BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC ĐÔNG Á

ThS.PHẠM THỊ NGỌC MINH

GIÁO TRÌNH
XÁC SUẤT THỐNG KÊ

LƯU HÀNH NỘI BỘ

Đà Nẵng, 2013

CHƯƠNG.1. KHÁI NIỆM CƠ BẢN VỀ XÁC SUẤT

BÀI - 1. PHÉP THỬ NGẪU NHIÊN VÀ CÁC LOẠI BIẾN CỐ

1.1.1. PHÉP THỬ NGẪU NHIÊN

Trong tự nhiên và xã hội, mỗi hiện tượng đều gắn liền với một nhóm các điều kiện cơ bản và các hiện tượng đó chỉ có thể xảy ra khi nhóm các điều kiện cơ bản gắn liền với nó được thực hiện. Do đó, khi muốn nghiên cứu một hiện tượng ta cần thực hiện nhóm các điều kiện cơ bản ấy.

Việc thực hiện một nhóm các điều kiện cơ bản để quan sát một hiện tượng nào đó có xảy ra hay không được gọi là thực hiện một phép thử.

Ví dụ 1.1:

Tung 1 đồng xu là 1 phép thử.

Ném 1 phi tiêu vào bia là 1 phép thử.

1.1.2. KHÔNG GIAN MẪU

Tập tất cả các kết quả có thể xảy ra trong 1 phép thử được gọi là không gian mẫu.

Ví dụ 1.2:

Tung 1 đồng xu là 1 phép thử, không gian mẫu là tập gồm 2 kết quả: sấp, ngửa.

Ném 1 phi tiêu vào bia là 1 phép thử, không gian mẫu gồm 2 kết quả: trúng bia hoặc không trúng bia.

1.1.3. BIẾN CỐ

Hiện tượng có thể xảy ra trong kết quả của phép thử được gọi là biến cố.

Ví dụ 1.3:

Tung 1 đồng xu là 1 phép thử, đồng xu lật sấp hay ngửa là 1 biến cố.

Tung một con xúc xắc xuống đất là một phép thử, con xúc xắc lật lên một mặt nào đó là 1 biến cố.

1.1.3.2. Biến cố ngẫu nhiên

Là biến cố có thể xảy ra hoặc không xảy ra khi thực hiện một phép thử.

Các biến cố ngẫu nhiên được kí hiệu là A, B, C, . . . hoặc A_1, A_2, ..., A_n, B_1, B_2, ..., B_n.

Ví dụ 1.4:

- Tung một con xúc xắc, gọi A là biến cố “Xuất hiện mặt 1 chấm” khi đó A là biến cố ngẫu nhiên.

- Bắn phát súng vào bia, gọi B là biến cố “Trúng vòng 10” khi đó B là biến cố ngẫu nhiên.

1.1.3.3. Biến cố chắc chắn

Là biến cố nhất định xảy ra khi thực hiện một phép thử.

Biến cố chắc chắn được kí hiệu là U.

Ví dụ 1.5:

- Thực hiện phép thử tung đồng xu. Gọi U là biến cố “Xuất hiện mặt sấp hoặc mặt ngửa”. U là biến cố chắc chắn.

- Chấm điểm bài thi của một học sinh với thang điểm 10, gọi U là biến cố “Số điểm đạt được không lớn hơn 10” thì U là biến cố chắc chắn.

1.1.3.4. Biến cố không có thể

Là biến cố nhất định không xảy ra khi thực hiện phép thử.

Biến cố không thể có được kí hiệu là V.

Ví dụ 1.6:

Chọn một học sinh trong một lớp học không có nữ, thì biến cố “Chọn được một học sinh nữ” là biến cố không thể có.

Tất cả các biến cố mà chúng ta gặp trong thực tế đều thuộc một trong 3 loại biến cố kể trên, tuy nhiên các biến cố ngẫu nhiên là các biến cố thường gặp hơn cả.

Hai hay nhiều biến cố trong phép thử có khả năng xảy ra như nhau, được gọi là đồng khả năng.

Ví dụ 1.7:

- Tung một đồng xu cân đối đồng chất, ta có số trường hợp đồng khả năng là 2.

- Tung một con xúc xắc cân đối đồng chất, ta có số trường hợp đồng khả năng là 6.

BÀI - 2. QUAN HỆ GIỮA CÁC BIẾN CỐ

1.2.1. TỔNG, TÍCH CỦA 2 BIẾN CỐ

1.2.1.1. Tổng biến cố

Định nghĩa 1.1:

Biến cố C được gọi là tổng hai biến cố A và B, kí hiệu C = A + B nếu C chỉ xảy ra khi có ít nhất một trong hai biến cố A và B xảy ra.

Định nghĩa 1.2:

Biến cố A được gọi là tổng của n biến cố $\mathrm{A}_1, \mathrm{A}_2, ..., \mathrm{A}_n$ *nếu A xảy ra khi ít nhất có một trong n biến cố ấy xảy ra.*

Kí hiệu: $A = \sum_{i=1}^{n} A_i$

Ví dụ 1.8:

Hai người thợ săn cùng săn 1 con thú. Gọi A là biến cố "người thứ nhất bắn trúng con thú", B là biến cố "người thứ hai bắn trúng con thú", C là biến cố " con thú bị bắn trúng".

Ta có: $C = A + B$.

Ví dụ 1.9:

Kiểm tra n sản phẩm. Gọi A_i là biến cố "Sản phẩm thứ i là xấu", A là biến cố "Có ít nhất một sản phẩm xấu".

Ta có $A = A_1 + A_2 + ... + A_n$

1.2.1.2. Tích biến cố

Định nghĩa 1.3:

Biến cố C được gọi là tích của hai biến cố A và B nếu C xảy ra khi và chỉ khi cả hai biến cố A và B cùng đồng thời xảy ra, kí hiệu C = A . B.

Ví dụ 1.10:

Có hai hộp đựng một số quả cầu trắng và đen. Gọi A là biến cố "Lấy được quả cầu trắng ở hộp thứ nhất", B là biến cố "Lấy được quả cầu trắng ở hộp thứ hai". Gọi C là biến cố "Lấy được hai quả cầu trắng".

Ta có: $C = A.B$

Định nghĩa 1.4:

Biến cố A được gọi là tích của n biến cố $\mathrm{A}_1, \mathrm{A}_2, ..., \mathrm{A}_n$ *nếu A xảy ra khi cả n biến cố nói trên cùng đồng thời xảy ra.*

Kí hiệu: $A = \prod_{i=1}^{n} A_i$

Ví dụ 1.11:

Có n xạ thủ cùng bắn vào một mục tiêu, gọi A_i là biến cố "Người thứ i bắn trúng mục tiêu", i = 1, 2,..., n. Khi đó, biến cố $A = \prod_{i=1}^{n} A_i$ là biến cố "Cả n xạ thủ cùng bắn trúng".

1.2.2. BIẾN CỐ XUNG KHẮC, BIẾN CỐ ĐỘC LẬP, HỌ BIẾN CỐ ĐẦY ĐỦ

1.2.2.1. Biến cố xung khắc

Định nghĩa 1.5:

Hai biến cố A và B được gọi là xung khắc với nhau nếu chúng không thể đồng thời xảy ra trong một phép thử.

Trường hợp ngược lại, nếu hai phép thử có thể xảy ra đồng thời trong một phép thử được gọi là không xung khắc.

Ví dụ 1.12:

Có hai hộp đựng một số quả cầu trắng và đen. Lấy ở mỗi hộp một quả cầu. Gọi A là biến cố "Lấy được hai quả cầu cùng màu", B là biến cố "Lấy được hai quả khác màu". Khi đó A, B là hai biến cố xung khắc.

Ví dụ 1.13:

Hai người A, B cùng bắn vào một tấm bia, gọi A là biến cố "Người A bắn trúng", B là biến cố "Người B bắn trúng". Khi đó hai biến cố A, B là không xung khắc.

Định nghĩa 1.6:

Nhóm n biến cố $A_1, A_2, ..., A_n$ *được gọi là xung khắc từng đôi nếu bất kỳ hai biến cố nào trong nhóm này cũng xung khắc với nhau.*

Ví dụ 1.14:

Trong một cái hộp có 3 viên bi xanh, 4 viên bi vàng, 5 viên bi đỏ. Gọi A_1 là biến cố "Lấy được hai viên bi xanh", A_2 là biến cố "Lấy được hai viên bi vàng", A_3 là biến cố "Lấy được hai viên bi đỏ". Khi đó A_1, A_2, A_3 xung khắc nhau từng đôi một.

1.2.2.2. Nhóm biến cố đầy đủ

Định nghĩa 1.7:

Các biến cố $A_1, A_2, ..., A_n$ *được gọi là một nhóm đầy đủ các biến cố nếu trong kết quả của một phép thử sẽ xảy ra một và chỉ một trong các biến cố đó.*

Nói cách khác các biến cố trên sẽ tạo nên một nhóm đầy đủ các biến cố nếu chúng xung khắc từng đôi một và tổng của chúng là một biến cố chắc chắn.

Ví dụ 1.15:

Chọn ngẫu nhiên 1 sản phẩm trong 1 kho hàng chứa sản phẩm do 3 nhà máy sản xuất. Gọi A_i là biến cố "Sản phẩm lấy ra do nhà máy thứ i sản xuất", ta có A_1, A_2, A_3 là một nhóm đầy đủ các biến cố.

1.2.2.3. Biến cố đối lập

Định nghĩa 1.8:

Hai biến cố A và $\overline{A}$ gọi là đối lập với nhau nếu chúng tạo nên một nhóm đầy đủ các biến cố.

Ví dụ 1.16:

Bắn một phát đạn vào bia. Gọi A là biến cố "Bắn trúng bia", $\overline{A}$ là biến cố "Bắn trượt bia". A và $\overline{A}$ là hai biến cố đối lập.

1.2.2.4. Biến cố độc lập

Định nghĩa 1.9:

Hai biến cố A và B được gọi là độc lập với nhau nếu việc xảy ra hay không xảy ra của biến cố này không làm thay đổi xác suất xảy ra của biến cố kia và ngược lại.

Trong trường hợp việc xảy ra hay không xảy ra của biến cố này làm thay đổi xác suất xảy ra của biến cố kia thì hai biến cố đó được gọi là phụ thuộc nhau.

Chú ý: Tính độc lập của các biến cố có tính tương hỗ. Nếu A và B độc lập với nhau thì A và $\overline{B}$, $\overline{A}$ và B, $\overline{A}$ và $\overline{B}$ cũng độc lập với nhau.

Định nghĩa 1.10:

Các biến cố $A_1, A_2, ..., A_n$ được gọi là độc lập từng đôi với nhau nếu mỗi cặp hai trong n biến cố đó độc lập với nhau.

Ví dụ 1.17:

Tung một đồng xu 3 lần. Gọi A_i $(i = \overline{1,3})$ là biến cố "Được mặt sấp ở lần tung thứ i". Rõ ràng mỗi cặp hai trong ba biến cố đó độc lập với nhau.

BÀI - 3. XÁC SUẤT CỦA BIẾN CỐ VÀ CÁC QUY TẮC TÍNH XÁC SUẤT CƠ BẢN

1.3.1. ĐỊNH NGHĨA CỔ ĐIỂN VỀ XÁC SUẤT

1.3.1.1. Định nghĩa

Xác suất xuất hiện biến cố A trong một phép thử là tỉ số giữa số kết cục thuận lợi cho A và tổng số các kết cục duy nhất đồng khả năng có thể xảy ra khi thực hiện phép thử đó.

Nếu kí hiệu P(A) là xác suất của biến cố A, m là số kết cục thuận lợi cho biến cố A, n là số kết cục duy nhất đồng khả năng của phép thử, ta có công thức sau:

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

Ví dụ 1.18:

Gieo một con xúc xắc cân đối, đồng chất. Gọi A là biến cố "Xuất hiện mặt có số chấm chẵn", B là biến cố "Xuất hiện mặt có số chấm nhỏ hơn 3". Tính xác suất của A, B.

Giải:

Khi gieo con xúc xắc một cách ngẫu nhiên ta có tổng số kết cục duy nhất đồng khả năng là 6. Kết cục thuận lợi cho biến cố A xảy ra là 3 và kết cục thuận lợi cho biến cố B xảy ra là 2. Nên ta có:

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

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

1.3.1.2. Các tính chất của xác suất

a) Xác suất của biến cố ngẫu nhiên là một số dương lớn hơn 0 và nhỏ hơn 1

$$0 < P(A) < 1$$

b) Xác suất của biến cố chắc chắn bằng 1

$$P(U) = 1$$

c) Xác suất của biến cố không thể có bằng 0

$$P(V) = 0$$

Như vậy xác suất của một biến cố bất kỳ luôn luôn thoả mãn điều kiện

$$0 \le P(A) \le 1$$

Mệnh đề đảo của hai tính chất b, c chưa chắc đúng, tức là nếu một biến cố có xác suất bằng 1 thì chưa chắc là biến cố chắc chắn và nếu một biến cố có xác suất bằng 0 thì chưa chắc đã là biến cố không thể có.

1.3.2. ĐỊNH NGHĨA THỐNG KÊ VỀ XÁC SUẤT

1.3.2.1. Định nghĩa tần suất

Tần suất suất hiện biến cố trong n phép thử là tỉ số giữa số phép thử trong đó biến cố xuất hiện và tổng số phép thử được thực hiện.

Như vậy, nếu kí hiệu số phép thử là n, số lần xuất hiện biến cố A là k, tần xuất suất hiện biến cố A là f(A) thì :

$$f(A) = \frac{k}{n}$$

Cùng với khái niệm xác suất, khái niệm tần suất là một trong những khái niệm cơ bản của lí thuyết xác suất.

Ví dụ 1.19:

- Khi kiểm tra ngẫu nhiên 80 sản phẩm do một nhà máy sản xuất, người ta phát hiện ra 3 phế phẩm. Gọi A là biến cố "xuất hiện phế phẩm". Vậy tần suất xuất hiện phế phẩm bằng:

$$f(A) = \frac{3}{80}$$

- Để nghiên cứu khả năng xuất hiện mặt sấp khi gieo đồng tiền người ta tiến hành tung một đồng tiền nhiều lần và thu được kết quả như sau:

Người làm thí nghiệm	**Số lần tung (n)**	**Số lần được mặt sấp (k)**	**Tần suất f(A) = k/n**
Buffon	4040	2048	0,5069
Pearson	12000	6019	0,5016
Pearson	24000	12012	0,5005

Qua ví dụ trên ta thấy khi số phép thử tăng lên thì tần suất xuất hiện của mặt sấp sẽ dao động ngày càng ít hơn xung quanh giá trị không đổi là 0.5. Tính ổn định của tần suất là cơ sở để đưa ra định nghĩa thống kê về xác suất

1.3.2.2. Định nghĩa thống kê về xác suất

Xác suất xuất hiện biến cố A trong một phép thử là một số p không đổi mà tần suất f xuất hiện biến cố đó trong n phép thử sẽ hội tụ theo xác suất về p khi số phép thử tăng lên vô hạn

Tức là $f_n(A) \xrightarrow[n\to\infty]{\text{hội tụ theo xác suất}} P(A$

Như vậy về mặt thực tế, với số phép thử n đủ lớn ta có thể lấy P(A) = f(A)

1.3.3. CÁC QUY TẮC TÍNH XÁC SUẤT

1.3.3.1. Công thức cộng xác suất

Định lý:

Xác suất của tổng hai biến cố xung khắc bằng tổng xác suất của các biến cố đó.

Nếu hai biến cố A và B xung khắc với nhau thì

$$P(A+B) = P(A) + P(B)$$

Hệ quả 1.

Xác suất của tổng các biến cố xung khắc từng đôi $A_1, A_2, ..., A_n$ *bằng tổng xác suất của các biến cố đó.*

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

Hệ quả 2.

Nếu các biến cố $A_1, A_2, ..., A_n$ *tạo nên nhóm đầy đủ các biến cố thì tổng xác suất của chúng bằng 1.*

$$P(A_1) + P(A_2) + ... + P(A_n) = 1$$

Hệ quả 3.

Tổng xác suất của hai biến cố đối lập bằng 1.

$$P(A) + P(\overline{A}) = 1$$

Ví dụ 1.20:

Trong hòm có 10 chi tiết, trong đó có hai chi tiết hỏng. Tính xác suất để khi lấy ngẫu nhiên ra 6 chi tiết thì có không quá 1 chi tiết hỏng.

Giải:

Gọi A_0 là biến cố "Trong 6 chi tiết lấy ra không có chi tiết hỏng", A_1 là biến cố "trong 6 chi tiết lấy ra có 1 chi tiết hỏng", A là biến cố "trong 6 chi tiết lấy ra có không quá 1 chi tiết hỏng "

$$A = A_0 + A_1$$

Vì A_0, A_1 là hai biến cố xung khắc nên

$$P(A) = P(A_0 + A_1) = P(A_0) + P(A_1)$$

Dùng định nghĩa cổ điển về xác suất ta có

$$P(A_0) = \frac{C_8^6}{C_{10}^6} = \frac{2}{15}$$

$$P(A_1) = \frac{C_2^1.C_8^5}{C_{10}^6} = \frac{8}{15}$$

Vậy $\quad P(A) = \frac{2}{15} + \frac{8}{15} = \frac{2}{3}$

Ví dụ 1.21:

Trong một cái hộp có 3 viên bi vàng, 4 viên bi đỏ, 5 viên bi xanh. Chọn ngẫu nhiên một lần 2 bi từ trong hộp. Tính xác suất của các biến cố sau:

A = "2 bi chọn ra cùng màu"

B = "2 bi chọn ra khác màu"

C = "2 bi chọn ra có ít nhất 1 viên bi đỏ".

Giải:

Tổng số kết quả đồng khả năng có thể xảy khi thực hiện phép thử là tổ hợp chập 2 của 12: $n = C_{12}^2$

Gọi A_1 là biến cố "2 bi chọn ra đều có màu vàng"

A_2 là biến cố "2 bi chọn ra đều có màu đỏ"

A_3 là biến cố "2 bi chọn ra đều có màu xanh"

Số kết cục thuận lợi cho A_1, A_2, A_3 xảy ra tương ứng là: C_3^2, C_4^2, C_5^2

$$P(A_1) = \frac{C_3^2}{C_{12}^2} = \frac{3}{66}; \quad P(A_2) = \frac{C_4^2}{C_{12}^2} = \frac{6}{66}; \quad P(A_3) = \frac{C_5^2}{C_{12}^2} = \frac{10}{66};$$

a. Gọi A là biến cố "2 bi chọn ra cùng màu": $A = A_1 + A_2 + A_3$

$$\Rightarrow P(A) = P(A_1) + P(A_2) + P(A_3) = \frac{3}{66} + \frac{6}{66} + \frac{10}{66} = \frac{19}{66}$$

b. Vì A là biến cố "2 bi chọn ra cùng màu" nên $\overline{A}$ là biến cố "2 bi chọn ra khác màu"

$$\Rightarrow P(\overline{A}) = 1 - P(A) = 1 - \frac{19}{66} = \frac{47}{66}$$

c. Gọi C là biến cố "2 bi chọn ra có ít nhất một bi đỏ", $\overline{C}$ là biến cố "2 bi chọn ra không có bi đỏ". Trong hộp có 8 bi không có màu đỏ nên số kết cục thuận lợi cho $\overline{C}$ là $C_8^2 = 28$

$$\Rightarrow P(\overline{C}) = \frac{28}{66} = \frac{14}{33} \;\Rightarrow P(C) = 1 - P(\overline{C}) = 1 - \frac{14}{33} = \frac{19}{33}$$

1.3.3.2. Định lý nhân

Định lý:

Xác suất của tích hai biến cố độc lập bằng tích các xác suất thành phần

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

Hệ quả:

Xác suất của tích n biến cố độc lập toàn phần bằng tích các xác suất thành phần

$$P(\prod_{i=1}^{n} A_i) = \prod_{i=1}^{n} P(A_i)$$

Ví dụ 1.22:

Có hai hộp đựng chi tiết. Hộp thứ nhất đựng 5 cái ốc, trong đó có 4 cái tốt, hộp thứ hai đựng 6 cái vít trong đó có 5 cái tốt. Lấy ngẫu nhiên mỗi hộp một chi tiết. Tính xác suất để lấy được một bộ ốc vít tốt.

Giải:

A_1 là biến cố "Lấy được cái ốc tốt"

A_2 là biến cố "Lấy được các vít tốt"

Khi đó A = $A_1 A_2$

Vì A_1, A_2 là hai biến cố độc lập nên

P(A) = $P(A_1).P(A_2) = \frac{4}{5}.\frac{5}{6} = \frac{2}{3}$

1.3.3.3. Mở rộng định lý cộng và định lý nhân xác suất

Định lý :

Xác suất của tổng hai biến cố không xung khắc bằng tổng xác suất của các biến cố đó trừ đi xác suất của tích các biến cố đó.

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

Nếu các biến cố A, B độc lập thì công thức trên có dạng

P(A + B) = P(A) + P(B) – P(A)P(B)

Còn nếu A, B là hai biến cố phụ thuộc thì công thức trên có dạng

P(A + B) = P(A) + P(B) – P(A).P(B/A)

Nếu A, B là hai biến cố xung khắc thì tích AB là biến cố không thể có, do đó P(AB) = 0. Ta thu được công thức cộng xác suất đã xét ở phần trước.

Hệ quả 1:

Xác suất của tổng n biến cố không xung khắc được xác định bằng công thức

$$P(\sum_{i=1}^{n} A_i) = \sum_{i} P(A_i) - \sum_{i<j} P(A_i A_j) + \sum_{i<j<k} P(A_i A_j A_k) + ... + (-1)^{n-1}.P(A_1 A_2 ... A_n)$$

Hệ quả 2:

Xác suất của tích n biến cố được xác định bằng công thức

$$P(\prod_{i=1}^{n} A_i) = \sum_i P(A_i) - \sum_{i<j} P(A_i + A_j) + \sum_{i<j<k} P(A_i + A_j + A_k) - ... + (-1)^{n-1}.P(A_1 + A_2 + ... + A_n)$$

Ví dụ 1.23:

Hai máy bay cùng ném bom một mục tiêu, mỗi máy bay ném một quả với xác suất trúng mục tiêu tương ứng là 0,7 và 0,8. Tìm xác suất để mục tiêu bị ném trúng.

Giải:

Gọi A_1 là biến cố ‘Quả bom thứ nhất ném trúng mục tiêu”

Gọi A_2 là biến cố ‘Quả bom thứ hai ném trúng mục tiêu”

Gọi A là biến cố “Mục tiêu bị ném trúng”. Áp dụng định lý trên ta có, A_1 và A_2 là không xung khắc và độc lập nên

$$P(A) = 1 - P(\overline{A_1}).P(\overline{A_2}) = 1 - 0,3.0,2 = 0.94.$$

BÀI - 4. XÁC SUẤT CÓ ĐIỀU KIỆN

1.4.1. XÁC SUẤT CÓ ĐIỀU KIỆN

Định nghĩa 1.11:

Xác suất của biến cố A được tính với điều kiện biến cố B đã xảy ra gọi là xác suất có điều kiện của A và kí hiệu là P(A/B).

Ví dụ 1.24:

Trong hộp có 10 sản phẩm, trong đó có 3 phế phẩm. Lấy ngẫu nhiên lần lượt 2 sản phẩm. Tính xác suất để lần thứ hai lấy được chính phẩm, biết rằng lần thứ nhất lấy được chính phẩm.

Giải:

Gọi A là biến cố "Lần thứ nhất lấy được chính phẩm", B là biến cố "Lần thứ hai lấy được chính phẩm".

Sau khi lần thứ nhất lấy được chính phẩm (biến cố A đã xảy ra) trong bình chỉ còn lại 7 sản phẩm, trong đó có 4 chính phẩm. Vậy xác suất có điều kiện của B là:

$$P(B/A) = \frac{6}{9}$$

1.4.2. CÔNG THỨC

Xác suất của tích hai biến cố A và B bằng tích của một trong hai biến cố đó với xác suất có điều kiện của biến cố còn lại

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

Hệ quả 1:

Nếu P(B) > 0 thì xác suất của biến cố A với điều kiện biến cố B xảy ra được tính theo công thức

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

Còn nếu P(B) = 0 thì xác suất trên không xác định. Tương tự, nếu P(A) > 0 thì ta có

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

Hệ quả 2:

Xác suất của tích n biến cố phụ thuộc bằng tích xác suất của n biến cố đó, trong đó xác suất của mỗi biến cố tiếp sau đều được tính với điều kiện tất cả các biến cố trước đó đã xảy ra.

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

Nếu A và B độc lập thì

$$P(A/B) = P(A) \qquad \text{và} \qquad P(B/A) = P(B)$$

Ví dụ 1.25:

Một lớp có 40 học sinh, trong đó có 15 học sinh giỏi Toán, 20 học sinh giỏi Văn, 10 học sinh giỏi cả Toán và Văn. Chọn ngẫu nhiên một học sinh của lớp đó. Biết rằng học sinh đó đã giỏi Toán, tính xác suất học sinh đó giỏi Văn.

Giải:

Gọi A là biến cố "Chọn được học sinh giỏi Văn"

Gọi B là biến cố "Chọn được học sinh giỏi Toán"

A.B là biến cố "Chọn được học sinh giỏi Văn và Toán". Ta có xác suất cần tìm là P(A/B)

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

Ta có: $P(A.B) = \frac{C_{10}^1}{C_{40}^1} = \frac{10}{40}$; $\qquad P(B) = \frac{C_{15}^1}{C_{40}^1} = \frac{15}{40}$

$$\Rightarrow P(A/B) = \frac{10}{40} : \frac{15}{40} = \frac{10}{15} = \frac{2}{3}$$

Ví dụ 1.26:

Trong hộp có 8 quả cầu trắng và 6 quả cầu đỏ giống nhau về hình dạng, kích thước và trọng lượng. Lấy ngẫu nhiên lần lượt không hoàn lại 3 quả cầu từ trong hộp. Tính xác suất để 3 quả lấy ra đều màu trắng.

Giải:

Gọi A_i là biến cố "Quả cầu lấy lần thứ i có màu trắng"; i = 1, 2, 3.

Gọi A là biến cố "3 quả cầu lấy ra đều có màu trắng"

$$A = A_1.A_2.A_3$$

Ta có: $P(A) = P(A_1.A_2.A_3) = P(A_1).P(A_2/A_1).P(A_3/A_1A_2)$

Vì lấy ra không hoàn lại nên

$$P(A_1) = \frac{8}{14},\ P(A_2/A_1) = \frac{7}{13},\ P(A_3/A_1A_2) = \frac{6}{12},\ P(A) = \frac{8}{14}.\frac{7}{13}.\frac{6}{12} = \frac{2}{13}$$

Ví dụ 1.27:

Một xí nghiệp có 3 ô tô hoạt động độc lập. Xác suất để trong một ngày các ô tô bị hỏng là 0,1; 0,2; 0,15. Tìm xác suất để trong một ngày có đúng một ô tô bị hỏng.

Giải:

Gọi A_i là biến cố "Ô tô thứ i bị hỏng trong ngày", i = 1, 2, 3.

A là biến cố "Trong ngày có đúng một ô tô bị hỏng".

Khi đó

$$A = A_1\overline{A_2}\,\overline{A_3} + \overline{A_1}A_2\overline{A_3} + \overline{A_1}\,\overline{A_2}A_3$$

$$P(A) = P(A_1)P(\overline{A_2})P(\overline{A_3}) + P(\overline{A_1})P(A_2)P(\overline{A_3}) + P(\overline{A_1})P(\overline{A_2})P(A_3)$$

Vì $P(A_1) = 0{,}1;\ P(A_2) = 0{,}2;\ P(A_3) = 0{,}15$

Nên

$$P(\overline{A_1}) = 0{,}9;\ P(\overline{A_2}) = 0{,}8;\ P(\overline{A_3}) = 0{,}85$$

$$P(A) = 0{,}1.0{,}8.0{,}85 + 0{,}9.0{,}2.0{,}85 + 0{,}9.0{,}8.0{,}15 = 0{,}329$$

BÀI - 5. CÔNG THỨC XÁC SUẤT ĐẦY ĐỦ - CÔNG THỨC BAYES

1.5.1. CÔNG THỨC XÁC SUẤT ĐẦY ĐỦ

Nhóm $H_1, H_2, ..., H_n$ là nhóm đầy đủ các biến cố. Giả sử biến cố A có thể xảy ra đồng thời với một trong các biến cố H_i. Lúc đó xác suất của biến cố A được tính bằng công thức:

$$P(A) = \sum_{i=1}^{n} P(H_i)P(A/H_i)$$

Các biến cố $H_1, H_2, ..., H_n$ thường được gọi là các giả thuyết.

Ví dụ 1.28:

Một nhà máy có 3 phân xưởng cách biệt cùng sản xuất một loại sản phẩm. Tỉ lệ sản phẩm của phân xưởng 1, 2, 3 lần lượt là 35%, 25%, 40%. Tỉ lệ phế phẩm của phân xưởng 1, 2, 3 tương ứng là 2%, 1%, 3%. Lấy ngẫu nhiên 1 sản phẩm trong kho hàng của nhà máy. Tính xác suất để sản phẩm đó là phế phẩm, cho biết ý nghĩa của xác suất này.

Giải:

Gọi H_i là biến cố "Sản phẩm lấy ra là do phân xưởng i sản xuất", (i= 1, 2, 3).

Theo bài ta có

$P(H_1) = 35\% = 0,35;$

$P(H_2) = 0,25;$

$P(H_3) = 0,4$

H_1, H_2, H_3 là một nhóm đầy đủ các biến cố.

Gọi A là biến cố "Sản phẩm chọn ra là phế phẩm". Áp dụng công thức xác suất đầy đủ, ta có:

$$P(A) = \sum_{i=1}^{3} P(H_i)P(A/H_i)$$

với

$P(A/H_1) = 2\% = 0,02$;

$P(A/H_2) = 0,01$;

$P(A/H_3) = 0,03$

$\Rightarrow P(A) = 0,35.0,02 + 0,25.0,01 + 0,4.0,03 = 0,0215 = 2,15\%$

Vậy xác suất cần tìm là 0,0215. Xác suất này chính là tỉ lệ phế phẩm của nhà máy.

Ví dụ 1.29:

Trong một chiếc hộp kín có 3 quả cầu xanh và 5 quả cầu đỏ hoàn toàn giống nhau. Chọn ngẫu nhiên lần lượt không hoàn lại 2 quả cầu từ trong hộp. Tính xác suất để quả cầu lấy lần thứ 2 có màu xanh.

Giải:

Gọi A là biến cố "Quả cầu chọn lần thứ hai có màu xanh", A xảy ra đồng thời với một trong các biến cố:

H_1 là biến cố "Quả cầu chọn lần một có màu xanh".

H_2 là biến cố "Quả cầu chọn lần một có màu đỏ".

Ta có: $P(H_1) = \frac{3}{8}$; $P(H_2) = \frac{5}{8}$

H_1, H_2 là một nhóm đầy đủ các biến cố.

Áp dụng công thức xác suất đầy đủ ta có:

$$P(A) = P(H_1)P(A/H_1) + P(H_2)P(A/H_2)$$

với

$$P(A/H_1) = \frac{2}{7};$$

$$P(A/H_2) = \frac{3}{7}$$

$$\Rightarrow P(A) = \frac{3}{8}.\frac{2}{7} + \frac{5}{8}.\frac{3}{7} = \frac{3}{8}$$

1.5.2. CÔNG THỨC BAYES

Giả sử biến cố A có thể xảy ra đồng thời với một trong n biến cố $H_1, H_2, ..., H_n$ tạo nên một nhóm đầy đủ các biến cố. Lúc đó:

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

Các biến cố $H_1, H_2, ..., H_n$ thường được gọi là các giả thuyết. Các xác suất $P(H_1)$, $P(H_2)$, ..., $P(H_n)$ được xác định trước khi phép thử tiến hành, do đó thường được gọi là các xác suất tiên nghiệm. Còn các xác suất $P(H_1 / A)$, $P(H_2 / A)$, ..., $P(H_n / A)$ được xác định sau khi phép thử đã tiến hành và biến cố A đã xảy ra, do đó được gọi là các xác suất hậu nghiệm. Như vậy công thức Bayescho phép đánh giá lại xác suất xảy ra các giả thuyết sau khi đã biết kết quả của phép thử là biến cố A đã xảy ra.

Ví dụ 1.30:

Cho 3 hộp bi. Hộp thứ 1 có 2 bi trắng và 1 bi đen. Hộp thứ 2 có 3 bi trắng và 1 bi đen. Hộp thứ 3 có 2 bi trắng và 2 bi đen. Lấy ngẫu nhiên một hộp, và từ đó lấy hú họa một bi. Tìm xác suất để bi đó là trắng. Khi lấy bi ra thấy nó là trắng, tìm xác suất để nó thuộc hộp thứ 2.

Giải:

a) A là biến cố "Lấy được bi trắng". A xảy ra đồng thời với một trong các biến cố:

H_i: "Lấy được hộp thứ i" (i = 1, 2, 3)

Theo công thức xác suất đầy đủ, ta có:

$$P(A) = \sum_{i=1}^{3} P(H_i)P(A/H_i)$$

với $P(H_i) = \frac{1}{3}$; (i = 1, 2, 3) và $P(A/H_1) = \frac{2}{3}$; $P(A/H_1) = \frac{3}{4}$; $P(A/H_3) = \frac{1}{2}$

$$\Rightarrow P(A) = \frac{1}{3}.\frac{2}{3} + \frac{1}{3}.\frac{3}{4} + \frac{1}{3}.\frac{1}{2} = \frac{23}{36}$$

b) Áp dụng công thức Bayes ta có:

$$P(H_2/A) = \frac{P(H_2).P(A/H_2)}{P(A)} = \frac{\frac{1}{3}.\frac{3}{4}}{\frac{23}{36}} = \frac{9}{23}$$

Ví dụ 1.31:

Có hai lô sản phẩm, lô thứ nhất có tỷ lệ chính phẩm là 3/4, còn lô thứ hai có tỷ lệ chính phẩm 2/3. Lấy ngẫu nhiên một lô và từ đó lấy ngẫu nhiên một sản phẩm thấy nó là chính phẩm. Sản phẩm được bỏ trở lại và từ lô đó lấy tiếp một sản phẩm. Tìm xác suất để lần thứ hai cũng lấy được chính phẩm.

Giải:

Gọi A là biến cố "Sản phẩm lấy lần đầu là chính phẩm". Biến cố A xảy ra với một trong hai giả thuyết sau:

H_1: "Lấy được lô thứ I"

H_2: "Lấy được lô thứ II"

Theo công thức xác suất đầy đủ ta có $P(A) = \sum_{i=1}^{2} P(H_i)P(A/H_i)$

Theo điều kiện đầu bài: $P(H_1) = P(H_2) = \frac{1}{2}$

$$P(A/H_1) = \frac{3}{4}; P(A/H_2) = \frac{2}{3}$$

Do đó: $P(A) = \frac{1}{2}.\frac{3}{4} + \frac{1}{2}.\frac{2}{3} = \frac{17}{24}$

Sau khi biến cố A đã xảy ra, xác suất của các biến cố H_1, H_2 thay đổi theo công thức Bayes như sau:

$$P(H_1/A) = \frac{P(H_1).P(A/H_1)}{P(A)} = \frac{3}{8}.\frac{17}{24} = \frac{9}{17}$$

$$P(H_2/A) = \frac{P(H_2).P(A/H_2)}{P(A)} = \frac{1}{3}.\frac{17}{24} = \frac{8}{17}$$

Gọi B là biến cố "Sản phẩm lấy lần thứ hai là chính phẩm". B có thể xảy ra với một trong hai giả thyết H_1 và H_2. Do đó theo công thức xác suất đầy đủ:

$$P(B) = P(H_1/A)P(B/H_1A) + P(H_2/A)P(B/H_2A)$$

Vì sản phẩm thứ nhất được bỏ lại lô, do đó tỷ lệ chính phẩm ở các lô đó vẫn không thay đổi. Vì thế:

$$P(B/H_1A) = \frac{3}{4};\ P(B/H_2A) = \frac{2}{3}$$

$$P(B) = \frac{9}{17}.\frac{3}{4} + \frac{8}{17}.\frac{2}{3} = \frac{145}{204} = 0{,}71$$

BÀI - 6. PHÉP THỬ LẶP VÀ CÔNG THỨC BERNOULLY

1.6.1. PHÉP THỬ LẶP

Tiến hành n phép thử, giả sử trong mỗi phép thử chỏ có thể xảy ra một trong hai trường hợp: hoặc biến cố A xảy ra hoặc biến cố A không xảy ra. Xác suất để A xảy ra trong mỗi phép thử đểu bằng p.

Ví dụ 1.32:

Tung 1 đồng xu nhiều lần, mỗi lần xác suất xảy ra mặt ngửa là 0,5 như nhau.

1.6.2. CÔNG THỨC BERNOULLY

Lược đồ Bernoully:

- Thực hiện n phép thử độc lập

- Trong mỗi phép thử chỉ có hai trường hợp: hoặc biến cố A xảy ra hoặc biến cố A không xảy ra.

- Xác suất xảy ra của biến cố A trong mỗi phép thử đều bằng q = 1 - q.

Những bài toán thỏa mãn cả ba điều giả thiết trên gọi là tuân theo lược đồ Bernoully.

Khi đó xác suất để trong n phép thử độc lập nói trên, biến cố A xuất hiện đúng x lần, kí hiệu là:

$$P_n(x) = C_n^x p^x q^{n-x}, \quad x = 0, 1, 2, \ldots, n$$

Ví dụ 1.33:

Xét trò chơi rút quân bài 10 lần mà xảy ra x lần được quân cơ là có thưởng. Tìm xác suất để cả 10 lần đều rút được quân cơ.

Giải:

Gọi A là biến cố "Rút được quân cơ", mà trong bộ bài (52 quân) có 13 quân cơ nên ta có $P(A) = \frac{13}{52} = \frac{1}{4}$

Xác suất để cả 10 lần đều rút được quân cơ là :

$$P_{10}(10) = C_{10}^{10}\left(\frac{1}{4}\right)^{10}\left(\frac{3}{4}\right)^{0} = \frac{1}{4^{10}}$$

CHƯƠNG.2. ĐẠI LƯỢNG NGẪU NHIÊN VÀ QUY LUẬT PHÂN PHỐI XÁC SUẤT

BÀI - 1. ĐỊNH NGHĨA VÀ PHÂN LOẠI BIẾN NGẪU NHIÊN

2.1.1. ĐỊNH NGHĨA

Một biến số được gọi là ngẫu nhiên nếu trong kết quả của phép thử nó sẽ nhận một và chỉ một trong các giá trị có thể có của nó tùy thuộc vào sự tác động của các nhân tố ngẫu nhiên.

Các biến ngẫu nhiên kí hiệu là X, Y, Z,… hoặc X_1, X_2, …..,X_n; Y_1, Y_2, …Y_n, ... còn các giá trị có thể có của chúng được kí hiệu là x, y, z hay x_1, x_2, …,x_n; y_1, y_2, …,y_n...

Sở dĩ biến X nào đó gọi là ngẫu nhiên vì trước khi tiến hành phép thử ta chưa có thể nói một cách chắc chắn nó sẽ nhận giá trị bằng bao nhiêu, mà chỉ có thể dự đoán điều đó với một xác suất nhất định. Nói cách khác việc X nhận một giá trị nào đó ($X = x_i$), i = 1, 2, ..., n, về thực chất là các biến cố ngẫu nhiên. Hơn nữa vì trong kết quả của phép thử biến X nhất định nhận một và chỉ một trong các giá trị có thể có của nó. Do đó $\{X = x_i\}, i = \overline{1, n};$ tạo nên một nhóm đầy đủ các biến cố.

Ví dụ 2.1:

Gieo 2 đồng tiền, gọi X là “Số mặt ngửa xuất hiện”. X là BNN vì trong kết quả của phép thử nó sẽ nhận được một trong 3 giá trị có thể có là 0, 1, 2.

Ví dụ 2.2:

Ném phi tiêu, gọi Y là biến cho biết khoảng cách từ tâm của bia đến vị trí chạm của phi tiêu. Y khi đó là một BNN.

2.1.2. PHÂN LOẠI BIẾN NGẪU NHIÊN

2.1.2.1. Biến ngẫu nhiên rời rạc

BNN gọi là rời rạc nếu các giá trị có thể có của nó lập nên một tập hợp hữu hạn hoặc đếm được.

Nói cách khác, BNN sẽ là rời rạc nếu ta có thể liệt kê được tất cả các giá trị có thể có của nó.

Ví dụ 2.3:

- Biến ngẫu nhiên chỉ số mặt ngửa xuất hiện khi gieo 2 đồng tiền là BNN rời rạc.

- Biến ngẫu nhiên chỉ số tuổi của sinh viên một trường đại học là một BNN rời rạc.

2.1.2.2. Biến ngẫu nhiên liên tục

BNN gọi là liên tục nếu các giá trị có thể có của nó lấp đầy một khoảng trên trục số.

Đối với BNN liên tục ta không thể liệt kê được tất cả các giá trị có thể có của nó.

Ví dụ 2.4:

BNN chỉ chiều cao của sinh viên được chọn ra từ một trường đại học hay BNN chỉ khoảng cách từ tâm bia đến điểm bắn trúng là các BNN liên tục.

Có thể nói rằng gần như tất cả các đại lượng ta gặp trong thực tế đều là các BNN và chúng sẽ thuộc về một trong hai loại BNN đã kể trên.

BÀI - 2. QUY LUẬT PHÂN BỐ XÁC SUẤT

Qui luật phân phối xác suất của biến ngẫu nhiên là sự tương ứng giữa các giá trị có thể có của nó và các xác suất tương ứng với giá trị đó.

Người ta thường dùng 3 phương pháp để mô tả quy luật phân phối xác suất của BNN. Ta sẽ lần lượt nghiên cứu các phương pháp đó.

2.2.1. BẢNG PHÂN PHỐI XÁC SUẤT, HÀM PHÂN BỐ XÁC SUẤT

2.2.1.1. Bảng phân bố xác suất

Định nghĩa 2.1:

Giả sử BNNRR X có thể nhận một trong các giá trị có thể có là $x_1, x_2, ..., x_n$ với các xác suất tương ứng là $p_1, p_2, .., p_n$. Thì bảng phân phối xác suất của BNNRR X có dạng như sau:

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

Chú ý: các xác suất p_i phải thoả mãn điều kiện

$$\begin{cases} p_i \geq 0 \quad \forall i \\ \sum_{i=1}^{n} p_i = 1 \end{cases}$$

Ví dụ 2.5:

Bảng phân phối xác suất của BNN chỉ số mặt ngửa xuất hiện khi gieo ngẫu nhiên 2 đồng tiền là:

X	0	1	2
P	$\frac{1}{4}$	$\frac{2}{4}$	$\frac{1}{4}$

Như ta đã biết các giá trị có thể có của X là 0, 1, 2

Và $P_1 = P(X = 0) = P(SS) = \frac{1}{4}$

$P_2 = P(X = 1) = P(SN, NS) = \frac{2}{4}$

$P_3 = P(X = 2) = P(SS) = \frac{1}{4}$

Ví dụ 2.6:

Trong một hộp có 3 viên bi đỏ và 2 viên bi xanh. Chọn ngẫu nhiên không hoàn lại các viên bi trong hộp cho đến khi được bi xanh thì dừng lại. Gọi X là biến ngẫu nhiên chỉ số bi được chọn ra. Lập bảng phân phối xác suất của X.

Giải:

Các giá trị có thể có của X là: 1, 2, 3, 4.

Gọi A_i là biến cố chọn bi lần i là đỏ

B_i là biến cố chọn bi lần i là xanh, $i = \overline{1,4}$

Ta có: $p_1 = P(X = 1) = P(B_1) = \frac{2}{5} = 0,4$

$p_2 = P(X = 2) = P(A_1.B_2) = P(A_1).P(B_2/A_1) = \frac{3}{5}.\frac{2}{4} = 0,3$

$p_3 = P(X = 3) = P(A_1A_2B_3) = P(A_1).P(A_2/A_1).P(B_3/A_1A_2) = 0,2$

$p_4 = 0,1$

Vậy bảng phân phối xác suất của X là:

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

Ví dụ 2.7:

Trong hộp có 6 chính phẩm và 4 phế phẩm. Lấy ngẫu nhiên 2 sản phẩm. Xây dựng qui luật phân phối xác suất của số chính phẩm được lấy ra.

Giải:

Gọi X là "Số chính phẩm được lấy trong 2 sản phẩm". Khi đó X là BNNRR với các giá trị có thể có là 0, 1, 2.

Ta có $p_1 = P(X = 0) = \frac{C_4^2}{C_{10}^2} = \frac{2}{15}$

$p_2 = P(X = 1) = \frac{8}{15}$; $p_3 = P(X = 3) = \frac{5}{15}$

Như vậy qui luật phân phối xác suất của X có dạng:

X	0	1	2
P	$\frac{2}{15}$	$\frac{8}{15}$	$\frac{5}{15}$

2.2.1.2. Hàm phân bố xác suất

Định nghĩa 2.2:

Khái niệm hàm phân bố xác suất áp dụng được đối với cả BNN rời rạc và liên tục. Giả sử X là BNN, x là một số thực nào đó. Xét biến cố "BNN X nhận giá trị nhỏ hơn X", kí hiệu (X < x). Hiển nhiên x thay đổi thì P(X < x) cũng thay đổi theo. Như vậy, xác suất này là một hàm số của x. Ta có định nghĩa sau:

Hàm phân bố xác suất của biến ngẫu nhiên X, kí hiệu F(x), là xác suất để biến ngẫu nhiên X nhận giá trị nhỏ hơn x, với x là một số thực bất kỳ:

$$F(x) = P(X < x)$$

Đối với từng loại BNN hàm phân bố xác suất được tính theo công thức riêng. Chẳng hạn nếu X là BNN rời rạc thì hàm phân bố xác suất được xác định bằng công thức:

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

Ví dụ 2.8:

Biến ngẫu nhiên rời rạc X có bảng phân phối xác suất như sau:

X	1	3	4
P	0,1	0,5	0,4

Hãy xây dựng hàm phân bố xác suất của X và vẽ đồ thị.

Giải:

Nếu $x \le 1$ thì (X < x) = $\varnothing$

$$\Rightarrow F(x) = P(X < x) = 0$$

Nếu $1 < x \le 3$ thì biến cố (X < x) chỉ xảy ra khi (X = 1)

$$\Rightarrow F(x) = 0,1$$

Nếu $3 < x \le 4$ thì biến cố (X < x) xảy ra khi (X = 1) hoặc (X = 3)

$$\Rightarrow F(x) = 0,1 + 0,5 = 0,6$$

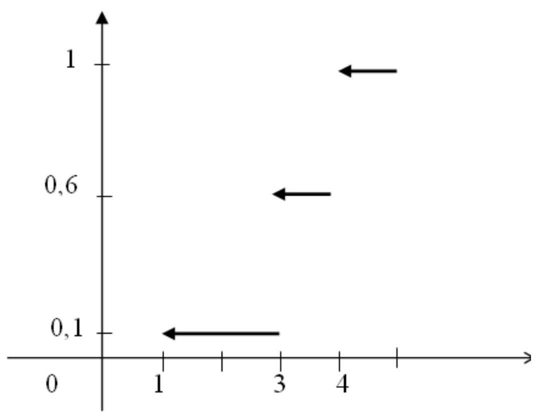
Nếu x > 4 thì thì biến cố (X < x) xảy ra khi (X = 1) hoặc (X = 3) hoặc (X = 4)

$$\Rightarrow F(x) = 0,1 + 0,5 + 0,4 = 1$$

Vậy hàm phân bố xác suất của x có dạng như sau:

$$F(x) = \begin{cases} 0 & khi \quad x \le 1 \\ 0,1 & khi \quad 1 < x \le 3 \\ 0,6 & khi \quad 3 < x \le 4 \\ 1 & khi \quad x > 4 \end{cases}$$

Đồ thị của hàm F(x) có dạng như sau:



2.2.1.3. Hàm mật độ xác suất

Đối với BNN liên tục X có thể dùng hàm phân bố xác suất để mô tả quy luật phân phối xác suất của nó. Tuy nhiên phương pháp này có hạn chế. Hàm phân bố xác suất không thể đặc trưng được xác suất để BNN liên tục X nhận một giá trị xác định. Vì thế đối với BNN liên tục người ta thường dùng hàm mật độ.

Định nghĩa 2.3:

Hàm mật độ xác suất của biến ngẫu nhiên liên tục X, kí hiệu f(x), là đạo hàm bậc nhất của hàm phân bố xác suất của biến ngẫu nhiên đó.

$$f(x) = F'(x)$$

Các tính chất

<u>Tính chất 1.</u>

Hàm mật độ xác suất luôn không âm.

$$f(x) \geq 0$$

<u>Tính chất 2.</u>

Xác suất để biến ngẫu nhiên liên tục X nhận giá trị trong khoảng (a,b) bằng tích phân xác định của hàm mật độ xác suất trong khoảng đó :

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

<u>Tính chất 3.</u>

Hàm phân bố xác suất F(x) của BNNLT X bằng tích phân suy rộng của hàm mật độ xác suất trong khoảng (-∞, x):

$$F(x) = \int_{-\infty}^{x} f(x)dx$$

<u>Tính chất 4.</u>

Tích phân suy rộng trong khoảng $(-\infty, \infty)$ của hàm mật độ xác suất bằng 1.

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

Chú ý:

Để hàm số f(x) có thể là hàm mật độ xác suất của BNN liên tục X thì nó phải thoã mãn hai tính chất cơ bản là tính chất 1 và tính chất 4, tức là:

$$F(x) = \begin{cases} f(x) \geq 0 & \forall x \\ \int_{-\infty}^{+\infty} f(x)dx = 1 & \end{cases}$$

Ví dụ 2.9:

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

$$F(x) = \begin{cases} 0 & khi \quad x \le 0 \\ ax^2 & khi \quad 0 < x \le 1 \\ 1 & khi \quad x > 1 \end{cases}$$

a. Tìm hệ số a

b. Tìm hàm mật độ xác suất f(x)

c. Tìm xác suất để biến ngẫu nhiên X nhận giá trị trong khoảng (0,25; 0,75)

Giải:

a. X là BNN liên tục nên F(x) là hàm liên tục. Hàm số F(x) liên tục $\forall x \neq 1$

F(x) liên tục tại x = 1

$$\Leftrightarrow \lim_{x \to 1^+} F(x) = F(1) = \lim_{x \to 1^-} F(x)$$

$$\Leftrightarrow a(1)^2 = 1$$

$$\Leftrightarrow a = 1$$

b. Từ định nghĩa hàm mật độ xác suất ta có:

$$f(x) = F'(x) = \begin{cases} 0 & khi \quad x \le 0 \\ 2x & khi \quad 0 < x \le 1 \\ 0 & khi \quad x > 1 \end{cases}$$

c. Theo tính chất của hàm phân bố xác suất

P(0,25 < X < 0,75) = F(0,75) – F(0,25)

= (0,75)2 – (0,25)2 = 0,5

BÀI - 3. CÁC THAM SỐ ĐẶC TRƯNG CỦA ĐẠI LƯỢNG NGẪU NHIÊN

2.3.1. KÌ VỌNG

2.3.1.1. Định nghĩa

Cho X là biến ngẫu nhiên. Kì vọng toán của BNN X là một số thực, kí hiệu E(X), được xác định như sau:

Nếu X là biến ngẫu nhiên rời rạc có bảng phân phối xác suất

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

Thì $$E(X)=\sum_{i=1}^{n} x_i p_i$$

Nếu X là biến ngẫu nhiên liên tục có hàm mật độ xác suất là f(x) thì

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

Ví dụ 2.10:

Tìm kì vọng toán của BNN

a. X là BNN rời rạc có bảng phân phối xác suất là

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

b. X là BNN liên tục có hàm mật độ là

$$f(x)=\begin{cases}0 & \text{khi} \quad x\notin[0,1]\\ 2x & \text{khi} \quad x\in[0,1]\end{cases}$$

Giải:

a. $$E(X)=0.\frac{1}{8}+1.\frac{3}{8}+2.\frac{3}{8}+3.\frac{1}{8}=1{,}5$$

b. $$E(X)=\int_{-\infty}^{\infty} xf(x)dx=\int_{-\infty}^{0} 0dx+\int_{0}^{1} 2x^2dx+\int_{1}^{\infty} 0dx=\frac{2}{3}$$

3.1.2. Các tính chất của kì vọng toán

<u>Tính chất 1.</u>

Kì vọng toán của một hằng số bằng chính hằng số đó

$$E(C) = C$$

<u>Tính chất 2.</u>

Kì vọng toán của tích giữa một hằng số với một BNN bằng tích giữa hằng số đó và kì vọng toán của BNN ấy

$$E(CX) = CE(X)$$

<u>Tính chất 3.</u>

Kì vọng toán của tổng hai BNN bằng tổng các kì vọng toán thành phần

$$E(X + Y) = E(X) + E(Y)$$

Hệ quả :

Kì vọng toán của tổng n BNN $X_1, X_2, .. , X_n$ bằng tổng các kì vọng toán thành phần.

$$E(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} E(X_i)$$

<u>Tính chất 4.</u>

Kì vọng toán của tích hai BNN độc lập bằng tích các kì vọng thành phần

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

Hệ quả :

Kì vọng toán của tích n BNN $X_1, X_2, .. , X_n$ độc lập lẫn nhau bằng tích các kì vọng thành phần

$$E\left(\prod_{i=1}^{n} X_i\right) = \prod_{i=1}^{n} E(X_i)$$

2.3.2. PHƯƠNG SAI

2.3.2.1. Định nghĩa

Phương sai của biến ngẫu nhiên X, kí hiệu V(X), là kì vọng toán của bình phương sai lệch của biến ngẫu nhiên so với kì vọng toán của nó.

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

a. Nếu X là BNN rời rạc thì phương sai được xác định theo công thức

$$V(X) = \sum_{i=}^{n} [x_i - E(X)]^2 p_i = \sum_{i=1}^{n} x_i^2 p_i - [E(X)]^2$$

b. Nếu X là BNN liên tục thì phương sai được xác định bằng công thức

$$V(X) = \int_{-\infty}^{\infty} [x - E(X)]^2 f(x)x = \int_{-\infty}^{\infty} x^2 f(x)dx - [E(X)]^2$$

Ví dụ 2.11:

Tìm phương sai của BNN sau:

a. X là BNN rời rạc có bảng phân phối xác suất là

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

b. X là BNN liên tục có hàm mật độ là

$$f(x) = \begin{cases} 0 & \text{khi} \quad x \notin [0,1] \\ 2x & \text{khi} \quad x \in [0,1] \end{cases}$$

Giải:

a. $$E(X) = 0.\frac{1}{8} + 1.\frac{3}{8} + 2.\frac{3}{8} + 3.\frac{1}{8} = 1{,}5$$

$$E(X^2) = 0^2.\frac{1}{8} + 1^2.\frac{3}{8} + 2^2.\frac{3}{8} + 3^2.\frac{1}{8} = 3$$

$$\Rightarrow V(X) = E(X^2) - [E(X)]^2 = 3 - (1{,}5)^2 = 0{,}75$$

b. $$E(X) = \int_{-\infty}^{\infty} xf(x)dx = \int_{-\infty}^{0} 0dx + \int_{0}^{1} 2x^2 dx + \int_{1}^{\infty} 0dx = \frac{2}{3}$$

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)dx = \int_{-\infty}^{0} 0dx + \int_{0}^{1} x^2\, 2xdx + \int_{1}^{\infty} 0dx = \frac{1}{2}$$

$$\Rightarrow V(X) = E(X^2) - [E(X)]^2 = \frac{1}{9}$$

Hoặc áp dụng công thức:

$V(X) = \int_{-\infty}^{\infty} x^2 f(x)dx - [E(X)]^2$ để tính phương sai V(X)

2.3.2.2. Các tính chất của phương sai

<u>Tính chất 1.</u>

Phương sai của một hằng số bằng 0: $V(C) = 0$

<u>Tính chất 2.</u>

Phương sai của tích giữa một hằng số và một BNN bằng tích giữa bình phương hằng số đó và phương sai của BNN ấy.

$$V(CX) = C^2V(X)$$

<u>Tính chất 3.</u>

Phương sai của tổng hai BNN độc lập bằng tổng các phương sai thành phần

$$V(X+Y) = V(X) + V(Y)$$

Hệ quả 1 :

Phương sai của tổng n BNN độc lập với nhau bằng tổng các phương sai thành phần

$$V\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} V(X_i)$$

Hệ quả 2 **:**

Phương sai của tổng một hằng số với một BNN bằng phương sai của BNN đó.

$$V(C + X) = V(X)$$

Hệ quả 3 **:**

Phương sai của hiệu hai BNN độc lập bằng tổng các phương sai thành phần

$$V(X - Y) = V(X) – V(Y)$$

2.3.3. ĐỘ LỆCH CHUẨN

Độ lệch chuẩn của BNN X, kí hiệu σ_x là căn bậc hai của phương sai:

$$\sigma_x = \sqrt{V(X)}$$

Ta thấy rằng đơn vị đo của phương sai bằng bình phương đơn vị đo của BNN. Vì vậy khi cần phải đánh giá mức độ phân tán của BNN theo đơn vị đo của nó người ta thường tính độ lệch tiêu chuẩn vì nó có cùng đơn vị đo với BNN cần nghiên cứu.

2.3.4. MỐT, TRUNG VỊ

2.3.4.1. Mốt

Mốt, ký hiệu là m_0, là giá trị của biến ngẫu nhiên tương ứng với:

- Xác suất lớn nhất nếu là biến ngẫu nhiên rời rạc
- Cực đại của hàm mật độ xác suất nếu là biến ngẫu nhiên liên tục.

Trong thực tế có thể gặp ngẫu nhiên không có giá trị Mốt hoặc ngược lại nhiều giá trị Mốt cùng một lúc.

2.3.4.2. Trung vị

Trung vị, ký hiệu là m_d là giá trị nằm ở chính giữa tập hợp các giá trị có thể có của biến ngẫu nhiên. Nói cách khác đó là giá trị chia phân phối của biến ngẫu nhiên thành hai phần bằng nhau.

Nếu X là biến ngẫu nhiên rời rạc thì giá trị X_i sẽ có trung vị m_d nếu thoả mãn điều kiện

$$F(X_i) \leq 0,5 < F(X_{i+1})$$

Còn nếu X là biến ngẫu nhiên liên tục thì trung vị m_d là giá trị thoả mãn điều kiện

$$\int_{-\infty}^{m_d} f(x)dx = 0{,}5$$

Ví dụ 2.12:

Tìm trung vị và Mốt của biến ngẫu nhiên có bảng phân phối xác suất như sau:

X	20	21	22	23	24	25
P	0,3	0,25	0,18	0,14	0,1	0,03

Giải:

Để tìm trung vị trước hết ta xây dựng hàm phân bố xác suất của X.

$$F(x) = \begin{cases} 0 & khi \quad x \le 20 \\ 0{,}3 & khi \quad 20 < x \le 21 \\ 0{,}55 & khi \quad 21 < x \le 22 \\ 0{,}73 & khi \quad 22 < x \le 23 \\ 0{,}87 & khi \quad 23 < x \le 24 \\ 1 & khi \quad x > 25 \end{cases}$$

Từ đó $m_d = 21$. Dễ thấy rằng $m_0 = 20$

BÀI - 4. CÁC QUY LUẬT PHÂN PHỐI XÁC SUẤT

2.4.1. PHÂN PHỐI NHỊ THỨC

2.4.1.1. Định nghĩa

Biến ngẫu nhiên rời rạc X nhận một trong các giá trị có thể có là 0, 1, 2, …, n với các xác suất tương ứng được tính bằng công thức (3.2) được gọi là phân phối theo qui luật nhị thức với các tham số n và p.

Quy luật nhị thức được kí hiệu là B(n, p).

Như vậy, bảng phân phối xác xuất của biến ngẫu nhiên X phân phối theo quy luật nhị thức có dạng:

X	0	1	...	x	...	n
P	$C_n^0p^0q^n$	$C_n^1p^xq^{n-1}$	...	$C_n^xp^xq^{n-x}$	...	$C_n^np^nq^0$

Trong thực tế, đôi khi ta phải tính xác suất để biến ngẫu nhiên X nhận giá trị trong khoảng [x, x+h] trong đó h là một số nguyên dương (h ≤ n – x). Lúc đó ta có thể tính xác suất này theo công thức:

$$P(x \le X \le x+h) = p_x + p_{x+1} + ... + p_{x+h} \qquad \textbf{(2-1)}$$

Trong đó mỗi xác suất thành phần được tính bằng công thức $P_x = C_n^xp^xq^{n-x}$

2.4.1.2. Các tham số đặc trưng của qui luật nhị thức

Giả sử biến ngẫu nhiên X phân phối theo qui luật nhị thức với các tham số n và p thì kì vọng toán: *E(X) = np* và phương sai: *V(X) = npq*.

Thật vậy, gọi X_i $(i = \overline{1,n})$ là số lần xuất hiện biến cố A trong thử thứ i. Lúc đó các phép thử tiến hành độc lập, các biến ngẫu nhiên X_i độc lập với nhau và mỗi X_i đều phân phối theo quy luật không - một với tham số là p. Như vậy số lần xuất hiện biến cố A trong n phép thử X bằng: $X = \sum_{i=1}^{n} X_i$

Theo tính chất của kỳ vọng toán và phương sai ta có:

$$E(X) = E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i)$$

Và $$V(X) = V\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} V(X_i)$$

Vì X_i $(i = \overline{1,n})$ cùng phân phối theo quy luật không - một với tham số p, do đó:

$$E(X_i) = p,\ i = \overline{1,n}$$

Và $$V(X_i) = pq$$

Từ đó: $E(X) = \sum_{i=1}^{n} E(X_i) = np$; $V(X) = \sum_{i=1}^{n} V(X_i) = npq$

Như vậy độ lệch tiêu chuẩn: $\sigma_x = \sqrt{V(X)} = \sqrt{npq}$

Ví dụ 2.13:

Một phân xưởng có 10 máy hoạt động, xác suất để một máy bị hỏng trong một ca là 0,2.

Tính xác suất để trong ca có không quá 2 máy bị hỏng.

Tính trung bình số máy bị hỏng trong ca.

Giải:

Gọi X là số máy bị hỏng của một ca, thì X tuân theo qui luật phân phối nhị thức với n = 10; p = 0,2.

a) Xác suất để trong ca có không quá 2 máy bị hỏng là:

$$P[X \le 2] = C_{10}^{0} 0{,}2^{0}.0{,}8^{10} + C_{10}^{1} 0{,}2^{1}.0{,}8^{9} + C_{10}^{2} 0{,}2^{2}.0{,}8^{8}$$

$$= 0{,}8^{10} + 10.0{,}2.0{,}8^{9} + 45.0{,}2^{2}.0{,}8^{8}$$

$$= 0{,}678$$

b) Trung bình số máy bị hỏng trong ca:

*E(X) = 10*0,2 = 2*

2.4.2. PHÂN PHỐI CHUẨN

2.4.2.1. Định nghĩa

Biến ngẫu nhiên liên tục X nhận các giá trị trong khoảng $(-\infty, +\infty)$ *gọi là phân phối theo quy luật chuẩn với các tham số* μ *và* σ^2*, nếu hàm mật độ xác suất của nó có dạng:*

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad \textbf{(2-2)}$$

Nếu tiến hành khảo sát hàm số trên và vẽ đồ thị của nó ta sẽ thu được các kết quả sau đây:

a. Hàm số xác định trên toàn trục Ox

b. Với mọi giá trị của x hàm số luôn luôn dương, như vậy, đồ thị của nó luôn nằm cao hơn trục Ox

c. Khi $x \to \pm\infty$ thì f(x) $\to$ 0 tức là trục Ox là đường tiệm cận ngang

d. Ta tìm đạo hàm bậc nhất $f'(x) = -\frac{x-\mu}{\sigma^3\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

Dễ dàng thấy rằng $f'(x) = 0$ khi $x = \mu$; $f'(x) > 0$ khi $x < \mu$, $f'(x) < 0$ khi $x > \mu$.

Như vậy khi x = μ hàm số có cực đại bằng $\frac{1}{\sigma\sqrt{2\pi}}$

e. Hiệu x - μ trong biểu thức của hàm f(x) nằm trong dạng bình phương, tức là hàm số đối xứng qua đường thẳng x = μ.

f. Ta tìm điểm uốn của hàm. Đạo hàm bậc hai

$$f''(x) = -\frac{1}{\sigma^3\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}\left[1-\frac{(x-\mu)^2}{\sigma^2}\right]$$

g. Dễ dàng thấy rằng khi x = $\mu+\sigma$ và x = $\mu-\sigma$ đạo hàm bậc hai bằng 0 và đi qua hai điểm đó nó đổi dấu (tại cả hai điểm đó hàm số đều bằng $\frac{1}{\sigma\sqrt{2\pi e}}$).

Như vậy các điểm: $\left(\mu-\sigma, \frac{1}{\sigma\sqrt{2\pi e}}\right)$ và $\left(\mu+\sigma, \frac{1}{\sigma\sqrt{2\pi e}}\right)$ là các điểm uốn.

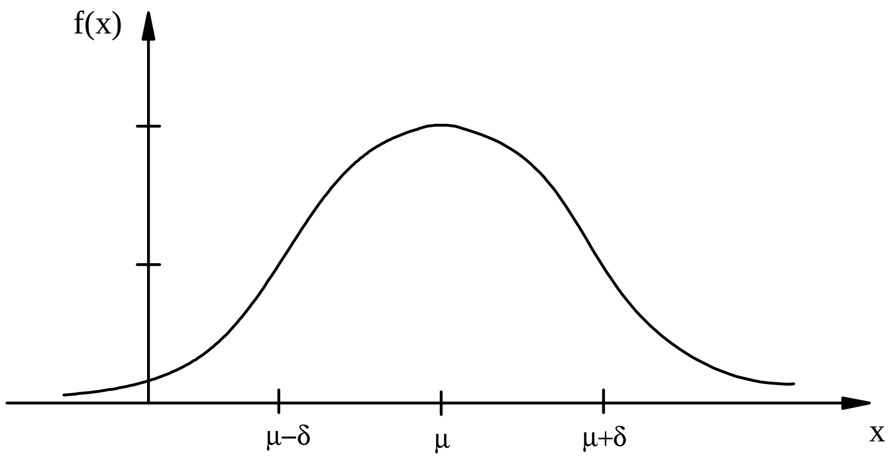
Vậy đồ thị của hàm mật độ xác suất của phân phối chuẩn có dạng như sau **(Hình 2-1)**

Hai tham số μ và σ có ý nghĩa rất quan trọng trong phân phối chuẩn (bản chất của nó sẽ được trình bày về sau). Khi μ và σ thay đổi, dạng đồ thị của hàm mật độ xác suất f(x) cũng thay đổi như sau:

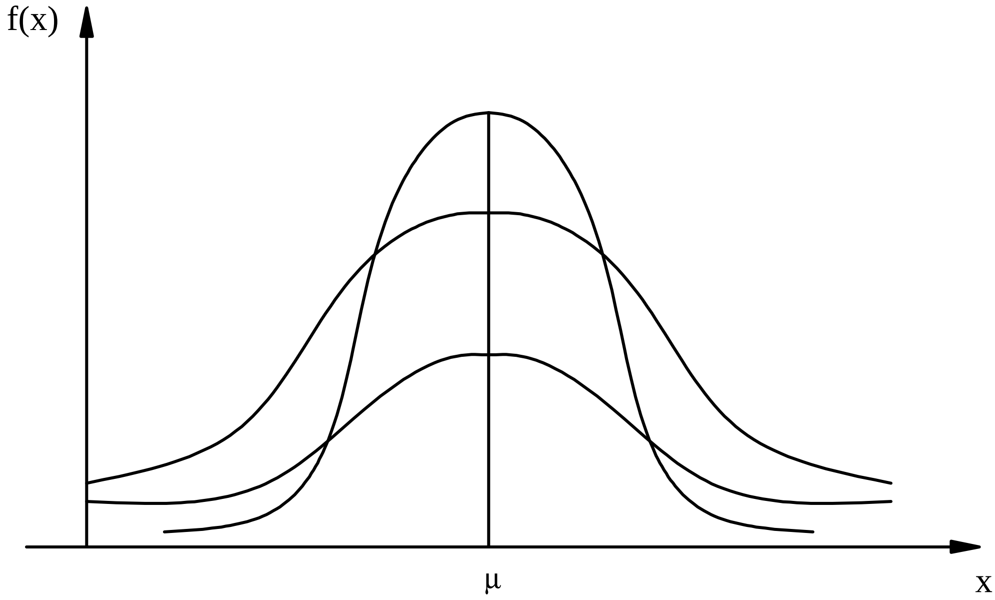
Khi μ thay đổi thì dạng của đường cong f(x) không thay đổi song nó sẽ chuyển dịch sang phải hoặc sang trái theo trục Ox. Khi μ tăng lên thì đồ thị sẽ dịch sang phải, còn khi μ giảm thì đồ thị sẽ dịch sang trái.

Khi σ thay đổi thì dạng của đồ thị sẽ thay đổi theo. Nếu σ tăng lên thì đồ thị sẽ thấp xuống và phình ra, còn khi σ giảm thì đồ thị sẽ cao và nhọn thêm.

Trên hình **(Hình 2-2)** ta minh họa đồ thị f(x) với ba giá trị khác nhau của σ.



Hình 2-1: Đồ thị hàm f(x) của phân phối chuẩn



Hình 2-2: Sự thay đổi của f(x) theo σ

Theo tính chất của hàm mật độ xác suất, ta có hàm phân bố của biến ngẫu nhiên X phân phối theo quy luật chuẩn được xác định bằng biểu thức:

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

2.4.2.2. Các tham số đặc trưng của quy luật chuẩn

Ta sẽ chứng minh rằng trong quy luật chuẩn thì μ chính là kỳ vọng toán còn σ chính là độ lệch chuẩn của X. Thật vậy, theo định nghĩa kỳ vọng toán của biến ngẫu nhiên liên tục ta có:

$$E(X) = \int_{-\infty}^{+\infty} xf(x)dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} xe^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

Ta thực hiện phép đổi biến số: $z = \dfrac{x-\mu}{\sigma}$

Từ đó $x = \sigma z + \mu$, $dx = \sigma dz$. Chú ý rằng khi đổi biến các cận lấy tích phân không thay đổi, ta có:

$$E(X) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} (\sigma z + \mu) e^{-\frac{z^2}{2}} dz = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \sigma z e^{-\frac{z^2}{2}} dz + \frac{\mu}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{z^2}{2}} dz$$

Tích phân thứ nhất bằng không do hàm dưới dấu tích phân là hàm lẻ mà cận lấy tích phân lại đối xứng. Còn tích phân thứ hai bằng: $\int_{-\infty}^{+\infty} e^{-\frac{z^2}{2}} dz = \sqrt{2\pi}$ (*tích phân Poisson*)

Do đó: *E(X)* = μ **(2-3)**

Theo định nghĩa phương sai của biến ngẫu nhiên liên tục và do *E(X)* = μ ta có:

$$V(X) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} (x-\mu)^2 e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

Ta thực hiện phép đổi biến số $z = \frac{x-\mu}{\sigma}$, từ đó $x - \mu = z\sigma$, $dx = \sigma dz$. Chú ý rằng cận lấy tích phân không thay đổi, ta có: $V(X) = \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} z^2 e^{-\frac{z^2}{2}} dz$

Lấy tích phân từng phần bằng cách đặt $u = z$, $dv = ze^{-\frac{z^2}{2}} dz$ ta tìm được $V(X) = \sigma^2$

Do đó: $$\sigma_x = \sqrt{V(X)} = \sigma \qquad \textbf{(2-4)}$$

Như vậy kỳ vọng toán của biến ngẫu nhiên X phân phối chuẩn là $E(X) = \mu$ và $V(X) = \sigma^2$. Phân phối chuẩn được ký hiệu $N(\mu, \sigma^2)$.

Có liên quan mật thiết với biến ngẫu nhiên phân phối chuẩn là một phân phối khác gọi là phân phối chuẩn hóa.

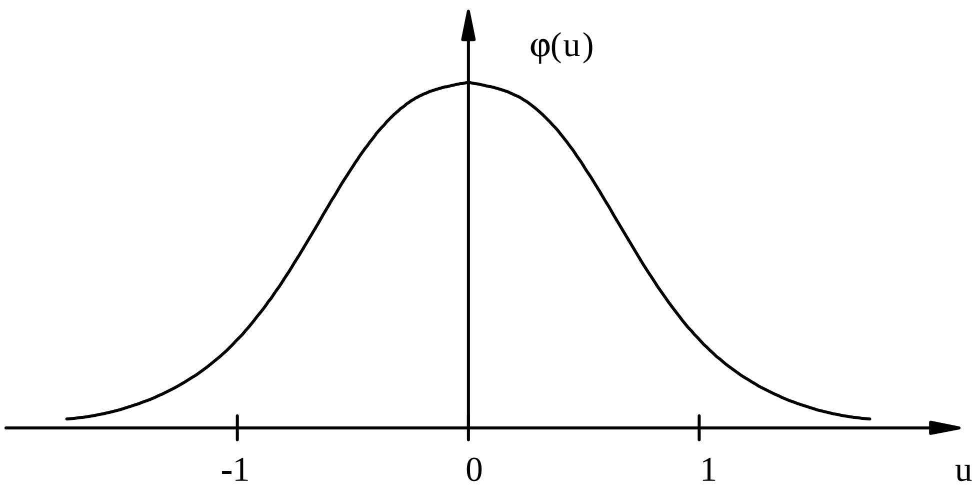
Giả sử biến ngẫu nhiên X phân phối chuẩn có kỳ vọng toán bằng μ và độ lệch chuẩn bằng σ. Xét biến ngẫu nhiên :

$$U = \frac{X - \mu}{\sigma}$$

2.4.2.3. Định nghĩa

Biến ngẫu nhiên U nhận các giá trị trong khoảng $(-\infty, +\infty)$ *gọi là tuân theo quy luật phân phối chuẩn hóa nếu hàm mật độ xác suất của nó có dạng:*

$$\varphi(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} \qquad \textbf{(2-5)}$$



Hình 2-3: Đồ thị của hàm φ(u)

Đồ thị của hàm φ(u) có dạng như hình vẽ.

Đặc điểm của đồ thị này là nó lấy trục tung làm trục đối xứng. Các giá trị của hàm $\varphi(u)$ được tính sẵn thành bảng **(Phụ lục 1)**.

Hàm phân bố xác suất của biến ngẫu nhiên U phân phối chuẩn hóa có dạng:

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

Ta tìm các tham số đặc trưng của biến ngẫu nhiên U phân phối chuẩn hóa:

$$E(U) = E\left(\frac{X - \mu}{\sigma} \right)$$

Theo tính chất của kỳ vọng toán ta có:

$$E(U) = \frac{1}{\sigma} E(X - \mu) = \frac{1}{\sigma}[E(X) - \mu]$$

Song *E(X) = μ*, do đó *E(U) = 0*.

$$V(U) = V\left(\frac{X - \mu}{\sigma} \right)$$

Cũng theo tính chất của phương sai ta có:

$$V(U) = \frac{1}{\sigma^2} V(X - \mu) = \frac{1}{\sigma^2} V(X)$$

Song $V(X) = \sigma^2$, do đó *V(U) = 1*.

Phân phối chuẩn hóa được ký hiệu là *N(0, 1)*.

Ngoài các tham số đặc trưng là kỳ vọng toán μ và phương sai σ^2, trong phân phối chuẩn có một tham số khác có nhiều ứng dụng trong thực tế, đó là giá trị tới hạn chuẩn.

2.4.2.4. Định nghĩa

Giá trị tới hạn chuẩn mức mức α, ký hiệu là u_α, là giá trị của biến ngẫu nhiên U có phân phối chuẩn hóa thỏa mãn điều kiện $P(U > u_\alpha) = \alpha$.

Vì U chuẩn hóa nên theo **(2-5)** ta có hàm mật độ của U là:

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

Theo tính chất hàm mật độ thì

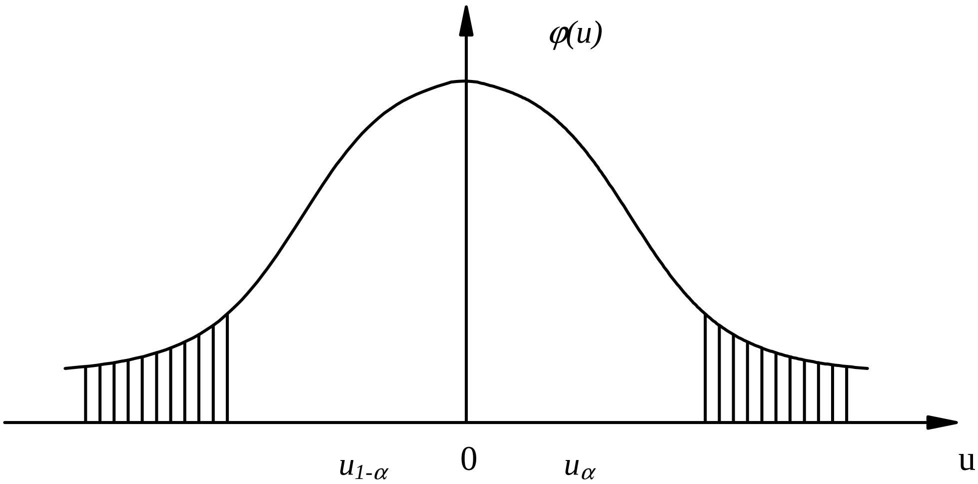
$$P(U > u_\alpha) = \int_{u_\alpha}^{+\infty} \varphi(u) du$$

Do đó: $$P(U > u_\alpha) = \frac{1}{\sqrt{2\pi}} \int_{u_\alpha}^{+\infty} e^{-\frac{u^2}{2}} = \alpha = \alpha.$$

Cho trước α, dựa vào biểu thức trên người ta tính được u_α và ngược lại.

Các giá trị của u_α được tính sẵn thành bảng **(Phụ lục 3)**.

Trên đồ thị giá trị tới hạn chuẩn u_α là giá trị sao cho diện tích giới hạn bởi đường cong phân phối chuẩn hóa, trục *Ou* và đường thẳng $u = u_\alpha$ bằng α.



Hình 2-4: Giá trị tới hạn chuẩn u_α

Từ hình vẽ ta thấy ngay giá trị tới hạn chuẩn có tính chất sau đây: $u_{1-\alpha} = -u_\alpha$

Sau đây ta sẽ xây dựng một số công thức có nhiều ứng dụng trong việc giải các bài toán thực tế.

2.4.2.5. Công thức tính xác xuất để biến ngẫu nhiên X phân phối chuẩn nhận giá trị trong khoảng (a, b)

Ta biết rằng nếu biến ngẫu nhiên liên tục X có hàm mật độ xác xuất là f(x) thì xác xuất để X nhận giá trị trong khoảng (a, b) sẽ được tính theo công thức

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

Giả sử X phân phối chuẩn. Lúc đó:

$$P(a < X < b) = \frac{1}{\sigma\sqrt{2\pi}} \int_a^b e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

Ta thay biến mới $z = \dfrac{x-\mu}{\sigma}$, từ đó $x = z\sigma + \mu$, $dx = \sigma dz$.

Ta tìm cận tích phân khi đổi biến.

Khi $x = a$ thì $z = \dfrac{a-\mu}{\sigma}$ và khi $x = b$ thì $z = \dfrac{b-\mu}{\sigma}$. Ta có

$$\Phi_0(u) = \frac{1}{\sqrt{2\pi}} \int_0^u e^{-\frac{z^2}{2}} dz$$

Trong đó:

- Giá trị của hàm $\Phi_0(\mathrm{u})$ được tính sẵn thành bảng **(Phụ lục 2)**.

Chú ý rằng hàm $\Phi_0(u)$ có các tính chất sau:

* $\Phi_0(-u) = -\Phi_0(u)$

* Với mọi *u > 5* thì $\Phi_0(u) \approx \Phi_0(5) = 0{,}5$

Các tính chất trên được vận dụng khi tra bảng giá trị hàm $\Phi_0(u)$.

Ta thu được công thức:

$$P(a < X < b) = \Phi_0\left(\frac{b-\mu}{\sigma}\right) - \Phi_0\left(\frac{a-\mu}{\sigma}\right) \qquad \textbf{(2-6)}$$

Chú ý: Do tính chất $\int_{-\infty}^{+\infty} \varphi(u)du = 1$ và hàm đồ thị hàm $\varphi(u)$ đối xứng nhau qua trục tung nên $\int_{-\infty}^{0} \varphi(u)du = \int_{0}^{+\infty} \varphi(u)du = 0{,}5$. Như vậy $P(-\infty < U < 0) = 0{,}5$

Từ đó: $\Phi(u) = P(-\infty < U < 0) + P(0 < U < u) = 0{,}5 + \Phi_0(u)$.

Vậy ta có mối liên hệ sau: $\Phi(u) = 0{,}5 + \Phi_0(u)$

Ví dụ 2.14:

Nghiên cứu chiều cao của nam giới khi trưởng thành ở một vùng dân cư, người ta nhận thấy rằng chiều cao đó tuân theo quy luật phân phối chuẩn với trung bình là 165cm và độ lệch chuẩn là 5cm.

Tìm tỷ lệ nam trưởng thành có tầm vóc trên 180cm

Tìm tỷ lệ nam trưởng thành có chiều cao từ 158cm đến 175cm

Những người có chiều cao dưới 155cm gọi là bị lùn. Tính xác suất để trong 4 người bất kỳ thì có ít nhất 1 người bị lùn.

Giải:

Gọi X là biến ngẫu nhiên chỉ chiều cao của nam giới khi trưởng thành. Theo giả thiết X có phân phối chuẩn với $\mu = 165;\ \sigma = 5$.

Tỷ lệ nam trưởng thành có tầm vóc trên 180cm là:

$$\mathrm{P}(\mathrm{X} > 180) = \Phi_0(+\infty) - \Phi_0\left(\frac{180-165}{5}\right) = 0,5 - \Phi_0(3)$$
$$= 0,5 - 0,4987 = 0,0013$$

Tỷ lệ nam trưởng thành có chiều cao từ 158cm đến 175cm là:

$$\mathrm{P}(158 < \mathrm{X} < 170) = \Phi_0\left(\frac{175-165}{5}\right) - \Phi_0\left(\frac{158-165}{5}\right)$$
$$= \Phi_0(2) - \Phi_0(-1{,}4) = \Phi_0(2) + \Phi_0(1{,}4)$$
$$= 0{,}4772 + 0{,}4192 = 0{,}8964$$

Xác suất để một người là bị lùn:

$$P(X < 155) = \Phi_0\left(\frac{155-165}{5}\right) - \Phi_0(-\infty) = -\Phi_0(2) + 0,5$$

$$= -0,4772 + 0,5 = 0,0228$$

Gọi Y là số người bị lùn trong số 4 người bất kỳ (các giá trị có thể có của Y = 0, 1, 2, 3, 4)

Xác suất để trong 4 người bất kỳ có ít nhất 1 người bị lùn là:

$$P(Y \geq 1) = 1 - P(Y < 1) = 1 - P(Y = 0)$$

$$= 1 - C_4^0 0,0228^0.0,9772^4$$

$$= 1 - 0,9772^4$$

$$\approx 0,0881$$

2.4.3. PHÂN PHỐI POISSON

2.4.3.1. Định nghĩa

Biến ngẫu nhiên rời rạc X nhận một trong các giá trị có thể có X = 0, 1, … với các xác suất tương ứng được tính bằng công thức (3.11) gọi là phân phối theo quy luật Poisson với tham số là λ.

Quy luật Poisson được ký hiệu là $P(\lambda)$.

Như vậy, bảng phân phối xác suất của biến ngẫu nhiên X phân phối theo quy luật Poisson có dạng:

X	0	1	…	x	…
P	$e^{-\lambda}\frac{\lambda^0}{0!}$	$e^{-\lambda}\frac{\lambda^1}{1!}$	…	$e^{-\lambda}\frac{\lambda^x}{x!}$	…

Nếu phải tìm xác suất để trong n phép thử biến ngẫu nhiên X phân phối theo quy luật Poisson nhận giá trị trong khoảng *[x, x + h]* trong đó h là một số nguyên dương và $h \leq n - x$. Lúc đó ta có thể tính xác suất này theo công thức:

$$P(x \leq X \leq x + h) = P_x + P_{x+1} + ... + P_{x+h} \qquad \textbf{(2-7)}$$

trong đó mỗi xác suất thành phần được tính bằng công thức **(2-6)**.

Giữa các xác suất P_x và P_{x-1} có mối liên hệ truy chứng sau đây:

$$P_x = \frac{\lambda}{x} P_{x-1} \qquad \textbf{(2-8)}$$

Thật vậy, xét tỷ số:

$$\frac{P_x}{P_{x-1}} = \frac{\frac{e^{-\lambda}\lambda^x}{x!}}{\frac{e^{-\lambda}\lambda^{x-1}}{(x-1)!}} = \frac{\lambda}{x}$$

Từ đó suy ra **(2-8)**.

Ví dụ 2.15:

Một lô hàng có tỷ lệ phế phẩm là 4%. Người ta kiểm tra 150 sản phẩm của lô hàng đó và nếu trong đó có không quá 2 phế phẩm thì lô hàng được chấp nhận. Tìm xác suất để lô hàng được chấp nhận.

Giải:

Bài toán thỏa mãn lược đồ Bernoulli song vì n = 150 > 20 và p = 0,04 < 0,1. Do đó nếu gọi X là số phế phẩm của lô hàng thì X là biến ngẫu nhiên rời rạc và có thể coi như phân phối theo quy luật Poisson với tham số là λ = np = 150.0,04 = 6. Xác suất để lô hàng được chấp nhận chính là xác suất để biến ngẫu nhiên X nhận giá trị trong khoảng [0, 2]. Theo công thức (3.12) ta có:

$$P(0 \le X \le 2) = P_0 + P_1 + P_2$$

$$P_0 = \frac{\lambda^0}{0!}e^{-\lambda} = \frac{1^0}{0!}(2{,}71)^{-6} = (2{,}71)^{-6}$$

$$P_1 = \frac{\lambda^1}{1!}e^{-\lambda} = \frac{1^1}{1!}(2{,}71)^{-6} = (2{,}71)^{-6}$$

$$P_2 = \frac{\lambda^2}{2!}e^{-\lambda} = \frac{1^2}{2!}(2{,}71)^{-6} = \frac{1}{2}(2{,}71)^{-6}$$

Do đó: $$P(0 \le X \le 2) = \left[1 + 1 + \frac{1}{2}\right](2{,}71)^{-6} = \frac{2{,}5}{2{,}71^6} \approx 0{,}0063$$

2.4.3.2. Các tham số đặc trưng của quy luật Poisson

Giả sử X phân phối theo quy luật Poisson. Ta sẽ chứng minh rằng:

$$E(X) = \lambda \qquad \textbf{(2-9)}$$

Thật vậy, theo định nghĩa của kỳ vọng toán, ta có:

$$E(X) = \sum_{x=0}^{\infty} xP_x = \sum_{x=0}^{\infty} xe^{-\lambda}\frac{\lambda^x}{x!} = e^{-\lambda}\sum_{x=1}^{\infty}\frac{\lambda^x}{(x-1)!} = \lambda e^{-\lambda}\sum_{x=1}^{\infty}\frac{\lambda^{x-1}}{(x-1)!}$$

Song ta lại có: $\sum_{x=1}^{\infty}\frac{\lambda^{x-1}}{(x-1)!} = e^{\lambda}$ do đó $E(X) = \lambda e^{-\lambda}e^{\lambda} = \lambda$

Bằng cách tính tương tự có thể tìm được $E(X^2) = \lambda^2 + \lambda$.

Do đó: $V(X) = \lambda^2 + \lambda - \lambda^2 = \lambda$

Vậy $$V(X) = \lambda \qquad \textbf{(2-10)}$$

Như vậy là trong quy luật Poisson cả kỳ vọng toán và phương sai đều bằng λ. Đó là tính chất đặc biệt của quy luật Poisson.

2.4.4. PHÂN PHỐI STUDENT

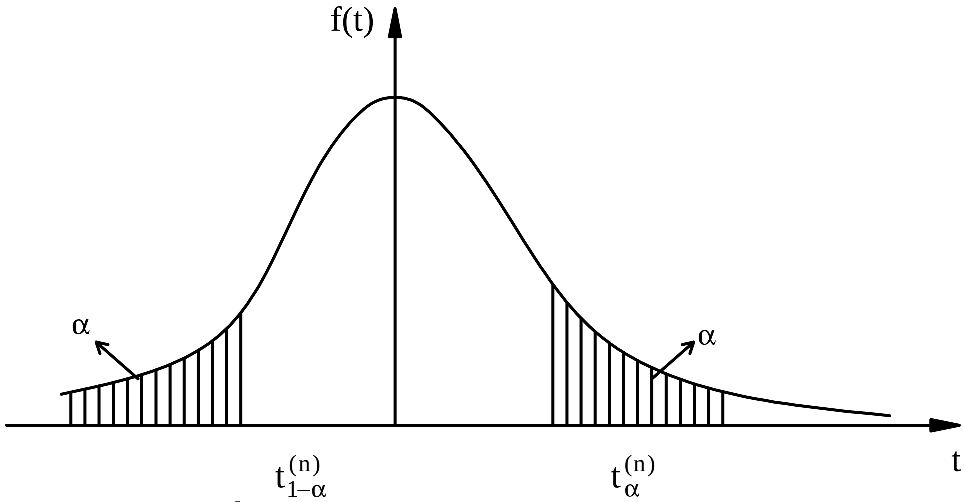
Biến ngẫu nhiên liên tục T gọi là phân phối theo quy luật Student với n bậc tự do nếu hàm mật độ xác suất của nó được xác định bằng biểu thức sau:

$$f(t) = \frac{\Gamma\left(\frac{n}{2}\right)}{\sqrt{\pi(n-1)}\Gamma\left(\frac{n-1}{2}\right)}\left[1+\frac{t^2}{n-1}\right]^{-\frac{n}{2}} \quad ,\forall t \qquad \textbf{(2-11)}$$

Trong đó $\Gamma(x)$ là hàm Gamma.

Đồ thị của hàm f(t) có dạng như ở **hình 2-5**.

Có thể chứng minh được rằng nếu biến ngẫu nhiên T phân phối theo quy luật Student với n bậc tự do thì kỳ vọng toán E(T) = 0 và phương sai $V(T) = \frac{n}{n-2}$



Hình 2-5: Đồ thị hàm f(t) của quy luật Student

Giá trị tới hạn Student, ký hiệu $t_{\alpha}^{(n)}$, là giá trị của biến ngẫu nhiên T phân phối theo quy luật Student với n bậc tự do, thỏa mãn điều kiện:

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

Giá trị tới hạn Student có tính chất sau đây:

$$t_{\alpha}^{(n)} = -t_{1-\alpha}^{(n)}$$

Ý nghĩa của nó được thể hiện trên **hình 2-5**.

Khi số bậc tự do tăng lên, phân phối Student sẽ hội tụ rất nhanh về phân phối chuẩn hóa. Do đó nếu n khá lớn (n > 30) có thể dùng phân phối chuẩn hóa thay cho phân phối Student.

Tuy nhiên cần phải nhấn mạnh rằng số bậc tự do nhỏ (n < 30) việc thay thế quy luật Student bằng quy luật chuẩn có thể dẫn đến những sai sót rất lớn. Chẳng hạn với n

= 4 và α = 0,05 thì giá trị tới hạn Student $t_{0,005}^{(4)} = 4,604$ trong khi đó $u_{0,005} = 2,58$, tức là lệch nhau 4,604 – 2,58 = 2,024.

Trong thực tế quy luật Student thường được sử dụng trong trường hợp sau đây: Giả sử có U là biến ngẫu nhiên phân phối chuẩn hóa N(0, 1) và biến ngẫu nhiên V độc lập với U, phân phối theo quy luật khi bình phương với n bậc tự do.

Nếu xét biến ngẫu nhiên $T = \dfrac{U}{\sqrt{\dfrac{V}{n}}}$

thì biến ngẫu nhiên T sẽ phân phối theo quy luật Student với n bậc tự do.

CHƯƠNG.3. TỔNG THỂ VÀ MẪU

BÀI - 1. KHÁI NIỆM TỔNG THỂ VÀ MẪU

3.1.1. TỔNG THỂ

Khi nghiên cứu một vấn đề người ta thường khảo sát trên một dấu hiệu nào đó, các dấu hiệu này thể hiện trên nhiều phần tử. Tập tất cả các phần tử mang dấu hiệu được gọi là tổng thể.

3.1.2. MẪU

Từ tổng thể lấy ra n phần tử.

Mẫu ngẫu nhiên kích thước n là tập hợp của n biến ngẫu nhiên độc lập X_1, X_2,…, X_n *được thành lập từ biến ngẫu nhiên X trong tổng thể nghiên cứu và có cùng quy luật phân phối xác suất với X.*

Mẫu ngẫu nhiên thường được ký hiệu là $W = (X_1, X_2, ..., X_n)$

3.1.3. CÁC PHƯƠNG PHÁP CHỌN MẪU

3.1.3.1. Mẫu đơn giản

Là loại mẫu được chọn trực tiếp từ danh sách đã được đánh số của tổng thể. Từ tổng thể kích thước N người ta dùng cách rút thăm đơn giản ra n phần tử của mẫu theo một bảng số ngẫu nhiên nào đó. Như vậy bất cứ phần tử nào của của tổng thể đều có thể được lấy vào mẫu với khả năng như nhau. Phương pháp này có ưu điểm cho phép thu được mẫu có tính đại diện cao, có thể suy rộng các kết quả của mẫu cho tổng thể với một sai số xác định, song để vận dụng phải có toàn bộ danh sách của tổng thể nghiên cứu, do vậy phương pháp này thường được sử dụng đối với tổng thể có quy mô nhỏ.

Chẳng hạn để điều tra sản lượng nuôi trồng thuỷ sản trong một thôn, người ta lập danh sách các hộ gia đình nuôi trồng thuỷ sản dựa trên sổ theo dõi nhân khẩu của thôn, sau đó rút thăm ngẫu nhiên không lặp lại từ danh sách đã lập để chọn ra các hộ cần điều tra.

3.1.3.2. Mẫu hệ thống

Là loại mẫu đã được đơn giản hoá trong cách chọn, trong đó chỉ có phần tử đầu tiên được chọn một cách ngẫu nhiên, sau đó dựa trên danh sách đã được đánh số của tổng thể để chọn các phần tử tiếp theo vào mẫu theo một thủ tục nào đó.

Chẳng hạn trên danh sách được đánh số thứ tự gồm các hộ nuôi trồng thuỷ sản của thôn, ta chọn ngẫu nhiên chủ hộ đầu tiên, giả sử có thứ tự là 3 trong danh sách, các hộ tiếp theo được chọn điều tra có số thứ tự cách nhau 4 đơn vị: 7, 11, 15,…

Nhược điểm của phương pháp này là dễ mắc sai số hệ thống nếu như tổng thể không được sắp xếp theo thứ tự ngẫu nhiên mà theo một thứ tự chủ quan nào đó. Tuy vậy do cách thức đơn giản của nó, phương pháp chọn mẫu ngẫu nhiên hệ thống thường được dùng ở cấp chọn mẫu cuối cùng khi tổng thể tương đối thuần nhất.

3.1.3.3. Mẫu phân nhóm

Trong chọn mẫu phân nhóm người ta chia tổng thể ra thành các nhóm có độ thuần nhất cao để chọn ra các phần tử đại diện cho từng nhóm. Việc phân nhóm có hiệu quả khi tổng thể nghiên cứu không thuần nhất theo dấu hiệu nghiên cứu. Sau khi đã phân nhóm thì kích thước mẫu được phân bổ cho mỗi nhóm theo một quy tắc nào đó, chẳng hạn tỷ lệ thuận với kích thước của mỗi nhóm.

Thí dụ như khi điều tra sản lượng khai thác thủy sản, người ta thường phân loại tàu thuyền theo nhóm kích thước, theo nghề,… vì đó là yếu tố ảnh hưởng đến loài và sản lượng đánh bắt, sau đó sử dụng hai phương pháp chọn mẫu ngẫu nhiên ở trên để chọn ra các đơn vị điều tra cuối cùng.

3.1.3.4. Mẫu chùm

Trong một số trường hợp, để tiện cho việc nghiên cứu người ta muốn quy diện nghiên cứu gọn về một khu vực nhất định chứ không để cho các phần tử của mẫu phân tán quá rộng. Người ta chia tổng thể nghiên cứu thành nhiều khối đơn vị và từ đó chọn ngẫu nhiên một số khối và điều tra tất cả các phần tử trong khối đã chọn. Theo phương pháp này tổng thể phải được chia thành các khối theo nguyên tắc:

- Mỗi phần tử của tổng thể chỉ được phân vào một khối

- Mỗi khối chứa nhiều phần tử khác nhau về dấu hiệu nghiên cứu, sao cho nó có độ phân tán cao như của tổng thể.

- Phân chia các khối tương đối đồng đều nhau về quy mô.

Chẳng hạn trong điều tra sản lượng khai thác thuỷ sản, do các cảng cá (bến cá) nằm phân tán ở ven biển nên cách chọn mẫu này tỏ ra là một lựa chọn phù hợp. Người ta sẽ chia các cảng cá (bến cá) thành các khối theo từng khu vực địa lý, sau đó chỉ điều tra một số khối. Cách chọn mẫu điều tra này phải thoả mãn hai tiêu chí:

- Nhóm được chọn điều tra phải có số lượng tàu thuyền hoạt động đủ lớn và có nhiều nghề hoạt động.

- Thuận lợi cho việc đi lại của người điều tra, tức là người điều tra có thể kết hợp đến một số hoặc toàn bộ các bến cá trong thời gian ngắn nhất.

Phương pháp chọn mẫu chùm có ưu điểm là tiết kiệm thời gian và chi phí đi lại và không cần phải lập danh sách tất cả các đơn vị trong tổng thể. Tuy nhiên phương pháp này cũng có nhược điểm là nếu các đơn vị mẫu tập trung, không phân bố đồng đều trong tổng thể sẽ làm giảm tính đại diện của mẫu nên sai số chọn mẫu sẽ tăng.

3.1.3.5. Mẫu nhiều cấp

Nếu các phần tử của tổng thể phân tán quá rộng và thiếu thông tin về chúng, người ta thường chọn mẫu theo nhiều cấp. Khi chọn nhiều cấp ta có các đơn vị mẫu ở mỗi cấp.

Chẳng hạn trong điều tra sản lượng khai thác, người ta thường chọn đơn vị mẫu cấp 1 là các bến cá và đơn vị mẫu cấp 2 là các tàu khai thác thuỷ sản.

Việc chọn mẫu ở mỗi cấp có thể tiến hành theo phương pháp chọn mẫu ngẫu nhiên đơn giản, chọn mẫu hệ thống, chọn mẫu chùm hay chọn mẫu phân nhóm.

BÀI - 2. MÔ HÌNH XÁC SUẤT CỦA TỔNG THỂ MẪU

3.2.1. CÁC THAM SỐ ĐẶC TRƯNG

3.2.1.1. Trung bình tổng thể

Giả sử trong tổng thể kích thước N dấu hiệu định lượng χ nhận các giá trị $x_1, x_2, \ldots, x_N$. Trung bình tổng thể, ký hiệu là m, là trung bình số học của các giá trị của dấu hiệu trong tổng thể.

$$m = \frac{1}{N}\sum_{i=1}^{N} x_i$$

3.2.1.2. Phương sai tổng thể

Phương sai tổng thể, ký hiệu là σ^2, là trung bình số học của bình phương các sai lệch giữa các giá trị của dấu hiệu trong tổng thể và trung bình tổng thể.

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - m)^2$$

Trong thực tế, để tiện cho tính toán, phương sai tổng thể thường được tính bằng công thức sau: $\sigma^2 = \frac{1}{N}\sum_{i=1}^{k} N_i x_i^2 - m^2$

3.2.2. MẪU NGẪU NHIÊN

Lấy n phần tử của tổng thể theo phương pháp hoàn lại để quan sát. Gọi giá trị X_i là giá trị của X^* đo được trên phần tử thứ $i(i=\overline{1,n})$ thì $X_1, X_2, ..., X_n$ là các đại lượng ngẫu nhiên độc lập có cùng phân phối như X. Khi đó bộ $(X_1, X_2, ..., X_n)$ được gọi là mẫu ngẫu nhiên kích thước n được tạo nên từ đại lượng ngẫu nhiên gốc X.

Kí hiệu $W_X = (X_1, X_2, ..., X_n)$

Giả sử Xi nhận giá trị $x_i(i=\overline{1,n})$. Khi đó $(x_1, x_2, ...,x_n)$ là một giá trị cụ thể của mẫu ngẫu nhiên W_X, được gọi là mẫu cụ thể. Kí hiệu $w_X = (x_1, x_2, ..., x_n)$.

Ví dụ 3.1:

Kết quả môn Toán của một lớp gồm 100 sinh viên cho bởi bảng sau:

Điểm	3	4	5	6	7
Số sinh viên có điểm tương ứng	25	20	40	10	5

Giải:

Gọi X là điểm môn Toán của một sinh viên được chọn ngẫu nhiên trong danh sách lớp thì X là đại lượng ngẫu nhiên có phân phối

X	3	4	5	6	7
P	0,25	0,20	0,40	0,10	0,05

Chọn ngẫu nhiên 5 sinh viên trong danh sách lớp để xem điểm. Gọi Xi là điểm của sinh viên thứ i. Ta có mẫu ngẫu nhiên kích thước n=5 được xây dựng từ đại lượng ngẫu nhiên X.

$$W_X = (X_1, X_2, ..., X_n)$$

Giả sử sinh viên thứ nhất được 4 điểm, thứ hai được 3 điểm, thứ 3 được 6 điểm, thứ tư được 7 điểm và thứ năm được 5 điểm. Ta được mẫu cụ thể:

$$w_X = (4, 3,6,7,5)$$

BÀI - 3. THỐNG KÊ

3.3.1. TRUNG BÌNH MẪU

Trung bình mẫu là một thống kê, ký hiệu là $\overline{X}$ và là trung bình số học của các giá trị mẫu: $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

Khi mẫu ngẫu nhiên nhận một giá trị cụ thể $w = (x_1, x_2, ..., x_n)$ thì trung bình mẫu cũng nhận giá trị cụ thể bằng: $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

Nếu biến ngẫu nhiên gốc X có $E(X) = m$ và $V(X) = \sigma^2$ thì $E(\overline{X}) = m$ và $V(\overline{X}) = \frac{\sigma^2}{n}$.

3.3.2. PHƯƠNG SAI MẪU, PHƯƠNG SAI HIỆU CHỈNH MẪU

Phương sai mẫu là một thống kê, ký hiệu S^2 được xác định bằng công thức sau:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} X_i^2 - n\overline{X}^2\right]$$

trong đó $\overline{X}$ là trung bình của mẫu ngẫu nhiên.

Phương sai hiệu chỉnh mẫu, kí hiệu S^{*2} được xác định bằng biểu thức sau:

$$S^{*2} = \frac{1}{n}\sum_{i=1}^{n}(X_i - m)^2$$, trong đó *m* là trung bình của tổng thể.

Khi mẫu ngẫu nhiên nhận một giá trị cụ thể $w = (x_1, x_2, ..., x_n)$ thì phương sai mẫu cũng nhận giá trị cụ thể bằng:

$$s^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - n\overline{x}^2\right] = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 n_i - \frac{\left(\sum_{i=1}^{n} x_i n_i\right)^2}{n}\right]$$

Nếu biến ngẫu nhiên gốc X có $E(X) = m$ và $V(X) = \sigma^2$ thì $E(S^2) = \sigma^2$ và $V(S^{*2}) = \sigma^2$

3.3.3. ĐỘ LỆCH CHUẨN MẪU

Độ lệch chuẩn mẫu, ký hiệu S, là căn bậc hai của phương sai mẫu.

Như vậy: $S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2}$

Giá trị của nó trên một mẫu cụ thể là một số xác định, ký hiệu là s.

CHƯƠNG.4. ƯỚC LƯỢNG THAM SỐ CỦA ĐẠI LƯỢNG NGẪU NHIÊN

BÀI - 1. PHƯƠNG PHÁP ƯỚC LƯỢNG ĐIỂM

4.1.1. PHƯƠNG PHÁP HÀM ƯỚC LƯỢNG

Định nghĩa 4.1:

Giả sử cần ước lượng tham số θ của biến ngẫu nhiên gốc X, từ tổng thể lập mẫu ngẫu nhiên kích thước n:

$$W = (X_1, X_2, \ . \ . \ . \ , X_n)$$

Chọn thống kê $\theta^* = f(X_1, X_2, \ . \ . \ . \ , X_n)$, ta gọi là hàm ước lượng của X.

Thực hiện phép thử ta được mẫu cụ thể $w = (x_1, x_2, \ . \ . \ . \ , x_n)$. Khi đó ước lượng điểm của θ là giá trị $\theta^*_0 = f(x_1, x_2, \ . \ . \ . \ , x_n)$.

4.1.2. CÁC TIÊU CHUẨN LỰA CHỌN HÀM ƯỚC LƯỢNG

4.1.2.1. Ước lượng không chệch

Định nghĩa 4.2:

Thống kê θ* của mẫu được gọi là ước lượng không chệch của tham số θ của biến ngẫu nhiên gốc nếu

$$E(\theta^*) = \theta$$

Ngược lại, nếu $E(\theta^*) \neq \theta$ thì θ^* được gọi là ước lượng chệch của θ

Chú ý:

θ* là ước lượng không chệch của θ không có nghĩa là mọi giá trị của θ* đều bằng với θ mà chỉ có nghĩa: trung bình các giá trị của θ* bằng θ. Từng giá trị của θ* có thể sai lệch rất lớn so với θ.

4.1.2.2. Ước lượng hiệu quả

Định nghĩa 4.3:

Thống kê của mẫu được gọi là ước lượng hiệu quả nhất của tham số θ của biến ngẫu nhiên gốc X nếu nó là ước lượng không chệch và có phương sai nhỏ nhất so với mọi ước lượng không chệch khác được xây dựng trên cùng mẫu đó.

Nếu θ* đã là một ước lượng không chệch của θ thì trong nhiều trường hợp giá trị nhỏ nhất của phương sai V(θ*) có thể tìm được dựa vào bất đẳng thức Cramer – Rao được phát biểu như sau:

Cho mẫu ngẫu nhiên $W = (X_1, X_2, \ . \ . \ . \ , X_n)$ được xây dựng từ biến ngẫu nhiên gốc X có hàm mật độ xác suất f(x,θ) thỏa mãn một số điều kiện nhất định (thường được thỏa mãn trong thực tế, là các phân phối xác suất đã xét ở chương trước) và θ* là một ước lượng không chệch bất kỳ của θ thì

$$V(\theta^*) \geq \frac{1}{ne\left[\dfrac{\partial(\ln f(x,\theta))}{\partial\theta}\right]}$$

Người ta chứng minh được trung bình mẫu $\overline{X}$ là ước lượng hiệu quả nhất của kỳ vọng toán μ của biến ngẫu nhiên gốc X trong tổng thể khi X phân phối chuẩn N(μ, σ^2).

4.1.2.3. Ước lượng vững.

Định nghĩa 4.4:

Thống kê θ* của mẫu được gọi là ước lượng vững của tham số θ của biến ngẫu nhiên gốc X nếu θ* hội tụ theo xác suất đến θ khi n → ∞.

Tức là với mọi ε dương bé tùy ý ta luôn có:

$$\lim_{n\to\infty} P(|\theta^* - \theta| < \varepsilon) = 1$$

BÀI - 2. PHƯƠNG PHÁP ƯỚC LƯỢNG BẰNG KHOẢNG TIN CẬY

4.2.1. KHÁI NIỆM

Định nghĩa 4.5:

Khoảng (G_1,G_2) của thống kê G được gọi là khoảng tin cậy của tham số θ nếu với xác suất bằng (1-α) cho trước thỏa mãn điều kiện :$P(G_1 < \theta < G_2) = 1- \alpha$

Xác suất (1-α) được gọi là độ tin cậy của ước lượng, còn I = $G_2 - G_1$ được gọi là độ dài khoảng tin cậy.

4.2.2. ƯỚC LƯỢNG TRUNG BÌNH

Giả sử trong tổng thể biến ngẫu nhiên gốc X phân phối chuẩn N(μ,σ^2) nhưng chưa biết tham số μ của nó. Để ước lượng μ từ tổng thể ta lập mẫu ngẫu nhiên kích thước n : W = (X_1, X_2, ..., X_n).

Để chọn thống kê G thích hợp ta xét hai trường hợp sau

4.2.2.1. Đã biết phương sai σ^2 của biến ngẫu mhiên gốc X trong tổng thể

Lúc đó ta chọn thống kê $G = U = \dfrac{\overline{X}-\mu}{Se(\overline{X})} = \dfrac{(\overline{X}-\mu)\sqrt{n}}{\sigma}$

Khoảng tin cậy đối xứng: $\alpha_1 = \alpha_2 = \alpha/2$.

Lúc này khoảng tin cậy của μ là $\left(\overline{X}-\varepsilon, \overline{X}+\varepsilon\right)$

Với: $\varepsilon = \dfrac{\sigma}{\sqrt{n}}U_{\alpha/2}$, $\overline{X}$ là trung bình của mẫu cụ thể

ε được gọi là độ chính xác của ước lượng. Nó phản ánh mức độ sai lệch của trung bình mẫu so với trung bình của tổng thể với xác xuất (1- α) cho trước.

- Khoảng tin cậy bên phải: $\alpha_1 = 0$ và $\alpha_2 = \alpha$. Khi đó $U_{\alpha_1} = U_0 = +\infty$

Khoảng tin cậy của μ là $\overline{X} - \dfrac{\sigma}{\sqrt{n}}U_{\alpha};+\infty$

Biểu thức này được dùng để ước lượng giá trị tối thiểu của μ.

- Khoảng tin cậy bên trái: $\alpha_2 = 0$ và $\alpha_1 = \alpha$. Khi đó $U_{\alpha} = U_0 = +\infty$

Khoảng tin cậy của μ là $-\infty, \overline{X} + \dfrac{\sigma}{\sqrt{n}}U_{\alpha}$

Biểu thức này được dùng để ước lượng giá trị tối đa của μ.

Ví dụ 4.1:

Trọng lượng của một loại sản phẩm là biến ngẫu nhiên phân phối theo quy luật chuẩn với độ lệch chuẩn là 1 gam. Cân thử 25 sản phẩm loại này ta thu được kết quả như sau

Trọng lượng (gam)	18	19	20	21
Số sản phẩm tương ứng	3	5	15	2

Với độ tin cậy 0,95 hãy tìm khoảng tin cậy đối xứng của trọng lượng trung bình của loại sản phẩm nói trên.

Giải:

Gọi X là "Trọng lượng sản phẩm" theo giải thiết X phân phối chuẩn với σ =1. Vậy trọng lượng trung bình của sản phẩm chính là tham số μ. Đây là bài toán ước lượng bằng khoảng tin cậy đối xứng giá trị của tham số μ của phân phối N(μ, σ^2) khi đã biết phương sai của nó. Vậy ta có khoảng tin cậy

$$\left(\overline{X} - \frac{\sigma}{\sqrt{n}}U_{\alpha/2}; \overline{X} + \frac{\sigma}{\sqrt{n}}U_{\alpha/2}\right)$$

Lấy từ tổng thể ra một mẫu ngẫu nhiên kích thước n = 25, gọi X_i là trọng lượng của sản phẩm thứ i (i = $\overline{1,25}$) ta có

$$W = (X_1, X_2, \ldots, X_{25})$$

Từ đó $\overline{X} = \frac{1}{25}\sum_{i=1}^{25} X_i$

Với độ tin cậy 1 - α = 0,95 thì α/2 = 0,025

Tra bảng giá trị tới hạn chuẩn có $U_{0,025} = 1,96$

Vậy khoảng tin cậy đối xứng của μ là

$$\left(\overline{X} - \frac{1}{\sqrt{25}}1,96; \overline{X} + \frac{1}{\sqrt{25}}.1,96\right) = (\overline{X} - 0,392;\ \overline{X} + 0,392)$$

Từ bảng số liệu tìm được trung bình mẫu cụ thể

$$\overline{x} = \frac{3.18 + 3.19 + 15.20 + 2.21}{25} = 19,64$$

Vậy với độ tin cậy 0,95 qua mẫu cụ thể này, khoảng tin cậy đối xứng của μ là

(19,64 – 0,392; 19,64 + 0,392) hay (19,248 < μ < 20,032)

Kết quả thu được cho biết 95% qua mẫu kích thước n = 25 sẽ chứa đựng tham số μ trong khoảng (19,248; 20,032).

4.2.2.2. Chưa biết phương sai σ2 của biến ngẫu nhiên gốc X trong tổng thể và kích thước mẫu n < 30

Lúc đó thống kê G = T = $\frac{(\overline{X} - \mu)\sqrt{n}}{S}$ với độ tin cậy bằng (1- α).

- Khoảng tin cậy đối xứng khi $\alpha_1 = \alpha_2 = \alpha/2$: $\left(\overline{X}-\varepsilon, \overline{X}+\varepsilon\right)$, với $\overline{X}$ là trung bình của mẫu cụ thể , $\varepsilon = \frac{S}{\sqrt{n}} T_{\alpha/2}^{(n-1)}$, S là độ lệch tiêu chuẩn điều chỉnh của mẫu cụ thể.

- Khoảng tin cậy bên phải khi $\alpha_1 = 0$; $\alpha_2 = \alpha$: $\left(\overline{X} - \frac{S}{\sqrt{n}} T_{\alpha}^{(N-1)}; +\infty\right)$

- Khoảng tin cậy bên trái khi $\alpha_2 = 0$; $\alpha_1 = \alpha$: $\left(-\infty; \overline{X} + \frac{S}{\sqrt{n}} T_{\alpha}^{(n-1)}\right)$

Trong trường hợp này độ dài khoảng tin cậy I cũng là ngắn nhất khi khoảng tin cậy là đối xứng, do đó nó cũng bằng hai lần độ chính xác và được xác định bằng biểu thức $I = 2\varepsilon = \frac{2S}{\sqrt{n}} T_{\alpha/2}^{(n-1)}$

Ví dụ 4.2:

Để xác định trọng lượng trung bình của các bao bột trong kho, người ta đem cân ngẫu nhiên 15 bao trong kho đó và tìm được $\overline{x}$ = 39,8 kg, s_2 = 0,144. Hãy tìm khoảng tin cậy đối xứng của trọng lượng trung bình của các bao bột trong kho với yêu cầu độ tin cậy của việc ước lượng là 99%. Giả thiết trọng lượng đóng bao của các bao bột là biến ngẫu nhiên phân phối chuẩn.

Giải:

Gọi X là "Trọng lượng bột đóng bao", theo giả thiết X phân phối chuẩn. Vậy trọng lượng đóng bao trung bình chính là giá trị μ. Đây là bài toán ước lượng bằng khoảng tin cậy đối xứng giá trị tham số μ của phân phối $N(\mu, \sigma^2)$ khi chưa biết σ^2 của X. Vậy ta có khoảng tin cậy :

$$(\overline{X} - \frac{S}{\sqrt{n}} T_{\alpha/2}^{(n-1)}; \overline{X} + \frac{S}{\sqrt{n}} T_{\alpha/2}^{(n-1)})$$

Cân ngẫu nhiên 15 bao bột, gọi X_i ($i = \overline{1,15}$) là trọng lượng của bao thứ i ta có mẫu ngẫu nhiên W = (X_1,. . . , X_{15}).

Với độ tin cậy 1 - α = 0,99 thì α/2 = 0,005 tra bảng phân phối Student có

$T_{0,005}^{14}$ = 2,977. Vậy với độ tin cậy 0,99 khoảng tin cậy đối xứng của μ là:

$$(\overline{X} - \frac{S}{\sqrt{15}} 2{,}977; \overline{X} + \frac{S}{\sqrt{15}} 2{,}977)$$

Với mẫu cụ thể ta tính được $\overline{x}$ = 39,8, s^2 = 0,144 nên s = 0,397. Vậy với độ tin cậy 0,99 qua mẫu cụ thể khoảng tin cậy đối xứng của μ là :

$$(39{,}8 - \frac{0{,}379}{\sqrt{15}} 2{,}977; 39{,}8 + \frac{0{,}379}{\sqrt{15}} 2{,}977) = (39{,}5023; 40{,}0977).$$

4.2.2.3. Chưa biết phương sai σ2 của biến ngấu nhiên gốc X trong tổng thể và kích thước mẫu n $\geq$ 30

Trường hợp này kích thước mẫu lớn (n $\geq$ 30**)**, có thể dùng ước lượng S thay cho σ^2 chưa biết và làm như trường hợp **4.2.2.1**.

4.2.3. ƯỚC LƯỢNG PHƯƠNG SAI

4.2.3.1. Đã biết kỳ vọng toán μ

Chọn thống kê G = χ2 = $\dfrac{nS^{*2}}{\sigma^2}$

- Nếu $\alpha_1 = \alpha_2 = \alpha$ khoảng tin cậy có dạng: $\left(\dfrac{nS^{*2}}{\chi^{2(n)}_{\alpha/2}};\ \dfrac{nS^{*2}}{\chi^{2(n)}_{1-\alpha/2}}\right)$

Khoảng tin cậy trên không đối xứng.

- Nếu $\alpha_1 = 0$, $\alpha_2 = \alpha$, ta có khoảng tin cậy bên phải $\left(\dfrac{nS^{*2}}{\chi^{2(n)}_{\alpha}};\ +\infty\right)$

- Nếu $\alpha_2 = 0$, $\alpha_1 = \alpha$, ta có khoảng tin cậy bên trái $\left(0;\ \dfrac{nS^{*2}}{\chi^{2(n)}_{1-\alpha}}\right)$

Ví dụ 4.3:

Mức hao phí cho nguyên liệu cho một đơn vị sản phẩm là biến ngẫu nhiên phân phối chuẩn với trung bình là 20 gam. Để ước lượng mức độ phân tán của mức hao phí này người ta cân thử 25 sản phẩm và thu được kết quả như sau:

Hao phí nguyên liệu (gam)	19,5	20	20,5
Số sản phẩm tương ứng	5	18	2

Với độ tin cậy 1 - α = 0,90 hãy ước lượng σ^2 nếu $\alpha_1 = \alpha_2 = \alpha/2 = 0{,}05$.

Giải:

Gọi X là mức hao phí nguyên liệu cho 1 đơn vị sản phẩm. X phân phối chuẩn với kỳ vọng toán đã biết μ = 20. Đây là bài toán ước lượng phương sai của phân phối $N(\mu,\sigma^2)$ khi đã biết μ. Vậy ta có khoảng tin cậy của σ^2 là :

$$\left(\frac{nS^{*2}}{\chi^{2(n)}_{\alpha/2}};\ \frac{nS^{*2}}{\chi^{2(n)}_{1-\alpha/2}}\right)$$

Tra bảng giá trị của χ^2 ta có

$$\chi^{2(25)}_{\alpha/2} = \chi^{2(25)}_{0,05} = 37{,}65;\quad \chi^{2(25)}_{\alpha/2} = \chi^{2(25)}_{0,95} = 14{,}61$$

Để tìm s*^{2} ta lập bảng $\quad s^{*2} = \dfrac{1}{n}\sum_{i=1}^{n} n_i(x_i-\mu)^2 = \dfrac{1{,}75}{25} = 0{,}07$

Với độ tin cậy 0,90 khoảng tin cậy của σ^2 là

(0,0464; 0,1198)

4.2.3.2. Chưa biết kỳ vọng toán μ

Ta chọn thống kê G = $\chi^2 = \dfrac{(n-1)S^2}{\sigma^2}$

- Nếu $\alpha_1 = \alpha_2 = \alpha/2$, khoảng tin cậy của σ^2 là $\left(\dfrac{(n-1)S^2}{\chi^{2(n-1)}_{\alpha/2}}; \dfrac{(n-1)S^2}{\chi^{2(n-1)}_{1-\alpha/2}}\right)$

x_i	n_i	$(x_i - \mu)$	$(x_i - \mu)^2$	$n_i(x_i - \mu)^2$
19,5	5	0,5	0,25	1,25
20,0	18	0,0	0,00	0,00
20,5	2	0,5	0,25	0,50
	n = 25			$\sum$ = 1,75

- Nếu $\alpha_1 = 0$; $\alpha_2 = \alpha$, ta có các khoảng tin cậy bên phải của σ^2 : $\left(\dfrac{(n-1)S^2}{\chi^{2(n-1)}_{\alpha}}; +\infty\right)$

- Nếu $\alpha_2 = 0$; $\alpha_1 = \alpha$, ta có khoảng tin cậy bên trái của σ^2 : $\left(0; \dfrac{(n-1)S^2}{\chi^{2(n-1)}_{1-\alpha}}\right)$

Ví dụ 4.4:

Với độ tin cậy 0,95 hãy ước lượng phương sai của kích thước các chi tiết cho ở bảng sau:

Kích thước chi tiết (cm)	Số chi tiết tương ứng
54,795 – 54,805	6
54,805 – 54,815	14
54,815 – 54,825	33
54,825 – 54,835	47
54,835 – 54,845	45
54,845 – 54,855	33
54,855 – 54,865	15
54,865 – 54,875	7
	n = 200

Giải:

Đây là bài toán ước lượng phương sai của phân phối N(μ, σ^2) khi chưa biết μ. Vậy khoảng tin cậy của σ^2 có dạng

Nếu $\alpha_1 = \alpha_2 = \alpha/2$, khoảng tin cậy của σ^2 là

$$\left(\frac{(n-1)S^2}{\chi^{2(n-1)}_{\alpha/2}};\frac{(n-1)S^2}{\chi^{2(n-1)}_{1-\alpha/2}}\right)$$

Qua mẫu cụ thể ta tìm được: $S^2 = 0.0002689$; $n = 200$

Tra bảng χ^2: $\chi^{2(199)}_{0,975} \approx 198,98; \chi^{2(199)}_{0,025} \approx 284,8$

Vậy với độ tin cậy 0,95 qua mẫu cụ thể này khoảng tin cậy của σ^2 là:

$$\left(\frac{199.0,0002689}{284,8};\frac{199.0,0002689}{198,98}\right)$$

Hay ($0,000188 < \sigma^2 < 0,000269$).

4.2.4. ƯỚC LƯỢNG TỶ LỆ

Chọn thống kê $U = \dfrac{(f-p)\sqrt{n}}{\sqrt{p(p-1)}}$

- Khoảng tin cậy đối xứng khi $\alpha_1 = \alpha_2 = \dfrac{\alpha}{2}$: $\left(f-\varepsilon, f+\varepsilon\right)$,

$$\text{với } \varepsilon = \frac{\sqrt{f(1-f)}}{\sqrt{n}}u_{\alpha/2}$$

- Khoảng tin cậy bên phải khi $\alpha_1 = 0; \alpha_2 = \alpha$: $\left(f-\dfrac{\sqrt{f(1-f)}}{\sqrt{n}}u_\alpha;+\infty\right)$

-Khoảng tin cậy bên trái khi $\alpha_1 = \alpha; \alpha_2 = 0$: $\left(-\infty;\dfrac{\sqrt{f(1-f)}}{\sqrt{n}}u_\alpha\right)$

Độ dài khoảng tin cậy ngắn nhất trong trường hợp khoảng tin cậy đối xứng:

$$I = 2\varepsilon = \frac{2\sqrt{f(1-f)}}{\sqrt{n}}u_{\alpha/2}$$

Ví dụ 4.5:

Kiểm tra ngẫu nhiên 400 sản phẩm do một máy sản xuất thấy có 20 phế phẩm. Với độ tin cậy 0,95 hãy ước lượng tỷ lệ phế phẩm tối đa của máy đó.

Giải:

Gọi p là tỷ lệ phế phẩm của máy đó. Như vậy p là cơ cấu của tập hợp sản phẩm do máy đó sản xuất theo dấu hiệu "phế phẩm".

Đây là bài toán ước lượng tham số p của qui luật phân phối A(p) bằng khoảng tin cậy bên trái.

Vậy khoảng tin cậy của p có dạng **(6.48)**: $\left(-\infty;\frac{\sqrt{f(1-f)}}{\sqrt{n}}u_\alpha\right)$

Qua mẫu cụ thể ta có f = $\frac{20}{400}=0,05.$ Với 1-α =0,95 -> $u_{0,05}=1,645$. Vậy với độ tin cậy 0,95 qua mẫu cụ thể này khoảng tin cậy của p là:

$$(-\infty;0,05+\frac{\sqrt{0,05.0,95}}{\sqrt{400}}1,645)$$

Hay p < 0.0679, hay tỷ lệ phế phẩm tối đa của máy đó là 6,79%.

CHƯƠNG.5. KIỂM ĐỊNH GIẢ THIẾT THỐNG KÊ

BÀI - 1. KHÁI NIỆM CHUNG

5.1.1. GIẢ THIẾT THỐNG KÊ

Khi nghiên cứu về lĩnh vực nào đó trong thực tế ta thường đưa ra các nhận xét khác nhau về các đối tượng quan tâm. Những nhận xét như vậy thường được coi là các giả thiết, chúng có thể đúng hoặc sai.

Định nghĩa 5.1:

Giả thuyết thống kê là giả thuyết về dạng phân phối xác suất của biến ngẫu nhiên, về các tham số đặc trưng của biến ngẫu nhiên hoặc về tính độc lập của các biến ngẫu nhiên

Giả thiết thống kê đưa ra được ký hiệu là H_0 và được gọi là *giả thuyết gốc*. Khi đưa ra một giả thuyết thống kê, người ta còn nghiên cứu kèm theo nó mệnh đề mâu thuẫn với nó gọi là giả thuyết đối và ký hiệu là H_1 để khi giả thuyết H_0 bị bác bỏ thì thừa nhận giả thuyết H_1. H_0 và H_1 tạo nên cặp giả thuyết thống kê.

Chẳng hạn ta nghiên cứu nhu cầu thị trường về một loại hàng hóa nào đó. Ta có thể đưa ra cặp giả thiết thống kê sau:

- H_0: Nhu cầu trung bình về loại hàng hoá này là $\mu = 1000$ đơn vị/ tháng, lúc đó các giả thuyết đối tương ứng của nó là:

- H_1: $\mu > 1000$; H_1: $\mu < 1000$; H_1: $\mu \neq 1000$.

Vì các giả thuyết thống kê có thể đúng hoặc sai nên cần kiểm định, tức là tìm ra kết luận về tính thừa nhận được hay không của giả thuyết đó. Việc kiểm định này gọi là *kiểm định thống kê* vì nó dựa vào thông tin thực nghiệm của mẫu để kết luận.

Từ biến ngẫu nhiên gốc X trong tổng thể lập mẫu ngẫu nhiên kích thước n

$$W = (X_1, X_2, \ldots, X_n)$$

và chọn thống kê

$$G = f(X_1, X_2, \ldots, X_n, \theta_0)$$

trong đó θ_0 là tham số liên quan đến giả thuyết cần kiểm định. Điều kiện đặt ra đối với thống kê G là nếu H_0 đúng thì quy luật phân phối xác suất của G hoàn toàn xác định. Thống kê G được gọi là *tiêu chuẩn kiểm định*.

Sau khi đã chọn được tiêu chuẩn kiểm định G, do quy luật phân phối xác suất của G đã biết nên với một xác suất khá bé bằng α cho trước có thể tìm được miền W_α tương ứng sao cho với điều kiện giả thiết H_0 đúng xác suất để G nhận giá trị thuộc miền W_α bằng α. Điều kiện này được viết như sau:

$$P(G \in W_\alpha / H_0) = \alpha$$

Giá trị α gọi là *mức ý nghĩa* của kiểm định. W_α được gọi là *miền bác bỏ* giả thuyết H_0 với mức ý nghĩa α.

Thực hiện một phép thử đối với mẫu ngẫu nhiên W = (X_1, X_2, …, X_n) thu được một mẫu cụ thể w = (x_1, x_2, …, x_n) và qua đó tính được một giá trị cụ thể của tiêu chuẩn kiểm định G: $G_{qs} = f(x_1, x_2, \ldots, x_n, \theta_0)$. Giá trị này được gọi giá trị quan sát của tiêu chuẩn kiểm định.

- Nếu giá trị quan sát của tiêu chuẩn kiểm định thuộc miền bác bỏ ($G_{qs} \in W_\alpha$) bác bỏ H_0 thừa nhận H_1.

- Nếu giá trị quan sát của tiêu chuẩn kiểm định không thuộc miền bác bỏ ($G_{qs} \notin W_\alpha$) thì điều đó chưa khẳng định rằng H_0 đúng mà chỉ có nghĩa là chưa khẳng định được H_0 sai. Do đó qua mẫu cụ thể này chưa có cơ sở để bác bỏ H_0 (trên thực tế vẫn thừa nhận H_0).

5.1.2. SAI LẦM LOẠI MỘT VÀ SAI LẦM LOẠI 2

Khi kiểm định giả thiết thống kê, có thể mắc một trong hai loại sai lầm sau:

Sai lầm loại 1: Bác bỏ giả thuyết H_0 trong khi H_0 đúng.

Xác suất mắc phải sai lầm này đúng bằng mức ý nghĩa α. Sai lầm này có thể sinh ra do kích thước mẫu quá nhỏ, do phương pháp lấy mẫu v.v…

Sai lầm loại 2: Thừa nhận H_0 trong khi H_0 sai.

Trong thực tế người ta tiến hành như sau: Sau khi đã ấn định một mức ý nghĩa α và với mẫu kích thước n xác định thì trong vô số các miền bác bỏ W_α tương ứng có thể tìm được, ta chọn ra miền bác bỏ W_α sao cho xác suất mắc sai lầm loại 2 là nhỏ nhất.

BÀI - 2. KIỂM ĐỊNH THAM SỐ

5.2.1. KIỂM ĐỊNH GIẢ THUYẾT VỀ GIÁ TRỊ TRUNG BÌNH

5.2.1.1. Khi biết phương sai

Giả sử biến ngẫu nhiên gốc X trong tổng thể phân phối theo quy luật chuẩn $N(\mu, \sigma^2)$ với phương sai đã biết nhưng chưa biết kỳ vọng toán μ. Nếu có cơ sở để giả thiết rằng giá trị của nó bằng μ_0 ta đưa ra giả thiết thống kê:

$$H_o: \mu = \mu_0$$

Để kiểm định giả thiết trên từ tổng thể lập mẫu kích thước n:

$$W = (X_1, X_2, \ldots, X_n)$$

Vì đã biết phương sai σ^2 của biến ngẫu nhiên gốc X trong tổng thể nên tiêu chuẩn kiểm định được chọn là thống kê

$$G = U = \frac{(\overline{X} - \mu_0)\sqrt{n}}{\sigma}$$

a) $\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu > \mu_0 \end{cases}$

Lúc đó với mức ý nghĩa α cho trước có thể tìm được giá trị tới hạn chuẩn u_α sao cho

$$P(G \in W_\alpha / H_0) = P(U > u_\alpha) = \alpha$$

Ta thu được miền bác bỏ bên phải W_α được xác định bằng biểu thức

$$W_\alpha = \left\{ U = \frac{(\overline{X} - \mu_0)\sqrt{n}}{\sigma}; U > u_\alpha \right\}$$

b) $\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu < \mu_0 \end{cases}$

Lúc đó với mức ý nghĩa α cho trước có thể tìm được giá trị tới hạn chuẩn $u_{1-\alpha}$ sao cho

$$P(G \in W_\alpha / H_0) = P(U < u_{1-\alpha}) = P(U < -u_\alpha) = \alpha$$

Ta thu được miền bác bỏ bên trái W_α được xác định bằng biểu thức

$$W_\alpha = \left\{ U = \frac{(\overline{X} - \mu_0)\sqrt{n}}{\sigma}; U < -u_\alpha \right\}$$

c) $\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{cases}$

Lúc đó với mức ý nghĩa α cho trước có thể tìm được hai giá trị tới hạn chuẩn là $u_{1-\alpha/2}$ và $u_{\alpha/2}$ sao cho

$$\begin{aligned} P(G \in W_\alpha / H_0) &= P(U < u_{1-\alpha/2}) + P(U > u_{\alpha/2}) \\ &= P(U < -u_{\alpha/2}) + P(U > u_{\alpha/2}) \\ &= P(|U| > u_{\alpha/2}) = \alpha \end{aligned}$$

Ta thu được miền bác bỏ hai phía được xác định bằng biểu thức

$$W_\alpha = \left\{ U = \frac{(\overline{X} - \mu_0)\sqrt{n}}{\sigma}; |U| > u_{\alpha/2} \right\}$$

Lập mẫu cụ thể w = (x_1, x_2, …, x_n) và tính giá trị quan sát tiêu chuẩn kiểm định

$$U_{qs} = \frac{(\overline{x} - \mu_0)\sqrt{n}}{\sigma}$$

và so sánh với W_α để kết luận:

- Nếu $U_{qs} \in W_\alpha$ thì bác bỏ H_0, thừa nhận H_1;
- Nếu $U_{qs} \notin W_\alpha$ thì chưa có cơ sở để bác bỏ H_0.

Ví dụ 5.1:

Gạo được đóng gói 20kg một bao trên máy tự động. Trọng lượng các bao gạo tuân theo quy luật chuẩn với độ lệch chuẩn 2kg. Nghi ngờ máy hoạt động không bình thường nên các bao gạo có xu hướng bị đóng thừa. Người ta cân thử 100 sản phẩm và thu được kết quả sau:

Trọng lượng gạo (kg)	19	20	21	22	23
Số bao tương ứng	10	60	20	5	5

Với mức ý nghĩa α = 0,05 hãy kết luận điều nghi ngờ nói trên.

Giải:

Gọi X là trọng lượng bao gạo. Theo giả thiết X là biến ngẫu nhiên tuân theo quy luật chuẩn với σ = 2. Bài toán yêu cầu kiểm định tham số μ của biến ngẫu nhiên phân phối chuẩn khi đã biết phương sai của tổng thể.

- Cặp giả thiết thống kê:

$$\begin{cases} H_0 : \mu = 20 \\ H_1 : \mu > 20 \end{cases}$$

- Tiêu chuẩn kiểm định:

$$U = \frac{(\overline{X} - 20)\sqrt{100}}{2}$$

Trong đó $\overline{X}$ là trung bình mẫu ngẫu nhiên kích thước n = 100.

Với $\alpha = 0,05$ ta có $u_\alpha = u_{0,05} = 1,65$.

- Miền bác bỏ: $W_\alpha = (1,65; +\infty)$

Từ mẫu cụ thể ta có:

$$\bar{x} = \frac{19.10 + 20.60 + 21.20 + 22.5 + 23.5}{10 + 60 + 20 + 5 + 5} = 20,35$$

- Giá trị quan sát của tiêu chuẩn kiểm định:

$$U_{qs} = \frac{(\bar{x} - 20)\sqrt{100}}{2} = \frac{(20,35 - 20)\sqrt{100}}{2} = 1,75$$

Như vậy $U_{qs} \in W_\alpha$. Bác bỏ H_0 thừa nhận H_1, tức là máy có xu hướng đóng thừa trọng lượng với mức ý nghĩa $\alpha = 0,05$.

5.2.1.2. Khi chưa biết phương sai

Tiêu chuẩn kiểm định là thống kê $G = T = \dfrac{(\bar{X} - \mu_0)\sqrt{n}}{S}$

a) $\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu > \mu_0 \end{cases}$

Lúc đó với mức ý nghĩa α cho trước có thể tìm được giá trị tới hạn Student $t_\alpha^{(n-1)}$ sao cho

$$P(G \in W_\alpha / H_0) = P(T > t_\alpha^{(n-1)}) = \alpha$$

Ta thu được miền bác bỏ bên phải W_α được xác định bằng biểu thức

$$W_\alpha = \left\{ T = \frac{(\bar{X} - \mu_0)\sqrt{n}}{S}; T > t_\alpha^{(n-1)} \right\}$$

b) $\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu < \mu_0 \end{cases}$

Lúc đó với mức ý nghĩa α cho trước có thể tìm được giá trị tới hạn Student $t_{1-\alpha}^{(n-1)}$ sao cho

$$P(G \in W_\alpha / H_0) = P(T < t_{1-\alpha}^{(n-1)}) = P(T < -t_\alpha^{(n-1)}) = \alpha$$

Ta thu được miền bác bỏ bên trái W_α được xác định bằng biểu thức

$$W_\alpha = \left\{ T = \frac{(\bar{X} - \mu_0)\sqrt{n}}{S}; T < -t_\alpha^{(n-1)} \right\}$$

c) $\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu \neq \mu_0 \end{cases}$

Lúc đó với mức ý nghĩa α cho trước có thể tìm được hai giá trị tới hạn Student là $t_{\alpha/2}^{(n-1)}$ và $t_{1-\alpha/2}^{(n-1)}$ sao cho

$$\begin{aligned} P(G \in W_\alpha / H_0) &= P(T < t_{1-\alpha/2}^{(n-1)}) + P(T > t_{\alpha/2}^{(n-1)}) \\ &= P(T < -t_{\alpha/2}^{(n-1)}) + P(T > t_{\alpha/2}^{(n-1)}) \\ &= P(|T| > t_{\alpha/2}^{(n-1)}) = \alpha \end{aligned}$$

Ta thu được miền bác bỏ hai phía được xác định bằng biểu thức

$$W_\alpha = \left\{ T = \frac{(\overline{X} - \mu_0)\sqrt{n}}{S}; |T| > t_{\alpha/2}^{(n-1)} \right\}$$

Lập mẫu cụ thể w = (x_1, x_2, …, x_n) tính được $\overline{x}$, s và giá trị quan sát tiêu chuẩn kiểm định

$$T_{qs} = \frac{(\overline{x} - \mu_0)\sqrt{n}}{s}$$

và so sánh với W_α để kết luận:

- Nếu $T_{qs} \in W_\alpha$ thì bác bỏ H_0, thừa nhận H_1;
- Nếu $T_{qs} \notin W_\alpha$ thì chưa có cơ sở để bác bỏ H_0.

5.2.2. KIỂM ĐỊNH GIẢ THUYẾT PHƯƠNG SAI

Giả sử trong tổng thể biến ngẫu nhiên gốc X phân phối $N(\mu, \sigma^2)$ với σ^2 chưa biết song có cơ sở để giả thiết rằng giá trị của nó bằng σ_0^2. Người ta đưa ra giả thuyết thống kê H_0: $\sigma^2 = \sigma_0^2$.

Để kiểm định giả thuyết trên từ tổng thể lập mẫu ngẫu nhiên kích thước n

$$W = (X_1, X_2, \ldots, X_n)$$

Và chọn tiêu chuẩn kiểm định là thống kê

$$G = \chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$$

a) $\begin{cases} H_0 : \sigma^2 = \sigma_0^2 \\ H_1 : \sigma^2 > \sigma_0^2 \end{cases}$

Miền bác bỏ bên phải W_α được xác định bằng biểu thức

$$W_\alpha = \left\{ \chi^2 = \frac{(n-1)S^2}{\sigma_0^2}; \chi^2 > \chi_\alpha^{2(n-1)} \right\}$$

b) $\begin{cases} H_0: \sigma^2 = \sigma_0^2 \\ H_1: \sigma^2 < \sigma_0^2 \end{cases}$

Miền bác bỏ bên trái W_α được xác định bằng biểu thức

$$W_\alpha = \left\{ \chi^2 = \frac{(n-1)S^2}{\sigma_0^2}; \chi^2 < \chi_{1-\alpha}^{2(n-1)} \right\}$$

c) $\begin{cases} H_0: \sigma^2 = \sigma_0^2 \\ H_1: \sigma^2 \neq \sigma_0^2 \end{cases}$

Miền bác bỏ hai phía được xác định bằng biểu thức

$$W_\alpha = \left\{ \chi^2 = \frac{(n-1)S^2}{\sigma_0^2}; \chi_{]}^{2} < \chi_{1-\alpha/2}^{2(n-1)} \qquad \chi^2 > \chi_{\alpha/2}^{2(n-1)} \right\}$$

Với mẫu cụ thể w tính được ta tìm được giá trị cụ thể s^2 và tính được giá trị quan sát của tiêu chuẩn kiểm định

$$\chi_{qs}^2 = \frac{(n-1)s^2}{\sigma_0^2}$$

và so sánh với W_α để kết luận:

- Nếu $\chi_{qs}^2 \in W_\alpha$ thì bác bỏ H_0, thừa nhận H_1;

- Nếu $\chi_{qs}^2 \notin W_\alpha$ thì chưa có cơ sở để bác bỏ H_0.

Ví dụ 5.2:

Để kiểm tra độ chính xác của một chiếc máy người ta đo ngẫu nhiên kích thước của 15 chi tiết do máy đó sản xuất và tính được $s^2 = 14{,}6$. Với mức ý nghĩa $\alpha = 0{,}01$ hãy kết luận máy móc có hoạt động bình thường không, biết rằng kích thước chi tiết là biến ngẫu nhiên phân phối chuẩn có dung sai theo thiết kế là $\sigma^2 = 12$.

Giải:

Gọi X là kích thước chi tiết, theo giả thiết X phân phối chuẩn.

- Cặp giả thuyết thống kê

$\begin{cases} H_0: \sigma^2 = 12 \\ H_1: \sigma^2 > 12 \end{cases}$

- Tiêu chuẩn kiểm định $\quad G = \chi^2 = \dfrac{(n-1)S^2}{\sigma_0^2} = \dfrac{(14).S^2}{12}$

Với $\alpha = 0{,}01$ suy ra $\chi_{0,01}^{(14)} = 29{,}14$

- Miền bác bỏ $W_\alpha = (29,14\ ;\ +\infty)$

- Giá trị quan sát của tiêu chuẩn kiểm định

$$\chi^2_{qs} = \frac{(14).14,6}{12} = 17,033$$

$\chi^2_{qs} \notin W_\alpha$ do đó với mức ý nghĩa $\alpha = 0,01$ chưa có cơ sở để bác bỏ H_0, tức là có thể nói máy móc vẫn hoạt động bình thường.

5.2.3. KIỂM ĐỊNH GIẢ THUYẾT VỀ TỶ LỆ

Giả sử trong tổng thể nghiên cứu biến ngẫu nhiên gốc X phân phối không - một với tham số là tỷ lệ p. Nếu chưa biết p song có cơ sở giả thiết rằng giá trị của nó bằng p_0, ta đưa ra giả thuyết thống kê

$$H_0: p = p_0$$

Từ tổng thể lập mẫu ngẫu nhiên kích thước n

$$W = (X_1, X_2, ..., X_n)$$

Người ta chứng minh được rằng với $n > 5$ và $\dfrac{\left|\sqrt{\dfrac{p}{1-p}} - \sqrt{\dfrac{1-p}{p}}\right|}{\sqrt{n}} < 0,3$

thì thống kê $G = U = \dfrac{(f - p_0)\sqrt{n}}{\sqrt{p_0(1-p_0)}}$

phân phối xấp xỉ N(0,1). Do đó với mức ý nghĩa α và tùy thuộc vào giả thuyết đối H_1, các miền bác bỏ được xác định như sau.

a) $\begin{cases} H_0: p = p_0 \\ H_1: p > p_0 \end{cases}$

$$W_\alpha = \left\{ U = \frac{(f - p_0)\sqrt{n}}{\sqrt{p_0(1-p_0)}};\ U > u_\alpha \right\} \quad \textbf{(5-1)}$$

b) $\begin{cases} H_0: p = p_0 \\ H_1: p < p_0 \end{cases}$

$$W_\alpha = \left\{ U = \frac{(f - p_0)\sqrt{n}}{\sqrt{p_0(1-p_0)}};\ U < -u_\alpha \right\} \quad \textbf{(5-2)}$$

c) $\begin{cases} H_0: p = p_0 \\ H_1: p \neq p_0 \end{cases}$

$$W_\alpha = \left\{ U = \frac{(f - p_0)\sqrt{n}}{\sqrt{p_0(1-p_0)}}; |U| > u_{\alpha/2} \right\} \qquad \textbf{(5-3)}$$

Với mẫu cụ thể w = (x_1, x_2, …, x_n) tìm được giá trị quan sát U_{qs} của tiêu chuẩn kiểm định, so sánh với W_α và kết luận.

Ví dụ 5.3:

Tỷ lệ khách hàng tiêu dùng một loại sản phẩm ở địa phương A là 60%. Sau một chiến dịch quảng cáo người ta muốn đánh giá xem chiến dịch quảng cáo này có thực sự mang lại hiệu quả hay không. Để làm điều đó người ta đã phỏng vấn ngẫu nhiên 400 khách hàng thì thấy có 250 người tiêu dùng loại sản phẩm nói trên. Với mức ý nghĩa 0,05 hãy kết luận về hiệu quả của chiến dịch quảng cáo đó.

Giải:

Gọi p là tỷ lệ khách hàng tiêu dùng loại sản phẩm đó ở địa phương A. Đây là bài toán kiểm định tham số p của phân phối A(p).

- Cặp giả thiết thống kê

$$\begin{cases} H_0 : p = 0{,}6 \\ H_1 : p > 0{,}6 \end{cases}$$

vì n > 5 và $\dfrac{\left|\sqrt{\dfrac{0{,}6}{0{,}4}} - \sqrt{\dfrac{0{,}4}{0{,}6}}\right|}{\sqrt{400}} = 0{,}02 < 0{,}3$ nên

-Tiêu chuẩn kiểm định

$$U = \frac{(f - p_0)\sqrt{n}}{\sqrt{p_0(1-p_0)}}$$

Với α = 0,05 suy ra $u_\alpha = u_{0,05}$ = 1,65

- Miền bác bỏ là

$$W_\alpha = (1{,}65\ ;\ +\infty)$$

Với f = $\dfrac{250}{400} = 0{,}625$ ta có

- Giá trị quan sát của tiêu chuẩn kiểm định

$$U_{qs} = \frac{(0{,}625 - 0{,}6)\sqrt{400}}{\sqrt{0{,}6.0{,}4}} = 1{,}02$$

$U_{qs} \notin W_\alpha$ nên chưa có cơ sở bác bỏ H_0, tức là chưa thể nói chiến dịch quảng cáo có hiệu quả với mức ý nghĩa 5%.

MỤC LỤC

Trang

DANH MỤC CÁC ĐỊNH NGHĨA

Trang

DANH MỤC CÁC VÍ DỤ

Trang

DANH MỤC CÁC HÌNH VẼ

Trang

TÀI LIỆU THAM KHẢO

[1] Nguyễn Cao Văn, Thái Văn Ninh, *Lý thuyết xác suất và thống kê toán,* NXB Thống kê, 2005.

[2] Đào Hữu Hồ, *Xác suất thống kê*, NXB ĐHQGHN, 2008.

[3] Đào Hữu Hồ, *Bài tập xác suất thống kê*, NXB ĐHQGHN, 2008.

[4] John A.Rice, Mathematical statictis and data analysic, Berkely University.

[5] Tống Đình Quỳ, *Hướng dẫn giải bài tập xác suất thống kê*, NXB Bách khoa HN, 2007.

[6] Đặng Hùng Thắng, *Bài tập thống kê*, NXB GD, 2008.

[7] Đặng Hùng Thắng, *Thống kê và ứng dụng*, NXB GD, 2008.

[8] Nguyễn Văn Hộ, *Xác suất thống kê*, NXB GD 2005.

[9] Tô Văn Ban, *Xác suất thông kê*, NXB GD, 2010.

[10] Đặng Hùng Thắng, *Bài tập xác suất*, NXB GD, 2009