PGS. TS. PHẠM XUÂN KIỀU

Giáo trình XÁC SUẤT và THỐNG KÊ

DÙNG CHO SINH VIÊN CÁC NGÀNH SINH HỌC, NÔNG - LÂM - NGƯ NGHIỆP, KINH TẾ VÀ QUẢN LÝ KINH TẾ, TÂM LÝ - GIÁO DỤC HỌC

NHÀ XUẤT BẢN GIÁO DỤC

PGS. TS. PHẠM VĂN KIỀU

Giáo trình
XÁC SUẤT VÀ THỐNG KÊ

(Dùng cho sinh viên các ngành Sinh học, Nông – Lâm – Ngư nghiệp, Kinh tế và Quản lý kinh tế, Tâm lý – Giáo dục học)

NHÀ XUẤT BẢN GIÁO DỤC

Lời nói đầu

Trong các lĩnh vực kinh tế, quân sự và các bộ môn khoa học thực nghiệm như vật lý, hoá học, sinh vật học, nông, lâm, ngư nghiệp, tâm lý, xã hội học v.v... người ta xử lý các kết quả thí nghiệm bằng phương pháp thống kê toán học hoặc biểu diễn các quy luật ngẫu nhiên bằng mô hình toán học. Do tình hình phát triển khoa học và kĩ thuật và kinh tế của nước ta đã đặt ra cho các trường đại học, cao đẳng phải đưa vào chương trình đào tạo dạy cho sinh viên tất cả các ngành nghề một số mô hình xác suất và thống kê toán học. Mức độ nội dung và phương pháp truyền thụ tuỳ thuộc vào nhu cầu của từng ngành mà đưa vào nhiều hay ít. Nhưng bất luận thế nào thì vấn đề này vẫn rất cần thiết và cấp bách. Qua nhiều năm giảng dạy cho sinh viên và Cao học các ngành sinh vật, nông – lâm nghiệp, địa lý, dân số, tâm lý học v.v... chúng tôi biên soạn cuốn Giáo trình xác suất và thống kê với thời lượng từ 45 – 60 tiết dùng cho sinh viên các trường đại học và cao đẳng kĩ thuật, kinh tế, nông lâm, ngư nghiệp, tin học, quản lý v.v... Để tiếp thu được nội dung của giáo trình này và có khả năng vận dụng tốt cho ngành nghề của mình độc giả phải hiểu biết về giải tích tổ hợp, phép tính vi tích phân, ma trận và hệ phương trình đại số tuyến tính.

Nội dung của cuốn sách được chia thành 8 chương. Chương 0 trình bày một số nội dung của giải tích tổ hợp. Chương 1, 2, 3 trình bày các khái niệm xác suất, biến ngẫu nhiên, hàm phân phối và các số đặc trưng và một số định lý giới hạn thuộc luật số lớn và định lý giới hạn trung tâm. Nó làm cơ sở khoa học để nghiên cứu về thống kê ở các chương tiếp sau.

Chương 4, 5, 6, 7 trình bày các vấn đề quan trọng của thống kê toán học như : Mẫu ngẫu nhiên, hàm phân phối mẫu, các số đặc trưng mẫu, lý thuyết ước lượng, lý thuyết kiểm định giả thiết, hồi quy và tương quan.

Sau mỗi chương đều có bài tập ứng dụng và phần hướng dẫn trả lời. Nội dung cuốn sách khá phong phú vì đã được đúc rút có chọn lọc từ các bài giảng trong nhiều năm cho sinh viên các ngành Sinh học, Nông nghiệp, Địa lý, Kinh tế, Xã hội học, Dân số học v.v... chúng tôi hy vọng cuốn giáo trình này sẽ rất bổ ích cho các độc giả trong nghiên cứu bộ môn này hoặc sử dụng xác suất thống kê vào nghiên cứu chuyên môn của mình.

Chúng tôi rất mong nhận được sự góp ý xây dựng của độc giả để lần tái bản sách được hoàn thiện hơn. Một lần nữa chúng tôi xin tỏ lòng cám ơn các đồng chí trong ban biên tập và ấn loát đã giúp chúng tôi nhiều để cuốn sách sớm được đến tay độc giả.

Tác giả

Hà Nội ngày 25 - 12 - 2004

Phần một

XÁC SUẤT

Chương 0

GIẢI TÍCH TỔ HỢP

> ***Các nội dung chính :***
> *Hoán vị, chỉnh hợp, tổ hợp, nhị thức Newton.*

0.1. HOÁN VỊ

Cho tập hợp M gồm n phần tử.

Định nghĩa 0.1. Mỗi cách sắp xếp của n phần tử của tập hợp M theo một thứ tự nhất định được gọi là một hoán vị của n phần tử đã cho.

Ký hiệu số các hoán vị khác nhau của n phần tử là P_n.

Định lý 0.1. $P_n = n!$; $(n! = 1.2.3...n)$

Ví dụ 0.1. Có bao nhiêu số khác nhau gồm 4 chữ số được thiết lập từ $\{1, 2, 3, 4\}$.

Giải : $P_4 = 4! = 24$ số.

0.2. CHỈNH HỢP

0.2.1. Chỉnh hợp không lặp

Cho tập hợp M có n phần tử, k là một số nguyên dương $1 \le k \le n$.

Định nghĩa 0.2

Mỗi cách sắp xếp k phần tử của tập hợp M theo một thứ tự nhất định được gọi là một chỉnh hợp không lặp chập k của n phần tử đã cho.

Ký hiệu số các chỉnh hợp không lặp chập k khác nhau của n phần tử đã cho là P_n^k.

Hai chỉnh hợp không lặp chập k của n phần tử là khác nhau nếu nó có ít nhất một phần tử khác nhau, hoặc chúng có thứ tự khác nhau.

Định lý 0.2

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

Ví dụ 0.2. Có bao nhiêu số khác nhau gồm 3 chữ số được thiết lập từ (1, 2, 3, 4, 5).

Giải :

Một số gồm 3 chữ số được thiết lập từ (1, 2, 3, 4, 5) tương ứng với 1 chỉnh hợp không lặp chập 3 của 5 chữ số 1, 2, 3, 4, 5.

Vậy số các số khác nhau gồm 3 chữ số được thiết lập từ 1, 2, 3, 4, 5 bằng :

$$P_5^3 = \frac{5!}{(5-3)!} = 60.$$

Ví dụ 0.3. Có mấy cách chọn ba cuốn sách từ tập hợp gồm 7 cuốn sách và xếp lên giá sách có ba chỗ.

Giải :

Mỗi cách xếp 3 cuốn sách trong 7 cuốn sách là một chỉnh hợp chập 3 của 7 phần tử đã cho.

Vậy số cách sắp xếp khác nhau 3 cuốn sách trong 7 cuốn lên giá sách có 3 chỗ trống là : $P_7^3 = \dfrac{7!}{(7-3)!} = 210$.

0.2.2. Chỉnh hợp lặp

Cho tập hợp M gồm n phần tử.

Định nghĩa 0.3. Gọi chỉnh hợp lặp chập k của n phần tử của tập M là một tập hợp có thứ tự gồm k phần tử lấy từ tập M, mà mỗi phần tử của nó có thể có mặt tới k lần.

Ví dụ 0.4. M = {1, 2}.

Lập tất cả các chỉnh hợp lặp chập 3 của 2 phần tử đã cho.

Giải :

{1, 1, 1}, {1, 1, 2}, {1, 2, 2}, {1, 2, 1}, {2, 1, 1}, {2, 1, 2}, {2, 2, 1}, {2, 2, 2}

nghĩa là có $2^3 = 8$ chỉnh hợp lặp chập 3 khác nhau của 2 phần tử đã cho.

Định lý 0.3. Số các chỉnh hợp lặp chập k khác nhau của n phần tử đã cho bằng n^k.

Ví dụ 0.5. Có bao nhiêu cách phân ngẫu nhiên 12 tặng phẩm cho 3 người.

Giải :

Mỗi cách phân 12 tặng phẩm cho 3 người là một chỉnh hợp chập 12 của 3.

Vậy số cách phân ngẫu nhiên 12 tặng phẩm cho ba người bằng 3^{12}.

0.3. TỔ HỢP

Cho tập hợp M gồm có n phần tử.

Định nghĩa 0.4. Một tập con (không kể thứ tự) gồm k phần tử ($k \leq n$) của tập M được gọi là một tổ hợp chập k của n phần tử đã cho.

Hai tổ hợp chập k của n phần tử được gọi là khác nhau nếu chúng có ít nhất 1 phần tử khác nhau.

Ký hiệu C_n^k là số các tổ hợp chập k khác nhau của n phần tử đã cho.

Định lý 0.4. $C_n^k = \dfrac{n!}{k!(n-k)!} = \dfrac{(n)_k}{k!}$

Trong đó, ký hiệu $(n)_k = n(n-1)\ldots(n-k+1)$. Quy ước $0! = 1$.

Ví dụ 0.6. Có bao nhiêu cách phân công 5 người đi lao động của lớp có 50 học sinh.

Giải :

Mỗi cách chọn ngẫu nhiên 5 người trong 50 người là 1 tổ hợp chập 5 của 50.

Vậy số cách phân công khác nhau 5 người trong 50 đi lao động là :

$$C_{50}^5 = \frac{50!}{5!(50-5)!} = \frac{50!}{5!45!} = 2118760$$

Ví dụ 0.7. Có bao nhiêu cách phân ngẫu nhiên 15 hành khách lên ba toa tầu mà toa thứ nhất có đúng 3 hành khách.

Giải :

Ta thấy có C_{15}^3 cách lấy 3 hành khách trong 15 hành khách cho vào toa thứ nhất. Số cách phân 12 hành khách còn lại lên 2 toa thứ hai và thứ ba bằng số các chỉnh hợp lặp chập 12 của 2, nghĩa là bằng 2^{12}. Vậy số cách phân ngẫu nhiên 15 hành khách lên 3 toa tàu mà toa thứ nhất có đúng 3 hành khách bằng : $C_{15}^3 \times 2^{12}$

Người ta còn mở rộng khái niệm tổ hợp chập k của x (x là một số thực bất kỳ).

Ta ký hiệu $(x)_r = x(x-1)\ldots(x-r+1)$ với r là số nguyên dương.

Đặt $C_x^0 = 1$, nếu $r < 0$ thì đặt $C_x^r = 0$.

Nếu r là số không nguyên thì C_x^r không tồn tại.

Nếu $r = 0$ thì đặt $(x)_0 = 1$.

Khi đó C_x^r được xác định bởi công thức :

$$C_x^r = \frac{(x)_r}{r!} = \frac{x(x-1)\ldots(x-r+1)}{r!}$$

Ví dụ 0.8. Tính C_{-1}^r , C_{-2}^r.

Giải :

Ta có : $C_{-1}^r = \frac{(-1)_r}{r!} = \frac{(-1)(-2)\ldots(-r)}{r!} = (-1)^r \frac{r!}{r!} = (-1)^r$.

Và $C_{-2}^r = \frac{(-2)_r}{r!} = \frac{(-2)(-3)\ldots(-(r+1))}{r!} = (-1)^r \frac{(r+1)!}{r!} = (-1)^r(r+1)$.

0.4. CÔNG THỨC NHỊ THỨC NEWTON

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

Chứng minh công thức này bằng quy nạp. Việc chứng minh công thức này được xem như bài tập.

Xét các trường hợp đặc biệt.

Ví dụ 0.9. Chứng minh rằng :

a) $0 = \sum_{k=0}^{n} (-1)^{n-k} C_n^k$

b) $2^n = \sum_{k=0}^{n} C_n^k$

Giải :

a) Từ công thức $(a+b)^n = \sum_{k=0}^{n} C_n^k a^k b^{n-k}$, đặt a = 1, b = – 1 ta có :

$$0 = \sum_{k=0}^{n} (-1)^{n-k} C_n^k .$$

b) Lấy a = b = 1 ta có : $2^n = \sum_{k=0}^{n} C_n^k$.

Bài tập chương 0

1. Chứng minh rằng :

a) $C_n^r = C_n^{n-r}$;   b) $C_{n+1}^r = C_n^{r-1} + C_n^r$

2. Chứng minh rằng :

$$C_n^r = \sum_{k=0}^{n} C_{n-m}^k C_m^{r-k}$$

3. Tìm n từ các phương trình :

a) $C_n^2 = 45$;   b) $\dfrac{P_n^4}{C_{n-1}^3} = 60$;   c) $C_n^8 = C_n^{12}$

4. a) Có mấy cách phân phối ngẫu nhiên 20 tặng phẩm cho 4 người.

b) Có mấy cách phân phối ngẫu nhiên 20 tặng phẩm cho 4 người sao cho người thứ nhất có đúng 3 tặng phẩm.

c) Có mấy cách phân phối ngẫu nhiên 20 tặng phẩm cho 4 người sao cho mỗi người có 5 tặng phẩm.

5. Trên mặt phẳng có 20 điểm (không có 3 điểm nào cũng nằm trên một đường thẳng). Qua mỗi cặp điểm ta vẽ được một đường thẳng. Hỏi có bao nhiêu đường thẳng như vậy.

6. Một học sinh phải thi 4 môn trong 10 ngày (mỗi ngày thi một môn). Có mấy cách lập chương trình thi.

7. Có bao nhiêu số khác nhau gồm 5 chữ số (chữ số đầu tiên khác không) được lập từ các chữ số từ 0 đến 9 ?

8. Có mấy cách lập một hội đồng gồm 3 người lấy trong số 4 cặp vợ chồng nếu :

a) Trong hội đồng có thể tham gia mỗi một trong 8 người trên.

b) Trong hội đồng phải có hai nữ một nam.

9. Các số 1, 2..., n lập thành một hàng ngang. Hỏi có mấy cách sắp xếp sao cho :

a) Hai chữ số 1 và 2 đứng cạnh nhau ;

b) Ba chữ số 1, 2 và 3 đứng cạnh nhau.

10. Trong hộp có 100 sản phẩm gồm có 90 sản phẩm tốt và 10 phế phẩm. Hỏi :

a) Có bao nhiêu cách lấy 10 sản phẩm từ hộp gồm 100 sản phẩm.

b) Có bao nhiêu khả năng lấy từ 100 sản phẩm ra 10 sản phẩm trong đó có 8 sản phẩm tốt và 2 phế phẩm.

11. Từ thành phố A có ba con đường đi đến thành phố B và từ B có 2 con đường đi tới thành phố C.

Hỏi có mấy cách đi từ A đến C (phải qua B).

12. Trên một vòng tròn có 12 điểm. Có mấy cách vẽ dây cung có các mút là các điểm đã cho. Có mấy tam giác nhận các điểm là các đỉnh.

13. Phân ngẫu nhiên 12 hành khách lên 3 toa tầu.

a) Có mấy cách phân ngẫu nhiên 12 hành khách lên 3 toa tầu.

b) Có mấy cách phân ngẫu nhiên 12 hành khách lên 3 toa tầu mà toa thứ nhất có đúng 4 hành khách.

c) Có mấy cách phân ngẫu nhiên 12 hành khách lên 3 toa tầu mà mỗi toa có 4 hành khách.

14. Một lô hàng có 12 sản phẩm trong đó có 8 sản phẩm tốt và 4 phế phẩm.

a) Có mấy cách lấy ngẫu nhiên 4 sản phẩm trong 12 sản phẩm đó.

b) Có mấy cách lấy ngẫu nhiên 4 sản phẩm trong 12 sản phẩm, trong đó có 3 sản phẩm tốt và 1 phế phẩm.

Trả lời bài tập chương 0

1. a) Khai triển $C_n^r = \dfrac{n!}{r!(n-r)!}$; $C_n^{(n-r)} = \dfrac{n!}{(n-r)!r!}$. Từ đó rút ra kết luận.

b) $C_n^r + C_n^{r-1} = \dfrac{n!}{r!(n-r)!} + \dfrac{n!}{(r-1)!(n-r+1)!} = \dfrac{(n+1)!}{r!(n+1-r)!} = C_{n+1}^r$

2. Kết quả thu được bằng cách khai triển hai vế của đẳng thức sau :

$$(1+t)^n = (1+t)^{n-m}(1+t)^m$$

$$\sum_{r=0}^{n} C_n^r t^r = \sum_{r=0}^{n}\sum_{k=0}^{n} C_{n-m}^k C_m^{r-k} t^r .$$

Từ đó ta suy ra đẳng thức phải chứng minh.

3. a) Từ $C_n^2 = 45$ suy ra $\dfrac{n!}{(n-2)!2!} = \dfrac{n(n-1)}{2} = 45$, suy ra $n = 10$.

b) $\dfrac{P_n^4}{C_{n-1}^3} = 6n = 60$. Vậy $n = 10$.

c) Từ $C_n^8 = C_n^{12}$ ta suy ra $12!(n-12)! = 8!(n-8)!$. Vậy $n = 20$.

4. a) 4^{20} (chính là số các chỉnh hợp lặp chập 20 của 4).

b) $C_{20}^3 \times 3^{17}$

c) $C_{20}^5 \times C_{15}^5 \times C_{10}^5 \times C_5^5$

5. Hai điểm ta vẽ được một đường thẳng. Vậy số các đường thẳng khác nhau bằng số các tổ hợp chập 2 của 20, nghĩa là có C_{20}^2 đường thẳng.

$$C_{20}^2 = 190.$$

6. Mỗi chương trình thi là 1 chỉnh hợp không lặp chập 4 của 10. Vậy số cách lập chương trình thi là : $P_{10}^4 = \dfrac{10!}{(10-4)!} = 5040$.

7. $P_{10}^5 - P_9^4$

8. a) $C_8^3 = 56$; b) $C_4^2 \times C_4^1 = 24$.

9. a) $(n-1)!$ b) $(n-2)!$

10. a) C_{100}^{10} b) $C_{90}^{8} \times C_{10}^{2}$.

11. Có 2.3 = 6 cách đi từ A đến C qua B.

12. Số dây cung vẽ được là C_{12}^2, và số tam giác là C_{12}^3.

13. a) 3^{12} ; b) $C_{12}^4 \times 2^8$; c) $C_{12}^4 \times C_8^4 \times C_4^4$.

14. a) C_{12}^4 ; b) $C_8^3 \times C_4^1$.

Chương I

BIẾN CỐ NGẪU NHIÊN VÀ XÁC SUẤT

Những nội dung chính của chương :

** Khái niệm biến cố ngẫu nhiên và xác suất.*

** Quy tắc tính xác suất của tổng các biến cố.*

** Xác suất có điều kiện, công thức xác suất của một tích các biến cố.*

** Sự độc lập các biến cố, của các phép thử.*

** Công thức xác suất nhị thức.*

Những kiến thức chuẩn bị học chương này :

** Một số vấn đề về tập hợp và giải tích tổ hợp.*

1.1. MỞ ĐẦU

Trong nhiều trường hợp, việc lặp lại một thí nghiệm với những điều kiện bên ngoài giống hệt nhau nhưng không dẫn tới cùng một kết quả.

Ví dụ 1.1. Khi gieo ngẫu nhiên một đồng tiền, ta không thể đoán trước được kết quả là mặt sấp hay mặt ngửa xuất hiện.

Ví dụ 1.2. Khi bắn một viên đạn vào bia, ta không đoán trước được viên đạn trúng bia hay không trúng bia. Vậy hiện tượng mà khi biết các điều kiện ban đầu của một thí nghiệm không đủ xác định kết quả của nó thì ta gọi hiện tượng đó là hiện tượng ngẫu nhiên.

Rất nhiều hiện tượng trong sinh học, kinh tế, kỹ thuật là các hiện tượng ngẫu nhiên (các quy luật của Men đen là một ví dụ quan trọng). Tất cả các phép đo lường đều chứa đựng sai số ngẫu nhiên. Việc nghiên cứu một đám đông dựa vào một mẫu với kích thước hạn chế cũng chứa đựng sai số ngẫu nhiên. Bởi vì mẫu không phải là hình ảnh chính xác của một đám đông ; thành phần của mẫu phụ thuộc vào việc chọn ngẫu nhiên các cá thể trong đám đông.

Việc nghiên cứu các hệ thống những hiện tượng ngẫu nhiên để từ đó rút ra các quy luật ngẫu nhiên, đó là mục tiêu của môn xác suất và thống kê toán học. Trong các ngành khoa học thực nghiệm như vật lý, hoá học, sinh học, nông, lâm nghiệp, thuỷ, hải sản, giáo dục, xã hội học, kinh tế học, ... đều sử dụng tích cực các mô hình xác suất và thống kê toán học.

1.2. BIẾN CỐ NGẪU NHIÊN

1.2.1. Phép thử ngẫu nhiên

Phép thử ngẫu nhiên là một khái niệm cơ bản không định nghĩa. Ta có thể mô tả như sau : Phép thử ngẫu nhiên là sự thực hiện một nhóm các điều kiện xác định và có thể được lặp lại nhiều lần. Kết quả của nó ta không đoán định được trước. Ta ký hiệu phép thử bằng chữ G.

Các ví dụ về phép thử :

– Gieo một lần con xúc xắc được xem như tiến hành một phép thử "Gieo xúc xắc". Kết quả của phép thử này là tập hợp $\Omega = \{B_1, B_2, B_3, B_4, B_5, B_6\}$, B_i là sự kiện mặt trên của con xúc xắc có i chấm.

– Việc lai giống bò A với giống bò B là một phép thử "lai con bò A với B".

– Gieo một hạt đậu tương được xem như tiến hành một phép thử "gieo hạt đậu tương". Kết quả của phép thử này là một tập hợp Ω = {nảy mầm, không nảy mầm}.

– Một bà mẹ sinh một con được xem như tiến hành thử một phép thử "bà mẹ sinh một con". Kết quả của phép thử này là tập hợp Ω = {Trai, gái}.

– Kiểm tra một học sinh về một môn học nào đó cũng được xem như tiến hành một phép thử "Kiểm tra một học sinh". Kết quả của phép thử này là tập hợp Ω = {Đạt, không đạt}.v.v... Vậy mỗi phép thử G tương ứng với một không gian biến cố sơ cấp Ω, các kết quả loại trừ nhau của phép thử là phần tử của Ω.

1.2.2. Biến cố ngẫu nhiên

Trong kết quả của phép thử, đặc trưng định tính của kết quả được gọi là biến cố ngẫu nhiên ; đặc trưng định lượng của kết quả của phép thử được gọi là biến ngẫu nhiên. Trở về phép thử "Bà mẹ sinh một con", đặc trưng định tính của kết quả của phép thử là con trai hoặc con gái, ta thường gọi là biến cố sinh con trai hay biến cố sinh con gái. Còn X là số con trai trong một lần sinh một con là biến ngẫu nhiên.

– Biến cố ngẫu nhiên liên kết với phép thử G là sự kiện có thể xảy ra hay không xảy ra tuỳ thuộc vào kết quả của phép thử G.

Ký hiệu biến cố ngẫu nhiên bằng chữ in hoa A, B, C ...

– Biến cố sơ cấp là sự kiện xảy ra khi và chỉ khi có một kết quả cụ thể trong số những kết quả loại trừ nhau của phép thử G. Ký hiệu là ω.

– Tập tất cả các biến cố sơ cấp tương ứng với phép thử G được gọi là không gian biến cố sơ cấp (hoặc không gian mẫu) – ký hiệu là Ω.

– Sự kiện không xảy ra trong một phép thử được gọi là biến cố không và ký hiệu là $\varnothing$.

– Sự kiện nhất định xảy ra trong một phép thử được gọi là biến cố chắc chắn, ký hiệu là Ω. Biến cố chắc chắn gồm tất cả các biến cố sơ cấp. Ta thường coi nó là không gian biến cố sơ cấp.

Ví dụ 1.3. Một bà mẹ sinh 1 con. Ký hiệu A là biến cố sinh con trai, B là biến cố sinh con gái. Không gian biến cố sơ cấp $\Omega = \{A, B\}$, còn A, B đều là biến cố sơ cấp.

Ví dụ 1.4. Gieo một lần con xúc xắc. Gọi B_i là biến cố "Mặt trên của nó có i chấm", $i = \overline{1,6}$.

Không gian biến cố sơ cấp $\Omega = \{B_1, B_2, B_3, B_4, B_5, B_6\}$.

Các $B_1, B_2 ..., B_6$ là những biến cố sơ cấp.

Mọi biến cố sơ cấp đều là biến cố ngẫu nhiên. Ngược lại biến cố ngẫu nhiên nói chung không là biến cố sơ cấp.

Ví dụ 1.5. Gọi A là biến cố "Mặt trên con xúc xắc có số chấm là số chẵn". Ta thấy $A = \{B_2, B_4, B_6\}$.

1.2.3. Quan hệ giữa các biến cố

– Biến cố A được gọi là kéo theo biến cố B nếu sự xảy ra của A kéo theo sự xảy ra của B và ký hiệu $A \subset B$.

Ví dụ 1.6. B_1 = Biến cố "Mặt trên của con xúc xắc có 1 chấm"

B = Biến cố "Mặt trên của con xúc xắc có số chấm là số lẻ".

Nghĩa là $B = \{B_1, B_3, B_5\}$.

Ta thấy : $B_1 \subset B$, nghĩa là B_1 kéo theo B.

– Hai biến cố A và B được gọi là bằng nhau nếu A kéo theo B và B kéo theo A và ký hiệu A = B.

nghĩa là $A = B \Leftrightarrow A \subset B \;\&\; B \subset A$.

– Các biến cố không đồng thời xảy ra nếu sự xuất hiện của một trong chúng loại trừ sự xuất hiện của những biến cố khác trong cùng một phép thử.

– Các biến cố đồng thời xảy ra nếu chúng có thể cùng xuất hiện trong một phép thử (còn gọi là các biến cố tương thích).

– Các biến cố được gọi là đồng khả năng nếu sự xuất hiện của biến cố này hay biến cố khác với khả năng như nhau.

Ví dụ 1.7. Gieo một lần con xúc xắc cân đối và đồng chất. Biến cố xuất hiện các mặt B_1 (một chấm), B_2 (hai chấm), B_3 (ba chấm), B_4 (bốn chấm), B_5 (năm chấm), B_6 (sáu chấm) là như nhau.

1.2.4. Các phép toán trên biến cố

– Biến cố tổng (phép cộng)

Tổng của hai biến cố A và B là biến cố mà nó xảy ra nếu ít nhất một trong hai biến cố A hoặc B xảy ra và ký hiệu $A \cup B$

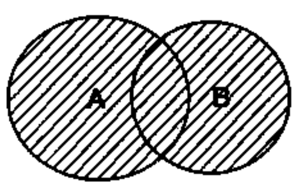
Trong ví dụ 1.4 : Đặt $A = \{B_2, B_4, B_6\}$ và $B = \{B_1, B_3, B_5\}$.

Ta có $A \cup B = \{B_1, B_2, B_3, B_4, B_5, B_6\}$.

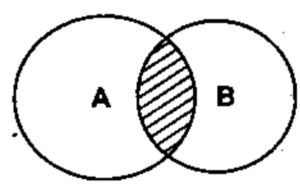
– Biến cố tích (phép nhân)

Tích của hai biến cố A và B là biến cố mà nó xảy ra nếu cả A và B đồng thời xảy ra và ký hiệu $A \cap B$ (hoặc AB).

Theo ví dụ trên ta có $A \cap B = \varnothing$.



Hình 1.1. $A \cup B$



Hình 1.2. $A \cap B$

Các tính chất của phép cộng và phép nhân :

$A \cup A = A$; $A \cup \Omega = \Omega$; $A \cup \varnothing = A$.

$A \cap A = A$; $A \cap \Omega = A$; $A \cap \varnothing = \varnothing$.

$A \cup B = B \cup A$; $A \cap B = B \cap A$.

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

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

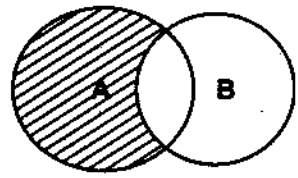
– Sự xung khắc :

Hai biến cố A và B được gọi là xung khắc với nhau nếu $A \cap B = \varnothing$.

Dãy các biến số $B_1, B_2, ..., B_n$ được gọi là xung khắc từng đôi nếu $B_i \bigcap_{i \neq j} B_j = \varnothing$, i, j = 1, ..., n .

– Biến cố hiệu (phép trừ)

Hiệu của biến cố A trừ cho biến cố B là biến cố xảy ra nếu biến cố A xảy ra và biến cố B không xảy ra và ký hiệu $A \setminus B$ (h.1.3)



Hình 1.3. A\B

Ví dụ 1.8. Đặt A = {1, 2, 3} và B = {3, 4, 5}

Ta có $A \setminus B = \{1, 2\}$, và $B \setminus A = \{4, 5\}$

Phép trừ không có tính chất giao hoán : $A \setminus B \neq B \setminus A$

– Biến cố đối lập :

Gọi hiệu $\Omega \setminus A$ là biến cố đối lập của biến cố A và ký hiệu là $\overline{A} = \Omega \setminus A$

Tính chất của biến cố đối lập :

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

– Hệ đầy đủ các biến cố :

Dãy n biến cố $B_1, B_2 ... B_n$ lập thành một hệ đầy đủ các biến cố nếu nó thoả mãn các điều kiện sau đây :

* $B_1 \cup B_2 \cup ... \cup B_n = \Omega$.
* $B_i \bigcap_{i \neq j} B_j = \varnothing$, i, j = 1, 2, ..., n.

Các ví dụ về hệ đầy đủ các biến cố :

+ Gieo một lần con xúc xắc.

Đặt B_i = biến cố "mặt trên của con xúc xắc có i chấm", $i = \overline{1, 6}$.

Dãy $B_1, B_2, B_3, B_4, B_5, B_6$ lập thành hệ đầy đủ các biến cố. Vì nó có tính chất :

$B_1 \cup B_2 \cup ... \cup B_6 = \Omega$ và $B_i \bigcap_{i \neq j} B_j = \varnothing$; i, j = 1, 2, ..., 6.

+ Gieo một đồng tiền một lần.

Đặt A = biến cố "xuất hiện mặt sấp", $\overline{A}$ là biến cố "xuất hiện mặt ngửa". Ta thấy dãy A, $\overline{A}$ thành hệ đầy đủ các biến cố. Vì $A \cup \overline{A} = \Omega$ và $A \cap \overline{A} = \varnothing$.

Chú ý Hai biến cố đối lập với nhau thì chúng xung khắc với nhau.

Điều ngược lại nói chung không đúng. Hai biến cố xung khắc với nhau nhưng không đối lập của nhau. Ví dụ : Gieo một lần con xúc xắc. Hai biến cố B_1 và B_2 là xung khắc với nhau, nhưng $\overline{B_1} = \{B_2, B_3, B_4, B_5, B_6\} \supset B_2$

Dưới đây ta có thể chính xác hoá khái niệm biến cố ở trên bằng định nghĩa theo tiên đề sau :

Định nghĩa 1.1.

1. Cho tập hợp $\Omega \neq \varnothing$. Ω được gọi là không gian biến cố sơ cấp. Phần tử $\omega \in \Omega$ được gọi là biến cố sơ cấp.

2. Cho σ – đại số $\mathcal{F}$ (đọc là xích ma đại số $\mathcal{F}$) gồm các tập con của Ω, nghĩa là tập hợp $\mathcal{F}$ các tập con của Ω có các tính chất sau :

+ $\Omega \in \mathcal{F}$;

+ Nếu $A \in \mathcal{F}$ thì $\overline{A} \in \mathcal{F}$;

+ Nếu $A_1, A_2, ..., A_n, ... \in \mathcal{F}$ thì $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$.

Phần tử A của $\mathcal{F}$ được gọi là biến cố ngẫu nhiên.

Dựa vào tính chất của phép toán tập hợp và định nghĩa của σ – đại số $\mathcal{F}$ ta suy ra $\varnothing \in \mathcal{F}$ và nếu $A, B \in \mathcal{F}$ thì $A \cap B \in \mathcal{F}$.

1.3. CÁC ĐỊNH NGHĨA XÁC SUẤT

1.3.1. Định nghĩa xác suất cổ điển

Định nghĩa 1.2.

Giả sử phép thử G tương ứng với không gian biến cố sơ cấp gồm n biến cố $e_1, e_2 ... e_n$ đồng khả năng, $\Omega = \{e_1, e_2, ... e_n\}$. Khi đó nếu biến cố A gồm có m biến cố sơ cấp $e_{i_1}, ..., e_{i_m}$ nghĩa là :

$$A = e_{i_1} \cup e_{i_2} \cup ... \cup e_{i_m},\ 1 < i_1, ..., i_m < n,$$

thì tỷ số $\frac{m}{n}$ được gọi là xác suất của biến cố A và ký hiệu $P(A) = \frac{m}{n}$, $0 \le m \le n$.

Số m các phần tử của A được gọi là số khả năng thuận lợi cho biến cố A ; còn n được gọi là số khả năng có thể.

Vậy ta có thể viết lại xác suất của biến cố A như sau :

$$P(A) = \frac{\text{số khả năng thuận lợi cho A}}{\text{số khả năng có thể}}$$

Ví dụ 1.9. Gieo một lần con xúc xắc cân đối và đồng chất. Tìm xác suất để :

a) Mặt trên của nó có 1 chấm.

b) Mặt trên của nó có số chấm là số chẵn.

Giải :

a) Đặt A = Biến cố "mặt trên của con xúc xắc có 1 chấm".

Vì con xúc xắc cân đối và đồng chất nên khả năng xuất hiện các mặt $B_1, B_2, B_3, B_4, B_5, B_6$ là như nhau. Vậy số khả năng có thể n = 6 và số khả năng thuận lợi cho A bằng m = 1.

Vậy xác suất của biến cố A là : $P(A) = \frac{1}{6}$

b) Đặt $B = \{B_2, B_4, B_6\}$. Số khả năng thuận lợi cho B là 3.

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

Ví dụ 1.10. Xét một đặc tính do một cặp gen A và a gây ra. Trong việc lai tạo thì bố mẹ mỗi người cho một gen. Nếu cả hai người đều mang gen dị hợp tử, nghĩa là hợp tử Aa, thì các hợp tử của con sẽ là một trong bốn loại sau : AA, aA, Aa, aa.

Tìm xác suất để con có các kiểu gen : aa, aA, AA.

Giải :

Vì 4 biến cố aa, aA, Aa, AA là đồng khả năng, do đó xác suất để con có kiểu gen aa là 1/4 .

Xác suất để con có kiểu gen AA là 1/4.

Nếu gộp aA, Aa vào một kiểu gen, thì xác suất để con có kiểu gen aA bằng 2/4 = 1/2.

Ví dụ 1.11. Một lô sản phẩm gồm N sản phẩm trong đó có M sản phẩm tốt và N – M phế phẩm. Lấy ngẫu nhiên s sản phẩm từ lô hàng. Tìm xác suất để trong s sản phẩm lấy ra có đúng k sản phẩm tốt.

Giải :

Số khả năng lấy s sản phẩm trong N sản phẩm bằng C_N^s.

Số khả năng lấy k sản phẩm tốt trong M sản phẩm là C_M^k.

Số khả năng lấy s sản phẩm từ lô hàng trong đó có k sản phẩm tốt và s – k sản phẩm xấu (phế phẩm) là $C_M^k \times C_{N-M}^{s-k}$.

Vậy xác suất phải tìm là :

$$P(A) = \frac{C_M^k \times C_{N-M}^{s-k}}{C_N^s} \qquad (1.1)$$

Ví dụ 1.12. 25 hành khách lên ngẫu nhiên 5 toa tầu. Tìm xác suất để :

a) Toa thứ nhất có đúng 4 hành khách.

b) Mỗi toa có 5 hành khách.

Giải :

a) Số cách phân ngẫu nhiên 25 hành khách lên 5 toa tầu là 5^{25}.

Số cách phân ngẫu nhiên 25 hành khách lên 5 toa tầu mà toa thứ nhất có 4 hành khách là $C_{25}^4 \times 4^{21}$.

Xác suất phải tìm là $P(A) = \dfrac{C_{25}^4 \times 4^{21}}{5^{25}}$.

b) Số cách phân ngẫu nhiên 25 hành khách lên 5 toa tầu mà mỗi toa có 5 hành khách $C_{25}^5 \times C_{20}^5 \times C_{15}^5 \times C_{10}^5 \times C_5^5 = \dfrac{25!}{(5!)^5}$

Xác suất phải tìm là : $P(B) = \dfrac{25!/(5!)^5}{5^{25}} = \dfrac{25!}{(5!)^5.5^{25}}$.

Ví dụ 1.13. Lấy ngẫu nhiên lần lượt 3 chữ số từ tập hợp 5 chữ số {0, 1, 2, 3, 4}, xếp thành hàng ngang từ trái sang phải. Tìm xác suất để nhận được một số gồm 3 chữ số (không kể chữ số 0 đứng đầu).

Giải :

Số các số gồm 3 chữ số (kể cả chữ số 0 đứng đầu) có thể lập được là số các chỉnh hợp không lặp chập 3 của 5 :

$$P_5^3 = \frac{5!}{(5-3)!} = 60.$$

Số các số gồm 3 chữ số mà chữ số 0 đứng đầu là :

$$P_4^2 = \frac{4!}{(4-2)!} = 3.4 = 12.$$

Số các số có 3 chữ số mà không có chữ số 0 đứng đầu là :

$$P_5^3 - P_4^2 = 60 - 12 = 48$$

Xác suất phải tìm là $P(A) = \dfrac{P_5^3 - P_4^2}{P_5^3} = \dfrac{48}{60} = \dfrac{4}{5}$.

1.3.2. Định nghĩa xác suất theo tần suất

Xét phép thử ngẫu nhiên nào đó. Biến cố A được quan sát trong phép thử này. Ta lặp lại độc lập n lần phép thử này với điều kiện như nhau. Gọi k là số lần xuất hiện biến cố A trong n phép thử đó.

Tỷ số $\frac{k}{n}$ được gọi là tần suất xuất hiện biến cố A.

Nói chung tần suất $\frac{k}{n}$ bị thay đổi nếu ta thực hiện hàng loạt các phép thử khác từ n phép thử hoặc nếu số phép thử n thay đổi. Song thực nghiệm chứng tỏ rằng khi số phép thử n càng lớn, tỷ số $\frac{k}{n}$ dao động quanh số cố định và sự khác giữa chúng càng nhỏ đi. Hai nhà thống kê Piếcsơn và Búpphông tiến hành gieo nhiều lần đồng tiền cân đối và đồng chất. Kết quả các lần gieo được cho ở bảng dưới đây.

Bảng 1

	Số lần gieo	Số lần xuất hiện mặt ngửa	Tần suất
Búpphông	4040	2048	0,5080
Piếcsơn (đợt 1)	12000	6019	0,5016
Piếcsơn (đợt 2)	24000	12012	0,5005

Qua các thực nghiệm trên ta thấy rằng số lần gieo đồng tiền càng tăng thì tần suất xuất hiện mặt ngửa càng gần $\frac{1}{2}$.

Số $\frac{1}{2}$ được gọi là xác suất xuất hiện mặt ngửa.

Bây giờ ta phát biểu định nghĩa xác suất theo quan điểm thống kê như sau :

Định nghĩa 1.3. Nếu số phép thử n càng lớn mà tần suất xuất hiện biến cố A $\frac{k}{n}$ sai khác số cố định p nào đó càng bé thì ta nói rằng biến cố A ổn định ngẫu nhiên và số p được gọi là xác suất xuất hiện biến cố A.

Định nghĩa này có ưu điểm là : nó giải quyết được trường hợp không gian biến cố sơ cấp gồm vô hạn biến cố sơ cấp và không cần giả thiết tính đồng khả năng, trong khi đó định nghĩa xác suất cổ điển chỉ áp dụng trong phạm vi không gian biến cố sơ cấp gồm hữu hạn biến cố sơ cấp đồng khả năng. Song định nghĩa xác suất theo thống kê cũng có nhược điểm nhiều về đặc trưng toán học. Nó không phản ánh được nhiều về đặc trưng của biến cố mà tỷ số $\frac{k}{n}$ có tính ổn định.

1.3.3. Định nghĩa xác suất hình học

Định nghĩa 1.4. Cho miền Ω đo được (trong đường thẳng, mặt phẳng, không gian ba chiều v.v...) và miền con S đo được của Ω. Ta lấy ngẫu nhiên một điểm trong miền Ω. Đặt A = biến cố "$M \in S$" (đọc là điểm M thuộc miền S). Xác suất của biến cố A được xác định như sau :

$$P(A) = \frac{\text{độ đo của } S}{\text{độ đo của } \Omega}$$

(Miền Ω chính là không gian biến cố sơ cấp).

– Nếu miền Ω là đường cong hay đoạn thẳng thì "độ đo" của Ω là độ dài của nó.

– Nếu miền Ω là hình phẳng hay mặt cong thì "độ đo" của Ω là diện tích của nó.

– Nếu Ω là hình khối ba chiều thì "độ đo" của Ω là thể tích của nó. v.v...

Ví dụ 1.14. Tìm xác suất để điểm M rơi vào hình tròn nội tiếp hình vuông có cạnh 2m.

Giải :

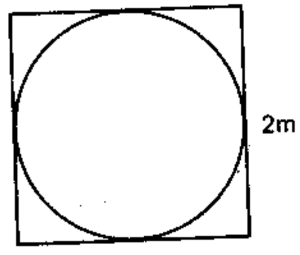
Hình tròn nội tiếp hình vuông có cạnh 2m có đường kính 2m (hình 1.4).

Vậy diện tích hình tròn đó là $\pi R^2 = \pi m^2$.

Diện tích hình vuông là $S = 2 \times 2 = 4m^2$

Xác suất phải tìm là :

$$P(A) = \frac{\text{diện tích hình tròn}}{\text{diện tích hình vuông}}$$



$$P(A) = \frac{\pi}{4}$$

Hình 1.4

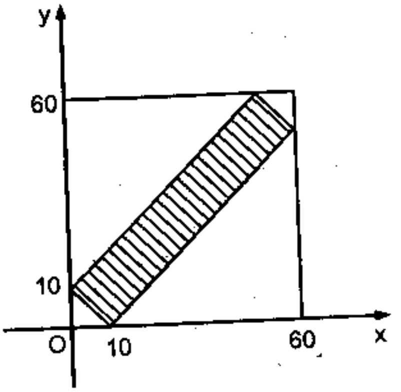
Ví dụ 1.15. Hai người hẹn gặp nhau ở một địa điểm xác định vào khoảng 8 đến 9 giờ. Người đến trước sẽ đợi người kia 10 phút ; sau đó nếu không gặp thì sẽ đi khỏi điểm hẹn. Hãy tìm xác suất để hai người gặp nhau, nếu biết rằng mỗi người có thể đến chỗ hẹn trong khoảng thời gian quy định một cách ngẫu nhiên và không tuỳ thuộc vào người kia đến vào lúc nào.

Giải :

Ký hiệu x là thời điểm người thứ nhất đến điểm hẹn, y là thời điểm người thứ hai đến điểm hẹn. Hai người gặp nhau khi và chỉ khi $|x - y| \le 10$.

Ta biểu diễn x, y như toạ độ các điểm trên mặt phẳng toạ độ Đề các vuông góc.

Đơn vị ở các trục là phút. Không gian biến cố sơ cấp Ω là hình vuông có cạnh là 60, còn biến cố sơ cấp thuận lợi cho việc hai người gặp nhau là những điểm trong miền gạch (hình 1.5).



Hình 1.5

Vậy xác suất phải tìm là :

$$P(A) = \frac{60^2 - 50^2}{60^2} = \frac{11}{36}$$

Để khắc phục các nhược điểm của các định nghĩa trên và đảm bảo độ chính xác về mặt toán học, Cômôgôrốp đã đưa ra định nghĩa xác suất bằng phương pháp tiên đề.

Định nghĩa 1.5. Hàm P xác định trên σ – đại số $\mathcal{F}$ và lấy giá trị trong $R_+ = [0, +\infty)$ được gọi là độ đo xác suất nếu thoả mãn các điều kiện sau đây :

1. $\mathcal{F} \ni A \mapsto P(A) \geq 0$.

2. $P(\Omega) = 1$

3. Nếu $A_1, A_2, \dots A_n, \in \mathcal{F}$ và $A_i \bigcap_{i \neq j} A_j = \varnothing$ thì $P(\bigcap_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n)$.

P(A) được gọi là xác suất của biến cố A. Bộ ba $(\Omega, \mathcal{F}, P)$ được gọi là không gian xác suất.

Các định nghĩa trình bày ở mục trên là trường hợp riêng của định nghĩa theo tiên đề.

1.4. CÁC TÍNH CHẤT CỦA XÁC SUẤT

1.4.1. Nếu $A, B \in \mathcal{F}$ và $A \subset B$ thì $P(A) \leq P(B)$

Thật vậy, vì $A \subset B$ nên $B = A \cup \overline{A}B$

A và $\overline{A}B$ xung khắc nên $P(B) = P(A \cup \overline{A}B) = P(A) + P(\overline{A}B)$

Vì $P(\overline{A}B) \geq 0$, do đó $P(B) \geq P(A)$

Từ định nghĩa cổ điển ta suy ra các tính chất sau :

1.4.2. Với A là biến cố bất kỳ ta có $P(A) \geq 0$

Thực vậy, vì $0 \leq m \leq n$ nên $P(A) = \dfrac{m}{n} \geq 0$

1.4.3. $P(\Omega) = 1$ vì $P(\Omega) = \dfrac{n}{n} = 1$

1.4.4. Nếu $A \cap B = \varnothing$ thì $P(A \cup B) = P(A) + P(B)$

Thật vậy, gọi m_A là số khả năng thuận lợi cho biến cố A, m_B là số khả năng thuận lợi cho biến cố B. Vì $A \cap B = \varnothing$ nên số khả năng thuận lợi cho biến cố tổng $A \cup B$ là $m_A + m_B$.

Vậy $P(A \cup B) = \dfrac{m_A + m_B}{n} = \dfrac{m_A}{n} + \dfrac{m_B}{n} = P(A) + P(B)$.

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

Vì $A \cap \overline{A} = \varnothing$ theo trên ta có $P(A \cup \overline{A}) = P(A) + P(\overline{A}) = P(\Omega) = 1$.

Vậy $P(A) + P(\overline{A}) = 1$. Ta suy ra điều phải chứng minh.

Hệ quả 1.1. $P(\varnothing) = 0$ Vì $\varnothing = (\overline{\Omega})$ nên $P(\varnothing) = P(\overline{\Omega})$.

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

1.4.6. Nếu A, B là hai biến cố bất kỳ thì

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

Chứng minh :

Ta có $A \cup B = A \cup \overline{A}B$. Vì A, $\overline{A}B$ xung khắc nên

$$P(A \cup \overline{A}B) = P(A) + P(\overline{A}B) \qquad (*)$$

Mặt khác $B = AB \cup \overline{A}B$. Do đó $P(B) = P(AB) + P(\overline{A}B)$

Từ đó suy ra : $P(\overline{A}B) = P(B) - P(AB)$.

Thay vào (*) ta nhận được $P(A \cup B) = P(A) + P(B) - P(AB)$.

Ta có thể mở rộng cho trường hợp tổng của n biến cố.

$$P(A_1 \cup A_2 \cup \ldots \cup A_n) =$$

$$= \sum_{i=1}^{n} P(A_i) - \sum_{i \neq j} P(A_i)P(A_j) + \sum_{i \neq j \neq k} P(A_iA_jA_k) - \ldots + (-1)^{n-1}P(A_1A_2 \ldots A_n)$$

1.4.7. Với A, B là hai biến cố bất kỳ ta có : $P(A \setminus B) = P(A) - P(AB)$

Chú ý : Từ định nghĩa xác suất theo tiên đề thì tính chất a, b, c, chính là tiên đề 1, 2, 3. Đồng thời ta sử dụng định nghĩa xác suất theo tiên đề cũng chứng minh được các tính chất còn lại.

1.5. XÁC SUẤT CÓ ĐIỀU KIỆN, CÔNG THỨC XÁC SUẤT CỦA TÍCH CÁC BIẾN CỐ, SỰ ĐỘC LẬP CỦA CÁC BIẾN CỐ

1.5.1. Xác suất có điều kiện

Định nghĩa 1.6. Giả sử A, B là hai biến cố bất kỳ và $P(A) > 0$.

Ta gọi tỷ số $\dfrac{P(AB)}{P(A)}$ là xác suất có điều kiện của biến cố B với điều kiện biến cố A đã xảy ra và ký hiệu :

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

Tương tự, nếu $P(B) > 0$ ta gọi tỷ số $\frac{P(AB)}{P(B)}$ là xác suất có điều kiện của biến cố A với điều kiện biến cố B đã xảy ra và ký hiệu $P(A / B) = \frac{P(AB)}{P(B)}$.

Ví dụ 1.16. Gieo một lần con xúc xắc cân đối và đồng chất.

Ký hiệu A là biến cố "mặt trên có 1 chấm hoặc 2 chấm hoặc 3 chấm". Và B là biến cố "Mặt trên có 3 chấm, hoặc 4 chấm hoặc 5 chấm". Tính xác suất của $A \cup B$, của A, của B, của AB, của $A \setminus B$.

Giải :

$$P(A) = \frac{3}{6} = \frac{1}{2}\ ;\ P(B) = \frac{3}{6} = \frac{1}{2}.$$

Ta có $AB = \{B_3\}$ (B_3 – mặt trên có 3 chấm).

$$P(AB) = \frac{1}{6}.$$

Theo quy tắc xác suất của tổng các biến cố ta có :

$$P(A \cup B) = P)A) + P(B) - P(AB) = \frac{1}{2} + \frac{1}{2} - \frac{1}{6} = \frac{5}{6}.$$

Tương tự $P(A \setminus B) = P(A) - P(AB) = \frac{1}{2} - \frac{1}{6} = \frac{3-1}{6} = \frac{1}{3}$.

Ví dụ 1.17. Một lô sản phẩm gồm có 12 sản phẩm trong đó có 8 sản phẩm tốt và 4 phế phẩm.

1. Rút ngẫu nhiên liên tiếp không hoàn lại hai sản phẩm từ lô hàng. Tìm xác suất để cả hai sản phẩm đó là sản phẩm tốt.

2. Rút ngẫu nhiên một sản phẩm từ lô hàng và không để ý tới sản phẩm đó. Sau đó rút tiếp sản phẩm thứ hai.

Tìm xác suất để sản phẩm lấy ra lần thứ hai là sản phẩm tốt.

Giải :

Theo định nghĩa xác suất có điều kiện ta có $P(B / A) = \frac{P(AB)}{P(A)}$.

Ta suy ra $P(AB) = P(A)\,P(B / A)$.

Đặt A = biến cố "sản phẩm lấy ra lần I là sản phẩm tốt".

và B = biến cố "sản phẩm lấy ra lần II là sản phẩm tốt".

Xác suất phải tìm là P(AB) = P(A) P(B / A)

Ta có $P(A) = \frac{8}{12}$ và $P(B/A) = \frac{7}{11}$.

Vậy $P(AB) = \frac{8}{12} \times \frac{7}{11} = \frac{14}{33}$.

2. Theo ký hiệu ở câu 1, ta có :

$$B = \Omega \cap B = (A \cup \overline{A})B = AB \cup \overline{A}B.$$

$$P(B) = P(AB) + P(\overline{A}B)$$

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

Ta có $P(A) = \frac{8}{12}$ và $P(B/A) = \frac{7}{11}$, $P(\overline{A}) = \frac{4}{12}$; $P(B/\overline{A}) = \frac{8}{11}$.

Vậy $P(B) = \frac{8}{12} \times \frac{7}{11} + \frac{4}{12} \times \frac{8}{11} = \frac{8.11}{12.11} = \frac{2}{3}$.

Từ định nghĩa xác suất có điều kiện ta suy ra công thức xác suất của tích hai biến cố :

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

Công thức này có thể mở rộng ra công thức xác suất của tích n biến cố.

$$P(A_1A_2\ldots A_n) = P(A_1)P(A_2/A_1)\ldots P(A_n/A_1, \ldots, A_{n-1})$$

Ví dụ 1.18. Một lô sản phẩm có 100 sản phẩm trong đó có 90 sản phẩm tốt và 10 phế phẩm. Kiểm tra ngẫu nhiên liên tiếp không hoàn lại 5 sản phẩm. Nếu có ít nhất 1 phế phẩm trong 5 sản phẩm kiểm tra đó thì không nhận lô hàng. Tìm xác suất để nhận lô hàng.

Giải :

Đặt A_i = biến cố "sản phẩm kiểm tra thứ i là sản phẩm tốt", $i = \overline{1,5}$

Đặt A = biến cố " nhận lô hàng". Ta thấy $A = A_1A_2A_3A_4A_5$.

Theo công thức xác suất của tích các biến cố ta có :

$$P(A) = P(A_1)P(A_2/A_1)P(A_3/A_1, A_2)P(A_4/A_1\ldots A_3)P(A_5/A_1, \ldots, A_4).$$

$$P(A) = \frac{90}{100} \times \frac{89}{99} \times \frac{88}{98} \times \frac{87}{97} \times \frac{86}{96}.$$

1.5.2. Công thức xác suất toàn phần và công thức Bayes

Mệnh đề : Giả sử A là biến cố bất kỳ và $B_1, B_2, \ldots B_n$ lập thành hệ đầy đủ các biến cố và $P(B_i) > 0$. Khi đó :

1. $$P(A) = \sum_{i=1}^{n} P(B_i)P(A/B_i). \qquad (1.2)$$

Đây là công thức xác suất toàn phần.

2. Nếu $P(A) > 0$ thì $$P(B_k/A) = \frac{P(B_k)P(A/B_k)}{\sum_{i=1}^{n} P(B_i)P(A/B_i)} \qquad (1.3).$$

Công thức này được gọi là công thức Bayes.

Chứng minh :

1. Ta có :

$$A = A \cap \Omega = A\left(B_1 \cup B_2 \cup \ldots \cup B_n\right) = AB_1 \cup AB_2 \cup \ldots \cup AB_n$$

Vì các $B_1, \ldots, B_n$ là xung khắc từng đôi nên $AB_1, AB_2, \ldots, AB_n$ cũng xung khắc từng đôi nên :

$$P(A) = P(AB_1) + P(AB_2) + \ldots + P(AB_n)$$

$$P(A) = P(B_1)P(A/B_1) + P(B_2)P(A/B_2) + \ldots + P(B_n)P(A/B_n).$$

2. Theo công thức xác suất tính ta có :

$$P(B_kA) = P(B_k)\,P(A/B_k) = P(A)\,P(B_k/A)$$

Từ đó suy ra :

$$P(B_k / A) = \frac{P(B_k)P(A / B_k)}{P(A)} = \frac{P(B_k)P(A / B_k)}{\sum_{i=1}^{n} P(B_i)P(A / B_i)}.$$

Mệnh đề được chứng minh hoàn toàn.

Ví dụ 1.19. Cho hai lô sản phẩm. Lô I có 20 sản phẩm trong đó có 15 sản phẩm tốt và 5 phế phẩm. Lô II có 20 sản phẩm trong đó có 10 sản phẩm tốt 10 phế phẩm. Lấy ngẫu nhiên 1 lô và từ lô đó chọn ngẫu nhiên 1 sản phẩm.

1. Tìm xác suất để sản phẩm lấy ra là sản phẩm tốt.

2. Giả sử sản phẩm lấy là sản phẩm tốt. Tìm xác suất để sản phẩm đó của lô thứ hai.

Giải :

1. Đặt A = biến cố "sản phẩm lấy ra là sản phẩm tốt".

và B_i = biến cố "sản phẩm lấy ra ở lô thứ i", i = 1, 2.

Ta suy ra B_1, B_2 lập thành hệ đầy đủ. Bởi vì sản phẩm lấy ra không của lô I thì của lô II. Đó là điều chắc chắn nghĩa là $B_1 \cup B_2 = \Omega$. 1 sản phẩm đã lấy ở lô I thì thôi lấy ở lô II và ngược lại, lấy ở lô 2 thì thôi lấy ở lô I, nghĩa là :

$$B_1 \cap B_2 = \varnothing.$$

Theo công thức xác suất toàn phần ta có :

$$P(A) = P(B_1)P(A/B_1) + P(B_2)P(A/B_2).$$

Theo đầu bài ta có :

$$P(B_1) = P(B_2) = \frac{1}{2}\ ; \quad P(A/B_1) = \frac{15}{20}\ ; \quad P(A/B_2) = \frac{10}{20}.$$

Vậy $P(A) = \frac{1}{2} \times \frac{15}{20} + \frac{1}{2} \times \frac{10}{20} = \frac{25}{40} = \frac{5}{8}$.

2. Xác suất phải tìm là :

$$P(B_2 / A) = \frac{P(B_2)P(A / B_2)}{P(A)} = \frac{\frac{1}{2} \times \frac{1}{2}}{\frac{5}{8}} = \frac{2}{5}$$

Ví dụ 1.20. Người ta biết rằng một cặp trẻ sinh đôi có thể là một cặp sinh đôi cùng trứng hoặc một cặp sinh đôi không cùng trứng.

Một cặp sinh đôi cùng trứng những đứa trẻ bao giờ cũng cùng giới tính ; còn sinh đôi không cùng trứng xác suất để chúng cùng giới tính bằng $\frac{1}{2}$. Giả sử cặp trẻ sinh đôi cùng trứng với xác suất bằng p.

Tìm xác suất để cặp trẻ sinh đôi cùng giới tính là cặp sinh đôi cùng trứng.

Giải :

Đặt B_1 = Biến cố "cặp trẻ sinh đôi là sinh đôi cùng trứng".

B_2 = biến cố "cặp trẻ sinh đôi là sinh đôi không cùng trứng".

B_1, B_2 lập thành hệ đầy đủ các biến cố.

Đặt A = biến cố "Cặp trẻ sinh đôi cùng giới tính".

Theo công thức xác suất toàn phần ta có :

$$P(A) = P(B_1)\,P(A/B_1) + P(B_2)\,P(A/B_2)$$

Theo giả thiết $P(A/B_1) = 1$; $P(A/B_2) = \frac{1}{2}$ và $P(B_1) = p$, $P(B_2) = 1 - p$.

Vậy $P(A) = p \times 1 + (1-p) \times \frac{1}{2} = \frac{1+p}{2}$.

Xác suất phải tìm là :

$$P(B_1 / A) = \frac{P(B_1)P(A / B_1)}{P(A)} = \frac{1 \times p}{\frac{1+p}{2}} = \frac{2p}{1+p}.$$

Ví dụ 1.17. Trong một đám đông người mà số đàn ông bằng nửa số đàn bà. Xác suất để đàn ông bị bạch bệnh tạng là 0,06 và xác suất để người đàn bà bị bệnh bạch tạng là 0,0036.

1. Tìm xác suất để một cá thể bất kỳ bị bệnh bạch tạng.

2. Tìm xác suất để một người bị bệnh bạch tạng trong đám đông đó là đàn ông.

Giải :

1. Gọi A là biến cố "Một cá thể trong đám đông bị bạch tạng".
và B_1 là biến cố "Một cá thể là đàn ông".

B_2 là biến cố "Một cá thể là đàn bà".

Ta thấy B_1, B_2 lập thành hệ đầy đủ các biến cố.

Theo công thức xác suất toàn phần ta có :

$$P(A) = P(B_1)\,P(A/B_1) + P(B_2)\,P(A/B_2)$$

Theo giả thiết $P(B_1) = \frac{1}{3}$, $P(B_2) = \frac{2}{3}$ và $P(A/B_1) = 0{,}06$;

$P(A/B_2) = 0{,}0036$.

Vậy $P(A) = \frac{1}{3} \times 0{,}06 + \frac{2}{3} \times 0{,}0036 = 0{,}0224$.

2. Xác suất để một cá thể bị bạch tạng là đàn ông chính là xác suất điều kiện của biến cố B_1 với điều kiện A đã xảy ra.

Theo công thức Bayes ta có :

$$P(B_1/A) = \frac{P(B_1)P(A/B_1)}{P(A)} = \frac{\frac{1}{3} \times 0,06}{0,0224} \approx 0,89$$

1.5.3. Sự độc lập các biến cố, sự độc lập của các phép thử

Định nghĩa 1.7. Hai biến cố A, B được gọi là độc lập với nhau nếu :

$$P(AB) = P(A)\,P(B).$$

Hệ quả 1.2. Hai biến cố A và B là độc lập khi và chỉ khi hoặc $P(B/A) = P(B)$ hoặc $P(A/B) = P(A)$.

Hệ quả 1.3. Hai biến cố A và B độc lập khi và chỉ khi hoặc A, $\overline{B}$ là độc lập, hoặc $\overline{A}$, B là độc lập hoặc $\overline{A}$, $\overline{B}$ là độc lập.

1.5.4. Dãy n biến cố độc lập

Định nghĩa 1.8. Dãy n biến cố A_1, A_2, ...A_n được gọi là độc lập nếu ta lấy ra một dãy con bất kỳ các biến cố từ n biến cố trên thì xác suất của tích các biến cố của dãy con đó bằng tích các xác suất của từng biến cố, nghĩa là.

Với tập con bất kỳ $I \subseteq \{1, 2, ..., n\}$ có $P(\bigcap_{i \in I} A_i) = \prod_{i \in I} P(A_i)$.

– Nếu dãy các biến cố thoả mãn định nghĩa 1.8 thì dãy đó được gọi là độc lập trong toàn thể.

– Nếu từng đôi một trong dãy đó mà độc lập với nhau thì dãy đó được gọi là độc lập từng đôi.

Từ đó suy ra rằng : Nếu dãy các biến cố A_1, ..., A_n độc lập trong toàn thể thì nó độc lập từng đôi. Song điều ngược lại nói chung không đúng.

Ví dụ 1.22. Gieo 2 con xúc xắc cân đối và đồng chất. Gọi A là biến cố "con xúc xắc thứ nhất xuất hiện mặt có số chấm là số chẵn". Và B là biến cố "con xúc xắc thứ II xuất hiện mặt có số chấm là số lẻ". C là biến cố "cả hai con xúc xắc xuất hiện mặt có số chấm là số chẵn hoặc lẻ". Xét xem ba biến cố A, B, C có độc lập từng đôi và độc lập trong toàn thể không ?

Giải :

Theo giả thiết ta có : $P(A) = P(B) = \frac{1}{2} = P(C)$.

Bởi vì $C = A\overline{B} \cup \overline{A}B$. Hai con xúc xắc gieo độc lập.

Do đó : $P(C) = P(A)P(\overline{B}) + P(\overline{A})P(B) = \frac{1}{2} \times \frac{1}{2} + \frac{1}{2} \times \frac{1}{2} = \frac{1}{2}$.

Và $\quad P(AB) = P(A)P(B) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$.

Và $\quad P(AC) = P[A(A\overline{B} \cup \overline{A}B)] = P(A\overline{B} \cup \varnothing) = P(A\overline{B}) + 0$

$$P(AC) = P(A)\,P(\overline{B}) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}.$$

Và $\quad P(BC) = P[B(A\overline{B} \cup \overline{A}B)] = P(\overline{A}B) = P(\overline{A})P(B)$.

$$P(BC) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}.$$

Từ các kết quả trên ta kết luận : Ba biến cố A, B, C là độc lập từng đôi.

Mặt khác : $ABC = \varnothing$. Do đó $P(ABC) = 0$.

Và $P(A)\,P(B)\,P(C) = \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{8} \neq P(ABC)$.

Điều đó chứng tỏ A, B, C không độc lập trong toàn thể.

1.5.5. Dãy các phép thử độc lập

Định nghĩa 1.9. Dãy n phép thử $G_1, G_2 ... G_n$ trong mỗi phép thử G_i tương ứng với không gian biến cố sơ cấp Ω_i gồm r biến cố sơ cấp A_1, A_2, ... A_r được gọi là độc lập nếu :

$$P(A_{i_1}^1 A_{i_2}^2 ... A_{i_n}^n) = P(A_{i_1}^1)P(A_{i_2}^2)...P(A_{i_n}^n).$$

Trong đó $A_{i_1}^1$ là một biến cố bất kỳ trong r biến cố A_1, ...A_r tương ứng với phép thử G_1.

....

$A_{i_n}^n$ là một biến cố bất kỳ trong r biến cố A_1, ..., A_r tương ứng với phép thử G_n.

Ví dụ về dãy phép thử độc lập :

* Bắn 20 viên đạn độc lập vào 1 mục tiêu.

Mỗi lần bắn 1 viên, được xem như tiến hành 1 phép thử. Không gian biến cố sơ cấp tương ứng với mỗi phép thử là $\Omega = \{$trúng đích (biến cố A), không trúng đích (biến cố $\overline{A}$)$\}$. 20 lần bắn độc lập là 20 phép thử độc lập.

* Gieo 10 lần con xúc xắc cân đối và đồng chất được xem như tiến hành 10 phép thử độc lập. Không gian biến cố sơ cấp tương ứng với mỗi phép thử là $\Omega_i = \{B_1, B_2, B_3, B_4, B_5, B_6\}$.

1.6. CÔNG THỨC XÁC SUẤT NHỊ THỨC

Định nghĩa 1.10. Dãy n phép thử $G_1, G_2, \ldots G_n$ được gọi là dãy n phép thử Bernoulli nếu nó thỏa mãn các điều kiện sau :

1. Dãy n phép thử đó là độc lập.

2. Trong mỗi phép thử G_i tương ứng với không gian biến cố sơ cấp $\Omega_i = \{A, \overline{A}\}$.

3. Xác suất của biến cố A là P(A) không thay đổi trong mọi phép thử. Đặt P(A) = p.

Bài toán : Tìm xác suất để trong dãy n phép thử Bernoulli biến cố A xuất hiện đúng k lần.

Giải :

Xét biến cố tích của n biến cố dạng : $A\,A\,\overline{A}\,\overline{A} \ldots A\,\overline{A}$ (1.4)

Trong tích này có k biến cố A và n – k biến cố $\overline{A}$. Vì dãy n phép thử này độc lập nên :

$P(A\,A\,\overline{A}\,\overline{A} \ldots A\,\overline{A}) = P(A)^k P(\overline{A})^{n-k} = p^k (1-p)^{n-k}$; với k = 0, 1, ... n.

Ta nhận thấy rằng : Biến cố "Trong dãy n phép thử Bernoulli, biến cố A xuất hiện đúng k lần" bằng tổng của C_n^k các biến cố tích xung khắc từng đôi dạng (1.4) mà mỗi hạng tử của tổng này đều có xác suất là $p^k(1-p)^{n-k}$. Nếu ký hiệu xác suất của biến cố này là $P_n(k)$ thì ta có :

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

Công thức này được gọi là công thức xác suất nhị thức. Nếu đặt $1 - p = q$ thì ta có :

$$P_n(k) = C_n^k p^k q^{n-k}, \ k = 0, 1, 2, ..., n.$$

Ví dụ 1.23. Gieo ngẫu nhiên 20 lần một đồng tiền cân đối và đồng chất. Tìm xác suất để :

1. Có đúng 1 lần xuất hiện mặt sấp.

2. Có ít nhất 2 lần xuất hiện mặt sấp.

Giải :

1. Xem việc gieo 20 lần một đồng tiền cân đối và đồng chất như là tiến hành dãy 20 phép thử Bernoulli, xác suất xuất hiện mặt sấp (biến cố A) luôn luôn bằng 1/2 trong một lần gieo.

Theo công thức xác suất nhị thức ta có :

$$P_n(k) = C_n^k p^k q^{n-k}$$

Ở đây $n = 20$; $k = 1$. Vậy :

$$P_{20}(1) = C_{20}^1 \left(\frac{1}{2}\right)^1 \left(1 - \frac{1}{2}\right)^{20-1} = 20 \times \frac{1}{2^{20}} = \frac{5}{2^{18}}.$$

2. Ta phải tính xác suất của biến cố $[k \geq 2]$, k là số lần xuất hiện biến cố A trong dãy n phép thử Bernoulli.

Ta có $P[k \geq 2] = 1 - P[k<2] = 1 - P_{20}(0) - P_{20}(1)$.

$$P[k \geq 2] = 1 - C_{20}^0 \left(\frac{1}{2}\right)^0 \left(1 - \frac{1}{2}\right)^{20-0} - C_{20}^1 \left(\frac{1}{2}\right)^1 \left(1 - \frac{1}{2}\right)^{20-1}$$

$$= 1 - \frac{1}{20^{20}} - \frac{5}{2^{18}} = 1 - \frac{21}{2^{20}}$$

Ví dụ 1.24. Một bà mẹ sinh 2 con (mỗi lần sinh một con). Giả sử xác suất sinh con trai là 0,51. Tìm xác suất để trong người con đó :

1. Có đúng 1 con trai.

2. Không có con trai.

3. Có 2 con trai.

Giải :

Trong thống kê người ta chứng minh được rằng : Các lần sinh là độc lập và xác suất sinh con trai là 0,51 trong mọi lần sinh. Theo công thức xác suất nhị thức ta có : Xác suất để trong 2 lần sinh đó (mỗi lần sinh 1 con) có k con trai là :

$$P_2(k) = C_2^k(0,51)^k(1-0,51)^{2-k} \text{ ; } k = 0, 1, 2.$$

1. Với k = 1 ta có : $P_2(1) = C_2^1(0,51)(0,49) = 0,4998$.

2. Với k = 0 ta có : $P_2(0) = C_2^0(0,51)^0(0,49)^2 = 0,2401$.

3. Với k = 2 ta có : $P_2(2) = C_2^2(0,51)^2(0,49)^0 = 0,2601$.

Qua ví dụ trên ta nhận thấy trong số những gia đình có 2 con thì số gia đình có 1 con trai và 1 con gái là đông hơn cả.

Ví dụ 1.25. Gieo 100 hạt đậu tương. Xác suất nảy mầm của mỗi hạt là 0,9. Tính xác suất để trong 100 hạt.

1. Có đúng 80 hạt nảy mầm.
2. Có ít nhất 1 hạt nảy mầm.
3. Có nhiều nhất 98 hạt nảy mầm.

Giải :

Gieo ngẫu nhiên 100 hạt đậu tương được xem như thực hiện 100 phép thử Bernoulli và P(A) = 0,9 (A là biến cố hạt nảy mầm).

Theo công thức xác suất nhị thức ta có :

$$P_{100}(k) = C_{100}^k(0,9)^k.(0,1)^{100-k}.$$

1. Với k = 80 có : $P_{100}(80) = C_{100}^{80}(0,9)^{80}.(0,1)^{20}$.

2. Gọi k là số hạt nảy mầm trong 100 hạt.

Ta cần tính xác suất $P[k \geq 1]$

Ta có :

$P[k \geq 1] = 1 - P[k = 0] = 1 - C_{100}^0(0,9)^0(0,1)^{100-0} = 1 - (0,1)^{100}$.

3. Lời giải của câu hỏi này là $P[\, k \leq 98]$.

Mà $P[\, k \leq 98] = 1 - P[k > 98] = 1 - C_{100}^{99}(0,9)^{99}(0,1) - C_{100}^{100}(0,9)^{100}$.

$P[\, k \leq 98] = 1 - (0,9)^{99}$.

Ví dụ 1.26. Một lô hàng chứa rất nhiều sản phẩm với tỷ lệ phế phẩm $p = 0,02$. Cần phải lấy một mẫu với cỡ bằng bao nhiêu, sao cho xác suất để có ít nhất một phế phẩm trong mẫu đó không bé hơn $R = 0,95$.

Giải :

Gọi A là biến cố : "trong mẫu có ít nhất một phế phẩm".

Gọi n là cỡ mẫu phải tìm. Đặt $q = 1 - p$.

Ta có : $P(A) = 1 - P(\bar{A}) = 1 - C_n^0 p^0 q^n = 1 - q^n \geq R = 0,95$.

Từ đó ta suy ra $(0,98)^n \leq 0,05$.

$$n \geq \frac{\ln(0,05)}{\ln(0,98)}$$

Ta khảo sát sự biến thiên của xác suất $P_n(k)$ khi cố định n và cho k biến thiên từ 0 đến n.

Muốn vậy, ta xét tỷ số :

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

* Xét trường hợp tỷ số này lớn hơn hoặc bằng 1. Ta suy ra $\frac{(n-k)p}{(k+1)q} \geq 1$

Ta rút ra $(n-k)p \geq kq + q$ hay $k \leq np - q$

Vậy khi k tăng từ 0 đến $np - q$ thì xác suất $P_n(k)$ tăng.

* Xét trường hợp tỷ số này nhỏ hơn hoặc bằng 1, nghĩa là $\frac{(n-k)p}{(k+1)q} \leq 1$

Ta rút ra $k \geq np - q$

Điều đó có nghĩa là xác suất $P_n(k)$ giảm khi k tăng từ $np - q$ đến n.

Chứng tỏ rằng khi $k = np - q$ thì xác suất $P_n(k)$ đạt cực đại.

Ta nhận thấy khi $k = np - q$ thì $\frac{P_n(k+1)}{P_n(k)} = 1$, nghĩa là $P_n(k+1) = P_n(k)$.

Song k chỉ nhận giá trị nguyên, do đó :

• Nếu $np - q$ là số nguyên thì k có 2 giá trị $k_1 = np - q$ và $k_2 = np - q + 1$ mà tại đó xác suất $P_n(k)$ đạt cực đại.

• Nếu $np - q$ là không nguyên thì k có 1 giá trị $k = [np - q] + 1$ mà tại đó xác suất $P_n(k)$ đạt cực đại ; trong đó [a] là ký hiệu phần nguyên của a.

Ví dụ 1.27. Một xạ thủ bắn ngẫu nhiên độc lập 14 viên đạn vào một mục tiêu với xác suất bắn trúng đích của mỗi viên đạn là 0,2. Tìm số viên đạn trúng đích với khả năng lớn nhất.

Giải :

Xem việc bắn độc lập 14 viên đạn vào một mục tiêu như là tiến hành 14 phép thử Bernoulli với xác suất trúng đích của mỗi viên đạn (biến cố A) là không đổi $p = 0{,}2$.

Ta có $n = 14$. Vậy $np - q = 14 . 0{,}2 - 0{,}8 = 2$, là số nguyên. Vậy có hai giá trị của k :

$k_1 = 2$; $k_2 = 3$ tại đó xác suất $P_n(k)$ đạt cực đại

Bài tập chương I

1. Kiểm tra theo thứ tự một lô hàng gồm N sản phẩm. Các sản phẩm đều thuộc một trong hai loại tốt hoặc xấu. Ký hiệu A_k ($k = 1, 2, ..., N$) là biến cố chỉ sản phẩm kiểm tra thứ k thuộc loại xấu. Viết bằng ký hiệu các biến cố sau đây :

   a) Cả N sản phẩm đều xấu.

   b) Có ít nhất một sản phẩm xấu.

   c) m sản phẩm kiểm tra đầu là tốt, các sản phẩm còn lại là xấu.

   d) Các sản phẩm kiểm tra theo thứ tự chẵn là xấu, còn các sản phẩm kiểm tra theo thứ tự lẻ là tốt.

   e) Không gian biến cố sơ cấp có mấy phần tử.

2. Bắn không hạn chế vào một mục tiêu cho đến khi có viên đạn trúng mục tiêu thì thôi bắn. Giả sử mỗi lần bắn chỉ có hai khả năng trúng bia (biến cố A) hoặc chệch bia (biến cố $\overline{A}$).

   a) Hãy mô tả không gian biến cố sơ cấp.

   b) Hãy nêu 1 hệ đầy đủ các biến cố.

3. Có n bệnh nhân. Gọi A_k là biến cố bệnh nhân thứ k khỏi bệnh. Hãy viết bằng ký hiệu các biến cố sau :

a) Tất cả các bệnh nhân đều khỏi bệnh.

b) Có ít nhất một người không khỏi bệnh.

c) Có đúng 1 người không khỏi bệnh.

d) Có đúng 2 người không khỏi bệnh.

4. Một dụng cụ điện tử gồm có 3 bóng đèn loại I và 4 bóng loại II. Gọi A_k (k = 1, 2, 3) là biến cố chỉ bóng đèn loại I thứ k tốt, còn B_j (j = 1, 2, 3, 4) là biến cố chỉ bóng đèn loại II thứ j tốt. Dụng cụ tiếp tục làm việc được nếu có ít nhất 1 bóng loại I tốt và không ít hơn 3 bóng loại II tốt.

a) Hãy biểu diễn biến cố (C) chỉ dụng cụ vẫn làm việc được qua các biến cố A_k và B_j và các biến cố đối của chúng.

b) Biểu diễn biến cố (D) "có 1 và chỉ 1 bóng đèn loại I tốt và có đúng 2 bóng đèn loại 2 tốt".

5. Cho hai biến cố A và B. Xét các biến cố A, $\overline{A}$, $\overline{A \cup B}$.

a) Các biến cố đó có xung khắc từng đôi không.

b) Các biến cố đó có lập thành hệ đầy đủ các biến cố không.

6. Chọn ngẫu nhiên 1 công nhân trong số các công nhân có mặt ở xí nghiệp. Gọi A là biến cố xảy ra khi người công nhân được chọn là nam và B là biến cố người công nhân được chọn ở khu tập thể ; C là biến cố người công nhân được chọn không hút thuốc lá.

a) Hãy mô tả biến cố $AB\overline{C}$.

b) Với điều kiện nào ta có $A\overline{B}C = A$.

c) Khi nào thì ta có $C = \overline{A}$.

7. Cho ba biến cố A, B, C. Viết biểu thức chỉ biến cố.

a) Chỉ có A xảy ra.

b) A và B xảy ra nhưng C không xảy ra.

c) Cả ba biến cố cùng xảy ra.

d) Có ít nhất một trong ba biến cố A, B, C xảy ra.

e) Có ít nhất hai biến cố cùng xảy ra.

g) Có một và chỉ một trong ba biến cố ấy xảy ra.

h) Chỉ có hai trong ba biến cố đó xảy ra.

i) Không có biến cố nào trong ba biến cố đó xảy ra.

k) Không có quá hai biến cố trong ba biến cố đó xảy ra.

8. Có 5 cuốn sách khác nhau A, B, C, D, E đặt trên giá sách. Rút lần lượt (không hoàn lại) 3 cuốn.

a) Không gian biến cố sơ cấp có mấy phần tử.

b) Số biến cố sơ cấp thuận lợi cho biến cố rút được cuốn sách A và biến cố không rút được cuốn sách A là bao nhiêu.

9. Một bà mẹ sinh hai con (mỗi lần sinh được một con hoặc trai hay gái).

a) Không gian biến cố sơ cấp có mấy phần tử.

b) Có bao nhiêu biến cố sơ cấp thuận lợi cho 2 con có 1 trai, 1 gái ?

10. Chia ngẫu nhiên 15 tặng phẩm cho 3 người.

a) Có bao nhiêu khả năng thuận lợi cho biến cố : "người thứ nhất được đúng 3 tặng phẩm".

b) Có bao nhiêu khả năng thuận lợi cho biến cố : "mỗi người được đúng 5 tặng phẩm".

Bài tập sử dụng định nghĩa xác suất :

11. Một lô hàng gồm 1000 sản phẩm, trong đó có 30 sản phẩm xấu. Lấy hú hoạ 1 sản phẩm từ lô hàng.

a) Tìm xác suất để sản phẩm lấy ra là sản phẩm tốt.

b) Lấy ngẫu nhiên (1 lần) 10 sản phẩm từ lô hàng. Tìm xác suất để trong 10 sản phẩm lấy ra có đúng 8 sản phẩm tốt.

12. Một hộp chứa 30 bi trắng, 7 bi đỏ và 15 bi xanh. Một hộp khác chứa 10 bi trắng, 6 bi đỏ và 9 bi xanh. Ta lấy ngẫu nhiên từ mỗi hộp 1 bi. Tìm xác suất để 2 bi rút ra cùng màu.

13. Gieo đồng thời hai con xúc xắc cân đối và đồng chất. Tìm xác suất sao cho :

a) Tổng số chấm ở mặt trên hai con xúc xắc bằng 8.

b) Hiệu số chấm ở mặt trên hai con xúc xắc có trị số tuyệt đối bằng 2.

c) Số chấm ở mặt trên hai con xúc xắc bằng nhau.

14. Một lô hàng có n sản phẩm trong đó có m sản phẩm xấu. Lấy ngẫu nhiên từ lô hàng k sản phẩm. Tìm xác suất để trong k sản phẩm lấy ra có đúng s sản phẩm xấu.

15. Chia 12 tặng phẩm cho ba người. Tìm xác suất để :

a) Người thứ nhất được đúng 3 tặng phẩm.

b) Mỗi người được 4 tặng phẩm.

16. Một khoá chữ được lập nên bởi 6 vòng ghép tiếp nhau quay quanh một trục. Mỗi vành đều chia thành 6 phần bằng nhau, trên mỗi phần có ghi 1 chữ số. Khoá được mở khi mỗi vòng đặt đúng vị trí xác định trước. Tìm xác suất để mở được khoá (giả sử các chữ số được lắp ghép một cách tuỳ ý).

17. Một loạt vé xổ số với tổng số tiền vé là n đồng. Giá mỗi vé là 300 đồng. Số vé trúng thưởng loại q_1 là m_1 đồng, loại q_2 là m_2 đồng ($q_1 > q_2$).

a) Tìm xác suất trúng thưởng không quá q_1 đồng của 1 vé.

b) Tìm xác suất trúng thưởng q_2 đồng của 1 vé.

18. 12 hành khách lên ngẫu nhiên 4 toa tàu.

a) Tìm xác suất để mỗi toa có 3 hành khách.

b) Tìm xác suất để trong một toa có 6 hành khách, một toa có 4 hành khách, hai toa còn lại mỗi toa 1 hành khách.

19. Một khách sạn có 6 phòng phục vụ khách, nhưng có tất cả 10 khách đến xin nghỉ trọ, trong đó có 6 nam và 4 nữ. Khách sạn phục vụ theo nguyên tắc "ai đến trước phục vụ trước và mỗi phòng nhận một người".

a) Tìm xác suất để cho cả 6 nam đều được nghỉ trọ.

b) Tìm xác suất để 4 nam và 2 nữ được nghỉ trọ.

c) Tìm xác suất sao cho ít nhất có 2 trong 4 nữ được nghỉ trọ.

Xác suất hình học :

20. Trên đường tròn tâm 0 bán kính r người ta lấy một điểm A cố định.

a) Lấy ngẫu nhiên điểm M trên đường tròn đó. Tìm xác suất để khoảng cách từ M đến A không vượt quá r.

b) Lấy ngẫu nhiên điểm N trong hình tròn đó. Tìm xác suất để khoảng cách từ N đến A không vượt quá r.

21. Cắt ngẫu nhiên đoạn dây dài 1 mét thành ba đoạn. Tìm xác suất để từ ba đoạn đó ta dựng được một tam giác.

22. Lấy ngẫu nhiên hai số dương (mỗi số không lớn hơn 1). Tìm xác suất sao cho tổng của chúng không lớn hơn 1, còn tích của chúng không lớn hơn $\frac{2}{9}$.

23. Các hệ số a, b của các phương trình :

a) $x^2 + ax + b^2 = 0$

b) $(a + 1)x^2 + 2bx - a + 1 = 0$

được lấy ngẫu nhiên giá trị trong đoạn $[-1 ; 1]$. Tìm xác suất để :

– Phương trình (a) có nghiệm thực.

– Phương trình (b) có nghiệm thực.

24. Trên đoạn OA có độ dài là s của trục Ox, đặt ngẫu nhiên hai điểm B (OB = x) và C (OC = y). Tìm xác suất sao cho độ dài BC bé hơn $\frac{s}{2}$.

Bài tập về xác suất tổng, tích, xác suất điều kiện :

25. Bắn ba viên đạn vào cùng một bia. Xác suất trúng đích của viên thứ nhất, viên thứ hai và viên thứ ba tương ứng bằng 0,4 ; 0,5 ; 0,7.

a) Tìm xác suất sao cho trong 3 viên đạn có đúng 1 viên trúng đích.

b) Tìm xác suất để có ít nhất 1 viên đạn trúng đích.

26. Một lô hàng gồm 150 sản phẩm có chứa 6% phế phẩm.

Người ta dùng phương pháp chọn mẫu để kiểm tra lô hàng và quy ước rằng : Kiểm tra lần lượt 6 sản phẩm, nếu có ít nhất 1 trong 6 sản phẩm đó là phế phẩm thì loại lô hàng. Tìm xác suất để chấp nhận lô hàng.

27. Bắn liên tiếp vào một mục tiêu cho đến khi nào có 1 viên đạn đầu tiên trúng mục tiêu thì ngừng bắn. Tìm xác suất sao cho phải bắn đến viên thứ 6. Biết rằng xác suất trúng đích của mỗi viên đạn là 0,2 và các lần bắn là độc lập.

28. Một máy bay gồm ba bộ phận có tầm quan trọng khác nhau. Muốn bắn rơi được máy bay, thì chỉ cần có 1 viên đạn trúng vào bộ phận thứ nhất, hoặc 2 viên đạn trúng bộ phận thứ hai hoặc 3 viên đạn trúng bộ phận thứ 3. Xác suất để 1 viên đạn trúng bộ phận thứ nhất, thứ hai, thứ ba với điều kiện viên đạn đó đã trúng máy bay tương ứng bằng 0,15 ; 0,30 ; 0,55. Tìm xác suất để máy bay bị bắn rơi khi :

a) Có 1 viên đạn trúng.

b) Có 2 viên đạn trúng.

c) Có 3 viên đạn trúng.

d) Có 4 viên đạn trúng.

29. Một nhà máy sản xuất bóng đèn. Máy A sản xuất 25% số bóng đèn, máy B sản xuất 35% số bóng đèn, còn máy C sản xuất 40%. Tỷ lệ sản phẩm hỏng của các máy đó trên tổng số sản phẩm do nhà máy đó sản xuất tương ứng bằng 5% (máy A), 4% (máy B), 2% (máy C). Lấy ngẫu nhiên 1 sản phẩm thì được sản phẩm xấu. Tìm xác suất để cho sản phẩm lấy ra là do :

a) Máy A sản xuất.

b) Máy B sản xuất.

c) Máy C sản xuất.

30. Hai đấu thủ A và B thi đấu một số lần, trong mỗi lần hoặc đấu thủ A thắng hoặc đấu thủ B thắng. Xác suất thắng của A trong mỗi lần bằng p. Trước lúc vào thi đấu đều có quy ước là mỗi đấu thủ phải thắng mấy lần mới được xem là thắng cuộc. Trò chơi bị huỷ nếu không có đấu thủ nào thắng đủ số lần quy định để thắng cuộc.

a) Tìm xác suất để A thắng cuộc, nếu giả sử A cần có 2 lần thắng, còn B phải có 3 lần thắng.

b) Tính xác suất thắng cuộc của A, nếu giả sử rằng A cần thắng m lần, còn B thắng n lần.

31. Một học sinh khi vào thi chỉ thuộc 18 trong 25 câu hỏi thi. Tìm xác suất để học sinh trả lời được 3 câu hỏi mà học sinh đó rút được.

32. Ba cậu bé chơi trò chơi gieo đồng tiền liên tiếp. Ai gieo được mặt sấp đầu tiên sẽ thắng cuộc. Tìm xác suất thắng cuộc của mỗi cậu bé.

Bài tập về công thức xác suất toàn phần và công thức Bayes :

33. Giả sử có 3 kiện hàng với số sản phẩm tốt tương ứng là 20, 15, 10. Lấy ngẫu nhiên 1 kiện hàng (giả sử 3 kiện hàng đó có cùng khả năng rút) rồi từ kiện đó lấy hú họa 1 sản phẩm. Biết rằng 3 kiện hàng đó đều có 20 sản phẩm.

a) Tìm xác suất để sản phẩm chọn ra là sản phẩm tốt.

b) Giả sử sản phẩm chọn ra là sản phẩm tốt. Tìm xác suất để sản phẩm đó thuộc kiện hàng thứ hai.

34. Với 3 kiện hàng được cho như trong bài 33, ta chọn ngẫu nhiên 1 kiện và từ kiện đó lấy hú họa 1 sản phẩm thấy là sản phẩm tốt. Trả sản phẩm đó lại kiện hàng vừa lấy ra, sau đó lại lấy tiếp 1 sản phẩm thì được sản phẩm tốt. Tìm xác suất để các sản phẩm được lấy từ kiện hàng thứ 3.

35. Tỷ số ô tô tải và ô tô con đi qua đường có trạm bơm dầu là $\frac{5}{2}$. Xác suất để 1 ô tô tải qua đường được nhận dầu là 0,1 còn xác suất để 1 ô tô con qua đường được nhận dầu là 0,2.

Có một ô tô qua đến trạm để nhận dầu. Tìm xác suất để ô tô đó là xe tải.

36. Cho n hộp bi, mỗi hộp chứa m bi trắng và k bi đỏ. Lấy hú họa 1 bi từ hộp thứ 1 và bỏ vào hộp thứ 2, sau đó lấy hú họa 1 bi ở hộp thứ 2 bỏ vào hộp thứ 3 ...

Tìm xác suất sao cho viên bi cuối cùng rút ra từ hộp thứ n là bi trắng.

37. Tiến hành 3 phép thử độc lập. Xác suất xuất hiện biến cố A trong mỗi phép thử là $p = 0,1$. Xác suất xuất hiện biến cố B tuỳ thuộc vào số lần xuất hiện của A. Nếu A xuất hiện i lần $i = 0, 1, 2, 3$, thì xác suất xuất hiện biến cố B tương ứng là 0, i.

Tìm số (chỉ số lần xuất hiện biến cố A) có khả năng nhất, nếu giả sử biến cố B đã xuất hiện.

Phép thử Bernuolli – Công thức xác suất nhị thức :

38. Một lô hàng chứa rất nhiều sản phẩm với tỷ lệ phế phẩm là $p = 0,02$. Cần phải lấy một mẫu với cỡ bằng bao nhiêu, sao cho xác suất để có ít nhất 1 phế phẩm trong mẫu đó không bé hơn 0,95.

39. Một bà mẹ sinh 3 người con (mỗi lần sinh một con). Giả sử xác suất sinh con trai là $\frac{1}{2}$. Tìm xác suất sao cho trong 3 con đó :

a) Có 2 con trai.

b) Có không quá 1 con trai.

c) Có không ít hơn 1 con trai.

40. Tỷ lệ học sinh trong trường bị cận thị là 1%. Hỏi cần lấy một mẫu cỡ bao nhiêu (chọn bao nhiêu học sinh) sao cho trong mẫu đó ít nhất 1 học sinh bị cận thị với xác suất không bé hơn 0,95.

41. Một nữ công nhân phụ trách 12 máy dệt. Xác suất để mỗi máy dệt trong khoảng thời gian t cần đến sự chăm sóc của nữ công nhân bằng $\frac{1}{3}$. Tính xác suất để :

a) Trong khoảng thời gian t có 4 máy cần đến sự chăm sóc của nữ công nhân.

b) Trong khoảng thời gian t số máy dệt cần đến sự chăm sóc của nữ công nhân không bé hơn 3 và không lớn hơn 6.

42. Trồng 2 hàng cây và mỗi hàng 4 cây. Xác suất để mỗi cây sống là 0,8. Trồng lần thứ nhất nếu cây nào chết thì trồng lại cây đó. Tìm xác suất để không phải trồng quá hai lần.

43. Bắn độc lập 14 viên đạn vào một mục tiêu. Xác suất bắn trúng đích của mỗi viên đạn bằng 0,2. Mục tiêu bị phá huỷ hoàn toàn, nếu có ít nhất 2 viên đạn trúng vào mục tiêu. Tìm xác suất để :

a) Mục tiêu bị phá huỷ hoàn toàn.

b) Mục tiêu bị phá huỷ một phần.

44. Trong một kho hàng có chứa k sản phẩm. Mỗi sản phẩm trong kho hàng đó là tốt hoặc là xấu với xác suất như nhau và bằng $\frac{1}{2}$.

Chọn ngẫu nhiên từ kho hàng đó ra n sản phẩm (chọn hoàn lại) và thấy có m sản phẩm tốt ($0 \leq m \leq n$). Tìm xác suất sao cho trong kiện hàng đó có đúng v phế phẩm.

45. Một quả cầu được đánh dấu ở trong hộp I với xác suất p và trong hộp II với xác suất $1 - p$ $(0 < p < 1)$. Xác suất rút được quả cầu đánh dấu từ hộp chứa quả cầu này bằng P $(P \neq 1)$.

Rút ngẫu nhiên liên tiếp (có hoàn lại) n quả cầu từ hai hộp đó. Hỏi cần rút từ mỗi hộp bao nhiêu quả cầu để xác suất rút được quả cầu được đánh dấu ít nhất 1 lần là lớn nhất.

46. Một tổng đài có liên lạc với 10 địa điểm, ở mỗi địa điểm có đặt một máy điện thoại. Các máy điện thoại này sử dụng một cách độc lập và thường xuyên như nhau với thời gian trung bình mỗi lần nói chuyện là 6 phút. Tìm xác suất sao cho một trong các máy điện thoại ở các địa điểm đã cho khi cần gọi thì tổng đài đang bận.

Bài tập làm thêm :

47. Giả sử có n ngồi thành 1 dãy ngang. Tìm xác suất để cho 2 người cho trước luôn luôn ngồi cách r người.

48. Cho n người ngồi thành vòng tròn $(n > 2)$.

a) Tìm xác suất để 2 người cho trước luôn luôn ngồi cạnh nhau.

b) Tìm xác suất để hai người cho trước luôn luôn ngồi cách nhau r người.

49. Giả sử một hộp chứa n bi giống nhau. Gọi A là biến cố rút ngẫu nhiên được một số chẵn bi và B là biến cố rút được số lẻ bi. Hỏi biến cố nào có khả năng xuất hiện nhiều hơn, nếu giả thiết các nhóm bi rút ở các lần đều đồng khả năng.

Hướng dẫn và trả lời bài tập chương I

1. a) $A = A_1A_2 \ldots A_N$; b) $B = A_1 \cup A_2 \cup \ldots \cup A_N$;

c) $C = \overline{A}_1\overline{A}_2 \ldots \overline{A}_m A_{m+1} \ldots A_n$; d) $D = \overline{A}_1 A_2 \overline{A}_3 A_4 \overline{A}_5 \ldots$

e) Không gian biến cố sơ cấp có 2^N phần tử.

2. a) Không gian biến cố sơ cấp gồm các phần tử .

$A, \overline{A}A, \overline{A}\,\overline{A}\,A, \ldots, \overline{A} \,..\, \overline{A}\,A, \ldots$

b) Gọi B_1 là biến cố xảy ra khi số lần bắn là lẻ,

và B_2 là biến cố xảy ra khi số lần bắn là chẵn.

Ta suy ra B_1, B_2 lập thành hệ đầy đủ các biến cố.

3. a) $A = A_1A_2 ... A_n$

b) $B = A_1 \cup A_2 \cup ... \cup A_n$

c) $C = \overline{A}_1 A_2 ... A_n \cup A_1 \overline{A}_2 A_3 ... A_n \cup ... \cup A_1 A_2 ... \overline{A}_n$

d) $D = \overline{A}_1 \overline{A}_2 A_3 A_4 ... A_n \cup \overline{A}_1 A_2 \overline{A}_3 A_4 ... A_n \cup ... \cup A_1 ... A_{n-2} \overline{A}_{n-1} \overline{A}_n$

4. a) $C = (A_1 \cup A_2 \cup A_3)[\overline{B}_1 B_2 B_3 B_4 \cup B_1 \overline{B}_2 B_3 B_4 \cup B_1 B_2 \overline{B}_3 B_4 \cup$
$\cup B_1 B_2 B_3 \overline{B}_4 \cup B_1 B_2 B_3 B_4]$

b) $D = (A_1 \overline{A}_2 \overline{A}_3 \cup \overline{A}_1 A_2 \overline{A}_3 \cup \overline{A}_1 \overline{A}_2 A_3)[B_1 B_2 \overline{B}_3 \overline{B}_4 \cup ... \cup \overline{B}_1 \overline{B}_2 B_3 B_4]$

5. a) Các biến cố đó xung khắc từng đôi.

b) Chúng lập thành hệ đầy đủ các biến cố.

6. a) $AB\overline{C}$ là biến cố người công nhân được chọn là nam ở trong khu tập thể không hút thuốc.

b) Khi $A \subset \overline{B}$, $A \subset C$ thì $A\overline{B}C = A$.

c) Các nữ công nhân trong đại hội không hút thuốc $\overline{A} = C$.

7. a) $A\overline{B}\overline{C}$

b) $AB\overline{C}$

c) $A \cup B \cup C$

d) ABC

e) $AB \cup AC \cup BC \cup ABC$

g) $A\overline{B}\overline{C} + \overline{A}B\overline{C} + \overline{A}\overline{B}C$

h) $AB\overline{C} + A\overline{B}C + \overline{A}BC$

i) $\overline{A}\overline{B}\overline{C}$

k) $\overline{ABC}$

8. a) Có $5 \times 4 \times 3 = 60$ phần tử.

b) $C_3^1 \times 6 = 18$ phần tử tương ứng với biến cố rút được A.

9. a) Không gian biến cố sơ cấp có 4 phần tử Ω = (TT, GT, TG, GG).

b) Có hai biến cố thuận lợi do biến cố trong hai con có một trai, một gái.

10. a) $C_{15}^3 \times 2^{12}$ khả năng thuận lợi cho biến cố "người thứ nhất được 3 tặng phẩm".

b) $C_{15}^5 \times C_{10}^5 \times C_5^5$ khả năng thuận lợi cho biến cố "mỗi người được 5 tặng phẩm".

11. a) $P = \dfrac{970}{1000} = 0,97$; b) $P = \dfrac{C_{970}^{8} \times C_{30}^{2}}{C_{1000}^{10}}$.

12. $p = \dfrac{30 \times 10 + 7 \times 6 + 15 \times 9}{52 \times 25} = \dfrac{207}{625}$

13. a) $P(A) = \dfrac{5}{36}$; b) $P(B) = \dfrac{8}{36} = \dfrac{2}{9}$; c) $P(C) = \dfrac{6}{36} = \dfrac{1}{6}$.

14. a) $p = \dfrac{C_m^s \times C_{n-m}^{k-s}}{C_n^k}$

15. a) $p = \dfrac{C_{12}^3 \times 2^9}{3^{12}}$; b) $p = \dfrac{C_{12}^4 \times C_8^4 \times C_4^4}{3^{12}} = \dfrac{12!}{(4!)^3.3^{12}}$

16. $p = \dfrac{1}{6^6}$

17. a) $p = \dfrac{300(m_1 + m_2)}{n}$; b) $p = \dfrac{300m_2}{n}$

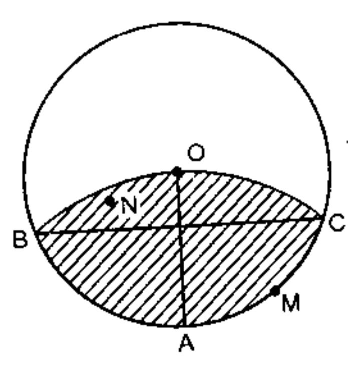
18. a) $p = \dfrac{12!}{(3!)^4.4^{12}}$; b) $p = \dfrac{12!}{6!4!.4^{12}}$

19. a) $p = \dfrac{1}{210}$; b) $p = \dfrac{3}{7}$; c) $p = \dfrac{37}{42}$

20. a) $p = \dfrac{1}{3}$; b) $p = 0,391$.

Hướng dẫn

a) Tập hợp những điểm M trên đường tròn mà khoảng cách từ M đến A không vượt quá r nằm trên cung BAC. Vì vậy xác suất để khi lấy ngẫu nhiên điểm M trên đường tròn mà khoảng cách từ M tới A không vượt quá r bằng :



$$P(A) = \frac{\text{độ dài cung } \overset{\frown}{BAC}}{\text{chu vi đường tròn}} =$$

$$= \frac{\frac{120}{360}.2\pi r}{2\pi r} = \frac{1}{3} \approx 0,33.$$

b) Xác suất phải tìm là :

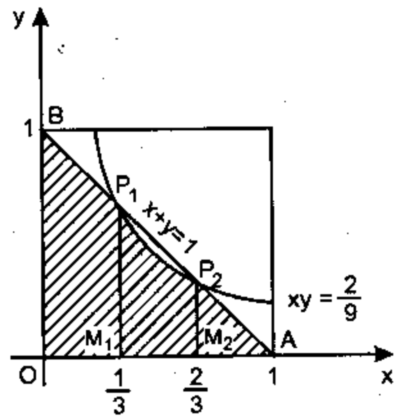
$$P(B) = \frac{\text{diện tích hình gạch}}{\text{diện tích hình tròn}} = \frac{2\left(\frac{\pi r^2}{3} - \frac{r}{2} \times \frac{r\sqrt{3}}{2}\right)}{\pi r^2} = 0{,}39.$$

21. Hướng dẫn : Gọi độ dài các đoạn cắt là x, y – x, 1 – y với $0 \le x < y \le 1$. Từ độ dài 1 cạnh của 1 tam giác nhỏ hơn tổng của hai cạnh kia và tính đối xứng ta suy ra đáp số p = 0,25.

22. Gọi x và y là hai số dương lấy được. Vậy ta có :

$0 \le x \le 1$; $0 \le y \le 1$; $x + y \le 1$; $0 \le xy \le \frac{2}{9}$

Không gian biến cố sơ cấp là hình vuông có cạnh 1.



Miền thuận lợi cho biến cố đang xét là hình gạch được giới hạn bởi các đường cong : $x + y = 1$ và $xy = \frac{2}{9}$

Miền đó là $G = \{(x, y) : x + y \le 1$ và $xy \le \frac{2}{9}, x \ge 0, y \ge 0\}$

Diện tích miền G = dt OBP_1M_1 + dt $M_1P_1P_2M_2$ + dt M_2P_2A

$$dtG = \frac{1}{3} + \frac{2}{3}\int_{\frac{1}{3}}^{\frac{2}{3}} \frac{dx}{x} = \frac{1}{3}\left(1 + \frac{2}{3}\ln 2\right) \approx 0{,}487$$

Xác suất phải tìm là : $p = \frac{dtG}{1} \approx 0{,}487.$

23. Phương trình a) p = 0,25 ; Phương trình b) p = 0,78.

24. $p = \frac{3}{4}$

Hướng dẫn : Gọi (x, y) là toạ độ của điểm M. M có thể chạy tuỳ ý trên hình vuông cạnh s. Những điểm M có toạ độ (x, y) thoả mãn $|y - x| \le \frac{s}{2}$ là miền thuận lợi cho biến cố đang quan tâm.

Vậy xác suất phải tìm là : $P = \dfrac{s^2 - \dfrac{s^2}{4}}{s^2} = \dfrac{3}{4}$.

25. a) $p = 0{,}36$;   b) $p = 0{,}91$.

26. $P(A) = \dfrac{141}{150} \times \dfrac{140}{149} \times \dfrac{139}{148} \times \dfrac{138}{147} \times \dfrac{137}{146} \times \dfrac{136}{145}$

27. $p = (0{,}8)^5 \times 0{,}2$.

28. Đặt A là biến cố "máy bay rơi"

a) P(A/1 viên trúng) = 0,15 ;   b) P(A/2 viên trúng) = 0,368.

c) P(A/3 viên trúng) = 0,728 ;   d) P(A/4 viên trúng) = 1.

29. H là biến cố "sản phẩm lấy ra là sản phẩm xấu"

a) $P(A / H) = \dfrac{25}{69}$;   b) $P(B / H) = \dfrac{28}{69}$;   c) $P(C / H) = \dfrac{16}{69}$

30. a) Đặt $q = 1 - p$, $P(A) = p^2(1 + 2q + 3q^2)$

b) $P(A) = p^m(1 + C_m^1 q + C_{m+1}^2 q^2 + ... + C_{m+n-1}^{n-1} q^{n-1})$.

31. $p = \dfrac{18}{25} \times \dfrac{17}{24} \times \dfrac{16}{23} = \dfrac{204}{575}$

32. $p_1 = \dfrac{4}{7}$; $p_2 = \dfrac{2}{7}$; $p_3 = \dfrac{1}{7}$

33. a) $P(A) = \dfrac{3}{4}$;   b) $P(B_2/A) = \dfrac{1}{3}$.

34. Đặt A = biến cố "các sản phẩm lấy ra đều tốt"

và B_i = biến cố "sản phẩm lấy ở lô thứ i", i = 1, 2, 3.

B_1, B_2, B_3 lập thành hệ đầy đủ các biến cố.

Ta áp dụng công thức xác suất toàn phần, tính được P(A). Sau đó áp dụng công thức Bayes tính được $P(B_3/A) = \dfrac{4}{29}$.

35. Đặt A = biến cố "có 1 ô tô đến nhận dầu"

và B_1 = biến cố "ô tô tải đi qua đường có trạm dầu"

B_2 = biến cố "ô tô con đi qua đường có trạm dầu"

B_1, B_2, lập thành hệ đầy đủ các biến cố.

Áp dụng công thức xác suất toàn phần, sau đó áp dụng công thức Bayes ta tính được $P(B_1/A) = \frac{5}{9}$.

36. $p = \dfrac{m}{m+k}$

37. Áp dụng công thức xác suất toàn phần và Bayes.

Đặt B_i = biến cố " A xuất hiện i lần" ; B_0, B_1, B_2, B_3 lập thành hệ đầy đủ các biến cố. Ta tính được $P(B) = \sum_{i=0}^{3} P(B_i)P(B/B_i)$ và tính được $P(B_0/B)$; $P(B_1/B)$; $P(B_2/B)$; $P(B_3/B)$. So sánh các xác suất trên ta nhận được số có khả năng nhất (nghĩa là có xác suất lớn nhất) là 1.

38. $n \geq \dfrac{\ln(0,05)}{\ln(0,98)}$.

39. a) $p_1 = C_3^2 \times \frac{1}{8} = \frac{3}{8}$; b) $p_2 = C_3^0 \times \frac{1}{8} + C_3^1 \times \frac{1}{8} = \frac{1}{2}$;

c) $p_3 = 1 - C_3^0 \times \frac{1}{8} = \frac{7}{8}$

40. $n \geq 300$.

41. a) $P_{12}(4) \approx 0,238$; b) $P_{12}(3 \leq m \leq 6) = 0,751$

42. A là biến cố "không phải trồng quá hai lần"

B_i là biến cố "trồng lần thứ nhất có i cây chết", $i = \overline{0\,;8}$.

$$P(A) = \sum_{i=0}^{8} P(B_i)P(A/B_i) = \sum_{i=0}^{8} C_8^i (0,2)^i (0,8)^{8-i} (0,8)^i = (0,8)^8 \sum_{i=0}^{8} C_8^i (0,2)^i$$

43. a) $P_{14}(m \geq 2) \approx \dfrac{73}{91}$; b) $P_{14}(1) = C_{14}^1 (0,2)(0,8)^{13} \approx \dfrac{2}{13}$.

44. Đặt A = biến cố "trong n sản phẩm lấy ra có đúng m phế phẩm"

B_v = biến cố "trong k sản phẩm có v phế phẩm", $v = 0, ..., k$.

Áp dụng công thức xác suất toàn phần, công thức xác suất nhị thức và công thức Bayes ta nhận được :

$$P(B_v/A) = \frac{C_k^v\left(\frac{v}{k}\right)^m\left(1-\frac{v}{k}\right)^{n-m}}{\sum_{v=1}^{k-1} C_k^v\left(\frac{v}{k}\right)^m\left(1-\frac{v}{k}\right)^{n-m}}.$$

45. $m = \frac{n}{2} + \frac{\ln\frac{1-p}{p}}{2\ln(1-p)}$

46. Mỗi máy điện thoại khi gọi xem như một phép thử Bernoulli với xác suất $p = \frac{1}{10}$. Xác suất cần tìm là $P_1 = 1 - P_{10}(0) = 0{,}651$.

47. $p = \frac{2(n-r-1)}{n(n-1)}$

48. a) $p = \frac{2}{n-1}$. (Khi biết vị trí của một người thì người kia có thể ngồi ở $n-1$ vị trí còn lại, trong đó 2 vị trí cạnh vị trí người thứ nhất).

b) Nếu $n = 2r + 2$ thì $p = \frac{1}{n-1}$.

Nếu $n \neq 2r + 2$ thì $p = \frac{2}{n-1}$.

49. Số nhóm có một số chẵn bi hay số biến cố sơ cấp thuận lợi cho A là :

$$C_n^2 + C_n^4 + ... = N_1$$

Số nhóm có một số lẻ bi hay số biến cố sơ cấp thuận lợi cho B là :

$$C_n^1 + C_n^3 + C_n^5 + ... = N_2$$

Vậy ta có : $N_2 - N_1 = 1 - (1-1)^n = 1$, tức là $P(B) - P(A) > 0$.

Chương II

BIẾN NGẪU NHIÊN VÀ HÀM PHÂN PHỐI

> ***Những nội dung chính trọng chương :***
> * *Khái niệm về biến ngẫu và hàm phân phối.*
> * *Các tính chất của hàm phân phối.*
> * *Sự độc lập của các biến ngẫu nhiên.*
> * *Phân phối xác xuất của hàm của biến ngẫu nhiên.*
> * *Kỳ vọng toán, phương sai, mômen và các số đặc trưng khác.*

2.1. KHÁI NIỆM BIẾN NGẪU NHIÊN, HÀM PHÂN PHỐI VÀ CÁC TÍNH CHẤT CỦA CHÚNG

2.1.1. Các khái niệm

a) Biến ngẫu nhiên

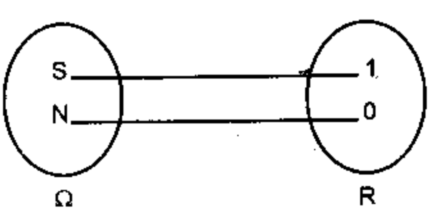
Một trong những khái niệm quan trọng trong lý thuyết xác suất là khái niệm biến ngẫu nhiên. Trước khi đưa vào định nghĩa chính xác của biến ngẫu nhiên, ta xét một số ví dụ sau đây.

Ví dụ 2.1. Gieo ngẫu nhiên một lần đồng tiền cân đối và đồng chất.

Ta xem 1 lần gieo đồng tiền như là tiến hành 1 phép thử ngẫu nhiên "gieo đồng tiền". Không gian biến cố sơ cấp tương ứng với phép thử này là $\Omega = \{S, N\}$. Gọi X là số lần xuất hiện mặt sấp trong một lần gieo đó. Ta thấy rằng X có thể nhận 2 giá trị 0 hoặc 1.

X = 0 nếu đồng tiền xuất hiện mặt ngửa. Điều này có nghĩa là ứng phần tử $N \in \Omega$ cho số 0 với xác suất $P(\{N\}) = \frac{1}{2}$.

X = 1 nếu đồng tiền xuất hiện mặt sấp. Điều này có nghĩa là ứng phần tử $S \in \Omega$ cho số 1 với xác suất $P(\{S\}) = \frac{1}{2}$.



Hình 2.1

Qua các ví dụ trên ta thấy đại lượng X liên quan với phép thử ngẫu nhiên mà ứng với mỗi kết quả của phép thử cho 1 số với một xác suất nào đó, được gọi là giá trị của X. Đại lượng X như thế được gọi là biến ngẫu nhiên (đại lượng ngẫu nhiên).

Bây giờ ta đi đến định nghĩa chính xác bằng toán học.

Định nghĩa 2.1. Hàm X xác định trên không gian biến cố sơ cấp Ω và nhận giá trị trong không gian R (R là trục số thực) nếu với $x \in R$ tập hợp $\{\omega : X(\omega) < x\}$ là biến cố ngẫu nhiên, $(\omega \in \Omega)$.

Ta thường ký hiệu biến ngẫu nhiên bằng chữ in hoa X, Y, Z... Giá trị của nó ký hiệu bằng chữ thường x, y, z ...

Các ví dụ về biến ngẫu nhiên :

– X là số con trai trong một lần sinh (1 con). X là biến ngẫu nhiên. Giá trị mà nó có thể nhận là 0 ; 1.

– X là số viên đạn trúng đích khi bắn liên tiếp n viên đạn độc lập vào một mục tiêu. Giá trị mà nó có thể nhận là 0, 1, ..., n.

– X là số sản phẩm tốt trong 10 sản phẩm được chọn ngẫu nhiên từ lô sản phẩm có 100 sản phẩm tốt và 50 phế phẩm. X cũng là biến ngẫu nhiên. Giá trị mà nó có thể nhận là 0, 1, ..., 10.

– X là số chấm ở mặt trên con xúc xắc khi gieo 1 lần con xúc xắc cân đối và đồng chất. X là biến ngẫu nhiên.

– X là số hạt nảy mầm khi gieo 10 hạt đậu tương. X là biến ngẫu nhiên.

– X là độ cao của một cây tại thời gian t nào đó. X là biến ngẫu nhiên.

Ta dừng lại ở ví dụ đầu : X là số con trai trong 1 lần sinh con. Ta thấy X thỏa mãn định nghĩa biến ngẫu nhiên ở trên.

Thực vậy, ở đây không gian biến cố sơ cấp $\Omega = \{T, G\}$, và X có thể nhận 2 giá trị 0 hoặc 1.

Với mỗi $x \in R$ ta chứng minh tập hợp $[\omega : X(\omega) < x]$ là biến cố ngẫu nhiên. Muốn vậy, ta phân tích tập hợp này khi x biến thiên từ $-\infty$ đến $+\infty$.

$$[\omega : X(\omega) < x] = \begin{cases} \varnothing & x \le 0 \\ \{G\} & 0 < x \le 1 \\ \Omega & x > 1 \end{cases}$$

Ba tập $\varnothing, \{G\}, \Omega$ đều là biến cố ngẫu nhiên. Vậy $[\omega : X(\omega) < x]$ là biến ngẫu nhiên.

Suốt giáo trình này ta chỉ xét những biến ngẫu nhiên thoả mãn điều kiện $P[\omega : X(\omega) < +\infty] = 1$.

Ta quan tâm nghiên cứu hai loại biến ngẫu nhiên : biến ngẫu nhiên rời rạc và biến ngẫu nhiên liên tục.

Định nghĩa 2.2. Biến ngẫu nhiên rời rạc là biến ngẫu nhiên mà các giá trị có thể nhận của nó là tập hợp hữu hạn hoặc vô hạn đếm được.

Trong các ví dụ trên, trừ ví dụ cuối cùng, đều là những biến ngẫu nhiên rời rạc.

Định nghĩa 2.3. Biến ngẫu nhiên liên tục là biến ngẫu nhiên mà giá trị có thể nhận của nó là tất cả mọi điểm trong khoảng (a, b) nào đó, a có thể bằng $-\infty$, b có thể bằng $+\infty$.

b) Hàm phân phân phối của biến ngẫu nhiên

Ta nhận thấy tập $[\omega : X(\omega) < x]$, $x \in R$, thay đổi nếu x thay đổi. Do đó xác suất $P[\omega : X(\omega) < x]$ cũng thay đổi, tức là xác suất này phụ thuộc vào x. Nó là hàm của x.

Định nghĩa 2.4. Gọi hàm $P[\omega : X(\omega) < x]$, $x \in R$, là hàm phân phối của biến ngẫu nhiên X và ký hiệu :

$$F(x) = P[\omega : X(\omega) < x],\ x \in R.$$

Ví dụ 2.2. Tìm hàm phân phối của biến ngẫu nhiên X chỉ số lần xuất hiện mặt sấp khi gieo một lần đồng tiền cân đối và đồng chất.

Giải :

Không gian biến cố sơ cấp tương ứng với phép thử "gieo đồng tiền" là $\Omega = \{S, N\}$. Vì X có thể nhận 2 giá trị 0 hoặc 1. Vì vậy :

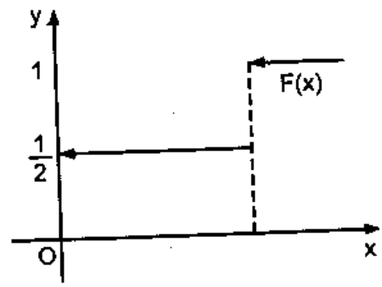
$$\text{Với } x \in R,\ [\omega : X(\omega) < x] = \begin{cases} \varnothing & x \le 0 \\ \{N\} & 0 < x \le 1 \\ \Omega & x > 1 \end{cases}$$

Vậy hàm phân phối của biến ngẫu nhiên X là :

$$F(x) = P[\omega : \; X(\omega) < x] \; = \begin{cases} P(\varnothing) & \text{với } x \le 0 \\ P(\{N\}) & \text{với } 0 < x \le 1 \\ P(\Omega) & \text{với } x > 1 \end{cases}$$

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

Đồ thị của hàm F(x) là (H.2.2).



Hình 2.2

2.1.2. Các tính chất của hàm phân phối

a) Hàm phân phối F(x) là hàm đơn điệu tăng, nghĩa là nếu $x_1 < x_2$ thì $F(x_1) \le F(x_2)$.

Thật vậy, vì $x_1 < x_2$ nên ta có thể viết :

$$[\omega : X(\omega) < x_2] = [\omega : X(\omega) < x_1] \cup [\omega : x_1 \le X(\omega) < x_2].$$

Hai biến cố ở vế phải là xung khắc với nhau. Do đó :

$$P[\omega : X(\omega) < x_2] = P[\omega : X(\omega) < x_1] \quad + P[\omega : x_1 \le X(\omega) < x_2]$$

Suy ra $F(x_2) = F(x_1) + P[\omega : x_1 \le X(\omega) < x_2]$.

Vì xác suất của biến cố bất kỳ không âm, cho nên ta có :

$$F(x_2) \ge F(x_1).$$

Từ chứng minh tính chất trên, ta suy ra công thức :

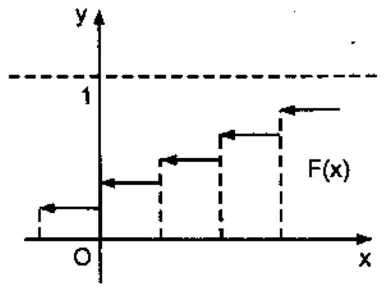
$$P[\omega : x_1 \le X(\omega) < x_2] = F(x_2) - F(x_1) \tag{2.1}$$

Và $P[\omega : X = x_1] = F(x_1 + 0) - F(x_1)$.

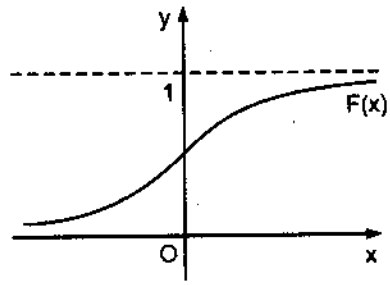
b) Hàm phân phối F(x) là hàm liên tục bên trái, nghĩa là : $\lim\limits_{x \to a-} F(x) = F(a)$.

c) $\lim\limits_{x \to -\infty} F(x) = 0$ *và* $\lim\limits_{x \to +\infty} F(x) = 1$

Đối với hàm phân phối của biến ngẫu nhiên rời rạc thì đồ thị của nó có dạng bậc thang (H.2.3), còn đồ thị của hàm phân phối của biến ngẫu nhiên liên tục tuyệt đối thì đồ thị có dạng (H.2.4).



Hình 2.3



Hình 2.4

2.1.3. Phân phối rời rạc và phân phối liên tục tuyệt đối

a) Phân phối rời rạc

Giả sử X là biến ngẫu nhiên rời rạc. Nó nhận các giá trị có thể với các xác suất tương ứng là $P[X = x_i] = p_i \geq 0$. Cụ thể là :

X	x_1	x_2	...	x_n	...
$P[X = x_i]$	p_1	p_2	...	p_n	...

$$\text{Với } \sum_{i=1}^{\infty} p_i = 1 \tag{2.2}$$

Định nghĩa 2.5. Bảng phân phối xác suất trên được gọi là bảng phân phối xác suất của biến ngẫu nhiên X.

Nếu ta sắp xếp các giá trị $x_1, x_2, ..., x_n, ...$ theo thứ tự tăng dần, ví dụ $x_1 < x_2 < ... < x_n < ...$ thì hàm phân phối của X có thể viết dưới dạng :

$$F(x) = \begin{cases} 0 & \text{với } x \leq x_1 \\ p_1 & \text{với } x_1 < x \leq x_2 \\ p_1 + p_2 & \text{với } x_2 < x \leq x_3 \\ \dots\dots\dots\dots & \\ p_1 + p_2 + \dots + p_k & \text{với } x_k < x \leq x_{k+1} \\ \dots\dots\dots\dots & \end{cases} \tag{2.3}$$

Nếu các giá trị $x_1, x_2, ..., x_n, ...$ ở vị trí bất kỳ thì ta có thể viết hàm phân phối dưới dạng : $F(x) = \sum_{x_i < x} p_i$; $x \in R$. (2.4)

Ví dụ 2.3. Gieo một lần con xúc xắc cân đối và đồng chất.

Ký hiệu X là biến ngẫu nhiên chỉ số chấm ở mặt trên của con xúc xắc. Tìm phân phối xác suất của X. Viết hàm phân phối của X. Tính xác suất $P[0 \le X < 3]$.

Giải :

Vì con xúc xắc là cân đối và đồng chất, nên các biến cố xuất hiện mặt có 1 chấm, 2 chấm, ..., 6 chấm là đồng khả năng và với xác suất là 1/6.

Vậy phân phối xác suất của X là :

X	1	2	3	4	5	6
p	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Hàm phân phối của X là :

$$F(x) = \begin{cases} 0 & \text{với } x \le 1 \\ \frac{1}{6} & \text{với } 1 < x \le 2 \\ \frac{1}{3} & \text{với } 2 < x \le 3 \\ \frac{1}{2} & \text{với } 3 < x \le 4 \\ \frac{2}{3} & \text{với } 4 < x \le 5 \\ \frac{5}{6} & \text{với } 5 < x \le 6 \\ 1 & \text{với } x > 6 \end{cases}$$

Xác suất $P[0 \le X < 3] = F(3) - F(0) = \frac{1}{3} - 0 = \frac{1}{3}$.

Ví dụ 2.4. Một lô sản phẩm gồm 12 sản phẩm trong đó có 8 sản phẩm tốt và 4 phế phẩm. Lấy ngẫu nhiên từ lô hàng 3 sản phẩm. Gọi X là số sản phẩm tốt trong 3 sản phẩm lấy ra. Tìm phân phối xác suất của X. Viết hàm phân phối. Tính xác suất $P[1 \le X < 3]$.

Giải :

Phân phối xác suất của X là : $P[X = k] = \dfrac{C_8^k \times C_4^{3-k}}{C_{12}^3}$, k = 0, 1, 2, 3.

Hoặc có thể viết dưới dạng bảng sau :

X	0	1	2	3
P[X = k]	$\dfrac{4}{220}$	$\dfrac{48}{220}$	$\dfrac{112}{220}$	$\dfrac{56}{220}$

Hàm phân phối của X là :

$$F(x) = \begin{cases} 0 & \text{với } x \le 0 \\ \dfrac{4}{220} & \text{với } 0 < x \le 1 \\ \dfrac{52}{220} & \text{với } 1 < x \le -2 \\ \dfrac{164}{220} & \text{với } 2 < x \le 3 \\ 1 & \text{với } x > 3 \end{cases}$$

Xác suất $P[\,1 \le X < 3] = F(3) - F(1) = \dfrac{164}{220} - \dfrac{4}{220} = \dfrac{160}{220} = \dfrac{8}{11}$.

Ví dụ 2.5. Bắn liên tiếp 3 viên đạn độc lập vào một mục tiêu. Xác suất trúng đích của mỗi viên đạn là 0,5. Gọi X là số viên đạn trúng đích trong 3 viên.

Tìm hàm phân phối xác suất của X. Viết hàm phân phối của X.

Tính xác suất $P[X \ge 1]$.

Giải :

Xem việc bắn 3 viên đạn độc lập vào 1 mục tiêu như tiến hành dãy 3 phép thử Bernoulli. Xác suất bắn trúng đích (biến cố A) của mỗi viên đạn là $p = \dfrac{1}{2}$.

Theo công thức xác suất nhị thức ta có :

$$P[X = k] = C_3^k \left(\frac{1}{2}\right)^k \left(1 - \frac{1}{2}\right)^{3-k}, \ k = 0, 1, 2, 3.$$

$P[X = k] = C_3^k\left(\frac{1}{2}\right)^3$, k = 0, 1, 2, 3, là phân phối xác suất của X.

Ta có thể viết dưới dạng bảng sau :

X	0	1	2	3
p	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

Hàm phân phối của X là :

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

Xác suất $P[X \ge 1] = 1 - P[X < 1] = 1 - F(1) = 1 - \frac{1}{8} = \frac{7}{8}$.

b) Phân phối liên tục tuyệt đối

Định nghĩa 2.6. Biến ngẫu nhiên X có phân phối liên tục tuyệt đối nếu hàm phân phối của nó có dạng : $F(x) = \int_{-\infty}^{x} f(u)du$, $x \in R$. (2.5)

Hàm dưới dấu tích phân f(x) được gọi là hàm mật độ của X. Theo tính chất của tích phân xác định ta có F'(x) = f(x) tại các điểm liên tục của f(x).

Từ tính chất của hàm phân phối ta suy ra tính chất của hàm mật độ là :

+ $f(x) \ge 0$.

Thật vậy, ta biết rằng $\frac{F(x + \Delta x) - F(x)}{\Delta x} \ge 0$, vì F(x) là đơn điệu tăng nên $F(x + \Delta x) - F(x)$ và Δx có cùng dấu.

Do đó $f(x) = \lim\limits_{\Delta x \to 0} \frac{F(x + \Delta x) - F(x)}{\Delta x} \ge 0$

$$+\int_{-\infty}^{+\infty} f(x)dx = 1, \text{ vì } \lim_{x\to+\infty} F(x) = 1 \tag{2.6}$$

$$+\ P[\omega : a \le X(\omega) < b] = \int_a^b f(x)dx$$

Thật vậy, $P[a \le X < b] = F(b) - F(a) = \int_{-\infty}^{b} f(x)dx - \int_{-\infty}^{a} f(x)dx = \int_a^b f(x)dx$

Ví dụ 2.6. Giả sử hàm phân phối của biến ngẫu nhiên X là :

$$F(x) = \frac{1}{\pi}\text{arctg}x + \frac{1}{2}$$

Tìm hàm mật độ của X và tính xác suất $P[-1 \le X < 1]$.

Giải :

Ta có hàm mật độ $f(x) = F'(x) = \dfrac{1}{\pi(1+x^2)}$

Xác suất $P[-1 \le X < 1] = F(1) - F(-1) =$

$$= \frac{1}{2} + \frac{1}{\pi}\text{arctg}(1) - \left(\frac{1}{2} + \frac{1}{\pi}\text{arctg}(-1)\right) = \frac{1}{\pi}\left[\frac{\pi}{4} - \left(-\frac{\pi}{4}\right)\right] = \frac{1}{2}$$

Ví dụ 2.7. Giả sử hàm mật độ của biến ngẫu nhiên X là :

$$f(x) = \begin{cases} Ae^{-\lambda x} & \text{với } x > 0;\ \ \lambda > 0 \\ 0 & \text{với } x \le 0 \end{cases}$$

Tìm A. Tìm hàm phân phối của X.

Giải :

Theo tính chất của hàm mật độ ta có : $\int_{-\infty}^{+\infty} f(x)dx = 1$.

Do đó $\int_0^{+\infty} Ae^{-\lambda x}dx = 1 \Rightarrow A\left(-\frac{1}{\lambda}e^{-\lambda x}\right)\Big|_0^{+\infty} = 1 \Rightarrow \frac{A}{\lambda} = 1 \Rightarrow A = \lambda.$

Hàm phân phối của biến ngẫu nhiên X là :

$$F(x) = \int_{-\infty}^{x} f(u)du = \begin{cases} 0 & \text{với } x \le 0 \\ \int_{0}^{x} \lambda e^{-\lambda u} du & \text{với } x > 0,\ \lambda > 0 \end{cases}$$

$$F(x) = \begin{cases} 0 & \text{với } x \le 0 \\ 1 - e^{-\lambda x} & \text{với } x > 0,\ \lambda > 0 \end{cases}$$

Ví dụ 2.8. Giả sử biến ngẫu nhiên X có phân phối đều trên đoạn [a, b]. Tìm hàm phân phối của X. Tính xác suất của $P[\, a \le X < \frac{a+b}{2}\,]$.

Giải :

Biến ngẫu nhiên X phân phối đều trên [a, b], nghĩa là hàm mật độ của X có dạng : $f(x) = \begin{cases} 0 & \text{với } x \notin [a;b] \\ \dfrac{1}{b-a} & \text{với } x \in [a;b] \end{cases}$

Hàm phân phối của biến ngẫu nhiên X là :

$$F(x) = \int_{-\infty}^{x} f(u)du = \begin{cases} 0 & \text{với } x \le a \\ \dfrac{x-a}{b-a} & \text{với } a < x \le b \\ 1 & \text{với } x > b \end{cases}$$

Xác suất phải tìm là :

$$P[a \le X < \frac{a+b}{2}] = F\left(\frac{a+b}{2}\right) - F(a) = \frac{\frac{a+b}{2} - a}{b-a} - 0 = \frac{1}{2}.$$

2.1.4. Phân phối xác suất của n biến ngẫu nhiên

a) Các khái niệm và tính chất của phân phối đồng thời của n biến ngẫu nhiên

Định nghĩa 2.7. Vectơ ngẫu nhiên n chiều là một bộ gồm n đại lượng ngẫu nhiên $(X_1, X_2, ..., X_n)$; trong đó X_i , $i = \overline{1,n}$ là biến ngẫu nhiên.

Định nghĩa 2.8. Hàm phân phối đồng thời của n biến ngẫu nhiên X_1, $X_2, ..., X_n$ là hàm :

$$F(x_1, x_2, ..., x_n) = P\left(\bigcap_{i=1}^{n} [\omega : X_i(\omega) < x_i]\right), (x_1, x_2, ..., x_n) \in R^n.$$

Tính chất của hàm phân phối đồng thời $F(x_1, x_2, ..., x_n)$:

+ $F(x_1, x_2, ..., x_n)$ là hàm đơn điệu tăng theo các biến $x_1, x_2, ..., x_n$.

+ $F(x_1, x_2, ..., x_n)$ là hàm liên tục bên trái theo các biến.

+ $\lim\limits_{\substack{x_1 \to +\infty \\ x_2 \to +\infty \\ \cdots \\ x_n \to +\infty}} F(x_1, x_2, ..., x_n) = 1$ và $\lim\limits_{x_i \to -\infty} F(x_1, x_2, ..., x_n) = 0$.

Từ định nghĩa và tính chất của hàm phân phối ta suy ra :

$$F_{X_iX_j}(x_i, x_j) = \lim_{\substack{x_1 \to +\infty \\ \cdots \\ x_{i-1} \to +\infty \\ x_{i+1} \to +\infty \\ \cdots \\ x_{j-1} \to +\infty \\ x_{j+1} \to +\infty \\ \cdots \\ x_n \to +\infty}} F(x_1, ..., x_i, ..., x_j, ..., x_n). \tag{2.7}$$

$F_{X_iX_j}(x_i, x_j)$ được gọi là phân phối biên duyên.

Ví dụ 2.9. Giả sử hàm phân phối đồng thời của 2 biến ngẫu nhiên X, Y là :

$$F(x, y) = \begin{cases} 1 - e^{-x} - e^{-y} + e^{-(x+y)} & \text{với} \quad x > 0, \ y > 0 \\ 0 & \text{trong trường hợp khác} \end{cases}$$

Tìm hàm phân phối của X, của Y.

Giải :

Hàm phân phối của X là :

$$F_X(x) = \lim_{y \to +\infty} F(x, y) = \begin{cases} 1 - e^{-x} & \text{với} \quad x > 0 \\ 0 & \text{với} \quad x \le 0 \end{cases}$$

Hàm phân phối của Y là :

$$F_Y(y) = \lim_{x \to +\infty} F(x, y) = \begin{cases} 1 - e^{-y} & \text{với} \quad y > 0 \\ 0 & \text{với} \quad y \le 0 \end{cases}$$

với n = 2, hàm phân phối đồng thời của (X, Y) là F(x, y). Tương tự như trong trường hợp một biến, người ta cũng chứng minh được công thức : Miền D là hình chữ nhật, $D = \{(x, y) : a \le x < b \text{ và } a' \le x < b'\}$.

$$P[\omega : (X, Y) \in D] = F(b, b') - F(a, b') - F(b, a') + F(a, a') \qquad (2.8)$$

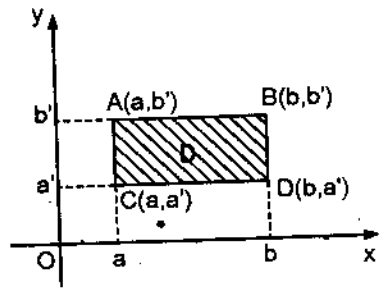
Ví dụ 2.10. Hàm phân phối đồng thời được cho trong ví dụ 2.9.

Tính xác suất $P[\omega : 1 \le X < 2 ; 1 \le Y < 2]$.

Giải :

Ta có :

$P[\omega : \ 1 \le X < 2 ; 1 \le Y < 2] =$



Hình 2.5

$= F(2,2) - F(1,2) - F(2,1) + F(1,1)$

$= 1 - e^{-2} - e^{-2} + e^{-4} - (1 - e^{-1} - e^{-2} + e^{-3}) - (1 - e^{-2} - e^{-1} + e^{-3}) + 1 - e^{-1} - e^{-1} + e^{-2}$

$= e^{-2} - 2e^{-3} + e^{-4}$.

b) Phân phối rời rạc

Ta xét trường hợp 2 biến.

Giả sử X và Y là hai biến ngẫu nhiên rời rạc. Chúng có thể nhận các giá trị

$X : x_1, x_2, \dots, \ x_n, \dots$

và $\quad Y : y_1, y_2, \dots \ y_m, \dots$

Đặt $P([\omega : X = x_i] \cap [\omega : Y = y_j]) = p_{ij} \ge 0$, i, j = 1, 2... ; $\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} p_{ij} = 1$.

Định nghĩa 2.9. Dãy các xác suất p_{ij} ; i, j = 1, 2, ... ; như trên được gọi là phân phối đồng thời của hai biến ngẫu nhiên X, Y.

Ta có thể liệt kê hết các xác suất p_{ij} dưới dạng bảng sau :

X \ Y	y_1	y_2	.....	y_m	....
x_1	p_{11}	p_{12}	......	p_{1m}	....
x_2	p_{21}	p_{22}	.....	p_{2m}	....
⋮	⋮	⋮	⋮	⋮	⋮
x_n	p_{n1}	p_{n2}	....	p_{nm}	...
⋮	....	....	....	....	....

(2.9)

Từ phân phối đồng thời p_{ij}, i, j = 1, 2, ..., của hai biến ngẫu nhiên (X, Y) ta có thể suy ra phân phối của X và của Y theo công thức : Phân phối xác suất của X là :

$$P[X = x_i] = \sum_{j=1}^{\infty} p_{ij} \ , i = 1, 2, \ldots \qquad (2.10)$$

Và phân phối xác suất của Y là :

$$P[Y = y_i] = \sum_{i=1}^{\infty} p_{ij} \ , j = 1, 2, \ldots \qquad (2.11)$$

Ví dụ 2.11. Giả sử phân phối đồng thời của hai biến ngẫu nhiên (X, Y) là :

X \ Y	1	2
1	0,10	0,06
2	0,30	0,18
3	0,20	0,16

Tìm phân phối xác suất của X, của Y, của X + Y.

Giải :

P [X = 1] = 0,10 + 0,06 = 0,16 ;

P [X = 2] = 0,30 + 0,18 = 0,48 ;

P [X = 3] = 0,20 + 0,16 = 0,36 ;

Vậy phân phối xác suất của X là

X	1	2	3
p	0,16	0,48	0,36

Phân phối xác suất của Y là :

P [Y = 1] = 0,1 + 0,30 + 0,20 = 0,60

P [Y = 2] = 0,06 + 0,18 + 0,16 = 0,40

Vậy phân phối xác suất của Y là :

Y	1	2
p	0,60	0,40

Phân phối xác suất của X + Y là :

X + Y có thể nhận các giá trị 2, 3, 4, 5.

$P[X + Y = 2] = P[X = 1 ; Y = 1] = 0{,}10$

$P[X + Y = 3] = P[X = 1 ; Y = 2] + P[X = 2 ; Y = 1] = 0{,}06 + 0{,}30 = 0{,}36$

$P[X + Y = 4] = P[X = 2 ; Y = 2] + P[X = 3 ; Y = 1] = 0{,}18 + 0{,}20 = 0{,}38$

$P[X + Y = 5] = P[X = 3 ; Y = 2] = 0{,}16$

Vậy phân phối xác suất của X + Y là :

X + Y	2	3	4	5
p	0,10	0,36	0,38	0,16

*** Phân phối đa thức**

Xét dãy n phép thử độc lập $G_1, G_2, \ldots . G_n$, trong mỗi phép thử G_i có r biến cố có thể xảy ra là $A_1, A_2, \ldots, A_r$. Gọi p_1 là xác suất xuất hiện biến cố A_1 trong mỗi phép thử ; p_2 là xác suất xuất hiện biến cố A_2 trong mỗi phép thử ;..., p_r là xác suất xuất hiện biến cố A_r trong mỗi phép thử ;

Gọi X_i là số lần xuất hiện biến cố A_i trong n phép thử, $i = \overline{1, r}$.

Khi đó phân phối đồng thời của $X_1, X_2, \ldots, X_r$ là :

$$P[X_1 = k_1, X_2 = k_2, \ldots, X_r = k_r] = \frac{n!}{k_1!k_2!\ldots k_r!} \times p_1^{k_1} p_2^{k_2} \ldots p_r^{k_r} \qquad (2.12)$$

Trong đó $n = k_1 + k_2 + \ldots + k_r$; $p_1 + p_2 + \ldots + p_r = 1$.

Phân phối xác suất ở trên được gọi là phân phối đa thức.

Ví dụ 2.12. Tìm xác suất để khi gieo ngẫu nhiên 20 lần một con xúc xắc cân đối và đồng chất có 4 lần xuất hiện mặt 1 chấm, 3 lần xuất hiện mặt 2 chấm, 5 lần xuất hiện mặt 3 chấm, 2 lần xuất hiện mặt 4 chấm, 2 lần xuất hiện mặt 5 chấm, 4 lần xuất hiện mặt 6 chấm.

Giải :

Vì con xúc xắc cân đối và đồng chất nên các mặt có 1, 2, 3, 4, 5, 6 chấm có khả năng xuất hiện như nhau. Do đó các xác suất của chúng đều bằng $\frac{1}{6}$, nghĩa là $p_1 = p_2 = \ldots = p_6 = 1/6$ (p_i là xác suất của biến cố A_i, $i = \overline{1, 6}$).

Vậy xác suất phải tìm là :

$$P[X_1 = 4, X_2 = 3, X_3 = 5, X_4 = 2, X_5 = 2, X_6 = 4]$$

$$= \frac{20!}{4!3!5!2!2!4!}\left(\frac{1}{6}\right)^4\left(\frac{1}{6}\right)^3\left(\frac{1}{6}\right)^5\left(\frac{1}{6}\right)^2\cdot\left(\frac{1}{6}\right)^2\left(\frac{1}{6}\right)^4$$

$$= \left(\frac{1}{6}\right)^{20}\frac{20!}{(4!)^2 3!5!(2!)^2}$$

c) Phân phối liên tục tuyệt đối (Xét trường hợp n = 2)

Định nghĩa 2.10. Hai biến ngẫu nhiên X, Y có phân phối đồng thời liên tục tuyệt đối nếu hàm phân phối đồng thời của chúng có dạng :

$$F(x, y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(u, v)dudv\ , (x, y) \in R^2 \qquad (2.13).$$

Hàm dưới dấu tích phân f(x, y) được gọi là hàm mật độ đồng thời của hai biến ngẫu nhiên X, Y.

Theo tính chất của tích phân xác định ta có : $f(x, y) = \dfrac{\partial^2 F(x,y)}{\partial x \partial y}$

Tính chất của hàm mật độ đồng thời :

+ $f(x, y) \geq 0$

+ $\displaystyle\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x, y)dxdy = 1$ (2.14)

+ $\displaystyle P[\omega : a \leq X < b\ ;\ c \leq Y < d] = \int_a^b\int_c^d f(x, y)dxdy$

Ví dụ 2.13. Giả sử hàm phân phối đồng thời của hai biến ngẫu nhiên X, Y là :

$$F(x, y) = \begin{cases} 1 - e^{-x} - e^{-y} + e^{-(x+y)} & \text{với } x > 0,\ y > 0 \\ 0 & \text{trong trường hợp khác.} \end{cases}$$

Tìm hàm mật độ đồng thời. Tìm hàm phân phối của X, của Y.

Tính xác suất $P[0 \leq X < 2\ ;\ 0 \leq Y < 2]$.

Giải :

$$\frac{\partial F(x,y)}{\partial x} = \begin{cases} 1 - e^{-x} - e^{-(x+y)} & \text{với} \quad x > 0, \quad y > 0 \\ 0 & \text{trong trường hợp khác.} \end{cases}$$

$$f(x, y) = \frac{\partial^2 F(x,y)}{\partial x \partial y} = \begin{cases} e^{-(x+y)} & \text{với} \quad x > 0, \quad y > 0 \\ 0 & \text{trong trường hợp khác.} \end{cases}$$

– Hàm phân phối của X là :

$$F_X(x) = \lim_{y \to +\infty} F(x,y) = \begin{cases} 1 - e^{-x} & \text{với } x > 0 \\ 0 & \text{với } x \le 0 \end{cases}$$

– Hàm phân phối của Y là :

$$F_Y(y) = \lim_{x \to +\infty} F(x,y) = \begin{cases} 1 - e^{-y} & \text{với } y > 0 \\ 0 & \text{với } y \le 0 \end{cases}$$

– Xác suất :

$$P[0 \le X < 2 \,;\, 0 \le Y < 2] = \int_0^2\int_0^2 f(x,y)dxdy = \int_0^2\int_0^2 e^{-(x+y)}dxdy$$

$$= \int_0^2 e^{-x}\left(\int_0^2 e^{-y}dy\right)dx = \int_0^2 e^{-x}(1 - e^{-2})dx$$

$$= (1 - e^{-2})\int_0^2 e^{-x}dx = (1 - e^{-2})^2$$

Ví dụ 2.14. Giả sử hàm mật độ đồng thời của hai biến ngẫu nhiên X, Y là :

$$f(x,y) = \begin{cases} Ae^{-(x+y)} & \text{với} \quad x > 0 \,; \quad y > 0 \\ 0 & \text{trong trường hợp khác.} \end{cases}$$

Tìm A. Viết hàm phân phối đồng thời của X, Y. Tìm hàm mật độ của X, của Y.

Giải :

Theo tính chất của hàm mật độ ta có : $\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)dxdy = 1$

Ta có : $\int_0^{+\infty}\int_0^{+\infty} Ae^{-(x+y)}dxdy = 1$

$$A\int_0^{+\infty} e^{-x}\left(\int_0^{+\infty} e^{-y}dy\right)dx = 1 \Leftrightarrow A.1 = 1. \text{ Vậy } A = 1.$$

– Hàm phân phối đồng thời của X, Y là :

$$F(x, y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(u,v)dudv = \begin{cases} 0 & \text{với } x, y \text{ nằm góc vuông thứ 2, 3, 4.} \\ \int_0^x\int_0^y e^{-(u+v)}dudv = & \text{với } x > 0 \text{ ; } y > 0 \end{cases}$$

$$= \begin{cases} 0 & \text{với } x \le 0 \text{ hoặc } y \le 0. \\ 1 - e^{-x} - e^{-y} + e^{-(x+y)} & \text{với } x > 0 \text{ ; } y > 0. \end{cases}$$

– Hàm mật độ của X :

$$f_X(x) = \int_{-\infty}^{+\infty} f(x, y)dy = \begin{cases} e^{-x} & \text{với } \quad x > 0. \\ 0 & \text{với } \quad x \le 0. \end{cases}$$

– Hàm mật độ của Y :

$$f_Y(y) = \int_{-\infty}^{+\infty} f(x, y)dx = \begin{cases} e^{-y} & \text{với } \quad y > 0. \\ 0 & \text{với } \quad y \le 0. \end{cases}$$

d) Dãy các biến ngẫu nhiên độc lập

Định nghĩa 2.11. Dãy n biến ngẫu nhiên $X_1, X_2, ..., X_n$ được gọi là độc lập nếu : $F(x_1, x_2, ..., x_n) = F_{X_1}(x_1)F_{X_2}(x_2)....F_{X_n}(x_n)$ (2.15)

Định lý 2.1. Nếu tồn tại hàm mật độ đồng thời $f(x_1, ... x_n)$ của n biến ngẫu nhiên $X_1, ..., X_n$ và hàm mật độ của từng biến $f_{X_1}(x_1), ..., f_{X_n}(x_n)$ thì điều kiện cần và đủ để n biến đó độc lập là :

$$f(x_1, x_2, ..., x_n) = f_{X_1}(x_1)f_{X_2}(x_2)....f_{X_n}(x_n) \qquad (2.16)$$

Trong ví dụ 2.14 : X, Y là hai biến ngẫu nhiên độc lập bởi vì :

$$f_X(x).f_Y(y) = \begin{cases} e^{-x}.e^{-y} & \text{với } x > 0, \ y > 0. \\ 0 & \text{trong trường hợp khác.} \end{cases} = f(x, y)$$

2.2. CÁC SỐ ĐẶC TRƯNG

2.2.1. Kỳ vọng toán

a) Các định nghĩa và ví dụ

Định nghĩa 2.12. Giả sử phân phối xác suất của biến ngẫu nhiên X là :

X	x_1 x_2 .... x_n ....
p	p_2 p_2 .... p_n ....

; $\sum_{i=0}^{\infty} p_i = 1$

Nếu tổng $\sum_{i=0}^{\infty} |x_i| p_i < +\infty$ thì tổng $\sum_{i=1}^{\infty} x_i p_i$ được gọi là kỳ vọng toán của biến ngẫu nhiên X, ký hiệu $E(X) = \sum_{i=0}^{\infty} x_i p_i$ (2.17)

Ví dụ 2.15. Cho phân phối xác suất của X là :

X	0	1
p	$\frac{1}{2}$	$\frac{1}{2}$

Tìm kỳ vọng toán của X.

Giải :

$$E(X) = 0 \times \frac{1}{2} + 1 \times \frac{1}{2} = \frac{1}{2}.$$

Ví dụ 2.16. Bắn liên tiếp ba viên đạn độc lập vào một mục tiêu. Gọi X là số viên đạn trúng đích trong 3 viên. Xác suất trúng đích của mỗi viên đan là 0,5. Tìm phân phối xác suất của X. Tính kỳ vọng toán của X.

Giải :

Coi việc bắn 3 viên đạn như thực hiện dãy 3 phép thử Bernoulli với xác suất $p = \frac{1}{2}$. Theo công thức xác suất nhị thức ta có :

$$P(X = k) = C_3^k \left(\frac{1}{2}\right)^k \left(1 - \frac{1}{2}\right)^{3-k} = C_3^k \times \frac{1}{8} \text{ ; } k = 0, 1, 2, 3.$$

Đây chính là phân phối xác suất của X. Ta có thể viết phân phối này dưới dạng :

X	0	1	2	3
p	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

$$E(X) = 0 \times \frac{1}{8} + 1 \times \frac{3}{8} + 2 \times \frac{3}{8} + 3 \times \frac{1}{8} = 1{,}5$$

Ví dụ 2.17. Một bà mẹ sinh 2 con (mỗi lần sinh 1 con). Giả sử xác suất sinh con trai là 0,51. Gọi X là số con trai trong 2 lần sinh. Tìm phân phối xác suất của X. Tính kỳ vọng của X.

Giải :

Xem việc sinh 2 con như thực hiện 2 phép thử Bernoulli với xác suất p = 0,51. Theo công thức xác suất nhị thức ta có :

$$P(X = k) = C_2^k(0,51)^k(1 - 0,51)^{2-k};\ k = 0, 1, 2.$$

$$P(X = k) = C_2^k(0,51)^k(0,49)^{2-k};\qquad k = 0, 1, 2.$$

Đây là phân phối xác suất của X.

Ta có thể viết nó dưới dạng :

X	0	1	2
P[X = k]	$(0{,}49)^2$	$2(0{,}51)(0{,}49)$	$(0{,}51)^2$

$$E(X) = 0 \times (0{,}49)^2 + 1 \times 2(0{,}51)(0{,}49) + 2 \times (0{,}51^2) = 1{,}02$$

Ví dụ 2.18. Một lô sản phẩm gồm 12 sản phẩm trong đó có 10 sản phẩm tốt và 2 phế phẩm. Lấy ngẫu nhiên 2 sản phẩm từ lô hàng. Gọi X là số phế phẩm trong 2 sản phẩm lấy ra. Tìm phân phối xác suất của X. Tính kỳ vọng của X.

Giải :

Xác suất để có k phế phẩm trong 2 sản phẩm chọn ra là :

$$P[X = k] = \frac{C_2^k C_{10}^{2-k}}{C_{12}^2} \text{ với } k = 0, 1, 2.$$

Đây chính là phân phối xác suất của X. Ta có thể viết nó dưới dạng :

X	0	1	2
P[X = k]	$\frac{45}{66}$	$\frac{20}{66}$	$\frac{1}{66}$

Kỳ vọng toán của X là :

$$E(X) = 0 + 1 \times \frac{20}{66} + 2 \times \frac{1}{66} = \frac{22}{66} = \frac{1}{3}$$

Ví dụ 2.19. Bắn liên tiếp độc lập không hạn định vào một mục tiêu. Bắn cho tới khi nào có 1 viên đầu tiên trúng mục tiêu thì dừng bắn. Giả sử xác suất để viên đạn trúng mục tiêu là p. Gọi X là số viên đạn cần bắn để lần đầu tiên có viên trúng mục tiêu. Tìm phân phối xác suất của X. Tính kỳ vọng toán của X.

Giải :

X có thể nhận vô hạn các giá trị : 1, 2, 3..., n.....

Xác suất để có k viên đạn cần bắn để lần đầu tiên có viên trúng mục tiêu là : $P[X = k] = (1-p)^{k-1}p$; k = 1, 2, 3, ..., n....... Đây là phân phối xác suất của X (Phân phối này được gọi là phân phối hình học).

Đặt q = 1 − p, ta có : $P[X = k] = q^{k-1}p$; k = 1, 2, 3, ..., n.......

Kỳ vọng toán của X là :

$$E(X) = \sum_{k=1}^{\infty} kq^{k-1}p = p\sum_{k=1}^{\infty} kq^{k-1}$$

Ta thấy $\sum_{k=1}^{\infty} q^k$ với $|q| < 1$ là chuỗi số hội tụ và $\sum_{k=1}^{\infty} kq^{k-1}p = (\sum_{k=1}^{\infty} q^k)'$.

Mặt khác $\sum_{k=1}^{\infty} q^k = \left(\frac{q}{1-q}\right)$ (cấp số nhân công bội q, số hạng đầu là q).

Vậy $\sum_{k=1}^{\infty} kq^{k-1} = (\frac{q}{1-q})' = \frac{1}{(1-q)^2} = \frac{1}{p^2}$.

Ta có $E(X) = p \times \frac{1}{p^2} = \frac{1}{p}$.

Định nghĩa 2.13. Nếu biến ngẫu nhiên X có hàm mật độ là f(x) và nếu $\int_{-\infty}^{+\infty} |x| f(x)dx < +\infty$ thì ta gọi tích phân $\int_{-\infty}^{+\infty} xf(x)dx$ là kỳ vọng toán của biến ngẫu nhiên X và ký hiệu $E(X) = \int_{-\infty}^{+\infty} xf(x)dx$ (2.18)

Ví dụ 2.20. Giả sử biến ngẫu nhiên X có phân phối đều trên đoạn [a, b]. Tìm kỳ vọng toán của X.

Giải :

X có phân phối đều trên [a, b], nghĩa là hàm mật độ của nó có dạng :

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

Vậy kỳ vọng của X là :

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

$$E(X) = 0 + \int_{a}^{b} x\frac{1}{b-a}dx + 0 = \frac{1}{b-a}\frac{x^2}{2}\Big|_a^b = \frac{a+b}{2}$$

Ví dụ 2.21. Tuổi thọ của người là đại lượng ngẫu nhiên X có phân phối mũ với hàm mật độ :

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

Tính tuổi thọ trung bình (tính kỳ vọng của X).

Giải :

$$E(X) = \int_{-\infty}^{+\infty} xf(x)dx = \int_{-\infty}^{0} xf(x)dx + \int_{0}^{+\infty} xf(x)dx = 0 + \int_{0}^{+\infty} x\lambda e^{-\lambda x}dx$$

Đặt $t = \lambda x$ ta có :

$$E(X) = \frac{1}{\lambda}\int_{0}^{+\infty} te^{-t}dt = \frac{1}{\lambda}(-te^{-t} + e^{-t})\Big|_0^{+\infty} = \frac{1}{\lambda}.$$

Hoặc ta có thể dùng hàm Gama để tính tích phân dạng ở trên như sau :

Hàm Gama là hàm có dạng $\Gamma(a) = \int_0^{+\infty} x^{a-1}e^{-x}dx$, a>0.

Nếu tính tích phân từng phần 1 lần ta có $\Gamma(a) = (a-1)\Gamma(a-1)$. (2.19)

Nếu a = n (số tự nhiên) thì $\Gamma(n) = (n-1)!$

Áp dụng công thức trên ta tính được $\int_0^{+\infty} te^{-t}dt = \Gamma(2) = 1! = 1$.

Vậy $E(X) = \frac{1}{\lambda}\int_0^{+\infty} te^{-t}dt = \frac{1}{\lambda}\Gamma(2) = \frac{1}{\lambda}$.

Ví dụ 2.22. Giả sử biến ngẫu nhiên X có hàm mật độ là :

$$f(x) = \begin{cases} 0 & \text{với} \quad x \le 0 \\ \dfrac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} x^{\frac{n}{2}-1}e^{-\frac{x}{2}} & \text{với} \quad x > 0 \end{cases}$$

Tính kỳ vọng toán của X.

Giải :

$$E(X) = \int_{-\infty}^{+\infty} xf(x)dx = \int_0^{+\infty} x\frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}x^{\frac{n}{2}-1}e^{-\frac{x}{2}}dx = \frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}\int_0^{+\infty} x^{\frac{n}{2}}e^{-\frac{x}{2}}dx$$

Đặt $t = \frac{x}{2}$ ta có $dt = \frac{dx}{2}$

$$\text{Và } E(X) = \frac{1}{\Gamma\left(\frac{n}{2}\right)}.2\int_0^{+\infty} t^{\frac{n}{2}}e^{-t}dt = \frac{2\Gamma\left(\frac{n}{2}+1\right)}{\Gamma\left(\frac{n}{2}\right)} = 2\frac{\frac{n}{2}\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} = n$$

Vậy E(X) = n.

b) Tính chất của kỳ vọng

+ E(C) = C (C là hằng số)

+ Nếu X, Y có kỳ vọng thì X ± Y cũng có kỳ vọng và

$$E(X \pm Y) = E(X) \pm E(Y)$$

Chứng minh :

Ở đây chỉ chứng minh trong trường hợp X, Y là biến ngẫu nhiên rời rạc với các giá trị tương ứng $x_1, x_2, ..., x_n, ...$ và $y_1, y_2, ..., y_m, ...$

Đặt $P([X = x_i] \cap [Y = y_j]) = p_{ij}$, $i = 1, 2, ... ; j = 1, 2, ...$

Theo định nghĩa kỳ vọng toán ta có :

$$E(X) + E(Y) = \sum_{i=1}^{\infty} x_i p_i + \sum_{j=1}^{\infty} y_j q_j$$

mà $q_j = \sum_{i=1}^{\infty} p_{ij}$; $p_i = \sum_{j=1}^{\infty} p_{ij}$

Do đó :

$$E(X) + E(Y) = \sum_{i=1}^{\infty} x_i \left(\sum_{j=1}^{\infty} p_{ij} \right) + \sum_{j=1}^{\infty} y_j \left(\sum_{i=1}^{\infty} p_{ij} \right) = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} x_i p_{ij} + \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} y_j p_{ij}$$

$$= \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} (x_i + y_j) p_{ij} = E(X + Y).$$ (Điều phải chứng minh).

+ Nếu X, Y là hai biến ngẫu nhiên độc lập và có kỳ vọng E(X), E(Y) thì E(XY) = E(X)E(Y).

Nếu Y = C thì E(CX) = CE(X)

Chứng minh :

Ta chỉ xét trong trường hợp X, Y là biến ngẫu nhiên rời rạc.

Ta có $E(X)E(Y) = \left(\sum_{i=1}^{\infty} x_i p_i \right) \left(\sum_{j=1}^{\infty} y_j q_j \right)$. Các chuỗi sẽ là hội tụ tuyệt đối nên : $E(X)E(Y) = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} x_i x_j p_i q_j$

X, Y là độc lập, nên $p_i q_j = p_{ij}$

Vậy $\quad E(X)E(Y) = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} x_i x_j p_{ij} = E(XY).$

Mệnh đề được chứng minh.

+ Giả sử biến ngẫu nhiên X có hàm mật độ là f(x) và biến ngẫu nhiên $Y = \psi(X)$ có kỳ vọng $E\psi(X)$.

Khi đó :
$$E\psi(X) = \int_{-\infty}^{+\infty} \psi(x)f(x)dx. \qquad (2.20)$$

Nếu X có phân phối rời rạc thì :
$$E\psi(X) = \sum_{i=1}^{\infty} \psi(x_i)p_i \qquad (2.21)$$

Ví dụ 2.23. Cho phân phối xác suất của biến ngẫu nhiên X là :

X	0	1
p	$\frac{1}{2}$	$\frac{1}{2}$

Tính $E(2X + 1)$, EX^3, $E(e^X)$.

Giải :

$$E(2X + 1) = 2E(X) + 1$$

Mà $\quad E(X) = 0 \times \frac{1}{2} + 1 \times \frac{1}{2} = \frac{1}{2}$

Vậy $\quad E(2X + 1) = 2 \times \frac{1}{2} + 1 = 2.$

$$E(X^3) = 0 \times \frac{1}{2} + 1^3 \times \frac{1}{2} = \frac{1}{2}$$

$$E(e^X) = e^0 \times \frac{1}{2} + e^1 \times \frac{1}{2} = \frac{1}{2}(e + 1)$$

Ví dụ 2.24. Giả sử X có phân phối đều trên đoạn [0 ;1].

Tính $E(X^3)$.

Giải :

X có phân phối đều trên đoạn [0 ; 1], nghĩa là hàm mật độ của nó có dạng :

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

Vậy $E(X^3) = \int\limits_{-\infty}^{+\infty} x^3 f(x)dx = 0 + \int\limits_0^1 x^3 dx + 0 = \frac{x^4}{4}\Big|_0^1 = \frac{1}{4}$.

*** Kỳ vọng toán của vectơ ngẫu nhiên n chiều**

Định nghĩa 2.14.

Gọi kỳ vọng toán của vectơ ngẫu nhiên $(X_1, X_2, ..., X_n)$ là vectơ $(EX_1, EX_2, ..., EX_n)$, trong đó EX_i, $i = \overline{1, n}$ là kỳ vọng toán của biến ngẫu nhiên X_i.

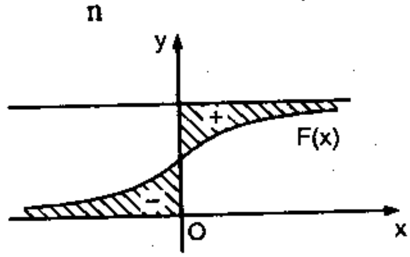
c) Ý nghĩa của kỳ vọng

– Kỳ vọng E(X) là trung bình có trọng lượng. Nếu lặp lại n lần độc lập phép đo đại lượng ngẫu nhiên X, ta nhận được kết quả $X_1, ..., X_n$. Dưới một số giả thiết nhất định thì $\overline{X} = \frac{X_1 + X_2 + ... + X_n}{n}$ sẽ hội tụ về kỳ vọng E(X) khi $n \to \infty$.

Vì vậy với n đủ lớn ta có thể xem $\overline{X} \approx E(X)$

– Ý nghĩa hình học của kỳ vọng (H.2.6)

$$E(X) = \int\limits_0^{+\infty} (1 - F(x))dx - \int\limits_0^{+\infty} F(x)dx$$



Hình 2.6

2.2.2. Phương sai

Định nghĩa 2.15. Gọi $E(X - E(X))^2$ là phương sai của biến ngẫu nhiên X và ký hiệu $D(X) = E(X - E(X))^2$.

Tính chất của phương sai :

+ $DX = \sum\limits_{i=1}^{\infty} (x_i - EX)^2 p_i$ (Nếu X có phân phối rời rạc).

+ $DX = \int\limits_{-\infty}^{+\infty} (x - E(X))^2 f(x)dx$ (Nếu X có hàm mật độ f(x)).

+ $DX = E(X^2) - (E(X))^2$.

+ Nếu X và Y là hai biến ngẫu nhiên độc lập và có DX, DY thì

$$D(X + Y) = DX + DY.$$

+ $D(cX) = c^2DX$.

Ví dụ 2.25. Cho hai biến ngẫu nhiên X, Y độc lập và có phân phối xác suất là :

X	0	1
p	$\frac{1}{2}$	$\frac{1}{2}$

và

Y	2	0	−2
p	$\frac{1}{6}$	$\frac{2}{3}$	$\frac{1}{6}$

Tính E(XY), E(X + Y), D(2X + Y + 3), $E(X^2)$, $E(Y^2)$.

Giải :

$E(X) = 0 \times \frac{1}{2} + 1 \times \frac{1}{2} = \frac{1}{2}$; $\qquad E(X^2) = 0^2 \times \frac{1}{2} + 1^2 \times \frac{1}{2} = \frac{1}{2}$

$$DX = E(X^2) - (E(X))^2 = \frac{1}{2} - \left(\frac{1}{2}\right)^2 = \frac{1}{4}$$

$$E(Y) = 2 \times \frac{1}{6} + 0 + (-2) \times \frac{1}{6} = 0$$

$$E(Y^2) = 2^2 \times \frac{1}{6} + 0 + 2^2 \times \frac{1}{6} = \frac{8}{6} = \frac{4}{3}$$

$$DY = \frac{4}{3} - 0 = \frac{4}{3}$$

Ta có : $E(XY) = E(X)E(Y) = 0 \times \frac{1}{2} = 0$

Và $E(X + Y) = E(X) + E(Y) = \frac{1}{2} + 0 = \frac{1}{2}$

$$D(2X + Y + 3) = 4DX + DY = 4 \times \frac{1}{4} + \frac{4}{3} = 1 + \frac{4}{3} = \frac{7}{3}$$

Ví dụ 2.26. Với giả thiết như ví dụ 2.24. Tìm phương sai của X.

Giải :

$$E(X^2) = \int_0^1 x^2dx = \frac{x^3}{3}\Big|_0^1 = \frac{1}{3} \text{ ; Ta có } E(X) = \frac{1}{2}$$

Phương sai của X là : $DX = E(X^2) - (EX)^2 = \frac{1}{3} - \left(\frac{1}{2}\right)^2 = \frac{1}{12}$

Ví dụ 2.27. Với giả thiết như trong ví dụ 2.22. Tìm phương sai DX.

Giải :

$$E(X^2) = \int_0^{+\infty} x^2 \frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} dx = \frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} \int_0^{+\infty} x^{\frac{n}{2}+1} e^{-\frac{x}{2}} dx$$

Đặt $t = \frac{x}{2}$. Ta có $dt = \frac{dx}{2}$

$$E(X^2) = \frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} 2^{\frac{n}{2}+1} 2 \int_0^{+\infty} t^{\frac{n}{2}+1} e^{-t} dt = \frac{4}{\Gamma\left(\frac{n}{2}\right)} \Gamma\left(\frac{n}{2}+2\right)$$

$$E(X^2) = \frac{4\left(\frac{n}{2}+1\right)\left(\frac{n}{2}\right)\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} = n^2 + 2n.$$ Ta biết $E(X) = n$.

Vậy $DX = E(X^2) - (E(X))^2 = n^2 + 2n - n^2 = 2n$.

2.2.3. Mômen

a) Mômen gốc bậc k

Định nghĩa 2.16. Gọi $E(X^k)$ là mômen gốc bậc k của biến ngẫu nhiên X.

$k = 1$, $E(X)$ là kỳ vọng của X.

b) Mômen trung tâm

Định nghĩa 2.17. Gọi $E(X - E(X))^k$ là mômen trung tâm bậc k của biến ngẫu nhiên X.

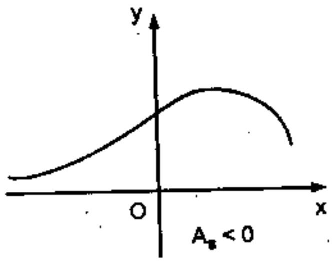
Nếu $k = 1$ thì $E(X - E(X)) = 0$.

Nếu $k = 2$ thì $E(X - E(X))^2$ là phương sai của X.

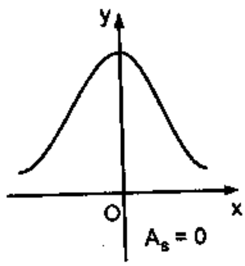
Nếu $k = 3$ thì $E(X - E(X))^3$, ta sử dụng đặc trưng này để đo độ lệch trái, lệch phải của đồ thị hàm mật độ.

Trong thực tế người ta thường dùng các đại lượng sau để đo độ lệch phải, lệch trái của đồ thị của hàm mật độ.

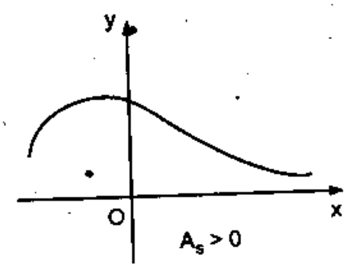
$$A_s = \begin{cases} \dfrac{1}{\sigma^3}\displaystyle\int_{x_{min}}^{x_{max}} \left(x - E(X)\right)^3 f(x)\,dx & \text{(X có hàm mật độ f(x).} \\ \dfrac{1}{\sigma^3}\displaystyle\sum_{i\geq 1}\left(x_i - E(X)\right)^3 p_i & \text{(X có phân phối rời rạc).} \end{cases} \quad (2.22)$$



Hình 2.7



Hình 2.8



Hình 2.9

Khi $A_s < 0$ thì phân phối có độ lệch phải (H.2.7).

Khi $A_s = 0$ thì phân phối đối xứng (H.2.8).

Khi $A_s > 0$ thì phân phối có độ lệch trái (H.2.9).

Nếu k = 4, $E(X - E(X))^4$ mô tả độ nhọn của đồ thị hàm mật độ. Trong thực tế người ta thường dùng đại lượng sau để mô tả độ nhọn của đồ thị hàm mật độ.

$$\varepsilon_X = \begin{cases} \dfrac{1}{\sigma^4}\displaystyle\int_{x_{min}}^{x_{max}} \left(x - E(X)\right)^4 f(x)\,dx - 3 & \text{(Với X có hàm mật độ f(x)).} \\ \dfrac{1}{\sigma^4}\displaystyle\sum_{i\geq 1}\left(x_i - E(X)\right)^4 p_i - 3 & \text{(Với X có phân phối rời rạc).} \end{cases} \quad (2.23)$$

2.2.4. Các số đặc trưng khác

a) Mốt (mod)

Định nghĩa 2.18. Mốt (mod) là giá trị của biến ngẫu nhiên X được ký hiệu là x_{mod} mà tại đó hàm mật độ f(x) đạt cực đại. Trường hợp X là biến ngẫu nhiên rời rạc, x_{mod} là giá trị tại số xác suất $P[X = x_{mod}]$ là lớn nhất.

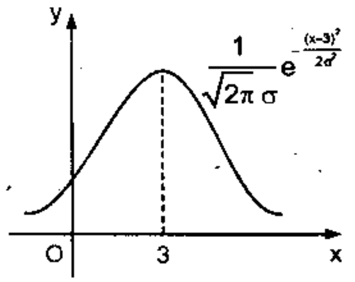
Ví dụ 2.28. Cho phân phối xác suất của biến ngẫu nhiên X là :

X	-1	0	1
p	$\frac{1}{6}$	$\frac{2}{3}$	$\frac{1}{6}$

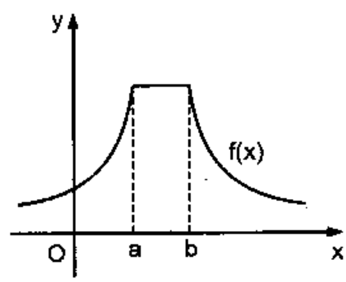
$x_{mod} = 0$ vì $P[X = 0] = 2/3$ (lớn nhất).

Ví dụ 2.29. Giả sử biến ngẫu nhiên X có hàm mật độ :

$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-3)}{2\sigma^2}}$. Ta thấy $x_{mod} = 3$ (H2.10) tại đó hàm mật độ đạt cực đại.



Hình 2.10



Hình 2.11

Chú ý : Mod có thể có vô số giá trị. Hình (H.2.11), cả khoảng (a, b) mà f(x) đạt cực đại, $a \le x \le b$.

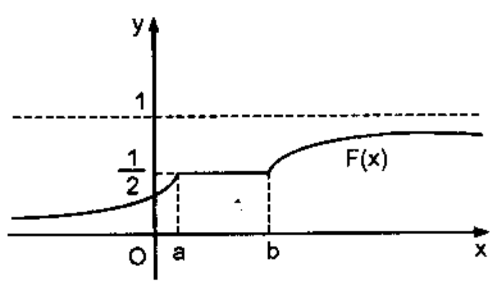
b) Trung vị (median)

Định nghĩa 2.19. Trung vị (median) là giá trị của biến ngẫu nhiên X, ký hiệu x_{Me} mà tại đó $F(x_{Me}) = 1/2$.

Ví dụ 2.30. Cho hàm phân phối của biến ngẫu nhiên X là :

$$F(x) = \begin{cases} 0 & \text{với} \quad x \le 0 \\ 1 & \text{với} \quad 0 < x \le 1 \\ 0 & \text{với} \quad x > 1 \end{cases}$$

Từ $F(x) = 1/2$ suy ra $x_{Me} = 1/2$.



Hình 2.12

Chú ý : Trung vị có thể có vô số giá trị hình (H.2.12).

c) Hệ số biến thiên

Định nghĩa 2.20. Tỷ số $M = \dfrac{\sqrt{DX}}{E(X)}$ được gọi là hệ số biến thiên.

Người ta dùng hệ số này để so sánh độ biến động của nhiều đám đông với nhau.

2.2.5. Hệ số tương quan và ma trận tương quan

a) Hệ số tương quan của hai đại lượng ngẫu nhiên

Định nghĩa 2.21. Gọi tỷ số $\rho(X, Y) = \dfrac{E(X - E(X))(Y - E(Y))}{\sqrt{DXDY}}$ (2.24)

là hệ số tương quan của hai đại lượng ngẫu nhiên X, Y.

Nếu một trong 2 đại lượng ngẫu nhiên X, Y là hằng số thì quy ước $\rho(X, Y) = 0$.

Ý nghĩa : Hệ số tương quan $\rho(X, Y)$ cho biết mức độ quan hệ giữa hai đại lượng X và Y.

Tính chất của hệ số tương quan :

$$-1 \le \rho(X, Y) \le 1 \qquad (2.25)$$

$|\rho(X, Y)| = 1$ khi và chỉ khi X và Y là phụ thuộc tuyến tính.

b) Ma trận tương quan của n đại lượng ngẫu nhiên

Định nghĩa 2.22.

Ma trận tương quan của n biến ngẫu nhiên $X_1, X_2, ..., X_n$ là ma trận có dạng :

$$D = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

trong đó : $a_{ij} = E(X_i - E(X_i))(X_j - E(X_j))$; $i, j = \overline{1,n}$

$a_{ii} = DX_i$

D là ma trận đối xứng.

Hệ số tương quan của X_i, X_j là :

$$\rho(X_i, X_j) = \frac{E(X_i - EX_i)(X_j - EX_j)}{\sqrt{DX_i DX_j}}$$

Ví dụ 2.31. Giả sử X, Y là hai biến ngẫu nhiên rời rạc có phân phối đồng thời là :

Y	X		
	$x_1 = 2$	$x_2 = 5$	$x_3 = 8$
$y_1 = 0,4$	0,15	0,30	0,35
$y_2 = 0,8$	0,05	0,12	0,03

Tìm phân phối xác suất của X, của Y. Tính ma trận tương quan D của X, Y. Tính hệ số tương quan ρ(X, Y).

Giải :

Phân phối xác suất của biến ngẫu nhiên X là : $p_j = \sum_{i=1}^{2} p_{ij}$

X	2	5	8
p_j	0,20	0,42	0,38

Phân phối xác suất của biến ngẫu nhiên Y là : $q_i = \sum_{j=1}^{3} p_{ij}$

Y	0,4	0,8
q_i	0,80	0,20

Kỳ vọng của biến ngẫu nhiên X là :

$$E(X) = 2 \times 0,20 + 5 \times 0,42 + 8 \times 0,38 = 5,54.$$

Kỳ vọng của biến ngẫu nhiên Y là :

$$E(Y) = 0,4 \times 0,8 + 0,8 \times 0,20 = 0,48.$$

Mômen bậc 2 của X, của Y là :

$$E(X^2) = 2^2 \times 0,2 + 5^2 \times 0,42 + 8^2 \times 0,38 = 35,62$$

$$E(Y^2) = 0,4^2 \times 0,80 + 0,8^2 \times 0,20 = 0,256$$

Phương sai của X, của Y là :

$$DX = E(X^2) - (EX)^2 = 35,62 - (5,54)^2 = 4,9084$$

$$DY = E(Y^2) - (EY)^2 = 2,256 - (0,48)^2 = 0,0256$$

$$a_{12} = a_{21} = E(X - EX)(Y - EY) = E(XY) - E(X)E(Y)$$

$$E(XY) = 2 \times 0,4 \times 0,15 + 5 \times 0,4 \times 0,30 + 8 \times 0,4 \times 0,35 + \\ + 2 \times 0,8 \times 0,05 + 5 \times 0,8 \times 0,12 + 8 \times 0,8 \times 0,03 \\ = 2,592$$

$$a_{12} = a_{21} = 2,5920 - 5,54 \times 0,48 = -0,0672.$$

Ma trận tương quan của hai biến ngẫu nhiên X, Y là :

$$D = \begin{bmatrix} 4,9084 & -0,0672 \\ -0,0672 & 0,0256 \end{bmatrix}$$

Hệ số tương quan của X, Y là :

$$\rho(X, Y) = \frac{E(X - EX)(Y - EY)}{\sqrt{DXDY}} = \frac{-0,0672}{\sqrt{4,9084 \times 0,0256}} = \frac{-0,0672}{0,3544}$$

$$\rho(X, Y) = -0,1895$$

2.2.6. Kỳ vọng có điều kiện

a) Phân phối có điều kiện và kỳ vọng điều kiện

Giả sử hai biến ngẫu nhiên X, Y có phân phối đồng thời là :

$$P([X = x_i] \cap [Y = y_j]) = p_{ij}$$

Định nghĩa 2.23. Phân phối có điều kiện của biến ngẫu nhiên Y với điều kiện $X = x_i$ là tỷ số $\frac{p_{ij}}{P[X = x_i]}$ và ký hiệu là :

$$P[Y = y_j / X = x_i] = \frac{p_{ij}}{P[X = x_i]}, \ j = 1, 2, ...$$

Tương tự : gọi tỷ số $\frac{P([X = x_i] \cap [Y = y_j])}{P[Y = y_j]} = P\left[X = x_i / Y = y_j\right]$, $i = 1, 2, ...$ là phân phối có điều kiện của biến ngẫu nhiên X với điều kiện $Y = y_j$.

Định nghĩa 2.24. Giả sử X, Y có hàm mật độ đồng thời là f(x, y). Gọi tỷ số $\frac{f(x,y)}{f_X(x)} = f(y/x)$ là mật độ có điều kiện của biến ngẫu nhiên Y với điều kiện X = x.

Tương tự : Gọi tỷ số $\frac{f(x,y)}{f_Y(y)} = f(x/y)$ là mật độ có điều kiện của biến ngẫu nhiên X với điều kiện Y = y.

Định nghĩa 2.25. Gọi tổng $\sum_{j=1}^{\infty} y_j P\left[Y = y_j / X = x\right]$ là kỳ vọng có điều kiện của biến ngẫu nhiên Y với điều kiện X = x và ký hiệu :

$$E(Y/X = x) = \sum_{j=1}^{\infty} y_j P\left[Y = y_j / X = x\right]$$

Tương tự, gọi tổng $\sum_{i=1}^{\infty} x_i P[X = x_i / Y = y] = E(X/Y = y)$ là kỳ vọng có điều kiện của biến ngẫu nhiên X với điều kiện Y = y.

Định nghĩa 2.26. Gọi tích phân $E(Y/X = x) = \int_{-\infty}^{+\infty} yf(y/x)dx$ là kỳ vọng có điều kiện của biến ngẫu nhiên Y với điều kiện X = x.

Tương tự, gọi tích phân $E(X/Y = y) = \int_{-\infty}^{+\infty} xf(x/y)dx$ là kỳ vọng có điều kiện của biến ngẫu nhiên X với điều kiện Y = y.

Ví dụ 2.32. Với số liệu như trong ví dụ 2.31. Tìm kỳ vọng điều kiện của Y với điều kiện X = 5.

Giải :

Trước hết ta phải tính phân phối có điều kiện của Y với điều kiện X = 5.

Ta biết $$P\left[Y = y_j / X = 5\right] = \frac{P\left([X = 5] \cap [Y = y_j]\right)}{P[X = 5]}.$$

$$P[Y = 0,4/X = 5] = \frac{0,30}{0,42} = \frac{5}{7}$$

$$P[Y = 0,8/X = 5] = \frac{0,12}{0,42} = \frac{2}{7}$$

Ta viết dưới dạng bảng :

Y	0,4	0,8
$P[Y = y_j/X = 5]$	$\frac{5}{7}$	$\frac{2}{7}$

Kỳ vọng điều kiện của Y với điều kiện X = 5.

$$E(Y/X = 5) = 0,4 \times \frac{5}{7} + 0,8 \times \frac{2}{7} = \frac{2}{7} + \frac{1,6}{7} = \frac{3,6}{7}.$$

Ví dụ 2.33. Cho hàm mật độ đồng thời của hai biến ngẫu nhiên X, Y là :

$$f(x, y) = \begin{cases} \dfrac{12}{(1 + x + y)^5} & \text{với } x > 0,\ y > 0 \\ 0 & \text{trong trường hợp khác.} \end{cases}$$

Tìm f(y/x) ; E(Y/X = x).

Giải :

Ta có $f(y/x) = \dfrac{f(x, y)}{f_X(x)}$, trong đó $f_X(x) = \int\limits_{-\infty}^{+\infty} f(x, y)dy$

Thay cụ thể biểu thức của f(x, y) ta có :

$$f_X(x) = \begin{cases} \int\limits_{0}^{+\infty} \dfrac{12}{(1 + x + y)^5} dy & \text{với } x > 0 \\ 0 & \text{với } x \le 0 \end{cases} = \begin{cases} \dfrac{x}{(1 + x)^4} & \text{với } x > 0 \\ 0 & \text{với } x \le 0 \end{cases}$$

$$\text{Vậy } f(y/x) = \frac{f(x, y)}{f_X(x)} = \begin{cases} \dfrac{12(1 + x)^4}{3(1 + x + y)^5} & \text{với } x > 0 \text{ ; } y > 0 \\ 0 & \text{trong trường hợp khác.} \end{cases}$$

Kỳ vọng điều kiện của biến ngẫu nhiên Y với điều kiện X = x là :

$$E\left(Y / X = x\right) = \int\limits_{-\infty}^{+\infty} yf(y/x)dy = \int\limits_{0}^{+\infty} y \frac{4(1 + x)^4}{(1 + x + y)^5} dy = \frac{1}{3}(1 + x)$$

b) Tính chất của kỳ vọng có điều kiện

+ Nếu c_1, c_2 là hai hằng số và Y_1, Y_2 có kỳ vọng điều kiện $E(Y_1/X)$, $E(Y_2/X)$ thì $E(c_1Y_1 + c_2Y_2/X) = c_1E(Y_1/X) + c_2E(Y_2/X)$

+ E(Y) = E(E(Y/X)).

+ DY = D(E(Y/X)) + E(D(Y/X)).

Trong đó $D(Y/X) = E[(Y - E(Y/X))^2/X]$ được gọi là phương sai điều kiện của Y với điều kiện X đã cho.

2.3. MỘT SỐ PHÂN PHỐI THÔNG DỤNG

2.3.1. Phân phối nhị thức

Định nghĩa 2.27. Biến ngẫu nhiên X có phân phối nhị thức với tham số (n, p) nên phân phối của nó có dạng :

$$P[X = k] = C_n^k p^k q^{n-k} \ ; k = 0, 1, ..., n \ ; , q = 1 - p$$

Khảo sát dáng điệu của xác suất P[X = k] khi k biến thiên từ 0 đến n.

Muốn vậy ta xét : $\dfrac{P[X = k + 1]}{P[X = k]} = \dfrac{C_n^{k+1}p^{k+1}q^{n-k-1}}{C_n^k p^k q^{n-k}} = \dfrac{(n - k)p}{(k + 1)q}$

Tỷ số này ≥ 1 tương đương với (n – k)p ≥ (k + 1)q hoặc tương đương k ≤ np – q.

Tương tự, tỷ số đó < 1 ứng với những giá trị của k > np – q.

Từ kết quả trên ta kết luận rằng :

Xác suất P[X = k] tăng khi k tăng từ 0 đến np – q và nó giảm khi k tiếp tục tăng từ np – q đến n. Vì k nhận giá trị nguyên nên ta có kết luận sau :

– Nếu np – q là nguyên thì xác suất $P_n(k) = P[X = k]$ đạt cực đại tại 2 điểm : $k_0 = np - q$ và $k_1 = np - q + 1$.

– Nếu np – q không nguyên thì chỉ có 1 giá trị của k là $k_0 = [np - q] + 1$, mà tại đó xác suất P[X = k] đạt cực đại.

Trong đó [np – q] ký hiệu phần nguyên của np – q.

a) Kỳ vọng toán của X

$$E(X) = \sum_{k=0}^{n} kC_n^k p^k q^{n-k} = np\sum_{k=1}^{n} \frac{(n-1)!\,p^{k-1}q^{n-1-(k-1)}}{(k-1)!(n-1-(k-1))!}$$

Đặt r = k − 1, ta có :

$$E(X) = np\sum_{r=0}^{n-1} C_{n-1}^{r}p^{r}q^{n-1-r} = np(p+q)^{n-1} = np$$

b) Phương sai của X

Trước hết ta tính $E(X^2) = \sum_{k=0}^{n} k^2 C_n^k p^k q^{n-k}$

Tính tương tự như phần trên, ta nhận được $E(X^2) = npq + n^2p^2$.

Phương sai của X là : $DX = E(X^2) - (EX)^2$

$$DX = npq + n^2p^2 - (np)^2 = npq.$$

Ví dụ 2.34. Gieo 1000 hạt đậu tương. Xác suất nảy mầm của mỗi hạt là p = 0,8. Gọi X là số hạt nảy mầm trong 1000 hạt. Tìm phân phối xác suất của X. Tính xác suất $P[X \geq 1]$. Tính kỳ vọng và phương sai của X.

Giải :

Gieo 1000 hạt đậu tương được xem như thực hiện 1000 phép thử Bernoulli với xác suất nảy mầm p = 0,8. Theo công thức xác suất nhị thức ta có :

$$P[X = k] = C_{1000}^{k}(0,8)^k(0,2)^{1000-k}, \ k = 0, 1, 2, ..., 1000$$

Đây chính là phân phối xác suất của X.

– Xác suất : $P[X \geq 1] = 1 - P[X = 0] = 1 - (0,2)^{1000}$.

– Kỳ vọng của X : $E(X) = np = 1000 \times 0,8 = 800$.

– Phương sai của X : $DX = npq = 1000 \times 0,8 \times 0,2 = 160$.

Ví dụ 2.35. Một lô sản phẩm có tỷ lệ phế phẩm là 5%. Lấy ngẫu nhiên liên tiếp (có hoàn lại) 30 sản phẩm từ lô hàng. Gọi X là số sản phẩm tốt trong 30 sản phẩm lấy ra. Tìm phân phối xác suất của X. Tính kỳ vọng toán và phương sai của X. Tính số sản phẩm tốt có khả năng nhất trong 30 sản phẩm lấy ra.

Giải :

Xem việc lấy liên tiếp (có hoàn lại) 30 sản phẩm từ lô hàng như việc thực hiện dãy 30 phép thử Bernoulli. Xác suất để 1 sản phẩm là sản phẩm tốt là p = 1 − 0,05 = 0,95.

Theo công thức xác suất nhị thức ta có :

$$P[X = k] = C_{30}^{k}(0,95)^{k}(0,05)^{30-k} \ ; k = 0, 1, ..., 30$$

Đây là phân phối xác suất của X.

Kỳ vọng toán của X : E(X) = np = 30 × 0,95 = 28,5.

Phương sai của X : DX = npq = 30 × 0,95 × 0,05 = 1,425.

2.3.2. Phân phối Poisson

Định nghĩa 2.28. Biến ngẫu nhiên X có phân phối Poisson với tham số $\lambda > 0$ nếu phân phối xác suất của nó có dạng :

$$P[X = k] = \frac{\lambda^k e^{-\lambda}}{k!}, k = 0, 1, 2, ...$$

Hàm phân phối của X : $F(x) = \sum_{k<x} \frac{\lambda^k e^{-\lambda}}{k!}$ với $x \in R$.

Kỳ vọng của X : $E(X) = \sum_{k=0}^{\infty} k \frac{\lambda^k e^{-\lambda}}{k!} = \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!}$

Đặt r = k − 1 ta có :

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

Phương sai của X là $DX = \lambda$.

Sự liên hệ giữa phân phối nhị thức với phân phối Poisson :

Định lý 2.2. Nếu $np \to \lambda$, $p \to 0$ khi $n \to \infty$ thì

$$P_n(k) \to \frac{\lambda^k e^{-\lambda}}{k!} \text{ khi } n \to \infty$$

Chứng minh :

Đặt $np = \lambda$ ta suy ra $p = \lambda/n$.

Ta có $P_n(k) = C_n^k p^k (1-p)^{n-k}$

$$= \frac{n(n-1)...(n-(k-1))}{k!} p^k (1-p)^{n-k}$$

$$= \frac{n(n-1)...(n-(k-1))}{k!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k}$$

giữ λ không đổi và cho $n \to \infty$ thì $p \to 0$ và ta có :

$$\lim P_n(k) = \lim_{n\to\infty} \frac{n(n-1)...(n-(k-1))}{k!\,n^k}\lambda^k\left(1-\frac{\lambda}{n}\right)^{n-k}$$

$$= \frac{\lambda^k}{k!}\lim_{n\to\infty}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)...\left(1-\frac{k-1}{n}\right)\left(1-\frac{\lambda}{n}\right)^{n-k}$$

$$= \frac{\lambda^k}{k!}\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{n}\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{-k} = \frac{\lambda^k e^{-\lambda}}{k!}$$

Định lý được chứng minh.

Từ định lý trên ta rút ra công thức xấp xỉ với n khá lớn và p khá bé ta có :

$$P_n(k) \approx \frac{\lambda^k e^{-\lambda}}{k!}\ ;\ \lambda \approx np. \qquad (2.26)$$

Ví dụ 2.36. Một bao thóc giống có tỷ lệ hạt lép là 0,0002. Chọn ngẫu nhiên liên tiếp có hoàn lại 10000 hạt. Tính gần đúng xác suất để trong 10000 hạt có đúng 5 hạt lép.

Giải :

Kiểm tra liên tiếp có hoàn lại 10000 hạt được xem như thực hiện 10000 phép thử Bernoulli, $p = 0{,}0002$; $k = 5$.

Theo công thức xấp xỉ ở trên ta có $\lambda = 10^4 \times 0{,}0002 = 2$ và xác suất phải tìm là :

$$P_{10^4}(5) \approx \frac{2^5 e^{-2}}{5!}$$

2.3.3. Phân phối siêu bội

Trong chương I ta đã giải bài toán sau :

Một lô sản phẩm gồm N sản phẩm, trong đó có M sản phẩm tốt và N – M phế phẩm. Lấy ngẫu nhiên 1 lần n sản phẩm. Tìm xác suất để trong n sản phẩm lấy ra có đúng k sản phẩm tốt.

Giải :

Xác suất phải tìm là :

$$q(k, N, M, n) = \frac{C_M^k \times C_{N-M}^{n-k}}{C_N^n}$$

Định nghĩa 2.29. Biến ngẫu nhiên X có phân phối siêu bội với tham số N, M, n nếu phân phối xác suất của nó có dạng :

$$P[X = k] = \frac{C_M^k \times C_{N-M}^{n-k}}{C_N^n} = q(k, N, M, n) \; ; \; k = 0, ..., M$$

Kỳ vọng của X là : $E(X) = \dfrac{nM}{N}$

Phương sai của X là : $DX = n \times \dfrac{M}{N}\left(\dfrac{N-M}{N}\right)\left(1 - \dfrac{n-1}{N-1}\right)$

Định lý 2.3. Nếu n cố định, còn N tăng lên vô hạn và tỷ số $\dfrac{M}{N} \to p$; $0 < p < 1$ thì phân phối siêu bội với tham số N, M, n sẽ tiến tới phân phối nhị thức với tham số (n, p), nghĩa là :

$$q(k, N, M, n) \to P_n(k) = C_n^k p^k (1-p)^{n-k} \text{ khi } N \to \infty.$$

2.3.4. Phân phối chuẩn

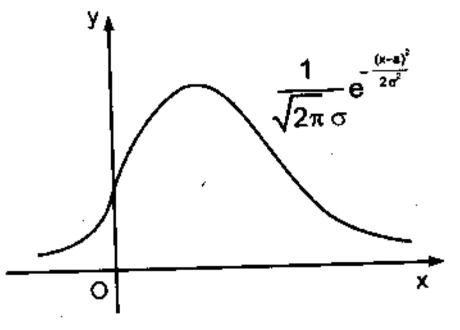
Định nghĩa 2.30. Biến ngẫu nhiên X có phân phối chuẩn dạng tổng quát $N(a, \sigma^2)$ nếu hàm mật độ của nó có dạng :

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

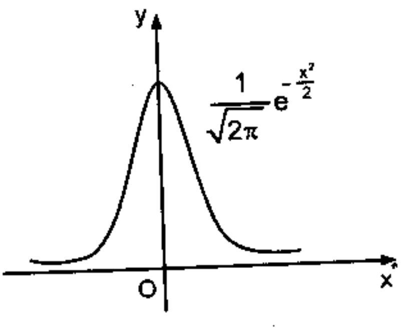
Đồ thị của nó ở hình (H.2.13).

Trường hợp a = 0, σ = 1 ta có : $f(x) = \dfrac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$

Đồ thị của nó ở hình (H.2.14).



Hình 2.13



Hình 2.14

Hàm phân phối chuẩn dạng N(0 ;1) :

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{u^2}{2}} du \text{ với } x \in R.$$

Từ tính chất của tích phân xác định ta suy ra $\Phi(-x) = 1 - \Phi(x)$.

Định lý 2.3. Nếu biến ngẫu nhiên X có phân phối chuẩn dạng $N(a, \sigma^2)$ thì :

– $Y = \dfrac{X - a}{\sigma}$ có phân phối chuẩn dạng N(0 ;1).

$$- P[\alpha \le X < \beta] = \Phi\left(\frac{\beta - a}{\sigma}\right) - \Phi\left(\frac{\alpha - a}{\sigma}\right) \tag{2.27}$$

Chứng minh tính chất này dành cho độc giả.

Ví dụ 2.37. Giả sử độ cao X của trẻ em có phân phối chuẩn dạng N(1,3 ; 0,01). Tính xác suất để trẻ em có độ cao nằm trong khoảng (1,2 ; 1,4).

Giải :

Theo công thức (2.27) ta có :

$$P[1,2 \le X < 1,4] = \Phi\left(\frac{1,4 - 1,3}{\sqrt{0,01}}\right) - \Phi\left(\frac{1,2 - 1,3}{\sqrt{0,01}}\right) = \Phi(1) - \Phi(-1)$$

$$= 2\Phi(1) - 1 = 0,6826$$

(vì tra bảng giá trị hàm phân phối chuẩn $\Phi(1) = 0,8413$)

2.3.5. Phân phối khi bình phương (χ^2)

Định nghĩa 2.31. Biến ngẫu nhiên X có phân phối χ^2 với n bậc tự do nếu hàm mật độ của nó có dạng :

$$f(x) = \begin{cases} 0 & \text{với } x \le 0 \\ \dfrac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} & \text{với } x > 0 \end{cases}$$

Tính chất của phân phối khi bình phương :

Nếu dãy các biến ngẫu nhiên độc lập $X_1, X_2, ..., X_n$ có phân phối chuẩn dạng N(0;1) thì $Z = \sum_{n=1}^{n} X_i^2$ có phân phối khi bình phương (χ^2) với n bậc tự do.

Bảng giá trị của Z có phân phối χ^2 được cho ở bảng 3 (phần Phụ lục).

2.3.6. Phân phối Student

Định nghĩa 2.32. Biến ngẫu nhiên X có phân phối Student với k bậc tự do nếu hàm mật độ của nó có dạng :

$$f(x) = \frac{1}{\sqrt{k}\beta\left(\frac{1}{2}, \frac{k}{2}\right)}\left(1 + \frac{x^2}{k}\right)^{-\frac{k+1}{2}}$$

trong đó $\beta(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}dx$, $a > 0$, $b > 0$, được gọi là hàm Bêta.

Tính chất đơn giản của phân phối Student :

Nếu hai biến ngẫu nhiên X và Y là độc lập, X có phân phối χ^2 với k bậc tự do và Y có phân phối chuẩn dạng N(0;1), thì $T = \frac{Y}{\sqrt{\frac{X}{k}}}$ có phân phối Student với k bậc tự do.

Giá trị của T có phân phối Student cho ở bảng 2 (phần Phụ lục).

2.3.7. Phân phối F (Fisher R.A – Snedecor G.W)

Định nghĩa 2.33. Biến ngẫu nhiên X có phân phối F với m, n bậc tự do nếu hàm mật độ của nó có dạng :

$$f(x) = \begin{cases} 0 & \text{với } x \le 0 \\ \dfrac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} m^{\frac{m}{2}} n^{\frac{n}{2}} \dfrac{x^{\frac{m}{2}-1}}{(n+mx)^{\frac{m+n}{2}}} & \text{với } x > 0 \end{cases}$$

Tính chất đơn giản của hàm phân phối F :

Nếu X, Y là hai biến ngẫu nhiên độc lập và có phân phối khi bình phương với số bậc tự do tương ứng là m và n, thì $F = \frac{X/m}{Y/n}$ có phân phối F với m, n bậc tự do.

Giá trị F có phân phối cho ở bảng 4 (phần Phụ lục).

Bài tập chương II

1. Một lô sản phẩm gồm 100 sản phẩm trong đó có 90 sản phẩm tốt và 10 phế phẩm. Chọn ngẫu nhiên 3 sản phẩm từ lô hàng (chọn 1 lần). Gọi X là số sản phẩm tốt trong 3 sản phẩm lấy ra. Tìm phân phối xác suất của X. Viết hàm phân phối của X. Tính kỳ vọng của X. Tính xác suất $P[X \geq 1]$.

2. Gieo 10 lần đồng tiền cân đối và đồng chất. Gọi X là số lần xuất hiện mặt sấp trong 10 lần gieo đó. Tìm phân phối xác suất của X. Viết hàm phân phối. Tính kỳ vọng và phương sai của X. Tính xác suất $P[X \geq 1]$; $P[0 \leq X \leq 8]$.

3. Bắn không hạn chế vào một bia. Bắn cho tới khi nào trúng đích thì dừng lại. Gọi X là số viên đạn cần bắn để lần đầu tiên trúng đích. Tìm phân phối xác suất của X ; viết hàm phân phối. Tính kỳ vọng của X.

Tính xác suất $P[X \geq 2]$; $P[X < 3]$.

4. Cho phân phối xác suất của biến ngẫu nhiên X là :

X	– 2	0	2
P[X = i]	$\frac{1}{6}$	$\frac{2}{3}$	$\frac{1}{6}$

Viết hàm phân phối. Tính xác suất $P[-1 \leq X < 1]$. Tính kỳ vọng và phương sai của X.

5. Gọi X là biến ngẫu nhiên bằng số kiểu gen aa thấy được trong 40 cá thể sinh ra từ sự giao hai dị hợp tử Aa (ta nhắc lại rằng trong sự giao đó, xác suất để thu được một cá thể có kiểu gen aa bằng 1/4). Tìm phân phối xác suất của biến ngẫu nhiên X. Tính xác suất để X có giá trị lớn nhất bằng 6.

Nếu Y là biến ngẫu nhiên bằng số kiểu gen aa thấy được trong số 400 cá thể. Tính giá trị của y sao cho biến cố $[Y \leq y]$ có cùng xác suất với biến cố $[X \leq 6]$.

6. Ta gọi X_1 và X_2 là những biến ngẫu nhiên bằng số kiểu gen aa thấy được trong hai nhóm 40 cá thể sinh ra từ sự giao của hai dị hợp tử Aa. Tìm phân phối xác suất của biến ngẫu nhiên $X_1 + X_2$. Tính xác suất của các biến cố :

$$[\omega : X_1 \leq 6 \text{ và } X_2 \leq 6] ; [\omega : X_1 + X_2 \leq 12].$$

7. Trong m phiếu của một cuộc bầu cử, số người bầu cho các ứng cử viên A, B, C lần lượt bằng m_1, m_2, m_3 ($m_1 + m_2 + m_3 = m$). Ta gọi X, Y, Z là các biến ngẫu nhiên bằng số phiếu bầu cho A, B và C khi kiểm tra n phiếu đầu tiên.

   a) Hãy xác định phân phối xác suất của biến ngẫu nhiên X và tính xác suất của biến cố $[X = r_1]$. Tìm phân phối xác suất của Y, của Z.

   b) Tìm phân phối đồng thời của X, Y, Z, nghĩa là tính xác suất $P[X = r_1 ; Y = r_2 ; Z = r_3]$, với bộ ba số nguyên có tổng bằng n.

8. Ta coi các sản phẩm do một máy chế tạo ra là kết quả của các phép thử độc lập và ta cho rằng mọi sản phẩm được chế tạo có xác suất bị hỏng là q (và như vậy xác suất không bị hỏng là $p = 1 - q$).

   Hãy xác định xác suất của biến cố A_r : = "r – 1 sản phẩm đầu không hỏng và sản phẩm thứ r bị hỏng". Tính xác suất của biến cố A bằng hợp của tất cả các biến cố A_r. Tính xác suất của biến cố đối lập $\overline{A}$ của A. Các kết quả đó có ý nghĩa như thế nào ?

   Hãy xác định xác suất của biến cố $B_{r,s}$: = biến cố "r – 1 sản phẩm đầu không hỏng, sản phẩm thứ r bị hỏng, s – 1 sản phẩm sau đó không hỏng và sản phẩm thứ r + s bị hỏng".

   Hãy chứng tỏ rằng : nếu X và X + Y là các số thứ tự của sản phẩm bị hỏng thứ nhất và sản phẩm bị hỏng thứ hai, thì các biến X và Y là độc lập và có cùng phân phối.

   Hãy xác định phân phối xác suất của biến ngẫu nhiên Z là chỉ số thứ tự của sản phẩm bị hỏng thứ k ; ở đây ta có nhận xét là biến cố Z = r có thể được xác định bởi : có k – 1 sản phẩm bị hỏng trong số r – 1 sản phẩm đầu và sản phẩm thứ r bị hỏng.

9. Ta giả thiết rằng : số N người con trong một gia đình (chọn một cách ngẫu nhiên) là biến ngẫu nhiên có phân phối Poisson với tham số λ (chẳng hạn $\lambda = 2{,}2$).

   Hãy tìm phân phối của biến ngẫu nhiên X (X là số con trai trong 1 gia đình chọn một cách ngẫu nhiên).

   Tính xác suất để N = n khi ta biết X = 3.

10. Cho hai biến ngẫu nhiên X, Y là độc lập và có phân phối xác suất tương ứng là :

X	−1	0	1	2
$P[X = x_i]$	0,2	0,3	0,3	0,2

Y	−1	0	1
$P[Y = y_i]$	0,3	0,4	0,3

Tìm phân phối xác suất của X^2, $X + Y$.

Tính kỳ vọng và phương sai của X^2, $X + Y$.

11. Cho hàm phân phối của biến ngẫu nhiên X là :

$$F(x) = \frac{1}{2} + \frac{1}{\pi}\operatorname{arctg} x$$

a) Tìm xác suất của biến cố $[\omega : 0 < X < 1]$.

b) Tìm hàm mật độ của X.

12. Cho hàm phân phối của biến ngẫu nhiên X

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{(u-a)^2}{2\sigma^2}}\,du,\ x \in R$$

a) Tìm hàm mật độ của X.

b) Tính xác suất $P[0 \le X < 0{,}92]$ khi $a = 0$, $\sigma = 1$.

13. Cho hàm phân phối của biến ngẫu nhiên X là :

$$F(x) = \begin{cases} 0 & \text{với} \quad x \le 0 \\ ax + b & \text{với} \quad 0 < x \le 1 \\ 1 & \text{với} \quad x > 1 \end{cases}$$

a) Tìm a, b để F(x) là hàm liên tục.

b) Tìm hàm mật độ của X và tính xác suất $P[0 < X < 1/2]$.

14. Cho hàm mật độ của biến ngẫu nhiên X là :

$$f(x) = \begin{cases} \frac{1}{\theta} e^{-\frac{x}{\theta}} & \text{với } x > 0,\ \theta > 0 \\ 0 & \text{với } x \le 0 \end{cases}$$

a) Tìm hàm phân phối của X và tính xác suất $P[0 \le X < \theta]$.

b) Tính kỳ vọng và phương sai của X.

15. Cho hàm mật độ của biến ngẫu nhiên X là :

$$f(x) = \begin{cases} Ax^n e^{-x} & \text{với } x > 0 \\ 0 & \text{với } x \le 0 \end{cases}$$

a) Tìm A. Viết hàm phân phối của X.

b) Tính xác suất $P[0 \le X < +\infty]$.

16. Cho hàm mật độ của biến ngẫu nhiên X là :

$$f(x) = \frac{a}{1 + x^2} \text{ với } -\infty < x < +\infty.$$

a) Tìm a, viết hàm phân phối của biến ngẫu nhiên X.

b) Tính xác suất $P[0 \le X < 1]$.

17. Cho hàm mật độ của biến ngẫu nhiên X là :

$$f(x) = \begin{cases} A|\sin x| & \text{với } x \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right) \\ 0 & \text{với } x \notin \left(-\frac{\pi}{2}, \frac{\pi}{2}\right) \end{cases}$$

Tìm A, viết hàm phân phối của biến ngẫu nhiên X.

18. Cho hàm mật độ của biến ngẫu nhiên X là :

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{với } x > 0, \lambda > 0 \\ 0 & \text{với } x \le 0 \end{cases}$$

Tìm hàm phân phối của X. Tính kỳ vọng và phương sai của X.

19. Giả sử hàm mật độ của biến ngẫu nhiên X có dạng :

$f(x) = Cx^2(1 - x)^3$ với $0 \le x \le 1$.

Tìm C, viết hàm phân phối của X.

20. Giả sử hàm mật độ của biến ngẫu nhiên X có dạng :

$$f(x) = \frac{4a}{e^x + e^{-x}}$$

a) Tìm a.

b) Tìm hàm phân phối của biến ngẫu nhiên X.

c) Tính xác suất $P[0 < X < 1]$.

21. Cho phân phối xác suất của biến ngẫu nhiên X là :

X	1	2	3	4	5	6
$P[X = x_i]$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

a) Tính kỳ vọng và phương sai của X.

b) Tính mômen bậc ba của X.

22. Một lô hàng gồm 12 sản phẩm trong đó có 10 sản phẩm tốt và 2 phế phẩm. Chọn ngẫu nhiên 2 sản phẩm từ lô hàng. Gọi X là số phế phẩm trong 2 sản phẩm chọn ra. Tìm phân phối xác suất của X. Tính kỳ vọng của X.

23. Giả sử N_t là số hạt phóng xạ trong khoảng thời gian t có phân phối Poisson với tham số $\lambda = 1/2$.

Tính kỳ vọng và phương sai của N_t.

24. Tìm phương sai của đại lượng ngẫu nhiên X chỉ số lần xuất hiện biến cố A trong hai phép thử độc lập, nếu như xác suất xuất hiện biến cố A trong mỗi phép thử là như nhau và biết rằng kỳ vọng của X bằng 1,2.

25. Đại lượng ngẫu nhiên rời rạc X chỉ nhận hai giá trị x_1 và x_2 ($x_2 > x_1$). Xác suất để X nhận giá trị x_1 là 0,6. Tìm phân phối xác suất của X và các giá trị x_1, x_2 mà X có thể nhận nếu như biết kỳ vọng EX = 1,4 và phương sai DX = 0,24.

26. Trong 1 lô hàng có 500 đơn vị hàng hoá. Tỷ lệ hàng kém phẩm chất là 5%. Lấy ngẫu nhiên liên tiếp (có hoàn lại) 50 đơn vị hàng hoá từ lô hàng. Gọi X là số hàng hoá kém phẩm chất trong 50 đơn vị hàng hoá chọn ra. Tìm kỳ vọng và phương sai của X.

27. Xác suất bắn trúng đích của một khẩu súng là p. Tiến hành bắn liên tiếp trong điều kiện không đổi cho đến khi có k phát trúng đích thì thôi bắn. Tính kỳ vọng của số lần bắn cần thiết.

28. Cho hàm mật độ của biến ngẫu nhiên X là :

$$f(x) = \begin{cases} \dfrac{1}{\pi\sqrt{a^2 - x^2}} & \text{với } x \in (-a; a) \\ 0 & \text{với } x \notin (-a; a) \end{cases}$$

Tìm kỳ vọng và phương sai của X.

29. Giả sử biến ngẫu nhiên X có hàm mật độ là :

$$f(x) = a\,e^{-|x|}$$

a) Tìm a.

b) Tìm hàm phân phối của biến ngẫu nhiên X.

c) Tính kỳ vọng và phương sai của X.

30. Cho phân phối đồng thời của 2 biến ngẫu nhiên X, Y là :

Y \ X	1	2
0,4	0,1	0,2
0,8	0,1	0,6

a) Tìm hệ số tương quan của X, Y.

b) Tính kỳ vọng điều kiện của X với điều kiện Y = 1.

31. Giả sử hàm mật độ đồng thời của X, Y là :

$$f(x, y) = \begin{cases} x + y & \text{với } 0 \le x \le 1 \text{ và } 0 \le y \le 1 \\ 0 & \text{trong trường hợp khác.} \end{cases}$$

a) Tìm hệ số tương quan của X, Y.

b) Tính kỳ vọng điều kiện của Y với điều kiện X = x.

32. Giả sử hai biến ngẫu nhiên X, Y là độc lập và có hàm mật độ tương ứng là :

$$f_X(x) = \begin{cases} 0 & \text{với } x \le 0 \\ 5e^{-5x} & \text{với } x > 0 \end{cases} \quad \text{và} \quad f_Y(y) = \begin{cases} 0 & \text{với } y \le 0 \\ 2e^{-2y} & \text{với } y > 0 \end{cases}$$

a) Tìm hàm mật độ đồng thời của X, Y.

b) Viết hàm phân phối đồng thời của X, Y.

c) Tính xác suất $P[\omega : 0 \le X \le 1 \,;\, 0 \le Y \le 2]$.

33. Giả sử hàm mật độ đồng thời của hai biến ngẫu nhiên X, Y là :

$$f(x, y) = \begin{cases} 4xye^{-x^2-y^2} & \text{với } x > 0 \,;\; y > 0 \\ 0 & \text{trong trường hợp khác.} \end{cases}$$

a) Tính kỳ vọng của X, của Y.

b) Tính phương sai của X, của Y.

c) X, Y có độc lập không ? Tại sao ?

Hướng dẫn và trả lời bài tập chương II

1. Phân phối xác suất của X là : $P[X = k] = \dfrac{C_{90}^{k} \times C_{10}^{3-k}}{C_{100}^{3}}$, k = 0, 1, 2, 3.

X	0	1	2	3
P	$\frac{12}{16170}$	$\frac{405}{16170}$	$\frac{4005}{16170}$	$\frac{11748}{16170}$

Hàm phân phối của X là :

$$F(x) = \begin{cases} 0 & \text{với } x \le 0 \\ \dfrac{12}{16170} & \text{với } 0 < x \le 1 \\ \dfrac{417}{16170} & \text{với } 1 < x \le 2 \\ \dfrac{4422}{16170} & \text{với } 2 < x \le 3 \\ 1 & \text{với } x > 3 \end{cases}$$

Kỳ vọng của X là : $E(X) = \dfrac{4851}{3916}$

Xác suất $P[X \ge 1] = 1 - P[x = 0] = 1 - \dfrac{12}{16170}$.

2. $P[X = k] = C_{10}^{k} \times \dfrac{1}{2^{10}}$, k = 0, 1, ..., 10.

$F(x) = \dfrac{1}{2^{10}} \sum\limits_{k<x} C_{10}^{k}$, $x \in R$.

$P[X \ge 1] = 1 - \dfrac{1}{2^{10}}$, $P[0 \le X \le 8] = 1 - \dfrac{11}{2^{10}}$.

3. $P[X = k] = q^{k-1}p$; p = 0,2 , q = 0,8.

Do đó $P[X = k] = (0,8)^{k-1}.0,2$ với k = 1, 2, ...

Hàm phân phối của X là : $F(x) = 0,2 \sum\limits_{k<x} (0,8)^{k-1}$, $x \in R$.

$P[X \ge 2] = 1 - P[X = 1] = 1 - 0,2 = 0,8$; $P[X < 3] = 0,36$.

4. $$F(x) = \begin{cases} 0 & \text{với } x \le -2 \\ \dfrac{1}{6} & \text{với } -2 < x \le 0 \\ \dfrac{5}{6} & \text{với } 0 < x \le 2 \\ 1 & \text{với } x > 2 \end{cases}$$

$$P[-1 \le X \le 1] = \frac{5}{6} - \frac{1}{6} = \frac{2}{3}$$

5. $P[X = k] = C_{40}^{k}\left(\dfrac{1}{4}\right)^{k}\left(\dfrac{3}{4}\right)^{40-k}$ với $k = 0, 1, 2, ..., 40.$

$$P[X \le 6] = P[X < 7] \approx \Phi\left(\frac{7 - \frac{1}{2} - 10}{\sqrt{40 \times \frac{1}{4} \times \frac{3}{4}}}\right) \approx \Phi(-1,277) =$$

$$= 1 - \Phi(1,277) \approx 0,101$$

$$P[Y \le y] = P[Y < y + 1] \approx \Phi\left(\frac{y + \frac{1}{2} - 100}{\sqrt{400 \times \frac{1}{4} \times \frac{3}{4}}}\right) \approx F(-1,277) \Rightarrow y = 88.$$

6. $P[X_1 + X_2 = k] = C_{80}^{k}\left(\dfrac{1}{4}\right)^{k}\left(\dfrac{3}{4}\right)^{80-k}, k = 0, 1, 2, ..., 80.$

$$P[X_1 \le 6, X_2 \le 6] = P[X_1 \le 6]P[X_2 \le 6] = 0,01.$$

$$P[X_1 + X_2 \le 12] = P[X_1 + X_2 < 13] \approx F\left(\frac{13 - \frac{1}{2} - 20}{\sqrt{80 \times \frac{1}{4} \times \frac{3}{4}}}\right) \approx 0,026.$$

7. a) $P[X = r_1] = \dfrac{C_{m_1}^{r_1} \times C_{m-m_1}^{n-r_1}}{C_m^n}.$

Phân phối xác suất của Y, Z cũng sẽ là phân phối siêu bội như X, nhưng chỉ khác nhau là thay m_1 bằng m_2 hoặc m_3 tương ứng.

b) $P[X = r_1, Y = r_2, Z = r_3] = \dfrac{C_{m_1}^{r_1} \times C_{m_2}^{r_2} \times C_{m_3}^{r_3}}{C_m^n}.$

8. $P(A_r) = p^{r-1} \times q$; $P(\bigcup_{r=1}^{\infty} A_r) = 1$; $P(\bar{A}) = 1 - 1 = 0$;

trong đó $A = \bigcup_{r=1}^{\infty} A_r$; $P(B_{rs}) = p^{r+s-2}q^2$;

X, Y là độc lập vì : $P[Y = s] = p^{s-1}q$; $P\left[\dfrac{Y = s}{X = r}\right] = p^{s-1}q$;

$$P[Z = r] = \frac{(r-1)!}{(k-1)!(r-k)!} p^{r-k}q^k$$

9. $P[X = r] = \sum_{n=0}^{\infty} P[N = n].P[X = r / N = n]$

P[N = n] có phân phối Poisson với tham số λ, còn :

$$P[X = r/N = n] = \begin{cases} 0 & \text{với } n < r \\ C_n^r p^r (1-p)^{n-r} & \text{với } n > r \end{cases}$$

$$P[X = r] = \sum e^{-\lambda} \frac{\lambda^n}{n!} \frac{n!}{r!(n-r)!} \left(\frac{1}{2}\right)^n = \frac{e^{-\lambda}}{r!} \sum_{n=r}^{\infty} \left(\frac{\lambda}{2}\right)^n \frac{1}{(n-r)!}$$

$$P[X = r] = \frac{e^{-\frac{\lambda}{2}}}{r!} \left(\frac{\lambda}{2}\right)^r.$$

$$P[X = n/X = 3] = \frac{e^{-\frac{\lambda}{2}}}{(n-3)!} \left(\frac{\lambda}{2}\right)^{n-3}.$$

10.

X^2	0	1	4
P_{X^2}	0,3	0,5	0,2

$X + Y$	-2	-1	0	1	2	3
P_{X+Y}	0,06	0,17	0,27	0,27	0,17	0,06

$E(X^2) = 1{,}3$; $E(X + Y) = EX + EY = 0{,}50.$

$D(X^2) = 2{,}01$; $D(X + Y) = DX + DY = 1{,}05 + 0{,}6 = 1{,}65$

11. a) $P[0 < X < 1] = \dfrac{\pi}{4}$; b) $f_X(x) = \dfrac{1}{\pi(1 + x^2)}$

12. a) $f(x) = \dfrac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}$; b) $P[0 \le X < 0{,}92] \approx \Phi_{(0,5)} - \Phi_{(0)} = 0{,}3212.$

13. a) $a = 1$; $b = 0$; b) $P[0 \le X < 1/2] = F(1/2) - F(0) = 1/2$.

14. a) $F(x) = \begin{cases} 0 & \text{với } x \le 0 \\ 1 - e^{-\frac{x}{\theta}} & \text{với } x > 0,\ \theta > 0. \end{cases}$

$P[0 \le X < \theta] = 1 - e^{-1}$.

b) $EX = \theta$; $DX = \theta^2$.

15. a) $A = \dfrac{1}{n!}$, $F(x) = \begin{cases} \dfrac{1}{n!}\displaystyle\int_0^x u^n e^{-u} du & \text{với } x > 0 \\ 0 & \text{với } x \le 0 \end{cases}$

b) $P[0 \le X < +\infty] = \dfrac{1}{n!}\displaystyle\int_0^{+\infty} u^n e^{-u} du = \dfrac{\Gamma(n+1)}{n!} = \dfrac{n!}{n!} = 1$

16. a) $a = \dfrac{1}{\pi}$; $F(x) = \dfrac{1}{2} + \dfrac{1}{\pi}\text{arctg}x$; b) $P[0 < X < 1] = 1/4$.

17. $A = \dfrac{1}{2}$; $F(x) = \begin{cases} 0 & \text{với } x \le -\dfrac{\pi}{2} \\ \dfrac{1}{2}\cos x & \text{với } -\dfrac{\pi}{2} < x \le 0 \\ 1 - \dfrac{1}{2}\cos x & \text{với } 0 < x \le \dfrac{\pi}{2} \\ 1 & \text{với } x > \dfrac{\pi}{2} \end{cases}$

18. a) $F(x) = \begin{cases} 0 & \text{với } x \le 0 \\ 1 - e^{-\lambda x} & \text{với } x > 0,\ \lambda > 0 \end{cases}$

b) $P[-1 < X < 1] = 1 - e^{-\lambda}$

19. $A = 60$; $F(x) = \begin{cases} 0 & \text{với } x \le 0 \\ 60\displaystyle\int_0^x u^2(1-u)^3 du & \text{với } 0 < x \le 1 \\ 1 & \text{với } x > 1 \end{cases}$

20. a) $a = \dfrac{1}{2\pi}$; b) $F(x) = \dfrac{2}{\pi}\text{arctg}e^x$;

c) $P[0 < X < 1] = \dfrac{2}{\pi}\text{arctg}e - \dfrac{1}{2}$

21. a) $E(X) = 3{,}5$; $DX = 2{,}916$; b) $E(X^3) = 73{,}5$.

22.

X	0	1	2
P_X	$\frac{45}{66}$	$\frac{20}{66}$	$\frac{1}{66}$

$$E(X) = 0 \times \frac{45}{66} + 1 \times \frac{20}{66} + 2 \times \frac{1}{66} = \frac{1}{3}$$

23.
$$P[N_t = k] = \frac{\left(\frac{1}{2}\right)^k e^{-\frac{1}{2}}}{k!} \text{ với } k = 0, 1, 2, \ldots$$

$$E(N_t) = \frac{1}{2} \text{ ; } D(N_t) = \frac{1}{2}$$

24. Ta thấy giá trị của X có thể xem là 0 và 1. Ta có 2 phép thử Bernoulli, $P(A) = p$ không đổi trong mọi phép thử. Vậy $E(X) = np = 2p = 1,2$. Ta suy ra $p = 0,6$. Do đó $q = 1 - p = 0,4$.

Hơn nữa $DX = np(1 - p) = 2 \times 0,6 \times 0,4 = 0,48$

25. Ta đưa đến hệ hai phương trình

$$0,6x_1 + 0,4x_2 = 1,4$$
$$0,6x_1^2 + 0,4x_2^2 = 2,2$$

Bằng cách sử dụng giả thiết $E(X) = 1,4$ và $DX = 0,24$. Giải hệ này ta được $x_1 = 1$ và $x_2 = 2$ và phân phối xác suất của X là :

X	x_1	x_2
p	0,6	0,4

26. Có thể xem rằng việc kiểm tra 50 đơn vị hàng hóa từ lô hàng là việc lấy từng đơn vị hàng hóa kiểm tra sau đó trả lại lô. Vậy ta có thể xem như thực hiện 50 phép thử Bernoulli. Xác suất để 1 đơn vị hàng hóa là phế phẩm bằng 0,05.

Theo công thức xác suất nhị thức ta có :

$$P[X = k] = C_{50}^k (0,05)^k (0,95)^{50-k} \qquad k = 0, \ldots, 50.$$

$$E(X) = np = 50 \times 0,05 = 2,5$$

27. $P[X = m] = C_{m-1}^{k-1}p^{k-1}q^{m-k}p = C_{m-1}^{k-1}p^{k}q^{m-k}$

Trong đó : X là số lần bắn cần thiết.

$P[X = m] = 0$ với $m < k$.

$$E(X) = \sum_{m=k}^{\infty} mC_{m-1}^{k-1}p^{k}q^{m-k} = \frac{kp^{k}}{(1-q)^{k+1}} = \frac{k}{p}$$

Hoặc có thể lý luận như sau :

Gọi X_1 là số viên đạn cần bắn để lần đầu tiên trúng đích : $E(X_1) = \frac{1}{p}$;

Gọi X_2 là số viên đạn cần bắn để lần đầu tiên có viên trúng đích sau viên mà lần thứ nhất trúng đích : $E(X_2) = \frac{1}{p}$, ...;

Gọi X_k là số viên đạn cần bắn để lần đầu tiên trúng đích sau k – 1 viên trúng đích trước đó : $E(X_k) = \frac{1}{p}$;

$X = X_1 + X_2 + ...X_k$. Vậy $E(X) = E(X_1) + ... + E(X_k) = \frac{k}{p}$

28. $E(X) = 0$; $DX = \frac{a^2}{2}$

29. a) $a = \frac{1}{2}$;

b) $F(x) = \frac{1}{2}\int_{-\infty}^{x} e^{-|u|}du = \begin{cases} \frac{1}{2}e^{x} & \text{với } x \le 0 \\ 1 - \frac{1}{2}e^{-x} & \text{với } x > 0 \end{cases}$

c) $E(X) = 0$, $DX = 2$

30. a) $\rho(x, y) = \frac{2}{9} \approx 0{,}2222$

b)

X	1	2
$P[X = x_i / Y = 1]$	$\frac{1}{2}$	$\frac{1}{2}$

$$E(X / Y = 1) = 1 \times \frac{1}{2} + 2 \times \frac{1}{2} = \frac{3}{2} = 1{,}5$$

31. $f_X(x) = \begin{cases} x + \frac{1}{2} & \text{với } 0 < x \le 1 \\ 0 & \text{với x khác.} \end{cases}$ $f_Y(y) = \begin{cases} y + \frac{1}{2} & \text{với } 0 < y \le 1 \\ 0 & \text{với y khác.} \end{cases}$

$EX = \frac{7}{12}$; $EY = \frac{7}{12}$; $DX = \frac{11}{144} = DY$; $E(XY) = \frac{1}{3}$; $\rho(X, Y) = \frac{-1}{11}$

$$f(y/x) = \frac{f(x,y)}{f_X(x)} = \begin{cases} \dfrac{x+y}{x+\frac{1}{2}} & \text{với } 0 < x \le 1,\ 0 < y \le 1 \\ 0 & \text{trong trường hợp khác.} \end{cases}$$

$$E(Y/X) = \int_{-\infty}^{+\infty} yf(y/x)dy = \frac{5}{3(2x+1)}$$

32. a) $f(x,y) = \begin{cases} 10e^{-(5x+2y)} & \text{với } x > 0, y > 0 \\ 0 & \text{trong trường hợp khác.} \end{cases}$

b) Hàm phân phối đồng thời của X, Y :

$$F(x, y) = \begin{cases} 1 - e^{-5x} - e^{-2y} + e^{-(5x+2y)} & \text{với } x > 0,\ y > 0 \\ 0 & \text{trong trường hợp khác.} \end{cases}$$

c) $P[\omega : 0 \le X \le 1; 0 \le Y \le 2] = \int_0^1\int_0^2 10e^{-(5x+2y)}dxdy =$

$$= 10\int_0^1 e^{-5x}\left(\int_0^2 e^{-2y}dy\right)dx = (1 - e^{-4})(1 - e^{-5})$$

33. a) $f_X(x) = 2xe^{-x^2}$ với $x > 0$ và $f_Y(y) = 2ye^{-y^2}$ với $y > 0$

$E(X) = \frac{\sqrt{\pi}}{2}$, tương tự $E(Y) = \frac{\sqrt{\pi}}{2}$

b) $DX = 1 - \frac{\pi}{4}$, tương tự $DY = 1 - \frac{\pi}{4}$

c) X, Y là độc lập. Vì $f(x, y) = f_X(x)f_Y(y)$

Chương III

LUẬT SỐ LỚN VÀ ĐỊNH LÝ GIỚI HẠN TRUNG TÂM

> ***Những nội dung chính trong chương :***
>
> ** Luật số lớn dạng Tsêbưsép và Khinchin.*
>
> ** Định lý giới hạn trung tâm của Moivie – Laplace và của Lindeberg.*

3.1. LUẬT SỐ LỚN DẠNG TSÊBƯSÉP VÀ KHINCHIN

Trước hết ta xét ví dụ gieo n đồng tiền cân đối và đồng chất. Gọi m_A là số lần xuất hiện mặt sấp trong n lần gieo đó. Tỷ số $\frac{m_A}{n}$ được gọi là tần xuất xuất hiện mặt sấp (biến cố A). Người ta thấy rằng nếu số lần gieo càng tăng thì $\frac{m_A}{n}$ càng gần tới $\frac{1}{2}$. Ta đi đến giải bài toán : với điều kiện nào thì $\frac{m_A}{n}$ hội tụ về xác suất của P(A). Bài toán này được giải nhờ các định lý thuộc loại Luật số lớn.

Định nghĩa 3.1. Dãy $(X_n,\ n \geq 1)$ được gọi là hội tụ theo xác suất về biến ngẫu nhiên X khi $n \to \infty$ nếu với $\varepsilon > 0$ cho trước tuỳ ý ta có :

$$\lim_{n\to\infty} P\left[\omega\ :\ |X_n - X| < \varepsilon\right] = 1$$

(hoặc tương đương $\lim_{n\to\infty} P\left[\omega : |X_n - X| \geq \varepsilon\right] = 0$ và ký hiệu $X_n \xrightarrow{P} X$ khi $n \to \infty$.

Định nghĩa 3.2. Dãy các biến ngẫu nhiên $X_1, X_2, ..., X_n, ...$ được gọi là tuân theo luật số lớn nếu với $\varepsilon > 0$ cho trước tùy ý ta có :

$$\lim_{n\to\infty} P\left[\omega : \left|\frac{X_1 + X_2 + ... + X_n}{n} - \frac{E(X_1) + E(X_2) + ... + E(X_n)}{n}\right| < \varepsilon\right] = 1 .$$

Bất đẳng thức Tsêbưsép.

Mệnh đề : Giả sử biến ngẫu nhiên X có kỳ vọng EX và phương sai DX. Khi đó với $\varepsilon > 0$ cho trước tuỳ ý ta có :

$$P\left[|X - EX| \geq \varepsilon\right] \leq \frac{1}{\varepsilon^2} DX \tag{3.1}$$

Chứng minh :

– Ta có :

$$DX = E(X - EX)^2 = E(X - EX)^2 I_{[|X-EX|\geq\varepsilon]}) + E(X - EX)^2 I_{[|X-EX|<\varepsilon]})$$

Bỏ số dương thứ hai ở vế phải ta có :

$$DX \geq E(X - EX)^2 I_{[|X-EX|\geq\varepsilon]} \geq \varepsilon^2 EI_{[|X-EX|\geq\varepsilon]} \geq \varepsilon^2 P\left[|X - EX| \geq \varepsilon\right]$$

Vậy $P\left[|X - EX| \geq \varepsilon\right] \leq \frac{1}{\varepsilon^2} DX$.

Một trong các áp dụng của bất đẳng thức này là tiêu chuẩn 3 xích ma.

Giả sử DX = σ. Chọn $\varepsilon = t\sigma$ ta có : $P\left[|X - EX| \geq t\sigma\right] \leq \frac{1}{t^2}$

(3.2)

Cho t = 1 ta có $P[|X - EX| \geq \sigma] \leq 1$

Cho t = 2 ta có $P[|X - EX| \geq 2\sigma] \leq \frac{1}{4}$

Cho t = 3 ta có $P[|X - EX| \geq 3\sigma] \leq \frac{1}{9}$

– Nếu đại lượng ngẫu nhiên X có phân phối chuẩn có kỳ vọng bằng a và phương sai bằng σ^2, thì ta cũng có tiêu chuẩn tương tự như trên.

Dựa trên bất đẳng thức $P[|X - EX| < \varepsilon] = 2\,\Phi\left(\frac{\varepsilon}{\sigma}\right)$

Trong đó $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{u^2}{2}} du$; $x \in R$.

Đặt $\varepsilon = t\sigma$ ta có : $P[|X - EX| < t\sigma] = 2\Phi(t)$.

(3.3)

Cho t = 1 ta có : $P[|X - EX| < \sigma] = 2\Phi(1) = 0{,}6827$.

Cho t = 2 ta có : $P[|X - EX| < 2\sigma] = 2\Phi(2) = 0{,}9545$

Cho t = 3 ta có : $P[|X - EX| < 3\sigma] = 2\Phi(3) = 0{,}9973$.

Định lý 3.1 (Định lý Tsêbưsép)

Nếu dãy biến ngẫu nhiên $X_1, X_2, ..., X_n$ độc lập và có phương sai bị chặn đều bởi cùng một hằng số C, nghĩa là :

$$DX_1 \le C, DX_2 \le C, ..., DX_n \le C,$$

thì dãy đó tuân theo luật số lớn, nghĩa là : với $\varepsilon > 0$ có :

$$\lim_{n\to\infty} P\left[\left|\frac{X_1 + X_2 + ... + X_n}{n} - \frac{1}{n}\sum_{i=1}^{n} EX_i\right| \ge \varepsilon\right] = 0$$

Chứng minh :

Đặt $Y_n = \dfrac{X_1 + X_2 + ... + X_n}{n}$. Ta có $EY_n = \dfrac{EX_1 + EX_2 + ... + EX_n}{n}$.

Áp dụng bất đẳng thức Tsêbưsép cho đại lượng Y_n ta có, với $\varepsilon > 0$ cho trước tuỳ ý :

$$P\left[|Y_n - EY_n| \ge \varepsilon\right] \le \frac{DY_n}{\varepsilon^2} = \frac{1}{\varepsilon^2 n^2}\sum_{i=1}^{n} DX_i \le \frac{nC}{n^2\varepsilon^2} = \frac{C}{n\varepsilon^2}$$

Vế phải $\dfrac{C}{\varepsilon^2 n} \to 0$ khi $n \to \infty$.

Vậy $\lim\limits_{n\to\infty} P[|Y_n - EY_n| \ge \varepsilon] = 0$. Từ đó rút ra kết luận của định lý :

$$\lim_{n\to\infty} P\left[\left|\frac{X_1 + X_2 + ... + X_n}{n} - \frac{1}{n}\sum_{i=1}^{n} EX_i\right| \ge \varepsilon\right] = 0$$

Hệ quả 3.1.

Nếu dãy biến ngẫu nhiên $X_1, X_2, ..., X_n$ độc lập, có $EX_i = a$, $i = \overline{1, n}$ và $DX_1 \le C, ..., DX_n \le C$ thì :

$$\frac{X_1 + X_2 + ... + X_n}{n} \xrightarrow{P} a \quad \text{khi } n \to \infty.$$

Hệ quả 3.2.

Giả sử k là số lần xuất hiện biến cố A trong dãy n phép thử Bernoulli và P(A) = p không đổi trong mọi phép thử, thì $\dfrac{k}{n} \xrightarrow{P} p$ khi $n \to \infty$.

Định lý 3.2 (Định lý Khinchin)

Nếu dãy biến ngẫu nhiên $X_1, X_2, ..., X_n$ độc lập, có phân phối như nhau và $EX_i = a$, $i = \overline{1, n}$ thì :

$$\frac{X_1 + X_2 + ... + X_n}{n} \xrightarrow{P} a \text{ khi } n \to \infty .$$

3.2. Định lý giới hạn trung tâm cổ điển dạng Moivie – Laplace, Lindeberg

Vào năm 1730 De Moivie A. đã tìm ra sự liên hệ giữa xác suất nhị thức và hàm mật độ chuẩn. Nội dung này được thể hiện dưới dạng định lý sau đây :

Định lý 3.3 (Định lý Moivie)

Nếu xác suất P(A) = p để biến cố A xuất hiện trong mỗi phép thử của dãy n phép thử Bernoulli là không đổi, $0 < p < 1$ thì :

$$P_n(k) - \frac{1}{\sqrt{npq}}\varphi(x) \xrightarrow{P} 0 \text{ khi } n \to \infty,$$

trong đó $P_n(k) = C_n^k p^k q^{n-k}$, $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, $x = \frac{k - np}{\sqrt{npq}}$.

* Với n khá lớn ta có công thức xấp xỉ : $P_n(k) \approx \frac{1}{\sqrt{npq}} . \varphi(x)$ (3.4)

Ví dụ 3.1. Bắn 400 phát đạn độc lập vào một mục tiêu. Xác suất trúng đích của mỗi viên đạn là p = 0,2. Tính gần đúng xác suất để trong 400 phát có đúng 80 lần bắn trúng đích.

Giải :

Áp dụng công thức xấp xỉ : $P_n(k) \approx \frac{1}{\sqrt{npq}} . \varphi(x)$

Ta có :

$$P_{400}(80) \approx \frac{1}{\sqrt{400.0,2.0,8}} . \frac{1}{\sqrt{2\pi}} e^{-\frac{(80-400\times0,2)^2}{2\times400\times0,2\times0,8}} \approx \frac{1}{8\sqrt{2\pi}} \approx 0,04986.$$

Mãi đến năm 1783 Laplace P.S mới mở rộng kết quả của Moivie và mang tên là Định lý giới hạn tích phân của Moivie – Laplace.

Định lý 3.4. (Định lý Moivie – Laplace)

Gọi k là số lần xuất hiện biến cố A trong dãy n phép thử Bernoulli, p là xác suất để biến cố A xuất hiện trong mỗi phép thử, $0 < p < 1$ và $q = 1 - p$, k_1 , $k_2 (k_1 < k_2)$ là hai số nguyên đã cho. Khi đó :

$$\lim_{n\to\infty} P\left[\frac{k - np}{\sqrt{np(1-p)}} < x\right] = \Phi(x), \text{ đều với mọi } x \in R.$$

Từ định lý này ta suy ra công thức xấp xỉ với n khá lớn :

$$P\left[k_1 \le k < k_2\right] \approx \Phi\left(\frac{k_2 - np}{\sqrt{npq}}\right) - \Phi\left(\frac{k_1 - np}{\sqrt{npq}}\right) \quad (3.5)$$

Trong đó $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{u^2}{2}} du$.

Ví dụ 3.2.

Gieo ngẫu nhiên 10 000 hạt đậu tương. Xác suất nảy mầm của mỗi hạt là 0,9. Tính gần đúng xác suất để số hạt nảy mầm không ít hơn 8000.

Giải :

Ta phải tính gần đúng xác suất $P[k \ge 8000]$.

Xác suất này bằng : $P[k \ge 8000] = 1 - P[k < 8000]$.

Theo công thức xấp xỉ ở trên ta có :

$$P[k < 8000] \approx \Phi\left(\frac{8000 - np}{\sqrt{npq}}\right) = \Phi\left(\frac{8000 - 10^4 \times 0,9}{\sqrt{10^4 \times 0,9 \times 0,1}}\right) \approx \Phi\left(\frac{-1000}{30}\right)$$

$$\approx \Phi(-33,3) = 1 - \Phi(33,3)$$

Ta biết rằng $\Phi(x) \approx 1$ với $x > 5$.

Vậy $P[k < 8000] \approx 1 - 1 = 0$.

Xác suất phải tìm là : $P[k \ge 8000] \approx 1$.

Chú ý : Để đạt độ chính xác cao hơn ta thay công thức (3.5) bởi công thức :

$$P\left[k_1 \le k < k_2\right] \approx \Phi\left(\frac{k_2 + \frac{1}{2} - np}{\sqrt{npq}}\right) - \Phi\left(\frac{k_1 - \frac{1}{2} - np}{\sqrt{npq}}\right) \quad (3.6)$$

Ví dụ 3.3.

Trong một lô hàng, tỷ lệ phế phẩm là 5%. Kiểm tra ngẫu nhiên liên tiếp (có hoàn lại) 100 sản phẩm từ lô hàng. Tính gần đúng xác suất để số sản phẩm tốt trong 100 sản phẩm kiểm tra nằm trong khoảng (90 ; 96).

Giải :

$$P[90 \le k < 96] \approx \Phi\left(\frac{96 + \frac{1}{2} - 100 \times 0{,}95}{\sqrt{100 \times 0{,}95 \times 0{,}05}}\right) - \Phi\left(\frac{90 - \frac{1}{2} - 100 \times 0{,}95}{\sqrt{100 \times 0{,}95 \times 0{,}05}}\right)$$

$$\approx \Phi(0{,}69) - \Phi(-2{,}53) = \Phi(0{,}69) + \Phi(2{,}53) - 1$$

$$\approx 0{,}7549 + 0{,}9943 - 1 = 0{,}7492.$$

Định lý giới hạn trung tâm cổ điển được tiếp tục mở rộng bởi Liapunốp A.M năm 1901, tiếp theo là Lindeberg năm 1922. Ta phát biểu nội dung định lý Lindeberg (không chứng minh) như sau :

Định lý 3.5.

Nếu dãy biến ngẫu nhiên $X_1, X_2, ..., X_n$ độc lập có kỳ vọng EX_i và phương sai DX_i, $i = \overline{1;n}$ và thỏa mãn điều kiện Lindeberg :

$$\lim_{n\to\infty} \frac{1}{B_n} \sum_{k=1}^{n} E\left(X_k - EX_k\right)^2 I_{[|X_k - EX_k| > \tau B_n]}(\omega) = 0,$$

trong đó $B_n^2 = \sum_{k=1}^{n} DX_k$, τ là hằng số dương, $I_A(x) = \begin{cases} 1 \text{ với } x \in A \\ 0 \text{ với } x \notin A \end{cases}$

thì tồn tại giới hạn :

$$\lim_{n\to\infty} P\left[\frac{\sum_{k=1}^{n}(X_k - EX_k)}{B_n} < x\right] = \Phi(x) \qquad (3.7)$$

đều với mọi $x \in R$.

Hệ quả của định lý này là định lý giới hạn trung tâm dạng Liapunốp và định lý giới hạn trung tâm của Moivie – Laplace.

Từ định lý của Moivie – Laplace ta suy ra công thức sau :

$$P\left[\left|\frac{k}{n} - p\right| < \varepsilon\right] = 2\Phi\left(\varepsilon \times \sqrt{\frac{n}{pq}}\right) - 1 \qquad (3.8)$$

Ví dụ 3.4. Cho $\varepsilon = 0{,}01$; $p = \frac{1}{2}$, $P[\left|\frac{k}{n} - p\right| < \varepsilon] = 0{,}95$. Tìm n.

Giải :

Theo (3.8) ta có : $2\Phi\left(\varepsilon \times \sqrt{\frac{n}{pq}}\right) - 1 = 0{,}95$ suy ra $\Phi\left(\varepsilon \times \sqrt{\frac{n}{pq}}\right) = 0{,}975$.

Tra bảng phân phối chuẩn vào biểu thức trên ta có $\Phi(1{,}96) = 0{,}975$. Vậy $\varepsilon \times \sqrt{\frac{n}{pq}} = 1{,}96$. Thay $p = \frac{1}{2} = q$ vào biểu thức trên ta tìm được $n = 9604$.

Ví dụ 3.5. Cho $\varepsilon = 0{,}01$, $n = 10000$, $p = \frac{1}{2}$. Tính xác suất $P[\left|\frac{k}{n} - p\right| < \varepsilon]$.

Giải :

Ta có : $\varepsilon \times \sqrt{\frac{n}{pq}} = 0{,}01\sqrt{\frac{10^4}{\frac{1}{2} \times \frac{1}{2}}} = 2$. Giá trị của hàm $\Phi(x)$ tại 2 là $\Phi(2) = 0{,}9772$.

Vậy $P[\left|\frac{k}{n} - p\right| < 0{,}01] = 2\Phi(2) - 1 = 0{,}9544$.

Bài tập chương III

1. Một lô bóng đèn điện tử gồm 10000 chiếc. Xác suất để mỗi bóng đèn bị hỏng là 0,005. Tìm xác suất để trong lô có 40 bóng đèn bị hỏng.
2. Xác suất sinh con trai bằng 0,51. Tìm xác suất sao cho trong 1000 lần sinh (mỗi lần sinh 1 con), số con trai ít hơn số con gái.
3. Gieo một con xúc xắc cân đối và đồng chất 12000 lần. Tìm xác suất để cho số lần xuất hiện mặt trên của con xúc xắc có 1 chấm nằm trong khoảng [1900, 2150).
4. Cho biến ngẫu nhiên X có hàm phân phối :

$$F(x) = \begin{cases} 0 & \text{với } x \le 0 \\ x & \text{với } 0 < x \le 1 \\ 1 & \text{với } x > 1 \end{cases}$$

Ước lượng xác suất sao cho trong 100 lần lặp lại độc lập phép đo đại lượng X có tối đa 70 lần nhận giá trị trong khoảng (0,2 ; 0,7).

5. Một căng tin phục vụ hai đợt ăn trưa liên tiếp cho 2n người, mỗi ngày 2n người đó chọn một trong hai đợt ăn và có xác suất bằng 1/2 để 1 người trong số đó chọn đợt thức ăn thứ nhất ; người ta cho rằng sự lựa chọn đó được quyết định vì nhiều lý do khác nhau và độc lập với nhau.

   Số chỗ ngồi s của căng tin phải bằng bao nhiêu để xác suất "không đủ chỗ cho những người đến ăn" nhỏ hơn 0,1. Hãy tính giá trị của s nếu 2n = 1000.

6. Ta cho rằng các sản phẩm do một nhà máy chế tạo có xác suất p bị hỏng ; nếu $p \leq 0{,}1$ thì chất lượng được coi là đảm bảo và nếu $p \geq 0{,}2$ thì số sản phẩm sản xuất ra được coi là kém phẩm chất.

   Để kiểm tra chất lượng của việc chế tạo, ta lấy một mẫu gồm 100 sản phẩm và xác định số n_1 các sản phẩm hỏng của mẫu đó.

   – Nếu $n_1 \geq 16$ thì sản phẩm sản xuất ra là kém phẩm chất.

   – Nếu $n_1 \leq 7$ thì sản phẩm được công nhận là tốt.

   – Nếu $7 < n_1 < 16$ thì người ta xác định số n_2 các sản phẩm hỏng trong mẫu thứ hai gồm 100 sản phẩm và sản phẩm được coi là kém phẩm chất hay tốt tùy theo $n_2 \geq 16$ hay $n_2 < 16$. Tìm xác suất P_1 và P_2 của sự sai lầm, nghĩa là xác suất P_1 khi p = 0,1 mà coi sản phẩm thuộc loại kém phẩm chất và xác suất P_2 khi p = 0,2 mà công nhận sản phẩm là tốt.

7. Gieo 3200 hạt lúa. Gọi X là số hạt nảy mầm trong 3200 hạt. Tìm xác suất sao cho giá trị của X nằm trong khoảng $(1600 + 5\sqrt{2}\ ;\ 1600 + 10\sqrt{2})$. Biết rằng xác suất nảy mầm của mỗi hạt là p = 0,5.

8. Dùng bất đẳng thức Tsêbưsép để ước lượng xác suất của biến cố $[\omega : |X - EX| < 0{,}2]$ nếu DX = 0,004.

9. Tiến hành 10 phép thử độc lập. Biến cố A xuất hiện trong thời gian (0 ; T) bằng 0,05. Nhờ bất đẳng thức Tsêbưsép, ước lượng xác suất sao cho giá trị tuyệt đối của hiệu giữa số lần xuất hiện biến cố A và kỳ vọng của nó :

   a) Nhỏ hơn 2 ; b) Không nhỏ hơn 2.

10. Các dãy biến ngẫu nhiên độc lập sau đây có tuân theo luật số lớn không, nếu phân phối xác suất tương ứng là :

a) $P[X_k = \pm 2^k] = \frac{1}{2}$.

b) $P[X_k = \pm 2^k] = 2^{-(2k+1)}$, $P[X_k = 0] = 1 - 2^{-2k}$, với $k = 1,2, \ldots, n$

Hướng dẫn và trả lời bài tập chương III

1. $P_{10^4}(40) \approx \frac{1}{\sqrt{npq}} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} = 0,0206$, trong đó $x = \frac{k - np}{\sqrt{npq}}$.

2. $P[X<500] \approx \Phi\left(\frac{500 - 510}{\sqrt{1000 \times 0,49 \times 0,51}}\right) \approx 0,2643$.

3. $$P[1900 < X < 2150] \approx \Phi\left(\frac{2150 - 12000/6}{\sqrt{12000 \times \frac{5}{6} \times \frac{1}{6}}}\right) - \Phi\left(\frac{1900 - 12000/6}{\sqrt{12000 \times \frac{5}{6} \times \frac{1}{6}}}\right)$$

$$\approx \Phi\left(\frac{3\sqrt{3}}{\sqrt{2}}\right) - \Phi(-\sqrt{6}) \approx 0,99$$

4. $P[0,2 \le X \le 0,7] = F(0,7) - F(0,2) = 1/2$.

$$P[k \le 70] = P[k < 71] \approx \Phi\left(\frac{71 - 100 \times \frac{1}{2}}{\sqrt{100 \times \frac{1}{2} \times \frac{1}{2}}}\right) \approx \Phi\left(\frac{21}{5}\right) \approx 0,999968.$$

5. X là số người ăn đợt I. X có phân phối nhị thức với tham số (2n ; p = 1/2). Số s các chỗ sẽ thiếu nếu $X > s$ hoặc $2n - X > s \Leftrightarrow X < 2n - s$ và cả khi $|X - n| > s - n$.

Tính s sao cho : $P[|X - s| > s - n] \approx 2\left(1 - \Phi\left(\frac{s - n}{\sqrt{2n \times \frac{1}{2} \times \frac{1}{2}}}\right)\right) \le 0,01$

Suy ra $s - n \ge 2,58\sqrt{\frac{n}{2}}$.

Đặc biệt nếu 2n = 1000, ta suy ra $s \geq 500 + 2{,}58\sqrt{250} = 540{,}8$.

Ta lấy s = 541.

6. Gọi A_1, A_2, A_3 và B_1, B_2 là các biến cố : $n_1 \geq 16$; $n_1 \leq 7$; $7 < n_1 < 16$; $n_2 \geq 16$; $n_2 < 16$.

– Việc đưa sản phẩm vào loại kém là biến cố $A_1 \cup (A_3B_1)$. Nếu p = 0,1 thì :

$$P(A_1) = P[\,X > 15] \approx \Phi\left(\frac{15 + \frac{1}{2} - 10}{\sqrt{100 \times 0{,}1 \times 0{,}9}}\right) \approx 0{,}031.$$

$$P(A_3) \;= P[8 \leq X \leq 15] \approx \Phi\left(\frac{15+\frac{1}{2}-10}{\sqrt{100\times0{,}1\times0{,}9}}\right) - \Phi\left(\frac{8-\frac{1}{2}-10}{\sqrt{100\times0{,}1\times0{,}9}}\right)$$

$$\approx 0{,}766.$$

$P(B_1) = P[X > 15] \approx 0{,}031.$

Vậy xác suất để đưa sản phẩm vào loại kém là :

$$P_1 \approx 0{,}031(1 + 0{,}766) = 0{,}055$$

– Sự kiện đưa sản phẩm vào loại tốt là $A_2 \cup (A_3B_2)$.

Tương tự $P(A_2) = P[X \leq 7] \approx \Phi\left(\dfrac{7 + \frac{1}{2} - 20}{\sqrt{100 \times 0{,}2 \times 0{,}8}}\right) \approx 0{,}001.$

$$P(A_3) \;= P[8 \leq X \leq 15] \approx \Phi\left(\frac{15+\frac{1}{2}-20}{\sqrt{100\times0{,}2\times0{,}8}}\right) - \Phi\left(\frac{8-\frac{1}{2}-20}{\sqrt{100\times0{,}2\times0{,}8}}\right)$$

$$\approx 0{,}190.$$

$P(B_2) = P[X > 15] \approx 0{,}191.$

Vậy xác suất để công nhận sản phẩm là sản phẩm tốt là :

$$P_2 = 0{,}001 + 0{,}19 \times 0{,}191 = 0{,}037.$$

7. $P[1600 + 5\sqrt{2} < X < 1600 + 10\sqrt{2}\,] = P\left[\dfrac{5\sqrt{2}}{\sqrt{800}} < \dfrac{X - 1600}{\sqrt{800}} < \dfrac{10\sqrt{2}}{\sqrt{800}}\right]$

$$\approx \Phi\left(\frac{1}{2}\right) - \Phi\left(\frac{1}{4}\right) = 0{,}0928\,.$$

8. $P[|X - EX| < 0{,}2] \geq 0{,}9$.

9. a) $P[|X - 0{,}5| < 2] \geq 0{,}88$; b) $P[|X - 0{,}5| \geq 2] \geq 0{,}12$.

10. a) Không tuân theo luật số lớn.

b) Tuân theo luật số lớn.

Hướng dẫn :

a) $EX_k = 0$. Ta chứng minh $P\left(\left|\frac{1}{n}\sum_{k=1}^{n} X_k\right| > \varepsilon\right)$ không tiến tới 0, nghĩa là dãy $X_1, ..., X_n, ...$ không tuân theo luật số lớn.

Đặt $\eta = \frac{1}{n}\sum_{k=1}^{n} X_k$.

Ta có $B = \{\omega : |\eta| \geq \varepsilon ;\ X_n = 2^n,\ X_{n-1} = 2^{n-1}\} \subset \{\omega : |\eta| \geq \varepsilon\} = A$.

Nếu chọn n sao cho $\frac{2^{n-1}}{n} > \varepsilon$ thì $P[B] = \frac{1}{4}$ vì khi $X_n = 2^n, X_{n-1} = 2^{n-1}$, ta có $|\eta| \geq \varepsilon$.

Vậy $P(A) > \frac{1}{4}$.

Chứng tỏ $\frac{1}{n}\sum_{k=1}^{n} X_k$ không tiến tới 0. Vậy $X_1, ..., X_n, ...$ không tuân theo luật số lớn.

b) $E(X_k) = 0$; $DX_k = 1, k = 1, 2, ...$

$$P\left\{\left|\frac{1}{n}\sum_{k=1}^{n} X_k\right| \geq \varepsilon\right\} \leq \frac{1}{n^2\varepsilon^2}\sum_{k=1}^{n} DX_k = \frac{1}{n\varepsilon^2} \to 0 \text{ khi } n \to \infty.$$

Vậy $X_1, ..., X_n, ...$ tuân theo luật số lớn.

Phần hai

THỐNG KÊ TOÁN HỌC

Chương IV

MẪU NGẪU NHIÊN, HÀM PHÂN PHỐI MẪU VÀ CÁC SỐ ĐẶC TRƯNG MẪU

Những nội dung chính trong chương :

** Khái niệm mẫu ngẫu nhiên.*

** Hàm phân phối mẫu.*

** Các số đặc trưng mẫu.*

4.1. MẪU NGẪU NHIÊN

4.1.1. Khái niệm về mẫu ngẫu nhiên

Giả sử ta cần nghiên cứu một tính chất nào đó của các cá thể trong một đám đông Q. Trên thực tế số phần tử của đám đông rất lớn hoặc vì một số khó khăn nào đó mà ta không thể khảo sát được tất cả các phần tử của nó, nhưng lại muốn có một kết luận đủ chính xác về tính chất của các cá thể trong đám đông đó. Để giải quyết vấn đề này ta phải chọn ra một tập hợp các phần tử đại diện cho đám đông đó. Tập các phần tử đại diện này được gọi là tập mẫu.

Để đi đến một định nghĩa chính xác về mẫu ngẫu nhiên, ta xét một ví dụ sau : Để đo một đại lượng chưa biết θ nào đó, ta tiến hành n thí nghiệm. Kết quả của n thí nghiệm này được đặc trưng bởi dãy n biến ngẫu nhiên $X_1, X_2,..., X_n$ mà phân phối đồng thời là $F(x_1, x_2, ..., x_n, \theta)$ phụ thuộc vào tham số θ. $(X_1, X_2,..., X_n)$ được gọi là một mẫu ngẫu nhiên từ phân phối $F(x, \theta)$. (Đôi khi người ta còn gọi là mẫu quan sát $(X_1,..., X_n)$).

Định nghĩa 4.1. Mẫu ngẫu nhiên là 1 dãy n biến ngẫu nhiên $(X_1,..., X_n)$ từ phân phối $F(x, \theta)$, n được gọi là kích thước mẫu. Các giá trị của mẫu được ký hiệu bằng chữ thường $x_1, ..., x_n$. Về mặt hình học, một mẫu

$(X_1,..., X_n)$ được xem như 1 điểm trong không gian R^n chiều. Không gian R^n được gọi là không gian mẫu.

Chú ý : Thông thường người ta hay xét mẫu ngẫu nhiên $(X_1,..., X_n)$ là dãy các biến ngẫu nhiên độc lập có cùng phân phối.

4.1.2. Phương pháp chọn mẫu

Tùy theo phương pháp thiết lập khác nhau mà ta được các loại mẫu khác nhau. Ta có thể đưa ra 5 loại mẫu thông dụng sau, còn các phương pháp khác, khi cần độc giả có thể tìm tài liệu viết riêng về các phương pháp đó.

a) Mẫu ngẫu nhiên hoàn lại. Từ tập hợp tổng quát (tập gồm các đối tượng cùng loại) gồm N phần tử ta chọn ngẫu nhiên 1 phần tử, khảo sát và ghi lại kết quả X_1 ; Sau đó trả phần tử đó vào tập tổng quát, rồi chọn ngẫu nhiên phần tử thứ hai khảo sát và ghi lại kết quả X_2. Sau đó ta trả lại phần tử này vào tập tổng quát và tiếp tục chọn ngẫu nhiên phần tử thứ ba v.v.. Lặp lại như thế đến n lần. Ta nhận được dãy kết quả $(X_1,..., X_n)$. Mẫu ngẫu nhiên này được gọi là mẫu ngẫu nhiên hoàn lại. Các X_i được chọn với xác suất như nhau và bằng $\frac{1}{N}$.

b) Mẫu ngẫu nhiên không hoàn lại. Từ tập tổng quát gồm N phần tử, ta chọn ngẫu nhiên 1 phần tử, khảo sát và ghi lại kết quả X_1. X_1 nhận được với xác suất $\frac{1}{N}$. Bỏ phần tử ấy ra ngoài, ta lại chọn ngẫu nhiên phần tử thứ hai từ tập còn lại, khảo sát và ghi kết quả X_2 ; X_2 nhận được với xác suất $\frac{1}{N-1}$. Sau đó ta bỏ phần tử đó ra ngoài và tiếp tục chọn ngẫu nhiên phần tử thứ 3 từ tập còn lại và cứ tiếp tục như thế đến lần thứ n. Ta nhận được mẫu $(X_1, X_2,..., X_n)$ với xác suất tương ứng $\frac{1}{N}, \frac{1}{N-1}, ..., \frac{1}{N-n+1}$. Mẫu ngẫu nhiên này được gọi là mẫu ngẫu nhiên không hoàn lại. Hai loại mẫu a) và mẫu b) được mang một tên chung là *mẫu ngẫu nhiên đơn giản.*

c) Mẫu được chọn theo phương pháp cơ học. Trường hợp tập tổng quát có vô hạn phần tử làm theo cách a) và b) rất khó khăn, ta phải tiến hành bằng phương pháp sau : Dùng máy để phân chia ngẫu nhiên tập tổng quát thành các tập nhỏ, rồi từ những tập nhỏ đó ta chọn ra một số phần tử đại diện rồi hợp nhất lại ta được một mẫu chung. Ngoài ra ta có thể dùng bảng số ngẫu nhiên để chọn ; tức là ta có thể đánh số thứ tự tất cả các phần tử của tập tổng quát (thường các số này đã được ghi từ trước). Ta chọn ngẫu nhiên 1 số trong bảng số ngẫu nhiên, sau đó chọn những phần tử nào của tập tổng quát có số thứ tự trùng với những số ta vừa lấy ra để khảo sát. Phương pháp chọn mẫu này được gọi là phương pháp cơ học.

d) Mẫu "điển hình". Mẫu "điển hình" là mẫu mà phần tử của nó không chọn từ toàn bộ tập tổng quát mà từ bộ phận "điển hình" của nó. Tất nhiên từ bộ phận "điển hình" ta chọn mẫu theo cách a) hoặc cách b).

e) Phương pháp phân dãy (series) chia tập tổng quát thành nhiều dãy, sau đó từ mỗi dãy lại chọn ra một mẫu con theo cách a) hoặc b). Hợp nhất các mẫu còn lại ta được mẫu chung.

4.1.3. Sắp xếp số liệu thực nghiệm

Từ mẫu ngẫu nhiên $(X_1, X_2, ..., X_n)$ thường có hai cách sắp xếp tiện lợi cho việc áp dụng các tiêu chuẩn thống kê.

a) Sắp xếp theo các giá trị khác nhau. Giả sử mẫu $(X_1, X_2, ..., X_n)$ có k quan sát khác nhau là $X_1, X_2, ..., X_k$, $k \le n$ và X_1 có tần số n_1, X_2 có tần số n_2, ..., X_k có tần số n_k ; $n_1 + n_2 + ... + n_k = n$.

Ví dụ 4.1. Kiểm tra ngẫu nhiên 50 học sinh bằng 1 bài thi viết. Kết quả cho ở bảng sau :

X	2	4	5	6	7	8	9	10
n_i	4	6	20	10	5	2	2	1

b) Sắp xếp dưới dạng khoảng. Nếu mẫu ngẫu nhiên $(X_1, X_2, ..., X_n)$ có nhiều quan sát khác nhau, khoảng cách giữa các quan sát không như nhau hoặc các X_i, $i = \overline{1, n}$ khác nhau rất ít, ta không thể sắp xếp theo phần a) mà sẽ sắp xếp chúng dưới dạng khoảng thì việc xử lý số liệu thống kê sẽ thuận tiện hơn.

Ta tiến hành như sau : Khoảng (x_{min}, x_{max}) chứa toàn bộ các quan sát $X_1, X_2, ..., X_n$. Bây giờ ta chia khoảng (x_{min}, x_{max}) thành nhiều khoảng nhỏ, độ dài mỗi khoảng không nhất thiết phải bằng nhau. Song để tiện tính toán ta chia thành các khoảng có độ dài như nhau. Người ta chứng minh được rằng : số khoảng được chọn tối ưu (đảm bảo độ chính xác) được tính theo công thức :

$$\text{Số khoảng} = 1 + 3,322 \lg n$$

và độ dài khoảng là : $h = \dfrac{x_{max} - x_{min}}{1 + 3,322 \lg n}$

Thường chọn đầu mút trái của khoảng đầu tiên là $x_{min} - \dfrac{h}{2} = a_1$, đầu mút kia $a_2 = a_1 + h$, $a_3 = a_2 + h$, ... tiếp tục làm cho tới lúc mút đầu của khoảng cuối cùng không bằng hoặc lớn hơn x_{max}.

Ví dụ 4.2. Cho mẫu quan sát dưới dạng khoảng (bảng 4.1) :

Bảng 4.1

Năng suất hàng năm tính r% ở dạng khoảng	Số lượng công nhân n_i	Tần suất $\frac{n_i}{n}$
80 – 90	8	8/117
90 – 100	15	15/117
100 – 110	46	46/117
110 – 120	29	29/117
120 – 130	13	13/117
130 – 140	3	3/117
140 – 150	3	3/117

4.2. HÀM PHÂN PHỐI MẪU, ĐA GIÁC TẦN SUẤT VÀ TỔ CHỨC ĐỒ TẦN SUẤT

4.2.1. Hàm phân phối mẫu. Cho mẫu ngẫu nhiên $(X_1, X_2, ..., X_n)$ từ phân phối F(x) (hoặc F(x, θ)

Định nghĩa 4.2. Hàm phân phối mẫu (hay hàm phân phối thực nghiệm) là tỷ số $\frac{m}{n}$, trong đó n là kích thước mẫu, m là số giá trị mẫu $X_i < x$; $x \in R$; và ký hiệu :

$$F_n(x) = \frac{m}{n}, x \in R.$$

Ví dụ 4.3. Kiểm tra ngẫu nhiên 20 học sinh trong một lớp bằng 1 bài toán. Kết quả cho ở bảng sau :

X	2	3	4	5	6	8
n_i	1	2	2	6	7	2

Hàm phân phối mẫu là :

$$F_n(x) = \frac{m}{n} = \begin{cases} 0 & \text{với } x \le 2 \\ 1/20 & \text{với } 2 < x \le 3 \\ 3/20 & \text{với } 3 < x \le 4 \\ 5/20 & \text{với } 4 < x \le 5 \\ 11/20 & \text{với } 5 < x \le 6 \\ 18/20 & \text{với } 6 < x \le 8 \\ 1 & \text{với } x > 8 \end{cases}$$

Tính chất của hàm phân phối mẫu :

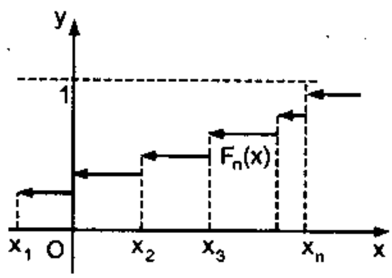
+ $0 \le F_n(x) \le 1$ vì $0 \le m \le n$.

+ $F_n(x)$ là hàm đơn điệu tăng theo x.

+ $F_n(x) = 0$ nếu $x \le \min(X_1,..., X_n)$.

$F_n(x) = 1$ nếu $x > \max(X_1,..., X_n)$.

+ $F_n(x) \to F(x)$ khi $n \to \infty$ theo nghĩa xác suất.



Hình 4.1

Hình ảnh thống kê của hàm phân phối mẫu (Hình 4.1).

4.2.2. Đa giác tần suất và tổ chức đồ tần suất

a) Đa giác tần suất

Giả sử X_1 có tần số n_1

X_2 có tần số n_2

....

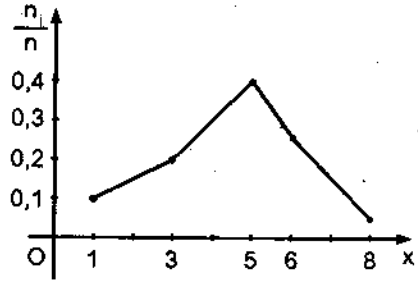
X_k có tần số n_k

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

Định nghĩa 4.3. Đường nối các điểm $\left(X_1, \frac{n_1}{n}\right), \left(X_2, \frac{n_2}{n}\right), ..., \left(X_k, \frac{n_k}{n}\right)$ được gọi là đa giác tần suất.

Ví dụ 4.4. Kiểm tra ngẫu nhiên 20 học sinh bằng một bài tập làm văn. Kết quả được cho ở bảng sau :

X_i	1	3	5	6	8
n_i	2	4	8	5	1
$\frac{n_i}{n}$	0,1	0,2	0,4	0,35	0,05



Hình 4.2

Đa giác tần suất mô tả đồ thị của mật độ thực nghiệm (số liệu cho dưới dạng giá trị khác nhau).

b) Tổ chức đồ tần suất. Dạng này cũng mô tả đồ thị mật độ phân bố thực nghiệm trên cơ sở mẫu quan sát cho dưới dạng khoảng. Tổ chức đồ tần suất là một hình bậc thang gồm nhiều hình chữ nhật có đáy trùng với trục hoành. Độ dài cạnh đáy hình chữ nhật thứ i là độ dài khoảng thứ i, còn chiều rộng vuông góc với trục hoành với độ lớn là $\frac{n_i}{n_h}$, trong đó h là độ dài 1 khoảng thứ i.

Diện tích hình chữ nhật thứ i là $\frac{n_i}{n_h} \times h = \frac{n_i}{n}$

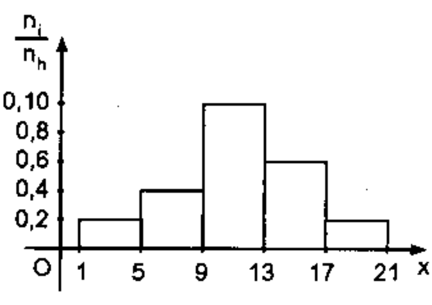
Diện tích cả hình bậc thang

$$h\frac{n_1}{n_h} + h\frac{n_2}{n_h} + ... + h\frac{n_k}{n_h} = 1$$

Ví dụ 4.5.

Bảng 4.2

Số TT	Khoảng h	Tần số n_i	$\frac{n_i}{n_h}$
1	1 – 5	40	0,02
2	5 – 9	80	0,04
3	9 – 13	200	0,10
4	13 – 17	120	0,06
5	17 – 21	40	0,02



Hình 4.3

4.3. CÁC SỐ ĐẶC TRƯNG MẪU

a) Trung bình mẫu

Định nghĩa 4.4. gọi $\overline{X} = \dfrac{X_1 + X_2 + ... + X_n}{n}$ là trung bình mẫu nếu mẫu ngẫu nhiên cho dưới dạng :

X	X_1	X_2	...	X_k
n_i	n_1	n_2	...	n_k

Thì trung bình mẫu được tính theo công thức :

$$\overline{X} = \frac{n_1X_1 + n_2X_2 + ... + n_kX_k}{n_1 + n_2 + ... + n_k}$$

Nếu mẫu ngẫu nhiên cho dưới dạng khoảng thì trung bình mẫu được tính theo công thức sau :

$$\overline{X} = \frac{n_1X_1^* + ... + n_kX_k^*}{n_1 + n_2 + ... + n_k}$$

trong đó $X_i^* = \dfrac{X_i + X_{i+1}}{n}$

X_i là mút trái của khoảng thứ i ;

X_{i+1} là mút phải của khoảng thứ i.

Số TT	Khoảng h	Tần số n_i
1	$x_1 - x_2$	n_1
2	$x_2 - x_3$	n_2
...	...	...
k	$x_k - x_{k+1}$	n_k

b) Phương sai mẫu : Có hai công thức để tính phương sai mẫu :

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

và $S_n^{*2}(X) = \dfrac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$

Ví dụ 4.6. Cho mẫu ngẫu nhiên :

X_i	1	2	3	4
n_i	20	15	10	5

Tìm $\overline{X}$, $S_n^2(X)$, $S_n^{*2}(X)$.

Giải :

$$\overline{X} = \frac{1\times 20 + 2\times 15 + 3\times 10 + 4\times 5}{20+15+10+5} = \frac{100}{50} = 2$$

$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 = \frac{20\times 1^2 + 15\times 2^2 + 10\times 3^2 + 5\times 4^2}{50} = \frac{250}{50} = 5$$

$$S_n^2(X) = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - (\overline{X})^2 = 5 - 2^2 = 1$$

$$S_n^{*2}(X) = \frac{n}{n-1} S_n^2(X) = \frac{50}{49}\times 1 = \frac{50}{49}$$

Ví dụ 4.7. Số liệu được cho dưới dạng trong ví dụ 4.5. Tính $\overline{X}$ và $S_n^2(X)$.

Giải :

$x_1^* = \dfrac{1+5}{2} = 3$; $x_2^* = \dfrac{5+9}{2} = 7$; $x_3^* = \dfrac{9+13}{2} = 11$;

$x_4^* = \dfrac{13+17}{2} = 15$; $x_5^* = \dfrac{17+21}{2} = 19$.

Ta có mẫu :

X	$x_1^* = 3$	$x_2^* = 7$	$x_3^* = 4$	$x_4^* = 15$	$x_5^* = 19$
n_i	40	80	200	120	40

$$\overline{X} = \frac{40\times 3 + 80\times 7 + 200\times 11 + 120\times 15 + 40\times 19}{40+80+200+120+40} = \frac{5440}{480}$$

$$\overline{X} = \frac{34}{3} \approx 11,33$$

$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 = \frac{40\times 3^2 + 80\times 7^2 + 200\times 11^2 + 120\times 15^2 + 40\times 19^2}{480}$$

$$= \frac{70020}{480} = \frac{1167}{8} \approx 145{,}8$$

$$S_n^2(X) = \frac{1167}{8} - \left(\frac{34}{3}\right)^2 = \frac{255}{72}$$

Chú ý 4.1 : Nếu mẫu ngẫu nhiên cho dưới dạng

$$\begin{array}{cccc}
A_1 & A_2 & \cdots & A_v \\
X_{11} & X_{12} & \cdots & X_{1v} \\
X_{21} & X_{22} & \cdots & X_{2v} \\
\vdots & \vdots & & \vdots \\
X_{n_1 1} & X_{n_2 2} & \cdots & X_{n_v v}
\end{array}$$

$n = n_1 + n_2 + \ldots + n_v$

Mẫu quan sát được chia thành v nhóm. Ta phải tính các trung bình và phương sai mẫu sau đây :

– Trung bình trong nhóm thứ j :

$$\overline{X}_j = \frac{1}{n_j}\sum_{i=1}^{n_j} X_{ij}$$

– Trung bình mẫu chung :

$$\overline{X} = \frac{1}{n}\sum_{j=1}^{v}\sum_{i=1}^{n_j} X_{ij}$$

– Phương sai mẫu trong nhóm j :

$$S_{nj}^2(X) = \frac{1}{n_j}\sum_{i=1}^{n_j}(X_{ij} - \overline{X}_j)^2 = \frac{1}{n_j}\sum_{k=1}^{r} n_{kj}(X_{kj} - \overline{X}_j)^2$$

(Trường hợp $(X_{1j}, X_{2j}, \ldots, X_{n_j j})$ có r khác nhau và tần số tương ứng $n_{1j}, \ldots, n_{rj}$, $n_j = n_{1j} + n_{2j} + \ldots + n_{rj}$).

– Phương sai mẫu nhóm trong (nt – nhóm trong) :

$$S_{nt}^2(X) = \frac{1}{n}\sum_{j=1}^{v} n_j S_{n_j}^2(X)$$

– Phương sai mẫu nhóm giữa : $S_G^2(X) = \frac{1}{n}\sum_{j=1}^{v} n_j(\overline{X}_j - \overline{X})^2$

– Phương sai chung : $S_c^2(X) = \frac{1}{n}\sum_{J=1}^{r}\sum_{i=1}^{n_j}(X_{iJ} - \overline{X})^2$

Ta có thể chứng minh được rằng $S_c^2(X) = S_G^2(X) + S_{nt}^2(X)$

Chú ý 4.2 : Nếu các dữ liệu được chia thành từng nhóm và nếu không còn giữ được kết quả ban đầu thì ta cho tất cả các kết quả của mỗi nhóm lấy cùng một giá trị là giá trị trung bình của nhóm.

Nếu các nhóm có cùng chiều dài h (các quan sát X_i cách đều với khoảng cách h) và nếu tần suất của các nhóm giảm nhanh về phía biên thì người ta hiệu chỉnh giá trị tính được của phương sai bằng cách trừ đi 1 lượng là $\frac{h^2}{12}$ (được gọi là hiệu chỉnh Sêpớt). Trong ví dụ 4.7, khoảng cách giữa các x_i^* là h = 4. Vậy $S_n^2(X) = \frac{1167}{8} - \left(\frac{34}{3}\right)^2 - \frac{4^2}{12} = \frac{1159}{72}$.

Ví dụ 4.8. Cho mẫu ngẫu nhiên :

Nhóm 1		Nhóm 2	
X_{i1}	n_{ir}	X_{i2}	n_{ir}
2	1	3	2
4	7	8	3
5	2		

Tìm $\overline{X}_1, \overline{X}_2, \overline{X}, S_{n_1}^2(X), S_{n_2}^2(X), S_{n_t}^2(X), S_G^2(X), S_c^2(X)$

Giải :

$$\overline{X}_1 = \frac{1\times 2 + 7\times 4 + 2\times 5}{10} = 4\ ;\ \overline{X}_2 = \frac{2\times 3 + 3\times 8}{5} = 6\ ;$$

$$\bar{X} = \frac{1}{n}\sum_{j=1}^{v}\sum_{i=1}^{n_j} X_{ij} = \frac{14}{3}$$

$$S^2_{n_1}(X) = \frac{1}{n_1}\sum n_{i_1}(X_{i_1} - \bar{X}_1) = \frac{1\times(2-4)^2 + 7(4-4)^2 + 2(5-4)^2}{10} = 0,6$$

Tương tự $S^2_{n_2}(X) = \dfrac{2(3-6)^2 + 3(8-6)^2}{5} = 6$

– Phương sai mẫu nhóm trong :

$$S^2_{n_t}(X) = \frac{n_1 S^2_{n_1}(X) + n_2 S^2_{n_2}(X)}{n} = \frac{10\times 0,6 + 5\times 6}{15} = \frac{12}{5}$$

– Phương sai mẫu nhóm giữa :

$$S^2_G(X) = \frac{n_1(\bar{X}_1 - \bar{X})^2 + n_2(\bar{X}_2 - \bar{X})^2}{n} = \frac{10\left(4 - \frac{14}{3}\right)^2 + 5\left(6 - \frac{14}{3}\right)^2}{15} = \frac{8}{9}$$

– Phương sai mẫu chung :

$$S^2_C(X) = \frac{1.\left(2-\frac{14}{3}\right)^2 + 7\left(4-\frac{14}{3}\right)^2 + 2\left(5-\frac{14}{3}\right)^2 + 2\left(3-\frac{14}{3}\right)^2 + 3\left(8-\frac{14}{3}\right)^2}{15}$$

$$= \frac{148}{45}$$

c) Hệ số tương quan mẫu

Cho mẫu ngẫu nhiên $(X_1, Y_1), (X_2, Y_2), ..., (X_n, Y_n)$ đối với cặp biến ngẫu nhiên (X, Y).

Hệ số tương quan mẫu của (X, Y) được tính theo công thức :

$$r = \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{S_n(X)S_n(Y)}$$

$$r = \frac{n\sum_{i=1}^{n} X_i Y_i - (\sum_{i=1}^{n} X_i)(\sum_{i=1}^{n} Y_i)}{\sqrt{\left[n\sum_{i=1}^{n} X_i^2 - (\sum_{i=1}^{n} X_i)^2\right]\left[n\sum_{i=1}^{n} Y_i^2 - (\sum_{i=1}^{n} Y_i)^2\right]}}$$

Ví dụ 4.9. Cho mẫu ngẫu nhiên với cặp biến ngẫu nhiên (X, Y) là :

X	2	3	4	5	6	7	8	9
Y	3	7	8	9	13	17	16	17

Tìm hệ số tương quan mẫu r.

Giải :

Từ dữ liệu trên ta tính được :

$$\sum_{i=1}^{n} X_i = 2 + 3 + 4 + 5 + 6 + 7 + 8 + 9 = 44$$

$$\sum_{i=1}^{n} Y_i = 3 + 7 + 8 + 9 + 13 + 15 + 16 + 17 = 88$$

$$\sum_{i=1}^{n} X_i^2 = 4 + 9 + 16 + 25 + 36 + 49 + 64 + 81 = 284$$

$$\sum_{i=1}^{n} Y_i^2 = 1142$$

$$\sum_{n=8} X_i Y_i = 2 \times 3 + 3 \times 7 + 4 \times 8 + 5 \times 9 + 6 \times 13 + 7 \times 15 +$$

$$+ 8 \times 16 + 9 \times 17 = 568.$$

$$r = \frac{8 \times 568 - 44 \times 88}{\sqrt{[8 \times 284 - 44^2][8 \times 1142 - 88^2]}} = 0,98$$

Ví dụ 4.10. Cho mẫu quan sát với cặp biến ngẫu nhiên (X, Y) là :

Y \ X	1	2	3
2	10	2	
3	1	8	
4		2	7

Tính hệ số tương quan mẫu r của X, Y.

Giải :

Ta lập các bảng phụ để tính $\sum X_i$, $\sum Y_i$, $\sum X_i^2$, $\sum Y_i^2$, $\sum X_i Y_i$

X_i	n_i	n_iX_i	$n_iX_i^2$
1	11	11	11
2	12	24	48
3	7	21	63
	n = 30	$\sum X_i = 56$	$\sum X_i^2 = 122$

Y_i	n_i	n_iY_i	$n_iY_i^2$
2	12	24	48
3	9	27	81
4	9	36	144
		$\sum Y_i = 87$	$\sum Y_i^2 = 273$

X_i	Y_i	n_i	$n_iX_iY_i$
1	2	10	20
1	3	1	3
2	2	2	8
2	3	8	48
2	4	2	16
3	4	7	84
			$\sum Y_i = 179$

$$r = \frac{30\times 179 - 56\times 87}{\sqrt{[30\times 122 - 56^2][30\times 273 - 87^2]}} \approx 0{,}87$$

Chú ý 4.3 : Nếu các quan sát X_i, Y_i là các số khá lớn hoặc dưới dạng số thập phân phức tạp gây khó khăn cho việc tính toán thì ta nên dùng phép biến đổi thích hợp để các số liệu nhỏ và đơn giản hơn. Chẳng hạn, nếu các X_i, $i = \overline{1, n}$, cách đều 1 khoảng có độ dài h_1, còn các $Y_i, i = \overline{1, n}$ cách đều nhau 1 khoảng có độ dài h_2. Khi đó ta dùng phép biến đổi :

$$u_i = \frac{X_i - X_o}{h_1},\ v_i = \frac{Y_i - Y_o}{h_2},\ i = \overline{1, n}$$

trong đó X_o, Y_o là số chọn tùy ý, sao cho nhận được các giá trị của u_i và v_i nhỏ và thuận tiện cho việc tính toán.

Lúc đó các số đặc trưng mẫu có dạng :

$$\overline{X} = x_o + h_1\overline{u},\ \overline{Y} = Y_o + h_2\overline{v},\ \overline{u} = \frac{1}{n}\sum_{i=1}^{n} u_i, \overline{v} = \frac{1}{n}\sum_{i=1}^{n} v_i$$

và $$S_n^2(X) = h_1^2\left[\frac{1}{n}\sum_{i=1}^{n} u_i^2 - (\overline{u})^2\right],\ S_n^2(Y) = h_2^2\left[\frac{1}{n}\sum_{i=1}^{n} v_i^2 - (\overline{v})^2\right]$$

và $$r = \frac{n \times \sum_{i=1}^{n} u_i v_i - (\sum_{i=1}^{n} u_i)(\sum_{i=1}^{n} v_i)}{\sqrt{[n \times \sum_{i=1}^{n} u_i^2 - (\sum_{i=1}^{n} u_i)^2][n \times \sum_{i=1}^{n} v_i^2 - (\sum_{i=1}^{n} v_i)^2]}}$$

Ví dụ 4.11. Cho mẫu ngẫu nhiên :

Y \ X	10	20	30	40	50	60
15	5	7				
25		20	23			
35			30	47	2	
45			10	11	20	6
55				9	7	3

Tính hệ số tương quan mẫu r của X, Y.

Giải : Chọn $h_1 = h_2 = 10$; $X_o = 40$, $Y_o = 35$.

Đặt $u_i = \frac{X_i - 40}{10}$, $v_i = \frac{Y_i - 35}{10}$

Ta nhận được bảng giá trị của u_i và v_i như sau :

v \ u	−3	−2	−1	0	1	2
−2	5	7				
−1		20	23			
0			30	47	2	
1			10	11	20	6
2				9	7	3

Lập các bảng phụ để tính các tổng : $\sum_{i=1}^{n} u_i, \sum_{i=1}^{n} v_i, \sum_{i=1}^{n} u_i^2, \sum_{i=1}^{n} v_i^2, \sum_{i=1}^{n} u_i v_i$

u_i	n_i	n_iu_i	$n_iu_i^2$
−3	5	−15	45
−2	27	−54	108
−1	63	−63	63
0	67	0	0
1	29	29	29
2	9	18	36
	n = 200	$\sum u_i = -85$	$\sum u_i^2 = 281$

v_i	n_i	n_iv_i	$n_iv_i^2$
−2	12	−24	48
−1	43	−43	43
0	79	0	0
1	47	47	47
2	19	38	76
		$\sum v_i = 18$	$\sum v_i^2 = 214$

u_i	v_i	n_i	$n_iu_iv_i$
−3	−2	5	30
−2	−2	7	28
−2	−1	20	40
−1	−1	23	23
−1	0	30	0
−1	1	10	−10
0	0	47	0
0	1	11	0
0	2	9	0
1	0	2	0
1	1	20	20
1	2	7	14
2	1	6	12
2	2	3	12
			$\sum u_iv_i = 169$

$$r = \frac{200 \times 169 - (-85)(18)}{\sqrt{[200 \times 281 - 85^2][200 \times 214 - 18^2]}} = 0{,}77$$

d) Mômen mẫu

– Mômen gốc mẫu bậc k của đại lượng ngẫu nhiên X được định nghĩa như sau :

$$m_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k = \frac{\sum_{i=1}^{s} n_i X_i}{\sum_{i=1}^{s} n_i} \ ; n = n_1 + n_2 + ... + n_s$$

– Mô men trung tâm mẫu bậc k của đại lượng ngẫu nhiên X :

$$a_k = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^k = \frac{\sum_{i=1}^{s} n_i (X_i - \overline{X})^k}{\sum_{i=1}^{s} n_i}$$

Chú ý : Trong trường hợp k = 4, nếu kể đến hiệu chỉnh Sêpớt ta phải cộng thêm $\frac{7h^4}{240}$, nghĩa là $a_k = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^k + \frac{7h^4}{240}$, h là khoảng cách giữa các quan sát X_i.

e) Trung vị mẫu

– Nếu mẫu ngẫu nhiên cho dưới dạng $(X_1, X_2, ..., X_n)$, thì ta sắp xếp các quan sát X_i, $i = \overline{1, n}$, theo thứ tự tăng dần :

$$X^{(1)} < X^{(2)} < ... < X^{(q-1)} < X^{(q)} < X^{(q+1)} < ... < X^{(n)}$$

+ Nếu n là số chẵn, tức là n = 2q thì trung vị là :

$$X_{Me} = \frac{X^{(q)} + X^{(q+1)}}{2}$$

+ Nếu n là số lẻ, tức là n = 2q – 1 thì :

$$X_{Me} = X^{(q)}$$

Ví dụ 4.12. Cho mẫu ngẫu nhiên khi tiến hành 100 thí nghiệm đo đại lượng X là :

X	1	2	3	4	5	6
n_i	4	6	22	16	36	16

Tìm số trung vị x_{Me}.

Giải : Ta thấy n = 100 là số chẵn. Do đó $x_{Me} = \frac{X^{(50)} + X^{(51)}}{2}$. Nhìn vào mẫu trên ta thấy có 48 giá trị X_i nhỏ hơn hoặc bằng 4 và 84 giá trị $X_i \le 5$. Vậy từ quan sát thứ 49 trở đi đến quan sát thứ 84 có giá trị bằng 5. Do đó $X^{(50)} = X^{(51)} = 5$.

Vậy $x_{Me} = \frac{5+5}{2} = 5$

– Nếu mẫu quan sát được cho dưới dạng khoảng thì số trung vị được tính theo công thức :

$$x_{Me} = A_{Me} + h \times \frac{\frac{n}{2} - m_{Me}}{n_{Me}},$$

trong đó : A_{Me} là đầu mút trái của khoảng trung vị ;

n_{Me} là số lần xuất hiện khoảng trung vị ;

m_{Me} số lần xuất hiện các khoảng trước khoảng trung vị.

Từ số liệu trong ví dụ ở bảng 4.1, ta có :

$$h = 10,\ A_{Me} = 100,\ n_{Me} = 46,\ m_{Me} = 23,\ n = 117$$

Vậy $$x_{Me} = 100 + 10 \times \frac{\frac{117}{2} - 23}{46} = 107{,}8.$$

g) Mốt (mod). Công thức tính mốt (mod) mẫu trong trường hợp số hiệu được cho dưới dạng khoảng là :

$$x_{Mod} = A_{Mo} + h \times \frac{n_{Mo} - n_{Mo-1}}{2n_{Mo} - n_{Mo-1} - n_{Mo+1}}$$

trong đó : A_{Mo} là mút trái của khoảng Mốt ;

n_{Mo} số lần xuất hiện khoảng Mốt ;

n_{Mo-1} số lần xuất hiện của khoảng trước khoảng Mốt ;

n_{Mo+1} là số lần xuất hiện của khoảng sau khoảng Mốt.

Từ số liệu ở bảng 4.1 ta có :

$$n_{Mo} = 46,\ A_{Mo} = 100,\ n_{Mo-1} = 15,\ n_{Mo+1} = 29,\ h = 10$$

Vậy $$x_{Mod} = 100 + 10 \times \frac{46 - 15}{2 \times 46 - 15 - 29} = 106{,}5$$

Nếu số liệu sắp xếp theo các giá trị khác nhau của các X_i thì x_{Mod} chính là giá trị của X mà tần suất xuất hiện giá trị đó là lớn nhất.

Ví dụ 4.13. Cho mẫu quan sát của đại lượng ngẫu nhiên X là :

X	1	2	3
n_i	1	4	1

Ta thấy tần suất xuất hiện giá trị 1, 2, 3 tương ứng là $\frac{1}{6}, \frac{2}{3}, \frac{1}{6}$. Tần suất xuất hiện X = 2 là $\frac{2}{3}$ lớn nhất trong các tần suất còn lại. Vậy $x_{Mod} = 2$.

Chương V

ƯỚC LƯỢNG THAM SỐ

> ***Những nội dung chính trong chương :***
> ** Các khái niệm ước lượng điểm.*
> ** Phương pháp hợp lý cực đại.*
> ** Khoảng ước lượng của xác suất, của kỳ vọng, hiệu hai xác suất, hiệu hai kỳ vọng.*

5.1. ƯỚC LƯỢNG ĐIỂM

5.1.1. Các khái niệm ước lượng điểm

Giả sử $(X_1, X_2,..., X_n)$ là mẫu ngẫu nhiên từ phân phối $f(x, \theta)$, tham số $\theta \in U$.

Định nghĩa 5.1. Ước lượng điểm của tham số θ là đại lượng ngẫu nhiên $T_n(X) = \varphi(X_1, ..., X_n)$ chỉ phụ thuộc vào các quan sát X_i và không phụ thuộc tham số θ.

Ví dụ 5.1. Giả sử $(X_1, X_2,..., X_n)$ là mẫu ngẫu nhiên độc lập từ phân phối chuẩn dạng $N(a ; \sigma^2)$.

* Đại lượng $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ là ước lượng điểm của kỳ vọng a. Ta thấy trong biểu thức của $\overline{X}$ không có mặt a mà chỉ chứa các $X_1, ..., X_n$.

* Đại lượng $S_n^2(X) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$

hoặc $S_n^{*2}(X) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ là các ước lượng điểm của σ^2. Ta thấy trong biểu thức của $S_n^2(X), S_n^{*2}(X)$ không có mặt σ^2, mà chỉ có các quan sát $X_1, ..., X_n$.

Định nghĩa 5.2. Ước lượng T_n được gọi là ước lượng không chệch của tham số θ nếu $ET_n = \theta$.

Ví dụ 5.2. Mẫu ngẫu nhiên cho trong ví dụ 5.1. Chứng minh rằng :

+ $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ là ước lượng không chệch của a.

+ $S_n^{*2}(X)$ là ước lượng không chệch của σ^2.

Giải :

+ Ta có $E(\overline{X}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{na}{n} = a$

Vậy $\overline{X}$ là ước lượng không chệch của a.

+ Ta có :

$$S_n^{*2}(X) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - a + (a - \overline{X}))^2$$

$$= \frac{1}{n-1}\sum_{i=1}^{n}\{(X_i - a)^2 + (X_i - a)(a - \overline{X}) + (a - \overline{X})^2\}$$

Ta biết $a - \overline{X} = -\frac{1}{n}\sum_{i=1}^{n}(X_i - a)$

$$(a - \overline{X})^2 = \frac{1}{n^2}[\sum_{i=1}^{n}(X_i - a)^2 + \sum_{i \neq j}(X_i - a)(X_j - a)]$$

Vậy $S_n^{*2}(X) = \frac{1}{n-1}\sum_{i=1}^{n}\{(X_i - a)^2 - \frac{2}{n}(X_i - a)^2 - \frac{2}{n}\sum_{\substack{k=1 \\ k \neq i}}^{n}(X_k - a)(X_i - a) +$

$$+ \frac{1}{n^2}[\sum_{i=1}^{n}(X_i - a)^2 + \sum_{i \neq j}(X_i - a)(X_j - a)]\}$$

$$ES^{*2}(X) = \frac{1}{n-1}\sum_{i=1}^{n}\{\sigma^2 - \frac{2}{n}\sigma^2 - 0 + \frac{1}{n^2}[n\sigma^2 + 0]\}$$

$$ES_n^{*2}(X) = \frac{1}{n-1}\sum_{i=1}^{n}\left\{\sigma^2 - \frac{\sigma^2}{n}\right\} = \frac{n\sigma^2 - \sigma^2}{n-1} = \sigma^2$$

Vậy $S_n^{*2}(X)$ là ước lượng không chệch của σ^2.

Ta cũng dễ dàng suy ra $S_n^2(X)$ không là ước lượng không chệch của σ^2. Vì :

$$ES_n^2(X) = \frac{n-1}{n} ES_n^{*2}(X) = \frac{n-1}{n}\sigma^2 \neq \sigma^2$$

Định nghĩa 5.3. Ước lượng T_n của tham số θ được gọi là ước lượng vững của tham số θ nếu $T_n \xrightarrow{P} \theta$ theo xác suất khi $n \to \infty$, nghĩa là với $\varepsilon > 0$ cho trước tùy ý có $\lim\limits_{n\to\infty} P[| T_n - \theta |< \varepsilon] = 1$.

Ví dụ 5.3. Giả sử $(X_1, ..., X_n)$ là mẫu ngẫu nhiên độc lập từ phân phối chuẩn dạng $N(a, \sigma^2)$. Chứng minh rằng $\overline{X} = \frac{1}{n}\sum\limits_{i=1}^{n} X_i$ là ước lượng vững của a.

Giải :

Vì $X_1, X_2,..., X_n$ là dãy biến ngẫu nhiên độc lập có $EX_1 = EX_2 = ... = EX_n = a$ và $DX_1 = \sigma^2, ..., DX_n = \sigma^2$. Theo hệ quả 1 của Định lý Tsêbưsép ta có $\overline{X} \xrightarrow{P} a$ khi $n \to \infty$. Theo định nghĩa ước lượng vững, $\overline{X}$ là ước lượng vững của a.

Ví dụ 5.4. Giả sử k là số lần xuất hiện biến cố A trong dãy n phép thử Bernoulli. P(A) là xác suất xuất hiện biến cố A trong mọi phép thử, không đổi và bằng p.

Chứng minh rằng $\frac{k}{n}$ là ước lượng vững của p.

Giải : Gọi X_i là số lần xuất hiện biến cố A trong phép thử thứ i. Vậy $k = X_1 + X_2 + ... + X_n$.

Ta biết rằng $X_1, X_2,..., X_n$ là độc lập và X_i có phân phối xác suất là :

X	0	1
	$1-p$	p

Ta có $EX_i = p$; $DX_i = p(1-p) \leq \frac{1}{4}$.

Vậy dãy $X_1, X_2,..., X_n$ độc lập $EX_1 = EX_2 = ... = EX_n = p$ và $DX_1 \le \frac{1}{4}$, $DX_2 \le \frac{1}{4}$, ..., $DX_n \le \frac{1}{4}$. Theo hệ quả 1 của định lý Tsêbưsep ta có $\frac{X_1 + ... + X_n}{n} \xrightarrow{P} p$ khi $n \to \infty$, nghĩa là $\frac{k}{n} \xrightarrow{P} p$ khi $n \to \infty$. Vậy $\frac{k}{n}$ là ước lượng vững của p.

Định nghĩa 5.4. Ước lượng T_n của tham số θ được gọi là ước lượng không chệch tốt nhất nếu :

+ $E(T_n) = \theta$.

+ $DT_n \le D\hat{T}_n$, $\hat{T}_n$ là ước lượng không chệch bất kỳ của θ.

5.1.2. Phương pháp hợp lý cực đại để tìm ước lượng

Định nghĩa 5.5. (Hàm hợp lý)

Giả sử $(X_1, X_2,..., X_n)$ là mẫu ngẫu nhiên độc lập từ phân phối $f(x, \theta)$, $\theta \in U$.

Gọi tích $L(X/\theta) = f(X_1, \theta)\, f(X_2, \theta) ..., f(X_n, \theta)$ là hàm hợp lý.

Định nghĩa 5.6. (Ước lượng hợp lý cực đại)

Ước lượng $\hat{\theta}(X_1, ..., X_n)$ được gọi là ước lượng hợp lý cực đại của tham số θ nếu $L(X/\hat{\theta}(X)) \ge L(X/\theta)$ với mọi $\theta \in U$.

Từ định nghĩa ước lượng hợp lý cực đại ta rút ra phương pháp tìm ước lượng như sau :

Tìm giá trị $\hat{\theta}(X)$ của θ sao cho $L(X/\theta)$ đạt cực đại tại $\hat{\theta}(X)$.

– Trường hợp θ là một số :

Ta tìm : $$\frac{\partial L(X/\theta)}{\partial \theta} = 0 \qquad (5.1)$$

Giải phương trình (5.1) tìm được $\hat{\theta}$. Sau đó xét dấu của L'_θ hoặc $L''_{\theta\theta}$ xem $\hat{\theta}$ có phải là điểm để $L(X/\theta)$ đạt cực đại hay không. Nếu tại $\hat{\theta}$ mà $L(X/\theta)$ đạt cực đại thì $\hat{\theta}$ là ước lượng phải tìm. Nếu $f(x, \theta) > 0$ thì $L(X/\theta) > 0$. Ta có thể viết :

$$\frac{1}{L}\cdot\frac{\partial L}{\partial \theta} = 0 \Leftrightarrow \frac{\partial \ln L}{\partial \theta} = 0 \Leftrightarrow \sum_{i=1}^{n} \frac{\partial \ln f(X_i, \theta)}{\partial \theta} = 0 \qquad (5.2)$$

Giải phương trình (5.2) ta được $\hat{\theta}(X)$. Cũng lý luận như trên ta nhận được ước lượng phải tìm.

Ví dụ 5.5. Giả sử $(X_1, ..., X_n)$ là mẫu ngẫu nhiên độc lập từ phân phối Poisson với tham số $\lambda > 0$.

Tìm ước lượng hợp lý cực đại của λ.

Giải :

Tính $f(X_i, \lambda) = \dfrac{\lambda^{X_i} e^{-\lambda}}{X_i\,!}$, $X_i = 0, 1, 2, ... \in$

$$\ln f(X_i, \lambda) = X_i \ln\lambda - \lambda - \ln X_i!$$

$$\frac{\partial \ln f(X_i;\lambda)}{\partial\lambda} = \frac{X_i}{\lambda} - 1$$

Thay vào phương trình (5.2) ta có $\sum_{i=1}^{n}\left(\dfrac{X_i}{\lambda} - 1\right) = 0$

Ta suy ra $\sum_{i=1}^{n} X_i - n\lambda = 0$. Vậy $\hat{\lambda} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$.

Xét cực đại của L :

Ta có $\dfrac{\partial^2 L}{\partial\lambda^2} = \sum_{i=1}^{n}\dfrac{X_i}{\lambda^2} = -\dfrac{\sum_{i=1}^{n} X_i}{\lambda^2} < 0$ với mọi λ, vì các $X_i \geq 0$ nên $\sum_{i=0}^{n} X_i \geq 0$. Vì vậy $\lambda = \dfrac{1}{n}\sum_{i=1}^{n} X_i$ là ước lượng phải tìm.

– Trường hợp $\theta = (\theta_1, \theta_2, ..., \theta_r)$:

Làm tương tự như trường hợp tham số là 1 số. Ở đây có r ẩn, do đó ta phải lập được hệ r phương trình :

$$\begin{cases} \dfrac{\partial L}{\partial\theta_1} = 0 \\ .... \\ \dfrac{\partial L}{\partial\theta_r} = 0 \end{cases} \tag{5.3}$$

giải hệ này ta tìm được $\hat{\theta} = (\hat{\theta}_1, ..., \hat{\theta}_r)$ thường hay gặp hàm mật độ có dạng hàm mũ. Do đó đạo hàm của hàm mũ lũy thừa vẫn còn dạng hàm số mũ.

Việc giải phương trình có chứa hàm số mũ thường phức tạp. Vì vậy, giống như trường hợp tham số là 1 số, ta cũng nhận được hệ r phương trình :

$$\begin{cases} \sum_{i=1}^{n} \dfrac{\partial \ln f(X_i, \theta_1, ..., \theta_r)}{\partial \theta_1} = 0 \\ .... \\ \sum_{i=1}^{n} \dfrac{\partial \ln f(X_i, \theta)}{\partial \theta_r} = 0 \end{cases} \tag{5.4}$$

Giải hệ (5.4) ta tìm được $\hat{\theta}(X) = (\hat{\theta}_1(X), ..., \hat{\theta}_r(X))$.

Ví dụ 5.6. Giả sử $(X_1, X_2, ..., X_n)$ là mẫu ngẫu nhiên độc lập từ phân phối chuẩn dạng $N(a ; \sigma^2)$. Tìm ước lượng hợp lý cực đại của $(a ; \sigma^2)$.

Giải :

$$\ln f(X_i ; a, \sigma^2) = \ln\left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(X_i - a)^2}{2\sigma^2}}\right)$$

$$= -\frac{(X_i - a)^2}{2\sigma^2} - \frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln\sigma^2$$

$$\frac{\partial \ln f}{\partial a} = \frac{X_i - a}{\sigma^2} ; \frac{\partial \ln f}{\partial(\sigma^2)} = \frac{(X_i - a)^2}{2(\sigma^2)^2} - \frac{1}{2\sigma^2}$$

Thay vào hệ phương trình (5.4) ta có :

$$\begin{cases} \sum_{i=1}^{n} \dfrac{(X_i - a)^2}{\sigma^2} = 0 \\ \sum_{i=1}^{n} \left(\dfrac{(X_i - a)^2}{2(\sigma^2)^2} - \dfrac{1}{2\sigma^2}\right) = 0 \end{cases} \Leftrightarrow \begin{cases} \hat{a} = \dfrac{1}{n}\sum_{i=1}^{n} X_i = \overline{X} \\ \hat{\sigma}^2 = \dfrac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 \end{cases}$$

5.2. ƯỚC LƯỢNG KHOẢNG

Giả sử $(X_1, X_2, ..., X_n)$ là mẫu ngẫu nhiên độc lập từ phân phối $f(x, \theta), \theta \in U$.

Định nghĩa 5.7. Khoảng $(\hat{\theta}_1(X_1, ..., X_n) ; \hat{\theta}_2(X_1, ..., X_n))$, $(\hat{\theta}_1 < \hat{\theta}_2)$, được gọi là khoảng ước lượng của tham số θ với độ tin cậy $1 - \alpha$ nếu :

$$P[\hat{\theta}_1(X) < \theta < \hat{\theta}_2(X)] = 1 - \alpha \tag{5.5}$$

Dựa vào định nghĩa này ta có thể tìm được các khoảng ước lượng của các tham số trong một số phân phối thông dụng.

5.2.1. Khoảng ước lượng của kỳ vọng a

Giả sử $(X_1, X_2,..., X_n)$ là mẫu ngẫu nhiên từ phân phối chuẩn dạng $N(a, \sigma^2)$. Tìm khoảng ước lượng của a với độ tin cậy $1 - \alpha$.

Giải :

– Trường hợp σ đã biết :

Xét xác suất $P[|\overline{X} - a| < \varepsilon] = 1 - \alpha = P\left[\left|\frac{\overline{X} - a}{\frac{\sigma}{\sqrt{n}}}\right| < \frac{\varepsilon\sqrt{n}}{\sigma}\right] = 1 - \alpha$

Vì $\overline{X}$ có sự phân phối chuẩn dạng $N\left(a; \frac{\sigma^2}{N}\right)$ nên $\frac{\overline{X} - a}{\frac{\sigma}{\sqrt{n}}}$ có phân phối chuẩn dạng N(0 ; 1). Đặt $x_\alpha = \frac{\varepsilon\sqrt{n}}{\sigma}$, ta có :

$$P[|\overline{X} - a| < \varepsilon] = P\left[\frac{|\overline{X} - a|\sqrt{n}}{\sigma} > x_\alpha\right] = 2\Phi(x_\alpha) - 1 = 1 - \alpha.$$

Vậy $\Phi(x_\alpha) = 1 - \frac{\alpha}{2}$. Từ đây ta tìm x_α bằng cách tra bảng giá trị hàm phân phối chuẩn N(0 ; 1).

Giải $\frac{|(\overline{X} - a)\sqrt{n}|}{\sigma} < x_\alpha$, ta nhận được

$$\overline{X} - x_\alpha \frac{\sigma}{\sqrt{n}} < a < \overline{X} + x_\alpha \frac{\sigma}{\sqrt{n}} \tag{5.6}$$

hoặc có thể viết $a = \overline{X} \pm x_\alpha \frac{\sigma}{\sqrt{n}}$.

Đây là khoảng ước lượng phải tìm.

– Trường hợp σ chưa biết :

Ta cũng xét xác suất $P[|\bar{X} - a| < \varepsilon] = 1 - \alpha$

Ta có $P[|\bar{X} - a| < \varepsilon] = P\left[\dfrac{|(\bar{X} - a)\sqrt{n}|}{S_n^*(X)} < t_\alpha\right] = 1 - \alpha$

trong đó $t_\alpha = \dfrac{\varepsilon\sqrt{n}}{S_n^*(X)}$. Đại lượng $\dfrac{(\bar{X} - a)\sqrt{n}}{S_n^*(X)}$ có phân phối Student với $n - 1$ bậc tự do. Vậy $P\left[\dfrac{|\bar{X} - a|\sqrt{n}}{S_n^*(X)} < t_\alpha\right] = 2\varphi(t, n - 1) - 1 = 1 - \alpha$

Vậy $\varphi(t ; n - 1) = 1 - \dfrac{\alpha}{2}$. t_α có thể tra ở bảng phân phối Student với $n - 1$ bậc tự do và mức ý nghĩa α (bảng tiêu chuẩn hai phía). Theo định lý giới hạn trung tâm $\dfrac{(\bar{X} - a)\sqrt{n}}{S_n^*(X)}$ có phân phối tiệm cận là phân phối chuẩn $N(0 ; 1)$. Vì vậy với n khá lớn, $n > 30$, ta có thể tra ở bảng phân phối chuẩn $N(0 ; 1)$ sao cho $\Phi(t_\alpha) = 1 - \dfrac{\alpha}{2}$.

Giải $\dfrac{|(\bar{X} - a)\sqrt{n}|}{S_n^*(X)} < t_\alpha$ ta được :

$$\bar{X} - \frac{t_\alpha S_n^*(X)}{\sqrt{n}} < a < \bar{X} + \frac{t_\alpha S_n^*(X)}{\sqrt{n}} \tag{5.7}$$

hoặc viết $a = \bar{X} \pm t_\alpha \dfrac{S_n^*(X)}{\sqrt{n}}$, $S_n^{*2} = \dfrac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$.

+ Nếu $n \geq 30$ thì t_α tra ở bảng phân phối chuẩn $N(0 ; 1)$ sao cho $\Phi(t_\alpha) = 1 - \dfrac{\alpha}{2}$.

+ Nếu $n < 30$ thì t_α tra ở bảng phân phối Student với $n - 1$ bậc tự do và mức ý nghĩa α (bảng tiêu chuẩn hai phía).

Ví dụ 5.7. Tìm khoảng ước lượng của kỳ vọng a với độ tin cậy 0,95. Biết rằng mẫu $(X_1, ..., X_n)$ độc lập từ phân phối chuẩn N(a ; 25) và $\overline{X} = 14$, n = 25.

Giải : Ở đây $1 - \alpha = 0,95$ ta suy ra $\alpha = 0,05$. Tra bảng giá trị của hàm phân phối chuẩn N(0 ; 1) ta tìm được $x_\alpha = 1,96$. Vậy khoảng ước của a với độ tin cậy 0,95 là :

$$14 - 1,96 \times \frac{5}{\sqrt{25}} < a < 14 + 1,96 \times \frac{5}{\sqrt{25}}$$

$12,04 < a < 15,96$

Ví dụ 5.8. Tiến hành kiểm tra ngẫu nhiên 10 học sinh bằng 1 bài vật lý. Kết quả điểm của 10 học sinh được cho bởi bảng sau.

X : 5   4   3   5   6   7   6   2   8   5

Giả sử các quan sát này có phân phối chuẩn dạng tổng quát $N(a ; \sigma^2)$. Tìm khoảng ước lượng của a với độ tin cậy 0,95.

Giải :

Ở đây σ chưa biết, n = 10. Vì vậy t_α tra ở bảng phân phối Student với 9 bậc tự do và mức $\alpha = 0,05$ (Bảng tiêu chuẩn 2 phía), t(5%, 9) = 2,26.

$$\overline{X} = \frac{5+4+3+5+6+7+6+2+8+5}{10} = 5,1$$

$$S_n^{*2}(X) = \frac{1}{9}[(5-5,1)^2 + (4-5,1)^2 + (3-5,1)^2 + 2(5-5,1)^2 +$$
$$+ 2(6-5,1)^2 + (7-5,1)^2 + (2-5,1)^2 + (8-5,1)^2] \approx 3,21$$

Vậy khoảng ước lượng của a với độ tin cậy 0,95 là :

$$5,1 - 2,26 \times \frac{\sqrt{3,21}}{\sqrt{10}} < a < 5,1 + 2,26 \times \frac{\sqrt{3,21}}{\sqrt{10}}$$

$3,71 < a < 6,39.$

5.2.2. Khoảng ước lượng của phương sai σ^2 trong mẫu từ phân phối chuẩn

Người ta tìm được khoảng ước lượng của phương sai σ^2 với độ tin cậy $1 - \alpha$ là :

$$\frac{(n-1)S_n^{*2}(X)}{t_2} < \sigma^2 < \frac{(n-1)S_n^{*2}(X)}{t_1} \tag{5.8}$$

trong đó t_1; t_2 tra trong bảng phân phối χ^2 với n – 1 bậc tự do sao cho $P[\chi^2 > t_1] = 1 - \frac{\alpha}{2}$; $P[\chi^2 > t_2] = \frac{\alpha}{2}$.

$$S_n^{*2}(X) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

Trong ví dụ 5.8, tìm khoảng ước lượng của tham số σ^2 với độ tin cậy 0,90.

Giải :

$$1 - \alpha = 0{,}90, \text{ ta suy ra } \alpha = 0{,}10.$$

tra bảng phân phối χ^2 ta tìm được t_1 (95% ; 9) = 3,325 ; t_2 = (5% ; 9) = 16,919.

Khoảng ước lượng của σ^2 với độ tin cậy 0,90 là :

$$\frac{9 \times 3{,}21}{16{,}919} < \sigma^2 < \frac{9 \times 3{,}21}{3{,}325}$$

$$1{,}5 < \sigma^2 < 8{,}6$$

5.2.3. Khoảng ước lượng của hiệu hai trung bình của hai mẫu độc lập từ phân phối chuẩn

Giả sử $(X_1, X_2,..., X_n)$ là mẫu ngẫu nhiên độc lập từ phân phối chuẩn $N(a, \sigma_1^2)$ và $(Y_1, Y_2,..., Y_m)$ là mẫu ngẫu nhiên độc lập từ phân phối chuẩn $N(a_2, \sigma_2^2)$. X, Y là độc lập.

Người ta tìm khoảng ước lượng của hiệu hai kỳ vọng $a_1 - a_2$ như sau :

– Trường hợp σ_1, σ_2 đã biết

Khoảng ước lượng của hiệu 2 trung bình $a_1 - a_2$ với độ tin cậy $1 - \alpha$ là :

$$a_1 - a_2 = \bar{X} - \bar{Y} \pm x_\alpha\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{n}} \tag{5.9}$$

x_α tra ở bảng phân phối chuẩn N(0 ; 1) sao cho $\Phi(x_\alpha) = 1 - \frac{\alpha}{2}$

– Trường hợp DX, DY chưa biết, ta phải giả thiết DX = DY.

Khi đó khoảng ước lượng của hiệu hai trung bình $a_1 - a_2$ với độ tin cậy $1 - \alpha$ là :

$$a_1 - a_2 = \overline{X} - \overline{Y} \pm t_\alpha S\sqrt{\frac{1}{n} + \frac{1}{m}} \qquad (5.10)$$

trong đó $S^2 = \dfrac{1}{n+m-2}\left[\sum_{i=1}^{n} X_i^2 - \dfrac{(\sum_{i=1}^{n} X_i)^2}{n} + \sum_{i=1}^{n} Y_i^2 - \dfrac{(\sum_{i=1}^{n} Y_i)^2}{n}\right]$

+ Nếu $n + m \leq 60$ thì t_α tra ở bảng phân phối Student với $n + m - 2$ bậc tự do và mức α. (Bảng tiêu chuẩn hai phía).

+ Nếu $n + m > 60$ thì t_α tra bảng phân phối chuẩn sao cho $\Phi(t_\alpha) = 1 - \dfrac{\alpha}{2}$.

Ví dụ 5.9. Cho hai mẫu ngẫu nhiên độc lập từ phân phối chuẩn :

X :	15,7	10,3	12,6	14,5	12,6	13,8	11,9
Y :	12,3	13,7	10,4	11,4	14,9	12,6	

Giả sử $EX = a_1$, $EY = a_2$, $DX = DY$.

Tìm khoảng ước lượng của hiệu hai trung bình $a_1 - a_2$ với độ tin cậy 0,90.

Giải :

Theo giả thiết của bài toán ta có $n = 7$, $m = 6$, $\alpha = 0{,}10$.

Tra bảng phân phối Student ta tìm được t (10%, 11) = 1,796. Ta tính được $\overline{X} = 13{,}06$, $\overline{Y} = 12{,}55$ và $S^2 = 2{,}89$.

Suy ra $S = \sqrt{2{,}89} = 1{,}7$

$$S\sqrt{\frac{1}{7} + \frac{1}{6}} = 1{,}7 \times 0{,}556297 = 0{,}9458$$

$$t_\alpha . S\sqrt{\frac{1}{m} + \frac{1}{n}} = 1{,}796 \times 0{,}9458 = 1{,}699$$

Vậy khoảng ước lượng của hiệu 2 trung bình $a_1 - a_2$ là :

$$a_1 - a_2 = |\overline{X} - \overline{Y}| = 0,51 \pm 1,699$$

5.2.4. Khoảng ước lượng của xác suất p trong phân phối nhị thức

Xét xác suất $P\left[\left|\frac{k}{n} - p\right| < \varepsilon\right] = 1 - \alpha$

$$= P\left[\left|\frac{k - np}{\sqrt{np(1-p)}}\right| < \varepsilon \times \sqrt{\frac{n}{p(1-p)}}\right] = 1 - \alpha$$

Đặt $x_\alpha = \varepsilon\sqrt{\frac{n}{p(1-p)}}$.

Ta biết $P\left[\frac{k - np}{\sqrt{np(1-p)}} < x_\alpha\right] \approx \Phi(x_\alpha)$ (Định lý giới hạn trung tâm của Laplace), $\Phi(x)$ là hàm phân phối chuẩn N(0 ; 1).

Vậy $P\left[\left|\frac{k - np}{\sqrt{np(1-p)}}\right| < x_\alpha\right] \approx 2\Phi(x_\alpha) - 1 = 1 - \alpha$

Ta suy ra $\Phi(x_\alpha) = 1 - \frac{\alpha}{2}$. Vậy x_α tra ở bảng phân phối chuẩn N(0 ;1) sao cho $\Phi(x_\alpha) = 1 - \frac{\alpha}{2}$. Thay p trong $\sqrt{np(1-p)}$ bằng tần suất $\hat{p} = \frac{k}{n}$ và giải $\left|\frac{k - np}{\sqrt{n\hat{p}(1-\hat{p})}}\right| < x_\alpha$ ta nhận được :

$$\hat{p} - x_\alpha\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \hat{p} + x_\alpha\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \qquad (5.11)$$

Ví dụ 5.10. Gieo 400 hạt giống đậu tương thấy có 5 hạt không nảy mầm. Hãy tìm khoảng ước lượng của xác suất không nảy mầm p của mỗi hạt với độ tin cậy 0,999.

Giải :

Theo trên ta có $\Phi(x_\alpha) = 1 - \frac{\alpha}{2}$, mà $1 - \alpha = 0,999$ nên $\alpha = 0,001$. Vậy $\Phi(x_\alpha) = 1 - 0,0005 = 0,9995$.

Tra bảng phân phối chuẩn N(0 ; 1) ta tìm được $x_\alpha = 3,3$.

Thay vào biểu thức (5.11) ta được :

$$\frac{5}{400}-3,3\sqrt{\frac{\frac{5}{400}\left(1-\frac{5}{400}\right)}{400}}<p<\frac{5}{400}+3,3\sqrt{\frac{\frac{5}{400}\left(1-\frac{5}{400}\right)}{400}}$$

$-0,058 < p < 0,0308$. Vì xác suất không âm nên ta lấy khoảng

$0 \le p < 0,0308$.

5.2.5. Khoảng ước lượng của hiệu hai xác suất trong hai dãy phép thử Bernoulli

Xét hai dãy phép thử Bernoulli. Dãy thứ nhất có n phép thử ; xác suất để biến cố A xuất hiện trong mỗi phép thử của dãy I là $P(A) = p_1$; X là số lần xuất hiện biến cố A trong dãy I. Dãy thứ II có m phép thử ; Y là số lần xuất hiện biến cố A trong dãy thứ II ; xác suất để biến cố A xuất hiện trong mỗi phép thử ở dãy II là $p_2 = P(A)$.

Người ta tìm được khoảng ước lượng của hiệu hai xác suất $p_1 - p_2$ với độ tin cậy $1-\alpha$ là :

$$\hat{p}_1-\hat{p}_2-x_\alpha\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n}+\frac{\hat{p}_2(1-\hat{p}_2)}{m}}<p_1-p_2<\hat{p}_1-\hat{p}_2+$$

$$+x_\alpha\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n}+\frac{\hat{p}_2(1-\hat{p}_2)}{m}} \qquad (5.12)$$

Ví dụ 5.11. Để đánh giá chất lượng sản phẩm do 2 nhà máy sản xuất ra, người ta kiểm tra ngẫu nhiên 200 sản phẩm ở nhà máy I thấy có 20 phế phẩm và 300 sản phẩm của nhà máy II thấy có 15 phế phẩm. Tìm khoảng ước lượng của hiệu hai xác suất để tìm sản phẩm là phế phẩm của 2 nhà máy với độ tin cậy 0,95.

Giải : $\alpha = 0,05$, tra bảng phân phối chuẩn ta tìm được $x_\alpha = 1,96$. $\hat{p}_1 = \frac{X}{n} = \frac{20}{200} = 0,1$, $\hat{p}_2 = \frac{15}{300} = 0,05$.

Thay vào (5.12) ta được :

$$0,1-0,05-0,0483 < p_1 - p_2 < 0,1-0,05+0,0483$$

$$0,0017 < p_1 - p_2 < 0,0983$$

Bài tập chương IV và chương V

1. Tìm hàm phân phối mẫu, trung bình mẫu, phương sai mẫu của đại lượng ngẫu nhiên X. Biết rằng mẫu ngẫu nhiên đã cho là.

a)

X_i	1	4	6
n_i	10	15	25

b)

X_i	2	5	7	8
n_i	1	3	2	4

c)

X_i	4	7	8
n_i	5	2	3

2. Xây dựng các đa giác tần suất tương ứng với các mẫu đã cho sau :

a)

X_i	1	4	5	7
n_i	20	10	14	6

b)

X_i	2	3	5	6
n_i	10	15	5	20

c)

X_i	15	20	45	30	35
n_i	10	15	30	20	25

3. Xây dựng tổ chức đồ tần suất tương ứng với các mẫu đã cho sau đây :

Số TT	Khoảng h	Tần số n_i
1	2 – 7	5
2	7 – 12	10
3	12 – 17	25
4	.17 – 22	6
5	22 – 27	4

Số TT	Khoảng h	Tần số n_i
1	10 – 15	2
2	15 – 20	4
3	20 – 25	8
4	25 – 30	4
5	30 – 35	2

Tìm trung bình mẫu và phương sai mẫu ứng với số liệu đã cho ở trên.

4. Tìm hệ số tương quan mẫu của (X, Y) trên cơ sở các mẫu cho sau đây :

a)

X	1	2	4	5	6	7
Y	2	3	8	12	12	13

b)

Y \ X	1	2	3
2	9	1	
3	2	8	
4		3	7

5. Cho mẫu ngẫu nhiên độc lập $(X_1, X_2, ..., X_n)$ từ phân phối

X	1	0
	p	1 − p

$, 0 < p < 1$

a) Tìm ước lượng hợp lý cực đại của p.

b) Xét xem ước lượng tìm được có phải là ước lượng không chệch, ước lượng vững không ?

6. Giả sử $(X_1, X_2, ..., X_n)$ là mẫu ngẫu nhiên độc lập từ phân phối

$$f(x, \theta) = \begin{cases} \frac{1}{\theta} e^{-\frac{x}{\theta}} & \text{với } x > 0, \theta > 0 \\ 0 & \text{với } x \le 0 \end{cases}$$

a) Tìm ước lượng hợp lý cực đại của θ.

b) Xét xem ước lượng tìm được có phải là ước lượng không chệch, ước lượng vững không ?

7. Ta giả sử rằng đại lượng ngẫu nhiên X bằng sản lượng tính ra tạ/ha của loại lúa đã cho trong một miền xác định có phân phối chuẩn, có kỳ vọng toán bằng a và phương sai là σ^2. Tìm ước lượng điểm của a, của σ^2. Tìm khoảng ước lượng của a, của σ^2 với độ tin cậy 0,90. Biết rằng các kết quả thu được trên 10 mảnh đất là :

Mảnh	1	2	3	4	5	6	7	8	9	10
Sản lượng	51	48	56	57	44	52	50	60	46	47

8. Độ cao của trẻ em tuân theo phân phối chuẩn dạng $N(a, \sigma^2)$. Tìm ước lượng điểm của a, của σ^2 nếu như đo ngẫu nhiên 10 em ta nhận được kết quả như sau : (đơn vị mét)

1	2	3	4	5	6	7	8	9	10
1,5	1,55	1,49	1,51	1,50	1,52	1,45	1,6	1,5	1,46

9. Gieo 60 hạt đậu tương thấy có 15 hạt không nảy mầm. Tìm ước lượng điểm của xác suất không nảy mầm p. Tìm khoảng ước lượng của p với độ tin cậy 0,95.

10. Bắn 10.000 viên đạn độc lập vào một mục tiêu thấy có 7.000 viên trúc đích. Tìm khoảng ước lượng của xác suất trúng đích p của mỗi viên đạn với độ tin cậy 0,95.

11. Áp dụng hai phương pháp điều trị. Theo phương pháp A, điều trị 180 người có 150 người khỏi bệnh ; Theo phương pháp B, điều trị 266 người có 160 người khỏi bệnh.

Hãy tìm khoảng ước lượng của hiệu hai xác suất khỏi bệnh của mỗi người theo hai phương pháp đó với độ tin cậy 95%.

12. Người ta đo ngẫu nhiên 10 sản phẩm của nhà máy I và 12 sản phẩm của nhà máy II. Kết quả như sau :

Kích thước	X_i	3,4	3,5	3,7	3,9
Tần số	n_i	2	3	4	1

Kích thước	Y_i	3,2	3,4	3,6
Tần số	n_i	2	2	8

Giả sử các X_i, Y_i độc lập và có phân phối chuẩn $EX = a_1$, $EY = a_2$, $DX = DY$. Tìm khoảng ước lượng của hiệu hai trung bình của 2 mẫu đã cho với độ tin cậy 0,95.

13. Cho hai mẫu ngẫu nhiên độc lập

X :	15,7	10,3	12,6	14,5	12,6	13,8	11,9
Y :	12,3	13,7	10,4	11,4	14,9	12,6	

Giả sử X, Y có phân phối chuẩn dạng $N(a_1, \sigma^2)$ và $N(a_2, \sigma^2)$ tương ứng. Tìm khoảng ước lượng của hiệu hai trung bình của hai mẫu với độ tin cậy 0,90.

Hướng dẫn và trả lời bài tập chương IV và V

4. a) $\sum X_i = 25$, $\sum Y_i = 50$, $\sum X_i^2 = 131$, $\sum Y_i^2 = 534$, $\sum X_i Y_i = 263$, $r = 0{,}97$.

b) $\sum X_i = 56$, $\sum Y_i = 90$, $\sum X_i^2 = 122$, $\sum Y_i^2 = 290$, $\sum X_i Y_i = 184$, $r = 0{,}86$.

5. Hướng dẫn $f(x_i, p) = p^{x_i}(1-p)^{1-x_i}$, $x_i = 0$ hoặc 1

Đáp số : a) $\hat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{k}{n}$

b) Nó là ước lượng không chệch, ước lượng vững của p.

6. a) $\hat{\theta}(X) = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}$

b) $\hat{\theta}(X) = \overline{X}$ là ước lượng không chệch và ước lượng vững của θ.

7. Ước lượng điểm của a là $\hat{a} = 51{,}5$; ước lượng điểm của σ^2 là

$$\hat{\sigma}^2 = 27{,}61, \hat{\sigma} = 5{,}255.$$

Khoảng ước lượng của a với độ tin cậy 0,90 là $a = 51{,}5 \pm 3{,}04$.

8. a) Ước lượng của a là $\hat{a} = 1{,}508$, ước lượng điểm của σ^2, $\hat{\sigma}^2 = 0{,}00184$.

b) Khoảng ước lượng của a với độ tin cậy 0,95 là $a = 1{,}508 \pm 0{,}0306$ khoảng ước lượng của σ^2 với độ tin cậy 0,95 là $0{,}00087 < \sigma^2 < 0{,}0061$.

9. Ước lượng điểm của p là $\hat{p} = 0{,}25$. Khoảng ước lượng của p với độ tin cậy 0,95 là $p = 0{,}25 \pm 0{,}1095$.

10. Ước lượng điểm của p là $\hat{p} = 0{,}7$. Khoảng ước lượng của p với độ tin cậy 0,95 là $p = 0{,}7 \pm 0{,}00898$.

11. Khoảng ước lượng của hiệu hai xác suất với độ tin cậy 0,95 :

$p_1 - p_2 = 0{,}2318 \pm 0{,}0801$.

12. Khoảng ước lượng của hiệu hai trung bình với độ tin cậy 0,95 là $a_1 - a_2 = 0{,}1 \pm 0{,}1748$.

13. Khoảng ước lượng của hiệu hai trung bình của kích thước sản phẩm do 2 nhà máy sản xuất ra là : $a_1 - a_2 = 0{,}51 \pm 1{,}702$.

Chương VI

KIỂM ĐỊNH GIẢ THIẾT

Những nội dung chính trong chương :

** Kiểm định về trung bình, về xác suất p trong phân phối nhị thức.*

** So sánh hai trung bình, so sánh hai xác suất.*

** Tiêu chuẩn χ^2 kiểm định về phân phối đã cho ; tiêu chuẩn χ^2 kiểm định tính độc lập và tính thuần nhất.*

** Tiêu chuẩn Cônmôgôrốp, Smiécnốp kiểm định về phân phối đã cho và kiểm định tính thuần nhất.*

** Tiêu chuẩn Wilcoxon, Mann–Whitney kiểm định tính thuần nhất của 2 mẫu độc lập.*

** Tiêu chuẩn Wilcoxon kiểm định tính thuần nhất của hai mẫu phụ thuộc.*

** Phân tích phương sai theo một yếu tố, hai yếu tố.*

6.1. THIẾT LẬP BÀI TOÁN

– Giả thiết thống kê : Giả sử đại lượng ngẫu nhiên X có phân phối F(x) (hoặc F(x, θ), $\theta \in U$).

Những giả thiết về phân phối F(x) được gọi là giả thiết thống kê và ký hiệu là H_o.

Những giả thiết cũng về phân phối F(x) nhưng khác với giả thiết H_o được gọi là đối thiết (hoặc giả thiết chọn), ký hiệu là K.

Khi phân phối F(x ; θ) phụ thuộc vào tham số θ thì những giả thiết về phân phối F(x ; θ) được chuyển sang giả thiết về tham số θ.

• **Ví dụ 6.1.** Nếu biến ngẫu nhiên X có phân phối chuẩn dạng N(a ; 1) thì giả thiết về phân phối chuẩn được chuyển sang giả thiết về tham số a.

Ví dụ 6.2. a) $H_o : a = 5$ với $K : a \neq 5$

b) $H_o : a = 5$ hoặc $a = 7$ với $K : 5 < a < 7$

c) $H_o : a \leq 5$ với $K : a > 5$

Nếu tập giả thiết H_o có một phần tử (H : tập hợp các mệnh đề) thì H_o được gọi là giả thiết đơn (bài toán a và c). Nếu tập H_o có ≥ 2 phần tử thì H_o được gọi là giả thiết hợp (bài toán b).

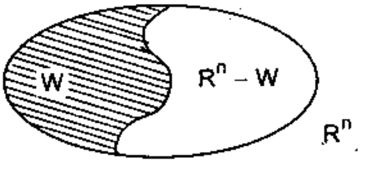
Tương tự, nếu tập K có 1 phần tử thì K được gọi là đối thiết đơn (bài toán b và c). Nếu tập K có ≥ 2 phần tử thì K được gọi là đối thiết hợp (bài toán a).

– Kiểm định giả thiết thống kê : Kiểm định giả thiết thống kê là việc chọn một trong hai quyết định là bác bỏ giả thiết H_o hoặc chấp nhận giả thiết H_o.

– Tiêu chuẩn kiểm định giả thiết : Để có được quyết định chấp nhận hoặc bác bỏ giả thiết H_0 ta phải dựa trên một tiêu chuẩn nào đó. Vậy tiêu chuẩn kiểm định giả thiết được hiểu như sau :

Tiêu chuẩn kiểm định giả thiết là một đại lượng ngẫu nhiên Z phụ thuộc vào các quan sát $X_1, X_2, ..., X_n$ và không phụ thuộc tham số θ, nghĩa là $Z = Z(X_1, ..., X_n)$ xác định trên không gian mẫu R^n, nhờ nó ta có thể kiểm định được giả thiết.

Vì Z(X) xác định trên không gian mẫu R^n, nên R^n được chia thành hai bộ phận. Một bộ phận W của R^n mà mẫu $(X_1, X_2, ..., X_n)$ rơi vào đó thì ta bác bỏ giả thiết H_o. W được gọi là miền tiêu chuẩn (hình 6.1).



Hình 6.1

Vậy miền tiêu chuẩn W là bộ phận của R^n mà ta bác bỏ giả thiết.

Tiêu chuẩn kiểm định giả thiết thể hiện ở một trong ba dạng sau đây :

Đặt $X = (X_1, ..., X_n)$

$[X \in W] \Leftrightarrow Z(X) > C_u$ bác bỏ giả thiết H_o.

$[X \in W] \Leftrightarrow Z(X) < C_v$ bác bỏ giả thiết H_o.

$[X \in W] \Leftrightarrow Z(X) > C_u$ hoặc $Z(X) < C_v$ bác bỏ giả thiết H_o.

C_u, C_v được gọi là điểm tiêu chuẩn.

Hai tiêu chuẩn đầu được gọi là tiêu chuẩn một phía (có 1 điểm tiêu chuẩn).

Tiêu chuẩn thứ ba được gọi là tiêu chuẩn hai phía.

C_u có thể là $+\infty$, C_v có thể là $-\infty$.

Muốn tìm miền tiêu chuẩn W hay tiêu chuẩn thống kê Z(X) ta dựa trên hai loại sai lầm sau đây :

– Sai lầm loại I. Nếu giả thiết H_o là giả thiết đúng mà bác bỏ H_o thì ta mắc sai lầm. Sai lầm đó được gọi là sai lầm loại I. $P[W/H_o$ đúng] là xác suất mắc sai lầm 1.

– Sai lầm loại II. Nếu giả thiết H_o là giả thiết sai mà chấp nhận giả thiết thì cũng mắc sai lầm. Sai lầm đó được gọi là sai lầm loại II. $P[R^n - W/H_o$ sai] là xác suất mắc sai lầm loại II.

Để tìm tiêu chuẩn kiểm định giả thiết ta phải đồng thời hạn chế tới mức tối thiểu khả năng mắc hai loại sai lầm trên, nghĩa là ta phải đồng thời cực tiểu hóa các xác suất sai lầm loại I và xác suất sai lầm loại II. Song việc làm này rất khó khăn. Thường trong thực tế người ta cho phép được mắc sai lầm loại I ở mức xác suất α nào đó (tùy theo tầm quan trọng của sai lầm loại I). Sau đó cực tiểu xác suất sai lầm loại II.

6.2. MỘT SỐ BÀI TOÁN KIỂM ĐỊNH GIẢ THIẾT

6.2.1. Kiểm định về xác suất p trong phân phối nhị thức (tỷ lệ phần trăm)

Bài toán. Giả sử trong dãy n phép thử Bernoulli biến cố A xuất hiện X lần. Gọi $p = P(A)$ là xác suất để A xuất hiện trong mỗi phép thử.

Kiểm định giả thiết $H_o : p = p_o$ với $K : p \neq p_o$ ở mức α.

Giải :

Người ta tìm được tiêu chuẩn để kiểm định giả thiết H_o với đối thiết K và tiêu chuẩn này được phát biểu như sau :

Giả thiết H_o bị bác bỏ ở mức α nếu :

$$|Z| = \frac{|X - np_o|}{\sqrt{np_o(1-p_o)}} > x_\alpha \qquad (6.1)$$

Còn $|Z| < x_\alpha$ thì chấp nhận H_o,

trong đó x_α tra ở bảng giá trị của hàm phân phối chuẩn N(0 ; 1) sao cho $\Phi(x_\alpha) = 1 - \frac{\alpha}{2}$.

Ví dụ 6.3. Gieo 300 hạt đậu tương. Kết quả là 261 hạt nảy mầm. Người ta nói rằng : tỷ lệ nảy mầm của đậu tương là 0,90. Điều nhận định đó có đúng không ? Tại sao ? Cho mức kiểm định $\alpha = 5\%$.

Giải : Ta xem việc gieo 300 hạt đậu như tiến hành 300 phép thử Bernoulli. Gọi p là xác suất nảy mầm của mỗi hạt đậu tương. Ta kiểm định giả thiết $H_o : p = 0{,}9$ với $K : p \neq 0{,}90$.

Ta thấy $\alpha = 0{,}05$ nên $\Phi(x_\alpha) = 1 - \dfrac{\alpha}{2} = 0{,}975$. Tra bảng chuẩn tìm được $x_\alpha = 1{,}96$.

Bây giờ tính $|Z| = \dfrac{|X - np_o|}{\sqrt{np_o(1 - p_o)}} = \dfrac{|261 - 300 \times 0{,}9|}{\sqrt{300 \times 0{,}9 \times 0{,}1}} \approx 1{,}73$

Ta nhận thấy $|Z| = 1{,}73 < x = 1{,}96$. Vậy ta phải chấp nhận giả thiết ; nghĩa là tỷ lệ nảy mầm của đậu tương là 0,90.

* Tiêu chuẩn một phía. Nếu tần suất $\dfrac{X}{n} > p_o$. Ta đi đến kiểm định giả thiết $H_o : p = p_o$ với $K : p > p_o$ ở mức α.

Giải :

Tiêu chuẩn này được phát biểu như sau :

Giả thiết H_o bị bác bỏ ở mức α nếu

$$Z = \frac{X - np_o}{\sqrt{np_o(1 - p_o)}} > x_\alpha \tag{6.2}$$

Còn $Z < x_\alpha$; chấp nhận giả thiết H_o ; trong đó x_α tra ở bảng giá trị hàm phân phối chuẩn N(0 ; 1) sao cho $\Phi(x_\alpha) = 1 - \alpha$.

– Nếu tần suất $\dfrac{X}{n} < p_o$ thì ta đi đến kiểm định giả thiết :

$H_o : p = p_o$ với $K : p < p_o$ ở mức α.

Tương tự như trường hợp đối thiết đơn $K : p > p_o$. Ta cũng tìm được tiêu chuẩn để kiểm định giả thiết $H_o : p = p_o$ với $K : p < p_o$ như sau :

Giả thiết H_o bị bác bỏ ở mức α nếu

$$Z = \frac{np_o - X}{\sqrt{np_o(1-p_o)}} > x_\alpha \tag{6.3}$$

nếu $Z < x_\alpha$ thì chấp nhận giả thiết H_o, trong đó x_α tra ở bảng phân phối chuẩn N(0 ; 1) sao cho $\Phi(x_\alpha) = 1 - \alpha$.

Ví dụ 6.4. Người ta cho biết tỷ lệ phế phẩm trong một lô hàng 0,02. Ta kiểm tra ngẫu nhiên 480 sản phẩm thấy có 12 phế phẩm. Xét xem tỷ lệ phế phẩm đã công bố có đúng không ? Tại sao ? Cho mức kiểm định $\alpha = 0{,}05$.

Giải : Ta nhận thấy n = 480 ; X = 12.

$$\frac{X}{n} = \frac{12}{480} = 0{,}025 > 0{,}02$$

Vậy ta đi đến bài toán kiểm định giả thiết :

H_o : p = 0,02 với K : p > 0,02 ở mức α = 0,05.

Tính $Z = \dfrac{X - np_o}{\sqrt{np_o(1-p_o)}} = \dfrac{12 - 480 \times 0{,}02}{\sqrt{480 \times 0{,}02 \times 0{,}98}} = 0{,}784$

$\Phi(x_\alpha) = 1 - \alpha = 1 - 0{,}05 = 0{,}95$, tra bảng phân phối chuẩn ta tìm được $x_\alpha = 1{,}65$. Ta nhận thấy $Z = 0{,}784 < x_\alpha$. Vậy ta phải chấp nhận giả thiết H_o, nghĩa là tỷ lệ phế phẩm trong lô hàng là 0,02.

Bài toán kiểm định về xác suất p trong phân phối nhị thức thường sử dụng tốt trong việc giải nhiều bài toán có liên quan đến dãy phép thử Bernoulli. Ví dụ : Xác định tỷ lệ phế phẩm trong một nhà máy sản xuất hàng hóa ; tỷ lệ con trai trên hành tinh của chúng ta ; tỷ lệ học sinh đỗ khi n thí sinh tham dự thi ; tỷ lệ khỏi bệnh của các bệnh nhân khi được điều trị theo phương pháp cải tiến nào đó v.v...

6.2.2. So sánh hai xác suất trong phân phối nhị thức

Tiêu chuẩn này chủ yếu áp dụng giải các bài toán so sánh tỷ lệ xuất hiện đặc tính A của các cá thể trong 2 đám đông khác nhau.

Bài toán : Xét hai dãy phép thử Bernoulli. Dãy I gồm n phép thử, X là số lần xuất hiện biến cố A trong dãy I, $p_1 = P(A)$ là xác suất để A xuất hiện trong mỗi phép thử của dãy I. Dãy II gồm m phép thử ; Y là số lần xuất hiện biến cố A trong dãy II, $p_2 = P(A)$ là xác suất để A xuất hiện trong mỗi phép thử của dãy II. Hãy so sánh hai xác suất p_1, p_2 ở mức α.

Ta đưa về kiểm định giả thiết :

$H_o : p_1 = p_2$ với $K : p_1 \neq p_2$ ở mức α.

Giải : Người ta chứng minh được rằng : tồn tại tiêu chuẩn để kiểm định giả thiết $H_o : p_1 = p_2$ với $K : p_1 \neq p_2$ ở mức α. Tiêu chuẩn này được phát biểu như sau :

Giả thiết H_o bị bác bỏ ở mức α nếu :

$$|Z| = \frac{\left|\frac{X}{n} - \frac{Y}{m}\right|}{\sqrt{\left(\frac{1}{n} + \frac{1}{m}\right)\left(\frac{X+Y}{n+m}\right)\left(1 - \frac{X+Y}{n+m}\right)}} > x_\alpha \qquad (6.4)$$

Còn nếu $|Z| < x_\alpha$ thì chấp nhận giả thiết H_o, trong đó x_α tra ở bảng phân phối chuẩn N(0 ; 1) sao cho $\Phi(x_\alpha) = 1 - \frac{\alpha}{2}$.

Ví dụ 6.5. Có hai phương pháp gieo hạt : theo phương pháp A gieo 100 hạt thấy có 80 hạt nảy mầm, theo phương pháp B gieo 125 hạt thấy có 90 hạt nảy mầm. Hãy so sánh hiệu quả của hai phương pháp ở mức $\alpha = 0,05$.

Giải : Ta thấy ở đây $n = 100$, $X = 80$, $m = 125$, $Y = 90$, $\alpha = 0,05$, $\Phi(x_\alpha) = 1 - \frac{\alpha}{2} = 0,975$. Tra bảng giá trị của hàm phân phối chuẩn N(0 ; 1) ta tìm được $x_\alpha = 1,96$.

Tính giá trị của |Z| trong (6.4) ta có :

$$|Z| = \frac{\left|\frac{80}{100} - \frac{90}{125}\right|}{\sqrt{\left(\frac{1}{100} + \frac{1}{125}\right)\left(\frac{80+90}{100+125}\right)\left(1 - \frac{80+90}{100+125}\right)}} = 1,387$$

Ta nhận thấy $|Z| = 1,387 < x_\alpha = 1,96$. Vậy chấp nhận giả thiết H_o, nghĩa là hiệu quả của hai phương pháp là tương đương.

– Tiêu chuẩn một phía :

+ Nếu $\frac{X}{n} > \frac{Y}{m}$ thì đưa về bài toán kiểm định giả thiết :

$H_o : p_1 = p_2$ với $K : p_1 > p_2$ ở mức α.

Giải : Đây là giả thiết đơn và đối thiết đơn. Người ta chứng minh được rằng tồn tại tiêu chuẩn để kiểm định giả thiết $H_o : p_1 = p_2$ với $K : p_1 > p_2$ và tiêu chuẩn này được phát biểu dưới dạng :

Giả thiết H_o bị bác bỏ ở mức α nếu :

$$Z = \frac{\frac{X}{n} - \frac{Y}{m}}{\sqrt{\left(\frac{1}{n} + \frac{1}{m}\right)\left(\frac{X+Y}{n+m}\right)\left(1 - \frac{X+Y}{n+m}\right)}} > x_\alpha \qquad (6.5)$$

Còn nếu $Z < x_\alpha$ thì chấp nhận giả thiết H_o ; trong đó x_α tra ở bảng giá trị của hàm phân phối chuẩn $N(0 ; 1)$ sao cho $\Phi(x_\alpha) = 1 - \alpha$.

+ Nếu $\frac{X}{n} < \frac{Y}{m}$ thì đưa về bài toán kiểm định giả thiết :

$H_o : p_1 = p_2$ với $K : p_1 < p_2$ ở mức α.

Tương tự như phần trên, người ta cũng chỉ ra được rằng tồn tại tiêu chuẩn để kiểm định giả thiết $H_o : p_1 = p_2$ với $K : p_1 < p_2$ và tiêu chuẩn này được phát biểu dưới dạng :

Giả thiết H_o bị bác bỏ ở mức α nếu :

$$Z = \frac{\frac{Y}{m} - \frac{X}{n}}{\sqrt{\left(\frac{1}{n} + \frac{1}{m}\right)\left(\frac{X+Y}{n+m}\right)\left(1 - \frac{X+Y}{n+m}\right)}} > x_\alpha \qquad (6.6)$$

Còn nếu $Z < x_\alpha$ thì chấp nhận giả thiết H_o, trong đó x_α tra ở bảng giá trị của hàm phân phối chuẩn $N(0 ; 1)$ sao cho $\Phi(x_\alpha) = 1 - \alpha$.

Ví dụ 6.6. Để đánh giá chất lượng sản phẩm do hai nhà máy sản xuất ra, người ta kiểm tra ngẫu nhiên 100 sản phẩm từ nhà máy I thấy có 20 phế phẩm và 150 sản phẩm từ nhà máy II thấy có 15 phế phẩm. Hãy so sánh chất lượng sản phẩm của hai nhà máy ở mức $\alpha = 5\%$.

Giải : $\dfrac{X}{n} = \dfrac{20}{100} = 0,2$, $\dfrac{Y}{m} = \dfrac{15}{150} = 0,1$. Ta thấy $\dfrac{X}{n} > \dfrac{Y}{m}$. Ta đi đến bài toán kiểm định giả thiết $H_o : p_1 = p_2$ với $K : p_1 > p_2$ ở mức $\alpha = 5\%$. Vì ở đây $\alpha = 0,05$ nên $\Phi(x_\alpha) = 1 - \alpha = 0,95$.

Tra trong bảng giá trị của hàm phân phối chuẩn N(0 ; 1) ta tìm được $x_\alpha = 1,65$.

Tính Z :

$$Z = \frac{\dfrac{X}{n} - \dfrac{Y}{m}}{\sqrt{\left(\dfrac{1}{n} + \dfrac{1}{m}\right)\left(\dfrac{X+Y}{n+m}\right)\left(1 - \dfrac{X+Y}{n+m}\right)}}$$

$$= \frac{0,2 - 0,1}{\sqrt{\left(\dfrac{1}{100} + \dfrac{1}{150}\right)\left(\dfrac{20+15}{100+150}\right)\left(1 - \dfrac{20+15}{100+150}\right)}}$$

$$Z = \frac{0,1}{\sqrt{0,0020}} = 2,232.$$

Ta nhận thấy $Z = 2,232 > x_\alpha = 1,65$. Vậy giả thiết H_o bị bác bỏ, nghĩa là tỷ lệ phế phẩm ở nhà máy I cao hơn tỷ lệ phế phẩm ở nhà máy II.

6.2.3. Kiểm định về trung bình

Giả sử $(X_1, X_2,..., X_n)$ là mẫu ngẫu nhiên độc lập từ phân phối chuẩn $N(a, \sigma^2)$.

Kiểm định giả thiết $H_o : a = a_o$ với $K : a \neq a_o$ ở mức α.

Giải : Người ta chứng minh được rằng : tồn tại tiêu chuẩn để kiểm định giả thiết $H_o : a = a_o$ với $K : a \neq a_o$ ở mức α.

Tiêu chuẩn này được phát biểu dưới dạng sau :

– Trường hợp σ đã biết :

Giả thiết H_o bị bác bỏ ở mức α nếu :

$$|Z| = \frac{|\overline{X} - a_o|\sqrt{n}}{\sigma} > x_\alpha \qquad (6.7)$$

Nếu $|Z| < x_\alpha$ thì chấp nhận giả thiết H_o ; trong đó x_α tra ở bảng giá trị của hàm phân phối chuẩn N(0 ; 1) sao cho $\Phi(x_\alpha) = 1 - \frac{\alpha}{2}$.

– Trường hợp σ chưa biết :

Giả thiết H_0 bị bác bỏ ở mức α nếu :

$$|Z| = \frac{|\overline{X} - a_o|\sqrt{n}}{S_n^*(X)} > t_\alpha \qquad (6.8)$$

Nếu $|Z| < t_\alpha$ thì chấp nhận giả thiết H_0,

trong đó $S_n^{*2}(X) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ còn t_α thì tra bảng như sau :

+ Nếu n > 30 thì t_α tra ở bảng giá trị hàm phân phối chuẩn N(0 ; 1) sao cho $\Phi(t_\alpha) = 1 - \frac{\alpha}{2}$.

+ Nếu n ≤ 30 thì t_α tra ở bảng phân phối Student với n – 1 bậc tự do và mức ý nghĩa α (bảng giá trị tiêu chuẩn hai phía).

Ví dụ 6.7. Giả sử $(X_1, X_2, ..., X_n)$ là mẫu ngẫu nhiên độc lập từ phân phối chuẩn N(a ; 4) và $\overline{X} = 15$, n = 100. Hãy kiểm định giả thiết H_o : a = 16,5 với K : a ≠ 16,5 ở mức α = 5%.

Giải : Ở đây ta biết σ = 2, α = 0,05, n = 100, $\overline{X} = 15$, theo công thức (6.7) ta có :

$$|Z| = \frac{|\overline{X} - a_o|\sqrt{n}}{\sigma} = \frac{|15 - 16{,}5|\sqrt{100}}{2} = 7{,}5$$

Tra bảng phân phối chuẩn với $\Phi(x_\alpha) = 1 - \frac{\alpha}{2} = 0{,}975$ ta tìm được $x_\alpha = 1{,}96$. Ta nhận thấy $|Z| = 7{,}5 > x_\alpha = 1{,}96$. Vậy ta bác bỏ giả thiết H_o, nghĩa là trung bình a ≠ 16,5.

Ví dụ 6.8. Sau một đợt huấn luyện, người ta kiểm tra ngẫu nhiên 15 học sinh với kết quả điểm cho ở bảng sau :

X : 2 3 7 6 9 7 6 6 1 2 7 8 6 5 6

Giả sử X tuân theo phân phối chuẩn dạng $N(a ; \sigma^2)$.

Hãy kiểm định giả thiết $H_o : a = 6$ với $K : a \neq 6$ ở mức $\alpha = 0{,}05$.

Giải : Ở đây σ chưa biết và n = 15 < 30, ta phải áp dụng công thức (6.8). Tra trong bảng giá trị tiêu chuẩn Student ta tìm được t(5% ; 14) = 2,14.

Tính trung bình mẫu :

$$\overline{X} = \frac{2+3+7+6+9+7+6+6+1+2+7+8+6+5+6}{15} = \frac{81}{15} = 5{,}4$$

Phương sai mẫu :

$$S_n^{*2}(X) = \frac{1}{14}[2(2-5{,}4)^2 + (3-5{,}4)^2 + 3(7-5{,}4)^2 + 5(6-5{,}4)^2 +$$
$$+ (9-5{,}4)^2 + (5-5{,}4)^2 + (1-5{,}4)^2 + (8-5{,}4)^2]$$

$$= \frac{77{,}2}{14} = 5{,}5$$

Ta suy ra $S_n^*(X) = \sqrt{5{,}5} = 2{,}35$

$$|Z| = \frac{(5{,}4-6)\sqrt{15}}{2{,}35} \approx 0{,}987$$

Ta thấy $|Z| = 0{,}987 < t_\alpha = 2{,}14$. Vậy ta phải chấp nhận giả thiết H_o, nghĩa là trung bình a = 6.

* Tiêu chuẩn một phía

• Nếu $\overline{X} > a_o$ thì ta đưa đến bài toán kiểm định giả thiết :

$H_o : a = a_o$ với $K : a > a_o$ ở mức α.

Người ta chứng minh được rằng tồn tại tiêu chuẩn để kiểm định giả thiết $H_o : a = a_o$ với $K : a > a_o$.

Tiêu chuẩn này được phát biểu dưới dạng :

– Trường hợp σ đã biết :

Giả thiết H_o bị bác bỏ ở mức α nếu :

$$Z = \frac{(\overline{X} - a_o)\sqrt{n}}{\sigma} > x_\alpha, \tag{6.9}$$

trong đó x_α tra ở bảng phân phối chuẩn N(0 ; 1) sao cho $\Phi(x_\alpha) = 1 - \alpha$.

Nếu $Z < x_\alpha$ thì chấp nhận H_o.

– Trường hợp σ chưa biết :

Giả thiết H_o bị bác bỏ ở mức α nếu :

$$Z = \frac{(\bar{X} - a_o)\sqrt{n}}{S_n^*(X)} > t_\alpha \qquad (6.10)$$

Nếu $Z < t_\alpha$ thì chấp nhận giả thiết H_o ;

trong đó $S_n^{*2}(X) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$, t_α tra như sau :

+ Nếu $n > 30$ thì t_α tra ở bảng giá trị hàm phân phối chuẩn sao cho $\Phi(t_\alpha) = 1 - \alpha$

+ Nếu $n < 30$ thì t_α tra ở bảng giá trị tiêu chuẩn Student (bảng tiêu chuẩn 1 phía) với $n - 1$ bậc tự do và mức ý nghĩa α.

• Nếu $\bar{X} < a_o$ ta đi đến bài toán kiểm định giả thiết :

$H_o : a = a_o$ với $K : a < a_o$ ở mức α.

Cũng tương tự như trường hợp trên, người ta cũng chứng minh được rằng : tồn tại tiêu chuẩn để kiểm định giả thiết $H_o : a = a_o$ với $K : a < a_o$ và tiêu chuẩn kiểm định được phát biểu dưới dạng :

– Trường hợp σ đã biết :

Giả thiết H_o bị bác bỏ ở mức α nếu :

$$Z = \frac{(a_o - \bar{X})\sqrt{n}}{\sigma} > x_\alpha \qquad (6.11)$$

Còn nếu $Z < x_\alpha$ thì chấp nhận giả thiết H_o.

x_α tra ở bảng phân phối chuẩn $N(0 ; 1)$ sao cho $\Phi(x_\alpha) = 1 - \alpha$.

– Trường hợp σ chưa biết :

Giả thiết H_o bị bác bỏ ở mức α nếu :

$$Z = \frac{(a_o - \bar{X})\sqrt{n}}{S_n^*(X)} > t_\alpha \qquad (6.12)$$

Còn nếu $Z < t_\alpha$ thì chấp nhận giả thiết H_o ;

trong đó t_α tra như sau :

+ Nếu n > 30 thì t_α tra ở bảng phân phối chuẩn sao cho $\Phi(t_\alpha) = 1 - \alpha$.

+ Nếu n < 30 thì t_α tra ở bảng giá trị tiêu chuẩn Student (bảng tiêu chuẩn một phía) với n – 1 bậc tự do và mức ý nghĩa α.

Ví dụ 6.9. Để kiểm tra chất lượng sản phẩm của một nhà máy, người ta kiểm tra ngẫu nhiên 20 sản phẩm. Kích thước của các sản phẩm là :

Kích thước sản phẩm X_i :	34,8	34,9	35	35,1	35,3
Tần số xuất hiện n_i	2	3	4	5	6

Hãy kiểm định giả thiết :

H_o : a = 35 với K : a > 35 ở mức $\alpha = 0,05$.

Giả sử các quan sát này độc lập và có phân phối chuẩn $N(a\ ;\ \sigma^2)$.

Giải :

Ở đây $\alpha = 0,05$, n = 20, σ chưa biết, ta áp dụng công thức (6.10). Tra bảng phân phối Student ta tìm được t(5%, 19) = 1,73 (tiêu chuẩn một phía).

Tính :

$$\bar{X} = \frac{2 \times 34,8 + 3 \times 34,9 + 4 \times 35 + 6 \times 35,1 + 5 \times 35,3}{20}$$

$$\bar{X} = 35,07$$

$$S_n^{*2} = \frac{1}{19}[2(34,8 - 35,07)^2 + 3(34,9 - 35,07)^2 + 4(35 - 35,07)^2 +$$
$$+ 6(35,1 - 35,07)^2 + 5(35,3 - 35,07)^2] = 0,027$$

$$S_n^* = \sqrt{0,027} \approx 0,16$$

$$Z = \frac{(\bar{X} - a_o)\sqrt{n}}{S_n^*} = \frac{(35,07 - 35)\sqrt{20}}{0,16} \approx 1,96$$

Ta thấy $Z = 1,96 > t_\alpha = 1,73$. Vậy giả thiết H_o bị bác bỏ ở mức α ; nghĩa là a > 35.

6.2.4. So sánh trung bình của hai mẫu độc lập

Ta cần nghiên cứu một tính trạng X của các cá thể trong hai đám đông. Ta chọn mẫu $(X_1, X_2,..., X_n)$ từ đám đông I và một mẫu $(Y_1, Y_2,...,$

Y_m) từ đám đông II. Giả sử hai mẫu này độc lập và có phân phối chuẩn dạng tổng quát tương ứng $N(a_1, \sigma_1^2)$ và $N(a_2, \sigma_2^2)$, nghĩa $EX = a_1$, $EY = a_2$, $DX = \sigma_1^2$, $DY = \sigma_2^2$.

Kiểm định giả thiết $H_o : a_1 = a_2$ với $K : a_1 \neq a_2$ ở mức α.

Giải :

Người ta chứng minh được rằng : tồn tại tiêu chuẩn để kiểm định giả thiết $H_o : p_1 = p_2$ với $K : p_1 \neq p_2$ và tiêu chuẩn này được phát biểu dưới dạng sau :

– Trường hợp phương sai DX, DY đã cho trước :

Giả thiết H_o bị bác bỏ ở mức α nếu :

$$|Z| = \frac{|\bar{X} - \bar{Y}|}{\sqrt{\dfrac{\sigma_1^2}{n} + \dfrac{\sigma_2^2}{m}}} > x_\alpha \qquad (6.13)$$

Còn nếu $|Z| < x_\alpha$ thì chấp nhận giả thiết H_o, trong đó x_α tra ở bảng phân phối chuẩn sao cho $\Phi(x_\alpha) = 1 - \dfrac{\alpha}{2}$.

– Trường hợp DX, DY chưa biết ta phải giả thiết DX = DY :

Giả thiết H_o bị bác bỏ ở mức α nếu :

$$|Z| = \frac{|\bar{X} - \bar{Y}|}{S\sqrt{\dfrac{1}{n} + \dfrac{1}{m}}} > t_\alpha \qquad (6.14)$$

Còn nếu $|Z| < t_\alpha$ thì chấp nhận giả thiết H_o ;

trong đó $S^2 = \dfrac{1}{n + m - 2}\left[\sum_{i=1}^{n}(X_i - \bar{X})^2 + \sum_{i=1}^{m}(Y_i - \bar{Y})^2\right]$

t_α tra như sau :

+ Nếu $n + m > 60$ thì t_α tra ở bảng phân phối chuẩn sao cho $\Phi(t_\alpha) = 1 - \dfrac{\alpha}{2}$.

+ Nếu $n + m < 60$ thì t_α tra ở bảng phân phối Student với $n + m - 2$ bậc tự do và mức α (bảng tiêu chuẩn hai phía).

Ví dụ 6.10. Giả sử $(X_1, X_2,..., X_n)$ là mẫu ngẫu nhiên độc lập từ phân phối chuẩn $N(a_1, \sigma_1^2)$ và $(Y_1, Y_2,..., Y_m)$ là mẫu ngẫu nhiên độc lập từ phân phối chuẩn $N(a_2, \sigma_2^2)$, X và Y là độc lập. Biết rằng $\overline{X} = 130$, $\overline{Y} = 140$, $DX = 80$, $DY = 100$, $n = 40$, $m = 50$. So sánh hai kỳ vọng EX và EY ở mức $\alpha = 0{,}01$.

Giải : Ta đi đến việc kiểm định giả thiết $H_o : a_1 = a_2$ với $K : a_1 \neq a_2$ vì $n + m = 40 + 50 = 90 > 60$, nên t_α tra ở bảng phân phối chuẩn sao cho $\Phi(t_\alpha) = 1 - \frac{\alpha}{2} = 0{,}995$ ta nhận được $t_\alpha = 2{,}58$.

Tính giá trị $|Z|$:

$$|Z| = \frac{|\overline{X} - \overline{Y}|}{\sqrt{\frac{DX}{n} + \frac{DY}{m}}} = \frac{|130 - 140|}{\sqrt{\frac{80}{40} + \frac{100}{50}}} = 5$$

Ta thấy $|Z| = 5 > t_\alpha = 2{,}58$. Vậy bác bỏ giả thiết $H_o : a_1 = a_2$ với $K : a_1 \neq a_2$; nghĩa là trung bình a_1, a_2 là khác nhau.

Ví dụ 6.11. Để đánh giá chất lượng sản phẩm của hai nhà máy, người ta kiểm tra ngẫu nhiên 7 sản phẩm của nhà máy I và 6 sản phẩm của nhà máy II. Kết quả được cho ở bảng sau đây :

X	15,7	10,3	12,6	14,5	12,6	13,8	11,9
Y	12,3	13,7	10,4	11,4	14,9	12,6	

Giả sử X, Y đều có phân phối chuẩn dạng $N(a_1, \sigma^2)$, $N(a_2, \sigma^2)$ nghĩa là $EX = a_1$, $EY = a_2$, $DX = DY = \sigma^2$.

Hãy so sánh hai trung bình của hai mẫu ở mức $\alpha = 0{,}1$.

Giải : Vì $n + m = 7 + 6 = 13 < 60$. Để áp dụng tiêu chuẩn (6.14), ta lần lượt tính các đại lượng :

$$\overline{X} = 13{,}06, \ \overline{Y} = 12{,}55$$

$$S^2 = \frac{1}{11}[(15{,}7 - 13{,}06)^2 + (10{,}3 - 13{,}06)^2 + 2(12{,}6 - 13{,}06)^2 + (14{,}5 - 13{,}06)^2$$
$$+ (13{,}8 - 13{,}06)^2 + (11{,}9 - 13{,}06)^2 + (12{,}3 - 12{,}55)^2 + (13{,}7 - 12{,}55)^2$$
$$+ (10{,}4 - 12{,}55)^2 + (11{,}4 - 12{,}55)^2 + (14{,}9 - 12{,}55)^2 + (12{,}6 - 12{,}55)^2]$$

$S^2 = 2{,}89.$

Tính $|Z|$:

$$|Z| = \frac{|\overline{X} - \overline{Y}|}{S\sqrt{\frac{1}{n} + \frac{1}{m}}} = \frac{|13,06 - 12,55|}{\sqrt{2,89\left(\frac{1}{7} + \frac{1}{6}\right)}} = 0,54$$

tra bảng giá trị tiêu chuẩn Student ta có t(0,1 ; 11) = 1,796. Ta nhận thấy

$$|Z| = 0,54 < t_\alpha = 1,796.$$

Vậy chấp nhận giả thiết H_o, nghĩa là kích thước trung bình của 2 mẫu là như nhau.

Chú ý : Khi nhận được các kết quả thí nghiệm là hai mẫu ngẫu nhiên độc lập $(X_1, ..., X_n)$ và $(Y_1, Y_2, ..., Y_m)$ từ phân phối chuẩn, nhưng chưa biết gì về phương sai DX, DY, để so sánh được sự bằng nhau của hai trung bình EX, EY bằng tiêu chuẩn (6.14) ta phải xét xem hai phương sai DX, DY có bằng nhau hay không. Muốn vậy ta phải giải bài toán kiểm định giả thiết H_o : DX = DY với K : DX ≠ DY ở mức α.

Giải : Giả thiết H_o bị bác bỏ ở mức α nếu :

$$Z = \frac{S_n^{*2}(X)}{S_m^{*}(Y)} > f_{bảng}\left(\frac{\alpha}{2}, n-1, m-1\right) \tag{6.15}$$

trong đó $f_{bảng}\left(\frac{\alpha}{2}, n-1, m-1\right)$ tra ở bảng phân phối F với n – 1, m – 1 bậc tự do và mức α.

Ví dụ 6.12. Để đánh giá chất lượng thức ăn cho gia súc, người ta kiểm tra ngẫu nhiên 10 con lợn được ăn thức ăn loại A và 12 con được ăn thức ăn loại B. Số liệu đo được sự tăng trọng tính ra kg với thời gian 1 tháng cho ở bảng dưới đây :

– Thức ăn loại A :

X_i	3,4	3,5	3,7	3,9
n_i	2	3	4	1

– Thức ăn loại B :

Y_i	3,2	3,4	3,6
m_i	2	2	8

Giả sử các quan sát X_i, Y_i độc lập và có phân phối chuẩn với $EX = a_1$, $EY = a_2$ và phương sai DX, DY chưa biết. Hãy so sánh hai trung bình EX, EY ở mức α = 0,02.

Giải :

Trước hết ta kiểm định giả thiết H_o : DX = DY với K : DX ≠ DY

Tính $\overline{X} = 3{,}6$, $\overline{Y} = 3{,}5$ và phương sai mẫu $S_n^{*2}(X) = 0{,}0267$, $S_m^{*2}(Y) = 0{,}0255$. Từ đó suy ra $Z = \dfrac{S_n^{*2}(X)}{S_m^{*2}(Y)} = \dfrac{0,0267}{0,0255} = 1{,}05$. Tra bảng phân phối F ta tìm được $f_{bảng}(0{,}01, 9, 11) = 4{,}63$. Ta thấy $Z = 1{,}05 < f_{bảng} = 4{,}63$. Vậy chấp nhận giả thiết H_o, nghĩa là DX = DY.

Bây giờ ta trở về bài toán so sánh hai trung bình. Trước hết ta tính :

$$|Z| = \frac{|\overline{X} - \overline{Y}|}{S\sqrt{\frac{1}{n} + \frac{1}{m}}} = \frac{|\overline{X} - \overline{Y}|}{\sqrt{(n-1)S_n^{*2}(X) + (m-1)S_m^{*2}(Y)}}\sqrt{\frac{nm(n+m-2)}{n+m}}$$

$$= \frac{3,6 - 3,5}{\sqrt{9 \times 0,0267 + 11 \times 0,0255}}\sqrt{\frac{10 \times 12(10 + 12 - 2)}{10 + 12}} \approx 1{,}45$$

Tra bảng giá trị tiêu chuẩn phân phối Student ta tìm được $t(0{,}02 ; 20) = 2{,}53$.

Ta nhận thấy $|Z| = 1{,}45 < t_\alpha = 2{,}53$

Vậy chấp nhận giả thiết H_o, nghĩa là chất lượng của hai loại thức ăn đó là tương đương.

6.2.5. So sánh hai trung bình của 2 mẫu phụ thuộc từ phân phối chuẩn

Giả sử $(X_1, X_2,..., X_n)$ và $(Y_1, Y_2,..., Y_m)$ là hai mẫu ngẫu nhiên từ phân phối chuẩn. X, Y là phụ thuộc. Hãy so sánh hai trung bình EX và EY ở mức α.

Để giải quyết vấn đề này ta đưa đến bài toán kiểm định giả thiết H_o : EX = EY với K : EX ≠ EY ở mức α.

Để giải bài toán này ta tiến hành theo các bước sau đây :

– Tính các $d_i = X_i - Y_i$, $i = \overline{1, n}$.

$$\overline{d} = \frac{1}{n}\sum_{i=1}^{n} d_i \text{ và } S_d^2 = \frac{1}{n-1}\sum_{i=1}^{n}(d_i - \overline{d})^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} d_i^2 - \frac{(\sum_{i=1}^{n} d_i)^2}{n}\right)$$

– Kết luận :

Giả thiết H_o bị bác bỏ ở mức α nếu :

$$|Z| = \frac{|\bar{d}\sqrt{n}|}{S_d} > t_\alpha \qquad (6.16)$$

Còn nếu $|Z| < t_\alpha$ thì chấp nhận giả thiết H_o, trong đó t_α tra ở bảng phân phối Student với n – 1 bậc tự do và mức α (bảng tiêu chuẩn hai phía).

Ví dụ 6.13. Cho mẫu quan sát đối với cặp biến ngẫu nhiên X, Y như sau :

X	2	4	7	2	5	6	7	6
Y	8	4	10	5	8	8	9	9

Hãy so sánh hai trung bình của 2 mẫu ở mức α = 0,05.

Giải :

$n = 8$, $d_1 = 2 - 8 = -6$, $d_2 = 4 - 4 = 0$, $d_3 = 7 - 10 = -3$,

$d_4 = 2 - 5 = -3$, $d_5 = 5 - 8 = -3$, $d_6 = 6 - 8 = -2$, $d_7 = 7 - 9 = -2$

$d_8 = 6 - 9 = -3$, $\sum_{i=1}^{8} d_i = -6 + 0 - 3 - 3 - 3 - 2 - 2 - 3 = -22$

$$\bar{d} = \frac{-22}{8} = -2{,}75$$

$$\sum_{i=1}^{8} d_i^2 = 36 + 9 + 9 + 9 + 4 + 4 + 9 = 80$$

$$S_d^2 = \frac{1}{7}\left(80 - \frac{(-22)^2}{8}\right) = 2{,}7857 \text{ ; } S_d = 1{,}669.$$

$$\text{Vậy } |Z| = \frac{|\bar{d}\sqrt{n}|}{S_d} = \frac{2{,}75 \times \sqrt{8}}{1{,}669} \approx 4{,}66.$$

Tra bảng giá trị tiêu chuẩn Student ta tìm được t(5% ; 7) = 2,36. Ta thấy $|Z| = 4{,}66 > t_\alpha = 2{,}36$. Vậy bác bỏ giả thiết H_o, nghĩa là trung bình EX và EY là khác nhau.

Tiêu chuẩn này có thể áp dụng để giải các loại bài toán, ví dụ : cần đánh giá hiệu quả của 1 phương pháp tác động đến các cá thể của một đám đông, hoặc tác động của 1 loại thức ăn gia súc..., mà X_i, $i = \overline{1, n}$, là kết quả đo lần thứ nhất khi chưa chịu tác động của phương pháp đó, hoặc chưa cho ăn loại thức ăn đó và Y_i, $i = \overline{1, n}$ là kết quả thí nghiệm (kiểm tra) chính các cá thể này trong đám đông sau khi chịu tác động của phương pháp đó... Ngoài ra nó còn có thể áp dụng cho bài toán so sánh hai đặc tính nào đó của 1 cá thể trong đám đông. Ví dụ : năng lực học toán X và năng lực học vật lý Y của học sinh ở khối nào đó.

6.2.6. Tiêu chuẩn X^2 (khi bình phương) kiểm định về phân phối

Giả sử đại lượng ngẫu nhiên X nhận giá trị trong không gian S nào đó. Ta chia không gian S thành q phần rời nhau, không nhất thiết phải bằng nhau $S_1, S_2, ..., S_q$ ($S_1 \cup S_2 \cup ... \cup S_q = S$).

Giả sử $(X_1, ..., X_n)$ là mẫu quan sát đối với đại lượng ngẫu nhiên X.

Gọi n_k số các giá trị mẫu X_i của biến ngẫu nhiên X rơi vào bộ phận S_k. Ta có $n_1 + n_2 + ... + n_k = n$.

Và p_k là xác suất để một giá trị x của X rơi vào bộ phận S_k (nghĩa là $p_k = P[\omega : X \in S_k]$).

Hãy kiểm định giả thiết :

$$H_o : p_1 = p_1^o, p_2 = p_2^o, ..., p_q = p_q^o$$

$$K : p_1 \neq p_1^o, p_2 \neq p_2^o, ..., p_q \neq p_q^o$$

ở mức α, trong đó các p_i^o có thể là các con số, hoặc có thể là các phân phối đã cho, chẳng hạn như phân phối chuẩn, phân phối nhị thức, phân phối Poisson v.v.

– Trường hợp các p_i^o, $i = \overline{1, q}$ là những con số. Người ta chứng minh được rằng : tồn tại tiêu chuẩn để kiểm định giả thiết H_o.

Tiêu chuẩn được phát biểu dưới dạng :

Giả thiết H_o bị bác bỏ ở mức α nếu :

$$Z = \sum_{k=1}^{q} \frac{(n_k - np_k^o)^2}{np_k^o} = -n + \sum_{k=1}^{q} \frac{n_k^2}{np_k^o} > C_\alpha \tag{6.17}$$

Còn nếu $Z < C_\alpha$ thì chấp nhận giả thiết H_o.

C_α là hằng số tra ở bảng phân phối χ^2 với $q - 1$ bậc tự do và mức ý nghĩa α.

– Trường hợp các p_k^o là những phân phối đã cho và giả sử $p_k^o = p_k^o(\theta)$, $\theta = (\theta_1, ..., \theta_r)$.

Trong trường hợp này ta phải tiến hành các bước sau :

+ Tìm ước lượng của tham số θ là $\hat{\theta}$ (có thể dùng phương pháp hợp lý cực đại).

+ Tính ước lượng của $p_k^o(\theta)$: $\hat{p}_k^o = p_k^o(\hat{\theta})$.

+ Kết luận :

Giả thiết H_o bị bác bỏ ở mức α nếu :

$$Z = \sum_{k=1}^{q} \frac{(n_k - n\hat{p}_k^o)^2}{n\hat{p}_k^o} = -n + \sum_{k=1}^{q} \frac{n_k^2}{n\hat{p}_k^o} > C'_\alpha \qquad (6.18)$$

Còn nếu $Z < C'_\alpha$ thì chấp nhận giả thiết H_o.

C'_α tra ở bảng phân phối χ^2 với $q - 1 - r$ bậc tự do và mức α.

Ví dụ 6.14. Gieo một con xúc xắc 6000 lần. Số lần xuất hiện các mặt có số chấm bằng 1, 2, 3, 4, 5, 6 tương ứng là :

Số chấm	1	2	3	4	5	6
n_i	1045	942	1016	1064	988	945

Có thể coi con xúc xắc đó là xúc xắc thật không ? Nghĩa là xác suất để mỗi mặt xuất hiện bằng $\frac{1}{6}$. Cho $\alpha = 5\%$.

Giải : Để giải bài toán này ta đi đến kiểm định giả thiết :

$H_o : p_1 = \frac{1}{6}, p_2 = \frac{1}{6}, p_3 = \frac{1}{6}, p_4 = \frac{1}{6}, p_5 = \frac{1}{6}, p_6 = \frac{1}{6}$

$K : p_i \neq \frac{1}{6}, i = \overline{1,6}$.

Theo giả thiết $n = 6000$, $n_1 = 1045$, $n_2 = 942$, $n_3 = 1016$, $n_4 = 1064$, $n_5 = 988$, $n_6 = 945$.

Tính giá trị của Z :

$$Z = \frac{\left(1045 - 6000 \times \frac{1}{6}\right)^2}{6000 \times \frac{1}{6}} + \frac{(942 - 1000)^2}{1000} + \frac{(1016 - 1000)^2}{1000} +$$

$$+ \frac{(1064 - 1000)^2}{1000} + \frac{(988 - 1000)^2}{1000} + \frac{(945 - 1000)^2}{1000} \approx 12{,}91$$

Tra bảng phân phối khi bình phương ta tìm được C(5% ; 5) = 11,07. Ta thấy Z = 12,91 > C_α = 11,07. Vậy giả thiết H_o bị bác bỏ, nghĩa là con xúc xắc đó là xúc xắc giả.

Ví dụ 6.15. Người ta điều tra ngẫu nhiên 1600 gia đình có 4 con. Kết quả cho ở bảng sau đây :

Số con trai trong một gia đình	0	1	2	3	4
Số gia đình	111	367	576	428	118

Hãy kiểm định sự đúng đắn của tập hợp hai giả thiết sau đây ở $\alpha = 0{,}05$:

$H_o^{'}$: "Xác suất để một trẻ em là con trai bằng 1/2".

$H_o^{''}$: "Giới tính các trẻ em trong cùng 1 gia đình có tính độc lập".

Nếu ta quyết định loại bỏ tập hợp hai giả thiết trên thì hãy kiểm định sự đúng đắn của giả thiết $H_o^{''}$ ở mức $\alpha = 0{,}05$.

Giải :

a) Ta hãy kiểm định sự đúng đắn của hai giả thiết $H_o^{'}$ và $H_o^{''}$.

Gọi p_i, i = 0, 1, 2, 3, 4, là xác suất để một gia đình 4 con có i con trai. Theo công thức xác suất nhị thức và giả thiết $H_o^{'}$, $H_o^{''}$ ta có :

$$p_i = C_4^i \left(\frac{1}{2}\right)^i \left(1 - \frac{1}{2}\right)^{4-i} = C_4^i \frac{1}{2^4}, i = 0, 1, 2, 3, 4.$$

Vậy giả thiết $H_o^{'}$ và $H_o^{''}$ tương đương với giả thiết H_o :

$$H_o : p_o = \frac{1}{16}, p_1 = \frac{1}{4}, p_2 = \frac{3}{8}, p_3 = \frac{1}{4}, p_4 = \frac{1}{16}.$$

Áp dụng tiêu chuẩn (6.17) ta có :

$$Z = \frac{(111-100)^2}{100} + \frac{(367-400)^2}{400} + \frac{(576-600)^2}{600} + \frac{(428-400)^2}{400}$$

$$+ \frac{(118-100)^2}{100} \approx 10{,}09.$$

Tra bảng phân phối khi bình phương ta tìm được C(5% ; 4) = 9,49. Ta nhận thấy $Z = 10{,}09 > C_2 = 9{,}49$. Vậy bác bỏ giả thiết H_o, nghĩa là tập hợp 2 giả thiết $H_o^{'}$ và $H_o^{''}$ bị bác bỏ.

b) Bây giờ ta kiểm định sự đúng đắn của giả thiết $H_o^{''}$. Trước hết ta giả sử xác suất sinh con trai là θ, $0 < \theta < 1$. Dưới giả thiết $H_o^{''}$ thì xác suất để trong gia đình 4 con có i con trai bằng $p_i = C_4^i \theta^i (1-\theta)^{4-i}$, i = 0, 1, 2, 3, 4.

Ta kiểm định giả thiết $H_o^{''}$:

$$p_i = C_4^i \theta^i (1-\theta)^{4-i},\ i = 0, 1, 2, 3, 4.$$

– Ta ước lượng θ bằng cách sau : (phương pháp hợp lý cực đại)

$$\hat{\theta} = \frac{1 \times 367 + 2 \times 576 + 3 \times 428 + 4 \times 118}{6400} \approx 0{,}51$$

Thay giá trị θ = 0,51 để tính các $\hat{p}_i$:

$\hat{p}_i = C_4^i (0{,}51)^i (0{,}49)^{4-i}$, i = 0, 1, 2, 3, 4. Ta có :

$\hat{p}_o = 0{,}057$; $\hat{p}_1 = 0{,}23$; $\hat{p}_2 = 0{,}37$; $\hat{p}_3 = 0{,}26$; $\hat{p}_4 = 0{,}07$.

Tính giá trị của Z :

$$Z = \sum_{i=0}^{4} \frac{(n_i - n\hat{p}_i)^2}{n\hat{p}_i} = \frac{(111-96)^2}{96} + \frac{(367-384)^2}{384} + \frac{(576-592)^2}{592} +$$

$$+ \frac{(428-416)^2}{416} + \frac{(118-112)^2}{112} = 4{,}18.$$

Tra bảng phân phối khi bình phương ta tìm được :

C(5% ; 3) = 7,81.

Ta thấy $Z = 4{,}18 < C_2^{'} = 7{,}81$. Vậy chấp nhận giả thiết $H_o^{''}$.

Tóm lại : Xác suất sinh con trai là 0,51 và giới tính của các trẻ em trong 1 gia đình có tính độc lập.

6.2.7. Tiêu chuẩn χ^2 (khi bình phương) kiểm định tính độc lập và tính thuần nhất

a) Kiểm định tính độc lập. Ta xét n phép thử độc lập, trong mỗi phép thử có một và chỉ một trong các biến cố $A_1, A_2, ..., A_r$ và một trong các biến cố $B_1, B_2, ..., B_v$ xảy ra.

Đặt $P(A_i, B_j) = p_{ij}$; $i = \overline{1, r}$; $j = \overline{1, v}$

Ta có thể viết dãy xác suất này dưới dạng bảng sau :

A \ B	B_1	B_2	...	B_v	Tổng
A_1	p_{11}	p_{12}	...	p_{1v}	$p_{1\bullet}$
A_2	p_{21}	p_{22}	...	p_{2v}	$p_{2\bullet}$
...	...	...	...	...	...
A_r	p_{r1}	p_{r2}	...	p_{rv}	$p_{r\bullet}$
Tổng	$p_{\bullet 1}$	$p_{\bullet 1}$	...	$p_{\bullet v}$	1

Ta có $P(A_i) = \sum_{j=1}^{v} p_{ij}$, $P(B_j) = \sum_{i=1}^{r} p_{ij}$. Để xét tính độc lập của A và B ta đi đến bài toán :

Kiểm định giả thiết :

$$H_o : P(A_i B_j) = P(A_i)P(B_j)$$

$$K : P(A_iB_j) \neq P(A_i)P(B_j) \qquad i = \overline{1, r}$$

$$j = \overline{1, v}$$

Gọi X_{ij} là số lần xuất hiện biến cố tích A_iB_j trong n phép thử.

Ta có bảng quan sát sau :

A \ B	B_1	B_2	...	B_v	Tổng
A_1	X_{11}	X_{12}	...	X_{1v}	$X_{1\bullet}$
A_2	X_{21}	X_{22}	...	X_{2v}	$X_{2\bullet}$
...	...	...	...	...	...
A_r	X_{r1}	X_{r2}	...	X_{rv}	$X_{r\bullet}$
Tổng	$X_{\bullet 1}$	$X_{\bullet 2}$	...	$X_{\bullet v}$	n

Người ta chứng minh được rằng : tồn tại tiêu chuẩn để kiểm định giả thiết $H_o : P(A_iB_j) = P(A_i)P(B_j)$ với

$K : P(A_iB_j) \neq P(A_i)P(B_j)$

Tiêu chuẩn được phát biểu dưới dạng :

Giả thiết H_o bị bác bỏ ở mức α nếu :

$$Z = n \times \sum_{i=1}^{r}\sum_{j=1}^{v} \frac{\left(X_{ij} - \frac{X_{i\bullet}X_{\bullet j}}{n}\right)^2}{X_{i\bullet}X_{\bullet j}} > C_\alpha((r-1)(v-1)) \qquad (6.19)$$

nếu $Z < C_\alpha$ thì chấp nhận giả thiết H_o.

C_α tra ở bảng phân phối χ^2 với $(r - 1)(v - 1)$ bậc tự do và mức ý nghĩa α.

Chú ý : – Tiêu chuẩn này áp dụng tốt trong trường hợp mẫu lớn.

– Đảm bảo điều kiện $\frac{X_{i\bullet}X_{\bullet j}}{n} \geq 5$.

Ví dụ 6.16. Để nghiên cứu sự phụ thuộc giữa việc chia học sinh các nhóm để dạy với tình trạng kiến thức của học sinh về môn Thể dục, người ta chia 30 học sinh thành 3 nhóm. Nhóm I gồm các em có năng khiếu đặc biệt ; nhóm II gồm các em học khá ; nhóm III gồm những em học lực trung bình. Sau một đợt huấn luyện, người ta tiến hành kiểm tra trên 30 học sinh ở 3 nhóm đó. Kết quả kiểm tra được cho ở bảng sau :

Tình trạng kiến thức		Trung bình	Khá	Giỏi	Tổng
Nhóm	I	3	3	4	10
	II	4	4	2	10
	III	2	8		10
Tổng		9	15	6	30

Hãy kiểm định giả thiết H_o : "Sự phân nhóm theo trình độ học sinh để dạy độc lập với tình trạng kiến thức của học sinh" ở mức $\alpha = 0,05$.

Giải : Tra bảng phân phối khi bình phương ta tìm được C(5% ; 4) = 9,49. Tính giá trị của Z.

$$Z = n \times \sum_{i=1}^{3}\sum_{j=1}^{3}\frac{\left(X_{ij} - \frac{X_{i\bullet}X_{\bullet j}}{n}\right)^2}{X_{i\bullet}X_{\bullet j}} = 30 \times \left[\frac{\left(3 - \frac{10\times 9}{30}\right)^2}{10\times 9} + \frac{\left(3 - \frac{10\times 15}{30}\right)^2}{10\times 15} + \right.$$

$$+\frac{\left(4 - \frac{10\times 6}{30}\right)^2}{10\times 6} + \frac{\left(4 - \frac{9\times 10}{30}\right)^2}{9\times 10} + \frac{\left(4 - \frac{15\times 10}{30}\right)^2}{15\times 10} + \frac{\left(2 - \frac{6\times 10}{30}\right)^2}{6\times 10} +$$

$$\left. + \frac{\left(2 - \frac{9\times 10}{30}\right)^2}{9\times 1} + \frac{\left(8 - \frac{15\times 10}{30}\right)^2}{15\times 10} + \frac{\left(0 - \frac{6\times 10}{30}\right)^2}{6\times 10}\right] = 7{,}4$$

Ta nhận thấy $Z = 7{,}4 < C(5\% \,;\, 4) = 9{,}49$. Vậy chấp nhận giả thiết H_o ; nghĩa là hai đặc tính A và B là độc lập. Điều này có nghĩa là việc phân chia nhóm học sinh theo nhóm cùng trình độ không ảnh hưởng đến tình trạng kiến thức của học sinh.

Ví dụ 6.17. Trong đám đông ốc sên rừng có hai đặc tính : đặc tính A về màu của vỏ : có hai biến dạng vàng (A_1) và hồng (A_2) ; đặc tính B là số vạch trên vỏ từ 0 vạch đến 5 vạch. Quan sát ngẫu nhiên 169 con ốc sên. Kết quả được cho ở bảng sau :

Số vạch / Màu	0	1	2	3	4	5
Vàng	35	6	13	32	4	25
Hồng	14	2	12	16	0	10

Có thể chấp nhận hay không giả thiết H_o : "trong đám đông ốc sên, màu của vỏ và số vạch trên vỏ độc lập với nhau". Mức kiểm định $\alpha = 0{,}05$.

Giải : Số ốc sên có 1 vạch và 4 vạch quá nhỏ. Để có thể sử dụng tiêu chuẩn khi bình phương đảm bảo độ chính xác theo ý muốn thì số $\frac{X_{i\bullet}X_{\bullet j}}{n} \geq 5$. Vậy trong bảng số liệu ở trên ta phải dồn 1 số cột vào 1 cột. Ví dụ : cột 1 và 2, cột 3 và 4 thành 1 cột, cụ thể là :

Số vạch / Màu	0	1–2	3–4	5	Tổng
Vàng	35	19	36	25	115
Hồng	14	14	16	10	54
Tổng	49	33	52	35	169

Trong ví dụ này r = 2, v = 4.

Tra bảng phân phối khi bình phương ta tìm được (5% ; 3) = 7,81.

Tính giá trị của Z :

$$Z = 169\left[\frac{\left(35 - \frac{49\times115}{169}\right)^2}{49\times115} + \frac{\left(19 - \frac{33\times115}{169}\right)^2}{33\times115} + \right.$$

$$+ \frac{\left(36 - \frac{52\times115}{169}\right)^2}{52\times115} + \frac{\left(25 - \frac{35\times115}{169}\right)^2}{35\times115} + \frac{\left(14 - \frac{49\times54}{169}\right)^2}{49\times54}$$

$$\left. + \frac{\left(14 - \frac{33\times54}{169}\right)^2}{33\times54} + \frac{\left(16 - \frac{52\times54}{169}\right)^2}{52\times54} + \frac{\left(10 - \frac{35\times54}{169}\right)^2}{35\times54}\right]$$

$Z = 2,13$

Ta nhận thấy $Z = 2,13 < C_\alpha = 7,81$. Vậy chấp nhận giả thiết H_o, nghĩa là màu của vỏ con ốc sên độc lập với số vạch ở trên vỏ.

b) Kiểm định tính thuần nhất

Nếu một thí nghiệm có thể thực hiện được trong những điều kiện khác nhau thì nó xác định các phép thử khác nhau : $G_1, G_2, ..., G_v$ (các quy luật xác suất của các phép thử đó khác nhau một cách tiên quyết). Vấn đề đặt ra là tìm xem các xác suất của một hệ sự kiện đầy đủ $A_1, A_2, ..., A_r$ có như nhau trong v phép thử, tức là kiểm định giả thiết.

H_o : "xác suất p_{ij} của biến cố A_i trong phép thử G_j là 1 số p_i không phụ thuộc vào j" ;

hoặc ta còn xét sự thuần nhất của nhiều tập con. Xét đám đông Q các cá thể có các biến dạng $A_1, A_2, ..., A_r$ của một đặc tính A và môt phép phân hoạch đám đông Q thành v đám đông nhỏ $Q_1, Q_2, ..., Q_v$. Trong các đám đông $Q_1, Q_2, ..., Q_v$ được gọi là thuần nhất đối với đặc tính A nếu việc khảo sát đặc tính đó không cho phép ta phân biệt được các đám đông nhỏ.

Chấp nhận hay bác bỏ sự thuần nhất của Q tức là kiểm định giả thiết H_o : "Xác suất p_{ij} của các cá thể có biến dạng A_i của đặc tính A trong đám đông Q_j là một số p_i không phụ thuộc j".

Một cách tổng quát, ta xét hai đặc tính A và B. Chia A thành r mức $A_1, A_2, ..., A_r$ và B thành v mức $B_1, ..., B_v$. Ký hiệu $P(A_i, B_j) = p_{ij}$, $i = \overline{1, r}$; $j = \overline{1, v}$.

Giả thiết H_o về sự thuần nhất đối với B là :

$H_o : p_{11} = p_{12} = ... = p_{1v}$ với k : $p_{ij} \neq p_{i2} \neq ... \neq p_{iv}$, $i = \overline{1, r}$.

....

$p_{r1} = p_{r2} = ... = p_{rv}$

Kiểm định giả thiết H_o ở mức α.

Người ta cũng tính được tiêu chuẩn để kiểm định giả thiết H_o như sau : giả thiết H_o bị bác bỏ ở mức α nếu :

$$Z = n \times \sum_{i=1}^{r}\sum_{j=1}^{v} \frac{\left(n_{ij} - \frac{X_{i\bullet}X_{\bullet j}}{n}\right)^2}{n_i . n_j} > C_\alpha \qquad (6.20)$$

Còn nếu $Z < C_\alpha$ thì chấp nhận giả thiết H_o.

C_α tra ở bảng phân phối χ^2 với $(r - 1)(v - 1)$ bậc tự do và mức α. Ta nhận thấy lời giải của bài toán kiểm định tính thuần nhất giống như lời giải của bài toán kiểm định tính độc lập.

Xét trường hợp đặc biệt r = v = z.

Giả thiết H_o bị bác bỏ ở mức α nếu :

$$Z = n \times \frac{(ad - bc)^2}{(a + c)(b + d)(a + b)(c + d)} > C_\alpha (1) \qquad (6.21)$$

A \ B	B_1	B_2	Tổng
A_1	a	b	a + b
A_2	c	d	c + d
Tổng	a + c	b + d	n

$C_\alpha(1)$ tra ở bảng phân phối χ^2 1 bậc tự do và mức α.

Ví dụ 6.18. Để nghiên cứu hiệu quả của một phương pháp điều trị mới, người ta chia bệnh nhân làm hai nhóm. Nhóm I chữa theo phương pháp cũ, còn nhóm II chữa theo phương pháp mới. Dùng các tiêu chuẩn xác định ta có thể chữa quá trình phát triển bệnh ra làm hai loại : nặng thêm và nhẹ đi. Kết quả theo dõi được cho ở bảng sau :

	Nặng thêm	Nhẹ đi	Tổng
Nhóm I	270	220	490
Nhóm II	120	210	330
Tổng	390	430	820

So sánh hai phương pháp điều trị đó ở mức $\alpha = 0{,}05$.

Giải : Tra bảng phân phối khi bình phương ta tìm được C(5% ; 1) = 3,84

$$\text{Tính } Z = \frac{820(270 \times 210 - 220 \times 120)^2}{390 \times 430 \times 490 \times 330} = 27{,}76$$

Ta nhận thấy $Z = 27{,}7 > C_\alpha = 3{,}84$. Ta bác bỏ giả thiết H_o, nghĩa là tỷ lệ khỏi bệnh của 2 phương pháp điều trị khác nhau. Tỷ lệ khỏi bệnh điều trị theo phương pháp mới cao hơn phương pháp cũ.

Chú ý : Nếu một số các X_{ij} trong bảng là quá nhỏ, sử dụng tiêu chuẩn χ^2 không chính xác, ta thay nó bằng tiêu chuẩn sau đây :

	B_1	B_2	Tổng
A_1	X_{11}	X_{12}	$X_{1\bullet}$
A_2	X_{21}	X_{22}	$X_{2\bullet}$
Tổng	$X_{\bullet 1}$	$X_{\bullet 2}$	n

Giả thiết H_o về tính độc lập của A, B (hoặc tính thuần nhất của 2 nhóm con A_1, A_2) bị bác bỏ ở mức α nếu :

$$\sum_{\xi_{11}} P[X_{11} = \xi_{11}] = \sum_{\xi_{11}} \frac{X_{\bullet 1}!X_{\bullet 2}!X_{1\bullet}!X_{2\bullet}!}{n!\xi_{11}!\xi_{12}!\xi_{21}!\xi_{22}!} < \frac{\alpha}{2} \qquad (6.22)$$

trong tổng trên ta lấy các giá trị ξ_{11} như sau :

Nếu $X_{11}X_{12} - X_{12}X_{21} < 0$ thì lấy các giá trị $\xi_{11} \leq X_{11}$.

Còn nếu $X_{11}X_{22} - X_{12}X_{21} > 0$ thì lấy giá trị $\xi_{11} \geq X_{11}$.

Còn các $\xi_{12}, \xi_{21}, \xi_{22}$ được xác định bởi các đẳng thức sau :

$$\xi_{11} + \xi_{12} = X_{1\bullet},\ \xi_{11} + \xi_{21} = X_{\bullet 1},\ \xi_{11} + \xi_{12} + \xi_{21} + \xi_{22} = n.$$

Ví dụ 6.19. Trong một nhà máy, người ta dùng hai loại máy để sản xuất, trong một khoảng thời gian đã cho (chẳng hạn 1 năm), người ta có các kết quả sau đây đối với số thợ làm việc trên các máy đó.

	Bị tai nạn	Không bị tai nạn	Tổng
Loại 1	5	245	250
Loại 2	1	299	300
Tổng	6	544	550

Có thể kết luận là một trong hai loại máy ít nguy hiểm hơn loại kia không ? Cho $\alpha = 0{,}05$.

Giải :

Ta chọn giả thiết H_o : "Sự nguy hiểm của hai loại máy là như nhau" nói cách khác đám đông nhỏ gồm các thợ làm việc trên hai loại máy đó, thuần nhất với đặc tính bị tai nạn.

Xét tích $X_{11}X_{22} - X_{12}X_{21} = 5 \times 299 - 1 \times 245 > 0$

Khi tính tổng các xác suất trong (6.22) ta lấy giá trị của $\xi \geq 5$, tức là giá trị 5 hoặc 6 (chỉ có thể là 2 giá trị đó thôi, vì tổng ở cột thứ I bằng 6).

$$\text{Vậy} \quad P[X_{11} = 5] = \frac{6!544!250!300!}{550!5!245!1!299!} = 0{,}0628$$

$$P[X_{11} = 6] = \frac{6!544!250!300!}{550!6!244!0!300!} = 0{,}0086$$

$$\text{Vậy} \quad P[X_{11} \geq \xi_{11}] = P[X_{11} = 5] + P[X_{11} = 6] = 0{,}0714.$$

$\frac{\alpha}{2} = 0{,}025$. Ta thấy $P[X_{11} > \xi_{11}] = 0{,}0714 > \frac{\alpha}{2} = 0{,}025$.

Vậy giả thiết H_o được chấp nhận, nghĩa là ta không thể quyết định cấm dùng loại máy thứ nhất hoặc là nên dùng loại máy nào.

6.2.8. Tiêu chuẩn Wilcoxon kiểm định tính thuần nhất của hai mẫu độc lập

Tiêu chuẩn này kiểm định tính thuần nhất của hai mẫu độc lập $(X_1, X_2, ..., X_n)$ và $(Y_1, ..., Y_m)$ với giả thiết X, Y độc lập và có phân phối liên tục. Tiêu chuẩn này thường dùng trong nghiên cứu tâm lý, xã hội học, khoa học giáo dục v.v..

Bài toán đặt ra như sau : Giả sử X có hàm phân phối $F_1(x)$ và Y có hàm phân phối $F_2(x)$; $F_1(x)$ và $F_2(x)$ là liên tục.

Kiểm định giả thiết : $H_o : F_1(x) = F_2(x)$ ở mức α.

$K : F_1(x) \neq F_2(x)$

Năm 1945, Wilcoxon đã tìm được tiêu chuẩn để kiểm định giả thiết trên H_o như sau :

a) Trường hợp kích thước n, m < 25

Để kiểm định giả thiết H_o với đối thiết K ta tiến hành theo các bước sau :

– Dồn hai mẫu X và mẫu Y thành một mẫu chung và sắp xếp chúng theo thứ tự tăng dần.

Giả sử kích thước của mẫu X nhỏ hơn kích thước của mẫu Y, nghĩa là $n < m$ (Nếu $m < n$ thì lại thay Y_i cho X_i). Ta gọi W là tổng các chỉ số thứ tự (được gọi là hạng) của các quan sát X_i trong mẫu chung.

Ta tính được $E(W) = \dfrac{n(n+m+1)}{2}$ và $DW = \dfrac{nm(n+m+1)}{12}$.

– Tìm điểm tiêu chuẩn trên $W_{trên}$ và điểm tiêu chuẩn dưới $W_{dưới}$ theo công thức :

$$W_{trên} = (n + m + 1)n - W_{dưới}$$

trong đó :

$W_{dưới}(Q ; n ; m)$ tìm trong bảng giá trị của hàm Q với $Q = \dfrac{\alpha}{2}$.

– Kết luận :

Nếu $W < W_{dưới}$ hoặc $W > W_{trên}$ thì bác bỏ giả thiết H_o.

Ngược lại, $W_{dưới} < W < W_{trên}$ thì chấp nhận H_o.

Ví dụ 6.20. Kiểm tra ngẫu nhiên 6 học sinh ở lớp thứ nhất và 8 học sinh ở lớp thứ hai bằng 1 bài thi toán. Kết quả cho ở bảng sau :

X	15	23	25	26	28	29		
Y	12	14	18	20	22	24	27	30

Xét xem lớp nào đạt khá hơn. Cho mức kiểm định $\alpha = 0,05$.

Giải :

Gọi $F_1(x)$ là hàm phân phối của X, $F_2(x)$ là hàm phân phối của Y. Ta kiểm định giả thiết :

$H_o : F_1(x) = F_2(y)$ với $K : F_1(x) \neq F_2(y)$ ở mức $\alpha = 0,05$.

Ta dồn 2 mẫu vào 1 mẫu chung và sắp xếp chúng theo thứ tự tăng dần :

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
12	14	15	18	20	22	23	24	25	26	27	28	29	30

Ta có $W = 3 + 7 + 9 + 10 + 12 + 13 = 54$

Tra bảng giá trị của hàm Q với $\alpha = 0,05$, $n = 6$, $m = 8$

ta có $W_{dưới}(0,025 ; 6 ; 8) = 29$.

$$W_{trên} = n(n + m + 1) - W_{dưới} = 6(6 + 8 + 1) - 29 = 61$$

Ta thấy $29 < W < 61$. Vậy chấp nhận giả thiết H_o nghĩa là, phân bố tính trạng kiến thức của học sinh về môn toán của hai lớp là như nhau.

+ Tiêu chuẩn một phía :

– Nếu xu thế $X < Y$, nghĩa là $F_1(x) > F_2(x)$ thì ta đi đến bài toán kiểm định giả thiết :

$H_o : F_1(x) = F_2(x)$ với $K : F_1(x) > F_2(x)$ ở mức α.

Lời giải : Giả thiết H_o bị bác bỏ ở mức α nếu $W < W_{dưới}$.

– Nếu xu thế $X > Y$, nghĩa là $F_1(x) < F_2(x)$, thì ta đưa về bài toán kiểm định giả thiết $H_o : F_1(x) = F_2(x)$ với $K : F_1(x) < F_2(x)$.

Lời giải : Giả thiết H_o bị bác bỏ ở mức α nếu :

$$W > W_{trên} \text{ ; trong đó } W_{trên} = n(n + m + 1) - W_{dưới}$$

b) Trường hợp kích thước mẫu n, m > 25

Vì W có phân phối nhị thức nên theo định lý giới hạn trung tâm của Laplace ta có $P[W < k] \approx \Phi\left(\dfrac{k - E(W)}{\sqrt{DW}}\right)$ với n, m lớn. Φ(x) là hàm phân phối chuẩn N(0 ; 1).

Tiêu chuẩn kiểm định giả thiết $H_o : F_1(x) = F_2(x)$ với $K : F_1(x) \neq F_2(x)$ giống như trường hợp a) nhưng $W_{dưới}$ được chọn như sau :

$$W_{dưới}(Q ; n ; m) = \left[\frac{n(n + m + 1) - 1}{2} - x_2\sqrt{\frac{nm(n + m + 1)}{12}}\right]$$

trong đó [a] là phần nguyên của a.

x_α tra trong bảng chuẩn sao cho $\Phi(x_\alpha) = 1 - \dfrac{\alpha}{2}$.

Còn trong trường hợp tiêu chuẩn 1 phía với đối thiết

$$K : F_1(x) > F_2(x)$$

hoặc $K : F_1(x) < F_2(x)$ thì x_α tra trong bảng phân phối chuẩn sao cho $F(x_\alpha) = 1 - \alpha$.

Ví dụ 6.21. Giả sử cho 2 mẫu $(X_1, ..., X_n)$ và $(Y_1, ..., Y_m)$ độc lập có kích thước n = 40, m = 50. Hãy kiểm định giả thiết :

$$H_o : F_1(x) = F_2(x) \text{ với } K : F_1(x) \neq F_2(x) \text{ ở mức } \alpha = 0{,}05.$$

Giả sử W = 1800.

Giải :

α = 0,05 ta suy ra $\Phi(x_2) = 1 - \dfrac{\alpha}{2} = 0{,}975$. Tra trong bảng phân phối chuẩn ta có $x_\alpha = 1{,}96$.

Tính $W_{dưới}$ như sau :

$$W_{dưới} = \left[\frac{(40 + 50 + 1)40 - 1}{2} - 1{,}96\sqrt{\frac{40 \times 50(40 + 50 + 1)}{12}}\right]$$

$$= 1578$$

$W_{trên} = (m + n + 1)n - W_{dưới} = (40 + 50 + 1)40 - 1578 = 2062.$

Ta thấy $W_{dưới} = 1578 < W = 1800 < W_{trên} = 2062$

Vậy chấp nhận giả thiết H_o.

• Tiêu chuẩn Mann – Whitney (năm 1947) :

Mann–Whitney giải bài toán kiểm định giả thiết H_o dựa trên thống kê U. Ta vẫn giả thiết kích thước n < m.

Đặt $U = nm + \dfrac{n(n+1)}{2} - W$

Ta suy ra $EU = \dfrac{nm}{2}$ và $DU = \dfrac{nm(n+m+1)}{12}$.

Giả thiết H_o bị bác bỏ ở mức α nếu :

$$|Z| = \frac{\left|U - \dfrac{nm}{2}\right|}{\sqrt{\dfrac{nm(n+m+1)}{12}}} > x_\alpha \qquad (6.23)$$

Còn nếu $|Z| < x_\alpha$ thì chấp nhận giả thiết H_o.

Chú ý : Nếu trong mẫu có một số quan sát trùng nhau, để đảm bảo độ chính xác thì các giá trị trùng nhau trong cùng 1 đoạn được gán cùng 1 hạng bằng số trung bình của các hạng của chúng. Trong biểu thức của phương sai của W hoặc phương sai của U trừ đi 1 lượng $\dfrac{\sum K}{n+m-1}$, $K = \dfrac{l^3 - l}{12}$, l là số các giá trị trùng nhau trong cùng 1 đoạn. $\sum K$ là tổng số trong các đoạn khác nhau.

Ví dụ : Cho mẫu :

2 2 2 3 3   4 4 4 4   5   6   7 7 7 8 9

mẫu trên có 4 đoạn có giá trị trùng nhau.

$l_1 = 3$; $K_1 = \dfrac{3^3 - 3}{12} = 2$; $l_2 = 2$, $K_2 = \dfrac{2^3 - 2}{12} = 0{,}5$;

$l_3 = 4$, $K_3 = \dfrac{4^3 - 4}{12} = 5$; $l_4 = 3$, $K_4 = \dfrac{3^3 - 3}{12} = 2$.

$\sum K = K_1 + K_2 + K_3 + K_4 = 2 + 0{,}5 + 5 + 2 = 9{,}5$

6.2.9. Tiêu chuẩn Wilcoxon kiểm định tính thuần nhất của hai mẫu phụ thuộc

Giả sử cho hai mẫu ngẫu nhiên $(X_1, X_2,..., X_n)$ và $(Y_1, Y_2,..., Y_n)$. X, Y là không độc lập, có phân phối liên tục và phân phối của nó có tính chất đối xứng, nghĩa là :

$$P[X < c - x] = P[X \geq c + x]$$

Ví dụ : Trong nghiên cứu về giáo dục, các quan sát X_i có thể được xem như kết quả lần đo thứ nhất của đặc tính A của các cá thể ; còn Y_i là kết quả đo cũng đặc tính đó trong lần đo thứ hai.

Giả thiết H_o : "X và Y có phân bố như nhau".

Ta có thể viết dưới dạng $H_o : P[X \leq Y] = P[X \geq Y]$

Đặt $D_i = Y_i - X_i$. Các cặp (X_i, Y_i) mà $X_i = Y_i$ thì không được kể tới. Gọi m số các cặp (X_i, Y_i) mà $X_i \neq Y_i$, $(m \leq n)$.

Ta sắp xếp các $|D_i|$, $i = \overline{1, m}$ theo thứ tự tăng dần theo độ lớn. $|D_i|$ nhỏ nhất được gán hạng 1, sau đó là hạng 2, ... và số cuối cùng là hạng m. Nếu có 1 số các $|D_i|$ trùng nhau thì những số đó được gán cùng 1 hạng là số trung bình của các hạng của chúng.

Sau đó mỗi hạng được mang dấu "+" nếu ứng với $D_i > 0$ và dấu "−" nếu ứng với $D_i < 0$. Giả thiết H_o được viết lại :

$H_o : P[D_i > 0] = P[D_i < 0]$, (tức là số trung vị bằng 0),

hoặc có thể viết dạng tương đương :

H_o : "số trung vị của $D_i = 0$"

Bài toán kiểm định giả thiết được phát biểu như sau :

Kiểm định giả thiết $H_o : P[D_i > 0] = P[D_i < 0]$

$$K : P[D_i > 0] \neq P[D_i < 0]$$

Ở mức α.

Ký hiệu R_i là hạng của $|D_i|$ và T là tổng của các hạng R_i của các $|D_i|$ mà $D_i > 0$, nghĩa là $T = \sum_{D_i > 0} R_i$.

Wilcoxon đã tìm được tiêu chuẩn để kiểm định giả thiết H_o với đối thiết K ở trên.

Tiêu chuẩn được phát biểu dưới dạng :

a) Trường hợp m < 20

Giả thiết H_o bị bác bỏ ở mức α nếu :

$$T < W_{\frac{\alpha}{2}} \text{ hoặc } T > W_{1-\frac{\alpha}{2}} \qquad (6.24)$$

còn $W_{\frac{\alpha}{2}} < T < W_{1-\frac{\alpha}{2}}$ thì chấp nhận giả thiết H_o, trong đó $W_{\frac{\alpha}{2}}$, $W_{1-\frac{\alpha}{2}}$ tra ở bảng tiêu chuẩn Wilcoxon với m < 20.

• Tiêu chuẩn một phía :

– Nếu X_i có xu thế trội hơn Y_i thì ta đưa về bài toán kiểm định giả thiết :

$$H_o : P[D_i > 0] = P[D_i < 0]$$

$$K : P[D_i > 0] < P[D_i < 0], \text{ ở mức } \alpha.$$

Lời giải :

Giả thiết H_o bị bác bỏ ở mức α nếu :

$$T < W_\alpha \qquad (6.25)$$

– Nếu Y_i có xu thế trội hơn X_i thì ta đưa về bài toán kiểm định giả thiết :

$$H_o : P[D_i > 0] = P[D_i < 0]$$

$$K : P[D_i > 0] > P[D_i < 0], \text{ ở mức } \alpha.$$

Lời giải :

Giả thiết H_o bị bác bỏ ở mức α nếu :

$$T > W_{1-\alpha} \qquad (6.26)$$

b) Trường hợp m > 20

Ta nhận xét sau : Phân phối giới hạn của thống kê T khi $n \to \infty$ là phân phối chuẩn (xem trong [3]).

Vì vậy khi tìm tiêu chuẩn kiểm định giả thiết :

$$H_o : P[D_i > 0] = P[D_i < 0]$$

$$K : P[D_i > 0] \neq P[D_i < 0], \text{ ở mức } \alpha \text{, ta chọn :}$$

$$W_\alpha = \frac{m(m+1)}{4} + x_{\frac{\alpha}{2}}\sqrt{\frac{m(m+1)(2m+1)}{24}} \qquad (6.27)$$

$$W_{1-\alpha} = \frac{m(m+1)}{2} - W_\alpha$$

trong đó $x_{\frac{\alpha}{2}}$ là số tra ở bảng phân phối chuẩn sao cho $\Phi\left(x_{\frac{\alpha}{2}}\right) = 1 - \frac{\alpha}{2}$.

Giả thiết H_o bị bác ở ở mức α nếu :

$$T < W_\alpha \text{ hoặc } T > W_{1-\alpha}$$

• Tiêu chuẩn một phía :

– Bài toán kiểm định giả thiết :

$H_o : P[D_i > 0] = P[D_i < 0]$

$K : P[D_i > 0] < P[D_i < 0]$, ở mức α.

Lời giải : Giả thiết H_o bị bác bỏ ở mức α nếu $T < W_\alpha$.

– Bài toán kiểm định giả thiết :

$H_o : P[D_i > 0] = P[D_i < 0]$

$K : P[D_i > 0] > P[D_i < 0]$, ở mức α.

Lời giải :

Giả thiết H_o bị bác bỏ ở mức α nếu :

$$T < W_{1-\alpha}$$

Ví dụ 6.22. Để đánh giá việc tuyên truyền vận động sinh đẻ có kế hoạch, người ta phỏng vấn ngẫu nhiên 12 người bằng 1 câu hỏi định sẵn. Sau đó mở một đợt huấn luyện về nội dung này. Cuối đợt huấn luyện, người ta kiểm tra lần thứ hai cũng trên 12 người này. Kết quả hai lần kiểm tra được cho ở bảng sau :

Số TT	1	2	3	4	5	6	7	8	9	10	11	12
Điểm KT lần I	3	8	5	4	2	1	6	7	1	3	4	2
Điểm KT lần 2	4	8	6	2	6	3	6	10	6	10	7	8
Hiệu số điểm	1	0	1	–2	4	2	0	3	5	7	3	6
Hạng R_i	1,5		1,5	–3,5	7	3,5		5,5	8	10	5,5	9

$$T = 1,5 + 1,5 + 7 + 3,5 + 5,5 + 8 + 10 + 5,5 + 9 = 51,5$$

$$m = 10,\ \alpha = 0,05\ ;\ W_{1-\alpha} = 44 \text{ suy ra } T > W_{1-\alpha}$$

Vậy bác bỏ giả thiết H_o ; nghĩa là điểm đợt hai cao hơn đợt một, tức là đợt huấn luyện đó có kết quả làm chuyển biến đáng kể nhận thức của học viên về vấn đề kế hoạch hóa gia đình.

Ví dụ 6.23. Vào đầu năm học giáo viên, kiểm tra trình độ ngữ pháp của học sinh bằng một bài chính tả. Sau đó giáo viên luyện tập cho học sinh về vấn đề đó trong một tháng. Hết đợt học, giáo viên cho học sinh làm bài kiểm tra đợt hai. Kết quả hai lần kiểm tra được cho ở bảng sau :

STT	Số lỗi lần I	Số lỗi lần 2	Hiệu số D_i	Hạng của $\lvert D_i \rvert$	R_i
1	7	0	–7	25	–25,0
2	8	2	–6	23	–23,0
3	6	2	–4	16,5	–16,5
4	4	5	1	2,5	2,5
5	2	4	2	7,5	7,5
6	5	3	–2	7,5	–7,5
7	3	1	–2	7,5	–7,5
8	10	4	–6	23,0	–23,0
9	8	5	–3	12,5	–12,5
10	7	2	–5	20,0	–20,0
11	6	4	–2	7,5	–7,5
12	10	5	–5	20,0	–20,0
13	3	3	0		o
14	1	4	3	12,5	12,5
15	5	1	–4	16,5	–16,5
16	8	0	–8	26,0	–26,0
17	7	4	–3	12,5	–12,5
18	6	4	–2	7,5	–7,5
19	4	5	1	2,5	2,5
20	10	6	–4	16,5	–16,5
21	9	7	–2	7,5	–7,5
22	0	3	3	12,5	12,5
23	5	6	1	2,5	2,5
24	7	2	–5	20,0	–20,0
25	8	2	–6	23,0	–23,0
26	3	4	1	2,5	2,5
27	5	9	4	16,5	16,5

Xét xem đợt huấn luyện ngữ pháp có hiệu quả tốt không ? Cho $\alpha = 0,05$.

Giải :

Ta đưa về kiểm định giả thiết $H_o : P[D_i > 0] = P[D_i < 0]$

$K : P[D_i > 0] < P[D_i < 0]$

Tính giá trị của T :

$$T = \sum_{D_i > 0} R_i = 2,5 + 7,5 + 12,5 + 2,5 + 12,5 + 2,5 + 2,5 + 16,5 = 59$$

m = 26, $\alpha = 0,05$. Tra bảng ta tìm được $x_\alpha = 1,64$.

$$W_\alpha = \frac{26(26+1)}{4} - 1,64\sqrt{\frac{26(26+1)(52+1)}{24}} \approx 110,93$$

Ta thấy $W_\alpha > T = 59$. Vậy bác bỏ giả thiết H_o, nghĩa là số lỗi chính tả đợt kiểm tra lần 2 ít hơn lần 1. Vậy đợt huấn luyện có hiệu quả tốt.

6.2.10. Tiêu chuẩn Cônmôgôrốp (kiểm định về phân phối)

Tiêu chuẩn này giúp ta tìm được hàm phân phối F_o trên cơ sở mẫu ngẫu nhiên $(X_1, X_2,..., X_n)$ đã cho. Tìm F_o bằng cách kiểm định giả thiết :

$$H_o : F(x) = F_o(x)$$

Bây giờ ta phát biểu bài toán như sau :

Giả sử $(X_1, X_2,..., X_n)$ là mẫu ngẫu nhiên từ phân phối F(x). Hàm phân phối mẫu của X là $F_n(x) = \frac{m}{n}$; $-\infty < x < +\infty$ trong đó m là số các $X_i < x$, n kích thước mẫu.

Giả sử $F_o(x)$ là hàm phân phối lý thuyết của X.

Hãy kiểm định giả thiết :

$H_o : F(x) = F_o(x)$

$K : F(x) \neq F_o(x)$

Với giả thiết $F_o(x)$ là hàm liên tục :

Đặt $D_m = \sup\limits_{-\infty < x < +\infty} |F_n(x) - F_o(x)|$

Cômôgônốp chứng minh được :

$$\lim_{n\to\infty} P[\sqrt{n}D_n(x) < x] = \sum_{k=-\infty}^{+\infty} (-1)^k e^{-2k^2x^2} = K(x)$$

Giá trị hàm K(x) được cho ở phần Phụ lục.

Tiêu chuẩn kiểm định giả thiết H_o được phát biểu dạng :

Giả thiết H_o bị bác bỏ ở mức α nếu :

$$\sqrt{n}D_n > C_\alpha \tag{6.28}$$

trong đó C_α tra trong bảng giá trị của hàm K(x) sao cho $K(C_\alpha) = 1 - \alpha$.

Nếu $n \le 100$ có thể sử dụng biểu thức (6.2) trong bảng thống kê toán Bônsep L.N và Smiêcnốp N.V (Nhà XB.M.1965) để tính D_α sao cho $P[D_n(x) > D_\alpha] = \alpha$.

Trong thực hành, để tính D_n ta sử dụng công thức sau đây :

$$D_n = \max(D_n^-, D_n^+)$$

trong đó : $$D_n^+ = \max_{1\le i\le n}\left(\frac{i}{n} - F_o(X_i)\right) = \max_{1\le i\le n}(F_n(X_{(i+1)}) - F_o(X_{(i)}))$$

$$D_n^- = \max_{1\le i\le n}\left(F_o(X_{(i)}) - \frac{i-1}{n}\right) = \max_{1\le i\le n}(F_o(X_{(i+1)}) - F_n(X_{(i)}))$$

$X_{(1)} < X_{(2)} < \ldots < X_{(n)}$ là mẫu thứ tự được lập từ $(X_1, X_2,\ldots, X_n)$.

Ví dụ 6.23. Kết quả đo 1000 chi tiết máy, được làm tròn với độ chính xác 0,5mm và được cho ở bảng sau đây, trong đó n_i là tần số xuất hiện $X_{(i)}$. Hãy dùng tiêu chuẩn Cônmôgôrốp để kiểm định giả thiết :

H_o : "phân phối của các quan sát là phân phối chuẩn N(100 ; 25 ; 1)".

i	$X_{(i)}$	n_i
1	98,0	21
2	98,5	47
3	99,0	87
4	99,5	158
5	100,0	181

i	$X_{(i)}$	n_i
6	100,5	201
7	101,0	142
8	101,5	97
9	102,0	41
10	102,5	2

Giải : Đặt $Y = \dfrac{X - E(X)}{\sqrt{DX}}$

Nếu X có phân phối chuẩn $N(a ; \sigma^2)$ thì Y có phân phối chuẩn N(0 ; 1) vậy ta có thể giả thiết :

$$F_o(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{u^2}{2}} du$$

Xác định hàm phân phối mẫu :

$$F_n(X_{(i)}) = \frac{1}{1000} \left(\sum_{j=1}^{i-1} n_j + 0,5 n_i\right)$$

Xây dựng bảng :

i	$X_{(i)} - 100,25$	$F_o(X_{(i)})$	$F_n(X_{(i)})$	$F_o(X_{(i)}) - F_n(X_{(i)})$
1	–2,25	0,0123	0,0105	0,0018
2	–1,75	0,0401	0,0445	–0,0044
3	–1,25	0,1056	0,1115	–0,00059
4	–0,75	0,2266	0,2340	–0,00074
5	–0,25	0,4013	0,4035	–0,00022
6	+0,25	0,5987	0,5945	0,0042
7	0,75	0,7734	0,7660	0,0074
8	1,25	0,8954	0,8855	$\boxed{0,0089}$
9	1,75	0,9599	0,9545	0,0054
10	2,25	0,9877	0,9875	0,0002

Khi đó $D_n(x) = \max\limits_{1 \le i \le 10} |F_o(X_{(i)}) - F_n(X_{(i)})| = 0,0089$ và

$\sqrt{n} D_n(X) = \sqrt{1000} \times 0,0089 = 0,281$; $C_{0,01} = 1,63$ ta thấy

$\sqrt{n} D_n = 0,281 < C_\alpha = 1,63$. Vậy ta chấp nhận giả thiết H_o.

6.2.11. Tiêu chuẩn Cônmôgôrốp – Smiêcnốp kiểm định tính thuần nhất của hai mẫu

Cho hai mẫu ngẫu nhiên độc lập $(X_1, X_2,..., X_n)$ và $(Y_1, Y_2,..., Y_m)$. Hàm phân phối của X là $F_1(x)$, của Y là $F_2(x)$. Giả thiết X, Y là độc lập ; $F_1(x)$, $F_2(x)$ là hàm phân phối liên tục.

Hãy kiểm định giả thiết :

$$H_o : F_1(x) = F_2(x) \text{ với } K : F_1(x) \neq F_2(x) \text{ ở mức } \alpha.$$

Đặt $D^*_{nm} = \sup\limits_{-\infty<x<+\infty} |F_{1n}(x) - F_{2m}(x)|$

trong đó $F_{1n}(x)$, $F_{2m}(x)$ là phân phối mẫu của X, của Y tương ứng. Với giả thiết H_o, Smiêcnốp chứng được rằng :

$$\lim_{n,m\to\infty} P\left[\left(\frac{nm}{n+m}\right)^{\frac{1}{2}} D^*_{n,m} < x\right] = K(x)$$

Do đó với n, m đủ lớn ta có thể tìm x_α sao cho :

$$P\left[\left(\frac{nm}{n+m}\right)^{\frac{1}{2}} D^*_{nm} > x_\alpha\right] = \alpha.$$

Số x_α tra ở bảng giá trị của hàm K(x) sao cho $K(x_\alpha) = 1 - \alpha$.

Vậy tiêu chuẩn kiểm định giả thiết H_o là :

Giả thiết H_o bị bác bỏ ở mức α nếu :

$$\left(\frac{nm}{n+m}\right)^{\frac{1}{2}} D^*_{nm} > x_\alpha$$

nếu có dấu bất đẳng thức ngược lại thì chấp nhận H_o.

x_α tra ở bảng giá trị của hàm K(x) sao cho $K(x_\alpha) = 1 - \alpha$.

Ví dụ 6.24. Hai máy cùng sản xuất ra một loại phụ tùng. Để kiểm định giả thiết H_o : "cả hai máy đó cho ta các phụ tùng cùng loại" (có cùng độ chính xác về kích thước). Người ta chọn trong mỗi đám sản phẩm của mỗi máy 60 phụ tùng và kết quả kiểm tra kích thước của mỗi loại được cho trong bảng sau đây. Hãy kiểm định giả thiết H_o ở mức $\alpha = 0,05$.

x	Số các $X_i < x$	Số các $Y_i < x$	$F_{1n}(x)$	$F_{2m}(x)$	$\lvert F_{1n}(x) - F_{2m}(x)\rvert$
71,95	1	0	0,1067	0	0,0167
72,11	1	1	0,1067	0,0167	0,0000
73,12	1	2	0,1067	0,0333	0,0166
...	....	...	...	....	...
72,54	46	52	0,7667	0,8667	0,1000
72,55	46	56	0,7667	0,9333	$\boxed{0,1667}$
72,56	49	57	0,8167	0,9500	0,1333
72,58	51	57	0,8500	0,9500	0,1000
....	...	...	...	.....	....
72,59	57	60	0,9500	1,000	0,0500
72,70	58	60	0,9667	1,000	0,0333
72,72	59	60	0,9833	1,000	0,0167
72,73	60	60	1,000	1,000	0,0000

Giải : Nhìn trong bảng số liệu ta thấy trong cột $F_{1n}(x) - F_{2m}(x)$ có số 0,1667 là lớn nhất. Vậy $D^*_{nm} = 0,1667$.

Từ đó suy ra :

$$\left(\frac{nm}{n+m}\right)^{\frac{1}{2}} D^*_{nm} = 0,9130$$

Tra trong bảng giá trị của hàm K(x) ta tìm được $x_\alpha = 1,36$.

Ta thấy $\left(\frac{nm}{n+m}\right)^{\frac{1}{2}} D^*_{nm} < x_\alpha = 1,36$. Vậy ta chấp nhận giả thiết H_o, nghĩa là "kích thước của các sản phẩm do hai nhà máy sản xuất ra là như nhau".

Bài tập chương VI

1. Đợt kiểm tra sức khỏe của trẻ em ở nhà trẻ, người ta khám ngẫu nhiên 100 cháu thấy có 20 cháu có hiện tượng còi xương do suy dinh dưỡng. Gọi p là xác suất để 1 trẻ em mắc chứng bệnh còi xương. Hãy kiểm định giả thiết $H_o : p = 0,15$ với $K : p \neq 0,15$ ở mức $\alpha = 5\%$.

2. TỦ lệ phế phẩm trong một kiện hàng được ghi là 0,02. Kiểm tra ngẫu nhiên 480 sản phẩm từ kiện hàng.thấy có 12 phế phẩm. Hỏi tỷ lệ phế phẩm ghi trên kiện hàng có đúng không ? Tại sao ? Cho mức kiểm định $\alpha = 0,05$.

3. Người ta cân ngẫu nhiên 10 trẻ em hai tuổi. Kết quả được cho ở bảng sau (tính ra kg) :

Khối lượng X_i	12,3	12,5	12,8	13,0	13,5
Tần số n_i	1	2	4	2	1

Giả sử các quan sát X_i có phân phối chuẩn $N(a ; \sigma^2)$. σ chưa biết. Hãy kiểm định giả thiết $H_o : a = 12$ với $K : a \neq 12$ ở mức 0,05.

4. Cho mẫu quan sát đối với biến ngẫu nhiên X từ phân phối chuẩn với độ lệch chuẩn $\sigma = 5,2$ và kích thước mẫu $n = 100$, trung bình mẫu $\bar{X} = 27,56$.

Hãy kiểm định giả thiết $H_o : a = 26$ với $K : a \neq 26$ ở mức $\alpha = 5\%$.

5. Sau một đợt bồi dưỡng nghiệp vụ sư phạm người ta kiểm tra ngẫu nhiên 70 học viên. Kết quả cho ở bảng sau (thang điểm bậc 10) :

X_i	5	6	7	8	9	10
n_i	5	10	15	20	12	8

Giả sử các X_i tuân theo phân phối chuẩn $N(a ; \sigma^2)$. Hãy kiểm định giả thiết $H_o : a = 8$ với $K : a \neq 8$ ở mức $\alpha = 0,05$.

6. Để kiểm tra các chất lượng sản phẩm do hai nhà máy sản xuất, người ta kiểm tra ngẫu nhiên 10 sản phẩm do nhà máy I sản xuất và 12 sản phẩm do nhà máy II sản xuất. Kết quả như sau :

Kích thước X_i	3,4	3,5	3,7	3,9
Tần số n_i	2	3	4	1

Kích thước Y_i	3,2	3,4	3,6
Tần số m_i	2	2	8

Giả sử các X_i, Y_i độc lập có phân phối chuẩn dạng tổng quát và $DX = DY$.

Hãy kiểm định giả thiết H_o : $EX = EY$ với K : $EX \neq EY$ ở mức $\alpha = 5\%$.

7. Muốn so sánh khối lượng óc ở những người trên và dưới 50 tuổi, người ta đã khảo sát một mẫu được ghi trong bảng sau đây (các dạng khối lượng được nhóm thành các lớp cách nhau 50 g, mỗi lớp được xác định bởi trung điểm của nó).

Tuổi	1175	1225	1275	1325	1375	1425	1475
Trên 50	6	15	27	25	28	18	8
Dưới 50	15	36	42	50	54	44	24

Ta có thể cho là khối lượng trung bình của óc người trên 50 tuổi và người dưới 50 tuổi là như nhau không ? cho $\alpha = 0,05$.

8. Người ta điều tra ngẫu nhiên 250 người ở xã A thấy có 140 nữ và 160 người ở xã B thấy có 80 nữ. Hãy so sánh tỷ lệ nữ ở hai xã với mức $\alpha = 0,05$.

9. Áp dụng hai phương pháp gieo hạt. Theo phương pháp A gieo 180 thì có 150 hạt nảy mầm ; theo phương pháp B gieo 256 hạt thấy có 160 hạt nảy mầm. Hãy so sánh hiệu quả của hai phương pháp ở mức $\alpha = 0,05$.

10. Để so sánh năng lực học toán và vật lý của học sinh, người ta kiểm tra ngẫu nhiên 8 em bằng hai bài toán và vật lý. Kết quả được cho ở bảng dưới đây (X là điểm toán, Y là điểm vật lý) :

X	15	20	16	22	24	18	20	14
Y	15	22	14	25	29	20	24	16

Giả sử X, Y có phân phối chuẩn $N(a_1, \sigma^2)$ và $N(a_2, \sigma^2)$ tương ứng. Hãy so sánh điểm trung bình a_1, a_2 ở mức $\alpha = 5\%$.

11. Trên hai cách phân tích khối lượng trong cùng một trình tự, người ta cân 10 mẫu hóa chất. Kết quả như sau :

X_i	25	30	28	50	20	40	32	36	42	38
Y_i	28	31	26	52	24	36	33	35	45	40

Giả sử các X, Y tuân theo phân phối chuẩn.

Hãy so sánh kết quả phân tích ở mức $\alpha = 0{,}01$.

12. Để xét phân bố tình trạng kiến thức của học sinh ở hai khối A và B, người ta tiến hành kiểm tra ngẫu nhiên 6 em ở khối A và 8 em ở khối B. Kết quả như sau :

X_i	3	4	6	10	13	17		
Y_i	1	2	5	8	16	20	22	23

Hãy kiểm định giả thiết $H_o : F_1(x) = F_2(x)$ với $K : F_1 \neq F_2(x)$ ở mức $\alpha = 5\%$, trong đó $F_1(x)$ là hàm phân phối của X, $F_2(x)$ là hàm phân phối của Y.

13. Áp dụng hai phương pháp I và II để sản xuất cùng 1 loại sản phẩm, người ta cân ngẫu nhiên 6 sản phẩm được sản xuất theo phương pháp I và 9 sản phẩm được sản xuất theo phương pháp II.

Kết quả là :

X_i	0,2	0,3	0,5	0,8	1,0	1,3			
Y_i	0,1	0,4	0,6	0,7	1,4	0,9	1,7	1,8	1,9

Hãy kiểm định giả thiết H_o : "Hiệu quả của hai phương pháp đó là như nhau", ở mức $\alpha = 5\%$.

14. Người ta nói rằng, những người ăn một lượng lớn cam quýt có số lần mắc bệnh cúm ít hơn người bình thường. Qua 5 năm thực hiện trên 12 người. Kết quả ghi được ở bảng sau :

5 năm bình thường	10	7	12	8	14	19	12	15	13	8	12	14
5 năm ăn lượng lớn cam, quýt	2	3	2	2	5	2	2	2	12	10	7	3

Theo anh (chị) điều nhận định đó có đúng không ? Cho mức kiểm định $\alpha = 0{,}01$.

15. Để so sánh chất lượng lốp xe, người ta đo số km đi được của từng loại lốp. Kết quả thực nghiệm trên hai loại lốp như sau :

Loại Y	41	50	33	59	46	54	58	53	54	55	59
Loại X	38	54	30	35	36	50	52	45	47	46	40

Hãy kiểm định giả thiết H_o : "Loại lốp (X) có chất lượng giống như loại lốp (Y)" ở mức $\alpha = 5\%$.

16. Ta xét hai cặp gen alen A, a và B, b, trong đó cả A và B đều có tính trội ; nhưng bây giờ ta khảo sát sự giao của hai dị hợp tử kép AaBb. Nếu giả thiết A và a cũng như B và b có khả năng sống sót như nhau và có sự độc lập của các đặc tính thì xác suất của 4 kiểu hình [AB], [Ab], [aB], [ab] bằng 9/16, 3/16, 3/16, 1/16.

Hãy kiểm định giả thiết H_o ở mức $\alpha = 0{,}05$.

$$H_o : P([AB]) = \frac{9}{16} \text{ ; } P([Ab]) = 3/16,\ P([aB]) = 3/16,\ P([ab]) = \frac{1}{16}$$

nếu được cho ở bảng sau đây :

Kiểu gen	[AB]	[Ab]	[aB]	[ab]
Tần số	328	122	77	33

17. Để dự báo sâu bệnh của cánh đồng ngô, người ta kiểm tra ngẫu nhiên 500 hốc, mỗi hốc hai cây. Kết quả kiểm tra như sau :

Số cây bị bệnh	2 cây bị bệnh	1 cây bị bệnh	0 cây bị bệnh
Số hốc	73	185	242

Hỏi tình hình sâu bệnh của ngô như thế đã cần báo động chưa ? Cho mức kiểm định $\alpha = 0{,}05$.

18. Trong kỳ thi tốt nghiệp đại học, hai nhóm thí sinh I và II đã đạt kết quả sau đây :

	I	II
Đỗ	92	119
Trượt	127	108

Hãy kiểm định giả thiết H_o : "Kết quả thi của thí sinh ở hai nhóm I và II là như nhau". Cho mức kiểm định $\alpha = 5\%$.

19. Để khảo sát xem màu của mắt và tóc có phải là các đặc tính độc lập hay không, người ta quan sát một mẫu gồm 3200 người và có kết quả sau đây :

		Màu tóc		
		Vàng	Nâu	Đen và hung
Mắt	Xanh da trời	872	380	112
	Xanh lá cây, Nâu	500	815	521

- Từ kết quả đó suy ra được điều gì ? cho mức kiểm định $\alpha = 0,05$.

20. Nghiên cứu tình hình lao động nông nghiệp ở vùng (X) người ta lấy số liệu ở hai năm 1975 và 1995 như sau :

	1975	1995
Số lao động nông nghiệp	28300	7840
Số lao động chung khác	10000	4260

Có thể nói gì về lao động nông nghiệp ở 2 thời điểm trên. Cho mức kiểm định $\alpha = 0,05$.

Hướng dẫn và trả lời bài tập chương VI

1. $n = 100$, $p_o = 0,15$, $X = 20$, $x_\alpha = 1,96$; $Z = 1,4 < x_\alpha = 1,96$
Chấp nhận H_o, tức là tỷ lệ còi xương của trẻ em là 15%.

2. $n = 480$, $X = 20$, $p_o = 0,02$, $x_\alpha = 1,96$, $Z = 0,78$.
Chấp nhận H_o, tức là tỷ lệ phế phẩm của kiện hàng là 0,02.

3. $S_n^{*2}(X) = 0,11111$, $S_n^*(X) = 0,33333$, $t(5\% ; 9) = 2,26$
$|Z| = 7,59$. Bác bỏ giả thiết H_o.

4. $Z = 3 > x_\alpha = 1,96$. Bác bỏ giả thiết H_o.

5. $S_n^{*2}(X) = 52,30$, $S_n^*(X) = 7,23$, $x_\alpha = 1,96$
$|Z| = 0,37$. Chấp nhận giả thiết H_o.

6. $\bar{X} = 3,6$; $\bar{Y} = 3,5$, $t(5\% ; 200) = 2,09$, $|Z| = 1,45$.

$|Z| < t_\alpha = 2,09$, chấp nhận giả thiết H_o.

7. Chấp nhận giả thiết H_o "khối lượng óc của người trên và dưới 50 tuổi là như nhau".

8. $x_\alpha = 1,65$, $Z = 1,188$, $Z < x_\alpha$; chấp nhận giả thiết H_o tức là số nữ ở xã A và số nữ ở xã B có tỷ lệ tương đương.

9. $x_\alpha = 1,65$, $Z = 4,725$. $Z > x_\alpha$. Bác bỏ giả thiết H_o, tức là phương pháp A có hiệu quả hơn phương pháp B.

10. $\sum d_i = -16$, $\sum d_i^2 = 66$, $S_d^2 = 4,85$; $t(5\% ; 7) = 2,36$

$Z = 2,57$; $Z > t_\alpha$; Bác bỏ giả thiết H_o.

11. $\sum d_i = 9$, $\sum d_i^2 = 65$, $S_d^2 = 6,32$; $Z = 1,1338$; $t(1\% ;9) = 3,25$; $Z < t_\alpha$. Chấp nhận giả thiết H_o.

12. $n = 6$; $m = 8$ $(n < m)$; $W = 41$; $W_{dưới} = 29$; $W_{trên} = 61$ ta có $W_{dưới} < W < W_{trên}$. Chấp nhận giả thiết H_o.

13. $W = 39$, $\alpha = 0,05$; $W_{dưới} = 31$; $W_{trên} = 65$.

$W_{dưới} < W < W_{trên}$. Chấp nhận giả thiết H_o.

14. Bác bỏ giả thiết H_o, tức là người ăn cam, quýt với số lượng lớn thì ít bị bệnh cúm hơn người bình thường.

15. Bác bỏ giả thiết H_o ; tức là loại lốp (X) có chất lượng tốt hơn loại (Y).

16. $C(5\% ; 3) = 7,81$, $Z = 10,87 > C_\alpha = 7,81$. Bác bỏ giả thiết H_o. Điều đó có nghĩa là số liệu điều tra không phù hợp với giả thiết H_o.

17. Bác bỏ giả thiết H_o. Bệnh ngô đang có lây lan. Do đó cần có biện pháp báo động ngăn chặn bệnh lây lan.

18. $C(5\% ; 1) = 3,84$; $z = 4,85 > C_\alpha = 3,84$. Bác bỏ giả thiết H_o. Tỷ lệ đỗ ở hai khối khác nhau.

19. $C(5\% ; 2) = 6$, $n = 3200$; $Z = 463,95 > C_\alpha = 6$. Bác bỏ giả thiết H_o, nghĩa là màu của tóc và màu của mắt có liên hệ, phụ thuộc nhau.

20. $n = 50400$, $C_\alpha = 3,84$, $Z = 375,06 > C_\alpha$; bác bỏ giả thiết H_o, tức là tỷ lệ lao động nông nghiệp trong năm 1975 cao hơn 1995.

Chương VII

PHÂN TÍCH PHƯƠNG SAI VÀ PHÂN TÍCH HỒI QUY

Những nội dung chính trong chương :

* *Phân tích phương sai theo 1 yếu tố.*
* *Phân tích phương sai theo 2 yếu tố.*
* *Hồi quy 1 biến, hồi quy tuyến tính nhiều biến.*
* *Hệ số tương quan hai biến, hệ số tương quan bội, hệ số tương quan riêng.*

7.1. ĐẶT VẤN ĐỀ

Trong nhiều bài toán kiểm định giả thiết có quan hệ tới giá trị trung bình của phân phối chuẩn và đặc biệt là giả thiết tuyến tính. Giả sử dãy $X_1, X_2, \ldots, X_n$ độc lập, có phân phối chuẩn với kỳ vọng $EX_i = \xi_i$ và phương sai như nhau, $DX_i = \sigma^2$, $i = \overline{1, n}$. Các trung bình ξ_i thỏa mãn phương trình

$\sum_{j=1}^{n} a_{ij}\xi_j = b_i$, $i = 1, 2, \ldots, b_i$; $i = \overline{1, n}$, là những hằng số chưa biết.

Giả thiết H_o : $\sum_{j=1}^{n} a_{ij}\xi_j = b_i$; $i = r + 1, \ldots, s$ $(r \le s < n)$. (Giả thiết H_o nói về giả thiết của tổ hợp tuyến tính của các trung bình ξ_i) được gọi là giả thiết tuyến tính. Trong trường hợp đặc biệt nó là giả thiết về các trung bình ξ_i, $i = 1, 2\ldots$ Người ta chứng minh được rằng tồn tại tiêu chuẩn để kiểm định giả thiết tuyến tính H_o.

Bài toán phân tích phương sai và phân tích hồi quy (tuyến tính) là mô hình đặc biệt của bài toán kiểm định giả thiết tuyến tính. Song về mặt ứng dụng của nó lại rất lớn và quan trọng, cho nên ta xét riêng nó trong chương này. Để độc giả không nghiên cứu chuyên về môn toán có thể hiểu được thực chất vấn đề và biết cách vận dụng vào lĩnh vực nghiên cứu của mình, chúng tôi giới thiệu sơ bộ như sau :

Trong những lĩnh vực nghiên cứu về : sinh vật, nông nghiệp, lâm nghiệp, kinh tế giáo dục .v.v... ta quan tâm tới năng suất của loại ngũ cốc nào đó, chất lượng sản phẩm do nhà máy sản xuất ra, chất lượng học tập của học sinh v.v.. Muốn đạt được yêu cầu đó ta phải khảo sát sự ảnh hưởng của các yếu tố có ảnh hưởng đến vấn đề đang xét. Bài toán phân tích phương sai giúp ta giải quyết vấn đề đó, nghĩa là đánh giá sự ảnh hưởng của các yếu tố đến năng suất của ngũ cốc, đến chất lượng sản phẩm, đến chất lượng học tập của học sinh v.v.. Quan điểm của phân tích phương sai là ở chỗ khai triển phương sai của biến ngẫu nhiên X thành tổng của những phương sai của những biến ngẫu nhiên thành phần độc lập, mà mỗi cái trong chúng đặc trưng sự ảnh hưởng của yếu tố này hay yếu tộ khác hoặc tác dụng chung giữa chúng. Sự so sánh những phương sai này cho phép ta đánh giá sự tồn tại ảnh hưởng của các yếu tố đối với biến ngẫu nhiên X đang xét.

Ví dụ : Giả sử X là biến ngẫu nhiên được xét và giả sử chịu tác động của hai yếu tố A và B, hiệu $X - m$ (trong đó m là kỳ vọng của X) được khai triển thành tổng :

$$X - m = U + V + W$$

trong đó : U là độ lệch chịu tác động bởi yếu tố A,

V là độ lệch chịu tác động bởi yếu tố B,

W là độ lệch chịu tác động bởi các yếu tố ngẫu nhiên khác.

Giả sử U, V, W là những biến ngẫu nhiên độc lập với phương sai tương ứng DU, DV, DW.

Ta có $DX = DU + DV + DW$

Sự so sánh DU hoặc DV với DW cho biết mức độ ảnh hưởng của A hoặc B đến X so với sự ảnh hưởng của yếu tố ngẫu nhiên khác tới X. Phân tích phương sai cho ta tiêu chuẩn đánh giá sự tồn tại ảnh hưởng của các yếu tố đó đến đại lượng ngẫu nhiên X đang xét.

Dưới đây ta chỉ xét vấn đề phân tích phương sai theo một và hai yếu tố.

7.2. PHÂN TÍCH PHƯƠNG SAI

7.2.1. Phân tích theo một yếu tố tác động

a) Phương pháp chọn ngẫu nhiên từng phần tử từ tập tổng quát

Giả sử xét mẫu ngẫu nhiên được tạo nên bởi một yếu tố A tác động gồm v mức $A_1, A_2, \dots A_v$ được cho bởi bảng sau :

A_1	$A_2 \ldots$	A_v
X_{11}	$X_{12} \ldots$	X_{1v}
X_{21}	$X_{22} \ldots$	X_{2v}
...	... ...	....
$X_{n_1 1}$	$X_{n_2 2} \ldots$	$X_{n_v v}$

Giả thiết rằng các quan sát X_{ij} ; $i = 1, \ldots, n_j$; $j = \overline{1, v}$, có phân phối chuẩn $N(m_j ; \sigma^2)$ với kỳ vọng $EX_{ij} = m_j$, $DX_{ij} = \sigma^2$. Nếu đặt $b_j = m_j - m$, trong đó $m = \frac{1}{n}\sum_{j=1}^{v} n_j m_j$. Ta suy ra $EX_{ij} = m + b_j$ và $X_{ij} = m + b_j + W_{ij}$, trong đó W_{ij} có phân phối chuẩn $EW_{ij} = 0$ và $DW_{ij} = \sigma^2$ (7.1)

Để khảo sát sự ảnh hưởng của yếu tố A, ta xét sự ảnh hưởng của từng mức $A_1, A_2, \ldots, A_v$. Nếu tác động ở từng mức $A_1, A_2, \ldots, A_v$ đến X khác nhau, thì có nghĩa là A có ảnh hưởng đến X. Vì vậy bài toán đưa về so sánh các trung bình của các quan sát trong từng mức với nhau. Nếu các trung bình này khác nhau thì có nghĩa là A có tác động đến X.

Kiểm định giả thiết $H_o : m_1 = m_2 = \ldots = m_v$

$K : m_1 \neq m_2 \neq \ldots \neq m_v$ ở mức α.

hoặc có thể viết dưới dạng tương đương :

$$H_o : b_1 = b_2 = \ldots = b_v = 0$$

$$K : b_1 \neq b_2 \neq \ldots \neq b_v \neq 0$$

Tính $\bar{X} = \frac{1}{n}\sum_{j=1}^{v}\sum_{i=1}^{n_j} X_{ij}$; $\bar{X}_{\bullet 1} = \frac{1}{n_1}\sum_{i=1}^{n_1} X_{i1}, \ldots, \bar{X}_{\bullet v} = \frac{1}{n_v}\sum_{i=1}^{n_v} X_{iv}$

$$n = n_1 + n_2 + \ldots + n_v.$$

Xét $Q = \sum_{j=1}^{v}\sum_{i=1}^{n_j}(X_{ij} - \bar{X})^2 = \sum_{j=1}^{v}\sum_{i=1}^{n_j}(X_{ij} - \bar{X}_{\bullet j} + \bar{X}_{\bullet j} - \bar{X})^2$

$$= \sum_{j=1}^{v}\sum_{i=1}^{n_j}(X_{ij} - \bar{X}_{\bullet j})^2 + \sum_{j=1}^{v}\sum_{i=1}^{n_j}(\bar{X}_{\bullet j} - \bar{X})^2 +$$

$$+ 2\sum_{j=1}^{v}\sum_{i=1}^{n_j}(X_{ij} - \bar{X}_{\bullet j})(\bar{X}_{\bullet j} - \bar{X})^2$$

Đặt $Q_A = \sum_{j=1}^{v} n_j(\bar{X}_{\bullet j} - \bar{X})^2$; $Q_R = \sum_{j=1}^{v}\sum_{i=1}^{n_j}(X_{ij} - \bar{X}_{\bullet j})^2$

Số hạng thứ 3 ở vế phải của đẳng thức trên bằng 0.

Vậy $Q = Q_A + Q_R$

trong đó Q được gọi là tổng bình phương độ lệch toàn thể.

Q_A được gọi là tổng bình phương độ lệch nhóm giữa.

Q_R được gọi là tổng bình phương độ lệch nhóm trong.

Dưới giả thiết H_o người ta chứng minh được rằng :

Q_A/σ^2 có phân phối χ^2 với v – 1 bậc tự do.

Q_R/σ^2 có phân phối χ^2 với n – v bậc tự do.

Q_A, Q_R là độc lập và $F_A = \dfrac{Q_A/(v-1)}{Q_R/(n-v)}$ có phân phối F với v – 1, n – v bậc tự do.

Tiêu chuẩn kiểm định giả thiết H_o.

Giả thiết H_o bị bác bỏ ở mức α nếu :

$$F_A > f_\alpha\,(v-1, n-v) \qquad (7.2)$$

trong đó f_α là số tra trong bảng phân phối F với v – 1, n – v bậc tự do và mức α.

Ta cũng nhận được 1 số ước lượng của các tham số cần thiết :

+ Ước lượng của m là $\bar{X}$.

+ Ước lượng của m_j là $\bar{X}_{\bullet j}$.

+ Ước lượng của b_j là $\bar{X}_{\bullet j} - \bar{X}$.

Ví dụ 7.1. Người ta tiến hành dạy thí điểm ba phương pháp bằng cách bố trí dạy song song 3 lớp cùng loại đối tượng như nhau (như nhau về trình độ xuất phát và thiết bị học tập). Sau khi dạy xong một phần chương trình, người ta kiểm tra cả 3 lớp, mỗi lớp 15 học sinh. (Điểm cho theo bậc 20). Kết quả cho ở bảng sau đây :

A	B	C
9	15	18
11	16	14
10	15	17
12	10	9
7	13	14
11	14	17
12	15	16
20	7	15

A	B	C
13	13	16
11	15	8
13	15	14
11	14	10
10	11	16
12	15	15
13	10	17

Xét sự ảnh hưởng của ba phương pháp đó đến tình trạng kiến thức của học sinh, cho mức kiểm định $\alpha = 0{,}01$.

Giải : Dựa vào bảng dữ liệu trên ta tính được $Q_A = 89{,}2$; $Q_R = 262$; $\overline{X} = 12{,}866$; $\overline{X}_{\bullet 1} = 11$; $\overline{X}_{\bullet 2} = 13{,}2$; $\overline{X}_{\bullet 3} = 14{,}4$; $\hat{b}_1 = -1{,}866$; $\hat{b}_2 = 0{,}334$; $\hat{b} = 1{,}534$; $F_A = \dfrac{Q_A/2}{Q_R/42} = 7{,}15$. Tra bảng F ta tìm được $f(0{,}01\,;\,2\,;\,42) = 5{,}15$. Ta thấy $F_A > f_\alpha$. Vậy bác bỏ giả thiết H_o, nghĩa là hiệu quả của 3 phương pháp đó khác nhau.

Ví dụ 7.2. Thời gian trong bụng mẹ của 16 con bò con thuộc giống của 3 con bò đực, tính ra ngày như sau :

Bò đực 1	Bò đực 2	Bò đực 3
292	281	282
289	289	295
299	292	292
296	281	287
302	286	292
289	280	285
288	291	284
293	288	284

Bò đực 1	Bò đực 2	Bò đực 3
293	298	287
294	280	297
284	285	287
286	289	289
287	288	290
291	287	286
301	292	278
290	288	290

Ta có thể coi thời gian trong bụng mẹ là một đặc tính có tính di truyền hay không ? (Ta có thể coi thời gian có chửa của 1 con bò cái là 1 biến ngẫu nhiên chuẩn có phương sai σ^2 độc lập với con bò cái đang xét). Cho mức kiểm định $\alpha = 0,05$.

Bò I	Bò 2	Bò 3
2	−9	−8
−1	−1	5
9	2	2
6	−9	−3
12	−4	2
−1	−10	−5
−2	1	−6
3	−2	−6
3	8	−3
4	−10	7
−6	−5	−3
−4	−1	−1
−3	−2	0
1	−3	−4
11	2	−12
0	−2	0

Giải : Ta cho rằng con của con bò đực thứ i có một thời gian ngẫu nhiên ở trong bụng mẹ có quy luật chuẩn $N(m_i ; \sigma^n)$. Nếu đặc tính thời gian trong bụng mẹ là không di truyền thì các kỳ vọng m_i phải bằng nhau (giả thiết H_o). Ta sẽ kiểm định giả thiết H_o và coi thời gian trong bụng mẹ là đặc tính di truyền, nếu ta bác bỏ giả thiết H_o. Để tính toán thuận tiện ta đặt $U_{ij} = X_{ij} - 290$, dữ liệu của U_{ij} được cho ở bảng bên, ta tính được.

$U_{\bullet 1} = 34$; $\bar{U}_{\bullet 1} = 2,125$; $U_{\bullet 2} = -45$; $\bar{U}_{\bullet 2} = -2,8125$; $\bar{U}_{\bullet 3} = -2,187$; $U_{\bullet 3} = -35$; $\sum_{i=1}^{16} U_{i1}^2 = 488$; $\sum_{i=1}^{16} U_{i2}^2 = 499$; $\sum_{i=1}^{r} U_{i3}^2 = 431$; $\sum_{i \neq j} U_{ij}^2 = 1418$;

$Q_A = 231,29$; $Q_R = 1142,63$.

$$\text{Tính } F_A = \frac{Q_A/2}{Q_R/(n-v)} = \frac{231,29/2}{1142,63/45} = 4,55$$

Tra bảng F ta tìm được $f(0,05 ; 2,45) = 3,21$

Ta thấy $F_A > f_\alpha$. Vậy bác bỏ giả thiết H_o, nghĩa là thời gian trong bụng mẹ của các bò con là một đặc tính di truyền.

b) Phương pháp khối (bloc) ngẫu nhiên

Trong mục a) ta lấy ngẫu nhiên từng đối tượng bất kỳ trong tập tổng quát để thí nghiệm. Đôi khi do yêu cầu thực tế hay do hoàn cảnh nhất định nào đó, người ta phân các đối tượng thí nghiệm thành từng khối trước khi thí nghiệm theo một tiêu chuẩn nào đó. Phương pháp làm thí nghiệm như thế được gọi là khối ngẫu nhiên.

Mỗi khối gồm các đối tượng thí nghiệm khác nhau. Kích thước của mỗi khối (bloc) bằng số các đối tượng thí nghiệm. Ví dụ có r khối, mỗi khối có v đối tượng thí nghiệm. Giả sử yếu tố A được phân làm v mức : $A_1, A_2, ..., A_v$. Kết quả thí nghiệm cho ở bảng sau đây :

Khối \ A	A_1	A_2	...	A_v
Khối 1	X_1	X_2	...	X_v
Khối 2	X_{21}	X_{22}	...	X_{2v}
...	...	...	..	...
Khối r	X_{r1}	X_{r2}	...	X_{rv}

Giả sử các quan sát X_{ij} có phân phối chuẩn $N(\mu_{ij}, \sigma^2)$ có kỳ vọng $EX_{ij} = \mu_{ij}$, $DX_{ij} = \sigma^2$.

Đặt $$\mu = \frac{1}{n}\sum_{i=1}^{r}\sum_{j=1}^{v}\mu_{ij} \ ; \ \mu_{i\bullet} = \frac{1}{v}\sum_{j=1}^{v}\mu_{ij} \ ; \mu_{\bullet j} = \frac{1}{r}\sum_{i=1}^{r}\mu_{ij}$$

và $a_i = \mu_{i\bullet} - \mu$, $b_j = \mu_{\bullet j} - \mu$.

Vậy $EX_{ij} = \mu + a_i + b_j$.

Suy ra $$X_{ij} = \mu + a_i + b_j + W_{ij} \tag{7.3}$$

trong đó W_{ij} có phân phối chuẩn $N(0 ; \sigma^2)$.

Trong mô hình (7.3), b_j là thành phần sinh ra do yếu tố A tác động ở mức A_j, còn a_i là phần do khối thứ i tạo ra.

Do đó muốn khảo sát sự ảnh hưởng của yếu tố A đến X, sự ảnh hưởng của việc phân khối để làm thí nghiệm đến X, cần phải kiểm định hai giả thiết sau :

$$H_o : b_1 = b_2 = ... = b_v = 0$$
$$H_o' : a_1 = a_2 = ... = a_r = 0$$

Trước khi đi đến tiêu chuẩn kiểm định các giả thiết trên, ta làm các phần bổ trợ sau :

Đặt $\bar{X} = \frac{1}{rv}\sum_{i=1}^{r}\sum_{j=1}^{v}X_{ij}$; $\bar{X}_{i\bullet} = \frac{1}{v}\sum_{j=1}^{v}X_{ij}$; $\bar{X}_{\bullet j} = \frac{1}{r}\sum_{i=1}^{r}X_{ij}$

Xét $Q = \sum_{i=1}^{r}\sum_{j=1}^{v}(X_{ij} - \bar{X})^2$.

Ta có thể tách Q thành tổng ba thành phần như sau :

$$Q = Q_A + Q_B + Q_R$$

trong đó : $Q_A = r\sum_{j=1}^{v}(\bar{X}_{\bullet j} - \bar{X})^2$; $Q_B = v\sum_{i=1}^{r}(\bar{X}_{i\bullet} - \bar{X})^2$

$$Q_R = Q - Q_A - Q_B = \sum_{i,j}(X_{ij} - \bar{X}_{i\bullet} - \bar{X}_{\bullet j} + \bar{X})^2 .$$

Người ta chứng minh được rằng :

Q_A/σ^2 có phân phối khi bình phương với $v - 1$ bậc tự do.

$\frac{Q_B}{\sigma^2}$ có phân phối khi bình phương với $r - 1$ bậc tự do.

Q_R/σ^2 có phân phối khi bình phương với $(r - 1)(v - 1)$ bậc tự do.

$F_A = \frac{Q_A/(v-1)}{Q_R/(r-1)(v-1)} = \frac{Q_A}{Q_R}(r-1)$ có phân phối F với $(v - 1)$, $(r - 1)(v - 1)$ bậc tự do.

$F_B = \frac{Q_B/(r-1)}{Q_R/(r-1)(v-1)} = \frac{Q_B}{Q_R}(v-1)$ có phân phối F với $(r - 1)$, $(r - 1)(v - 1)$ bậc tự do.

Các Q_A, Q_B, Q_R là độc lập với nhau.

Từ đó suy ra tiêu chuẩn kiểm định giả thiết H_o và H_o' như sau :

– Giả thiết H_o bị bác bỏ ở mức α nếu :

$$F_A > f_\alpha(v - 1, (r - 1)(v - 1)) \qquad (7.4)$$

trong đó f_α tra ở bảng phân phối F với $v - 1$, $(r - 1)(v - 1)$ bậc tự do và mức α.

Nếu $F_A < f_\alpha$ thì chấp nhận giả thiết H_o.

– Giả thiết H_o' bị bác bỏ ở mức α nếu :

$$F_B > f_\alpha'(r-1, (r-1)(v-1)) \quad (7.5)$$

Nếu $F_B < f_\alpha'$ thì chấp nhận H_o'.

f_α' tra ở bảng phân phối F với (r – 1), (r – 1)(v – 1) bậc tự do và mức α.

Ước lượng của μ là $\overline{X}$.

Ước lượng của a_i là $(\overline{X}_{i\bullet} - \overline{X})$.

Ước lượng của b_j là $(\overline{X}_{\bullet j} - \overline{X})$.

Ví dụ 7.3. Tiến hành dạy thí điểm ba phương pháp A_1, A_2, A_3. Học sinh được phân thành 15 mức $B_1, B_2, \ldots B_{15}$. Kết quả thí nghiệm được cho ở bảng sau (X_{ij} là điểm kiểm tra cho theo thang 100).

B_j \ A_i	A_1	A_2	A_3	Tổng	Trung bình $\overline{X}_{i\bullet}$
B_1	13	15	16	44	14,66
B_2	15	13	17	45	15
B_3	12	14	15	41	13,66
B_4	14	15	16	45	15
B_5	12	14	15	41	13,66
B_6	13	13	17	43	14,33
B_7	11	14	9	34	11,33
B_8	13	12	13	38	12,66
B_9	11	12	15	38	12,66
B_{10}	12	14	13	39	13
B_{11}	9	10	13	32	10,66
B_{12}	11	14	14	39	13
B_{13}	10	9	9	28	9,33
B_{14}	7	9	14	30	10
B_{15}	10	6	8	24	8
Tổng $X_{\bullet j}$	173	184	204	561	$\overline{X}$ = 12,46
$\overline{X}_{\bullet j}$	11,533	12,266	13,6		

Xét sự ảnh hưởng của ba phương pháp đến chất lượng học tập của học sinh. Cho mức kiểm định $\alpha = 0,01$.

Giải : Dựa vào bảng dữ liệu của đầu bài ta tìm được $\bar{X}_{\bullet 1}, \bar{X}_{\bullet 2}, \bar{X}_{\bullet 3}, \bar{X}_{1\bullet}, \bar{X}_{2\bullet}, \bar{X}_{3\bullet}, \ldots \bar{X}_{15\bullet}, \bar{X}$ như trong bảng và $Q_A = 32,9$, $Q_B = 195,2$; $Q = 303,2$; $Q_R = 75,1$.

$$F_A = \frac{Q_A/2}{Q_R/28} = 6,13\,,\ F_B = \frac{Q_B/14}{Q_R/28} = 5,2\,.$$

Tra bảng phân phối F ta tìm được $f_{0,01}(2\ ;\ 28) = 5,45$; $f'(0,01\ ;\ 14\ ;\ 28) = 2,8$.

Ta thấy :

– $F_A > f_\alpha$. Vậy giả thiết H_o bị bác bỏ, nghĩa là ba phương pháp A_1, A_2, A_3 có hiệu quả khác nhau. Ước lượng điểm trung bình của 3 lớp $\bar{X}_{\bullet 1} = 11,533$; $\bar{X}_{\bullet 2} = 12,266$; $\bar{X}_{\bullet 3} = 13,6$.

– $F_B > f'_\alpha$. Vậy H'_o bị bác bỏ ở mức α, nghĩa là việc chia các mức (trình độ), $B_1, \ldots, B_{15}$ cũng có ảnh hưởng đến chất lượng học của học sinh.

c) Hình vuông Latinh

Trong trường hợp có r loại đối tượng và số mức của hai yếu tố A và B đều bằng r. Đáng lý ra chúng ta phải cần đến $r \times r \times r = r^3$ đơn vị thí nghiệm để nghiên cứu tất cả các trường hợp có thể. Nhưng việc làm đó nhiều khi dẫn đến lãng phí công sức và tiền của làm thí nghiệm, song kết quả cũng không chính xác hơn bao nhiêu. Dưới đây ta sẽ trình bày phương pháp làm số thí nghiệm ít hơn nhưng vẫn đảm bảo được các kết luận có độ chính xác mong muốn. Đó là phương pháp bố trí thí nghiệm theo hình vuông Latinh.

Định nghĩa 7.1. Xét hình vuông những số $\{1, 2, \ldots, r\}$, trong đó r là số chiều của hình vuông. Gọi hình vuông Latinh là hình vuông mà mỗi số trong hình chỉ có mặt một lần trong mỗi hàng trong r hàng và một lần trong mỗi cột trong r cột của hình vuông đó.

Ví dụ : $\begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \\ 3 & 1 & 2 \end{bmatrix}$

Giả sử mẫu quan sát được cho dưới dạng hình vuông sau đây :

	Khối cột 1	Khối cột 2	...	Khối cột r
Khối hàng 1	$y_{11}(1)$	$y_{12}(2)$	...	$y_{1r}(r)$
Khối hàng 2	$y_{21}(2)$	$y_{22}(1)$	...	$y_{2r}(r-1)$
.....	.....	.....	...	...
Khối hàng r	$y_{r1}(r)$	$y_{r2}(r-1)$	...	$y_{rr}(1)$

Trong đó $y_{ij}(k)$ là quan sát xuất hiện ở ô (i,j) là làm trên đối tượng k.

Ví dụ 7.4. Có ba loại bò A, B, C chia thành 3 nhóm theo lứa tuổi và cho ăn theo ba chế độ 1, 2, 3. $y_{ij}(k)$ là lượng sữa của con bò k ở tuổi i và chế độ ăn j.

Ta có mẫu quan sát :

	Mức ăn 1	Mức ăn 2	Mức ăn 3
Tuổi 1	$y_{11}(C)$	$y_{12}(B)$	$y_{13}(A)$
Tuổi 2	$y_{21}(B)$	$y_{22}(A)$	$y_{23}(C)$
Tuổi 3	$y_{31}(A)$	$y_{32}(C)$	$y_{33}(B)$

Giả sử $y_{ij}(k)$ có phân phối chuẩn $N(\mu_{ij}(k)\ ;\ \sigma^2)$, $Ey_{ij}(k) = \mu_{ij}(k)$, $Dy_{ij}(k) = \sigma^2$.

Đặt $\mu = \dfrac{1}{r^2}\sum_{i,j(k)}\mu_{ij}(k)$, $\mu_{i\bullet\bullet} = \dfrac{1}{r}\sum_{j(k)}\mu_{ij}(k)$; $\mu_{\bullet j\bullet} = \dfrac{1}{r}\sum_{i(k)}\mu_{ij}(k)$;

$\mu_{\bullet\bullet}(k) = \dfrac{1}{r}\sum_{(i,j)}\mu_{ij}(k)$;

và đặt $a_i = \mu_{i\bullet\bullet} - \mu$, $b_j = \mu_{\bullet j\bullet} - \mu$; $C_{(k)} = \mu_{\bullet\bullet(k)} - \mu$.

Ta có thể viết $Ey_{ij(k)} = \mu + a_i + b_j + c_{(k)}$.

Từ đó suy ra $y_{ij(k)} = \mu + a_i + b_j + c_{(k)} + w_{ij\,(k)}$ (7.6)

trong đó $w_{ij(k)}$ có phân phối chuẩn với kỳ vọng bằng 0 và phương sai là σ^2.

Trong đó : μ là trung bình toàn thể.

a_i là hiệu quả riêng theo khối hàng i.

b_j là hiệu quả riêng theo khối cột j.

c(k) là hiệu quả riêng của đối tượng k.

$w_{ij(k)}$ do yếu tố ngẫu nhiên khác gây nên. Nó có phân phối chuẩn với kỳ vọng bằng 0 và phương sai bằng σ^2.

Ước lượng của μ là $\overline{y} = \frac{1}{r}\sum_i \sum_{j(k)} y_{ij(k)}$.

Ước lượng của a_i là $\overline{y}_{i\bullet\bullet} - \overline{y}$.

Ước lượng của b_j là $\overline{y}_{\bullet j\bullet} - \overline{y}$.

Ước lượng của $C_{(k)}$ là $\overline{y}_{\bullet\bullet(k)} - \overline{y}$.

Trong ví dụ 7.4 ta tính các trung bình mẫu :

$$\overline{y}_{\bullet\bullet A} = \frac{1}{3}(y_{13}(A) + y_{22}(A) + y_{31}(A))$$

$$\overline{y}_{\bullet 1\bullet} = \frac{1}{3}(y_{11}(C) + y_{21}(B) + y_{31}(A))$$

$$\overline{y}_{1\bullet\bullet} = \frac{1}{3}(y_{11}(C) + y_{12}(B) + y_{13}(A))$$

Việc xét sự ảnh hưởng của việc phân mức theo cột và theo hàng ta đã xét ở mục b) ; bây giờ ta chỉ xét sự ảnh hưởng đến từng loại đối tượng thí nghiệm cụ thể. Muốn giải quyết vấn đề này ta đi đến bài toán kiểm định giả thiết :

$$H_o : C_1 = C_2 = ... = C_r = 0.$$

Nếu chấp nhận giả thiết H_o này, nghĩa là hiệu quả tác động lên các đối tượng là như nhau.

Để giải quyết bài toán này ta tiến hành theo các bước sau :

Phân tích :

$$Q = \sum\sum(y_{ij(k)} - \overline{y})^2 = Q_H + Q_C + Q_Đ + Q_R$$

trong đó :

$$Q_H = r\sum_{i=1}^{r}(\overline{y}_{i\bullet\bullet} - \overline{y})^2 ; Q_C = r\sum_{j=1}^{r}(\overline{y}_{\bullet j\bullet} - \overline{y})^2$$

$$Q_Đ = r\sum(\overline{y}_{\bullet\bullet(k)} - \overline{y})^2$$

Q_H – đặc trưng ảnh hưởng của khối hàng.

Q_C – đặc trưng ảnh hưởng của khối cột.

$Q_Đ$ – đặc trưng ảnh hưởng đối với từng đối tượng.

$$Q_R = Q - Q_H - Q_C - Q_Đ = \sum_i \sum_{j(k)} (y_{ij(k)} - \overline{y}_{i\bullet\bullet} - \overline{y}_{\bullet j\bullet} - \overline{y}_{\bullet\bullet(k)} + \overline{y})^2$$

Người ta chứng minh được rằng :

$Q_Đ/\sigma^2$ có phân phối χ^2 với r – 1 bậc tự do.

Q_R/σ^2 có phân phối χ^2 với (r – 1)(r – 2) bậc tự do ; $Q_Đ$, Q_R là độc lập với nhau.

$\dfrac{Q_Đ/(r-1)}{Q_R/(r-1)(r-2)}$ có phân phối F với r – 1, (r – 1)(r – 2) bậc tự do.

Ta rút ra tiêu chuẩn kiểm định giả thiết H_o như sau : giả thiết H_o bị bác bỏ ở mức α nếu :

$$F > f_\alpha(r-1\ ;\ (r-1)(r-2)) \tag{7.7}$$

nếu $F < f_\alpha$ thì chấp nhận giả thiết H_o.

Trong đó $f_\alpha(r-1\ ;\ (r-1)(r-2))$ tra ở bảng phân phối F với r – 1, (r – 1)(r – 2) bậc tự do và mức α.

Ví dụ 7.5. Muốn so sánh chất lượng cao su của ba loại lốp người ta làm thí nghiệm như sau : Một trong ba loại lốp A, B, C được sử dụng ở một trong ba loại xe : ôtô con ; ôtô có tải trọng trung bình, ôtô có tải trọng lớn và cho đi trên một trong ba loại đường. Đi cho tới khi hỏng lốp. Người ta ghi được một số km đi được ở bảng sau :

	ôtô con	ôtô TB	ôtô lớn
Đường loại 1	A : 17	B : 22	C : 28
Đường loại 2	C : 24	A : 16	B : 17
Đường loại 3	B : 15	C : 23	A : 15

Để tiện cho việc tính toán ta làm phép biến đổi :

Đặt $Z_{ij(k)} = y_{ij(k)} - 17$.

Ta có $A:0 \quad B:5 \quad C:11$

$C:7 \quad A:-1 \quad B:0$

$B:-2 \quad C:6 \quad A:-2$

Ta tính được $Q = \sum\sum Z^2_{ij(k)} - 9\overline{Z}^2$; $\overline{Z} = \frac{24}{9}$.

$$Q = 0^2 + 25 + 121 + 49 + 1 + 0 + 4 + 36 + 4 - 9 \times \left(\frac{24}{9}\right)^2$$

$$Q = 240 - 64 = 176$$

$$Q_Đ = 3\,(\sum \overline{Z}^2_{\bullet\bullet k} - 3\overline{Z}^2)$$

$$\overline{Z}_{\bullet\bullet A} = \frac{1}{3}(0 - 1 - 2) = -1$$

$$\overline{Z}_{\bullet\bullet B} = \frac{1}{3}(-2 + 5 + 0) = 1$$

$$\overline{Z}_{\bullet\bullet C} = \frac{1}{3}(7 + 6 + 11) = 8$$

$$\sum \overline{Z}^2_{\bullet\bullet k} = (-1)^2 + 1^2 + 8^2 = 66$$

$$Q_Đ = 3\left(66 - \frac{64}{3}\right) = 134$$

$$Q_H = 3(\sum Z^2_{i\bullet\bullet} - 3\overline{Z}^2)$$

$$\overline{Z}_{1\bullet\bullet} = \frac{1}{3}(0 + 5 + 11) = \frac{16}{3}, \overline{Z}_{2\bullet\bullet} = \frac{1}{3}(7 - 1 + 0) = 2$$

$$\overline{Z}_{3\bullet\bullet} = \frac{1}{3}(-2 + 6 - 2) = \frac{2}{3}, Q_H = 3\left(4 + \frac{68}{9}\right) = 34{,}7$$

$$Q_C = 3(\sum_{j=1}^{3} \overline{Z}^2_{\bullet j\bullet} - 3\overline{Z}^2)$$

$$\overline{Z}_{\bullet 1\bullet} = \frac{1}{3}(+7 - 2) = \frac{5}{3};\ \ \overline{Z}_{\bullet 2\bullet} = \frac{1}{3}(5 - 1 + 6) = \frac{10}{3};$$

$$\overline{Z}_{\bullet 3\bullet} = \frac{1}{3}(11 + 0 - 2) = 3;\ \ \sum \overline{Z}^2_{\bullet j\bullet} = \frac{25}{9} + \frac{100}{9} + 9 = \frac{125}{9} + 9 = 22{,}89$$

$$Q_C = \frac{14}{3} \approx 4{,}7$$

$Q_R = Q - Q_H - Q_C - Q_Đ = 176 - 134 - 34,7 - 4,7 = 2,6$

Cuối cùng ta tính được $F = \dfrac{Q_Đ/2}{Q_R/2} = \dfrac{134/2}{2,6/2} = 51,5.$

Tra bảng phân phối F ta tìm được f(0,05 ; 2 ; 2) = 19.

Ta nhận thấy $F = 51,9 > f_\alpha = 19$.

Vậy bác bỏ giả thiết H_o, nghĩa là chất lượng cao su đối với từng loại lốp khác nhau.

Để định lượng sự khác nhau ta tìm ước lượng của C(A), C(B), C(C) tức là $\overline{y}_{\bullet\bullet(A)} - \overline{y}$, $\overline{y}_{\bullet\bullet(B)} - \overline{y}$, $\overline{y}_{\bullet\bullet(C)} - \overline{y}$.

7.2.2. Phân tích phương sai theo hai yếu tố tác động

Giả sử có hai yếu tố A, B tác động. Yếu tố A được phân làm r mức là $A_1, A_2, \ldots A_r$, còn yếu tố B được phân làm v mức $B_1, B_2, \ldots, B_v$.

Giả sử mỗi ô (i, j) chịu tác động bởi A_iB_j có số quan sát như nhau và bằng s.

Giả sử các quan sát X_{ijg}, i = 1, ..., r ; j = 1, 2, ..., v ; g = 1, 2, ..., s, có phân phối chuẩn $N(m_{ij}, \sigma^2)$ có kỳ vọng $EX_{ijg} = m_{ij}$ và $DX_{ijg} = \sigma^2$.

Đặt $m_{i\bullet} = \dfrac{1}{v}\sum_{j=1}^{v} m_{ij}$; $m_{\bullet j} = \dfrac{1}{r}\sum_{i=1}^{r} m_{ij}$; $m = \dfrac{1}{rv}\sum_{i=1}^{r}\sum_{j=1}^{v} m_{ij}$ và đặt

$a_i = m_{i\bullet} - m$; $b_j = m_{\bullet j} - m$

$F_{ij} = m_{ij} - m_{i\bullet} - b_j = m_{ij} - m_{\bullet j} - a_i$

Từ đó ta có $EX_{ijg} = m + a_i + b_j + F_{ij}$

và suy ra $X_{ijg} = m + a_i + b_j + F_{ij} + W_{ijg}$ (7.8)

trong đó a_i sinh ra do A_i tác động ; b_j sinh ra do B_j tác động ; F_{ij} sinh ra do sự tác động chung A_iB_j ;

W_{ijg} sinh ra do tác động của các yếu tố ngẫu nhiên khác. Nó có phân phối chuẩn với kỳ vọng bằng 0 và phương sai bằng σ^2.

Để biết được các yếu tố A, B có tác dụng đến X không và giữa A, B có sinh ra lực tác dụng có ý nghĩa đến X không, ta đi đến bài toán kiểm định ba giả thiết sau đây :

$$H_o : a_1 = a_2 = \ldots = a_r = 0$$

$$H_o' : b_1 = b_2 = \ldots = b_v = 0$$

$$H_o'' : F_{ij} = 0 \text{ với mọi } i = \overline{1,r},\ j = \overline{1,v}.$$

Đặt : $$\overline{X}_{ij} = \frac{1}{s}\sum_{j=1}^{s} X_{ijg},\ \overline{\overline{X}}_{i\bullet} = \frac{1}{v}\sum_{j=1}^{v}\overline{X}_{ij},\ \overline{\overline{X}}_{\bullet j} = \frac{1}{r}\sum_{i=1}^{r}\overline{X}_{ij},$$

$$\overline{\overline{X}} = \frac{1}{rv}\sum_{i=1}^{r}\sum_{j=1}^{v}\overline{X}_{ij}$$

Xét : $$Q = \sum_{i=1}^{r}\sum_{j=1}^{v}\sum_{g=1}^{s}(X_{ijg} - \overline{\overline{X}})^2$$

Ta phân tích Q thành tổng :

$$Q = Q_A + Q_B + Q_{AB} + Q_R$$

trong đó : $Q_A = sv\sum_{i=1}^{r}(\overline{\overline{X}}_{i\bullet} - \overline{\overline{X}})^2$ (đặc trưng sự ảnh hưởng của A).

$Q_B = sr\sum_{j=1}^{v}(\overline{\overline{X}}_{\bullet j} - \overline{\overline{X}})^2$ (đặc trưng sự ảnh hưởng của B).

$$Q_{AB} = s\sum_{i=1}^{r}\sum_{j=1}^{v}(\overline{X}_{ij} - \overline{\overline{X}}_{i\bullet} - \overline{\overline{X}}_{\bullet j} + \overline{\overline{X}})^2$$

(đặc trưng sự ảnh hưởng chung của AB)

Dưới giả thiết H_o, H_o', H_o'' người ta chứng minh được rằng :

Q_A, Q_B, Q_{AB}, Q_R là độc lập và :

Q_A/σ^2 có phân phối χ^2 với $r - 1$ bậc tự do ;

Q_B/σ^2 có phân phối χ^2 với $v - 1$ bậc tự do ;

Q_{AB}/σ^2 có phân phối χ^2 với $(r - 1)(v - 1)$ bậc tự do ;

Q_R/σ^2 có phân phối χ^2 với $rv(s - 1)$ bậc tự do.

$F_A = \dfrac{Q_A/(r-1)}{Q_R/rv(s-1)}$ có phân phối F với $r - 1$, $rv(s - 1)$ bậc tự do.

$F_B = \dfrac{Q_B/(v-1)}{Q_R/rv(s-1)}$ có phân phối F với $v - 1$, $rv(s - 1)$ bậc tự do.

$$F_{AB} = \frac{Q_{AB}/(v-1)(r-1)}{Q_R / rv(s-1)}$$ có phân phối F với $(r-1)(v-1)$, $rv(s-1)$ bậc tự do.

Ta suy ra các tiêu chuẩn sau :

* Giả thiết H_o bị bác bỏ ở mức α nếu :

$$F_A > f_\alpha(r-1, rv(s-1)) \tag{7.8}$$

* Giả thiết $H_o^{'}$ bị bác bỏ ở mức α nếu :

$$F_B > f_\alpha^{'}(v-1, rv(s-1)) \tag{7.9}$$

* Giả thiết $H_o^{''}$ bị bác bỏ ở mức α nếu :

$$F_{AB} > f_\alpha^{''}((r-1)(v-1), rv(s-1)) \tag{7.10}$$

trong đó f_α, $f_\alpha^{'}, f_\alpha^{''}$ là số tra từ bảng phân phối F với số bậc tự do tương ứng $r-1$, $rv(s-1)$,

$v-1$, $rv(s-1)$,

$(r-1)(v-1)$, $rv(s-1)$

và mức ý nghĩa α.

Ví dụ 7.6. Nghiên cứu sự tác động của hai yếu tố A và B đến X ở mức $\alpha = 0,05$, dựa trên bảng quan sát sau :

A \ B	B_1					B_2				
A_1	190	260	170	170	170	190	150	210	150	150
A_2	150	250	220	140	180	230	190	200	190	200
A_3	190	185	135	195	195	150	170	150	170	180

Giải :

Ở đây $r = 3$, $v = 2$, $s = 5$.

$\overline{X}_{11} = 192$; $\overline{X}_{12} = 170$; $\overline{\overline{X}}_{1\bullet} = 181$; $\overline{\overline{X}}_{2\bullet} = 195$; $\overline{\overline{X}}_{3\bullet} = 172$

$\overline{X}_{21} = 188$; $\overline{X}_{22} = 202$; $\overline{\overline{X}}_{\bullet 1} = 186,7$; $\overline{\overline{X}}_{\bullet 2} = 178,7$; $\overline{\overline{X}} = 182,7$

$\overline{X}_{31} = 180$; $\overline{X}_{32} = 164$.

$Q_A = 2686,6$; $Q_B = 480$; $Q_{AB} = 1860$; $Q_R = 22360$.

$F_A = 2,88$; $F_B = 0,26$; $F_{AB} = 0,999$. Cho $\alpha = 0,05$, tra bảng phân phối F ta tìm được f(0,05 ; 2 ; 24) = 3,4 ; f'(0,05 ; 1 ; 24) = 4,26 ; f'' (0,05 ; 2 ; 24) = 3,4.

Ta thấy $F_A < f_\alpha$, $F_B < f'_\alpha$, $F_{AB} < f''_\alpha$. Vậy các giả thiết H_o, H'_o, H''_o được chấp nhận, nghĩa là A, B, AB không có tác động có ý nghĩa tới X.

7.3. PHÂN TÍCH HỒI QUY

7.3.1. Khái niệm hàm hồi quy và phương pháp bình phương bé nhất

a) Có nhiều bài toán trong các lĩnh vực sinh học, kinh tế, giáo dục, khí tượng học v.v.. Khi biết được một số giá trị thực nghiệm của các đại lượng, ví dụ : X_1, Y_1, ..., X_n, Y_n, cần biểu thị đại lượng Y thông qua X (hoặc ngược lại), nghĩa là ta phải tìm hàm $y = f(x) + \varepsilon$ (hoặc $x = \varphi(y) + \varepsilon_1$). Hàm $y = f(x) + \varepsilon$ tìm được như thế, thường được gọi là hàm hồi quy của đại lượng ngẫu nhiên Y theo đại lượng ngẫu nhiên X. Trong mục này ta định nghĩa hàm hồi quy trên quan điểm kỳ vọng có điều kiện.

Định nghĩa 7.2. Gọi kỳ vọng điều kiện E(Y/X) là hàm hồi quy của đại lượng ngẫu nhiên Y theo đại lượng ngẫu nhiên X.

Ký hiệu $y = E(Y/X = x) = \varphi(x)$.

Tương tự : gọi kỳ vọng có điều kiện $E(X/Y=y) = \psi(y)$ là hàm hồi quy của đại lượng ngẫu nhiên X theo đại lượng ngẫu nhiên Y.

Định nghĩa 7.3. Hàm hồi quy của Y theo X được gọi là tuyến tính nếu $E(Y/X) = aX + b$, trong đó a, b là hằng số. Tương tự : Hàm hồi quy của X theo Y được gọi là tuyến tính nếu :

$E(X/Y) = cY + d$, trong đó c, d là hằng số.

b) Phương pháp bình phương bé nhất

Phương pháp này dựa trên mệnh đề sau :

Mệnh đề : $E(Y - \varphi(X))^2 \to \min$ khi $\varphi(X) = E(Y/X)$.

Chứng minh :

$$E(Y - \varphi(X))^2 = E(Y - E(Y/X) + E(Y/X) - \varphi(X))^2 =$$

$$= E(Y - E(Y/X))^2 + E(E(Y/X) - \varphi(X))^2 + 2E(Y - E(Y/X)(E(Y/X) - \varphi(X))$$

Số hạng thứ ba bằng 0, còn hai số hạng còn lại là ≥ 0. Ta bỏ đi số $E(E(Y/X) - \varphi(X))^2$, nói cách khác, nếu chọn $\varphi(X) = E(Y/X)$ thì số hạng đó triệt tiêu. Vậy ta có :

$E(Y - \varphi(X))^2 \to \min$ khi $\varphi(X) = E(Y/X)$

Từ mệnh đề này ta rút ra phương pháp tìm hàm hồi quy.

Tìm $\varphi(X)$ để sao cho $E(Y - \varphi(X))^2 \to \min$. Phương pháp này được gọi là phương pháp bình phương bé nhất.

Ta sẽ ứng dụng phương pháp này để tìm hàm hồi quy tuyến tính.

Tìm a và b sao cho $Q = E(Y - aX - b)^2$ đạt cực tiểu. Đạo hàm của Q theo a, theo b và cho bằng 0 ta được hệ phương trình :

$$\begin{cases} -2E(Y - aX - b)X = 0 \\ -2E(Y - aX - b) = 0 \end{cases} \Leftrightarrow \begin{cases} aE(X^2) + bE(X) = E(XY) \\ aE(X) + b = E(Y) \end{cases}$$

Giải hệ này ta tìm được :

$$a = \frac{E(XY) - E(X)E(Y)}{E(X^2) - (EX)^2},\ b = E(Y) - aE(X).$$

Ta suy ra $a = \rho(X, Y)\sqrt{\dfrac{DY}{DX}}$.

Tương tự ta cũng tính được hàm hồi quy tuyến tính của X theo Y :

$$\hat{x} = cy + d,$$

trong đó $c = \dfrac{E(XY) - E(X)E(Y)}{E(Y^2) - (EY)^2}$; $d = E(X) - cE(Y)$

Nếu cho mẫu ngẫu nhiên $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ thì ta ước lượng được các hệ số a, b, c, d như sau :

$$\hat{a} = \frac{\frac{1}{n}\sum_1^n X_iY_i - \overline{X}\overline{Y}}{\frac{1}{n}\sum_1^n X_i^2 - (\overline{X})^2} = \frac{n\sum_{i=1}^n X_iY_i - (\sum_1^n X_i)(\sum_1^n Y_i)}{n\sum_{i=1}^n X_i^2 - (\sum_1^n X_i)^2}$$

$$\hat{b} = \overline{Y} - \hat{a}\overline{X}$$

và hàm hồi quy tuyến tính mẫu của Y theo X là :

$$\hat{y} = \hat{a}x + \hat{b}$$

Tương tự, ta cũng nhận được hàm hồi quy tuyến tính mẫu của X theo Y là :

$\hat{x} = \hat{c}y + \hat{d}$, trong đó :

$$\hat{c} = \frac{\frac{1}{n}\sum_{1}^{n} X_i Y_i - \overline{X}\,\overline{Y}}{\frac{1}{n}\sum_{1}^{n} Y_i^2 - (\overline{Y})^2} = \frac{n\sum_{1}^{n} X_i Y_i - (\sum_{1}^{n} X_i)(\sum_{1}^{n} Y_i)}{n\sum_{1}^{n} Y_i^2 - (\sum_{1}^{n} Y_i)^2}$$

$$\hat{d} = \overline{X} - \hat{c}\overline{Y}$$

Chú ý : Ta có thể tìm trực tiếp $\hat{a}, \hat{b}, \hat{c}, \hat{d}$ bằng cách buộc điều kiện $\sum_{i=1}^{n}(Y_i - aX_i - b)^2$ và $\sum_{i=1}^{n}(X_i - cY_i - d)^2$ đạt cực tiểu.

7.3.2. Tỷ số tương quan và độ sai dự báo

Định nghĩa 7.4. Nếu $DY > 0$ thì gọi $\sqrt{\frac{D(E(Y/X))}{DY}}$ là tỷ số tương quan của Y đối với X và ký hiệu :

$$K_X(Y) = \sqrt{\frac{D(E(Y/X))}{DY}}$$

Người ta chứng minh được rằng :

$$0 \le K_X(Y) \le 1$$

Trong lĩnh vực dự báo, người ta gọi $E(Y/X)$ là hàm dự báo của Y theo X và gọi $\delta = E(Y - E(Y/X))^2$ là độ sai dự báo.

Từ tính chất của kỳ vọng có điều kiện và đẳng thức :

$DY = E(Y - E(Y/X))^2 + D(E(Y/X))$ ta suy ra :

$$\delta = DY(1 - K_X^2(Y)) \tag{7.11}$$

Trong trường hợp hồi quy tuyến tính $E(Y/X) = aX + b$ thì :

$$K_X(Y) = \rho(X, Y)$$

Thực vậy, ta biết $a = \rho(X, Y)\sqrt{\frac{DY}{DX}}$, $b = E(Y) - aE(X)$.

Ta suy ra :

$$\begin{aligned} E(Y - (aX + b))^2 &= E(Y - E(Y) - a(X - E(X))^2 \\ &= E(Y - E(Y))^2 - 2aE(Y - E(Y))(X - E(X)) + a^2E(X - E(X))^2 \\ &= DY - 2DY\rho^2(X, Y) + DY\rho^2(X, Y) \end{aligned}$$

$$E(Y - (aX + b))^2 = DY(1 - \rho^2(X, Y)) \qquad (7.12)$$

Từ (7.11) ta suy ra trong trường hợp $E(Y/X) = aX + b$ rằng :

$$E(Y - (aX + b))^2 = DY(1 - K_X^2(Y)) \qquad (7.13)$$

So sánh (7.12) và (7.13) ta nhận được $\rho(X, Y) = K_X(Y)$. Vậy khi hồi quy $E(Y/X) = aX + b$ thì độ sai dự báo là :

$$\delta = DY(1 - \rho^2(X, Y)) \qquad (7.14)$$

Khi biết mẫu quan sát $(X_1, Y_1), ..., (X_n, Y_n)$ ta ước lượng độ sai dự báo là :

$$\hat{\delta} = S_n^2(Y)(1 - r^2) \qquad (7.15)$$

Ví dụ 7.7. Với số liệu trong ví dụ 4.9. Tìm hàm hồi quy tuyến tính mẫu của Y theo X. Tính độ sai dự báo mẫu của Y theo X.

Ta đã tính được trong ví dụ 4.9 :

$\sum X_i = 44$; $\sum Y_i = 88$; $\sum X_i^2 = 284$; $\sum Y_i^2 = 1142$; $\sum X_i Y_i = 568$

$r = 0{,}98$. Ta tính $\hat{a}$ và $\hat{b}$:

$$\hat{a} = \frac{n\sum X_i Y_i - (\sum X_i)(\sum Y_i)}{n\sum X_i^2 - (\sum X_i)^2} = \frac{8 \times 568 - 44 \times 88}{8 \times 284 - 44^2} = 2{,}00$$

$\hat{b} = 0{,}00$

Hàm hồi quy tuyến tính mẫu của Y theo X là $\hat{y} = 2x$.

Độ sai dự báo : $\hat{\delta} = S_n^2(Y)(1 - r^2) = 21{,}75(1 - 0{,}98^2)$

$$\hat{\delta} = 0{,}8613$$

7.3.3. Khoảng ước lượng và kiểm định giả thiết về hệ số hồi quy

a) Ta biết rằng ước lượng điểm của hệ số hồi quy a là :

$$\hat{a} = \frac{S_n(Y)}{S_n(X)} r$$

với n khá lớn thì phân phối của $\hat{a}$ xấp xỉ với phân phối chuẩn có kỳ vọng bằng a và phương sai bằng $\sigma_n^2 = \left(\frac{S_n(Y)}{S_n(X)} \cdot \frac{1 - r^2}{\sqrt{n}}\right)^2$. Do đó khoảng ước lượng của a với độ tin cậy $1 - \alpha$ là :

$$\hat{a} - x_\alpha \frac{S_n(Y)}{S_n(X)} \cdot \frac{1 - r^2}{\sqrt{n}} < a < \hat{a} + x_\alpha \frac{S_n(Y)}{S_n(X)} \cdot \frac{1 - r^2}{\sqrt{n}} \qquad (7.16)$$

trong đó x_α tra ở bảng phân phối chuẩn sao cho $\Phi(x_\alpha) = 1 - \frac{\alpha}{2}$.

* Với kích thước mẫu bé $n < 30$ thì khoảng ước lượng của a với độ tin cậy $1 - \alpha$ là :

$$\hat{a} - t_\alpha \frac{S_n(Y)}{S_n(X)} \cdot \frac{\sqrt{1 - r^2}}{\sqrt{n-2}} < a < \hat{a} + t_\alpha \frac{S_n(Y)}{S_n(X)} \cdot \frac{\sqrt{1 - r^2}}{\sqrt{n-2}} \qquad (7.17)$$

trong đó t_α tra ở bảng phân phối Student với $n - 2$ bậc tự do.

Ví dụ 7.8. Với số liệu cho trong ví dụ 4.10. Viết phương trình hồi quy tuyến tính mẫu của Y theo X và tìm khoảng ước lượng của hệ số hồi quy a với độ tin cậy 95%.

Giải : Ta đã tính được trong ví dụ 4.10 :

$\Sigma X_i = 56$; $\Sigma Y_i = 87$; $\Sigma X_i^2 = 122$; $\Sigma Y_i^2 = 273$; $\Sigma X_i Y_i = 179$

$\bar{X} = 1{,}87$; $\bar{Y} = 2{,}9$; $r = 0{,}87$.

Ta tính $\hat{a} = \dfrac{30 \times 179 - 56 \times 87}{30 \times 122 - (56)^2} \approx 0{,}95$.

$$\hat{b} = 2{,}9 - 0{,}95 \times 1{,}87 = 1{,}1235$$

Phương trình hồi quy tuyến tính mẫu của Y theo X là :

$$\hat{y} = 0{,}95x + 1{,}12$$

Tính $S_n^2(Y) = \dfrac{1}{30}\left(273 - \dfrac{87^2}{30}\right) = 0{,}69$; $S_n(Y) = 0{,}83$.

$S_n^2(X) = \dfrac{1}{30}\left(122 - \dfrac{56^2}{30}\right) = 0{,}58$; $S_n(X) = 0{,}76$. Tra bảng ta có $x_\alpha = 1{,}96$.

Vậy khoảng ước lượng của hệ số hồi quy a với độ tin cậy 95% là :

$$a = 0{,}95 \pm 1{,}96 \frac{0{,}83}{0{,}76} \cdot \frac{1 - 0{,}87^2}{\sqrt{30}} = 0{,}95 \pm 0{,}095$$

b) Kiểm định giả thiết về hệ số hồi quy :

$H_o : a = a_o$ với $K : a \neq a_o$ ở mức α.

Lời giải : Giả thiết H_o bị bác bỏ ở mức α nếu :

$$|Z| = \frac{|(\hat{a} - a_o)\sqrt{M}|}{S_I} > t_\alpha \qquad (7.18)$$

trong đó $M = \sum_{i=1}^{n}(X_i - \overline{X})^2$

$$S_I^2 = \frac{1}{n-2}\sum_{i=1}^{n}(Y_i - \overline{Y} - \hat{a}(X_i - \overline{X}))^2$$

$$= \frac{1}{n-2}\left[\sum_{1}^{n} Y_i^2 - \frac{(\sum_{1}^{n} Y_i)^2}{n} - \hat{a}^2\left(\sum_{1}^{n} X_i^2 - \frac{(\sum_{1}^{n} X_i)^2}{n}\right)\right]$$

S_I^2 có phân phối χ^2 với $n - 2$ bậc tự do.

t_α tra trong bảng phân phối Student với $n - 2$ bậc tự do và mức α (bảng tiêu chuẩn hai phía).

7.3.4. Hồi quy không tuyến tính

Xét hàm hồi quy dạng :

$$\hat{y} = E(Y/X) = a_k X^k + a_{k-1}X^{k-1} + ... + a_1 X + a_o$$

Giả sử cho mẫu ngẫu nhiên $(X_1, Y_1), ..., (X_n, Y_n)$. Tìm $a_o, a_1, ..., a_n$ bằng phương pháp bình phương bé nhất.

Đặt $Q = \sum_{i=1}^{n}(Y_i - a_o - a_1 X_i - ... - a_k X_i^k)^2$

Đạo hàm của Q theo các biến $a_o, a_1, ..., a_k$ và cho bằng 0 ta được hệ phương trình để tìm $a_o, ..., a_k$.

$$\begin{cases} a_o n + a_1\sum_{1}^{n} X_i + ... + a_k\sum_{1}^{n} X_i^k = \sum_{1}^{n} Y_i \\ ... \\ a_o\sum_{1}^{n} X_i^k + a_1\sum_{1}^{n} X_i^{k+1} + ... + a_k\sum_{1}^{n} X_i^{2k} = \sum_{i=1}^{n} X_i^k Y_i \end{cases}$$

Giải hệ phương trình này ta tìm được $\hat{a}_o, ..., \hat{a}_k$ và phương trình hồi quy mẫu của Y theo X là :

$$\hat{y} = \hat{a}_o + \hat{a}_1 x + ... + \hat{a}_k x^k$$

Ví dụ 7.9. Cho mẫu quan sát đối với cặp biến ngẫu nhiên X, Y là :

Y \ X	1	2	3	4
5				1
4		3	6	6
3	3	8	6	3
2	5	4	3	
1	2			

Tìm hàm hồi quy mẫu của Y theo X dạng :

$$y = a_o + a_1x + a_2x^2$$

Giải : Để tính toán dễ dàng ta lập bảng phụ sau đây :

Y \ X	1	2	3	4	n_y	n_yy	$n_{xy}xy$	$n_{xy}x^2y$
5				1	1	5	20	80
4		3	6	6	15	60	192	648
3	3	8	6	3	20	60	147	411
2	5	4	3		12	24	44	96
1	2				2	2	2	2
n_x	10	15	15	10	50	151	405	1237
n_xx	10	30	45	40	125			
n_xx^2	10	60	135	160	365			
n_xx^3	10	120	405	640	1175			
n_xx^4	10	240	1215	2560	4025			

Trong trường hợp ví dụ này ta có hệ phương trình :

$$50a_o + 125a_1 + 365a_2 = 151$$

$$125a_o + 365a_1 + 1175a_2 = 405$$

$$365a_o + 1175a_1 + 4025a_2 = 1237$$

Giải ra ta tìm được $\hat{a}_o = 1,407$; $\hat{a}_1 = 0,815$; $\hat{a}_2 = -0,058$.

Vậy phương trình đường hồi quy mẫu của Y theo X là :

$$y = 1,407 + 0,815x - 0,058x^2$$

7.3.5. Hồi quy tuyến tính nhiều biến

Định nghĩa 7.5. Gọi kỳ vọng có điều kiện $E(Y/X_1, ..., X_n)$ là hàm hồi quy của biến ngẫu nhiên Y theo các biến ngẫu nhiên $X_1, ..., X_n$.

Định nghĩa 7.6. Hàm hồi quy của Y theo $X_1, X_2, ..., X_n$ được gọi là tuyến tính nếu $E(Y/X_1, ..., X_n) = a_o + a_1X_1 + a_2X_2 + ... + a_nX_n$ cũng như trường hợp hàm hồi quy của Y theo X, ta cũng có thể áp dụng phương pháp bình phương bé nhất để tìm các hệ số $a_o, a_1, ..., a_n$.

Giả sử cho mẫu ngẫu nhiên :

$(Y_1, X_{11}, ..., X_{n1}), ... (Y_N, X_{1N}, ..., X_{nN})$. Tìm $a_o, a_1, ..., a_n$ sao cho $Q_o = \sum_{i=1}^{N}(Y_i - a_o - a_1X_{1i} - ... - a_nX_{ni})^2$ đạt cực tiểu. Đạo hàm Q_o theo các biến $a_o, ..., a_n$ và cho bằng 0 ta được hệ phương trình để tìm $a_o, a_1, ..., a_n$.

$$\begin{cases} Na_o + a_1\sum_{i=1}^{N}X_{1i} + ... + a_n\sum_{i=1}^{N}X_{ni} = \sum_{i=1}^{N}Y_i \\ ..... \\ a_o\sum_{i=1}^{N}X_{ni} + a_1\sum_{i=1}^{N}X_{1i}X_{ni} + ... + a_n\sum_{i=1}^{N}X_{ni}^2 = \sum_{i=1}^{N}X_{ni}Y_i \end{cases}$$

Giải hệ này ta được $\hat{a}_o, \hat{a}_1, ..., \hat{a}_n$ và hàm hồi quy tuyến tính mẫu của Y theo $X_1, ..., X_n$ là :

$$y = \hat{a}_o + \hat{a}_1x_1 + ... + \hat{a}_nx_n$$

– Để tính toán đơn giản ta thường đổi biến số để đưa về phương trình dạng chính tắc và ta được dạng mới của phương trình hồi quy :

Đặt $z = \dfrac{Y - EY}{\sqrt{DY}}, t_1 = \dfrac{X_1 - EX_1}{\sqrt{DX_1}}, ..., t_n = \dfrac{X_n - EX_n}{\sqrt{DX_n}}$

Phương trình hồi quy tuyến tính đối với biến mới :

$$\hat{z} = \beta_1t_1 + \beta_2t_2 + ... + \beta_nt_n$$

Dùng phương pháp bình phương bé nhất để tìm $\beta_1, ..., \beta_n$,

ta có $a_i = \beta_i\sqrt{\dfrac{DY}{DX_i}}$, trong đó $\beta_i = \dfrac{D_i}{D}$,

$$D = \begin{vmatrix} 1 & r_{12} & \dots & r_{1n} \\ r_{21} & 1 & \dots & \\ \dots\dots & & & \\ r_{n1} & r_{n2} & \dots & 1 \end{vmatrix}$$, D_i nhận được từ D bằng cách bỏ đi cột i thay bằng cột hệ số tự do.

Ta có thể chi tiết hóa việc tìm các $\beta_i, \dots, \beta_n$ như sau :

Đặt $Q_o = \sum_{i=1}^{N}(Y_i - \hat{\beta}_1 t_{1i} - \dots - \hat{\beta}_n t_{ni})^2$

đạo hàm Q_o theo các $\hat{\beta}_i$ và cho bằng 0 ta được :

$$\begin{cases} \hat{\beta}_1 \sum_{i=1}^{N} t_{i1}^2 + \dots + \hat{\beta}_n \sum_{i=1}^{N} t_{ni} t_{1i} = \sum_{i=1}^{N} t_{1i} Z_i \\ \dots\dots \\ \hat{\beta}_1 \sum_{i=1}^{N} t_{1i} t_{ni} + \dots + \hat{\beta}_n \sum_{i=1}^{N} t_{ni}^2 = \sum_{i=1}^{N} t_{ni} Z_i \end{cases}$$

Đặt $h_j = \sum_{i=1}^{N} Z_i t_{ji}$; $g_{ij} = \sum_{k=1}^{N} t_{ik} t_{jk}$

hệ phương trình trên đưa về dạng :

$$\begin{cases} g_{11}\hat{\beta}_1 + \dots + g_{1n}\hat{\beta}_n = h_1 \\ \dots\dots \\ g_{n1}\hat{\beta}_1 + \dots + g_{nn}\hat{\beta}_n = h_n \end{cases}$$

Giải hệ này ta được $\hat{\beta}_1 = \sum_{i=1}^{N} a_{i1} h_1, \dots, \hat{\beta}_n = \sum_{i=1}^{N} a_{in} h_n$

trong đó a_{ij} thỏa mãn điều kiện :

$$\sum_{r=1} g_{rj} a_{ij} = \begin{cases} 1 & \text{với } r = i \\ 0 & \text{với } r \neq i \end{cases}$$

* Khoảng ước lượng của hệ số hồi quy β_i với độ tin cậy $1 - \alpha$ là :

$$\hat{\beta}_i - t_\alpha \sqrt{a_{ii}}\, S < \beta_i < \hat{\beta}_i + t_\alpha \sqrt{a_{ii}}\, S, i = \overline{1, n}$$

t_α tra ở bảng phân phối Student với $N - n$ bậc tự do.

$$S^2 = \frac{Q_o}{N - n}$$

* Kiểm định giả thiết về hệ số hồi quy β_i, $i = \overline{1, n}$

$$H_o : \beta_i = \beta_i^o \qquad i = \overline{1, n}$$

$$K : \beta_i \neq \beta_i^o$$

Lời giải :

Giả thiết H_o bị bác bỏ ở mức α nếu :

$$|Z| = \frac{|\hat{\beta}_i - \beta_i^o|}{\sqrt{a_{ii}}\,S} > t_\alpha$$

còn nếu $|Z| < t_\alpha$ thì ngược lại.

7.3.6. Hệ số tương quan

a) Hệ số tương quan của X, Y

Hệ số tương quan mẫu của X, Y :

$$r = \frac{\frac{1}{n}\sum_{i=1}^{n} X_i Y_i - \overline{X}\overline{Y}}{S_n(X)S_n(Y)} = \frac{n \times \sum_{i=1}^{n} X_i Y_i - (\sum_{i=1}^{n} X_i)(\sum_{i=1}^{n} Y_i)}{\sqrt{\left[n\sum_{i=1}^{n} X_i^2 - (\sum_{i=1}^{n} X_i)^2\right]\left[n\sum_{i=1}^{n} Y_i^2 - (\sum_{i=1}^{n} Y_i)^2\right]}}$$

Với kích thước mẫu n rất lớn, r có phân phối xấp xỉ phân phối chuẩn với kỳ vọng ρ và phương sai $d_r^2 = \dfrac{(1 - r^2)^2}{n}$.

– Khoảng ước lượng của hệ số tương quan với độ tin cậy $1 - \alpha$ là :

$$r - x_\alpha \frac{1 - r^2}{\sqrt{n}} < \rho < r + x_\alpha \frac{1 - r^2}{\sqrt{n}}$$

trong đó x_α tra ở bảng phân phối chuẩn N(0 ; 1) sao cho $\Phi(x_\alpha) = 1 - \dfrac{\alpha}{2}$

* Trường hợp kích thước mẫu bé $n < 30$:

Đặt $r = \text{th}Z$ (tang hyperbolic của Z), trong đó Z là biến ngẫu nhiên có phân phối không phụ thuộc vào ρ và n ; nếu $n \to \infty$ thì phân phối Z tiến tới phân phối chuẩn với kỳ vọng :

$$EZ = \frac{1}{2}\ln\frac{1 + \rho}{1 - \rho} + \frac{\rho}{2(n - 1)}$$

và phương sai $DZ \approx \dfrac{1}{n - 3}$ với $Z = \dfrac{1}{2}\ln\dfrac{1 + r}{1 - r}$ (hàm ngược).

Từ đó suy ra khoảng ước lượng của $\frac{1}{2}\ln\frac{1+\rho}{1-\rho}$ với độ tin cậy $1-\alpha$ là :

$$Z - \frac{r}{2(n-1)} - \frac{x_\alpha}{n-3} < \frac{1}{2}\ln\frac{1+\rho}{1-\rho} < Z - \frac{r}{2(n-1)} + \frac{x_\alpha}{n-3}$$

trong đó x_α tra ở bảng phân phối chuẩn N(0 ; 1) sao cho $\Phi(x_\alpha) = 1 - \frac{\alpha}{2}$.

– Kiểm định giả thiết về hệ số tương quan :

$H_o : \rho = 0$ với $K : \rho \neq 0$ ở mức α

Lời giải :

Giả thiết H_o bị bác bỏ ở mức α nếu :

$|Z| = \frac{|r|}{d_r} > x_\alpha$, còn $|Z| < x_\alpha$ thì chấp nhận H_o.

x_α tra ở bảng phân phối chuẩn N(0 ; 1) sao cho $\Phi(x_\alpha) = 1 - \frac{\alpha}{2}$.

Ví dụ 7.10. Với số liệu cho trong ví dụ 4.11. Tìm khoảng ước lượng của hệ số tương quan ρ với độ tin cậy 95%. Kiểm định giả thiết $H_o : \rho = 0$ với $K : \rho \neq 0$ ở mức $\alpha = 5\%$.

Giải :

* Trong ví dụ 4.11 ta đã tìm được r = 0,77 ; n = 200 và tra ở bảng phân phối chuẩn ta tìm được $x_\alpha = 1{,}96$. Vậy khoảng ước lượng của hệ số tương quan ρ với độ tin cậy 95% là :

$$0{,}77 - 1{,}96\,\frac{1-0{,}77^2}{\sqrt{200}} < \rho < 0{,}77 + 1{,}96\,\frac{1-0{,}77^2}{\sqrt{200}}$$

$$0{,}7136 < \rho < 0{,}8264$$

* Kiểm định giả thiết $H_o : \rho = 0$ với $K : \rho \neq 0$ ở mức $\alpha = 5\%$.

Nếu $|Z| = \frac{|r|}{d_r} > x_\alpha$ thì bác bỏ giả thiết H_o hoặc $|r| > x_\alpha dr$ thì bác bỏ giả thiết H_o.

Tính $x_\alpha d_r = 1{,}96.\,\frac{1-0{,}77^2}{\sqrt{200}} \approx 0{,}0564$

Ta thấy $r = 0{,}77 > x_\alpha d_r = 0{,}0564$. Vậy giả thiết H_o bị bác bỏ.

b) Hệ số tương quan bội (tương quan nhiều chiều)

Định nghĩa 7.7. Gọi căn bậc hai của tỷ số $\frac{D(E(Y / X_1, ..., X_n))}{DY}$ là hệ số tương quan bội của Y theo $X_1, ..., X_n$ và ký hiệu :

$$R = \sqrt{\frac{D(E(Y / X_1, ..., X_n))}{DY}}$$

Người ta chứng minh được rằng :

$+ \; 0 \le R \le 1$

$+ \; R = \sqrt{\beta_1 r_{YX_1} + \beta_2 r_{YX_2} + ... + \beta_n r_{YX_n}}$

$+ \; R = \sqrt{A^* / A}$

trong đó : $A = (-1)^{n+1}\begin{vmatrix} 1 & r_{12} & ... & r_{1n} \\ r_{21} & 1 & ... & r_{2n} \\ .... & & & \\ r_{n1} & r_{n2} & ... & 1 \end{vmatrix}$

Định thức A* nhận được từ A bằng cách bổ sung thêm 1 hàng ở trên cùng và 1 cột ở ngoài cùng bên trái phần tử của nó là những hệ số tương quan đôi một giữa Y và X_i, $i = \overline{1, n}$.

c) Hệ số tương quan riêng

Định nghĩa 7.8. Hệ số tương quan giữa Y và $X_1, ..., X_p$ khi cố định $X_{p+1}, ... X_n$ được gọi là hệ số tương quan riêng của $Y, X_1, ..., X_p$.

Ký hiệu $r_{yx_1...x_p.x_{p+1}...x_n}$ (hoặc viết tắt $r_{012...p.(p+1)...n}$)

Ta có công thức :

$$r_{ij.k} = \frac{r_{ij} - r_{ik}r_{jk}}{\sqrt{(1 - r_{ik}^2)(1 - r_{jk}^2)}}$$

Bài tập chương VII

1. Để khảo sát độ thuần nhất về mặt hướng của một cánh đồng lúa, người ta ngăn ra 3 băng song song và trong mỗi băng lấy ra 5 mảnh một cách ngẫu nhiên, mỗi mảnh có diện tích một a ($100m^2$). Vì lý do khác nhau (sương giá, cây cỏ mọc tự nhiên, ký sinh trùng), kết quả của 3 mảnh được coi là vô giá trị và bị loại bỏ. Các kết quả giữ lại là :

Sản lượng các mảnh						
Băng	1	46	42	48		
	2	57	53	43	54	48
	3	50	41	47	42	

Giả sử các quan sát trên có phân phối chuẩn, có phương sai chung σ^2 ; mỗi quan sát ở mỗi băng có cùng kỳ vọng toán. Hãy kiểm định giả thiết H_o : "Cánh đồng thuần nhất theo hướng vuông góc với các băng" ở mức $\alpha = 0{,}05$.

2. Để nghiên cứu hiệu quả của một loại thức ăn cho trẻ em, người ta chia làm 5 chế độ ăn khác nhau : A_1, A_2, A_2, A_4, A_5. Kết quả của thức ăn thể hiện ở độ tăng trọng của trẻ em. Số liệu đo được ở 14 cháu như sau :

Mức ăn				
A_1	A_2	A_3	A_4	A_5
7,3	5,4	6,4	7,9	7,1
7,6	7,1	8,1	9,5	
8,3	7,4		9,6	
8,3				
8,4				

Xét xem thức ăn có tác dụng tăng trọng cho trẻ em không ? Cho mức $\alpha = 0{,}05$.

3. Để nghiên cứu hiệu quả của ba phương pháp dạy trẻ ở một nhà trẻ, người ta chia các cháu làm 4 nhóm theo lứa tuổi. Sau một kỳ giảng

dạy, người ta tiến hành kiểm tra sự phát triển tư duy, phát triển vốn từ vựng. Kết quả kiểm tra được cho ở bảng sau :

Phương pháp		A	B	C
Nhóm	1	30	20	21
	2	36	24	22
	3	35	26	31
	4	31	30	34

Hỏi hiệu quả của ba phương pháp đó có giống nhau không ? Cho mức $\alpha = 5\%$.

4. Để nghiên cứu sự ảnh hưởng của trình độ văn minh, kiến thức văn hóa đến sự phát triển dân số, người ta thí nghiệm trên 4 lớp người có trình độ khác nhau A, B, C, D ở trên 4 vùng khác nhau, có mức sống kinh tế khác nhau. Kết quả thực nghiệm được cho ở bảng sau :

	Kinh tế I	Kinh tế II	Kinh tế III	Kinh tế IV
Vùng 1	A : 52	B : 40	C : 30	D : 65
Vùng 2	B : 49	C : 55	D : 40	A : 60
Vùng 3	C : 40	D : 40	A : 50	B : 45
Vùng 4	D : 35	A : 55	B : 36	C : 55

Sử dụng phương pháp hình vuông Latinh để kiểm định giả thiết H_0 : "Trình độ văn minh và kiến thức văn hóa, mức kinh tế không ảnh hưởng đến sự phát triển dân số". Cho mức $\alpha = 5\%$.

5. Để so sánh sản lượng sữa của 4 loại bò A, B, C, D, người ta làm thí nghiệm ở 4 vùng và cho ăn theo 4 mức khác nhau. Kết quả đo được (tính ra lýt) trong một khoảng thời gian nhất định được cho ở bảng sau đây :

	Thức ăn I	Thức ăn II	Thức ăn III	Thức ăn IV
Vùng 1	C : 219	D : 166	D : 238,3	A : 217,7
Vùng 2	D : 190	C : 214,3	A : 245,2	B : 203
Vùng 3	A : 232	B : 181,8	D : 221,8	C : 190
Vùng 4	B : 191,3	A : 219,5	C : 190,3	D : 203,3

Sử dụng phương pháp hình vuông Latinh để kiểm định giả thiết H_o : "Sản lượng sữa của 4 loại bò là như nhau". Cho $\alpha = 0,05$.

6. Để nghiên cứu năng suất của cà chua (tính ra kg) người ta xét hai yếu tố : giống và phân. Người ta thí nghiệm trên 2 loại giống A_1, A_2 và 2 mức phân B_1, B_2. Kết quả thí nghiệm được cho ở bảng sau đây :

Giống A \ Phân B	B_1	B_2
A_1	120 – 123	118 – 119
A_2	125 – 123	118 – 118

Xét sự ảnh hưởng của yếu tố giống (A) đến năng suất cà chua. Xét sự ảnh hưởng của phân (B) đến năng suất cà chua. Xét sự ảnh hưởng chung cả phân và giống đến năng suất cà chua. Cho mức kiểm định $\alpha = 5\%$.

7. Để nghiên cứu sự ảnh hưởng của thức ăn B và sự phân nhóm để chăn nuôi đến sự tăng trọng của lợn, người ta làm thí nghiệm trên 45 con lợn. Kết quả tăng trọng sau 3 tuần người ta cân được như sau :

Nhóm \ Thức ăn	B_1	B_2	B_3
A_1	8 6 4 5 9	5 4 4 6 3	2 1 0 1 2
A_2	5 6 5 5 6	4 3 4 4 5	3 4 3 4 4
A_3	7 6 7 7 8	4 3 2 3 4	2 0 1 2 1

Kiểm định các giả thiết sau :

H_o : "Thức ăn không ảnh hưởng đến sự tăng trọng của lợn".

H_o' : "Việc phân nhóm không ảnh hưởng đến sự tăng trọng của lợn".

H_o'' : "Việc phân nhóm và thức ăn không ảnh hưởng đến sự tăng trọng của lợn". Cho mức $\alpha = 0,05$.

Hồi quy và tương quan :

8. Tìm hàm hồi quy tuyến tính mẫu của Y theo X và tính hệ số tương quan mẫu của X, Y trên cơ sở các mẫu quan sát sau :

a)

X	2	3	7	6	5
Y	5	6	15	13	10

b)

Y \ X	1	3	5
2	10	2	
3	5	5	
4		8	10

c)

Y	X							
	5	10	15	20	25	30	35	40
100	2	1						
120	3	4	3					
140			5	10	8			
160						6	1	1
180							4	1

9. Tìm hồi quy bậc hai của Y theo X, $\hat{y} = ax^2 + bx + c$ và hệ số tương quan mẫu của X, Y dựa trên mẫu quan sát sau đây

Y \ X	2	3	5
25	20		
45		30	
110			48

10. Tìm hàm hồi quy tuyến tính mẫu của Z theo X, Y dựa trên mẫu quan sát sau đây :

$$Z = \begin{bmatrix} 1 \\ 3 \\ 2 \\ 3 \\ 0 \\ 2 \end{bmatrix}, \qquad (X, Y) = \begin{bmatrix} 1 & 2 \\ 2 & 3 \\ 3 & 2 \\ 4 & 2 \\ 2 & 1 \\ 2 & 2 \end{bmatrix}$$

Hướng dẫn và trả lời bài tập chương VII

1. $f_\alpha(5\% ; 2 ; 9) = 4{,}26$; $\overline{X}_{\bullet 1} = 45{,}33$, $\overline{X}_{\bullet 2} = 51{,}00$; $b_1 = -2{,}25$;
$b_2 = 3{,}417$; $Q = 294{,}92$; $Q_A = 100{,}25$; $Q_R = 194{,}67$;
$F_A = 2{,}32 < f_\alpha = 4{,}26$;
$\overline{X}_{\bullet 3} = 45{,}00$; $b_3 = -2{,}583$. Chấp nhận giả thiết H_o.

2. $f(5\% ; 4 ; 10) = 3{,}48$; $\overline{X} = 7{,}743$; $\overline{X}_{\bullet 1} = 7{,}98$; $\overline{X}_{\bullet 2} = 6{,}633$
$\overline{X}_{\bullet 3} = 7{,}25$; $\overline{X}_{\bullet 4} = 9{,}00$; $\overline{X}_{\bullet 5} = 7{,}10$; $b_1 = 0{,}237$; $b_2 = -1{,}11$;
$b_3 = -0{,}493$; $b_4 = 1{,}257$; $b_5 = -0{,}643$; $Q = 16{,}19$; $Q_A = 9{,}61$;
$Q_R = 6{,}58$; $F_A = 3{,}29 < f_\alpha = 3{,}48$. Chấp nhận giả thiết H_o.

3. $f_A(5\% ; 2 ; 6) = 5{,}14$; $f_B(5\% ; 3 ; 6) = 4{,}76$; $\overline{X} = 28{,}33$; $\overline{X}_{1\bullet} = 23{,}67$;
$\overline{X}_{2\bullet} = 27{,}33$; $\overline{X}_{3\bullet} = 30{,}67$; $\overline{X}_{4\bullet} = 31{,}67$; $\overline{X}_{\bullet 1} = 33$; $\overline{X}_{\bullet 2} = 25$;
$\overline{X}_{\bullet 3} = 27$; $b_1 = 4{,}67$; $b_2 = -3{,}33$; $b_3 = -1{,}33$; $a_1 = -4{,}63$;
$a_2 = -1{,}00$; $a_3 = 2{,}33$; $a_4 = 3{,}33$; $Q = 342{,}67$; $Q_A = 138{,}67$;
$Q_B = 118{,}00$; $Q_R = 86{,}00$.
$F_A = 4{,}84 < f_A = 5{,}14$. Chấp nhận giả thiết H_o.
$F_B = 2{,}74 < f_B = 4{,}76$. Chấp nhận giả thiết H_o'.

4. $f_Đ\,(5\% ; 3 ; 6) = 4{,}76$; $\overline{X} = 46{,}687$; $Q = 1475{,}44$; $Q_H = 117{,}19$;
$Q_C = 633{,}69$; $Q_Đ = 321{,}69$; $Q_R = 402{,}87$;
$F_Đ = 1{,}6 < f_Đ = 4{,}76$. Chấp nhận giả thiết H_o.

5. $\overline{X} = 207{,}719$; $Q = 7191{,}84$; $Q_H = 324{,}73$;
$Q_C = 1726{,}04$; $Q_Đ = 2505{,}96$; $Q_R = 2635{,}12$;
$F_Đ = 1{,}90 < f_Đ = 4{,}76$. Chấp nhận giả thiết H_o.

6. $n = 8$, $r = 2$, $v = 2$, $s = 2$

$\overline{X}_{11} = 121{,}5$	$\overline{X}_{12} = 118{,}5$	$\overline{\overline{X}}_{1\bullet} = 120$
$\overline{X}_{21} = 124$	$\overline{X}_{22} = 118$	$\overline{\overline{X}}_{2\bullet} = 121$
$\overline{\overline{X}}_{\bullet 1} = 122{,}75$	$\overline{\overline{X}}_{\bullet 2} = 118{,}25$	$\overline{\overline{X}} = 120{,}5$

$Q = 54$; $Q_B = 40{,}5$; $Q_A = 2{,}00$; $Q_{AB} = 4{,}5$; $Q_R = 7{,}0$;

$F_A = 1{,}14 < f_\alpha = 7{,}71$. Chấp nhận giả thiết H_o ;

$F_B = 23{,}14 > f'_\alpha = 7{,}71$. Bác bỏ giả thiết H'_o ;

$F_{AB} = 2{,}57 < f''_\alpha = 7{,}71$. Chấp nhận H''_o.

7. $r = 3$; $v = 3$; $s = 5$

$f_A(5\% ; 2 ; 36) = 3{,}27$; $f'_B(5\% ; 2 ; 36) = 3{,}27$; $f''_{AB}(5\% ; 4 ; 36) = 2{,}65$

$\overline{X}_{11} = 6{,}4$	$\overline{X}_{12} = 4{,}4$	$\overline{X}_{13} = 1{,}2$	$\overline{\overline{X}}_{1\bullet} = 4{,}0$
$\overline{X}_{21} = 5{,}4$	$\overline{X}_{22} = 4{,}0$	$\overline{X}_{23} = 3{,}6$	$\overline{\overline{X}}_{2\bullet} = 4{,}33$
$\overline{X}_{31} = 7{,}0$	$\overline{X}_{32} = 3{,}2$	$\overline{X}_{33} = 1{,}2$	$\overline{\overline{X}}_{3\bullet} = 3{,}8$
$\overline{\overline{X}}_{\bullet 1} = 6{,}27$	$\overline{\overline{X}}_{\bullet 2} = 3{,}87$	$\overline{\overline{X}}_{\bullet 3} = 2{,}0$	$\overline{\overline{X}} = 4{,}04$

$Q = 203{,}91$; $Q_B = 137{,}24$; $Q_A = 2{,}18$; $Q_{AB} = 27{,}29$; $Q_R = 37{,}2$

$F_A = 1{,}05 < f_\alpha = 3{,}27$. Chấp nhận giả thiết H_o ;

$F_B = 66{,}71 > f'_\alpha = 3{,}27$. Bác bỏ giả thiết H'_o ;

$F_{AB} = 6{,}60 < f''_\alpha = 2{,}65$. Bác bỏ giả thiết H''_o.

8. a) $\overline{X} = 4{,}6$; $\overline{Y} = 9{,}8$; $r = 0{,}99$; $\hat{y} = 2{,}07x + 0{,}28$

b) $\overline{X} = 2{,}75$; $\overline{Y} = 3{,}15$; $r = 0{,}78$; $\hat{y} = 0{,}43x + 1{,}98$

c) $\overline{X} = 21{,}02$; $\overline{Y} = 140{,}82$; $r = 0{,}91$; $\hat{y} = 1{,}92x + 100{,}36$.

9. $\hat{y} = 2{,}94x^2 + 7{,}27x - 1{,}25$.

10. $\hat{z} = 16{,}3x + 1{,}46y - 0{,}92$.

PHỤ LỤC

Bảng 1. Giá trị hàm $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{\frac{u^2}{2}} du$

x	Φ(x)	x	Φ(x)	x	Φ(x)	x	Φ(x)
0,00	0,0000	0,20	0,0793	0,40	0,1554	0,60	0,2257
0,01	0,0040	0,21	0,0832	0,41	0,1591	0,61	0,2291
0,02	0,0080	0,22	0,0871	0,42	0,1628	0,62	0,2324
0,03	0,0120	0,23	0,0910	0,43	0,1664	0,63	0,2357
0,04	0,0160	0,24	0,0948	0,44	0,1700	0,64	0,2389
0,05	0,0199	0,25	0,0987	0,45	0,1736	0,65	0,2422
0,06	0,0239	0,26	0,1026	0,46	0,1772	0,66	0,2454
0,07	0,0279	0,27	0,1064	0,47	0,1808	0,67	0,2486
0,08	0,0319	0,28	0,1103	0,48	0,1844	0,68	0,2517
0,09	0,0359	0,29	0,1141	0,49	0,1879	0,69	0,2549
0,10	0,0398	0,30	0,1179	0,50	0,1915	0,70	0,2580
0,11	0,0438	0,31	0,1217	0,51	0,1950	0,71	0,2611
0,12	0,0478	0,32	0,1255	0,52	0,1985	0,72	0,2642
0,13	0,0517	0,33	0,1293	0,53	0,2019	0,73	0,2673
0,14	0,0557	0,34	0,1331	0,54	0,2054	0,74	0,2703
0,15	0,0596	0,35	0,1368	0,55	0,2088	0,75	0,2734
0,16	0,0636	0,36	0,1406	0,56	0,2123	0,76	0,2764
0,17	0,0675	0,37	0,1443	0,57	0,2157	0,77	0,2794
0,18	0,0714	0,38	0,180	0,58	0,2190	0,78	0,2823
0,19	0,0753	0,39	0,1517	0,59	0,2224	0,79	0,2852

x	Φ(x)	x	Φ(x)	x	Φ(x)	x	Φ(x)
0.80	0.2881	1.01	0.3438	1.22	0.3883	1.43	0.4236
0.81	0.2910	1.02	0.3461	1.23	0.3907	1.44	0.4251
0.82	0.2939	1.03	0.3485	1.24	0.3925	1.45	0.4265
0.83	0.2967	1.04	0.3508	1.25	0.3944	1.46	0.4279
0.84	0.2995	1.05	0.3631	1.26	0.3962	1.47	0.4292
0.85	0.3023	1.06	0.3554	1.27	0.3980	1.48	0.4306
0.86	0.3051	1.07	0.3577	1.28	0.3997	1.49	0.4319
0.87	0.3078	1.08	0.3599	1.29	0.4015	1.50	0.4332
0.88	0.3106	1.09	0.3621	1.30	0.4032	1.51	0.4345
0.89	0.3133	1.10	0.3643	1.31	0.4049	1.52	0.4357
0.90	0.3159	1.11	0.3665	1.32	0.4066	1.53	0.4370
0.91	0.3186	1.12	0.3686	1.33	0.4082	1.54	0.4382
0.92	0.3212	1.13	0.3708	1.34	0.4099	1.55	0.4394
0.93	0.3238	1.14	0.3729	1.35	0.4115	1.56	0.4406
0.94	0.3264	1.15	0.3749	1.36	0.4131	1.57	0.4418
0.95	0.3289	1.16	0.3770	1.37	0.4147	1.58	0.4429
0.96	0.3315	1.17	.03790	1.38	0.4162	1.59	0.4441
0.97	0.3340	1.18	0.3010	1.39	0.4177	1.60	0.4452
0.98	0.3365	1.19	0.3830	1.40	0.4192	1.62	0.4463
0.99	0.3389	1.20	0.3849	1.41	0.4207	1.63	0.4474
1.00	0.3413	1.21	0.3869	1.42	0.4222	1.64	0.4484

x	Φ(x)	x	Φ(x)	x	Φ(x)	x	Φ(x)
1.64	0.4495	1.88	0.4699	2.24	0.4875	2.72	0.4967
1.65	0.4505	1.89	0.4706	2.26	0.4881	2.74	0.4969
1.66	0.4515	1.90	0.4713	2.28	0.4887	2.76	0.4971
1.67	0.4525	1.91	0.4719	2.30	0.4893	2.78	0.4973
1.68	0.4535	1.92	0.4726	2.32	0.4898	2.80	0.4974
1.69	0.4545	1.93	0.4732	2.34	0.4904	2.82	0.4976
1.70	0.4554	1.94	0.4738	2.36	0.4909	2.84	0.4977
1.71	0.4564	1.95	0.4744	2.38	04913	2.86	0.4979
1.72	0.4573	1.96	0.4750	2.40	0.4918	2.88	0.4980
1.73	0.4582	1.97	0.4756	2.42	0.4922	2.90	0.4981
1.74	0.4591	1.98	0.4762	2.44	0.4927	2.92	0.4982
1.75	0.4599	1.99	0.4767	2.46	0.4931	2.94	0.4984
1.76	0.4608	2.00	0.4772	2.48	0.4934	2.96	0.4985
1.77	0.4616	2.02	0.4783	2.50	0.4938	2.98	0.4986
1.78	0.4625	2.04	0.4793	2.52	0.4941	3.00	0.49865
1.79	0.4633	2.06	0.4803	2.54	0.4945	3.20	0.49931
1.80	0.4641	2.08	0.4812	2.56	0.4948	3.40	0.49966
1.81	0.4649	2.10	0.4821	2.58	0.4951	3.60	0.499841
1.82	0.4656	2.12	0.4830	2.60	0.4953	3.80	0.499928
1.83	0.4664	2.14	0.4838	2.62	0.4956	4.00	0.499968
1.84	0.4671	2.16	0.4846	2.64	0.4959	4.50	0.499997
1.85	0.4678	2.18	0.4854	2.66	0.4961	5.00	0.499997
1.86	0.4686	2.20	0.4861	2.68	0.4963		
1.87	0.4693	2.22	0.4868	2.70	0.4965		

Bảng 2. Phân phối Sudent P[|T| > T_α] = α

Số bậc tự do	Mức ý nghĩa α (tiêu chuẩn hai phía)					
k	0.10	0.05	0.02	0.01	0.002	0.001
1	6.31	12.7	31.82	63.7	318.3	637.0
2	2.92	4.3	6.97	9.92	22.33	31.6
3	2.33	3.18	4.54	5.84	10.22	12.9
4	2.13	2.78	3.75	4.60	7.17	8.61
5	2.01	2.57	3.37	4.03	5.89	6.86
6	1.94	2.45	3.14	3.71	5.21	5.96
7	1.89	2.36	3.00	3.50	4.79	5.40
8	1.86	2.31	2.90	3.36	4.50	5.04
9	1.83	2.26	2.82	3.25	4.30	4.78
10	1.81	2.23	2.76	3.17	4.14	4.59
11	1.80	2.18	2.72	3.11	4.03	4.44
12	1.78	2.16	2.68	3.05	3.93	4.32
13	1.77	2.14	2.65	3.01	3.85	4.22
14	1.76	2.13	2.62	2.98	3.79	4.14
15	1.75	2.12	2.60	2.95	3.73	4.07
16	1.75	2.11	2.58	2.92	3.69	4.01
17	1.71	2.10	2.57	2.90	3.65	3.96
18	1.73	2.09	2.55	2.88	3.61	3.92
19	1.73	2.09	2.54	2.86	3.58	3.88
20	1.73	2.08	2.53	2.85	3.55	3.85
21	1.72	2.07	2.52	2.83	3.53	3.82
22	1.72	2.07	2.51	2.82	3.51	3.79
23	1.71	2.06	2.50	2.81	3:49	3.77
24	1.71	2.06	2.49	2.80	3.47	3.74
25	1.71	2.06	2.49	2.79	3.45	3.72
26	1.71	2.05	2.48	2.78	3.44	3.71
27	1.71	2.05	2.47	2.77	3.42	3.69
28	1.70	2.05	2.46	2.76	3.40	3.66
29	1.70	2.04	2.46	2.76	3.40	3.66
30	1.70	2.02	2.42	2.75	3.39	3.65
40	1.68	2.02	2.42	2.70	3.31	3.55
60	1.67	2.00	2.39	2.66	3.23	3.46
120	1.66	1.98	2.36	2.62	3.17	3.37
∞	1.64	1.96	2.33	2.58	3.09	3.29
	0.05	0.025	0.01	0.005	0.001	0.0005

Mức ý nghĩa α (tiêu chuẩn một phía)

Bảng 3. Phân phối khi bình phương với n bậc tự do P [X > C_α] = α

Bậc tự do n	Xác suất															
	0,99	0,98	0,95	0,90	0,80	0,70	0,50	0,30	0,20	0,10	0,05	0,02	0,01	0,005	0,002	0,001
1	0,0001	0,0006	0,0009	0,016	0,064	0,148	0,455	1,07	1,64	2,7	3,81	5,5	6,6	7,9	9,5	10,83
2	0,020	0,040	0,103	0,211	0,446	0,713	1,386	2,41	3,22	4,6	6,0	7,8	9,2	11,6	12,4	13,8
3	0,115	0,185	0,352	0,584	1,005	0,424	2,366	3,66	4,64	6,3	7,8	9,8	11,3	12,8	14,8	16,3
4	0,30	0,43	0,71	0,06	1,65	2,19	3,36	4,9	6,0	7,8	9,5	11,7	13,3	14,9	16,9	18,5
5	0,55	0,75	1,14	1,61	2,34	3,00	4,35	6,1	7,3	9,2	11,1	13,4	15,1	16,3	18,9	20,5
6	0,187	1,13	1,63	2,20	3,07	3,83	5,35	7,2	8,6	10,6	12,6	15,0	16,8	18,6	20,7	22,5
7	1,24	1,56	2,17	2,83	3,82	4,67	6,34	8,4	9,8	12,0	14,1	16,6	18,5	20,3	2,6	24,3
8	1,65	2,03	2,73	3,49	4,59	5,53	7,34	9,5	11,0	13,4	15,5	18,2	20,1	21,9	24,3	21,6
9	2,09	2,53	3,32	4,17	5,38	6,39	8,35	10,7	12,2	14,7	16,9	19,7	21,7	23,6	26,1	27,9
10	2,56	3,06	3,94	4,86	6,18	7,27	9,34	11,8	13,4	16,0	18,3	21,2	23,2	25,2	27,7	29,6
11	3,1	3,6	4,6	5,6	7,0	8,1	10,3	12,9	14,6	17,3	19,7	22,6	24,7	26,8	29,4	31,3
12	3,6	4,2	5,2	6,3	7,8	9,0	11,3	14,0	15,8	18,5	21,0	24,1	26,2	28,3	31,0	32,9
13	4,1	4,8	5,9	7,0	8,6	9,9	12,3	15,1	17,0	19,8	22,4	25,5	27,7	29,8	32,5	34,5
14	4,7	5,4	6,6	7,8	9,5	10,8	13,3	16,2	18,2	21,1	23,7	26,9	29,1	31,0	34,0	36,1
15	5,2	6,0	7,3	8,5	10,3	11,7	14,3	17,3	19,3	22,3	25,0	28,3	30,6	32,5	35,5	37,7
16	5,8	6,6	8,0	9,3	11,2	12,6	15,3	18,4	20,5	23,5	26,3	29,6	32,0	34,0	37,0	39,2
17	6,4	7,3	8,7	10,1	12,0	13,5	16,3	19,5	21,6	24,8	27,6	31,0	33,4	35,5	38,5	40,8
18	7,0	7,9	9,4	10,9	12,9	14,4	17,3	20,6	22,8	26,0	28,9	32,3	34,8	37,0	40,0	42,3
19	7,6	8,6	10,1	11,7	13,7	15,4	18,3	21,7	23,9	27,2	30,1	33,7	36,2	38,5	41,5	43,8
20	8,3	9,2	10,9	12,4	14,6	16,3	19,3	22,8	25,0	28,4	31,4	35,0	37,6	40,0	43,0	45,3
21	8,9	9,9	11,6	13,2	15,4	17,2	20,3	23,9	26,2	29,6	32,7	36,3	38,9	41,5	44,5	46,8
22	9,5	10,6	12,3	14,0	16,3	18,1	21,3	24,9	27,3	30,8	33,9	37,7	40,3	42,0	46,0	48,3
23	10,2	11,3	13,1	14,8	17,2	19,0	22,3	26,0	28,4	32,0	35,2	39,0	41,6	44,0	47,5	49,7
24	10,9	12,0	13,8	15,7	18,1	19,9	23,3	27,1	29,6	33,2	36,4	40,3	43,0	45,5	48,5	51,2
25	11,5	12,7	14,6	16,5	18,9	20,9	24,3	28,1	30,7	34,4	37,7	41,6	44,3	47,0	50,0	52,6
26	12,2	13,4	15,4	17,3	19,8	21,8	25,3	29,3	31,8	35,6	38,9	42,9	45,6	48,0	51,5	54,1
27	12,9	14,1	16,2	18,1	20,7	22,7	26,3	30,3	32,9	36,7	40,1	44,1	47,0	49,5	53,0	55,5
28	13,6	14,8	16,9	18,9	21,6	23,6	27,3	31,4	34,0	37,9	41,3	45,4	48,3	51,0	54,5	56,9
29	14,3	15,6	17,7	19,8	22,5	24,6	28,3	32,5	35,1	39,1	42,6	46,7	49,6	52,5	56,0	58,3
30	15,0	16,3	18,5	20,6	23,4	25,5	29,3	33,5	36,3	40,3	43,8	48,0	50,9	54,0	57,5	59,7

Bảng 4. Phân phối F (Fisher – Snedecor), mức ý nghĩa α = 0,01

K_2 \ K_1	1	2	3	4	5	6	7	8	9	10	11	12
1	4052	4999	5403	56.25	57.64	5889	59.28	5981	6022	6056	6082	6106
2	98.49	99.01	99.17	99.25	99.30	99.33	99.34	99.36	99.38	99.40	99.41	99.42
3	34.12	30.81	20.48	27.71	23.24	27.91	27.67	27.49	27.34	27.23	27.13	27.05
4	21.20	18.00	16.69	15.98	15.52	14.21	14.98	14.80	14.66	14.55	14.45	14.37
5	16.26	13.27	12.06	11.30	10.97	10.67	10.45	10.27	10.15	10.05	9.96	9.89
6	13.74	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.79	7.72
7	12.25	9.55	9.45	7.89	7.46	7.19	7.00	6.84	6.71	6.62	6.54	6.47
8	11.26	8.65	7.59	7.01	6.63	6.37	6.19	6.03	5.91	5.82	5.74	5.67
9	10.56	8.02	6.99	6.42	6.06	5.80	5.62	5.47	5.35	5.26	5.18	5.11
10	10.04	7.56	6.55	5.99	5.64	5.39	5.21	5.06	4.95	4.85	4.78	4.71
11	9.85	7.20	6.22	5.67	5.32	5.07	4.88	4.74	4.63	4.54	4.46	4.40
12	9.33	6.93	5.95	5.41	5.06	4.82	4.65	4.50	4.39	4.30	4.22	4.16
13	8.86	6.51	5.56	5.03	4.69	4.46	4.28	4.14	4.03	3.94	3.86	3.80
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.61	3.55
20	8.10	5.85	4.94	4.43	4.10	3.87	3.71	3.56	3.45	3.37	3.30	3.23
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.25	3.17	3.09	3.03
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.06	2.98	2.90	2.84
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.88	2.80	2.73	2.66
50	7.17	5.06	4.20	3.72	3.41	3.18	3.02	2.88	2.78	2.70	2.62	2.56
70	7.01	4.92	4.08	3.60	3.29	3.07	2.91	2.77	2.67	2.59	2.51	2.45
100	6.90	4.82	3.98	3.51	3.20	2.99	2.82	2.69	2.59	2.51	2.43	2.36
200	6.76	4.71	3.88	3.41	3.11	2.90	2.73	2.60	2.50	2.41	2.34	2.28
400	6.70	4.66	3.83	3.36	3.06	2.85	2.69	2.55	2.46	2.37	2.29	2.23
x	6.64	4.60	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.24	2.18

F – Phân phối α = 0,01 (Fisher – Snedecor)

Tiếp bảng 4

K_2 \ K_1	14	16	20	24	30	40	50	75	100	200	500	∞
1	6142	6169	6208	6234	6258	6286	6302	6323	6334	6352	6361	6366
2	99.43	99.44	99.45	99.46	99.47	99.48	99.48	99.49	99.49	99.49	99.50	99.50
3	26.92	26.83	26.69	26.60	26.50	26.41	26.35	26.27	26.23	26.18	26.14	26.12
4	14.24	14.15	14.02	13.93	13.83	13.74	13.69	13.61	13.57	13.52	13.48	13.46
5	9.77	9.68	9.55	9.47	9.38	9.29	9.24	9.17	9.13	9.07	9.04	9.02
6	7.60	7.52	7.39	7.31	7.23	7.14	7.09	7.02	6.99	6.94	6.90	6.88
7	6.35	6.27	6.15	6.07	5.98	5.90	5.85	5.78	5.75	5.70	5.67	5.65
8	5.56	5.48	5.36	5.28	5.20	5.11	5.11	5.06	5.00	4.96	4.88	4.86
9	5.00	4.92	4.80	4.73	4.46	4.56	4.51	4.45	4.41	4.36	4.33	4.31
10	4.60	4.52	4.41	4.33	4.25	4.17	4.12	4.05	4.01	3.96	3.93	3.91
11	4.29	4.21	4.10	4.02	3.94	3.86	3.80	3.74	3.70	3.66	3.62	3.60
12	4.05	3.98	3.86	3.78	3.70	3.61	3.56	3.49	3.46	3.41	3.38	3.36
13	3.70	3.62	3.51	3.43	3.34	3.26	3.21	3.14	3.11	3.06	3.02	3.00
16	3.45	3.37	3.25	3.18	3.01	3.01	2.96	2.89	2.86	2.80	2.77	2.75
20	3.13	3.05	2.94	2.86	2.77	2.69	2.63	2.56	2.53	2.47	2.44	2.42
24	2.93	2.85	2.74	2.66	2.58	2.49	2.44	2.36	2.33	2.27	2.23	2.21
30	2.74	2.66	2.55	2.47	2.38	2.29	2.24	2.16	2.13	2.07	2.03	2.01
40	2.56	2.49	2.37	2.29	2.20	2.11	2.05	1.97	1.94	1.88	1.84	1.81
50	2.46	2.39	2.26	2.18	2.10	2.00	1.94	1.86	1.82	1.76	1.71	1.68
70	2.35	2.28	2.15	2.07	1.98	1.88	1.82	1.74	1.69	1.62	1.56	1.53
100	2.26	2.19	2.06	1.98	1.89	1.79	1.73	1.64	1.59	1.51	1.46	1.43
200	2.17	2.09	1.97	1.88	1.79	1.69	1.62	1.53	1.48	1.39	1.33	1.28
400	2.12	2.04	1.92	1.84	1.74	1.64	1.57	1.47	1.42	1.32	1.24	1.19
x	2.07	1.99	1.87	1.79	1.69	1.59	1.52	1.41	1.36	1.25	1.15	1.09

Bảng phân phối F (Fisher–Snedecor), mức ý nghĩa $\alpha = 0,05$

Tiếp bảng 4

K_2 \ K_1	1	2	3	4	5	6	7	8	9	10	11
1	161	200	216	225	230	2.34	237	239	241	242	243
2	18.5	19.00	19.16	19.25	19.30	19.33	19.36	19.37	19.38	19.39	19.40
3	10.13	9.55	9.28	9.12	9.01	8.94	8.88	8.84	8.81	8.78	8.76
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.93
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.78	4.74	4.70
6	5.99	5.14	4.76	4.53	4.39	4.21	4.15	4.10	4.10	4.06	4.03
7	5.59	4.74	4.35	4.12	3.97	3.79	3.73	3.68	3.68	3.63	3.60
8	5.32	4.46	4.07	3.84	3.69	3.50	3.44	3.39	3.39	3.34	3.31
9	5.12	4.26	3.86	3.63	3.48	3.29	3.23	3.18	3.18	3.13	3.10
10	4.96	4.10	3.71	3.48	3.22	3.14	3.07	3.02	3.02	2.97	2.94
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.86	2.82
12	4.75	3.88	3.49	3.26	3.11	3.00	2.92	2.85	2.80	2.76	2.72
14	4.60	3.74	3.34	3.11	2.96	2.85	2.77	2.70	2.65	2.60	2.56
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.45
20	4.35	3.49	3.10	2.87	2.71	2.60	2.52	2.45	2.40	2.35	2.31
24	4.26	3.40	3.01	2.78	2.62	2.51	2.43	2.36	2.30	2.26	2.22
30	4.17	3.32	2.92	2.69	2.53	2.42	2.34	2.26	2.21	2.16	2.12
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.07	2.04
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07	2.02	2.98
70	3.98	3.13	2.74	2.50	2.35	2.23	2.14	2.07	2.01	2.97	1.93
100	3.94	2.09	2.70	2.46	2.30	2.19	2.10	2.03	1.97	1.93	1.88
200	3.89	3.04	2.65	2.41	2.26	2.14	2.05	1.98	1.92	1.87	1.83
400	3.86	3.02	2.62	2.39	2.23	2.12	2.03	1.96	1.90	1.85	1.81
x	33.84	2.99	2.60	2.37	2.21	2.09	2.01	1.94	1.88	1.83	1.79

Bảng phân phối F (Fisher–Snedecor), mức ý nghĩa α = 0,05

Tiếp bảng 4

K_2 \ K_1	12	1.3	16	20	30	40	50	75	100	200	500	∞
1	244	245	246	248	250	251	252	253	253	254	254	2.54
2	1941	19.42	19.43	19.44	19.46	19.47	19.47	19.48	19.49	19.49	19.50	19.50
3	8.74	8.71	8.66	8.66	8.02	8.00	8.58	8.57	8.59	8.54	8.54	8.53
4	5.91	5.87	5.84	5.80	5.74	5.71	5.70	5.68	5.66	5.65	5.64	5.63
5	4.68	4.64	4.60	4.56	4.50	4.46	4.44	4.42	4.40	4.38	4.38	1.36
6	4.00	3.96	3.92	3.87	3.81	3.77	3.75	3.72	3.71	3.69	3.68	3.67
7	3.57	3.52	3.49	3.44	3.38	3.34	3.32	3.29	3.28	3.25	3.24	3.23
8	3.28	3.23	3.20	3.15	3.08	3.05	3.03	3.00	2.98	2.96	2.94	2.93
9	3.07	3.02	2.98	2.93	2.86	2.82	2.80	2.77	2.76	2.73	2.72	2.71
10	2.91	2.86	2.82	2.77	2.70	2.67	2.64	2.61	2.59	2.56	2.55	2.54
11	2.79	2.74	2.70	2.65	2.57	2.53	2.50	2.47	2.45	2.42	2.41	2.40
12	2.69	2.64	2.60	2.54	2.46	2.42	2.40	2.36	2.35	2.32	2.30	2.30
13	2.53	2.48	2.44	2.39	2.31	2.27	2.24	2.21	2.19	2.16	2.14	2.13
16	2.42	2.37	2.33	2.28	2.20	2.16	2.13	2.09	2.07	2.04	2.02	2.01
20	2.98	2.23	2.18	2.12	2.04	1.99	1.96	1.92	1.90	1.87	1.85	1.84
24	2.18	2.13	2.09	2.02	1.94	1.89	1.86	1.82	1.80	1.76	1.74	1.73
30	2.09	2.04	1.99	1.93	1.84	1.79	1.76	1.72	1.69	1.66	1.64	1.62
40	2.00	1.95	1.90	1.84	1.74	1.69	1.66	1.61	1.59	1.55	1.53	1.51
50	1.95	1.90	1.85	1.78	1.69	1.63	1.60	1.55	1.52	1.18	1.46	1.44
70	1.89	1.84	1.79	1.72	1.62	1.46	1.53	1.47	1.45	1.40	1.37	1.35
100	1.85	1.79	1.75	1.68	1.51	1.51	1.48	1.42	1.39	1.34	1.30	1.28
200	1.80	1.74	1.69	1.62	1.45	1.45	1.42	1.35	1.32	1.26	1.22	1.19
400	1.78	1.72	1.67	1.60	1.42	1.42	1.38	1.32	1.28	1.22	1.16	1.13
x	1.75	1.69	1.64	1.57	1.40	1.40	1.35	1.28	1.24	1.17	1.11	1.00

Bảng 5. Giá trị tiêu chuẩn Wilcoxon với $n \leq 20$ (so sánh cặp đôi). Tiêu chuẩn một phía

n	$\alpha = 0{,}05$		$\alpha = 0{,}025$		$\alpha = 0{,}01$		$\alpha = 0{,}005$	
	W_α	$W_{1-\alpha}$	W_α	$W_{1-\alpha}$	W_α	$W_{1-\alpha}$	W_α	$W_{1-\alpha}$
6	3	17	1	20	0	21	0	21
7	4	24	3	25	0	21	0	21
8	6	30	4	32	2	34	1	35
9	9	36	6	39	4	41	2	43
10	11	44	9	36	6	49	4	51
11	14	52	11	55	8	58	6	60
12	18	60	14	64	10	68	8	70
13	22	69	18	73	13	78	10	81
14	26	79	22	83	16	89	13	92
15	31	89	26	94	20	100	16	104
16	36	100	30	106	24	112	20	116
17	42	111	35	118	28	125	24	129
18	48	123	41	130	33	138	28	143
19	54	136	47	143	38	152	33	157
20	61	149	53	157	44	166	38	172

Tiêu chuẩn hai phía :

$\alpha = 0{,}1$		$\alpha = 0{,}05$		$\alpha = 0{,}01$	
$W_{\frac{\alpha}{2}}$	$W_{1-\frac{\alpha}{2}}$	$W_{\frac{\alpha}{2}}$	$W_{1-\frac{\alpha}{2}}$	$W_{\frac{\alpha}{2}}$	$W_{1-\frac{\alpha}{2}}$

***Bảng 6*. Điểm tiêu chuẩn của tiêu chuẩn Wilcoxon**

Kích thước mẫu		Q				Kích thước mẫu		Q			
n_1	n_2	0,005	0,01	0,025	0,05	n_1	n_2	0,005	0,01	0,025	0,05
6	6	23	24	26	28	7	20	52	56	62	67
	7	24	25	27	30		21	53	58	64	69
	8	25	27	29	31		22	55	59	66	72
	9	26	28	31	33		23	57	61	68	74
	10	27	29	32	35		24	58	63	70	76
	11	28	30	34	37		25	60	64	72	78
	12	30	32	35	38	8	8	43	45	49	51
	13	31	33	37	40		9	45	47	51	54
	14	32	34	38	42		10	47	49	53	56
	15	33	36	40	44		11	49	51	55	59
	16	34	37	42	46		12	51	53	58	62
	17	36	39	43	47		13	53	56	60	64
	18	37	40	45	49		14	54	58	62	67
	19	38	41	46	51		15	56	60	65	69
	20	39	43	48	53		16	58	62	67	72
	21	40	44	50	55		17	60	64	70	75
	22	42	45	51	57		18	62	66	72	77
	23	43	47	53	58		19	64	68	74	80
	24	44	48	54	60		20	66	70	77	83
	25	45	50	56	62		21	68	72	79	85
7	7	32	34	58	39		22	70	74	81	88
	8	34	35	38	41		23	71	76	84	90
	9	35	37	40	43		24	73	78	86	93
	10	37	39	42	45		25	75	81	89	96
	11	38	40	44	47	9	9	56	59	62	66
	12	40	42	46	49		10	58	61	65	69
	13	41	44	48	52		11	61	63	68	72
	14	43	45	50	54		12	63	66	71	75
	15	44	47	52	56		13	65	68	73	78
	16	46	49	54	58		14	67	71	76	81
	17	47	51	56	61		15	69	73	79	85
	18	49	52	58	63		16	72	76	82	87
	19	50	54	60	65						

Tiếp bảng 6

Kích thước mẫu		Q				Kích thước mẫu		Q			
n_1	n_2	0,005	0,01	0,025	0,05	n_1	n_2	0,005	0,01	0,025	0,05
9	17	74	78	84	90	11	20	114	119	128	135
	18	76	81	87	93		21	117	123	131	139
	19	78	83	90	96		22	120	126	135	143
	20	81	85	93	99		23	123	129	139	147
	21	83	88	95	102		24	126	132	142	151
	22	85	90	98	105		25	105	136	146	155
	23	88	93	101	108	12	12	112	109	115	120
	24	90	95	104	111		13	115	113	119	125
	25	92	98	107	114		14	119	116	123	129
10	10	71	74	78	82		15	122	120	127	133
	11	73	77	81	86		16	125	124	131	138
	12	75	79	84	89		17	129	127	135	142
	13	79	82	88	92		18	132	131	139	146
	14	81	85	91	96		19	136	134	143	150
	15	84	88	94	99		20	139	138	147	155
	16	86	91	97	103		21	142	142	151	159
	17	89	93	100	106		22	146	145	155	163
	18	92	96	103	110		23	149	149	159	168
	19	94	99	107	113		24	125	153	163	172
	20	97	102	110	117		25	129	156	167	176
	21	99	105	113	120	13	13	133	130	136	142
	22	102	108	116	123		14	136	134	141	147
	23	105	110	119	127		15	140	136	145	152
	24	107	113	122	130		16	144	142	150	156
	25	110	116	126	134		17	151	146	154	161
11	11	87	91	96	100		18	144	150	188	166
	12	90	94	99	104		19	148	154	163	171
	13	93	97	103	108		20	151	158	167	175
	14	96	100	106	112		21	155	162	171	180
	15	99	103	110	116		22	159	166	176	185
	16	102	107	113	120		23	163	170	180	189
	17	105	110	117	123		24	166	174	185	194
	18	108	113	121	127		25	170	178	189	199
	19	111	116	124	134						

Tiếp bảng 6

Kích thước mẫu		Q				Kích thước mẫu		Q			
n_1	n_2	0,005	0,01	0,025	0,05	n_1	n_2	0,005	0,01	0,025	0,05
14	14	147	152	160	166	16	20	215	223	234	243
	15	151	156	164	171		21	220	228	239	249
	16	155	161	169	176		22	225	233	245	255
	17	159	165	174	182		23	230	238	251	261
	18	163	170	179	187		24	235	240	256	267
	19	168	174	183	192		25	240	249	262	273
	20	172	178	188	197	17	17	223	230	240	249
	21	176	183	193	202		18	228	235	246	255
	22	180	187	198	207		19	234	241	252	262
	23	184	192	203	212		20	239	246	258	268
	24	188	196	207	218		21	244	252	264	274
	25	192	200	212	223		22	249	258	270	281
15	15	171	176	184	192		23	255	263	276	282
	16	175	181	190	197		24	260	269	282	294
	17	180	186	195	203		25	265	275	288	300
	18	184	190	200	208	18	18	252	259	270	280
	19	189	195	205	214		19	258	265	277	287
	20	193	200	210	220		20	263	271	283	294
	21	198	205	216	225		21	269	277	290	301
	22	202	210	221	231		22	275	283	296	307
	23	207	214	226	236		23	280	289	303	314
	24	211	219	231	242		24	286	295	309	321
	25	216	224	237	248		25	292	301	316	328
16	16	196	202	211	219	19	19	283	291	303	313
	17	201	207	217	225		20	289	297	309	320
	18	206	216	222	231		21	295	303	316	328
	19	210	218	228	237		22	301	310	323	335

Kích thước mẫu		Q				Kích thước mẫu		Q			
n_1	n_2	0,005	0,01	0,025	0,05	n_1	n_2	0,005	0,01	0,025	0,05
19	23	307	316	330	342	21	24	370	381	396	410
	24	313	323	337	350		25	377	388	404	418
	25	319	329	344	357	22	22	386	396	411	424
20	20	315	324	337	348		23	393	403	419	432
	21	322	331	344	356		24	400	411	427	441
	22	328	337	351	364		25	408	419	435	450
	23	335	344	359	371	23	23	424	434	451	465
	24	341	351	366	379		24	431	443	459	474
	25	348	358	373	387		25	439	451	468	483
21	21	349	359	373	385	24	24	464	475	492	507
	22	359	366	381	393		25	472	484	501	517
	23	363	373	388	401	25	25	505	517	536	552

Bảng 7. Phân phối giới hạn của thống kê $D_n(x)$ của Cômôgôrôp

$$\lim_{n\to\infty} P[\sqrt{n}D_n < x] = K(x)$$

x	K(x)	x	K(x)	x	K(x)
0.28	0.000001	0.58	0.110395	0.88	0.579070
0.29	0.000004	0.59	0.112760	0.89	0.593316
0.30	0.000009	0.60	0.135718	0.90	0.607270
0.31	0.000021	0.61	0.149229	0.91	0.620928
0.32	0.000046	0.62	0.163255	0.92	0.634286
0.33	0.000091	0.63	0.177753	0.93	0.647338
0.34	0.000171	0.64	0.192677	0.94	0.660082
0.35	0.000303	0.65	0.207987	0.95	0.672516
0.36	0.000511	0.66	0.223637	0.96	0.684636
0.37	0.000826	0.67	0.239582	0.97	0.696444
0.38	0.001285	0.68	0.255780	0.98	0.707940
0.39	0.001929	0.69	0.272189	0.99	0.709126
0.40	0.003972	0.70	0.288765	1.00	0.730000
0.41	0.003972	0.71	0.305471	1.01	0.740566
0.42	0.005476	0.72	0.322265	1.02	0.750826
0.43	0.007377	0.73	0.339133	1.03	0.760780
0.44	0.009730	0.74	0.355981	1.04	0.770434
0.45	0.012590	0.75	0.372833	1.05	0.779794
0.46	0.016005	0.76	0.389640	1.06	0.788860
0.47	0.020022	0.77	0.406372	1.07	0.797636
0.48	0.024682	0.78	0.423002	1.08	0.806128
0.49	0.030017	0.79	0.439505	1.09	0.814342
0.50	0.036055	0.78	0.455857	1.10	0.814342
0.51	0.042814	0.81	0.472041	1.11	0.822282
0.52	0.050306	0.82	0.488030	1.12	0.829950
0.53	0.058354	0.83	0.503808	1.13	0.844502
0.54	0.067497	0.84	0.519366	1.14	0.851394
0.55	0.077183	0.85	0.534682	1.15	0.848038
0.56	0.087577	0.86	0.549744	1.16	0.864442
0.57	0.098656	0.87	0.564546	1.17	0.870612

Tiếp bảng 7

x	K(x)	x	K(x)	x	K(x)
0.18	0.876548	1.47	0.973448	2.04	0.999516
1.19	0.882258	1.48	0.974970	2.06	0.999588
1.20	0.887750	1.49	0.976412	2.08	0.999650
1.21	0.893030	1.50	0.977782	2.10	0.999705
1.22	0.898104	1.52	0.980310	2.12	0.999750
1.23	0.902972	1.54	0.982578	2.14	0.999790
1.24	0.707648	1.56	0.984610	2.16	0.999822
1.25	0.912132	1.58	0.986426	2.18	0.999852
1.25	0.912132	1.60	0.988048	2.20	0.999874
1.26	0.916432	1.62	0.989492	2.22	0.999896
1.27	0.920556	1.64	0.990777	2.24	0.999912
1.28	0.924505	1.66	0.991917	2.26	0.999926
1.29	0.928288	1.68	0.992928	2.28	0.999940
1.30	0.931908	1.70	0.993223	2.30	0.999949
1.31	0.935370	1.72	0.994612	2.32	0.999958
1.32	0.938662	1.74	0.995309	2.34	0.999965
1.33	0.941848	1.76	0.995922	2.36	0.999970
1.34	0.944872	1.78	0.996460	2.38	0.999976
1.35	0.947756	1.80	0.996932	2.40	0.999980
1.36	0.950512	1.82	0.997346	2.42	0.999984
1.37	0.953142	1.84	0.997707	2.44	0.999987
1.38	0.955650	1.86	0.998023	2.48	0.999991
1.39	0.958040	1.88	0.998297	2.50	0.99999335
1.40	0.960318	1.90	0.998556	2.55	0.99999956
1.41	0.962486	1.92	0.998744	2.60	0.99999974
1.43	0.966516	1.94	0.998924	2.65	0.99999984
1.44	0.968382	1.96	0.999079	2.70	0.99999990
1.45	0.970158	2.00	0.999329	2.90	0.999999990
1.46	0.971846	2.02	0.999428	3.00	0.9999999997

TÀI LIỆU THAM KHẢO

1. *Phạm Văn Kiều*.
Lý thuyết xác suất và Thống kê toán học. NXB KH & KT, 1998.
2. *Lehmann E*. Testing statistical hypotheses. New York, 1959.
3. *Joroslav Hájek* Zbyněk, Šidák. Theory of Rank tests, Prague, 1967.
4. *Guy Lefort*. Mathematiques pour les sciencos, Billogiques et agronoiques, 1967.
5. *Mustafa Benyaklef*. Probabilites et statistique mathematique, 1977.

MỤC LỤC

Chịu trách nhiệm xuất bản :

Chủ tịch HĐQT kiêm Tổng Giám đốc NGÔ TRẦN ÁI

Phó Tổng Giám đốc kiêm Tổng biên tập VŨ DƯƠNG THỤY

Biên tập nội dung :

NGUYỄN HỒNG ÁNH

Trình bày bìa :

TÀO THANH HUYỀN

Sửa bản in :

NGUYỄN THU HUYỀN

Chế bản :

TRẦN THU HƯƠNG

GIÁO TRÌNH XÁC SUẤT VÀ THỐNG KÊ

Mã số : 7B608M5 – DAI

In 1.000 cuốn, khổ 16 x 24 cm, tại Công ty Cổ phần In Phúc Yên.

Số xuất bản: 1205/568 – 03.

In xong và nộp lưu chiểu quý IV năm 2005.

CÔNG TY CỔ PHẦN SÁCH ĐẠI HỌC - DẠY NGHỀ
HEVOBCO
Địa chỉ : 25 Hàn Thuyên, Hà Nội

8 934980 534021

Giá: 24.500đ