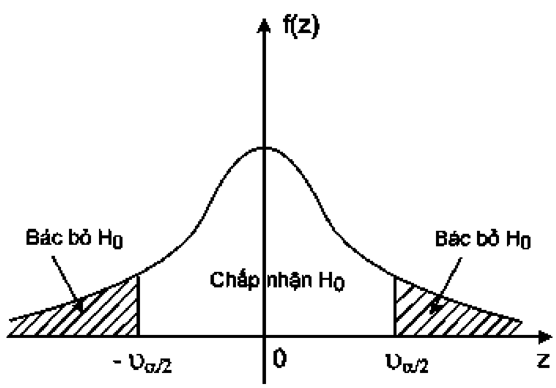
BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC NÔNG NGHIỆP I
****

Ths.LÊ ĐỨC VĨNH

GIÁO TRÌNH
XÁC SUẤT THỐNG KÊ



HÀ NỘI - 2006

Chương 1 : Phép thử . Sự kiện

Những kiến thức về giải tích tổ hợp sinh viên đã được học trong chương trình phổ thông. Tuy nhiên để giúp người học dễ dàng tiếp thu kiến thức của những chương kế tiếp chúng tôi giới thiệu lại một cách có hệ thống những kiến thức này. Phép thử ngẫu nhiên và sự kiện ngẫu nhiên là bước khởi đầu để người học làm quen với môn học Xác suất. Trong chương này chúng tôi trình bày những kiến thức tối thiểu về sự kiện ngẫu nhiên, các phép toán về các sự kiện ngẫu nhiên, hệ đầy đủ các sự kiện đồng thời chỉ ra cách phân chia một sự kiện ngẫu nhiên theo một hệ đầy đủ. Những kiến thức này là cần thiết để người học có thể tiếp thu tốt những chương tiếp theo.

I. Giải tích tổ hợp

1.Qui tắc nhân: Trong thực tế nhiều khi để hoàn thành một công việc, người ta phải thực hiện một dãy liên tiếp k hành động.

Hành động thứ nhất: có 1 trong n_1 cách thực hiện

Hành động thứ hai: có 1 trong n_2 cách thực hiện

. . ..

Hành động thứ k: có 1 trong n_k cách thực hiện

Gọi n là số cách hoàn thành công việc nói trên, ta có:

$$n = n_1 n_2 .. n_k$$

Qui tắc trên gọi là qui tắc nhân.

Ví dụ: Để đi từ thành phố A tới thành phố C phải qua thành phố B. Có một trong bốn phương tiện để đi từ A tới B là: đường bộ, đường sắt, đường không và đường thuỷ. Có một trong hai phương tiện để đi từ B tới C là đường bộ và đường thuỷ. Hỏi có bao nhiêu cách đi từ A tới C?

Để thực hiện việc đi từ A tới C ta phải thực hiện một dãy liên tiếp hai hành động.

Hành động thứ nhất: chọn phương tiện đi từ A tới C có $n_1= 4$ cách

Hành động thứ hai: chọn phương tiện đi từ B tới C có $n_2 = 2$ cách

Vậy theo qui tắc nhân, số cách đi từ A tới C là n= 4.2 = 8 cách

2.Qui tắc cộng:

Để hoàn thành công việc người ta có thể chọn một trong k phương án.

Phương án thứ nhất: có 1 trong n_1 cách thực hiện

Phương án thứ hai: có 1 trong n_2 cách thực hiện

. . . .

Phương án thứ k: có 1 trong n_k cách thực hiện

Gọi n là số cách hoàn thành công việc nói trên, ta có:

$$n = n_1 + n_2 + ... + n_k$$

Qui tắc trên gọi là qui tắc cộng

Ví dụ: Một tổ sinh viên gồm hai sinh viên Hà Nội, ba sinh viên Nam Định và ba sinh viên Thanh Hoá. Cần chọn hai sinh viên cùng tỉnh tham gia đội thanh niên xung kích. Hỏi có bao nhiêu cách chọn.

Phương án thứ nhất: Chọn hai sinh viên Hà Nội có $n_1 = 1$ cách

Phương án thứ hai: Chọn hai sinh viên Nam Định có $n_2 = 3$ cách

Phương án thứ ba: Chọn hai sinh viên Thanh Hoá có $n_3 = 3$ cách

Theo qui tắc cộng ta có số cách chọn hai sinh viên theo yêu cầu:

$$n = 1 + 3 + 3 = 7 \text{ cách}$$

3.Hoán vị

Trước khi đưa ra khái niệm một hoán vị của n phần tử ta xét ví dụ sau:.

Ví dụ: Có ba học sinh A,B,C được sắp xếp ngồi cùng một bàn học. Hỏi có bao nhiêu cách sắp xếp?

Có một trong các cách sắp xếp sau:

ABC, ACB, BAC, BCA, CAB, CBA.

Nhận thấy rằng: Đổi chỗ bất kỳ hai học sinh nào cho nhau ta được một cách sắp xếp khác. Từ một cách sắp xếp ban đầu, bằng cách đổi chỗ liên tiếp hai học sinh cho nhau ta có thể đưa về các cách sắp xếp còn lại. Mỗi một cách sắp xếp như trên còn được gọi là một hoán vị của ba phần tử A, B, C. Tổng quát với tập hợp gồm n phần tử ta có định nghĩa sau:

3.1 Định nghĩa: Một hoán vị của n phần tử là một cách sắp xếp có thứ tự n phần tử đó.

3.2 Số hoán vị của n phần tử: Với một tập gồm n phần tử đã cho. Số tất cả các hoán vị của n phần tử ký hiệu là P_n.Ta cần xây dựng công thức tính P_n.

Để tạo ra một hoán vị của n phần tử ta phải thực hiện một dãy liên tiếp n hành động.

Hành động thứ nhất: Chọn 1 phần tử xếp đầu có n cách chọn

Hành động thứ hai: Chọn 1 phần tử xếp thứ 2 có n-1 cách chọn

. . . .

Hành động cuối: Chọn phần tử còn lại xếp cuối có 1 cách chọn

Theo qui tắc nhân, số cách tạo ra 1 hoán vị của n phần tử là

$$P_n = n.(n-1) \ldots.2.1 = n!$$

4. Chỉnh hợp không lặp

4.1 Định nghĩa: Một chỉnh hợp không lặp chập k của n phần tử là một cách sắp xếp có thứ tự gồm k phần tử khác nhau lấy từ n phần tử đã cho.

Ví dụ: Có 5 chữ số 1, 2, 3, 4, 5. Hãy lập tất cả các số gồm 2 chữ số khác nhau

Các số đó là: 12, 13, 14, 15, 21, 23, 24, 25, 31, 32, 34, 35, 41, 42, 43, 45, 51, 52, 53, 54.

Mỗi một số trên chính là một cách sắp xếp có thứ tự gồm hai phần tử khác nhau lấy từ năm phần tử là năm chữ số đã cho. Vậy mỗi số là chỉnh hợp không lặp chập hai của năm phần tử.

4.2 Số các chỉnh hợp không lặp: Số các chỉnh hợp không lặp chập k của n phần tử kí hiệu là A_n^k. Ta xây dựng công thức tính A_n^k.

Để tạo ra một chỉnh hợp không lặp chập k của n phần tử ta phải thực hiện một dãy liên tiếp k hành động.

Hành động thứ nhất: chọn 1 trong n phần tử để xếp đầu: có n cách

Hành động thứ hai: chọn 1 trong n-1 phần tử để xếp thứ 2: có n -1 cách

. . . .

Hành động thứ k: chọn 1 trong n-k+1 phần tử để xếp cuối: có n-k+1 cách

Theo qui tắc nhân: Số cách tạo ra một chỉnh hợp không lặp chập k của n phần tử là :

$$A_n^k = n(n-1)..\ ....(n-k+1)$$

Để dễ nhớ ta sử dụng công thức sau:

$$A_n^k = n.(n-1)....(n-k+1) = n.(n-1)...(n-k+1).\frac{(n-k).......2.1}{(n-k)......2.1} = \frac{n!}{(n-k)!}$$

5. Chỉnh hợp lặp: Để hiểu thế nào là một chỉnh hợp lặp ta xét ví dụ sau:

Ví dụ: Hãy lập các số gồm 2 chữ số từ 4 chữ số: 1, 2, 3, 4.

Các số đó là: 11, 12, 13, 14, 21, 22, 23, 24, 31, 32, 33, 34, 41, 42, 43, 44.

Mỗi số trong các số nói trên là một cách sắp xếp có thứ tự gồm hai chữ số, mỗi chữ số có thể có mặt đến hai lần lấy từ bốn chữ số đã cho. Mỗi cách sắp xếp như vậy còn gọi là một chỉnh hợp lặp chập hai của bốn phần tử. Tổng quát hoá ta có định nghĩa sau:

5.1 Định nghĩa: Một chỉnh hợp lặp chập k của n phần tử là một cách sắp xếp có thứ tự gồm k phần tử mà mỗi phần tử lấy từ n phần tử đã cho có thể có mặt nhiều lần.

5.2 Số các chỉnh hợp lặp chập k:

Số các chỉnh hợp lặp chập k của n phần tử được ký hiệu là $\hat{A}_n^k$. Ta sẽ đưa ra công thức tính $\hat{A}_n^k$.

Để tạo ra một chỉnh hợp lặp chập k của n phần tử ta phải thực hiện một dãy liên tiếp k hành động.

Hành động thứ nhất: chọn 1 trong n phần tử xếp đầu có n cách

Hành động thứ hai: chọn 1 trong n phần tử xếp thứ 2 có n cách

. . . .

Hành động thứ k: chọn 1 trong n phần tử xếp thứ k có n cách

Theo qui tắc nhân ta có: $\hat{A}_n^k = n^k$

6.Tổ hợp: Các khái niệm trên luôn để ý đến trật tự của tập hợp ta đang quan sát. Tuy nhiên trong thực tế có nhiều khi ta chỉ cần quan tâm tới các phần tử của tập con của một tập hợp mà không cần để ý đến cách sắp xếp tập con đó theo một trật tự nào. Từ đây ta có khái niệm về tổ hợp như sau

6.1 Định nghĩa: Một tổ hợp chập k của n phần tử là một tập con gồm k phần tử lấy từ n phần tử đã cho.

Ví dụ: Cho tập hợp gồm bốn phần tử {a,b,c,d}. Hỏi có bao nhiêu tập con gồm hai phần tử?

Các tập con đó là {a,b},{a,c},{a,d},{b,c},{b,d},{c,d}

Vậy tập hợp gồm bốn phần tử {a,b,c,d} có sáu tập con vừa nêu.

6.2: Số tổ hợp chập k của n phần tử có ký hiệu là C_n^k

Bằng cách đổi chỗ các phần tử cho nhau, một tổ hợp chập k của n phần tử có thể tạo ra k! chỉnh hợp không lặp chập k của n phần tử.

Có C_n^k tổ hợp chập k của n phần tử tạo ra A_n^k chỉnh hợp không lặp chập k của n phần tử.

Vậy ta có : $C_n^k = \frac{A_n^k}{k!} = \frac{n!}{k!(n-k)!}$

7.Tổ hợp lặp:

7.1 Định nghĩa: Một tổ hợp lặp chập k của n phần tử là một nhóm không phân biệt thứ tự gồm k phần tử, mỗi phần tử có thể có mặt đến k lần lấy từ n phần tử đã cho.

Ví dụ: Cho tập {a,b,c} gồm 3 phần tử

Các tổ hợp lặp của tập hợp trên là {a,a},{a,b},{a,c},{b,b},{b,c},{c,c}

7.2 Số các tổ hợp lặp chập k của n phần tử ký hiệu là:. $\hat{C}_n^k$

Việc tạo ra một tổ hợp lặp chập k của n phần tử tương đương với việc xếp k quả cầu giống nhau vào n ngăn kéo đặt liền nhau, hai ngăn liên tiếp cùng chung một vách ngăn. Các vách ngăn trừ vách ngăn đầu và cuối có thể xê dịch và đổi chỗ cho nhau. Mỗi cách sắp xếp k quả cầu giống nhau vào n ngăn là một cách bố trí n+k-1 phần tử (gồm k quả cầu và n-1 vách ngăn) theo thứ tự từ phải sang trái. Cách bố trí không đổi khi các quả cầu đổi chỗ cho nhau hoặc các vách ngăn đổi chỗ cho nhau. Cách bố trí thay đổi khi các quả cầu và các vách ngăn đổi chỗ cho nhau. Ta có (n+k-1)! cách bố trí n+k-1 phần tử (gồm k quả cầu và n-1 vách ngăn). Số cách đổi chỗ k quả cầu là k! , số cách đổi chỗ n-1 vách ngăn là (n-1)! . Vậy ta có số các tổ hợp lặp chập k của n phần tử là:

$$\hat{C}_n^k = \frac{(n+k-1)!}{k!(n-1)!} = C_{n+k-1}^k$$

Ví dụ: Tại một trại giống gà có ba loại gà giống A, B, C. Một khách hàng vào định mua 10 con. Hỏi có bao nhiêu cách mua (giả sử rằng số lượng các giống gà A, B, C mỗi loại của trại đều lớn hơn 10).

Ta thấy mỗi một cách mua 10 con gà chính là một tổ hợp lặp chập 10 của 3 phần tử. Vậy số cách mua là: $\hat{C}_3^{10} = C_{12}^{10} = 66$

8. Nhị thức Newton

Ta có: $(a+b)^2 = a^2 + 2ab + b^2 = C_2^0 a^2 b^0 + C_2^1 a^1 b^1 + C_2^1 a^0 b^2$

$(a+b)^3 = a^3 + 3a^2 b + 3ab^2 + b^3 = C_3^0 a^3 b^0 + C_3^1 a^2 b^1 + C_3^2 a^1 b^2 + C_3^3 a^0 b^3$

Mở rộng ra:

$$(a+b)^n = C_n^0 a^n b^0 + C_n^1 a^{n-1} b^1 + \ldots\ldots + C_n^k a^{n-k} b^k + \ldots\ldots\ldots\ldots + C_n^n a^0 b^n$$

Công thức trên gọi là công thức nhị thức Newton.

Ta chứng minh công thức nhị thức Newton theo qui nạp..

Với n = 2 ta có công thức đúng.

Giả sử công thức đúng với n = m tức là:

$$(a+b)^m = C_m^0 a^m b^0 + C_m^1 a^{m-1} b^1 + \ldots\ldots + C_m^m a^0 b^m$$

Ta sẽ chứng minh:

$$(a+b)^{m+1} = C_{m+1}^0 a^{m+1} b^0 + C_{m+1}^1 a^m b^1 + \ldots\ldots\ldots + C_{m+1}^{m+1} a^0 b^{m+1}$$

Thật vậy:

$$(a+b)^{m+1} = (a+b)^m (a+b) = (C_m^0 a^n b^0 + \ldots + C_m^k a^{m-k} b^k + \ldots + C_m^m a^0 b^m)(a+b) \quad =>$$

$$(a+b)^{m+1} = (C_m^0 + C_m^1) a^{m+1} b^0 + \ldots + (C_m^{k-1} + C_m^k) a^{m+1-k} b^k + \ldots + (C_m^{m-1} + C_m^m) a^0 b^{m+1}$$

Mặt khác: $C_m^{k-1} + C_m^k = C_{m+1}^k$ suy ra:

$$(a+b)^{m+1} = C_{m+1}^0 a^{m+1} b^0 + C_{m+1}^1 a^m b^1 + \ldots\ldots\ldots + C_{m+1}^{m+1} a^0 b^{m+1}.$$

Theo nguyên lý qui nạp công thức nhị thức Newton được chứng minh.

Ví dụ: Tìm hệ số của x^{12} trong khai triển: $(x + \frac{1}{x^2})^{20}$

Ta có: $(x + \frac{1}{x})^{20} = C_{20}^0 x^{20} + \ldots\ldots\ldots + C_{20}^k x^{20-2k} + \ldots\ldots + C_{20}^{20} \frac{1}{x^{20}}$.

Xét 20 - 2 k = 12

=> k = 4 Vậy hệ số của x^{12} là: $C_{20}^4 = 4745$

II. Phép thử, sự kiện

1.Phép thử ngẫu nhiên và không ngẫu nhiên

Một phép thử có thể coi là một thí nghiệm, một quan sát các hiện tượng tự nhiên, các hiện tượng xã hội và các vấn đề kĩ thuật với cùng một hệ điều kiện nào đó.

Trong các loại phép thử có những phép thử mà khi bắt đầu tiến hành thực hiện ta đã biết được kết quả sẽ xảy ra sau khi thử như đun nước ở điều kiện bình thường (dưới áp suất 1 atmotphe) thì đến 100^oC nước sẽ sôi, hoặc cho dung dịch NaOH không dư vào dung dịch HCl cũng không dư ta thu được muối ăn NaCl và nước H_2O.

Những phép thử mà khi bắt đầu tiến hành thử ta biết được những kết quả nào sẽ xảy ra sau khi thử được gọi là các phép thử không ngẫu nhiên.

Tuy nhiên có rất nhiều loại phép thử mà ngay khi bắt đầu tiến hành phép thử ta không thể biết được những kết quả nào sẽ xảy ra sau khi thử chẳng hạn như khi gieo 100 hạt đậu giống, số hạt nảy mầm sau một thời gian gieo có thể là từ 0 đến 100 hoặc khi cho ấp 10 quả trứng thì số trứng gà có thể nở ra gà con là từ 0 đến 10 con. Những phép thử loại này gọi là những phép thử ngẫu nhiên.

Trong giáo trình này chúng ta chỉ quan tâm tới những phép thử ngẫu nhiên, đó là những phép thử mà khi bắt đầu tiến hành thử ta chưa thể biết những kết quả nào sẽ xảy ra. Để đơn giản từ đây trở đi khi nói tới phép thử ta phải hiểu đấy là phép thử ngẫu nhiên

2. Sự kiện:

Các kết quả có thể có của một phép thử ứng với một bộ các điều kiện xác định nào đó gọi là các sự kiện ngẫu nhiên hoặc đơn giản gọi là các sự kiện hoặc các biến cố.

Ta thường lấy các chữ cái A, B, C, D. . .. . . hoặc A_i, B_j, C_k, D_n.. . . để chỉ các sự kiện.

Ví dụ 1: Tung một con xúc xắc cân đối và đồng chất có thể có các sự kiện sau:

A: Sự kiện xuất hiện mặt chẵn

B: Sự kiện xuất hiện mặt lẻ

A_i: Sự kiện xuất hiện mặt có i chấm.

Ví dụ 2: Trong một giỏ đựng hoa quả có chứa 1 quả cam, 1 quả quýt, 1 quả đào và 1 quả lê. Chọn ngẫu nhiên ra 2 quả có thể có các sự kiện sau:

A: Hai quả được chọn gồm 1 cam 1 quýt

B: Hai quả được chọn gồm 1 cam 1 đào

C: Hai quả được chọn gồm 1 cam 1 lê

D: Hai quả được chọn gồm 1 quýt 1 lê

E: Hai quả được chọn gồm 1 quýt 1 đào

G: Hai quả được chọn gồm 1 đào 1 lê

3. Sự kiện tất yếu và sự kiện không thể có

Sự kiện tất yếu hoặc sự kiện chắc chắn là sự kiện nhất thiết phải xảy ra sau khi phép thử được thực hiện. Ta kí hiệu sự kiện này là Ω..

Sự kiện không thể có hoặc sự kiện bất khả hoặc sự kiện rỗng là sự kiện không bao giờ xảy ra sau khi thử. Ta kí hiệu sự kiện này là ϕ.

Ví dụ: Đứng tại Hà Nội ném một hòn đá

Sự kiện đá rơi xuống địa giới Việt Nam là sự kiện tất yếu

Sự kiện đá rơi xuống Đại Tây Dương là sự kiện bất khả.

4. Quan hệ giữa các sự kiện, hai sự kiện bằng nhau

Sự kiện A được gọi là kéo theo sự kiện B nếu A xảy ra thì B cũng xảy ra và kí hiệu

$A \subset B$ (hoặc $A \Rightarrow B$).

Nếu A kéo theo B và B kéo theo A thì ta nói A bằng B và viết A = B. Trong xác suất hai sự kiện bằng nhau được coi là một

Ví dụ: Một học sinh thi hết một môn học

A là sự kiện học sinh đó đỗ (đạt điểm từ 5 tới 10)

B là sự kiện học sinh đó đỗ trung bình hoặc khá (đạt điểm từ 5 tới 8)

C là sự kiện học sinh đó đỗ khá hoặc giỏi

G là sự kiện học sinh đó đỗ giỏi (đạt điểm 9, 10)

K là sự kiện học sinh dố đỗ khá (đạt điểm 7, 8)

TB là sự kiện học sinh đó đỗ trung bình (đạt điểm 5, 6)

A_i là sự kiện học sinh đó đạt i điểm (i = 0, 1, ...,9, 10).

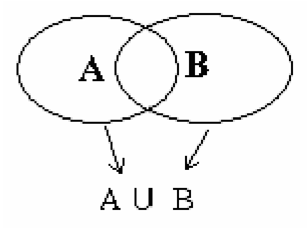
Ta có: $G \Rightarrow A; B \Rightarrow A; C \Rightarrow A; A_6 \Rightarrow A; A_9 \Rightarrow G; A_7 \Rightarrow B; A_7 \Rightarrow K; A_5 \Rightarrow TB$...

5.Các phép tính về sự kiện

5.1 Phép hợp: Hợp của 2 sự kiện A và B là sự kiện C, sự kiện C xảy ra khi A xảy ra hoặc B xảy ra.

Kí hiệu: $A \cup B = C$ và đọc là A hợp B bằng C

Ta có thể mô tả hợp của 2 sự kiện A và B bằng hình vẽ sau:



Hình 1

Dựa vào hình vẽ trên có thể thấy C xảy ra khi:

- A xảy ra và B không xảy ra.
- B xảy ra và A không xảy ra.
- Cả A và B cùng xảy ra.

Vì vậy có thểnói hợp của hai sự kiện A và B là một sự kiện C xảy ra khi ít nhất 1 trong 2 sự kiện A, B xảy ra.

Ví dụ: Một sinh viên thi hết một môn học

Gọi : A là sự kiện sinh viên đó không phải thi lại (điểm thi từ 5 đến 10)

B là sự kiện sinh viên đó đạt điểm trung bình khá (điểm thi từ 5 đến 8)

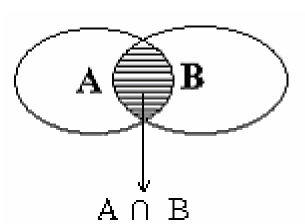
C là sự kiện sinh viên đó đạt điểm khá giỏi (điểm thi từ 7 đến 10)

Ta có: A = $B \cup C$.

5.2 Phép giao: Giao của 2 sự kiện A và B là sự kiện D, sự kiện D xảy ra khi cả A và B cùng xảy ra.

Kí hiệu: $A \cap B = D$ hoặc AB = D và đọc là A giao B bằng D hoặc A nhân B bằng D

Hình vẽ sau mô tả giao của 2 sự kiện A và B



Hình 2

Ví dụ: Quay lại ví dụ ở mục 5.1

Gọi K là sự kiện sinh viên đó đạt điểm khá (điểm thi từ 7 đến 8)

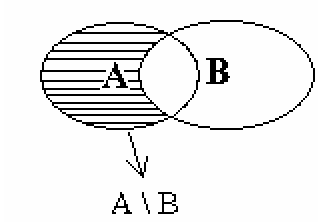
Ta có: K = $B \cap C$

Nếu $A \cap B = \phi$ ta nói A và B là 2 sự kiện xung khắc với nhau. Khi A xung khắc với B thì hợp của 2 sự kiện A và B được kí hiệu là A + B và đọc là A cộng B.

5.3 Phép trừ. Sự kiện đối lập: Hiệu của sự kiện A trừ sự kiện B là sự kiện E, sự kiện E xảy ra khi A xảy ra và B không xảy ra.

Kí hiệu: A\B= E và đọc là A trừ B bằng E

Ta cũng có thể mô tả hiệu của sự kiện A trừ sự kiện B bằng hình vẽ sau:



Hình 3

Dễ nhận thấy rằng: Nếu $A \cap B = \phi$ thì A \ B = A

Sự kiện : $\Omega \setminus A$ Gọi là sự kiện đối lập của sự kiện A và kí hiệu là $\overline{A}$.

Từ định nghĩa sự kiện đối lập của sự kiện A ta thấy:

* A và $\overline{A}$. xung khắc với nhau

* Nếu A không xảy ra thì $\overline{A}$ xảy ra và ngược lại

Hai sự kiện đối lập nhau xung khắc với nhau "mạnh mẽ" theo kiểu có anh thì không có tôi nhưng không có anh thì phải có tôi.

Ví dụ: Một tổ học sinh gồm 3 học sinh nam 3 học sinh nữ. Chọn ngẫu nhiên 2 người.

Gọi : A là sự kiện 2 học sinh được chọn là cùng giới

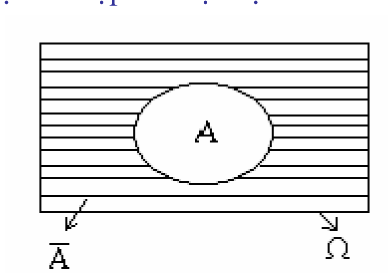
B là sự kiện 2 học sinh được chọn đều là nam

C là sự kiện 2 học sinh được chọn đều là nữ

D là sự kiện 2 học sinh được chọn có một nam một nữ

Ta có A \ B = C, D = $\overline{A}$.

Hình sau mô tả sự kiện đối lập của sự kiện A



Hình 4

5.4 Tính chất

1/ $\phi \Rightarrow A\ ;\ A \Rightarrow \Omega\ \ \forall A$

2/ $A \cup \phi = A\ ;\ A\phi = \phi\ ;\ A \cup \Omega = \Omega\ ;\ A\Omega = A$

3/ Nếu $A \Rightarrow B\ ;\ B \Rightarrow C$ thì $A \Rightarrow C$

4/ $A \cup B = B \cup A\ ;\ AB = BA$

5/ $A\cup(B\cup C)=(A\cup B)\cup C$; $A(BC)=(AB)C$

6/ $A(B\cup C)=AB\cup AC$; $A\cup(BC)=(A\cup B)(A\cup C)$

7/ $A\setminus B=A\overline{B}$

8/ $\overline{A\cup B}=\overline{A}\,\overline{B}$; $\overline{AB}=\overline{A}\cup\overline{B}$

Việc chứng minh các tính chất trên khá dễ dàng xin dành cho bạn đọc. Chúng tôi chỉ chứng minh tính chất 8 phần 1 như là một ví dụ minh hoạ cho việc chứng minh các sự kiện bằng nhau:

Ta chứng minh: $\overline{A\cup B}=\overline{A}\,\overline{B}$

Giả sử $\overline{A\cup B}$ xảy ra theo định nghĩa của sự kiện đối lập => $A\cup B$ không xảy ra, theo định nghĩa của hợp hai sự kiện => A không xảy ra và B không xảy ra, lại theo định nghĩa của sự kiện đối lập => $\overline{A}$ xảy ra và $\overline{B}$ xảy ra, theo định nghĩa của phép giao hai sự kiện

=> $\overline{A}\,\overline{B}$. xảy ra.

Vậy ta có: $\overline{A\cup B}\Rightarrow\overline{A}\,\overline{B}$ (1)

Ngược lại giả sử $\overline{A}\,\overline{B}$ xảy ra, theo định nghĩa của phép giao, => $\overline{A}$ xảy ra và $\overline{B}$ xảy ra, lại theo định nghĩa của sự kiện đối lập => A không xảy ra và B không xảy ra, theo định nghĩa của hợp hai sự kiện => $A\cup B$. không xảy ra, theo định nghĩa của sự kiện đối lập

=> $\overline{A\cup B}$ xảy ra. Vậy ta cũng có: $\overline{A}\,\overline{B}\Rightarrow\overline{A\cup B}$ (2)

Từ (1) và (2) => $\overline{A\cup B}=\overline{A}\,\overline{B}$

6. Sự kiện có thể phân chia được, sự kiện sơ cấp cơ bản

6.1 Sự kiện có thể phân chia được

Sự kiện A được gọi là có thể phân chia được nếu tồn tại hai sự kiện $B\neq\phi$, $C\neq\phi$, $BC=\phi$ và $A=B+C$. Khi đó ta nói A phân chia được thành hai sự kiện B và C.

Ví dụ: Trong một con xúc xắc cân đối và đồng chất.

Gọi A là sự kiện xuất hiện mặt có số chấm chia hết cho 3.

Gọi A_i là sự kiện xuất hiện mặt i chấm

Sự kiện A có thể phân chia được vì tồn tại A_3; $A_6\neq\phi$; $A_3A_6=\phi$ và $A=A_3+A_6$.

6.2 Sự kiện sơ cấp cơ bản: Sự kiện khác rỗng và không thể phân chia được gọi là sự kiện sơ cấp cơ bản.

Ví dụ: Quay lại ví dụ ở mục 6.1. Các sự kiện A_1, A_2, A_3, A_4, A_5, A_6 là các sự kiện sơ cấp cơ bản.

Ta nhận thấy rằng các sự kiện sơ cấp cơ bản là các sự kiện mà sau một phép thử chỉ có một trong các sự kiện này xảy ra.

7. Hệ đầy đủ các sự kiện

7.1 Hệ đầy đủ các sự kiện: Hệ các sự kiện A_1, A_2,. ... A_n gọi là một hệ đầy đủ các sự kiện nếu:

1/ $A_i \neq \phi$ với mọi i = 1, 2 . . . . n

2/ $A_i A_j = \phi$ với mọi i khác j

3/ $A_1 + A_2 + \ldots\ldots + A_n = \Omega$

Ví dụ: Đem hai cá thể ở thế hệ F_1 mang gen Aa, Aa lai với nhau. Các cá thể con ở thế hệ F_2 có thể có 1 trong 4 kiểu gien AA, Aa, aA và aa. Chọn 1 cá thể con trong các cá thể nói trên.

Gọi: A là sự kiện cá thể con là đồng hợp tử (mang gen AA hoặc aa)

B là sự kiện cá thể con là dị hợp tử (mang gen Aa hoặc aA)

C là sự kiện cá thể con có mang gen trội (AA, Aa, aA)

A_1 là sự kiện cá thể con chỉ mang gen trội (AA)

A_2 là sự kiện cá thể con chỉ mang gen lặn (aa)

Ta có: A, B là một hệ đầy đủ các sự kiện

C, A_2 cũng là một hệ đầy đủ các sự kiện

B, A_1, A_2 cũng là một hệ đầy đủ các sự kiện

Như vậy: với một phép thử đã cho có thể có nhiều hệ đầy đủ các sự kiện khác nhau.

7.2 Phân chia một sự kiện theo hệ đầy đủ.

Giả sử A_1, A_2, ...A_n là một hệ đầy đủ các sự kiện. A là một sự kiện khác rỗng nào đó. Ta có:

$$A = A\Omega = A(A_1 + A_2 + \ldots\ldots\ldots + A_n) = AA_1 + \ldots.AA_i + \ldots..AA_n$$

Khi đó ta nói A được phân chia gián tiếp nhờ hệ đầy đủ các sự kiện: A_1, A_2 , A_3 ,..., A_n.

Như đã biết với mỗi phép thử có thể lập ra nhiều hệ đầy đủ các sự kiện vì vậy mỗi sự kiện khác rỗng A cũng có thể phân chia theo nhiều cách khác nhau. Mục đích của việc phân chia sự kiện A ra một số sự kiện đơn giản hơn nhằm đánh giá khả năng xảy ra của sự kiện A nhờ các sự kiện đơn giản này.

8. Đại số và σ - đại số các sự kiện

Xét Ω là một tập hợp khác rỗng mà ta gọi là sự kiện chắc chắn. C là một họ các tập con nào đó của Ω.Mỗi tập con A của Ω, $A \in C$ gọi là một sự kiện. Họ C được gọi là $\sigma -$đại số các sự kiện nếu:

1/ $\phi \in C$.

2/ Nếu $A \in C$ thì $\overline{A} \in C$

3/ Nếu A_1, A_2. . . . . . A_n. . .là các sự kiện thuộc C thì $\bigcup_{n=1}^{\infty} A_n \in C$

Họ C được gọi là đại số các sự kiện nếu yêu cầu 1, 2 nêu trên thoả mãn và hợp của một số hữu hạn các sự kiện thuộc C cũng là một sự kiện thuộc C. Ta nhận thấy rằng nếu C là $\sigma -$ đại số các sự kiện thì C cũng là một đại số các sự kiện.

Ví dụ: Tung đồng thời 2 đồng tiền, các sự kiện sơ cấp cơ bản là:

SS, SN, NS, NN. Xét Ω = SS + SN +NS +NN.

Tập tất cả các tập hợp con của Ω là một đại số các sự kiện.và cũng là một $\sigma -$ đại số các sự kiện

Bài tập chương 1

1. Một đoạn gen gồm 2 gen X, 2 gen Y, 2 gen Z, 2 gen T liên kết với nhau theo một hàng dọc.

a. Hỏi có bao nhiêu cách liên kết 8 gen nói trên?

b. Hỏi có bao nhiêu cách liên kết để 2 gen X đứng liền nhau?

c. Hỏi có bao nhiêu cách liên kết để có 3 gen XYZ đứng liền nhau theo thứ tự trên.

2. Có 10 người xếp theo một hàng dọc

a. Có bao nhiêu cách sắp xếp để 2 người A và B đứng liền nhau?

b. Có bao nhiêu cách sắp xếp để 2 người A và B đứng cách nhau đúng 3 người?

3. Có thể lập được bao nhiêu số gồm 10 chữ số khác nhau sao cho:

a.Không có 2 chữ số chẵn nào đứng liền nhau

b. Không có 2 chữ số lẻ nào đứng liền nhau

c. Các chữ số chẵn đứng liền nhau

d. Các chữ số lẻ đứng liền nhau

4. Cho 6 chữ số: 1, 2, 3, 4, 5, 6

a. Có thể lập được bao nhiêu số gồm 8 chữ số sao cho chữ số 1 và chữ số 2 mỗi chữ số có mặt đúng 2 lần, các chữ số còn lại có mặt đúng 1 lần.

b. Có thể lập được bao nhiêu số chẵn gồm 8 chữ số trong đó chữ số 2 có mặt đúng 3 lần, các chữ số còn lại có mặt đúng một lần.

c. Có thể lập được bao nhiêu số lẻ gồm 8 chữ số trong đó chữ số 1 có mặt đúng 3 lần, các chữ số còn lại có mặt đúng 1 lần.

5*. Trong một kì thi tin học quốc tế tại một khu vực gồm 6 phòng thi đánh số từ 1 đến 6 dành cho ba đoàn Việt nam , Mĩ và Nga mỗi đoàn gồm 4 thí sinh. Mỗi phòng thi có 2 máy tính (không đánh số) dành cho 2 thí sinh. Việc xếp 2 thí sinh vào mỗi phòng thi theo nguyên tắc hai thí sinh cùng một quốc tịch không được xếp cùng một phòng. Hỏi có bao nhiêu cách sắp xếp các thí sinh của ba đoàn vào 6 phòng?

6*. Dọc theo hai bên đường vào một trường trong học người ta dự định trồng mỗi bên 3 cây bàng, 3 cây phượng và 3 cây bằng lăng.

a. Hỏi có bao nhiêu cách trồng để các cây cùng loại trồng đối diện nhau?

b. Hỏi có bao nhiêu cách trồng để không có hai cây cùng loại nào trồng đối diện nhau?

7*. Vòng chung kết giải vô địch bóng đá châu Âu gồm 16 đội trong đó có đội chủ nhà và đội vô địch bốn năm trước.

a. Có bao nhiêu cách chia 16 đội vào bốn bảng A, B, C, D.

b, Có bao nhiêu cách chia 16 đội vào bốn bảng A, B, C, D sao cho đội chủ nhà và đội vô địch bốn năm trước không cùng bảng.

c. Giải bài toán trên trong trường hợp không để ý tới vai trò của các bảng.

8. Một đàn gà gồm 4 con gà mái và 6 con gà trống. Trong 4 con gà mái có 2 con màu vàng, 2 con màu đen. Trong 6 con gà trống có 3 con màu vàng và 3 con màu đen. Chọn ngẫu nhiên 2 con gà

a. Có bao nhiêu cách chọn để được 1 con trống 1 con mái

b. Có bao nhiêu cách chọn để được 2 con màu vàng

c. Có bao nhiêu cách chọn để được1 con trống 1 con mái cùng màu

9. Một tổ sinh viên gồm 6 nam 4 nữ. Trong 6 nam có 2 sinh viên Hà Nội và 4 sinh viên tỉnh Hà Tây. Trong 4 nữ có 2 nữ sinh Hà Nội và 2 nữ sinh Thái Bình. Chọn ngẫu nhiên ra 3 người

a. Có bao nhiêu cách chọn ra 3 sinh viên nam?

b. Có bao nhiêu cách chọn ra 2 sinh viên nam 1 sinh viên nữ?

c. Có bao nhiêu cách chọn ra 3 sinh viên gồm đủ 3 tỉnh?

10. Cho đa giác đều gồm 2n cạnh

a. Hỏi có thể lập được bao nhiêu hình chữ nhật có 4 đỉnh là 4 đỉnh của đa giác đều này?

b. Hỏi đa giác đều nói trên có bao nhiêu đường chéo?

11. Cho tập A = $\{0,1,2,3,4,5,6,7,8,9,10\}$

a. A có bao nhiêu tập con có ít nhất 2 chữ số nhỏ hơn 6

b. A có bao nhiêu tập con có ít nhất 2 chữ số lớn hơn 6

12. Có 4 viên bi giống nhau được bỏ vào 3 cái hộp. Hỏi có bao nhiêu cách bỏ?

13*. Có 4 hành khách đợi tàu tại nhà ga A để đi tới B. Một đoàn tàu gồm 4 toa chuẩn bị rời ga A để đi tới B.

a. Có bao nhiêu cách lên tàu của 4 hành khách trên.

b. Có bao nhiêu cách lên tàu của 4 hành khách trên sao cho mỗi người lên một toa.

c. Có bao nhiêu cách để 4 hành khách trên lên hai toa mỗi toa 2 người.

14. Trong khai triển $(x-\frac{2}{x^2})^{50}$.

a. Tìm số hạng không chứa x

b. Tìm hệ số của x^{20}

c. Tìm hệ số của x^{-40}

15. Chứng minh các đồng nhất thức:

a. $C_n^0 + C_n^1 + \dots\dots\dots\dots + C_n^k + \dots\dots + C_n^n = 2^n$

b. $C_n^1 + 2C_n^2 + \dots\dots\dots + kC_n^k + \dots\dots + nC_n^n = n2^{n-1}$

c. $C_n^0 + \frac{1}{2}C_n^1 + ... + \frac{1}{k+1}C_n^k + ... + \frac{1}{n+1}C_n^n = \frac{2^{n+1}-1}{n+1}$

d. $C_{2n}^0 + C_{2n}^2 + ... + C_{2n}^{2k} + ... + C_{2n}^{2n} = C_{2n}^1 + C_{2n}^3 + ... + C_{2n}^{2k+1} + ... + C_{2n}^{2n-1}$

16. Cho p, q > 0, p + q = 1. Tìm số hạng lớn nhất trong dãy số sau:

$C_n^0p^0q^n; C_n^1pq^{n-1}; ....; C_n^kp^{n-k}q^k; ....; C_n^np^nq^0$

17. Xếp 3 người theo một hàng dọc. Nêu các sự kiện sơ cấp cơ bản

18. Từ 4 người A, B, C, D lấy ngẫu nhiên 2 người. Nêu tập các sự kiện sơ cấp cơ bản.

19. Hai cá thể sinh vật có cùng kiểu gen Aa Bb đem lai với nhau. Hãy nêu các kiểu gen có thể có của các cá thể con.

20. Từ hai nhóm học sinh, nhóm thứ nhất gồm 4 học sinh nam A, B, C, D nhóm thứ hai gồm 4 học sinh nữ X, Y, Z, T. Chọn mỗi nhóm ra 2 học sinh.

a. Chỉ ra tập các sự kiện sơ cấp cơ bản ứng với phép thử trên

b. Chỉ ra hai hệ đầy đủ các sự kiện.

21. Tung một lần 3 đồng tiền.

a.Hãy chỉ ra các sự kiện sơ cấp cơ bản.

b.Hãy chỉ ra một hệ đầy đủ các sự kiện chỉ gồm hai sự kiện

22. Tung đồng thời hai con xúc xắc.

a. Có bao nhiêu sự kiện sơ cấp cơ bản

b. Hãy chỉ ra một hệ đầy đủ các sự kiện gồm 11 sự kiện

23. Một đa giác đều gồm 2n cạnh (n > 2). Chọn ngẫu nhiên bốn đỉnh.

a. Có bao nhiêu sự kiện sơ cấp cơ bản?

b. Có bao nhiêu sự kiện để bốn đỉnh được chọn lập thành hình chữ nhật?

Khi n = 3 Chọn ngẫu nhiên 3 đỉnh của một lục giác đều.

c. Có bao nhiêu sự kiện sơ cấp cơ bản?

d. Có bao nhiêu sự kiện ba đỉnh được chọn lập thành tam giác đều?

25. Chứng minh các tính chất về các phép toán của các sự kiện.

Chương 2 : Xác suất

Việc đưa ra những số đo thích hợp đánh giá khả năng khách quan xảy ra của mỗi sự kiện được trình bày trong phần đầu của chương này. Các dạng định nghĩa xác suất từ các định nghĩa cổ điển tới định nghĩa xác suất theo hệ tiên đề giúp người học hình dung được sự phát triển và tính phong phú, đa dạng của môn xác suất. Các tính chất các định lý về xác suất được trình bày ở mức tối thiểu để người học khỏi cảm thấy nặng nề khi tiếp thu chúng. Những ví dụ đưa ra giúp người học thấy được những áp dụng thực thực tế của môn xác suất và qua các ví dụ này người học có thể hiểu cách làm các bài toán xác suất.

I. Các định nghĩa của xác suất

1. Mở đầu: Khi tiến hành một phép thử, có thể có một trong nhiều sự kiện sẽ xảy ra, mỗi sự kiện là một đặc tính định tính, việc chỉ ra "số đo" khả năng xảy ra của mỗi một sự kiện là điều cần thiết. Ta có thể hiểu xác suất của mỗi sự kiện là "số đo" khả năng xảy ra của sự kiện đó. Việc gắn cho mỗi sự kiện một "số đo" khả năng xảy ra của nó phải đảm bảo tính khách quan, tính hợp lý và tính phi mâu thuẫn. Trong mục này chúng ta sẽ đưa ra các định nghĩa của xác suất. Mỗi dạng có những ưu và nhược điểm nhất định. Tuy vậy, qua các dạng định nghĩa này có thể hình dung ra sự phát triển của môn xác suất, một môn học có nguồn gốc xuất phát từ những sòng bạc nhưng nhờ sự tự hoàn thiện trong quá trình phát triển nên môn xác suất không những có đầy đủ các yếu tố cơ bản của một ngành khoa học chính xác mà còn là một trong những ngành của Toán học có thể hỗ trợ cho tất cả các lĩnh vực khoa học khác từ khoa học tự nhiên đến khoa học kĩ thuật và kể cả những ngành tưởng như xa lạ với Toán học đó là các ngành khoa học xã hội.

2. Định nghĩa xác suất theo quan niệm đồng khả năng.

2.1 Phép thử đồng khả năng: Một phép thử đồng khả năng là một phép thử mà các kết quả trực tiếp (còn gọi là sự kiện sơ cấp) ứng với phép thử này có khả năng xuất hiện như nhau sau khi thử. Chẳng hạn khi ta gieo một con xúc xắc cân đối và đồng chất thì việc xuất hiện một trong các mặt có số chấm từ 1 đến 6 là có khả năng như nhau hoặc khi chọn ngẫu nhiên hai trong năm người A, B, C, D, E thì việc chọn được AB hoặc CD . . . DE là có khả năng xuất hiện như nhau.

2.2 Định nghĩa xác suất theo quan niệm đồng khả năng:

Xét một phép thử đồng khả năng. Giả sử sau phép thử này có một trong n sự kiện sơ cấp có thể xảy ra và có một trong n_A sự kiện sơ cấp xảy ra kéo theo A xảy ra. Ta thấy lấy $\frac{n_A}{n}$ làm số đo khách quan xảy ra sự kiện A là hợp lý. Vì vậy ta có định nghĩa sau:

Định nghĩa: Xác suất của sự kiện A là số $P(A) = \frac{n_A}{n}$

* n là số kết quả đồng khả năng sau phép thử

* n_A là số kết quả xảy ra kéo theo A xảy ra hoặc số kết quả thuận lợi cho sự kiện A hay số kết quả hợp thành sự kiện A

Việc tính xác suất dựa trên định nghĩa trên phải thực hiện theo trình tự sau:

* Xét phép thử đang quan sát có phải là phép thử đồng khả năng không
* Nếu phép thử là đồng khả năng thì phải tìm số sự kiện đồng khả năng n
* Để tính xác suất của sự kiện A ta phải tìm số kết quả kéo theo A sau đó sử dụng định nghĩa

$$P(A) = \frac{n_A}{n}$$

2.3 Các ví dụ

Ví dụ 2.1: Gieo hai đồng tiền cân đối và đồng chất. Tính xác suất để cả hai cùng xuất hiện mặt quốc huy.

Gọi A là sự kiện cả hai đồng tiền cùng xuất hiện mặt quốc huy.

Ta có: Số sự kiện đồng khả năng: n = 4

Số sự kiện kéo theo A: $n_A = 1$ .Vậy $P(A) = \frac{1}{4}$

Ví dụ 2.2: Một đàn gà có bốn con gà ri gồm hai mái hai trống và sáu con gà tam hoàng gồm hai trống bốn mái. Chọn ngẫu nhiên hai con gà

Gọi A là sự kiện hai con gà được chọn đều là trống

B là sự kiện hai con gà được chọn gồm một trống một mái

C là sự kiện hai con gà được chọn là gà mái ri

Hãy tính xác suất của các sự kiện A, B, C

Ta có: Số sự kiện đầy khả năng là $C_{10}^2 = 45$

Số sự kiện kéo theo A là $C_4^2 = 6$

Số sự kiện kéo theo B là $C_4^1 C_6^1 = 24$

Số sự kiện kéo theo C là $C_2^2 = 1$

Vậy: $P(A) = \frac{6}{45} = \frac{2}{15}$, $P(B) = \frac{24}{45} = \frac{8}{15}$, $P(C) = \frac{1}{45}$

Ví dụ 2.3: Có ba gen X, Y, Z và ba gen x, y, z xếp ngẫu nhiên theo một dãy dọc. Tính xác suất để các gen x, y, z xếp liền nhau.

Gọi A là sự kiện cần tính xác suất

Số sự kiện đồng khả năng: n = 6! = 720

Số sự kiện kéo theo A: $n_A = 3!4! = 144$. Vậy: $P(A) = \frac{144}{720} = \frac{1}{5}$

Ví dụ 2.4: Hai cá thể bố và mẹ cùng có kiểu gen AaBb. Tính xác suất để cá thể con có kiểu gen giống kiểu gen của bố mẹ. Ta có bảng liên kết gen sau:

Bố \ Mẹ	AB	Ab	aB	ab
AB	AABB	AABb	AaBB	AaBb
Ab	AABb	AAbb	AaBb	Aabb
aB	AaBB	AaBb	aaBB	aaBb
ab	AaBb	Aabb	aaBb	aabb

Dựa vào bảng trên ta có: Số sự kiện đồng khả năng n = 16

Số sự kiện kéo theo A: $n_A = 4$. Vậy $P(A) = \frac{4}{16} = \frac{1}{4}$

3- Định nghĩa xác suất theo tần suất

Định nghĩa xác suất theo quan niệm đồng khả năng có ưu điểm là chỉ ra cách tính xác suất của một sự kiện rõ ràng và đơn giản. Tuy nhiên định nghĩa này chỉ áp dụng được với loại phép thử đồng khả năng và số kết quả sau phép thử là hữu hạn. Trong thực tế thường gặp những loại phép thử không có tính chất trên, để khắc phục hạn chế này ta có thể định nghĩa xác suất theo quan điểm thống kê.

3.1 Tần suất của sự kiện: Giả sử ta tiến hành n phép thử với cùng một hệ điều kiện thấy có n_A lần xuất hiện sự kiện A. Số n_A được gọi là tần số xuất hiện sự kiện A và tỉ số:

$f_n(A) = \frac{n_A}{n}$ gọi là tần suất xuất hiện sự kiện A.

Ta nhận thấy rằng khi n thay đổi n_A thay đổi vì thế $f_n(A)$ cũng thay đổi. Ngay cả khi tiến hành dãy n phép thử khác với cùng một điều kiện thì tần số và tần suất của n lần thử này cũng có thể khác tần số và tần suất của n lần thử trước. Tuy nhiên tần suất có tính ổn định nghĩa là khi số phép thử n khá lớn tần suất biến đổi rất nhỏ xung quanh một giá trị xác định. Để minh chứng cho nhận xét trên ta xét một ví dụ kinh điển về xác định tần số và tần suất việc xuất hiện mặt sấp (mặt không có chữ) của một đồng tiền do Buffon và Pearson thực hiện

Người làm thí nghiệm	Số lần tung 1 đồng tiền	Tần số mặt sấp	Tần suất mặt sấp
Buffon	4040	2040	0.5080
Pearson	12000	6010	0.5010
Pearson	24000	12012	0.5005

Ta nhận thấy rằng khi số lần tung tiền n tăng lên, tần suất xuất hiện mặt sấp ổn định dần về giá trị 0,5 được lấy làm xác xuất xuất hiện mặt sấp khi tung một đồng tiền cân đối và đồng chất.

3.2 Định nghĩa: Xác suất của một sự kiện là trị số ổn định của tần suất khi số phép thử tăng lên vô hạn.

Việc khẳng định tần suất của một sự kiện ổn định (hay tiến tới) một giá trị xác định khi số phép thử tăng lên vô hạn được đảm bảo bởi định lý Bernoulli sẽ được phát biểu và chứng minh trong chương sau. Tuy định nghĩa xác suất bằng tần suất không chỉ ra giá trị cụ thể xác suất của sự kiện nhưng trong thực tế khi số lần thử n là lớn ta thường lấy tần xuất $f_n(A)$ thay cho xác suất của sự kiện A. Vào cuối thế kỷ 19 nhà toán học Laplace theo dõi các bản thống kê về dân số trong vòng 10 năm của London, Peterbua, Berlin và nước Pháp ông ta tìm ra tần suất sinh con trai của ba vùng trên và cả nước Pháp là $\frac{22}{43}$. Khi xem xét tỉ lệ sinh con trai của Paris ông tìm được tần suất $\frac{25}{49}$, tần suất này nhỏ hơn $\frac{22}{43}$. Ngạc nhiên về sự khác nhau đó, Laplace điều tra thêm và tìm ra hai điều thú vị sau:

Một là: Vào thời bấy giờ các trẻ em đẻ ra không ghi tên cha trong giấy khai sinh thì dù sinh ở Marseille, Bordeaux hay bất cứ ở nơi nào trên đất Pháp đều có trong bản thông kê trẻ sinh ở Paris.

Hai là: Phần lớn những đứa trẻ nói trên đều là con gái.

Sau khi loại những đứa trẻ không sinh ở Paris ra khỏi danh sách này thì tỉ lệ trẻ trai ở Paris trở về con số $\frac{22}{43}$.

Qua ví dụ nêu trên chúng tôi muốn các nhà nông học tương lai khi quan sát hoặc thí nghiệm thấy có một số liệu nào đó khác với số liệu đã biết thì cần phải tìm nguyên do sự khác biệt này xuất phát từ đâu, rất có thể qua đó ta có thể phát hiện được những điều bổ ích phục vụ cho chuyên môn.

4. Định nghĩa xác suất bằng hình học

Với những phép thử đồng khả năng mà số kết quả sau một phép thử là vô hạn thì việc sử dụng định nghĩa xác suất ở mục 2 để tính xác suất của một sự kiện là không thực hiện được. Để khắc phục hạn chế này người ta đưa ra định nghĩa xác suất bằng hình học.

4.1 Độ đo của một miền: Giả sử D là một miền hình học nào đó chẳng hạn D là một đoạn thẳng, một hình phẳng hay một khối không gian. Số đo độ dài, diện tích, thể tích tương ứng được gọi là độ đo của miền D và kí hiệu là m(D)

4.2. Định nghĩa :

Xét một phép thử với vô hạn kết quả đồng khả năng, giả sử có thể thiết lập sự tương ứng một - một mỗi kết quả với một điểm thuộc miền G có độ đo là m(G) . Mỗi kết quả kéo theo sự kiện A tương ứng với mỗi điểm thuộc miền $D \subset G$ có độ đo m(D).

Xác suất của sự kiện A là số $P(A) = \frac{m(D)}{m(G)}$

Ví dụ 1: Một đường dây cáp quang nối Hà Nội với thành phố Hồ Chí Minh dài 1800 km gặp sự cố kĩ thuật làm tắc nghẽn việc thông tin liên lạc. Sự cố kĩ thuật có thể xảy ra ở bất cứ một vị trí nào trên đường cáp quang trên với cùng một khả năng. Tính xác suất để sự cố kĩ thuật xảy ra cách Hà Nội không quá 300km.

Miền G ở đây là đường cáp quang nối Hà Nội- thành phố Hồ Chí Minh có m(G) = 1800. Miền D tương ứng với sự kiện cần tính xác suất là đoạn cáp quang từ Hà nội tới vị trí cách Hà Nội 300 km, m(D) = 300.

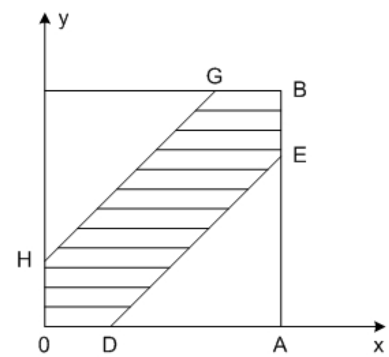
Vậy xác suất cần tính $P = \frac{300}{1800} = \frac{1}{6}$.

Ví dụ 2: Hai người A, B hẹn gặp nhau tại một địa điểm trong quãng thời gian từ 12 giờ đến 13 giờ theo qui ước, người đến trước đợi người đến sau không quá 15 phút. Tính xác suất để hai người gặp được nhau. Biết rằng mỗi người có thể đến điểm hẹn vào bất cứ thời điểm nào trong quãng thời gian nói trên.

Gọi x là thời điểm A đến chỗ hẹn, y là thời điểm B đến chỗ hẹn, $0 \le x, y \le 60$

Việc hai người đến chỗ hẹn tương ứng với điểm M(x, y) thuộc hình vuông OABC có cạnh dài 60 đơn vị dài. Hai người gặp được nhau

$\Leftrightarrow |x-y| \le 15 \Leftrightarrow x-15 \le y \le x+15 \Leftrightarrow$ M(x, y) thuộc hình ODEBGH.



Hình 1

Ta có miền G là hình vuông OABC, miền D là hình ODEBGH.

$m(G) = 60^2$, $m(D) = 60^2 - 45^2$.

Vậy xác suất cần tính $P = \frac{m(D)}{m(G)} = \frac{60^2 - 45^2}{60^2} = 1 - \frac{9}{16} = \frac{7}{16}$

Một số bài toán thực tế như quá trình thụ phấn, quá trình thụ tinh .... có thể áp dụng như bài toán gặp gỡ nói trên.

5. Hệ tiên đề Kolmogorop

Mặc dù ra đời từ thế kỉ 17 nhưng do nguồn gốc xuất phát và những khái niệm được nêu ra có tính mô tả thiếu những luận cứ khoa học nên cả một quãng thời gian dài từ thế kỉ 17 đến trước những năm 30 của thế kỉ 20 xác suất không được coi là một ngành toán học chính thống. Mãi tới năm 1933 khi nhà toán học Nga A.N Kolmogorop xây dựng hệ tiên đề cho lý thuyết xác suất thì xác suất mới được công nhận là một ngành toán học chính thống sánh ngang hàng với nhiều ngành toán học khác như số học, hình học, đại số, giải tích...

Tuy được chấp nhận muộn màng nhưng xác suất đã có mặt trong hầu hết các lĩnh vực khoa học từ khoa học tự nhiên , khoa học kĩ thuật dến khoa học xã hội. Vì là một giáo trình dành cho các ngành không chuyên về toán chúng tôi chỉ có ý định trình bày sơ lược hệ tiên đề về lý thuyết xác suất do A.N Kolmogorop đưa ra

Xét C là một σ - đại số các sự kiện . Xác suất P là một hàm xác định trên C thoả mãn :

1/ $P(A) \ge 0 \;\; \forall A \in C$

2/ $P(\Omega) = 1$

3/ Nếu $A_1, A_2, \ldots, A_n, \ldots$ . . . . xung khắc từng đôi, $A_n \in C$, $n = 1,2,\ldots$ thì

$$P(\sum_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$$

Bộ ba (Ω, A, P) được gọi là không gian xác suất.

II Các tính chất và các định lý

1. Các tính chất.

Để đơn giản, ta chỉ sử dụng định nghĩa theo quan điểm đồng khả năng để chứng minh các tính chất sẽ nêu trong mục này. Tuy nhiên các tính chất đó cũng đúng với mọi dạng định nghĩa xác suất khác.

1/ $0 \le P(A) \le 1$ vì $0 \le n_A \le n \Rightarrow 0 \le \dfrac{n_A}{n} = P(A) \le \dfrac{n}{n} = 1$

2/ $P(\phi) = 0, P(\Omega) = 1$ vì $n_\phi = 0, n_\Omega = n$ suy ra điều cần chứng minh.

3/ Nếu $A \cap B = \phi$ thì $P(A+B) = P(A) + P(B)$

Gọi n_A là số sự kiện kéo theo A, n_B là số sự kiện kéo theo B do A xung khắc với B nên số sự kiện kéo theo A + B là

$$n_{A+B} = n_A + n_B \Rightarrow P(A+B) = \frac{n_{A+B}}{n} = \frac{n_A + n_B}{n} = \frac{n_A}{n} + \frac{n_B}{n} = P(A) + P(B)$$

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

Gọi n_A là số sự kiện kéo theo A, n_B là số sự kiện kéo theo B, n_{AB} là số sự kiện kéo theo AB, $n_{A\cup B}$ là số sự kiện kéo theo $A \cup B$. Ta có

$$n_{A\cup B} = n_A + n_B - n_{AB} \Rightarrow P(A \cup B) = \frac{n_{A\cup B}}{n} = \frac{n_A + n_B - n_{AB}}{n}$$

$$\Rightarrow P(A \cup B) = \frac{n_A}{n} + \frac{n_B}{n} - \frac{n_{AB}}{n} = P(A) + P(B) - P(AB)$$

Hệ quả 1: $P(\overline{A}) = 1 - P(A)$. Thật vậy ta có

$A + \overline{A} = \Omega \Rightarrow P(A + \overline{A}) = P(\Omega) \Rightarrow P(A) + P(\overline{A}) = 1 \Rightarrow$ điều cần chứng minh.

Hệ quả 2: Nếu $A_1, A_2, \ldots A_n$ xung khắc từng đôi thì $P(\sum_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i)$

Áp dụng nhiều lần tính chất 1.3 ta có hệ quả trên.

5/ Nếu $A \subset B \Rightarrow P(A) \le P(B)$

Vì $A \subset B \Rightarrow n_A \le n_B \Rightarrow P(A) = \dfrac{n_A}{n} \le \dfrac{n_B}{n} = P(B)$

2. Xác suất có điều kiện

Xét hai sự kiện A và B trong một phép thử được tiến hành ứng với một bộ điều kiện nào đó. Việc xuất hiện sự kiện này đôi khi ảnh hưởng đến xác suất xuất hiện của sự kiện kia và ngược lại .

Chẳng hạn trong một hộp có 3 bi trắng và 2 bi đỏ, rút lần lượt 2 bi. Lần đầu rút được bi trắng hay không rõ ràng ảnh hưởng đến xác suất xuất hiện bi trắng ở lần thứ hai.

2.1. Định nghĩa: Xác suất của sự kiện A với giả thiết sự kiện B đã xảy ra là xác suất có điều kiện của A với điều kiện B.

Ta kí hiệu xác suất này là P(A/B) hoặc $P_B(A)$

Ví dụ 2.1: Quay lại ví dụ vừa nêu trên. Gọi B là sự kiện lần đầu rút được bi trắng , A là sự kiện lần sau cũng rút được bi trắng. Ta có P(A/B)=$\frac{2}{4}=\frac{1}{2}$ còn $P(A/\overline{B})=\frac{3}{4}$. Rõ ràng việc xuất hiện hay không xuất hiện B ảnh hưởng tới xác suất xuất hiện A.

Ví dụ 2.2: Tính trạng hoa vàng gen A là tính trạng trội, hoa trắng gen a là tính trạng lặn. Hai cây đậu hoa vàng dị hợp tử (cùng mang gen Aa) đem lai với nhau các cá thể con có các kiểu gen AA, Aa, aA, aa vơi cùng một khả năng. Chọn một cá thể con thì thấy cá thể này có hoa màu vàng. Tính xác suất để cá thể đó là đồng hợp tử

Gọi B là sự kiện cá thể con có hoa màu vàng, A là sự kiện cá thể con có gen đồng hợp tử.

Ta có: $P(A/B)=\frac{1}{3}$

2.2 Công thức xác suất có điều kiện

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

Thật vậy gọi n_B là số sự kiện kéo theo B(do giả thiết B đã xảy ra nên $n_B\neq 0$, gọi n_{AB} là sự kiện kéo theo AB

Ta có $P(A/B)=\frac{n_{AB}}{n_B}=\frac{\frac{n_{AB}}{n}}{\frac{n_B}{n}}=\frac{P(AB)}{P(B)}$

3. Công thức nhân xác suất

Từ $P(A/B)=\frac{P(AB)}{P(B)}\Rightarrow P(AB)=P(B)P(A/B)$ (1)

Thay đổi vai trò của A và B cho ta có P(AB) = P(A)P(B/A)

Mở rộng ta có: $P(A_1A_2...A_n)=P(A_1)P(A_2/A_1)...P(A_n/A_1A_2...A_{n-1})$ (2)

Công thức trên gọi là công thức nhân xác suất. Áp dụng liên tiếp công thức (1) nhiều lần ta có công thức (2)

Ví dụ 3.1: Có 6 cây đậu hoa vàng và 2 cây đậu hoa trắng lấy lần lượt 2 cây đậu. Tính xác suất để cả 2 cây đậu lấy ra là cây đậu hoa vàng.

Gọi A là sự kiện cả 2 cây lấy ra là đậu hoa vàng

A_1 là sự kiện cây lấy ra lần đầu màu vàng

A_2 là sự kiện cấy lấy ra lần hai màu vàng

Ta có: $A = A_1A_2$ từ đó suy ra

$$P(A) = P(A_1A_2) = P(A_1)P(A_2/A_1) = \frac{6}{8}.\frac{5}{7} = \frac{30}{56} = \frac{15}{28}$$

Sử dụng định nghĩa xác suất theo quan điểm đồng khả năng ta cũng có kết quả trên.

Ví dụ 3.2: Một giống lúa mới tại một trại lai tạo giống trước khi đưa ra sản xuất đại trà phải tiến hành liên tiếp ba lần kiểm định do ba trung tâm khảo cứu giống cấp một, cấp hai, cấp ba tiến hành. Nếu giống lúa được chấp nhận ở trung tâm cấp dưới thì được chuyển lên trung tâm cấp trên để kiểm định tiếp. Qua thống kê cho thấy giống của trại trên được trung tâm cấp một chấp nhận với xác suất 0,7. Sau khi chuyển lên trung tâm cấp hai nó được chấp nhận với xác suất 0,8. Nếu được chuyển lên trung tâm cấp ba nó được chấp nhận với xác suất 0,9. Tính xác suất để giống lúa được đưa ra sản xuất đại trà.

Gọi: A là sự kiện giống lúa được đưa ra sản xuất đại trà.

A_i là sự kiện giống lúa được chấp nhận ở trung tâm cấp i.

Ta có: $A = A_1A_2A_3$

$\Rightarrow P(A) = P(A_1A_2A_3) = P(A_1)P(A_2/A_1)P(A_3/A_1A_2) = 0{,}7.0{,}8.0{,}9 = 0{,}486$

4. Các sự kiện độc lập.

4.1 Hai sự kiện độc lập: Sự kiện A được gọi là độc lập với sự kiện B nếu:

$$P(A/B) = P(A)$$

Từ định nghĩa trên ta có

* Nếu A độc lập với B thì P(AB)=P(A)P(B)

Thật vậy P(AB)=P(B)P(A/B) =P(B)P(A)

* Nếu A độc lập với B thì B cũng độc lập với A

Do P(AB)=P(A)P(B/A) =P(B)P(A)$\Rightarrow$P(B/A)=P(B). Do vậy B cũng độc lập với A.

* A độc lập với B $\Leftrightarrow$ P(AB)= P(B)P(A)

4.2. Hệ độc lập từng đôi và độc lập hoàn toàn

Hệ: $A_1, A_2, ..., A_n$ được gọi là độc lập từng đôi nếu A_i độc lập A_j $\forall i \neq j$

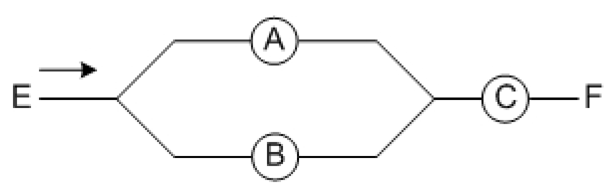
Hệ: $A_1, A_2, ..., A_n$ được gọi là độc lập hoàn toàn nếu

$P(A_i / A_{j_1}A_{j_2}...A_{j_k}) = P(A_i)\ \forall \{A_{j_1}, A_{j_2}, ...A_{j_k}\} \subset \{A_1, A_2, ..., A_n\}$

Từ định nghĩa trên ta thấy hệ độc lập hoàn toàn thì độc lập từng đôi nhưng điều ngược lại nói chung không đúng.

4.3. Các ví dụ

Ví dụ 4.1: Một mạng cấp nước như hình vẽ



Hình 2

Nước được cấp từ E đến F qua ba trạm bơm tăng áp A, B, C. Các trạm bơm làm việc độc lập với nhau. Xác suất để các trạm bơm A,B,C có sự cố sau một thời gian làm việc lần lượt là: 0,1; 0,1; 0,05. Tính xác suất để vùng F mất nước

Gọi: F là sự kiện vùng F mất nước

A là sự kiện trạm A có sự cố

B là sự kiện trạm B có sự cố

C là sự kiện trạm C có sự cố

Ta có: $F = (A \cap B) \cup C \Rightarrow P(F) = P[(A \cap B) \cup C]$

$= P(AB)+P(A)-P(ABC) = P(A)P(B)+P(C)-P(A)P(B)P(C)$

$= 0,01 + 0,05 - 0,005 = 0,055$

Ví dụ 4.2: Có hai lồng gà giống. Lồng thứ nhất có 2 gà trống, 4 gà mái. Lồng thứ hai có 4 gà trống, 2 gà mái. Lấy ngẫu nhiên từ mỗi lồng ra 1 con. Tính xác suất để 2 con gà lấy ra đều là gà mái

Gọi : A_1 là sự kiện con gà lấy ra ở lồng một là gà mái

A_2 là sự kiện con gà lấy ra ở lồng hai là gà mái

Ta có: $P(A_1A_2) = P(A_1)P(A_2) = \frac{4}{6}.\frac{2}{6} = \frac{2}{9}$

5. Dãy phép thử độc lập: Trong thực tế nhiều khi ta gặp những phép thử hợp gồm một dãy liên tiếp các phép thử như nhau được lặp đi lặp lại n lần và để ý đến sự xuất hiện của một sự kiện A nào đó trong n lần thử này. Chẳng hạn khi gieo một đồng tiền cân đối và đồng chất n lần hoặc tung một con xúc xắc cân đối và đồng chất n lần thì những phép thử thuộc loại này chính là dãy phép thử độc lập.

5.1. Lược đồ Bernoulli. Tiến hành một dãy n phép thử mà phép thử sau độc lập với các phép thử trước đó, xác suất xuất hiện sự kiện A ở mỗi phép thử là như nhau và bằng p

($p \neq 0$, $p \neq 1$). Dãy n phép thử độc lập loại này còn được gọi là một lược đồ Bernoulli.

5.2. Công thức Bernoull: Trong một lược đồ Bernoulli sự kiện A có thể xuất hiện từ 0 đến n lần. Gọi B_k là sự kiện A xuất hiện đúng k lần trong lược đồ Bernoulli. ta xây dựng công thức tính $P(B_k)$

Gọi A_i là sự kiện A xuất hiện ở lần thứ i trong n lần thử

Ta có $B_k = A_1A_2...A_k\overline{A}_{k+1}...\overline{A}_n + ... + \overline{A}_1...\overline{A}_{n-k}A_{n-k+1}...A_n$. Mỗi sự kiện của tổng các sự kiện trên gồm tích của n sự kiện trong đó A xuất hiện k lần và $\overline{A}$ xuất hiện n-k lần. Mỗi tích trên tương ứng với việc chọn ra k phép thử (A xuất hiện) từ n phép thử đã cho, theo lý thuyết tổ hợp có tất cả C_n^k tích như vậy.

Do n phép thử là độc lập $P(A_i) = p$, $P(\overline{A}_j) = 1 - p = q$

nên $P(A_1A_2...A_k\overline{A}_{k+1}...\overline{A}_n) = ... = P(\overline{A}_1...\overline{A}_{n-k}A_{n-k+1}...A_n) = p^kq^{n-k}$

Suy ra: $P(B_k) = C_n^k\, p^kq^{n-k}$

Đây là công thức Bernoulli cho ta biết xác suất A xuất hiện k lần trong một lược đồ Bernoulli

Gọi: $P_n(k)$ là xác suất để sự kiện A xuất hiện k lần trong một lược đồ Bernoulli và

$P_n(k_1, k_2)$ là xác suất để A xuất hiện trong khoảng từ k_1 đến k_2 lần ($k_1 < k_2$)

Ta có: $P_n(k) = P(B_k) = C_n^k\, p^k q^{n-k}$

$$P_n(k_1,k_2) = \sum_{k=k_1}^{k_2} P_n(k) = \sum_{k=k_1}^{k_2} C_n^k p^k q^{n-k}$$

Ví dụ 5.1: Xác suất để một quả trứng gà đem ấp nở ra gà con là 0,8. Đem ấp 5 quả trứng. Tính xác suất để có 3 quả nở ra gà con?

Ta có một lược đồ Bernoulli với n = 5, p = 0,8. Xác suất cần tính là

$$P_5(3) = C_5^3 0{,}8^3 0{,}2^2 = 0{,}2048$$

Ví dụ 5.2: Tỉ lệ đậu hoa vàng đồng hợp tử gen AA, hoa vàng dị hợp tử gen Aa và hoa trắng gen aa là 1 : 2 : 1. Chọn10 hạt đậu đem gieo

1/Tính xác suất để có 4 cây đậu hoa vàng là đồng hợp tử

2/ Tính xác suất để có 5 cây đậu hoa vàng

Nếu chỉ xét tới các cây đậu hoa vàng đồng hợp tử trong số cây đậu ta có lược đồ Bernoullie với

$p_1 = \frac{1}{4}, q_1 = \frac{3}{4}$. Vậy xác suất cần tính là

$$P_{10}(4) = C_{10}^4 (\frac{1}{4})^4.(\frac{3}{4})^6$$

Trong trường hợp thứ 2 ta có $p_2 = \frac{3}{4}, q_1 = \frac{1}{4}$ và xác suất cần tính

$$P_{10}(5) = C_{10}^5 (\frac{3}{4})^5.(\frac{1}{4})^5$$

5.3. Số lần xuất hiện chắc nhất: Xét một lược đồ Bernoullie với số lần thử n và xác suất xuất hiện sự kiện A, P(A) = p .

k_0 được gọi là số lần xuất hiện chắc nhất hoặc số lần xuất hiện có khả năng nhất nếu:

$P_n(k_0) \geq P_n(k)\ \forall\ k = 0, 1..., n$. Để tìm k_0 ta chỉ cần xét dãy số $P_n(0), P_n(1),...P_n(k),...P_n(n)$ xem số nào lớn nhất thì k ứng với số đó chính là k_0 cần tìm. Tuy nhiên việc tính tất cả các số trong dãy số trên sẽ mất nhiều thời gian. Vì vậy ta đưa ra thuật toán tìm số lần xuất hiện chắc nhất từ nhận xét sau. Trong dãy số $u_1, u_2,... u_n$ nếu $\frac{u_{k+1}}{u_k}$ còn lớn hơn 1 thì dãy số còn tăng đến khi nào $\frac{u_{k+1}}{u_k}$ nhỏ hơn 1 thì dãy số bắt đầu giảm. Số k_0 mà từ đó dãy chuyển từ tăng sang giảm là số cần tìm. Áp dụng nhận xét trên ta xét

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

$$\Rightarrow P_n(k+1) > P_n(k) \Leftrightarrow np - kp > kq + q \Leftrightarrow np - q > k(p+q) \Leftrightarrow np - q > k$$

Do np - q là một hằng số nên khi k còn nhỏ hơn np - q dãy còn tăng tới khi k vượt qua np – q thì dãy bắt đầu giảm.

Nếu np - q không phải là số nguyên số lần xuất hiện chắc nhất là: $k_0 = [np-q]+1$

Chú ý: Phần nguyên của số thực x là số nguyên lớn nhất nhỏ thua hoặc bằng x, kí hiệu là $[x]$

Nếu np - q là số nguyên thì số lần xuất hiện chắc nhất là $k_0 = np - q$ và k_0+1.

Khi đó ta có $P_n(k_0) = P_n(k_0+1) \geq P_n(k)\ \forall k = \overline{0,n}$

Ví dụ 5.3: Xác suất để mỗi cây sống sau thời gian trồng là 0,8. Trồng 1000 cây, Tìm số cây có khả năng sống cao nhất

Ta có n =1000, p = 0,8, q = 0,2 $\Rightarrow$ np - q = 799,8

Vậy số cây có khả năng sống cao nhất $k_0 = 800$

6. Công thức xác suất toàn phần

Xét $A_1, A_2, \ldots, A_n$ là một hệ đầy đủ các sự kiện, A là một sự kiện nào đó.

Ta có:

$$A = A\Omega = A(A_1 + A_2 + \ldots + A_n) = AA_1 + AA_2 + \ldots + AA_n$$

$$\Rightarrow P(A) = P(AA_1 + AA_2 + \ldots + AA_n)$$

Sử dụng công thức cộng và nhân xác suất ta có

$$P(A) = P(A_1)P(A/A_1) + P(A_2)P(A/A_2) + \ldots + P(A_n)P(A/A_n)$$

Công thức trên được gọi là công thức xác suất toàn phần

Ví dụ 6.1: Một kho hàng có 10 kiện hàng trong đó có 4 kiện do máy A sản xuất, 3 kiện do máy B sản xuất và 3 kiện còn lại do máy C sản xuất. Tỉ lệ sản phẩm loại hai do các máy sản xuất lần lượt là 0,02; 0,03; 0,05. Lấy ngẫu nhiên từ kho ra một kiện hàng rồi từ đó lấy ra một sản phẩm. Tính xác suất để sản phẩm lấy ra là sản phẩm loại hai

Gọi: A là sự kiện sản phẩm lấy ra là sản phẩm loại hai

A_i là sự kiện sản phẩm lấy ra do máy i sản xuất

Khi đó A_1, A_2, A_3 là một hệ đầy đủ $\Rightarrow A = AA_1 + AA_2 + AA_3$

Theo công thức xác suất toàn phần ta có

$$P(A) = P(A_1)P(A/A_1) + P(A_2)P(A/A_2) + P(A_3)P(A/A_3)$$

$$= \frac{4}{10}.0{,}02 + \frac{3}{10}.0{,}03 + \frac{3}{10}.0{,}05 = 0{,}032$$

Ví dụ 6.2: Một loài sinh vật có các kiểu gen AA, Aa, aa theo tỉ lệ: 1 : 2 : 1.

Nếu cá thể bố (mẹ) có kiểu gen AA lai với các thể mẹ (bố) có kiểu gen AA thì các cá thể con đều có kiểu gen AA.

Nếu cá thể bố (mẹ) có kiểu gen AA lai với các thể mẹ (bố) có kiểu gen Aa thì cá thể con có kiểu gen AA, Aa theo tỉ lệ 1 : 1.

Nếu cá thể bố (mẹ) có kiểu gen AA lai với các thể mẹ (bố) có kiểu gen aa thì cá thể con chỉ có các kiểu Aa. Chọn một cá thể con từ cá thể mẹ có kiểu gen AA.

1/ Tính xác suất để cá thể con có kiểu gen AA

2/ Tính xác suất để cá thể con có kiểu gen Aa

Gọi: A là sự kiện cá thể con có kiểu gen AA

B là sự kiện cá thể con có kiểu gen Aa

A_1 là sự kiện cá thể bố có kiểu gen AA

A_2 là sự kiện cá thể bố có kiểu gen Aa

A_3 là sự kiện cá thể bố có kiểu gen aa

Theo đầu bài :

$$P(A_1)=\frac{1}{4}\ ;\ P(A_2)=\frac{2}{4}\ ;P(A_3)=\frac{1}{4}\ ;\ P(A/A_1)=1\ ;\ P(A/A_2)=\frac{1}{2}\ ;P(A/A_3)=0$$

$$P(B/A_1)=0\ ;\ P(B/A_2)=\frac{1}{2}\ ;\ P(B/A_3)=1$$

Hệ: A_1, A_2, A_3 là hệ đầy đủ.

$$A = AA_1 + AA_2 + AA_3$$

suy ra $P(A) = P(A_1)P(A/A_1)+ P(A_2)P(A/A_2)+ P(A_3)P(A/A_3)$

$$=\frac{1}{4}.1+\frac{2}{4}.\frac{1}{2}+\frac{1}{4}.0=\frac{1}{4}+\frac{1}{4}=\frac{1}{2}$$

$$B = BA_1 + BA_2 + BA_3$$

suy ra $P(B) = P(A_1)P(B/A_1)+ P(A_2)P(B/A_2)+ P(A_3)P(B/A_3)$

$$=\frac{1}{4}.0+\frac{2}{4}.\frac{1}{2}+\frac{1}{4}.1=\frac{1}{4}+\frac{1}{4}=\frac{1}{2}$$

7. Công thức Bayes

Giả sử $A_1, A_2,...A_i. . . . A_n$ là một hệ đầy đủ các sự kiện. A là một sự kiện nào đó có $P(A) \neq 0$

Theo công thức xác suất toàn phần ta có:

$$P(A) = P(A_1)P(A/A_1)+ P(A_2)P(A/A_2)+... +P(A_i)P(A/A_i)+...+P(A_n)P(A/A_n)$$

Xét A_j là một sự kiện nào đó trong hệ các sự kiện đã cho

Ta có $P(A_j/A)=\dfrac{P(A_jA)}{P(A)}=\dfrac{P(A_j)P(A/A_j)}{\sum_{i=1}^{n}P(A_i)P(A/A_i)}$

Công thức trên được gọi là công thức Bayes. Các xác suất $P(A_j/A)$ gọi là các xác suất hậu nghiệm để phân biệt với các xác suất tiên nghiệm $P(A_i)$

Ví dụ 7.1: Cặp trẻ sinh đôi có thể là sinh đôi thật (do cùng một trứng sinh ra) trong trường hợp này chúng luôn cùng giới. Trường hợp cặp sinh đôi do hai trứng sinh ra gọi là giả sinh đôi. Nếu cặp sinh đôi do hai trứng sinh ra thì xác suất để chúng cùng giới là 1/2. Biết xác suất để cặp sinh đôi do cùng một trứng sinh ra là p. Một cặp trẻ sinh đôi ra đời biết chúng cùng giới. Tính xác suất để chúng là sinh đôi thật.

Gọi: A là sự kiện cặp trẻ sinh đôi là cùng giới

A_1 là sự kiện cặp trẻ sinh đôi là sinh đôi thật

A_2 là sự kiện cặp trẻ sinh đôi là giả sinh đôi

A_1, A_2 là hệ đầy đủ, $P(A_1) = p$; $P(A_2) = 1-p$; $P(A/A_1) = 1$; $P(A/A_2) = \frac{1}{2}$

Theo công thức Bayes ta có xác suất cần tính là:

$$P(A_1/A) = \frac{P(A_1)P(A/A_1)}{P(A_1)P(A/A_1) + P(A_2)P(A/A_2)} = \frac{p.1}{p.1 + (1-p)\frac{1}{2}} = \frac{2p}{1+p}$$

Ví dụ 7.2: Tỉ lệ người đến khám tại một bệnh viện mắc bệnh A là 60%, trong số những người mắc bệnh A có 50% mắc cả bệnh B, còn trong số những người không mắc bệnh A có 70% mắc bệnh B.

1/ Khám cho một người thì thấy người đó mắc bệnh B. Tính xác suất để người được khám cũng mắc bệnh A.

2/ Nếu người được khám không mắc bệnh B tìm xác suất để người đó không mắc bệnh A.

Gọi : A là sự kiện người chọn đi khám mắc bệnh A

B là sự kiện người chọn đi khám mắc bệnh B

Ta có A và $\overline{A}$ là một hệ đầy đủ, $P(A) = 0,6$; $P(\overline{A}) = 0,4$

Vì vậy ta có: $B = BA + B\overline{A}$

Xác suất cần tính ở phần 1 là $P(A/B) = \frac{P(A)P(B/A)}{P(B)}$

Xác suất cần tính ở phần 2 là $P(\overline{A}/\overline{B}) = \frac{P(\overline{A})P(\overline{B}/\overline{A})}{P(\overline{B})}$

Ta có: $P(B/A) = 0,5$; $P(\overline{B}/\overline{A}) = 0,3$.

Suy ra: $P(B) = P(A)P(B/A) + P(\overline{A})P(B/\overline{A})$

$= 0,6.0,5 + 0,4.0,7 = 0,58 \Rightarrow P(\overline{B}) = 0,42$

Vậy: $P(A/B) = \frac{0,30}{0,58} = \frac{15}{29}$, $\quad P(\overline{A}/\overline{B}) = \frac{0,4.0,3}{0,42} = \frac{2}{7}$

Ví dụ 7.3: Để gây đột biến cho một tính trạng người ta tìm cách tác động lên hai gen A, B bằng phóng xạ. Xác suất đột biến của tính trạng do gen A là 0,4; do gen B là 0,5 và do cả hai gen là 0,9.

1/ Tính xác suất để có đột biến ở tính trạng đó biết rằng phóng xạ có thể tác động lên gen A với xác suất 0,7 và lên gen B với xác suất 0,6.

2/ Tính trạng có dấu hiệu đột biến. Xác định vai trò đóng góp của từng gen

Gọi : C là sự kiện có đột biến ở tính trạng đang xét

A là sự kiện phóng xạ tác dụng lên gen A

B là sự kiện phóng xạ tác dụng lên gen B

C_1 là sự kiện phóng xạ chỉ tác động lên gen A

C_2 là sự kiện phóng xạ chỉ tác dụng lên gen B

C_3 là sự kiện phóng xạ tác dụng lên cả 2 gen

C_4 là sự kiện phóng xạ không tác dụng lên gen nào

Khi đó hệ C_1, C_2, C_3, C_4 là một hệ đầy đủ

$C_1 = A\overline{B}$, $C_2 = \overline{A}B$, $C_3 = AB$, $C_4 = \overline{A}\,\overline{B}$. Mặt khác A, B là độc lập nên

$P(C_1) = P(A)P(\overline{B}) = 0{,}28$, $P(C_2) = P(\overline{A})P(B) = 0{,}18$

$P(C_3) = P(A)P(B) = 0{,}42$; $P(C_4) = P(\overline{A})P(\overline{B}) = 0{,}12$

Mặt khác $P(C/C_1) = 0{,}4$; $P(C/C_2) = 0{,}5$; $P(C/C_3) = 0{,}9$ và $P(C/C_4) = 0$

Theo công thức xác suất toàn phần ta có

$P(C) = 0{,}28.0{,}4 + 0{,}18.0{,}5 + 0{,}42.0{,}9 + 0{,}12.0 = 0{,}58$

Vai trò đóng góp của riêng gen A cho sự đột biến là

$$P(C_1/C) = \frac{P(C_1)P(C/C_1)}{P(C)} = \frac{112}{580} \approx 0{,}1931$$

Vai trò đóng góp của riêng gen B cho sự đột biến là

$$P(C_2/C) = \frac{P(C_2)P(C/C_2)}{P(C)} = \frac{90}{580} \approx 0{,}1552$$

Vai trò đóng góp của cả hai gen cho sự đột biến là

$$P(C_3/C) = \frac{P(C_3)P(C/C_3)}{P(C)} = \frac{378}{580} \approx 0{,}6517$$

Bài tập chương II

1.Tung đồng thời 3 đồng tiền. Tính xác suất để cả 3 đồng tiền cùng xuất hiện mặt sấp.

2. Một tổ học sinh có 4 nam, 4 nữ. Chọn ngẫu nhiên 3 học sinh.

a.Tính xác suất để trong 3 người được chọn có 2 nam, 1 nữ.

b. Tính xác suất để trong 3 người được chọn đều là nữ.

3. Có 6 tấm thẻ đánh số từ 1 đến 6. Rút lần lượt 3 tấm rồi đặt từ trái qua phải

a.Tính xác suất để số lập được là số chẵn

b. Tính xác suất để số lập được chia hết cho 3

c. Tính xác suất để số lập được chia hết cho 5

4. Có n người xếp theo một hàng dọc ($n > 5$)

a.Tính xác suất để 2 người A, B đứng liền nhau

b.Tính xác suất để 2 người A, B đứng cách nhau đúng 3 người

5. Một học sinh có 5 quyển sách Toán, 3 quyển sách Văn và 2 quyển sách Ngoại ngữ. Học sinh này xếp ngẫu nhiên các quyển sách này trên một ngăn của giá sách.

a. Tính xác suất để 5 quyển sách Toán đứng liền nhau

b. Tính xác suất để không có 2 quyển sách Toán nào xếp liền nhau

6. Chọn ngẫu nhiên 10 học sinh, tính xác suất để không có 2 học sinh nào có cùng sinh nhật.

7. Cho một lô hàng có n sản phẩm trong đó có m phế phẩm. Lấy ngẫu nhiên k sản phẩm ($k < n$, $k < m$). Tính xác suất để trong k sản phẩm lấy ra có l phế phẩm ($l < k$).

8*. Một đoàn tàu vào ga gồm có 4 toa, trên sân ga có 8 hành khách đợi lên tàu. Các hành khách này có lên một trong bốn toa trên một cách ngẫu nhiên

a. Tính xác suất để mỗi toa có đúng 2 hành khách mới lên

b. Tính xác suất để mỗi toa có đúng 4 hành khách mới lên

c. Tính xác suất để một toa có đúng 5 hành khách mới lên 3 toa còn lại mỗi toa có 1 hành khách lên.

9. Tại một trại lợn giống có 4 con lợn nái thuộc các loài A, B , C, D cho phối giống với 4 lợn đực cũng thuộc 4 loài trên một cách ngẫu nhiên .

a. Tính xác suất để các cặp lợn cùng loài phối giống với nhau

b. Tính xác suất để không có cặp nào cùng loài phối giống với nhau

10. Trong 10 hạt đậu giống có 4 hạt đậu hoa vàng thuần chủng, 3 hạt đậu hoa vàng không thuần chủng và 3 hạt đậu hoa trắng. Chọn ngẫu nhiên 3 hạt đậu

a. Tính xác suất để 3 hạt đậu được chọn gồm 3 loại khác nhau

b. Tính xác suất để 3 hạt đậu được chọn là đậu cho hoa màu vàng

c. Tính xác suất để 3 hạt đậu được chọn có ít nhất một hạt cho hoa màu trắng

11. Một đoạn thẳng có chiều dài 2l được bẻ ngẫu nhiên thành 3 đoạn. Tính xác suất để 3 đoạn này lập thành một tam giác.

12. Do thiếu kinh nghiệm một nhân viên thụ tinh nhân tạo cho bò chỉ chuẩn đoán được bò sẽ rụng trứng trong khoảng 0h sáng đến 24h cùng ngày. Biết rằng trứng và tinh trùng có thể sống trong tử cung không quá t giờ ($t < 12$).

a. Kỹ thuật viên tiến hành thụ tinh nhân tạo vào lúc 12h. Tính xác suất để việc thụ tinh thành công.

b. Kĩ thuật viên tiến hành việc thụ tinh nhân tạo một cách ngẫu nhiên trong quãng thời gian từ 10h đến 14h. Tính xác suất để việc thụ tinh thành công.

13. Lai gà lông màu nâu với gà lông màu trắng gà con ở thế hệ F_1 có lông màu nâu, màu xám và màu trắng theo tỉ lệ: 1 : 2 : 1. Chọn ngẫu nhiên 5 quả trứng ở thế hệ F_1 đem ấp và cả 5 quả trứng đều nở. Tính xác suất để:

a. Có đúng 3 gà con có lông màu nâu.

b. Có 2 gà con có lông màu nâu và 3 gà con có lông màu xám.

c. Có 1 gà con có lông màu nâu, 2 gà con có lông màu xám, 2 gà con có lông màu trắng.

14. Biết tỉ lệ người có nhóm máu O, A, B và AB trong cộng đồng tương ứng là:

34%, 37%, 21%, 8%. Người có nhóm máu O, A, B chỉ có thể nhận máu của người cùng nhóm với mình hoặc nhận từ người có nhóm máu O, còn người có nhóm máu AB có thể nhận máu từ bất cứ một người có nhóm máu nào. Có một người cần tiếp máu và một người cho máu. Việc truyền máu đã được thực hiện.

a. Tính xác suất để người nhận máu có nhóm máu A

b. Tính xác suất để người nhận máu có nhóm máu B

15. Một công ty có hai phòng chức năng. Phòng A gồm 3 nhân viên nam, 2 nhân viên nữ. Phòng B gồm 3 nhân viên nam, 3 nhân viên nữ. Để kiểm tra năng lực làm việc của mỗi phòng, giám đốc công ty quyết định chọn mỗi phòng 2 nhân viên để kiểm tra chuyên môn. Biết rằng mỗi nhân viên ở phòng A có thể vượt qua kỳ kiểm tra với xác suất 0,8 đối với nam và 0,7 đối với nữ. Mỗi nhân viên phòng B có thể vượt qua kỳ kiểm tra với xác suất 0,7 đối với nam và 0,8 đối với nữ.

a. Tính xác suất để 4 nhân viên được chọn đều là nam.

b. Tính xác suất để 4 nhân viên được chọn đều qua kì kiểm tra

c. Khả năng vượt qua kì kiểm tra của phòng nào cao hơn?

16. Một nhóm bệnh nhân gồm 6 người trong đó có 4 người mắc bệnh A và 5 người mắc bệnh B.

a. Tìm số bệnh nhân mắc cả hai loại bệnh

b. Chọn ngẫu nhiên 2 trong số 6 bệnh nhân nói trên. Tính xác suất để 2 người đó mắc cả hai loại bệnh.

c. Người ta định sử dụng một loại biệt dược X để điều trị cho nhóm bệnh nhân trên. Xác suất để một bệnh nhân chỉ mắc một loại bệnh khi sử dụng biệt dược X khỏi bệnh là 0,8. Xác suất để một bệnh nhân mắc cả hai loại bệnh khi sử dụng biệt dược X khỏi bệnh là 0,6. Chọn ngẫu nhiên hai bệnh nhân trong 6 bệnh nhân nói trên rồi cho dùng biệt dược X. Tính xác suất để cả hai bệnh nhân khỏi bệnh.

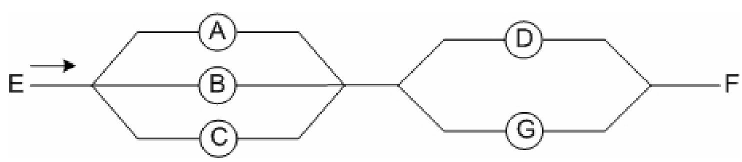
17. Ba phòng thí nghiệm được giao nhiệm vụ tạo giống lúa mới. Ba phòng làm việc độc lập, xác suất thành công tương ứng là 0,4; 0,3; 0,2.

a. Tính xác suất để có đúng một phòng thành công.

b. Tính xác suất để có ít nhất một phòng thành công.

c. Trong một năm nếu phòng nào thành công trong việc tạo ra giống lúa mới thì được coi là hoàn thành nhiệm vụ. Nếu thất bại được làm thêm một lần nữa và nếu lần này thành công thì cũng được coi là hoàn thành nhiệm vụ. Tính xác suất để cả ba phòng cùng hoàn thành nhiệm vụ.

18. Một mạng cung cấp điện như hình vẽ



Hình 3

Điện được cung cấp từ E tới khu tiêu dùng F qua năm trạm biến áp A, B, C, D, G. Các trạm biến áp này làm việc độc lập, xác suất để mỗi trạm biến áp A, B, C có sự cố kĩ thuật sau một thời gian hoạt động là 0,1. Xác suất trên với hai trạm D, G là 0,05.

a. Tính xác suất để F mất điện.

b. Biết F bị mất điện.Tính xác suất để cả 2 trạm D, G có sự cố.

19. Cho A, B là hai sự kiện có $P(A) = 0{,}45$; $P(B) = 0{,}30$; $P(A \cup B) = 0.60$. Hãy tính các xác suất sau:

a. $P(\overline{A}\overline{B})$; b. $P(B/A)$; c. $P(AB)$; d. $P(A/B)$

20. Cho $P(A) = 3/14$; $P(B) = 1/6$; $P(C) = 1/3$; $P(AC) = 1/7$; $P(B/C) = 5/21$.

Tính: a. $P(A/C)$; b. $P(C/A)$; c. $P(BC)$; d. $P(C/B)$

21. Nếu một cơn bão xuất hiện ở bờ biển Philippin thì cơn bão đó sẽ đổ bộ vào Việt Nam với tỉ lệ p_1. Kinh nghiệm cho biết xác suất để một cơn bão xuất hiện ở vùng biển này trong tháng Tám là p_2.

a. Tính xác suất để một cơn bão sẽ xuất hiện ở bờ biển Philippin và sẽ đổ bộ vào Việt Nam trong tháng Tám năm nay.

b. Nếu cơn bão hình thành ở vùng biển Philippin mà được làm nhẹ đi bằng kĩ thuật phun hoá chất khi bão qua vùng biển Trường sa thì khả năng nó đổ bộ vào Việt Nam sẽ giảm đi 1/4. Tính xác suất ở phần a trong trường hợp này. Biết rằng các cơn bão xuất hiện ở vùng biển Philippin khi đổ bộ vào đất liền luôn đi qua quần đảo Trường sa.

22. Một nhà phân tích thị trường chứng khoán xem xét triển vọng của các chứng khoán của nhiều công ty đang phát hành. Một năm sau 25% số chứng khoán tỏ ra tốt hơn nhiều so với trung bình của thị trường, 25% số chứng khoán tỏ ra xấu hơn nhiều so với trung bình của thị trường và 50% bằng trung bình của thị trường. Trong số những chứng khoán trở nên tốt có 40% được nhà phân tích đánh giá là mua tốt, 20% số chứng khoán là trung bình cũng được đánh giá là mua tốt và 10% số chứng khoán trở nên xấu cũng được đánh giá là mua tốt.

a. Tính xác suất để một chứng khoán được đánh giá là mua tốt sẽ trở thành tốt.

b. Tính xác suất để một chứng khoán được đánh giá là mua tốt sẽ trở thành xấu.

23. Một đại lý tại Hà Nội kinh doanh đồ uống do ba công ty A, B, C sản xuất theo tỉ lệ 2 : 3 : 5. Tỷ lệ đồ uống có ga tương ứng ở ba công ty trên là 70%, 60% và 50%.

a. Chọn ngẫu nhiên một kiện hàng tại kho của đại lý. Tính xác suất để kiện đồ uống được chọn là đồ uống có ga.

b. Biết kiện hàng được chọn là đồ uống có ga. Tính xác suất để kiện hàng đó do công ty A sản xuất.

24. Trong một kho số lượng rượu loại A và loại B là như nhau. Người ta chọn ngẫu nhiên từ trong kho ra một chai rượu và đưa cho 5 người sành rượu nếm thử để xem đây là loại rượu nào. Giả sử xác suất đoán đúng của mỗi người là 0,7. Có 3 người kết luận là rượu loại A, 2 người kết luận là rượu loại B. Tính xác suất để chai rượu trên là rượu loại A.

25. Một trung tâm phân phối giống cây trồng nhận cây giống từ 3 cơ sở khác nhau theo tỉ lệ: 2 : 3 : 5. Tỷ lệ cây giống xấu tương ứng là 5%, 3% và 2%.

a. Chọn ngẫu nhiên một cây giống của trung tâm. Tính xác suất để cây giống được chọn là cây xấu.

b. Biết cây giống được chọn là cây giống xấu. Khả năng cây giống đó thuộc cơ sở nào là cao nhất? Tại sao?

26. Cho lai gà lông xám thuần chủng (tính trạng trội) với gà lông trắng thuần chủng (tính trạng lặn) ở thế hệ F_1 tất cả gà con đều có lông màu xám , ở thế hệ F_2 gà có lông màu xám và màu trắng theo tỉ lệ 3 : 1. Biết tỉ lệ gà lông xám thuần chủng , gà lông xám không thuần chủng, gà lông trắng thuần chủng trong một đàn gà là 1 : 2 : 1. Một quả trứng gà của một gà mẹ lông xám không thuần chủng sắp nở ra một chú gà con.

a. Tính xác suất để gà con nở ra có lông màu trắng.

b. Biết gà con nở ra có lông màu xám. Tính xác suất để gà bố có lông màu trắng.

27. Tỉ lệ người có kí sinh trùng sốt rét trong máu của mỗi người dân vùng cao là 0,2.

a. Chọn ngẫu nhiên 4 người. Tính xác suất để trong 4 người được chọn có 3 người trong máu có kí sinh trùng sốt rét.

b. Lấy máu của 100 người đem thử. Tính xác suất để có ít nhất một người có kí sinh trùng sốt rét trong máu.

28. Có hai tổ học sinh. Tổ thứ nhất có 4 nam 5 nữ, tổ thứ hai có 5 nam 6 nữ. Chon ngẫu nhiên ra mỗi tổ 3 học sinh rồi ghép mỗi học sinh tổ này với mỗi học sinh của tổ kia làm một nhóm học tập

a.Tính xác suất để các nhóm học tập đều cùng giới.

b.Tính xác suất để các nhóm học tập đều khác giới

29. Nhân ngày quốc tế phụ nữ, sinh viên A vào cửa hàng hoa tại cổng trường mua ngẫu nhiên 3 bông hoa để tặng cho 3 bạn nữ mỗi người 1 bông. Sinh viên B cũng vào cửa hàng hoa này và cũng mua ngẫu nhiên 3 bông hoa để tặng cho 3 bạn nữ nói trên mỗi người 1 bông. Cửa hàng hoa chỉ bán 3 loại hoa là hồng bạch, hồng vàng và hồng nhung.
a. Tính xác suất để mỗi bạn nữ được tặng 2 bông hoa cùng màu.
b. Tính xác suất để mỗi bạn nữ được tặng 2 bông hoa gồm 2 màu khác nhau.

30. Một hộp đậu giống gồm 2 hạt đậu trắng và 3 hạt đậu đỏ. Một hộp khác gồm 3 hạt đậu trắng và 2 hạt đậu đỏ. Tỉ lệ nảy mầm là 0,8 đối với mỗi hạt đậu trắng, là 0,7 đối với mỗi hạt đậu đỏ. Lấy ngẫu nhiên từ mỗi hộp ra 2 hạt đem gieo.
a. Tính xác suất để cả 4 hạt đều nảy mầm.
b. Biết 4 hạt đem gieo đều nảy mầm. Tính xác suất để 4 hạt này đều là đậu đỏ.

31. Lai hai giống hoa màu hồng và màu đỏ thuần chủng, các cây con F_1 có thể cho hoa màu hồng, màu đỏ hoặc màu cánh sen với tỉ lệ 1: 1: 2. Chọn ngẫu nhiên 5 hạt hoa F_1 đem gieo. Tính xác suất để:
a. Có đúng 3 cây cho hoa màu đỏ.
b. Có 2 cây hoa màu đỏ, 3 cây màu hồng.
c. Có 1 cây màu đỏ, 1 cây màu hồng và 3 cây màu cánh sen.

32. Dưới tác động của phóng xạ các nhiễm sắc thể của một tế bào bị gãy làm hai mảnh trong đó chỉ có một mảnh chứa tâm động. Các mảnh gãy theo thời gian sẽ tự ghép lại với nhau một cách ngẫu nhiên và tế bào sẽ sống sót nếu mỗi cặp mảnh ghép với nhau chỉ chứa một tâm động. Tìm xác suất để tế bào sống sót, biết rằng tế bào đó có n nhiễm sắc thể bị gãy.

33. (Bài toán Buffon) Trên mặt phẳng có một dải các đường thẳng song song cách đều nhau một khoảng 2a. Gieo ngẫu nhiên một cái kim có chiều dài 2l (l < a). Tính xác suất để cái kim cắt một trong những đường thẳng trên.

34. Hai tầu thuỷ cập vào một cảng để trả hàng một cách độc lập trong vòng 24h. Biết rằng thời gia bốc dỡ hàng của tàu thứ nhất là 2h, của tàu thứ hai là 3h Tính xác suất để một trong hai tàu trên phải chờ để cập bến.

35. Có n viên bi bỏ ngẫu nhiên vào m cái hộp (n > m).

a. Tính xác suất để có đúng 1 hộp không chứa viên bi nào.

b. Tính xác suất để có 1 hộp chứa cả n viên bi.

c. Tính xác suất để mỗi hộp có ít nhất 1 viên bi.

36. (Bài toán Banach) Một nhà toán học có 2 bao diêm, mỗi bao có n que. Ông ta để mỗi bên túi 1 bao. Khi sử dụng nhà bác học rút ngẫu nhiên 1 bao rồi rút ra 1 que để dùng. Tìm xác suất để khi ông phát hiện 1 bao đã hết diêm thì bao kia còn k que.

37. Có k thùng hạt giống gồm k loại khác nhau được gửi đến một trung tâm bảo quản giống. Trung tâm này có k phòng được đánh số từ 1 đến k mỗi phòng bảo quản một loại hạt giống. Do người phụ trách kĩ thuật của trung tâm vắng mặt, nhân viên bảo vệ đành xếp tạm mỗi thùng hạt giống vào một phòng.

a. Tính xác suất để không thùng nào để đúng vị trí.

b. Tính xác suất để các thùng đều để đúng vị trí.

38*. Trên một toa tàu có 30 hành khách. Đến ga tiếp theo mỗi hành khách có thể xuống tàu với xác suất 0,3. Tại ga này mỗi hành khách mới có thể lên toa tàu trên với xác suất 0,5. Tính xác suất để khi ra khỏi ga toa tàu vẫn còn đủ 30 hành khách.

39. Một cửa hàng bán một loại sản phẩm trong đó 30% do nhà máy A sản xuất, 40% do nhà máy B sản xuất, 30% do nhà máy C sản xuất.

Tỷ lệ sản phẩm loại một của ba nhà máy trên lần lượt là: 0,9 ; 0,8 , 0,9.

a. Mua ngẫu nhiên một sản phẩm tại cửa hàng. Tĩm xác suất để sản phẩm mua được là loại một.

b. Biết sản phẩm mua đượclà loại một. Tính xác suất để sản phẩm đó do nhà máy A sản suất.

40. Tại một vùng dân cư, tỷ lệ người nghiện hút thuốc lá là 0,2. Biết rằng tỷ lệ viêm họng trong số người nghiện thuốc lá là 0,7 và với người không nghiện là 0,2. Khám ngẫu nhiên 1 người thì thấy người đó bị viêm họng. Tính xác suất để người đó nghiện thuốc lá.

41. Có 8 người rút thăm để chọn các căn hộ trong một chung cư từ tầng 8 đến tầng 15, mỗi tầng có 8 căn hộ.

a. Tính xác suất để để cả 8 người trên đều nhận được các căn hộ trong cùng một tầng.

b. Tính xác suất để 8 người trên nhận được 8 căn hộ trên 8 tầng khác nhau.

Chương 3 Biến ngẫu nhiên

Định nghĩa chính xác mang tính toán học thuần tuý về biến ngẫu nhiên vượt khỏi yêu cầu của giáo trình. Định nghĩa được trình bày ở đây mang tính mô tả, tuy nhiên nó cũng giúp cho người học hiểu được thế nào là biến ngẫu nhiên, biến ngẫu nhiên rời rạc, biến nhiên liên tục. Các khái niệm khác như bảng phân phối xác suất hàm phân phối cũng như hàm mật độ xác suất đều được trình bày với những kiến thức đơn giản nhất. Các số đặc trưng quan trọng nhất của biến ngãu nhiên như kì vọng, phương sai, độ lệch chuẩn được trình bày kĩ hơn các số đặc trưng khác.Các biến ngẫu nhiên rời rạc và liên tục thường gặp trong thực tế cũng như các số đặc trưng của chúng được giới thiệu khá kĩ. Khái niệm véctơ ngẫu nhiên được gới thiệu một cách sơ lược. Các ví dụ liên quan tới các kiến thức lý thuyết cũng như các ứng dụng thực tế giúp người học hiểu và có hứng thú hơn đối với môn học. Luật số lớn, một số định lí về luật số lớn và một số định lý giới hạn được giới thiệu sơ lược trong chương này.

I Biến ngẫu nhiên

Khi tiến hành một phép thử ngẫu nhiên, các kết quả của phép thử thường là các đặc tính định tính. Tuy nhiên trong nhiều phép thử mỗi một kết quả của phép thử thường được gán tương ứng với một giá trị định lượng nào đó. Chẳng hạn khi chơi các trò chơi ăn tiền mỗi kết quả của một lần chơi được gán tương ứng với một số tiền (đặc tính định lượng) mà người chơi được hay mất hoặc khi nhằm bắn một phát đạn vào bia, mỗi kết quả của việc bắn tương ứng với điểm số (đặc tính định lượng) mà xạ thủ đạt được.

1.Định nghĩa:

Biến ngẫu nhiên (thực) là biến nhận giá trị là các số thực phụ thuộc vào kết quả của phép thử ngẫu nhiên .

Ta thường dùng các chữ cái hoa X, Y, Z... để chỉ các biến ngẫu nhiên và các chữ cái thường x, y, z...hoặc x_i , y_j.... . để chỉ các giá trị cụ thể mà biến ngẫu nhiên đó nhận.

2. Các ví dụ:

Ví dụ 1: Tung đồng thời hai con xúc xắc. Gọi X là tổng số chấm ở hai mặt trên, X là một biến ngẫu nhiên và có thể nhận một trong các giá trị từ 2 đến 12

Ví dụ 2: Một người nhằm bắn vào bia cho tới khi trúng bia thì ngừng. Gọi Y là số đạn cần dùng. Y là một biến ngẫu nhiên nhận các giá trị:

1, 2, ..., n, ...

Ví dụ 3: Thắp sáng liên tục một bóng đèn điện cho tới khi dây tóc của bóng đèn bị cháy. Gọi Z là thời gian bóng đèn sáng. Z là một biến ngẫu nhiên

Qua ba ví dụ trên ta thấy có hai loại biến ngẫu nhiên:

Loại thứ nhất là loại biến ngẫu nhiên chỉ nhận một số hữu hạn hay vô hạn đếm được các giá trị.

*Một tập được gọi là vô hạn đếm được nếu tồn tại một phép tương ứng một - một tới tập các số tự nhiên N.

Loại thứ hai là loại biến ngẫu nhiên mà nó có thể nhận các giá trị trong một khoảng hoặc một số khoảng thực nào đó. Loại biến ngẫu nhiên thứ nhất gọi là biến ngẫu nhiên rời rạc. Loại biến ngẫu nhiên thứ hai gọi là biến ngẫu nhiên liên tục.

Việc đưa ra một định nghĩa thuần tuý toán học về biến ngẫu nhiên không được trình bày ở đây. Người đọc muốn biết có thể tham khảo ở các tài liệu dẫn ra ở cuối giáo trình này.

3. Bảng phân phối xác suất của một biến ngẫu nhiên rời rạc.

3.1. Định nghĩa: Bảng phân phối xác suất của một biến ngẫu nhiên rời rạc X là một bảng gồm hai dòng

Dòng trên ghi các giá trị có thể có của biến ngẫu nhiên X, dòng dưới ghi các xác suất tương ứng. Nếu X nhận 1 số hữu hạn các giá trị thì bảng phân phối xác suất của X là:

X	x_1	x_2 ...	x_i	...	x_n
P	p_1	p_2 ...	p_i	...	p_n

Nếu X nhận 1 số vô hạn đếm các giá trị thì bảng phân phối xác suất của X là

X	x_1	x_2 ...	x_i	...	x_n ...
P	p_1	p_2 ...	p_i	...	p_n ...

$p_i = P(X = x_i)$ là xác suất để X nhận giá trị là x_i.

Do X nhận và chỉ nhận một trong các giá trị x_i nên ta có $\sum_{i=1}^{n} p_i = 1$ đối với bảng thứ nhất và $\sum_{i=1}^{\infty} p_i = 1$ đối với bảng thứ hai.

3.2. Các ví dụ

Ví dụ 1: Một người chơi trò chơi ăn tiền bằng cách tung đồng thời 2 đồng tiền cân đối và đồng chất. Nếu cả hai xuất hiện mặt sấp anh ta được 100 đồng, nếu cả hai xuất hiện mặt ngửa anh ta mất 40 đồng còn xuất hiện một sấp một ngửa anh ta mất 30 đồng. Gọi X là số tiền anh ta nhận được sau một ván chơi. Lập bảng phân phối xác suất của X

Nhận thấy X có thể nhận các giá trị - 40, - 30, 100 tương ứng với việc mất 40 đồng , mất 30 đồng và được 100 đồng.

Ta có $P(X = -40) = \frac{1}{4}, \quad P(X = -30) = \frac{1}{2}, \quad P(X = 100) = \frac{1}{4}$

Vậy bảng phân phối xác suất của X là

X	- 40	- 30	100
P	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

Ví dụ2 : Một phòng thí nghiệm được cấp ba triệu đồng để tiến hành thí nghiệm tìm một chủng vi rút trong gia cầm. Một lần thí nghiệm chi phí một triệu đồng. Nếu phát hiện ra chủng vi rút này thì ngừng thí nghiệm. Nếu không phát hiện ra thì làm thí nghiệm cho tới khi phát hiện ra chủng vi rút trên hoặc hết kinh phí thì dừng. Gọi Y là số tiền mà phòng thí nghiệm trên tiết kiệm được. Lập bảng phân phối xác suất của Y biết các thí nghiệm là độc lập với nhau và xác suất để tìm ra chủng vi rút ở mỗi lần thí nghiệm là 0,3.

Ta thấy Y có thể nhận một trong ba giá trị 0, 1, 2. Với xác suất tương ứng

$P(Y=0) = 0,7^2 = 0,49$; $P(Y = 1) = 0,7.0,3 = 0,21$; $P(Y = 2) = 0,3$.

Vậy bảng phân phối xác suất của Y là

Y	0	1	2
P	0,49	0,21	0,3

Ví dụ 3: Một người nhằm bắn vào một mục tiêu cho tới khi trúng đích thì dừng. Các lần bắn độc lập, xác suất trúng đích của mỗi lần bắn là p

($0 < p < 1$). Gọi Z là số đạn phải dùng. Lập bảng phân phối xác suất của Z

Nhận thấy Z có thể nhận các giá trị 1, 2, ..., n,...

$P(Z = n) = q^{n-1}p$ ($q = 1 - p$). Vậy bảng phân phối xác suất của Z là

Z	1	2	...	i	...	n	...
P	p	qp		$q^{i-1}p$		$q^{n-1}p$	...

4. Hàm phân phối xác suất

4.1. Định nghĩa: Hàm phân phối xác suất của biến ngẫu nhiên X là hàm số F(x) hoặc $F_X(x)$ cho bởi $F(x) = P(X<x)$ với mọi $x \in R$

Từ định nghĩa ta thấy mọi biến ngẫu nhiên đều có hàm phân phối xác suất. Nếu X là một biến ngẫu nhiên rời rạc thì

$F(x) = P(X = x_1) + ... + P(X = x_{i-1})$ với $x_{i-1} < x \le x_i$ và $F(x) = 0$ nếu $x \le x_1$

4.2. Các ví dụ

Ví dụ 1: Biến ngẫu nhiên X có bảng phân phối xác suất

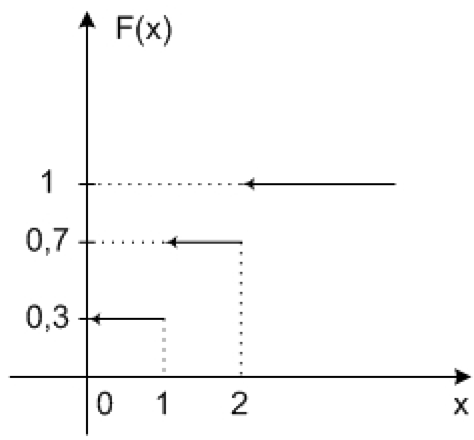
X	0	1	2
P	0,3	0,4	0,3

1/ Lập hàm phân phối xác suất của X

2/ Vẽ đồ thị của F(x)

Ta có $F(x) = \begin{cases} 0 & khi\ x \le 0 \\ 0,3 & khi\ 0 < x \le 1 \\ 0,7 & khi\ 1 < x \le 2 \\ 1 & khi\ x > 2 \end{cases}$

Đồ thị của F(x)



Hình 1

Ví dụ 2: Để thử sức chịu nén của một loại vật liệu người ta tiến hành theo ba mức sau:

Mức 1: Tiến hành thử với áp lực $200\,kG/cm^2$. Nếu vật liệu chịu được áp lực này thì chuyển sang mức hai

Mức 2: Tiến hành thử với áp lực $230\,kG/cm^2$. Nếu vật liệu chịu được áp lực này thì chuyển sang mức ba

Mức 3: Tiến hành thử với áp lực $250\,kG/cm^2$

Biết các lần thử độc lập với nhau và xác suất để loại vật liệu chịu được các mức thử trên tương ứng là 0,90; 0,60; 0,40. Gọi X là số lần thử. Y là số lần thử thành công. Hãy tìm hàm phân phối xác suất của X và của Y

Ta thấy: X có thể nhận các giá trị 1, 2, 3

Y có thể nhận các giá trị 0, 1, 2, 3.

$P(X = 1) = 0{,}1$; $P(X = 2) = 0{,}9.0{,}4 = 0{,}36$; $P(X = 3) = 0{,}9.0{,}6 = 0{,}54$

Vậy hàm phân phối xác suất của X là

$$F_X(x) = \begin{cases} 0 & \text{khi } x \le 1 \\ 0{,}1 & \text{khi } 1 \le x < 2 \\ 0{,}46 & \text{khi } 2 \le x < 3 \\ 1 & \text{khi } x > 3 \end{cases}$$

$P(Y = 0) = 0{,}1$; $P(Y = 1) = 0{,}9.0{,}4 = 0{,}36$; $P(Y = 2) = 0{,}9.0{,}6.0{,}6 = 0{,}324$

$P(Y = 3) = 0{,}9.0{,}6.0{,}4 = 0{,}216$

$$F_Y(x) = \begin{cases} 0 & khi \quad x \le 0 \\ 0,1 & khi \quad 0 \le x < 1 \\ 0,46 & khi \quad 1 \le x < 2 \\ 0,784 & khi \quad 2 \le x < 3 \\ 1 & khi \quad x > 3 \end{cases}$$

5. Các tính chất của hàm phân phối: Hàm phân phối F(x) của biến ngẫu nhiên X có các tính chất sau

Tính chất 5.1: $0 \le F(x) \le 1$

Tính chất này suy ra từ định nghĩa của hàm phân phối xác suất .

Tính chất 5.1: $\lim\limits_{x \to -\infty} F(x) = 0 \quad ; \quad \lim\limits_{x \to \infty} F(x) = 1$

Việc chứng minh tính chất này vượt ra khỏi kiến thức của giáo trình này. Tuy nhiên nếu đặt

$$\lim_{x \to -\infty} F(x) = F(-\infty) \; ; \; \lim_{x \to \infty} F(x) = F(\infty)$$

thì ta có thể hiểu khi $x \to -\infty$ sự kiện X < x trở thành sự kiện không thể có còn khi $x \to +\infty$ sự kiện X < x trở thành sự kiện tất yếu. Từ đó suy ra kết quả của tính chất 2

Tính chất 5.3: Hàm phân phối xác suất F(x) là hàm không giảm

Thật vậy: $\forall\, x_1, x_2 \in R,\ x_1 < x_2$

Xét $F(x_2) = P(X < x_2) = P[(X < x_1) \cup [x_1 \le X < x_2]]$

$= P(X < x_1) + P[x_1 \le X < x_2]] = F(x_1) + P[x_1 \le X < x_2] \ge F(x_1)$

Tính chất 5.4: $P(a \le X < b) = F(b) - F(a)$

Thay b và x_2, a vào x_1 trong chứng minh tính chất 3 ta có

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

Tính chất 5.5: Hàm phân phối xác suất là hàm liên tục trái. Ta công nhận tính chất này

Người ta cũng đã chứng minh được rằng nếu một hàm thực F(x) thoả mãn các tính chất 2, tính chất 3 và tính chất 5 thì nó cũng là hàm phân phối xác suất của một biến ngẫu nhiên X nào đó.

Từ các tính chất trên ta nhận thấy khi biết hàm phân phối xác suất của biến ngẫu nhiên X thì ta có thể tính xác suất để X nhận giá trị trong một khoảng bất kì. Vì vậy biết hàm phân phối xác suất của X cũng là biết được qui luật xác suất của X.

Ví dụ: Cho $F(x) = \begin{cases} 0 & \text{khi} \quad x \le 0 \\ 1 - \cos x & \text{khi} \quad 0 < x \le \frac{\pi}{2} \\ 1 & \text{khi} \quad x > \frac{\pi}{2} \end{cases}$

1/ Chứng minh rằng F(x) là hàm phân phối xác suất của biến ngẫu nhiên X nào đó

2/ Gọi X là biến ngẫu nhiên có hàm phân phối F(x).

Tính: $P(0 \le X < \frac{\pi}{4})$

Ta có: $\lim_{x \to -\infty} F(x) = 0$; $\lim_{x \to \infty} F(x) = 1$

Vậy hàm F(x) thoả mãn tính chất 2. Dễ thấy F(x) liên tục với mọi x $\in$ R, vậy F(x) thoả mãn tính chất 5

$\forall$ x $\le$ 0, F(x) = 0 $\Rightarrow$ F(x) không giảm

$\forall$ x $\in [0, \frac{\pi}{2}]$, F(x) = 1 - cosx là hàm tăng. Với $\forall$ x $> \frac{\pi}{2}$, F(x) = 1 cũng không giảm.Từ những kết quả nêu trên ta thấy F(x) không giảm trên R. Theo các yêu cầu để một hàm số với biến số thực trở thành hàm phân phối xác suất của một biến ngẫu nhiên nào đó, hàm F(x) thoả mãn các yêu câù này. Vậy nó là hàm phân phối xác suất của một biến ngẫu nhiên.Gọi X là biến ngẫu nhiên có hàm phân phối xác suất F(x) nêu trên, theo tính chất 4 ta có:

$$P(0 \le x < \frac{\pi}{4}) = F(\frac{\pi}{4}) - F(0) = 1 - \frac{\sqrt{2}}{2} - 0 = \frac{2-\sqrt{2}}{2}$$

6. Hàm mật độ xác suất

Định nghĩa: Nếu hàm phân phối F(x) của biến ngẫu nhiên X có đạo hàm $\forall$ $x \in R$ thì $F'(x) = f(x)$ được gọi là hàm mật độ xác suất của biến ngẫu nhiên X

Khi đó biến ngẫu nhiên X gọi là biến ngẫu nhiên liên tục tuyệt đối.

Từ định nghĩa trên ta thấy nếu X là biến ngẫu nhiên rời rạc thì hàm phân phối xác suất của X là hàm gián đoạn nên F(x) không khả vi (không có đạo hàm) tại những điểm gián đoạn. Vì vậy biến ngẫu nhiên rời rạc không có hàm mật độ xác suất.

7. Các tính chất

Tính chất 7.1: Hàm mật độ xác suất của biến ngẫu nhiên X là một hàm không âm

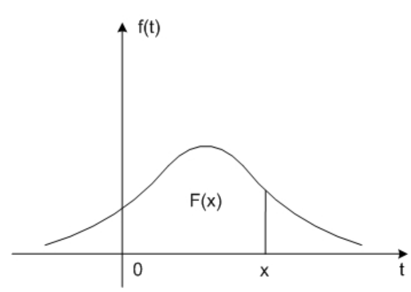
Thật vậy vì F(x) là hàm không giảm nên $f(x) = F'(x) \ge 0$

Tính chất 7.2: Hàm phân phối xác suất $F(x) = \int_{-\infty}^{x} f(t)dt$ ở đó f(x) là hàm mật độ xác suất của X

Thật vậy do F(x) là một nguyên hàm của f(x) nên:

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

Tính chất trên được minh hoạ bởi hình sau:



Hình 2

Tính chất 7.3: $P(a \le X < a) = \int_a^b f(x)dx$

Thật vậy do F(x) là một nguyên hàm của f(x) nên

$$\int_a^b f(x)dx = F(x)\Big|_a^b = F(b) - F(a) = P(a \le X < b)$$

Tính chất 7.4: Nếu X có hàm mật độ f(x) thì P(X = a) = 0 $\forall$ x $\in$ R

Vì P(X =a) = $\lim\limits_{b \to a^+} P(a \le X < b) = \lim\limits_{b \to a^+} [F(b) - F(a) = 0$

Do F(x) khả vi nên F(x) liên tục

Từ hai tính chất trên ta có hệ quả sau: Nếu biến ngẫu nhiên X có hàm mật độ xác suất f(x) thì

P[a≤ X < b] = P[a <X ≤ b]= P[a< X < b]= P[a≤ X ≤ b] = $\int_a^b f(x)dx$

Tính chất 7.5: $\int_{-\infty}^{+\infty} f(x)dx = 1$ Vì:

$$\int_{-\infty}^{+\infty} f(x)dx = F(x)\Big|_{-\infty}^{+\infty} = F(+\infty) - F(-\infty) = 1$$

Từ tính chất này ta thấy diện tích của hình giới hạn bởi hàm mật độ xác suất của và trục Ox bằng 1.

Người ta cũng đã chứng minh được rằng nếu một hàm thực f(x) không âm và $\int_{-\infty}^{+\infty} f(x)dx = 1$ thì nó cũng là hàm mật độ xác suất của một biến ngẫu nhiên X nào đó

8. Các ví dụ:

Ví dụ 1: Cho $f(x) = \begin{cases} 0 & \text{nếu } x \notin [0,\pi] \\ A\sin x & \text{nếu } x \in [0,\pi] \end{cases}$

1/ Tìm A để f(x) là hàm mật độ xác suất

2/ Tìm hàm phân phối xác suất tương ứng

Để f(x) trở thành hàm mật độ xác suất thì f(x) ≥ 0 và $\int_{-\infty}^{+\infty} f(x)dx = 1$.

Từ f(x) ≥ 0 suy ra A ≥ 0.

Xét $\int_{-\infty}^{+\infty} f(x)dx = 1 \Leftrightarrow \int_0^{\pi} A\sin x dx = 1 \Leftrightarrow -A\cos x\,|_0^{\pi} = 1 \Leftrightarrow 2A = 1 \Leftrightarrow A = \frac{1}{2}$

Có F(x) = $\int_{-\infty}^{x} f(t)dt = \begin{cases} 0 & \text{nếu } x \le 0 \\ \frac{1}{2} - \frac{1}{2}\cos x & \text{nếu } 0 < x \le \pi \\ 1 & \text{nếu } x > \pi \end{cases}$

Ví dụ 2: Cho $f(x) = \frac{A}{1+x^2}$

1/ Tìm A để f(x) là hàm mật độ xác suất của biến ngẫu nhiên X nào đó

2/ Tìm hàm phân phối xác suất tương ứng

3/ Tính P[0<X<1]

Để f(x) là hàm mật độ xác suất $\Rightarrow A \ge 0$ và

$$\int_{-\infty}^{+\infty} \frac{A}{1+x^2}dx = 1 \Leftrightarrow A\text{arctg}x \,|_{-\infty}^{+\infty} = 1 \Leftrightarrow A\pi = 1 \Leftrightarrow A = \frac{1}{\pi}$$

F(x) = $\int_{-\infty}^{x} \frac{1}{\pi}.\frac{dt}{1+t^2} = \frac{1}{\pi}\text{arctg}t \,|_{-\infty}^{x} = \frac{1}{\pi}(\text{arctg}x + \frac{\pi}{2})$

Xác suất: P[0 <X <1] = $\frac{1}{\pi}\text{arctg}1 = \frac{1}{4}$

II. Các số đặc trưng

Khi biết bảng phân phối xác suất hay hàm phân phối xác suất đối với biến ngẫu nhiên rời rạc, biết hàm phân phối xác suất hay hàm mật độ xác suất đối với biến ngẫu nhiênliên tục là hoàn toàn xác định được qui luật xác suất của biến ngẫu nhiên. Tuy nhiên, trong thực tế,để giải quyết một vấn đề nào đó nhiều khi không cần phải biết một trong các loại hàm nêu trên mà chỉ cần biết một số giá trị đặc trưng tương ứng với biến ngẫu nhiên đang xét. Các giá trị đặc trưng này được chia thành hai nhóm một nhóm đặc trưng cho vị trí và một nhóm đặc trưng cho mức phân tán của biến ngẫu nhiên.

1. Kì vọng

1.1.Các định nghĩa

Nếu biến ngẫu nhiên X có bảng phân phối xác suất

X	x_1 x_2 ... x_i ....x_n
P	p_1 p_2 ... p_i .... p_n

thì kì vọng toán (hoặc vọng số) của X là số kí hiệu là M(X) hay E(X) cho bởi:

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

Nếu biến ngẫu nhiên X nhận vô hạn đếm được các giá trị có bảng phân phối xác suất

X	x_1 x_2 ... x_i ... x_n ...
P	p_1 p_2 ... p_i ... p_n ...

và nếu $\sum_{n=1}^{\infty} p_n \mid x_n \mid$ hội tụ thì kì vọng toán của X là số M(X) hoặc E(X) cho bởi

$$E(X) = \sum_{n=1}^{\infty} p_n x_n$$

Nếu biến ngẫu nhiên X có hàm mật độ xác suất f(x) và nếu $\int_{-\infty}^{+\infty} |x| f(x)dx$ hội tụ thì kì vọng toán của X là số E(X) = $\int_{-\infty}^{+\infty} x f(x)dx$

Từ các định nghĩa trên ta nhận thấy:

* Định nghĩa chỉ ra cách tính kì vọng toán của biến ngẫu nhiên .

*Các biến ngẫu nhiên rời rạc nhận một số hữu hạn các giá trị luôn có kì vọng toán.

* Các biến ngẫu nhiên rời rạc nhận một số vô hạn đếm được hoặc không đếm được các giá trị có thể không có giá trị kì vọng .

* Kì vọng của biến ngẫu nhiên X là giá trị đặc trưng cho vị trí (trọng tâm hoặc trung tâm) của biến ngẫu nhiên .

* Kì vọng còn đưọc gọi là trung bình số học của biến ngẫu nhiên.

1.2 Các ví dụ

Ví dụ 1: Một nhóm 10 người trong đó ba người cao 1,62 m, hai người cao 1,66m, hai người cao 1,70m và ba người cao 1,74m. Chọn ngẫu nhiên một người trong nhóm người trên. Gọi X là chiều cao của người được chọn. Tính E(X)

Ta có bảng phân phối xác suất của X

X	1,62	1,66	1,70	1,74
P	0,3	0,2	0,2	0,3

Vậy E(X) = 0,3.1,62 + 0,2.1,66 + 0,2.1,70 + 0,3.1,74 = 1,68m

Ví dụ 2: Sau một năm bán hàng, một cửa hàng kinh doanh hoa tươi tại Hà nội nhận thấy số lẵng hoa X bán ra trong ngày theo tỉ lệ (xác suất) sau:

X	9	10	11	12	13	14	15
P	0,05	0,10	0,15	0,25	0,20	0,15	0,10

Một lẵng hoa tươi mua vào 60.000 đồng bán ra 100000 đồng, nếu trong ngày bán không hết số còn lại bị vứt bỏ. Số lẵng hoa cần mua vào là bao nhiêu để lợi nhuận trung bình thu được là cao nhất?

Để thực hiên bài toán trên ta lập bảng sau:

Hàng đầu của bảng ghi số lẵng hoa Y dự định mua vào trong ngày.

Cột đầu của bảng ghi số lẵng hoa X có thể bán ra trong ngày.

Cột cuối ghi xác suất bán được số lẵng hoa tương ứng.

Ô giao giữa dòng i và cột j là tiền lời (trăm nghìn) thu được khi mua vào j lẵng bán ra i lẵng.

Y \ X	9	10	11	12	13	14	15	P
9	360	300	240	180	120	60	0	0,05
10	360	400	340	280	220	160	100	0,10
11	360	400	440	380	320	260	200	0,15
12	360	400	440	480	420	360	300	0,25
13	360	400	440	480	520	460	400	0,20
14	360	400	440	480	520	560	500	0,15
15	360	400	440	480	520	560	600	0,10

Việc quyết định mua các lẵng hoa hàng ngày có thể thực hiện theo các phương án sau:

Phương án 1: Mua vào 9 lẵng, tiền lời trung bình là:

$E_1 = 360$

Phương án 2: Mua vào 10 lẵng, tiền lời trung bình là:

$E_2 = 300.\ 0{,}05 + 400.\ 0{,}95 = 395$

Phương án 3: Mua vào 11 lẵng, tiền lời trung bình là:

$E_3 = 240.\ 0{,}05 + 340.\ 0{,}10 + 440.\ 0{,}85 = 420$

Phương án 4: Mua vào 12 lẵng, lợi nhuận là:

$E_4 = 180.\ 0{,}05 + 280.\ 0{,}1 + 380.\ 0{,}15 + 480.\ 0{,}70 = 430$

Phương án 5:Mua vào 13 lẵng, lợi nhuận là:

$E_5 = 120.\ 0{,}05 + 220.\ 0{,}1 + 320.\ 0{,}15 + 420.\ 0{,}25 + 520.\ 0{,}45 = 415$

Phương án 6: Mua vào 14 lẵng, lợi nhuận trung bình là:

$E_6 = 60.\ 0{,}05 + 160.\ 0{,}1 + 260.\ 0{,}15 + 360.\ 0{,}25 + 460.\ 0{,}20 + 560.\ 0{,}25 = 380$

Phương án 7: Mua vào 15 lẵng, lợi nhuận trung bình là:

$E_7 = 0.\ 0{,}05 + 100.\ 0{,}1 + 200.\ 0{,}15 + 300.\ 0{,}25 + 400.\ 0{,}2 + 500.\ 0{,}15 + 600.\ 0{,}1 = 330$

Từ các kết quả trên ta thấy khi cửa hàng mua vào 12 lãng hoa thì lợi nhuận trung bình là cao nhất.

Ví dụ 3: Một xạ thủ nhằm bắn vào một mục tiêu cho tới khi trúng mục tiêu thì dừng. Các lần bắn độc lập, xác suất trúng mục tiêu của mỗi lần bắn là 0,8. Gọi X là lượng đạn phải dùng. Tính kì vọng của X

X có bảng phân phối xác suất sau

X	1	2	...	n	...
P	0,8	0,2.0,8	...	$0{,}2^{n-1}0{,}8$	...

Do X chỉ nhận các giá trị nguyên dương nên nếu chuỗi $\sum_{n=1}^{\infty} n.0{,}2^{n-1}.0{,}8$ hội tụ thì giá trị đó chính là kì vọng toán của X

Xét $f(x) = \sum_{n=1}^{\infty} x^n$ với $\forall$ x $\in (0,1)$. Do chuỗi hội tụ đều trong miền đang xét nên

$$f'(x)=\sum_{n=1}^{\infty} nx^{n-1}$$

Mặt khác $\sum_{n=1}^{\infty} x^n = \frac{x}{1-x} \Rightarrow \sum_{n=1}^{\infty} nx^{n-1} = (\frac{x}{1-x})' = \frac{1-x+x}{(1-x)^2} = \frac{1}{(1-x)^2}$

$$\Rightarrow \sum_{n=1}^{\infty} n.0{,}2^{n-1} = f'(0{,}2) = \frac{1}{0{,}8^2} \Rightarrow E(X) = \frac{0{,}8}{0{,}8^2} = 1{,}25$$

Ví dụ 4: Đại biến ngẫu nhiên X có hàm mật độ $f(x)=\frac{1}{\pi(1+x^2)}$

(phân phối Cauchy). Tính E(X)

Ta có: $\int_{-\infty}^{+\infty} |x| f(x)dx = \frac{2}{\pi}\int_{-\infty}^{+\infty} \frac{x}{1+x^2}dx$

Do tích phân này phân kì nên X không có kì vọng.

Ví dụ 5: Biến ngẫu nhiên X có hàm mật độ

$$f(x)=\begin{cases} 0 & \text{nếu } x \notin [0,\pi] \\ \frac{1}{2}\sin x & \text{nếu } x \in [0,\pi] \end{cases}$$

Ta có E(X) = $\int_{-\infty}^{+\infty} xf(x)dx = \frac{1}{2}\int_0^{\pi} x\sin x dx = \frac{\pi}{2}$

1.3 Tính chất

1/ Kì vọng của hằng số bằng chính nó

Thật vậy ta có thể coi hằng số là biến ngẫu nhiên chỉ nhận giá trị C với xác suất bằng 1 nên E(C) = 1.C = C

2/ Hằng số có thể đưa ra ngoài dấu kì vọng

Xét: Y= kX, nếu X là biến ngẫu nhiên rời rạc với $P(X = x_i) = p_i$ thì

$P(Y = kx_i) = p_i$. Vậy E(Y) = $\sum p_i(kx_i) = k\sum p_i x_i = kE(X)$

3/ Kì vọng của một tổng bằng tổng các kì vọng

Ta chứng minh trong trường hợp X, Y là các biến ngẫu nhiên rời rạc.

Gọi: Z = X + Y với $z_{ij} = x_i + y_j$, $P(Z = z_{ij}) = p_{ij}$

Ta có:

$$E(Z) = \sum p_{ij}z_{ij} = \sum p_{ij}(x_i + y_j) = \sum p_{ij}x_i + \sum p_{ij}y_j = \sum x_i \sum p_{ij} + \sum y_j \sum p_{ij}$$

$$= \sum p_{i\bullet}x_i + \sum p_{\bullet j}y_j = E(X) + E(Y)$$

$$p_{i\bullet} = P(X = x_i) = \sum_{j=1} p_{ij} \quad ; \quad p_{\bullet j} = P(Y = y_j) = \sum_{i=1} p_{ij}$$

Hệ quả 1: E(aX + b) = aE(X) + b

Hệ quả 2: $E(\sum_{i=1}^{n} a_i X_i) = \sum_{i=1}^{n} a_i E(X_i)$

Áp dụng nhiều lần tính chất 2 và tính chất 3 ta có hai hệ quả trên

4/ Nếu X độc lập với Y thì E(XY) = E(X). E(Y)

Đặt: $Z = XY$, $z_{ij} = x_i y_j$. Giả sử $P(X = x_i) = p_{i\bullet}$, $P(Y = y_j) = p_{\bullet j}$

$$\Rightarrow P(Z = z_{ij}) = p_{i\bullet}.p_{\bullet j} \Rightarrow E(Z) = \sum p_{i\bullet}.p_{\bullet j} z_{ij} = \sum\sum p_{i\bullet}.x_i.p_{\bullet j} y_j$$

$$= \sum p_{i\bullet}.x_i \sum p_{\bullet j} y_j = E(X).E(Y)$$

5/ Nếu $Y = \varphi(X)$ với φ là một hàm số xác định nào đó. Nếu X là biến ngẫu nhiên rời rạc với $P(X = x_i) = p_i$ thì $P[Y = \varphi(x_i)] = p_i$.

Vậy $E(Y) = \sum p_i \varphi(x_i)$

Nếu X có hàm mật độ f(x) thì Y có hàm mật độ f(x) vậy

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

2. Phương sai

Xét biến ngẫu nhiên X có kì vọng E(X)

2.1 . *Định nghĩa*: Phương sai của biến ngẫu nhiên X là số kí hiệu là D(X) hoặc VarX và $D(X) = E[X-E(X)]^2$

Từ định nghĩa trên ta thấy:

* Nếu X có phương sai thì X phải có kì vọng

* Phương sai còn được gọi là độ lệch bình phương trung bình của X đối với kì vọng của nó.

* Phương sai càng nhỏ thì X càng tập trung xung quanh kì vọng E(X)

Vậy phương sai là đại lượng đặc trưng cho mức độ phân tán của biến ngẫu nhiênquanh giá trị trung bình lí thuyết của nó.

Xét: $D(X) = E[X-E(X)]^2 = E\{X^2-2XE(X) + [E(X)]^2\}$

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

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

2.2 Các ví dụ

Ví dụ 1: Biến ngẫu nhiên X có bảng phân phối xác suất

X	0	1	2
P	0,3	0,4	0,3

Tính D(X).

Có: $D(X) = E(X^2) - E^2(X)$

$E(X) = 0.0,3 + 1.0,4 + 2.0,3 = 1$

$E(X^2) = 0.0,3 + 1^2.0,4 + 2^2.0,3 = 1,6$. Vậy $D(X) = 1,6 - 1^2 = 0,6$

Ví dụ 2: Một xạ thủ nhằm bắn vào một mục tiêu cho tới khi trúng đích thì dừng. Các lần bắn là độc lập, xác suất trúng đích ở mỗi lần bắn là 0,8. Gọi X là lượng đạn cần dùng.Tính D(X)

Nhận thấy X là biến ngẫu nhiên nhận các giá trị tự nhiên dương với $P(X = k) = 0,2^{k-1}.0,8$

$$E(X) = \sum_{k=1}^{\infty} k.0,2^{k-1}.0,8 = 0,8\sum_{k=1}^{\infty} k.0,2^{k-1}$$

Xét chuỗi: $\sum_{k=1}^{\infty} x^k = f(x)$ với $x \in (0,1) \Rightarrow f'(x) = \sum_{k=1}^{\infty} kx^{k-1}$

Mặt khác $f(x) = \dfrac{x}{1-x} \Rightarrow f'(x) = \dfrac{1-x+x}{(1-x)^2} = \dfrac{1}{(1-x)^2} \quad \Rightarrow \sum_{k=1}^{\infty} k.0,2^{k-1} = f'(0,2) = \dfrac{1}{0,8^2}$

$$\Rightarrow E(X) = \frac{1}{0,8} = 1,25$$

$$E(X^2) = \sum_{k=1}^{\infty} k^2.0,2^{k-1}.0,8 = 0,8\sum_{k=1}^{\infty} k^2.0,2^{k-1}$$

$$= 0,8\sum_{k=1}^{\infty} k(k-1).0,2^{k-1} + 0,8\sum_{k=1}^{\infty} k.0,2^{k-1}$$

$$= 0,8.0,2\sum_{k=1}^{\infty} k(k-1).0,2^{k-2} + \frac{1}{0,8}$$

Từ $\sum_{k=1}^{\infty} x^k = f(x)$ với $x \in (0,1)$; $f'(x) = \sum_{k=1}^{\infty} kx^{k-1} \Rightarrow f''(x) = \sum_{k=1}^{\infty} k(k-1)x^{k-2}$

$$\Rightarrow f''(x) = \sum_{k=1}^{\infty} k(k-1)x^{k-2}$$

$$f'(x) = \frac{1}{(1-x)^2} \Rightarrow f''(x) = \frac{2}{(1-x)^3} \Rightarrow f''(0,2) = \frac{2}{0,8^3} = \sum k(k-1).0,2^{k-2}$$

$$\Rightarrow \quad E(X^2) = 0,8.0,2\frac{2}{0,8^3} + \frac{1}{0,8} = \frac{0,4}{0,8^2} + \frac{1}{0,8} = \frac{1,2}{0,8^2}$$

$$\Rightarrow D(X) = E(X^2) - E^2(X) = \frac{1,2}{0,8^2} - \frac{1}{0,8^2} = \frac{0,2}{0,8^2} \approx 0,3125$$

Ví dụ 3: Cho biến ngẫu nhiên X có hàm mật độ xác suất

$$f(x) = \begin{cases} 0 & \text{nếu } x \notin [0,\pi] \\ \dfrac{1}{2}\sin x & \text{nếu } x \in [0,\pi] \end{cases} \qquad \text{Tính } D(X).$$

Ta có: $E(X) = \int_{-\infty}^{+\infty} xf(x)dx = \dfrac{1}{2}\int_0^{\pi} x\sin x dx = \dfrac{\pi}{2}$

$$E(X^2) = \int_{-\infty}^{+\infty} x^2 f(x)dx = \frac{1}{2}\int_0^{\pi} x^2 \sin x dx = \frac{\pi^2}{2} - 1$$

Vậy: $D(X) = \frac{\pi^2}{2} - 1 - \frac{\pi^2}{4} = \frac{\pi^2}{4} - 1$

2.3 Các tính chất của phương sai

1/ Phương sai của hằng số bằng 0

Thật vậy: $D(C) = E(C^2) - E^2(C) = C^2 - C^2 = 0$

2/ Hằng số đưa ra ngoài dấu phương sai phải bình phương lên

Vì: $D(kX) = E(k^2X^2) - [E(kX)^2] = k^2E(X^2) - k^2E^2(X)$

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

3/ Nếu X độc lập với Y thì $D(X+Y) = D(X) + D(Y)$

Ta có: $D(X + Y) = E[(X+Y)^2] - [E(X +Y)]^2$

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

$= E(X^2) + 2E(XY) + E(Y^2) - E^2(X) - 2E(X)E(Y) - E^2(Y)$

$= E(X^2) + E(Y^2) - E^2(X) - E^2(Y) = D(X) +D(Y)$

Hệ quả 1: $D(aX +b) = a^2D(X)$

Hệ quả 2: Nếu $X_1, X_2, \ldots, X_n$ độc lập thì $D(\sum k_i X_i) = \sum k_i^2 D(X_i)$

Hai hệ quả trên suy trực tiếp từ ba tính chất vừa nêu. Minh hoạ cho sự hữu ích của hệ quả 2 ta xét ví dụ sau.

Ví dụ: Gieo đồng thời 10 con xúc xắc cân đối và đồng chất

Gọi X là tổng số chấm ở các mặt trên. Tính D(X)

Gọi X_i là số chấm ở mặt trên của con xúc xắc thứ i

$\Rightarrow X = \sum_{i=1}^{10} X_i$ Do các X_i là độc lập với nhau nên $D(X) = \sum_{i=1}^{10} D(X_i)$

$P(X_i=1) = P(X_i=2) = \ldots = P(X_i=6) = \frac{1}{6}$

$\Rightarrow E(X_i) = \frac{1}{6}(1 + 2 + \ldots + 6) = 3,5$

$E(X_i^2) = \frac{1}{6}(1 + 4 + \ldots + 36) = \frac{91}{6}; D(X_i) = \frac{91}{6} - (\frac{21}{6})^2 = \frac{546-441}{36} = \frac{105}{36} = \frac{35}{12}$

Vậy: $D(X) = \frac{1050}{36} = \frac{350}{12} = \frac{175}{6}$

3. Độ lệch chuẩn

Việc dùng phương sai để đo mức độ phân tán của biến ngẫu nhiên quanh giá trị kì vọng của nó sẽ trở nên không thích hợp đối với biến ngẫu nhiên có thứ nguyên (có đơn vị đo đi kèm) bởi nếu X có thứ nguyên bậc nhất thì phương sai D(X) lại có thứ nguyên bậc hai.

Để khắc phục nhược điểm này người ta đưa ra một giá trị cũng đặc trưng cho mức độ phân tán của X nhưng có cùng thứ nguyên với X đó là độ lệch chuẩn

3.1 .*Định nghĩa*: Độ lệch chuẩn của biến ngẫu nhiên X là số

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

* Độ lệch chuẩn cùng thứ nguyên với X nếu X là biến có thứ nguyên

* Biến ngẫu nhiên X có độ lệch chuẩn khi và chỉ khi X có phương sai

3.2. Tính chất

1/ $\sigma(X) \geq 0$

2/ $\sigma(kX) = |k|\sigma(X)$

3/ Nếu X độc lập với Y thì :

$$\sigma(X \pm Y) = \sqrt{D(X) + D(Y)} \leq \sigma(X) + \sigma(Y)$$

Việc chứng minh các tính chất trên khá dễ dàng nên dành cho người đọc

4. Một số giá trị đặc trưng khác

Ngoài các giá trị đặc trưng là kì vọng, phương sai hoặc độ lệch chuẩn người ta còn đưa ra một số giá trị đặc trưng khác

4.1 Mômen: Mômen cấp k của biến ngẫu nhiên X đối với a là số

$$\mu_k(a) = E(X - a)^k$$

* Nếu a = 0 thì $\mu_k(0)$ gọi là mô men gốc cấp k của X

* Nếu a = E(X) thì $\mu_k(a)$ gọi là mô men trung tâm cấp k của X và kí hiệu là μ_k

Ta có $\mu(0) = E(X)$, $\mu_2 = D(X)$, $\mu_1 = 0$ với mọi biến ngẫu nhiên

Mặt khác khi X có phân phối đối xứng qua kì vọng E(X) thì $\mu_k = 0$ với k là số tự nhiên lẻ. Vì vậy ta có thể sử dụng μ_3 để xét xem phân phối xác suất của X có đối xứng hay không. Tuy nhiên, nếu X là đại lượng có thứ nguyên thì μ_3 có thứ nguyên bậc ba so với X, vì vậy người ta dùng $H_3 = \dfrac{\mu_3}{\sigma^3}$ làm số đo cho tính chất đối xứng hay không đối xứng của X và H_3 gọi là hệ số bất đối xứng của X. Người ta dùng hệ số $H_4 = \dfrac{\mu_4}{\sigma^4} - 3$ làm hệ số nhọn của phân phối xác suất của X, σ^2 là phương sai của X.

4.2 Mode: Nếu X là biến ngẫu nhiên rời rạc thì mode của đại lượng ngẫu nhiên X là số kí hiệu là ModX thoả mãn $P(X = modX) \geq P(X = x_i)\ \forall\ x_i$

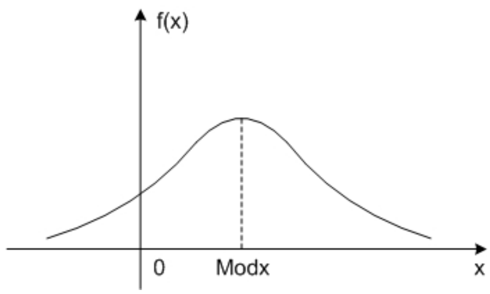
Nếu X là biến ngẫu nhiên liên tục thì ModX là điểm mà tại đó hàm mật độ f(x) đạt giá trị cực đại.

Ví dụ 1: Cho X có bảng phân phối xác suất

X	1	2	3	4	5
P	0,3	0,2	0,1	0,3	0,1

ModX = 1 hoặc ModX = 4. X có 2 Mod

Ví dụ 2: Cho X có hàm mật độ f(x) cho bởi đồ thị sau:



Hình 3

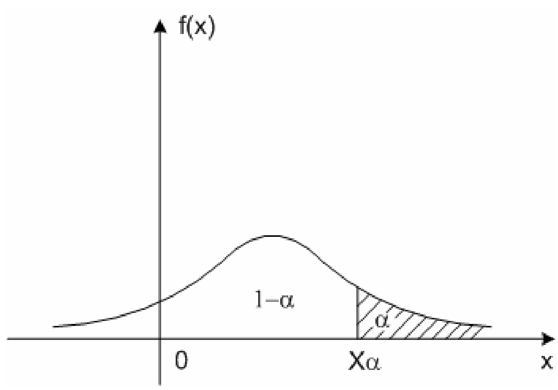
ModX là hoành độ điểm cực đại của hàm mật độ f(x)

4.3. Phân vị:

Phân vị mức 1 - α của biến ngẫu nhiên X là số X_α xác định bởi:

$$P(X < X_\alpha) \leq 1 - \alpha \leq P(X \leq X_\alpha)$$

Nếu α = 0,5 thì $X_{0,5}$ gọi là trung vị của X



Hình 4

III. Một số qui luật xác suất rời rạc thường gặp

1. Phân phối nhị thức

1.1 Định nghĩa

Biến ngẫu nhiên X được gọi là có phân phối nhị thức nếu X có bảng phân phối xác suất sau:

X	0	1 . . .	k . . .	n
P	$C_n^0 p^0 q^n$	$C_n^1 p\ q^{n-1}$	$C_n^k p^k q^{n-k}$	$C_n^n p^n q^0$

Với $0 < p < 1$, $q = 1 - p$

Khi X có phân phối nhị thức ta kí hiệu $X \sim B(n, p)$, n, p gọi là các tham số của phân phối

Phân phối nhị thức với n = 1 còn gọi là phân phối Bernoulli

Từ định nghĩa trên ta thấy: Số lần xuất hiện sự kiện A trong một lược đồ Bernulli có phân phối nhị thức.

1.2 Kì vọng và phương sai

$$X \sim B(n,p) \text{ ta có } E(X) = \sum_{k=0}^{n} k\, C_n^k p^k q^{n-k} = p\sum_{k=1}^{n} k\, C_n^k p^{k-1} q^{n-k}$$

Xét: $\left(\sum_{k=1}^{n} C_n^k x^k q^{n-k}\right)' = \sum_{k=1}^{n} k\, C_n^k x^{k-1} q^{n-k}$

Mặt khác $\sum_{k=0}^{n} C_n^k x^k q^{n-k} = (q+x)^n$ nên $\left(\sum_{k=0}^{n} C_n^k x^k q^{n-k}\right)' = n(q+x)^{n-1}$

$$\Rightarrow \sum_{k=1}^{n} k\, C_n^k x^{k-1} q^{n-k} = n(q+x)^{n-1}$$

Thay x = p ta có: $E(X) = p\sum_{k=1}^{n} k\, C_n^k p^{k-1} q^{n-k} = p.\, n(q+p)^{n-1} = np$

Phương sai: $D(X) = E(X^2) - E^2(X)$

$$E(X^2) = \sum_{k=0}^{n} k^2\, C_n^k p^k q^{n-k} = p^2\sum_{k=1}^{n} k\,(k-1)\, C_n^k p^{k-2} q^{n-k} + \sum_{k=0}^{n} k\, C_n^k p^k q^{n-k}$$

Xét: $(q+x)^n = \sum_{k=0}^{n} C_n^k x^k q^{n-k} \Rightarrow n(n-1)(q+x)^{n-2} = \sum_{k=1}^{n} k\,(k-1)\, C_n^k x^{k-2} q^{n-k}$

Thay x = p ta có: $n(n-1) = \sum_{k=1}^{n} k\,(k-1)\, C_n^k p^{k-2} q^{n-k}$.

Mặt khác $\sum_{k=0}^{n} k\, C_n^k p^k q^{n-k} = E(X) = np$. vậy ta có $E(X^2) = n(n-1)p^2 + np$

$$\Rightarrow D(X) = n^2p^2 - np^2 + np - (np)^2 = np(1-p) = npq$$

Ví dụ 1: Xác suất để một cây sống sau một thời gian trồng là 0,8. Trồng 1000 cây. Gọi X là số cây sống sau một thời gian trồng. Tính E(X), D(X).

Ta có X~ B(1000; 0,8) vậy E(X) = 1000.0,8 =800,

D(X) =1000.0,8.0,2 =160

2. Phân phối siêu bội

2.1 Định nghĩa: Xét một đám đông gồm N cá thể trong đó có M cá thể có đặc tính A. Chọn ngẫu nhiên n cá thể ($n \le M$, $n \le N - M$). Gọi X là số cá thể có đặc tính A trong n cá thể được chọn. X là biến ngẫu nhiên có phân phối siêu bội.

Nhận thấy X có thể nhận 1 trong các giá trị từ 0, 1, ... ,n

$P(X = k) = \frac{C_M^k C_{N-M}^{n-k}}{C_N^n}$ Như vậy quy luật siêu bội phụ thuộc vào ba tham số N, M, n.

2.2 Bảng phân phối xác suất của quy luật siêu bội, kì vọng, phương sai của phân phối siêu bội.

Từ công thức xác suất: P(X=không = $\frac{C_M^k C_{N-M}^{n-k}}{C_N^n}$ ta có bảng phân phối xác suất của phân phối siêu bội là:

X	0	1	. . .	k	. . .	n
P	$\frac{C_M^0 C_{N-M}^{n}}{C_N^n}$	$\frac{C_M^1 C_{N-M}^{n-1}}{C_N^n}$	...	$\frac{C_M^k C_{N-M}^{n-k}}{C_N^n}$	...	$\frac{C_M^n C_{N-M}^{0}}{C_N^n}$

Từ bảng phân phối các xác suất của phân phối siêu bội xét tỉ số của

P(X = k+1) và P(X = k) ta có:

Nếu $\frac{(N+1)(n+1)}{N+2}$ không phải là số nguyên thì ModX = $[\frac{(N+1)(n+1)}{N+2}]$

Còn nếu $\frac{(N+1)(n+1)}{N+2}$ = k_0 là số nguyên thì ModX là k_0 hoặc k_0 -1

Ấp dụng công thức tính kì vọng và phương sai của X ta có:

$$E(X) = \frac{nM}{N} \text{ và } D(X) = \frac{nM}{N}.\frac{N-M}{N}(1-\frac{n-1}{N-1})$$

Nhận xét 1: Nếu trong đám đông gồm N cá thể trong đó có M cá thể có đặc tính A, chọn ngẫu nhiên lần lượt n cá thể có hoàn lại. Gọi Y là số cá thể có đặc tính A trong n cá thể thì

Y~ B(n,p) với p = $\frac{M}{N}$ Theo công thức tính kì vọng và phương sai của phân phối nhị thức ta có:

$$E(Y) = np = \frac{nM}{N}, D(Y) = \frac{nM}{N}.\frac{N-M}{N}$$

Từ đây ta có E(X) =E(Y), D(X) < D(Y). Như vậy phân phối siêu bội và phân phối nhị thức (tương ứng với việc lấy mẫu không hoàn lại và có hoàn lại từ một đám đông sẽ đề cập tới trong phần thống kê) có cùng kì vọng. Tuy vậy phương sai của phân phối siêu bội nhỏ hơn phương sai của phân phối nhị thức. Sự khác biệt này càng cao nếu n càng lớn. Kết quả này phù hợp với trực quan là mẫu không lặp lại ít phân tán hơn mẫu có lặp lại. Tuy nhiên nếu số lượng các cá thể trong đám đông N là rất lớn so với n thì sự sai khác giữa phân phối siêu bội và phân phối nhị thức là không đáng kể. Điều này có nghĩa là nếu X và Y là phân phối siêu bội và phân phối nhị thức tương ứng nói ở trên khi N lớn so với n thì:

$P(X = k) \approx P(Y = k) \ \forall k = 0, 1 ..., n$

Nhận xét 2: Điều kiện $n \le N - M$ có thể bỏ qua, khi đó X có thể có thể không nhận một số giá trị đầu hoặc một số giá trị cuối còn công thức tính xác suất, kì vọng, phương sai của X vẫn tính như cũ.

2.1 Các ví dụ

Ví dụ 1: Từ một đàn gà gồm 10 con trong đó có 5 con mắc bệnh A. Chọn ngẫu nhiên ra 3 con. Gọi X là số gà mắc bệnh A trong 3 con gà được chọn, X là biến ngẫu nhiên có phân phối siêu bội

Ví dụ 2: Một tổ học sinh gồm 5 nam 4 nữ, chọn ngẫu nhiên 3 người, gọi Y là số nữ sinh trong 3 người được chọn. Y là biến ngẫu nhiên có phân phối siêu bội.

3. Phân phối hình học.

3.1 Định nghĩa: Biến ngẫu nhiên X có phân phối hình học với tham số thực p > 0 nếu X có bảng phân phối xác suất sau:

X	1	2	3 …………k……………n………………
P	p	p(1-p)	$p(1-p)^2$ $p(1-p)^{k-1}$ $p(1-p)^{n-1}$……….

3.2 Các số đặc trưng:

Đặt q = 1 – p => $E(X) = \sum_{k=1}^{\infty} k.q^{k-1}.p = p\sum_{k=1}^{\infty} k.p^{k-1}$

Xét chuỗi: $\sum_{k=1}^{\infty} x^k = f(x)$ với $x \in (0,1) \Rightarrow f'(x) = \sum_{k=1}^{\infty} kx^{k-1}$

Mặt khác $f(x) = \dfrac{x}{1-x} \Rightarrow f'(x) = \dfrac{1-x+x}{(1-x)^2} = \dfrac{1}{(1-x)^2} \quad \Rightarrow \sum_{k=1}^{\infty} k.q^{k-1} = f'(q) = \dfrac{1}{p^2}$

$$\Rightarrow E(X) = \frac{1}{p}$$

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

$$= p\sum_{k=1}^{\infty} k(k-1).q^{k-1} + p\sum_{k=1}^{\infty} k.q^{k-1}$$

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

Từ $\sum_{k=1}^{\infty} x^k = f(x)$ với $x \in (0,1)$; $f'(x) = \sum_{k=1}^{\infty} kx^{k-1} \Rightarrow f''(x) = \sum_{k=1}^{\infty} k(k-1)x^{k-2}$

$$\Rightarrow f''(x) = \sum_{k=1}^{\infty} k(k-1)x^{k-2}$$

$$f'(x) = \frac{1}{(1-x)^2} \Rightarrow f''(x) = \frac{2}{(1-x)^3} \Rightarrow f''(q) = \frac{2}{p^3} = \sum k(k-1).q^{k-2}$$

$$\Rightarrow \quad E(X^2) = pq\frac{2}{p^3} + \frac{1}{p} = \frac{2q}{p^2} + \frac{1}{p} = \frac{2q+p}{p^2}$$

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

Ví dụ 3: Tiến hành liên tiếp các thí nghiệm tìm nguyên tố vi lượng trong các mẫu đất cho tới khi thí nghiệm thànhcông thì dừng. Các thí nghiệm thực hiện độc lập, Xác suất thàng công ở mỗi thí nghiệm là 0,6. Gọi X là số lần phải thực hiện thí nghiệm.

1/ Chỉ ra qui luật xác suất của X.

2/ Tính E(X), D(X).

Từ giả thuyết suy ra: X nhận các giá trị tự nhiên dương, $P(X=n)=0{,}6.0{,}4^{n-1}$. Vẫy có phân phối hình học với tham số p = 0,6.

$$E(X) = \frac{1}{p} = \frac{1}{0{,}6} \approx 1{,}6666 \; ; \quad D(X) = \frac{q}{p^2} = \frac{0.4}{0.6^2} \approx 1{,}1111$$

4. Phân phối Poisson

4.1 Định nghĩa: Biến ngẫu nhiên X có phân phối Poisson với tham số thực $\lambda > 0$ nếu X có bảng phân phối xác suất sau:

X	0	1	2	...	n	...
P	$e^{-\lambda}$	$e^{-\lambda}\frac{\lambda}{1!}$	$e^{-\lambda}\frac{\lambda^2}{2!}$	...	$e^{-\lambda}\frac{\lambda^n}{n!}$	...

Ta kí hiệu X~ P_λ đọc là X có phân phối Poisson với tham số λ

4.2 Các số đặc trưng

Xét: $\frac{P(X=k+1)}{P(X=k)} = \frac{\lambda}{k} \geq 1 \Leftrightarrow \lambda \geq k$ Từ đây ta có

Nếu λ không là số nguyên thì ModX = [λ]

Nếu λ là số nguyên thì ModX = λ hoặc ModX = λ-1

$$E(X) = \sum_{n=1}^{\infty} ne^{-\lambda}.\frac{\lambda^n}{n!} = e^{-\lambda}.\lambda\sum_{n=1}^{\infty}\frac{\lambda^{n-1}}{(n-1)!} = e^{-\lambda}.\lambda e^{\lambda} = \lambda$$

Tương tự trên ta cũng có D(X) = λ. Như vậy phân phối Poisson rất đặc biệt là cả kì vọng và phương sai đều bằng tham số λ

Nhận xét: Gọi X(t) là số lần xảy ra sự kiện A trong quãng thời gian [0,t). Hiện tượng ngẫu nhiên trên được gọi là một quá trình Poisson nếu nó thoả mãn 3 yêu cầu sau:

* Nếu t < t' thì qui luật của biến X(t) - X(t') chỉ phụ thuộc vào $\Delta t = t - t'$ chứ không phụ thuộc vào t. Yêu cầu này được gọi là yêu cầu thuần nhất theo thời gian.

* Nếu t < $t' \leq t_1 < t_1'$ thì các phép thử tương ứng với việc khảo sát các hiện tượng trong quãng thời gian $\Delta t = t - t'$ và $\Delta t_1 = t_1 - t_1'$ là độc lập với nhau

* Nếu $P_1(\Delta t)$ và $P_2(\Delta t)$ là các xác suất tương ứng xảy ra một lần sự kiện A và ít nhất 2 lần sự kiện A trong quãng thời gian Δt thì

$$\lim_{\Delta t \to 0} \frac{P_1(\Delta t)}{\Delta t} = \lambda; \lim_{\Delta t \to 0} \frac{P_2(\Delta t)}{\Delta t} = 0$$

Việc nhận điện thoại tại một trạm điện thoại là một quá trình Poisson. Số lần các cuộc điện thoại gọi đến trạm điện thoại này là biến ngẫu nhiên có phân phối Poisson. Quá trình phân rã các nguyên tố phóng xạ cũng là một quá trình Poisson và số nguyên tử bị phân huỷ trong một quãng thời gian cũng là một biến ngẫu nhiên có phân phối Poisson.

5. Phân phối đa thức

5.1 Định nghĩa: Xét một dãy n phép thử độc lập , sau mỗi phép thử có thể có một trong k sự kiện $A_1, A_2, \ldots, A_k$ xảy ra. Xác suất $P(A_i) = p_i$

$P(A_i) = p_i$, $\sum_{i=1}^{k} p_i = 1$. Gọi X_i là số lần xuất hiện sự kiện A_i trong n lần thử. Luật phân phối của các biến $X_1, X_2, \ldots, X_k$ được gọi là luật đa thức.

Ta thấy $X_1 + X_2 + \ldots + X_k = n$, khác với các qui luật đã nêu ở trên qui luật đa thức là qui luật của k - 1 biến ngẫu nhiên. Biến thứ k phụ thuộc vào các biến còn lại bởi:

$X_k = n - (X_1 + X_2 + \ldots + X_{k-1})$

Nếu k = 2 luật đa thức trở thành luật nhị thức

5.2 Công thức xác suất của luật đa thức

Giả sử $X_1, X_2, \ldots, X_k$ là các biến ngẫu nhiên có luật đa thức

Ta đi tính xác suất $P(X_1 = n_1, X_2 = n_2, \ldots, X_k = n_k)$ với $\sum_{i=1}^{k} n_i = n$

Gọi A_{ij} là sự kiện A_i xuất hiện ở lần thử thứ j, $i = \overline{1,k}$, $j = \overline{1,n}$ nên một trong những kết quả của dãy n phép thử thoả mãn $X_i = n_i$ có xác suất xuất hiện là: $p_1^{n_1} p_2^{n_2} \ldots p_k^{n_k}$

Có tất cả $\frac{n!}{n_1! n_2! \ldots n_k!}$ .Vậy xác suất cần tính là:

$$P(X_1 = n_1, X_2 = n_2, \ldots, X_k = n_k) = \frac{n!}{n_1! n_2! \ldots n_k!} . p_1^{n_1} p_2^{n_2} \ldots p_k^{n_k}$$

IV Các phân phối liên tục

1. Phân phối đều

1.1. Định nghĩa: Biến ngẫu nhiên X có hàm mật độ

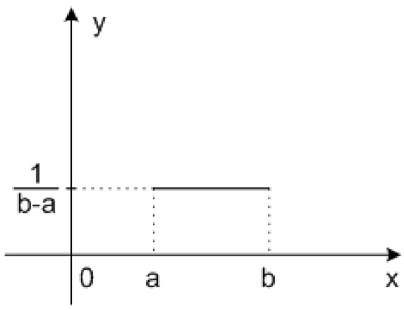
$$f(x) = \begin{cases} 0 & \text{nếu } x \notin [a,b] \\ \dfrac{1}{b-a} & \text{nếu } x \in [a,b] \end{cases}$$

được gọi là biến ngẫu nhiên có phân phối đều trên đoạn [a,b]

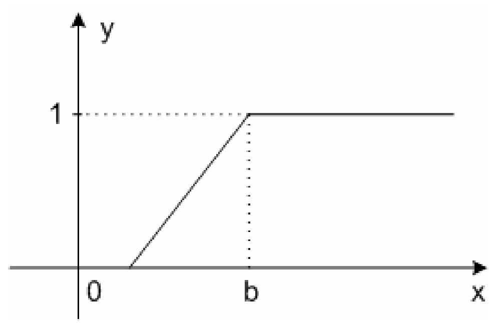
Nếu X có phân phối đều trên đoạn [a,b] thì hàm phân phối xác suất của X là

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

Đồ thị của hàm mật độ và hàm phân phối xác suất của phân phối đều trên [a, b] là các hình sau:



Hình 5



Hình 6

1.2. Kì vọng và phương sai của phân phối đều

Kì vọng : $E(X)=\int_{-\infty}^{+\infty}xf(x)dx=\int_a^b\frac{x}{b-a}dx=\frac{a+b}{2}$

Phương sai: $D(X)=E(X^2)-E^2(X)$

Có: $E(X^2)=\int_{-\infty}^{+\infty}x^2f(x)dx=\int_a^b\frac{x^2}{b-a}dx=\frac{1}{3}.(b^2+ab+a^2)$

Vậy $D(X)=\frac{1}{3}(b^2+ab+a^2)-\frac{(a+b)^2}{4}=\frac{(b-a)^2}{12}$

2. Phân phối mũ

2.1. Định nghĩa: Biến ngẫu nhiên có hàm mật độ

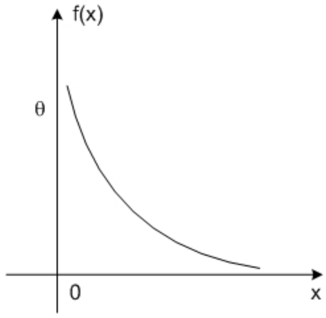
$$f(x)=\begin{cases}0 & \text{nếu } x<0\\ \theta e^{-\theta x} & \text{nếu } x\ge 0\end{cases}\qquad \theta>0$$

được gọi là biến ngẫu nhiên có phân phối mũ với tham số θ .

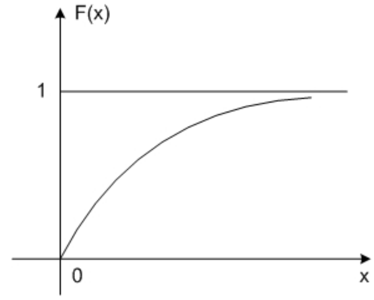
Hàm phân phối xác suất của phân phối mũ là

$$F(x)=\begin{cases}0 & \text{nếu } x<0\\ 1-e^{-\theta x} & \text{nếu } x\ge 0\end{cases}$$

Đồ thị của hàm mật độ xác suất và hàm phân phối xác suất của phân phối mũ là các hình sau:



Hình 7



Hình 8

2.2.Kì vọng và phương sai

Kì vọng: $E(X) = \int_{-\infty}^{+\infty} xf(x)dx = \int_{0}^{+\infty} xf(x)dx = \frac{1}{\theta}$

Tương tự ta có phương sai: $D(X) = \frac{1}{\theta^2}$

Phân phối mũ là phân phối đặc trưng cho thời gian làm việc của các linh kiện điện tử hoặc bán dẫn.

3.Phân phối chuẩn

Phân phối chuẩn là phân phối thường gặp trong tự nhiên cũng như trong kĩ thuật, nó đóng một vai trò quan trọng trong lí thuyết xác suất cũng như các ứng dụng của phân phối này trong xử lí số liệu.Chính vì vậy chúng ta sẽ đề cập đến phân phối chuẩn kĩ càng hơn.

3.1 Định nghĩa: Biến ngẫu nhiên có hàm mật độ

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

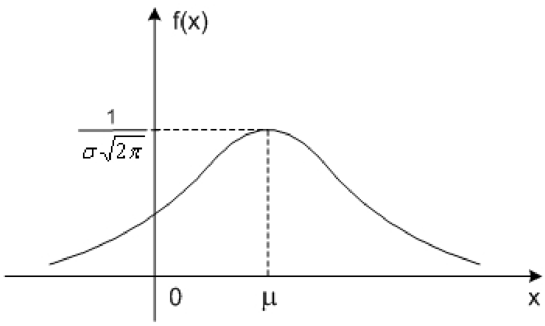
được gọi là biến ngẫu nhiên có phân phối chuẩn với 2 tham số μ và σ^2 và kí hiệu

$X \sim N(\mu;\sigma^2)$

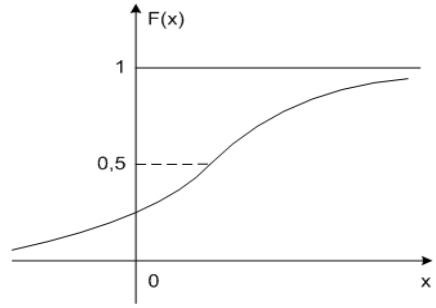
Hàm phân phối xác suất của phân phối chuẩn $F(x) = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt$

Do hàm mật độ của phân phối chuẩn không có nguyên hàm sơ cấp nên ta không thể biểu diễn hàm phân phối F(x) bởi một hàm số sơ cấp.

Đồ thị của hàm mật độ và hàm phân phối xác suất của phân phối chuẩn cho bởi các hình sau:



Hình 9



Hình 10

Đồ thị của hàm mật độ f(x) có dạng hình chuông nên phân phối chuẩn còn được gọi là phân phối hình chuông.

3.2. Kì vọng và phương sai :

Trong giáo trình giải tích ta đã biết $\int_{-\infty}^{+\infty} e^{-\frac{z^2}{2}}dt = \sqrt{2\pi}$. Đây là tích phân Euler - Poisson.

Bây giờ chúng ta sẽ sử dụng tích phân Euler - Poisson để tính kì vọng và phương sai của phân phối chuẩn

$E(X) = \int_{-\infty}^{+\infty} x.\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx$. Đặt $z = \frac{x-\mu}{\sigma} \Rightarrow dz = \frac{dx}{\sigma}$

Khi x $\to \pm\infty$ thì z $\to \pm\infty$, x = $\mu + \sigma z$. Thay vào công thức tính kì vọng ta có :

$$E(X) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}(\mu+\sigma z)e^{-\frac{z^2}{2}}dz = \frac{\mu}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}e^{-\frac{z^2}{2}}dz + \frac{\sigma}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}ze^{-\frac{z^2}{2}}dz$$

Do hàm trong dấu tích phân ở số hạng thứ hai là hàm lẻ vậy tích phân thứ hai bằng 0. Tích phân thứ nhất bằng $\sqrt{2\pi}$. Vậy E(X) = μ.

Xét: $E(X^2) = \int_{-\infty}^{+\infty} x^2.\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx$. Cũng đổi biến như trên ta có

$$E(X^2) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}(\mu+\sigma z)^2 e^{-\frac{z^2}{2}}dz$$

$$= \frac{\mu^2}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}e^{-\frac{z^2}{2}}dz + \frac{2\mu\sigma}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}ze^{-\frac{z^2}{2}}dz + \frac{\sigma^2}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}z^2e^{-\frac{z^2}{2}}dz$$

$$= \mu^2 + \frac{\sigma^2}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}z^2e^{-\frac{z^2}{2}}dz \quad (1)$$

Xét tích phân : $\int_{-\infty}^{+\infty} z^2e^{-\frac{z^2}{2}}dz = I$. Đặt $z = u \Rightarrow dz = du,\ ze^{-\frac{z^2}{2}}dz = dv \Rightarrow v = -e^{-\frac{z^2}{2}}$.

Ta có $I = \lim\limits_{a\to+\infty}(-ze^{-\frac{z^2}{2}})|_{-\infty}^{a} + \int\limits_{-\infty}^{+\infty} e^{-\frac{z^2}{2}}dz = 0 + \sqrt{2\pi}$ .Thay vào (1) ta có

$$E(X^2) = \mu^2 + \sigma^2 \Rightarrow D(X) = \sigma^2.$$

Như vậy hai tham số μ và σ^2 chính là kì vọng và phương sai của phân phối chuẩn. Tới đây ta có thể khẳng định phân phối chuẩn hoàn toàn xác định khi biết kì vọng và phương sai của nó.

3.3. Phân phối chuẩn tắc: Nếu biến ngẫu nhiên X có phân phối chuẩn với kì vọng $\mu = 0$ và phương sai $\sigma^2 = 1$ thì X được gọi là biến ngẫu nhiên có phân phối chuẩn tắc hoặc phân phối Gauss. Hàm mật độ của phân phối chuẩn tắc được kí hiệu là $\varphi(x)$ còn gọi là hàm Gauss, hàm phân phối được kí hiệu là $\phi(x)$ còn gọi là hàm Laplace. Giá trị của hàm $\varphi(x)$ và $\phi(x)$ được cho bởi bảng 1 và bảng 2 ở phần phụ lục cuối sách.

Hàm $\varphi(x)$ là hàm chẵn, $\varphi(-x) = \varphi(x)$. Trong khoảng $(0, +\infty)$, $\varphi(x)$ đơn điệu giảm

$$\varphi(0) = 0{,}3989;\ \varphi(1) = 0{,}2420;\ \varphi(2) = 0{,}0540;\ \varphi(3) = 0{,}0044;\ \varphi(4) = 0{,}0001$$

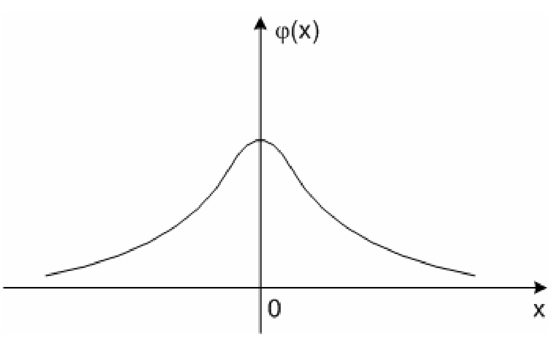
Nếu: $x \geq 4$ thì $\varphi(x) \approx 0$

Hàm $\phi(x) = \int\limits_{-\infty}^{x} \varphi(t)dt$, $\phi(-x) = 1 - \phi(x)$ với $x > 0$

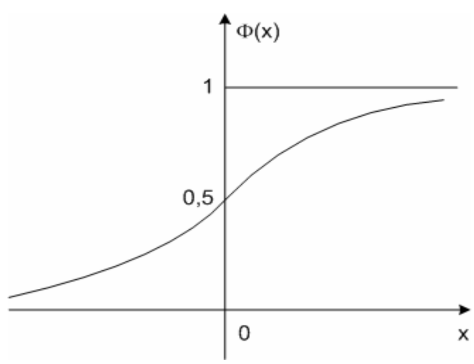
$$\phi(0) = 0{,}5;\ \phi(1) = 0{,}2420;\ \phi(2) = 0{,}0540;\ \phi(3) = 0{,}0044;\ \phi(3{,}9) = 0{,}0001$$

Nếu: $x \geq 4$ thì $\phi(x) \approx 1$ .Nếu $x < -4$ thì $\phi(x) \approx 0$.

Đồ thị của hàm $\varphi(x)$ và $\phi(x)$ là các hình sau:



Hình 11



Hình 12

3.4 Tính xác suất : Giả sử $X \sim N(\mu; \sigma^2)$.

$$P[a \leq X \leq b] = \int\limits_{a}^{b} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$. Bằng cách đổi biến:

$z = \dfrac{x-\mu}{\sigma}$ ta có $P[a \leq X \leq b] = \int\limits_{\frac{a-\mu}{\sigma}}^{\frac{b-\mu}{\sigma}} \varphi(z)dz = \phi\left(\dfrac{b-\mu}{\sigma}\right) - \phi\left(\dfrac{a-\mu}{\sigma}\right)$

Ví dụ: Cho X~ N(3 , 4). Tính P[2≤ X ≤ 4].

Áp dụng công thức trên ta có:

$$P[\,2 \le X \le 4] = \phi\left(\frac{4-3}{2}\right) + \phi\left(\frac{2-3}{2}\right) = 2\phi(0,5) - 1 = 0,3828.$$

3.5 Qui tắc 3σ:

Xét biến ngẫu nhiên X với kì vọng μ và phương sai σ^2.

$$P[\,|X-\mu| < \varepsilon\,] = P\left[\left|\frac{X-\mu}{\sigma}\right| < \frac{\varepsilon}{\sigma}\right] = 2\phi\left(\frac{\varepsilon}{\sigma}\right) - 1.$$

Với $\varepsilon = \sigma$ ta có: $P[\,|X-\mu| < \sigma\,] = 2\phi(1) - 1 = 0,6826$

Với $\varepsilon = 2\sigma$ ta có: $P[\,|X-\mu| < 2_{\sigma}\,] = 2\phi(2) - 1 = 0,9544$

Với $\varepsilon = 3\sigma$, $P[\,|X-\mu| < 3\sigma\,] = 2\phi(3) - 1 = 0,9973.$

Như vậy nếu X ~ N($\mu;\sigma^2$) thì $P[\,|X-\mu| < \varepsilon\,] = 1$ khi $\varepsilon > 3\sigma$. Điều này có nghĩa là nếu biến ngẫu nhiên X có phân phối chuẩn với kì vọng μ và phương sai σ^2 thì gần như chắc chắn rằng X sẽ nhận giá trị trong khoảng $[\mu - 3\sigma, \mu + 3\sigma]$.

Ví dụ: Trọng lượng của mỗi con gà trong một trại gà là biến ngẫu nhiên X có phân phối chuẩn với kì vọng μ = 3 kg và độ lệch chuẩn σ = 0,5 kg. Khi bắt ngẫu nhiên một con gà trong trại, gọi X là trọng lượng con gà con gà vừa bắt, ta có: $\mu - 3\sigma = 1,5$kg ; $\mu + 3\sigma = 4,5$kg. vậy ta có thể tin rằng con gà này có trọng lượng từ 1,5 kg đến 4,5kg.

V. Véc tơ ngẫu nhiên

1. Véc tơ ngẫu nhiên:

Cho U = (X_1, X_2, ... ,X_n) là một véc tơ thuộc không gian thực R^n. Nếu X_i là các biến ngẫu nhiên thì U gọi là một véc tơ ngẫu nhiên n chiều, các biến X_1, X_2, ... ,X_n là các thành phần ngẫu nhiên của véc tơ ngẫu nhiên U

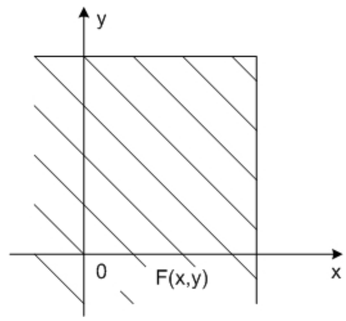
Ví dụ: Chọn ngẫu nhiên một người từ đám đông. Gọi X là chiều cao, Y là trọng lượng của người được chọn. Véc tơ U = (X, Y) là một véc tơ ngẫu nhiên hai chiều.

Để đơn giản trong giáo trình này ta chỉ đề cập đến véc tơ ngẫu nhiên hai chiều. Các kết quả về véc tơ ngẫu nhiên nhiều chiều có thể suy ra từ véc tơ ngẫu nhiên hai chiều. Nếu các thành phần X, Y của véc tơ ngẫu nhiên (X, Y) là các biến ngẫu nhiên rời rạc thì (X, Y) gọi là véc tơ ngẫu nhiên rời rạc. Nếu các thành phần đó là các biến ngẫu nhiên liên tục thì (X, Y) gọi là véc tơ ngẫu nhiên liên tục.

2. Qui luật xác suất:

2.1. Hàm phân phối: Cho véc tơ ngẫu nhiên (X, Y).

Hàm F(x,y) = P(X< x , Y< y) gọi là hàm phân phối xác suất đồng thời của véc tơ (X, Y)



Hình 13

Về mặt hình học F(x, y) là xác suất để điểm ngẫu nhiên M(X, Y) rơi vào miền gạch hình vẽ trên.

2.2. Tính chất

1/ $\lim\limits_{\substack{x\to+\infty \\ y\to+\infty}} F(x,y)=1 \qquad \lim\limits_{x\to-\infty} F(x,y)=\lim\limits_{y\to-\infty} F(x,y)=0$

2/ $\lim\limits_{y\to+\infty} F(x,y)=F_X(x)$ gọi là hàm phân phối xác suất của thành phần X

3/ $\lim\limits_{x\to+\infty} F(x,y)=F_Y(y)$ gọi là hàm phân phối xác suất của thành phần Y

4/ Hàm phân phối không giảm theo từng biến, nghĩa là

$F(x_1, y) \le F(x_2, y)$ với mọi $x_1 < x_2$

$F(x, y_1) \le F(x, y_2)$ với mọi $y_1 < y_2$

5/ $P(a \le X < b, c \le Y < d) = F(b, d) - F(a, d) - F(b, c) + F(a, c)$

Việc chứng minh các tính chất trên xin dành cho người đọc.

Từ các tính chất trên ta nhận thấy rằng: Khi biết hàm phân phối xác suất đồng thời của véc tơ ngẫu nhiên (X, Y) ta hoàn toàn xác định được qui luật xác suất của nó cũng như các qui luật của các thành phần tương ứng.

3 Bảng phân phối xác suất đồng thời của véc tơ ngẫu nhiên rời rạc:

Đối với véc tơ ngẫu nhiên rời rạc (X, Y) việc cho bảng phân phối xác suất đồng thời tỏ ra thuận lợi hơn việc cho hàm phân phối xác suất

3.1 Bảng phân phối xác suất đồng thời của véc tơ (X, Y) là bảng sau:

X \ Y	y_1		y_j		y_m
x_1	p_{11}		p_{1j}		p_{1m}
x_i	p_{i1}		p_{ij}		p_{im}
x_k	p_{k1}		p_{kj}		p_{km}

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

Ta có $\sum_{i=1}^{k}\sum_{j=1}^{m} p_{ij} = 1, \sum_{j=1}^{m} p_{ij} = p_{i\bullet} = P(X = x_i), \sum_{i=1}^{k} p_{ij} = p_{\bullet j} = P(X = y_j)$

3.2. Bảng phân phối xác suất của các thành phần

Bảng phân phối xác suất của các thành phần X, Y là các bảng sau:

X	x_1	x_2 ...	x_i ...	x_k
P	$p_{1.}$	$p_{2.}$ ...	$p_{i.}$ ...	$p_{k.}$

Y	y_1	y_2 ...	y_j ...	y_m
P	$p_{.1}$	$p_{.2}$ ...	$p_{.j}$ ...	$p_{.m}$

Ví dụ: Véc tơ ngẫu nhiên (X, Y) có bảng phân phối xác suất đồng thời sau:

X \ Y	0	1	2
0	0,1	0,2	0,3
1	0,1	0,1	0,2

Ta có: P(X = 0) = 0,6; P(X = 1) = 0,4

P(Y = 0) = 0,2; P(Y = 1) = 0,3; P(Y = 2) = 0,5

Vậy bảng phân phối xác suất của thành phần X, Y lần lượt là:

X	0	1
P	0,6	0,4

Y	1	2	3
P	0,2	0,3	0,5

4. Hàm mật độ xác suất: Cho véc tơ ngẫu nhiên (X, Y) có hàm phân phối xác suất F(x, y)

4.1. Định nghĩa: Nếu tồn tại hàm f(u, v) sao cho F(x, y) = $\int_{-\infty}^{x}\int_{-\infty}^{y} f(u,v)dudv$ với mọi (x, y) $\in R^2$ thì hàm f(x, y) gọi là hàm mật độ xác suất đồng thời của véc tơ ngẫu nhiên (X, Y)

Từ định nghĩa trên ta thấy: nếu (X, Y) là véc tơ ngẫu nhiên rời rạc thì (X, Y) không có hàm mật độ xác suất

4.2. Tính chất:

1/ $f(x, y) \geq 0$ với mọi $(x, y) \in R^2$

2/ $f(x, y) = \dfrac{\partial F(x,y)}{\partial y \partial x}$

3/ $\int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty} f(x,y)dxdy = 1$

4/ $P[\,(X, Y) \in D\,] = \iint\limits_{D} f(x,y)dxdy$

4.3. Hàm mật độ của các biến thành phần

Cho véc tơ ngẫu nhiên (X, Y) với hàm mật độ đồng thời f(x, y)

Hàm $f_X(x) = \int\limits_{-\infty}^{+\infty} f(x,y)dy$ gọi là hàm mật độ của thành phần X

Hàm $f_Y(y) = \int\limits_{-\infty}^{+\infty} f(x,y)dx$ gọi là hàm mật độ của thành phần Y

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

$$f(x, y) = \frac{1}{\pi^2(1+x^2)(1+y^2)}$$

1/ Tính $P(0 \leq X \leq 1, 0 \leq Y \leq 1)$

2/ Tìm $f_X(x)$ và $f_Y(y)$

Ta có : $P(0 \leq X \leq 1, 0 \leq Y \leq 1) =$

$$= \frac{1}{\pi^2}\int\limits_0^1\int\limits_0^1 \frac{1}{(1+x^2)(1+y^2)}dxdy = \frac{1}{\pi^2}\int\limits_0^1 \frac{dx}{1+x^2}\int\limits_0^1 \frac{dy}{1+y^2} = \frac{1}{16}$$

$$f_X(x) = \int\limits_{-\infty}^{+\infty} \frac{dy}{\pi^2(1+x^2)(1+y^2)} = \frac{1}{\pi(1+x^2)} \quad \text{tương tự} \quad f_Y(y) = \frac{1}{\pi(1+y^2)}$$

5. Một số đặc trưng: Đối với véc tơ ngẫu nhiên hai chiều (X, Y) ta thường quan tâm đến các đặc trưng sau:

5.1.Hiệp phương sa i(covariance hoặc mô men tương quan):

Hiệp phương sai của véc tơ (X, Y) là số: $cov(X, Y) = E\{[X - E(X)][Y - E(Y)]\}$

Dựa vào tính chất của kỳ vọng ta có:

$$cov(X, Y) = E(XY) - E(X)E(Y)$$

Vậy cov(X, X) = var(X), cov(Y,Y) = var(Y).

Nếu X độc lập với Y thì cov(X, Y) = 0

Điều ngược lại nói chung không đúng

5.2. Hệ số tương quan:

Hệ số tương quan giữa X, Y là số $\rho(X,Y) = \dfrac{cov(X,Y)}{\sqrt{D(X)D(Y)}}$

Hệ số tương quan có các tính chất sau:

1/ Nếu X độc lập với Y thì $\rho = 0$, điều ngược lại nói chung không đúng.

2/ $-1 \le \rho(X,Y) \le 1$.

3/ $\rho(X,Y) = \pm 1$ khi Y= AX + B.

4/ $\rho(aX+b, cY+d) = \begin{cases} \rho(X,Y) \ khi & ac > 0 \\ -\rho(X,Y) \ khi & ac < 0 \end{cases}$

Việc chứng minh các tính chất trên dành cho người đọc.

6. Phân phối có điều kiện.

Giả sử (X,Y) là vectơ ngẫu nhiên rời rạc, ta biết

$$P\left(X = x_i / Y = y_j\right) = \frac{P(X = x_i, Y = y_j)}{P(Y = y_j)} = \frac{p_{ij}}{p_{.j}}$$

Bảng phân phối xác suất

X	x_1	x_2	x_i	x_k
$P\left(X/_{Y=y_j}\right)$	$\frac{p_{1j}}{p.j}$	$\frac{p_{2j}}{p.j}$	$\frac{p_{ij}}{p.j}$	$\frac{p_{kj}}{p.j}$

được gọi là bảng phân phối xác suất có điều kiện của X với điều kiện $Y=y_j$.

Tương tự bảng phân phối xác suất

Y	y_1	y_2	y_j	y_m
$P\left(Y/_{X=x_i}\right)$	$\frac{p_{i1}}{p_{i.}}$	$\frac{p_{i2}}{p_{i.}}$	$\frac{p_{ij}}{p_{i.}}$	$\frac{p_{im}}{p_{i.}}$

được gọi là bảng phân phối xác suất có điều kiện của Y với điều kiện $X = x_i$.

$\sum_{i=1}^{k} \frac{p_{ij} x_i}{p_{.j}} = E(X/_{Y=y_j})$ được gọi là kỳ vọng có điều kiện của X biết $Y = y_j$

$\sum_{j=1}^{m} \frac{p_{ij} y_j}{p_{i.}} = E(Y/_{X=x_i})$ được gọi là kỳ vọng có điều kiện của Y biết $X = x_i$.

Nếu (X,Y) là vectơ ngẫu nhiên liên tục có hàm mật độ đồng thời f(x,y) thì hàm

$\varphi(x/y) = \frac{f(x,y)}{f_Y(y)}$ gọi là hàm mật độ có điều kiện của X với điều kiện Y = y.

Tương tự hàm:

$\varphi(y/x) = \frac{f(x,y)}{f_x(x)}$ gọi là hàm mật độ có điều kiện của Y với điều kiện X = x.

$\int_{-\infty}^{+\infty} x\varphi(x/y)dx = E(X/_{Y=y})$ là kì vọng có điều kiện của X với điều kiệnY = y

$\int_{-\infty}^{+\infty} y\psi(y/x)dy = E(Y/_{X=x})$ là kì vọng có điều kiện của Y với điều kiện X = x.

Ví dụ 1: Vectơ ngẫu nhiên (X,Y) có bảng phân phối xác suất đồng thời sau:

Y \ X	0	1	2
0	0,1	0,2	0,1
2	0,2	0,1	0,1
4	0,05	0,1	0,05

Phân phối xác suất của thành phần của X là :

X	0	1	2
P	0,35	0,40	0,25

Kì vọng: E(X) = 0,9

Phân phối xác suất của thành phần Y là:

Y	0	2	4
P	0,4	0,4	0,2

Kì vọng: E(Y) = 1,6

Phân phối xác suất của X với điều kiện Y = 1 là:

X	0	2	4
P(X/Y=1)	0,40	0,40	0,20

Kì vọng E(X/Y=1) =1,5.

Y	0	1	2
P(Y/X=2)	0,5	0,25	0,25

Kì vọng E(Y/X=2) = 0,75.

Ví dụ 2: Vectơ ngẫu nhiên (X,Y) có hàm mật độ xác suất đồng thời :

$$f(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2+\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2-2\rho\frac{(x-\mu_X)(y-\mu_y)}{\sigma_X\sigma_Y}\right]}$$

được gọi là có phân phối chuẩn hai chiều.

Nhận thấy:

$$f_X(x) = \frac{1}{\sigma_X{}^2\sqrt{2\pi}} e^{-\frac{(x-\mu_X)^2}{2\sigma^2{}_X}} \quad ; \quad f_Y(y) = \frac{1}{\sigma_Y{}^2\sqrt{2\pi}} e^{-\frac{(y-\mu_Y)^2}{2\sigma^2{}_Y}}$$

$$\varphi(x/y) = \frac{1}{2\pi\sigma_X\sqrt{1-\rho^2}} e^{-\frac{1}{2\sigma_X(1-\rho^2)}\left[x-\mu_X-\rho\frac{\sigma_X}{\sigma_Y}(y-\mu_Y)\right]^2}$$

$$\varphi(y/x) = \frac{1}{2\pi\sigma_Y\sqrt{1-\rho^2}} e^{-\frac{1}{2\sigma_Y(1-\rho^2)}\left[y-\mu_Y-\rho\frac{\sigma_Y}{\sigma_X}(x-\mu_X)\right]^2}$$

$$E(Y/_{X=x}) = \rho\frac{\sigma_X}{\sigma_Y}(x-\mu_X)+\mu_Y$$

$$E(X/_{Y=y}) = \rho\frac{\sigma_Y}{\sigma_X}(y-\mu_Y)+\mu_X$$

Với $E(X) = \mu_X$, $E(Y) = \mu_Y$, $D(X) = \sigma^2_X$, $D(Y) = \sigma^2_Y$, $\rho(X,Y) = \rho$

Từ những kết quả trên chúng ta thấy :

* Các thành phần của phân phối chuẩn hai chiều là phân phối chuẩn.

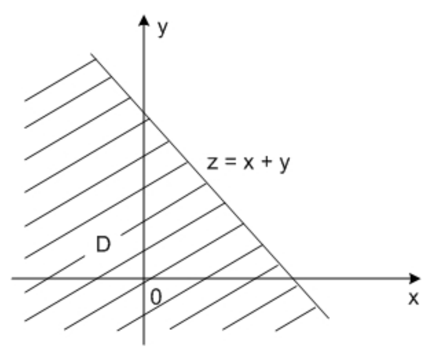
* Nếu $\rho = 0$ thì $f(x,y) = f_X(x)f_Y(y)$. Vậy trong phân phối chuẩn hai chiều $\rho = 0$ khi và chỉ khi X và Y độc lập.

7. Phân phối của tổng hai biến ngẫu nhiên .

Cho vectơ ngẫu nhiên (X,Y) có hàm mật độ phân phối đồng thời f(x, y).

Xét vectơ ngẫu nhiên Z = X + Y,chúng ta đi tìm qui luật phân phối xác suất của Z. Gọi $F_Z(z)$ là hàm phân phối xác suất của biến Z. Ta có:

$F_Z(z) = P[Z < z] = P[X + Y < z]$. Nếu coi M(X ,Y) là điểm ngẫu nhiên trong mặt phẳng Oxy thì P(X + Y< z) là xác suất để M(X , Y) rơi vào miền bên trái đường thẳng x + y = z



Hình 14

Vậy $F_Z(z) = P[M(X,Y) \in D] = \int_{-\infty}^{+\infty}[\int_{-\infty}^{z-x} f(x,y)dy]dx$

Lấy đạo hàm của tích phân theo biến z ta có :

$F'_Z(z) = f_z(z) = \int_{-\infty}^{+\infty} f(x, z-x)dx$ đây là công thưc xác định hàm mật độ xác suất của Z. Thay đổi vai trò của x và y trong công thức tích phân kép trên ta cũng có

$$F'_Z(z) = f_Z(z) = \int_{-\infty}^{+\infty} f(x, z-x)dx = \int_{-\infty}^{+\infty} f(z-y, y)dy$$

Trường hợp X độc lập với Y ta có

$$f_Z(z) = \int_{-\infty}^{+\infty} f_X(x) f_Y(z-x)dx = \int_{-\infty}^{+\infty} f_Y(y) f_X(z-y)dy$$

VI. Luật số lớn và các định lý giới hạn.

Việc trình bày chính xác các kết quả của phần này sẽ vượt qua khuôn khổ và yêu cầu của giáo trình. Tuy nhiên vì tầm quan trọng của nó, nhất là những ứng dụng trong thống kê chúng tôi sẽ cố gắng trình bày những kiến thức tối thiểu và đơn giản nhất.

1. Hội tụ của dãy biến ngẫu nhiên.

1.1. Hội tụ hầu chắc chắn: Dãy các biến ngẫu nhiên $\{X_n\}$ gọi là hội tụ hầu chắc chắn tới biến X nếu P[$\lim\limits_{n\to\infty} X_n \neq X$] = 0 khi đó ta kí hiệu:

$$X_n \xrightarrow{h.c.c} X.$$

1.2 Hội tụ theo xác suất: Dãy các biến ngẫu nhiên $\{X_n\}$ gọi là hội tụ theo xác suất tới biến X nếu $\forall \varepsilon > 0, \lim\limits_{n\to\infty} P[|X_n - X| \leq \varepsilon] = 1$. Khi đó ta ký hiệu:

$$X_n \xrightarrow{P} X.$$

1.3 Hội tụ theo quy luật : Dãy các biến ngẫu nhiên $\{X_n\}$ gọi là hội tụ theo quy luật tới biến X nếu dãy hàm phân phối $\{F_n(x)\}$ của dãy $\{X_n\}$ hội tụ tới hàm phân phối F(x) của X.

Người ta chứng minh được rằng nếu một dãy biến ngẫu nhiên hội tụ hầu chắc chắn tới X thì nó cũng hội tụ theo xác suất tới X. Một dãy biến ngẫu nhiênhội tụ theo xác suất tới X thì cũng hội tụ theo quy luật tới X.

2 Luật số lớn và các định lý về luật số lớn.

2.1 Sơ qua về luật số lớn: Dãy các biến *ngẫu* nhiên $\{X_n\}$ gọi là tuân theo luật số lớn nếu $\overline{X} - \overline{a} \xrightarrow{P} 0$; $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$; $\overline{a} = \frac{1}{n}\sum_{i=1}^{n} EX_i$.

Dãy các biến ngẫu nhiên $\{X_n\}$ gọi là tuân theo luật mạnh số lớn nếu :

$$\overline{X} - \overline{a} \xrightarrow{hcc} 0; \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \ ; \overline{a} = \frac{1}{n}\sum_{i=1}^{n} EX_i .$$

3.2 *Bất đẳng thức Chebyshev:* Cho biến ngẫu nhiên X có E(X) = a, D(X) = b^2.

Với mọi $\varepsilon > 0$ ta có P[| X- a | $< \varepsilon$] $\geq 1 - \frac{b^2}{\varepsilon^2}$.

Bất đẳng thức trên do nhà toán học người Nga Chebyshev đưa ra từ thế kỷ 19.
Ta chứng minh bất đẳng thức cho trường hợp X có hàm mật độ xác suất f(x).

$P[|X-a| < \varepsilon] = 1 - P[|X-a| \geq \varepsilon]$ (1)

$$P[|X-a| \geq \varepsilon] = \int_{|x-a|\geq\varepsilon} f(x)dx = \int_{(x-a)^2 \geq \varepsilon^2} f(x)dx$$

$$\leq \frac{1}{\varepsilon^2}\int_{(x-a)^2\geq\varepsilon^2}(x-a)^2 f(x)dx \leq \frac{1}{\varepsilon^2}\int_{-\infty}^{+\infty}(x-a)^2 f(x)dx = \frac{b^2}{\varepsilon^2} \quad (2)$$

Từ (1) và (2) suy ra $P[|X-a| < \varepsilon] \geq 1 - \frac{b^2}{\varepsilon^2}$, điều phải chứng minh

3.3 *Định lý Chebyshev:* Dãy các biến ngẫu nhiênđộc lập $\{X_n\}$ có

$E(X_n) = \mu$, $D(X_n) = \sigma^2$ tuân theo luật số lớn.

Đặt $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ ta có $E(\overline{X}) = \mu$, $D(\overline{X}) = \frac{\sigma^2}{n}$.

Áp dụng bất đẳng thức Chebyshev với biến $\overline{X}$ ta có:

$1 \geq P[|\overline{X} - \mu| < \varepsilon] \geq 1 - \frac{\sigma^2}{\varepsilon^2 n}$. Do $\lim_{n\to\infty}(1 - \frac{\sigma^2}{\varepsilon^2 n}) = 1$ nên

$\lim_{n\to\infty} P[|\overline{X} - \mu| < \varepsilon] = 1 \Rightarrow$ điều phải chứng minh.

Định lý Chebyshev là cơ sở toán học cho phép đo các đại lượng vật lý. Khi đo một đại lượng vật lý nào đó có độ đo chưa biết, người ta tiến hành đo nhiều lần được các giá trị $X_1, X_2, \ldots, X_n$ sau đó lấy $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ làm giá trị của đại lượng cần đo μ.

2.4 Định lý Bernoulli: Tiến hành một dãy n phép thử độc lập, xác suất xuất hiện sự kiện A ở mỗi phép thử là p. Gọi $f_n(A) = \frac{n_A}{n}$ là tần suất của sự kiện A.

Dãy $f_n(A)$ tuân theo luật số lớn.

Thật vậy : ta có $E(f_n(A)) = p$, $D(f_n(A)) = \frac{pq}{n}$ (q = 1- p).

áp dụng bất đẳng thức Chebyshev với đại lượng $f_n(A)$ ta có:

$1 \geq P[|f_n(A) - p| < \varepsilon] \geq 1 - \frac{pq}{\varepsilon^2 n}$. Từ đây suy ra định lý được chứng minh.

3. Các định lý giới hạn.

Các định lý về luật số lớn liên quan tới sự hội tụ theo xác suất của dãy các đại lượng ngẫu nhiên. Phần này chúng tôi trình bày không chứng minh các định lý về sự hội tụ theo quy luật của dãy các biến ngẫu nhiên.

3.1 .Định lý giới hạn điạ phương Moivre -Laplace.

Xét một lược đồ Bernoulli, xác suất xuất hiện sự kiện A ở mỗi phép thử là p. Xác suất để A xuất hiện đúng k lần trong n lần thử là $P_n(k)$.

Xác suất để A xuất hiện trong khoảng từ k_1 đến k_2 lần trong n lần thử là $P_n(k_1, k_2)$. Theo công thức Bernoulli ta có:

$$P_n(k) = C_n^k p^k q^{n-k}, \quad P_n(k_1, k_2) = \sum_{k=k_1}^{k_2} C_n^k p^k q^{n-k}.$$

Việc sử dụng công thức Bernoulli để tính các xác suất đòi hỏi phải thực hiện một số lượng khá lớn các phép toán nên khi n lớn ta sử dụng công thức xấp xỉ được suy ra từ các định lý sau.

Định lý 1: Định lý giới hạn địa phương Moivre -Laplace:

$$\lim_{n\to\infty}\left[P_n(k) - \frac{1}{\sqrt{npq}}\varphi(x_k)\right] = 0, \text{ ở đó } x_k = \frac{k-np}{\sqrt{npq}}.$$

Điều này có nghĩa là khi n khá lớn ta cũng có công thức xấp xỉ sau:

$$P_n(k) \approx \frac{1}{\sqrt{npq}}\varphi\left(\frac{k-np}{\sqrt{npq}}\right).$$

Định lý 2: Định lý giới hạn tích phân Moivre -Laplace:

$$\lim_{n\to\infty}\left[P_n(k_1, k_2) - \int_{x_{k_1}}^{x_{k_2}}\varphi(t)dt\right] = 0$$

ở đó $x_{k_1} = \dfrac{k_1 - np}{\sqrt{npq}}$, $x_{k2} = \dfrac{k_2 - np}{\sqrt{npq}}$, $\varphi(t) = \dfrac{1}{\sqrt{2\pi}}e^{-t^2/2}$.

Khi n khá lớn ta cũng có công thức xấp xỉ sau

$$P_n(k_1, k_2) \approx \phi\left(\frac{k_2-np}{\sqrt{npq}}\right) - \phi\left(\frac{k_1-np}{\sqrt{npq}}\right).$$

Ví dụ 1: Xác suất để mỗi cây sống sau thời gian trồng là 0,8. Trồng 400 cây

1/ Tính xác suất để có đúng 320 cây sống.

2/ Tính xác suất để số cây sống nằm trong khoảng từ 300 đến 360 cây.

Giải: Xác suất cần tính ở phần 1 là:

$$P_{400}(320) \approx \frac{1}{\sqrt{400.0,8.0,2}}\varphi\left(\frac{320-320}{\sqrt{400.0,8.0,2}}\right) = \frac{1}{8}\varphi(0) = 0,05.$$

Xác suất cần tính ở phần 2 là:

$$P_{400}(300, 360) = \phi(5,0) - \phi(-2,5) = \phi(2,5) = 0,99379$$

Ví dụ 2 : Tại một khu nghỉ mát có 2n người đến nghỉ. Khách được phục vụ ăn trưa bởi hai đợt liên tiếp từ 10h 30 đến 11h 30 và từ 11h 30 đến 12h 30 . Mỗi khách nghỉ có thể đến quán ăn của khu nghỉ mát vào một trong hai đợt với cùng một khả năng. Hãy tìm số chỗ ngồi tối thiểu tại quán ăn để với xác suất không nhỏ hơn 0,95 có thể tin rằng không có khách nào không được phục vụ ăn trưa.

Gọi X là số người đến quán ăn trong lần phục vụ đầu: Từ 10h 30 đến 11h 30 .

$X \sim B(n; 0{,}5)$. Gọi k_0 là số chỗ ngồi tại quán ăn, ta có:

$$P[\, X \le k_0,\ 2n - X \le k_0] \ge 0{,}95 \Leftrightarrow P\,[\, 2n - k_0 \le X \le k_0\,] \ge 0{,}95$$

$$\Leftrightarrow \phi\left(\frac{k_0 - n}{0{,}5\sqrt{2n}}\right) - \phi\left(\frac{n - k_0}{0{,}5\sqrt{2n}}\right) \ge 0{,}95 \Leftrightarrow 2\,\phi\left(\frac{k_0 - n}{0{,}5\sqrt{2n}}\right) - 1 \ge 0{,}95$$

$$\Leftrightarrow \phi\left(\frac{k_0 - n}{0{,}5\sqrt{2n}}\right) \ge 0{,}975 = \phi(1{,}96) \Leftrightarrow k \ge n + 0{,}98$$

Với $n = 50 \Rightarrow k_0 \ge 50 + 9{,}8 = 59{,}8$. Vậy nếu khu nghỉ có 100 khách thì phòng ăn cần bố trí 60 chỗ.

3.2 Định lý giới hạn Poisson:

Khi sử dụng công thức xấp xỉ nêu trên để tính xác suất của phân phối nhị thức độ chính xác sẽ không cao nếu: P(A) = p gần 0 hoặc 1. Khi p gần 0 hoặc 1 ta sử dụng định lý sau:

Định lý 3: Xét một dãy vô hạn các phép thử độc lập, xác suất xuất hiện sự kiện A ở phép thử thứ n là p_n thoả mãn $\lim\limits_{n\to\infty} np_n = \lambda$.

Gọi X là số lần xuất hiện A trong dãy phép thử nói trên ta có:

$$\lim_{n\to\infty} P(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}$$

Áp dụng định lý nói trên khi $X \sim B(n, p)$ mà p gần 0 hoặc 1

Ta có: $$P(X = k) \approx e^{-np}\frac{(np)^k}{k!}.$$

Ví dụ : Tỉ lệ người có kí sinh trùng sốt rét trong máu ở một vùng đồng bằng sông Cửu Long là 0,01. Lấy mẫu máu của 100 người. Tính xác suất để có 5 người có ký sinh trùng sốt rét trong máu.

Gọi X là số người có kí sinh trùng sốt rét trong máu của 100 người.

$X \sim B(100; 0{,}01)$. Ta có $P(\, X = 5\,) \approx e^{-1}\dfrac{1}{5!} = 0{,}03066$.

3.4 Định lý giới hạn trung tâm:

Hai định lý nêu ở mục 3.1 chỉ là các trường hợp riêng của định lý giới hạn trung tâm. Ta biết rằng: Nếu biến ngẫu nhiên X có kì vọng a và phương sai b^2 thì biến ngẫu nhiên $Z = \dfrac{X - a}{b}$ gọi là biến ngẫu nhiên được chuẩn hoá. Những định lý đề cập tới dãy các biến ngẫu nhiên độc lập sau khi đã được chuẩn hoá, hội tụ theo quy luật tới phân phối chuẩn tắc được gọi là định lý giới hạn trung tâm.Trong mục này chúng tôi nêu hai trong những định lý nói trên.

Xét dãy các biến ngẫu nhiên độc lập: $X_1, X_2. \ldots . X_n \ldots$

có cùng kì vọng a và phương sai b^2. $\overline{X} = \dfrac{1}{n}\sum\limits_{i=1}^{n} X_i$, $Z_n = \dfrac{\overline{X} - a}{b}\sqrt{n}$

Định lý 4: Dãy các biến ngẫu nhiên $Z_1 \ldots Z_2 \ldots \ldots Z_n \ldots$ hội tụ theo qui luật tới phân phối chuẩn tắc.

Định lý vừa nêu là dạng đơn giản của định lý giới hạn trung tâm cổ điển, nó là trường hợp đặc biệt của định lí sau:

Định lý 5 (Lindeberg): Cho dãy các biến ngẫu nhiên độc lập: $X_1, X_2, \ldots\ldots, X_n\ldots.$

X_i có kì vọng a_i và phương sai b_i^2 .

$$Y_n = \frac{1}{n}\sum_{i=1}^{n} X_i\ ,\ A_n = \sum_{i=1}^{n} a_i\ ,\ B_n^2 = \sum_{i=1}^{n} b_i^2\ ,\ Z_n = \frac{Y_n - A_n}{B_n}\ .$$

Nếu giả thiết sau được thỏa mãn: $\lim\limits_{n\to\infty}\frac{1}{B_n^2}\sum\limits_{k=1}^{n}\int\limits_{|x|\ge\varepsilon B_n} x^2 dF_k(x) = 0$ với mọi $\varepsilon > 0$ thì dãy các biến ngẫu nhiên Z_1 , Z_2 , ….., Z_n…… hội tụ theo qui luật tới phân phối chuẩn tắc.

Người đọc muốn tìm hiểu kĩ hơn về các định lý giới hạn có thể đọc các sách tham khảo được nêu ở cuối giáo trình.

Bài tập chương III.

1. Một người tung đồng tiền 3 lần, gọi X là số mặt sấp ở phía trên của 3 đồng tiền vừa tung.

a. Lập bảng phân phối xác suất của X.

b. Tìm hàm phân phối xác suất của X.

c. Tính E(X), D(X).

2. Một người tung đồng thời hai con xúc xắc cân đối và đồng chất. Gọi tổng số chấm ở hai mặt trên là X. Nếu $X \leq 4$ anh ta mất 100 đồng. Nếu $5 \leq X < 10$ anh ta mất 50 đồng. Nếu $X > 10$ anh ta được 200 đồng. Gọi Y là số tiền anh ta nhận được sau một lần chơi.

a. Lập bảng phân phối xác suất củaY

b. Tìm hàm phân phối xác suất của Y

c, Tính kì vọng của Y. Có kết luận gì về trò chơi này

3. Một người có 3 viên đạn nhằm bắn vào bia cho tới khi trúng bia hoặc hết đạn thì dừng. Các lần bắn độc lập, xác suất trúng bia ở mỗi lần bắn là 0,8. Gọi Y là số đạn phải dùng

a. Lập bảng phân phối xác suất của Y

b. Tìm hàm phân phối xác suất của Y và vẽ đồ thị của hàm phân phối.

c. Tính E(Y), D(Y).

4. Hai cầu thủ bóng rổ lần lượt ném bóng vào rổ cho tới khi có một người ném trúng hoặc cả hai người ném hết 3 lượt thì dừng. Xác suất ném trúng rổ của người thứ nhất là 0,8, của người thứ hai là 0,6, các lần ném độc lập. Gọi X là số lần ném của người thứ nhất, Y là số lần ném của người thứ hai, $Z = X + Y$.

a. Hãy lập bảng phân phối xác suất của X, Y, Z.

b. Tính E(X), E(Y), E(Z)và D(X), D(Y), D(Z).

5. Một kho hàng có 6 lô hàng do nhà máy A sản xuất và 4 lô hàng do nhà máy B sản xuất. Lấy ngẫu nhiên từ kho hàng ra 3 lô hàng. Gọi X là số lô hàng do nhà máy A sản xuất trong 3 lô hàng được lấy ra.

a. Lập bảng phân phối xác suất của X.

b. Tính E(X), D(X).

6. Khi lai đậu hoa đỏ thuần chủng với đậu hoa trắng thuần chủng ở thế hệ F_1 các cây đậu đều có hoa màu đỏ. ở thế hệ F_2 các cây đậu có hoa màu đỏ và màu trắng theo tỷ lệ 3: 1. Chọn ngẫu nhiên 4 cây đậu ở thế hệ F_2. Gọi X là số cây đậu có hoa màu đỏ trong 4 cây trên.

a. Lập bảng phân phối xác suất của X.

b. Tính E(X), D(X).

7. Một cửa hàng kinh doanh thực phẩm tươi sống tại một khu dân cư qua một thời gian điều tra thị trường nhận thấy lượng tiêu thu thịt lợn X (kg/ngày) có bảng phân phối xác suất sau:

X	100	105	110	115	120	125
P	0,1	0,20	0,20	0,30	0,12	0,08

Giá thịt lợn mua vào 25000 đồng/kg bán ra 30000 đồng/ kg cuối ngày tiêu thụ không hết phải bán với giá 15000 đồng/ kg . Nếu là chủ cửa hàng, anh (chị) nên quyết định hàng ngày nên nhập lượng thịt lợn là bao nhiêu để có lợi nhuận trung bình là cao nhất.

8. Điều tra lượng sữa tươi đóng chai X bán ra hàng ngày của một cửa hàng được bảng phân phối xác suất như sau:

X	7	8	9	10	11	12
P	0,1	0,15	0,20	0,30	0,15	0,1

Mỗi chai sữa vào 3.000 đồng bán ra 5.000 đồng . Hết ngày không bán được thì vứt bỏ. Theo anh (chị) cửa hàng mỗi ngày nên nhập vào bao nhiêu chai sữa để có lợi nhuận trung bình cao nhất.

9. Một người trồng 2 giò phong lan và 2 chậu địa lan. Mỗi giò phong lan có xác suất sống là 0,6 nếu chết trồng lại lần thứ hai. Mỗi chậu địa lan có xác suất sống là 0,7 nếu chết trồng lại lần thứ hai. Phong lan và địa lan nếu sống sẽ ra hoa. Giá thành giống phong lan là 15.000 ^{đ} /giò, giá thành giống địa lan là 12.000 ^{đ} / chậu. Tiền công chăm sóc phong lan và địa lan tương ứng là 20.000 ^{đ} và 15.000^{đ} . Tiền bán mỗi giò phong lan là 200.000^{đ} , mỗi chậu địa lan là 150.000 ^{đ} . Hãy tính tiền lời trunh bình mà người trồng hoa thu được.

10. Khi cho ấp trứng vịt trong lò ấp thì 85% trứng nở. Trong số vịt con mới ra đời có 60% vịt mái.

a. Cho ấp 5000 quả trứng, nhiều khả năng nhất có bao nhiêu vịt mái.

b. Khi nuôi xác suất sống của vịt mái là 0,9 , của vịt đực là 0,8. Nếu tiền trứng và tiền ấp là 1200 đồng /1 quả, tiền thức ăn và tiền công chăm sóc là 1400 đồng/ 1 con(kể cả con sống và con chết). Tiền bán vịt mái là 6000 đồng, tiền bán vịt đực là 5000 đồng/1 con thì chủ lò lãi trung bình là bao nhiêu?.

11 . Ba vận động viên bóng bàn A, B, C có trình độ ngang nhau tham gia một giải đấu theo qui tắc: A và B gặp nhau ở trận đầu nếu ai thắng sẽ đấu tiếp với C. Người thắng ở trận thứ hai sẽ gặp người thua ở trận đầu. Giải đấu sẽ kết thúc khi có một người thắng liên tiếp 2 trận. Hãy tính xác suất thắng trận của từng người.

12. Một nhân viên tại một trung tâm trò chơi điện tử phụ trách 20 máy. Xác suất để mỗi máy cần sự giúp đỡ của nhân viên trong một giờ là 0,1. Các máy hoạt động độc lập.

a. Gọi X là số máy cần sự giúp đỡ của nhân viên trong 1 giờ. Tìm qui luật phân phối xác suất của X

b. Tìm xác suất để trong 1 giờ làm việc có ít nhất 1 máy cần đến sự giúp đỡ của nhân viên.

13. Một lô hàng có 8 sản phẩm loại A, 2 sản phẩm loại B. Rút lần lượt từ lô hàng ra 1 sản phẩm để kiểm tra cho tới khi gặp sản phẩm loại B thì dừng. Gọi X là số sản phẩm rút ra.

a. Lập bảng phân phối xác suất của X.

b. Tính E(X), D(X).

14. Tỉ lệ người có ký sinh trùng sốt rét trong máu tại một vùng miền núi là 0,1. Người ta tiến hành xét nghiệm mẫu máu cho dân ở vùng này cho tới khi phát hiện ra mẫu có kết quả dương tính thì dừng. Gọi X là số mẫu máu cần xét nghiệm.

a. Tính E(X), D(X).

b. Tính ModX.

15. Tiến hành một dãy các phép thử độc lập, xác suất xuất hiện sự kiện A ở mỗi phép thử P(A) = p. Gọi X là số lần thất bại trước thành công đầu tiên.

a. Hãy xác định quy luật phân phối xác suất của X.

b. Nếu Y là số lần thất bại trước thành công thứ k. Hãy xác định quy luật phân phối xác suất của Y.

16. Cho $f(x) = \begin{cases} 0 & \text{khi } x < 0 \\ Axe^{-x} & \text{khi } x \geq 0 \end{cases}$

a. Tìm A để f(x) trở thành hàm mật độ xác suất.

b. Tìm hàm phân phối xác suất tương ứng.

c. Tính E(X).

17. Cho $f(x) = Ae^{-|x|}$

a. Tìm A để f(x) là hàm mật độ xác suất của biến ngẫu nhiênX.

b. Tìm hàm mật độ xác suất của $Y = X^2$.

c. Tính xác suất $P[\,Y \geq 1\,]$.

18. Cho X, Y là 2 biến ngẫu nhiên độc lập có bảng phân phối xác suất sau:

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

Y	0	1	2
P	0,3	0,5	0,2

a. Hãy lập bảng phân phối xác suất của Z = X + Y.
b. Hãy lập bảng phân phối xác suất của U = X.Y

c. Hãy lập bảng phân phối xác suất đồng thời của (X, Y)

19. Tỉ lệ hạt thóc không nảy mầm là 0.001. Tính xác suất sao cho chọn ngẫu nhiên 1000 hạt thì

a. Có không quá 10 hạt không nảy mầm.

b. Có đúng 10 hạt không nảy mầm.

20. Cho $X_1, X_2, \ldots, X_i, \ldots X_n$ là dãy các biến ngẫu nhiên độc lập có $D(X_n) \leq M \; \forall n$. Chứng minh rằng dãy các biến ngẫu nhiên trên tuân theo luật số lớn.

21. Cho $X_1, X_2, \ldots, X_i, \ldots X_n$ là dãy các biến ngẫu nhiên độc lập có $\lim\limits_{n\to\infty}\frac{1}{n}\sum\limits_{i=1}^{n} D(X_i) = 0$.

Chứng minh rằng dãy các biến ngẫu nhiên trên tuân theo luật số lớn.

22. Các thế hệ con ở thế hệ F_2 có các kiểu gen AA, Aa, aa theo tỉ lệ 1 : 2 : 1. Xét 10 cá thể ở thế hệ F_2.

a. Tính xác suất để có 3 cá thể mang gen AA, 5 cá thể mang gen Aa và 2 cá thể mang gen aa.

b. Tính xác suất để số cá thể mang gen cùng kiểu nhiều hơn số cá thể mang gen khác kiểu.

23. Một phòng thí nghiệm được cấp kinh phí để tìm một nguyên tố vi lượng trong mẫu đất. Khả năng thành công ở thí nghiệm đầu là 0,5 với chi phí 10 triệu. Nếu thí nghiệm đầu thất bại kinh phí được tăng lên 15 triệu để thêm các loại hoá chất mới nên xác suất thành công là 0,6. Nếu thí nghiệm lần này cũng thất bại thì kinh phí vẫn được cấp như làn thứ 2 và xác suất thành công cũng là 0,6. Nếu lần này thất bại việc thí nghiệm kết thúc và phòng này được coi là không hoàn thành nhiệm vụ. Gọi X là số tiền phòng phải dùng.Tính E(X).

24. Cho: $f(x)=\begin{cases} 0 & x<0 \\ Ax^2e^{-2x} & x \geq 0 \end{cases}$

a. Xác định A để f(x) là hàm mật độ xác suất củamột biến ngẫu nhiên X nào đó.

b. Tìm hàm phân phối xác suất tương ứng.

c. Tính: $P(0 < X < 1)$

d. Tính kì vọng và phương sai của X

25. Cho: $f(x)=\begin{cases}0 & x<0\\ ke^{\frac{-x}{2\theta}} & x\geq 0\end{cases}$

a. Tìm k để f(x) là hàm mật là hàm mật độ xác suất của biến ngẫu nhiênY nào đó.

b. Tìm hàm phân phối xác suất tương ứng.

c. Tìm kì vọng và phương sai của Y.

26. Tỉ lệ người có ký sinh trùng sốt rét ở mỗi người dân vùng cao là 0,2. Thử máu 900 người dân ở vùng cao nói trên.

a. Tính xác suất để số mẫu máu có ký sinh trùng sốt rét nằm trong khoảng từ 156 đến 204.

b. Tính xác suất để số mẫu máu có ký sinh trùng sốt rét không nhỏ hơn 120

c. Tính xác suất để số mẫu máu có ký sinh trùng sốt rét không lớn hơn 150.

27. Xác suất của một loại hạt giống nảy mầm sau khi gieo là 0,8.

a. Phải gieo ít nhất bao nhiêu hạt để với xác suất không nhỏ hơn 0,977 có thể tin rằng có trên 100 hạt nảy mầm.

b. Phải gieo ít nhất bao nhiêu hạt để với xác suất không nhỏ hơn 0,977 có thể tin rằng tần suất nảy mầm lệch khỏi xác suất nảy mầm không quá 0,01 về giá trị tuyệt đối.

28. Trọng lượng X của mỗi con bò trong một đàn bò là biến ngẫu nhiêncó phân phối chuẩn với kì vọng 300 kg và độ lệch chuẩn 50 kg. Chon ngẫu nhiên 1 con bò trong chuồng. Tính xác suất để con bò được chọn :

a. Có trọng lượng trên 350 kg.

b. Có trọng lượng từ 250 đến 350 kg.

c. Chọn ngẫu nhiên 4 con bò trong đàn bò nói trên. Tính xác suất để 2 trong 4 con bò được chọn có trong lượng từ 250 đến 350 Kg.

29. Trọng lượng mỗi con cá chép trong hồ là một biến ngẫu nhiêncó phân phối chuẩn với kì vọng 2 kg và độ lệch chuẩn 0,4 kg. Cá có trọng lượng đến 1,5 kg là cá loại ba, cá có trọng lượng từ 1,5 kg đến 2,5 kg là cá loại hai, cá có trọng lượng trên 2,5 kg là cá loại một.

a. Bắt một con cá trong hồ, tính xác suất để con cá bắt được là cá loại hai. Bắt 1000 con cá trong hồ thì khả năng nhất có bao nhiêu cá loại hai

b. Có người bắt ngẫu nhiên 3 con cá trong hồ thì khả năng bắt được 3 con cá cùng loại lớn hơn khả năng bắt được 2 con cá cùng loại. Theo anh (chị) ý kiến trên là đúng hay sai? Tại sao?

c. Bắt 5 con cá trong hồ, tính xác suất để 1 con loại một, 3 con loại hai và 1 con loại ba.

30. Một người chuẩn bị 10 hốc để trồng bí, mỗi hốc gieo một hạt xác suất nảy mầm của mỗi hạt là 0,9.

a. Tính xác suất để có ít nhất 1 hốc không có hạt nảy mầm.

b. Biết mỗi hốc có hạt nảy mầm cây bí sẽ sống và ra quả đem bán được 50.000 ^{đ}. Tiền lời thu được từ 10 hốc bí trên là bao nhiêu?

c. Một người góp ý mỗi hốc bí nên gieo 3 hạt, nếu mỗi hốc có hơn 1 cây non thì nhổ bỏ chỉ để lại 1 cây với lập luận là làm như thế xác suất để cả 10 hốc đều có cây non sẽ lớn hơn 0,90. Anh (chị) có đồng ý với lập luận trên không ? Tại sao?

31. Năng suất lúa X ở một địa phương là biến ngẫu nhiên có phân phối chuẩn với kì vọng 60 tạ và độ lệch chuẩn là 5 tạ/ha. Gặt ngẫu nhiên 5 thửa ruộng.

a. Tính xác suất để trong 5 thửa ruộng được gặt có 2 thửa ruộng có năng suất lệch khỏi kì vọng không quá 2 tạ.

b.Tính xác suất để trong 5 thửa ruộng được gặt có ít nhất 1 thửa ruộng có năng suất lệch khỏi kì vọng không quá 2 tạ.

32. Các khách hàng mua xe hơi tại một đại lý nếu vì lý do nào đó không vừa ý, chiếc xe đã mua được trả lại trong vòng 2 ngày sau khi mua và được lấy nguyên tiền. Mỗi xe bị trả lại như vậy đại lý bị thiệt hại 5 triệu đồng. Tỉ lệ xe bị trả lại là 10 %. Có 100 xe mới được bán ra.

a. Tìm kì vọng và độ lệch chuẩn số xe bị trả lại.

b. Tìm kì vọng và độ lệch chuẩn của số tiền đại lý bị thiệt hại.

33. Biến ngẫu nhiên X có phân phối Poisson với $\lambda = 10$. Tính xác suất để X không vượt quá 3.

34. Số lần trục trặc của một mạng máy tính ở một trường đại học trong một tháng có phân phối Poisson với $\lambda = 2$. Tính xác suất để trong một tháng có 5 máy bị trục trặc.

35. Biến ngẫu nhiênX có phân phối mũ với kì vọng là 2. Hãy tính xác suất để X nhận giá trị nhỏ hơn 3.

36. Một cái bia hình chữ nhật có chiều dài 6m, chiều rộng 4m. Một người nhằm bắn vào bia, Gọi X, Y lần lượt là khoảng cách giữa điểm chạm của viên đạn với mặt phẳng chứa bia tới 2 trục đối xứng của bia. Biết X có phân phối chuẩn với kì vọng là 2 m và độ lệch chuẩn 0,9m. Y có phân phối chuẩn với kì vọng là 1,5m và độ lệch chuẩn 0,6m và X độc lập với Y. Tính xác suất để viên đạn được bắn trúng bia.

37. Cho vectơ ngẫu nhiên 2 chiều có bảng phân phối xác suất đồng thời sau:

Y \ X	0	1	2	3
0	0,05	0,10	0,05	0,10
1	0,01	0,05	0,10	0,10
2	0,05	0,10	0,05	0,15

a Tìm phân phối của các thành phần X,Y.

b. Tìm Cov(X,Y) , ρ(X, Y).

c. Tìm phân phối của $X/_{Y=2}$ và tính E($X/_{Y=2}$).

d. Tìm phân phối của $Y/_{X=1}$ và tính E($Y/_{X=1}$).

38. Xác suất sinh con trai là 0,5 với mỗi người mẹ. Một gia đình dự định có 3 con. Gọi X là biến ngẫu nhiên chỉ số con gái trong gia đình có 3 con. Y là biến ngẫu nhiên chỉ dãy các trẻ em có giới tính liền nhau, chẳng hạn nếu có 3 đứa trẻ đều là gái hoặc đều là trai thì Y = 1. Nếu đứa đầu là gái, đứa thứ hai là trai, đứa thứ ba là gái thì Y = 3.

a. Hãy lập bảng phân phối xác suất đồng thời của vectơ (X,Y).

b. Gọi Z là chi phí về quần áo biết Z = X +Y +10 . Hãy lập hàm phân phối xác suất của Z.

c. Tính kì vọng và phương sai của Z.

39. Xác suất để mỗi cây sống sau một thời gian trồng là 0,8. Trồng 400 cây.

a. Tính xác suất để có từ 300 cây đến 340 cây sống.

b. Tính xác suất để số cây sống không nhỏ hơn 300 cây.

c. Tính xác suất để số cây sống không quá 340 cây.

40. Tỉ lệ người có kí sinh trùng sốt rét tại một vùng đồng bằng là 0,001. Xét nghiệm 4000 mẫu máu.

a. Tính xác suất để có 8 mẫu có kí sinh trùng sốt rét.

b. Tính xác suất để để không có quá 5 mẫu máu có kí sinh trùng sốt rét

Chương 4:
Những khái niệm cơ bản mở đầu về thống kê.

Thống kê toán học có thể coi là một phương pháp khoa học phân tích và xử lý dữ liệu có được nhờ các thí nghiệm, các cuộc điều tra nghiên cứu các hiện tượng tự nhiên, các vấn đề kỹ thuật cũng như các vấn đề xã hội. Những dữ liệu ở đây có thể là những đặc tính định tính, cũng có thể là những đặc tính định lượng. Từ những dữ liệu thu thập được, dựa vào các quy luật xác suất để đưa ra những quyết định, những đánh giá và các dự báo về những hiện tượng đang được thí nghiệm hoặc đang được quan sát là mục đích của thống kê toán học.

I. Tổng thể và mẫu.

1.Tổng thể và kích thước của tổng thể.

Khi nghiên cứu hoặc quan sát một hoặc một số đặc tính của các phần tử trong một tập hợp nào đó thì tập hợp các phần tử này được gọi là tổng thể hoặc đám đông, mỗi phần tử thuộc tập hợp này được gọi là các cá thể của tổng thể hoặc cá thể của đám đông. Chẳng hạn như khi nghiên cứu chiều cao của thanh niên Việt Nam (có tuổi từ 15 đến 30) thì tổng thể gồm tất cả các thanh niên Việt Nam ở thời điểm bắt đầu nghiên cứu, đặc tính mà ta nghiên cứu là một đặc tính định lượng đó là chiều cao của mỗi cá thể trong tập hợp thanh niên Việt Nam. Khi điều tra một loại bệnh mới xuất hiện trên gia cầm tại các tỉnh thuộc đồng bằng Bắc Bộ thì tổng thể là toàn bộ gia cầm đang được chăn nuôi tại đồng bằng Bắc Bộ. Đặc tính mà ta quan tâm tới trong trường hợp này là một đặc tính định tính xét mỗi cá thể gia cầm trong tổng thể có hoặc không có loại bệnh mà ta quan tâm. Số lượng các cá thể trong một tổng thể được gọi là kích thước của tổng thể . Người ta thường dùng chữ hoa N để chỉ kích thước của tổng thể .Thông thường kích thuớc tổng thể rất lớn. Chính vì vậy việc xét một hoặc một số các đặc tính của tất cả các cá thể trong tổng thể sẽ tốn kém về thời gian, công sức cũng như tiền của. Trong nhiều trường hợp việc nghiên cứu tất cả các cá thể của tổng thể là không khả thi vì số lượng cá thể ở tổng thể có mặt trong thời gian nghiên cứu luôn biến động. Chẳng hạn việc xét tuổi thọ của tất cả công dân Việt Nam là một việc làm không khả thi. Trong nhiều trường hợp khác việc nghiên cứu tất cả các cá thể của tổng thể lại là một việc làm không có ý nghĩa. Nếu cần đánh giá chất lượng bia của nhà máy bia Hà Nội sản xuất trong một tháng mà đem mở tất cả các chai bia này ra để kiểm tra thì không thể chấp nhận được vì sau khi kiểm tra sẽ không còn bia để bán.

2. Mẫu và kích thước mẫu.

Một tập hợp các cá thể lấy ra từ tổng thể gọi là mẫu. Số lượng cá thể có mặt trong một mẫu gọi là kích thước của mẫu. . Người ta thường dùng chữ n để chỉ kích thước mẫu. Kích thước của mẫu thường nhỏ hơn rất nhiều so với kích thước của tổng thể. Từ tổng thể đã cho ta có thể lấy ra nhiều mẫu khác nhau với cùng một kích thước n. Tập hợp tất cả các mẫu có thể lấy ra được từ tổng thể được gọi là không gian mẫu. Nếu đặc tính mà ta nghiên cứu là đặc tính định lượng X thì với một mẫu cụ thể có kích thước n, cá thể thứ i trong mẫu có đặc tính $X = x_i$ thì bộ số

$(x_1, x_2, ..., x_i, ... x_n)$ cũng được gọi là một mẫu. Tập hợp tất cả các bộ n số tương ứng với tập tất cả các mẫu có cùng kích thước n cũng được gọi là không gian mẫu. Trong trường hợp này không gian mẫu là một tập con của không gian R^n.
Thay vì nghiên cứu tất cả các cá thể có mặt trong tổng thể ta chuyển sang nghiên cứu một bộ phận của tổng thể là mẫu vì vậy mẫu phải đại diện một cách khách quan nhất cho tổng thể. Để đảm bảo yêu cầu trên, người ta đưa ra các phương pháp lấy mẫu sau.

3. Các phương pháp lấy mẫu

3.1 Lấy mẫu ngẫu nhiên không hoàn lại: Đánh số các cá thể trong tổng thống thể từ 1 đến N. Rút ngẫu nhiên lần lượt n cá thể đưa vào một mẫu. Từ phương pháp lấy mẫu này ta thấy: xác suất để cá thể đầu tiên có mặt trong mẫu là $\frac{1}{N}$, xác suất để cá thể thứ i có mặt trong mẫu là $\frac{1}{N-i+1}$.

3.2 Lấy mẫu ngẫu nhiên có hoàn lại: Đánh số các cá thể trong tổng thể từ 1 đến N. Rút ngẫu nhiên từ tổng thể ra 1 cá thể, ghi đặc tính của cá thể này rồi trả cá thể đó về tổng thể, đặc tính vừa ghi được coi là phần tử đầu của mẫu. Việc xác định các phần tử tiếp theo của mẫu cũng được làm tương tự như trên. Từ phương pháp lấy mẫu ngẫu nhiên có hoàn lại ta thấy: xác suất để mỗi cá thể có mặt trong mẫu là $\frac{1}{N}$. Mỗi cá thể có mặt nhiều lần trong mẫu.
Ta nhận thấy rằng với kích thước n, số lượng các mẫu trong trường hợp lấy mẫu không lặp là A^n_N, số lượng các mẫu trong trường hợp lặp N^n . Khi N lớn hơn rất nhiều so với n thì A^n_N và N^n sai khác nhau không đáng kể vì vậy việc lấy mẫu có hoàn lại cũng gần giống như việc lấy mẫu không hoàn lại.

3.3. Lấy mẫu theo các lớp: Để mẫu đại diện một cách khách quan nhất cho tổng thể là ưu tiên hàng đầu của việc lấy mẫu. Trong nhiều trường hợp người ta chia tổng thể ra làm k lớp rồi từ mỗi lớp lấy ngẫu nhiên ra một số cá thể đưa vào mẫu. Nếu số lượng các cá thể ở lớp thứ i là N_i thì số cá thể được chọn đưa vào mẫu của lớp này là n_i cần thoả mãn $\frac{n_i}{n} \approx \frac{N_i}{N}$.

3.4: Lấy mẫu theo chu kỳ : Trong việc kiểm tra chất lượng các sản phẩm công nghiệp được sản xuất theo dây chuyền việc lấy mẫu ngẫu nhiên sẽ gặp khó khăn và tốn kém. Phương pháp lấy mẫu theo chu kỳ tỏ ra thuận lợi trong nền sản xuất công nghiệp hiện đại. Cứ sau một chu kỳ gồm T sản phẩm ta lấy ra một sản phẩm để đưa vào mẫu. Để tránh sự trùng lặp của chu kì sản xuất ra các sản phẩm tốt, xấu của dây chuyền với chu kỳ lấy mẫu ta có thể thay đổi chu kỳ T trong các đợt lấy mẫu khác nhau với mục đích mẫu phải đại diện một cách khách quan nhất cho tổng thể .
Các phương pháp lấy mẫu trên là các phương pháp phổ biến trong việc thu thập các dữ liệu.Việc lấy mẫu tốt, xấu theo nghĩa có khách quan hay không ảnh hưởng rất lớn tới việc đưa ra kết luận có chính xác hay không về các đặc tính có mặt trong tổng thể .

II. Bố trí mẫu và phân phối mẫu.

Khi nghiên cứu đặc tính định lượng X ở tổng thể bằng các phương pháp lấy mẫu ta nhận được một mẫu có kích thước n gồm n số liệu:

$x_1, x_2, \ldots, x_i, \ldots, x_n$.

Để khai thác và xử lý các thông tin chứa đựng trong dãy số liệu này ta cần sắp xếp số liệu nhằm dễ dàng nhận ra các đặc trưng của dãy số liệu đó.

1. Sắp xếp số liệu

Thông thường ta sắp xếp số liệu theo thứ tự tăng dần. Dãy số liệu này ưu điểm hơn dãy số liệu ban đầu, ta có thể dễ dàng nhận biết giá trị nhỏ nhất và giá trị lớn nhất của các số liệu mẫu. Biết được biên độ dao động của các số liệu mẫu. Với cách sắp xếp này ta dễ dàng nhận biết các số liệu có mặt trong mẫu hơn một lần vì các số liệu bằng nhau được xếp liền nhau.

1.2 Sắp xếp mẫu theo bảng tần số và tần suất mẫu.

Bảng tần số: Là bảng hai dòng sau:

x_i	x_1	x_2 ...	x_i ...	x_k
n_i	n_1	n_2 ...	n_i ...	n_k

Dãy trên ghi các giá trị có thể có của mẫu theo thứ tự tăng dần, dòng dưới ghi tần số tương ứng. Tần số mẫu là số cá thể có đặc tính $X = x_i$ trong mẫu. Bảng tần số mẫu cho ta nhiều thông tin hơn dãy số liệu được sắp xếp theo thứ tự tăng dần. Ngoài những thông tin có được như dãy số liệu được sắp xếp theo thứ tự tăng dần, qua bảng tần số ta có thể biết được số liệu nào có mặt nhiều nhất, số liệu nào có mặt ít nhất trong mẫu.

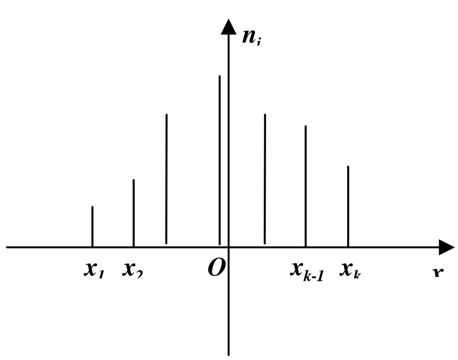
Bảng tần suất: Giá trị $f_i = \dfrac{n_i}{n}$ được gọi là tần suất của cá thể có $X = x_i$ trong mẫu. Bảng tần suất mẫu là bảng sau:

x_i	x_1	x_2 ...	x_i ...	x_k
f_i	f_1	f_2 ...	f_i ...	f_k

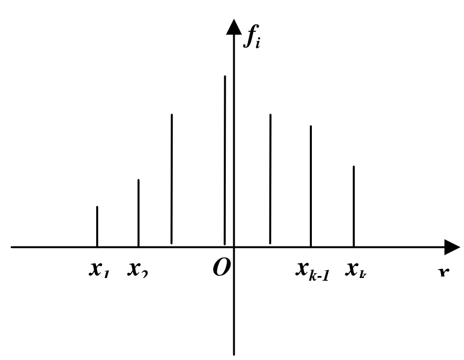
Ngoài những thông tin có được như bảng tần số mẫu ta còn biết được tỷ lệ phần trăm đóng góp của số liệu mẫu.

1.3 Biểu diễn bằng đồ thị:

Trong mặt phẳng Oxy với Oy là trục chỉ tần số n_i, Ox chỉ giá trị mẫu x_i. Xét tập các điểm $M_i(x_i, n_i)$, nối M_i với điểm nằm trên trục hoành có hoành độ x_i ta có biểu đồ hình gậy biểu diễn tần số. Cũng trong mặt phẳng Oxy với Oy là trục chỉ tần suất f_i, Ox chỉ giá trị mẫu x_i. Xét tập các điểm $N_i(x_i, n_i)$, nối N_i với x_i ta có biểu đồ hình gậy biểu diễn tần suất. Các biểu đồ này thể hiện bởi hai hình sau. Lấy tỷ lệ trên trục Oy trong biểu đồ tần suất gấp n lần trên biểu đồ tần số ta được hai biểu đồ giống nhau.

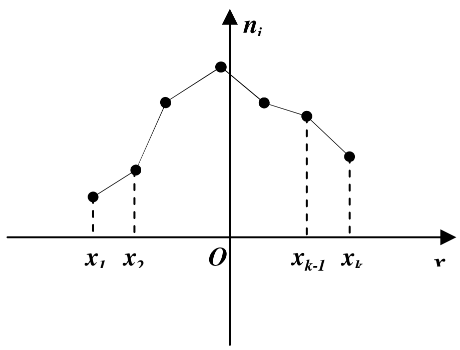


Hình 1

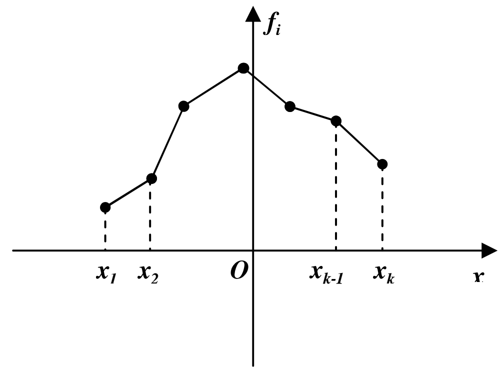


Hình 2

Nối điểm M_i với điểm M_{i+1}, N_i với điểm N_{i+1} với i = 1, k-1 ta được các đường gấp khúc gọi là các đa giác tần số và đa giác tần suất. Nếu tỷ lệ trên trục tung của đa giác tần suất gấp n lần tỷ lệ trên trục tung của đa giác tần số thì hai đa giác này là như nhau.



Hình 3



Hình 4

Hình 3 và hình 4 là các hình vẽ đa giác tần số và đa giác tần suất.

2. Sắp xếp số liệu theo các lớp: Khi kích thước mẫu lớn để thuận tiện cho việc phân tích và xử lý số liệu ta có thể chia mẫu ra làm k lớp kế tiếp với cùng độ rộng

$$l = \frac{\text{giá trị lớn nhất của mẫu - giá trị nhỏ nhất của mẫu}}{k}$$

Khi đó mẫu được sắp xếp theo bảng sau:

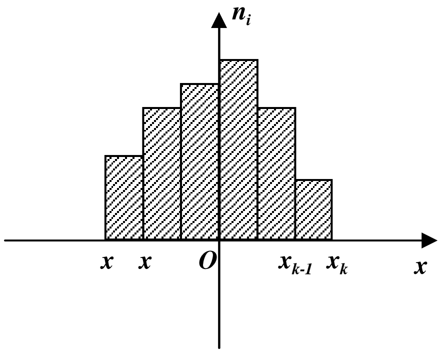
Lớp	$[x_1, x_{1+l})$	$[x_{1+l}, x_{1+2l})$	...	$[x_{1+(i-1)l}, x_{1+il})$	...	$[x_{1+(k-1)l}, x_k]$
Tần số n_i	n_1	n_2		n_i		n_k
Tần suất f_i	f_1	f_2		f_2		f_k

n_i = số cá thể có đặc tính X thoả mãn $x_{1+(i-1)l} \leq$
$X <\ x_{1+il}$ có trong mẫu

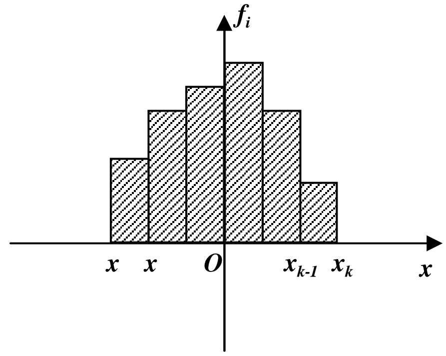
$f_i = \frac{n_i}{n}$ là tần suất mẫu của các cá thể mẫu thuộc lớp thứ i.

Các hình sau gọi là biểu đồ tương ứng với mẫu được bố trí theo lớp. Nếu coi mỗi lớp có độ rộng bằng một đơn vị dài thì mỗi hình chữ nhật có đáy $\lfloor x_{1+(i-1)l} ; x_{1+il} \rfloor$ trong hình 5 có diện tích bằng n_i . Mỗi hình chữ nhật với đáy là $\lfloor x_{1+(i-1)l}\ ;\ x_{1+il} \rfloor$ trong hình 6 có diện tích là f_i. Nếu nối trung điểm các cạnh phía trên của các hình chữ nhật liền kề ta được đa giác tần số và đa giác tần suất tương ứng .

x_i	x_1	x_2		x_i		x_k
f_i	f_1	f_2		f_i		f_k



Hình 5



Hình 6

3. Hàm phân phối mẫu.

Với mẫu cho bởi bảng tần suất:

Hàm phân phối mẫu là hàm cho bởi:

$$F_n(x) = \begin{cases} 0 & khi \quad x < x_1 \\ f_1 & khi \quad x_1 \le x < x_2 \\ \quad ..... & \\ f_1 + f_2 + ... + f_i & khi \quad x_i \le x < x_{i+1} \\ \quad ..... & \\ 1 & khi \quad x \ge x_k \end{cases}$$

Nhận thấy rằng: hàm phân phối mẫu phụ thuộc vào mẫu và kích thước của mẫu. Với hai mẫu có cùng kích thước có thể nhận được hai hàm phân phối mẫu khác nhau. Giả sử đặc tính định lượng X ở tổng thể có hàm phân phối F(x) .

Định lý Glivenco-Katelli đã chỉ ra rằng: với mọi số thực x, Sup $|F_n(x) - F(x)|$ hội tụ hầu chắc chắn tới 0 khi n tiến ra vô cùng. Điều này có nghĩa là với n đủ lớn ta có thể lấy phân phối mẫu thay cho phân phối xác suất chưa biết của tổng thể.

III. Mẫu ngẫu nhiên và các đặc trưng mẫu.

1. **Mẫu ngẫu nhiên.**

Giả sử đặc trưng biến X ở mỗi cá thể ở tổng thể là một đại lượng ngẫu nhiên có hàm phân phối xác suất F(x), ta tiến hành một phép lấy mẫu ngẫu nhiên có kích thước n. Gọi X_i là biến ngẫu nhiên chỉ giá trị X của cá thể thứ i trong mẫu, ta thấy các X_i là các biến ngẫu nhiên có cùng phân phối xác suất với X. Với mỗi mẫu cụ thể X_i sẽ có giá trị xác định là x_i. Do việc lấy mẫu là độc lập nên dãy

$X_1, X_2, ..., X_n$ là các đại lượng ngẫu nhiên độc lập.

Ta có thể hiểu véctơ ngẫu nhiên ($X_1, X_2, .. X_i,.., X_n$) trong đó X_i độc lập với X_j khi $i \neq j$ và các X_i có cùng phân phối xác suất với X được gọi là một mẫu ngẫu nhiên lấy từ tổng thể đã cho.

Với mỗi mẫu cụ thể ta có một bộ n số thực ($x_1, x_2, ..., x_i, ..., x_n$), bộ số này này gọi là một thể hiện của mẫu ngẫu nhiên ($X_1, X_2, ...X_i,.., X_n$) ứng với mẫu đang xét.

Ví dụ : Xét tổng thể là tập sinh viên Việt Nam, đặc trưng X là chiều cao của mỗi sinh viên. Xét một mẫu có kích thước n = 10, véc tơ ngẫu nhiên:
($X_1, X_2, ..., X_{10}$) với X_i là chiều cao của sinh viên thứ i trong mẫu là một mẫu ngẫu nhiên. Với một mẫu cụ thể

$x_1 = 1,65$; $x_2 = 1,52$; $x_3 = 1,60$; $x_4 = 1,65$; $x_5 = 1,70$

$x_6 = 1,58$; $x_7 = 1,68$; $x_8 = 1,72$; $x_9 = 1,52$; $x_{10} = 1,60$

Bộ số (1,65; 1,52; 1,60; 1,65; 1,70; 1,58; 1,68; 1,72; 1,52; 1,60) là một thể hiện của mẫu ngẫu nhiên ($X_1, X_2, ..., X_{10}$).

2. Các đặc trưng mẫu:

2.1 Hàm mẫu (thống kê): Hàm g($X_1, X_2, .., X_n$) với ($X_1, X_2, ., X_n$) là một mẫu ngẫu nhiên gọi là một hàm mẫu hay một thống kê .
Vì mẫu ($X_1,X_2, ..., X_n$) là một véc tơ ngẫu nhiên nên g($X_1, X_2, ..., X_n$) là một biến ngẫu nhiên. Với mẫu cụ thể biến ngẫu nhiên $X_i = x_i$ với i = 1, 2 . . . , n
g($x_1, x_2, ... , x_n$), là giá trị cụ thể mà thống kê g($X_1, X_2, ..., X_n$) nhận tương ứng với mẫu đã cho. Phân phối xác suất của thống kê g ($X_1, X_2, ..., X_n$) phụ thuộc vào phân phối xác suất của biếnngẫu nhiên X ở tổng thể .

2.2 Trung bình và phương sai mẫu.

Xét mẫu ngẫu nhiên ($X_1, X_2, ..., X_n$)

Thống kê $\overline{X} = \frac{1}{n}$($X_1 + X_2 + ...+ X_n$) gọi là trung bình mẫu. Với mẫu cụ thể:

($x_1, x_2, ..., x_n$), $\overline{x} = \frac{1}{n}(x_1 + x_2 + ...+ x_n)$ là giá trị mà trung bình mẫu nhận được ứng với mẫu đã cho. Ta coi $\overline{x}$ là một thể hiện của $\overline{X}$ với mẫu đã cho.

Thống kê: $\hat{S}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$ gọi là phương sai mẫu chưa hiệu chỉnh,

$\hat{s}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$ cũng được gọi là phương sai mẫu chưa hiệu chỉnh với ý nghĩa nó là một thể hiện của $\hat{S}^2$ ứng với mẫu cụ thể.

Thống kê: $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ gọi là phương sai mẫu đã hiệu chỉnh,

$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$ cũng được gọi là phương sai mẫu đã hiệu chỉnh tương ứng với mẫu đã cho.

Để phân biệt trong phần còn lại của giáo trình ta sử dụng các chữ viết hoa chỉ các thống kê của mẫu ngẫu nhiên và các chữ không viết hoa chỉ các thể hiện tương ứng,
Thống kê : $\hat{S}$ gọi là độ lệch chuẩn mẫu chưa hiệu chỉnh và $\hat{s}$ là thể hiện của $\hat{S}$ với mẫu đã cho.
Thống kê : S gọi là độ lệch chuẩn mẫu đã hiệu chỉnh và s là thể hiện của S với mẫu đã cho.
Khái niệm chưa hiệu chỉnh và đã hiệu chỉnh của phương sai mẫu sẽ giải thích ở chương sau. Ta có thể tính các thống kê trên theo công thức sau:

$$\hat{S}^2 = \overline{X^2} - \overline{X}^2 ; S^2 = \frac{n}{n-1}\hat{S}^2 \text{ với } \overline{X^2} = \frac{1}{n}\sum_{i=1}^{n}x_i^2$$

Ví dụ : Một mẫu cho bởi bảng tần số sau:

x_i	50	52	54	56
n_i	2	3	3	2

Ta có : $\overline{x} = \frac{1}{10}(2.50 + 3.52 + 3.54 + 2.56) = 53$

$$\hat{s}^2 = \frac{1}{10}[2(50-53)^2 + 3(52-53)^2 + 3(54-53)^2 + 2(56-53)^2] = 4,2$$

$$s^2 = \frac{1}{9}[2(50-53)^2 + 3(52-53)^2 + 3(54-53)^2 + 2(56-53)^2] = \frac{14}{3}$$

$\hat{s} = \sqrt{4,2} \approx 2,05$, $s = 2,16$.

Trong thực hành chúng ta chỉ tính các thể hiện của thống kê ứng với một mẫu cụ thể, để thuận lợi cho việc tính toán ta đưa ra một công thức khác để tính

$$\hat{s}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})$$

$$\Rightarrow \hat{s}^2 = \frac{1}{n}\sum_{i=1}^{n}x^2{}_i - \frac{2\overline{x}}{n}\sum_{i=1}^{n}x_i + \frac{1}{n}\sum_{i=1}^{n}\overline{x}^2) = \frac{1}{n}\sum_{i=1}^{n}x^2{}_i - \overline{x}^2$$

Đặt $\overline{x^2} = \frac{1}{n}\sum_{i=1}^{n}x^2{}_i$ ta có $\hat{s}^2 = \overline{x^2} - \overline{x}^2$, ta cũng có

$s^2 = \frac{n}{n-1}\hat{s}^2$ suy ra $s^2 = \frac{n}{n-1}(\overline{x^2} - \overline{x}^2)$

2.3 Một số đặc trưng mẫu khác: Do các đặc trưng nêu ra dưới đây không được sử dụng phổ biến trong phần còn lại của chương trình nên chúng tôi chỉ nêu ra các thể hiện của các đặc trưng mẫu ngẫu nhiên tương ứng với một mẫu cụ thể.

Mô men mẫu cấp k là số $m_k(0) = \frac{1}{n}\sum_{i=1}^{n}x_i^k$

Mô men trung tâm mẫu cấp k là số $m_k = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^k$

Hệ số bất đối xứng mẫu là số: $h_3 = \frac{m_k}{\hat{s}^3}$

Hệ số nhọn mẫu là số : $h_4 = \frac{m_4}{\hat{s}^4} - 3$

IV Một số phân phối xác suất thường gặp trong thống kê .

Phân phối chuẩn là một phân phối khá phổ biến, các biến ngẫu nhiên gặp trong nhiều lĩnh vực thường có phân phối chuẩn hoặc xấp xỉ chuẩn. Các biểu thức được trình bày trong phần còn lại của thống kê đều có liên quan mật thiết đến phân phối chuẩn. Mục này giới thiệu (không chứng minh) một số định lý về phân phối chuẩn và các phân phối có liên quan đến phân phối chuẩn.

1. Các định lý về phân phối chuẩn.

1.1 Định lý 1: Nếu biến ngẫu nhiên X có phân phối chuẩn với kỳ vọng μ và phương sai σ^2 thì $Z = \dfrac{X-\mu}{\sigma}$ có phân phối chuẩn tắc.

Một biến ngẫu nhiên X có phân phối bất kỳ với kỳ vọng a và phương sai là b^2 thì biến $Z = \dfrac{X-a}{b}$ gọi là biến đã được chuẩn hoá. Như vậy chuẩn hoá một biến ngẫu nhiên có phân phối chuẩn ta được một biến có phân phối chuẩn tắc.

1.2 Định lý 2: Nếu X và Y là hai biến chuẩn độc lập, $E(X) = \mu_X$, $D(X) = \sigma^2_X$; $E(Y) = \mu_Y$, $D(Y) = \sigma^2_Y$ thì :

1/ $Z = X + Y$ có phân phối chuẩn với kỳ vọng là tổng của hai kỳ vọng và phương sai là tổng của hai phương sai.

2/ $Z = X - Y$ có phân phối chuẩn với kỳ vọng là hiệu của hai kỳ vọng, và phương sai là tổng của hai phương sai.

Tổng, hiệu của hai biến chuẩn độc lập cũng là một biến chuẩn.

1.3 Định lý 3: Nếu $X_1, X_2, ..., X_n$ là dãy các biến chuẩn độc lập có cùng kỳ vọng μ và phương sai σ^2 thì $Z = \dfrac{\overline{X}-\mu}{\sigma}\sqrt{n}$ có phân phối chuẩn tắc.

ở đó $\overline{X} = \dfrac{1}{n}(X_1 + X_2 + ... + X_n)$ là biến trung bình của dãy biến ngẫu nhiên chuẩn $X_1, X_2, ..., X_n$.

1.3 Định lý 4: Nếu $X_1, X_2, ..., X_n$; $Y_1, Y_2, ..., Y_m$ là dãy các biến ngẫu nhiên chuẩn, độc lập, $X_i \sim N(\mu_X, \sigma^2_X)\ \forall\ i = 1, n$; $Y_j \sim N(\mu_Y, \sigma^2_y)\ \forall\ j = 1, m$ thì biến

$$Z = \frac{\overline{X} - \overline{Y} - (\mu_X - \mu_Y)}{\sqrt{\dfrac{\sigma^2_X}{n} + \dfrac{\sigma^2_Y}{m}}}$$ có phân phối chuẩn tắc.

$\overline{X}$ là trung bình của dãy các biến: $X_1, X_2, ..., X_n$

$\overline{Y}$ là trung bình của dãy các biến: $Y_1, Y_2, ..., Y_m$

2.Phân phối khi bình phương χ^2.

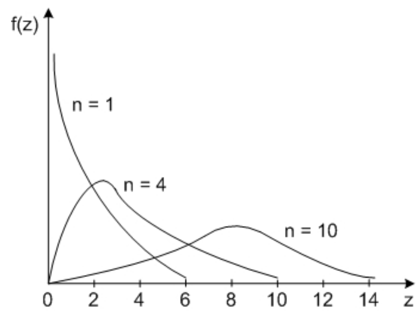
2.1. Định nghĩa: Nếu $X_1, X_2, ..., X_n$ là các biến ngẫu nhiên độc lập có phân phối chuẩn tắc thì biến ngẫu nhiên $Z = \sum_{i=1}^{n} X_i^2$ được gọi là biến ngẫu nhiên có phân phối khi bình phương với n bậc tự do. Ký hiệu $Z \sim \chi^2_n$

2.2 Định lý 1 : Nếu $Z \sim \chi^2_n$ thì hàm mật độ của Z có dạng:

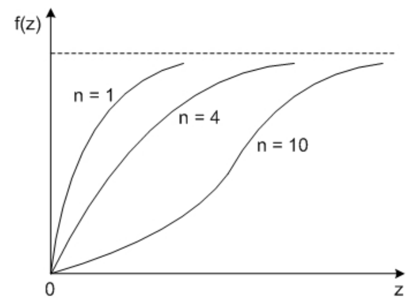
$$f(z) = \begin{cases} 0 & \text{khi } z \le 0 \\ \dfrac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} z^{\frac{n}{2}-1} e^{-\frac{z}{2}} & \text{khi } z > 0 \end{cases},$$

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

2.3 Định lý 2: Nếu $Z \sim \chi^2_n$ thì E(Z) = n và D(Z) = 2n.
Hàm mật độ và hàm phân phối xác suất của phân phối χ^2_n phụ thuộc vào bậc tự do của phân phối . Hai hình sau cho ta đồ thị hàm mật độ và hàm phân phối xác suất khi bình phương ứng với các bậc tự do n = 1, n = 4, n = 10.



Hình 7



Hình 8

2.4 Định lý 3: Nếu $X_1, X_2, \ldots, X_i, \ldots X_n$ là các biến ngẫu nhiên chuẩn, độc lập, $X_i \sim N(\mu, \sigma^2)$ thì biến $Z = \frac{n-1}{\sigma^2}S^2$ là biến ngẫu nhiên có phân phối χ^2_{n-1}

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2, \quad \overline{X} = \frac{1}{n}(X_1 + X_2 + \ldots + X_n)$$

Khi coi vec tơ ngẫu nhiên $(X_1, X_2, \ldots, X_n)$ là mẫu ngẫu nhiên của biến X ở tổng thể có phân phối chuẩn với kỳ vọng μ và phương sai σ^2 thì S^2 là phương sai mẫu đã hiệu chỉnh và $\overline{X}$ là trung bình mẫu.

3.Phân phối Student

3.1 Định nghĩa: Nếu X và Y là hai biến ngẫu nhiên độc lập , X có phân phối chuẩn tắc, Y có phân phối χ^2_n thì $Z = \frac{X}{\sqrt{Y}}\sqrt{n}$ là biến ngẫu nhiên có phân phối Student với n bậc tự do.

Khi đó ta ký hiệu $Z \sim S_n$ hoặc $Z \sim T_n$.

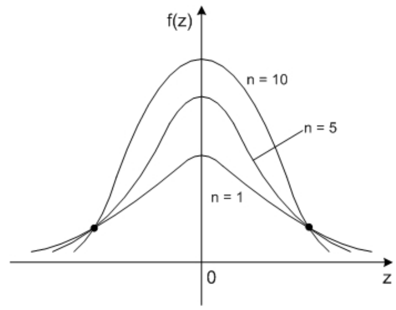
3.2 Định lý 1: Nếu $Z \sim T_n$ thì hàm mật độ của Z là

$$f(z) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\sqrt{\pi n}} \frac{1}{\left(1 + \frac{t^2}{n}\right)^{\frac{n+1}{2}}}$$

3.3 *Định lý 2:* Nếu $Z \sim T_1$ thì Z không có kỳ vọng và phương sai .
Nếu $Z \sim T_2$ thì E(Z) = 0, D(Z) không tồn tại.
Nếu $Z \sim T_n$ với n > 2 thì E(Z) = 0 , $D(Z) = \frac{n}{n-2}$.
Do hàm mật độ của phân phối T_n là hàm chẵn nên nếu $Z \sim T_n$ thì

$P[|Z| \geq \alpha] = 2P[0 \leq Z \leq \alpha], \forall \alpha > 0$.

Hàm mật độ và hàm phân phối T_n cũng phụ thuộc vào bậc tự do n. Hình vẽ sau là đồ thị của hàm mật độ phân phối T_n trong các trường hợp n = 1, n = 5 và n = 10.



Hình 9

3.4 Định lý 3: Nếu $X_1, X_2, ..., X_n$ là các biến ngẫu nhiên độc lập, X_i có phân phối chuẩn tắc với kỳ vọng μ và phương sai σ^2 thì biến ngẫu nhiên

$Z = \dfrac{\overline{X} - \mu}{S}\sqrt{n}$ có phân phối Student với n - 1 bậc tự do.

3.5 Định lý 4: Nếu $X_1, X_2, ..., X_n$; $Y_1, Y_2, ..., Y_m$ là dãy các biến ngẫu nhiên độc lập, $E(X_i) = \mu_X$, $\forall$ i= 1, n; $E(Y_j) = \mu_Y$, $\forall$ j = 1, m,

$D(X_i) = D(Y_j)$, $\forall$ i= 1, n, $\forall$ j = 1, m thì biến ngẫu nhiên $Z = \dfrac{\overline{X} - \overline{Y} - (\mu_X - \mu_Y)}{S\sqrt{\dfrac{1}{n} + \dfrac{1}{m}}}$ có

phân phối Student với n + m - 2 bậc tự do.

$\overline{X} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$ là biến ngẫu nhiên trung bình của dãy các biến:

$X_1, X_2, ...,X_i …, X_n$

$\overline{Y} = \dfrac{1}{m}\sum_{j=1}^{m} Y_j$ là biến ngẫu nhiên trung bình của dãy các biến

$Y_1, Y_2, ...,Y_m$

$$S^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2 + \sum_{j=1}^{m}(Y_j - \overline{Y})^2}{n + m - 2}, S = \sqrt{S^2}$$

4.Phân phối Fisher- Snedecor.

4.1 Định nghĩa: Nếu X và Y là hai biến ngẫu nhiên độc lập.

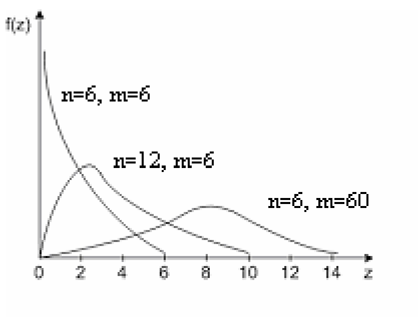
$X \sim \chi^2_n$; $Y \sim \chi^2_m$ thì $Z = \dfrac{X}{n}.\dfrac{m}{Y}$ được gọi là biến ngẫu nhiên có phân phối Fisher- Snedecor với n, m bậc tự do. Kí hiệu $Z \sim F_{n,m}$.

4.2 Định lý 1: Nếu $Z \sim F_{n,m}$ thì hàm mật độ của Z là :

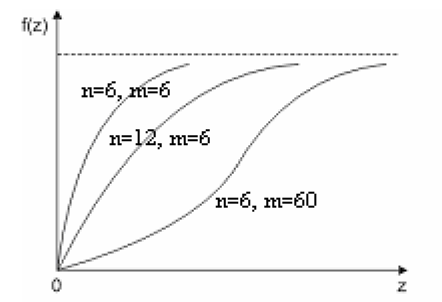
$$f(z)=\begin{cases}0 \quad \text{nếu} \quad z\le 0\\ \dfrac{\Gamma\left(\frac{n+m}{2}\right)\left(\frac{n}{m}\right)^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{m}{2}\right)}.z^{\frac{n-2}{2}}\left(1+\frac{n}{m}z\right)^{-\frac{n+m}{2}} \text{ nếu } z>0\end{cases}$$

4.3 Định lý 2: Với $m > 4$, $E(Z) = \dfrac{m}{m-2}, D(Z)=\dfrac{2m^2(n+m-2)}{n(m-4)(m-2)^2}$

Khi $m \le 2$, E(Z) và D(Z) không tồn tại, khi $2\le m <4$ thì $E(Z) = \dfrac{m}{m-2}$ còn D(Z) không tồn tại. Hàm mật độ và hàm phân phối của phân phối $F_{n,m}$ phụ thuộc vào n và m. Hình vẽ sau cho hàm mật độ của phân phối Fisher - Senedecor trong các trường hợp n = 6, m = 6; n = 12, m = 6; n = 6 ,m = 60.



Hình 10



Hình 11

4.4 Định lý 3: Nếu $Z \sim F_{n,m}$ thì $\dfrac{1}{Z} \sim F_{m,n}$.

4.5 Định lý 4: Nếu $X_1, X_2, ..., X_n$; $Y_1, Y_2, ..., Y_m$ là dãy các biến ngẫu nhiên chuẩn độc lập,

$X_i \sim N(\mu_X;\sigma_X^2)\ \forall\, i = 1, n$; $Y_j \sim N(\mu_Y;\sigma_Y^2)\ \forall\, j = 1, m$

$$S_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2,\ S_Y^2=\frac{1}{m-1}\sum_{j=1}^{m}(Y_i-\overline{Y})^2$$ thì biến ngẫu nhiên

$Z = \dfrac{S_X^2}{S_Y^2}$ có phân phối Fisher - Senedecor với n-1, m-1 bậc tự do.

5. Phân vị mức 1-α.

Cho X là biến ngẫu nhiên có hàm phân phối F(x), α là một số thực thuộc khoảng (0, 1) . Phân vị mức 1-α của biến ngẫu nhiên X là một số X_α thoả mãn

$$P(X<X_\alpha)\le 1-\alpha \le P(X\le X_\alpha)$$

Từ tính chất của hàm phân phối ta có:

$F(X_\alpha)\le 1-\alpha\le F(X_\alpha^{+0})$,vì lý do đó X_α cũng là phân vị mức 1-α của hàm phân phối F(x). $X_{0,5}$ chính là trung vị của biến ngẫu nhiên X.

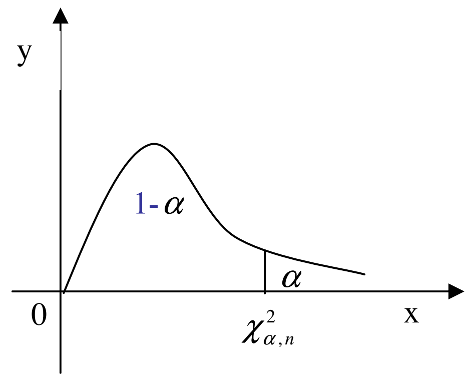
Nhận thấy rằng nếu X là biến ngẫu nhiên liên tục với hàm mật độ f(x) thì

$$F((X_\alpha) = \int_{-\infty}^{X_\alpha} f(x)dx = 1 - \alpha$$

Nếu biến ngẫu nhiên X có phân phối chuẩn tắc thì phân vị mức 1-α của X được kí hiệu là U_α hay U(α).

Nếu biến ngẫu nhiên X có phân phối khi bình phương với n bậc tự do thì phân vị mức1-α của X được kí hiệu là $\chi^2_{\alpha,n}$ hoặc $\chi^2(\alpha, n)$

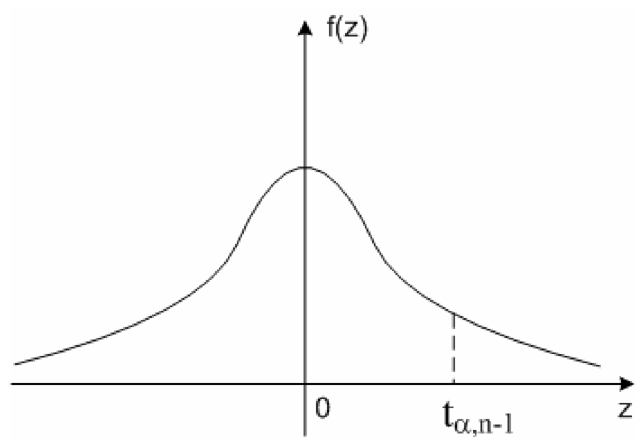
Chú ý 1: Một số tác giả dùng kí hiệu: $\chi^2_{n,\alpha}$ hoặc $\chi^2(\alpha, n)$



Hình 12

Nếu biến ngẫu nhiên X có phân phối Student với n bậc tự do thì phân vị mức 1-α của X được kí hiệu là $t_{\alpha,n}$ hoặc t(α, n)

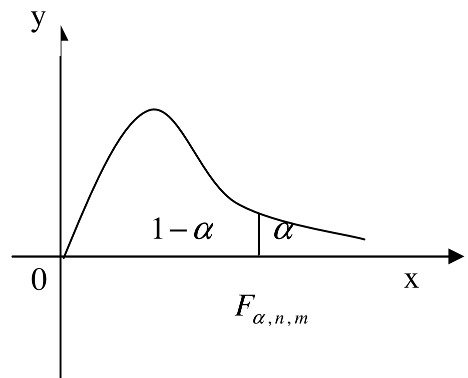
Chú ý 2: Một số tác giả dùng kí hiệu: $t_{n,\alpha}$ hoặc $t(n, \alpha)$



Hình 13

Nếu biến ngẫu nhiên X có phân phối Fisher -Snederco thì phân vị mức 1-α của X được kí hiệu là $F_{\alpha,n,m}$ hoặc F(α, n, m)

Chú ý 3: Một số tác giả dùng kí hiệu: $F_{n,m,\alpha}$ hoặc $F(n, m, \alpha)$



Hình 14

6.Giới thiệu cách tra bảng trong phụ lục

Để tìm U_α ta sử dụng bảng 2 là bảng cho giá trị của hàm Laplace.

* $\forall\alpha \in (0, 1)$, nhìn trong bảng giá trị của hàm $\phi(x)$ xem giá trị nào gần với $1-\alpha$ nhất.

* Hàng chứa giá trị này chỉ phần đơn vị và phần mười của U_α.

* Cột chứa giá trị này chỉ phần trăm của U_α.

Ví dụ 1 : Với $\alpha = 0{,}01$, giá trị gần $1-\alpha = 0{,}99$ nhất trong bảng giá trị hàm $\phi(x)$ là 0,99010 . Giá trị này nằm ở hàng 2,3 và cột 3. Vậy $U_{0,01} = 2{,}33$.

Với $\alpha = 0{,}025$, giá trị gần $1-\alpha = 0{,}975$ này nằm ở hàng 1,9 và cột 6. Vậy $U_{0,025} = 1{,}96$.

Để tìm $\chi^2_{\alpha,n}$ ta tra bảng 3.

* $\chi^2_{\alpha,n}$ là số nằm trên hàng n và cột α.

Ví dụ 2 : $\chi_{0,05,10} = 18{,}31$ là phân vị mức $1-\alpha = 0{,}95$ của phân phối χ^2_{10}

$\chi^2_{0,975,18} = 8{,}33$ là phân vị mức 0,025 của phân phối χ^2_{18}

Để tìm $t_{\alpha,n}$ tra bảng 4.

* $t_{\alpha,n}$ là số nằm trên hàng n và cột có giá tri là 2α.

Ví dụ 3: $t_{0,025,12} = 2{,}179$ là phân vị mức $1-\alpha = 0{,}975$ của phân phối T_{12}

$t_{0,01,30} = 2{,}457{,}33$ là phân vị mức $1-\alpha = 0{,}99$ của phân phối T_{30}.

Để tìm $F_{\alpha,n,m}$ ta tra bảng 5.1 và bảng 5.2

Nếu $\alpha = 0{,}05$ ta tra bảng 5.1. Giá trị $F_{n,m,0,05}$ là giao của cột n và hàng m trong bảng 5.1

Ví dụ 4: $F_{0,05,8,16} = 2{,}54$ là phân vị mức 0,95 của phân phối $F_{8\,,\,16}$

$F_{0,05,14,12} = 2{,}64$ là phân vị mức 0,95 của phân phối $F_{14\,,\,12}$

Nếu $1-\alpha = 0{,}01$ ta tra bảng 5.2 .Cách tra tương tự bảng 5.1.

Ví dụ 5: $F_{0,01,9,16} = 3{,}87$ là phân vị mức 0,99 của phân phối $F_{9\,,\,16}$

$F_{0,01,30,10} = 4{,}25$ là phân vị mức 0,99 của phân phối $F_{30\,,\,10}$.

Bài tập chương IV

1. Điều tra năng suất lúa X (tạ/ha) trên 10 thửa ruộng ta được bảng số liệu sau:

Thửa số	1	2	3	4	5	6	7	8	9	10
X	45,5	46,0	47,0	45,5	47,5	46,0	48,0	48,0	44,0	49,5

a. Lập bảng tần số và tần suất mẫu
b. Lập hàm phân phối mẫu (hàm tần số cộng dồn)
c. Lập biểu đồ hình que tần số và tần suất mẫu
d. Vẽ đồ thị của hàm phân phối mẫu

2. Trọng lượng X của 12 con lợn khi xuất chuồng là (đơn vị kg)
108; 112; 108; 120; 112; 114; 115; 112; 115; 118; 116; 110.
a. Lập hàm tần số và tần suất mẫu.
b. Lập hàm phân phối mẫu
c. Tính trung bình mẫu, phương sai mẫu chưa hiệu chỉnh và đã hiệu chỉnh. Tính độ lệch mẫu chưa hiệu chỉnh và đã hiệu chỉnh.

3. Trọng lượng X của 20 con lợn xuất chuồng cho bởi bảng sau:

X	85-94	95-104	105-114	115- 126	125-134	135-144
n_i	10	30	45	80	30	5

a. Tính trung bình và phương sai mẫu.
b. Lập biểu đồ tần số và tần suất

4. Đo chiều cao X của 100 sinh viên được chọn ngẫu nhiên ta được bảng số liệu sau

X	[1.50, 1.55)	[1.55, 1.60)	[1.60, 1,65)	[1.65, 1.70)	[1.70, 1.75]
n_i	5	18	42	27	8

a. Lập biểu đồ tần số và đa giác tần số.
b.Tính trung bình mẫu và phương sai mẫu theo các đại diện.

5 Một mẫu gồm số tiền mua hàng X (đơn vị USD) của 50 khách hàng tại một siêu thị là:

1	5	22	21	39	32	29	47	14	37
21	14	10	28	26	16	20	35	41	30
36	6	27	18	17	31	25	23	54	10
49	29	24	8	28	12	24	25	3	49
20	23	10	13	45	16	22	58	43	32

a. Xếp mẫu trên vào 6 lớp có bươc nhảy h = 10.
b. Tính trung bình mẫu theo các đại diện.
c. So sánh trung bình tính theo đại diện với các trung bình thực.
d. Tính phương sai mẫu chưa hiệu chỉnh và đã hiệu chỉnh theo các đại diện.

6. Cho Z ~ N(0, 1) Tìm phân vị mức 1- α của Z nếu
a. $1 - \alpha = 9,95$; b. $1 - \alpha = 0,99$; c. $1-\alpha = 0,40$; d. $1 - \alpha = 0,975$

7. Cho $Z \sim \chi^2_n$. Tìm phân vị mức 1 - α của Z nếu

a. $1 - \alpha = 0,95$, $n = 10$; $1 - \alpha = 0,99$; $n = 10$; b. $1 - \alpha = 0,05$; $n = 16$

c. $1 - \alpha = 0,95$; $n = 20$; d. $1 - \alpha = 0,01$; $n = 8$

8. Cho $Z \sim T_n$. Tìm phân vị mức 1 - α của Z nếu

a. $1 - \alpha$ 0,975 ; $n = 14$; b. $1 - \alpha = 0,99$; $n = 20$

c. $1 - \alpha = 0,05$; $n = 16$; d. $1 - \alpha = 0,10$; $n = 19$

9. Cho $Z \sim F_{n,\, m}$. Tìm phân vị mức 0,95 của Z nếu

a. n = 12; m = 14; b. n = 6; m = 15

c. n = 14; m = 24; d. n = 10; m = 6.

10. Cho $Z \sim F_{n,\, m}$. Tìm phân vị mức 0,99 của Z nếu

a. n = 8; m = 6; b. n =12; m =16

c. n = 6; m = 4; d. n =20; m =25.

Chương 5 Ước lượng tham số

Ước lượng tham số là một trong những bài toán cơ bản của thống kê toán học. Khi nghiên cứu đặc tính X của của mỗi cá thể của tổng thể, nếu xác định được qui luật xác suất của X thì việc đưa ra các đánh giá cũng như các dự báo về sự biến động của tổng thể liên quan đến đặc tính này sẽ chính xác và khách quan. Tuy nhiên không phải lúc nào chúng ta cũng xác định được qui luật xác suất của X. Trong một số trường hợp, ta chỉ biết được dạng toán học của hàm phân phối hoặc hàm mật độ của biến định lượng X mà chưa biết các tham số có mặt trong chúng. Vì vậy để xác định qui luật xác suất của X trước hết phải đưa ra những đánh giá về các tham số này. Bài toán ước lượng tham số sẽ giúp ta giải quyết vấn đề trên. Giả sử biến định lượng X ở tổng thể có hàm phân phối: $F(x, \theta_1, \theta_2, ..., \theta_k)$ ở đó dạng toán học của hàm phân phối đã biết còn các tham số $\theta_1, \theta_2, ..., \theta_k$ chưa biết. Từ mẫu ngẫu nhiên kích thước n (X_1, X_2, ..., X_n) đưa ra nhận định các tham số trên nhận giá trị nào hoặc các tham số trên nhận giá trị trong một khoảng nào là nội dung của bài toán ước lượng tham số.

I. Ước lượng điểm

1. Định nghĩa: Nếu lấy thống kê $\hat{G} = \hat{G}($ X_1, X_2, ..., X_n) thay cho tham số chưa biết θ thì G được gọi là một ước lượng điểm của θ

Do $\hat{G} = \hat{G}($ X_1, X_2, ..., X_n) là một thống kê nên một ước lượng điểm của tham số θ cũng là một biến ngẫu nhiên .

* Với mẫu cụ thể (x_1, x_2, ..., x_n), $\hat{g} = \hat{G}$ (x_1, x_2, ..., x_n) là một gía trị cụ thể của ước lượng điểm.

* Trong không gian mẫu ngẫu nhiên có thể xác định nhiều thống kê khác nhau vì vậy mỗi tham số θ có nhiều ước lượng điểm.

* Để ước lượng điểm là một thống kê thay thế tốt cho θ ta phải đưa ra các yêu cầu đánh giá tính tốt đó. Khi một thống kê thoả mãn các yêu cầu này ta có thể yên tâm dùng nó làm ước lượng của θ

2. Ước lượng không chệch: Thống kê $\hat{G} = \hat{G}($ X_1, X_2, ..., X_n) được gọi là một ước lượng không chệch của tham số θ nếu E($\hat{G}$) = θ

Ví dụ 1: Giả sử đặc trưng X ở tổng thể có kì vọng E(X) = μ

(X_1, X_2, ..., X_n) là một mẫu ngẫu nhiên. Khi đó thống kê $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ là một ước lượng không chệch của μ .

Thật vậy: Vì X_i có cùng phân phối xác suất với X nên E(X_i) = μ $\forall$ i = 1,n.

Xét: $E(\overline{X}) = E(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \mu$

Ví dụ 2: Giả sử đặc trưng X ở tổng thể có E(X) = μ , D(X) = σ^2 thống kê

$$\hat{S}^2 = \frac{1}{n}\sum(X_i - \overline{X})^2$$ là một ước lượng chệch của σ^2

Ta có $\hat{S}^2 = \frac{1}{n}\sum(X_i - \overline{X})^2 = \frac{1}{n}\sum[(X_i - \mu) - (\overline{X} - \mu)]^2$

$$= \frac{1}{n}[\sum(X_i - \mu)^2 - (\overline{X} - \mu)\sum(2X_i - 2\mu - \overline{X} + \mu)]$$

$$= \frac{1}{n}\sum(X_i - \mu)^2 - \frac{1}{n}(\overline{X} - \mu)[2\sum X_i - n\overline{X} + n\mu)]$$

$$= \frac{1}{n}\sum(X_i - \mu)^2 - \frac{1}{n}(\overline{X} - \mu)[2n\overline{X} - n\overline{X} + n\mu]$$

$$= \frac{1}{n}\sum(X_i - \mu)^2 - (\overline{X} - \mu)^2$$

Xét: $E(\hat{S}^2) = E[\frac{1}{n}\sum(X_i - \mu)^2 - (\overline{X} - \mu)^2] = \frac{1}{n}\sum E(X_i - \mu)^2 - E(\overline{X} - \mu)^2$

Do $E(X_i - \mu)^2 = D(X) = \sigma^2$, $E(\overline{X} - \mu)^2 = \frac{\sigma^2}{n}$ nên $E(\hat{S}^2) = \sigma^2 - \frac{\sigma^2}{n} = \frac{n-1}{n}\sigma^2 \neq \sigma^2$ vậy $\hat{S}$ là ước lượng chệch của $D(X) = \sigma^2$

Ta có $S^2 = \frac{n-1}{n}\hat{S}^2 \Rightarrow E(S^2) = E(\frac{n-1}{n}\hat{S}^2) = \sigma^2$. Vậy S^2 là một ước lượng không chệch của σ^2. Đây chính là lý do để $\hat{S}^2$ và S^2 được gọi là phương sai mẫu chưa hiệu chỉnh và đã hiệu chỉnh.

Một ước lượng không chệch của θ là một ước lượng "tốt" theo nghĩa nó có trung bình lý thuyết bằng tham số cần ước lượng.

3.Ước lượng vững: Thống kê $\hat{G} = \hat{G}(X_1,X_2,\ldots X_n)$ được gọi là một ước lượng vững của tham số θ nếu với mọi $\varepsilon > 0$ cho trước ta có $\lim_{n\to\infty} P\left[|\hat{G} - \theta| < \varepsilon\right] = 1$.

Từ định nghĩa ta nhận thấy: Nếu thống kê $\hat{G}$ là một ước lượng vững của θ thì khi n lớn (kích thước mẫu lớn) sự sai khác giữa $\hat{G}$ và θ là không đáng kể.

Ví dụ 1: Cho đại lượng ngẫu nhiên X với kì vọng $E(X) = \mu$, $D(X) = \sigma^2$ và $(X_1,X_2,\ldots X_n)$ là mẫu ngẫu nhiên khi đó $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ là một ước lượng không chệch và vững của μ

Tính không chệch đã được khẳng định trong ví dụ ở trên.Ta có $D(\overline{X}) = \frac{\sigma^2}{n}$

Với mọi $\varepsilon > 0$, áp dụng bất đẳng thức Trêbưsep với biến $\overline{X}$ ta có:

$P[|\overline{X} - \mu| < \varepsilon] \geq 1 - \frac{\sigma^2}{n\varepsilon^2}$ ta có $\lim_{n\to\infty}(1 - \frac{\sigma^2}{n\varepsilon^2}) = 1, \forall \varepsilon > 0$.

Suy ra $\lim_{n\to\infty} P[|\overline{X} - \mu| < \varepsilon] = 1$.Vậy $\overline{X}$ là ước lượng vững của μ.

Ví dụ 2: Xét một dãy n phép thử độc lập, xác suất xuất hiện sự kiện A ở mỗi phép thử là P(A) = p. Gọi n_A là số lần xuất hiện A trong n lần thử, tần suất $f_n(A) = \frac{n_A}{n}$ là ước lượng không chệch và vững của p = P(A). Khẳng định trên có từ định lý Bernoulli.

4.Ước lượng hiệu quả: Thống kê $\hat{G} = \hat{G}(X_1, X_2, \ldots X_n)$ được gọi là ước lượng hiệu quả của tham số θ nếu:

1/ $\hat{G}$ là một ước lượng không chệch của θ.

2/ Nếu $\hat{H} = \hat{H}(X_1, X_2 \ldots X_n)$ là một ước lượng không chệch bất kỳ của θ thì $E(\hat{G} - \theta)^2 \leq E(\hat{H} - \theta)^2$

Do $E(\hat{G}) = \theta$, $E(\hat{H}) = \theta$ nên $E(\hat{G} - \theta)^2 = D(\hat{G})$, $E(\hat{H} - \theta)^2 = D(\hat{H})$. Vì vậy ước lượng hiệu quả còn được gọi là ước lượng không chệch với phương sai bé nhất của θ.
Giả sử biến X của tổng thể có hàm mật độ xác suất $f(x, \theta)$ với tham số θ có thể nhận giá trị trong một miền nào đó. $\hat{G} = \hat{G}(X_1, X_2, \ldots X_n)$ là một ước lượng không chệch của θ ta luôn có:

$$E(\hat{G} - \theta)^2 = D(\hat{G}) \geq \frac{1}{nE\left[\frac{\partial \ln f(x,\theta)}{\partial \theta}\right]^2} \quad (*).$$

Nếu $D(\hat{G}) = \frac{1}{nE\left[\frac{\partial \ln f(x,\theta)}{\partial \theta}\right]^2}$ thì $\hat{G}$ là ước lượng hiệu quả của θ.

Bất đẳng thức (*) gọi là bất đẳng thức Cramer – Rao. Ta có thể chứng minh được rằng nếu biến X của tổng thể có phân phối chuẩn với kì vọng μ và phương sai là σ^2 thì $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ là một ước lượng hiệu quả của μ. Tần suất $f_n(A)$ trong một lược đồ Bernoulli cũng là một ước lượng hiệu quả của xác suất P(A) = p.
Có nhiều phương pháp tìm ước lượng điểm của tham số chưa biết có mặt trong quy luật phân phối xác suất của tổng thể. Trong giáo trình này trình bày hai phương pháp tìm ước lượng điểm của tham số đó là :
" Phương pháp hợp lý nhất" và " Phương pháp mô men".
Đây là hai phương pháp thông dụng để tìm ước lượng điểm của tham số.

5.Ước lượng điểm theo phương pháp hợp lý nhất.

Ước lượng điểm của tham số chưa biết θ theo phương pháp hợp lý nhất được dựa trên quan điểm "giá trị của θ trong thực tế chính là giá trị ứng với xác suất xảy ra lớn nhất".
Khi biến ngẫu nhiên X ở tổng thể có hàm mật độ xác suất $f(x, \theta)$ thì hàm mật độ xác suất đồng thời của mẫu ngẫu nhiên $(X_1, X_2, \ldots X_n)$ là

$$\varphi(x_1, x_2, \ldots x_n, \theta) = f(x_1, \theta).\, f(x_2, \theta). \ldots f(x_n, \theta) \quad (1)$$

Hàm $f(x_i, \theta)$ là hàm mật độ xác suất của thành phần X_i trong mẫu ngẫu nhiên: $(X_1, X_2, \ldots X_i \ldots X_n)$

Khi X là biến ngẫu nhiên rời rạc nhận giá trị xác suất $P(X = \alpha_j) = P_j(\theta)$ thì xác suất để $X_i = x_{ji}$ với $x_{ji} \in \{\alpha_1, \alpha_2, \ldots \alpha_j \ldots \alpha_n\}$ là $P(X_i = x_{ji}) = P_{ji}(\theta)$.

Xác suất để $(X_1,X_2 \ldots X_i \ldots X_n) = (x_{j1}, x_{j2}, \ldots, x_{ji}, \ldots x_{jn})$ là

$$\varphi(x_{j1}, x_{j2}, \ldots, x_{ji}, \ldots, x_{jn}) = p_{j1}(\theta)p_{j2}(\theta)\ldots p_{ji}(\theta)\ldots p_{jn}(\theta) \qquad (2)$$

Việc tìm một thống kê $\hat{G} = \hat{G}(X_1,X_2,\ldots X_i,\ldots,X_n)$ thay cho θ sao cho với mẫu cụ thể $(x_1, x_2, \ldots, x_i, \ldots x_n)$ đã cho $\hat{g} = \hat{G}(x_1, x_2, \ldots, x_i, \ldots, x_n)$ thoả mãn

$$f(x_1, \hat{g})\, f(x_2, \hat{g}) \ldots f(x_n, \hat{g}) \geq f(x_1,\theta) f(x_2,\theta) \ldots f(x_n,\theta) \qquad (3)$$

hoặc: $p_{j1}(\hat{g})p_{j2}(\hat{g})\ldots p_{ji}(\hat{g})\ldots p_{jn}(\hat{g}) \geq p_{j1}(\theta)p_{j2}(\theta)\ldots p_{ji}(\theta)\ldots p_{jn}(\theta)$ (4)

với mọi θ thuộc một miền thực nào đó là hợp lý theo quan điểm nêu trên. Trong trường hợp biến ngẫu nhiên X ở tổng thể là liên tục ta lập hàm:

$$\varphi(X_1,X_2,\ldots X_n,\theta) = f(X_1,\theta)f(X_2,\theta)\ldots f(X_n,\theta) \qquad (5)$$

Hàm này được gọi là hàm hợp lý mẫu. Việc tìm điểm cực đại của (5) tương đương với tìm điểm cực đại của:

$$\ln\varphi(X_1,X_2,\ldots X_n,\theta) = \sum_{i=1}^{n} \ln f(X_i,\theta) \quad (6)$$

Thống kê $\hat{G} = \hat{G}(X_1,X_2,\ldots X_n)$ để (6) cực đại thoả mãn phương trình

$$\frac{d\ln\varphi}{d\theta} = 0 \quad (7).$$

Phương trình (7) được gọi là phương trình hợp lý, mọi nghiệm của phương trình này để (6) cực đại là ước lượng hợp lý nhất của θ.

Ví dụ 1: Giả sử X ở tổng thể có phân phối chuẩn $N(\mu;\sigma^2)$ với σ^2 đã biết. . Ta tìm ước lượng điểm của μ

Xét hàm hợp lý mẫu: $\varphi(X_1,X_2,\ldots X_n, \mu) = \dfrac{1}{(\sigma\sqrt{2\pi})^n} e^{-\frac{1}{2\sigma^2}\sum (X_i-\mu)^2}$

$$\Rightarrow \ln\varphi = n\ln\frac{1}{\sigma\sqrt{2\pi}} - \frac{1}{2\sigma^2}\sum (X_i-\mu)^2$$

Phương trình hợp lý $\dfrac{d\ln\varphi}{d\mu} = \dfrac{1}{\sigma^2}\sum (X_i-\mu) = 0$

$$\Leftrightarrow \sum (X_i-\mu) = 0 \Leftrightarrow \mu = \frac{1}{n}\sum X_i = \overline{X}$$

Vậy $\hat{G} = \overline{X}$. Ta có $\dfrac{d^2\ln\varphi}{d\mu^2} = -\dfrac{n}{\sigma^2} < 0$. Vậy tại $\mu = \overline{X}$ hàm hợp lý đạt cực đại, $\overline{X}$ là ước lượng hợp lý nhất của μ.

Trường hợp đại lượng ngẫu nhiên X ở tổng thể là rời rạc ta lập hàm :

$\varphi(X_1,X_2,\ldots X_n,\theta) = p_{j1}(\theta)p_{j2}(\theta)\ldots p_{jn}(\theta)$ hàm này cũng được gọi là hàm hợp lý mẫu. Cũng như trên $\varphi(X_1,X_2,\ldots X_n,\theta)$ cực đại khi và chỉ khi $\ln\varphi = \sum_{i=1}^{n}\ln p_{j_i}(\theta)$ cực đại tương đương với θ là nghiệm của phương trình

$$\frac{d\ln\varphi}{d\theta} = \sum \frac{d\ln p_{j_i}(\theta)}{d\theta} = 0 \ (8)$$

Phương trình (8) cũng gọi là phương trình hợp lý

Ví dụ 2: Xác suất xuất hiện sự kiện A ở một phép thử độc lập P(A) = p. Tiến hành n phép thử độc lập. gọi X là số lần xuất hiện A trong n lần, X_i là số lần xuất hiện A ở lần thứ i, ta có: $X = \sum_{i=1}^{n} X_i$. Hàm hợp lý

$$\varphi(X_1, X_2, \ldots X_n, p) = C_n^x p^x (1-p)^{n-x} \Rightarrow \ln\varphi = \ln C_n^x + x\ln p + (n-x)\ln(1-p)$$

$$\frac{d\ln\varphi}{dp} = \frac{X}{p} - \frac{n-X}{1-p} = 0 \Leftrightarrow p = \frac{X}{n} = f_n(A),$$

$f_n(A)$ là tần suất xuất hiện A trong n phép thử. Vậy hàm ước lượng hợp lý nhất của xác suất p chính là tần suất $f_n(A)$

6. Ước lượng tham số theo phương pháp mômen

Giả sử đặc trưng X ở tổng thể có hàm phân phối xác suất $F(x, \theta_1, \theta_2, ..., \theta_m)$ ở đó kiểu dạng của hàm phân phối đã biết, các tham số $\theta_1, \theta_2, ..., \theta_m$ chưa biết. Từ một mẫu ngẫu nhiên $X_1, X_2, ..., X_n$ ta sẽ tìm các ước lượng điểm của các tham số $\theta_1, \theta_2, ..., \theta_m$

Ta có : $\mu_k(0) = EX^k = g_k(\theta_1, \theta_2, ..., \theta_m)$ là mômen gốc cấp k của X.

Thay $\mu_i(0)$ bởi các mômen gốc của mẫu tương ứng có hệ phương rình sau:

$$\begin{cases} g_1(\theta_1, \theta_2, ..., \theta_m) = \frac{1}{n}\sum X_i \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ g_k(\theta_1, \theta_2, ..., \theta_m) = \frac{1}{n}\sum X_i^k \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ g_m(\theta_1, \theta_2, ..., \theta_m) = \frac{1}{n}\sum X_i^m \end{cases}$$

Các nghiệm của hệ phương trình trên là các ước lượng của các tham số $\theta_1, \theta_2, ..., \theta_m$.

Phương pháp tìm các ước lượng điểm vừa nêu là phương pháp mômen.

Ví dụ: Giả sử đặc trưng X của tổng thể có phân phối chuẩn N($\mu;\sigma^2$) với ($X_1, X_2, ..., X_n$)là một mẫu ngẫu nhiên tương ứng. Ta biết $\mu_1(0) = \mu$, $\mu_2(0) = \mu^2 + \sigma^2$.

Mômen gốc mẫu cấp 1 là $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

Mômen gốc mẫu cấp 2 là $\overline{X^2} = \frac{1}{n}\sum_{i=1}^{n} X_i^2$

Từ đây ta có hệ phương trình sau :

$$\begin{cases} \mu = \overline{X} \\ \mu^2 + \sigma^2 = \overline{X^2} \end{cases} \Rightarrow \begin{cases} \mu = \overline{X} \\ \sigma^2 = \overline{X^2} - \overline{X}^2 = \hat{S}^2 \end{cases}$$

Vậy $\overline{X}$ là là ước lượng điểm của μ , $\hat{S}^2$ là ước lượng điểm của σ^2 tìm được theo phương pháp mômen.

II. Ước lượng khoảng

Trong thực tế nhiều khi ta muốn đánh giá một tham số chưa biết θ nhận giá trị trong một khoảng nào đó. Bài toán ước lượng khoảng giải quyết vấn đề này

1.Khoảng tin cậy. Độ tin cậy

Giả sử $\hat{G}_1 = \hat{G}_1$ $(X_1, X_2, \ldots X_n)$, $\hat{G}_2 = \hat{G}_2$ $(X_1, X_2, \ldots X_n)$ là hai thống kê có từ mẫu ngẫu nhiên $(X_1, X_2, \ldots, X_n)$, θ là một tham số có mặt trong phân phối xác suất của tổng thể, P là số thoả mãn $0 < P < 1$. Khoảng $[\hat{G}_1, \hat{G}_2]$ thoả mãn $P[\hat{G}_1 \le \theta \le \hat{G}_2] = P$ được gọi là khoảng tin cậy của θ với độ tin cậy P

$\hat{G}_2 - \hat{G}_1$ được gọi là bề rộng của khoảng tin cậy

Mong muốn của người làm thống kê là với mẫu đã cho tìm được khoảng tin cậy $[\hat{G}_1, \hat{G}_2]$ sao cho

* Bề rộng của khoảng tin cậy càng nhỏ càng tốt
* Độ tin cậy P càng lớn càng tốt

Tuy nhiên với kích thước mẫu cố định khi bề rộng của khoảng tin cậy giảm thì độ tin cậy P cũng giảm theo và ngược lại. Vì vậy trong thống kê người ta thường cố định độ tin cậy P và tìm một khoảng tin cậy $[\hat{G}_1, \hat{G}_2]$ ứng với độ tin cậy này sao cho nó có bề rộng càng nhỏ càng tốt. Thông thường, người ta chọn độ tin cậy P ở các mức:

$$P = 0{,}95; \quad P = 0{,}99; \quad P = 0{,}999$$

Để tìm $\hat{G}_1$ và $\hat{G}_2$ ứng với độ tin cậy P ta thực hiện theo các bước sau:

* Lập thống kê $\hat{G} = \hat{G}(X_1, X_2, \ldots, X_n, \theta)$ sao cho phân phối xác suất của $\hat{G}$ xác định
*Từ phân phối xác suất của G tìm cặp số a, b thoả mãn:
* $P(a \le \hat{G} \le b) = P$ (1), có thể tìm được vô số cặp a, b thoả mãn (1)
*Thực hiện phép biến đổi tương đương đưa (1) về:
* $P[\hat{G}_1 \le \theta \le \hat{G}_2] = P$ (2)

Khi đó khoảng $[\hat{G}_1, \hat{G}_2]$ là khoảng tin cậy của θ với độ tin cậy P. Trong tất cả các cặp số a, b thoả mãn (1), nếu tìm được a_1, b_1 sao cho $b_1 - a_1$ nhỏ nhất thì khoảng $[\hat{G}_1, \hat{G}_2]$ tìm được ứng với cặp số a_1, b_1 này cũng có bề rộng là nhỏ nhất.

2.Ước lượng kì vọng của phân phối chuẩn: Giả sử $X \sim N(\mu; \sigma^2)$, với mẫu ngẫu nhiên $(X_1, X_2, \ldots X_i, \ldots, X_n)$ và với độ tin cậy P đã cho ta tìm ước lượng khoảng của kì vọng μ.

2.1 Trường hợp σ^2 đã biết

Xét thống kê $Z = \dfrac{\overline{X} - \mu}{\sigma}\sqrt{n} \Rightarrow Z \sim N(0,1)$.

Với độ tin cậy P đã cho có thể tìm được vô số cặp a, b để $P(a \le Z \le b) = P$. Do phân phối chuẩn tắc là phân phối đối xứng, nếu lấy $b = U_{\frac{\alpha}{2}}$, $a = -U_{\frac{\alpha}{2}}$ với $\alpha = 1 - P$ thì $b - a = 2U_{\frac{\alpha}{2}}$ là nhỏ nhất. Khi đó ta có

$$P[-U_{\frac{\alpha}{2}} \le \frac{\overline{X} - \mu}{\sigma}\sqrt{n} \le U_{\frac{\alpha}{2}}] = P \Leftrightarrow P[\overline{X} - U_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \le \mu \le \overline{X} + U_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}] = P$$

Vậy khoảng tin cậy của kì vọng μ khi σ^2 đã biết là:

$$[\overline{X} - U_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}};\ \overline{X} + U_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}] \quad (3).$$

Chú ý: với mẫu cụ thể ($x_1, x_2, \ldots, x_n$) thì $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ là giá trị cụ thể của $\overline{X}$ ứng với mẫu đã cho, khi đó khoảng:

$$[\overline{x} - U_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}};\ \overline{x} + U_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}]$$ (3') cũng được gọi là khoảng tin cậy của kỳ vọng μ tương ứng với mẫu ($x_1, x_2, \ldots, x_n$).

Ví dụ: Giả sử biến X ở tổng thể có phân phối chuẩn với kì vọng chưa biết μ và phương sai $\sigma^2 = 4$. Từ một mẫu có kích thước n = 16 ta tìm được $\overline{x} = 12$.
Với độ tin cậy P = 0,95, hãy tìm ước lượng khoảng của μ.
Giải: ta có: 1 - P = 0,05 = α ; $U_{0,025} = 1{,}96$.

$$\overline{x} - U_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} = 12 - 1{,}96.\frac{1}{2} = 11{,}02;\ \overline{x} + U_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} = 12 + 1{,}96.\frac{1}{2} = 12{,}98.$$

Vậy khoảng tin cậy của μ với độ tin cậy P = 0,95 là [11,02; 12,98]

2.2 Trường hợp σ^2 chưa biết.

Xét thống kê $Z = \frac{\overline{X} - \mu}{S}\sqrt{n} \Rightarrow Z \sim T_{n-1}$ ở đó $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$. Với độ tin cậy P đã cho ta cùng tìm được vô số cặp số a, b để P [$a \le Z \le b$] = P. Do phân phối T_{n-1} là phân phối đối xứng nếu lấy $b = t_{\frac{\alpha}{2},n-1}$ và $a = -t_{\frac{\alpha}{2},n-1}$ thì $b - a = 2t_{\frac{\alpha}{2},n-1}$ nhỏ nhất, khi đó :

$$P[-t_{\frac{\alpha}{2},n-1} \le \frac{\overline{X} - \mu}{S}\sqrt{n} \le t_{\frac{\alpha}{2},n-1}] = P$$

$$\Leftrightarrow P[\overline{X} - t_{\frac{\alpha}{2},n-1}\frac{S}{\sqrt{n}} \le \mu \le \overline{X} + t_{\frac{\alpha}{2},n-1}\frac{S}{\sqrt{n}}] = P$$

Vậy khoảng tin cậy của kì vọng μ trong trường hợp phương sai chưa biết là:

$$[\overline{X} - t_{\frac{\alpha}{2},n-1}\frac{S}{\sqrt{n}};\ \overline{X} + t_{\frac{\alpha}{2},n-1}\frac{S}{\sqrt{n}}] \qquad (4)$$

Với mẫu cụ thể ($x_1, x_2, \ldots, x_n$) thì khoảng:

$$[\overline{x} - t_{\frac{\alpha}{2},n-1}\frac{s}{\sqrt{n}};\ \overline{x} + t_{\frac{\alpha}{2},n-1}\frac{s}{\sqrt{n}}]$$ (4') cũng gọi là khoảng tin cậy của μ khi σ^2 chưa biết.

Ví dụ: điều tra năng suất của một giống lúa trên 10 thửa ruộng ta có kết qủa sau:

x_i	4,4	4,5	4,6	4,8	4,9	5,0
n_i	1	1	3	3	1	1

Biết năng suất X(tấn /ha) , X ~ N($\mu;\sigma^2$). Với độ tin cậy P = 0,95 tìm ước lượng khoảng của μ.

Ta có: $\bar{x} = \frac{1}{10}(1.4,4+ 1.4.5 +3.4,6 +3.4,8+ 1.4,9+ 1.5,0) = 4,7.$

$s^2 = \frac{1}{9}(1.(-0,3)^2+1.(-0,2)^2+3.(-0,1)^2+3.(0,1)^2+1.0,1^2+0,3^2) = 0,071$

$s = \sqrt{s^2} = 0,266$; $1-P = 0,05$; $t_{0,025,9}, = 2,26$

$\bar{x} - t_{\frac{\alpha}{2},n-1}\frac{s}{\sqrt{n}} = 4,7-2,26.\frac{0,266}{\sqrt{10}} = 4,5$; $\bar{x} + t_{\frac{\alpha}{2},n-1}\frac{s}{\sqrt{n}} = 4,7+ 2,26.\frac{0,266}{\sqrt{10}} = 4,9$

Vậy khoảng tin cậy của năng suất μ với độ tin cậy P = 0,95 là [4,5; 4,9].

3.Ước lượng xác suất

Tiến hành một dãy n phép thử độc lập có n_A lần xuất hiện A. Với độ tin cậy P đã cho hãy ước lượng khoảng của tham số p = p(A).

Ta có tần suất xuất hiện A là $f = \frac{n_A}{n}$. Xét thống kê $Z = \frac{f-p}{\sqrt{p(1-p)}}\sqrt{n}$, Z hội tụ theo quy luật tới phân phối N(0,1). Tương tự như bài toán tìm ước lượng khoảng của kỳ vọng μ của phân phối chuẩn, khi n khálớn thì:

$$P[-U_{\frac{\alpha}{2}} \le \frac{f-p}{\sqrt{p(1-p)}}\sqrt{n} \le U_{\frac{\alpha}{2}}] = P \quad (5).$$

Xét bất đẳng thức: $-U_{\frac{\alpha}{2}} \le \frac{f-p}{\sqrt{p(1-p)}}\sqrt{n} \le U_{\frac{\alpha}{2}} \Leftrightarrow$

$n(f-p) \le p(1-p)\ U^2_{\frac{\alpha}{2}} \Leftrightarrow \quad p_1 \le p \le p_2$ với p_1, p_2 cho bởi

$$p_1 = \frac{nf + \frac{1}{2}U_{\frac{\alpha}{2}}^2 - U_{\frac{\alpha}{2}}\sqrt{nf(1-f) + \frac{1}{4}U_{\frac{\alpha}{2}}^2}}{n + U_{\frac{\alpha}{2}}^2}$$

$$p_2 = \frac{nf + \frac{1}{2}U_{\frac{\alpha}{2}}^2 + U_{\frac{\alpha}{2}}\sqrt{nf(1-f) + \frac{1}{4}U_{\frac{\alpha}{2}}^2}}{n + U_{\frac{\alpha}{2}}^2}$$

Vậy (5) $\Leftrightarrow$ $P(p_2 \le p \le p_2) = P$. Ước lượng khoảng của p là

$$\left[\frac{nf + \frac{1}{2}U_{\frac{\alpha}{2}}^2 - U_{\frac{\alpha}{2}}\sqrt{nf(1-f) + \frac{1}{4}U_{\frac{\alpha}{2}}^2}}{n + U_{\frac{\alpha}{2}}^2} ; \frac{nf + \frac{1}{2}U_{\frac{\alpha}{2}}^2 + U_{\frac{\alpha}{2}}\sqrt{nf(1-f) + \frac{1}{4}U_{\frac{\alpha}{2}}^2}}{n + U_{\frac{\alpha}{2}}^2}\right] \quad (5')$$

Ví dụ: Điều tra 2000 gia đình giáo viên ở các tỉnh đồng bằng Bắc Bộ thấy có1600 gia đình có ít nhất 1 con đang học hoặc đã tốt nghiệp đại học. Với độ tin cậy P = 0,95, hãy tìm ước lượng khoảng tỉ lệ p các gia đình giáo viên có con đang học hoặc đã tốt nghiệp đại học.

Ta có: $f = \frac{1600}{2000} = 0,8$, $1- f = 0,2$; $1- p = 0,05$; $U_{0,025} = 1,96$.

Mút trái của khoảng tin cậy là $\dfrac{1600+\dfrac{1{,}96^2}{2}-1{,}96\sqrt{320+\dfrac{1{,}96^2}{4}}}{2000+1{,}96^2}=0{,}78$

Mút phải của khoảng tin cậy là $\dfrac{1600+\dfrac{1{,}96^2}{2}+1{,}96\sqrt{320+\dfrac{1{,}96^2}{4}}}{2000+1{,}96^2}=0{,}82$

Vậy khoảng tin cậy của tỉ lệ p với độ tin cậy P = 0,95 là [0,78; 0,82].

Việc sử dụng công thức (5) để tìm khoảng tin cậy của p = P(A) khá cồng kềnh và phức tạp. Trong thực hành nếu cỡ mẫu lớn thì thống kê $Z=\dfrac{f-p}{\sqrt{f(1-f)}}\sqrt{n}$ có phân phối xấp xỉ chuẩn. Từ đó ta có khoảng tin cậy của p là:

$$[f - U_{\frac{\alpha}{2}}\sqrt{\frac{f(1-f)}{n}}\ ;\ f + U_{\frac{\alpha}{2}}\sqrt{\frac{f(1-f)}{n}}\ ] \qquad (6)$$

áp dụng (6) để tìm khoảng ước lượng của tỉ lệ p trong ví dụ vừa nêu.

Ta có: $f - U_{0,025}\sqrt{\dfrac{f(1-f)}{n}} = 0{,}8 - 1{,}96\sqrt{\dfrac{0{,}8.0{,}2}{2000}} = 0{,}78$

$f + U_{0,025}\sqrt{\dfrac{f(1-f)}{n}} = 0{,}8 + 1{,}96\sqrt{\dfrac{0{,}8.0{,}2}{2000}} = 0{,}82$

Khi cỡ mẫu n lớn, sử dụng công thức(5), (6) để tìm khoảng tin cậy của p ta có kết quả tương tự nhau.

4.Ước lượng phương sai của phân phối chuẩn.

Xét một mẫu ngẫu nhiên từ tổng thể có phân phối chuẩn N($\mu;\sigma^2$)

Thống kê $Z=\dfrac{(n-1)S^2}{\sigma^2}$ có phân phối χ^2_{n-1} với $S^2=\dfrac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$.

Với độ tin cậy P đã cho có vô số cặp số a, b để $P[a \le \dfrac{(n-1)S^2}{\sigma^2} \le b] = P$ (5)

Lấy a = $\chi^2_{n-1;1-\frac{\alpha}{2}}$; b = $\chi^2_{n-1;\frac{\alpha}{2}}$; α = 1- P, ta có

$$P[\chi^2_{1-\frac{\alpha}{2},n-1} \le \frac{(n-1)S^2}{\sigma^2} \le \chi^2_{\frac{\alpha}{2},n-1}] = P$$

$$\Leftrightarrow P[\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}n-1}} \le \sigma^2 \le \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2},n-1}}] = P \qquad (6)$$

Từ (6) ta có : khoảng tin cậy của σ^2 với độ tin cậy P là:

$$[\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2},n-1}} \le \sigma^2 \le \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2},n-1}}] \qquad (7)$$

Chú ý: $\left[\sqrt{\dfrac{n-1}{\chi^2_{\frac{\alpha}{2},n-1}}}S \le \sigma \le \sqrt{\dfrac{n-1}{\chi^2_{1-\frac{\alpha}{2},n-1}}}S\right]$ (7') là khoảng tin cậy của σ với độ tin cậy P

Với mẫu cụ thể : (x_1, x_2,....x_n) , $s^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$ khoảng

$[\dfrac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2},n-1}} \le \sigma^2 \le \dfrac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2},n-1}}]$ (7'') cũng gọi là khoảng tin cậy của σ^2 với độ tin cậy P ứng với mẫu đã cho.

$\left[\sqrt{\dfrac{n-1}{\chi^2_{\frac{\alpha}{2},n-1}}}s \le \sigma \le \sqrt{\dfrac{n-1}{\chi^2_{1-\frac{\alpha}{2}n-1}}}s\right]$ (7''') là khoảng tin cậy của σ tương ứng với mẫu đã cho.

Ví dụ: Biết tỉ lệ phần trăm X của một nguyên tố vi lượng trong các mẫu đất thuộc châu thổ sông Hồng có phân phối chuẩn N($\mu;\sigma^2$). Người ta tiến hành phân tích 15 mẫu đất và tìm được s^2 = 0,1%. Với độ tin cậy P = 0,95. Tìm khoảng tin cậy của σ^2.

P = 0,95 nên $\alpha = 0{,}05$, $\chi^2_{\ 0,975,14} = 5{,}73$; $\chi^2_{\ 0,025,14} = 26{,}12$

$\dfrac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2},n-1}} = \dfrac{14.0,001}{26,12} = 0{,}000536$; $\dfrac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2},n-1}} = \dfrac{14.0,001}{5,63} = 0{,}002487$

Vậy khoảng tin cậy của σ^2 ứng với mẫu đã cho là [0,000536 ; 0,002487].

5. Ước lượng số cá thể có đặc tính A trong đám đông gồm N cá thể

Đây là bài toán thường gặp trong thực tế, chẳng hạn cần đưa ra ước lượng về số người mắc một loại bệnh trong một khu vực dân cư có N người hoặc cần ước lượng về số phế phẩm trong một kho hàng gồm N sản phẩm,….

Gọi m là số cá thể có đặc tính A trong đám đông gồm N cá thể. Lấy từ đám đông ra một mẫu ngẫu nhiên (không hoàn lại) gồm n cá thể gọi X là số cá thể có đặc tính A trong n cá thể . X có phân phối siêu bội với $E(X) = \dfrac{nM}{N}$ và $D(X) = \dfrac{nM}{N}.\dfrac{N-M}{N}.(1-\dfrac{n-1}{N-1})$.

Biến ngẫu nhiên $Z = \dfrac{X - E(X)}{\sqrt{D(X)}}$ có phân phối giới hạn là chuẩn tắc.

Từ quy luật phân phối giới hạn của Z với độ tin cậy P đã cho ta có thể tìm được khoảng tin cậy của E(X) từ đó suy ra khoảng tin cậy của M. Tuy nhiên việc tìm khoảng tin cậy của E(X) theo phương pháp trên là khá phức tạp nên ta không đi theo hướng này. Ta cũng biết rằng nếu N lớn hơn rất nhiều so với n thì phân phối siêu bội xấp xỉ phân phối nhị thức B(n, p) với $p = \dfrac{M}{N}$. Giả sử n_A là số cá thể có đặc tính A trong n cá thể, $f = \dfrac{n_A}{n}$, khoảng tin cậy của p với độ tin cậy P là :

$$f - U_{\frac{\alpha}{2}}\sqrt{\frac{f(1-f)}{n}} \le p \le f + U_{\frac{\alpha}{2}}\sqrt{\frac{f(1-f)}{n}}$$

$$\Rightarrow f - U_{\frac{\alpha}{2}}\sqrt{\frac{f(1-f)}{n}} \le \frac{M}{N} \le f + U_{\frac{\alpha}{2}}\sqrt{\frac{f(1-f)}{n}}$$

$$N[f - U_{\frac{\alpha}{2}}\sqrt{\frac{f(1-f)}{n}}] \le M \le N[f + U_{\frac{\alpha}{2}}\sqrt{\frac{f(1-f)}{n}}] \qquad (8)$$

Ví dụ : Tại một vùng núi khu vực Tây Nguyên gồm 10000 người. Tiến hành xét nghiệm tìm kí sinh trùng sốt rét của 200 người thấy có 40 người có kí sinh trùng sốt rét trong máu. Hãy tìm khoảng tin cậy của số người có kí sinh trùng sốt rét trong máu với độ tin cậy P = 0,95

Ta có $f = \frac{40}{200} = 0{,}2;\ 1 - f = 0{,}8;\ \sqrt{\frac{f(1-f)}{n}} = 0{,}028; U_{0,025} = 1{,}96$

$$f - U_{\frac{\alpha}{2}}\sqrt{\frac{f(1-f)}{n}} = 0{,}2 - 1{,}96.0{,}028 = 0{,}1451$$

$$f + U_{\frac{\alpha}{2}}\sqrt{\frac{f(1-f)}{n}} = 0{,}2 + 1{,}96.0{,}028 = 0{,}2549$$

$$N[f - U_{\frac{\alpha}{2}}\sqrt{\frac{f(1-f)}{n}}] = 10000.0{,}1451 = 1451$$

$$N[f + U_{\frac{\alpha}{2}}\sqrt{\frac{f(1-f)}{n}}] = 10000.0{,}2549 = 2549$$

Với độ tin cậy P = 0,95 , khoảng tin cậy về số người có kí sinh trùng sốt rét trong nhóm máu ở vùng núi Tây Nguyên là [1451 ; 2549]

6. Ước lượng kích thước tổng thể

Để xác định số lượng N của một loài vật hoang dã hoặc số lượng cá có trong một cái hồ người ta tiến hành theo phương pháp sau:

Bắt M cá thể trong N cá thể, đánh dấu từng cá thể này (chẳng hạn mỗi con thú, mỗi con chim hoặc mỗi con cá bắt được cho gắn với một vòng nhôm), sau đó thả M cá thể vào môi trường mà chúng đã sinh sống. Sau một thời gian ta tìm bắt n cá thể (n < M), từ số cá thể X đã được đánh dấu có mặt trong n cá thể vừa bắt được ta sẽ tìm cách xác định số lượng cá thể N.

Ta biết X là đại lượng ngẫu nhiên có phân phối siêu bội. Dựa vào quy luật của X ta có thể tìm được khoảng tin cậy của $\frac{M}{N}$ với độ tin cậy P, từ đó suy ra khoảng tin cậy của N.

Để đơn giản ta xấp xỉ phân phối siêu bội bởi phân phối nhị thức. Khoảng tin cậy của $p = \frac{M}{N}$ với độ tin cậy P là

$$f - U_{\frac{\alpha}{2}}\sqrt{\frac{f(1-f)}{n}} \le \frac{M}{N} \le f + U_{\frac{\alpha}{2}}\sqrt{\frac{f(1-f)}{n}}$$

$$\Rightarrow \frac{M}{f+U_{\frac{\alpha}{2}}\sqrt{\frac{f(1-f)}{n}}} \le N \le \frac{M}{f-U_{\frac{\alpha}{2}}\sqrt{\frac{f(1-f)}{n}}} \qquad (9)$$

Bất đẳng thức cho bởi (9) cho ta khoảng tin cậy của N với độ tin cậy P.

Ví dụ: Để điều tra số lượng một loài cá trong hồ người ta đánh bắt 400 con cá rồi đánh dấu mỗi con cá này bằng một vòng nhôm nhỏ sau đó thả vào hồ. Sau một thời gian người ta bắt lại 150 con thấy trong đó có 50 con đã được bắt ở lần trước. Với độ tin cậy P = 0,95 hãy tìm khoảng tin cậy số lượng n của loài cá này.

Ta có $f = \frac{50}{150} = \frac{1}{3};\ 1-f = \frac{2}{3};\ \ U_{0,025} = 1,96; n = 150; M = 400$

$$f - U_{\frac{\alpha}{2}}\sqrt{\frac{f(1-f)}{n}} = \frac{1}{3} - 1,96.\sqrt{\frac{2}{150.9}} = 0,3246$$

$$f + U_{\frac{\alpha}{2}}\sqrt{\frac{f(1-f)}{n}} = \frac{1}{3} + 1,96.\sqrt{\frac{2}{150.9}} = 0,3420$$

$$\frac{M}{f+U_{\frac{\alpha}{2}}\sqrt{\frac{f(1-f)}{n}}} = \frac{400}{0,3420} \approx 1170$$

$$\frac{M}{f-U_{\frac{\alpha}{2}}\sqrt{\frac{f(1-f)}{n}}} = \frac{400}{0,3246} \approx 1232$$

Kết luận: Với độ tin cậy P = 0,95 ta có thể tin rằng loài cá trên ở trong hồ vào khoảng từ 1170 con đến 1232 con.

7. Kích thước mẫu cần thiết

Ta biết rằng bề rộng của khoảng tin cậy phụ thuộc và độ tin cậy P và kích thước mẫu n. Gọi 2l là bề rộng của khoảng tin cậy, 2l tỉ lệ thuận với P và tỉ lệ nghịch với n.

Bề rộng của khoảng tin cậy thể hiện sự chính xác cao hay thấp của ước lượng, khoảng tin cậy càng hẹp, độ chính xác ước lượng càng cao. Trong thực tế, để đảm bảo tính chính xác của ước lượng ta thường yêu cầu bề rộng của khoảng tin cậy nhỏ thua 2ε với ε là số dương cho trước. Trong mục này chúng ta đưa ra cách tìm kích thước mẫu tối thiểu để bề rộng của khoảng tin cậy $2l \le 2\varepsilon$.

Trong trường hợp σ^2 đã biết, kì vọng μ của phân phối chuẩn có bề rộng khoảng tin cậy

$2l = 2U_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$. Để $2l \le 2\varepsilon \Leftrightarrow U_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \le \varepsilon \Leftrightarrow \sqrt{n} \ge U_{\frac{\alpha}{2}}\frac{\sigma}{\varepsilon} \Leftrightarrow n \ge U^2_{\frac{\alpha}{2}}\frac{\sigma^2}{\varepsilon^2}$ (8)

Chỉ cần lấy n là số tự nhiên nhỏ nhất thoả mãn (8) ta có $2l \le 2\varepsilon$.

Ví dụ 1: Biết X ~ N(μ, 0,16). Hãy tìm kích thước mẫu n để với độ tin cậy P = 0,95 độ rộng của ước lượng khoảng $2l \le 0,2$.Ta có $U_{0.025} = 1,96$,

$\varepsilon = 0,1$, áp dụng (8) $\Rightarrow n \ge 1,96^2.\frac{0,16}{0,01} \Leftrightarrow n \ge 61,4656$. Vậy để $2l \le 0,2$ thì tối thiểu phải tiến hành điều tra 62 mẫu.

Độ rộng của khoảng tin cậy của xác suất p trong phân phối B(n, p) là:

$$2l = 2U_{\frac{\alpha}{2}}\sqrt{\frac{f(1-f)}{n}} \le \frac{U_{\frac{\alpha}{2}}}{\sqrt{n}} \quad \text{vì } f(1-f) \le \frac{1}{4} \quad \text{(Bất đẳng thức Caushy)}$$

Xét: $\frac{U_{\frac{\alpha}{2}}}{\sqrt{n}} \le 2\varepsilon \Leftrightarrow \sqrt{n} \ge \frac{U_{\frac{\alpha}{2}}}{2\varepsilon} \Leftrightarrow n \ge \frac{U^2_{\frac{\alpha}{2}}}{4\varepsilon^2}$ (9)

Vậy để khoảng ước lượng của xác suất p có độ rộng $2l \le 2\varepsilon$ thì số lần thử n phải thoả mãn (9).

Ví dụ 2: Phải tiến hành bao nhiêu phép thử độc lập để với độ tin cậy P = 0,99 bề rộng của khoảng tin cậy của xác suất P(A) = p nhỏ hơn 0,1

Ta có $U_{0.005} = 2{,}26$, $\varepsilon = 0{,}05$. Áp dụng (9) $\Rightarrow n \ge \frac{2{,}26^2}{4.0{,}05^2} = 510{,}76$

Vậy để thoả mãn yêu cầu phải tiến hành tối thiểu 511 phép thử .

Bài tập chương V

1. Cho mẫu ngẫu nhiên ($X_1, X_2, \ldots. X_n$) biết $X_i \sim N(\mu;\sigma^2)$

a. Chứng minh rằng: $Z_1 = X_1$, $Z_2 = \dfrac{X_1 + X_2}{2}$, $Z_i = \dfrac{X_1 + X_2 + \ldots + X_i}{i}$

$Z_n = \dfrac{X_1 + X_2 + \ldots + X_n}{n}$ đều là các ước lượng không chệch của μ.

b. Trong các ước lượng trên ước lượng nào là tốt nhất.

2. Cho mẫu ngẫu nhiên $X_1, X_2, \ldots. X_n$ biết X_i có phân phối mũ với tham số θ hãy chứng minh rằng: $Z = \dfrac{n-1}{n\overline{X}}$ với $\overline{X} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$ là ước lượng không chệch của θ.

3. Cho $X_1, X_2, \ldots. X_n$ là một mẫu ngẫu nhiên lấy từ một tổng thể có kì vọng μ và phương sai σ^2. Xét hai thống kê

$Z_1 = 2\dfrac{X_1 + 2X_2 + \ldots + nX_n}{n(n+1)}, Z_2 = \overline{X} = \dfrac{X_1 + X_2 + \ldots + X_n}{n}$

a. Chứng minh rằng cả hai thống kê trên đều là ước lượng không chệch của μ.

b. Trong hai ước lượng trên ước lượng nào tốt hơn.

4. Năng suất ngô X (tạ /ha) là đại lượng ngẫu nhiên có phân phối chuẩn N($\mu;\sigma^2$). Điều tra năng suất ngô của 130 thửa ruộng ta có kết quả sau:

Năng suất x_i	40	45	46	49	51	53	57
Số thửa n_i	2	5	9	35	43	22	14

a. Hãy tìm khoảng tin cậy của năng suất ngô trung bình μ với độ tin cậy $P = 0{,}95$.

b. Hãy tìm khoảng tin cậy của σ^2 với độ tin cậy $P = 0{,}98$.

5. Biết trọng lượng X (g/ quả) của mỗi quả trứng có phân phối chuẩn N(μ, 25) Cân một mẫu gồm 100 quả trứng ta có kết quả sau:

x_i	150	160	165	170	175	180	185
n_i	4	12	14	25	25	14	6

a. Với độ tin cậy $P = 0{,}95$ tìm khoảng tin cậy của trọng lượng trứng trung bình μ.

b. Trứng có khối lượng lớn hơn 170 g là trứng loại một. Với độ tin cậy $P = 0{,}95$ hãy tìm khoảng tin cậy của tỷ lệ trứng loại một.

6. Biết khối lượng X(kg/con) của mỗi con gà tại một trại gà có phân phối chuẩn N($\mu;\sigma^2$). Bắt ngẫu nhiên 20 con gà đem cân ta có kết quả sau:

x_i	2,1	2,3	2,4	2,6	2,7	2,9	3,1	3,3
n_i	1	2	3	3	5	3	2	1

Với độ tin cậy P = 0,95

a.Tìm khoảng tin cậy của trung bình μ

b.Tìm khoảng tin cậy của phương sai σ^2.

7. Sức chịu nén tối đa của một loại vật liệu là một biến chuẩn $N(\mu;\sigma^2)$. Thử 10 mẫu vật liệu nói trên ta có kết quả sau:

Sức chịu nén tối đa X(kg/cm^2)	250	270	300	330	350
n_i	1	2	4	2	1

Tìm khoảng tin cậy của μ với độ tin cậy P = 0,95.

8. Theo dõi doanh thu hàng tháng của 10 cửa hàng kinh doanh thóc giống tại một tỉnh ta có kết quả sau:

Doanh thu X(triệu đồng)	30	31	33	35	37	39	40
n_i	1	1	2	2	2	1	1

Biết X có phân phối chuẩn $N(\mu;\sigma^2)$. Hãy tìm khoảng tin cậy của μ với độ tin cậy P = 0,98.

9. Trọng lượng X của các gói mì ăn liền tuân theo phân phối chuẩn. Kiểm tra 20 gói mì ta có $\overline{x} = 78,0$, s = 2,5 g. Với độ tin cậy P = 0.95 hãy tìm khoảng tin cậy của E(X).

10. Kiểm tra 1000 mẫu máu một loại gia cầm có 120 mẫu chứa vi rút gây bệnh A. Hãy tìm khoảng tin cậy của tỉ lệ gia cầm chứa vi rút gây bệnh A trong máu với độ tin cậy P = 0,95.

11. Đo độ chịu lực X của 250 mẫu bê tông ta có kết quả sau:

X	180-190	190-200	200-210	210-220	220-230	230-240	240-250	250-260
Tần số n_i	12	15	30	58	65	35	20	15

Biết độ chịu lực X(kg/cm^2) tuân theo qui luật phân phối chuẩn. Hãy tìm khoảng tin cậy của E(X) với độ tin cậy P = 0,95.

12.Đo một đại lượng 15 lần bằng một dụng cụ đo không có sai số hệ thốngta có $s^2 = 0,4$. Biết sai số X có phân phối chuẩn, hãy tìm khoảng tin cậy của phương sai với độ tin cậy P = 0,95.

13. Trọng lượng X của một giống lợn khi xuất chuồng là một biến ngẫu nhiên có phân phối chuẩn. Một mẫu ngẫu nhiên gồm 9 con lợn đến thời gian xuất chuồng có trọng lượng cho bởi bảng sau:

129,8 ; 121,2 ; 138,6 ; 125,4 ; 122,6 ; 139,8 ; 129,9 ; 130,3 ; 125,8

Hãy tìm khoảng tin cậy của trọng lượng trung bình E(X) với độ tin cậy P = 0,95.

14. Để khảo sát mức tiêu thụ xăng trung bình của một loại ô tô người ta cho chạy thử 20 xe loại này trên đoạn đường 100km. Mức xăng tiêu thu tương ứng cho bởi bảng sau:

Mức xăng X	7,5	8,0	8,5	9,0	9,5
Số xe n_i	3	4	6	5	2

Hãy tìm khoảng tin cậy của mức xăng tiêu thụ trung bình với độ tin cậy P = 0,95.

15. Một công ty than có 10000 công nhân làm việc trực tiếp tại các hầm lò. Để xác định số công nhân mắc các bệnh về phổi người ta tiến hành kiểm tra 820 người thấy có 120 người mắc bệnh về phổi. Với độ tin cậy P = 0,95 hãy tìm khoảng tin cậy của số công nhân mắc bệnh về phổi trong tổng công ty.

16. Đểu ước lượng số lượng cò tại một vườn cò lớn ở đồng bằng sông Cửu Long người ta bắt ngẫu nhiên 800 con cò và cho mỗi con đeo một vòng nhôm nhỏ sau đó thả lại vườn. Một tháng sau bắt lại 320 con thấy có 80 con có đeo vòng nhôm. Hãy ước lượng số cò trong vườn với độ tin cậy P = 0,95.

17. Một kho hàng chứa 12000 sản phẩm. Để ước lượng số phế phẩm trong kho hàng người ta kiểm tra 500 sản phẩm thấy có 50 phế phẩm. Hãy ước lượng số phế phẩm trong kho với độ tin cậy P = 0.95.

18. Để ước lượng số người nghiện ma tuý trong một vùng người ta người ta ghi danh 1000 người được trả về cộng đồng sau khi cai nghiện. Một năm sau trở lại các trung tâm cai nghiện chọn ngẫu nhiên 800 người thấy có 480 người trong số 1000 người được trở về cộng đồng năm trước trở lại trại. Hãy ước tính số người nghiện trong vùng với độ tin cậy P = 0,95.

19. Đại lượng ngẫu nhiên X có phân phối chuẩn với phương sai 0,04. Tối thiểu phải điều tra bao nhiêu mẫu để với độ tin cậy P = 0,95 độ rộng của khoảng tin cậy không quá 0,4.

20. Phải kiểm tra ít nhất bao nhiêu mẫu bệnh phẩm để với độ tin cậy P = 0,95 độ rộng của khoảng tin cậy tỉ lệ người mắc bệnh ≤ 0,05.

21 Xét nghiệm 400 mẫu máu của những người dân tại một vùng cao phía bắc thấy 45 mẫu máu có kí sinh trùng sốt rét trong máu. Với độ tin cậy P = 0,95 hãy tìm khoảng tin cậy của tỉ lệ người dân có kí sinh trùng sốt rét trong máu ở vùng cao nói trên.

22. Kiểm tra 200 con gà tại một trại thấy có 80 con mắc bệnh A. Hãy tìm khoảng tin cậy của tỉ lệ gà mắc bệnh A ở trại gà nói trên với độ tin cậy P = 0,95.

23. Biết đặc trưng X có phân phối chuẩn N(μ ; 0,09). Hỏi dung lượng mẫu tối thiểu là bao nhiêu để với độ tin cậy P = 0,95 có thể tin rằng độ rộng của khoảng tin cậy của μ không vượt quá 0,5.

24. Tỉ lệ người có nhóm máu O ở một tộc người là p. Phải kiểm tra ít nhất bao nhiêu người để với độ tin cậy P = 0,95 độ rộng của khoảng tin cậy của p không vượt quá 0,02.

Chương 6: Kiểm định giả thuyết thống kê

Việc xác định qui luật xác suất của các biến có mặt trong tổng thể là một điều cần thiết trong xử lí số liệu. Bài toán ước lượng tham số mới giải quyết việc ước lượng tham tham số có mặt trong phân phối xác suất của tổng thể. Trong chương này chúng ta sẽ xây dựng các qui tắc đánh giá giả thuyết về các tham số, giả thuyết về các qui luật xác suất dựa trên mẫu ngẫu nhiên Qua các qui tắc kiểm định, người học có thể biết được cách xây dựng các giả thuyết và đối thuyết trong từng trường hợp cụ thể. Bài toán kiểm định giả thuyết thống kê là một bài toán lớn và quan trọng của thống kê toán học. Vì thời lượng chương trình có hạn, giáo trình chỉ đề cập tới một số qui tắc kiểm định thông dụng nhất. Một số qui tắc phi tham số giới thiệu trong giáo trình được đơn giản hóa bằng cách thay các thống kê dùng để kiểm định các qui tắc này bởi các qui luật xấp xỉ tương ứng.

I. Giả thuyết - Đối thuyết

1. Giả thuyết: Một mệnh đề (một câu khẳng định) về một vấn đề chưa biết nào đó được gọi là một giả thuyết. Các mệnh đề sau đều được gọi là các giả thuyết:

Vào năm 2010 con người sẽ có mặt trên sao hoả

Tham số $\theta=\theta_0$

Tham số $\theta\in[\theta_1;\theta_2]$

$X \sim N(\mu;\sigma^2)$

Sự kiện A độc lập với sự kiện B

Ta thường dùng H_0 để chỉ một giả thuyết. Giả thuyết là một mệnh đề nên có thể đúng hoặc không đúng. Tuy nhiên để kiểm tra tính đúng của một mệnh đề ta phải dựa trên tiêu chí thế nào là một mệnh đề đúng. Để khẳng định tính đúng sai của một mệnh đề ta thường kiểm tra mệnh đề này có thoả một số yêu cầu nào đó hay không hoặc đưa ra một mệnh đề khác trái với mệnh đề đã cho, trên cơ sở thực tế ta đưa ra quyết định coi mệnh đề ban đầu là đúng hoặc mệnh đề mới đưa ra là đúng. Trong thống kê ta sẽ theo hướng thứ hai.

2. Đối thuyết: Một mệnh đề trái với giả thuyết được gọi là một đối thuyết. Ta thường dùng H_1 để chỉ đối thuyết.

Ví dụ 1: H_0: Vào năm 2010 con người sẽ có mặt trên sao hoả.

Các mệnh đề sau là đối thuyết của giả thuyết H_0

H_1: Vào năm 2020 con người mới có mặt trên sao hoả

H_1: Vào năm 2010 con người chưa thể có mặt trên sao hoả

Ví dụ 2: H_0: $X \sim N(\mu;\sigma^2)$

Các đối thuyết của giả thuyết trên có thể là

H_1: $X \sim B(n, p)$ hoặc H_1: X không có phân phối chuẩn $N(\mu;\sigma^2)$

Nhận xét:

* Giả thuyết H_0 có thể đứng độc lập

* Đối thuyết phải đi kèm với mệnh đề trước đó được gọi là giả thuyết

* Mỗi giả thuyết có thể có nhiều đối thuyết khác nhau

* Một mệnh đề là giả thuyết trong trường hợp này có thể là đối thuyết trong trường hợp khác.

3. Giả thuyết thống kê và đối thuyết thống kê

Những giả thuyết và đối thuyết nói tới tham số có mặt trong qui luật xác suất của các đặc trưng có mặt trong tổng thể hoặc đề cập đến qui luật phân phối xác suất của những đặc trưng này được gọi là các giả thuyết và đối thuyết thống kê.

Ví dụ: H_0 : $X \sim N(\mu;\sigma^2)$

H_1 : $X \sim B(n, p)$

Hoặc H_0 : $\theta = \theta_0$

H_1 : $\theta \neq \theta_1$

là các giả thuyết và đối thuyết thống kê

4. Giả thuyết và đối thuyết tham số

Các giả thuyết và đối thuyết nói về tham số có mặt trong qui luật phân phối xác suất của tổng thể được gọi là các giả thuyết và đối thuyết tham số.

Ví dụ: Biết đặc trưng X ở tổng thể có phân phối chuẩn $N(\mu;\sigma^2)$

H_0 : $\mu=\mu_0$

H_1 : $\mu=\mu_1$ $(\mu_1 \neq \mu_2)$ hoặc: H_1 : $\mu \neq \mu_0$

là các giả thuyết và đối thuyết tham số

4.1 Giả thuyết đơn - Đối thuyết đơn

Giả thuyết đơn là giả thuyết trong đó tham số nhận một giá trị cụ thể nào đó.

Đối thuyết đơn là đối thuyết trong đó tham số nhận một giá trị cụ thể nào đó.

Ví dụ: Biết X~ B(n, p)

H_0 : $p = p_0$ là giả thuyết đơn

H_1 : $p = p_1$; ($p_1 \neq p_0$) là đối thuyết đơn của giả thuyết vừa nêu

4.2 Giả thuyết hợp - Đối thuyết hợp

Các giả thuyết hoặc đối thuyết trong đó tham số nhận hơn một giá trị gọi là giả thuyết hợp và đối thuyết hợp.

Ví dụ: Biết: $X \sim N(\mu;\sigma^2)$

H_0 : $\mu \in [\mu_1;\mu_2]$ là giả thuyết hợp

H_1 : $\mu < \mu_1$ hoặc $\mu > \mu_2$ là các đối thuyết hợp tương ứng với giả thuyết H_0

5. Giả thuyết và đối thuyết phi tham số: Những giả thuyết và đối thuyết thống kê không phải là các giả thuyết và đối thuyết tham số được gọi là các giả thuyết và đối thuyết phi tham số.

Ví dụ: H_0 : $X \sim N(\mu;\sigma^2)$

H_1 : $X \sim B(n, p)$

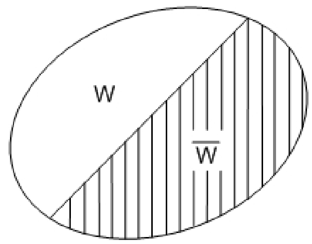
Hoặc: H_0 : A độc lập với B

H_1 : A không độc lập với B

là các giả thuyết và đối thuyết phi tham số

6. Kiểm định giả thuyết thống kê

Từ mẫu đã cho ta xây dựng một qui tắc chấp nhận giả thuyết H_0 (tương ứng với việc bác bỏ đối thuyết H_1) hoặc bác bỏ giả thuyết H_0 (tương ứng với việc chấp nhận đối thuyết H_1) được gọi là bài toán kiểm định một giả thuyết thống kê. Việc đưa ra một qui tắc chấp nhận hoặc bác bỏ giả thuyết H_0 dựa trên mẫu đã cho tương đương với việc xây dựng một qui tắc chia không gian mẫu V ra làm hai phần W và $\overline{W}$



Hình 1

Nếu mẫu ($X_1, X_2, \ldots , X_n$) $\in$ W ta quyết định bác bỏ giả thuyết H_0

Nếu mẫu ($X_1, X_2, \ldots , X_n$) $\in \overline{W}$ ta quyết định chấp nhận giả thuyết H_0

Kiểm định một giả thuyết thống kê không phải là một phép chứng minh về tính đúng hoặc không đúng của giả thuyết. Kiểm định một giả thuyết thống kê thực chất là xây dựng một qui tắc hành động dựa vào mẫu đã có đưa ra quyết định lựa chọn giả thuyết H_0 hoặc đối thuyết H_1

7. Các loại sai lầm: Với một qui tắc hành động chấp nhận hay bác bỏ H_0 ta có thể mắc phải các loại sai lầm sau:

7.1. Sai lầm loại 1: Bác bỏ giả thuyết H_0 khi H_0 đúng

Điều này có nghĩa là giả thuyết H_0 đúng nhưng mẫu ($X_1, X_2, \ldots , X_n$) $\in$ W nên ta bác bỏ H_0. Tương ứng với sai lầm loại 1 là *xác suất sai lầm loại 1:* $P(W/H_0) = \alpha$

7.2. Sai lầm loại 2: Chấp nhận giả thuyết H_0 khi H_0 sai. Điều này cũng có nghĩa là:

Khi H_0 sai (tức là coi H_1 đúng) nhưng mẫu ngẫu nhiên ($X_1, X_2, \ldots , X_n$) $\in \overline{W}$ nên ta chấp nhận H_0. Tương ứng với sai lầm loại 2 là *xác suất sai lầm loại 2:* $P(\overline{W}/H_1) = \beta$.

Ta có nhận xét*:*

Xác suất sai lầm loại 1: $P(W/H_0) = \alpha$ là xác suất bác bỏ H_0 mà thực ra H_0 đúng được gọi là mức ý nghĩa của bài toán kiểm định. Khi đó xác suất để chấp nhận H_0 khi H_0 đúng là $1 - \alpha$.

Xác suất sai lầm loại 2: $P(\overline{W}/H_1) = \beta$ là xác suất chấp nhận H_0 khi H_0 sai. Vậy xác suất bác bỏ H_0 khi H_0 sai là $1 - \beta$. Giá trị $1 - \beta$ được gọi là lực lượng của phép kiểm định. Mong muốn của người làm thống kê là xây một qui tắc chấp nhận hoặc bác bỏ một giả thuyết sao cho xác suất cả hai loại sai lầm càng nhỏ càng tốt. Tuy nhiên ta có

$$P(W/H_0) + P(\overline{W}/H_0) = 1 \quad ; \quad P(W/ H_1) + P(\overline{W}/H_1) = 1$$

Từ đây suy ra khi α giảm thì β tăng và ngược lại. Với mẫu có kích thước cố định, để xây dựng một qui tắc hành động chấp nhận hoặc bác bỏ giả thuyết ta có thể đi theo một trong hai hướng sau:

Hướng thứ nhất: Cố định xác suất sai lầm loại 1 xây dựng một qui tắc sao cho xác suất sai lầm loại 2 là nhỏ nhất hoặc có thể chấp nhận được.

Hướng thứ hai ngược lại với hướng thứ nhất

Do đối thuyết H_1 thường là mệnh đề hợp (là hợp của các mệnh đề) nên việc cố định xác suất sai lầm loại hai là phức tạp và khó khả thi. Trong giáo trình này chúng ta sẽ đi theo hướng thứ nhất để xây dựng qui tắc kiểm định giả thuyết. Với mỗi cặp giả thuyết và đối thuyết đã cho, không phải lúc nào cũng tồn tại hoặc tìm được một qui tắc sao cho lực lượng của phép kiểm định $1 - \beta$ là lớn nhất. Những qui tắc đưa ra trong giáo trình này là những qui tắc thông dụng.

II. Kiểm định các giả thuyết tham số

1. Giả thuyết đơn - Đối thuyết đơn

Cặp giả thuyết: H_0: $\theta = \theta_0$

Đối thuyết H_1: $\theta = \theta_1 \quad (\theta_1 \neq \theta_0)$

là cặp giả thuyết và đối thuyết đơn

* Một qui tắc kiểm định với cặp giả thuyết - đối thuyết đơn được gọi là mạnh nhất nếu nó có lực lượng của phép kiểm định là lớn nhất

* Định lý Neyman - Pearson đã chỉ ra rằng: Nếu đặc trưng X ở tổng thể có hàm mật độ $f(x, \theta)$ thì tồn tại qui tắc mạnh nhất kiểm định cặp giả thuyết - đối thuyết đơn vừa nêu. Việc phát biểu và chứng minh định lý Neyman - Pearson không được nêu ra trong giáo trình này. Người đọc muốn biết có thể tham khảo ở các sách đã dẫn ra ở cuối giáo trình.

2. Giả thuyết đơn - Đối thuyết hợp.

Giả thuyết H_0 : $\theta=\theta_0$

Với đối thuyết: H_1 : $\theta \in D$; (D là một miền không chứa θ_0)

được gọi là cặp giả thuyết đơn với đối thuyết hợp.

Nhận thấy rằng: Với mỗi $\theta \in D$, sai lầm loại 2: $\beta = \beta(\theta)$ là một hàm số xác định trên D. Qui tắc kiểm định cặp giả thuyết, đối thuyết trên sao cho $\beta(\theta)$ cực tiểu $\forall\ \theta \in D$ được gọi là qui tắc kiểm định mạnh đều nhất.

Ví dụ : Biết biến X ở tổng thể có phân phối chuẩn $N(\mu;\sigma^2)$ với σ^2 đã biết. Xét cặp giả thuyết, đối thuyết đơn

H_0 : $\mu=\mu_0$

H_1 : $\mu=\mu_1 \quad (\mu_1 > \mu_0)$

Với mức ý nghĩa α, qui tắc mạnh nhất để kiểm định cặp giả thuyết, đối thuyết trên là:

** Bác bỏ H_0 nếu:* $\dfrac{\overline{X}-\mu_0}{\sigma}\sqrt{n} > U_\alpha$

** Chấp nhận H_0 nếu:* $\dfrac{\overline{X}-\mu_0}{\sigma}\sqrt{n} \leq U_\alpha$

Bởi qui tắc kiểm định vừa nêu không phụ thuộc vào μ_1 mà chỉ cần yêu cầu $\mu_1 > \mu_0$ nên qui tắc trên cũng là qui tắc mạnh đều nhất kiểm định cặp giả thuyết đơn, đối thuyết hợp

H_0 : $\mu=\mu_0$

H_1 : $\mu>\mu_0$ với mức ý nghĩa α.

3. Đối thuyết một phía và hai phía

Xét giả thuyết đơn : H_0 : $\theta=\theta_0$

* Mệnh đề: H_1 : $\theta\neq\theta_0$ gọi là đối thuyết hai phía của H_0

* Mệnh đề: H_1 : $\theta<\theta_0$ gọi là đối thuyết phía trái của H_0

* Mệnh đề: H_1 : $\theta>\theta_0$ gọi là đối thuyết phía phải của H_0.

Không phải lúc nào cũng tồn tại qui tắc mạnh đều nhất để kiểm định giả thuyết H_0 với một trong ba đối thuyết vừa nêu. Các qui tắc kiểm định được giới thiệu trong giáo trình này hoặc là qui tắc mạnh đều nhất hoặc là qui tắc tốt và thông dụng trong thống kê . Qui tắc "tốt" ở đây có thể hiểu theo nghĩa: Qui tắc mạnh đều nhất là tối ưu toàn cục thì qui tắc " tốt" là tối ưu bộ phận

4. Kiểm định kì vọng của phân phối chuẩn khi phương sai đã biết.

Giả sử đặc trưng X ở tổng thể có phân phối chuẩn $N(\mu;\sigma^2)$ với σ^2 đã biết.

Từ mẫu ngẫu nhiên $(X_1, X_2,....X_n)$ ta xây dựng qui tắc kiểm định giả thuyết

H_0 : $\mu=\mu_0$ trong các trường hợp sau:

4.1. Trường hợp 1: Đối thuyết H_1 : $\mu\neq\mu_0$

Thống kê $Z=\dfrac{\overline{X}-\mu}{\sigma}\sqrt{n}$ có phân phối chuẩn tắc.

Với $t\in(0, 1)$ xác suất $P[|\dfrac{\overline{X}-\mu}{\sigma}\sqrt{n}|>U_{\frac{t}{2}}]=t$. Nếu giả thuyết H_0 đúng thì xác suất

$P[|\dfrac{\overline{X}-\mu_0}{\sigma}\sqrt{n}|>U_{\frac{t}{2}}]=t$. Nếu lấy $t=\alpha$ là mức ý nghĩa của bài toán kiểm định thì

$P[|\dfrac{\overline{X}-\mu_0}{\sigma}\sqrt{n}|>U_{\frac{\alpha}{2}}]=\alpha$.

Bất đẳng thức $\left|\dfrac{\overline{X}-\mu_0}{\sigma}\sqrt{n}\right|>U_{\frac{\alpha}{2}}\Leftrightarrow$ Mẫu ngẫu nhiên $(X_1, X_2,....X_n)\in W$.

Điều này có nghĩa là: $P[\left|\dfrac{\overline{X}-\mu_0}{\sigma}\sqrt{n}\right|>U_{\frac{\alpha}{2}}]=P(W/H_0)=\alpha$

Vậy qui tắc kiểm định: Giả thuyết H_0: $\mu=\mu_0$

Đối thuyết H_1: $\mu\neq\mu_0$.

với mức ý nghĩa α là:

Qui tắc 1: Nếu: $Z_T=\left|\dfrac{\overline{X}-\mu_0}{\sigma}\sqrt{n}\right|>U_{\frac{\alpha}{2}}$ quyết định bác bỏ H_0.

Nếu: $Z_T=\left|\dfrac{\overline{X}-\mu_0}{\sigma}\sqrt{n}\right|\leq U_{\frac{\alpha}{2}}$ quyết định chấp nhận H_0.

Với mẫu cụ thể $(x_1, x_2,....x_n)$ đã cho ta thực hiện bài toán kiểm định theo các bước sau:

Bước 1: Tính đại lượng $Z_T = \frac{|\overline{x} - \mu_0|}{\sigma}\sqrt{n}$

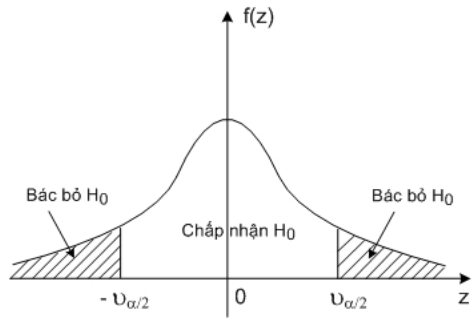
Bước 2: Tìm $U_{\frac{\alpha}{2}}$

Bước 3: So sánh hai giá trị trên rồi đưa ra quyết định:

Nếu: $Z_T > U_{\frac{\alpha}{2}}$ quyết định bác bỏ H_0.

Nếu: $Z_T \leq U_{\frac{\alpha}{2}}$ quyết định chấp nhận H_0.

Hình vẽ sau mô tả miền chấp nhận và miền bác bỏ giả thuyết H_0



Hình 2

Các giá trị $-U_{\frac{\alpha}{2}}$ và $U_{\frac{\alpha}{2}}$ là các ngưỡng so sánh khi quyết định chấp nhận hay bác bỏ H_0.

Ví dụ 1: Từ tổng thể có phân phối chuẩn N(μ, 4) ta lấy mẫu có kích thước n = 9 và tìm được $\overline{x}$ = 21,20.Với mức ý nghĩa α = 0,05 hãy kiểm định giả thuyết :

H_0: $\mu = 20$

H_1: $\mu \neq 20$.

Tính $Z_T = \frac{|\overline{x} - \mu_0|}{\sigma}\sqrt{n} = \frac{|21,2 - 20,0|}{2}3 = 1,8$; Tìm $U_{0,025}$ = 1,96.

Vì Z_T =1,80 < 1,96 = $U_{0,025}$ ta quyết định chấp nhận H_0.

4.2 Trường hợp 2: Đối thuyết $\mu > \mu_0$

Tương tự như trường hợp trên thống kê $Z = \frac{\overline{X} - \mu}{\sigma}\sqrt{n}$ ta có phân phối chuẩn tắc. Dựa vào thống kê này ta có qui tắc kiểm định giả thuyết :

H_0 : $\mu = \mu_0$.

H_1 : $\mu > \mu_0$ với mức ý nghĩa α là:

Qui tắc 2: Nếu : $\frac{\overline{X} - \mu_0}{\sigma}\sqrt{n} > U_\alpha$ quyết định bác bỏ H_0.

Nếu: $\frac{\overline{X} - \mu_0}{\sigma}\sqrt{n} \le U_\alpha$ quyết định chấp nhận H_0.

Với mẫu cụ thể (x_1, x_2,....x_n) ta cũng thực hiện bài toán kiểm định thao các bước:

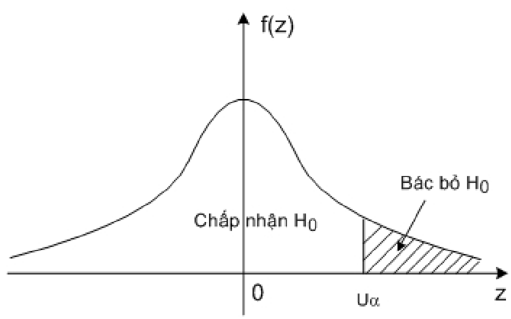
Bước 1: Tính $Z_T = \frac{\overline{x} - \mu_0}{\sigma}\sqrt{n}$

Bước 2: Tìm U_α

Bước 3: Nếu: $Z_T > U_\alpha$ quyết định bác bỏ H_0.

Nếu: $Z_T \le U_\alpha$ quyết định chấp nhận H_0.

Miền chấp nhận và miền bác bỏ được mô tả bởi hình vẽ sau:



. Hình 3

4.3 Trường hợp 3: Đối thuyết $\mu < \mu_0$

Tương tự như trường hợp trên thống kê Z = $\frac{\overline{X} - \mu}{\sigma}\sqrt{n}$ ta có phân phối chuẩn tắc. Dựa vào thống kê này ta có qui tắc kiểm định giả thuyết :

H_0 : $\mu = \mu_0$

H_1 : $\mu < \mu_0$ với mức ý nghĩa α là:

Qui tắc 3: Nếu: $Z_T = \frac{\overline{X}-\mu_0}{\sigma}\sqrt{n} < -U_\alpha$ quyết định bác bỏ H_0.

Nếu: $Z_T = \frac{\overline{X}-\mu_0}{\sigma}\sqrt{n} \geq -U_\alpha$ quyết định chấp nhận H_0.

Với mẫu cụ thể ($x_1, x_2,....x_n$) ta cũng thực hiện bài toán kiểm định thao các bước:

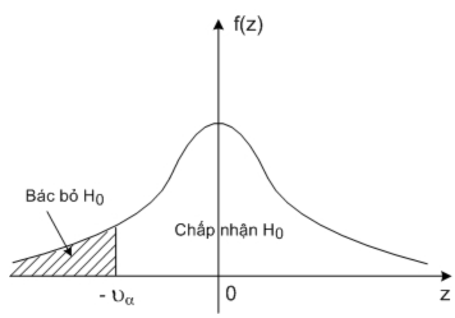
Bước 1: Tính $Z_T = \frac{\overline{x}-\mu_0}{\sigma}\sqrt{n}$

Bước 2: Tìm U_α .

Bước 3: Nếu $Z_T < -U_\alpha$ quyết định bác bỏ H_0.

Nếu $Z_T \geq -U_\alpha$ quyết định chấp nhận H_0.

Miền chấp nhận và miền bác bỏ được mô tả bởi hình vẽ sau:



. Hình 4

Ví dụ 3: Trọng lượng X gói mì ăn liền tuân theo qui luật chuẩn N(μ, 25).Từ mẫu 25 gói mì ăn liền ta tìm được $\overline{x} = 82$ gam. Với mức ý nghĩa $\alpha = 0,05$ hãy kiểm định giả thuyết

H_0 : $\mu = 80$

H_1 : $\mu > 80$.

Ta có $Z_T = \frac{\overline{x}-\mu_0}{\sigma}\sqrt{n} = \frac{82-80}{5}.5 = 2,0$; $U_{0,05} = 1,68$.

Từ $Z_T > U_{0,05}$, áp dụng qui tắc 2 ta quyết định bác bỏ H_0

5. Kiểm định kỳ vọng của phân phối chuẩn khi không biết phương sai

Giả sử biến X ở tổng thể có phân phối chuẩn N($\mu;\sigma^2$).

Từ mẫu ngẫu nhiên (X_1, X_2, ..., X_n) xây dựng quy tắc kiểm định giả thuyết

H_0: $\mu = \mu_0$ trong các trường hợp đối thuyết hai phía và đối thuyết một phía

5.1 *Kiểm định giả thuyết:* H_0: $\mu = \mu_0$

H_1: $\mu \neq \mu_0$

Xét thống kê $Z = \frac{\overline{X} - \mu}{S}\sqrt{n}$, với $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$

Do Z ~ T_{n-1} nên $\forall$ $t \in (0,1)$ ta có

$$P[\left|\frac{\overline{X} - \mu}{S}\sqrt{n}\right| > t_{\frac{t}{2},n-1}] = t \qquad (1)$$

Nếu H_0 đúng thì:

$$P[\left|\frac{\overline{X} - \mu_0}{S}\sqrt{n}\right| > t_{\frac{t}{2},n-1}] = t \qquad (2)$$

Lấy t = α là mức ý nghĩa của bài toán kiểm định ta có:

$$P[\left|\frac{\overline{X} - \mu_0}{S}\sqrt{n}\right| > t_{\frac{\alpha}{2},n-1}] = \alpha \qquad (3)$$

Bất đẳng thức:

$$P\left[\left|\frac{\overline{X} - \mu_0}{S}\right| > t_{\frac{\alpha}{2},n-1}\right] = \alpha \qquad \Leftrightarrow (X_1, X_2, ..., X_n) \in W \Rightarrow P(W/H_0) = \alpha.$$

Từ đây có quy tắc kiểm định

H_0: $\mu = \mu_0$

H_1: $\mu \neq \mu_0$ với mức ý nghĩa α là

Qui tắc 4: Nếu $\left|\frac{\overline{X} - \mu_0}{S}\sqrt{n}\right| > t_{\frac{\alpha}{2},n-1}$ quyết định bác bỏ H_0

Nếu $\left|\dfrac{\overline{X}-\mu_0}{S}\sqrt{n}\right| \le t_{\frac{\alpha}{2},n-1}$ ta chấp nhận H_0

Cũng như qui tắc 1: Với mẫu cụ thể (x_1, x_2,....x_n) ta tiến hành bài toán kiểm định theo các bước sau:

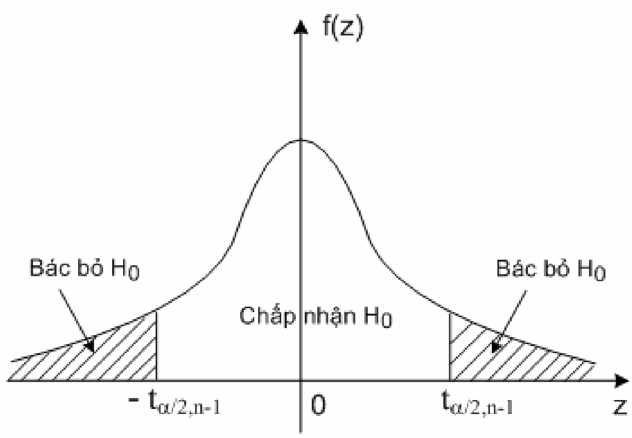
Bước 1: Tính $Z_T = \left|\dfrac{\overline{x}-\mu_0}{s}\sqrt{n}\right|$

Bước 2: Tìm $t_{\frac{\alpha}{2},n-1}$

Bước 3: Nếu $Z_T > t_{n-1,\frac{\alpha}{2}}$ quyết định bác bỏ H_0

Nếu $Z_T \le t_{\frac{\alpha}{2},n-1}$ quyết định chấp nhận H_0

Miền chấp nhận và miền bác bỏ H_0 cho bởi hình sau:



Hình 6

Ví dụ 1: Năng suất lúa là một đại lượng ngẫu nhiên X có phân phối chuẩn N($\mu;\sigma^2$). Điều tra năng suất giống lúa trên ở 200 ruộng ta được bảng các số liệu sau:

Năng suất tạ/ha	46	48	49	50	51	53	54	58
Số thửa ruộng	17	18	35	45	42	23	10	10

Với mức ý nghĩa α = 0,05. Hãy kiểm định cặp giả thuyết đối thuyết:

H_0: $\mu = 52$

H_1: $\mu \ne 52$

Để tính $\overline{x}$ và s ta lập bảng sau:

x_i	n_i	$x_i n_i$	$n_i x_i^2$
46	17	782	35972
48	18	804	41472
49	35	1715	84035
50	45	2250	112500
51	42	2142	109242
53	23	1219	64607
54	10	540	29160
58	10	580	33640
$\sum$	200	. 10032	. 510628

$\overline{x}$ = 50,16, $\hat{s}^2 = \overline{x^2} - \overline{x}^2 = 37{,}1144, s^2 = \dfrac{200\hat{s}^2}{199} = 37{,}3009$; s = 6,11.

$Z_T = \dfrac{|50{,}16 - 52{,}00|}{6{,}11}\sqrt{200} = 4{,}26$, $t_{0,025\,,\,199} = U_{0,025} = 1{,}96$

$Z_T = 4{,}26 > 1{,}96 \Rightarrow$ bác bỏ H_0

5.2 *Trường hợp 2.* Đối thuyết H_1: $\mu < \mu_0$

Tương tự như trường hợp 1. Từ thống kê $Z = \dfrac{\overline{X} - \mu}{S}\sqrt{n}$ có phân phối T_{n-1} nếu H_0 đúng ta có qui tắc kiểm định cặp giả thuyết đối thuyết ở mức ý nghĩa α là

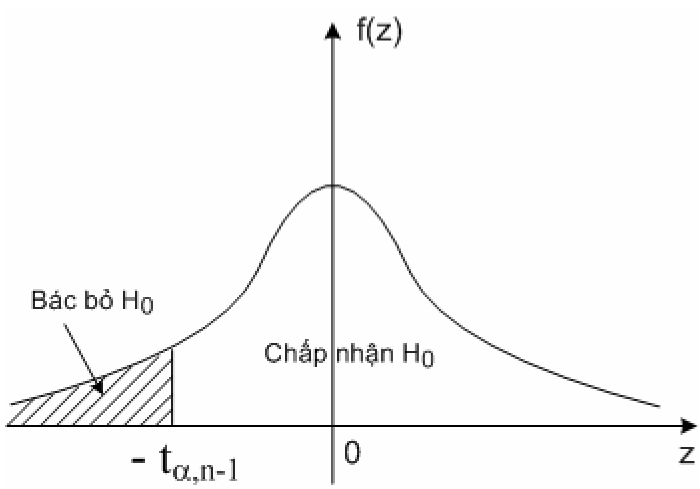
H_0: $\mu = \mu_0$

H_1: $\mu < \mu_0$

Qui tắc 5: Nếu $Z = \dfrac{\overline{X} - \mu_0}{S}\sqrt{n} < -t_{\alpha, n-1}$ bác bỏ H_0

Nếu $Z = \dfrac{\overline{X} - \mu_0}{S}\sqrt{n} \geq -\ t_{\alpha, n-1}$ chấp nhận H_0

Hình vẽ sau cho miền chấp nhận và bác bỏ H_0



Hình 7

Dựa vào qui tắc 5, các bước thực hiện bài toán kiểm định như trường hợp 1

5.3 *Đối thuyết H_1:* $\mu > \mu_0$

Tương tự như trường hợp 2, qui tắc kiểm định cặp giả thuyết đối thuyết

H_0: $\mu = \mu_0$

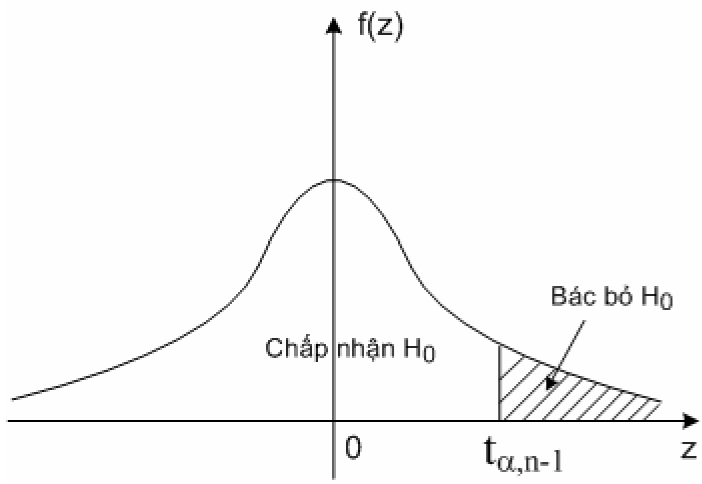
H_1: $\mu > \mu_0$

ở mức ý nghĩa α

Qui tắc 6: Nếu: $Z_T = \dfrac{\overline{X} - \mu_0}{S}\sqrt{n} > t_{\alpha, n-1}$, bác bỏ H_0

Nếu: $Z = \dfrac{\overline{X} - \mu_0}{S}\sqrt{n} \leq t_{\alpha, n-1}$ chấp nhận H_0

Hình vẽ sau cho miền chấp nhận và bác bỏ H_0



Hình 8

6.Kiểm định phương sai của phân phối chuẩn

Từ mẫu ngẫu nhiên (X_1, X_2, ... ,X_n) của tổng thể có phân phối chuẩn N($\mu;\sigma^2$), với mức ý nghĩa α xây dựng qui tắc kiểm định cặp giả thuyết , đối thuyết :

$$H_0: \quad \sigma^2 = \sigma_0^2$$

$$H_1: \quad \sigma^2 > \sigma_0^2$$

Thống kê $Z = \dfrac{(n-1)S^2}{\sigma^2}$ có phân phối χ^2_{n-1}, với t ∈ (0,1) ta có

$$P[\frac{(n-1)S^2}{\sigma^2} > \chi^2_{t,n-1}] = t \qquad (1)$$

Nếu H_0 đúng thì $P[\dfrac{(n-1)S^2}{\sigma_0^{\ 2}} > \chi^2_{t,n-1}] = t$ (2)

Cho t = α là mức ý nghĩa của bài toán kiểm định có:

$$P[\frac{(n-1)S^2}{\sigma_0^{\ 2}} > \chi^2_{\alpha,n-1}] = \alpha \qquad (3)$$

Bất đẳng thức:

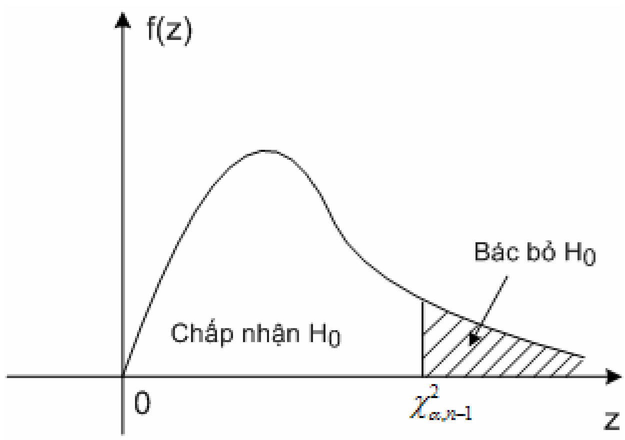
$\dfrac{(n-1)S^2}{\sigma_0^{\ 2}} > \chi^2_{\alpha,n-1}$ tương đương với (X_1, X_2, ... ,X_n) ∈ W $\Leftrightarrow P(W/H_0) = \alpha$

Qui tắc kiểm định H_0 là:

Qui tắc 7: Nếu $\dfrac{(n-1)S^2}{\sigma_0{}^2} > \chi^2_{\alpha,n-1}$ bác bỏ H_0

Nếu $\dfrac{(n-1)S^2}{\sigma_0{}^2} \le \chi^2_{\alpha,n-1}$ chấp nhận H_0

Hình vẽ sau chỉ miền chấp nhận và miền bác bỏ H_0



Hình 9

Từ mẫu cụ thể (x_1, x_2,....x_n) theo qui tắc ta thực hiện bài toán kiểm định cặp giả thuyết , đối thuyết: H_0: $\sigma^2 = \sigma_0^2$

H_1: $\sigma^2 > \sigma_0^2$ ở mức ý nghĩa α theo các bước sau:

Bước 1: Tính $Z_T = \dfrac{(n-1)s^2}{\sigma_0{}^2}$

Bước 2: Tìm $\chi^2_{\alpha,n-1}$

Bước 3: Nếu $Z_T > \chi^2_{\alpha,n-1}$ ta bác bỏ H_0

Nếu $Z_T \le \chi^2_{\alpha,n-1}$ ta chấp nhận H_0

Ví dụ: Một máy sản xuất các tấm chất dẻo được thường xuyên theo dõi về độ dày của sản phẩm Biết độ dày X của các tấm chất dẻo tuân theo qui luật chuẩn $N(\mu;\sigma^2)$. Nếu độ lệch chuẩn vượt quá 0,3mm thì chất lượng sản phẩm không được đảm bảo về kĩ thuật. Người ta chọn ngẫu nhiên 10 tấm chất dẻo rồi đo độ dày của mỗi tấm và được kết quả sau (đơn vị đo mm)

22,0 ; 22,6 ; 23,2 ; 22,7 ; 22,5 ; 22,8 ; 22,5 ; 22,8 ; 22,9 ; 23,0.

Từ yêu cầu của thực tế với mức ý nghĩa α = 0,05 hãy lập cặp giả thuyết và đối thuyết thích hợp đánh giá tình trạng làm việc của máy sản xuất các tấm chất dẻo trên

Độ lệch chuẩn ở mức cho phép không vượt quá 0,3 mm tương ứng với phương sai σ^2 không vượt quá 0,09 mm^2. Ta có cặp giả thuyết đối thuyết sau:

H_0: $\sigma^2 = 0{,}09$

H_1 : $\sigma^2 > 0{,}09$

Với mẫu đã cho ta có: $\bar{x} = 2{,}27$, $s^2 = 0{,}1089$, $Z_T = \dfrac{(n-1)s^2}{\sigma_0^2} = \dfrac{9.0{,}1089}{0{,}09} = 10{,}89$

ở mức ý nghĩa 0,05 ta có:

$\chi^2_{9,0,05} = 16{,}92$; $Z_T = 10{,}89 < 16{,}92 = \chi^2_{0,05,9}$ quyết định chấp nhận H_0, điều này có nghĩa là máy sản xuất các tấm dẻo vẫn hoạt động bình thường.

Với cặp giả thuyết H_0: $\sigma^2 = \sigma_0^2$

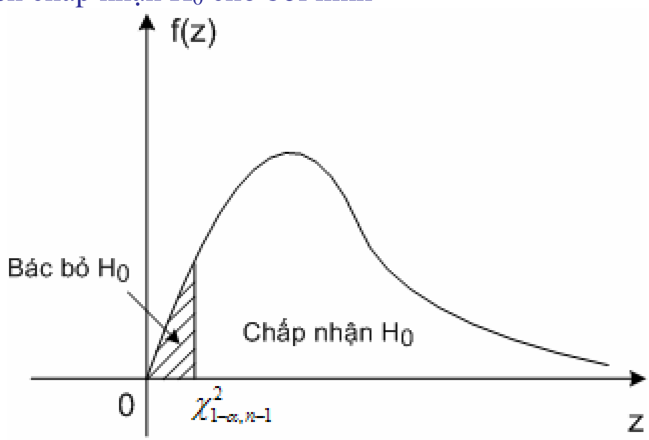
H_1: $\sigma^2 < \sigma_0^2$

Ta có qui tắc kiểm định là

Qui tắc 8: Nếu $\dfrac{(n-1)s^2}{\sigma_0^2} < \chi^2_{1-\alpha,n-1}$ ta bác bỏ H_0

Nếu $\dfrac{(n-1)s^2}{\sigma_0^2} \geq \chi^2_{1-\alpha,n-1}$ ta chấp nhận H_0

Miền bác bỏ và miền chấp nhận H_0 cho bởi hình



Hình 10

Với cặp giả thuyết : H_0: $\sigma^2 = \sigma_0^2$

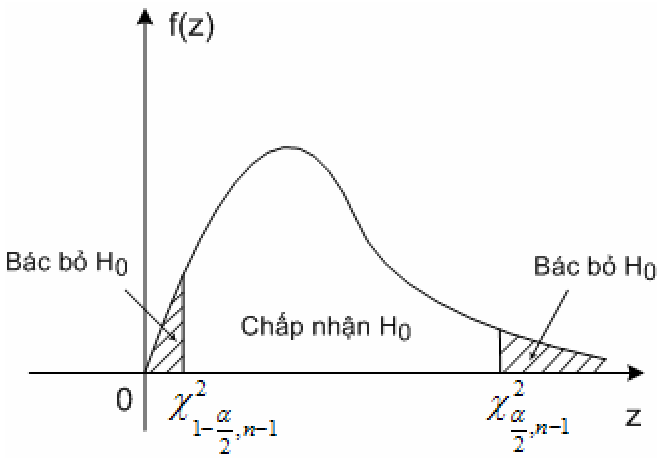
H_1: $\sigma^2 \neq \sigma_0^2$

Ta có qui tăc kiểm định là:

Qui tắc 9: Nếu $\dfrac{(n-1)s^2}{\sigma_0^2} < \chi^2_{1-\frac{\alpha}{2},n-1}$ hoặc $\dfrac{(n-1)s^2}{\sigma_0^2} > \chi^2_{\frac{\alpha}{2},n-1}$ bác bỏ H_0

Nếu $\chi^2_{1-\frac{\alpha}{2},n-1} \leq \dfrac{(n-1)s^2}{\sigma_0^2} \leq \chi^2_{\frac{\alpha}{2},n-1}$ chấp nhận H_0.

Miền bác bỏ và miền chấp nhận H_0 cho bởi hình



Hình 11

Nhận xét: Do phương sai là số do đặc trưng cho sai số nên các bài toán kiểm định về phương sai người ta thường kiểm định với đối thuyết một phía.

7. Kiểm định xác suất .

Xác suất xuất hiện sự kiện A ở mỗi phép thử P(A) = p. Tiến hành n phép thử độc lập có n_A lần xuất hiện A. Với mức ý nghĩa α ta xây dựng qui tắc kiểm định cặp giả thuyết, đối thuyết :

H_0: $p = p_0$

H_1: $p \neq p_0$

Xét thống kê $Z = \frac{f - p}{\sqrt{pq}}\sqrt{n}$, $f = \frac{n_A}{n}$, q =1 - p, thống kê này theo định lý giới hạn trung tâm có phân phối xấp xỉ chuẩn tắc. Nếu H_0 đúng thì

$Z = \frac{f - p_0}{\sqrt{p_0 q_0}}\sqrt{n}$ có phân phối xấp xỉ chuẩn tắc. Giống như bài toán kiểm định kỳ vọng của phân phối chuẩn, ta có qui tắc kiểm định cặp giả thuyết đối thuyết trên là:

Qui tắc 10: Nếu: $\left|\frac{f - p_0}{\sqrt{p_0 q_0}}\sqrt{n}\right| > U_{\frac{\alpha}{2}}$ ta bác bỏ H_0

Nếu: $\left|\frac{f - p_0}{\sqrt{p_0 q_0}}\sqrt{n}\right| \leq U_{\frac{\alpha}{2}}$ ta chấp nhận H_0.

Ví dụ: Xét nghiệm 1000 mẫu máu của những người dân ở vùng Tây Nguyên ta thấy có 232 mẫu máu có ký sinh trùng sốt rét. Hãy kiểm định :

H_0: $p = 0{,}2$

H_1: $p \neq 0{,}2$, mức ý nghĩa $\alpha = 0{,}05$.

Ta có $f = \frac{232}{1000} = 0{,}0232$, $Z_T = \left|\frac{f - p_0}{\sqrt{p_0 q_0}}\sqrt{n}\right| = \left|\frac{f - 0{,}2}{\sqrt{0{,}2.0{,}8}}\sqrt{1000}\right| = 2{,}53$.

$U_{0,025} = 1{,}96$; $Z_T = 2{,}53 > 1{,}96$ vậy giả thuyết H_0 bị bác bỏ.

Chú ý 1: Với cặp giả thuyết đối thuyết

H_0: $p = p_0$

H_1: $p > p_0$

Có qui tắc kiểm định sau:

Qui tắc 11: Nếu $\frac{f - p_0}{\sqrt{p_0 q_0}}\sqrt{n} > U_\alpha$ bác bỏ H_0

. Nếu $\frac{f - p_0}{\sqrt{p_0 q_0}}\sqrt{n} \le U_\alpha$ chấp nhận H_0.

Chú ý 2: Với cặp giả thuyết đối thuyết

H_0: $p = p_0$

H_1: $p < p_0$

Ta có qui tắc kiểm định sau:

Qui tắc 12: Nếu $\frac{f - p_0}{\sqrt{p_0 q_0}}\sqrt{n} < - U_\alpha$ bác bỏ H_0

. Nếu $\frac{f - p_0}{\sqrt{p_0 q_0}}\sqrt{n} \ge U_\alpha$ chấp nhận H_0.

Ví dụ: Một kho bảo quản hạt giống được xem là chưa đảm bảo kỹ thuật nếu tỷ lệ hạt nảy mầm dưới 70%. Người ta lấy 500 hạt trong kho đem gieo và thấy có 340 hạt nảy mầm.Với mức ý nghĩa $\alpha = 0{,}05$ hãy lập giả thuyết và đối thuyết thích hợp để xét xem kho bảo quản hạt giống trên đã đảm bảo kỹ thuật hay chưa.

Yêu cầu trên tương đương với việc kiểm định cặp giả thuyết đối thuyết :

H_0: $p = 0{,}7$

H_1: $p < 0{,}7$

Ta có $f = 0{,}68$, $1 - p_0 = 0{,}3$, $Z_T = \frac{f - p_0}{\sqrt{p_0(1-p_0)}}\sqrt{n} = \frac{0{,}68 - 0{,}7}{\sqrt{0{,}7.0{,}3}}\sqrt{500} = -1{,}79$

$U_{0,025} = 1{,}96$; $Z_T = -1{,}79 > -U_{0,025} = -1{,}96$

Với kết quả trên ta chưa đủ cơ sở để kết luận tiêu chuẩn kĩ thuật của kho bảo quản hạt giống là có vấn đề.

8. Kiểm định sự bằng nhau của hai kỳ vọng của hai phân phối chuẩn.

Từ hai mẫu ngẫu nhiên độc lập lấy từ hai tổng thể có phân phối chuẩn $N(\mu_X;\sigma_X^2)$ $N(\mu_Y;\sigma_Y^2)$) ta xây dựng qui tắc kiểm định cặp giả thuyết, đối thuyết:

H_0: $\mu_X = \mu_Y$

H_1: $\mu_X \ne \mu_Y$

8.1: Trường hợp σ_X^2 , σ_Y^2 đã biết.

Thống kê $Z = \frac{\overline{X} - \overline{Y} - (\mu_X - \mu_Y)}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}}$ có phân phối chuẩn tắc

Nếu H_0 đúng thì $Z = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}}$ có phân phối chuẩn tắc, Khi đó: $P[\frac{|\overline{X}-\overline{Y}|}{\sqrt{\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}}} > U_{\frac{\alpha}{2}}] = \alpha$

Từ đây có qui tắc bác bỏ H_0 là:

Qui tắc 13: Nếu $\frac{|\overline{X}-\overline{Y}|}{\sqrt{\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}}} > U_{\frac{\alpha}{2}}$ quyết định bác bỏ H_0

Nếu $\frac{|\overline{X}-\overline{Y}|}{\sqrt{\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}}} \le U_{\frac{\alpha}{2}}$ quyết định chấp nhận H_0.

Với hai mẫu cụ thể $(x_1, x_2, ..., x_n)$ và $(y_1, y_2,...y_m)$ thực hiện bài toán kiểm định cặp giả thuyết , đối thuyết trên theo các bước sau:

Bước 1: Tính $Z_T = \frac{|\overline{x}-\overline{y}|}{\sqrt{\frac{\sigma_x^2}{n}+\frac{\sigma_y^2}{m}}}$

Bước 2: Tìm $U_{\frac{\alpha}{2}}$.

Bước 3: Nếu $Z_T > U_{\frac{\alpha}{2}}$ bác bỏ H_0

Nếu $Z_T \le U_{\frac{\alpha}{2}}$ chấp nhận H_0

8.2. Trường hợp σ_X^2 ; σ_Y^2 chưa biết nhưng n, m $\ge$ 30.(gọi là mẫu lớn)

Do thống kê $Z = \frac{\overline{X}-\overline{Y}-(\mu_X-\mu_Y)}{\sqrt{\frac{\hat{S}_X^2}{n}+\frac{\hat{S}_Y^2}{m}}}$ có phân phối xấp xỉ phân phối chuẩn tắc. Nếu H_0 đúng thì

$$Z = \frac{\overline{X}-\overline{Y}}{\sqrt{\frac{\hat{S}_X^2}{n}+\frac{\hat{S}_Y^2}{m}}}$$ có phân phối xấp xỉ chuẩn tắc

Cũng như ở trường hợp trên ta có qui tắc chấp nhận hoặc bác bỏ giả thuyết H_0 với đối thuyết hai phía là:

Qui tắc 14: Nếu: $\frac{|\overline{X}-\overline{Y}|}{\sqrt{\frac{\hat{S}_x^2}{n}+\frac{\hat{S}_y^2}{m}}} > U_{\frac{\alpha}{2}}$ quyết định bác bỏ H_0

Nếu: $\frac{|\overline{X}-\overline{Y}|}{\sqrt{\frac{\hat{S}_x^2}{n}+\frac{\hat{S}_y^2}{m}}} \le U_{\frac{\alpha}{2}}$ quyết định chấp nhận H_0.

8.3. Trường hợp σ_X^2, σ_Y^2 chưa biết nhưng biết $\sigma_X^2 = \sigma_Y^2$.

Thống kê $Z = \dfrac{\overline{X} - \overline{Y} - (\mu_X - \mu_Y)}{S\sqrt{\dfrac{1}{n} + \dfrac{1}{m}}}$ có phân phối T_{n+m-2}

$$S^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2 + \sum_{j=1}^{m}(Y_j - \overline{Y})^2}{n+m-2} = \frac{(n-1)S_x^2 + (m-1)S_y^2}{n+m-2}$$

Nếu H_0 đúng thì $Z = \dfrac{\overline{X} - \overline{Y}}{S\sqrt{\dfrac{1}{n} + \dfrac{1}{m}}}$ có phân phối T_{n+m-2} và có qui tắc kiểm định cặp giả thuyết, đối thuyết ở mức ý nghĩa α là:

H_0 : $\mu_X = \mu_Y$

H_1 : $\mu_X \neq \mu_Y$

Qui tắc 15: Nếu $\dfrac{|\overline{X} - \overline{Y}|}{S\sqrt{\dfrac{1}{n} + \dfrac{1}{m}}} > t_{\frac{\alpha}{2} n+m-2}$ quyết định bác bỏ H_0

. Nếu $\dfrac{|\overline{X} - \overline{Y}|}{S\sqrt{\dfrac{1}{n} + \dfrac{1}{m}}} \leq t_{\frac{\alpha}{2}, n+m-2}$ quyết định chấp nhận H_0

Với hai mẫu cụ thể $(x_1, x_2, ..., x_n)$ và $(y_1, y_2, ... y_m)$ thực hiện bài toán kiểm định cặp giả thuyết , đối thuyết trên theo các bước sau:

Bước 1: Tính $Z_T = \dfrac{|\overline{x} - \overline{y}|}{s\sqrt{\dfrac{1}{n} + \dfrac{1}{m}}}$ với $s^2 = \dfrac{(n-1)s_X^2 + (m-1)s_Y^2}{n+m-2}$

Bước 2: Tìm $t_{\frac{\alpha}{2}, n+m-2}$

Bước 3: Nếu $Z_T > t_{\frac{\alpha}{2}, n+m-2}$ quyết định bác bỏ H_0

. Nếu $Z_T \leq t_{\frac{\alpha}{2}, n+m-2}$ quyết định chấp nhận H_0

Ví dụ: Điều tra năng suất của 8 thửa ruộng trồng giống lúa A và năng suất của 10 thửa ruộng trồng giống lúa B với cùng một điều kiện canh tác ta có kết quả sau:

X	40	38	40	42	44	41	36	39		
Y	41	44	38	42	40	45	39	37	43	41

Biết rằng X~N(μ_X;σ^2), Y~N(μ_Y;σ^2). Với mức ý nghĩa α = 0,05 hãy kiểm định:

H_0: $\mu_X = \mu_Y$

H_1: $\mu_X \neq \mu_Y$

Ta có $\overline{x} = 40$, $\overline{y} = 41$, $s_X^2 = \dfrac{42}{7} = 6$, $s_Y^2 = 6{,}66$, $s^2 = 6{,}375$, $s = 2{,}52$.

$$Z_T = \frac{|\overline{x} - \overline{y}|}{s\sqrt{\frac{1}{n} + \frac{1}{m}}} = \frac{1}{2{,}52\sqrt{\frac{1}{8} + \frac{1}{10}}} = 0{,}835.$$

$t_{0,025\,,\,16} = 2{,}12$; $Z_T = 0{,}835 < t_{0,025\,,\,16} = 2{,}12$. Chấp nhận H_0.

Chú ý 1: Bài toán kiểm định cặp giả thuyết và đối thuyết trên còn được gọi là bài toán so sánh sự bằng nhau và khác nhau hai kỳ vọng của hai phân phối chuẩn.

Chú ý 2: Trong thực tế nhiều khi ta cần so sánh kỳ vọng của biến ngẫu nhiên X với kỳ vọng của biến ngẫu nhiên Y cộng thêm một hằng số μ_0 khi đó ta có bài toán kiểm định cặp giả thuyết đối thuyết sau:

H_0 : $\mu_X = \mu_Y + \mu_0$

H_1 : $\mu_X \neq \mu_Y + \mu_0$

Khi gặp bài toán này có các qui tắc kiểm định sau:

* Nếu biết σ_x^2, σ_y^2 qui tắc kiểm định cặp giả thuyết, đối thuyết trên là:

Qui tắc 16: Nếu: $\dfrac{|\overline{X} - \overline{Y} - \mu_0|}{\sqrt{\dfrac{\sigma_x^2}{n} + \dfrac{\sigma_y^2}{m}}} > U_{\frac{\alpha}{2}}$ quyết định bác bỏ H_0

Nếu: $\dfrac{|\overline{X} - \overline{Y} - \mu_0|}{\sqrt{\dfrac{\sigma_x^2}{n} + \dfrac{\sigma_y^2}{m}}} \le U_{\frac{\alpha}{2}}$ quyết định chấp nhận H_0.

* Nếu chưa biết σ_x^2, σ_y^2 nhưng mẫu n, m $\ge$ 30 có qui tắc sau:

Qui tắc 17: Nếu: $\dfrac{|\overline{X} - \overline{Y} - \mu_0|}{\sqrt{\dfrac{\hat{S}_x^2}{n} + \dfrac{\hat{S}_y^2}{m}}} > U_{\frac{\alpha}{2}}$ quyết định bác bỏ H_0

Nếu: $\dfrac{|\overline{X} - \overline{Y} - \mu_0|}{\sqrt{\dfrac{\hat{S}_x^2}{n} + \dfrac{\hat{S}_y^2}{m}}} \le U_{\frac{\alpha}{2}}$ quyết định chấp nhận H_0.

* Chưa biết σ_X^2, σ_Y^2 nhưng biết $\sigma_x^2 = \sigma_y^2$. Ta có qui tắc sau:

Qui tắc 18: Nếu: $\dfrac{|\overline{X} - \overline{Y} - \mu_0|}{S\sqrt{\dfrac{1}{n} + \dfrac{1}{m}}} > t_{\frac{\alpha}{2}, n+m-2}$ quyết định bác bỏ H_0

Nếu: $\dfrac{|\overline{X} - \overline{Y} - \mu_0|}{S\sqrt{\dfrac{1}{n} + \dfrac{1}{m}}} \le t_{\frac{\alpha}{2}, n+m-2}$ quyết định chấp nhận H_0.

Chú ý 3: Nếu không có thông tin gì về σ_x^2 và σ_y^2 mà kích thước của mẫu n, m < 30 ta tiến hành theo một trong hai cách sau:

* *Cách 1:* Thu thập thêm dữ liệu mẫu để n, m ≥ 30 sau đó sử dụng qui tắc kiểm định như đã nêu trong mục 4.2.
* *Cách 2:* Xây dựng cặp giả thuyết đối thuyết phụ:

H'_0: $\sigma_x^2 = \sigma_y^2$

H'_1: $\sigma_x^2 \neq \sigma_y^2$ với cùng mức ý nghĩa α.

Qui tắc kiểm định cặp giả thuyết đối thuyết này sẽ trình bày ở mục sau.
Nếu H'_0 được chấp nhận ta sử dụng qui tắc như đã nêu ở mục 4.3.
Nếu H'_0 bị bác bỏ thì về mặt lý thuyết việc so sánh hai kỳ vọng trên chưa giải quyết được. (Bài toán Behrens-Fisher).

9. Kiểm định sự bằng nhau của hai kỳ vọng với đối thuyết một phía.

Từ hai mẫu ngẫu nhiên độc lập:

$(X_1, X_2, \ldots X_i, \ldots X_n)$; $X_i \sim N(\mu_x, \sigma_x^2)$,

$(Y_1, Y_2, \ldots Y_j, \ldots Y_n)$; $Y_j \sim N(\mu_y, \sigma_y^2)$.

Ta xây dựng qui tắc kiểm định cặp giả thuyết đối thuyết :

H_0: $\mu_X = \mu_Y$

H_1: $\mu_X > \mu_Y$

9.1 Trường hợp 1: Khi σ_x^2, σ_y^2 đã biết.

Thống kê $Z = \dfrac{\overline{X} - \overline{Y} - (\mu_X - \mu_Y)}{\sqrt{\dfrac{\sigma_X^2}{n} + \dfrac{\sigma_Y^2}{m}}}$ có phân phối chuẩn tắc.

Nếu H_0 đúng thì $Z = \dfrac{\overline{X} - \overline{Y}}{\sqrt{\dfrac{\sigma_X^2}{n} + \dfrac{\sigma_Y^2}{m}}}$ có phân phối chuẩn tắc.

Từ đây ta có qui tắc kiểm định cặp giả thuyết đối thuyết trên là:

Qui tắc 19: Nếu: $\dfrac{\overline{X} - \overline{Y}}{\sqrt{\dfrac{\sigma_X^2}{n} + \dfrac{\sigma_Y^2}{m}}} > U_\alpha$ quyết định bác bỏ H_0

Nếu: $\dfrac{\overline{X} - \overline{Y}}{\sqrt{\dfrac{\sigma_X^2}{n} + \dfrac{\sigma_Y^2}{m}}} \leq U_\alpha$ quyết định chấp nhận H_0

9.2. Trường hợp 2: Khi σ_x^2, σ_y^2 chưa biết nhưng m, n ≥ 30.

Tương tự như trên ta có qui tắc kiểm định là:

Qui tắc 20: Nếu: $\dfrac{\overline{X} - \overline{Y}}{\sqrt{\dfrac{\hat{S}_X^2}{n} + \dfrac{\hat{S}_Y^2}{m}}} > U_\alpha$ quyết định bác bỏ H_0

Nếu: $\dfrac{\overline{X}-\overline{Y}}{\sqrt{\dfrac{\hat{S}_X^2}{n}+\dfrac{\hat{S}_Y^2}{m}}} \le U_\alpha$ quyết định chấp nhận H_0

9.3 Trường hợp 3: Khi σ_x^2, σ_y^2 chưa biết nhưng biết $\sigma_x^2 = \sigma_y^2$

Do biến $Z = \dfrac{\overline{X}-\overline{Y}-(\mu_X-\mu_Y)}{S\sqrt{\dfrac{1}{n}+\dfrac{1}{m}}}$ có phân phối T_{n+m-2}

Nếu H_0 đúng thì $Z = \dfrac{\overline{X}-\overline{Y}}{S\sqrt{\dfrac{1}{n}+\dfrac{1}{m}}}$ có phân phối T_{n+m-2}

Từ đây ta có qui tắc chấp nhận và bác bỏ H_0 là:

Qui tắc 21: Nếu $\dfrac{\overline{X}-\overline{Y}}{S\sqrt{\dfrac{1}{n}+\dfrac{1}{m}}} > t_{\alpha,n+m-2}$ quyết định bác bỏ H_0

Nếu $\dfrac{\overline{X}-\overline{Y}}{S\sqrt{\dfrac{1}{n}+\dfrac{1}{m}}} \le t_{\alpha,n+m-2}$ quyết định chấp nhận H_0.

Chú ý: Để kiểm định cặp giả thuyết đối thuyết

H_0 : $\mu_X = \mu_Y + \mu_0$

H_1 : $\mu_X > \mu_Y + \mu_0$

* Trường hợp σ_x^2, σ_y^2 đã biết qui tắc kiểm định là:

Qui tắc 22 : Nếu $\dfrac{\overline{X}-\overline{Y}-\mu_0}{\sqrt{\dfrac{\sigma_X^2}{n}+\dfrac{\sigma_Y^2}{m}}} > U_\alpha$ quyết định bác bỏ H_0

Nếu: $\dfrac{\overline{X}-\overline{Y}-\mu_0}{\sqrt{\dfrac{\sigma_X^2}{n}+\dfrac{\sigma_Y^2}{m}}} \le U_\alpha$ quyết định chấp nhận H_0

Trường hợp chưa biết σ_x^2, σ_y^2 nhưng n, m ≥ 30 ta có qui tắc kiểm định sau:

Qui tắc 23: Nếu: $\dfrac{\overline{X}-\overline{Y}-\mu_0}{\sqrt{\dfrac{\hat{S}_X^2}{n}+\dfrac{\hat{S}_Y^2}{m}}} > U_\alpha$ quyết định bác bỏ H_0

Nếu: $\dfrac{\overline{X}-\overline{Y}-\mu_0}{\sqrt{\dfrac{\hat{S}_X^2}{n}+\dfrac{\hat{S}_Y^2}{m}}} \le U_\alpha$ quyết định chấp nhận H_0

Trường hợp chưa biết σ_x^2, σ_y^2 nhưng biết $\sigma_x^2 = \sigma_y^2$ ta có qui tắc kiểm định sau:

Qui tắc 24: Nếu : $\dfrac{\overline{X}-\overline{Y}-\mu_0}{S\sqrt{\dfrac{1}{n}+\dfrac{1}{m}}} > t_{\alpha,n+m-2}$ quyết định bác bỏ H_0

Nếu: $\dfrac{\overline{X}-\overline{Y}-\mu_0}{S\sqrt{\dfrac{1}{n}+\dfrac{1}{m}}} \le t_{\alpha,n+m-2}$ quyết định chấp nhận H_0.

Ví dụ: Theo dõi 10 thửa ruộng trồng giống lúa A với năng suất X ta được:

$\bar{x} = 6{,}2$ tấn/ ha, $s_x^2 = 0{,}16$.

Theo dõi 8 thửa ruộng trồng giống lúa B với năng suất Y ta được:

$\bar{y} = 5{,}4$ tấn/ ha, $s_y^2 = 0{,}20$.

Biết chi phí canh tác giống lúa A tốn kém hơn giống lúa B qui ra thóc là 0,4 tấn/ ha. Với mức ý nghĩa $\alpha = 0{,}05$, xây dựng giả thuyết và đối thuyết thích hợp để khẳng định giống lúa A có hiệu quả kinh tế hơn giống lúa B không. Biết rằng năng suất của hai giống lúa này đều có phân phối chuẩn với phương sai bằng nhau và giá thành của hai loại lúa này là như nhau.

Yêu cầu trên tương ứng với việc kiểm định cặp giả thuyết đối thuyết sau:

H_0 : $\mu_X = \mu_Y + 0{,}4$

H_1: $\mu_X > \mu_Y + 0{,}4$

Ta có $s^2 = \dfrac{(n-1)s_X^2 + (m-1)s_Y^2}{n+m-2} = \dfrac{9.0{,}16 + 7.0{,}20}{16} = 0{,}18;\ s = 0{,}42$

$$Z_T = \frac{\bar{x}-\bar{y}-\mu_0}{s\sqrt{\dfrac{1}{n}+\dfrac{1}{m}}} = \frac{6{,}2-5{,}4-0{,}4}{0{,}42\sqrt{\dfrac{1}{10}+\dfrac{1}{8}}} = 2{,}01\ ;\ t_{0{,}05{,}16} = 1{,}75$$

Do $Z_T = 2{,}01 > 1{,}75 = t_{0{,}05{,}16}$ quyết định bác bỏ H_0. Điều này có nghĩa là giống lúa A có hiệu quả kinh tế hơn giống lúa B.

10.Phương pháp so sánh cặp đôi

Xét n cặp mẫu ngẫu nhiên $(X_1,Y_1),(X_2,Y_2),....,(X_n,Y_n)$

X_i có cùng phân phối với X có phân phối chuẩn $N(\mu_X,\sigma_X^2)$.Y_i có cùng phân phối với Y có phân phối chuẩn $N(\mu_Y,\sigma_Y^2)$, X độc lập với Y.

Đặt D = X – Y , $D_i = X_i - Y_i$, $\mu_V = \mu_X - \mu_Y$, $\sigma_V^2 = \sigma_X^2 + \sigma_Y^2$

Khi đó $D \sim N(\mu_V,\sigma_V^2)$; $\overline{D} = \dfrac{1}{n}\sum_{i=1}^{n} D_i = \overline{X}-\overline{Y}$; $S_D^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(D_i - \overline{D})$

Cặp giả thuyết đối thuyết:

H_0 : $\mu_X = \mu_Y + \mu_0$

H_1 : $\mu_X \ne \mu_Y + \mu_0$

tương đương với cặp giả thuyết đối thuyết:

H_0': $\mu_V = \mu_0$

H_1': $\mu_V \ne \mu_0$

Sử dụng quy tắc 4 ta có:

Nếu $Z_T = \dfrac{\left|\overline{D} - \mu_0\right|}{S_D}\sqrt{n} > t_{\frac{\alpha}{2},n-1}$ ta quyết định bác bỏ H_0.

Nếu $Z_T \le t_{\frac{\alpha}{2},n-1}$ ta quyết định chấp nhận H_0.

Cặp giả thuyết đối thuyết:

$$H_0: \quad \mu_X = \mu_Y + \mu_0$$
$$H_1: \quad \mu_X > \mu_Y + \mu_0$$

tương đương với cặp giả thuyết đối thuyết:

$$H_0': \quad \mu_V = \mu_0$$
$$H_1': \quad \mu_V > \mu_0$$

Sử dụng *quy tắc 6* ta có:

Nếu $Z_T = \dfrac{\overline{D} - \mu_0}{S_V}\sqrt{n} > t_{\alpha,n-1}$ ta quyết định bác bỏ H_0.

Nếu $Z_T \le t_{\alpha,n-1}$ ta quyết định chấp nhận H_0.

Ví dụ: Tại một câu lạc bộ thẩm mỹ người ta quảng cáo rằng sau một khóa tập luyện giảm béo người tham gia tập luyện có thể giảm trọng lượng hơn 5 kg. Một mẫu ngẫu nhiên gồm 10 người được đo trọng lượng trước và sau khi tập luyện cho bởi bảng sau:

X	85	90	96	93	86	89	82	84	100	102
Y	80	86	90	86	81	81	78	81	91	93

Hãy lập cặp giả thuyết đối thuyết thích hợp để kiểm tra tính đúng đắn của việc quảng cáo nói trên. Biết rằng $X \sim N(\mu_X, \sigma_X^2)$, $Y \sim N(\mu_Y, \sigma_Y^2)$.

Xét cặp giả thuyết đối thuyết:

$$H_0: \quad \mu_X = \mu_Y + 5$$
$$H_1: \quad \mu_X > \mu_Y + 5$$

Đặt D = X – Y , $D_i = X_i - Y_i$, D_i nhận các giá trị sau:

V_i: 5 4 6 7 5 8 4 3 9 9

Ta có : $\overline{d} = 6,\ s_v^2 = \dfrac{40}{9} \Rightarrow s_d = 2{,}11$, $t_{0{,}05,9} = 1{,}87$

$$Z_T = \frac{\overline{d} - 5}{s_v}\sqrt{n} = \frac{6-5}{2{,}11}\sqrt{10} = 1{,}50$$

$Z_T = 1{,}50 < 1{,}87$ quyết định chấp nhận giả thuyết H_0. Điều này có nghĩa với mức ý nghĩa α = 0,05 trọng lượng của những người tập luyện giảm béo chỉ ở mức giảm đi nhỏ hơn hoặc bằng 5 kg.

Phương pháp so sánh cặp đôi thường được sử dụng trong các thí nghiệm khoa học. Để xem một phương pháp canh tác, một chế độ cho gia súc ăn, một loại thuốc mới … có tốt hơn loại cũ hay không, người ta sẽ bố trí n cặp thí nghiệm, một thực hiện theo phương pháp cũ một thực hiện theo phương pháp mới, sau thời gian thí nghiệm thu được n cặp dữ liệu từ đó đưa ra kế luận.

11.Phương pháp loại bỏ các sai số thô

Khi thu thập và kiểm tra các số liệu mẫu, do nhiều nguyên nhân chủ quan và khách quan, việc gặp phải các sai số là điều không thể tránh khỏi. Các loại sai số thường gặp trong việc điều tra và thu thập các số liệu là :

11.1. Sai số thô: Là sai số xuất hiện do vi phạm các nguyên tắc cơ bản của việc đo đạc hoặc do những sơ suất mà người thu thập số liệu gây ra một cách cố ý hoặc vô ý.

11.2. Sai số hệ thống: Là những sai số do các dụng cụ đo gây ra hoặc do kĩ thuật viên không nắm được qui tắc vận hành dụng cụ đo. Sai số loại này dễ phát hiện để loại bỏ.

11.3. Sai số ngẫu nhiên: Là sai số chịu tác động của nhiều nguyên nhân, các sai số này thường nhỏ và không chịu sự tác động của người thu thập số liệu.

Trước khi tiến hành phân tích và xử lý số liệu, việc loại bỏ các số liệu dị thường (các sai số thô) ra khỏi tập các số liệu cần xử lý là điều cần chú ý, có như vậy các thông tin thu được sau xử lý mới đảm bảo tính chính xác với độ tin cậy cao.

11.4. Phương pháp loại bỏ sai số thô:

Khi tiến hành loại bỏ sai số thô (số liệu lạ) ta cần chú ý:

- Trước tiên cần kiểm tra xem có sơ suất hoặc có vi phạm các nguyên tắc cơ bản khi thu thập số liệu không?
- Thử loại bỏ x_0 là số liệu bị nghi ngờ rồi tiến hành xử lý số liệu xem kết luận có khác so với khi giữ lại x_0 hay không? Nếu không có sai khác đáng kể thì nên giữ lại số liệu x_0.
- Nên tham khảo các tài liệu chuyên môn liên quan có thể giải thích cho việc xuất hiện số liệu lạ này sau đó mới quyết định nên giữ hay nên bỏ.

Giả sử ta có dãy số liệu: $x_0, x_1, ..., x_n$ ở đó x_0 bị nghi ngờ là số dị thường (giá trị nhỏ nhất hoặc lớn nhất) trong dãy số trên. Khi đó ta xét đại lượng: $Z_T = \dfrac{|x_0 - \overline{x}|}{s}$. Nếu

$Z_T > t_{\frac{\alpha}{2}, n-1}$ ta quyết định loại bỏ giá trị x_0 ra khỏi dãy các số liệu trên.

Nếu $Z_T \le t_{\frac{\alpha}{2}, n-1}$ ta kết luận dãy số liệu trên không có số dị thường. Trong thực tế tùy yêu cầu chính xác của việc xử lý số liệu người ta thường lấy α ở các mức 0,05 hoặc 0,01.

Việc đưa ra tiêu chuẩn loại bỏ sai số thô nói trên dựa trên giả thiết các số liệu mẫu $x_0, x_1, ..., x_n$ lấy từ tổng thể có phân phối chuẩn $N(\mu, \sigma^2)$.

Ví dụ: Người ta đo 10 trục thép do một dây chuyền cơ khí sản xuất tính được $\overline{x} = 1{,}98$, $s = 0{,}04$ trong đó có trục có đường kính lớn nhất là 2,03. Với mức $\alpha = 0{,}05$ hỏi trục có đường kính nêu trên có phải là trục dị thường không?

Ta có: $Z_T = \dfrac{|x_0 - \overline{x}|}{s}\sqrt{n} = \dfrac{|1{,}98 - 2{,}03|}{0{,}04}\sqrt{10} \approx 3{,}64$

$t_{0,025,9} = 2{,}26$, $Z_T = 3{,}64 > t_{0,025,9} = 2{,}26$ trục có đường kính 2,03 nên loại khỏi mẫu.

Chú ý: Để kiểm tra giá trị nhỏ nhất và giá trị lớn nhất của mẫu có phải là sai số thô hay không trong trường hợp kích thước mẫu không lớn ta có thể thực hiện theo qui tắc sau:

*Tính: $Z_{TM} = \dfrac{x_{max} - \overline{x}}{\hat{s}}$ hoặc $Z_{Tm} = \dfrac{\overline{x} - x_{min}}{\hat{s}}$

* Từ mức α ta tìm C_α từ bảng 7

* So sánh Z_{TM} và Z_{Tm} với C_α. Nếu $Z_{TM} > C_\alpha$ ta loại x_{max}. Nếu $Z_{Tm} > C_\alpha$ ta loại bỏ x_{min}

12.Kiểm định sự bằng nhau của hai xác suất

Xác suất xuất hiện sự kiện A trong một dãy n phép thử độc lập P(A) = p_1

Xác suất xuất hiện sự kiện B trong một dãy n phép thử độc lập P(B) = p_2

Ta xây dựng qui tắc kiểm định giả thuyết

$H_0: \quad p_1 = p_2$

12.1 Trường hợp đối thuyết hai phía $H_1 : p_1 \neq p_2$

Giả sử sau dãy n phép thử độc lập có n_A lần sự kiện A xuất hiện

Trong m phép thử khác có m_B lần sự kiện B xuất hiện

Đặt $f_1 = \frac{n_A}{n}, f_2 = \frac{m_B}{m}, f = \frac{n_A + m_B}{n+m}$

Thống kê $Z = \frac{f_1 - f_2 - (p_1 - p_2)}{\sqrt{f(1-f)(\frac{1}{n}+\frac{1}{m})}}$ có phân phối xấp xỉ chuẩn tắc

Nếu H_0 đúng thì $Z = \frac{f_1 - f_2}{\sqrt{f(1-f)(\frac{1}{n}+\frac{1}{m})}}$ có phân phối xấp xỉ chuẩn tắc nên

$$P\left[\frac{|f_1 - f_2|}{\sqrt{f(1-f)(\frac{1}{n}+\frac{1}{m})}} > U_{\frac{\alpha}{2}}\right] = \alpha.$$

Từ đây ta có qui tắc kiểm định đối với cặp giả thuyết đối thuyết:

$H_0: \quad p_1 = p_2$

$H_1: \quad p_1 \neq p_2$ ở mức ý nghĩa α là

Qui tắc 25: Nếu $\frac{|f_1 - f_2|}{\sqrt{f(1-f)(\frac{1}{n}+\frac{1}{m})}} > U_{\frac{\alpha}{2}}$ bác bỏ H_0

Nếu $\frac{|f_1 - f_2|}{\sqrt{f(1-f)(\frac{1}{n}+\frac{1}{m})}} \le U_{\frac{\alpha}{2}}$ chấp nhận H_0

Ví dụ: Điều tra một loại bệnh ở hai trại gà ta có kết quả sau:

Trại thứ nhất: Kiểm tra 500 con có 60 con mắc bệnh

Trại thứ hai: Kiểm tra 400 con có 50 con mắc bệnh

Gọi p_1, p_2 là xác suất để mỗi con gà ở trại thứ nhất và trại thứ hai mắc bệnh. Hãy kiểm định giả thuyết , đối thuyết

$H_0: \quad p_1 = p_2$

$H_1: \quad p_1 \neq p_2$ ở mức ý nghĩa $\alpha = 0{,}05$

Ta có: $f_1 = 0{,}12$, $f_2 = 0{,}125$, $f = 0{,}122$

$$Z_T = \frac{|f_1 - f_2|}{\sqrt{f(1-f)(\frac{1}{n}+\frac{1}{m})}} = \frac{0,005}{\sqrt{0,122.0,878(\frac{1}{500}+\frac{1}{400})}} \approx 0,23 \ ; \ U_{0,025} = 1,96$$

$Z_T = 0,23 < U_{0,025} = 1,96$; giả thuyết được chấp nhận.

12.2 Trường hợp đối thuyết một phía: H_1: $p_1 > p_2$

Tương tự như trường hợp trên ta có qui tắc kiểm định là

Qui tắc 26: Nếu $\frac{f_1 - f_2}{\sqrt{f(1-f)(\frac{1}{n}+\frac{1}{m})}} > U_\alpha$ bác bỏ H_0

Nếu $\frac{f_1 - f_2}{\sqrt{f(1-f)(\frac{1}{n}+\frac{1}{m})}} \le U_\alpha$ chấp nhận H_0

13.Kiểm định sự bằng nhau của hai phương sai

Giả sử X~ N($\mu_X ;\sigma_X^2$)), ($X_1, X_2, ..., X_n$) là một mẫu ngẫu nhiên tương ứng.

Y ~ N($\mu_Y ;\sigma_Y^2$)), ($Y_1, Y_2, ..., Y_n$) là một mẫu ngẫu nhiên tương ứng.

Ta xây dựng qui tắc kiểm định giả thuyết H_0: $\sigma_X^2 = \sigma_Y^2$ ở mức ý nghĩa α

13.1 Trường hợp đối thuyết một phía: H_1: $\sigma_X^2 > \sigma_Y^2$

Thống kê $Z = \frac{S_X^2}{\sigma_X^2}.\frac{\sigma_Y^2}{S_Y^2}$ có phân phối $F_{n-1,\, m-1}$. Nếu H_0 đúng thì $Z = \frac{S_X^2}{S_Y^2}$ có phân phối $F_{n-1,\, m-1}$. $\Rightarrow P(\frac{S_X^2}{S_Y^2} > F_{\alpha, n-1,\, m-1}) = \alpha$

Từ đây ta có qui tắc kiểm định cặp giả thuyết đối thuyết

H_0: $\sigma_X^2 = \sigma_Y^2$

H_1: $\sigma_X^2 > \sigma_Y^2$ ở mức ý nghĩa α là

Qui tắc 27: Nếu $\frac{S_X^2}{S_Y^2} > F_{\alpha, n-1,\, m-1}$ ta bác bỏ H_0

Nếu $\frac{S_X^2}{S_Y^2} \le F_{\alpha, n-1,\, m-1}$ ta chấp nhận H_0

13.2 Trường hợp đối thuyết 2 phía H_1: $\sigma_X^2 \neq \sigma_Y^2$

Với cặp giả thuyết đối thuyết

. H_0: $\sigma_X^2 = \sigma_Y^2$

H_1: $\sigma_X^2 \neq \sigma_Y^2$

ở mức ý nghĩa α ta có qui tắc kiểm định sau:

Qui tắc 28: Nếu: $\frac{S_X^2}{S_Y^2} \notin \left[F_{1-\frac{\alpha}{2},n-1,\,m-1}; F_{\frac{\alpha}{2},n-1,\,m-1}\right]$ ta bác bỏ H_0

Nếu: $\frac{S_X^2}{S_Y^2} \in \left[F_{1-\frac{\alpha}{2},n-1,\,m-1}; F_{\frac{\alpha}{2},n-1,\,m-1}\right]$ ta chấp nhận H_0

Ví dụ: Một mẫu gồm 17 phần tử lấy từ tổng thể có phân phối chuẩn N(μ_X ;σ_X^2) ta được: s_X^2=123,5.

Một mẫu ngẫu nhiên khác gồm 15 phần tử cũng lấy từ tổng thể có phân phối chuẩn N(μ_Y ;σ_Y^2) ta được s_Y^2 = 60,4. Với mức ý nghĩa α = 0,05 hãy kiểm định cặp giả thuyết đối thuyết

.

H_0: $\sigma_X^2 = \sigma_Y^2$

H_1: $\sigma_X^2 > \sigma_Y^2$

Ta có $Z_T = \frac{s_X^2}{s_Y^2} = \frac{123,5}{60,4} = 2,04$, $F_{0,05,16,14} = 2,48$, ta chấp nhận giả thuyệ̣t :H_0: $\sigma_X^2 = \sigma_Y^2$

14. Kiểm định sự bằng nhau của nhiều kì vọng

Việc so sánh sự bằng nhau của nhiều kì vọng là một yêu cầu khá phổ biến trong nông học, sinh học cũng như trong lâm học. Chẳng hạn như so sánh như so sánh năng suất của k giống lúa khác nhau hay so sánh chiều cao trung bình của k chủng người khác nhau...Những vấn đề vừa nêu sẽ được đề cập kĩ hơn trong giáo trình thống kê nâng cao. Trong mục này ta xây dựng phương pháp kiểm định giả thuyết đối thuyết

H_0: $\mu_1 = ... = \mu_i = ... = \mu_k$

H_1: $\exists$ i ≠ j để $\mu_i \neq \mu_j$

Giả sử :

($X_{11}, X_{12}, ..., X_{1n_1}$) là mẫu ngẫu nhiên lấy từ tổng thể có phân phối chuẩn N(μ_1 ;σ^2)

($X_{i1}, X_{i2}, ..., X_{in_i}$) là mẫu ngẫu nhiên lấy từ tổng thể có phân phối chuẩn N(μ_i ;σ^2)

($X_{k1}, X_{k2}, ..., X_{kn_k}$) là mẫu ngẫu nhiên lấy từ tổng thể có phân phối chuẩn N(μ_k ;σ^2)

$$\overline{X}_1 = \frac{1}{n_1}\sum_{j=1}^{n_1} X_{1j}, ..., \quad \overline{X}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} X_{ij}, ..., \quad \overline{X}_k = \frac{1}{n_k}\sum_{j=1}^{n_k} X_{kj}$$

$\overline{X}_i$ là trung bình của mẫu ngẫu nhiên thứ i

$SS_i = \sum_{j=1}^{n_i}(X_{ij} - \overline{X}_i)^2$ tổng bình phương độ lệch của nhóm thứ i

$SSW = \sum_{i=1}^{k} SS_i$ là tổng bình phương độ lệch bên trong các nhóm

$\overline{X} = \frac{1}{n}\sum_{i=1}^{k} n_i\overline{X}_i$ là trung bình chung, $n = \sum_{i=1}^{k} n_i$ là kích thước mẫu

$SSB = \sum n_i(\overline{X}_i - \overline{X})^2$ là tổng bình phương độ lệch giữa các nhóm

$SST = \sum_{i,j=1}^{k,n_i}(X_{ij} - \overline{X})^2$ là tổng bình phương độ lệch toàn phần

Ta có SST = SSW + SSG

$MSW = \dfrac{SSW}{n-k}$ gọi là bình phương trung bình bên trong các nhóm. Đây là một ước lượng không chệch của phương sai σ^2.

$MSB = \dfrac{SSB}{k-1}$ gọi là bình phương trung bình giữa các nhóm. Đây cũng là một ước lượng không chệch của phương sai σ^2.

Khi giả thuyết H_0 đúng người ta thường chứng minh được rằng biến $Z = \dfrac{MSB}{MSW}$ có phân phối $F_{k-1,n-k}$

Từ khẳng định trên ta có qui tắc kiểm định cặp giả thuyết đối thuyết

H_0: $\mu_1 = ... = \mu_i = ... = \mu_k$

H_1: $\exists i \neq j$ để $\mu_i \neq \mu_j$

ở mức ý nghĩa α là

Qui tắc 29: Nếu $\dfrac{MSG}{MSW} > F_{k-1;n-k;\alpha}$ ta bác bỏ H_0

Nếu $\dfrac{MSB}{MSW} \leq F_{\alpha,k-1;n-k}$ ta chấp nhận H_0

Với mẫu cụ thể

X_1	X_2		X_i		X_k
x_{11}	x_{21}		x_{i1}		x_{k1}
x_{12}	x_{22}		x_{i2}		x_{k2}
x_{1n_1}	x_{2n_2}		x_{in_i}		x_{kn_k}

Ta tiến hành thực hiện bài toán kiểm định cặp giả thuyết , đối thuyết

H_0: $\mu_1 = ... = \mu_i = ... = \mu_k$

H_1: $\exists i \neq j$ để $\mu_i \neq \mu_j$

ở mức ý nghĩa α theo các bước sau:

Bước 1: Tính các số liệu ứng với mẫu đã cho:

$$\overline{x}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} x_{ij}\ ,\ \overline{x} = \frac{1}{n}\sum_{i=1}^{k} n_i\overline{x}_i\ ,\ ss_i = \sum_{j=1}^{n_i}(x_{ij} - \overline{x}_i)^2,\ ssw = \sum_{i=1}^{k} ss_i\ ,\ ssb = \sum_{i=1}^{k} n_i(\overline{x}_i - \overline{x})^2$$

$$msb = \frac{ssg}{k-1},\ \ msw = \frac{ssw}{n-k},\ \ Z_T = \frac{ssb}{ssw}$$

Bước 2: Tìm $F_{\alpha,k-1,\,n-k}$

Bước 3: Nếu $Z_T > F_{\alpha, k-1, n-k}$ quyết định bác bỏ H_0

Nếu $Z_T \leq F_{\alpha, k-1, n-k}$ quyết định chấp nhận H_0

Ví dụ: Để so sánh năng suất của 3 giống lúa A, B, C người ta thực hiện thí nghiệm trên 20 thửa ruộng, sau khi thu hoạch ta có kết quả sau:

Giống lúa A trồng trên 7 thửa ruộng năng suất X_1 của từng thửa ruộng là:

54 57 55 58 61 54 52.

Giống lúa B trồng trên 6 thửa ruộng năng suất X_2 của từng thửa ruộng là:

50 55 58 54 61 52

Giống lúa C trồng trên 7 thửa ruộng năng suất X_3 của từng thửa ruộng là

58 60 54 56 61 60 57

ở mức ý nghĩa α = 0,05, hãy kiểm định giả thuyết

H_0: Năng suất của 3 giống lúa trên là như nhau

Biết rằng X_1, X_2, X_3 là các biến chuẩn có cùng phương sai

Ta có $\overline{x}_1 = \frac{1}{7}(54 + 57 + 55 + 58 + 61 + 54 + 52) = 56.$

$$\overline{x}_2 = \frac{1}{6}(50 + 55 + 58 + 54 + 61 + 52) = 55.$$

$$\overline{x}_3 = \frac{1}{7}(58 + 60 + 54 + 56 + 61 + 60 + 57) = 58.$$

$$\overline{x} = \frac{1}{20}(7.56 + 6.55 + 7.58) = 56{,}4.$$

$ss_1 = (54 - 56)^2+(57 - 56)^2 +(55 - 56)^2 +(58 - 56)^2 +(61 - 56)^2 +(54 - 56)^2 +(52 - 56)^2 = 55$

$ss_2 = (50 - 55)^2 + (55 - 55)^2 + (58 - 55)^2 + (54 - 55)^2 + (61 - 55)^2 + (52 - 55)^2 = 80$

$ss_3 = (58 - 58)^2+(60 - 58)^2+(54 - 58)^2 +(56 - 58)^2 +(61 - 58)^2 +(60 - 58)^2 +(57 - 58)^2 = 38$

$ssw = ss_1 + ss_2 + ss_3 = 173$

$ssb = 7.(56 - 56{,}4)^2 + 6.(55 - 56{,}4)^2 + 7.(58 - 56{,}4)^2 = 30{,}8$

$$msb = \frac{ssg}{k-1} = \frac{30{,}8}{2} = 15{,}40, \quad msw = \frac{ssw}{n-k} = \frac{173}{17} = 10{,}18$$

$$Z_T = \frac{msb}{msw} = \frac{15{,}40}{10{,}18} = 1{,}51 \quad F_{k-1,n-k,0{,}05} = F_{2,17,0{,}05} = 3{,}59$$

$Z_T = 1{,}51 < 3{,}59 = F_{0{,}05,2,17}$

Giả thuyết được chấp nhận, không có sự khác nhau về năng suất của ba giống lúa trên.

III . Kiểm định giả thuyết phi tham số

1. Kiểm định một phân phối xác suất

Theo dõi một dãy n phép thử độc lập ta có kết quả sau:

Sự kiện	A_1	A_2	...	A_i	...	A_k
Tần số	n_1	n_2	...	n_i	...	n_k

Ta tiến hành kiểm định cặp giả thuyết đối thuyết ở mức ý nghĩa α

H_0: $P(A_1) = p_1, \ \dots \ , P(A_i) = p_i, \ \dots \ , P(A_k) = p_k$

H_1: $\exists j$ để $P(A_j) \neq p_j$

Nhận thấy rằng: H_0 đúng thì $n_i \approx np_i$, để kiểm định cặp giả thuyết đối thuyết trên cần đưa ra một thống kê thích hợp để khi thống kê này vượt quá ngưỡng cho phép nào đó thì ta nói có sự khác biệt giữa n_i và np_i. Khi đó ta sẽ quyết định bác bỏ giả thuyết còn nếu ngược lại ta chấp nhận giả thuyết hay nói đúng hơn là mẫu đã cho phù hợp với giả thuyết.

1.1.Trường hợp các p_i đã biết

Người ta chứng minh được rằng nếu H_0 đúng thì thống kê $Z = \sum_{i=1}^{k}\frac{(n_i - np_i)^2}{np_i}$ có phân phối giới hạn χ^2_{k-1}. Khi thống kê này vượt qua ngưỡng $\chi^2_{k-1,\alpha}$ ta quyết định bác bỏ giả thuyết.

Qui tắc kiểm định cặp giả thuyết đối thuyết

H_0: $P(A_1) = p_1, \ \dots \ , P(A_i) = p_i, \ \dots \ , P(A_k) = p_k$

H_1: $\exists j$ để $P(A_j) \neq p_j$

ở mức ý nghĩa α là

Qui tắc 1: Nếu $\sum_{i=1}^{k}\frac{(n_i - np_i)^2}{np_i} > \chi^2_{\alpha,k-1}$ ta bác bỏ H_0

Nếu $\sum_{i=1}^{k}\frac{(n_i - np_i)^2}{np_i} \le \chi^2_{\alpha,k-1}$ ta chấp nhận H_0

Ví dụ 1: Hai cá thể ở thế hệ F_1 (cùng mang kiểu gen Aa) đem lai với nhau. Các cá thể ở thế hệ F_2 có một trong 3 kiểu gen AA, Aa, aa. Điều tra 200 cá thể ở thế hệ F_2 có:

Kiểu gen	AA	Aa	aa
Số cá thể	40	105	55

ở mức ý nghĩa $\alpha = 0{,}05$ kiểm định cặp giả thuyết , đối thuyết

H_0: Kiểu gen ở thế hệ F_2 tuân theo luật Mendel

H_1: Kiểu gen ở thế hệ F_2 không tuân theo luật Mendel

Giả thuyết H_0 tương ứng với $P(AA) = \frac{1}{4}$, $P(Aa) = \frac{2}{4}$, $P(aa) = \frac{1}{4}$

Sử dụng qui tắc 1 ta thực hiện bài toán kiểm định theo các bước sau:

Bước 1: Tính $Z_T = \sum_{i=1}^{3}\frac{(n_i - np_i)^2}{np_i} = \frac{(40-50)^2}{50} + \frac{(105-100)^2}{100} + + \frac{(55-50)^2}{50} = 2{,}75$

Bước 2: Tìm $\chi^2_{0,05\,,\,2} = 5{,}99$

Bước 3: $Z_T = 2{,}75 < 5{,}99 = \chi^2_{0,05\,,\,2}$. Theo qui tắc1 giả thuyết H_0 được chấp nhận điều này có nghĩa là mẫu đã cho phù hợp với qui luật Mendel

Chú ý: Xét $Z_T = \sum_{i=1}^{k}\frac{(n_i - np_i)^2}{np_i} = \sum_{i=1}^{k}\frac{n_i^2 - 2n_i np_i + n^2 p_i^2}{np_i}$

$$= \sum_{i=1}^{k}\frac{n_i^2}{np_i} - 2\sum_{i=1}^{k} n_i + \sum_{i=1}^{k} np_i = \sum_{i=1}^{k}\frac{n_i^2}{np_i} - n \quad (1)$$

Khi sử dụng qui tắc1 ta có thể tính Z_T theo công thức (1)

1.2 Khi các p_i phụ thuộc vào r tham số chưa biết (r < k-1)

Giả sử $p_i = p_i(\theta_1, \theta_2, ..., \theta_r)$ là các hàm phụ thuộc vào r tham số $\theta_1, \theta_2, ..., \theta_r$. Để thực hiện bài toán kiểm định trong trường hợp này trước hết ta cần tìm các ước lượng điểm của θ_i theo phương pháp hợp lý nhất. Nếu $\hat{\theta}_i$ là một ước lượng điểm của θ_i thì

$\hat{p}_i = p_i(\hat{\theta}_1, \hat{\theta}_2, ..., \hat{\theta}_r)$ là ước lượng điểm của p_i. Tương tự như qui tắc 1, thống kê $Z = \sum_{i=1}^{k} \frac{(n_i - n\hat{p}_i)^2}{n\hat{p}_i}$ có phân phối giới hạn χ^2_{k-r-1} nếu H_0 đúng. Từ đây ta có qui tắc kiểm định cặp giả thuyết, đối thuyết:

H_0: $P(A_1) = p_1, \ ... \ , P(A_i) = p_i, \ ... \ , P(A_k) = p_k$

H_1: $\exists j$ để $P(A_j) \neq p_j$

ở mức ý nghĩa α là

Qui tắc 2: Nếu $\sum_{i=1}^{k} \frac{(n_i - n\hat{p}_i)^2}{n\hat{p}_i} > \chi^2_{\alpha, k-r-1}$ ta bác bỏ H_0

Nếu $\sum_{i=1}^{k} \frac{(n_i - n\hat{p}_i)^2}{n\hat{p}_i} \leq \chi^2_{\alpha, k-r-1}$ ta chấp nhận H_0

Ví dụ 2: Để xem có sự lây lan của " bệnh nấm mầm" từ cây này sang cây khác ở các cây cọ dầu hay không người ta trồng 500 cặp cây cọ dầu vào 500 hốc tại một vườn ươm cây. Sau một thời gian kiểm tra ta thu được kết quả sau:

Cả 2 cây bị bệnh	1 cây bị bệnh	0 cây bị bệnh
73	185	242

Với mức ý nghĩa $\alpha = 0{,}05$ kiểm định cặp giả thuyết đối thuyết

H_0: Không có sự lây bệnh từ cây này sang cây khác

H_1: Có sự lây bệnh từ cây này sang cây khác

Gọi p là xác suất để mỗi cây cọ dầu bị " bệnh nấm mầm "

Nếu giả thuyết H_0 đúng thì:

Xác suất để cả hai cây bị mắc bệnh $p_2 = p^2$.

Xác suất để một trong hai cây bị mắc bệnh $p_1 = 2p(1 - p)$.

Xác suất để không cây nào mắc bệnh $p_2 = (1 - p)^2$

Giả thuyết H_0 tương ứng với giả thuyết $p_2 = p^2$, $p_1 = 2p(1 - p)$, $p_0 = (1 - p)^2$. Bài toán kiểm định thực hiện theo các bước:

Bước 1: Ước lượng xác suất p bởi tần suất $f = \frac{2.73 + 185}{1000} = 0{,}331$

$\hat{p}_0 = (1 - 0{,}331)^2 = 0{,}44754$; $\hat{p}_1 = 2.0{,}331(1 - 0{,}331) = 0{,}4429$; $\hat{p}_2 = 0{,}331^2 = 0{,}10956$

$$\frac{(73 - 500\hat{p}_0)^2}{500\hat{p}_0} + \frac{(185 - 500\hat{p}_1)^2}{500\hat{p}_1} + \frac{(242 - 500\hat{p}_2)^2}{500\hat{p}_2} = \frac{18{,}22^2}{54{,}78} + \frac{36{,}45^2}{22{,}15} + \frac{18{,}23^2}{223{,}77} = 12{,}55$$

Bước 2: Tìm $\chi^2_{0.05,1} = 3{,}84$

Bước 3: $Z_T = 12{,}55 > 3{,}84 = \chi^2_{0.05,1}$

Giả thuyết H_0 bị bác bỏ, có sự lây lan " bệnh nấm mầm" từ cây này sang cây khác.

Chú ý 1: Khi sử dụng hai qui tắc vừa nêu ta phải thực hiện yêu cầu np_i hoặc $n\hat{p}_i$ ít nhất phải bằng 5

Chú ý 2: Để kiểm định giả thuyết

H_0: $X \sim F(x, \theta_1, \theta_2, ..., \theta_r)$

Khi mẫu đã cho được phân chia thành k lớp

Lớp	$<x_1$	x_1- x_2		x_{i-1}- x_i		$\geq x_{k-1}$
Tần số	n_1	n_2		n_i		n_k

Trường hợp 1: Nếu các tham số $\theta_1, \theta_2, ..., \theta_r$ đã biết.

Ta có: $p_1 = F(x_1, \theta_1, \theta_2, ..., \theta_r)$ $p_2 = F(x_2, \theta_1, \theta_2, ..., \theta_r) - F(x_1, \theta_1, \theta_2, ..., \theta_r)$

$p_i = F(x_i, \theta_1, \theta_2, ..., \theta_r) - F(x_{i-1}, \theta_1, \theta_2, ..., \theta_r)$; $p_k = 1 - (p_1 + p_2 + ... + p_{k-1})$

Nếu $np_i \geq 5$ với mọi i = 1, k ta sử dụng qui tắc 1 để thực hiện bài toán kiểm định.
Nếu có những lớp mà $np_i < 5$ ta phải thực hiện ghép lớp này vào các lớp liền kề để các $np_i \geq 5$ sau đó sử dụng qui tắc 1.

Trường hợp 2: Nếu các tham số $\theta_1, \theta_2, ..., \theta_r$ chưa biết. Ta phải tìm các ước lượng điểm của $\theta_1, \theta_2, ..., \theta_r$ theo phương pháp hợp lý nhất. Giả sử $\hat{\theta}_i$ là ước lượng điểm của θ_i.

Ta có $\hat{p}_1 = F(x_1, \hat{\theta}_1, ...\hat{\theta}_r)$, $\hat{p}_2 = F(x_2, \hat{\theta}_1, ...\hat{\theta}_r) - F(x_1, \hat{\theta}_1, ...\hat{\theta}_r)$

$\hat{p}_i = F(x_i, \hat{\theta}_1, ...\hat{\theta}_r) - F(x_{i-1}, \hat{\theta}_1, ...\hat{\theta}_r)$, $\hat{p}_k = 1 - (\hat{p}_1 + \hat{p}_2 + ... + \hat{p}_{k-1})$

Nếu với mọi i =1, k mà $n\hat{p}_i \geq 5$ ta sử dụng qui tắc 2 để thực hiện bài toán kiểm định.
Nếu có những lớp mà $np_i < 5$ ta phải thực hiện ghép lớp này vào các lớp liền kề để các $np_i \geq 5$ sau đó mới sử dụng qui tắc 2.

Chú ý 3: Việc phân các số liệu mẫu vào các lớp nếu thoả mãn các yêu cầu sau thì lực lượng của phép kiểm định sẽ lớn (xác suất sai lầm loại 2 nhỏ)

*Xác suất để X nhận giá trị trong các lớp xấp xỉ nhau.

*Nếu kích thước mẫu nhỏ hơn 100 thì số lớp k phải lớn nhất thoả mãn

$np_i \geq 5$; $n\hat{p}_i \geq 5$

*Nếu kích mẫu lớn hơn 100 thì số lớp k phải xấp xỉ $3{,}2n^{2/5}$ và yêu cầu np_i hoặc $n\hat{p}_i \geq 5$ vẫn được bảo đảm.

Ví dụ: Sản lượng của loại đậu xám(tạ/ ha) trên 36 mảnh đất gần nhau được cho bởi bảng sau:

19,2	17,7	22,0	21,1	18,5	21,0	19,3	19,0	18,2
17,1	19,2	19,1	20,1	14,3	19,5	17,3	16,3	19,6
17,5	19,1	19,7	16,0	16,7	16,4	20,0	18,8	20,8
19,3	16,0	17,4	17,2	17,6	11,4	16,3	11,5	16,1

ở mức ý nghĩa α = 0,05 có thể xem mẫu đã cho có phù hợp với giả thuyết :

H_0 : Sản lượng loại đậu xám có phân phối chuẩn $N(\mu; \sigma^2)$

Ta ước lượng μ bởi $\bar{x}$ = 17,95, ước lượng σ^2 bởi s^2 = 5,617.

Lấy các $\hat{p}_i$ bằng nhau ta có số lớp k ở đây là 7. Vậy các $\hat{p}_i$ đều bằng $\frac{1}{7}$

x_1 là số thoả mãn $\phi(\frac{x_1 - \overline{x}}{s}) = \frac{1}{7} \Rightarrow x_1 = 15{,}42$

x_2 là số thoả mãn $\phi(\frac{x_2 - \overline{x}}{s}) = \frac{2}{7} \Rightarrow x_2 = 16{,}61$

x_3 là số thoả mãn $\phi(\frac{x_3 - \overline{x}}{s}) = \frac{3}{7} \Rightarrow x_3 = 17{,}53$

x_4 là số thoả mãn $\phi(\frac{x_4 - \overline{x}}{s}) = \frac{4}{7} \Rightarrow x_4 = 18{,}37$

x_5 là số thoả mãn $\phi(\frac{x_5 - \overline{x}}{s}) = \frac{5}{7} \Rightarrow x_5 = 19{,}29$

x_6 là số thoả mãn $\phi(\frac{x_6 - \overline{x}}{s}) = \frac{6}{7} \Rightarrow x_6 = 20{,}48$

Các số liệu mẫu được xếp vào các lớp theo bảng sau:

Lớp	$< x_1$	$[x_1 ; x_2)$	$[x_2 ; x_3)$	$[x_3 ; x_4)$	$[x_4 ; x_5)$	$[x_5 ; x_6)$	$\geq x_6$
n_i	3	6	6	3	7	7	4

Sử dụng *qui tắc 2* tiến hành kiểm định theo các bước sau:

Bước 1: Tính $Z_T = \sum_{i=1}^{k} \frac{(n_i - n\hat{p}_i)^2}{n\hat{p}_i} =$

$$\frac{(3-\frac{36}{7})^2}{\frac{36}{7}} + \frac{(6-\frac{36}{7})^2}{\frac{36}{7}} + \frac{(6-\frac{36}{7})^2}{\frac{36}{7}} + \frac{(3-\frac{36}{7})^2}{\frac{36}{7}} + \frac{(7-\frac{36}{7})^2}{\frac{36}{7}} + \frac{(7-\frac{36}{7})^2}{\frac{36}{7}} + \frac{(4-\frac{36}{7})^2}{\frac{36}{7}} = 3{,}666$$

Bước 2: Do phải ước lượng hai tham số μ và σ^2 nên số bậc tự do là 4, $\chi^2_{0,05,4} = 9{,}448$.

Bước 3: $Z_T = 3{,}666 < 9{,}488 = \chi^2_{0,05,4}$ mẫu đã cho phù hợp với giả thuyết.

Chú ý: $Z_T = \sum_{i=1}^{k} \frac{(n_i - n\hat{p}_i)^2}{n\hat{p}_i} = \sum_{i=1}^{k} \frac{n_i^2 - 2n_i n\hat{p}_i + n^2\hat{p}_i^2}{n\hat{p}_i}$

$$= \sum_{i=1}^{k} \frac{n_i^2}{n\hat{p}_i} - 2\sum_{i-1}^{k} n_i + \sum_{i=1}^{k} n_i\hat{p}_i = \sum_{i=1}^{k} \frac{n_i^2}{n\hat{p}_i} - n \quad (2)$$

Khi sử dụng qui tắc 2 ta có thể tính Z_T theo công thức cho bởi (2)

2. Kiểm định tính độc lập của hai đặc tính định tính.

Xét một đám đông mỗi cá thể ta để ý tới hai đặc tính định tính A và B. Giả sử đặc tính A được chia thành k mức A_1, A_2,... A_k, đặc tính B được chia thành m mức B_1, B_2...B_m.Từ đám đông lấy ra một mẫu ngẫu nhiên có kích thước n ta có kết quả sau:

B / A	B_1	B_2		B_j		B_m
A_1	n_{11}	n_{12}		n_{1j}		n_{1m}
A_2	n_{21}	n_{22}		n_{2i}		n_{2m}
A_i	n_{i1}	n_{i2}		n_{ij}		n_{im}
A_k	n_{k1}	n_{k2}		n_{kj}		n_{km}

n_{ij} là số cá thể có đặc tính $A = A_i$ và đặc tính $B = B_j$ trong mẫu.

Từ mẫu trên xây dựng qui tắc kiểm định cặp giả thuyết, đối thuyết

H_0 : A độc lập với B.

H_1 : A không độc lập với B ở mức ý nghĩa α.

Ta có $\sum_{j=1}^{m} n_{ij} = n_{i.}$ là số cá thể có đặc tính $A = A_i$

$\sum_{i=1}^{k} n_{ij} = n_{.j}$ là số cá thể có đặc tính $B = B_j$

$$\sum_{i,j=1}^{k,m} n_{ij} = \sum_{i=1}^{k} n_{i.} = \sum_{j=1}^{m} n_{.j} = \mathrm{n}$$

Để đơn giản ta qui ước: $\sum_{i=1}^{k}\sum_{j=1}^{m} n_{ij} = \sum_{i,j=1}^{k,m} n_{ij}$)

$f_{ij} = \dfrac{n_{ij}}{n}$ là ước lượng của xác suất $P(A_iB_j)$

$f_{i.} = \dfrac{n_{i.}}{n}$ là ước lượng của xác suất $P(A_i)$

$f_{.j} = \dfrac{n_{.j}}{n}$ là ước lượng của xác suất $p(B_j)$

Nếu giả thuyết đúng thì A_i độc lập với B_j vì vậy có $P(A_iB_j) = P(A_i)P(B_j)$

$$\Rightarrow f_{ij} \approx f_{i\bullet}f_{\bullet j} \Rightarrow n_{ij} \approx \frac{n_{i\bullet}n_{\bullet j}}{n}$$

Ta đưa ra thống kê Z thích hợp để khi thống kê này vượt qua một giá trị xác định nào đó thì ta khẳng định có sự khác biệt giữa n_{ij} và $\dfrac{n_{i\bullet}n_{\bullet j}}{n}$, từ đó đưa ra quyết định bác bỏ giả thuyết. Người ta đã chứng minh được thống kê

$Z = \sum_{i,j=1}^{k,m} \dfrac{\left(n_{ij} - \dfrac{n_{i\bullet}n_{\bullet j}}{n}\right)^2}{\dfrac{n_{i\bullet}n_{\bullet j}}{n}}$ có phân phối giới hạn là $\chi^2_{(k-1)(m-1)}$ nếu giả thuyết đúng. Từ đây ta có qui tắc bác bỏ giả thuyết ở mức ý nghĩa α là:

Qui tắc 3: Nếu $\sum_{i,j=1}^{k,m}\frac{(n_{ij}-\frac{n_{i\bullet}n_{\bullet j}}{n})^2}{\frac{n_{i\bullet}n_{\bullet j}}{n}} > \chi^2_{\alpha,(k-1)(m-1)}$. ta quyết định bác bỏ H_0

Nếu $\sum_{i,j=1}^{k,m}\frac{(n_{ij}-\frac{n_{i\bullet}n_{\bullet j}}{n})^2}{\frac{n_{i\bullet}n_{\bullet j}}{n}} \le \chi^2_{\alpha,(k-1)(m-1)}$ ta quyết định chấp nhận H_0 hay mẫu đã cho phù hợp với giả thuyết A độc lập với B.

Chú ý: Khi sử dụng qui tắc 3 để kiểm định tính độc lập của hai đặc tính A, B cần đáp ứng yêu cầu $\frac{n_{i\bullet}n_{\bullet j}}{n} \ge 5$.

Ví dụ: Xét một đàn ốc sên rừng, đặc tính A là màu vỏ gồm màu vàng (A_1) và màu hồng (A_2). Đặc tính B là số vạch trên vỏ gồm : 0 vạch(B_0), 1 hoặc 2 vạch (B_1), 3 hoặc 4 vạch (B_2) và 5 vạch (B_3). Bắt ngẫu nhiên 169 con ốc sên rừng thuộc đàn ốc sên nói trên ta có bảng sau:

Số vạch / Màu vỏ	0 (B_0)	1-2 (B_1)	3-4 (B_2)	5 (B_3)
Vàng(A_1)	35	19	36	25
Hồng(A_2)	14	14	16	10

Với mức ý nghĩa α = 0,05 hãy kiểm định giả thuyết

H_0: Màu vỏ độc lập về di truyền với số vạch trên vỏ

H_1 : Màu vỏ không độc lập về di truyền với số vạch trên vỏ

Ta có: $n_{1\bullet} = 115$, $n_{2\bullet} = 54$, $n_{\bullet 1} = 49$, $n_{\bullet 2} = 33$, $n_{\bullet 3} = 52$, $n_{\bullet 4} = 35$, $n = 169$

$$Z_T = \sum_{i,j=1}^{2,3}\frac{(n_{ij}-\frac{n_{i\bullet}n_{\bullet j}}{n})^2}{\frac{n_{i\bullet}n_{\bullet j}}{n}} = \frac{(35-\frac{115.49}{169})^2}{\frac{115.49}{169}} + \frac{(19-\frac{115.33}{169})^2}{\frac{115.33}{169}} + \frac{(36-\frac{115.52}{169})^2}{\frac{115.52}{169}}$$

$$+\frac{(25-\frac{115.35}{169})^2}{\frac{115.35}{169}} + \frac{(14-\frac{54.49}{169})^2}{\frac{54.49}{169}} + \frac{(14-\frac{54.33}{169})^2}{\frac{54.33}{169}} + \frac{(16-\frac{54.52}{169})^2}{\frac{54.52}{169}} + \frac{(10-\frac{54.35}{169})^2}{\frac{54.35}{169}} = 2{,}13$$

$$\chi^2{}_{\alpha,(k-1)(m-1)} = \chi^2{}_{0,05,3} = 7{,}81$$

$Z_T = 2{,}13 < 7{,}81 = \chi^2{}_{0,05,3}$. Ta quyết định chấp nhận H_0 màu vỏ và số vạch trên vỏ độc lập với nhau về di truyền.

Chú ý 1: $Z_T = \sum_{i,j=1}^{k,m}\frac{(n_{ij}-\frac{n_{i.}n_{.j}}{n})^2}{\frac{n_{i.}n_{.j}}{n}} = \left[\sum_{i,j=1}^{k,m}\frac{nn_{ij}^2}{n_{i.}n_{.j}} - 2\sum_{i,j=1}^{k,m}n_{ij} + \frac{1}{n}\sum_{i,j=1}^{k,m}n_{i.}n_{.j}\right]$

$$= n\sum_{i,j=1}^{k,m}\frac{n_{ij}^2}{n_{i.}n_{.j}} - 2n + \frac{1}{n}\sum_{i=1}^{k}n_{i.}\sum_{j=1}^{m}n_{.j} = n\left[\sum_{i,j=1}^{k,m}\frac{n_{ij}^2}{n_{i.}n_{.j}} - 1\right] \qquad (3)$$

Khi sử dụng qui tắc 3 có thể tính Z_T bằng công thức cho bởi (3)

Chú ý 2: Việc xây dựng qui tắc kiểm định tính thuần nhất của đám đông cũng được trình bày như tiêu chuẩn vừa nêu. Tiêu chuẩn đưa ra cũng giống như tiêu chuẩn vừa nêu.

3.Quy tắc dấu

Xét n cặp mẫu ngẫu nhiên : $(X_1, Y_1), (X_2, Y_2), ..., (X_n, Y_n)$

X_i có cùng phân phối với X có hàm mật độ f(x)

Y_i có cùng phân phối với Y có hàm mật độ g(x)

Nếu X, Y là các biến chuẩn thì việc so sánh kì vọng của X và Y đã được trình bày trong phương pháp so sánh cặp đôi. Bây giờ ta đưa ra quy tắc kiểm định trong trường hợp tổng quát cặp giả thuyết đối thuyết.

H_0: X có cùng phân phối với Y

H_1: X và Y có phân phối khác nhau.

Đặt D = X - Y , $D_i = X_i - Y_i$.

Nếu H_0 đúng người ta có thể chứng minh rằng P(D > 0) = P(D < 0) = 0,5.

Gọi M là số các giá trị mà $D_i > 0$ ta thấy M có phân phối nhị thức B(n, $\frac{1}{2}$). Cặp giả thuyết đối thuyết nêu trên tương đương với cặp giả thuyết đối thuyết.

H_0': M có phân phối nhị thức B(n, $\frac{1}{2}$)

H_1': M không có phân phối nhị thức B(n, $\frac{1}{2}$)

Sử dụng định lý giới hạn: Biến Z = $\frac{M - 0,5n}{0,5\sqrt{n}}$ có phân phối giới hạn chuẩn tắc ta có quy tắc kiểm định cặp giả thuyết H_0 và H_1 là :

Qui tắc 5: Nếu $Z_T = \frac{|M - 0,5n|}{0,5\sqrt{n}} > U_{\frac{\alpha}{2}}$ bác bỏ H_0

Nếu $Z_T \le U_{\frac{\alpha}{2}}$ chấp nhận H_0

Trong thực hành khi gặp các cặp số liệu (x_i, y_i) mà $x_i = y_i$ ta loại bỏ cặp số liệu này ra khỏi mẫu.

Ví dụ: Chiều cao X của người bố và chiều cao Y của con trai tương ứng từ mẫu gồm 20 cặp bố con được cho ở bảng sau:

X	1,72	1,70	1,62	1,58	1,64	1,68	1,67	1,73	1,57	1,63
Y	1,74	1,68	1,65	1,55	1,61	1,70	1,67	1,74	1,59	1,60
X	1,74	1,76	1,58	1,67	1,55	1,68	1,71	1,58	1,75	1,65
Y	1,72	1,73	1,60	1,64	1,62	1,66	1,65	1,62	1,77	1,61

Với mức ý nghĩa α = 0,05, hãy kiểm định cặp giả thuyết đối thuyết

H_0: X có cùng phân phối xác suất với Y

H_1: X không có cùng phân phối xác suất với Y

Ta loại bỏ mẫu thứ bảy do chiều cao của cặp cha con này như nhau.
Đặt: $D = X - Y$, $d_i = x_i - y_i$
Số mẫu có d_i dương m = 11 , kích thước mẫu n = 19.

$$Z_T = \frac{|m - 0,5n|}{0,5\sqrt{n}} = 0,96 \;\; ; U_{0,025} = 1,96 \Rightarrow \text{quyết định chấp nhận } H_0.$$

4.Quy tắc Wilcoxon

4.1 Thứ tự của dãy số

Cho dãy số : $x_1, x_2, \ldots\ldots, x_n$
Gọi $u_i = rank(x_i)$ là thứ hạng của số x_i khi xếp dãy số trên theo thứ tự tăng dần.
Nếu trong dãy số $x_1, x_2, \ldots\ldots, x_n$ có các giá trị bằng nhau được xếp từ thứ tự thứ k đến thứ k +m-1 thì thứ hạng của các số giống nhau này cùng bằng $k + \frac{m}{2}$.

Ví dụ: Cho dãy số: 1,4; 1,1; 1,4; 1,1; 1,5; 1,4; 1,6; 1,8; 1,7; 1,8.
Xếp dãy số trên theo thứ tự tăng dần ta có:
1,1; 1,1; 1,4; 1,4; 1,4; 1,5; 1,6; 1,7; 1,8; 1,8.
Khi đó:
rank(1,1) = 1,5, rank(1,4) = 4, rank(1,5) = 6, rank(1,6) =7, rank(1,7)=8, rank(1,8) = 9,5.

4.2 Quy tắc Wilcoxon

Dựa vào thứ tự của dãy số mẫu, Wilcoxon đưa ra quy tắc kiểm định cặp giả thuyết đối thuyết:

H_0: X có cùng phân phối xác suất với Y
H_1: X và Y có phân phối khác nhau

Wilcoxon giải quyết bài toán trên trong trường hợp mẫu gồm n cặp:
(X_1, Y_1); (X_2, Y_2) ;……;(X_n, Y_n)
Mann và Whitney giải quyết bài toán trên trong trường hợp tổng quát với hai mẫu
$(X_1, X_2, \ldots, X_n)$ và $(Y_1, Y_2, \ldots, Y_m)$.
Gọi V_i là thứ tự của X_i trong dãy gồm n + m số:
$X_1, X_2, \ldots, X_n, Y_1, Y_2, \ldots, Y_m$

Đặt $V = \sum_{i=1}^{n} V_i$,nếu H_0 đúng có thể chứng minh rằng

$$E(V) = \frac{n(n+m+1)}{2}; \;\; D(V) = \frac{nm(n+m+1)}{12}$$

Khi đó thống kê $Z = \dfrac{V - \frac{n(n+m+1)}{2}}{\sqrt{\frac{nm(n+m+1)}{12}}}$ có phân phối xấp xỉ chuẩn tắc .

Từ đây ta có quy tắc kiểm định cặp giả thuyết đối thuyết

H_0: X có cùng phân phối xác suất với Y
H_1: X và Y có phân phối khác nhau

ở mức ý nghĩa α là :

Quy tắc 6: Nếu $Z_T = \dfrac{\left|V - \dfrac{n(n+m+1)}{2}\right|}{\sqrt{\dfrac{nm(n+m+1)}{12}}} > U_{\frac{\alpha}{2}}$ ta bác bỏ H_0.

Nếu $Z_T \le U_{\frac{\alpha}{2}}$ ta chấp nhận H_0.

Ví dụ: Theo dõi doanh thu X của 10 cửa hàng thóc giống tại Hà Tây và doanh thu Y của 12 cửa hàng thóc giống tại Thái Bình ta có kết quả sau:

X(triệu đồng/tháng): 32, 36, 28, 24, 30, 25, 32, 33, 26, 27

Y(triệu đồng/tháng): 31, 35, 27, 31, 26, 28, 34, 32, 30, 31, 26, 29

Với mức ý nghĩa α = 0,05 hãy kiểm định cặp giả thiết đối thuyết:

H_0: X có cùng phân phối xác suất với Y

H_1: X và Y có phân phối khác nhau

Ta có tổng các thứ hạng của các x_i là v = 107,5

$$\frac{n(n+m+1)}{2} = 115 \;\; ; \;\; \sqrt{\frac{nm(n+m+1)}{12}} = 15{,}165$$

$$Z_T = \frac{\left|v - \dfrac{n(n+m+1)}{2}\right|}{\sqrt{\dfrac{nm(n+m+1)}{12}}} = 0{,}45 \;\; ; \;\; U_{0,025} = 1{,}96$$

$Z_T = 0{,}45 < U_{0,025} = 1{,}96$ ta quyết định chấp nhận H_0.

4.3 Quy tắc Kruskal-Wallis

Các dữ liệu thu được từ các cuộc điều tra trong sinh học, nông học, lâm học và y học thường được thu thập từ nhiều vùng khác nhau. Ta cần kiểm tra xem các dữ liệu này có cùng xuất phát từ một tập cơ bản (cùng một tổng thể) hay không? Giả sử mẫu được thu thập từ k vùng ($k \ge 3$) và giả sử rằng dãy các giá trị mẫu:

$x_{11}, x_{12}, \ldots, x_{1n_1}$ lấy từ vùng I, có đặc tính X_1

$x_{21}, x_{22}, \ldots, x_{2n_2}$ lấy từ vùng II, có đặc tính X_2

....

$x_{k1}, x_{k2}, \ldots, x_{kn_k}$ lấy từ vùng K, có đặc tính X_k

Kích thước mẫu $n = \sum_{j=1}^{k} n_j$.

Ta gọi n_{ij} là thứ tự của số liệu x_{ij} trong n số liệu trên, $1 \le i \le k$, $1 \le j \le n_i$. Đặt $R_i = \sum_{j=1}^{n_i} n_{ij}$

Xét thống kê: $Z = \dfrac{12}{n(n+1)} \sum_{i=1}^{k} \dfrac{R_i^2}{n_i} - 3(n+1)$

Nếu $k \ge 3$, $n_i \ge 6$ thì Z có phân phối xấp xỉ phân phối khi bình phương với k-1 bậc tự do. Dựa vào quy luật phân phối xấp xỉ của biến Z với mức ý nghĩa α ta có quy tắc kiểm định cặp giả thuyết đối thuyết

H_0: Dãy các số liệu trên thu thập từ một tập cơ bản

H_1: Dãy các số liệu trên không thu thập từ một tập cơ bản

Quy tắc 7: Nếu $Z_T = \dfrac{12}{n(n+1)}\sum_{i=1}^{k}\dfrac{R_i^2}{n_i} - 3(n+1) > \chi^2_{\alpha,k-1}$ bác bỏ H_0

Nếu $Z_T \le \chi^2_{\alpha,k-1}$ chấp nhận H_0.

Quy tắc trên được gọi là quy tắc Kruskal - Wallis.

Ví dụ: Nghiên cứu tác động của 3 loại thức ăn gia súc khác nhau đối với sự tăng trọng của một loài lợn người ta tiến hành thử nghiệm trên 20 con lợn.

Gọi: X_1 là mức tăng trọng trong một tháng ở mỗi con trong nhóm 6 con lợn dùng thức ăn loại A là:

17,5 13,5 9,0 12,5 11,0 16,5

X_2 là mức tăng trọng trong một tháng ở mỗi con lợn trong nhóm 7 con lợn dùng thức ăn loại B là:

16,0 14,5 11,5 8,5 12,0 15,0 10,5

X_3 là mức tăng trọng trong một tháng ở mỗi con lơn trong nhóm 7 con lợn dùng thức ăn loại C là:

17,0 9,5 14,0 13,0 10,0 15,5 8,0

Với mức ý nghĩa α = 0,05 hãy kiểm định cặp giả thuyết đối thuyết

H_0: Ba loại thức ăn có tác dụng như nhau với sự tăng trọng của lợn

H_1: Ba loại thức ăn có tác dụng khác nhau với sự tăng trọng của lợn

Giả thuyết H_0 tương đương với các số liệu mẫu trên lấy từ một đám đông thuần nhất.

Ta có: k = 3, n_1 = 6, $n_2 = n_3$ = 7, n = 20

. $R_1 = 65,\ R_2 = 71,\ R_3 = 69$

$$Z_T = \frac{12}{n(n+1)}\sum_{i=1}^{k}\frac{R_i^2}{n_i} - 3(n+1) = 4{,}77\ ;\ \chi^2_{0{,}05\,,\,2} = 5{,}99$$

. $Z_T = 4{,}77 < 5{,}99 \Rightarrow$ giả thuyết H_0 được chấp nhận, điều nay có thể hiểu là 3 loại thức ăn trên có tác dụng như nhau với việc tăng trọng của lợn.

Chú ý: Các qui tắc kiểm định phi tham số có ưu điểm là không cần biết trước kiểu dạng phân phối xác suất của các đặc trưng ở tổng thể, nhưng do lượng lượng thông tin thu được từ tổng thể không nhiều nên lực lượng của phép kiểm định của các qui tắc này không cao.

Bài tập chương VI

1. Biết độ chịu lực X của các mẫu bê tông có phân phối chuẩn N($\mu;\sigma^2$). Đo độ chịu lực của 210 mẫu bê tông ta có kết quả sau:

Độ chịu lực X_i(kg/cm^2)	195	205	215	225	235	245
Số mẫu bê tông n_i	13	18	46	74	34	15

Với mức ý nghĩa α = 0,05, hãy kiểm định giả thuyết, đối thuyết:

H_0 : $\mu = 230$

H_1: $\mu \neq 230$ hoặc H_1: $\mu < 230$

2.Trọng lượng của mỗi gói mì ăn liền X (g/gói) do một nhà máy sản xuất là biến chuẩn với phương sai bằng 2,25. Lấy ngẫu nhiên 20 gói mì do nhà máy trên sản xuất đem cân ta có trọng lượng trung bình
$\overline{x}$ = 78,2. Với mức ý nghĩa α = 0,05 hãy kiểm định cặp giả thuyết, đối thuyết

H_0: $\mu = 80$; H_1: $\mu \neq 80$

3. Năng suất X của một giống lúa trong vùng là một biến chuẩn. Điều tra năng suất lúa trên 36 mảnh ruộng ta có kết quả sau:

X_i(tấn/ha)	5,0	5,2	5,4	5,6	5,8	6,0
Số mảnh n_i	3	5	10	9	6	3

Với mức ý nghĩa α = 0,05 hãy kiểm định cặp giả thuyết , đối thuyết

a. H_0: $\mu = 5,5$; H_1: $\mu \neq 5,5$

b. H_0: $\sigma^2 = 0,8$; H_1: $\sigma^2 > 0,8$

4. Một mẫu ngẫu nhiên gồm 600 học sinh lớp 12 các vùng nông thôn khu vực phía Bắc thấy có 122 nói sẽ nộp đơn thi vào trương Đại Học Nông nghiệp I.
Với mức ý nghĩa α = 0,05 hãy kiểm định cặp giả thuyết, đối thuyết

H_0: Tỉ lệ học sinh thi vào ĐHNNI $p = 0,20$

H_1: Tỉ lệ học sinh thi vào ĐHNNI $p > 0,20$.

5. Để so sánh năng suất của hai giống lúa A (năng suất X), giống lúa B (năng suất Y), người ta trồng từng cặp trên các loại đất khác nhau sau thu hoạch ta được kết quả sau:

Giống A(năng suất X tấn / ha)	6	7	6,5	5,5	4,3	6,6	5,8	4,9	5,3	6,5
Giống B(năng suất Y tấn / ha)	5	4	7,5	5,5	5,5	5,6	6,8	4,2	6,3	4,5

Biết X và Y là các biến chuẩn. Với mức ý nghĩa 0,05 có thể coi năng suất hai giống lúa trên là khác nhau không? Sử dụng phương pháp so sánh cặp đôi . Hãy xét trong trường hợp lấy mẫu độc lập.

6. Để xét ảnh hưởng của hai loại phân bón A, B đối với một giống lúa người ta dùng phân A bón cho lúa trên 5 thửa ruộng. Dùng phân B bón cho lúa trên 6 thửa ruộng. Sau thu hoạch ta có kết quả:

X(tạ/ha)Năng suất lúa sử dụng phân A	45	47	43	44	46	
Y(tạ/ha)Năng suất lúa sử dụng phân B	46	49	43	46	50	44

Với mức ý nghĩa 0,05 có thể coi ảnh hưởng của hai loại phân trên đối với năng suất lúa là như nhau được không? Thực hiện như bài 5.

7.Để so sánh trọng lượng của con rạ (sinh từ lần thứ hai trở đi) và trọng lượng con so (sinh lần đầu) qua thống kê ở một nhà hộ sinh ta được kết quả sau:

Trọng lượng(g)	1700-2000	2000-2300	2300-2600	2600-2900	2900-3200
Số con rạ n_i	9	13	18	42	18
Số con so m_i	5	10	22	40	45

Với mức ý nghĩa 0,05 có thể coi trọng lượng con so lớn hơn trọng lượng con rạ không?

8. Theo dõi doanh thu X , Y hàng tháng của 8 cửa hàng bán giống cây trồng tại Nam Định và 10 cửa hàng bán giống cây trồng tại Thái Bình ta được kết quả sau:

X(triệu đồng/tháng)	32	36	28	24	30	25	32	33		
Y(triệu đồng/tháng)	31	35	27	36	31	26	28	34	32	30

Với mức ý nghĩa 0,05 có thể coi doanh thu của các cửa hàng bán giống cây trồng ở hai địa phương trên là khác nhau không?

9. Một nông trường bò sữa nhập ba giống bò A, B, C. Người ta thống kê sản lượng sữa của chúng theo ba mức: ít, trung bình và nhiều sữa. Từ bảng số liệu về sự phân bố ba giống bò trên theo ba mức:

Giống bò	A	B	C
Ít sữa	92	53	75
Trung bình	37	15	19
Nhiều sữa	46	19	12

Với mức ý nghĩa 0,05 hãy nhận định xem sản lượng sữa của 3 giống bò có khác nhau không?

10. Để điều tra mức độ xem phim của nhân dân một tỉnh người ta chia mức độ xem phim thành ba cấp (nhiều , vừa, ít). Kết quả điều tra 300 hộ như sau:

Mức độ / Vùng	Nhiều	Vừa	ít
Thành phố	48	26	26
Ven nội	38	34	28
Huyện	16	10	74

Có thể coi mức độ xem phim ở ba vùng là như nhau được không? Mức ý nghĩa 0,05.

11. Khảo sát màu mắt và màu tóc của 6800 người Pháp ta được kết quả sau:

Màu tóc / Màu mắt	Vàng	Nâu	Đen	Hung
Xanh	1768	807	189	47
Đen	946	1387	746	53
Nâu	115	438	288	16

Với mức ý nghĩa 0,05 hãy kiểm định giả thuyết:

H_0: Màu tóc độc lập với màu mắt.
H_1: Màu tóc không độc lập với màu mắt.

12. Để nghiên cứu mối liên hệ giữa việc nghiện thuốc lá (đặc tính A) và huyết áp (đặc tính B) người ta tiến hành điều tra 200 người kết quả cho bởi:

B \ A	A_0(không nghiện)	A_1(nghiện nhẹ)	A_2(nghiện nặng)
B_0(huyết áp bt)	50	25	28
B_1(huyết áp cao)	30	35	32

Với mức ý nghĩa 0,05 hãy kiểm định giả thuyết :
H_0: A độc lập với B
H_1: A không độc lập với B

13. Một loài hoa có 3 giống A, B, C. Mỗi giống hoa có thể cho hoa đỏ hoặc hoa trắng. Từ số liệu thống kê:

Màu\ Loài	A	B	C
Hoa đỏ	58	102	65
Hoa trắng	102	118	75

Với mức ý nghĩa 0,05. Hay kiểm định các giả thuyết:
a. Màu hoa và giống hoa độc lập với nhau
b. Trong giống hoa B tỉ lệ giữa hoa đỏ và hoa trắng là 1 : 1

14. Điều tra 100 gia đình có hai con ta được kết quả sau:

Số con trai / Số gia đình	0	1	2
n_i	20	56	24

Với mức $\alpha=0,05$ hãy kiểm định giả thuyết:
a. H_0: Số con trai trong mỗi gia đình tuân theo phân phối nhị thức B(2 ; 0,5)
b. H_0: Số con trai trong mỗi gia đình tuân theo phân phối nhị thức B(2 ; p)

15. Một loại cây có gen A chỉ lá quăn, gen a chỉ lá phẳng, gen B hạt trắng, gen b chỉ hạt đỏ. Khi lai hai cây thuần chủng lá quăn hạt đỏ và lá thẳng hạt trắng ta được thế hệ F_1. Cho hai cá thể ở thế hệ F_1 lai với nhau ở thế hệ F_2 ta có kết quả sau:
1160 cây lá quăn hạt đỏ ; 380 cây lá quăn hạt trắng
350 cây lá thẳng hạt đỏ ; 110 cây lá thẳng hạt trắng
Với các số liệu trên ở mức ý nghĩa 0,05 hãy kiểm định cặp giả thuyết đối thuyết :
H_0: Kết quả phù hợp với qui luật phân li tính trạng 9 : 3 : 3 : 1
H_1: Trái với H_0.

16. Xét mối liên quan giữa vợ chồng và thể trạng ta có bảng số liệu sau:

Vợ / Chồng	Gầy	Béo	Trung bình
Gầy	24	12	12
Béo	10	40	15
Trung bình	20	12	115

Với mức ý nghĩa: $\alpha = 0{,}05$ hãy kiểm định cặp giả thuyết đối thuyết:

H_0: Thể trạng và mối quan hệ vợ chồng độc lập với nhau.

H_1: Thể trạng và mối quan hệ vợ chồng có liên quan với nhau.

17. Một gói mì ăn liền đạt yêu cầu về trọng lượng nếu có trọng lượng 80 gam. Kiểm tra mẫu gồm 20 gói mì được $\bar{x} = 78{,}5$, s = 2,5. Với mức ý nghĩa 0,05 hãy xây dựng giả thuyết và đối thuyết thích hợp về khâu đóng gói mì ăn liền của nhà máy đạt yêu cầu không?

18. Đo chỉ số mỡ sữa X của 130 con bò lai F_1 ta được kết quả sau

X	3,0- 3,6	3,6- 4,2	4,2– 4,8	4,8 –5,4	5,4 –6,0	6,0 – 6,6	6,6 –7,2
n_i	2	8	35	43	22	15	5

Biết chỉ số mỡ sữa trung bình của giống bò thuần chủng là 4,95. Với mức ý nghĩa 0,01. Hãy đưa ra kết luận về việc lai tạo giống biết rằng chỉ số mỡ sữa X có phân phối chuẩn.

19. Phân tích hàm lượng mùn trong một loại đất theo hai phương pháp ta có kết quả sau:

Phương pháp 1: 27,5 27,0 27,3 27,6 27,8 (đơn vị %)

Phương pháp 2: 27,9 27,2 26,5 26,3 27,0 27,4 27,3 26,8 (đơn vị %)

Với mức ý nghĩa 0,05 hãy xây dựng giả thuyết và đối thuyết thích hợp và đưa ra kết luận.

20. Người ta chiếu xạ liều 3000 Rơnghen vào một quần thể ruồi dấm thấy trong số 805 con ở thế hệ F_1 có 80 con bị đột biến. Trong khi đó cũng chiếu xạ vào một quần thể ruồi dấm khác có cho ăn kèm theo một loại đường thì trong số 2756 con ở thế hệ F_1 có 357 con bị đột biến . Với mức ý nghĩa 0,05 hãy xây dựng cặp giả thuyết đối thuyết thích hợp và đưa ra kết luận.

21. Để so sánh hai loại thức ăn đối với việc tăng trọng của lợn người ta đã tiến hành thí nghiệm trên hai mẫu :

Mẫu I cho 8 con lợn ăn loại thức ăn A sau 1 tháng được kết quả sau:

X : 12,3 13,4 14,6 11,0 16,1 11,3 12,9 10,7

Mẫu II cho 7 con lợn ăn loại thức ăn B sau 1 tháng được kết quả sau:

Y : 13,2 14,3 16,8 13,1 14,5 15,7 14,5

Với mức ý nghĩa 0,05 hãy đưa ra cặp giả thuyết đối thuyết thích hợp rồi đưa ra kết luận.

22. Để khảo sát tác dụng của việc bón phân cho ngô 70 đơn vị đạm/ha, người ta trồng liền nhau mảnh đối chứng (không bón đạm) và mảnh thực nghiệm trên 15 thửa ruộng sau khi thu hoạch được kết quả sau:

Trọng lượng mảnh đối chứng X				55,8	53,3	30,1	51,0	37,8	68,8
Trọng lượng mảnh thực nghiệm Y				60,4	58,7	28,9	48,0	39,7	68,8
X	57,7	59,1	49,4	35,4	42,7	21,2	28,3	57,3	42,4
Y	56,8	40,6	57,3	44,3	32,2	47,7	77,0	55,1	66,1

Biết X, Y là các biến chuẩn. Với mức ý nghĩa α = 0,05. Hãy xây dựng cặp giả thuyết đối thuyết thích hợp và đưa ra kết luận.

23. Điều tra 320 gia đình có 5 con ta có các số liệu sau:

Số con trai X	5	4	3	2	1	0
Số gia đình n_i	18	56	110	88	40	8

Với mức ý nghĩa α = 0,05 hãy kiểm định giả thuyết đối thuyết

H_0: Số con trai X ~ B(5, 0,5)

H_1: Trái với H_0

24. Số tai nạn giao thông xảy ra mỗi ngày X tại một thành phố được ghi trong bảng sau:

X	0	1	2	3	4	5	6	7	8
n_i	10	32	46	35	20	9	2	1	1

Với mức ý nghĩa α = 0,05 hãy kiểm định giả thuyết : Số tai nạn giao thông không xảy ra trong ngày tuân theo luật Poisson.

25. Chiều cao X của cây dầu sau 6 tháng tuổi quan sát được cho ở bảng sau:

X	24 - 30	30 - 36	36 - 42	42 - 48	48 - 54	54 - 60	60 - 66
n_i	12	24	35	47	43	32	7

Với mức ý nghĩa α = 0.05 hãy kiểm định giả thuyết X có phân phối chuẩn.

26. Một loài hoa hồng có 4 màu : đỏ, hồng, bạch và vàng. Với mẫu gồm 200 bông hoa hồng thuộc loài hoa trên ta có bảng số liệu sau:

Màu hoa	đỏ	hồng	bạch	vàng
Số hoa	27	65	75	33

Với mức ý nghĩa 0,05 hãy kiểm định giả thuyết H_0 : Các màu hoa đỏ, hồng, bạch, vàng theo tỉ lệ 1 : 2 : 2 : 1.

27. Chi phí về văn hoá X (Đơn vị 100000đ/năm) và chi phí về đi lại Y (Đơn vị 100000 đồng/năm) của 15 gia đình cho bởi bảng sau:

X	12	6,5	6,2	8,8	4,5	7,0	7,1	20	15	7,5	8,5	10,9	8,2	8	10,5
Y	5,9	6,7	4,5	4,8	10	5,5	5,2	15	7,0	4,0	5,5	8,2	5,4	8,4	7,0

Sử dụng tiêu chuẩn về dấu kiểm định giả thuyết: X và Y có cùng qui luật xác suất với mức ý nghĩa 0,05.

28. Mức tiêu thụ xăng của 3 loại xe A, B, C (lít/100km) lần lượt là X , Y, Z. Người ta cho chạy thử 7 xe A, 7 xe B và 8 xe C các số liệu thu được cho ở bảng sau:

X : 10,5 8,7 7,5 9,6 8,4 9,0 8,7

Y : 9,4 7,5 6,9 8,9 9,4 10 8,1

Z : 7,1 8,4 7,0 9,8 8,7 10 7,9 8,2

Với mức ý nghĩa 0,05 sử dụng tiêu chuẩn Kruskal – Wallis hãy kiểm định giả thuyết: Mức tiêu thụ xăng của 3 loại xe nói trên có cùng qui luật xác suất

29. Một mẫu điều tra lương của công nhân một nhà máy may X_1, lương của công nhân nhà máy chế biến hải sản X_2, lương của công nhân nhà máy sản xuất dày da xuất khẩu X_3 và lương vủa công nhân nhà máy chế biến hàng nông sản X_4 tại một khu chế suất cho bởi bảng số liệu sau: (Đơn vị 100000 đồng/tháng)

X_1 : 8,5 8,8 7,9 8,5 9,2 9,5 8,3

X_2 : 9,0 9,1 8,7 8,6 9,4 9,2 8,5 9,1

X_3 : 10 9,4 9,2 8,6 8,7 8,1 9,9

X_4 : 8,1 8,8 8,6 9,0 9,2 7,8 8,7 8,9 9,1

Ở mức ý nghĩa 0,05 sử dụng tiêu chuẩn Kruskal – Wallis hãy kiểm định giả thuyết:

Mức lương của công nhân bốn nhà máy trên là như nhau.

30. Chiều cao X của một mẫu ngẫu nhiên của 12 sinh viên nam tại Hà nội và 14 sinh viên nam tại thành phố Hồ Chí Minh cho bởi bảng số liệu sau:

X: 1,65 1,72 1,60 1,68 1,59 1,75 1,77 1,66 1,78 1,80 1,56 1,70

Y: 1,59 1,61 1,64 1,70 1,68 1,57 1,55 1,78 1,72 1,77 1,60 1,64 1,62 1,77

Ở mức ý nghĩa 0,05 sử dụng tiêu chuẩn Mann – Whitney hãy kiểm định giả thuyết:

X và Y có cùng qui luật phân phối.