \begin{cases} k.\cos x & -\dfrac{\pi}{2} \le x \le \dfrac{\pi}{2} \\ 0 & |x| > \dfrac{\pi}{2} \end{cases}$$

a) Tìm k
b) Tìm hàm phân phối F(x).
c) Tính xác suất P(0 < X < π/4).

Giải.

a) Tìm k.

$$\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} k.\cos t dt = 1 \Rightarrow \quad k.2 = 1 \quad \Rightarrow \quad k = ½$$

b) Tìm hàm phân phối F(x).
Với x∈[−π/2, π/2] ta có

$$F(x) = \int_{-\frac{\pi}{2}}^{x} \frac{1}{2}.\cos t dt = \frac{1}{2}(\sin x + 1)$$

Suy ra hàm phân phối

$$F(x) = \begin{cases} 0 & , x \le -\dfrac{\pi}{2} \\ \dfrac{1}{2}(\sin x + 1) & , -\dfrac{\pi}{2} < x \le +\dfrac{\pi}{2} \\ 1 & , +\dfrac{\pi}{2} < x \end{cases}$$

c) Tính xác suất P(0 < X < π/4).

$$P(0 < X < \pi/4) = F(\pi/4) - F(0) = \frac{1}{2}\left(\frac{1}{\sqrt{2}}+1\right)-\frac{1}{2}=\frac{1}{2\sqrt{2}}$$

(iv) Tìm k để hàm

$$f(t)=\begin{cases} k.\dfrac{1}{\sqrt{t(1-t)}} & ,0<t<1 \\ 0 & ,t\notin(0,1)\end{cases}$$

là hàm mật độ của biến ngẫu nhiên

Giải.

Ta có $f(t) \sim t^{-1/2}$ khi $t \to 0_+$ và $f(t) \sim (1-t)^{-1/2}$ khi $t \to 1_-$ kéo theo tích phân $\int_0^1 f(t)dt$ hội tụ.

Đổi biến $t = \sin^2(u)$ ta có

$$\int_0^1 f(t)dt = k\int_0^1 \frac{1}{\sqrt{t(1-t)}}dt = k\int_0^{\frac{\pi}{2}} \frac{2.\sin u.\cos u}{\sqrt{\sin^2 u.(1-\sin^2 u)}}du = k\int_0^{\frac{\pi}{2}} 2du = k\pi$$

$\Rightarrow$ k.π = 1 $\Rightarrow$ k = 1/π.

Vậy f(t) là hàm mật độ của biến ngẫu nhiên khi và chỉ khi k = 1/π.

2. Các tham số đặc trưng

a) Kỳ vọng.

Cho biến ngẫu nhiên X liên tục, với hàm mật độ f(t), trên không gian xác suất (Ω,**B**,P).

• Định nghĩa. Kỳ vọng của biến ngẫu nhiên liên tục X là đại lượng

$$E(X) = \int_{-\infty}^{\infty} tf(t)dt$$

với điều kiện tích phân hội tụ tuyệt đối.

+ *Ví dụ*. Cho X là biến ngẫu nhiên tuân theo luật phân phối đều trên [a,b]. Ta có

$$E(X) = \int_a^b \frac{t}{b-a}dt = \left[\frac{1}{b-a}\frac{t^2}{2}\right]_{t=a}^{b} = \frac{a+b}{2}$$

• *Định lý*. Cho X là biến ngẫu nhiên với hàm mật độ f(t), φ:R→R là hàm sao cho φ(X) cũng là biến ngẫu nhiên. Khi đó

$$E[\varphi(X)] = \int_{-\infty}^{\infty} \varphi(t).f(t)dt$$

nếu tích phân hội tụ tuyệt đối.

CM. Xem phần Hàm biến ngẫu nhiên.

• *Tính chất*. Cho X, Y là các biến ngẫu nhiên liên tục, C là hằng. Khi đó

(i) E(C) = C

(ii) E(C.X) = C.E(X)

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

(iv) Nếu X, Y là các biến ngẫu nhiên độc lập, thì

$$E(X.Y) = E(X).E(Y)$$

b) *Mode và trung vị*.

• *Định nghĩa*. Cho biến ngẫu nhiên X với hàm mật độ f(t).

Mode của X là điểm M(X) mà tại đó f(t) đạt cực đại.

Trung vị của X là điểm m(X) thoả

$$P(X < m(X)) = P(X > m(X)) = ½$$

+ *Ví dụ*. Cho X tuân theo luật phân phối đều trên [a,b]. Khi đó mọi điểm $x \in [a,b]$ là *mode* của X và trung vị

$$m(X) = (a+b)/2$$

c) *Phương sai và độ lệch quân phương*.

• *Định nghĩa*. Cho biến ngẫu nhiên X với hàm mật độ f(t) có kỳ vọng E(X) = a.

Phương sai của X và ký hiệu là D(X) là

$$D(X) = E(X-a)^2 = \int_{-\infty}^{\infty} (t-a)^2 . f(t)dt$$

nếu tích phân hội tụ.

Nếu X có phương sai D(X) thì đại lượng

$$\sigma(X) = \sqrt{D(X)}$$

gọi là *độ lệch quân phương* của biến ngẫu nhiên X.

• *Công thức Koenig−Huyghens*.

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

CM.

$$D(X) = E(X-a)^2 = E(X^2 - 2.a.X + a^2)$$

$$= E(X^2) - 2E(X).E(X) + E(X)^2 = E(X^2) - [E(X)]^2$$

+ *Ví dụ*. Cho X tuân theo luật phân phối đều trên [a,b]. Ta có

$$D(X) = \int_a^b \frac{t^2}{b-a} dt - E(X)^2 = \left[\frac{1}{b-a} \frac{t^3}{3} \right]_{t=a}^{b} - \left(\frac{a+b}{2} \right)^2 = \frac{b^3 - a^3}{3(b-a)} - \frac{a^2 + 2ab + b^2}{4} = \frac{(b-a)^2}{12}$$

Suy ra

$$\sigma(X)=\sqrt{D(X)}=\frac{b-a}{2\sqrt{3}}$$

• *Tính chất*. Cho X, Y là các biến ngẫu nhiên liên tục, C là hằng. Khi đó
(i) D(C) = 0
(ii) D(C.X) = C^2.D(X)
(iii) σ(C.X) = |C|.σ(X)
(iv) Nếu X, Y là các biến ngẫu nhiên độc lập, thì

$$D(X + Y) = D(X) + D(Y)$$

d) Mômen

Cho biến ngẫu nhiên liên tục X với hàm mật độ f(t), a ∈ R, k ∈ N.

• *Mômen cấp k đối với a* của X là giá trị

$$\nu_k(a) = E[(X-a)^k] = \int_{-\infty}^{\infty}(t-a)^k.f(t)dt$$

• *Mômen gốc cấp k* của X là giá trị

$$\nu_k = \nu_k(0) = E(X^k)$$

➯ *Mômen gốc cấp 1* của X là kỳ vọng E(X) (nếu tồn tại).

• *Mômen trung tâm cấp k* của X, với điều kiện X có kỳ vọng, là giá trị

$$\mu_k = \nu_k(E(X)) = E[(X-E(X))^k]$$

➯ *Mômen trung tâm cấp 2* của X là phương sai D(X) (nếu tồn tại).

• *Định lý*. Cho r, s ∈ N , r ≤ s, và biến ngẫu nhiên liên tục X. Khi đó, nếu tồn tại mômen gốc cấp s của X, thì tồn tại mômen gốc cấp r của X.

CM

Với |t| ≤ 1, ta có $|t|^r \le 1$, và với |t| > 1, ta có $|t|^r \le |t|^s$. Suy ra

$$|t|^r \le 1 + |t|^s \quad \forall t\in R.$$

Từ đó sự hội tụ của tích phân $\int_{-\infty}^{\infty} f(t)dt$ và $\int_{-\infty}^{\infty}|t|^s\, f(t)dt$ kéo theo sự hội tu của tích phân $\int_{-\infty}^{\infty}|t|^r\, f(t)dt = \nu_r$

• *Hệ quả*. Cho r, s ∈ N , r ≤ s, và biến ngẫu nhiên liên tục X. Khi đó, nếu tồn tại mômen gốc cấp s của X, thì tồn tại mômen cấp r tại a ∈ R bất kỳ của X.

CM

Sử dụng định lý và đẳng thức sau

$$\gamma_k(a) = \mathrm{E}[(\mathrm{X - a})^{\mathrm{r}}] = \mathrm{E}\left[\sum_{i=0}^{\mathrm{n}} \mathrm{C(n,k).(\text{-}1)}^{\mathrm{r\text{-}i}}.a^{r-i}x^i\right] = \sum_{i=0}^{\mathrm{r}} \mathrm{C(n,k).(\text{-}1)}^{\mathrm{r\text{-}i}}.a^{r-i}.\mathrm{E(X^i)}$$

• *Mômen trung tâm cấp 3*.

μ_3 $= E[X - E(X)]^3 = E[X^3 - 3.X^2.E(X) + 3.X.E(X)^2 - E(X)^3] =$
$= \nu_3 - 3.\nu_1\nu_2 + 3.\nu_1^3 - \nu_1^3 = \nu_3 - 3.\nu_1\nu_2 + 2.\nu_1^3$

Bây giờ ta giả thiết rằng f(t) đối xứng qua E(X). Khi đó mọi mômen trung tâm cấp lẻ đều bằng 0.

• *Hệ số bất đối xứng*. Đại lượng

$$s_3 = \frac{\mu_3}{\sigma^3}$$

đặc trưng cho tính bất đối xứng của X, gọi là *hệ số bất đối xứng* của biến ngẫu nhiên X.

• *Mômen trung tâm cấp 4*.

μ_3 $= E[X - E(X)]^4 = E[X^4 - 4.X^3.E(X) + 6.X^2.E(X)^2 - 4.X.E(X)^3 + E(X)^4] =$
$= \nu_4 - 4.\nu_3\nu_1 + 6.\nu_1^2.\nu_2 - 4.\nu_1^3.\nu_1 + \nu_1^4 = \nu_4 - 4.\nu_3\nu_1 + 6.\nu_1^2.\nu_2 - 3.\nu_1^4$

• *Hệ số nhọn*. Đại lượng

$$N_4 = \frac{\mu_4}{\sigma^4} - 3$$

đặc trưng cho độ nhọn của đồ thị hàm mật độ f(t) so với phân phối chính qui, gọi là *hệ số nhọn* của biến ngẫu nhiên X.

Phân phối chính qui có độ nhọn $N_4 = 0$. Các đường cong nhọn hơn phân phối chính qui có $N_4 > 0$, và tù hơn phân phối chính qui có $N_4 < 0$.

3. Một số luật phân phối cơ bản

a) Luật phân phối chính qui (Laplace−Gauss) N(a,σ)

• *Định nghĩa*. biến ngẫu nhiên X gọi là tuân theo luật *phân phối chính qui*, hay *phân phối chuẩn*, N(a,σ), σ>0, nếu nó có hàm mật độ dạng

$$f(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t-a)^2}{2\sigma^2}} \quad \forall t \in R$$

↳ Ghi chú: Để chứng minh f(t) là hàm mật độ, ta sử dụng tích phân Euler

$$\int_{-\infty}^{\infty} e^{-\frac{t^2}{2}} dt = \sqrt{2\pi}$$

+ Kỳ vọng:

$$E(X) = \int_{-\infty}^{\infty} t\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(t-a)^2}{2\sigma^2}}dt$$

$$= \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty}\sigma(a+\sigma u)e^{-\frac{u^2}{2}}du$$ (đổi biến u=(t−a)/σ)

$$= \frac{a}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-\frac{u^2}{2}}du + \frac{\sigma}{\sqrt{2\pi}}\int_{-\infty}^{\infty}u.e^{-\frac{u^2}{2}}du$$

$$= \frac{a}{\sqrt{2\pi}}\sqrt{2\pi} + \frac{\sigma}{\sqrt{2\pi}}.0 = a$$

+ Phương sai: Ta có

$$E(X^2) = \int_{-\infty}^{\infty} t^2\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(t-a)^2}{2\sigma^2}}dt$$

$$= \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty}\sigma(a+\sigma.u)^2e^{-\frac{u^2}{2}}du$$ (đổi biến u=(t−a)/σ)

$$= \frac{a^2}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-\frac{u^2}{2}}du + \frac{2.a.\sigma}{\sqrt{2\pi}}\int_{-\infty}^{\infty}u.e^{-\frac{u^2}{2}}du + \frac{\sigma^2}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-\frac{u^2}{2}}du$$

$$= \frac{a^2}{\sqrt{2\pi}}\sqrt{2\pi} + \frac{2.a.\sigma}{\sqrt{2\pi}}.0 + \frac{\sigma^2}{\sqrt{2\pi}}\sqrt{2\pi} = a^2 + \sigma^2$$

Suy ra

$$D(X) = E(X^2) - E(X)^2 = a^2 + \sigma^2 - a^2 = \sigma^2$$

+ Độ lệch quân phương:

$$\sigma(X) = \sqrt{D(X)} = \sigma$$

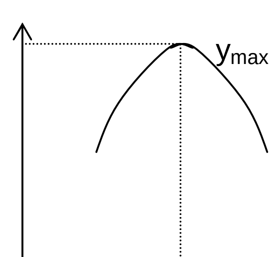
+ Mode:

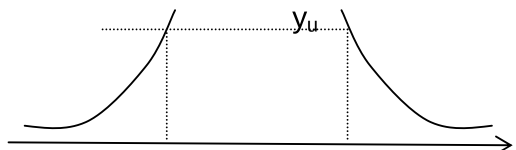
$$M(X) = a$$

+ Trung vị:

$$m(X) = a$$

+ Đồ thị của hàm f(t) gọi là đường cong phân phối chính qui. Đường cong có dạng hình chuông đối xứng qua đường x=a, đạt cực đại $y_{max}=\frac{1}{\sigma\sqrt{2\pi}}$ tại x=a, đạt giá trị $y_u = \frac{1}{\sigma\sqrt{2\pi e}}$ tại 2 điểm uốn tại x=a−σ và x=a+σ.





a−σ a

a+σ x

Khi a thay đổi, đường cong dịch chuyển song song với trục Ox còn dạng thì giữ nguyên. Khi σ tăng thì đường cong dẹp xuống, khi σ giảm thì đường cong lồi lên. Điều này hoàn toàn phù hợp với ý nghĩa của tham số σ là độ lệch quân phương, thước đo độ tản mát của giá trị của biến ngẫu nhiên.

• *Định lý 1*. Mọi biến ngẫu nhiên X có phân phối chính qui N(a,σ) có thể biểu diễn dưới dạng

$$X = a + \sigma.Z$$

trong đó Z là biến ngẫu nhiên có phân phối chuẩn N(0,1).

CM.

Đặt Z = (X − a)/σ. Hiển nhiên Z là biến ngẫu nhiên và hàm phân phối Φ(z), z∈ R, có dạng

$$\Phi(z) = P(Z \le z) = P(X \le a + \sigma.z) = \int_{-\infty}^{a+\sigma.z} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t-a)^2}{2\sigma^2}} dt$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{u^2}{2}} du$$ (đổi biến u=(t−a)/σ)

Suy ra hàm mật độ của Z là

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

Vậy Z tuân theo luật phân phối chuẩn N(0,1).

đpcm.

• *Định lý 2*. Biến ngẫu nhiên Z có phân phối chuẩn N(0,1) khi và chỉ khi biến ngẫu nhiên −Z có phân phối chuẩn N(0,1).

CM.

Giả sử Z có phân phối chuẩn N(0,1) với hàm phân phối Φ(z) và hàm mật độ φ(z).

Ta tính hàm phân phối F(z), z ∈ R, của biến ngẫu nhiên −Z

$$F(z) = P(-Z \le z) = P(Z \ge -z) = 1 - \Phi(-z)$$

Suy ra hàm mật độ f(z) của −Z là

$$f(z) = F'(z) = [1 - \Phi(-z)]' = \varphi(-z)$$

Mặt khác, φ(t) là hàm đối xứng qua x=0, tức là hàm chẵn. Vì vậy ta có

$$f(z) = \varphi(z) \qquad \forall z \in R,$$

tức −Z có phân phối chuẩn N(0,1).

đpcm.

• *Định lý 3*. Cho biến ngẫu nhiên X có phân phối chính qui N(a,σ). Khi đó với mọi A,B ∈ R (B≠0), biến ngẫu nhiên

$$Y = B.X + A$$

có phân phối chuẩn N(B.a + A, |B|.σ).

CM.

Theo định lý 1 ta có thể viết X = σ.Z + a, với Z có phân phối chuẩn N(0,1). Suy ra

$$Y = B.X + A = B.\sigma.Z + (B.a + A)$$

Trường hợp B > 0: theo định lý 1, Y có phân phối chuẩn N(B.a + A, B.σ).

Trường hợp B < 0: Ta đặt

$$Y = |B|.\sigma.(-Z) + (B.a + A)$$

Theo định lý 2, −Z có phân phối chuẩn N(0,1) và theo định lý 1, Y có phân phối chuẩn N(B.a + A, |B|.σ).

đpcm

• *Định lý 4*. Cho biến ngẫu nhiên X có phân phối chính qui N(a,σ). Khi đó với mọi α,β∈R, α<β, xác suất X rơi vào khoảng (α,β) là

$$P(\alpha < X < \beta) = \phi\left(\frac{\beta - a}{\sigma}\right) - \phi\left(\frac{\alpha - a}{\sigma}\right)$$

Trong đó

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{u^2}{2}} du \qquad \forall z \in R$$

CM. Suy ra từ định lý 1.

• *Phương pháp tra bảng tính xác suất của phân phối chuẩn N(0,1)*

Cho X là biến ngẫu nhiên có luật phân phối chuẩn N(0,1). Khi đó X có

- Hàm mật độ:

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \qquad \forall z \in R$$

- Hàm phân phối:

$$\Phi(z) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z} e^{-\frac{u^2}{2}}du \qquad \forall z \in R$$

do $\varphi(z)$ là hàm chẵn nên ta có

$$\Phi(-z) = 1 - \Phi(z) \qquad \forall z > 0$$

- Hàm Laplace:

$$\Phi_L(z) = \frac{1}{\sqrt{2\pi}}\int_{0}^{z} e^{-\frac{u^2}{2}}du \qquad \forall z \in R^+$$

✎ Lưu ý: Ta có

$$\Phi(z) = 0.5 + \Phi_L(z) \quad \text{và} \quad \Phi_L(z) = \Phi(z) - 0.5 \quad \forall z \geq 0$$

(i) Cho x, tra bảng $\Phi(x)$:

+ Trường hợp $x \geq 0$:

− Nếu $\exists x_0 \in$ Bảng, $x = x_0$, thì ta có ngay

$$\Phi(x) = \Phi(x_0)$$

− Nếu x không có trong Bảng, thì ta tìm cận x_1, x_2 trong bảng gần x nhất,

$$x_1 < x < x_2$$

và tính $\Phi(x)$ theo công thức nội suy tuyến tính

$$\Phi(x) = \Phi(x_1) + [\Phi(x_2) - \Phi(x_1)]\frac{x - x_1}{x_2 - x_1}$$

✎ Lưu ý: Nếu chỉ có bảng giá trị hàm Laplace $\Phi_L(x)$, thì ta tra hàm $\Phi_L(x)$, sau đó đặt $\Phi(x) = \Phi_L(x) + 0.5$.

+ Trường hợp $x < 0$:

Theo trên ta tra bảng tìm $\Phi(-x)$. Sau đó ta có $\Phi(x) = 1 - \Phi(-x)$.

(ii) Cho $y \in (0, 1)$, tra bảng tìm x thoả $\Phi(x) = y$.

+ Trường hợp $y \geq 0.5$:

− Nếu $\exists x_0 \in$ Bảng, $\Phi(x_0) = y$, thì ta có ngay $x = x_0$.

− Nếu y không có trong bảng, thì ta tìm cận y_1, y_2 trong bảng gần y nhất,

$$y_1 < y < y_2$$

và có $\Phi(x_1) = y_1$, $\Phi(x_2) = y_2$. Ta tính x theo công thức nội suy tuyến tính

$$x = x_1 + (x_2 - x_1)\frac{y - y_1}{y_2 - y_1}$$

✎ Lưu ý: Nếu chỉ có bảng giá trị hàm Laplace Φ_L, thì ta có

$$x = \Phi_L^{-1}(y - 0.5)$$

+ Trường hợp $y < 0.5$:
Tra bảng tính $z = \Phi^{-1}(1 - y)$ như trên, sau đó đặt $x = -z$.

+ *Ví dụ*.
(i) Tính $\Phi(0.74)$: $\Phi(0.75)$; $\Phi(0.746)$.
Tra bảng ta có

$$\Phi(0.74) = 0.7704;\ \Phi(0.75) = 0.7734;$$

Vậy

$$\Phi(0.746) = 0.7704 + (0.7734 - 0.7704).\frac{0.746 - 0.74}{0.75 - 0.74} = 0.7704 + 0.0018 = 0.7722$$

(ii) Tính $\Phi^{-1}(0.9128)$; $\Phi^{-1}(0.9115)$; $\Phi^{-1}(0.9131)$;
Tra bảng ta có

$$\Phi^{-1}(0.9115) = 1.35;\ \Phi^{-1}(0.9131) = 1.36;$$

Vậy

$$\Phi^{-1}(0.9128) = 1.35 + (1.36 - 1.35).\frac{0.9128 - 0.9115}{0.9131 - 0.9115} = 1.35 + 0.008 = 1.358$$

• *Mệnh* đề. Cho X là biến ngẫu nhiên có luật phân phối chuẩn $N(a,\sigma)$. Với $\alpha > 0$, ta có xác suất

$$P(|X - a| < \alpha) = 2.\Phi(\alpha/\sigma) - 1 = 2.\Phi_L(\alpha/\sigma)$$

CM.
Ta có

$$P(|X - a| < \alpha) = P(-\alpha/\sigma < (X - a)/\sigma < \alpha/\sigma)$$

$$= \Phi(\alpha/\sigma) - \Phi(-\alpha/\sigma) = 2.\Phi(\alpha/\sigma) - 1 = 2.\Phi_L(\alpha/\sigma)$$

+ *Ví dụ*. Người ta tiện một loại chi tiết có độ dài a. Biết độ dài X của chi tiết tuân theo luật phân phối $N(a,\sigma)$ với $\sigma = 0.2$cm.
(i) Hãy tính xác suất để độ dài chi tiết không lệch quá *a* dung sai là 0.3cm.
(ii) Muốn đảm bảo tỉ lệ phế phẩm không quá 5% thì phải chọn dung sai α bằng bao nhiêu ?
Giải.
(i) Xác suất cần tính là

$$P(|X - a| < 0.3) = 2.\Phi_L(0.3/0.2) = 2.\Phi_L(1.5) = 2 \times 0.4332 = 0.8664$$

Như vậy tỉ lệ phế phẩm là $1 - 0.8664 = 13\%$.

(ii) Dung sai α phải thoả

$$P(|X - a| < \alpha) \geq 0.95$$

$$\Leftrightarrow \quad 2.\Phi_L(\alpha/0.2) \geq 0.95 \quad \Leftrightarrow \quad \Phi_L(\alpha/0.2) \geq 0.475 \quad \Leftrightarrow \quad \alpha \geq 0.2 * \Phi_L^{-1}(0.475)$$

Tra bảng ta có $\Phi_L^{-1}(0.475) = 1.96$. Vậy

$$\alpha \geq 0.2 * 1.96 \quad \Leftrightarrow \quad \alpha \geq 0.392$$

• *Qui tắc 3σ.*

Trong công thức

$$P(|X - a| < \alpha) = 2.\Phi(\alpha/\sigma) - 1 = 2.\Phi_L(\alpha/\sigma)$$

nếu chọn α = 3σ thì ta có

$$P(|X - a| < 3\sigma) = 2.\Phi_L(3) = 0.9973$$

Xác suất này rất gần 1 nên có thể coi sự kiện |X − a| < 3.σ là hầu như chắc chắn. Vậy ta có qui tắc 3σ:

"*Nếu X có phân phối chuẩn N(a,σ) thì hầu như chắc chắn X lấy trị số trong khoảng (a − 3σ, a + 3σ)*"

+ *Ví dụ*. Khi đo lực chịu nén của một loại xà người ta thấy lực chịu bình quân là 320kg, sai số quân phương là 5kg. Hỏi muốn đảm bảo an toàn thì tải trọng tối đa cho phép đặt lên xà là bao nhiêu ?

Giải.

Gọi X là lực chịu nén của xà thì hầu như chắc chắn

$$|X - 320| < 3 \times 5 = 15 \text{ (kg)} \Leftrightarrow \quad 305 \text{ (kg)} < X < 335 \text{ (kg)}$$

Vậy muốn đảm bảo an toàn thì tải trọng tối đa cho phép đặt lên xà là 305kg.

b) *Luật phân phối mũ* E(λ).

Trong nhiều ứng dụng, đặc biệt trong lý thuyết phục vụ đám đông, người ta thường gặp các biến ngẫu nhiên tuân theo luật phân phối mũ.

• *Định nghĩa*. biến ngẫu nhiên X gọi là tuân theo luật phân phối mũ E(λ), λ > 0, nếu nó có hàm mật độ

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

+ Hàm phân phối:

$$F(x) = \begin{cases} 1 - e^{-\lambda.x} & \text{với } x \geq 0 \\ 0 & \text{với } x < 0 \end{cases}$$

+ Kỳ vọng và phương sai :

$$E(X) = \frac{1}{\lambda} \quad \text{và} \quad D(X) = \frac{1}{\lambda^2},\ \sigma(X) = \frac{1}{\lambda}$$

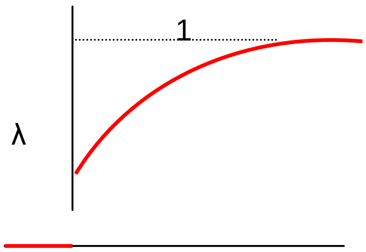
CM.

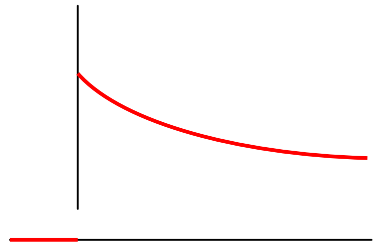
$$E(X) = \int_0^\infty t.\lambda.e^{-\lambda.t}dt = \left[-t.e^{-\lambda.t}\right]_{t=0}^\infty + \int_0^\infty e^{-\lambda.t}dt = \left[-\frac{1}{\lambda}e^{-\lambda.t}\right]_{t=0}^\infty = \frac{1}{\lambda}$$

$$E(X^2) = \int_0^\infty t^2.\lambda.e^{-\lambda.t}dt = \left[-t^2.e^{-\lambda.t}\right]_{t=0}^\infty + 2\int_0^\infty t.e^{-\lambda.t}dt = 0 + \frac{2}{\lambda}E(X) = \frac{2}{\lambda^2}$$

$\Rightarrow$ $D(X) = E(X^2) - E(X)^2 = \dfrac{1}{\lambda^2}$

+ Đồ thị:





hàm phân phối F(x) hàm mật độ f(t)

+ *Ví dụ*. Thời gian hoạt động T của một loại bóng đèn điện tử tuân theo luật phân phối mũ E(λ) với kỳ vọng E(T) = 400 giờ. Tính xác suất sao cho thời gian hoạt động của bóng đèn không dưới 600 giờ.

Giải. Ta có

$$E(T) = 400 = \frac{1}{\lambda} \quad \Rightarrow \quad \lambda = 1/400$$

Suy ra

$$P(T \geq 600) = 1 - P(T \leq 600) = 1 - F(600) = 1 - (1 - e^{-600/400}) = e^{-3/2} = 0.2231$$

➷ Ghi chú: Phân phối mũ là trường hợp riêng của phân phối Vâybun V(λ, n), λ > 0 và n ∈ **N**:

$$\begin{cases} f(x) = n.\lambda.x^{n-1}.e^{-\lambda.x} & \text{với } x \geq 0 \\ 0 & \text{với } x < 0 \end{cases}$$

CHƯƠNG 3
BIẾN NGẪU NHIÊN NHIỀU CHIỀU

I. BIẾN NGẪU NHIÊN 2 CHIỀU

1. Khái niệm.

Từ trước đến nay ta mới chỉ xét đến các biến ngẫu nhiên 1 chiều, tức là các biến ngẫu nhiên có các giá trị thực. Tuy nhiên trong thực tế chúng ta phải xét đến các hệ thống có niều biến ngẫu nhiên.

+ *Ví dụ 1*. Một thùng có a quả cầu trắng và b quả cầu đen. Rút liên tiếp 2 quả cầu (không trả lại vào thùng). Ký hiệu X_i, i=1,2, là các biến ngẫu nhiên nhận giá trị 1 nếu lần rút i được quả cầu trắng, nhận giá trị 0 nếu lần rút i được quả cầu đen. Khi đó cặp (X_1, X_2) tạo thành biến ngẫu nhiên 2 chiều.

+ *Ví dụ 2*. Gọi X, Y, Z tương ứng là chiều dài , chiều rộng, chiều cao của một loại thùng hình hộp chữ nhật. Nếu X, Y, Z là các biến ngẫu nhiên thì bộ ba (X, Y, Z) tạo thành một biến ngẫu nhiên ba chiều.

• *Định nghĩa*. Cho không gian xác suất (Ω, **B**, P), $n \in \mathbf{N}$, n > 1. Ánh xạ

$$X: \Omega \to R^n,\ X(\omega) = (X_1(\omega), X_2(\omega), \dots, X_n(\omega))$$

gọi là *biến ngẫu nhiên n chiều* hay *vectơ ngẫu nhiên n chiều* , nếu $X_1, X_2, \dots, X_n$ là các biến ngẫu nhiên trên không gian (Ω, **B**, P). Ký hiệu

$$X = (X_1, X_2, \dots, X_n)$$

Để đơn giản ta chỉ cần xét kỹ các biến ngẫu nhiên 2 chiều, sau đó mở rộng cho biến ngẫu nhiên n chiều.

2. Luật phân phối của cặp biến ngẫu nhiên rời rạc.

• *Định nghĩa*. Cho biến ngẫu nhiên Z = (X, Y), trong đó (X, Y) nhận các cặp giá trị

$$(x_i, y_j),\ i \in I,\ j \in J$$

với I, J là các tập chỉ số dạng hữu hạn {1,2,…,n} hoặc vô hạn {1,2,…, ∞}.
Khi đó luật phân phối của Z là dãy

$$\{[(x_i, y_j), p_{ij}] \mid i \in I, j \in J\}$$

trong đó

$$p_{ij} = P(X = x_i, Y = y_j),\ i \in I, j \in J$$

Nếu Z hữu hạn, tức I = {1, 2, …, m} và J = {1, 2, … , n} , ta có thể biểu diễn luật phân phối của Z dạng bảng như sau:

Y / X	y_1	y_2	…	y_j	…	y_n
x_1	p_{11}	p_{12}	…	p_{1j}	…	p_{1n}
x_2	p_{21}	p_{22}	…	p_{2j}	…	p_{2n}
:	:	:	…	:	…	:
x_i	p_{i1}	p_{i2}	…	p_{ij}	…	p_{in}
:	:	:	…	:	…	:
x_m	p_{m1}	p_{m2}	…	p_{mj}	…	p_{mn}

• *Định lý*.

(i) $\sum_{i\in I}\sum_{j\in J} p_{ij} = 1$

(ii) Biến ngẫu nhiên rời rạc X có luật phân phối

$$\{(x_i, p_i) \mid i \in I\}$$

trong đó

$$p_i = \sum_{j\in J} p_{ij} \qquad \forall\, i \in I$$

(iii) Biến ngẫu nhiên rời rạc Y có luật phân phối

$$\{(y_j, q_j) \mid j \in J\}$$

trong đó

$$q_j = \sum_{i\in I} p_{ij} \qquad \forall\, j \in J$$

CM.

(i) Tập

$$\{(x_i, y_j),\ i \in I,\ j \in J\}$$

tạo thành nhóm đầy đủ các sự kiện, vì vậy

$$\sum_{i\in I}\sum_{j\in J} p_{ij} = 1$$

(ii) Ta có

$$p_i = P(X = x_i) = \sum_{j\in J} P(X = x_i, Y = y_j) \quad = \quad \sum_{j\in J} p_{ij} \quad \forall\, i \in I$$

(iii) Ta có

$$q_j = P(Y = y_j) = \sum_{i\in I} P(X = x_i, Y = y_j) \quad = \sum_{i\in I} p_{ij} \quad \forall\, j \in J$$

\+ *Ví dụ 1*. Trở lại ví dụ 1 mục 1. Một thùng có a quả cầu trắng và b quả cầu đen. Rút liên tiếp 2 quả cầu (không trả lại vào thùng). Ký hiệu X_i, i=1,2, là các biến ngẫu nhiên nhận giá trị 1 nếu lần rút i được quả cầu trắng, nhận giá trị 0 nếu lần rút i được quả cầu đen. Khi đó cặp (X_1, X_2) tạo thành biến ngẫu nhiên 2 chiều.

$$p_{11} = P(X_1 = 1; X_2 = 1) = P(X_1 = 1).P(X_2 = 1/X_1 = 1) = \frac{a}{a+b}\frac{a-1}{a+b-1}$$

$$p_{10} = P(X_1 = 1; X_2 = 0) = P(X_1 = 1).P(X_2 = 0/X_1 = 1) = \frac{a}{a+b}\frac{b}{a+b-1}$$

$$p_{01} = P(X_1 = 0; X_2 = 1) = P(X_1 = 0).P(X_2 = 1/X_1 = 0) = \frac{b}{a+b}\frac{a}{a+b-1}$$

$$p_{00} = P(X_1 = 0; X_2 = 0) = P(X_1 = 0).P(X_2 = 0/X_1 = 0) = \frac{b}{a+b}\frac{b-1}{a+b-1}$$

$\Rightarrow$ Bảng phân phối của vectơ ngẫu nhiên (X_1, X_2) có dạng

X_1 \ X_2	1	0
1	$\frac{a}{a+b}\frac{a-1}{a+b-1}$	$\frac{a}{a+b}\frac{b}{a+b-1}$
0	$\frac{b}{a+b}\frac{a}{a+b-1}$	$\frac{b}{a+b}\frac{b-1}{a+b-1}$

Từ đó suy ra luật phân phối của X_1 và X_2

$$p_1 = P(X_1 = 1) = p_{11} + p_{10} = \frac{a}{a+b}$$

$$p_0 = P(X_1 = 0) = p_{01} + p_{00} = \frac{b}{a+b}$$

$$q_1 = P(X_2 = 1) = p_{11} + p_{01} = \frac{a}{a+b}$$

$$q_0 = P(X_2 = 0) = p_{10} + p_{00} = \frac{b}{a+b}$$

+ *Ví dụ 2*. Một thùng có n quả cầu đánh số từ 1 đến n , $n \in$ **N**, $n > 0$. Rút liên tiếp (có trả lại) 2 quả cầu. Gọi X là số nhỏ, Y là số lớn trong 2 số rút được. Xác định luật phân phối của vectơ ngẫu nhiên (X,Y), các biến ngẫu nhiên X và Y.

Giải.

Các cặp giá trị của (X,Y) là $\{(i,j) \mid i,j = 1,2,\dots,n\}$. Ta có

- Trường hợp $i > j$: $P(X = i; Y = j) = 0$
- Trường hợp $i = j$: $P(X = i; Y = j) = 1/n^2$
- Trường hợp $i < j$: Gọi

A_i , A_j tương ứng là sự kiện lần rút thứ nhất được số i, số j

B_i , B_j tương ứng là sự kiện lần rút thứ hai được số i, số j

Ta có

$$P(X = i; Y = j) = P(A_i.B_j) + P(A_j.B_i) = 1/n^2 + 1/n^2 = 2/n^2$$

Suy ra bảng phân phối xác suất của (X,Y) như sau

X \ Y	1	2	…	i	…	n
1	$1/n^2$	$2/n^2$	…	$2/n^2$	…	$2/n^2$
2	0	$1/n^2$	…	$2/n^2$	…	$2/n^2$
:	:	:	…	:	…	:
i	0	0	…	$1/n^2$	…	$2/n^2$
:	:	:	…	:	…	:
n	0	0	…	0	…	$1/n^2$

Từ đó suy ra luật phân phối xác suất của X và Y như sau:

$$p_i = P(X = i) = \frac{2(n-i)+1}{n^2} \quad \forall\ i=1, 2, \ldots, n$$

$$q_i = P(Y = i) = \frac{2(i-1)+1}{n^2} = \frac{2i-1}{n^2} \quad \forall\ i=1, 2, \ldots, n$$

3. Hàm phân phối xác suất

• *Định nghĩa*. Hàm phân phối xác suất của vectơ ngẫu nhiên (X,Y) là hàm 2 biến

$$F(x,y) = P(X \le x, Y \le y) \qquad \forall\ (x,y) \in R^2$$

• *Tính chất*.

(i) Ký hiệu F_X và F_Y tương ứng là các hàm phân phối của biến ngẫu nhiên X và Y. Ta có

$$F_X(x) = \lim_{y\to+\infty} F(x,y) = F(x, +\infty) \qquad \forall\ x \in R$$

$$F_Y(y) = \lim_{x\to+\infty} F(x,y) = F(+\infty, y) \qquad \forall\ y \in R$$

(ii)

$$\lim_{\substack{x\to+\infty \\ y\to+\infty}} F(x,y) = 1$$

(iii)

$$\lim_{x\to-\infty} F(x,y) = \lim_{y\to-\infty} F(x,y) = \lim_{\substack{x\to-\infty \\ y\to-\infty}} F(x,y) = 0$$

(iv) Hàm F(x,y) không giảm theo từng đối số

$$F(x_1, y_1) \le F(x_2, y_2)\ \forall\ (x_1, y_1), (x_2, y_2) \in R^2, x_1 \le x_2\ \&\ y_1 \le y_2$$

(v) Cho a, b, c, d $\in$ R, a < b, c < d. Ta có

$$P(a < X \leq b; c < Y \leq d) = F(b, d) - F(a, d) - F(b, c) + F(a, c)$$

CM.

Các tính chất (i) − (iv) suy ra từ định nghĩa.
Để chứng minh tính chất (v) ta biểu diễn sự kiện

$$(a < X \leq b; c < Y \leq d) = (X \leq b ; Y \leq d) - (X \leq a ; Y \leq d) \cup (X \leq b ; Y \leq c)$$

và từ đó, áp dụng công thức tính xác suất của sự kiện tổng, suy ra

$$\begin{aligned} P(a < X \leq b; c < Y \leq d) &= F(b, d) - [F(a, d) + F(b, c) - F(a, c)] \\ &= F(b, d) - F(a, d) - F(b, c) + F(a, c) \end{aligned}$$

đpcm

4. Hàm mật độ phân phối xác suất

• *Định nghĩa*. Hàm mật độ phân phối xác suất của vectơ ngẫu nhiên liên tục (X,Y) là đạo hàm hỗn hợp cấp 2 của hàm phân phối F(x,y)

$$f(x,y) = \frac{\partial^2 F(x,y)}{\partial x \partial y} \qquad \forall (x,y) \in R^2$$

• *Tính chất*.

(i) Hàm mật độ không âm

$$f(x,y) \geq 0 \qquad \forall (x,y) \in R^2$$

(ii) Hàm phân phối

$$F(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(u,v)dudv \qquad \forall (x,y) \in R^2$$

(iii) Tích phân kép

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)dxdy = 1$$

(iv) Với mọi miền D $\subset$ R^2, xác suất

$$P((x,y) \in D) = \iint_D f(x,y)dxdy$$

CM.

(i) hiển nhiên
(ii) Đặt

$$G(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(u,v)dudv \qquad \forall (x,y) \in R^2$$

Suy ra

$$\frac{\partial^2 G(x,y)}{\partial x \partial y} = f(x,y) \qquad \forall (x,y) \in R^2$$

và

$$\lim_{\substack{x\to-\infty \\ y\to-\infty}} G(x,y) = 0$$

Vậy

$$F(x,y) = G(x,y) \qquad \forall\ (x,y) \in R^2$$

(iii) suy ra từ (ii).
(iv) Nếu D là miền chữ nhật thì (iv) suy ra từ (ii). Nếu D là miền bất kỳ ta chia D thành các hình chữ nhật nhỏ, và bằng phương pháp giới hạn ta suy ra (iv).

+ *Ví dụ*. Cho vectơ ngẫu nhiên (X,Y) có mật độ

$$f(x,y) = \frac{1}{\pi^2(1+x^2)(1+y^2)} \qquad \forall\ (x,y) \in R^2$$

Ta có

$$F(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} \frac{1}{\pi^2(1+u^2)(1+v^2)}dudv = \frac{\left(\arctan x + \frac{\pi}{2}\right)\left(\arctan y + \frac{\pi}{2}\right)}{\pi^2}$$

Cho miền chữ nhật

$$D = \{(x,y) \mid 1 < x < \sqrt{3}\ \&\ 0 < y < 1\}$$

ta tính xác suất (X,Y) rơi vào D:

$$P((x,y) \in D) = F(\sqrt{3},1) - F(1,1) - F(\sqrt{3},0) + F(1,0)$$

$$= \frac{1}{\pi^2}\left[\left(\frac{\pi}{3}+\frac{\pi}{2}\right)\left(\frac{\pi}{4}+\frac{\pi}{2}\right) - \left(\frac{\pi}{4}+\frac{\pi}{2}\right)^2 - \left(\frac{\pi}{3}+\frac{\pi}{2}\right)\frac{\pi}{2} + \left(\frac{\pi}{4}+\frac{\pi}{2}\right)\frac{\pi}{2}\right]$$

$$= \frac{1}{\pi^2}\left[\frac{\pi^2.5.3}{6.4} - \frac{\pi^2.9}{16} - \frac{\pi^2.5}{6.2} + \frac{\pi^2.3}{4`2}\right] = \frac{30-27-20+18}{48} = \frac{1}{48}$$

• Hàm mật độ của các toạ độ.

Cho hàm mật độ f(x,y) của vectơ ngẫu nhiên (X,Y). Ta xác định hàm mật độ f_X và f_Y của các biến ngẫu nhiên X và Y.

Từ mục 2, 3 ta suy ra các hàm phân phối F_X và F_Y của X và Y như sau

$$F_X(x) = F(x, +\infty) = \int_{-\infty}^{x}\left[\int_{-\infty}^{+\infty} f(u,v)dv\right]du$$

$$F_Y(y) = F(+\infty, y) = \int_{-\infty}^{y}\left[\int_{-\infty}^{+\infty} f(u,v)du\right]dv$$

Vậy

$$f_X(x) = F_X'(x) = \int_{-\infty}^{+\infty} f(x,y)dy$$

và

$$f_Y(y) = F_Y'(y) = \int_{-\infty}^{+\infty} f(x,y)dx$$

+ *Ví dụ*. Cho vectơ ngẫu nhiên (X,Y) có hàm mật độ sau

$$f(x,y) = \begin{cases} \dfrac{1}{6\pi} & , \dfrac{x^2}{9} + \dfrac{y^2}{4} \le 1 \\ 0 & , \dfrac{x^2}{9} + \dfrac{y^2}{4} > 1 \end{cases}$$

Tìm hàm mật độ f_X và f_Y của X và Y.
Giải.

Khi |x| ≤ 3 ta có

$$f_X(x) = \frac{1}{6\pi} \int_{-2\sqrt{1-\frac{x^2}{9}}}^{2\sqrt{1-\frac{x^2}{9}}} dy = \frac{2}{9\pi}\sqrt{9-x^2}$$

còn nếu |x| > 3 thì hiển nhiên $f_X(x) = 0$.

Vậy

$$f_X(x) = \begin{cases} \dfrac{2}{9\pi}\sqrt{9-x^2} & , |x| \le 3 \\ 0 & , |x| > 3 \end{cases}$$

Tương tự ta có

$$f_Y(y) = \begin{cases} \dfrac{1}{2\pi}\sqrt{4-y^2} & , |y| \le 2 \\ 0 & , |y| > 2 \end{cases}$$

5. Hiệp phương sai và hệ số tương quan.

Cho vectơ ngẫu nhiên (X,Y). Giả sử tồn tại kỳ vọng

a = E(X) và b = E(Y).

• *Hiệp phương sai* của X và Y , ký hiệu $\mu_{X,Y}$, là trị

$$\mu_{X,Y} = \mathrm{cov}(X,Y) = E[(X - a).(Y - b)]$$

nếu kỳ vọng ở vế phải tồn tại.

• Trường hợp (X,Y) là vectơ ngẫu nhiên rời rạc với luật phân phối:

$$\{[(x_i, y_j), p_{ij}] \mid i \in I, j \in J\}$$

ta có

$$\mu_{X,Y} = \sum_{i\in I}\sum_{j\in J}(x_i - a)(y_j - b)p_{ij}$$

• Trường hợp (X,Y) là vectơ ngẫu nhiên liên tục với hàm mật độ f(x,y) ta có

$$\mu_{X,Y} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x-a)(y-b)f(x,y)dxdy$$

➯ Ghi chú: xem bài Hàm biến ngẫu nhiên.

• Công thức Koenig−Huyghens:

$$\mu_{X,Y} = E(X.Y) - E(X).E(Y)$$

CM.

$$\mu_{X,Y} = E[(X-a).(Y-b)] = E(X.Y) - bE(X) - aE(Y) + a.b = E(X.Y) - E(X).E(Y)$$

• *Bổ đề*. Cho biến ngẫu nhiên X và Y có độ lệch quân phương σ_X và σ_Y . Khi đó hiệp phương sai $\mu_{X,Y}$ tồn tại và

$$|\mu_{X,Y}| \le \sigma_X.\sigma_Y$$

CM.

Sử dụng bất đẳng thức

$$x.y \le \frac{x^2+y^2}{2}$$

và các công thức tính hiệp phương sai ở trên ta suy ra sự tồn tại của $\mu_{X,Y}$.

Bây giờ ta xét biểu thức không âm

$$0 \le g(\lambda) = D(\lambda.X + Y) = cov(\lambda.X + Y, \lambda.X + Y) = \lambda^2\sigma_X^2 + 2.\lambda.\mu_{X,Y} + \sigma_Y^2$$

- Trường hợp $\sigma_X = 0$:

$$g(\lambda) = 2.\lambda.\mu_{X,Y} + \sigma_Y^2 \ge 0 \quad \forall\lambda \quad \Rightarrow \quad \mu_{X,Y} = 0 \quad \Rightarrow \quad |\mu_{X,Y}| \le \sigma_X.\sigma_Y$$

- Trường hợp $\sigma_X > 0$: Tam thức $g(\lambda) \ge 0$ $\forall\lambda$ nên

$$\Delta' = \mu_{X,Y}^2 - \sigma_X^2\sigma_Y^2 \le 0 \quad \Rightarrow \quad |\mu_{X,Y}| \le \sigma_X.\sigma_Y$$

• *Hệ số tương quan* của X và Y , ký hiệu $\rho_{X,Y}$, là trị

$$\rho_{X,Y} = \frac{\mu_{X,Y}}{\sigma_X\sigma_Y}$$

trong đó σ_X và σ_Y ($\neq 0$) là độ lệch quân phương của X và Y.

Từ bổ đề suy ra

$$|\rho_{X,Y}| \leq 1$$

✎ Hai biến ngẫu nhiên X và Y gọi là *tương quan* nếu $\rho_{X,Y} \neq 0$, ngược lại, nếu $\rho_{X,Y} = 0$ ta nói X và Y không tương quan.

Nếu $\rho_{X,Y} > 0$ thì biến này "tăng" kéo theo biến kia "tăng".

Nếu $\rho_{X,Y} < 0$ thì biến này "tăng" kéo theo biến kia "giảm".

✎ Ta xét trường hợp $\sigma_X \neq 0$ và $|\mu_{X,Y}| = \sigma_X.\sigma_Y$.

Khi đó, theo chứng minh bổ đề, $\Delta' = 0$, tức $g(\lambda)$ có nghiệm kép λ_0. Như vậy biến ngẫu nhiên $\lambda_0 X + Y$ có phương sai bằng 0. Từ đó suy ra $\lambda_0 X + Y = C$ (hằng) hầu như khắp nơi (trừ vùng có xác suất bằng 0). Vậy có thể viết

$$Y = -\lambda_0.X + C$$

tức Y là hàm tuyến tính của X.

Như vậy, $\sigma_X \neq 0$, $\sigma_Y \neq 0$ và $|\rho_{X,Y}| = 1$ thì ta nói X và Y *tương quan tuyến tính*.

+ *Ví dụ 1*. Trở lại ví dụ 1 mục 1. Một thùng có a quả cầu trắng và b quả cầu đen. Rút liên tiếp 2 quả cầu (không trả lại vào thùng). Ký hiệu X_i, i=1,2, là các biến ngẫu nhiên nhận giá trị 1 nếu lần rút i được quả cầu trắng, nhận giá trị 0 nếu lần rút i được quả cầu đen. Khi đó cặp (X_1, X_2) tạo thành biến ngẫu nhiên 2 chiều.

Ta có

$$E(X_1) = \frac{a}{a+b}\ ,\ E(X_2) = \frac{a}{a+b}\ ,\ E(X_1.X_2) = \frac{a(a-1)}{(a+b)(a+b-1)}$$

$$\Rightarrow \quad \mu_{X_1,X_2} = E(X_1.X_2) - E(X_1).E(X_2) = \frac{a(a-1)}{(a+b)(a+b-1)} - \left(\frac{a}{a+b}\right)^2$$

$$= \frac{a(a-1)(a+b) - a^2(a+b-1)}{(a+b)^2(a+b-1)} = -\frac{ab}{(a+b)^2(a+b-1)}$$

Mặt khác

$$D(X_1) = D(X_2) = \frac{ab}{(a+b)^2}$$

$$\Rightarrow \quad \rho_{X_1,X_2} = \frac{\mu_{X_1,X_2}}{\sigma_{X_1}.\sigma_{X_2}} = -\frac{1}{a+b-1}$$

+ *Ví dụ 2*. Luật phân phối chính qui hai chiều.

Vectơ ngẫu nhiên (X,Y) gọi là tuân theo luật phân phối chính qui 2 chiều nếu có hàm mật độ dạng

$$f(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}.e^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-a}{\sigma_X}\right)^2 + \left(\frac{y-b}{\sigma_Y}\right)^2 - 2.\rho\frac{(x-a)(y-b)}{\sigma_X\sigma_Y}\right]}$$

trong đó

a, b là kỳ vọng của X, Y

σ_X , σ_Y là độ lệch quân phương của X, Y

ρ là hệ số tương quan của X và Y

🖎 Ghi chú: Nếu ρ = 0 thì

$$f(x,y) = \frac{1}{\sqrt{2\pi}.\sigma_X}.e^{-\frac{1}{2}\left(\frac{x-a}{\sigma_X}\right)^2}.\frac{1}{\sqrt{2\pi}.\sigma_Y}.e^{-\frac{1}{2}\left(\frac{y-b}{\sigma_Y}\right)^2} = f_X(x).f_Y(y)$$

với f_X , f_Y là các hàm mật độ của X, Y có phân phối chuẩn $N(a,\sigma_X)$, $N(b,\sigma_Y)$.

II. PHÂN PHỐI CÓ ĐIỀU KIỆN - BIẾN NGẪU NHIÊN ĐỘC LẬP

1. Trường hợp vectơ ngẫu nhiên rời rạc.

Cho vectơ ngẫu nhiên rời rạc (X,Y) với luật phân phối

$$\{[(x_i, y_j), p_{ij}] \mid i \in I, j \in J\}$$

Với mọi x_i, y_j, $i \in I$ và $j \in J$, xác suất để $X = x_i$ với giả thiết $Y = y_j$ gọi là xác suất có điều kiện điều kiện $P(x_i / y_j)$.

Xác suất để $Y = y_j$ với giả thiết $X = x_i$ gọi là xác suất có điều kiện điều kiện $P(y_j / x_i)$.

Cố định y_j, $j \in J$, phân phối

$$\{(x_i, P(x_i / y_j)) \mid i \in I\}$$

gọi là *phân phối có điều kiện* của biến ngẫu nhiên X với điều kiện $Y = y_j$.

Cố định x_i, $i \in I$, phân phối

$$\{(y_j, P(y_j / x_i)) \mid j \in J\}$$

gọi là *phân phối có điều kiện* của biến ngẫu nhiên Y với điều kiện $X = x_i$.

Ta có

$$P(x_i / y_j) = \frac{P(X = x_i; Y = y_j)}{P(Y = y_j)} = \frac{p_{ij}}{\sum_{i\in I} p_{ij}}$$

$$P(y_j / x_i) = \frac{P(X = x_i; Y = y_j)}{P(X = x_i)} = \frac{p_{ij}}{\sum_{j\in J} p_{ij}}$$

➯ Ghi chú: Hiển nhiên ta có

$$\sum_{i\in I} P(x_i / y_j) = \frac{\sum_{i\in I} p_{ij}}{\sum_{i\in I} p_{ij}} = 1$$

$$\sum_{j\in J} P(y_j / x_i) = \frac{\sum_{j\in J} p_{ij}}{\sum_{j\in J} p_{ij}} = 1$$

• *Kỳ vọng có điều kiện*

Kỳ vọng có điều kiện của X với điều kiện $Y = y_j$, $j \in J$, là đại lượng

$$E(X / y_j) = \sum_{i\in I} x_i P(x_i / y_j)$$

Kỳ vọng có điều kiện của Y với điều kiện $X = x_i$, $i \in I$, là đại lượng

$$E(Y / x_i) = \sum_{j\in J} y_j P(y_j / x_i)$$

• *Biến ngẫu nhiên độc lập*.

Hai biến ngẫu nhiên X và Y gọi là độc lập với nhau nếu luật phân phối có điều kiện của mỗi biến trùng với luật phân phối (không điều kiện) của nó, tức là

$$P(x_i / y_j) = P(X = x_i) = p_i \qquad \forall i \in I, j \in J$$
$$P(y_j / x_i) = P(Y = y_j) = q_j \qquad \forall i \in I, j \in J$$

Từ định nghĩa ta thấy biến ngẫu nhiên X và Y độc lập khi và chỉ khi

$$p_{ij} = p_i.q_j \qquad \forall\, i \in I, j \in J$$

+ *Ví dụ 1*. Một thùng có a quả cầu trắng và b quả cầu đen. Rút ngẫu nhiên 1 quả cầu, trả lại vào thùng, sau đó rút 1 quả nữa. Ký hiệu X_i, i=1,2, là các biến ngẫu nhiên nhận giá trị 1 nếu lần rút i được quả cầu trắng, nhận giá trị 0 nếu lần rút i được quả cầu đen. Khi đó cặp (X_1, X_2) tạo thành biến ngẫu nhiên 2 chiều và ta dễ dàng thấy rằng X_1 và X_2 độc lập với nhau.

+ *Ví dụ 2*. Trở lại ví dụ 1 mục 1. Một thùng có a quả cầu trắng và b quả cầu đen. Rút liên tiếp 2 quả cầu (không trả lại vào thùng). Ký hiệu X_i, i=1,2, là các biến ngẫu nhiên nhận giá trị 1 nếu lần rút i được quả cầu trắng, nhận giá trị 0 nếu lần rút i được quả cầu đen. Khi đó cặp (X_1, X_2) tạo thành biến ngẫu nhiên 2 chiều.

Hãy xác định luật phân phối có điều kiện của X_2 đối với $X_1 = 1$.

Ta có

$$P(X_2 = 1 / X_1 = 1) = \frac{p_{11}}{p_1} = \frac{\dfrac{a(a-1)}{(a+b)(a+b-1)}}{\dfrac{a}{a+b}} = \frac{a-1}{a+b-1}$$

$$P(X_2 = 0 / X_1 = 1) = \frac{p_{01}}{p_1} = \frac{\dfrac{a.b}{(a+b)(a+b-1)}}{\dfrac{a}{a+b}} = \frac{b}{a+b-1}$$

Ta thấy phân phối có điều kiện của X_2 đối với $X_1 = 1$ khác phân phối không điều kiện của X_2. Vậy X_1 và X_2 không độc lập.

Kỳ vọng có điều kiện của X_2 đối vơi $X_1 = 1$ là

$$E(X_2 / X_1 = 1) = \frac{a-1}{a+b-1}$$

Trong khi đó

$$E(X_2) = \frac{a}{a+b}$$

2. Trường hợp vectơ ngẫu nhiên liên tục

Cho vectơ ngẫu nhiên liên tục (X, Y) có mật độ f(x,y).

Với mọi y $\in$ R thoả $f_Y(y) \neq 0$, hàm

$$f_X(x/y) = \frac{f(x,y)}{f_Y(y)} \qquad , x \in R,$$

gọi là hàm mật độ phân phối có điều kiện của X với điều kiện Y = y.

Tương tự, với mọi x $\in$ R thoả $f_X(x) \neq 0$, hàm

$$f_Y(y/x) = \frac{f(x,y)}{f_X(x)} \qquad , y \in R,$$

gọi là hàm mật độ phân phối có điều kiện của Y với điều kiện X = x.

Hiển nhiên ta có

$$f_X(x/y) \geq 0 \quad \forall x \in R \qquad , \int_{-\infty}^{+\infty} f_X(t/y)dt = 1$$

$$f_Y(y/x) \geq 0 \quad \forall y \in R \qquad , \int_{-\infty}^{+\infty} f_Y(t/x)dt = 1$$

• Biến ngẫu nhiên X và Y gọi là *độc lập* với nhau nếu

$$f_X(x/y) = f_X(x) \ \forall y \in R, \ f_Y(y) \neq 0, \ \forall x \in R$$

hoặc

$$f_Y(y/x) = f_Y(y) \ \forall x \in R, \ f_X(x) \neq 0, \ \forall y \in R$$

Suy ra X và Y gọi là *độc lập* với nhau nếu

$$f(x,y) = f_X(x).f_Y(y) \qquad \forall (x, y) \in R$$

• *Kỳ vọng có điều kiện* của X với điều kiện Y = y, $f_Y(y) \neq 0$, là giá trị

$$E(X / y) = \int_{-\infty}^{+\infty} t.f_X(t/y)dt$$

Và kỳ vọng có điều kiện của Y với điều kiện X = x, $f_X(x) \neq 0$, là giá trị

$$E(Y / x) = \int_{-\infty}^{+\infty} t.f_Y(t/x)dt$$

+ *Ví dụ*. Cho r > 0 và vectơ ngẫu nhiên (X,Y) có mật độ

$$f(x,y) = \begin{cases} \dfrac{1}{\pi.r^2} & , x^2 + y^2 \leq r^2 \\ 0 & , x^2 + y^2 > r^2 \end{cases}$$

Với |y| < r ta có

$$f_Y(y) = \int_{-\infty}^{+\infty} f(x,y)dx = \frac{1}{\pi.r^2}\int_{-\sqrt{r^2-y^2}}^{+\sqrt{r^2-y^2}} dx = \frac{2\sqrt{r^2-y^2}}{\pi.r^2}$$

Với $|y| \geq r$ ta có $f_Y(y) = 0$.

Vậy, với $|y| < r$, ta có

$$f_X(x/y) = \frac{f(x,y)}{f_Y(y)} = \begin{cases} \dfrac{1}{2.\sqrt{r^2-y^2}} & , x^2 \leq r^2 - y^2 \\ 0 & , x^2 > r^2 - y^2 \end{cases}$$

Suy ra $f_X(./y) \neq f_X(.)$, vậy X và Y không độc lập với nhau.
Vì hàm mật độ chẵn, nên kỳ vọng có điều kiện

$$E(X / y) = \int_{-\sqrt{r^2-y^2}}^{+\sqrt{r^2-y^2}} x.dx = 0$$

III. HÀM ĐỐI SỐ NGẪU NHIÊN

1. Hàm 1 đối số ngẫu nhiên.

$$Y = \varphi(X)$$

a) Trường hợp X là biến ngẫu nhiên rời rạc.

Cho X có luật phân phối

$$\{(x_i, p_i) \mid i \in I\}$$

Ký hiệu

$$\{y_j \mid j \in J\}$$

là tập hợp các trị khác nhau của $\varphi(x_i)$, $i \in I$.

Đặt

$$q_j = \sum_{\varphi(x_i)=y_j} p_i \quad \forall j \in J$$

Khi đó biến ngẫu nhiên $Y = \varphi(X)$ có luật phân phối

$$\{(y_j, q_j) \mid j \in J\}$$

+ *Ví dụ*.

Cho $n \in \mathbf{N}$, $n > 0$, X là biến ngẫu nhiên có phân phối

$$\left\{\left(\frac{k}{n}, \frac{1}{2n+1}\right) \middle| k = -n, -(n-1), \ldots, -1, 0, 1, \ldots, (n-1), n\right\}$$

Khi đó biến ngẫu nhiên $Y = |X|$ có phân phối

$$\left\{\left(\frac{k}{n}, p_k\right) \middle| k = 0, 1, \ldots, (n-1), n\right\}$$

với

$$p_0 = 1/(2n+1) \text{ và } p_k = 2/(2n+1) \qquad \forall k = 1, 2, \ldots, n$$

b) Trường hợp X là biến ngẫu nhiên liên tục.

Cho X có hàm mật độ f(x).

(i) Giả thiết $\varphi(x)$ khả vi đơn điệu tăng trong (a,b) chứa các giá trị của X. Ta có

$$F_Y(y) = P(Y \le y) = P(\varphi(X) \le y) = P(X \le \varphi^{-1}(y)) = F_X(\varphi^{-1}(y))$$

Suy ra

$$\frac{dF_Y(y)}{dy} = \frac{dF_X(\varphi^{-1}(y))}{dx} \cdot \frac{d\varphi^{-1}(y)}{dy} = f_X(\varphi^{-1}(y)) \cdot \frac{d\varphi^{-1}(y)}{dy}$$

$\Rightarrow$

$$f_Y(y) = f_X(\varphi^{-1}(y)) \cdot \frac{d\varphi^{-1}(y)}{dy}$$

(ii) Giả thiết $\varphi(x)$ khả vi, đơn điệu giảm trong (a,b) chứa các giá trị của X. Ta có

$F_Y(y) = P(Y \le y) = P(\varphi(X) \le y) = P(X \ge \varphi^{-1}(y)) = 1 - F_X(\varphi^{-1}(y))$

Suy ra

$$\frac{dF_Y(y)}{dy} = -\frac{dF_X(\varphi^{-1}(y))}{dx}.\frac{d\varphi^{-1}(y)}{dy} = -f_X(\varphi^{-1}(y)).\frac{d\varphi^{-1}(y)}{dy}$$

$\Rightarrow$

$$f_Y(y) = f_X(\varphi^{-1}(y)).\left|\frac{d\varphi^{-1}(y)}{dy}\right|$$

Hợp nhất cả hai trường hợp ta có

$$f_Y(y) = f_X(\varphi^{-1}(y)).\left|\frac{d\varphi^{-1}(y)}{dy}\right|$$

+ *Ví dụ 1*. Cho biến ngẫu nhiên X có phân phối chuẩn N(0,1). Tìm luật phân phối của $Y = X^3$.

Giải.

Hàm mật độ phân phối của X là

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

Hàm $y = \varphi(x) = x^3$ là hàm đơn điệu tăng và khả vi trong $(-\infty, +\infty)$. Ta có

$$x = \varphi^{-1}(y) = \sqrt[3]{y} \qquad \Rightarrow \qquad \frac{dx}{dy} = \frac{d\varphi^{-1}(y)}{dy} = \frac{1}{3\sqrt[3]{y^2}}$$

Vậy hàm mật độ của Y là

$$g(y) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\sqrt[3]{y^2}}.\frac{1}{3\sqrt[3]{y^2}} = \frac{1}{3\sqrt{2\pi}}\frac{1}{\sqrt[3]{y^2}}e^{-\frac{1}{2}\sqrt[3]{y^2}}.$$

+ *Ví dụ 2*. Cho biến ngẫu nhiên X có hàm mật độ f(x). Tìm hàm mật độ của biến ngẫu nhiên $Y = a.X + b$, $a, b \in R$, $a \neq 0$.

Giải.

Hàm $y = \varphi(x) = a.x + b$ đơn điệu. Ta có

$$x = \varphi^{-1}(y) = \frac{y-b}{a} \qquad \Rightarrow \qquad \frac{dx}{dy} = \frac{d\varphi^{-1}(y)}{dy} = \frac{1}{a}$$

Vậy hàm mật độ của Y là

$$g(y) = f\left(\frac{y-b}{a}\right).\frac{1}{|a|}$$

(iii) Trường hợp hàm $\varphi(x)$ không đơn điệu.

Giả sử biến ngẫu nhiên X lấy giá trị thuộc khoảng (a,b), còn $\varphi(x)$ đơn điệu từng khúc trên (a,b), nghĩa là có thể chia (a,b) thành n khoảng con J_i, i=1, 2,…,n, rời nhau sao cho $\varphi(x)$ đơn điệu trên mỗi khoảng con đó.

Ký hiệu $\varphi_i : J_i \to (a,b)$, $\varphi_i(x) = \varphi(x)\ \forall x \in J_i$, i=1, 2, …,n.

Cho khoảng $(y, y+\Delta y) \subset (a,b)$. Ký hiệu

$$(x_i, x_i + \Delta x_i) = \varphi_i^{-1}((y, y+\Delta y))\ \forall i = 1, 2, \ldots, n.$$

(có thể Δx_i có dấu bất kỳ). Ta có

$$P(y < Y < y + \Delta y) = \sum_{i=1}^{n} P(X \in (x_i, x_i + \Delta x_i)$$

Vậy, hàm mật độ của Y là

$$g(y) = \lim_{\Delta y \to 0} \frac{P(y < Y < y + \Delta y)}{\Delta y} = \lim_{\Delta y \to 0} \frac{\sum_{i=1}^{n} P(X \in (x_i, x_i + \Delta x_i)}{\Delta y}$$

$$= \sum_{i=1}^{n} \lim_{\Delta y \to 0} \left[\frac{P(X \in (x_i, x_i + \Delta x_i)}{|\Delta x_i|} . \left| \frac{\Delta x_i}{\Delta y} \right| \right]$$

$$= \sum_{i=1}^{n} f(x_i). \left| \frac{d\varphi_i^{-1}(y)}{dy} \right| = \sum_{i=1}^{n} f\left(\varphi_i^{-1}(y)\right). \left| \frac{d\varphi_i^{-1}(y)}{dy} \right|$$

+ *Ví dụ*. Cho biến ngẫu nhiên X có phân phối chuẩn N(0,1). Tìm hàm mật độ của biến ngẫu nhiên $Y = X^2$.

Giải.

Hàm $y = \varphi(x) = x^2$ có hai nhánh đơn điệu trên $J_1 = (0, +\infty)$ và $J_2 = (-\infty, 0)$. Các hàm

$$\varphi_1 : J_1 \to R^+, \varphi_1(x) = x^2 \quad \text{và} \quad \varphi_2 : J_2 \to R^+, \varphi_2(x) = x^2$$

có các hàm ngược tương ứng là $x = \varphi_1^{-1}(y) = \sqrt{y}$ trên R^+ và $x = \varphi_2^{-1}(y) = -\sqrt{y}$ trên R^+. Hàm mật độ của X là

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

Áp dụng công thức trên ta nhận được

$$g(y) = f\left(\varphi_1^{-1}(y)\right). \left| \frac{d\varphi_1^{-1}(y)}{dy} \right| + f\left(\varphi_2^{-1}(y)\right). \left| \frac{d\varphi_2^{-1}(y)}{dy} \right|$$

$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\sqrt{y}\right)^2} . \left| \frac{1}{2\sqrt{y}} \right| + \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(-\sqrt{y}\right)^2} . \left| -\frac{1}{2\sqrt{y}} \right| = \frac{1}{\sqrt{2\pi . y}} e^{-\frac{1}{2}y} \quad \forall\ y > 0$$

2. Hàm nhiều đối số ngẫu nhiên rời rạc.

a) Cho (X, Y) là vectơ ngẫu nhiên với luật phân phối

$$\{((x_i, y_j), p_{ij}) \mid i \in I, j \in J\}$$

Cho $\varphi(x,y)$ là hàm hai biến. Xét biến ngẫu nhiên $Z = \varphi(X,Y)$.
Ký hiệu

$$\{ z_k \mid k \in K \}$$

là tập hợp các giá trị khác nhau của tập

$$\{\varphi(x_i,y_j) \mid i \in I, j \in J\}$$

Với mỗi $k \in K$, ký hiệu

$$r_k = \sum_{(i,j)\in IxJ,\varphi(x_i,y_j)=z_k} p_{ij}$$

Khi đó Z có luật phân phối

$$\{(z_k, r_k) \mid k \in K\}$$

Cho X và Y có luật phân phối tương ứng

$$\{(x_i, p_i) \mid i \in I\}$$

và

$$\{(y_j, q_j) \mid j \in J\}$$

Nếu X và Y độc lập với nhau thì

$$r_k = \sum_{(i,j)\in IxJ,\varphi(x_i,y_j)=z_k} p_i q_j$$

+ *Ví dụ*. Một thùng chứa n quả cầu đánh số từ 1 đến n. Rút ngẫu nhiên quả cầu thứ nhất, không bỏ lại thùng, sau đó rút quả thứ hai. Gọi X và Y là số của quả cầu thứ nhất và số của quả cầu thứ hai. Hãy tìm luật phân phối của X+Y.

Giải.

Ta dễ dàng tính được xác suất

$$p_{ij} = P(X = i, Y = j) = \begin{cases} 0 & ,i = j \\ \dfrac{1}{n(n-1)} & ,i \neq j \end{cases}$$

Hiển nhiên là X+Y nhận các giá trị từ 3 đến 2n−1. Cho $k \in \{3, 4, \ldots, 2n-1\}$. Xét các trường hợp sau:

(i) $3 \le k \le n$ và k lẻ. Ta có

$$P(X+Y = k) = \sum_{i+j=k} p_{ij} = \sum_{i=1}^{k-1} p_{i,k-i} = \frac{k-1}{n(n-1)}$$

(vì k lẻ nên không thể xảy ra $i=k-i \Rightarrow p_{i,k-1} = 1/n(n-1) \quad \forall\, i = 1, 2, \ldots, k-1$)

(ii) $3 \le k \le n$ và k chẵn. Ta có

$$P(X+Y = k) = \sum_{i+j=k} p_{ij} = \sum_{i=1}^{k-1} p_{i,k-i} = \frac{k-2}{n(n-1)}$$

(vì k chẵn nên $p_{i,k-1} = 1/n(n-1) \quad \forall\, i = 1, 2, \ldots, k-1 \;\&\; i \neq k/2$)

(iii) n + 1 ≤ k ≤ 2n − 1 và k lẻ.

Xét cặp (i, j), i + j = k. Ta có j = k − i , và

$$1 \le i \le n \ \& \ 1 \le k - i \le n \Rightarrow \quad k - n \le i \le n$$

Từ đó suy ra

$$P(X+Y = k) = \sum_{i+j=k} p_{ij} = \sum_{i=k-n}^{n} p_{i,k-i} = \frac{2n-k+1}{n(n-1)}$$

(iv) n + 1 ≤ k ≤ 2n − 1 và k chẵn.

Với i = k/2, $p_{i,k-i}$ = 0. Suy ra

$$P(X+Y = k) = \sum_{i+j=k} p_{ij} = \sum_{i=k-n}^{n} p_{i,k-i} = \frac{2n-k}{n(n-1)}$$

b) Tính ổn định của phân phối nhị thức và phân phối Poisson.

• *Định lý 1*. Cho các biến ngẫu nhiên X và Y độc lập, tuân theo luật phân phối nhị thức B(m,p) và B(n,p), 0<p<1. Khi đó tổng X+Y có phân phối nhị thức B(m+n,p).
CM.

Miền giá trị của X+Y là 0, 1, 2, …, m+n. Cho k ∈ N, 0 ≤ k ≤ m+n, ký hiệu q=1−p, ta có

$$P(X+Y = k) = \sum_{i+j=k} P(X=i).P(Y=j) = \sum_{i=0}^{k} C(m,i).p^i.q^{m-i}C(n,k-i).p^{k-i}.q^{n-k+i}$$

$$= p^k.q^{m+n-k}\sum_{i=1}^{k} C(m,i).C(n,k-i) = p^k.q^{m+n-k}.C(m+n,k)$$

Vậy X+Y có phân phối nhị thức B(m+n,p).

• *Định lý 2*. Cho các biến ngẫu nhiên X và Y độc lập, tuân theo luật phân phối Poisson P(λ) và P(μ), λ>0 và μ>0. Khi đó tổng X+Y có phân phối Poisson P(λ+μ).
CM.

Miền giá trị của X+Y là N. Cho n ∈ N, ta có

$$P(X+Y = n) = \sum_{i+j=n} P(X=i).P(Y=j) = \sum_{i=0}^{n} e^{-\lambda}\frac{\lambda^i}{i!}e^{-\mu}\frac{\mu^{n-i}}{(n-i)!}$$

$$= \frac{e^{-(\lambda+\mu)}}{n!}\sum_{i=0}^{n}\frac{n!}{i!.(n-i)!}\lambda^i.\mu^{n-i} = e^{-(\lambda+\mu)}.\frac{(\lambda+\mu)^n}{n!}$$

Vậy X+Y có phân phối Poisson P(λ+μ).

3. Hàm nhiều đối số ngẫu nhiên liên tục.

a) Tổng hai biến ngẫu nhiên.

Cho vectơ ngẫu nhiên (X,Y) có hàm mật độ f(x,y). Xét biến ngẫu nhiên Z = X+Y.

Gọi G(z) là hàm phân phối của Z. Ký hiệu

$$D_z = \{(x,y) \mid x+y \leq z\}.$$

Ta có

$$G(z) = P(Z \leq z) = P(X+Y \leq z) = P((X,Y) \in D_z)$$

$$= \iint_{D_z} f(x,y)dxdy = \int_{-\infty}^{\infty}\left[\int_{-\infty}^{z-x} f(x,y)dy\right]dx \quad \forall z \in R$$

Lấy đạo hàm G(z) ta nhận được hàm mật độ g(z) của Z.

$$g(z) = G'(z) = \int_{-\infty}^{\infty} f(x, z-x)dx$$

Tương tự ta có

$$g(z) = \int_{-\infty}^{\infty} f(z-y, y)dy$$

• Trường hợp X và Y độc lập ta có

$$g(z) = \int_{-\infty}^{\infty} f_X(x).f_Y(z-x)dx = \int_{-\infty}^{\infty} f_X(z-y).f_Y(y)dy$$

trong đó f_X và f_Y là các hàm mật độ của X và Y.

✎ Ghi chú: Các phép toán với f_X và f_Y như trên gọi là *tích chập* và ký hiệu là

$$f_X * f_Y$$

+ *Ví dụ 1*. Cho X và Y là các biến ngẫu nhiên độc lập, phân phối đều trên [0, 1]. Tìm phân phối của X+Y.

Giải.

Hàm mật độ f_X và f_Y của X và Y như sau:

$$f_X(t) = f_Y(t) = \begin{cases} 1 & , t \in [0;1] \\ 0 & , t \notin [0;1] \end{cases}$$

Suy ra hàm mật độ g(z) , $z \in R$, theo công thức

$$g(z) = \int_{-\infty}^{\infty} f_X(x).f_Y(z-x)dx = \int_0^1 f_Y(z-x)dx$$

Ta xét các trường hợp sau.

(i) $z \leq 0$: Ta có

$$z - x < 0 \ \forall x \in (0;1) \Rightarrow f_Y(z-x) = 0 \quad \forall x \in (0;1) \Rightarrow \quad g(z) = 0$$

(ii) z > 2: Ta có

$$z - x > 1 \ \forall x \in (0; 1) \Rightarrow f_Y(z-x) = 0 \quad \forall x \in (0; 1) \Rightarrow \qquad g(z) = 0$$

(iii) 0 < z ≤ 1: Ta có

$$f_Y(z-x) = \begin{cases} 1 & ,0 \le x \le z \\ 0 & ,z < x < 1 \end{cases} \qquad \Rightarrow \qquad g(z) = \int_0^z dx = z$$

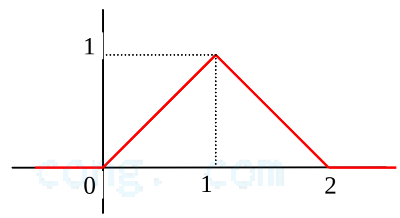
(iv) 1 < z ≤ 2: Ta có

$$f_Y(z-x) = \begin{cases} 1 & ,z-1 \le x \le 1 \\ 0 & ,0 < x \le z-1 \end{cases} \qquad \Rightarrow \qquad g(z) = \int_{z-1}^{1} dx = 1-(z-1) = 2-z$$

Cuối cùng ta có

$$g(z) = \begin{cases} 0 & ,z \le 0 \\ z & ,0 < z \le 1 \\ 2-z & ,1 < z \le 2 \\ 0 & ,2 < z \end{cases}$$

và đồ thị:



+ *Ví dụ 2*. Cho X và Y là các biến ngẫu nhiên độc lập, có phân phối mũ E(λ), λ>0. Khi đó X+Y có phân phối Gamma Γ(b,τ) với b=1/λ và τ=2.

➯ Ghi chú: Luật phân phối Gamma Γ(b,τ) có hàm mật độ

$$f(t) = \begin{cases} \dfrac{1}{b^{\tau}\Gamma(\tau)} t^{\tau-1} e^{-\frac{t}{b}} & ,t > 0 \\ 0 & ,t \le 0 \end{cases}$$

trong đó hàm Gamma $\Gamma(\tau) = \int_0^{\infty} x^{\tau-1}.e^{-x}dx$ (có tính chất Γ(τ+1) = τ.Γ(τ)).

Giải.

Hàm mật độ f_X và f_Y của X và Y như sau:

$$f_X(t) = f_Y(t) = \begin{cases} \lambda e^{-\lambda.t} & ,0 < t \\ 0 & ,t \le 0 \end{cases}$$

Suy ra hàm mật độ g(z) , z∈R, theo công thức

$$g(z) = \int_{-\infty}^{\infty} f_X(x).f_Y(z-x)dx = \begin{cases} \int_0^z f_X(x).f_Y(z-x)dx & ,z > 0 \\ 0 & ,z \le 0 \end{cases}$$

Ta có

$$\int_0^z f_X(x).f_Y(z-x)dx = \int_0^z \lambda.e^{-\lambda.x}.\lambda.e^{-\lambda.(z-x)}dx = \lambda^2.z.e^{-\lambda.z}$$

Vậy

$$g(z) = \begin{cases} \lambda^2.z.e^{-\lambda.z} & ,z > 0 \\ 0 & ,z \le 0 \end{cases}$$

+ *Ví dụ 3*. Cho X_1 và X_2 là các biến ngẫu nhiên độc lập có phân phối chính qui $N(m_1,\sigma_1)$ và $N(m_2,\sigma_2)$. Hãy chứng minh rằng $X_1 + X_2$ có phân phối chính qui

$N(m, \sigma)$ với $m = m_1 + m_2$ và $\sigma^2 = \sigma_1^2 + \sigma_2^2$.

CM.

Ta biểu diễn các biến ngẫu nhiên X_1 và X_2 như sau

$X_1 = \sigma_1.N_1 + m_1$ và $X_2 = \sigma_2.N_2 + m_2$

trong đó N_1 và N_2 là các biến ngẫu nhiên có phân phối chuẩn N(0, 1).

Suy ra

$$X_1 + X_2 = \sigma_1.N_1 + \sigma_2.N_2 + m_1 + m_2$$

$\Rightarrow$

$$\frac{X_1 + X_2}{\sqrt{\sigma_1^2+\sigma_2^2}} = \frac{\sigma_1}{\sqrt{\sigma_1^2+\sigma_2^2}}N_1 + \frac{\sigma_2}{\sqrt{\sigma_1^2+\sigma_2^2}}N_2 + \frac{m_1+m_2}{\sqrt{\sigma_1^2+\sigma_2^2}} = \alpha.N_1 + \beta.N_2 + a$$

với

$$\alpha = \frac{\sigma_1}{\sqrt{\sigma_1^2+\sigma_2^2}};\beta = \frac{\sigma_2}{\sqrt{\sigma_1^2+\sigma_2^2}};a = \frac{m_1+m_2}{\sqrt{\sigma_1^2+\sigma_2^2}}$$

Bây giờ ta chứng minh rằng $Y = \alpha.N_1 + \beta.N_2$ có phân phối chuẩn N(0, 1).

Hàm mật độ của $\alpha.N_1$ và $\beta.N_2$ tương ứng là

$$f_1(t) = \frac{1}{\alpha\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{t}{\alpha}\right)^2} \quad \text{và} \quad f_2(t) = \frac{1}{\beta\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{t}{\beta}\right)^2}$$

Gọi g là hàm mật độ của Y. Ta có

$$g(x) = \int_{-\infty}^{\infty} f_1(x-t).f_2(t)dt = \frac{1}{\alpha\beta.2\pi}\int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(\frac{x-t}{\alpha}\right)^2-\frac{1}{2}\left(\frac{t}{\beta}\right)^2}dt$$

Biến đổi biểu thức

$$Q(t) = \frac{1}{2}\left(\frac{x-t}{\alpha}\right)^2 + \frac{1}{2}\left(\frac{t}{\beta}\right)^2 = \frac{x^2}{2\alpha^2} + \frac{t^2}{2\alpha^2} - \frac{xt}{\alpha^2} + \frac{t^2}{2\beta^2} = \frac{t^2}{2\alpha^2\beta^2} - \frac{xt}{\alpha^2} + \frac{x^2}{2\alpha^2}$$

$$= \frac{1}{2\alpha^2\beta^2}\left(t^2 - 2xt.\beta^2 + x^2\beta^4\right) - \frac{x^2\beta^2}{2\alpha^2} + \frac{x^2}{2\alpha^2} = \frac{1}{2\alpha^2\beta^2}\left(t - x.\beta^2\right) + \frac{x^2}{2}$$

Suy ra

$$\text{g(x)} = \frac{1}{\alpha\beta.2\pi}\int_{-\infty}^{\infty} e^{-Q(t)}dt = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}\int_{-\infty}^{\infty}\frac{1}{\alpha\beta\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{t-x.\beta^2}{\alpha\beta}\right)^2}dt = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}$$

là mật độ của phân phối chuẩn N(0, 1).

Cuối cùng từ

$$X_1 + X_2 = \sqrt{\sigma_1^2 + \sigma_2^2}.Y + m_1 + m_2$$

suy ra đpcm.

b) Hiệu hai biến ngẫu nhiên.

Cho vectơ ngẫu nhiên (X,Y) có hàm mật độ f(x,y). Xét biến ngẫu nhiên Z = X−Y.

Gọi G(z) là hàm phân phối của Z. Ký hiệu

$$D_z = \{(x,y) \mid x-y \le z\}.$$

Ta có

$$G(z) = P(Z \le z) = P(X-Y \le z) = P((X,Y) \in D_z)$$

$$= \iint_{D_z} f(x,y)dxdy = \int_{-\infty}^{\infty}\left[\int_{x-z}^{\infty} f(x,y)dy\right]dx \qquad \forall z \in R$$

Lấy đạo hàm G(z) ta nhận được hàm mật độ g(z) của Z.

$$g(z) = G'(z) = \int_{-\infty}^{\infty} f(x, x-z)dx$$

Tương tự ta có

$$g(z) = \int_{-\infty}^{\infty} f(y+z, y)dy$$

• Trường hợp X và Y độc lập ta có

$$g(z) = \int_{-\infty}^{\infty} f_X(x).f_Y(x-z)dx = \int_{-\infty}^{\infty} f_X(y+z).f_Y(y)dy$$

trong đó f_X và f_Y là các hàm mật độ của X và Y.

+ *Ví dụ* . Cho X và Y là các biến ngẫu nhiên độc lập, phân phối đều trên [0, 1]. Tìm phân phối của X−Y.

Giải.

Hàm mật độ f_X và f_Y của X và Y như sau:

$$f_X(t) = f_Y(t) = \begin{cases} 1 & ,t \in [0;1] \\ 0 & ,t \notin [0;1] \end{cases}$$

Suy ra hàm mật độ g(z) , z∈R, theo công thức

$$g(z) = \int_{-\infty}^{\infty} f_X(x).f_Y(x-z)dx = \int_0^1 f_Y(x-z)dx$$

Ta xét các trường hợp sau.

(i) $z < -1$: Ta có

$x - z > 1 \ \forall x \in (0;1) \Rightarrow f_Y(x-z) = 0 \quad \forall x \in (0;1) \Rightarrow \quad g(z) = 0$

(ii) $z > 1$: Ta có

$x - z < 0 \ \forall x \in (0;1) \Rightarrow f_Y(x-z) = 0 \quad \forall x \in (0;1) \Rightarrow \quad g(z) = 0$

(iii) $-1 \le z \le 0$: Ta có

$$f_Y(x-z) = \begin{cases} 1 & ,0 \le x \le 1+z \\ 0 & ,1+z < x < 1 \end{cases} \Rightarrow \quad g(z) = \int_0^{1+z} dx = 1+z$$

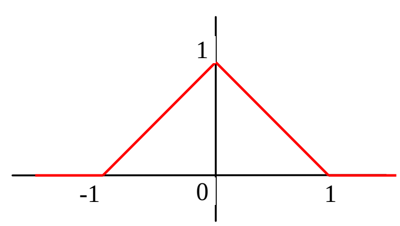
(iv) $0 < z \le 1$: Ta có

$$f_Y(x-z) = \begin{cases} 1 & ,z \le x \le 1 \\ 0 & ,0 < x < z \end{cases} \Rightarrow \quad g(z) = \int_z^1 dx = 1-z$$

Cuối cùng ta có

$$g(z) = \begin{cases} 0 & ,z \le -1 \\ 1+z & ,-1 < z \le 0 \\ 1-z & ,0 < z \le 1 \\ 0 & ,1 < z \end{cases}$$

và đồ thị:



c) Tích hai biến ngẫu nhiên.

Cho vectơ ngẫu nhiên (X,Y) có hàm mật độ f(x,y). Xét biến ngẫu nhiên Z = X.Y.

Gọi G(z) là hàm phân phối của Z. Ký hiệu

$$D_z = \{(x,y) \mid x.y \le z\}.$$

Ta có

$$G(z) = P(Z \le z) = P(X.Y \le z) = P((X,Y) \in D_z)$$

$$= \iint_{D_z} f(x,y)dxdy = \int_0^{\infty}\left[\int_{-\infty}^{\frac{z}{x}} f(x,y)dy\right]dx + \int_{-\infty}^{0}\left[\int_{\frac{z}{x}}^{\infty} f(x,y)dy\right]dx$$

$\forall$ z $\in$ R

Lấy đạo hàm G(z) ta nhận được hàm mật độ g(z) của Z.

$$g(z) = G'(z) = \int_0^{\infty} f\left(x,\frac{z}{x}\right).\frac{1}{x}dx + \int_{-\infty}^{0} -f\left(x,\frac{z}{x}\right).\frac{1}{x}dx = \int_{-\infty}^{\infty} f\left(x,\frac{z}{x}\right).\frac{1}{|x|}dx$$

Tương tự ta có

$$g(z) = \int_{-\infty}^{\infty} f\left(\frac{z}{y},y\right).\frac{1}{|y|}dy$$

• Trường hợp X và Y độc lập ta có

$$g(z) = \int_{-\infty}^{\infty} f_X(x).f_Y\left(\frac{z}{x}\right).\frac{1}{|x|}dx = \int_{-\infty}^{\infty} f_X\left(\frac{z}{y}\right).f_Y(y).\frac{1}{|y|}dy$$

trong đó f_X và f_Y là các hàm mật độ của X và Y.
+ *Ví dụ* . Cho X và Y là các biến ngẫu nhiên độc lập, phân phối đều trên [0, 1]. Tìm phân phối của X.Y.
Giải.

Hàm mật độ f_X và f_Y của X và Y như sau:

$$f_X(t) = f_Y(t) = \begin{cases} 1 & ,t \in [0;1] \\ 0 & ,t \notin [0;1] \end{cases}$$

Suy ra hàm mật độ g(z) của X.Y, z$\in$R, theo công thức

$$g(z) = \int_{-\infty}^{\infty} f_X(x).f_Y\left(\frac{z}{x}\right)\frac{1}{|x|}dx = \int_0^1 f_Y\left(\frac{z}{x}\right)\frac{1}{|x|}dx$$

Ta xét các trường hợp sau.

(i) z < 0: Ta có

z/x < 0 $\forall$x$\in$(0; 1) $\Rightarrow$ f_Y(z/x) = 0 $\forall$x$\in$(0; 1) $\Rightarrow$ g(z) = 0

(ii) z > 1: Ta có

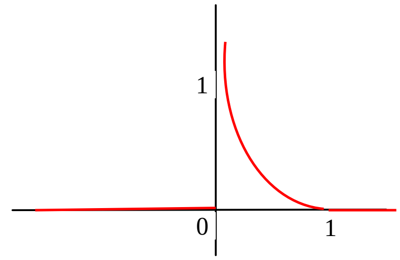
z/x > 1 $\forall$x$\in$(0; 1) $\Rightarrow$ f_Y(z/x) = 0 $\forall$x$\in$(0; 1) $\Rightarrow$ g(z) = 0

(iii) 0 ≤ z ≤ 1: Ta có

$$f_Y(z/x) = \begin{cases} 1 & ,z \le x \le 1 \\ 0 & ,0 < x < z \end{cases} \quad \Rightarrow \quad g(z) = \int_z^1 \frac{1}{x}dx = -\ln z$$

Cuối cùng ta có

g(z) = $\begin{cases} 0 & ,z \le 0 \\ -\ln z & ,0 < z \le 1 \\ 0 & ,1 < z \end{cases}$ và đồ thị:



c) Thương hai biến ngẫu nhiên.

Cho vectơ ngẫu nhiên (X,Y) có hàm mật độ f(x,y). Xét biến ngẫu nhiên Z = X/Y.

Gọi G(z) là hàm phân phối của Z. Ký hiệu

$$D_z = \{(x,y) \mid x/y \le z\}.$$

Ta có

$$G(z) = P(Z \le z) = P(X/Y \le z) = P((X,Y) \in D_z)$$

$$= \iint_{D_z} f(x,y)dxdy = \int_0^{\infty}\left[\int_{-\infty}^{z.y} f(x,y)dx\right]dy + \int_{-\infty}^{0}\left[\int_{z.y}^{\infty} f(x,y)dy\right]dx$$

$\forall z \in R$

Lấy đạo hàm G(z) ta nhận được hàm mật độ g(z) của Z.

$$g(z) = G'(z) = \int_0^{\infty} f(z.y,y).ydy + \int_{-\infty}^{0} -f(z.y,y).ydy = \int_{-\infty}^{\infty} f(z.y,y).|y|dy$$

Tương tự ta có

$$g(z) = \int_{-\infty}^{\infty} f\left(x,\frac{x}{z}\right).\frac{|x|}{z^2}dx$$

• Trường hợp X và Y độc lập ta có

$$g(z) = \int_{-\infty}^{\infty} f_X(z.y).f_Y(y).|y|dy = \int_{-\infty}^{\infty} f_X(x).f_Y\left(\frac{x}{z}\right).\frac{|x|}{z^2}dx$$

trong đó f_X và f_Y là các hàm mật độ của X và Y.

\+ *Ví dụ* . Cho X và Y là các biến ngẫu nhiên độc lập, phân phối đều trên [0, 1]. Tìm phân phối của X/Y.

Giải.

Hàm mật độ f_X và f_Y của X và Y như sau:

$$f_X(t) = f_Y(t) = \begin{cases} 1 & , t \in [0;1] \\ 0 & , t \notin [0;1] \end{cases}$$

Suy ra hàm mật độ g(z) của X.Y, z∈R, theo công thức

$$g(z) = \int_{-\infty}^{\infty} f_X(z.y).f_Y(y)|y|\,dy = \int_0^1 f_X(z.y)y\,dy$$

Ta xét các trường hợp sau.

(i) z < 0: Ta có

$z.y < 0\ \forall y \in (0; 1) \Rightarrow f_Y(z.y) = 0 \quad \forall x \in (0; 1) \Rightarrow \quad g(z) = 0$

(ii) 0 ≤ z ≤ 1: Ta có

$0 \le z.y \le 1\ \forall y \in (0; 1) \Rightarrow f_Y(z.y) = 1 \quad \forall x \in (0; 1) \Rightarrow \quad g(z) =$

$$\int_0^1 y\,dy = \left[\frac{y^2}{2}\right]_{y=0}^1 = \frac{1}{2}$$

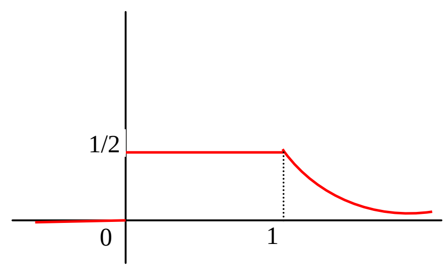
(iii) 1 < z : Ta có

$$f_Y(z.y) = \begin{cases} 1 & , 0 \le y \le \frac{1}{z} \\ 0 & , \frac{1}{z} < y \le 1 \end{cases} \Rightarrow \quad g(z) = \int_0^{\frac{1}{z}} y\,dy = \frac{1}{2.z^2}$$

Cuối cùng ta có

$$g(z) = \begin{cases} 0 & , z \le 0 \\ \frac{1}{2} & , 0 < z \le 1 \\ \frac{1}{2.z^2} & , 1 < z \end{cases}$$

và đồ thị:



IV. BIẾN NGẪU NHIÊN n CHIỀU

1. Khái niệm.

• *Định nghĩa*. Cho không gian xác suất (Ω, **B**, P), $n \in$ **N**, n > 1. Ánh xạ

$$X: \Omega \to R^n,\ X(\omega) = (X_1(\omega), X_2(\omega), \ldots, X_n(\omega))$$

gọi là *biến ngẫu nhiên n chiều* hay *vectơ ngẫu nhiên n chiều* , nếu $X_1, X_2, \ldots, X_n$ là các biến ngẫu nhiên trên không gian (Ω, **B**, P). Ký hiệu

$$X = (X_1, X_2, \ldots, X_n)$$

Hàm phân phối của vectơ ngẫu nhiên $(X_1, X_2, \ldots, X_n)$ là hàm n biến

$$F(x_1, x_2, \ldots, x_n) = P(X_1 \le x_1, X_2 \le x_2, \ldots, X_n \le x_n)$$

• Trường hợp $X_1, X_2, \ldots, X_n$ là các biến ngẫu nhiên rời rạc:
Luật phân phối của $(X_1, X_2, \ldots, X_n)$ như sau

$$\left\{\left(\left(x_{i_1}, \ldots, x_{i_n}\right), p_{i_1,\ldots,i_n}\right) \;\middle|\; i_1 \in I_1, \ldots, i_n \in I_n\right\}$$

với

$$p_{i_1,\ldots,i_n} = P\left(X_1 = x_{i_1}, \ldots, X_n = x_{i_n}\right) \qquad \forall\ (i_1, \ldots, i_n) \in I_1 \times \ldots \times I_n$$

trong đó $I_1, \ldots, I_n$ là các tập chỉ số đếm được hữu hạn hoặc vô hạn.

• Trường hợp $X_1, X_2, \ldots, X_n$ là các biến ngẫu nhiên liên tục:
Hàm mật độ phân phối của $(X_1, X_2, \ldots, X_n)$ là đạo hàm riêng hỗn hợp

$$f(x_1, x_2, \ldots, x_n) = \frac{\partial^n F(x_1, \ldots, x_n)}{\partial x_1 \ldots \partial x_n}$$

Cho $D \in R^n$. Xác suất $(X_1, X_2, \ldots, X_n) \in D$ được tính theo công thức

$$P((X_1, X_2, \ldots, X_n) \in D) = \int_D \ldots\ldots \int f(t_1, \ldots, t_n)\, dt_1 \ldots dt_n$$

2. Luật phân phối của các biến ngẫu nhiên thành phần.

a) Vectơ ngẫu nhiên rời rạc.

• *Xác suất của biến ngẫu nhiên thành phần X_k , $1 \le k \le n$.*

$$p_k = P(X_k = x_k) = \sum_{i_1 \in I_1} \ldots \sum_{i_{k-1} \in I_{k-1}} \sum_{i_{k+1} \in I_{k+1}} \ldots \sum_{i_n \in I_n} p_{i_1,\ldots,i_{k-1},k,i_{k+1},\ldots,i_n}$$

• *Xác suất của vectơ ngẫu nhiên thành phần $(X_1, \ldots, X_k)$, $1 < k \le n$.*

$$p_{i_1,\ldots,i_k} = P\left(X_1 = x_{i_1},\ldots,X_k = x_{i_k}\right) = \sum_{i_{k+1}\in I_{k+1}} \ldots \sum_{i_n \in I_n} p_{i_1,\ldots,i_{k-1},i_k,i_{k+1},\ldots,i_n}$$

b) Vectơ ngẫu nhiên liên tục.

• *Mật độ của biến ngẫu nhiên thành phần X_k , $1 \le k \le n$.*

$$f_k(t) = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} f\left(t_1,\ldots,t_{k-1},t,t_{k+1},\ldots,t_n\right)dt_1\ldots dt_{k-1}dt_{k+1}\ldots dt_n$$

• *Mật độ của vectơ ngẫu nhiên thành phần ($X_1, \ldots, X_k$), $1 < k \le n$.*

$$f_{1,\ldots,k}(t_1, \ldots, t_k) = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} f\left(t_1,\ldots,t_k,t_{k+1},\ldots,t_n\right)dt_{k+1}\ldots dt_n$$

3. Phân phối có điều kiện.

a) Vectơ ngẫu nhiên rời rạc.

Cho vectơ ngẫu nhiên ($X_1, X_2, \ldots, X_n$), $n \in \mathbf{N}$, $n > 1$, với luật phân phối

$$\left\{\left(\left(x_{i_1},\ldots,x_{i_n}\right), p_{i_1,\ldots,i_n}\right) \quad \middle| i_1 \in I_1,\ldots,i_n \in I_n\right\}$$

với

$$p_{i_1,\ldots,i_n} = P\left(X_1 = x_{i_1},\ldots,X_n = x_{i_n}\right) \qquad \forall\ (i_1, \ldots, i_n) \in I_1 \times \ldots \times I_n$$

trong đó $I_1, \ldots, I_n$ là các tập chỉ số đếm được hữu hạn hoặc vô hạn.

Luật *phân phối có điều kiện* của ($X_1, X_2, \ldots, X_k$), $1 \le k \le n$, với điều kiện $X_{k+1}=x_{k+1}, \ldots, X_n=x_n$ là phân phối

$$\left\{\left(\left(x_{i_1},\ldots,x_{i_k}\right), p_{i_1,\ldots,i_k / x_{k+1},\ldots,x_n}\right) \quad \middle| i_1 \in I_1,\ldots,i_k \in I_k\right\}$$

với

$$p_{i_1,\ldots,i_k / x_{k+1},\ldots,x_n} = \frac{p_{i_1,\ldots,i_k,i_{k+1},\ldots,i_n}}{p_{i_{k+1},\ldots,i_n}} \quad \forall\ (i_1, \ldots, i_k) \in I_1 \times \ldots \times I_k$$

b) Vectơ ngẫu nhiên liên tục.

Cho vectơ ngẫu nhiên liên tục ($X_1, X_2, \ldots, X_n$), $n \in \mathbf{N}$, $n > 1$, với hàm mật độ

$$f(x_1, x_2, \ldots, x_n)$$

Hàm *mật độ có điều kiện* của ($X_1, X_2, \ldots, X_k$), $1 \le k \le n$, với điều kiện $X_{k+1}=x_{k+1}, \ldots, X_n=x_n$, $f_{k+1,\ldots,n}(x_{k+1}, \ldots, x_n) \neq 0$, là

$$\left\{\left(\left(x_{i_1},\ldots,x_{i_k}\right), p_{i_1,\ldots,i_k / x_{k+1},\ldots,x_n}\right) \quad \middle| i_1 \in I_1,\ldots,i_k \in I_k\right\}$$

với

$$f\left(t_1,\dots,t_k / x_{k+1},\dots,x_n\right)=\frac{f\left(t_1,\dots,t_k,x_{k+1},\dots,x_n\right)}{f_{k+1,\dots,n}\left(x_{k+1},\dots,x_n\right)}$$

4. Biến ngẫu nhiên độc lập.

• *Định nghĩa*. Các biến ngẫu nhiên $X_1, X_2, \dots, X_n$ gọi là độc lập với nhau nếu luật phân phối của mỗi hệ con trích từ $(X_1, X_2, \dots, X_n)$ không phụ thuộc vào giá trị các biến ngẫu nhiên còn lại.

• *Định lý 1*.

Cho vectơ ngẫu nhiên rời rạc $(X_1, X_2, \dots, X_n)$, $n \in \mathbf{N}$, $n > 1$, với luật phân phối

$$\left\{\left(\left(x_{i_1},\dots,x_{i_n}\right), p_{i_1,\dots,i_n}\right) \;\middle|\; i_1 \in I_1,\dots,i_n \in I_n\right\}$$

với

$$p_{i_1,\dots,i_n} = P\left(X_1 = x_{i_1},\dots,X_n = x_{i_n}\right) \qquad \forall\ (i_1, \dots, i_n) \in I_1 \times \dots \times I_n$$

trong đó $I_1, \dots, I_n$ là các tập chỉ số đếm được hữu hạn hoặc vô hạn.

Các biến ngẫu nhiên $X_1, X_2, \dots, X_n$ độc lập với nhau khi và chỉ khi

$$p_{i_1,\dots,i_n} = P\left(X_1 = x_{i_1}\right)\dots P\left(X_n = x_{i_n}\right) \qquad \forall\ (i_1, \dots, i_n) \in I_1 \times \dots \times I_n$$

• *Định lý 2*.

Cho vectơ ngẫu nhiên liên tục $(X_1, X_2, \dots, X_n)$, $n \in \mathbf{N}$, $n > 1$, với hàm mật độ

$$f(x_1, x_2, \dots, x_n)$$

Các biến ngẫu nhiên $X_1, X_2, \dots, X_n$ độc lập với nhau khi và chỉ khi

$$f(x_1, x_2, \dots, x_n) = f_1(x_1). \dots . f_n(x_n) \qquad \forall\ (x_1, \dots, x_n) \in R^n$$

V. KỲ VỌNG VÀ PHƯƠNG SAI

• *Định lý 1*.

Cho vectơ ngẫu nhiên rời rạc $Z = (X_1, X_2, \ldots, X_n)$, $n \in \mathbf{N}$, $n > 1$, với luật phân phối

$$\left\{\left(\left(x_{i_1},\ldots,x_{i_n}\right), p_{i_1,\ldots,i_n}\right) \;\middle|\; i_1 \in I_1,\ldots,i_n \in I_n\right\}$$

với

$$p_{i_1,\ldots,i_n} = P\left(X_1 = x_{i_1},\ldots,X_n = x_{i_n}\right) \qquad \forall\, (i_1, \ldots, i_n) \in I_1 \times \ldots \times I_n$$

trong đó $I_1, \ldots, I_n$ là các tập chỉ số đếm được hữu hạn hoặc vô hạn.

Cho ánh xạ $\varphi: R^n \to R$ sao cho $\varphi(Z)$ là biến ngẫu nhiên.

Khi đó kỳ vọng $E[\varphi(Z)]$ của $\varphi(Z)$ là

$$E[\varphi(Z)] = \sum_{i_1 \in I} \ldots \sum_{i_n \in I_n} \varphi\left(x_{i_1},\ldots,x_{i_n}\right).p_{i_1,\ldots,i_n}$$

với điều kiện chuỗi hội tụ tuyệt đối.

CM.

Ta chứng minh cho trường hợp n=2. Trường hợp n > 2 chứng minh tương tự.

Ký hiệu lại $Z = (X, Y)$ là vectơ ngẫu nhiên với luật phân phối

$$\{[(x_i, y_j), p_{ij}] \mid i \in I, j \in J\}$$

với I, J là các tập chỉ số dạng hữu hạn {1,2,…,n} hoặc vô hạn {1,2,…, ∞}.

Ký hiệu

$$V = \{ z_k \mid k \in K \},$$

trong đó K là tập chỉ số đếm được, là tập giá trị khác nhau của $\varphi(Z)$.

Khi đó, với mọi $k \in K$ ta đặt

$$r_k = P(\varphi(Z) = z_k) = \sum_{(i,j) \in I_k} p_{ij}$$

trong đó

$$I_k = \{ (i, j) \in I \times J \mid \varphi(x_i, y_j) = z_k \}$$

Ta thấy I_k, $k \in K$, rời nhau từng cặp và

$$\bigcup_{k \in K} I_k = I \times J$$

Suy ra

$$E[\varphi(Z)] = \sum_{k \in K} z_k.r_k = \sum_{k \in K} \sum_{(i,j) \in I_k} \varphi\left(x_i, y_j\right).p_{ij} = \sum_{(i,j) \in I x J} \varphi\left(x_i, y_j\right).p_{ij} = \sum_{i \in I} \sum_{j \in J} \varphi\left(x_i, y_j\right).p_{ij}$$

• *Định lý 2*.

Cho vectơ ngẫu nhiên liên tục $(X_1, X_2, \ldots, X_n)$, $n \in \mathbf{N}$, $n > 1$, với hàm mật độ

$$f(x_1, x_2, \ldots, x_n)$$

Cho ánh xạ φ: $R^n \to R$ sao cho φ(Z) là biến ngẫu nhiên.
Khi đó kỳ vọng E[φ(Z)] của φ(Z) là

$$E[\varphi(Z)] = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} \varphi(t_1,\ldots,t_n).f(t_1,\ldots,t_n)dt_1\ldots dt_n$$

với điều kiện tích phân hội tụ tuyệt đối.
CM.
Suy ra từ định nghĩa và phép đổi biến của tích phân Lebesgue (công nhận).

Các định lý sau là hệ quả của 2 định lý trên.

• *Định lý 3*. Nếu $X_1, X_2, \ldots, X_n$ là các biến ngẫu nhiên có kỳ vọng thì

$$X_1 + X_2 + \ldots + X_n$$

có kỳ vọng và

$$E(X_1 + X_2 + \ldots + X_n) = E(X_1) + E(X_2) + \ldots + E(X_n)$$

• *Định lý 4*. Nếu $X_1, X_2, \ldots, X_n$ là các biến ngẫu nhiên độc lập có kỳ vọng thì

$$X_1.X_2 \ldots . X_n$$

có kỳ vọng và

$$E(X_1.X_2 \ldots . X_n) = E(X_1).E(X_2)\ldots E(X_n)$$

• *Hệ quả*. Nếu X , Y độc lập và có phương sai thì

$$\text{cov}(X,Y) = 0$$

• *Định lý 5*. Nếu $X_1, X_2, \ldots, X_n$ là các biến ngẫu nhiên có phương sai thì

$$X_1 + X_2 + \ldots + X_n$$

có phương sai và

$$D(X_1 + X_2 + \ldots + X_n) = \sum_{i=1}^{n} D(X_i) + 2.\sum_{1\le i\le j\le n} \text{cov}(X_i, X_j)$$

• *Hệ quả*. Nếu $X_1, X_2, \ldots, X_n$ là các biến ngẫu nhiên độc lập có phương sai thì

$$X_1 + X_2 + \ldots + X_n$$

có phương sai và

$$D(X_1 + X_2 + \ldots + X_n) = D(X_1) + D(X_2) + \ldots + D(X_n)$$

+ *Ví dụ 1*. Một thùng có a quả cầu trắng và b quả cầu đen. Rút liên tiếp 2 quả cầu (không trả lại vào thùng). Ký hiệu X_i, i=1,2, là các biến ngẫu nhiên nhận giá trị 1 nếu lần rút i được quả cầu trắng, nhận giá trị 0 nếu lần rút i được quả cầu đen. Khi đó cặp (X_1, X_2) tạo thành biến ngẫu nhiên 2 chiều.

Theo ví dụ 1, mục 5, bài I, ta có bảng phân phối của vectơ ngẫu nhiên (X_1, X_2)

X_1 \ X_2	1	0
1	$\frac{a}{a+b}\frac{a-1}{a+b-1}$	$\frac{a}{a+b}\frac{b}{a+b-1}$
0	$\frac{b}{a+b}\frac{a}{a+b-1}$	$\frac{b}{a+b}\frac{b-1}{a+b-1}$

luật phân phối của X_1 và X_2

$$p_1 = P(X_1 = 1) = p_{11} + p_{10} = \frac{a}{a+b}$$

$$p_0 = P(X_1 = 0) = p_{01} + p_{00} = \frac{b}{a+b}$$

$$q_1 = P(X_2 = 1) = p_{11} + p_{01} = \frac{a}{a+b}$$

$$q_0 = P(X_2 = 0) = p_{10} + p_{00} = \frac{b}{a+b}$$

kỳ vọng, phương sai và hiệp phương sai

$$E(X_1) = E(X_2) = \frac{a}{a+b} \text{ \& } D(X_1) = D(X_2) = \frac{a.b}{(a+b)^2}$$

$$\text{cov}(X_1, X_2) = E(X_1.X_2) - E(X_1).E(X_2)$$

$$= \frac{a}{a+b}\frac{a-1}{a+b-1} - \frac{a^2}{(a+b)^2} = -\frac{a.b}{(a+b)^2(a+b-1)}$$

Suy ra

$$D(X_1 + X_2) = \frac{2.a.b}{(a+b)^2} - \frac{2.a.b}{(a+b)^2(a+b-1)} = \frac{2.a.b.(a+b-2)}{(a+b)^2(a+b-1)}$$

$$D(X_1 - X_2) = \frac{2.a.b}{(a+b)^2} + \frac{2.a.b}{(a+b)^2(a+b-1)} = \frac{2.a.b.(a+b)}{(a+b)^2(a+b-1)}$$

+ *Ví dụ 2*. Cho bảng phân phối xác suất của vectơ ngẫu nhiên (X, Y) như sau

X \ Y	−1	0	1
−1	p/4	q/4	p/4
0	q/4	0	q/4
1	p/4	q/4	p/4

với $0 < p < 1$ và $q = 1 - p$.

Ta có

$$E(X.Y) = p/2 - p/2 = 0$$

Suy ra X và Y không tương quan.

Mặt khác vì

$$P(X = 0; Y = 0) = 0 \neq P(X = 0).P(Y = 0) = q^2/4$$

nên X và Y không độc lập.

+ *Ví dụ 3*. Cho vectơ ngẫu nhiên (X, Y) có phân phối đều trên hình tròn bán kính R > 0, tức (X, Y) có hàm mật độ

$$f(x,y) = \begin{cases} \dfrac{1}{\pi.r^2} & , x^2 + y^2 \le r^2 \\ 0 & , x^2 + y^2 > r^2 \end{cases}$$

Ta có

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

$$= \iint\limits_{x^2+y^2\le r^2} x.y.dxdy - \iint\limits_{x^2+y^2\le r^2} x.dxdy.\iint\limits_{x^2+y^2\le r^2} y.dxdy = 0 ,$$

vì các hàm dưới dấu tích phân đều đối xúng qua gốc toạ đọ trên miền lấy tích phân.

Suy ra X và Y không tương quan.

Mặt khác, theo ví dụ ở mục 2, bài II, X và Y không độc lập.

CHƯƠNG 4
HỘI TỤ NGẪU NHIÊN

I. HỘI TỤ XÁC SUẤT

1. Khái niệm hội tụ xác suất.

• *Định nghĩa*. Cho X là biến ngẫu nhiên và $(X_n)_{n\geq 1}$ là dãy biến ngẫu nhiên trên không gian xác suất (Ω, **B**, P). Ta nói rằng $(X_n)_{n\geq 1}$ *hội tụ xác suất* đến X, ký hiệu

$$X_n \underset{n\to\infty}{\overset{P}{\to}} X$$

nếu

$$\forall\ \varepsilon > 0,\ P(|X_n - X| \geq \varepsilon) \to 0 \text{ khi } n \to \infty$$

⇔

$$\forall\ \varepsilon > 0,\ P(|X_n - X| < \varepsilon) \to 1 \text{ khi } n \to \infty$$

+ *Ví dụ*. Người ta cho vô số quả cầu vào 3 thùng với xác suất như nhau. Với mọi n > 0, gọi Y_n là tần số và X_n là tần suất quả cầu rơi vào thùng thứ nhất trong số n quả cầu thả vào ba thùng. Chứng minh rằng

$$X_n \underset{n\to\infty}{\overset{P}{\to}} \frac{1}{3}$$

Giải.

Với mọi n > 0, Y_n có phân phối nhị thức B(n,1/3) với

$$E(Y_n) = n.\frac{1}{3} \text{ và } D(Y_n) = n.\frac{2}{9}$$

Với ε > 0 bất kỳ ta có

$$P\left(\left|X_n - \frac{1}{3}\right| \geq \varepsilon\right) = P\left(\left|Y_n - n.\frac{1}{3}\right| \geq n.\varepsilon\right) = \sum_{\left|k-\frac{n}{3}\right|\geq n.\varepsilon} P(Y_n = k) \leq$$

$$\leq \sum_{\left|k-\frac{n}{3}\right|\geq n.\varepsilon} \left(\frac{k-\frac{n}{3}}{n.\varepsilon}\right)^2 P(Y_n = k) \leq \frac{1}{(n.\varepsilon)^2}\sum_{k=0}^{n}\left(k-\frac{n}{3}\right)^2 P(Y_n = k)$$

$$= \frac{1}{(n.\varepsilon)^2} D(Y_n) = \frac{2}{9.n.\varepsilon^2}$$

Từ đó suy ra

$$P\left(\left|X_n - \frac{1}{3}\right| \geq \varepsilon\right) \xrightarrow[n\to\infty]{} 0 \qquad \Rightarrow \qquad X_n \xrightarrow[n\to\infty]{P} \frac{1}{3}$$

+ *Ví dụ 2*. Cho dãy biến ngẫu nhiên $(X_n)_{n\geq 1}$, $\lambda > 0$, X_n có phân phối mũ $E(n.\lambda)$ với mọi $n \geq 1$. Chứng minh rằng $X_n \xrightarrow[n\to\infty]{P} 0$.

Giải.

Với $\varepsilon > 0$ bất kỳ ta có

$$P(|X_n - 0| \geq \varepsilon) = \int_{\varepsilon}^{\infty} n.\lambda.e^{-n.\lambda.t}dt = \left[-e^{-n.\lambda.t}\right]_{\varepsilon}^{\infty} = e^{-n.\lambda.\varepsilon}$$

Từ đó suy ra

$$P(|X_n - 0| \geq \varepsilon) \xrightarrow[n\to\infty]{} 0 \qquad \Rightarrow \qquad X_n \xrightarrow[n\to\infty]{P} 0$$

2. Bất đẳng thức Trebưsep.

Cho biến ngẫu nhiên X có phương sai D(X). Khi đó

$$\forall\ \varepsilon > 0,\ P(|X - E(X)| \geq \varepsilon) \leq \frac{D(X)}{\varepsilon^2}$$

CM.

(i) Trường hợp X là biến ngẫu nhiên rời rạc với luật phân phối

$$\{ (x_i, p_i) \mid i \in I \}.$$

Với $\varepsilon > 0$, ta có

$$P(|X - E(X)| \geq \varepsilon) = \sum_{|x_i - E(X)|\geq\varepsilon} p_i \leq \sum_{|x_i - E(X)|\geq\varepsilon} \left(\frac{x_i - E(X)}{\varepsilon}\right)^2 p_i$$

$$\leq \frac{1}{\varepsilon^2}\sum_{i\in I}(x_i - E(X))^2.p_i = \frac{D(X)}{\varepsilon^2}$$

(ii) Trường hợp X là biến ngẫu nhiên liên tục với hàm mật độ f(t).

Với $\varepsilon > 0$, ta có

$$P(|X - E(X)| \geq \varepsilon) = \int_{|t-E(X)|\geq\varepsilon} f(t)dt \leq \int_{|t-E(X)|\geq\varepsilon} \left(\frac{t - E(X)}{\varepsilon}\right)^2 f(t)dt$$

$$\leq \frac{1}{\varepsilon^2}\int_{-\infty}^{\infty}(t - E(X))^2.f(t)dt = \frac{D(X)}{\varepsilon^2}$$

3. Luật số lớn yếu.

a) Định lý Trebưsep.

Cho $X_1, X_2, \ldots, X_n, \ldots$ là dãy biến ngẫu nhiên không tương quan từng đôi và có phương sai bị chặn bởi hằng số C nào đó. Ký hiệu

$$Y_n = \frac{\sum_{i=1}^{n}\left(X_i - E(X_i)\right)}{n} \qquad \forall\ n = 1, 2, \ldots$$

Khi đó

$$Y_n \xrightarrow[n\to\infty]{P} 0$$

CM.

Ta có

$E(Y_n) = \frac{1}{n}\sum_{i=1}^{n} E\left[X_i - E(X_i)\right] = 0$ và $D(Y_n) = \frac{1}{n^2}\sum_{i=1}^{n} D(X_i) \le \frac{n.C}{n^2} = \frac{C}{n}$

Theo bất đẳng thức Trebưsep suy ra: với mọi $\varepsilon > 0$

$$P(|Y_n| \ge \varepsilon) \le \frac{C}{n.\varepsilon^2}$$

và từ đó ta có

$$P\left(|Y_n| \ge \varepsilon\right) \xrightarrow[n\to\infty]{} 0 \qquad \Rightarrow \qquad Y_n \xrightarrow[n\to\infty]{P} 0$$

đpcm

Sau đây là các trường hợp riêng của định lý Trebưsep.

b) Định lý Khin-Chin.

Cho $X_1, X_2, \ldots, X_n, \ldots$ là dãy biến ngẫu nhiên không tương quan từng đôi, có phương sai bị chặn và có cùng kỳ vọng a. Ký hiệu

$$Y_n = \frac{\sum_{i=1}^{n} X_i}{n} \qquad \forall\ n = 1, 2, \ldots$$

Khi đó

$$Y_n \xrightarrow[n\to\infty]{P} a$$

b) Định lý Bernoulli.

Cho $X_1, X_2, \ldots, X_n, \ldots$ là dãy biến ngẫu nhiên độc lập có phân phối Bernoulli tham số p, $0 < p < 1$. Ký hiệu

$$Y_n = \frac{\sum_{i=1}^{n} X_i}{n} \qquad \forall\ n = 1, 2, \ldots$$

Khi đó

$$Y_n \xrightarrow[n\to\infty]{P} p$$

➪ Ghi chú: Trong trường hợp này ta có

$$P(|Y_n - p| \ge \varepsilon) \le \frac{p(1-p)}{n.\varepsilon^2} \le \frac{1}{4.n.\varepsilon^2}$$

Bất đẳng thức này có thể được sử dụng để tìm khoảng tin cậy của tham số p.

II. HỘI TỤ THEO LUẬT

1. Khái niệm hội tụ theo luật.

• *Định nghĩa*. Cho X là biến ngẫu nhiên và $(X_n)_{n\geq 1}$ là dãy biến ngẫu nhiên trên không gian xác suất (Ω, **B**, P). Ta nói rằng $(X_n)_{n\geq 1}$ *hội tụ theo luật* đến X, ký hiệu

$$X_n \xrightarrow[n\to\infty]{F} X$$

nếu

$$\forall x \in R,\ F_X \text{ liên tục tại } x \Rightarrow F_{X_n}(x) \to F_X(x) \text{ khi } n \to \infty,$$

trong đó F_X , F_{X_n} là các hàm phân phối của X , X_n , n=1, 2, …

➯ Từ định nghĩa suy ra:

Nếu $X_n \xrightarrow[n\to\infty]{F} X$ và a, b là các điểm liên tục của F_X thì

$$P(a < X_n \leq b) \to P(a < X \leq b) \text{ khi } n \to \infty.$$

+ *Ví dụ*. Cho X_n, n≥ 1, có phân phối

$$\left\{\left(\frac{k}{n}, \frac{1}{n+1}\right) \quad \middle| k = 0,1,\dots,n\right\}$$

Chứng minh rằng dãy $(X_n)_{n\geq 1}$ hội tụ theo luật tới biến ngẫu nhiên X có phân phối đều trên [0; 1].

CM.

Ta có

$$\frac{n.x-1}{n+1} \leq F_{X_n}(x) = \frac{[n.x]}{n+1} \leq \frac{n.x}{n+1} \qquad \forall x \in (0; 1)$$

và

$$F_{X_n}(0) = 1/(n+1),\ F_{X_n}(x) = 0\ \forall x < 0 \qquad \text{và} \qquad F_{X_n}(x) = 1 \quad \forall x \geq 1$$

Từ đó suy ra

$$\forall x \in R,\ F_{X_n}(x) \to F_X(x) \text{ khi } n \to \infty,$$

⇒

$$X_n \xrightarrow[n\to\infty]{F} X$$

• *Định lý 1*. Cho X là biến ngẫu nhiên rời rạc và $(X_n)_{n\geq 1}$ là dãy biến ngẫu nhiên rời rạc trên không gian xác suất (Ω, **B**, P). Giả sử X và dãy $(X_n)_{n\geq 1}$ có cùng miền giá trị rời rạc $\{ v_k \mid k \in K \}$, K là tập chỉ số đếm được hữu hạn hay vô hạn. Khi đó

$$X_n \xrightarrow[n\to\infty]{F} X \quad \Leftrightarrow \quad \forall k \in K,\ P(X_n = v_k) \to P(X = v_k) \text{ khi } n \to \infty.$$

✎ Gợi ý chứng minh. Mệnh đề suy ra từ định nghĩa và từ nhận xét rằng với mọi k∈K ta có

$$P(X = v_k) = F_X(b) - F_X(a) \quad \forall\ a, b \text{ thoả } (a; b) \cap \{ v_k \mid k \in K \} = \{ v_k \}$$

• *Định lý 2*. Cho X là biến ngẫu nhiên và $(X_n)_{n\geq 1}$ là dãy biến ngẫu nhiên trên không gian xác suất (Ω, **B**, P). Khi đó

$$X_n \xrightarrow[n\to\infty]{P} X \quad \Rightarrow \quad X_n \xrightarrow[n\to\infty]{F} X$$

2. Hội tụ theo luật của phân phối nhị thức đến phân phối Poisson.

• *Định lý*. Cho dãy biến ngẫu nhiên $(X_n)_{n\geq 1}$ có phân phối nhị thức $B(n, p_n)$ thoả $n.p_n \to \lambda > 0$ khi $n \to \infty$. Khi đó $X_n \xrightarrow[n\to\infty]{F} X$, trong đó X có phân phối Poisson P(λ).

✎ Ứng dụng.

Trong thực tế, khi $n \geq 30$, $p \leq 0.1$ và $n.p < 10$, thì ta có thể coi B(n, p) xấp xỉ phân phối Poisson P(n.p).

+ *Ví dụ*. Cho phép thử α có sự kiện A với xác suất P(A) = 0.05. Thực hiện phép thử 100 lần. Tính xác suất sự kiện A xảy ra 2 lần.

Giải.

Ký hiệu Y là biến ngẫu nhiên chỉ tần số xuất hiện A trong 100 phép thử. Y tuân theo luật phân phối nhị thức B(100, 0.05).

(i) Tính chính xác:

$$P(Y = 2) = C(100, 2) * 0.05^2 * 0.95^{98} \approx 0.081$$

(ii) Tính xấp xỉ:

Vì số lần thử 100 > 30 và xác suất p = 0.05 < 0.1 và n.p = 5 < 10, nên có thể coi Y tuân theo luật phân phối Poisson P(λ) với λ = n.p = 5. Vậy

$$P(Y = 2) \approx e^{-5}\frac{5^2}{2!} \approx 0.084.$$

Như vậy có thể coi đây là xấp xỉ không tồi.

III. ĐỊNH LÝ GIỚI HẠN TRUNG TÂM

1. Định lý.

Cho $(X_n)_{n\geq 1}$ là dãy biến ngẫu nhiên độc lập tương hỗ và có cùng luật phân phối với kỳ vọng a và phương sai $\sigma^2 > 0$. Đặt

$$M_n^* = \frac{\sqrt{n}(M_n - a)}{\sigma} \quad \text{với } M_n = \frac{1}{n}\left(X_1 + X_2 + ... + X_n\right)$$

Khi đó dãy (M_n^*) hội tụ theo luật đến biến ngẫu nhiên có phân phối chuẩn N(0,1).
CM. Công nhận.

↳ Ghi chú: So sánh với luật số lớn yếu: Giả thiết của luật số lớn yếu yếu hơn giả thiết của định lý giới hạn trung tâm, nhưng kết luận của định lý giới hạn trung tâm mạnh hơn.

2. Hội tụ theo luật của luật nhị thức đến luật phân phối chuẩn.

Ta đã chỉ ra rằng luật nhị thức B(n, p) là tổng của n luật nhị thức B(1,p) độc lập. Luật B(1,p) có kỳ vọng a = p và phương sai $\sigma^2 = p.(1-p)$. Như vậy, áp dụng định lý giới hạn trung tâm ta nhận được

• *Định lý*. Cho $(X_n)_{n\geq 1}$ là dãy biến ngẫu nhiên có luật phân phối B(n,p), 0 < p < 1. Khi đó

$$\frac{X_n - n.p}{\sqrt{n.p.(1-p)}} \xrightarrow[n\to\infty]{F} N(0,1)$$

↳ Ứng dụng. Trên thực tế, khi $n \geq 30$ và $p \approx 0.5$ ta có thể coi luật

$$B(n, p) \approx N(np, \sqrt{n.p.(1-p)}).$$

+ *Ví dụ*.

Cho X ~ B(100; 0.4). Vì n = 100 > 30 và p = 0.4 ≈ 0.5, có thể coi X xấp xỉ luật

$$N(100*0.4; \sqrt{100*0.4*0.6}) = N(40; 2.\sqrt{6})$$

Áp dụng tính xác suất

$$P(X \leq 50) = P\left(\frac{X-40}{2\sqrt{6}} \leq \frac{50-40}{2\sqrt{6}}\right) = P\left(\frac{X-40}{2\sqrt{6}} \leq 2.04\right) \approx 0.9793$$

3. Hội tụ theo luật của luật Poisson đến luật phân phối chuẩn.

Ta đã chỉ ra rằng luật Poisson P(n.λ) là tổng của n luật Poisson P(λ) độc lập. Luật P(λ) có kỳ vọng $a = \lambda$ và phương sai $\sigma^2 = \lambda$. Như vậy, áp dụng định lý giới hạn trung tâm ta nhận được

• *Định lý*. Cho $(X_n)_{n\geq 1}$ là dãy biến ngẫu nhiên có luật phân phối Poisson P(n.β), β>0. Khi đó

$$\frac{X_n - n.\beta}{\sqrt{n.\beta}} \xrightarrow[n\to\infty]{F} N(0,1)$$

↳ Ứng dụng. Trên thực tế, khi $\lambda \geq 10$, ta có thể xấp xỉ

$$P(\lambda) \approx N(\lambda, \sqrt{\lambda}).$$

+ *Ví dụ*. Số khách hàng X của một quầy hàng trong 1 giờ tuân theo luật phân phối Poisson P(30). Tính xác suất có không quá 20 khách hàng trong 1 giờ.
Giải.

Vì tham số λ = 30 > 10 nên có thể coi X có phân phối chuẩn N(30, $\sqrt{30}$). Từ đó suy ra

$$P(X \leq 20) = P\left(\frac{X-30}{\sqrt{30}} \leq \frac{20-30}{\sqrt{30}}\right) = P\left(\frac{X-30}{\sqrt{30}} \leq -1.826\right)$$
$$= \Phi_{0,1}(-1.826) = 1 - \Phi_{0,1}(1.826) = 1 - 0.9661 = 0.0339$$

4. Cách tính xác suất điểm của phân phối nhị thức và phân phối Poisson bằng luật phân phối chuẩn.

Cho X là biến ngẫu nhiên có phân phối nhị thức hoặc phân phối Poisson xấp xỉ phân phối chuẩn N(a, σ). Để tính xác suất P(X = k), $k \in$ **N**, ta có thể sử dụng một hai công thức xấp xỉ sau.

• Công thức 1. Dùng cho phân phối nhị thức và phân phối Poisson.

$$P(X = k) = P(X \leq k) - P(X \leq k-1) = F_X(k) - F_X(k-1) = \Phi_{a,\sigma}(k) - \Phi_{a,\sigma}(k-1)$$

• Công thức 2: Dùng cho phân phối nhị thức B(n, p), 0 < p <1, $n \in$ **N**.

$$P(X = k) = \begin{cases} P(X \leq 0.5) = \Phi_{a,\sigma}(0.5) & ,k = 0 \\ P(k - 0.5 \leq X < k + 0.5) = \Phi_{a,\sigma}(k+0.5) - \Phi_{a,\sigma}(k-0.5) & ,1 \leq k \leq n \\ P(X \geq n - 0.5) = 1 - \Phi_{a,\sigma}(n-0.5) & ,k = n \end{cases}$$

+ *Ví dụ*. Cho X có phân phối nhị thức B(n, p), n=30 và p=0.4. Tính P(X = 10).
Giải.

Vì $n \geq 30$ và $p \approx 0.5$ nên có thể coi X xấp xỉ phân phối N(30*0.4; $\sqrt{30*0.4*0.6}$) = N(12; 2.68).

- Công thức 1:

$$P(X = 10) = \Phi_{a,\sigma}(10) - \Phi_{a,\sigma}(10 - 1) = \Phi_{0,1}\left(\frac{10-12}{2.68}\right) - \Phi_{0,1}\left(\frac{9-12}{2.68}\right)$$

$$= \Phi_{0,1}(-0.75) - \Phi_{0,1}(-1.12) = 0.0925$$

- Công thức 2:

$$P(X = 10) = \Phi_{a,\sigma}(10 + 0.5) - \Phi_{a,\sigma}(10 - 0.5) = \Phi_{0,1}\left(\frac{10.5-12}{2.68}\right) - \Phi_{0,1}\left(\frac{9.5-12}{2.68}\right)$$

$$= \Phi_{0,1}(-0.56) - \Phi_{0,1}(-0.93) = 0.1115$$

CHƯƠNG 5
THỐNG KÊ MÔ TẢ

I. KHÔNG GIAN MẪU

Để nghiên cứu tính chất nào đó của các vật thể của một tập hợp lớn, người ta thường lấy một số vật thể để nghiên cứu, rồi từ đó rút ra kết luận cho tất cả vật thể trong tập hợp.

+ *Ví dụ*. Để xác định tuổi thọ của một loại bóng đèn, người ta không thể thử nghiệm tất cả bóng đèn, mà chỉ thử nghiệm một số bóng rồi suy ra tuổi thọ chung (tất nhiên với độ tin cậy nào đó).

• *Định nghĩa*. Tập hợp tất cả vật thể ban đầu gọi là *tập tổng thể*. *Mẫu* là tập con các vật thể lấy ra từ tập tổng thể. Số phần tử của mẫu gọi là *cỡ mẫu*.

Bằng phương pháp nào đó có thể lấy ra nhiều mẫu khác nhau cùng cỡ mẫu. Tập hợp tất cả các mẫu cùng cỡ mẫu của một tập tổng thể gọi là *không gian mẫu* , và mỗi mẫu được coi là một điểm của không gian mẫu.

Muốn cho từ mẫu lấy được có thể suy ra chính xác tính chất của tập tổng thể thì mẫu phải tiêu biểu. Mẫu được coi là *tiêu biểu* nếu người ta lấy mấu một cách ngẫu nhiên, tức là mọi phần tử của tập tổng thể có thể rơi vào mẫu với xác suất như nhau (có thể chọn hú hoạ hoặc sinh số ngẫu nhiên bằng máy tính).

Mẫu có hai tính chất: lặp hoặc không lặp và có thứ tự hoặc không có thứ tự. Gọi N là số tất cả vật thể, *n* là cỡ mẫu.

Mẫu có lặp có thứ tự là một chỉnh hợp lặp chập n từ N phần tử và số mẫu là

$$N^n$$

Mẫu không lặp có thứ tự là một chỉnh hợp không lặp chập n từ N phần tử và số mẫu n là

$$A(N, n) = N(N-1) \dots (N-n+1)$$

Mẫu có lặp không thứ tự là một tổ hợp lặp chập n từ N phần tử và số mẫu là

$$C(N+n-1, n)$$

Mẫu không lặp không thứ tự là một tổ hợp chập n từ N phần tử và số mẫu là

$$C(N, n)$$

Nếu N lớn và n nhỏ thì tỉ lệ số mẫu lặp và không lặp xấp xỉ 1, như vậy việc lấy mẫu lặp và không lặp cũng cho kết quả gần như nhau.

Bây giờ giả sử tính chất của vật thể cần nghiên cứu là đại lượng ngẫu nhiên X. Khi đó mỗi mẫu cỡ n sẽ cho kết quả là bộ (X_1, X_2, …, X_n). Ta nói là đã lấy mẫu

$$(X_1, X_2, \ldots, X_n)$$

từ đại lượng ngẫu nhiên X.

Mẫu $(X_1, X_2, \ldots, X_n)$ được phân lớp theo một trong hai cách sau:

(i) *Phân lớp đơn*:

$$\{(x_i, n_i) \mid 1 \le i \le k\}$$

với $x_1 < x_2 < \ldots < x_k$ và n_i là tần số xuất hiện x_i, $i=1,\ldots,k$, $\sum n_i = n$

(ii) *Phân lớp ghép*:

$$\{([a_i, a_{i+1}), n_i) \mid 1 \le i \le k\}$$

với $a_1 < a_2 < \ldots < a_k$ và n_i là số x_i rơi vào khoảng $[a_i; a_{i+1})$, $i=1,\ldots,k$, $\sum n_i = n$.

↳ Ghi chú: Phân lớp ghép chỉ áp dụng cho X là biến ngẫu nhiên liên tục.

II. BIỂU DIỄN PHÂN PHỐI MẪU

1. Trường hợp phân lớp đơn.

Cho đại lượng ngẫu nhiên X, n ∈ **N**. Giả sử ta có mẫu cỡ n với phân lớp đơn

$$\{(x_i, n_i) \mid 1 \le i \le k\},$$

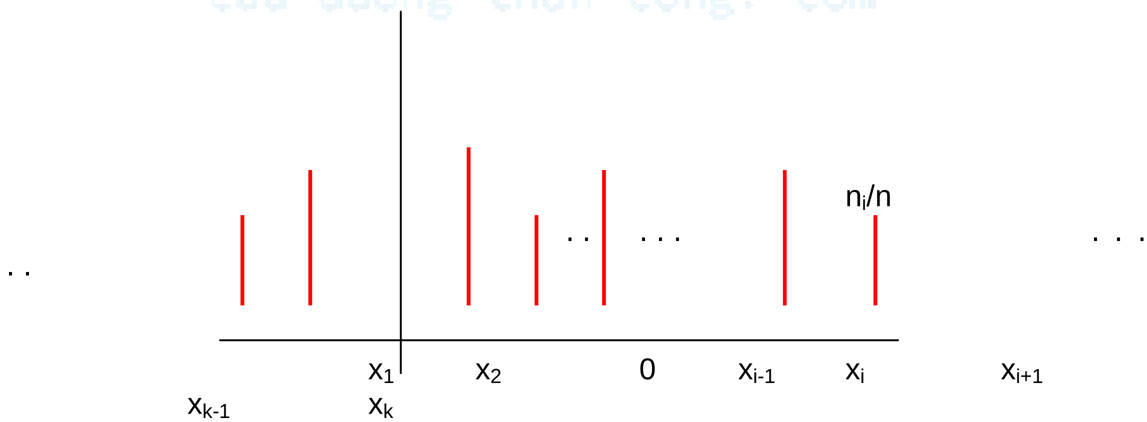
với $x_1 < x_2 < \ldots < x_k$ và n_i là tần số xuất hiện x_i , i=1,…,k, $\sum n_i = n$.

• *Tần suất* của x_i là đại lượng $\frac{n_i}{n}$, i=1,…,k.

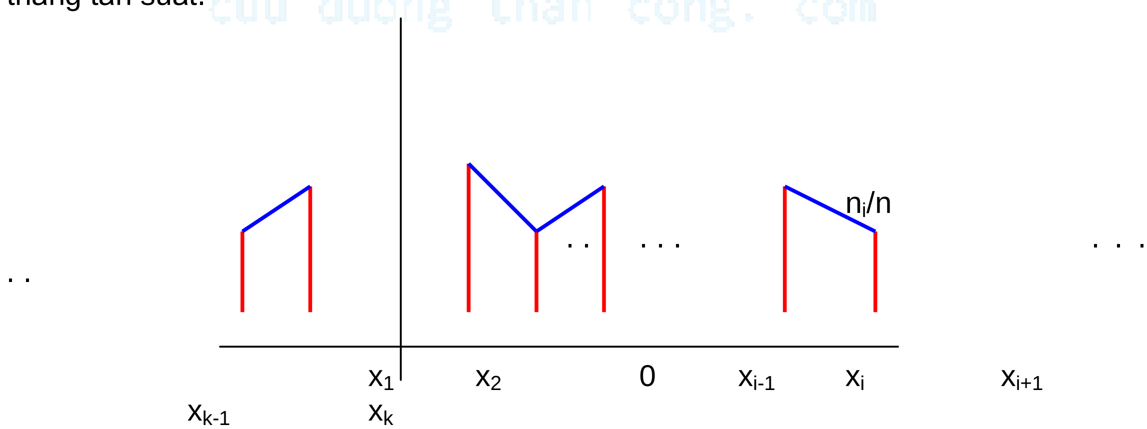
Bảng phân phối tần suất của X có dạng

x_1	x_2	…	x_i	…	x_k
$\frac{n_1}{n}$	$\frac{n_2}{n}$	…	$\frac{n_i}{n}$	…	$\frac{n_k}{n}$

• *Biểu đồ tần suất* được biểu diễn trên mặt phẳng toạ độ bằng các đoạn thẳng biểu diễn tần suất.



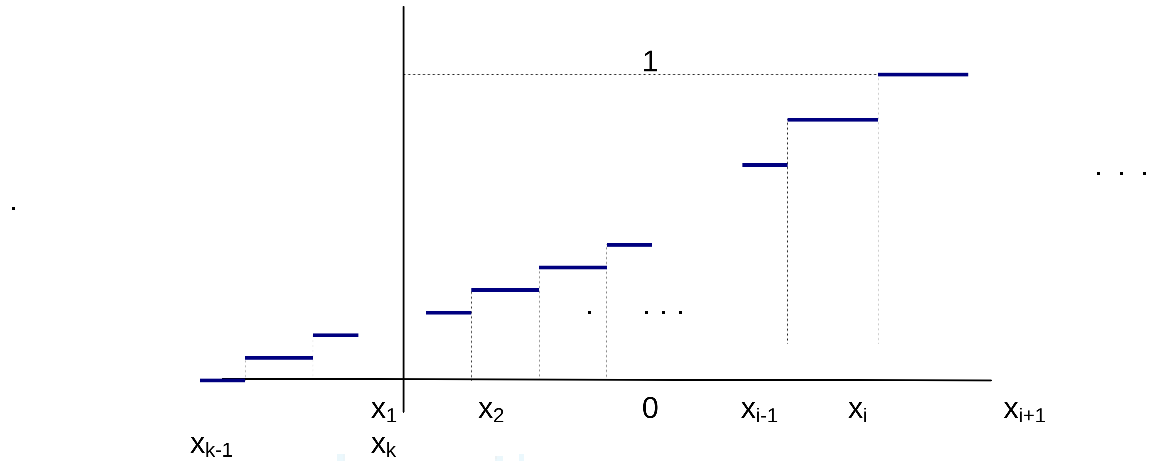
• *Đa giác tần suất* là đường gấp khúc (màu xanh) nối các đỉnh trên của các đoạn thẳng tần suất.



• *Tần suất tích luỹ* là hàm phân phối mẫu sau:

$$F_n(x) = \begin{cases} 0 & , x < x_1 \\ \sum_{i=1}^{j} \frac{n_i}{n} & , x_j \le x < x_{j+1}, j = 1,\ldots,k-1 \\ 1 & , x \ge x_k \end{cases}$$

Đồ thị có dạng bậc thang



Ghi chú: $F_n(x)$ là tần suất sự kiện X ≤ x, còn hàm phân phối F(x) là xác suất sự kiện X ≤ x. Vậy theo luật số lớn yếu (Định lý Bernoulli) ta có

$$F_n(x) \xrightarrow[n\to\infty]{P} F(x) \qquad \forall x \in R,$$

tức là

$$\forall \varepsilon > 0, \forall x \in R, P(|F_n(x) - F(x)| < \varepsilon) \to 1 \text{ khi } n \to \infty.$$

2. Trường hợp phân lớp ghép.

Cho đại lượng ngẫu nhiên liên tục X, n ∈ **N**. Giả sử ta có mẫu cỡ n với phân lớp ghép

$$\{([a_i, a_{i+1}), n_i) \mid 1 \le i \le k\}$$

với $a_1 < a_2 < \ldots < a_k$ và n_i là số x_i rơi vào khoảng $[a_i; a_{i+1})$, i=1,…,k, $\sum n_i = n$.

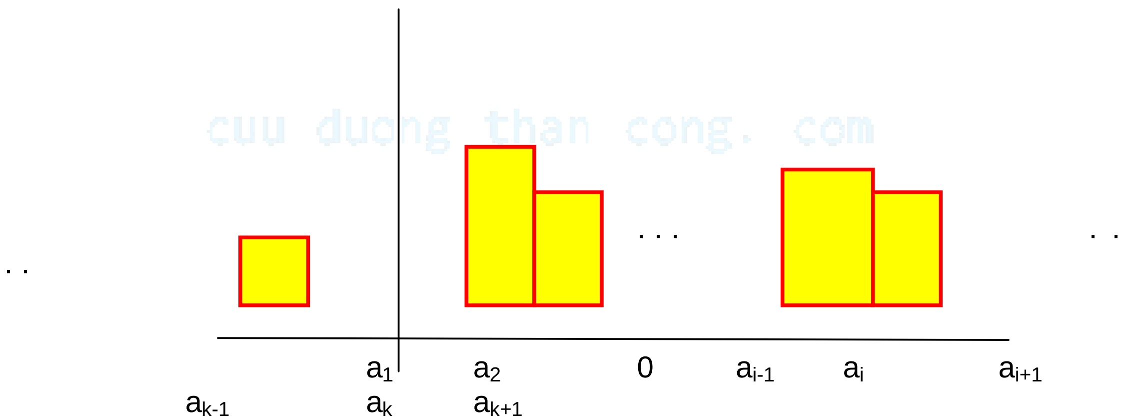
• *Tần suất* của lớp ghép i, tức khoảng $[a_i ; a_{i+1})$ là đại lượng $\frac{n_i}{n}$, i=1,…,k.

Các giá trị trong lớp $[a_i ; a_{i+1})$ được xấp xỉ bằng trị trung bình $\frac{a_i + a_{i+1}}{2}$

Bảng phân phối tần suất của X có dạng

$[a_i; a_{i+1})$	$\frac{a_i + a_{i+1}}{2}$	n_i	$\frac{n_i}{n}$
$[a_1; a_2)$ ⋮ $[a_k; a_{k+1})$	$\frac{a_1 + a_2}{2}$ ⋮ $\frac{a_k + a_{k+1}}{2}$	n_1 ⋮ n_k	$\frac{n_1}{n}$ ⋮ $\frac{n_k}{n}$

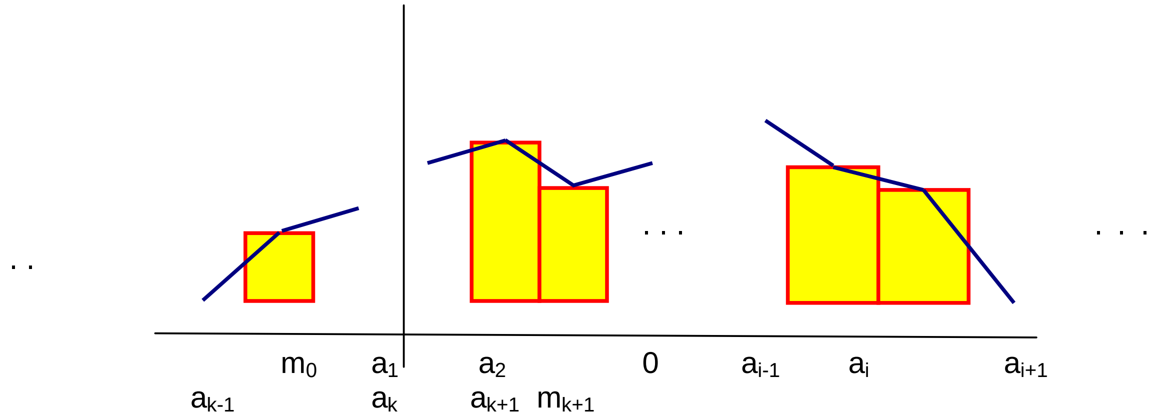
• *Tổ chức đồ tần suất* là cách biểu diễn tần suất trên mặt phẳng toạ độ trong đó tần suất $\frac{n_i}{n}$ được biểu diễn bằng hình chữ nhật đáy $[a_i; a_{i+1})$ và chiều cao là $\frac{n_i}{n(a_{i+1} - a_i)}$, i = 1, …, k.



• *Đa giác tần suất* là đường gấp khúc (màu xanh) nối các trung điểm đáy trên của các hình chữ nhật kề nhau trên tổ chức đồ tần suất.

Đoạn ngoài cùng bên trái nối trung điểm $[a_1; a_2)$ với điểm m_0 trên trục hoành cách a_1 một khoảng bằng nửa đoạn $[a_1; a_2)$.

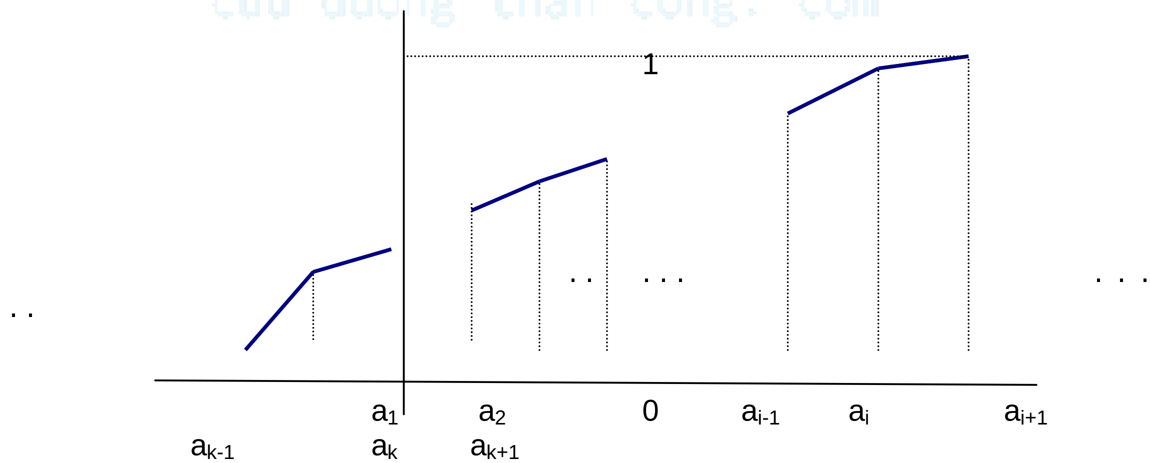
Đoạn ngoài cùng bên phải nối trung điểm $[a_k; a_{k+1})$ với điểm m_{k+1} trên trục hoành cách a_{k+1} một khoảng bằng nửa đoạn $[a_k; a_{k+1})$.



• *Hàm tần suất tích luỹ* là hàm phân phối mẫu có đường cong tần suất tích luỹ là đường gấp khúc nối các điểm

$$(a_1, 0),\ (a_2, \frac{n_1}{n}),\ (a_3, \frac{n_1+n_2}{n}),\ \ldots,\ (a_{j+1}, \sum_{i\le j}\frac{n_i}{n}),\ \ldots,\ (a_{k+1}, 1)$$

Đồ thị có dạng



III. CÁC THAM SỐ ĐẶC TRƯNG

1. Các tham số vị trí.

Cho đại lượng ngẫu nhiên X, $n \in$ **N**, và mẫu cỡ n của X.

a) Trị trung bình mẫu.

(i) Trường hợp mẫu phân lớp đơn

$$\{(x_i, n_i) \mid 1 \le i \le k\},$$

với $x_1 < x_2 < \ldots < x_k$ và n_i là tần số xuất hiện x_i , i=1,…,k, $\sum n_i = n$.

Ký hiệu tần suất của x_i là $f_i = \frac{n_i}{n}$, i=1,…,k. Ta định nghĩa các trị trung bình sau:

− *Trung bình cộng* hay *kỳ vọng mẫu*:

$$m_a = \overline{x} = \frac{1}{n}\sum_{i=1}^{k} n_i x_i = \sum_{i=1}^{k} f_i x_i$$

− *Trung bình hình học* :

$$m_g = \sqrt[n]{\prod_{i=1}^{k} x_i^{n_i}} = \prod_{i=1}^{k} x_i^{f_i}$$

− *Trung bình điều hoà*:

$$m_h = \frac{1}{\frac{1}{n}\sum_{i=1}^{k}\frac{n_i}{x_i}} = \frac{1}{\sum_{i=1}^{k}\frac{f_i}{x_i}}$$

− *Trung bình bình phương*:

$$m_q = \sqrt{\frac{1}{n}\sum_{i=1}^{k} n_i . x_i^2} = \sqrt{\sum_{i=1}^{k} f_i . x_i^2}$$

(i) Trường hợp mẫu phân lớp ghép.

$$\{([a_i, a_{i+1}), n_i) \mid 1 \le i \le k\}$$

với $a_1 < a_2 < \ldots < a_{k+1}$ và n_i là số x_i rơi vào khoảng $[a_i; a_{i+1})$, i=1,…,k, $\sum n_i = n$.

Ký hiệu tần suất của lớp ghép i, tức khoảng $[a_i ; a_{i+1})$ là $f_i = \frac{n_i}{n}$, i=1,…,k. Ta định nghĩa các trị trung bình tương tự như trường hợp mẫu phân lớp đơn với x_i thay bằng $c_i = \frac{a_i + a_{i+1}}{2}$.

− *Trung bình cộng* hay *kỳ vọng mẫu*:

$$m_a = \bar{x} = \frac{1}{n}\sum_{i=1}^{k} n_i c_i = \sum_{i=1}^{k} f_i c_i$$

b) Trung vị mẫu.

(i) Trường hợp mẫu phân lớp đơn

$$\{(x_i, n_i) \mid 1 \le i \le k\},$$

với $x_1 < x_2 < \ldots < x_k$ và n_i là tần số xuất hiện x_i , i=1,…,k, $\sum n_i = n$.

Trung vị mẫu , ký hiệu *med*, là số đứng giữa dãy $x_1, x_2, \ldots, x_k$ xác định như sau.

Xếp n trị x_i theo thứ tự như sau

$$\underbrace{x_1, x_1, \ldots, x_1}_{n_1}, \ldots, \underbrace{x_i, x_i, \ldots, x_i}_{n_i}, \ldots, \underbrace{x_k, x_k, \ldots, x_k}_{n_k}$$

Khi đó, nếu n = 2.m+1 lẻ thì *med* là phần tử ở vị trí thứ m+1, nếu n = 2.m chẵn thì *med* là trung bình cộng của phần tử ở vị trí thứ m và phần tử ở vị trí thứ m+1

+ *Ví dụ 1*: Cho mẫu cỡ 9 sau

3; 4; 4; 5; 6; 8; 8; 10; 11

Ở đây n = 9 = 2*4 + 1. Vậy *med* là phần tử thứ 5 (=4+1), tức *med* = 6

+ *Ví dụ 2*: Cho mẫu cỡ 100 sau

171; …; 171; 174; …; 174; 177; …; 177; 180; …; 180; 183; …; 183

6 17 41

27 9

Ở đây n = 100 = 2*50. Vậy *med* là trung bình cộng của phần tử thứ 50 và phần tử thứ 51, tức *med* = (177+177)/2 = 177.

(ii) Trường hợp mẫu phân lớp ghép

$$\{([a_i; a_{i+1}), n_i) \mid 1 \le i \le k\},$$

với $a_1 < a_2 < \ldots < a_{k+1}$ và n_i là số x_i rơi vào khoảng $[a_i; a_{i+1})$, $f_i = \frac{n_i}{n}$, i=1,…,k, $n = \sum n_i$.

Trung vị mẫu , ký hiệu *med*, là giá trị mà tại đó hàm tần suất tích luỹ F bằng ½, tức F(med) = ½.

med được xác định như sau:

− Tìm khoảng $[a_{h;}\ a_{h+1})$ chứa *med* thoả

$$p_{h-1} = \sum_{i \le h-1} f_i < \frac{1}{2} \le \sum_{i \le h} f_i = p_h$$

− Trung vị *med* được tính từ phương trình

$$\frac{med - a_h}{a_{h+1} - a_h} = \frac{0.5 - p_{h-1}}{p_h - p_{h-1}} = \frac{0.5 - p_{h-1}}{f_h}$$

⇒

$$med = a_h + \frac{0.5 - p_{h-1}}{f_h}\left(a_{h+1} - a_h\right)$$

+ *Ví dụ*: Cân 100 thanh niên ta có bảng tần suất lớp ghép sau

$[a_i;\ a_{i+1})$	59.5 − 62.5	62.5 − 65.5	65.5 − 68.5	68.5 − 71.5	71.5 − 74.5
f_i	5%	18%	42%	27%	8%

Vì

$$p_2 = 5\% + 18\% < ½ < 5\% + 18\% + 42\% = 65\% < p_3$$

nên khoảng chứa *med* là khoảng thứ 3

$$[a_3;\ a_4) = [65.5;\ 68.5).$$

Suy ra

$$med = 65.5 + \frac{0.5 - 23\%}{42\%}(68.5 - 65.5) = 65.5 + (27/42).3 = 67.4 \text{ (kg)}$$

c) Mode mẫu.

(i) Trường hợp mẫu phân lớp đơn

$$\{(x_i,\ n_i) \mid 1 \le i \le k\},$$

với $x_1 < x_2 < \ldots < x_k$ và n_i là tần số xuất hiện x_i , i=1,…,k, $\sum n_i = n$.

Mode mẫu là x_m $(1\le m\le k)$ có tần số n_m lớn nhất (có thể có nhiều *mode*)ẫu

+ *Ví dụ*. Mẫu cỡ 13

x_i	2	5	7	9	10	11	18
n_i	2	1	1	3	2	3	1

có hai *mode* là 9 và 11.

(ii) Trường hợp mẫu phân lớp ghép

$$\{([a_i; a_{i+1}), n_i) \mid 1 \le i \le k\},$$

với $a_1 < a_2 < \ldots < a_{k+1}$ và n_i là số x_i rơi vào khoảng $[a_i; a_{i+1})$, $f_i = \frac{n_i}{n}$, i=1,…,k, $n = \sum n_i$.

mode được xác định như sau:

− Tìm khoảng $[a_h; a_{h+1})$ có tần số lớn nhất (có thể có nhiều khoảng như vậy).

− *mode* được tính theo công thức

$$mode = a_h + \frac{n_h - n_{h-1}}{(n_h - n_{h-1}) + (n_h - n_{h+1})}(a_{h+1} - a_h)$$

+ *Ví dụ*: Cân 100 thanh niên ta có bảng tần suất lớp ghép sau

$[a_i; a_{i+1})$	59.5 − 62.5	62.5 − 65.5	65.5 − 68.5	68.5 − 71.5	71.5 − 74.5
f_i	5%	18%	42%	27%	8%

Vì lớp [65.5; 68.5) có tần suất lớn nhất nên mode được tính như sau

$$\text{Mode} = 65.5 + \frac{42-18}{42-18+42-27}(68.5 - 65.5) = 67.34$$

2. Các tham số phân tán

Cho đại lượng ngẫu nhiên X, n ∈ **N**. Giả thiết X có mẫu cỡ n hoặc phân lớp đơn

$$\{(x_i, n_i) \mid 1 \le i \le k\},$$

với $x_1 < x_2 < \ldots < x_k$ và n_i là tần số xuất hiện x_i , i=1,…,k, $\sum n_i = n$, hoặc phân lớp ghép

$$\{([a_i; a_{i+1}), n_i) \mid 1 \le i \le k\},$$

với $a_1 < a_2 < \ldots < a_{k+1}$ và n_i là số x_i rơi vào khoảng $[a_i; a_{i+1})$, $c_i = \frac{a_i + a_{i+1}}{2}$, i=1,…,k, $n = \sum n_i$.

a) Độ trải rộng.

Độ trải rộng của mẫu là hiệu

$x_k - x_1$ cho mẫu phân lớp đơn

và

$a_{k+1} - a_1$ cho mẫu phân lớp ghép.

b) Phương sai mẫu và *độ lệch chuẩn*

$$\hat{S}^2 = \frac{1}{n}\sum_{i=1}^{k} n_i\left(x_i - \overline{x}\right)^2$$ cho mẫu phân lớp đơn

và

$$\hat{S}^2 = \frac{1}{n}\sum_{i=1}^{k} n_i\left(c_i - \overline{x}\right)^2$$ cho mẫu phân lớp ghép.

✎ Ghi chú: Trong trường hợp phân lớp ghép, nếu các khoảng $[a_i; a_{i+1})$ bằng nhau và bằng c, thì có thể sử dụng phương sai hiệu chỉnh

$$\hat{S}_{hc}^2 = \hat{S}^2 - \frac{c^2}{12}$$ ($\frac{c^2}{12}$ gọi là hiệu chỉnh Shepard)

• Đại lượng $\hat{S} = \sqrt{\hat{S}^2}$ gọi là *độ lệch chuẩn*.

c) Độ lệch trung bình

$$e = \frac{1}{n}\sum_{i=1}^{k} n_i.\left|x_i - \overline{x}\right|$$ cho mẫu phân lớp đơn

và

$$e = \frac{1}{n}\sum_{i=1}^{k} n_i.\left|c_i - \overline{x}\right|$$ cho mẫu phân lớp ghép

d) Momen mẫu

• Momen mẫu bậc a ($a \in N$):

$$m_a = \frac{1}{n}\sum_{i=1}^{k} x_i^a$$ cho mẫu phân lớp đơn

và

$$m_a = \frac{1}{n}\sum_{i=1}^{k} c_i^a$$ cho mẫu phân lớp ghép

• Momen trung tâm mẫu bậc a ($a \in N$):

$$\mu_a = \frac{1}{n}\sum_{i=1}^{k} n_i.\left(x_i - \overline{x}\right)^a$$ cho mẫu phân lớp đơn

và

$\mu_a = \frac{1}{n}\sum_{i=1}^{k} n_i.\left(c_i - \overline{x}\right)^a$ cho mẫu phân lớp đơn

• Momen trung tâm rút gọn bậc a:

$$\alpha_a = \frac{\mu_a}{\hat{s}^a}$$

3. Các tham số hình dạng

a) Hệ số bất đối xứng mẫu

$$\gamma_1 = \alpha_3 = \frac{\mu_3}{\hat{s}^3}$$

b) Hệ số nhọn mẫu

$$\gamma_2 = \alpha_4 - 3$$

4. Các điểm phần tư

a) Trường hợp mẫu phân lớp đơn

$$\{(x_i, n_i) \mid 1 \le i \le k\},$$

với $x_1 < x_2 < \ldots < x_k$ và n_i là tần số xuất hiện x_i, i=1,…,k, $\sum n_i = n$.

Ký hiệu tần suất của x_i là $f_i = \frac{n_i}{n}$, i=1,…,k.

• *Điểm phần tư* của mẫu, ký hiệu q_1, là trị x_i nhỏ nhất thoả $\sum_{j=1}^{i} f_j \ge \frac{1}{4}$.

• *Điểm ba phần tư* của mẫu, ký hiệu q_3, là trị x_i nhỏ nhất thoả $\sum_{j=1}^{i} f_j \ge \frac{3}{4}$.

• Khoảng $[q_1; q_3]$ gọi là *khoảng phần tư* và trị $\delta = q_3 - q_1$ gọi là *độ lệch phần tư.*

\+ *Ví dụ*. Cho mẫu

x_i	1	2	3	4	5	6	7	8	9	10	11	12
n_i	1	3	8	12	13	16	12	14	9	5	5	2
f_i	0.01	0.03	0.08	0.12	0.13	0.16	0.12	0.14	0.09	0.05	0.05	0.02
$\sum f_j$	0.01	0.04	0.12	0.24	0.37	0.53	0.65	0.79	0.88	0.93	0.98	1

Điểm phần tư $q_1 = 5$, vì $\sum_{j=1}^{4} f_j = 0.24 \le \frac{1}{4} \le 0.37 = \sum_{j=1}^{5} f_j$.

Điểm ba phần tư $q_3 = 8$, vì $\sum_{j=1}^{7} f_j = 0.65 \le \frac{3}{4} \le 0.79 = \sum_{j=1}^{8} f_j$.

Khoảng phần tư là $[q_1; q_3] = [5; 8]$
Độ lệch phần tư là $q_3 - q_1 = 8 - 5 = 3$

b) Trường hợp mẫu phân lớp ghép

$$\{([a_i; a_{i+1}), n_i) \mid 1 \le i \le k\},$$

với $a_1 < a_2 < \ldots < a_{k+1}$ và n_i là số x_i rơi vào khoảng $[a_i; a_{i+1})$, $f_i = \frac{n_i}{n}$, i=1,…,k, $n = \sum n_i$.

• *Hàm tần suất tích luỹ* là hàm có đồ thị là đường gấp khúc nối các điểm

$$(a_i; F(a_i)), \text{ với } F(a_i) = \sum_{j=1}^{i-1} f_j, \quad i=1,\ldots,k,k+1.$$

• *Điểm phần tư* là điểm q_1 thoả $F(q_1) = 1/4$
• *Điểm ba phần tư* là điểm q_3 thoả $F(q_3) = 3/4$
• Khoảng $[q_1; q_3]$ gọi là *khoảng phần tư* và trị $\delta = q_3 - q_1$ gọi là *độ lệch phần tư*.

IV. PHÂN TÍCH THỐNG KÊ BIẾN NGẪU NHIÊN 2 CHIỀU

1. Tổng quát

a) Mẫu phân lớp đơn

Cho đại lượng ngẫu nhiên hai chiều (X, Y), n $\in$ **N**. Mẫu phân lớp đơn cỡ n của (X, Y) có dạng như sau

$$\{((x_i, y_j), n_{ij}) \mid 1 \le i \le r \;\&\; 1 \le j \le s\}$$

trong đó r, s $\in$ N, $n_{ij} \in$ N, $\sum n_{ij} = n$.

Mẫu trên có thể biểu diễn dạng bảng như sau

X \ Y	y_1	y_2	…	y_j	…	y_s
x_1	n_{11}	n_{12}	…	n_{1j}	…	n_{1s}
x_2	n_{21}	n_{22}	…	n_{2j}	…	n_{2s}
:	:	:	…	:	…	:
x_i	n_{i1}	n_{i2}	…	n_{ij}	…	n_{is}
:	:	:	…	:	…	:
x_r	n_{r1}	n_{r2}	…	n_{rj}	…	n_{rs}

• *Tần suất* cặp (x_i, y_j) là đại lượng $f_{ij} = \frac{n_{ij}}{n}$, $1 \le i \le r$, $1 \le j \le s$.

Ký hiệu

$$n_{i,\bullet} = \sum_{j=1}^{s} n_{ij}\ ,\ 1 \le i \le r$$

và

$$n_{\bullet,j} = \sum_{i=1}^{r} n_{ij}\ ,\ 1 \le j \le s$$

Ta có hai mẫu của X và Y là

$$\{(x_i, n_{i,\bullet}) \mid 1 \le i \le r\}$$

và

$$\{(y_j, n_{\bullet,j}) \mid 1 \le j \le s\}$$

• *Tần suất có điều kiện* của x_i với điều kiện y_j là đại lượng $\frac{n_{ij}}{n_{\bullet,j}}$

• *Tần suất có điều kiện* của y_j với điều kiện x_i là đại lượng $\dfrac{n_{ij}}{n_{i,\bullet}}$

• Các tham số đặc trưng.

- Trị trung bình mẫu:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{r} n_{i,\bullet}.x_i \qquad \text{và} \qquad \overline{y} = \frac{1}{n}\sum_{j=1}^{s} n_{\bullet,j}.y_j$$

- Phương sai mẫu:

$$\hat{S}_X^2 = \frac{1}{n}\sum_{i=1}^{r} n_{i,\bullet}\left(x_i - \overline{x}\right)^2 \qquad \text{và} \qquad \hat{S}_Y^2 = \frac{1}{n}\sum_{j=1}^{s} n_{\bullet,j}\left(y_j - \overline{y}\right)^2$$

- Hiệp phương sai mẫu:

$$\hat{S}_{XY} = \frac{1}{n}\sum_{i=1}^{r}\sum_{j=1}^{s} n_{i,j}\left(x_i - \overline{x}\right)\left(y_j - \overline{y}\right) = \frac{1}{n}\sum_{i=1}^{r}\sum_{j=1}^{s} n_{i,j}.x_i.y_j - \overline{x}.\overline{y}$$

- Hệ số tương quan mẫu:

$$R_{XY} = \frac{\hat{S}_{XY}}{\hat{S}_X.\hat{S}_Y}$$

b) Mẫu phân lớp ghép

Cho đại lượng ngẫu nhiên hai chiều (X, Y), n ∈ **N**. Mẫu phân lớp ghép cỡ n của (X, Y) có dạng như sau

$$\{(([a_i;a_{i+1}), [b_j;b_{j+1})), n_{ij}) \mid 1 \le i \le r \;\&\; 1 \le j \le s\}$$

trong đó r, s ∈ N, n_{ij} ∈ N, $\sum n_{ij} = n$.

Mẫu trên có thể biểu diễn dạng bảng như sau

Y \ X	$[b_1;b_2)$	$[b_2;b_3)$	…	$[b_j;b_{j+1})$	…	$[b_s;b_{s+1})$
$[a_1;a_2)$	n_{11}	n_{12}	…	n_{1j}	…	n_{1s}
$[a_2;a_3)$	n_{21}	n_{22}	…	n_{2j}	…	n_{2s}
:	:	:	…	:	…	:
$[a_i;a_{i+1})$	n_{i1}	n_{i2}	…	n_{ij}	…	n_{is}
:	:	:	…	:	…	:
$[a_r;a_{r+1})$	n_{r1}	n_{r2}	…	n_{rj}	…	n_{rs}

• *Tần suất* cặp lớp $([a_i;a_{i+1}), [b_j;b_{j+1}))$ là đại lượng $f_{ij} = \dfrac{n_{ij}}{n}$, $1 \le i \le r$, $1 \le j \le s$.

Ký hiệu

$$n_{i,\bullet} = \sum_{j=1}^{s} n_{ij}, \ 1 \le i \le r$$

và

$$n_{\bullet,j} = \sum_{i=1}^{r} n_{ij}, \ 1 \le j \le s$$

Ta có hai mẫu của X và Y là

$$\{([a_i;a_{i+1}), n_{i,\bullet}) \mid 1 \le i \le r\}$$

và

$$\{([b_j;b_{j+1}), n_{\bullet,j}) \mid 1 \le j \le s\}$$

• *Tần suất có điều kiện* của $[a_i;a_{i+1})$ với điều kiện $[b_j;b_{j+1})$ là đại lượng $\dfrac{n_{ij}}{n_{\bullet,j}}$

• *Tần suất có điều kiện* của $[b_j;b_{j+1})$ với điều kiện $[a_i;a_{i+1})$ là đại lượng $\dfrac{n_{ij}}{n_{i,\bullet}}$.

Ký hiệu $x_i = (a_i + a_{i+1})/2$, $i=1,\dots,r$ và $y_j = (b_j + b_{j+1})/2$. Ta định nghĩa các tham số đặc trưng tương tự như trường hợp phân lớp đơn.

• Các tham số đặc trưng.

- Trị trung bình mẫu:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{r} n_{i,\bullet}.x_i \qquad \text{và} \qquad \overline{y} = \frac{1}{n}\sum_{j=1}^{s} n_{\bullet,j}.y_j$$

- Phương sai mẫu:

$$\hat{S}_X^2 = \frac{1}{n}\sum_{i=1}^{r} n_{i,\bullet}\left(x_i - \overline{x}\right)^2 \qquad \text{và} \qquad \hat{S}_Y^2 = \frac{1}{n}\sum_{j=1}^{s} n_{\bullet,j}\left(y_j - \overline{y}\right)^2$$

- Hiệp phương sai mẫu:

$$\hat{S}_{XY} = \frac{1}{n}\sum_{i=1}^{r}\sum_{j=1}^{s} n_{i,j}\left(x_i - \overline{x}\right)\left(y_j - \overline{y}\right) = \frac{1}{n}\sum_{i=1}^{r}\sum_{j=1}^{s} n_{i,j}.x_i.y_j - \overline{x}.\overline{y}$$

- Hệ số tương quan mẫu:

$$R_{XY} = \frac{\hat{S}_{XY}}{\hat{S}_X.\hat{S}_Y}$$

+ *Ví dụ.*

Để xác định mối quan hệ giữa chi phí quảng cáo và doanh số bán hàng người ta thống kê số liệu trong 10 tháng như sau:

i	1	2	3	4	5	6	7	8	9	10
p_i	480	450	480	540	570	420	390	520	470	480
c_i	22	18	20	24	24	22	14	22	18	16

Ở đây p_i và c_i tương ứng là số sản phẩm bán ra và chi phí quảng cáo trong tháng i, i=1,…,10.

Từ bảng trên ta suy ra mẫu thống kê của số sản phẩm bán ra X như sau:

x_i	390	420	450	470	480	520	540	570
$n_{i,\bullet}$	1	1	1	1	3	1	1	1

và mẫu thống kê của chi phí Y như sau:

y_j	14	16	18	20	22	24
$n_{\bullet,j}$	1	1	2	1	3	2

Từ đó ta tính được

$\overline{x}$ = 480; $\overline{y}$ = 20; $\hat{S}_X^2$ = 2600; $\hat{S}_Y^2$ = 10.4; $\hat{S}_{XY}$ = 118; R_{XY} = 0.72

+ *Ví dụ 2*. Bảng sau cho mẫu thống kê điểm 2 môn toán (X) và tin (Y) thang điểm 20 của 100 sinh viên

Y \ X	[0;4)	[4;8)	[8;12)	[12;16)	[16;20)
[0;4)	2	5	2		
[4;8)	1	12	10	3	
[8;12)		3	28	12	1
[12;16)		1	5	10	2
[16;20)				1	2

Mẫu thống kê của X là

$[a_i;a_{i+1})$	[0;4)	[4;8)	[8;12)	[12;16)	[16;20)
$n_{i,\bullet}$	9	26	44	18	3
$x_i=(a_i+a_{i+1})/2$	2	6	10	14	18

và Y là

$[b_j;b_{j+1})$	[0;4)	[4;8)	[8;12)	[12;16)	[16;20)
$n_{\bullet,j}$	3	21	45	26	5
$y_j=(b_j+b_{j+1})/2$	2	6	10	14	18

Từ đó ta tính được

$\overline{x} = 9.20$; $\overline{y} = 10.36$; $\hat{S}_X^2 = 14.08$; $\hat{S}_Y^2 = 12.5104$; $\hat{S}_{XY} = 8.608$; $R_{XY} = 0.65$

c) Đám mây điểm

Để biểu diễn mẫu 2 chiều người ta dùng khái niệm *đám mây điểm*.
Cho mẫu phân lớp đơn cỡ n của (X, Y)

$$\{((x_i, y_j), n_{ij}) \mid 1 \le i \le r \ \& \ 1 \le j \le s\}$$

trong đó $r, s \in N$, $n_{ij} \in N$, $\sum n_{ij} = n$.

Mỗi cặp (x_i, y_j) với tần suất n_{ij} được biểu diễn bằng n_{ij} điểm tụ xung quah điểm $M_{ij}(x_i, y_j)$ hoặc bằng hình tròn tâm $M_{ij}(x_i, y_j)$ bán kính tỉ lệ thuận với n_{ij}.

Hình tạo ra gọi là đám mây điểm biểu diễn mẫu 2 chiều. Điểm $G(\overline{x}, \overline{y})$ gọi là *tâm điểm* của đám mây điểm.

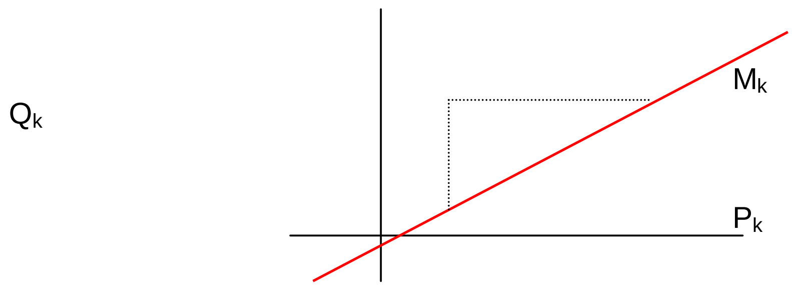
Khái niệm đám mây điểm biểu diễn mẫu 2 chiều phân lớp ghép cũng định nghĩa tương tự .

2. Điều chỉnh tuyến tính

Điều chỉnh tuyến tính là tìm đường thẳng điều chỉnh đám mây điểm biểu diễn phân phối mẫu của vectơ ngẫu nhiên (X,Y).

Ta áp dụng phương pháp bình phương nhỏ nhất.

Ký hiệu Δ là đường thẳng có phương trình $y = a.x + b$ ($a \neq 0$). Với mỗi điểm $M_k(x_k, y_k)$ trên đám mây điểm ta ký hiệu $P_k(x_k, a.x_k+b)$, $Q_k((y_k-b)/a, y_k)$ là các điểm chiếu của M_k lên Δ theo Ox và Oy.



a) Đường thẳng hồi qui của y theo x.

Đường thẳng hồi qui của y theo x là đường thẳng có hệ số a, b làm cực tiểu tổng

$$S(a,b) = \sum_{k=1}^{n} \overline{M_k P_k}^2 = \sum_{k=1}^{n} (y_k - a.x_k - b)^2$$

Giải hệ sau theo a và b

$$\begin{cases} \dfrac{\partial S}{\partial a} = 2b\sum_{k=1}^{n} x_k - 2\sum_{k=1}^{n} x_k (y_k - a.x_k) & = 0 \\ \dfrac{\partial S}{\partial b} = 2nb - 2\sum_{k=1}^{n} (y_k - a.x_k) & = 0 \end{cases}$$

Khử b ta có

$$a = \frac{\dfrac{1}{n}\sum_{k=1}^{n} x_k y_k - \overline{x}.\overline{y}}{\dfrac{1}{n}\sum_{k=1}^{n} x_k^2 - \overline{x}^2} = \frac{\hat{S}_{XY}}{\hat{S}_X^2}$$

Từ đó suy ra

$$b = \overline{y} - a.\overline{x} = \overline{y} - \frac{\hat{S}_{XY}}{\hat{S}_X^2}\overline{x}$$

Vì đây là điểm duy nhất có các đạo hàm triệt tiêu và S(a,b) > 0 bị chặn dưới nên nó cũng là điểm cực tiểu. Vậy phương trinh đường thẳng hồi qui Δ của y theo x là

$$y-\overline{y}=\frac{\hat{S}_{XY}}{\hat{S}_X^2}\left(x-\overline{x}\right)$$

b) Đường thẳng hồi qui của x theo y.

Đường thẳng hồi qui của x theo y là đường thẳng có hệ số a, b làm cực tiểu tổng

$$S(a,b) = \sum_{k=1}^{n}\overline{M_kQ_k}^2=\sum_{k=1}^{n}\left(x_k-\frac{y_k-b}{a}\right)^2$$

Tương tự như trên ta tính được phương trinh đường thẳng hồi qui Δ' của x theo y là

$$x-\overline{x}=\frac{\hat{S}_{XY}}{\hat{S}_Y^2}\left(y-\overline{y}\right)$$

➯ Ghi chú:

- Trong trường hợp phân lớp đơn ta coi lớp (x_{ij}, n_{ij}) có n_{ij} điểm trùng nhau và các phương trình đường thẳng hồi qui vẫn đúng.
- Các đường thẳng Δ và Δ' giao nhau tại điểm G($\overline{x}$, $\overline{y}$) và có các hệ số góc cùng dấu với $\hat{S}_{XY}$.
- Các đường thẳng Δ và Δ' trùng nhau khi và chỉ khi

$$\frac{\hat{S}_{XY}}{\hat{S}_X^2}=\frac{\hat{S}_Y^2}{\hat{S}_{XY}} \quad\Leftrightarrow\quad \hat{S}_{XY}^2=\hat{S}_X^2.\hat{S}_Y^2 \quad\Leftrightarrow\quad R_{XY}^2=1$$

Nếu R_{XY} gần 1, thì ta nói X và Y *tương quan tốt*.

V. PHÂN PHỐI XÁC SUẤT CỦA ĐẠI LƯỢNG THỐNG KÊ TRÊN KHÔNG GIAN MẪU

1. Khái niệm phân phối xác suất của đại lượng thống kê

Cho đại lượng ngẫu nhiên X có mật độ f(x). Giả sử $(x_1, x_2, \ldots, x_n)$ là mẫu của X và có mật độ $f(x_1). \ldots . f(x_n)$. Một hàm $Y = g(x_1, x_2, \ldots, x_n)$ bất kỳ gọi là đại lượng thống kê trên không gian mẫu. $Y = g(x_1, x_2, \ldots, x_n)$ cũng là đại lượng ngẫu nhiên.

+ *Ví dụ*. kỳ vọng mẫu $\overline{x}$ và phương sai mẫu $\hat{S}^2$ là các đại lượng thống kê.

Vấn đề đặt ra là tìm hàm phân phối H(y) của Y.

Ta có

$$H(y) = \int_{G_y} f(x_1)\ldots f(x_n)dx_1\ldots dx_n$$

với

$$G_y = \{ (x_1, x_2, \ldots, x_n) \mid g(x_1, x_2, \ldots, x_n) \leq y \}$$

2. Phân phối xác suất của một số đại lượng thống kê

a) Phân phối xác suất của kỳ vọng mẫu

• *Định lý 1*. Nếu mẫu $(x_1, x_2, \ldots, x_n)$ được lấy từ đại lượng ngẫu nhiên X có phân phối chuẩn $N(\theta, \sigma^2)$, thì

(i) $\overline{x} = \frac{1}{n}\sum_{k=1}^{n} x_k$ có phân phối chuẩn $N(\theta, \frac{\sigma^2}{n})$

(ii) $\frac{\overline{x} - \theta}{\sigma}\sqrt{n}$ có phân phối chuẩn N(0, 1)

b) Phân phối χ^2 (khi bình)

• *Định nghĩa*. Nếu X_i, $1 \leq i \leq n$, là các đại lượng ngẫu nhiên có cùng phân phối chuẩn N(0, 1), thì biến ngẫu nhiên

$$U = \sum_{k=1}^{n} X_k^2$$

có phân phối *khi bình với n bậc tự do* có ký hiệu là χ_n^2.

• *Định lý 2*. Cho biến ngẫu nhiên U có phân phối χ_n^2. Khi đó

(i) Hàm mật độ của U là

$$f(u) = \begin{cases} \dfrac{1}{2^{n/2}\Gamma\left(\dfrac{n}{2}\right)} u^{\frac{n}{2}-1}.e^{-\frac{u}{2}} & ,u > 0 \\ 0 & ,u \leq 0 \end{cases}$$

(ii) E(U) = n; D(U) = 2.n

➯ Ghi chú. $\Gamma(a) = \int_0^{\infty} x^{a-1}e^{-x}dx$ (a > 0).

• *Định lý 3*. Cho biến ngẫu nhiên X có phân phối chuẩn $N(\theta, \sigma^2)$ và $(x_1, x_2, \ldots, x_n)$ là mẫu của X. Khi đó đại lượng thống kê

$$\frac{n-1}{\sigma^2}s^2$$

có phân phối χ^2_{n-1}, trong đó

$$s^2 = \frac{1}{n-1}\sum_{k=1}^{n}\left(x_k - \overline{x}\right)^2$$

c) Phân phối student

• *Định nghĩa*. Cho biến ngẫu nhiên Z có phân phối chuẩn N(0,1) và U có phân phối χ^2_n (n≥ 1) độc lập với nhau. Khi đó biến ngẫu nhiên

$$t = \frac{Z}{\sqrt{U}}\sqrt{n}$$

tuân theo luật *phân phối student với n bậc tự do*.

• *Định lý 4*. Cho t tuân theo luật phân phối student với n bậc tự do (n≥ 1). Khi đó

(i) Hàm mật độ của t là

$$f(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\pi.n}\,\Gamma\left(\frac{n}{2}\right)}.\frac{1}{\left(1+\frac{t^2}{n}\right)^{\frac{n+1}{2}}} \quad \forall t \in (-\infty, +\infty)$$

(ii) Với n > 1: E(t) = 0 (f(t) là hàm chẵn).

Với n > 2: $D(t) = \frac{n}{n-2}$

• *Định lý 5*. Cho X tuân theo luật phân phối chuẩn $N(\theta, \sigma^2)$ và $(x_1, x_2, \ldots, x_n)$ (n≥ 1) là mẫu của X. Khi đó đại lượng thống kê

$$t = \frac{\overline{x}-\theta}{s}.\sqrt{n}$$

có phân phối student với n-1 bậc tự do, trong đó

$$s^2 = \frac{1}{n-1}\sum_{k=1}^{n}\left(x_k - \overline{x}\right)^2$$

CM.

Suy ra từ định lý 3 và định nghĩa vì

$$t = \frac{\overline{x}-\theta}{s}.\sqrt{n} = \frac{\dfrac{\overline{x}-\theta}{\sigma}\sqrt{n}}{\dfrac{\sqrt{n-1}.s}{\sigma}}.\sqrt{n-1}$$

d) Phân phối Fisher

• *Định nghĩa*. Cho các biến ngẫu nhiên độc lập U_1 có phân phối χ^2_{n1} và U_2 có phân phối χ^2_{n2} (n1, n2 ≥ 1). Khi đó biến ngẫu nhiên

$$F = \frac{\dfrac{U_1}{n1}}{\dfrac{U_2}{n2}}$$

tuân theo luật *phân phối Fisher với cặp bậc tự do (n1, n2)*, ký hiệu là $F_{n1,n2}$.

• *Định lý 6*. Cho biến ngẫu nhiên X có phân phối $F_{n1,n2}$. Khi đó

(i) Hàm mật độ của X là

$$f(t) = \begin{cases} \dfrac{\Gamma\left(\dfrac{n1+n2}{2}\right)\left(\dfrac{n1}{n2}\right)^{\frac{n1}{2}}}{\Gamma\left(\dfrac{n1}{2}\right)\Gamma\left(\dfrac{n2}{2}\right)} t^{\frac{n1}{2}-1}.\left(1+\dfrac{n1}{n2}t\right)^{-\frac{n1+n2}{2}} & ,t > 0 \\ 0 & ,t \le 0 \end{cases}$$

(ii) $E(X) = \dfrac{n2}{n2-2}$ ∀n2 > 2; $D(X) = \dfrac{2.n2^2.(n1+n2-2)}{n1.(n2-4).(n2-2)^2}$ ∀n2 > 4

Bây giờ ta cho $(x_1, x_2, \ldots, x_{n1})$ là mẫu của X , $(y_1, y_2, \ldots, y_{n2})$ là mẫu của Y và

$$\overline{x} = \frac{1}{n1}\sum_{i=1}^{n1} x_i \ ; \ s_1^2 = \frac{1}{n1-1}\sum_{k=1}^{n1}\left(x_k - \overline{x}\right)^2$$

và

$$\overline{y} = \frac{1}{n2}\sum_{i=1}^{n2} y_i \ ; \ s_2^2 = \frac{1}{n2-1}\sum_{k=1}^{n2}\left(y_k - \overline{y}\right)^2$$

• *Định lý 7*. Cho X và Y là các biến ngẫu nhiên độc lập có phân phối chuẩn cùng phương sai (D(X) = D(Y)). Khi đó đại lượng thống kê

$$F = \frac{s_1^2}{s_2^2}$$

có phân phối Fisher $F_{n1-1,n2-1}$.

CM.

Suy ra từ định lý 3 và định nghĩa.

VI. PHÂN PHỐI TIỆM CẬN CHUẨN CỦA ĐẠI LƯỢNG THỐNG KÊ

Theo các định lý giới hạn, khi cỡ mẫu n tăng đến vô cùng thì có thể chứng minh nhiều đại lượng thống kê có hàm phân phối xác suất tiến tới hàm phân phối chuẩn. Các phân phối đó gọi là phân phối tiệm cận chuẩn.

• *Định lý 1*. Cho đại lượng ngẫu nhiên X với E(X) = θ và D(X) = σ^2 và $(x_1, x_2, \ldots, x_n)$ là mẫu của X. Khi đó

$$\frac{\overline{x}-\theta}{\sigma}.\sqrt{n}$$

có phân phối tiến tới phân phối chuẩn N(0,1) khi $n \to +\infty$.

Từ định lý 1 suy ra

• *Định lý 2*. Cho sự kiện A của phép thử α có xác suất p và n ≥ 1. Giả sử phép thử α được thực hiện n lần một cách độc lập và sự kiện A xuất hiện m lần. Khi đó

$$\frac{\frac{m}{n}-p}{\sqrt{p.(1-p)}}.\sqrt{n}$$

có phân phối tiến tới phân phối chuẩn N(0,1) khi $n \to +\infty$.

• *Định lý 3*. Nếu đại lượng ngẫu nhiên U có phân phối χ_n^2, thì các đại lượng

$$\frac{U-n}{\sqrt{2.n}} \quad \text{và} \quad \left(\sqrt{2.U}-\sqrt{2.n-1}\right)$$

có phân phối tiến tới phân phối chuẩn N(0,1) khi $n \to +\infty$.

• *Định lý 4*. Nếu đại lượng ngẫu nhiên t có phân phối student với n bậc tự do , thì phân phối xác suất của t tiến tới phân phối chuẩn N(0,1) khi $n \to +\infty$.

✎ Ghi chú. Với n ≥ 30 phân phối student được coi là trùng với phân phối chuẩn N(0,1).

CHƯƠNG 6
LÝ THUYẾT ƯỚC LƯỢNG

0. BÀI TOÁN ƯỚC LƯỢNG

Cho X là đại lượng ngẫu nhiên có luật phân phối P(x, θ). Giả thiết dạng của P đã biết, nhưng tham số θ chưa biết và ta cần tìm cách ước lượng θ. Có hai phương pháp tiếp cận: *ước lượng điểm* và *ước lượng khoảng*.

1. Ước lượng điểm

Ước lượng điểm là dựa trên mẫu $(x_1, x_2, \ldots, x_n)$ của X, ta tìm đại lượng thống kê

$$\hat{\theta}(x_1, x_2, \ldots, x_n)$$

thay cho θ với độ chính xác nào đó.

Đại lượng $\hat{\theta}(x_1, x_2, \ldots, x_n)$ gọi là *hàm ước lượng* của θ.

2. Ước lượng khoảng

Ước lượng khoảng là dựa trên mẫu $(x_1, x_2, \ldots, x_n)$ của X, ta tìm khoảng

$$[\hat{\theta}_1, \hat{\theta}_2]$$

trong đó $\hat{\theta}_1 = \hat{\theta}_1(x_1, x_2, \ldots, x_n)$ và $\hat{\theta}_2 = \hat{\theta}_2(x_1, x_2, \ldots, x_n)$, sao cho có thể coi

$$\hat{\theta}_1 \leq \theta \leq \hat{\theta}_2$$

với độ tin cậy nào đó.

1. ƯỚC LƯỢNG ĐIỂM

1. Hàm ước lượng của một tham số

Cho biến ngẫu nhiên X với luật phân phối P(x, θ) và mẫu (x_1, x_2, …, x_n) của X.

• *Định nghĩa 1*. Đại lượng thống kê $\hat{\theta}$(x_1, x_2, …, x_n) được chọn sử dụng thay cho θ gọi là *hàm ước lượng* của θ.

+ *Ví dụ 1*. Giả sử E(X) = μ và D(X) = σ^2. Ta có thể coi $\overline{x} = \frac{1}{n}\sum_{k=1}^{n} x_k$ là ước lượng của μ và $\hat{S}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2$ là ước lượng của σ^2.

Ứng với mỗi tham số θ có thể có nhiều hàm ước lượng khác nhau. Vấn đề đặt ra là phải chọn hàm ước lượng theo tiêu chuẩn nào để có thể coi là tốt.

• *Định nghĩa 2*. Hàm ước lượng $\hat{\theta}$(x_1, x_2, …, x_n) của θ gọi là *ước lượng không chệch* nếu

$$E[\hat{\theta}(x_1, x_2, \ldots, x_n)] = \theta$$

với mọi θ trong khoảng xác định H nào đó.

Nếu coi $\hat{\theta}$(x_1, x_2, …, x_n) − θ là sai số của ước lượng thì điều kiện trên chứng tỏ kỳ vọng sai số bằng 0.

+ *Ví dụ 2*. Kỳ vọng mẫu $\overline{x}$ là ước lượng không chệch của kỳ vọng μ. Thật vậy, ta có

$$E(\overline{x}) = E\left(\frac{1}{n}\sum_{k=1}^{n} x_k\right) = \frac{1}{n}\sum_{k=1}^{n} E(x_k) = \mu$$

+ *Ví dụ 3*. đại lượng thống kê $\hat{S}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2$ là ước lượng chệch của phương sai σ^2. Thật vậy, ta có

$$E\left(\hat{S}^2\right) = \frac{1}{n}E\left[\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2\right] = \frac{1}{n}E\left[\sum_{i=1}^{n}\left(\left(x_i - \overline{x}\right) - \left(\overline{x} - E\left(\overline{x}\right)\right)\right)^2\right]$$

$$\Rightarrow \quad E\left(\hat{S}^2\right) = \frac{1}{n}E\left[\sum_{i=1}^{n}\left(\left(x_i - \overline{x}\right)^2 + \left(\overline{x} - E\left(\overline{x}\right)\right)^2 - 2.\left(x_i - \overline{x}\right).\left(\overline{x} - E\left(\overline{x}\right)\right)\right)^2\right]$$

$$\Rightarrow \quad E\left(\hat{S}^2\right) = \frac{1}{n}\left[\sum_{i=1}^{n}\left(E\left(x_i - \overline{x}\right)^2 + E\left(\overline{x} - E\left(\overline{x}\right)\right)^2 - 2.E\left(x_i - \overline{x}\right).\left(\overline{x} - E\left(\overline{x}\right)\right)\right)^2\right]$$

$\Rightarrow \qquad E\left(\hat{S}^2\right) = \sigma^2 - \dfrac{\sigma^2}{n} = \dfrac{n-1}{n}\sigma^2 \neq \sigma^2$

↳ Ghi chú. Vì (n−1)/n →1 khi n→+∞, nên với n > 50 ta có thể coi

$$\hat{S}^2 \approx s^2 = \frac{1}{n-1}\sum_{k=1}^{n}\left(x_k - \bar{x}\right)^2$$

• *Độ chính xác của các ước lượng không chệch.*

Giả sử $\hat{\theta}(x_1, x_2, \ldots, x_n)$ là ước lượng không chệch của θ và

$$D[\hat{\theta}(x_1, x_2, \ldots, x_n)] = \delta^2$$

Khi đó, theo bất đẳng thức Trebưsep, với mọi ε > 0, ta có

$$P\{|\hat{\theta}(x_1, x_2, \ldots, x_n) - \theta| < \varepsilon.\delta\} \geq 1 - \frac{\delta^2}{\varepsilon^2.\delta^2} = 1 - \frac{1}{\varepsilon^2}$$

Nếu chọn ε = 3 thì

$$P\{|\hat{\theta}(x_1, x_2, \ldots, x_n) - \theta| < 3.\delta\} \geq 1 - 1/9 \approx 0.889$$

Công thức trên đúng với mọi phân phối xác suất của X. Nếu $\hat{\theta}(x_1, x_2, \ldots, x_n)$ có phân phối chuẩn $N(\theta, \delta^2)$ thì ta có

$$P\{|\hat{\theta}(x_1, x_2, \ldots, x_n) - \theta| < 3.\delta\} \approx 0.997$$

Trong thực tế người ta viết

$$|\hat{\theta}(x_1, x_2, \ldots, x_n) - \theta| < 3.\delta$$

và gọi đó là công thức 3δ.

Ở ước lượng khoảng ta sẽ nghiên cứu độ chính xác triệt để hơn.

• *Định nghĩa 3*. Ước lượng không chệch $\hat{\theta}(x_1, x_2, \ldots, x_n)$ của θ gọi là *ước lượng hiệu quả* trên khoảng H của θ, nếu với mọi ước lượng không chệch $T(x_1, x_2, \ldots, x_n)$ của θ ta có

$$D[\hat{\theta}(x_1, x_2, \ldots, x_n)] \leq D[T(x_1, x_2, \ldots, x_n)] \qquad \forall\, \theta \in H$$

• *Định lý 1* (bất đẳng thức Cramer-Rao). Cho biến ngẫu nhiên X có mật độ f(x, θ), $(x_1, x_2, \ldots, x_n)$ là mẫu cỡ n của X thoả một số điều kiện nhất định và $\hat{\theta}(x_1, x_2, \ldots, x_n)$ là hàm ước lượng không chệch của θ. Khi đó

$$D[\hat{\theta}(x_1, x_2, \ldots, x_n)] \geq \frac{1}{n.E\left[\frac{\partial \ln f(x,\theta)}{\partial \theta}\right]^2}$$

+ *Ví dụ 4*. Cho biến ngẫu nhiên X có phân phối chuẩn $N(\mu, \sigma^2)$. Ta chỉ ra rằng $\bar{x}$ là ước lượng hiệu quả của μ. Thật vậy, vì

$$\frac{\partial \ln f(x,\mu)}{\partial \mu} = \frac{x-\mu}{\sigma^2}$$

nên

$$\frac{1}{n.E\left[\frac{\partial \ln f(x,\mu)}{\partial \mu}\right]^2} = \frac{1}{n.E\left[\frac{(x-\mu)^2}{\sigma^4}\right]^2} = \frac{\sigma^2}{n}$$

Mặt khác ta biết rằng $D(\bar{x}) = \frac{\sigma^2}{n}$. Suy ra $\bar{x}$ thoả bất đẳng thức Cramer-Rao. Vậy $\bar{x}$ là ước lượng hiệu quả của μ.

• *Định nghĩa 4*. Hàm ước lượng $\hat{\theta}(x_1, x_2, \ldots, x_n)$ của θ gọi là *ước lượng vững* nếu

$$\forall \varepsilon > 0\ \forall\ \theta \in H,\ \lim_{n\to+\infty} P\{\,|\hat{\theta}(x_1, x_2, \ldots, x_n) - \theta| < \varepsilon\,\} = 1$$

trong đó xác suất P được tính theo θ.

• *Định lý 2*. Cho $\hat{\theta}(x_1, x_2, \ldots, x_n)$ là hàm ước lượng của θ thoả

(i) $\hat{\theta}(x_1, x_2, \ldots, x_n)$ là ước lượng không chệch của θ hoặc

$$\lim_{n\to+\infty} [E(\hat{\theta}(x_1, x_2, \ldots, x_n)) - \theta] = 0$$

(ii) $\lim_{n\to+\infty} D[\hat{\theta}(x_1, x_2, \ldots, x_n)] = 0$

Khi đó $\hat{\theta}(x_1, x_2, \ldots, x_n)$ là ước lượng vững của θ.

CM.

Áp dụng bất đẳng thức Trebưsep ta có

$$\forall\ \varepsilon > 0,\ P\{|\hat{\theta}(x_1, x_2, \ldots, x_n) - E[\hat{\theta}(x_1, x_2, \ldots, x_n)]\,| < \varepsilon\,\} \geq 1 - \frac{D\left[\hat{\theta}(x_1,\ldots,x_n)\right]}{\varepsilon^2}$$

Từ đó, sử dụng (ii), suy ra

$$\forall\ \varepsilon > 0,\ \lim_{n\to+\infty} P\{|\hat{\theta}(x_1, x_2, \ldots, x_n) - E[\hat{\theta}(x_1, x_2, \ldots, x_n)]\,| < \varepsilon\,\} = 1 \quad (*)$$

Nếu $\hat{\theta}(x_1, x_2, \ldots, x_n)$ là ước lượng không chệch, tức $E[\hat{\theta}(x_1, x_2, \ldots, x_n)] = \theta$, thì theo định nghĩa nó là ước lượng vững.

Ta xét trường hợp

$$\lim_{n\to+\infty} [E(\hat{\theta}(x_1, x_2, \ldots, x_n)) - \theta] = 0$$

Cho $\varepsilon > 0$ bất kỳ. Tồn tại n_ε thoả

$$|E(\hat{\theta}(x_1, x_2, \ldots, x_n)) - \theta| < \varepsilon/2, \qquad \forall\ n \geq n_\varepsilon$$

Mặt khác, từ bất đẳng thức

$$|\hat{\theta}(x_1, x_2, \ldots, x_n) - \theta| \leq |\hat{\theta}(x_1, x_2, \ldots, x_n) - E[\hat{\theta}(x_1, x_2, \ldots, x_n)]| + |E(\hat{\theta}(x_1, x_2, \ldots, x_n)) - \theta|$$

suy ra: Với mọi $n \geq n_\varepsilon$, sự kiện

$$|\hat{\theta}(x_1, x_2, \ldots, x_n) - E[\hat{\theta}(x_1, x_2, \ldots, x_n)]| < \varepsilon/2$$

kéo theo sự kiện

$$|\hat{\theta}(x_1, x_2, \ldots, x_n) - \theta| \leq \varepsilon/2 + \varepsilon/2 = \varepsilon$$

Vậy

$$P\{|\hat{\theta}(x_1, x_2, \ldots, x_n) - \theta| < \varepsilon\} \geq P\{|\hat{\theta}(x_1, x_2, \ldots, x_n) - E[\hat{\theta}(x_1, x_2, \ldots, x_n)]| < \varepsilon/2\}$$

Từ đó, theo (*), suy ra

$$\lim_{n\to+\infty} P\{|\hat{\theta}(x_1, x_2, \ldots, x_n) - \theta| < \varepsilon\} = 1$$

Vậy $\hat{\theta}(x_1, x_2, \ldots, x_n)$ là ước lượng vững của θ.

+ *Ví dụ 5*. Xét ước lượng $\bar{x}$ của $\mu = E(X)$. Theo ví dụ 1, $\bar{x}$ là ước lượng không chệch của μ. Tiếp theo

$$D(\bar{x}) = D\left[\frac{1}{n}\sum_{i=1}^{n} x_i\right] = \frac{1}{n^2}\sum_{i=1}^{n} D(x_i) = \frac{\sigma^2}{n}$$

Vậy theo định lý trên, $\bar{x}$ là ước lượng vững của μ.

+ *Ví dụ 6*. Trong một lô sản phẩm, cứ lấy 1 sản phẩm thì xác suất lấy phải phế phẩm là p. Người ta lấy n sản phẩm, thì có m phế phẩm. Tìm ước lượng không chệch của p.

Giải.

Gọi X_i, i=1, 2, …, n, là số phế phẩm xuất hiện trong lần lấy sản phẩm thứ i. Rõ ràng X_i là đại lượng ngẫu nhiên có luật phân phối

$$P(X_i = 1) = p \text{ và } P(X_i = 0) = 1 - p$$

Ta có

$$E(X_i) = p \;\&\; D(X_i) = p(1-p) \quad \forall\, i = 1, \ldots, n$$

Như vậy việc lấy n sản phẩm tương đương với việc lấy mẫu có lặp (x_1, x_2, …, x_n). Vậy theo ví dụ 2,

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{m}{n}$$

là ước lượng không chệch của p.

Theo ví dụ 5, *m/n* cũng là ước lượng vững của *p*.

➯ Ghi chú: *m/n* là ước lượng hiệu quả của *p*.

+ *Ví dụ 7*. Trong một xí nghiệp, để biết số đơn vị nguyên liệu cần thiết sản xuất ra 1 thành phẩm người ta lấy mẫu cỡ 20:

3.0; 3.8; 3.1; 3.2; 3.5; 3.2; 3.5; 3.6; 3.3; 3.8 3.5; 3.2; 4.0; 3.6; 3.4; 3.5; 4.3; 3.5; 3.0; 4.0

Gọi X là đại lượng ngẫu nhiên chỉ số lượng đơn vị nguyên liệu cần thiết để sản xuất 1 thành phẩm. Ta cần ước lượng $\mu = E(X)$. Theo ví dụ 2 và ví dụ 3, ta lấy

$$\overline{x} = \frac{1}{20}\sum_{i=1}^{20} x_i = 3.5$$

làm ước lượng của μ và lấy

$$s^2 = \frac{1}{19}\sum_{i=1}^{20}\left(x_i - \overline{x}\right)^2$$

làm ước lượng của $\sigma^2 = D(X)$.

Từ đó ta có thể xấp xỉ phương sai của $\overline{x}$

$$D(\overline{x}) = \sigma^2/n \approx s^2/n = \delta^2.$$

Ta có $\delta = \frac{s}{\sqrt{n}} = 0.8$. Vậy theo công thức 3δ ta được

$$\mu = 3.5 \pm 3{*}0.8 = 3.5 \pm 0.24$$

với xác suất 0.889.

• Kết quả:

Tham số	**Hàm ước lượng** $\hat{\theta}(x_1, x_2, \ldots, x_n)$	**E** $\hat{\theta}(x_1, x_2, \ldots, x_n)$	**D** $\hat{\theta}(x_1, x_2, \ldots, x_n)$	**Tính chất của** $\hat{\theta}(x_1, x_2, \ldots, x_n)$
Kỳ vọng $\mu = E(X)$	$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$	μ	$\frac{\sigma^2}{n}$	- không chệch - vững - hiệu quả, nếu X phân phối

				chuẩn
Xác suất p	m/n	p	p(1−p)/n	- không chệch - vững - hiệu quả
Phương sai σ^2	$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2$	σ^2	$\frac{1}{n}\left(\mu_4 - \frac{n-3}{n-1}\sigma^4\right)$ với $\mu_4 = E(X-\mu)^4$	- không chệch - vững

2. Phương pháp hợp lý cực đại (R.A.Fisher)

Giả sử đại lượng ngẫu nhiên X có hàm mật độ $f(x, \theta)$ với dạng của f đã biết, nhưng θ chưa biết. Để ước lượng θ ta lấy mẫu $(x_1, x_2, \ldots, x_n)$ và lập hàm

$$L(\theta) = f(x_1, \theta) \times \ldots \times f(x_n, \theta) \qquad (1)$$

$L(\theta)$ gọi là hàm hợp lý của mẫu, nó phụ thuộc $x_1, \ldots, x_n$ và θ nhưng coi $x_1, \ldots, x_n$ là hằng và θ là biến. Vấn đề đặt ra là tìm $\hat{\theta}(x_1, x_2, \ldots, x_n)$ sao cho

$$L(\hat{\theta}(x_1, x_2, \ldots, x_n)) \geq L(\theta) \;\; \forall\, \theta \in H \qquad (2)$$

Đặt $\Psi(\theta) = \ln[L(\theta)]$, điều kiện trên tương đương

$$\Psi(\hat{\theta}(x_1, x_2, \ldots, x_n)) \geq \Psi(\theta) \qquad \forall\, \theta \in H \qquad (3)$$

Ước lượng $\hat{\theta}(x_1, x_2, \ldots, x_n)$ xác định bởi điều kiện trên gọi là *ước lượng hợp lý cực đại* của θ.

Nếu $\Psi(\theta)$ khả vi theo θ thì tại $\hat{\theta}(x_1, x_2, \ldots, x_n)$ ta có

$$\frac{d\Psi}{d\theta} = 0 \qquad (4)$$

Phương trình này gọi là *phương trình hợp lý* và mọi nghiệm của nó, nếu thoả (2) hoặc (3) đều là ước lượng hợp lý cực đại của θ.

+ *Ví dụ 1.* Cho biến ngẫu nhiên X có phân phối chuẩn $N(\mu, \sigma^2)$, σ^2 đã biết, μ chưa biết và $(x_1, x_2, \ldots, x_n)$ là mẫu cỡ n của X. Hãy tìm ước lượng hợp lý cực đại của μ.
Giải.

Ta có

$$L(\mu) = \frac{1}{\left(\sqrt{2\pi}\sigma\right)^n} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}$$

$\Rightarrow$

$$\Psi(\mu) = \ln[L(\mu)] = -\, n.\ln\left(\sqrt{2\pi}\sigma\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(x_i - \mu\right)^2$$

$\Rightarrow$

$$\frac{d\Psi(\mu)}{d\mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)$$

Vậy, phương trình hợp lý là

$$\sum_{i=1}^{n}(x_i-\mu) = 0$$

Giải phương trình này ta được ước lượng hợp lý cực đại của μ là

$$\hat{\mu}(x_1, x_2, \ldots, x_n) = \frac{1}{n}\sum_{i=1}^{n}x_i = \bar{x}$$

(vì $\frac{d^2\Psi(\mu)}{d\mu^2} = -\frac{n}{\sigma^2} < 0$, nên tại $\hat{\mu}$ hàm $\Psi(\mu)$ đạt giá trị lớn nhất).

➯ Ghi chú: Lý thuyết trên có thể mở rộng cho trường hợp $\theta = (\theta_1, \ldots, \theta_k)$, trong đó hệ phương trình hợp lý là

$$\frac{\partial\Psi(\theta)}{\partial\theta_i} = 0 \text{ , } i=1, \ldots, k$$

+ *Ví dụ 2*. Cho biến ngẫu nhiên X có phân phối chuẩn $N(\mu, \sigma^2)$, σ^2 và μ đều chưa biết và $(x_1, x_2, \ldots, x_n)$ là mẫu cỡ n của X. Hãy tìm ước lượng hợp lý cực đại của μ.
Giải.

Hệ phương trình hợp lý là

$$\frac{\partial\Psi(\mu,\sigma^2)}{\partial\mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu) = 0$$

$$\frac{\partial\Psi(\mu,\sigma^2)}{\partial\sigma^2} = \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2 - \frac{n}{2\sigma^2} = 0$$

Giải ra ta có

$$\hat{\mu} = \bar{x} \quad \text{và} \quad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2$$

Đạo hàm riêng cấp 2 là

$$\frac{\partial\Psi^2(\mu,\sigma^2)}{\partial\mu^2} = -\frac{n}{\sigma^2} \ ; \qquad \frac{\partial\Psi^2(\mu,\sigma^2)}{\partial\mu\partial\sigma^2} = -\frac{1}{\sigma^4}\sum_{i=1}^{n}(x_i-\mu)$$

$$\frac{\partial\Psi^2(\mu,\sigma^2)}{\partial(\sigma^2)^2} = \frac{n}{2\sigma^4} - \frac{1}{\sigma^6}\sum_{i=1}^{n}(x_i-\mu)^2$$

Thế $\mu = \hat{\mu}$ và $\sigma^2 = \hat{\sigma}^2$ vào các đạo hàm riêng ta có

A = $\dfrac{\partial \Psi^2(\hat{\mu},\hat{\sigma}^2)}{\partial \mu^2} = -\dfrac{n}{\hat{\sigma}^2}$; B = $\dfrac{\partial \Psi^2(\hat{\mu},\hat{\sigma}^2)}{\partial \mu \partial \sigma^2} = -\dfrac{1}{\hat{\sigma}^4}\sum_{i=1}^{n}\left(x_i - \hat{\mu}\right)$ = 0

C = $\dfrac{\partial \Psi^2(\hat{\mu},\hat{\sigma}^2)}{\partial\left(\sigma^2\right)^2} = \dfrac{n}{2\hat{\sigma}^4} - \dfrac{1}{\hat{\sigma}^6}\sum_{i=1}^{n}\left(x_i - \hat{\mu}\right)^2 = \dfrac{n}{2\hat{\sigma}^4} - \dfrac{n.\hat{\sigma}^2}{\hat{\sigma}^6} = -\dfrac{n}{2\hat{\sigma}^4}$

$\Rightarrow$

$B^2 - A.C = -\dfrac{n^2}{2\hat{\sigma}^6}$ và A < 0

Vậy ($\hat{\mu}$, $\hat{\sigma}^2$) là ước lượng hợp lý cực đại.

✎ Ghi chú:

- Trường hợp X là biến ngẫu nhiên rời rạc, ta cũng định nghĩa tương tự khái niệm ước lượng hợp lý cực đại.

- Ước lượng hợp lý cực đại là ước lượng vững (CM). Khi n khá lớn, nó có phân phối tiệm cận chuẩn và khá gần ước lượng hiệu quả.

- Khái niệm ước lượng hợp lý cực đại định nghĩa theo (2) hoặc (3) dựa trên quan điểm “giá trị của θ trong thực tế là giá trị ứng với xác suất xảy ra lớn nhất” (vì vậy nó là hợp lý nhất).

2. ƯỚC LƯỢNG KHOẢNG

• *Định nghĩa*.

Cho biến ngẫu nhiên X có phân phối phụ thuộc tham số θ, $\theta \in$ H, và mẫu (x_1, x_2, …, x_n). Khoảng $[\hat{\theta}_1, \hat{\theta}_2]$, trong đó $\hat{\theta}_1 = \hat{\theta}_1(x_1, x_2, \ldots, x_n)$ và $\hat{\theta}_2 = \hat{\theta}_2(x_1, x_2, \ldots, x_n)$, gọi là *khoảng ước lượng (tin cậy)* của tham số θ với *độ tin cậy* γ, $0 < \gamma < 1$, nếu

$$P\{\hat{\theta}_1 \le \theta \le \hat{\theta}_2\} = \gamma$$

• *Bài toán 1*. Cho biến ngẫu nhiên X có phân phối chuẩn $N(\mu,\sigma^2)$ với σ đã biết và μ chưa biết. Cho mẫu (x_1, x_2, …, x_n) và γ, $0<\gamma<1$. Hãy xác định khoảng tin cậy của μ với độ tin cậy γ.

Giải.

Đại lượng $\frac{\bar{x}-\mu}{\sigma}\sqrt{n}$ có phân phối chuẩn N(0,1). Gọi $\Phi(u)$ là hàm phân phối chuẩn N(0,1), tức

$$\Phi(u) = \int_{-\infty}^{u} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

Ta tìm $u_\gamma > 0$ thoả

$$\gamma = P\{-u_\gamma \le \frac{\bar{x}-\mu}{\sigma}\sqrt{n} \le u_\gamma\} = \Phi(u_\gamma) - \Phi(-u_\gamma) = \Phi(u_\gamma) - [1 - \Phi(u_\gamma)] = 2.\Phi(u_\gamma) - 1$$

Từ đó suy ra

$$u_\gamma = \Phi^{-1}\left(\frac{1+\gamma}{2}\right)$$

Với u_γ ta có

$$\gamma = P\{\bar{x} - u_\gamma \frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + u_\gamma \frac{\sigma}{\sqrt{n}}\}$$

Vậy

$$[\bar{x} - u_\gamma \frac{\sigma}{\sqrt{n}} \ , \ \bar{x} + u_\gamma \frac{\sigma}{\sqrt{n}}]$$

là khoảng ước lượng (tin cậy) của μ với độ tin cậy γ.

+ *Ví dụ*. Đo 25 lần chi tiết máy. Giả sử không có sai số hệ thống và sai số ngẫu nhiên có phân phối chuẩn $N(\mu,\sigma^2)$ với $\sigma = 10$. Biết $\bar{x} = 100$, hãy tìm khoảng tin cậy của chiều dài chi tiết máy với độ tin cậy $\gamma = 0.99$.

Giải.

Ta có

$$u_{0.99} = \Phi^{-1}(0.5 + 0.99/2) = \Phi^{-1}(0.995) = 2.575$$

$\Rightarrow$

$$\bar{x} - u_\gamma \frac{\sigma}{\sqrt{n}} = 100 - 2.575.\frac{10}{\sqrt{25}} = 100 - 5.15 = 94.85 ;$$

$$\overline{x}+u_\gamma\frac{\sigma}{\sqrt{n}} = 100 + 2.575.\frac{10}{\sqrt{25}} = 100 + 5.15 = 105.15$$

Vậy [94.85 ; 105.15] là khoảng tin cậy của μ với độ tin cậy 0.99.

• *Bài toán 2*. Cho biến ngẫu nhiên X có phân phối chuẩn N(μ,σ^2) với σ đã biết và μ chưa biết. Cho d > 0 và $\gamma \in$ R, 0<γ<1, phải lấy mẫu cỡ (n) nhỏ nhất là bao nhiêu để ước lượng $\overline{x}$ của μ không sai khác với μ quá d đơn vị với độ tin cậy γ (P{|$\overline{x}$ − μ| ≤ d} ≥ γ) ?

Giải.

Với

$$u_\gamma = \Phi^{-1}\left(\frac{1+\gamma}{2}\right)$$

ta có

$$\gamma = P\{\ \overline{x}-u_\gamma\frac{\sigma}{\sqrt{n}} \le \mu \le \overline{x}+u_\gamma\frac{\sigma}{\sqrt{n}}\}$$

⇒

$$\gamma = P\{\ |\overline{x} - \mu| \le u_\gamma\frac{\sigma}{\sqrt{n}}\ \}$$

Vậy để

$$P\{|\overline{x} - \mu| \le d\} \ge \gamma$$

n phải thoả

$$u_\gamma\frac{\sigma}{\sqrt{n}} \le d \quad \Rightarrow \quad n \ge u_\gamma^{\,2}\frac{\sigma^2}{d^2}$$

Suy ra n nhỏ nhất là

$$n = \left\lceil \frac{u_\gamma^{\,2}\sigma^2}{d^2} \right\rceil$$

trong đó $\lceil x \rceil$ ký hiệu số nguyên nhỏ nhất ≥ x.

+ *Ví dụ*. Cho biến ngẫu nhiên X có phân phối chuẩn N(μ,σ^2) với σ^2 = 25. Phải lấy mẫu cỡ (n) nhỏ nhất là bao nhiêu để ước lượng $\overline{x}$ của μ không sai khác với μ quá d = 1 đơn vị với độ tin cậy γ = 0.95.

Giải.

Theo trên n nhỏ nhất là

$$n = \left\lceil \frac{u_\gamma^{\,2}\sigma^2}{d^2} \right\rceil = \left\lceil \frac{u_{0.95}^{\,2}.25}{1} \right\rceil = \lceil 1.96^2.25 \rceil = \lceil 96.04 \rceil = 97$$

• *Bài toán 3*. Cho biến ngẫu nhiên X có phân phối chuẩn N(μ,σ^2) với σ và μ chưa biết. Cho mẫu (x_1, x_2, …, x_n) và γ, 0<γ<1. Hãy xác định khoảng tin cậy của μ với độ tin cậy γ.

Giải.

Ta xét đại lượng thống kê

$$\frac{\overline{x}-\mu}{S}\sqrt{n} \qquad (s^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i-\overline{x}\right)^2)$$

Đại lượng này có hàm phân phối Student với n−1 bậc tự do, ký hiệu là $T_k(t)$. Tương tự như bài toán 1, với

$$t_{n-1,\gamma} = T_{n-1}^{-1}\left(\frac{1+\gamma}{2}\right)$$

ta có

$$\gamma = P\{ \overline{x}-t_{n-1,\gamma}\frac{s}{\sqrt{n}} \le \mu \le \overline{x}+t_{n-1,\gamma}\frac{s}{\sqrt{n}} \}$$

Vậy

$$[\overline{x}-t_{n-1,\gamma}\frac{s}{\sqrt{n}} \ , \ \overline{x}+t_{n-1,\gamma}\frac{s}{\sqrt{n}}]$$

là khoảng tin cậy của μ với độ tin cậy γ.

+ *Ví dụ*. Một giống lúa gieo trên 10 miếng đất thí nghiệm có điều kiện giống nhau, cho sản lượng tính theo cùng đơn vị như sau

25.4; 28.0; 20.1; 27.4; 25.6; 23.9; 24.8; 26.4; 27.0; 25.4

Hãy xác định khoảng tin cậy của sản lượng giống lúa với độ tin cậy γ = 0.95, biết sản lượng lúa là đại lượng ngẫu nhiên có phân phối chuẩn $N(\mu,\sigma^2)$ với γ và σ^2 chưa biết.

Giải.

Ta tính được

$\overline{x}$ = 25.4 và s = 2.24.

và tra bảng ta có $t_{9,\,0.95}$ = 2.262. Từ đó ta tính các cận của khoảng tin cậy

$$\mu_1 = 25.4 - 2.262\,.\frac{2.24}{\sqrt{10}} = 23.8 \quad \text{và} \quad \mu_1 = 25.4 + 2.262\,.\frac{2.24}{\sqrt{10}} = 27$$

Vậy khoảng tin cậy của sản lượng giống lúa với độ tin cậy γ = 0.95 là

[23.8; 27]

• *Bài toán 4*. Cho biến ngẫu nhiên X có phân phối chuẩn $N(\mu,\sigma^2)$ với σ và μ chưa biết. Cho mẫu $(x_1, x_2, \ldots, x_n)$ và γ, 0<γ<1. Hãy xác định khoảng tin cậy của σ^2 với độ tin cậy γ.

Giải.

Theo Định lý 3, bài V, chương 5 (Thống kê mô tả), đại lượng thống kê

$$\frac{n-1}{\sigma^2}s^2$$

có phân phối χ^2 với $n - 1$ bậc tự do.

Ký hiệu $\chi_k(u)$ là hàm phân phối của phân phối χ^2 với k bậc tự do. Ta tìm 2 số dương u_1 và u_2 sao cho

$$P\left(u_1 \le \frac{n-1}{\sigma^2}s^2 \le u_2\right) = \chi_{n-1}(u_2) - \chi_{n-1}(u_1) = \gamma$$

Trong các số u_1 và u_2 thoả điều kiện trên, người ta thường chọn sao cho

$$\chi_{n-1}(u_1) = \frac{1-\gamma}{2} \quad \Rightarrow \quad u_1 = \chi_{n-1}^{-1}\left(\frac{1-\gamma}{2}\right)$$

và

$$1 - \chi_{n-1}(u_2) = \frac{1-\gamma}{2} \quad \Rightarrow \quad u_2 = \chi_{n-1}^{-1}\left(\frac{1+\gamma}{2}\right)$$

Suy ra

$$P\left(\frac{n-1}{u_2}s^2 \le \sigma^2 \le \frac{n-1}{u_1}s^2\right) = \gamma$$

Vậy khoảng tin cậy của σ^2 với độ tin cậy γ là

$$\left[\frac{n-1}{u_2}s^2; \frac{n-1}{u_1}s^2\right]$$

+ *Ví dụ*. Xét ví dụ ở bài toán 3. Xác định khoảng tin cậy của phương sai sản lượng lúa σ^2 với độ tin cậy $\gamma = 0.9$.

Giải.

Ta có $s^2 = 2.24^2 = 5.02$. Tra bảng hàm phân phối χ_9^2 ta được

$$u_1 = \chi_{n-1}^{-1}\left(\frac{1-\gamma}{2}\right) = \chi_9^{-1}\left(\frac{1-0.9}{2}\right) = \chi_9^{-1}(0.05) = 3.33$$

và

$$u_2 = \chi_{n-1}^{-1}\left(\frac{1+\gamma}{2}\right) = \chi_9^{-1}\left(\frac{1+0.9}{2}\right) = \chi_9^{-1}(0.95) = 16.92$$

Từ đó suy ra

$$\sigma_1^2 = \frac{n-1}{u_2}s^2 = \frac{10-1}{16.92}5.02 = 2.67$$

và

$$\sigma_2^2 = \frac{n-1}{u_1}s^2 = \frac{10-1}{3.33}5.02 = 13.57$$

Vậy khoảng tin cậy của phương sai sản lượng lúa σ^2 với độ tin cậy $\gamma = 0.9$ là

[2.67; 13.57]

• *Bài toán 5*. Ước lượng tham số với mẫu cỡ lớn.

Trong các bài toán trước ta giả thiết X có phân phối chuẩn và sử dụng hàm phân phối chính xác của các đại lượng thống kê.

Tuy nhiên, nếu cỡ mẫu lớn, ta có thể sử dụng phân phối tiệm cận chuẩn để tìm khoảng tin cậy cho đơn giản.

+ *Ví dụ 1*. Giả sử sự kiện A của phép thử α có xác suất p. Thực hiện phép thử α n lần với n khá lớn. Giả sử A xuất hiện m lần. Hãy tìm khoảng tin cậy của p với độ tin cậy γ, $0 < \gamma < 1$.

Giải.

Theo định lý 2, bài VI, chương 5 (Thống kê mô tả), ta có thể coi đại lượng

$$\frac{\frac{m}{n} - p}{\sqrt{p(1-p)}}\sqrt{n}$$

có phân phối tiệm cận chuẩn N(0,1).

Để tìm khoảng tin cậy của p ta xấp xỉ

$$p(1-p) \approx \frac{m}{n}\left(1 - \frac{m}{n}\right)$$

Từ đó suy ra

$$P\left(-u_\gamma \le \frac{\frac{m}{n} - p}{\sqrt{\frac{m(n-m)}{n^2}}}\sqrt{n} \le u_\gamma\right) = \gamma$$

$\Leftrightarrow$

$$P\left(\frac{m}{n} - u_\gamma\sqrt{\frac{m(n-m)}{n^3}} \le p \le \frac{m}{n} + u_\gamma\sqrt{\frac{m(n-m)}{n^3}}\right) = \gamma$$

Vậy ta có khoảng tin cậy của p với độ tin cậy γ là

$$\left[\frac{m}{n} - u_\gamma\sqrt{\frac{m(n-m)}{n^3}}; \frac{m}{n} + u_\gamma\sqrt{\frac{m(n-m)}{n^3}}\right]$$

Ví dụ, cho

$$n = 4000; \quad m = 3200; \quad \gamma = 0.95$$

ta tính được

$$m/n = 0.8; \quad \sqrt{\frac{m(n-m)}{n^3}} = \sqrt{0.8\frac{800}{4000^2}} = \frac{\sqrt{6.4}}{400} = 0.0063$$

Tra bảng ta có

$$u_{0.95} = 1.96$$

Vậy khoảng tin cậy là

$$[0.8 - 1.96 \times 0.0063; 0.8 + 1.96 \times 0.0063] = [0.788 ; 0.812]$$

+ *Ví dụ 2*. Giả thiết như ở ví dụ 1 và cho d > 0. Hỏi phải thực hiện ít nhất bao nhiêu lần để m/n không sai khác p quá d với độ tin cậy γ, tức $P(|m/n - p| \le d) = \gamma$.

Giải.

Với n lớn ta có thể coi đại lượng

$$\frac{\frac{m}{n} - p}{\sqrt{p(1-p)}}\sqrt{n}$$

có phân phối tiệm cận chuẩn N(0,1).

Ta có

$$P\left(-u_\gamma \le \frac{\frac{m}{n} - p}{\sqrt{p(1-p)}}\sqrt{n} \le u_\gamma\right) = \gamma$$

$\Rightarrow$

$$P\left(\left|\frac{m}{n} - p\right| \le \frac{u_\gamma\sqrt{p(1-p)}}{\sqrt{n}}\right) = \gamma$$

Vậy để

$$P(|m/n - p| \le d) = \gamma$$

n cần thoả

$$\frac{u_\gamma\sqrt{p(1-p)}}{\sqrt{n}} \le d \qquad \Rightarrow \qquad n \ge \frac{u_\gamma^2}{d^2}p(1-p)$$

Vì $p(1 - p) \le \frac{1}{4}$, nên ta chỉ cần lấy n nhỏ nhất $\ge \frac{1}{4}\frac{u_\gamma^2}{d^2}$, tức

$$n = \left\lceil \frac{u_\gamma^2}{4d^2} \right\rceil$$

Chẳng hạn, cần ước lượng tỉ lệ phế phẩm p trong lô hàng với độ tin cậy γ = 0.95 và sai số không quá d = 0.01. Hỏi phải lấy cỡ mẫu ít nhất là bao nhiêu.

Ta có $u_{0.95} = 1.96$. Vậy cỡ mẫu ít nhất là

$$n = \left\lceil \frac{u_\gamma^2}{4d^2} \right\rceil = \left\lceil \frac{1.96^2}{4 \cdot 0.01^2} \right\rceil = 9604$$

+ *Ví dụ 3.* Cho biến ngẫu nhiên X có phân phối chuẩn $N(\mu,\sigma^2)$ với σ và μ chưa biết. Cho mẫu $(x_1, x_2, \ldots, x_n)$ với n khá lớn và γ, $0<\gamma<1$. Hãy xác định khoảng tin cậy của μ với độ tin cậy γ.

Giải.

Với n lớn (n>30), đại lượng thống kê

$$\frac{\overline{x}-\mu}{S}\sqrt{n} \qquad \left(s^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i-\overline{x}\right)^2 \right)$$

có phân phối tiệm cận chuẩn N(0,1).

Ta có

$$P\left(-u_\gamma \le \frac{\overline{x}-\mu}{s}\sqrt{n} \le u_\gamma\right) = \gamma$$

$\Rightarrow$

$$P\left(\overline{x}-u_\gamma\frac{s}{\sqrt{n}} \le \mu \le \overline{x}+u_\gamma\frac{s}{\sqrt{n}}\right) = \gamma$$

Vậy khoảng tin cậy của μ với độ tin cậy γ là

$$\left[\overline{x}-u_\gamma\frac{s}{\sqrt{n}};\overline{x}+u_\gamma\frac{s}{\sqrt{n}}\right]$$

Chẳng hạn, cho

n = 1376; $\overline{x}$ = 70.4; s = 0.37; γ = 0.99.

Tra bảng ta có $u_{0.99}$ = 2.58. Vậy khoảng tin cậy là

$$\left[70.4-2.58\frac{0.37}{\sqrt{1376}};70.4+2.58\frac{0.37}{\sqrt{1376}}\right] = [70.375; 70.425]$$

• *Bài toán 6*. Ước lượng trong trường hợp tập tổng thể hữu hạn và mẫu không lặp.

Cho N phần tử, trong đó có M phần tử có tính chất α. Hãy ước lượng tỉ lệ

$$p = \frac{M}{N}$$

với độ tin cậy γ.

Giải.

Chọn ngẫu nhiên n phần tử, gọi m là số phần tử có tính chất α. đại lượng ngẫu nhiên m có phân phối siêu hình học

$$P(m = k) = \frac{C(M,k).C(N-M,n-k)}{C(N,n)}, \qquad k=0, 1, \ldots, k_0 \ (k_0 = \min\{n,M\})$$

với

$$E\left(\frac{m}{n}\right) = p \qquad \text{và} \qquad D\left(\frac{m}{n}\right) = \frac{N-n}{N-1}.\frac{p(1-p)}{n}$$

Vậy $\hat{p} = \frac{m}{n}$ là ước lượng không chệch của p.

Nếu N và n đủ lớn thì

$$\frac{\frac{m}{n}-p}{\sqrt{\frac{N-n}{N-1}p(1-p)}}\sqrt{n}$$

có phân phối tiệm cận chuẩn N(0,1).

Từ đó suy ra

$$P\left(-u_\gamma \le \frac{\frac{m}{n}-p}{\sqrt{\frac{N-n}{N-1}p(1-p)}}\sqrt{n} \le u_\gamma\right) = \gamma$$

⇔

$$P\left(\frac{m}{n}-u_\gamma\frac{\sqrt{\frac{N-n}{N-1}p(1-p)}}{\sqrt{n}} \le p \le \frac{m}{n}+u_\gamma\frac{\sqrt{\frac{N-n}{N-1}p(1-p)}}{\sqrt{n}}\right) = \gamma$$

trong đó $u_\gamma = \Phi^{-1}\left(\frac{1+\gamma}{2}\right) = \Phi_L^{-1}\left(\frac{\gamma}{2}\right)$

Vì tham số p ở hai cận trên chưa biết nên có thể thay nó bằng giá trị $\frac{m}{n}$, tức là

$$P\left(\frac{m}{n}-u_\gamma\frac{\sqrt{\frac{N-n}{N-1}\frac{m}{n}\left(1-\frac{m}{n}\right)}}{\sqrt{n}} \le p \le \frac{m}{n}+u_\gamma\frac{\sqrt{\frac{N-n}{N-1}\frac{m}{n}\left(1-\frac{m}{n}\right)}}{\sqrt{n}}\right) = \gamma$$

Suy ra khoảng ước lượng [p_1; p_2] của $p = \frac{M}{N}$ với độ tin cậy γ có cận dưới

$$p_1 = \frac{m}{n}-u_\gamma\frac{\sqrt{\frac{N-n}{N-1}\frac{m}{n}\left(1-\frac{m}{n}\right)}}{\sqrt{n}}$$

và cận trên

$$p_2 = \frac{m}{n}+u_\gamma\frac{\sqrt{\frac{N-n}{N-1}\frac{m}{n}\left(1-\frac{m}{n}\right)}}{\sqrt{n}}$$

(i) Trường hợp biết N, ước lượng M:

Từ $p_1 \le \frac{M}{N} \le p_2$ với độ tin cậy γ suy ra

$$p_1.N \le M \le p_2.N \quad \text{với độ tin cậy } \gamma$$

+ *Ví dụ*. Trong lô 3000 hộp thịt lấy ra 200 hộp thì có 40 hộp kém chất lượng. Tìm khoảng ước lượng của số hộp kém chất lượng với độ tin cậy $\gamma = 90\%$.
Giải.

Ta có

$$N = 3000; \quad n = 200; \quad m = 40; \quad \gamma = 90\%;$$

Suy ra

$$m/n = 40/200 = 0.2; \quad u_\gamma = 1.65;$$

$$u_\gamma \frac{\sqrt{\frac{N-n}{N-1}\frac{m}{n}\left(1-\frac{m}{n}\right)}}{\sqrt{n}} = 1.65.\sqrt{\frac{3000-200}{3000-1}0.2(1-0.2)\frac{1}{200}} = 0.045$$

$$p_1 = 0.2 - 0.045 = 0.155; \quad p_2 = 0.2 + 0.045 = 0.245;$$

và

$$p_1.N = 0.155 \times 3000 = 465; \quad p_2.N = 0.245 \times 3000 = 735;$$

Vậy

$$465 \le M \le 735 \quad \text{với độ tin cậy } 90\%.$$

(ii) Trường hợp biết M, n không đáng kể so với N, ước lượng N:

Có thể coi $\frac{N-n}{N-1} \approx 1$. Suy ra khoảng ước lượng $[p_1; p_2]$ của $p = \frac{M}{N}$ với độ tin cậy γ có cận dưới

$$p_1 = \frac{m}{n} - u_\gamma \frac{\sqrt{\frac{m}{n}\left(1-\frac{m}{n}\right)}}{\sqrt{n}}$$

và cận trên

$$p_2 = \frac{m}{n} + u_\gamma \frac{\sqrt{\frac{m}{n}\left(1-\frac{m}{n}\right)}}{\sqrt{n}}$$

Từ $p_1 \le \frac{M}{N} \le p_2$ với độ tin cậy γ suy ra

$$\frac{M}{p_2} \le N \le \frac{M}{p_1} \quad \text{với độ tin cậy } \gamma$$

+ *Ví dụ*. Để ước lượng số cá trong hồ người ta bắt 1000 con đánh dấu rồi thả lại xuống hồ. Sau đó bắt lần thứ hai 1000 con thì thấy có 100 con đánh dấu. Hãy tìm khoảng ước lượng số cá trong hồ với độ tin cậy 0.9.
Giải.

Gọi N là số cá trong hồ

M = 1000 là số cá đánh dấu
n = 1000 là số cá bắt lần thứ hai
m = 100 là số cá đánh dấu bắt được lần thứ hai.

Ta có

$$m/n = 100/1000 = 0.1; \qquad \gamma = 90\% ; \qquad u_\gamma = 1.65$$

và

$$u_\gamma \frac{\sqrt{\frac{m}{n}\left(1-\frac{m}{n}\right)}}{\sqrt{n}} = 1.65\frac{\sqrt{0.1\times 0.9}}{\sqrt{1000}} = \frac{1.65\times 0.3}{10\sqrt{10}} = \frac{0.495}{31.6} = 0.0157$$

$\Rightarrow$

$p_1 = 0.1 - 0.0157 = 0.0843$; $\qquad$ $p_2 = 0.1 + 0.0157 = 0.1157$;

và

$$\frac{M}{p_2} = 1000/0.1157 = 8643 \text{ và } \qquad \frac{M}{p_1} = 1000/0.0843 = 11862$$

Vậy

$$8643 \le N \le 11862 \text{ với độ tin cậy } 90\%.$$

• *Bài toán 7. Ứng dụng bất đẳng thức Trebưsep.*

Cho X là biến ngẫu nhiên có phân phối bất kỳ với phương sai σ^2 đã biết, (x_1, x_2, …, x_n) là mẫu của X. Tìm khoảng ước lượng của kỳ vọng μ = E(X) với độ tin cậy γ (0 < γ < 1).

Giải.

Ta đã biết đại lượng

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

có $E(\bar{x}) = \mu$ và $D(X) = \frac{\sigma^2}{n}$. Theo bất đẳng thức Trebưsep, với ε > 0 ta có

$$P\{|\bar{x} - \mu| \le \varepsilon\} \ge 1 - \frac{\sigma^2}{n.\varepsilon^2}$$

Vậy với hệ số tin cậy γ cho trước ta chỉ cần xác định $\varepsilon_\gamma > 0$ thoả

$$1 - \frac{\sigma^2}{n.\varepsilon_\gamma^2} = \gamma \Rightarrow \qquad \varepsilon_\gamma = \frac{\sigma}{\sqrt{n.(1-\gamma)}}$$

Khi đó khoảng ước lượng của μ với độ tin cậy γ là

$$[\bar{x} - \varepsilon_\gamma ; \bar{x} + \varepsilon_\gamma]$$

+ *Ví dụ*. Cân 100 em bé 18 tháng ta được trọng lượng trung bình là 12.8 kg. Biết trọng lượng em bé 18 tháng là đại lượng ngẫu nhiên có độ lệch quân phương σ = 2.1. Hãy xác định khoảng ước lượng của kỳ vọng với độ tin cậy γ = 0.95.

Giải.

Ta có

$$\varepsilon_\gamma = \frac{\sigma}{\sqrt{n.(1-\gamma)}} = \frac{2.1}{\sqrt{100.(1-0.95)}} = 0.939$$

Vậy khoảng ước lượng là

$$[12.8 - 0.939; 12.8 + 0.939] = [11.861 ; 13.739]$$