HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG

PGS.TS. Lê Bá Long

Bài giảng

LÝ THUYẾT XÁC SUẤT VÀ THỐNG KÊ

(*Dành cho sinh viên hệ đại học chuyên ngành Kinh tế*)

Hà Nội, 2013

LỜI NÓI ĐẦU

Các hiện tượng diễn ra trong tự nhiên, xã hội hoặc có tính chất tất định (có tính quy luật, có thể biết trước kết quả) hoặc có tính chất ngẫu nhiên (không biết trước kết quả). Mặc dù không thể nói trước một hiện tượng ngẫu nhiên xảy ra hay không xảy ra khi thực hiện một phép thử, tuy nhiên nếu tiến hành quan sát khá nhiều lần một hiện tượng ngẫu nhiên trong các phép thử như nhau, ta có thể rút ra được những kết luận khoa học về hiện tượng này. Lý thuyết xác suất nghiên cứu khả năng xuất hiện của các hiện tượng ngẫu nhiên và ứng dụng chúng vào thực tế.

Lý thuyết xác suất cũng là cơ sở để nghiên cứu Thống kê – môn học nghiên cứu các phương pháp thu thập thông tin, chọn mẫu, xử lý thông tin, nhằm rút ra các kết luận hoặc đưa ra quyết định cần thiết. Ngày nay, với sự hỗ trợ tích cực của máy tính điện tử và công nghệ thông tin, lý thuyết xác suất thống kê ngày càng được ứng dụng rộng rãi và hiệu quả trong mọi lĩnh vực khoa học tự nhiên và xã hội. Chính vì vậy lý thuyết xác suất thống kê được giảng dạy cho hầu hết các nhóm ngành ở đại học.

Tập bài giảng lý thuyết xác suất và thống kê toán được biên soạn lại theo chương trình qui định của Học viện Công nghệ Bưu Chính Viễn Thông dành cho hệ đại học chuyên ngành kinh tế với hình thức đào tạo theo tín chỉ. Nội dung của cuốn sách bám sát các giáo trình của các trường đại học khối kinh tế và theo kinh nghiệm giảng dạy nhiều năm của tác giả. Chính vì thế, giáo trình này cũng có thể dùng làm tài liệu học tập, tài liệu tham khảo cho sinh viên của các trường đại học và cao đẳng khối kinh tế.

Nội dung của tập bài giảng có 6 chương tương ứng với 3 tín chỉ:

Chương 1: Biến cố ngẫu nhiên và xác suất.

Chương 2: Biến ngẫu nhiên.

Chương 3: Biến ngẫu nhiên hai chiều.

Chương 4: Cơ sở lý thuyết mẫu.

Chương 5: Ước lượng các tham số của biến ngẫu nhiên.

Chương 6: Kiểm định giả thiết thống kê.

Ba chương đầu thuộc về lý thuyết xác suất, ba chương còn lại là những vấn đề cơ bản của lý thuyết thống kê. Điều kiện tiên quyết của môn học này là hai môn Toán cao cấp 1 và Toán cao cấp 2 trong chương trình toán đại cương khối kinh tế. Mặc dù tác giả rất có ý thức trình bày một cách tương đối đầy đủ và chặt chẽ. Tuy nhiên, vì sự hạn chế của chương trình toán dành cho khối kinh tế nên nhiều kết quả và định lý chỉ được phát biểu, minh họa và không có đủ kiến thức cơ sở để chứng minh chi tiết.

Giáo trình được viết cho đối tượng là sinh viên các trường đại học khối kinh tế, vì vậy tác giả cung cấp nhiều ví dụ minh họa tương ứng với từng phần lý thuyết và có nhiều ví dụ ứng dụng vào bài toán kinh tế. Ngoài ra tác giả cũng có ý thức trình bày thích hợp đối với người tự học. Trước khi nghiên cứu các nội dung chi tiết, người học nên xem phần giới thiệu của mỗi chương, để thấy được mục đích, ý nghĩa, yêu cầu chính của chương đó. Trong mỗi chương, mỗi nội dung, người học có thể tự đọc và hiểu được cặn kẽ thông qua cách diễn đạt và chỉ dẫn rõ ràng. Đặc biệt học viên nên chú ý đến các nhận xét, bình luận, để hiểu sâu sắc hơn hoặc mở rộng tổng quát hơn các kết quả và hướng ứng dụng vào thực tế.

Các ví dụ là để minh hoạ trực tiếp khái niệm, định lý hoặc các thuật toán, vì vậy sẽ giúp người học dễ tiếp thu bài hơn. Sau mỗi chương đều có các câu hỏi luyện tập và các bài tập tự luận. Có khoảng từ 20 đến 30 bài tập cho mỗi chương, tương ứng với 3 -5 câu hỏi cho mỗi tiết lý thuyết. Hệ thống câu hỏi này bao trùm toàn bộ nội dung vừa được học. Có những câu hỏi kiểm tra trực tiếp các kiến thức vừa được học, nhưng cũng có những câu đòi hỏi học viên phải vận dụng một cách tổng hợp và sáng tạo các kiến thức đã học để giải quyết. Vì vậy, việc giải các bài tập này giúp học viên nắm chắc hơn lý thuyết và tự kiểm tra được mức độ tiếp thu lý thuyết của mình. Có đáp án và hướng dẫn giải các bài tập ở cuối cuốn sách. Tuy nhiên tác giả khuyên học viên nên cố gắng tự mình giải các bài tập này và chỉ đối chiếu hoặc tham khảo kết quả khi thực sự cần thiết.

Tuy tác giả đã rất cố gắng, song do thời gian bị hạn hẹp, nên các thiếu sót còn tồn tại trong tập bài giảng là điều khó tránh khỏi. Tác giả rất mong nhận được sự đóng góp ý kiến của bạn bè, đồng nghiệp, các học viên xa gần. Xin chân thành cám ơn.

Tác giả xin bày tỏ lời cám ơn tới PGS.TS Phạm Ngọc Anh, TS. Vũ Gia Tê, Ths. Lê Bá Cầu, TS. Nguyễn Thị Nga đã đọc bản thảo và cho những ý kiến phản biện quý giá.

Cuối cùng, tác giả xin bày tỏ sự cám ơn đối với Ban Giám đốc Học viện Công nghệ Bưu Chính Viễn Thông, bạn bè đồng nghiệp đã khuyến khích, động viên, tạo nhiều điều kiện thuận lợi để chúng tôi hoàn thành tập tài liệu này.

Hà Nội, 2013

TÁC GIẢ

MỤC LỤC

CHƯƠNG 1: BIẾN CỐ NGẪU NHIÊN VÀ XÁC SUẤT

Các hiện tượng trong tự nhiên hay xã hội xảy ra một cách ngẫu nhiên (không biết trước kết quả) hoặc tất định (biết trước kết quả sẽ xảy ra). Chẳng hạn một vật nặng được thả từ trên cao chắc chắn sẽ rơi xuống đất, trong điều kiện bình thường nước sôi ở 100^0C ... Đó là những hiện tượng diễn ra có tính quy luật, tất định. Trái lại khi tung đồng xu ta không biết mặt sấp hay mặt ngửa sẽ xuất hiện. Ta không thể biết trước có bao nhiêu cuộc gọi đến tổng đài, có bao nhiêu khách hàng đến điểm phục vụ trong khoảng thời gian nào đó. Ta không thể xác định trước chỉ số chứng khoán trên thị trường chứng khoán ở một thời điểm khớp lệnh trong tương lai … Đó là những hiện tượng ngẫu nhiên. Tuy nhiên, nếu tiến hành quan sát khá nhiều lần một hiện tượng ngẫu nhiên trong những hoàn cảnh như nhau, thì trong nhiều trường hợp ta có thể rút ra những kết luận có tính quy luật về những hiện tượng này. Lý thuyết xác suất nghiên cứu các qui luật của các hiện tượng ngẫu nhiên. Việc nắm bắt các quy luật này sẽ cho phép dự báo các hiện tượng ngẫu nhiên đó sẽ xảy ra như thế nào. Chính vì vậy các phương pháp của lý thuyết xác suất được ứng dụng rộng rãi trong việc giải quyết các bài toán thuộc nhiều lĩnh vực khác nhau của khoa học tự nhiên, kỹ thuật và kinh tế-xã hội.

Chương này trình bày một cách có hệ thống các khái niệm cơ bản và các kết quả chính về lý thuyết xác suất:

- Các khái niệm phép thử, biến cố.
- Quan hệ giữa các biến cố.
- Các định nghĩa về xác suất: định nghĩa xác suất theo cổ điển, theo thống kê.
- Các tính chất của xác suất: công thức cộng và công thức nhân xác suất, xác suất của biến cố đối.
- Xác suất có điều kiện, công thức nhân trong trường hợp không độc lập. Công thức xác suất đầy đủ và định lý Bayes.

Khi đã nắm vững các kiến thức về đại số tập hợp (một trường hợp cụ thể của đại số Boole) như hợp, giao tập hợp, tập con, phần bù của một tập con … học viên sẽ dễ dàng trong việc tiếp thu, biểu diễn hoặc mô tả các biến cố.

Để tính xác suất các biến cố theo phương pháp cổ điển đòi hỏi phải tính số các trường hợp thuận lợi đối với biến cố và số các trường hợp đồng khả năng có thể. Vì vậy học viên cần nắm vững các phương pháp đếm - giải tích tổ hợp (đã được học ở lớp 12 và trong chương 1 của môn đại số). Tuy nhiên để thuận lợi cho người học chúng tôi sẽ nhắc lại các kết quả chính về phương pháp đếm trong mục 1.2.2.

Một trong những khó khăn của bài toán xác suất là xác định được biến cố và sử dụng đúng các công thức thích hợp. Bằng cách tham khảo các ví dụ và giải nhiều bài tập sẽ rèn luyện tốt kỹ năng này.

1.1 PHÉP THỬ VÀ BIẾN CỐ

1.1.1 Phép thử

Trong thực tế ta thường gặp nhiều thí nghiệm, quan sát mà các kết quả của nó không thể dự báo trước được. Ta gọi chúng là các *phép thử ngẫu nhiên*.

Phép thử ngẫu nhiên thường được ký hiệu bởi chữ $\mathcal{C}$. Tuy không biết kết quả sẽ xảy ra như thế nào, nhưng trong nhiều trường hợp ta có thể liệt kê được hoặc biểu diễn tất cả các kết quả của phép thử $\mathcal{C}$.

Ví dụ 1.1:

- Phép thử tung đồng xu có hai khả năng xảy ra là mặt sấp, ký hiệu S, hoặc mặt ngửa, ký hiệu N. Ta gọi S, N là các biến cố sơ cấp. Tập các biến cố sơ cấp được gọi là không gian mẫu. Vậy không gian mẫu của phép thử là $\Omega = \{S, N\}$.

- Với phép thử gieo xúc xắc 6 mặt, có thể xem các biến cố sơ cấp là số các chấm trên mỗi mặt xuất hiện. Vậy không gian mẫu $\Omega = \{1, 2, 3, 4, 5, 6\}$.

- Phép thử tung đồng thời 2 đồng xu có không gian mẫu là:

$$\Omega = \{(S,S),(S,N),(N,S),(N,N)\}.$$

Chú ý rằng bản chất của các biến cố sơ cấp không có vai trò đặc biệt gì trong lý thuyết xác suất. Chẳng hạn có thể mã hóa và xem không gian mẫu của phép thử tung đồng tiền là $\Omega = \{0, 1\}$, trong đó 0 là biến cố sơ cấp chỉ mặt sấp xuất hiện và 1 để chỉ mặt ngửa xuất hiện.

1.1.2 Biến cố

Với phép thử $\mathcal{C}$ ta có thể xét các biến cố (còn gọi là sự kiện) mà việc xảy ra hay không xảy ra hoàn toàn được xác định bởi kết quả của $\mathcal{C}$. Các biến cố ngẫu nhiên được ký hiệu bằng các chữ in hoa *A*, *B*, *C*, … Mỗi kết quả ω (biến cố sơ cấp) của phép thử $\mathcal{C}$ được gọi là kết quả thuận lợi cho biến cố A nếu A xảy ra khi kết quả của phép thử $\mathcal{C}$ là ω.

Ví dụ 1.2: Nếu gọi A là biến cố "số chấm xuất hiện là chẵn" trong phép thử gieo xúc xắc ở ví dụ 1.1 thì A có các kết quả thuận lợi là các mặt có 2, 4, 6 chấm, vì biến cố A xuất hiện khi kết quả của phép thử là mặt 2 chấm, 4 chấm hoặc 6 chấm. Mặt 1 chấm, 3 chấm, 5 chấm không phải là kết quả thuận lợi đối với A.

Tung hai đồng xu, biến cố xuất hiện một mặt sấp một mặt ngửa (xin âm dương) có các kết quả thuận lợi là (S,N) và (N,S).

Như vậy có thể xem mỗi biến cố A là một tập con của không gian mẫu Ω có các phần tử là các kết quả thuận lợi đối với A.

Cần chú ý rằng mỗi biến cố chỉ có thể xảy ra khi một phép thử được thực hiện, nghĩa là gắn với không gian mẫu nào đó.

Có hai biến cố đặc biệt sau:

• ***Biến cố chắc chắn***: là biến cố luôn luôn xảy ra khi thực hiện phép thử. Không gian mẫu Ω là một biến cố chắc chắn.

• ***Biến cố không thể***: là biến cố nhất định không xảy ra khi thực hiện phép thử. Biến cố không thể được ký hiệu $\varnothing$.

Tung một con xúc xắc, biến cố xuất hiện mặt có số chấm nhỏ hơn hay bằng 6 là biến chắc chắn, biến cố xuất hiện mặt có 7 chấm là biến cố không thể.

1.2 ĐỊNH NGHĨA XÁC SUẤT

Một biến cố ngẫu nhiên xảy ra hay không trong kết quả của một phép thử là điều không thể biết hoặc đoán trước được. Tuy nhiên bằng những cách khác nhau ta có thể định lượng khả năng xuất hiện của biến cố, đó là xác suất xuất hiện của biến cố.

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

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

Trường hợp các kết quả của phép thử xuất hiện đồng khả năng thì xác suất của một biến cố có thể được xác định bởi tỉ số của số trường hợp thuận lợi đối với biến cố và số trường hợp có thể. Với cách tiếp cận này ta có *định nghĩa xác suất theo phương pháp cổ điển*.

Trường hợp các kết quả của phép thử không đồng khả năng xuất hiện nhưng có thể thực hiện phép thử lặp lại nhiều lần độc lập, khi đó tần suất xác định khả năng xuất hiện của biến cố. Vì vậy ta có thể tính xác suất của biến cố thông qua tần suất xuất hiện của biến cố đó. Với cách tiếp cận này ta có *định nghĩa xác suất theo thống kê*.

1.2.1 Định nghĩa cổ điển về xác suất

1.2.1.1 Định nghĩa và ví dụ

Giả sử phép thử C thoả mãn hai điều kiện sau:

(i) Không gian mẫu có một số hữu hạn phần tử.

(ii) Các kết quả xảy ra đồng khả năng.

Khi đó xác suất của biến cố A được xác định và ký hiệu

$$P(A)=\frac{\text{sè trường hîp thuËn lîi đèi víi } A}{\text{sè trường hîp cã thÓ}} \tag{1.1}$$

Nếu xem biến cố A như là tập con của không gian mẫu Ω thì

$$P(A)=\frac{\text{sè phÇn tö cña } A}{\text{sè phÇn tö cña } \Omega}=\frac{|A|}{|\Omega|} \tag{1.2}$$

Ví dụ 1.3: Biến cố A xuất hiện mặt chẵn trong phép thử gieo con xúc xắc ở ví dụ 1.1 có 3 trường hợp thuận lợi ($|A|=3$) và 6 trường hợp có thể ($|\Omega|=6$). Vậy $P(A)=\frac{3}{6}=\frac{1}{2}$.

Biến cố xuất hiện một mặt sấp và một mặt ngửa khi gieo đồng thời hai đồng xu có 2 kết quả thuận lợi và 4 kết quả đồng khả năng có thể, vậy có xác suất xuất hiện của biến cố đó là $\frac{1}{2}$.

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

1.2.1.2 Các qui tắc đếm

A. Qui tắc cộng

Nếu có m_1 cách chọn loại đối tượng x_1, m_2 cách chọn loại đối tượng x_2, ... , m_n cách chọn loại đối tượng x_n. Các cách chọn đối tượng x_i không trùng với cách chọn x_j nếu $i \neq j$ thì có $m_1 + m_2 + \cdots + m_n$ cách chọn một trong các đối tượng đã cho.

Chẳng hạn để biết số SV có mặt của một lớp đông ta có thể lấy tổng số SV có mặt của các tổ do tổ trưởng cung cấp.

B. Qui tắc nhân

Giả sử công việc H gồm nhiều công đoạn liên tiếp $H_1, H_2, ..., H_k$.

Có n_1 cách thực hiện công đoạn H_1, ứng với mỗi công đoạn H_1 có n_2 cách thực hiện công đoạn H_2 … Vậy có tất cả $n_1 \cdot n_2 \cdot \cdots \cdot n_k$ cách thực hiện công việc H .

Ví dụ 1.4: Một nhân viên có 4 chiếc áo sơ mi và 3 quần dài đồng phục, thì anh ta có $4.3 = 12$ cách chọn áo sơ mi và quần đồng phục.

Ví dụ 1.5: Tung một con xúc xắc (6 mặt) hai lần. Tìm xác suất để trong đó có 1 lần ra 6 chấm.

Giải: Theo quy tắc nhân ta có số các trường hợp có thể khi tung con xúc xắc 2 lần là 6.6 = 36.

Gọi A là biến cố “ trong 2 lần tung con xúc xắc có 1 lần được mặt 6”. Nếu lần thứ nhất ra mặt 6 thì lần thứ hai chỉ có thể ra các mặt từ 1 đến 5, do đó có 5 trường hợp. Tương tự cũng có 5 trường hợp chỉ xuất hiện mặt 6 ở lần tung thứ hai. Áp dụng quy tắc cộng ta suy ra biến cố “chỉ có một lần ra mặt 6 khi 2 tung xúc xắc” có 10 trường hợp thuận lợi. Vậy xác suất cần tìm là $\frac{10}{36}$.

Ví dụ 1.6:

a. Có bao nhiêu số có 4 chữ số.

b. Có bao nhiêu số có 4 chữ số khác nhau.

c. Có bao nhiêu số có 4 chữ số khác nhau và chữ số cuối là 0.

Giải: **a.** Có 9 cách chọn chữ số đầu tiên (vì chữ số đầu tiên khác 0) và các chữ số còn lại có 10 cách chọn cho từng chữ số. Vậy có 9.10.10.10=9000 số cần tìm.

b. Có 9 cách chọn chữ số đầu tiên (vì chữ số đầu tiên khác 0), 9 cách chọn chữ số thứ hai, 8 cách chọn chữ số thứ ba và 7 cách chọn chữ số thứ tư. Vậy có 9.9.8.7=4536 số cần tìm.

c. Vì chữ số thứ tư là số 0 và các chữ số này khác nhau do đó có 9 cách chọn chữ số đầu tiên, 8 cách chọn chữ số thứ hai, 7 cách chọn chữ số thứ ba. Vậy có 9.8.7=504 số cần tìm.

C. Hoán vị

Mỗi cách đổi chỗ của n phần tử hoặc mỗi cách sắp xếp n phần tử vào n vị trí trong một hàng được gọi là phép hoán vị n phần tử. Sử dụng quy tắc nhân ta có thể tính được:

Có $n!$ hoán vị n phần tử. Quy ước $0! = 1$.

Ví dụ 1.7:

a. Có bao nhiêu cách bố trí 5 nam SV và 4 nữ SV theo một hàng.

b. Có bao nhiêu cách bố trí 5 nam SV và 4 nữ SV theo một hàng, sao cho các nữ SV ở vị trí số chẵn.

Giải: **a.** Số cách bố trí 9 SV (gồm 5 nam SV và 4 nữ SV) theo một hàng là 9!= 362880.

b. Có 5! cách bố trí nam SV, ứng với mỗi cách bố trí nam SV có 4! cách bố trí nữ SV vào vị trí chẵn tương ứng. Vậy có 5!4!=2880 cách bố trí theo yêu cầu.

Ví dụ 1.8: (***Hoán vị vòng tròn***) Có n người ($n \geq 3$), trong đó có hai người là anh em.

a. Có bao nhiêu cách sắp xếp n người ngồi xung quanh một bàn tròn.

b. Có bao nhiêu cách sắp xếp n người ngồi xung quanh một bàn tròn, trong đó có hai người là anh em ngồi cạnh nhau.

c. Có bao nhiêu cách sắp xếp n người ngồi xung quanh một bàn tròn, trong đó có hai người là anh em không ngồi cạnh nhau.

Giải: **a.** Có 1 người ngồi ở vị trí bất kỳ, vì vậy $n-1$ người còn lại có $(n-1)!$ cách chọn vị trí ngồi. Vậy có $(n-1)!$ cách sắp xếp n người ngồi xung quanh một bàn tròn.

b. Người anh ngồi ở một vị trí tùy ý, người em ngồi vào 1 trong 2 chỗ cạnh người anh (có 2 cách) và $n-2$ người còn lại còn lại ngồi tùy ý vào $n-2$ chỗ còn lại (có $(n-2)!$ cách). Vậy số các cách sắp xếp theo yêu cầu là $2.(n-2)!$.

c. Sử dụng kết quả phần a. và b. ta suy ra số cách sắp xếp n người ngồi xung quanh một bàn tròn, trong đó có hai người là anh em không ngồi cạnh nhau là

$$(n-1)!-2.(n-2)!=(n-2)!\left[(n-1)-2\right].$$

Ví dụ 1.9: Xếp ngẫu nhiên 6 cuốn sách toán và 4 sách lý vào 1 giá sách. Tính xác suất 3 cuốn sách toán đứng cạnh nhau.

Giải: Số trường hợp có thể là số cách sắp xếp 10 cuốn sách vào giá sách đó là 10!.

Ta xem 3 cuốn sách toán đứng cạnh nhau như là một cuốn sách lớn. Như vậy ta cần sắp xếp 8 cuốn sách vào giá sách (có 8! cách), ngoài ra 3 cuốn sách toán đứng cạnh nhau có 3! cách sắp xếp. Do đó số các trường hợp thuận lợi là 8!3!. Vậy xác suất 3 cuốn sách toán đứng cạnh nhau là $P=\dfrac{8!3!}{10!}=\dfrac{1}{15}$.

D. Chỉnh hợp

Chọn lần lượt k ($1 \leq k \leq n$) phần tử không hoàn lại trong tập n phần tử ta được một chỉnh hợp chập k của n phần tử. Sử dụng quy tắc nhân ta có thể tính được số các chỉnh hợp chập k của n phần tử là

$$A_n^k = n(n-1)...(n-k+1)=\frac{n!}{(n-k)!} \qquad \textbf{(1.3)}$$

Ví dụ 1.10: Có $A_{10}^4 = 10.9.8.7 = 5040$ cách bố trí 10 người ngồi vào 4 chỗ.

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

Giải: Gọi A là biến cố "quay ngẫu nhiên một lần được đúng số cần gọi". Số các trường hợp có thể là số các cặp hai chữ số khác nhau từ 10 chữ số từ 0 đến 9. Nó bằng số các chỉnh hợp chập 2 của 10 phần tử. Vậy số các trường hợp có thể là $A_{10}^2 = 10 \cdot 9 = 90$.

Số các trường hợp thuận lợi của A là 1. Vậy $P(A) = \frac{1}{90}$.

Cũng có thể tính trực tiếp số trường hợp có thể của biến cố A như sau: Có 10 khả năng cho con số ở hàng chục và với mỗi con số hàng chục có 9 khả năng cho con số ở hàng đơn vị khác với hàng chục. Áp dụng quy tắc nhân ta được số các trường hợp có thể là $10 \cdot 9 = 90$.

E. Tổ hợp

Chọn đồng thời k phần tử của tập n phần tử ta được một tổ hợp chập k của n phần tử. Cũng có thể xem một tổ hợp chập k của n phần tử là một tập con k phần tử của tập n phần tử.

Hai chỉnh hợp chập k của n phần tử là khác nhau nếu:

- có ít nhất 1 phần tử của chỉnh hợp này không có trong chỉnh hợp kia.
- các phần tử đều như nhau nhưng thứ tự khác nhau.

Do đó với mỗi tổ hợp chập k có $k!$ chỉnh hợp tương ứng. Mặt khác hai chỉnh hợp khác nhau ứng với hai tổ hợp khác nhau là khác nhau.

Vậy số các tổ hợp chập k của n phần tử là C_n^k thỏa mãn:

$$k!C_n^k = A_n^k \Rightarrow C_n^k = \frac{A_n^k}{k!} = \frac{n!}{k!(n-k)!} \tag{1.4}$$

Một vài trường hợp cụ thể

$$C_n^0 = 1;\ C_n^1 = n;\ C_n^2 = \frac{n(n-1)}{2};\ C_n^3 = \frac{n(n-1)(n-2)}{6};\ C_n^k = C_n^{n-k}. \tag{1.5}$$

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

a. Hai người trúng tuyển là nam

b. Hai người trúng tuyển là nữ

c. Có ít nhất 1nữ trúng tuyển.

Giải: Số trường hợp có thể là số tổ hợp chập 2 của 6 phần tử, vậy $|\Omega| = C_6^2 = \frac{6 \cdot 5}{2} = 15$.

a. Chỉ có 1 trường hợp cả 2 nam đều trúng tuyển do đó xác suất tương ứng là $P = \frac{1}{15}$.

b. Có $C_4^2 = \frac{4\cdot 3}{2} = 6$ cách chọn 2 trong 4 nữ, vậy xác suất tương ứng $P = \frac{6}{15}$.

c. Trong 15 trường hợp có thể chỉ có 1 trường hợp cả 2 nam được chọn, vậy có 14 trường hợp ít nhất 1 nữ được chọn. Do đo xác suất tương ứng $P = \frac{14}{15}$.

Ta cũng có thể tính số trường hợp thuận lợi của biến cố "có ít nhất 1 nữ được chọn" như sau.

Vì chỉ chọn 2 ứng viên nên biến cố có ít nhất 1 nữ trúng tuyển được chia thành 2 loại:

- Có 2 nữ được chọn: Có 6 cách
- Có 1 nữ và 1 nam được chọn: Có 4 . 2 cách chọn

Sử dụng quy tắc cộng ta được 14 trường hợp ít nhất 1 nữ được chọn.

Ví dụ 1.13: Một hộp có 8 bi màu đỏ, 3 bi trắng và 9 bi màu xanh. Lấy ngẫu nhiên 3 bi từ hộp. Tính xác suất trong các trường hợp sau:

a. 3 bi lấy được cùng màu đỏ

b. 2 đỏ và 1 trắng

c. Ít nhất 1 trắng

d. Mỗi màu 1 bi

e. Nếu lấy lần lượt không hoàn lại 3 bi, tính xác suất lấy được mỗi màu 1 bi.

Giải: **a.** $P = \frac{C_8^3}{C_{20}^3} = \frac{14}{285} = 0{,}0491$

b. $P = \frac{C_8^2 C_3^1}{C_{20}^3} = \frac{7}{95} = 0{,}0737$

c. $P = \frac{C_3^1 C_{17}^2 + C_3^2 C_{17}^1 + C_3^3}{C_{20}^3} = \frac{23}{57}$ hoặc $P = 1 - \frac{C_{17}^3}{C_{20}^3} = 1 - \frac{34}{57} = \frac{23}{57} = 0{,}4035$

d. $P = \frac{C_8^1 C_3^1 C_9^1}{C_{20}^3} = \frac{18}{95} = 0{,}1895$.

e. $P = \frac{8.3.9}{20.19.18} = \frac{3}{95} = 0{,}0316$.

Nhận xét 1.1:

Hoán vị, chỉnh hợp, tổ hợp có thể liên hệ với nhau như sau:

- Có thể xem mỗi hoán vị *n* phần tử là một cách sắp xếp *n* phần tử này thành một hàng.
- Mỗi chỉnh hợp chập *k* của *n* phần tử là một cách sắp xếp *k* phần tử từ *n* phần tử này thành một hàng.

- Khi sắp xếp các phần tử thành một hàng ta ngầm hiểu từ trái sang phải, vì vậy trường hợp hoán vị vòng quanh cần chọn một phần tử làm điểm xuất phát do đó có $(n-1)!$ cách hoàn vị vòng quanh của *n* phần tử.
- Có thể xem mỗi tổ hợp chập k của n vật là một cách sắp xếp *n* vật thành một hàng, trong đó có k vật loại 1 giống nhau và $n-k$ vật loại 2 còn lại cũng giống nhau.

  Có $n!$ cách sắp xếp *n* vật thành một hàng.

Vì các vật cùng loại giống nhau không phân biệt được, do đó nếu số cách sắp xếp các vật thỏa mãn yêu cầu trên là *N* thì ứng với mỗi một cách sắp xếp trong *N* cách ở trên có $k!$ hoán vị vật loại 1, $(n-k)!$ hoán vị vật loại 2 được đếm trong tổng số $n!$ cách.

$$\text{Vậy } k!(n-k)!N = n! \Rightarrow N = \frac{n!}{k!(n-k)!}.$$

Ta có thể mở rộng kết quả này như sau.

Công thức tổ hợp mở rộng

Số cách sắp xếp $n = n_1 + n_2 + \cdots + n_k$ vật theo một hàng: trong đó có n_1 vật loại 1 giống nhau, n_2 vật loại 2 giống nhau, ... , n_k vật loại k giống nhau là

$$\frac{n!}{n_1!n_2!...n_k!} \tag{1.6}$$

Công thức này có thể giải thích như sau:

Có $n!$ cách sắp xếp $n = n_1 + n_2 + \cdots + n_k$ vật khác nhau theo một hàng.

Vì các vật cùng loại giống nhau không phân biệt được, do đó nếu số cách sắp xếp các vật thỏa mãn yêu cầu trên là *N* thì ứng với mỗi một cách sắp xếp trong *N* cách ở trên có $n_1!$ hoán vị vật loại 1, $n_2!$ hoán vị vật loại 2, ..., $n_k!$ hoán vị vật loại k được đếm trong tổng số $n!$ cách. Vì vậy $n_1!n_2!...n_k!N = n! \Rightarrow N = \dfrac{n!}{n_1!n_2!...n_k!}$

Ví dụ 1.14: Cần sắp xếp 4 cuốn sách toán, 6 sách lý và 2 sách hóa khác nhau trên cùng một giá sách. Có bao nhiêu cách sắp xếp trong mỗi trường hợp sau:

a. Các cuốn sách cùng môn học phải đứng cạnh nhau.

b. Chỉ cần các sách toán đứng cạnh nhau.

c. Nếu các cuốn sách trong mỗi môn học giống nhau thì có bao nhiêu cách sắp xếp.

Giải: **a.** Có 4! cách sắp xếp các cuốn sách toán, 6! cách sắp xếp các cuốn sách lý, 2! cách sắp xếp các cuốn sách hóa và 3! cách sắp xếp 3 nhóm toán, lý, hóa.

Vậy số cách sắp xếp theo yêu cầu là 4!6!2!3!=207.360.

b. Ta ghép 4 sách toán thành 1 cuốn sách to. Như vậy có 9 cuốn sách cần sắp xếp, do đó có 9! cách sắp xếp. Trong mỗi trường hợp này các cuốn sách toán luôn đứng bên nhau, nhưng có 4! cách sắp xếp 4 cuốn sách toán.

Vậy số cách sắp xếp theo yêu cầu là $9!4!=8.709.120$.

c. Vì các cuốn sách cùng loại không phân biệt do đó có thể áp dụng công thức (1.6) và số cách sắp xếp là $\dfrac{12!}{4!6!2!}=13.860$.

1.2.2 Định nghĩa thống kê về xác suất

Định nghĩa xác suất theo cổ điển trực quan, dễ hiểu. Tuy nhiên khi phép thử có không gian mẫu vô hạn hoặc các kết quả không đồng khả năng thì cách tính xác suất cổ điển không áp dụng được. Trong trường hợp này người ta sử dụng phương pháp thông kê như sau.

Giả sử phép thử $\mathcal{C}$ *có thể được thực hiện lặp lại nhiều lần độc lập trong những điều kiện giống hệt nhau. Nếu trong* n *lần thực hiện phép thử* $\mathcal{C}$ *biến cố* A *xuất hiện* $k_n(A)$ *lần (gọi là* ***tần số*** *xuất hiện) thì tỉ số:*

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

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

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

$$P(A)=\lim_{n\to\infty} f_n(A) \tag{1.7}$$

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

Ví dụ 1.15: Một công ty bảo hiểm muốn xác định xác suất để một thanh niên 25 tuổi sẽ bị chết trong năm tới, người ta theo dõi 100.000 thanh niên và thấy rằng có 798 người bị chết trong vòng 1 năm sau đó. Vậy xác suất cần tìm xấp xỉ bằng $\dfrac{798}{100.000}\approx 0,008$.

Ví dụ 1.16: Thống kê cho thấy tần suất sinh con trai xấp xỉ 0,513. Vậy xác suất để bé trai ra đời lớn hơn bé gái.

Nhận xét 1.2: Định nghĩa xác suất theo thống kê khắc phục được hạn chế của định nghĩa cổ điển, phương pháp này hoàn toàn dựa trên các thí nghiệm hoặc quan sát thực tế để tìm xác suất của biến cố. Tuy nhiên định nghĩa thống kê về xác suất cũng chỉ áp dụng cho các phép thử mà có thể lặp lại được nhiều lần một cách độc lập trong những điều kiện giống hệt nhau. Ngoài ra để xác định một cách tương đối chính xác giá trị của xác suất thì cần tiến hành một số lần n đủ lớn các phép thử, mà việc này đôi khi không thể làm được vì hạn chế về thời gian và kinh phí.

Ngày nay với sự trợ giúp của công nghệ thông tin, người ta có thể mô phỏng các phép thử ngẫu nhiên mà không cần thực hiện các phép thử trong thực tế. Điều này cho phép tính xác suất theo phương pháp thống kê thuận tiện hơn.

1.3 QUAN HỆ CỦA CÁC BIẾN CỐ

Một cách tương ứng với các phép toán của tập hợp, trong lý thuyết xác suất người ta xét các quan hệ sau đây của các biến cố ***trong cùng một phép thử***.

1.3.1 Quan hệ biến cố đối

Với mỗi biến cố A, luôn luôn có biến cố gọi là biến cố đối của A, ký hiệu $\overline{A}$ và được xác định như sau: Biến cố A xảy ra khi và chỉ khi biến cố đối $\overline{A}$ không xảy ra.

Ví dụ 1.17: Bắn một phát đạn vào bia. Gọi A là biến cố "bắn trúng bia".

Biến cố đối của A là $\overline{A}$ "bắn trượt bia".

1.3.2 Tổng của các biến cố

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

Biến cố tổng $A \cup B$ xảy ra khi và chỉ khi có ít nhất A hoặc B xảy ra.

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

Biến cố tổng xảy ra khi ***có ít nhất một trong các biến cố*** A_i ***xảy ra***, với $i = 1, ..., n$.

Ví dụ 1.18: Một mạng điện gồm hai bóng đèn mắc nối tiếp. Gọi A_1 là biến cố "bóng đèn thứ nhất bị cháy", A_2 là biến cố "bóng đèn thứ hai bị cháy". Gọi A là biến cố "mạng mất điện". Ta thấy rằng mạng bị mất điện khi ít nhất một trong hai bóng bị cháy.

Vậy $A = A_1 \cup A_2$.

1.3.3 Tích của các biến cố

Tích của hai biến cố A, B là biến cố được ký hiệu $A \cap B$.

Biến cố tích $A \cap B$ xảy ra khi cả hai biến cố A, B đồng thời cùng xảy ra.

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

Biến cố tích xảy ra khi tất cả các biến cố A_i ***đồng thời cùng xảy ra***, với mọi $i = 1, ..., n$.

Ví dụ 1.19: Một mạng điện gồm hai bóng đèn mắc song song. Gọi A_1 là biến cố "bóng đèn thứ nhất bị cháy", A_2 là biến cố "bóng đèn thứ hai bị cháy". Gọi A là biến cố "mạng mất điện".

Ta thấy rằng mạng mắc song song bị mất điện khi cả hai bóng bị cháy. Vậy $A = A_1 \cap A_2$.

1.3.4 Biến cố xung khắc

Hai biến cố A, B gọi là xung khắc nếu hai biến cố này không thể đồng thời cùng xảy ra. Nói cách khác biến cố tích $A \cap B$ là biến cố không thể, nghĩa là $A \cap B = \varnothing$.

Đôi khi người ta còn ký hiệu tổng của hai biến cố xung khắc A và B là $A + B$.

Ví dụ 1.20: Một bình có 3 loại cầu: cầu mầu trắng, mầu đỏ và mầu xanh. Lấy ngẫu nhiên 1 cầu từ bình. Gọi A_t, $A_đ$, A_x lần lượt là biến cố quả cầu rút được là cầu trắng, đỏ, xanh. Các biến cố này xung khắc từng đôi một, vì mỗi quả cầu chỉ có 1 mầu.

1.3.5 Hệ đầy đủ các biến cố

Dãy các biến cố $\{A_1, A_2, ..., A_n\}$ được gọi là một hệ đầy đủ các biến cố nếu thỏa mãn hai điều kiện sau:

(i) Xung khắc từng đôi một, nghĩa là $A_i \cap A_j = \varnothing$ với mọi $i \neq j$; $i = 1, ..., n$; $j = 1, ..., n$

(ii) Tổng của chúng là biến cố chắc chắn, nghĩa là $A_1 \cup A_2 \cup ... \cup A_n = \Omega$.

Đặc biệt với mọi biến cố A, hệ hai biến cố $\{A, \overline{A}\}$ là hệ đầy đủ.

Ví dụ 1.21: Một nhà máy có ba phân xưởng sản xuất ra cùng một loại sản phẩm. Giả sử rằng mỗi sản phẩm của nhà máy chỉ do một trong ba phân xưởng này sản xuất. Chọn ngẫu nhiên một sản phẩm, gọi A_1, A_2, A_3 lần lượt là biến cố sản phẩm được chọn do phân xưởng thứ nhất, thứ hai, thứ ba sản xuất. Khi đó hệ ba biến cố $\{A_1, A_2, A_3\}$ là hệ đầy đủ.

Hệ ba biến cố $\{A_t, A_đ, A_x\}$ trong ví dụ 1.20 cũng là đầy đủ.

1.3.6 Tính độc lập của các biến cố

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

Trường hợp tổng quát: hệ các biến cố $A_1, A_2, ..., A_n$ được gọi là độc lập nếu việc xảy ra hay không xảy ra của một nhóm bất kỳ k biến cố, trong đó $1 \le k \le n$, không làm ảnh hưởng tới việc xảy ra hay không xảy ra của một nhóm nào đó các biến cố còn lại.

Ví dụ 1.22: Ba xạ thủ A, B, C mỗi người bắn một viên đạn vào mục tiêu. Gọi A, B, C lần lượt là biến cố A, B, C bắn trúng mục tiêu.

a. Hãy mô tả các biến cố: $A \cap B \cap C$, $\overline{A} \cap \overline{B} \cap \overline{C}$, $A \cup B \cup C$.

b. Biểu diễn các biến cố sau theo A, B, C:

- D: Có ít nhất 2 xạ thủ bắn trúng.
- E: Có nhiều nhất 1 xạ thủ bắn trúng.
- F: Chỉ có xạ thủ C bắn trúng.
- G: Chỉ có 1 xạ thủ bắn trúng.

c. Các biến cố A, B, C có xung khắc không, có độc lập không ?

Giải:

a. $A \cap B \cap C$: là biến cố cả 3 đều bắn trúng.

$\overline{A} \cap \overline{B} \cap \overline{C}$: là biến cố cả 3 đều bắn trượt.

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

b. Biến cố có ít nhất hai xạ thủ bắn trúng:

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

Vì chỉ có 3 xạ thủ bắn vào bia nên biến cố "có nhiều nhất một xạ thủ bắn trúng" cũng là biến cố "có ít nhất hai xạ thủ bắn trượt", vậy:

$$E=(\overline{A}\cap\overline{B})\cup(\overline{B}\cap\overline{C})\cup(\overline{C}\cap\overline{A}).$$

Biến cố chỉ có C bắn trúng: $F=\overline{A}\cap\overline{B}\cap C$.

Biến cố chỉ có một xạ thủ bắn trúng: $G=(A\cap\overline{B}\cap\overline{C})\cup(\overline{A}\cap B\cap\overline{C})\cup(\overline{A}\cap\overline{B}\cap C)$.

c. Ba biến cố A,B,C độc lập nhưng không xung khắc.

Nhận xét 1.3:

- Từ ví dụ trên cho thấy tính chất xung khắc hoặc độc lập của các biến cố được suy từ ý nghĩa của phép thử.
- Nếu A,B độc lập thì các cặp biến cố: $A,\overline{B}$; $\overline{A},B$; $\overline{A},\overline{B}$ cũng độc lập.
- Một số tài liệu ký hiệu tổng, tích của hai biến cố A,B là $A+B$ và AB. Mỗi cách ký hiệu có những thuận lợi riêng. Nhưng ký hiệu theo cách này rất khó biểu diễn các tính chất dạng đại số Boole của các biến cố, chẳng hạn tính chất phân phối của tổng đối với tích và tích đối với tổng của các biến cố được xét trong chú ý sau.
- Chú ý rằng các biến cố với phép toán tổng, tích và lấy biến cố đối tạo thành đại số Boole, do đó các phép toán được định nghĩa ở trên có các tính chất như các phép toán hợp, giao, lấy phần bù đối với các tập con của không gian mẫu. Chẳng hạn phép toán tổng, phép toán tích các biến cố có tính giao hoán, kết hợp, tổng phân bố đối với tích, tích phân bố đối với tổng, thỏa mãn luật De Morgan …

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

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

1.4 CÁC ĐỊNH LÝ VÀ TÍNH CHẤT XÁC SUẤT

1.4.1 Xác suất chắc chắn và xác suất không thể

Các định nghĩa trên của xác suất thỏa mãn các tính chất sau:

1. Với mọi biến cố A:

$$0\le P(A)\le 1. \qquad \textbf{(1.8)}$$

2. Xác suất của biến cố không thể bằng 0, xác suất của biến cố chắc chắn bằng 1.

$$P(\varnothing)=0,\ P(\Omega)=1 \qquad \textbf{(1.9)}$$

1.4.2 Qui tắc cộng xác suất

1.4.2.1 Trường hợp xung khắc

Nếu $A,\ B$ là hai biến cố xung khắc thì

$$P(A\cup B)=P(A)+P(B). \tag{1.10}$$

Tổng quát hơn, nếu $\{A_1,A_2,...,A_n\}$ là dãy các biến cố xung khắc từng đôi một thì

$$P\left(\bigcup_{i=1}^{n}A_i\right)=\sum_{i=1}^{n}P(A_i). \tag{1.11}$$

Từ công thức (1.9) và (1.11) ta có hệ quả:

Nếu $\{A_1,A_2,...,A_n\}$ là một hệ đầy đủ thì

$$\sum_{i=1}^{n}P(A_i)=1 \tag{1.12}$$

1.4.2.2 Trường hợp không xung khắc

- Nếu A,B là hai biến cố bất kỳ thì

$$P(A\cup B)=P(A)+P(B)-P(A\cap B) \tag{1.13}$$

- Nếu A,B,C là ba biến cố bất kỳ thì

$$P(A\cup B\cup C)=P(A)+P(B)+P(C)-P(A\cap B)-P(B\cap C)-P(C\cap A)+P(A\cap B\cap C) \tag{1.14}$$

- Nếu $\{A_1,A_2,...,A_n\}$ là dãy n biến cố bất kỳ

$$P\left(\bigcup_{i=1}^{n}A_i\right)=\sum_{i=1}^{n}P(A_i)-\sum_{1\le i<j\le n}P(A_i\cap A_j)+\sum_{1\le i<j<k\le n}P(A_i\cap A_j\cap A_k)\cdots+(-1)^{n-1}P(A_1\cap A_2\cap...\cap A_n) \tag{1.15}$$

Ví dụ 1.23: Một lô hàng có 25% sản phẩm loại I, 55% sản phẩm loại II và 20% sản phẩm loại III. Sản phẩm được cho là đạt chất lượng nếu thuộc loại I hoặc loại II. Chọn ngẫu nhiên 1 sản phẩm tìm xác suất để sản phẩm này đạt tiêu chuẩn chất lượng.

Giải: Gọi A_1,A_2,A_3 lần lượt là biến cố sản phẩm được chọn thuộc loại I, II, III. Ba biến cố này xung khắc từng đôi một.

$$P(A_1)=0{,}25,\ P(A_2)=0{,}55,\ P(A_3)=0{,}20.$$

Gọi A là biến cố sản phẩm được chọn đạt tiêu chuẩn chất lượng, ta có $A=A_1\cup A_2$.

Vậy xác suất tìm được sản phẩm đạt tiêu chuẩn chất lượng là:

$$P(A)=P(A_1)+P(A_2)=0{,}25+0{,}55=0{,}8.$$

Ví dụ 1.24: *Sơ đồ cây*

Nhiều phép thử có tính chất nối tiếp lập thành dãy, chẳng hạn phép thử tung liên tiếp đồng xu ba lần, quan sát chỉ số chứng khoán trong năm ngày liên tiếp, hoặc tám ký số liên tiếp nhận được của một bộ nhận thông tin ... Trong trường hợp này ta có thể biểu diễn không gian mẫu và các biến cố tương ứng dưới dạng sơ đồ cây.

Từ sơ đồ cây (hình 1.1) ta có

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

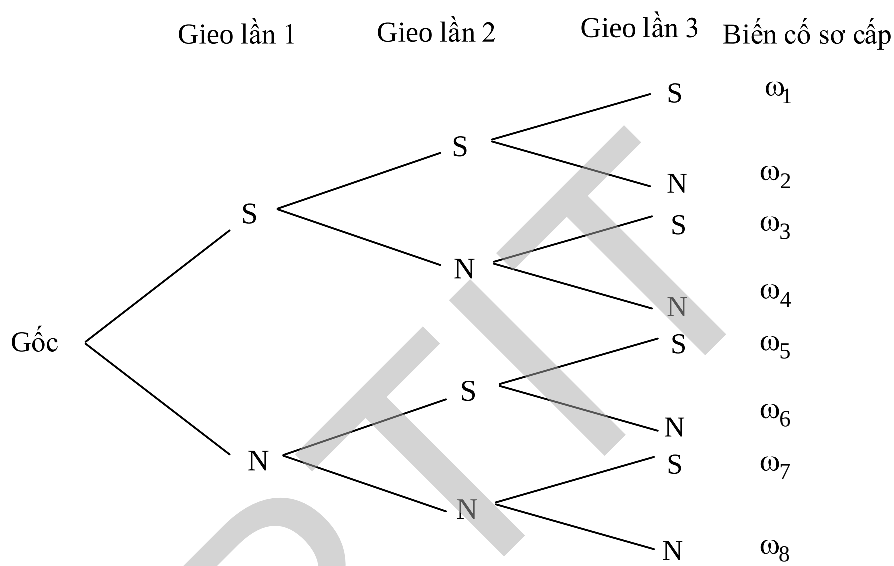
$A \cap B = \{\omega_3, \omega_4\}$, do đó $P(A \cap B) = \dfrac{1}{4}$. Áp dụng quy tắc cộng ta được

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{1}{2} + \frac{1}{2} - \frac{1}{4} = \frac{3}{4}.$$

Ta cũng có thể tính trực tiếp bằng cách xác định $A \cup B = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_7, \omega_8\}$. Vậy cũng có

$$P(A \cup B) = \frac{6}{8} = \frac{3}{4}.$$

Gieo liên tiếp một đồng xu 3 lần, ta có thể biểu diễn không gian mẫu như sau.



Hình 1.1: *Sơ đồ cây của phép thử gieo đồng xu liên tiếp 3 lần*

Ví dụ 1.25: Xét hai biến cố A, B trong cùng một phép thử có xác suất $P(A) = 0,7$, $P(B) = 0,6$.

a. Hai biến cố A, B có xung khắc không?

b. Giả sử $A \cup B$ là biến cố chắc chắn, tìm $P(A \cap B)$.

Giải : **a.** Theo công thức 1.8 và 1.13 ta có

$$1 \geq P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0,7 + 0,6 - P(A \cap B) \Rightarrow P(A \cap B) \geq 0,3$$

Vậy hai biến cố A, B không xung khắc.

b. Trường hợp $A \cup B$ là biến cố chắc chắn thì

$$P(A \cap B) = P(A) + P(B) - P(A \cup B) = 0,3.$$

1.4.3 Quy tắc xác suất của biến cố đối

Áp dụng công thức (1.12) cho hệ đầy đủ $\{A, \overline{A}\}$ ta được quy tắc tính xác suất biến cố đối:

Với mọi biến cố A

$$P(\overline{A}) = 1 - P(A) \text{ và cũng có } P(A) = 1 - P(\overline{A}). \qquad \textbf{(1.16)}$$

Ví dụ 1.26: Gieo con xúc xắc hai lần, tính xác suất ít nhất một lần xuất hiện mặt 6 chấm.

Giải : Gọi A là biến cố có ít nhất một xuất hiện mặt 6 chấm, khi đó biến cố đối $\overline{A}$ không có lần nào xuất hiện mặt 6 chấm..

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

Ví dụ 1.27: Trong phòng có n người ($n<365$; một năm có 365 ngày).

a. Tính xác suất có ít nhất hai người có cùng ngày sinh?

b. Tính xác suất này khi $n=10$.

Giải : **a.** Gọi A là biến cố có ít nhất hai người trong phòng có cùng ngày sinh. Biến cố đối $\overline{A}$ là biến cố mọi người không trùng ngày sinh.

Mọi người đều đồng khả năng được sinh ra vào một ngày bất kỳ trong năm do đó số các trường hợp có thể là 365^n. Số trường hợp thuận lợi đối với biến cố đối $\overline{A}$ là số chỉnh hợp chập n của 365. Vậy

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

b. Khi $n=10$ thì $P(\overline{A})=\dfrac{A_{365}^{10}}{365^{10}}=0,883,\ P(A)=1-0,883=0,117.$

Ví dụ 1.28: Giả sử phép thử C có không gian mẫu $\Omega=\{a,b,c,d\}$ với xác suất

$$P(a)=0,2,\ P(b)=0,3,\ P(c)=0,4,\ P(d)=0,1.$$

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

Tính xác suất của các biến cố $P(A)$; $P(B)$; $P(\overline{A})$; $P(A\cup B)$ và $P(A\cap B)$.

Giải: $P(A)=P(a)+P(b)=0,2+0,3=0,5$; $P(B)=P(b)+P(c)+P(d)=0,3+0,4+0,1=0,8$

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

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

$A\cap B=\{b\}$ do đó $P(A\cap B)=P(b)=0,3$.

1.4.4 Xác suất có điều kiện

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

Tính chất 1.1:

➢ Nếu $P(A)>0$ thì:

$$P(B|A)=\frac{P(A\cap B)}{P(A)}. \tag{1.17}$$

➢ Khi cố định A với $P(A)>0$ thì xác suất có điều kiện $P(B|A)$ có tất cả các tính chất của xác suất thông thường (công thức (1.8)-(1.16)) đối với biến cố B.

Chẳng hạn:

$$P(\overline{B}|A)=1-P\left(B\middle|A\right),\ P\left(B_1\cup B_2\middle|A\right)=P\left(B_1\middle|A\right)+P\left(B_2\middle|A\right)-P\left(B_1\cap B_2\middle|A\right)\ldots \quad \textbf{(1.18)}$$

Nhận xét 1.4: Ta có thể tính xác suất có điều kiện $P(B|A)$ bằng cách áp dụng công thức (1.17) hoặc tính trực tiếp.

Ví dụ 1.29: Gieo đồng thời hai con xúc xắc cân đối (6 mặt). Tính xác suất để tổng số chấm xuất hiện trên hai con xúc xắc ≥ 10 biết rằng ít nhất một con đã ra mặt có 5 chấm.

Giải: Gọi A là biến cố " ít nhất một con ra 5 chấm", bằng cách tính sử dụng xác suất biến cố đối tương tự ví dụ 1.27 ta có

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

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

Biến cố $A\cap B$ có 3 kết quả thuận lợi là (5,6), (6,5), (5,5). Vậy

$$P(A\cap B)=\frac{3}{36}\ \Rightarrow\ P\left(B\middle|A\right)=\frac{3}{36}\Big/\frac{11}{36}=\frac{3}{11}.$$

Ta cũng có thể tính trực tiếp như sau.

Có 11 trường hợp ít nhất một con xúc xắc xuất hiện mặt 5 chấm:

$$(5,1);(5,2);(5,3);(5,4);\mathbf{(5,5)};\mathbf{(5,6)};(1,5);(2,5);(3,5);(4,5);\mathbf{(6,5)}$$

trong đó có 3 trường hợp tổng số chấm ≥ 10.

Vậy $P(B|A)=\dfrac{3}{11}$

Ví dụ 1.30: Xét phép thử gieo đồng xu liên tiếp 3 lần ở ví dụ 1.24

Gọi A là biến cố lần thứ nhất ra mặt sấp.

B là biến cố lần thứ hai ra mặt ngửa.

C là biến cố số lần mặt sấp xuất hiện nhiều hơn hoặc bằng số lần mặt ngửa

$$P(A)=\frac{1}{2};\ P(A\cap B)=\frac{1}{4}\ \Rightarrow P(B|A)=\frac{1}{4}\Big/\frac{1}{2}=\frac{1}{2}.$$

$$A\cap C=\{\omega_1,\omega_2,\omega_3\}\ \Rightarrow P(A\cap C)=\frac{3}{8}\Rightarrow P(C|A)=\frac{3}{8}\Big/\frac{1}{2}=\frac{3}{4}.$$

Ví dụ 1.31: Có hai phân xưởng của nhà máy sản xuất cùng một loại sản phẩm. Phân xưởng I sản xuất được 1000 sản phẩm trong đó có 100 phế phẩm. Phân xưởng II sản xuất được 2000 sản phẩm trong đó có 150 phế phẩm. Lấy ngẫu nhiên một sản phẩm để kiểm tra và đó là phế phẩm. Tính xác suất phế phẩm này do phân xưởng thứ I sản xuất.

Giải: Gọi B là biến cố sản phẩm được chọn để kiểm tra là phế phẩm. Gọi A là biến cố sản phẩm được chọn để kiểm tra do phân xưởng I sản xuất. Ta cần tính xác suất có điều kiện $P(A|B)$.

Biến cố AB có 100 kết quả thuận lợi đồng khả năng do đó $P(A\cap B)=\dfrac{100}{3000}=\dfrac{1}{30}$.

Trong 3000 sản phẩm sản xuất ra có 250 phế phẩm, do đó $P(B)=\dfrac{250}{3000}=\dfrac{1}{12}$.

Áp dụng công thức (1.17) ta được

$$P(A\,|\,B)=\frac{1/30}{1/12}=2/5=0,4\,.$$

Ta có thể tính trực tiếp xác suất $P(A\,|\,B)$ như sau:

Có 250 trường hợp đồng khả năng có thể lấy được phế phẩm của nhà máy nhưng chỉ có 100 kết quả thuận lợi đối với biến cố phế phẩm do phân xưởng I sản xuất. Vậy xác suất để lấy được phế phẩm do phân xưởng thứ I sản xuất trong số các phế phẩm là

$$P(A\,|\,B)=\frac{100}{250}=\frac{2}{5}=0,4\,.$$

1.4.5 Quy tắc nhân xác suất

1.4.5.1 Trường hợp độc lập

- Nếu A, B là hai biến cố độc lập thì xác suất của biến cố B không phụ thuộc vào A có xảy ra hay không (xem mục 1.5.7), nghĩa là $P(B\,|\,A)=P(B)$. Theo (1.17) ta có

$$P(A\cap B)=P(A)P(B)\,. \tag{1.19}$$

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

$$P\left(A_1\cap A_2\cap ...\cap A_n\right)=P\left(A_1\right)P\left(A_2\right)...P\left(A_n\right) \tag{1.20}$$

Thông thường tính độc lập của các biến cố được suy ra từ ý nghĩa thực tế. Chẳng hạn nếu A là biến cố xạ thủ thứ nhất bắn trúng mục tiêu và B là biến cố xạ thủ thứ hai bắn trúng mục tiêu (xem ví dụ 1.14) thì A, B là hai biến cố độc lập.

1.4.5.2 Trường hợp không độc lập

- Với hai biến cố A, B bất kỳ, áp dụng công thức (1.17) ta có

$$P(A\cap B)=P(A)P(B\,|\,A) \tag{1.21}$$

- Với n biến cố bất kỳ $A_1,\, A_2,\, ...,\, A_n$:

$$P\left(A_1\cap A_2\cap ...\cap A_n\right)=P\left(A_1\right)P\left(A_2\middle|A_1\right)P\left(A_3\middle|A_1\cap A_2\right)...P\left(A_n\middle|A_1\cap A_2\cap ...\cap A_{n-1}\right) \tag{1.22}$$

Để chứng minh công thức (1.22) ta chỉ cần áp dụng lần lượt công thức (1.17) vào vế phải của công thức (1.22) và giản ước cuối cùng được vế trái.

Ví dụ 1.32: Túi I chứa 3 bi trắng, 7 bi đỏ, 15 bi xanh.

Túi II chứa 10 bi trắng, 6 bi đỏ, 9 bi xanh.

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

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

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

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

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

Biến cố 2 bi được rút cùng mầu là $(A_t \cap B_t) \cup (A_đ \cap B_đ) \cup (A_x \cap B_x)$

Vậy xác suất cần tìm:

$$P\big((A_t \cap B_t) \cup (A_đ \cap B_đ) \cup (A_x \cap B_x)\big) = P(A_t \cap B_t) + P(A_đ \cap B_đ) + P(A_x \cap B_x)$$

(theo công thức 1.11)

$$= P(A_t)P(B_t) + P(A_đ)P(B_đ) + P(A_x)P(B_x) \text{ (theo công thức 1.19)}$$

$$= \frac{3}{25}\cdot\frac{10}{25} + \frac{7}{25}\cdot\frac{6}{25} + \frac{15}{25}\cdot\frac{9}{25} = \frac{207}{625} \approx 0,331.$$

Ví dụ 1.33: Hai máy bay ném bom 1 mục tiêu. Mỗi máy bay ném 1 quả với xác suất trúng mục tiêu tương ứng là $0,7$, $0,8$ và độc lập với nhau. Tìm xác suất để mục tiêu bị trúng bom.

Giải: Gọi A_1, A_2 lần lượt tương ứng là biến cố "máy bay thứ nhất và máy bay thứ hai ném trúng mục tiêu". A là biến cố "mục tiêu bị đánh trúng".

Rõ ràng $A = A_1 \cup A_2$ và A_1, A_2 độc lập.

Do đó $P(A) = P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$

$$= P(A_1) + P(A_2) - P(A_1)P(A_2) = 0,7 + 0,8 - 0,7.0,8 = 0,96.$$

Ví dụ 1.34: Một hộp đựng 100 sản phẩm trong đó có 20 phế phẩm. Lấy ngẫu nhiên lần lượt và không hoàn lại 2 sản phẩm ở trong hộp.

a. Tính xác suất sản phẩm lấy được lần đầu là phế phẩm.

b. Tính xác suất sản phẩm lấy được lần thứ hai là phế phẩm biết rằng sản phẩm lấy lần đầu cũng là phế phẩm.

c. Tính xác suất cả hai sản phẩm lấy được đều là phế phẩm.

***Giải*: a**. Gọi A_1 là biến cố sản phẩm lấy được lần đầu là phế phẩm, ta có

$$P(A_1) = \frac{20}{100} = 0,2.$$

b. Gọi A_2 là biến cố sản phẩm lấy được lần thứ hai là phế phẩm. Vậy xác suất sản phẩm lấy được lần thứ hai là phế phẩm biết rằng sản phẩm lấy lần đầu cũng là phế phẩm:

$$P(A_2 \mid A_1) = \frac{19}{99} = 0,192.$$

c. $P(A_1 \cap A_2) = P(A_1)P(A_2 \mid A_1) = \dfrac{20}{100}\cdot\dfrac{19}{99} = 0,0384.$

Ví dụ 1.35: Rút lần lượt ngẫu nhiên không hoàn lại 3 quân bài từ cỗ bài tú lơ khơ. Tính xác suất trong các trường hợp sau:

a. Cả 3 quân bài rút được không phải là quân bích.

b. Lần thứ nhất rút được không phải quân bích và lần thứ hai rút được quân bích.

c. Hai lần đầu rút được không phải quân bích và lần thứ ba rút được quân bích.

Giải: : Gọi A_1, A_2, A_3 lần lượt tương ứng là biến cố lần thứ nhất, lần thứ hai và lần thứ ba rút được quân bài không phải là bích.

a. Biến cố cả 3 quân bài rút được không phải là quân bích là $A_1 \cap A_2 \cap A_3$.

Vậy xác suất cần tìm là $P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2 \mid A_1)P(A_3 \mid A_1 \cap A_2)$.

$$P(A_1) = \frac{39}{52},\ P(A_2 \mid A_1) = \frac{38}{51},\ P(A_3 \mid A_1 \cap A_2) = \frac{37}{50}.$$

$$P(A_1 \cap A_2 \cap A_3) = \frac{39}{52} \cdot \frac{38}{51} \cdot \frac{37}{50}.$$

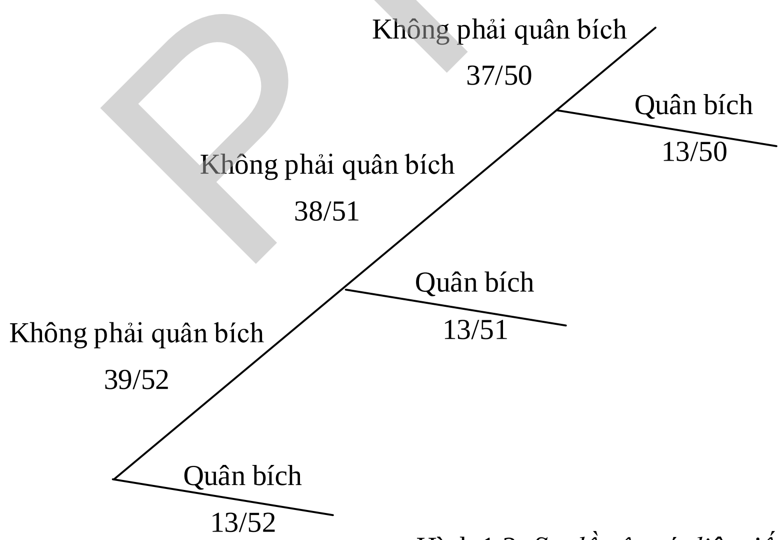
b. Xác suất lần thứ nhất rút được không phải quân bích và lần thứ hai rút được quân bích là

$$P\left(A_1 \overline{A_2}\right) = P\left(A_1\right) P\left(\overline{A_2} \mid A_1\right) = \frac{39}{52} \cdot \frac{13}{51}.$$

c. Xác suất hai lần đầu rút được không phải quân bích và lần thứ ba rút được quân bích là

$$P\left(A_1 \cap A_2 \cap \overline{A_3}\right) = P\left(A_1\right) P\left(A_2 \mid A_1\right) P\left(\overline{A_3} \mid A_1 \cap A_2\right) = \frac{39}{52} \cdot \frac{38}{51} \cdot \frac{13}{50}.$$

Tương tự ví dụ 1.12 và ví dụ 1.24 ta có thể biểu diễn các biến cố và xác suất tương ứng của phép thử rút liên tiếp 3 quân bài dưới dạng sơ đồ cây



Hình 1.2: *Sơ đồ cây rút liên tiếp 3 quân bài*

Ví dụ 1.36: Một thủ kho có một chùm chìa khóa gồm 9 chiếc, bề ngoài chúng giống hệt nhau nhưng trong đó chỉ có đúng 2 chiếc mở được kho. Anh ta thử ngẫu nhiên từng chìa (chìa nào không trúng thì bỏ ra). Tính xác suất để đến lần thử thứ ba mới mở được kho.

Giải: Ký hiệu A_i là biến cố "thử đúng chìa ở lần thứ i"; $i = 1, ..., 8$.

Ký hiệu B là biến cố "đến lần thử thứ ba mới mở được kho".

Ta có $B = \overline{A_1} \cap \overline{A_2} \cap A_3$.

Các biến cố này không độc lập, áp dụng công thức 1.18 ta có

$$P\left(\overline{A_1} \cap \overline{A_2} \cap A_3\right) = P\left(\overline{A_1}\right) P\left(\overline{A_2} \,\middle|\, \overline{A_1}\right) P\left(A_3 \,\middle|\, \overline{A_1} \cap \overline{A_2}\right).$$

Từ giả thiết ta có thể tính được

$$P\left(\overline{A_1}\right) = \frac{7}{9},\ P\left(\overline{A_2} \,\middle|\, \overline{A_1}\right) = \frac{6}{8},\ P\left(A_3 \,\middle|\, \overline{A_1} \cap \overline{A_2}\right) = \frac{2}{7}$$

Do đó

$$P\left(\overline{A_1} \cap \overline{A_2} \cap A_3\right) = \frac{7}{9} \cdot \frac{6}{8} \cdot \frac{2}{7} = \frac{1}{6}.$$

1.4.6 Công thức xác suất đầy đủ

Định lý 1.3: Giả sử $\{A_1, A_2, ..., A_n\}$ là một hệ đầy đủ các biến cố, khi đó với mọi biến cố B (trong cùng phép thử) ta có

$$P(B) = \sum_{i=1}^{n} P(A_i \cap B) = \sum_{i=1}^{n} P(A_i) P\left(B \,\middle|\, A_i\right) \tag{1.23}$$

Ví dụ 1.37: Một túi đựng 4 bi trắng và 6 bi đen. Người thứ nhất lấy ngẫu nhiên từ túi 3 bi (không hoàn lại), người thứ hai lấy tiếp 2 bi. Tính xác suất để người thứ hai lấy được 1 bi trắng.

Giải: Gọi lần lượt A_0, A_1, A_2, A_3 là biến cố người thứ nhất lấy được 0, 1, 2, 3 bi trắng.

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

Ta có: $P(A_0) = \dfrac{C_6^3}{C_{10}^3} = \dfrac{1}{6}$, $P(A_1) = \dfrac{C_4^1 C_6^2}{C_{10}^3} = \dfrac{1}{2}$, $P(A_2) = \dfrac{C_4^2 C_6^1}{C_{10}^3} = \dfrac{3}{10}$, $P(A_3) = \dfrac{C_4^3}{C_{10}^3} = \dfrac{1}{30}$.

Ta có bảng tổng hợp của các kết quả sau khi người thứ nhất chọn ngẫu nhiên 3 bi:

Biến cố A_k xảy ra	A_0	A_1	A_2	A_3
Số bi màu trắng người thứ nhất lấy được	0	1	2	3
Số bi màu trắng còn lại sau khi người thứ nhất lấy	4	3	2	1
Số bi màu đen còn lại sau khi người thứ nhất lấy	3	4	5	6

Từ đó ta tính được các xác suất có điều kiện

$$P(B|A_0) = \frac{C_4^1 C_3^1}{C_7^2} = \frac{12}{21},\ P(B|A_1) = \frac{C_3^1 C_4^1}{C_7^2} = \frac{12}{21},\ P(B|A_2) = \frac{C_2^1 C_5^1}{C_7^2} = \frac{10}{21},\ P(B|A_3) = \frac{C_1^1 C_6^1}{C_7^2} = \frac{6}{21}.$$

Vậy $P(B) = \dfrac{1}{6} \cdot \dfrac{12}{21} + \dfrac{1}{2} \cdot \dfrac{12}{21} + \dfrac{3}{10} \cdot \dfrac{10}{21} + \dfrac{1}{30} \cdot \dfrac{6}{21} = \dfrac{56}{105}$.

Ví dụ 1.38: Một người tham gia thi đấu cờ vua với một nhóm các đấu thủ chia làm ba loại: loại I chiếm $1/2$ số đấu thủ, loại II chiếm $1/4$ số đấu thủ và loại III chiếm $1/4$ số đấu thủ còn lại. Xác suất anh ta thắng đấu thủ loại I là 0,3, thắng đấu thủ loại II là 0,4 và thắng đấu thủ loại III là 0,5. Anh ta thi đấu ngẫu nhiên với một trong các đấu thủ loại I, loại II hoặc loại III. Tính xác suất anh ta thắng cuộc.

Giải: Gọi A_1, A_2, A_3 lần lượt là biến cố anh ta thi đấu với một trong các đấu thủ thuộc loại I, loại II, hoặc loại III. Ta có

$$P(A_1)=0,5, \quad P(A_2)=0,25, \quad P(A_3)=0,25.$$

Gọi B là biến cố anh ta đánh thắng, theo giả thiết ta có

$$P(B\,|\,A_1)=0,3; \qquad P(B\,|\,A_2)=0,4; \qquad P(B\,|\,A_3)=0,5.$$

Áp dụng công thức xác suất đầy đủ (1.23) ta được

$$P(B)=P(A_1)P\left(B\middle|A_1\right)+P(A_2)P\left(B\middle|A_2\right)+P(A_3)P\left(B\middle|A_3\right)$$

$$=0,5\cdot 0,3+0,25\cdot 0,4+0,25\cdot 0,5=0,375.$$

Ví dụ 1.39: Gieo xúc xắc. Nếu mặt 1 chấm hoặc 2 chấm xuất hiện ta gieo tiếp lần nửa và ngừng nếu ngược lại. Tính xác suất để tổng số chấm xuất hiện ít nhất là 5.

Giải: Gọi A_k là biến cố lần gieo thứ nhất xuất hiện k chấm, ta có

$$P(A_k)=\frac{1}{6} \text{ với mọi } k=1,2,3,4,5,6.$$

Gọi B là biến cố tổng số chấm xuất hiện ít nhất là 5.

Giả sử biến cố A_1 xảy ra, khi đó tổng số chấm ít nhất là 5 khi kết quả của lần gieo thứ hai là 4 chấm, 5 chấm hoặc 6 chấm. Tương tự, nếu biến cố A_2 xảy ra, khi đó tổng số chấm ít nhất là 5 khi kết quả của lần gieo thứ hai là 3, 4, 5 hoặc 6 chấm. Vậy

$$P(B\,|\,A_1)=\frac{3}{6}, \qquad P(B\,|\,A_2)=\frac{4}{6}$$

Nếu biến cố A_3, A_4, A_5 hoặc A_6 xảy ra thì dừng lại không gieo tiếp lần thứ hai, do đó

$$P(B\,|\,A_3)=P(B\,|\,A_4)=0, \quad P(B\,|\,A_5)=P(B\,|\,A_6)=1.$$

Áp dụng công thức xác suất đầy đủ ta được

$$P(B)=\frac{1}{6}\cdot\frac{3}{6}+\frac{1}{6}\cdot\frac{4}{6}+\frac{1}{6}\cdot 0+\frac{1}{6}\cdot 0+\frac{1}{6}\cdot 1+\frac{1}{6}\cdot 1=\frac{19}{36}.$$

1.4.7 Công thức Bayes

Từ công thức (1.21) ta có

$$P(A_k)P(B\,|\,A_k)=P(A_k\cap B)=P(B)P(A_k\,|\,B).$$

Dựa vào đẳng thức này ta được kết quả sau.

Định lý 1.4: Giả sử $\{A_1, A_2, ..., A_n\}$ là một hệ đầy đủ các biến cố, khi đó với mọi biến cố B (trong cùng phép thử) sao cho $P(B)>0$ ta có công thức Bayes:

$$P\left(A_k \middle| B\right)=\frac{P(A_k)P\left(B \middle| A_k\right)}{P(B)}=\frac{P(A_k)P\left(B \middle| A_k\right)}{\sum_{i=1}^{n} P(A_i)P\left(B \middle| A_i\right)};\ k=1,2,...,n \qquad \textbf{(1.24)}$$

Ví dụ 1.40: Một nhà máy có ba phân xưởng I, II, III cùng sản xuất ra một loại sản phẩm. Phân xưởng I, II, III sản xuất tương ứng 36%, 34%, 30% sản lượng của nhà máy, với tỷ lệ phế phẩm tương ứng là $0,12$; $0,1$; $0,08$.

a. Tìm tỷ lệ phế phẩm chung của nhà máy.

b. Lấy ngẫu nhiên một sản phẩm kiểm tra và đó là phế phẩm. Tính xác suất để phế phẩm đó là do phân xưởng I sản xuất.

Giải: Lấy ngẫu nhiên một sản phẩm của nhà máy để kiểm tra. Gọi B là biến cố "sản phẩm kiểm tra là phế phẩm".

Gọi A_1, A_2, A_3 lần lượt là biến cố sản phẩm lấy ra kiểm tra do phân xưởng I, II, III sản xuất. Theo giả thiết thì hệ 3 biến cố $\{A_1,\ A_2,\ A_3\}$ đầy đủ (xem thêm ví dụ 1.14) và có các xác suất.

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

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

a. Xác suất của biến cố B cũng là tỉ lệ phế phẩm chung của nhà máy.

Áp dụng công thức xác suất đầy đủ (1.23) ta có

$$P\left(B\right)=P\left(A_1\right)P\left(B \middle| A_1\right)+P\left(A_2\right)P\left(B \middle| A_2\right)+P\left(A_3\right)P\left(B \middle| A_3\right)=0,1012$$

b. Áp dụng công thức Bayes ta được

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

Ví dụ 1.41: Người ta dùng một thiết bị để kiểm tra một loại sản phẩm nhằm xác định sản phẩm có đạt yêu cầu không. Biết rằng sản phẩm có tỉ lệ phế phẩm là p. Thiết bị có khả năng phát hiện đúng sản phẩm là phế phẩm với xác suất α và phát hiện đúng sản phẩm đạt chất lượng với xác suất β. Kiểm tra ngẫu nhiên một sản phẩm, tìm xác suất sao cho sản phẩm này:

a. Được kết luận là phế phẩm.

b. Được kết luận là đạt chất lượng thì lại là phế phẩm.

c. Được kết luận là đúng với thực chất của nó.

Giải: Gọi A là biến cố được kết luận là phế phẩm.

Gọi H là biến cố "sản phẩm được chọn là phế phẩm".

Theo giả thiết ta có:

$$P(H)=p,\ P\left(A \middle| H\right)=\alpha,\ P\left(\overline{A} \middle| \overline{H}\right)=\beta.$$

a. Áp dụng công thức đầy đủ của biến cố A với hệ đầy đủ $\left\{H,\overline{H}\right\}$ ta có:

$$P(A)=P(H)P\left(A\middle|H\right)+P\left(\overline{H}\right)P\left(A\middle|\overline{H}\right)=p\alpha+(1-p)(1-\beta).$$

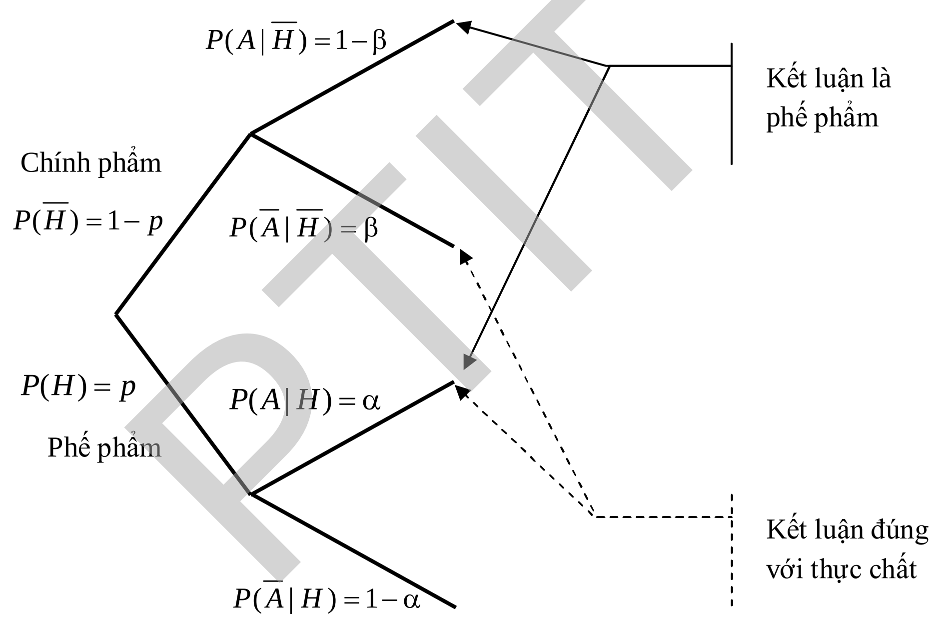
b. Biến cố sản phẩm kiểm tra được kết luận đạt chất lượng nhưng là phế phẩm là biến cố H với điều kiện $\overline{A}$. Áp dụng công thức Bayes ta được

$$P\left(H\mid\overline{A}\right)=\frac{P\left(H\cap\overline{A}\right)}{P\left(\overline{A}\right)}=\frac{P(H)P\left(\overline{A}\mid H\right)}{P\left(\overline{A}\right)}=\frac{p(1-\alpha)}{p(1-\alpha)+(1-p)\beta}.$$

c. Biến cố kết luận là đúng với thực chất của nó là $AH\cup\overline{A}\,\overline{H}$, có xác suất

$$P\left(A\cap H\right)+P\left(\overline{A}\cap\overline{H}\right)=P(H)P\left(A\middle|H\right)+P\left(\overline{H}\right)P\left(\overline{A}\middle|\overline{H}\right)=p\alpha+(1-p)\beta.$$

Ta có thể biểu diễn kết quả dưới dạng cây biểu đồ như sau



Hình 1.3: *Sơ đồ cây xác suất đầy đủ*

Ví dụ 1.42: Giả sử hai biến cố A, B có xác suất $P(A)=2/5$, $P(B)=1/3$ và $P(AB)=1/6$. Hãy tính

a. $P(A|B)$ **b**. $P(A\cup B)$ **c.** $P(A\cap\overline{B})$ **d.** $P(\overline{B}\mid\overline{A})$.

Giải: **a.** $P(A\mid B)=\dfrac{P(A\cap B)}{P(B)}=\dfrac{1/6}{1/3}=\dfrac{1}{2}$

b. $P(A\cup B)=P(A)+P(B)-P(A\cap B)=\dfrac{2}{5}+\dfrac{1}{3}-\dfrac{1}{6}=\dfrac{17}{30}$.

c. $P(B\mid A)=\dfrac{P(B\cap A)}{P(A)}=\dfrac{1/6}{2/5}=\dfrac{5}{12}\Rightarrow P(A\cap\overline{B})=P(A)P(\overline{B}\mid A)=\dfrac{2}{5}.\left(1-\dfrac{5}{12}\right)=\dfrac{7}{30}$.

d. $P(A\,|\,\overline{B})=\dfrac{P(A\cap\overline{B})}{P(\overline{B})}=\dfrac{7/30}{2/3}=\dfrac{7}{20}\Rightarrow P(\overline{B}\,|\,\overline{A})=\dfrac{P(\overline{B})P(\overline{A}\,|\,\overline{B})}{P(\overline{A})}=\dfrac{\frac{2}{3}.\left(1-\frac{7}{20}\right)}{\frac{3}{5}}=\dfrac{13}{18}$.

Ví dụ 1.43: Trước khi đưa sản phẩm ra thị trường người ta đã phỏng vấn ngẫu nhiên 200 khách hàng về sản phẩm đó và thấy có 34 người trả lời "sẽ mua", 97 người trả lời "có thể sẽ mua" và 69 người trả lời "không mua". Kinh nghiệm cho thấy tỷ lệ khách hàng thực sự sẽ mua sản phẩm tương ứng với những cách trả lời trên tương ứng là 70%, 30% và 1%.

a. Hãy đánh giá thị trường tiềm năng của sản phẩm đó.

b. Trong số khách hàng thực sự mua sản phẩm thì có bao nhiêu phần trăm trả lời "sẽ mua".

Giải: Gọi B là biến cố "người được phỏng vấn sẽ mua sản phẩm".

Gọi A_1, A_2, A_3 lần lượt là 3 biến cố tương ứng với 3 cách trả lời của khách hàng được phỏng vấn:

A_1 - người đó trả lời "sẽ mua"

A_2 - người đó trả lời "có thể mua"

A_3 - người đó trả lời "không mua"

$\{A_1, A_2, A_3\}$ là một hệ đầy đủ các biến cố với xác suất tương ứng $\dfrac{34}{200}, \dfrac{97}{200}, \dfrac{69}{200}$.

Các xác suất điều kiện $P\left(B\middle|A_1\right)=0,7$; $P\left(B\middle|A_2\right)=0,3$; $P\left(B\middle|A_3\right)=0,01$.

a. Theo công thức xác suất đầy đủ

$$P(B)=\frac{34}{200}\cdot 0,7+\frac{97}{200}\cdot 0,3+\frac{69}{200}\cdot 0,01=0,268$$

Vậy thị trường tiềm năng của sản phẩm đó là 26,8%.

b. Theo công thức Bayes

$$P\left(A_1\middle|B\right)=\frac{P(A_1)P\left(B\middle|A_1\right)}{P(B)}=\frac{0,17\cdot 0,7}{0,268}=0,444=44,4\% .$$

Nhận xét 1.5: Trong thực tế các xác suất $\{P(A_1), P(A_2), ..., P(A_n)\}$ đã biết và được gọi là các *xác suất tiền nghiệm*. Sau khi quan sát biết được biến cố B xảy ra, các xác suất của A_k được tính trên thông tin này (xác suất có điều kiện $P(A_k\,|\,B)$) được gọi là *xác suất hậu nghiệm*. Vì vậy công thức Bayes còn được gọi là công thức xác suất hậu nghiệm.

1.5 DÃY PHÉP THỬ BERNOULLI

Một phép thử có thể lặp lại độc lập và trong mỗi phép thử ta xét sự xuất hiện của biến cố A không đổi với $P(A)=p$, $(0<p<1)$ được gọi là phép thử Bernoulli.

p là xác suất thành công trong mỗi lần thử.

Một dãy lặp lại cùng một phép thử Bernoulli được gọi là dãy phép thử Bernoulli.

Kí hiệu H_k là biến cố “ A xuất hiện ra đúng k lần trong n phép thử”.

Đặt $P_n(k;p)=P(H_k)$.

Định lý 1.4: Xác suất của biến cố “ A xuất hiện ra đúng k lần trong n phép thử” là:

$$P_n(k;p)=C_n^k p^k (1-p)^{n-k};\ k=0,1,...,n. \qquad \textbf{(1.25)}$$

Chứng minh: H_k là tổng của C_n^k các biến cố xung khắc từng đôi nhận được bằng cách hoán vị các chữ A và $\overline{A}$ trong biến cố tích sau (xem nhận xét 1.1):

$$\underbrace{A\cap...\cap A}_{k\ \text{lÇn}}\cap\underbrace{\overline{A}\cap...\cap\overline{A}}_{n-k\ \text{lÇn}}$$

Từ tính chất độc lập suy ra xác suất của mỗi biến cố dạng này bằng

$$P(\underbrace{A\cap...\cap A}_{k\ \text{lÇn}}\cap\underbrace{\overline{A}\cap...\cap\overline{A}}_{n-k\ \text{lÇn}})=p^k(1-p)^{n-k}.$$

Vậy $\quad P_n(k;p)=C_n^k p^k (1-p)^{n-k}$.

Khi p và n không đổi thì xác suất $P_n(k;p)$ phụ thuộc k và đạt giá trị lớn nhất thỏa mãn điều kiện sau.

Định lý 1.5: Thực hiện một dãy n phép thử Bernoulli với xác suất thành công trong mỗi lần thử là p. Ta có các kết quả sau:

(i). $$P_n(k;p)=\frac{(n-k+1)p}{kq}P_n(k-1;p) \qquad \textbf{(1.26)}$$

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

$$(n+1)p-1\le m\le (n+1)p \qquad \textbf{(1.27)}$$

Như vậy, $P_{\max}=P_n(m;p)$

- Khi $(n+1)p$ không nguyên thì $m=\left[(n+1)p\right]$ (là phần nguyên của $(n+1)p$).
- Khi $(n+1)p$ nguyên thì $m=(n+1)p-1$ hoặc $m=(n+1)p$

$$P_{\max}=P_n(m-1;p)=P_n(m;p) \qquad \textbf{(1.28)}$$

Chứng minh: $\dfrac{P_n(k;p)}{P_n(k-1;p)}=\dfrac{\dfrac{n!}{k!(n-k)!}p^k q^{n-k}}{\dfrac{n!}{(k-1)!(n-k+1)!}p^{k-1}q^{n-k+1}}=\dfrac{(n-k+1)p}{kq}$, từ đó có (1.26).

(1.26) $\Rightarrow\ \dfrac{P_n(k;p)}{P_n(k+1;p)}=\dfrac{(k+1)(1-p)}{(n-k)p}$. Do đó

$$\frac{P_n(k;p)}{P_n(k+1;p)}<1 \Leftrightarrow \frac{(k+1)(1-p)}{(n-k)p}<1 \Leftrightarrow (k+1)-(k+1)p<np-kp \Leftrightarrow k+1<(n+1)p .$$

Vậy: $P_n(k;p)<P_n(k+1;p)$ khi $k<(n+1)p-1 \Rightarrow P_n(k;p)<P_n(m;p)$, $\forall k<(n+1)p-1$.

và $P_n(k;p)>P_n(k+1;p)$ khi $k\ge(n+1)p \Rightarrow P_n(k;p)<P_n(m;p)$, $\forall k>(n+1)p$,

trong đó m là số tự nhiên thỏa mãn $(n+1)p-1\le m\le(n+1)p$.

Khi $m=(n+1)p$ thì $\dfrac{P_n(m-1;p)}{P_n(m;p)}=\dfrac{(n+1)(1-p)p}{\left(n-(n+1)p+1\right)p}=\dfrac{(n+1)(1-p)p}{\left(n+1-(n+1)p\right)p}=1$

$$\Rightarrow P_n(m-1;p)=P_n(m;p).$$

Định nghĩa 1.3: m *xác định bởi công thức* (1.27) *hoặc* (1.28) *được gọi là số lần xuất hiện có khả năng nhất hay giá trị có khả năng xảy ra lớn nhất.*

Ví dụ 1.44: Bắn 7 viên đạn vào bia. Xác suất trúng đích của mỗi viên là $0,6$. Tìm xác suất trong các trường hợp sau:

a. Có đúng 3 viên trúng bia.

b. Có ít nhất 6 viên trúng bia.

c. Có ít nhất 1 viên trúng bia.

d. Tìm số viên đạn trúng bia có khả năng lớn nhất và tính xác suất tương ứng.

Giải: Có thể xem bắn mỗi viên đạn vào bia là thực hiện một phép thử Bernoulli mà xác suất thành công của phép thử là xác suất bắn trúng bia, theo giả thiết là 0,6. Bắn 7 viên là thực hiện 7 lần phép thử. Vậy:

a. Xác suất để có đúng 3 viên trúng bia là

$$P_7(3;0,6)=C_7^3(0,6)^3(0,4)^4=0,1935.$$

b. Xác suất để có ít nhất 6 viên trúng bia là

$$P_7(6;0,6)+P_7(7;0,6)=C_7^6(0,6)^6(0,4)+C_7^7(0,6)^7=0,1586.$$

c. Xác suất để có ít nhất 1 viên trúng bia là

$$1-P_7(0;0,6)=1-C_7^0(0,6)^0(0,4)^7=1-(0,4)^7=0,998.$$

d. $(n+1)p=(7+1)(0,6)=4,8$. Vậy số viên đạn có khả năng trúng bia nhất là 4.

$$P_7(4;0,6)=C_7^4(0,6)^4(0,4)^3=0,2903.$$

Ví dụ 1.45: Tín hiệu thông tin được phát đi 3 lần độc lập nhau. Xác suất thu được mỗi lần là 0.4.

a. Tìm xác suất để nguồn thu nhận được thông tin đúng 2 lần.

b. Tìm xác suất để nguồn thu nhận được thông tin đó.

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

Giải: Có thể xem mỗi lần phát tin là một phép thử Bernoulli mà sự thành công của phép thử là nguồn thu nhận được tin, theo giả thiết xác suất thành công của mỗi lần thử là 0,4. Vậy:

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

$$P_3(2;0,4)=C_3^2\left(0,4\right)^2\left(0,6\right)=0,288.$$

b. Xác suất để nguồn thu nhận được thông tin là

$$P=1-P_3(0;0,4)=1-\left(0,6\right)^3=0,784.$$

c. Xác suất để nguồn thu nhận được thông tin khi phát n lần là $P=1-\left(0,6\right)^n$.

Vậy nếu muốn xác suất thu được tin $\ge 0,9$ thì phải phát đi ít nhất n lần sao cho:

$1-\left(0,6\right)^n\ge 0,9 \Leftrightarrow \left(0,6\right)^n\le 0,1 \;\Leftrightarrow n\ge\dfrac{\lg\left(0,1\right)}{\lg\left(0,6\right)}=\dfrac{-1}{-1+0,778}=4,504$. Chọn $n=5$.

1.6 NGUYÊN LÝ XÁC SUẤT LỚN, XÁC SUẤT NHỎ

Biến cố không thể (biến cố $\varnothing$) có xác suất bằng 0. Một biến cố có xác suất gần bằng 0 vẫn có thể xảy ra khi thức hiện một số lớn các phép thử. Tuy nhiên qua thực nghiệm và quan sát thực tế, người ta thấy rằng các biến cố có xác suất nhỏ sẽ không xảy ra khi ta chỉ thực hiện một phép thử hay một vài phép thử. Từ đó ta thừa nhận nguyên lý sau đây, gọi là "Nguyên lý xác suất nhỏ": *Nếu một biến cố có xác suất rất nhỏ thì thực tế có thể cho rằng trong một phép thử biến cố đó sẽ không xảy ra.*

Khi tung đồng xu vẫn có khả năng đồng xu không xảy ra mặt sấp hoặc mặt ngửa mà ở trạng thái thẳng đứng. Tuy nhiên khả năng này rất khó xảy ra, vì vậy trong thực tế ta luôn giả thiết chỉ có hai khả năng là mặt sấp hoặc mặt ngửa xuất hiện.

Chẳng hạn mỗi chiếc máy bay đều có một xác suất rất nhỏ bị xảy ra tai nạn. Nhưng trên thực tế ta vẫn không từ chối đi máy bay vì tin tưởng rằng trong chuyến bay ta đi sự kiện máy bay rơi không xảy ra.

Hiển nhiên việc quy định một mức xác suất thế nào được gọi là nhỏ sẽ phụ thuộc vào từng bài toán cụ thể. Chẳng hạn nếu xác suất để máy bay rơi là 0,01 thì xác suất đó chưa thể được coi là nhỏ. Song nếu xác suất một chuyến tàu khởi hành chậm là 0,01 thì có thể coi rằng xác suất này là nhỏ.

Mức xác suất nhỏ này được gọi là *mức ý nghĩa*. Nếu α là mức ý nghĩa thì số $\beta=1-\alpha$ gọi là *độ tin cậy*. Khi dựa trên nguyên lý xác suất nhỏ ta khẳng định rằng: "Biến cố A có xác suất nhỏ (tức là $P(A)\le\alpha$) sẽ không xảy ra trên thực tế" thì độ tin cậy của kết luận trên là β. Tính đúng đắn của kết luận chỉ xảy ra trong $100\cdot\beta\%$ trường hợp.

Tương tự như vậy ta có thể đưa ra "Nguyên lý xác suất lớn": "*Nếu biến cố A có xác suất gần bằng 1 thì trên thực tế có thể cho rằng biến cố đó sẽ xảy ra trong một phép thử*". Cũng như trên, việc quy định một mức xác suất thế nào được gọi là lớn sẽ tùy thuộc vào từng bài toán cụ thể.

CÂU HỎI ÔN TẬP VÀ BÀI TẬP CHƯƠNG 1

1.1 Ta có thể có hai không gian mẫu Ω các biến cố sơ cấp cho cùng một phép thử $\mathcal{C}$?

Đúng ☐ Sai ☐.

1.2 Các biến cố A và $\overline{A} \cup \overline{B}$ luôn xung khắc.

Đúng ☐ Sai ☐.

1.3 Hai biến cố A và B là xung khắc thì $P(A \cup B) = P(A) + P(B)$.

Đúng ☐ Sai ☐.

1.4 Có thể xét tổng của 2 biến cố ứng với 2 phép thử khác nhau.

Đúng ☐ Sai ☐.

1.5 Hai biến cố xung khắc là hai biến cố độc lập.

Đúng ☐ Sai ☐.

1.6 Các biến cố đối của hai biến cố độc lập cũng là độc lập.

Đúng ☐ Sai ☐.

1.7 Xác suất của tổng hai biến cố độc lập bằng tổng xác suất của hai biến cố này.

Đúng ☐ Sai ☐.

1.8 Xác suất của tích 2 biến cố xung khắc bằng tích xác suất của 2 biến cố này.

Đúng ☐ Sai ☐.

1.9 Hệ 2 biến cố $\left\{A, \overline{A}\right\}$ là hệ đầy đủ.

Đúng ☐ Sai ☐.

1.10 Cho $\Omega = \left\{a, b, c, d\right\}$ trong đó các biến cố sơ cấp là đồng khả năng. Biến cố $A = \left\{a, b\right\}$ và $B = \left\{a, c\right\}$ là phụ thuộc vì chúng cùng xảy ra khi biến cố sơ cấp a xảy ra.

Đúng ☐ Sai ☐.

1.11 Trong thành phố có 5 hòm thư đều có đường liên lạc với nhau. Người bưu tá đi đưa thư theo một trình tự nào đó. Hỏi có bao nhiêu cách đi?

1.12 Trong một hòm đựng 10 chi tiết đạt tiêu chuẩn và 5 chi tiết là phế phẩm**.** Lấy đồng thời 3 chi tiết. Tính xác suất:

a. Cả 3 chi tiết lấy ra thuộc loại đạt tiêu chuẩn.

b. Trong số 3 chi tiết lấy ra có 2 chi tiết đạt tiêu chuẩn.

1.13 Một hộp có 10 bi màu đỏ, 30 trắng, 20 xanh và 15 vàng**.** Lấy ngẫu nhiên từ hộp 1 bi, tính xác suất bi lấy được trong các trường hợp sau:

a. màu vàng hoặc đỏ

b. không phải màu và không phải màu xanh

c. màu trắng

d. màu đỏ hoặc trắng hoặc xanh

1.14 Một hộp có 2 bi màu đỏ và 3 bi màu xanh**.** Lấy ngẫu nhiên từ hộp 2 bi, tính xác suất 2 bi lấy được trong các trường hợp sau:

a. cả hai cùng màu xanh

b. cả hai cùng màu đỏ

c. 1 bi màu đỏ và 1 bi màu xanh

1.15 Thang máy của một tòa nhà 7 tầng xuất phát từ tầng một với 3 khách. Tìm xác suất để:

a. Tất cả cùng ra ở tầng bốn.

b. Tất cả cùng ra ở một tầng

c. Mỗi người ra một tầng khác nhau.

1.16 Một người gọi điện thoại cho bạn nhưng lại quên mất 3 chữ số cuối và chỉ nhớ rằng chúng khác nhau. Tìm xác suất để người đó quay số một lần được đúng số điện thoại của bạn.

1.17 Ta kiểm tra theo thứ tự một lô hàng có 10 sản phẩm. Mỗi sản phẩm thuộc một trong hai loại: Tốt hoặc Xấu. Ký hiệu A_k $(k=1,...,10)$ là biến cố chỉ sản phẩm kiểm tra thứ k thuộc loại tốt. Biểu diễn các biến cố sau theo A_k:

a. Cả 10 sản phẩm đều tốt.

b. Có ít nhất một sản phẩm tốt.

c. Có 6 sản phẩm kiểm tra đầu là tốt, các sản phẩm còn lại là xấu.

d. Có 6 sản phẩm kiểm tra đầu là tốt.

1.18 Hai người cùng bắn vào một mục tiêu. Khả năng bắn trúng của từng người là 0,8 và 0,9. Tìm xác suất:

a. Chỉ có một người bắn trúng mục tiêu.

b. Có người bắn trúng mục tiêu.

c. Cả hai người bắn trượt.

1.19 Cơ cấu chất lượng sản phẩm của nhà máy như sau: 40% là sản phẩm loại I, 50% là sản phẩm loại II, còn lại là phế phẩm. Lấy ngẫu nhiên một sản phẩm của nhà máy. Tính xác suất sản phẩm lấy ra không phải là phế phẩm.

1.20 Tín hiệu thông tin được phát đi 3 lần độc lập nhau. Xác suất thu được tin của mỗi lần phát là 0,4. Tính xác suất để thu được thông tin đó.

1.21 Có 1000 vé số trong đó có 20 vé trúng thưởng. Một người mua 30 vé, tìm xác suất để người đó trúng 5 vé.

1.22 Rút ngẫu nhiên 5 quân bài từ cỗ bài tú lơ khơ. Tính xác suất trong các trường hợp sau:

a. Có 4 quân Át

b. 4 quân Át và 1quân K

c. 3 quân 10 và 2 quân J

d. 10, J, Q, K và Át

e. Đồng chất

f. 10, J, Q, K, Át và đồng chất

g. Có ít nhất 1 quân Át.

1.23 Tính xác suất rút lần lượt được 3 quân Át từ cỗ bài tú lơ khơ trong hai trường hợp sau:

a. Rút có hoàn lại

b. Rút không hoàn lại.

1.24 Một cỗ máy sản xuất 12.000 sản phẩm trong một ngày với tỉ lệ phế phẩm trung bình 3%. Chọn ngẫu nhiên 600 sản phẩm để kiểm tra, tính xác suất có 12 phế phẩm trong 600 sản phẩm này.

1.25 Để được nhập kho, sản phẩm của nhà máy phải qua 3 vòng kiểm tra chất lượng độc lập nhau. Xác suất phát hiện ra phế phẩm ở các vòng lần lượt theo thứ tự là 0,8; 0,9 và 0,99. Tính xác suất phế phẩm được nhập kho.

1.26 Một thủ kho có một chùm chìa khóa gồm 9 chiếc trông giống hệt nhau trong đó chỉ có một chiếc mở được kho. Anh ta thử ngẫu nhiên từng chìa khóa một, chiếc nào thử không đúng thì loại ra. Tính xác suất anh ta mở được cửa ở lần thử thứ 4.

1.27 Một lô hàng có 9 sản phẩm. Mỗi lần kiểm tra chất lượng lấy ngẫu nhiên 3 sản phẩm. Sau khi kiểm tra xong trả lại vào lô hàng. Tính xác suất để sau 3 lần kiểm tra lô hàng, tất cả các sản phẩm đều được kiểm tra.

1.28 Hai biến cố A, B có xác suất $P(A)=0,3$, $P(A\cup B)=0,65$. Giả sử A, B độc lập nhưng không xung khắc. Tính $P(B)$.

1.29 Giả sử hai biến cố A, B có xác suất $P(A)=1/2$, $P(B)=1/3$ và $P(A\cap B)=1/4$. Hãy tính

a. $P(A|B)$ **b.** $P(B|A)$ **c.** $P(A\cup B)$ **d.** $P(A\cap\overline{B})$

e. $P(\overline{A}\cap B)$ **f.** $P(\overline{B}|\overline{A})$ **g.** $P(\overline{A}|\overline{B})$ **h.** $P(\overline{A}\cap\overline{B})$.

1.30 Chọn ngẫu nhiên lần lượt không hoàn lại 2 số từ các số $\{0,1,...,9\}$. Tính xác suất số thứ hai chọn được là số 4.

1.31 Một lô hàng có 4 sản phẩm loại I và 5 sản phẩm loại II. Mỗi người lấy ngẫu nhiên 1 sản phẩm và không hoàn lại. Biết rằng người thứ hai lấy được sản phẩm loại I, tính xác suất người thứ nhất cũng lấy được sản phẩm loại I.

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

a. Tìm tỷ lệ phế phẩm chung của nhà máy.

b. Lấy ngẫu nhiên một sản phẩm kiểm tra và được sản phẩm là phế phẩm. Tính xác suất để phế phẩm đó là do phân xưởng I, II, III sản xuất.

1.33 Có bốn nhóm xạ thủ tập bắn. Nhóm thứ nhất có 5 người, nhóm thứ hai có 7 người, nhóm thứ ba có 4 người và nhóm thứ tư có 2 người. Xác suất bắn trúng đích của mỗi người trong nhóm thứ nhất, nhóm thứ hai, nhóm thứ ba và nhóm thứ tư theo thứ tự là 0,8; 0,7; 0,6 và 0,5. Chọn ngẫu nhiên một xạ thủ và biết rằng xạ thủ này bắn trượt. Hãy xác định xem xạ thủ này có khả năng ở trong nhóm nào nhất.

1.34 Bắn hai lần độc lập với nhau mỗi lần một viên đạn vào cùng một bia. Xác suất trúng đích của viên đạn thứ nhất là 0,7 và của viên đạn thứ hai là 0,4. Tìm xác suất để chỉ có một viên đạn trúng bia (biến cố A). Sau khi bắn, quan trắc viên báo có một vết đạn ở bia. Tìm xác suất để vết đạn đó là vết đạn của viên đạn thứ nhất.

1.35 Một trạm chỉ phát hai tín hiệu A và B với xác suất tương ứng 0.85 và 0.15 . Do có nhiễu trên đường truyền nên $1/7$ tín hiệu A bị méo và thu được như tín hiệu B còn $1/8$ tín hiệu B bị méo và thu được như A.

a. Tìm xác suất thu được tín hiệu A.

b. Giả sử đã thu được tín hiệu A. Tìm xác suất thu được đúng tín hiệu lúc phát.

1.36 Lô hàng có 3 sản phẩm loại I và 2 sản phẩm loại II, lô hàng thứ hai có 2 sản phẩm loại I và 5 sản phẩm loại II. Lấy ngẫu nhiên một sản phẩm từ một trong hai lô hàng này và đó là sản phẩm loại I. Tính xác suất sản phẩm loại I nhận được là từ lô hàng thứ nhất.

1.37 Một nhà máy sản xuất một chi tiết của điện thoại di động có tỷ lệ sản phẩm đạt tiêu chuẩn chất lượng là 85%. Trước khi xuất xưởng người ta dùng một thiết bị kiểm tra để kết luận sản phẩm có đạt yêu cầu chất lượng hay không. Thiết bị có khả năng phát hiện đúng sản phẩm đạt tiêu chuẩn với xác suất là 0,9 và phát hiện đúng sản phẩm không đạt tiêu chuẩn với xác suất là 0,95. Tìm xác suất để 1 sản phẩm được chọn ngẫu nhiên sau khi kiểm tra:

a. Được kết luận là đạt tiêu chuẩn.

b. Được kết luận là đạt tiêu chuẩn thì lại không đạt tiêu chuẩn.

c. Được kết luận đúng với thực chất của nó.

CHƯƠNG 2: BIẾN NGẪU NHIÊN

Trong chương này ta khảo sát các biến cố của đại lượng nhận các giá trị nào đó, khi các giá trị này thay đổi ta được biến ngẫu nhiên. Khái niệm biến ngẫu nhiên (còn được gọi là đại lượng ngẫu nhiên) và các đặc trưng của chúng là những khái niệm rất quan trọng của lý thuyết xác suất.

Đối với biến ngẫu nhiên ta chỉ quan tâm đến vấn đề biến ngẫu nhiên này nhận một giá trị nào đó hoặc nhận giá trị trong một khoảng với xác suất bao nhiêu. Các biến ngẫu nhiên trong các phép thử khác nhau có thể có các phân bố xác suất như nhau, nghĩa là cùng quy luật phân bố xác suất. Phân bố xác suất của biến ngẫu nhiên X có thể được khảo sát thông qua hàm phân bố xác suất $F_X(x) = P\{X \le x\}$. Khi ta biết qui luật phân bố xác suất của một biến ngẫu nhiên thì ta có thể tính các xác suất liên quan đến biến ngẫu nhiên này.

Trường hợp biến ngẫu nhiên chỉ nhận các giá trị rời rạc thì hàm phân bố xác suất hoàn toàn được xác định bởi hàm khối lượng xác suất hoặc bảng phân bố xác suất, đó là bảng ghi các giá trị mà biến ngẫu nhiên nhận với xác suất khác không tương ứng. Đối với biến ngẫu nhiên nhận giá trị liên tục thì hàm phân bố xác suất có thể được xác định bởi hàm mật độ xác suất.

Một số quy luật phân bố xác suất quan trọng sau:

- **Phân bố Bernoulli.**

Phân bố này thường gặp trong các bài toán xét sự xuất hiện của biến cố A nào đó trong phép thử mà xác suất xuất hiện là p. Trong thống kê ta sử dụng biến ngẫu nhiên này để biểu diễn dấu hiệu nghiên cứu có tính định tính trong đó tham số p là tần suất của tổng thể, ví dụ tỉ lệ cử tri bỏ phiếu cho ứng cử viên A nào đó, tỉ lệ phế phẩm của một lô sản phẩm, tỉ lệ nẩy mầm của lô hạt giống ….

- **Phân bố nhị thức**

Phân bố này thường gặp trong dãy phép thử Bernoulli, là tổng của các biến ngẫu nhiên độc lập có phân bố Bernoulli.

- **Phân bố Poisson**

Phân bố này thường gặp trong bài toán về quá trình đếm sự xuất hiện biến cố A nào đó trong một khoảng thời gian xác định. Chẳng hạn: số cuộc gọi đến một tổng đài, số khách hàng đến 1 điểm phục vụ, số tai nạn (xe cộ), số các sự cố xảy ra ở một địa điểm … trong một khoảng thời gian xác định nào đó.

- **Phân bố đều**

Phân bố đều trong một khoảng là phân bố xác suất của biến ngẫu nhiên liên tục và đồng khả năng lấy giá trị trong khoảng đó. Phân bố đều có ứng dụng rộng trong thống kê toán. Nó có ý nghĩa to lớn trong các bài toán sử dụng phương pháp phi tham số.

- **Phân bố chuẩn**

Phân bố chuẩn thường được gặp trong các bài toán về sai số khi đo đạc các đại lượng trong vật lý, thiên văn ... Trong thực tế, nhiều biến ngẫu nhiên tuân theo phân bố chuẩn hoặc tiệm cận chuẩn (định lý giới hạn trung tâm) chẳng hạn: trọng lượng, chiều cao của một nhóm người nào đó, điểm thi của thí sinh, năng suất cây trồng, mức lãi suất của một công ty, nhu cầu tiêu thụ của một mặt hàng nào đó ...

- **Phân bố "khi bình phương"**

Phân bố "khi bình phương" được ứng dụng để kiểm định các giả thiết về xác suất của hệ đầy đủ.

- **Phân bố Student**

Trong thống kê phân bố chuẩn và phân bố Student được ứng dụng để giải quyết các bài toán về giá trị trung bình.

Ngoài phương pháp sử dụng hàm phân bố để xác định biến ngẫu nhiên, trong nhiều trường hợp bài toán chỉ cần đòi hỏi khảo sát những đặc trưng cơ bản của biến ngẫu nhiên.

Các đặc trưng của biến ngẫu nhiên được chia thành hai loại sau:

❖ Các đặc trưng cho vị trí trung tâm, giá trị trung bình của biến ngẫu nhiên như: Kỳ vọng, Trung vị, Mốt.

❖ Các đặc trưng cho độ phân tán của biến ngẫu nhiên như: Phương sai, Độ lệch chuẩn, Hệ số biến thiên, Hệ số bất đối xứng và Hệ số nhọn.

Trong các bài toán thực tế kỳ vọng được sử dụng dưới dạng lợi nhuận kỳ vọng còn phương sai để tính mức độ rủi ro của quyết định. Trong kỹ thuật độ lệch chuẩn biểu diễn sai số của phép đo.

Để học tốt chương này học viên phải nắm vững định nghĩa xác suất, biến cố và các tính chất của chúng đã được học ở chương 1.

Các đặc trưng của biến ngẫu nhiên được xác định thông qua tính tổng của các số hạng nào đó (trường hợp biến ngẫu nhiên rời rạc) hoặc tính tích phân xác định (trường hợp biến ngẫu nhiên liên tục). Vì vậy học viên cần ôn tập về tổng của chuỗi và tích phân xác định.

2.1 ĐỊNH NGHĨA VÀ PHÂN LOẠI BIẾN NGẪU NHIÊN

Gieo một con xúc xắc 6 mặt. Ký hiệu A_1, A_2, A_3, A_4, A_5, A_6 lần lượt là biến cố "mặt 1 chấm xuất hiện", "mặt 2 chấm xuất hiện", …, "mặt 6 chấm xuất hiện".

Thay vì xét các biến cố như trên, ta xét đại lượng X là số chấm xuất hiện khi gieo con xúc xắc. Khi đó X có thể nhận các giá trị 1, 2, 3, 4, 5, 6 một cách ngẫu nhiên.

X nhận giá trị k là biến cố A_k, nghĩa là

$$\{X = k\} = A_k \text{, với } k = 1, 2, ..., 6 .$$

Ta gọi X là một biến ngẫu nhiên có miền giá trị $R_X = \{1, 2, ..., 6\}$.

Một cách tổng quát ta có khái niệm biến ngẫu nhiên như sau.

2.1.1 Định nghĩa biến ngẫu nhiên

Định nghĩa 2.1: *Biến ngẫu nhiên X là đại lượng nhận các giá trị nào đó phụ thuộc vào các* *yếu tố ngẫu nhiên. Đặc biệt với mọi giá trị thực* $x \in \mathbb{R}$: *" X nhận giá trị nhỏ hơn bằng* x *", ký hiệu* $\{X \le x\}$, *là một biến cố ngẫu nhiên.*

Đối với biến ngẫu nhiên người ta chủ yếu quan tâm xem nó nhận một giá trị nào đó hoặc nhận giá trị trong một khoảng với một xác suất bao nhiêu.

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

Ví dụ 2.1: Các đại lượng sau là biến ngẫu nhiên

- Tuổi thọ của một thiết bị đang hoạt động.
- Số khách hàng vào một điểm phục vụ trong khoảng thời gian nào đó.
- Số cuộc gọi đến một tổng đài trong khoảng thời gian nào đó.
- Sai số khi đo lường một đại lượng vật lý …

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

2.1.2 Phân loại

Để dễ dàng hơn trong việc nghiên cứu người ta chia các biến ngẫu nhiên thành hai loại:

❖ ***Biến ngẫu nhiên rời rạc***

Biến ngẫu nhiên X *là biến ngẫu nhiên rời rạc* nếu nó chỉ nhận một số hữu hạn hoặc vô hạn đếm được các giá trị. Nghĩa là các giá trị của miền giá trị R_X có thể liệt kê thành một dãy dạng $x_1, x_2, ...$.

❖ ***Biến ngẫu nhiên liên tục***

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

Ví dụ 2.2:

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

2.2 PHÂN BỐ XÁC SUẤT CỦA BIẾN NGẪU NHIÊN

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

Chẳng hạn xác suất bắn trúng bia của một xạ thủ là 0,8. Xạ thủ này bắn 10 viên, gọi X là số viên bắn trúng bia thì xác suất để xạ thủ bắn trúng k viên là

$$P\{X=k\}=C_{10}^{k}(0,8)^{k}(0,2)^{10-k},\ 0\le k\le 10 \text{ (xem dãy phép thử Bernoulli 1.5).}$$

Tương tự, giả sử tỷ lệ chính phẩm của lô hàng là 0,8. Chọn 10 sản phẩm kiểm tra, gọi Y là số chính phẩm phát hiện được thì xác suất chọn được k chính phẩm là

$$P\{Y=k\}=C_{10}^{k}(0,8)^{k}(0,2)^{10-k},\ 0\le k\le 10.$$

Vậy $P\{X=k\}=P\{Y=k\}$, với mọi $k=0,1,\dots,10$.

Nói cách khác quy luật phân bố xác suất của X và Y như nhau, mặc dù X và Y là hai biến ngẫu nhiên được xét trong hai phép thử khác nhau.

Quy luật phân bố xác suất được nghiên cứu thông qua hàm phân bố xác suất định nghĩa như sau.

2.2.1 Hàm phân bố xác suất

Định nghĩa 2.3: *Hàm phân bố xác suất (cumulative distribution function, viết tắt CDF) của biến ngẫu nhiên* X *là hàm số* $F_X(x)$ *xác định với mọi* $x\in\mathbb{R}$ *bởi công thức*:

$$F_X(x)=P\{X\le x\};\ -\infty<x<\infty \qquad \textbf{(2.1)}$$

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

Hàm phân bố có các tính chất sau:

a. $0\le F_X(x)\le 1$ với mọi $x\in\mathbb{R}$, **(2.2)**

b. $F_X(x)$ là hàm không giảm, liên tục bên phải.

$$F_X(a^+)=F_X(a) \text{ với } F_X(a^+)=\lim_{x>a,x\to a}F_X(x) \qquad \textbf{(2.3)}$$

Nếu X là biến ngẫu nhiên liên tục thì $F_X(x)$ là hàm liên tục.

c. $F_X(-\infty)=\lim\limits_{x\to-\infty}F_X(x)=0;\ F_X(+\infty)=\lim\limits_{x\to+\infty}F_X(x)=1$, **(2.4)**

d. $P\{a<X\le b\}=F_X(b)-F_X(a)$. **(2.5)**

e. $P\{X>a\}=1-F_X(a)$; $P\{X<a\}=F_X(a^-)$ với $F_X(a^-)=\lim\limits_{x<a,x\to a}F_X(x)$ **(2.6)**

Nhận xét 2.1: Một số tài liệu coi $G_X(x)=P\{X<x\};\ -\infty<x<\infty$ là hàm phân bố của biến ngẫu nhiên X. Mỗi cách định nghĩa có thuận lợi riêng, tuy nhiên trong giáo trình này ta sử dụng hàm phân bố $F_X(x)$ của biến ngẫu nhiên X theo công thức (2.1).

Có một số khác biệt giữa hai định nghĩa này. Chẳng hạn tính chất liên tục phải của $F_X(x)$ được thay bằng liên tục trái của $G_X(x)$, công thức (2.5) sẽ là

$$P\{a\le X<b\}=G_X(b)-G_X(a)\dots$$

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

Biến ngẫu nhiên rời rạc có miền giá trị là một tập hữu hạn hoặc vô hạn đếm được. Phân bố xác suất của biến ngẫu nhiên chỉ tập trung tại các giá trị thuộc miền giá trị, ta gọi là hàm khối lượng xác suất của biến ngẫu nhiên và được định nghĩa như sau.

Giả sử biến ngẫu nhiên rời rạc X có miền giá trị $R_X = \{x_1, x_2, ...\}$.

Đặt

$$p_X(x) = P\{X = x\} \tag{2.7}$$

Hàm $p_X(x)$ được *gọi là **hàm khối lượng xác suất** (probability mass function, viết tắt PMF) của biến ngẫu nhiên rời rạc* X.

Tính chất của hàm khối lượng xác suất $p_X(x)$:

- $p_X(x_k) > 0$ với mọi $x_k \in R_X$ **(2.8)**
- $\sum\limits_{x_k \in R_X} p_X(x_k) = 1$ **(2.9)**
- $p_X(x) = 0$ với mọi $x \notin R_X$ **(2.10)**

Hàm phân bố của X được xác định từ hàm khối lượng xác suất theo công thức:

$$F_X(x) = P\{X \le x\} = \sum_{x_k \le x;\, x_k \in R_X} p_X(x_k) \tag{2.11}$$

- Nếu biến ngẫu nhiên rời rạc X nhận vô hạn các giá trị $x_1, x_2, ...$ thì hàm phân bố xác suất có dạng:

$$F_X(x) = \begin{cases} 0 & \text{nÕu} \quad x < x_1 \\ p_X(x_1) + \cdots + p_X(x_{k-1}) & \text{nÕu} \quad x_{k-1} \le x < x_k\,,\ \forall k > 1 \end{cases} \tag{2.12}$$

- Nếu X chỉ nhận một số hữu hạn các giá trị $x_1, x_2, ..., x_n$ thì các biến cố

$$\{X = x_1\}, \{X = x_2\}, ..., \{X = x_n\} \tag{2.13}$$

lập thành hệ đầy đủ các biến cố và hàm phân bố xác suất có dạng:

$$F_X(x) = \begin{cases} 0 & \text{nÕu} \quad x < x_1 \\ p_X(x_1) + \cdots + p_X(x_{k-1}) & \text{nÕu} \quad x_{k-1} \le x < x_k \\ 1 & \text{nÕu} \quad x \ge x_n \end{cases} \tag{2.14}$$

Để trực quan hơn chúng ta biểu diễn hàm khối lượng xác suất của biến ngẫu nhiên rời rạc thông qua ***bảng phân bố xác suất***. Đó là bảng có hai hàng, hàng trên ghi các giá trị của miền giá trị R_X theo thứ tự tăng dần, hàng dưới là giá trị của hàm khối lượng xác suất tương ứng.

Bảng phân bố xác suất của X có dạng sau:

X	x_1	x_2	…
P	$p_X(x_1)$	$p_X(x_2)$	…

Ví dụ 2.3: Xét phép thử tung đồng thời 2 đồng xu (ví dụ 1.1). Không gian mẫu của phép thử là

$\Omega = \{(S,S),(S,N),(N,S),(N,N)\}$ gồm 4 kết quả đồng khả năng. Gọi X là số mặt sấp xuất hiện, khi đó X là một biến ngẫu nhiên rời rạc.

Hàm khối lượng xác suất

$$p_X(x) = \begin{cases} 0 & \text{nếu } \; x \neq 0,1,2 \\ 1/4 & \text{nếu } \; x = 0 \text{ hoặc } x = 2 \\ 1/2 & \text{nếu } \; x = 1 \end{cases}$$

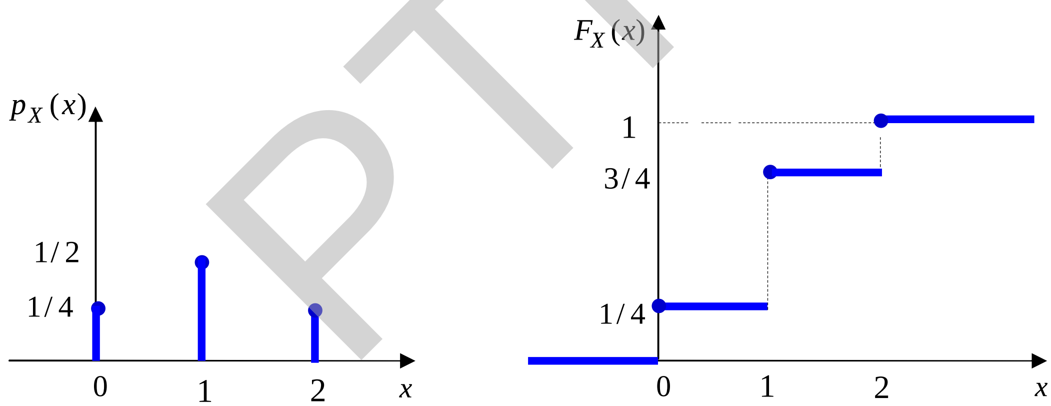
Bảng phân bố xác suất

X	0	1	2
P	$1/4$	$2/4$	$1/4$

Hàm phân bố xác suất

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

Đồ thị của hàm khối lượng xác suất và hàm phân bố



(*a*) (*b*)

Hình 2.1

Ví dụ 2.4: Chọn ngẫu nhiên 3 bi từ một túi có 6 bi đen, 4 bi trắng. Gọi X là số bi trắng trong 3 bi vừa chọn thì X là một biến ngẫu nhiên rời rạc. Tìm bảng phân bố xác suất và hàm phân bố xác suất.

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

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

Hàm khối lượng xác suất

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

$$p_X(x)=0 \text{ với mọi } x \text{ khác } 0,1,2,3.$$

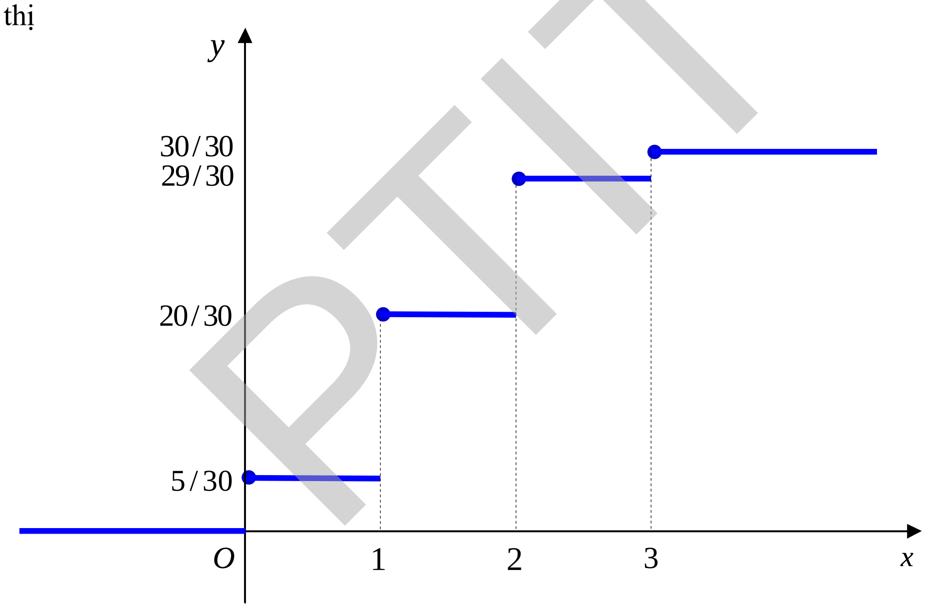
Bảng phân bố xác suất

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

Hàm phân bố xác suất

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

Đồ thị



Hình 2.2

Ví dụ 2.5: *Mô hình giá trị tài sản nhị phân (Binomial Asset Pricing Model)*

Mô hình giá trị tài sản nhị phân là một công cụ hữu ích để hiểu được lý thuyết giá cả chứng khoán và lý thuyết xác suất. Chúng ta xét mô hình với thời gian rời rạc và ở mỗi bước chuyển, giá cổ phiếu thay đổi theo một trong hai giá trị có thể. Giả sử khởi điểm ta có giá trị S_0 và hai hằng số dương d và u thỏa mãn

$$0<d<u$$

Khi đó trong chu kỳ tiếp giá cổ phiếu sẽ là dS_0 hoặc uS_0. Thông thường chúng ta chọn d và u thỏa mãn $0<d<1<u$. Như vậy giá cổ phiếu thay đổi từ S_0 thành dS_0 biểu thị

sự vận động đi xuống và chuyển từ S_0 thành uS_0 biểu thị sự vận động đi lên.

Chẳng hạn xét trường hợp cụ thể $d = \dfrac{1}{u}$.

Trong thực tế sự vận động đi lên hoặc đi xuống không thể biết trước được và ta coi đó là hiện tượng ngẫu nhiên. Để mô phỏng mô hình này ta gieo đồng xu, nếu mặt sấp xuất hiện (biến cố S) ta được giá cổ phiếu vận động đi lên và nếu mặt ngửa xuất hiện (biến cố N) thì giá cổ phiếu giảm xuống.

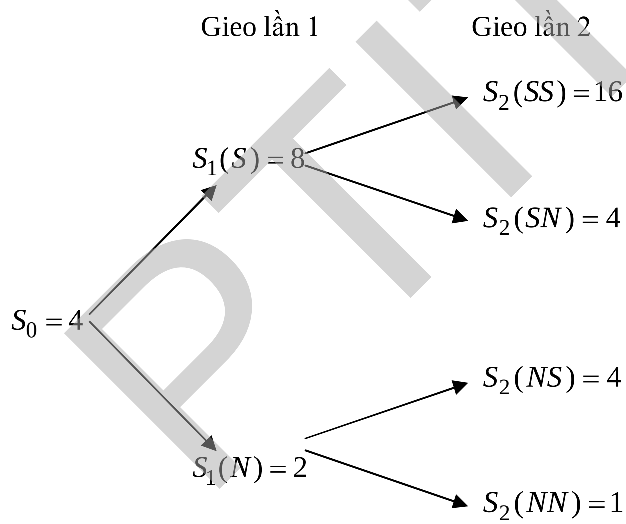
Ta ký hiệu giá cổ phiếu tại thời điểm 1 là S_1.

$$\text{Vậy } S_1(S) = uS_0 \text{ và } S_1(N) = dS_0 .$$

Tương tự, ta ký hiệu giá cổ phiếu tại thời điểm 2 là S_2:

$$S_2(SS) = uS_1(S) = u^2 S_0 , \qquad S_2(SN) = dS_1(S) = duS_0 ,$$

$$S_2(NS) = uS_1(N) = udS_0 , \qquad S_2(NN) = dS_1(N) = d^2 S_0 ,$$



Hình 2.3: *Cây nhị phân của giá cổ phiếu với* $S_0 = 4$, $u = 1/d = 2$

Tiếp tục ký hiệu giá cổ phiếu tại thời điểm 3 là S_3:

$$S_3(SSS) = uS_2(SS) = u^3 S_0 , \qquad S_3(SSN) = dS_2(SS) = du^2 S_0 ,$$

$$S_3(SNS) = uS_2(SN) = u^2 dS_0 , \qquad S_3(SNN) = dS_2(SN) = d^2 uS_0 ,$$

$$S_3(NSS) = uS_2(NS) = u^2 dS_0 , \qquad S_3(NSN) = dS_2(NS) = d^2 uS_0 ,$$

$$S_3(NNS) = uS_2(NN) = ud^2 S_0 , \qquad S_3(NNN) = dS_2(NN) = d^3 S_0 .$$

Như vậy S_1, S_2, S_3, … là các biến ngẫu nhiên.

Với trường hợp $S_0 = 4$, $u = 1/d = 2$ ta có các bảng phân bố xác suất của S_1, S_2, S_3

S_1	2	8
P	1/2	1/2

S_2	1	4	16
P	1/4	2/4	1/4

S_3	1/2	2	8	32
P	1/8	3/8	3/8	1/8

2.2.3 Hàm mật độ phân bố xác suất của biến ngẫu nhiên liên tục

Công thức (2.11) cho thấy hàm phân bố xác suất của biến ngẫu nhiên rời rạc có thể được xác định qua hàm khối lượng xác suất. Tuy nhiên xác suất của biến ngẫu nhiên liên tục lấy giá trị tại từng điểm bằng 0. Vì vậy đối với biến ngẫu nhiên liên tục ta không thể xét hàm khối lượng xác suất mà được thay bằng hàm mật độ xác suất. Dấu tổng của công thức 2.12 xác định hàm phân bố từ hàm khối lượng xác suất được chuyển thành dấu tích phân của hàm mật độ xác suất trong định nghĩa sau.

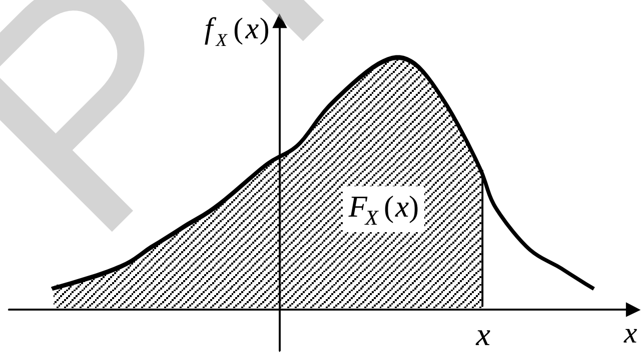
Định nghĩa 2.4: *Giả sử X là một biến ngẫu nhiên liên tục có hàm phân bố xác suất $F_X(x)$, nếu tồn tại hàm $f_X(x)$ sao cho*

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt \text{, với mọi } x \in \mathbb{R} \quad \textbf{(2.15)}$$

thì $f_X(x)$ được gọi là ***hàm mật độ xác suất*** *của biến ngẫu nhiên X (probability density function, viết tắt PDF).*

Như vậy giá trị của hàm $F_X(x)$ bằng diện tích hình phẳng giới hạn bởi đồ thị hàm mật độ xác suất $f_X(x)$, trục hoành và đường thẳng song song với trục tung có hoành độ là x.

Hàm phân bố $F_X(x)$ là một nguyên hàm của hàm mật độ $f_X(x)$.



Hình 2.4

Tính chất của hàm mật độ xác suất $f_X(x)$

1. $\dfrac{d}{dx}F_X(x) = f_X(x)$ tại các điểm x mà $f_X(x)$ liên tục. **(2.16)**

2. $f_X(x) \ge 0$ với mọi $x \in \mathbb{R}$, **(2.17)**

3. $\displaystyle\int_{-\infty}^{\infty} f_X(x)dx = 1$, **(2.18)**

4. $P\{a < X < b\} = P\{a \le X \le b\} = P\{a \le X < b\} = P\{a < X \le b\} = \displaystyle\int_a^b f_X(x)dx$. **(2.19)**

Nhận xét 2.2: Công thức 2.15 chỉ ra rằng hàm phân bố $F_X(x)$ là một nguyên hàm của hàm mật độ $f_X(x)$, vì vậy khi biết hàm mật độ ta có thể tìm hàm phân bố bằng cách tính tích phân. Ngược lại công thức 2.16 cho phép tìm hàm mật độ xác suất từ hàm phân bố xác suất bằng cách lấy đạo hàm của hàm phân bố xác suất. Các ví dụ sau minh họa điều này.

Ví dụ 2.6: Hàm phân bố xác suất của biến ngẫu nhiên liên tục X có dạng

$$F_X(x)=\begin{cases}0 & \text{víi } x<0\\ kx^2 & \text{víi } 0\le x<1\\ 1 & \text{víi } x\ge 1\end{cases}$$

a. Xác định hệ số k;

b. Tìm hàm mật độ xác suất $f_X(x)$.

Giải: **a.** Vì biến ngẫu nhiên X liên tục do đó hàm phân bố xác suất $F_X(x)$ cũng liên tục.

Xét tính liên tục của $F_X(x)$ tại $x=1$

$$\begin{cases}\lim\limits_{x\to 1^-} F_X(x)=\lim\limits_{x\to 1^-} kx^2=k\\ F_X(1)=1\end{cases} \Rightarrow k=1.$$

b. Theo công thức (2.16) của hàm mật độ xác suất ta có

$$f_X(x)=\frac{d}{dx}F_X(x)=\begin{cases}0 & \text{víi } x\le 0\\ 2x & \text{víi } 0<x<1\\ 0 & \text{víi } x\ge 1\end{cases}$$

Ví dụ 2.7: Biến ngẫu nhiên liên tục X với hàm mật độ xác suất có dạng

$$f_X(x)=\begin{cases}0 & \text{víi } x<1\\ \dfrac{k}{x^2} & \text{víi } x\ge 1\end{cases}$$

Hãy xác định:

a. Hệ số k;

b. Hàm phân bố xác suất $F_X(x)$;

c. Tính xác suất $P\{2<X<3\}$;

d. Thực hiện 4 lần phép thử độc lập biến ngẫu nhiên X, tính xác suất biến ngẫu nhiên X không lấy giá trị trong khoảng $(2;\,3)$.

Giải: **a.** Theo tính chất (2.18) ta có $1=\int\limits_{-\infty}^{\infty} f_X(x)dx=\int\limits_{1}^{\infty}\frac{k}{x^2}dx=-\lim\limits_{a\to\infty}\left(\frac{k}{x}\Big|_1^a\right)=k$, từ đó $k=1$.

b. Từ công thức (2.15) ta có

$$F_X(x)=\int\limits_{-\infty}^{x} f_X(t)dt=\begin{cases}0 & \text{víi } x<1\\ 1-\dfrac{1}{x} & \text{víi } x\ge 1\end{cases}$$

c. Từ công thức (2.19) và (2.5) ta có

$$P\{2 < X < 3\} = F_X(3) - F_X(2) = \left(1 - \frac{1}{3}\right) - \left(1 - \frac{1}{2}\right) = \frac{1}{2} - \frac{1}{3} = \frac{1}{6}.$$

d. Theo kết quả c. ta suy ra xác suất để X không lấy giá trị trong khoảng $(2;3)$ trong một phép thử bằng $1 - \frac{1}{6} = \frac{5}{6}$.

Vậy xác suất để trong 4 phép thử độc lập biến ngẫu nhiên X đều không lấy giá trị trong khoảng $(2;3)$ bằng $\left(\frac{5}{6}\right)^4 \approx 0{,}48$.

2.3 CÁC THAM SỐ ĐẶC TRƯNG CỦA BIẾN NGẪU NHIÊN

Ngoài phương pháp sử dụng hàm phân bố để xác định biến ngẫu nhiên, trong nhiều trường hợp bài toán chỉ cần đòi hỏi khảo sát những đặc trưng cơ bản của biến ngẫu nhiên như: Kỳ vọng, phương sai, độ lệch chuẩn, phân vị, trung vị …

2.3.1 Kỳ vọng

2.3.1.1 Định nghĩa

Với mọi biến ngẫu nhiên X ta ký hiệu $\mathrm{E}\,X$ hoặc $\mathrm{E}(X)$ hoặc $\mathrm{E}[X]$ và xác định như sau:

(i) Trường hợp X rời rạc với miền giá trị R_X với hàm khối lượng xác suất $p_X(x_k)$

$$\mathrm{E}\,X = \sum_{x_k \in R_X} x_k p_X(x_k) \qquad \textbf{(2.20)}$$

(ii) Trường hợp X liên tục có hàm mật độ xác suất $f_X(x)$ thì

$$\mathrm{E}\,X = \int_{-\infty}^{\infty} x f_X(x) dx \qquad \textbf{(2.21)}$$

Nếu chuỗi (2.20) hội tụ tuyệt đối (trường hợp X rời rạc) hoặc tích phân (2.21) hội tụ tuyệt đối (trường hợp X liên tục) thì ta gọi $\mathrm{E}\,X$ *là kỳ vọng của biến ngẫu nhiên* X (expected value), trường hợp ngược lại ta nói X không tồn tại kỳ vọng.

Kỳ vọng mang ý nghĩa là giá trị trung bình (average, mean value) của biến ngẫu nhiên X.

Ví dụ 2.8: Tính kỳ vọng của biến ngẫu nhiên X cho ở ví dụ 2.4.

Giải: $\mathrm{E}\,X = 0 \cdot \frac{5}{30} + 1 \cdot \frac{15}{30} + 2 \cdot \frac{9}{30} + 3 \cdot \frac{1}{30} = \frac{6}{5}$.

Ví dụ 2.9: Theo thống kê việc một thanh niên 25 tuổi sẽ sống thêm trên một năm có xác suất là 0,992, xác suất để người đó chết trong vòng một năm tới là 0,008 (xem ví dụ 1.10). Một chương trình bảo hiểm kinh doanh bảo hiểm sinh mạng trong 1 năm cho thanh niên độ tuổi 25 với số tiền chi trả 1000 đô la, tiền mua bảo hiểm là 10 đô la. Hỏi lợi nhuận trung bình của công ty bảo hiểm nhận được trên mỗi khách hàng là bao nhiêu?

Giải: Rõ ràng lợi nhuận là biến ngẫu nhiên X với 2 giá trị là $+10$ đô la (nếu người mua bảo hiểm không chết) và -990 đô la (nếu người mua bảo hiểm chết). Bảng phân bố xác suất tương ứng:

X	-990	$+10$
P	$0,008$	$0,992$

Do đó kỳ vọng $\mathrm{E}\,X=(-990)\cdot 0,008+10\cdot 0,992=2$ đô la.

Ta thấy lợi nhuận trung bình là một số dương vì vậy công ty bảo hiểm có thể làm ăn có lãi.

Ví dụ 2.10: Kỳ vọng của các biến ngẫu nhiên S_1, S_2, S_3 của ví dụ 2.5

$\mathrm{E}\,S_1=2\cdot\frac{1}{2}+8\cdot\frac{1}{2}=5$; $\mathrm{E}\,S_2=1\cdot\frac{1}{4}+4\cdot\frac{2}{4}+16\cdot\frac{1}{4}=\frac{25}{4}$; $\mathrm{E}\,S_3=\frac{1}{2}\cdot\frac{1}{8}+2\cdot\frac{3}{8}+8\cdot\frac{3}{8}+32\cdot\frac{1}{8}=\frac{125}{16}$.

Ví dụ 2.11: Tuổi thọ của một loại côn trùng nào đó là một biến ngẫu nhiên X (đơn vị là tháng) với hàm mật độ xác suất như sau:

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

Tìm hàm phân bố xác suất và tìm tuổi thọ trung bình của loài côn trùng trên.

Giải: Vì $\int\limits_0^4 x^2(4-x)dx=\left.4\frac{x^3}{3}-\frac{x^4}{4}\right|_0^4=\frac{64}{3}\;\Rightarrow\; k=\frac{3}{64}$. Vậy hàm phân bố xác suất sẽ là

$$F_X(x)=\int\limits_{-\infty}^{x} f_X(t)dt=\begin{cases} 0 & \text{nÕu } x\le 0\\ \dfrac{3x^3}{64}\left(\dfrac{4}{3}-\dfrac{x}{4}\right) & \text{nÕu } 0<x\le 4\\ 1 & \text{nÕu } x>4\end{cases}$$

Tuổi thọ trung bình:

$$\mathrm{E}\,X=\frac{3}{64}\int\limits_0^4 x^3(4-x)dx=\left.\frac{3}{64}\left(x^4-\frac{x^5}{5}\right)\right|_0^4=\frac{12}{5}\text{ (tháng).}$$

2.3.1.2 Ý nghĩa của kỳ vọng

Kỳ vọng mang ý nghĩa là giá trị trung bình mà biến ngẫu nhiên nhận được. Ta minh họa điều này trong trường hợp rời rạc sau.

Giả sử biến ngẫu nhiên rời rạc X nhận các giá trị $x_1,x_2,...,x_m$ với các tần số tương ứng $r_1,r_2,...,r_m$. Khi đó

r_ix_i là tổng giá trị X nhận được ứng với giá trị x_i.

$r_1x_1+r_2x_2+\cdots+r_mx_m$ là tổng tất cả các giá trị X nhận được.

Vậy giá trị trung bình của X

$$\frac{r_1x_1+r_2x_2+\cdots+r_mx_m}{n}$$

trong đó $r_1 + r_2 + \cdots + r_m = n$.

Ký hiệu $f_i = \dfrac{r_i}{n}$ là tần suất nhận giá trị x_i của X.

Khi đó giá trị trung bình của X có thể viết lại

$$\frac{r_1x_1 + r_2x_2 + \cdots + r_mx_m}{n} = f_1x_1 + f_2x_2 + \cdots + f_mx_m.$$

Tần suất f_i cũng chính là xác suất trường hợp X nhận giá trị x_i.

Chẳng hạn một nhân viên có thu nhập trong một năm như sau

Tháng	1	2	3	4	5	6	7	8	9	10	11	12
Thu nhập (triệu đồng)	5	5	5	6	6	7	7	7	6	6	8	20

Như vậy nhân viên này có 3 tháng thu nhập 5 triệu, 4 tháng thu nhập 6 triệu, 3 tháng thu nhập 7 triệu, 1 tháng thu nhập 8 triệu và cuối năm được thưởng 20 triệu.

Vậy tổng thu nhập cả năm

$$S = (5\,\text{triÖu})\cdot 3 + (6\,\text{triÖu})\cdot 4 + (7\,\text{triÖu})\cdot 3 + (8\,\text{triÖu})\cdot 1 + (20\,\text{triÖu})\cdot 1$$

Thu nhập bình quân một tháng

$$\overline{S} = (5\,\text{triÖu})\cdot\frac{3}{12} + (6\,\text{triÖu})\cdot\frac{4}{12} + (7\,\text{triÖu})\cdot\frac{3}{12} + (8\,\text{triÖu})\cdot\frac{1}{12} + (20\,\text{triÖu})\cdot\frac{1}{12}.$$

Trường hợp biến ngẫu nhiên liên tục phép tính tổng của giá trị trung bình được thay bằng phép tính tích phân xác định.

Khái niệm kỳ vọng được áp dụng rộng rãi trong nhiều lĩnh vực. Trong kinh doanh và quản lý thì kỳ vọng được ứng dụng dưới dạng lợi nhuận kỳ vọng hay doanh số kỳ vọng.

Ví dụ 2.12: Giả sử một cửa hàng sách dự định nhập một số cuốn niên giám thống kê. Nhu cầu hàng năm về loại sách này được cho trong bảng phân bố xác suất sau:

Nhu cầu j (cuốn)	20	21	22	23	24	25
Xác suất P_j	0,30	0,25	0,18	0,14	0,10	0,03

Cửa hàng mua với giá 7 USD/cuốn bán với giá 10 USD/cuốn. Song đến cuối năm phải hạ giá bán hết với giá 4 USD/cuốn. Cửa hàng muốn xác định số lượng cần nhập sao cho lợi nhuận kỳ vọng lớn nhất.

Giải: Gọi i là số lượng sách dự định nhập, j là nhu cầu.

Lúc đó lợi nhuận có điều kiện tương ứng được xác định bởi:

$$\mathrm{E}_{ij} = \begin{cases} 10j - 7i + 4(i-j) & \text{nÕu } j < i \\ 10i - 7i & \text{nÕu } j \ge i \end{cases} = \begin{cases} 6j - 3i & \text{nÕu } j < i \\ 3i & \text{nÕu } j \ge i \end{cases}$$

Các giá trị E_{ij} được cho trong bảng sau:

Nhu cầu j

Lượng Hàng Nhập i

i \ j	20	21	22	23	24	25
20	60	60	60	60	60	60
21	57	63	63	63	63	63
22	54	60	66	66	66	66
23	51	57	63	69	69	69
24	48	54	60	66	72	72
25	45	51	57	63	69	75
P_j	0,3	0,25	0,18	0,14	0,10	0,03

Với mỗi số lượng nhập i, lợi nhuận trung bình được tính theo công thức $\mathrm{E}_i = \sum_j P_j \mathrm{E}_{ij}$.

Kết quả

Số lượng nhập i (cuốn)	20	21	22	23	24	25
Lợi nhuận kỳ vọng E_i	60	61,2	60,9	59,52	57,3	54,48

Như vậy nếu cửa hàng nhập 21 cuốn thì lợi nhuận trung bình sẽ cao nhất.

2.3.1.3 Tính chất kỳ vọng

1) $\mathrm{E}(C) = C$ với mọi hằng số C. **(2.22)**

2) $\mathrm{E}(CX) = C\mathrm{E}(X)$ với mọi hằng số C. **(2.23)**

3) $\mathrm{E}(X_1 + \cdots + X_n) = \mathrm{E}(X_1) + \cdots + \mathrm{E}(X_n)$ **(2.24)**

4) Cho hàm số $\varphi(x)$, xét biến ngẫu nhiên $Y = \varphi(X)$ thì

$$\mathrm{E}Y = \begin{cases} \sum\limits_{x_i \in R_X} \varphi(x_i) p_X(x_i) & \text{nếu } X \text{ rời rạc có hàm khối lượng } p_X(x_i) \\ \int\limits_{-\infty}^{\infty} \varphi(x) f(x) dx & \text{nÕu } X \text{ liªn tôc cã hµm mËt ®é } f_X(x) \end{cases} \quad \textbf{(2.25)}$$

Đặc biệt ta có các đẳng thức sau nếu tổng hoặc tích phân sau tương ứng hội tụ:

$$\mathrm{E}X^2 = \begin{cases} \sum\limits_{x_i \in R_X} x_i^2 p_X(x_i) & \text{nếu } X \text{ rời rạc có hàm khối lượng } p_X(x_i) \\ \int\limits_{-\infty}^{\infty} x^2 f(x) dx & \text{nÕu } X \text{ liªn tôc cã hµm mËt ®é } f_X(x) \end{cases} \quad \textbf{(2.26)}$$

5) Nếu $X_1, ..., X_n$ độc lập thì $\mathrm{E}(X_1 \cdots X_n) = \mathrm{E}(X_1) \cdots \mathrm{E}(X_n)$. **(2.27)**

Ví dụ 2.13: Chọn ngẫu nhiên 3 bi từ một túi có 6 bi đen, 4 bi trắng.

a. Nếu chọn được 1 bi trắng sẽ được thưởng 200\$. Gọi Y là số tiền nhận được. Tính kỳ vọng của Y.

b. Nếu chọn được 1 bi trắng sẽ được thưởng 200\$ và chọn được 1 bi đen sẽ được thưởng 300\$. Gọi Z là số tiền nhận được. Tính kỳ vọng của Z.

Giải: **a.** Gọi X là số bi trắng trong 3 bi vừa chọn (xem ví dụ 2.4) thì $Y = \varphi(X) = 200X$ là một biến ngẫu nhiên rời rạc có bảng phân bố sau:

$Y = \varphi(X)$	0	200	400	600
P	$5/30$	$15/30$	$9/30$	$1/30$

$$\mathrm{E}\,Y = 0.\frac{5}{30} + 200.\frac{15}{30} + 400.\frac{9}{30} + 600.\frac{1}{30} = 240\,\$.$$

Mặt khác, theo công thức (2.23) và ví dụ 2.8 ta cũng được

$$\mathrm{E}\,Y = 200\mathrm{E}\,X = 200.\frac{6}{5} = 240\,\$.$$

b. $Z = 200X + 300(3 - X) = 900 - 100X$.

Áp dụng công thức 2.22, 2.23, 2.24 ta được

$$\mathrm{E}\,Z = \mathrm{E}\left(900 - 100X\right) = 900 - 100\mathrm{E}X = 900 - 100\cdot\frac{6}{5} = 780\$.$$

2.3.2 Phương sai

2.3.2.1 Định nghĩa

Phương sai (variance) *hay độ lệch* (deviation) *bình phương trung bình của biến ngẫu nhiên* X *là đại lượng đo sự phân tán bình phương trung bình của* X *xung quanh giá trị trung bình* $\mathrm{E}\,X$. *Nói cách khác phương sai của* X *là kỳ vọng của* $\left(X - \mathrm{E}\,X\right)^2$.

Phương sai của X được ký hiệu $\mathrm{D}\,X$ hoặc $\mathrm{Var}\,X$ và xác định như sau

$$\mathrm{D}\,X = \mathrm{E}\left(X - \mathrm{E}\,X\right)^2 \tag{2.28}$$

$\sigma_X = \sqrt{\mathrm{D}X}$ được gọi là ***độ lệch chuẩn*** (deviation) của X. **(2.29)**

Ta có $\left(X - \mathrm{E}\,X\right)^2 = X^2 - \left(2\mathrm{E}\,X\right)X + \left(\mathrm{E}\,X\right)^2$

Áp dụng các công thức 2.22, 2.23, 2.24 của kỳ vọng ta có thể tính phương sai theo công thức sau:

$$\mathrm{D}\,X = \mathrm{E}\,X^2 - \left(\mathrm{E}\,X\right)^2 \tag{2.29}$$

Từ công thức (2.29) ta có công thức tính phương sai tương ứng với từng trường hợp rời rạc và liên tục:

(i). Nếu X rời rạc với miền giá trị R_X và hàm khối lượng xác suất $p_X(x_k)$ thì

$$\mathrm{E}\,X^2=\sum_{x_i\in R_X}x_i^2p_X(x_i)\,;\ \mathrm{D}\,X=\mathrm{E}\,X^2-(\mathrm{E}\,X)^2=\sum_{x_i\in R_X}x_i^2p_X(x_i)-\left(\sum_{x_i\in R_X}x_ip_X(x_i)\right)^2 \quad \textbf{(2.30)}$$

(ii). Nếu X liên tục có hàm mật độ xác suất $f_X(x)$ thì

$$\mathrm{E}\,X^2=\int_{-\infty}^{\infty}x^2f_X(x)dx\,;\ \mathrm{D}\,X=\mathrm{E}\,X^2-(\mathrm{E}\,X)^2=\int_{-\infty}^{\infty}x^2f_X(x)dx-\left(\int_{-\infty}^{\infty}xf_X(x)dx\right)^2 \quad \textbf{(2.31)}$$

Ví dụ 2.14: Tính phương sai của biến ngẫu nhiên xét trong ví dụ 2.9.

Giải: $\mathrm{E}\,X^2=(-990)^2\cdot 0,008+10^2\cdot 0,992=7940$.

$$\Rightarrow\ \mathrm{D}X=\mathrm{E}X^2-(\mathrm{E}X)^2=7940-4=7936\ \Rightarrow \sigma_X=\sqrt{\mathrm{D}X}=\sqrt{7936}\approx 89,08\,.$$

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

Ví dụ 2.15: Tính phương sai của biến ngẫu nhiên xét trong ví dụ 2.11.

Giải: $\mathrm{E}\,X^2=\dfrac{3}{64}\displaystyle\int_0^4 x^4(4-x)dx=\dfrac{3}{64}\left(\dfrac{4x^5}{5}-\dfrac{x^6}{6}\right)\Bigg|_0^4=\dfrac{32}{5}$

$$\Rightarrow\ \mathrm{D}X=\mathrm{E}X^2-(\mathrm{E}X)^2=\frac{32}{5}-\left(\frac{12}{5}\right)^2=\frac{16}{25}\ \Rightarrow \sigma_X=\frac{4}{5}\,.$$

2.3.2.2 Ý nghĩa và ứng dụng thực tế của phương sai

Phương sai của biến ngẫu nhiên X là độ lệch bình phương trung bình quanh giá trị trung bình $\mathrm{E}\,X$. Trong kỹ thuật phương sai đặc trưng cho mức độ phân tán của các chi tiết gia công hay sai số của thiết bị. Trong quản lý và kinh doanh thì phương sai đặc trưng cho mức độ rủi ro của các quyết định.

Ví dụ 2.9 cho thấy đầu tư bảo hiểm cho những người 25 tuổi là có lãi, nhưng ví dụ 2.14 cho thấy rủi ro của bảo hiểm rất lớn.

Ví dụ 2.16: Một nhà đầu tư đang cân nhắc giữa việc đầu tư vào hai dự án A và B trong hai lĩnh vực độc lập nhau. Khả năng thu hồi vốn sau 2 năm (tính bằng $\%$) của hai dự án là các biến ngẫu nhiên có bảng phân bố sau:

Dự án A

X_A	65	67	68	69	70	71	73
P	0,04	0,12	0,16	0,28	0,24	0,08	0,08

Dự án B

X_B	66	68	69	70	71
P	0,12	0,28	0,32	0,20	0,08

Từ các bảng phân bố xác suất trên ta tìm được

$\mathrm{E}\,X_A = 69{,}16\%$; $\mathrm{D}X_A = 3{,}0944$;

$\mathrm{E}\,X_B = 68{,}72\%$; $\mathrm{D}X_B = 1{,}8016$;

Như vậy nếu chọn phương án đầu tư sao cho tỷ lệ thu hồi vốn kỳ vọng cao hơn thì chọn phương án *A*, song nếu cần chọn phương án có độ rủi ro thu hồi vốn thấp hơn thì chọn *B*.

2.3.2.3 Tính chất của phương sai

1) $\mathrm{D}(aX) = a^2\mathrm{D}(X)$ với mọi hằng số a. **(2.32)**

2) $\mathrm{D}(aX+b) = a^2\mathrm{D}(X)$ với mọi hằng số a, b. **(2.33)**

3) Nếu $X_1, ..., X_n$ độc lập có các phương sai hữu hạn thì

$$\mathrm{D}\left(a_1X_1+\cdots+a_nX_n\right) = a_1^2\mathrm{D}\left(X_1\right)+\cdots+a_n^2\mathrm{D}\left(X_n\right). \quad \textbf{(2.34)}$$

Nói riêng: Nếu X, Y độc lập và $\mathrm{D}X$, $\mathrm{D}Y$ hữu hạn thì $\mathrm{D}\left(X \pm Y\right) = \mathrm{D}X + \mathrm{D}Y$.

Ví dụ 2.17: Tung con xúc xắc n lần. Tìm kỳ vọng của tổng số chấm thu được.

Giải: Gọi $X_i\ (i=1,...,n)$ là số chấm thu được ở lần tung thứ i, gọi X là tổng số chấm thu được trong n lần tung. Như vậy $X = \sum_{i=1}^{n} X_i$. Theo công thức (2.24) ta có $\mathrm{E}\,X = \sum_{i=1}^{n}\mathrm{E}\,X_i$.

Các biến ngẫu nhiên X_i đều có bảng phân bố xác suất như sau

X_i	1	2	3	4	5	6
P	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$

Do đó $\mathrm{E}\,X_i = \dfrac{1}{6}\left(1+2+3+4+5+6\right) = \dfrac{7}{2} \Rightarrow \mathrm{E}\,X = \sum_{i=1}^{n}\mathrm{E}\,X_i = \dfrac{7}{2}n$.

$$\mathrm{E}\,X_i^2 = \frac{1}{6}\left(1^2+2^2+3^2+4^2+5^2+6^2\right) = \frac{91}{6} \Rightarrow \mathrm{D}\,X_i = \frac{91}{6} - \frac{7^2}{2^2} = \frac{35}{12}.$$

Vậy $\mathrm{D}\,X = \sum_{i=1}^{n}\mathrm{D}\,X_i = \dfrac{35}{12}n$.

Ví dụ 2.18: Cho n biến ngẫu nhiên $X_i\ (i=1,...,n)$ độc lập trong cùng một phép thử, có các kỳ vọng bằng nhau và các phương sai bằng nhau

$$\mathrm{E}\,X_1 = \mathrm{E}\,X_2 = ... = \mathrm{E}\,X_n = \mu\ ;\ \mathrm{D}\,X_1 = \mathrm{D}\,X_2 = ... = \mathrm{D}\,X_n = \sigma^2$$

Tìm kỳ vọng và phương sai của $\overline{X} = \dfrac{X_1+X_2+\cdots+X_n}{n}$.

Giải: $\mathrm{E}\,\overline{X} = \mathrm{E}\left(\dfrac{X_1+X_2+\cdots+X_n}{n}\right) = \dfrac{\mathrm{E}\,X_1+\mathrm{E}\,X_2+\cdots+\mathrm{E}\,X_n}{n} = \dfrac{n\mu}{n} = \mu$.

$$\mathrm{D}\,\overline{X} = \mathrm{D}\left(\frac{X_1+X_2+\cdots+X_n}{n}\right) = \frac{\mathrm{D}\,X_1+\mathrm{D}\,X_2+\cdots+\mathrm{D}\,X_n}{n^2} = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}.$$

2.3.3 Phân vị, Trung vị

2.3.3.1 Phân vị

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

$$P\{X < v_\alpha\} \le \alpha \le P\{X \le v_\alpha\} \tag{2.35}$$

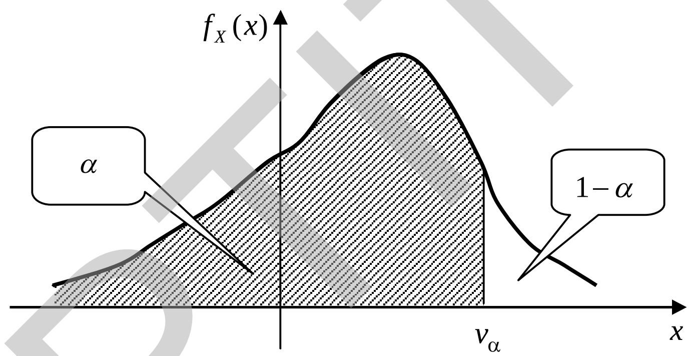
Nghĩa là

$$F_X(v_\alpha^-) \le \alpha \le F_X(v_\alpha) \tag{2.36}$$

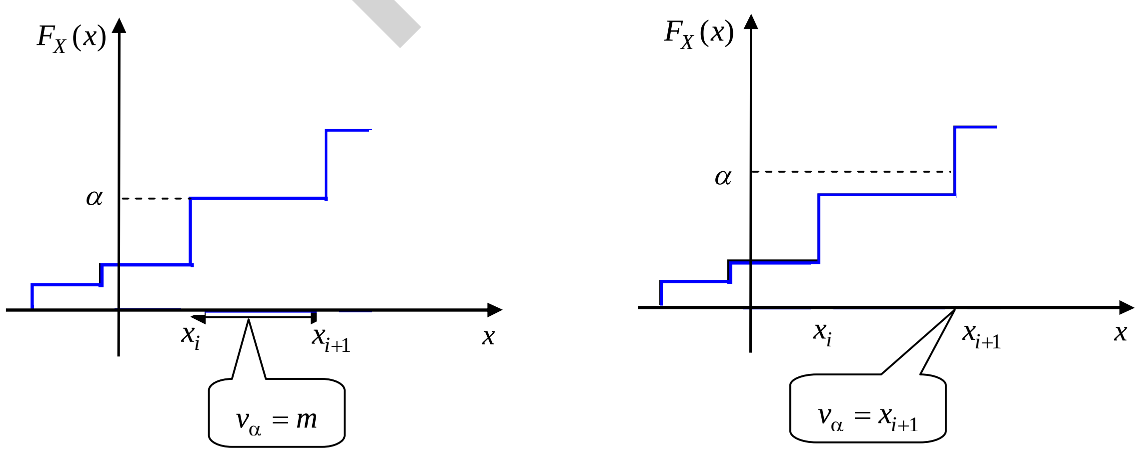
- Trường hợp biến ngẫu nhiên X liên tục với hàm phân bố xác suất $F_X(x)$, phân vị v_α là điểm phân chia miền giá trị R_X của X thành 2 miền với xác suất tương ứng là α và $1-\alpha$. Vậy v_α là nghiệm duy nhất của phương trình $F_X(x) = \alpha$.

$$v_\alpha = F_X^{-1}(\alpha) \tag{2.37}$$

Phân vị mức α là giá trị tới hạn mức $1-\alpha$ (hình 2.5). **(2.38)**



Hình 2.5 Phân vị mức α của biến ngẫu nhiên liên tục



Hình 2.6: Phân vị mức α của biến ngẫu nhiên rời rạc

- Trường hợp biến ngẫu nhiên X rời rạc có miền giá trị $R_X = \{x_1, x_2, ...\}$ và hàm khối lượng xác suất: $p_X(x) = P\{X = x\}$

Với mọi $x_i \in R_X$, $F_X(x_i) = p_X(x_1) + \cdots + p_X(x_i)$. Khi đó

$$v_\alpha = \begin{cases} m, \ \ \forall m \in [x_i, x_{i+1}] & \text{nếu } \alpha = F_X(x_i) \\ x_{i+1} & \text{nếu } F_X(x_i) < \alpha < F_X(x_{i+1}) \end{cases} \tag{2.39}$$

Phân vị mức α của biến ngẫu nhiên liên tục là duy nhất, nhưng của biến ngẫu nhiên rời rạc có thể là vô số.

2.3.3.2 Trung vị

Phân vị mức 1/2 được gọi là ***median*** hay ***trung vị*** của X , ký hiệu $\operatorname{Med} X$. Như vậy trung vị là điểm phân chia phân bố xác suất thành hai phần bằng nhau.

2.3.4 Mốt

Mốt (Mode) của biến ngẫu nhiên X là giá trị mà biến ngẫu nhiên X nhận với xác suất lớn nhất. Một biến ngẫu nhiên có thể có nhiều Mốt.

- Mốt của biến ngẫu nhiên rời rạc X với bảng phân bố xác suất:

X	x_1	x_2	…
P	$p_X(x_1)$	$p_X(x_2)$	…

được xác định như sau:

$$x_{i_0} = \operatorname{Mod} X \ \Leftrightarrow \ p_X(x_{i_0}) = \max\{p_X(x_1), p_X(x_2), \ldots\} \tag{2.40}$$

- Nếu X liên tục có hàm mật độ xác suất $f_X(x)$ thì

$$c = \operatorname{Mod} X \ \Leftrightarrow \ f_X(c) = \max\{f_X(x),\ x \in \mathbb{R}\}. \tag{2.41}$$

Ví dụ 2.19: Biến ngẫu nhiên X ở ví dụ 2.4 có Mốt và trung vị $\operatorname{Mod} X = \operatorname{Med} X = 1$.

Ví dụ 2.20: Tìm trung vị và Mốt của biến ngẫu nhiên rời rạc có bảng phân bố xác suất

X	20	21	22	23	24
P	0,3	0,25	0,18	0,14	0,13

Giải: Dễ thấy rằng $\operatorname{Mod} X = 20$.

Hàm phân bố xác suất của X : $F_X(x) = \begin{cases} 0 & \text{nếu } x < 20 \\ 0,3 & \text{nếu } 20 \le x < 21 \\ 0,55 & \text{nếu } 21 \le x < 22 \\ 0,73 & \text{nếu } 22 \le x < 23 \\ 0,87 & \text{nếu } 23 \le x < 24 \\ 1 & \text{nếu } x \ge 24 \end{cases}$

Từ đó suy ra $\operatorname{Med} X = 21$; các giá trị $x \in [22; 23]$ là phân vị mức 0,73 của X .

Ví dụ 2.21: Tìm $\operatorname{Med} X$ và $\operatorname{Mod} X$ của biến ngẫu nhiên liên tục X xét trong ví dụ 2.6

Giải: Med X là nghiệm của phương trình $F_X(x) = x^2 = \frac{1}{2} \Rightarrow$ Med $X = \frac{1}{\sqrt{2}}$.

Hàm mật độ $f_X(x) = \begin{cases} 0 & \text{víi} \quad x \le 0 \\ 2x & \text{víi} \quad 0 < x \le 1 \\ 0 & \text{víi} \quad x > 1 \end{cases}$ đạt cực đại tại $x = 1$, vậy Mod $X = 1$.

Ví dụ 2.22: Tìm Med X và Mod X của biến ngẫu nhiên liên tục X có hàm mật độ xác định như sau

$$f_X(x) = \begin{cases} \frac{3}{4}x\left(2-x\right) & \text{víi} \quad 0 \le x \le 2 \\ 0 & \text{nÕu tr¸i l¹i} \end{cases}$$

Giải: Hàm phân bố xác suất xác suất

$$F_X(x) = \begin{cases} 0 & \text{víi} \quad x < 0 \\ \frac{3}{4}\left(x^2 - \frac{x^3}{3}\right) & \text{víi} \quad 0 \le x < 2 \\ 1 & \text{víi} \quad x \ge 2 \end{cases}$$

Med X là nghiệm của phương trình

$$F_X(x) = \frac{1}{2} \Leftrightarrow \begin{cases} x^3 - 3x^2 + 2 = 0 \\ 0 \le x < 2 \end{cases} \Leftrightarrow \begin{cases} (x-1)\left[(x-1)^2 - 3\right] = 0 \\ 0 \le x < 2 \end{cases}.$$

Giải phương trình ta được: Med $X = 1$.

Hàm mật độ xác suất $f_X(x)$ có đạo hàm

$$\frac{d}{dx} f_X(x) = \begin{cases} \frac{3}{2}\left(1-x\right) & \text{víi} \quad 0 < x < 2 \\ 0 & \text{nÕu tr¸i l¹i} \end{cases}$$

đổi dấu từ dương sang âm khi đi qua $x = 1$, do đó hàm mật độ đạt cực đại tại $x = 1$.

Vậy Mod $X = 1$.

2.3.5 Mô men, hệ số bất đối xứng, hệ số nhọn

1) Mô men gốc cấp k: $m_k = \mathrm{E}X^k$; $k = 1, 2, \ldots$ **(2.42)**

2) Mô men quy tâm cấp k: $\mu_k = \mathrm{E}\left(X - \mathrm{E}X\right)^k$; $k = 1, 2, \ldots$ **(2.43)**

3) Hệ số bất đối xứng: $\alpha_3 = \frac{\mu_3}{\sigma^3}$ với $\sigma = \sqrt{\mathrm{D}X}$. **(2.44)**

4) Hệ số nhọn: $\alpha_4 = \frac{\mu_4}{\sigma^4}$. **(2.45)**

Nhận xét 2.3:

- $m_1 = \mathrm{E}X$, $\mu_1 = 0$, $\mu_2 = \mathrm{D}X$.

- α_3 đo mức độ bất đối xứng của luật phân bố :

Nếu $\alpha_3 < 0$ thì phân bố xác suất và đồ thị của hàm mật độ sẽ lệch về bên trái hơn.

$\alpha_3 = 0$ thì phân bố xác suất và đồ thị của hàm mật độ đối xứng.

$\alpha_3 > 0$ thì phân bố xác suất và đồ thị của hàm mật độ sẽ lệch về bên phải hơn.

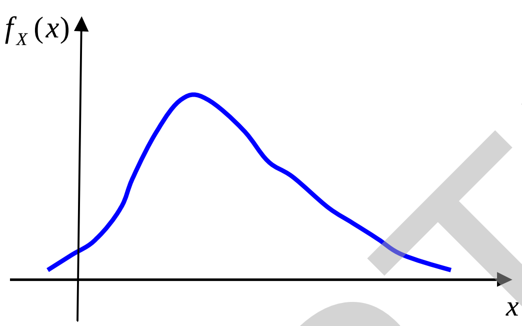
- Hệ số nhọn α_4 đặc trưng cho độ nhọn của đồ thị hàm mật độ so với đồ thị hàm mật độ của phân bố chuẩn.

Với biến ngẫu nhiên có phân bố chuẩn thì $\alpha_4 = 3$.

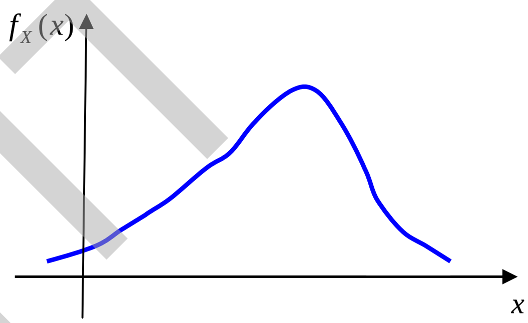
$\alpha_4 > 3$ thì đồ thị hàm mật độ sẽ nhọn hơn so với đồ thị hàm mật độ chuẩn.

$\alpha_4 < 3$ thì đồ thị hàm mật độ sẽ tù hơn so với đồ thị hàm mật độ chuẩn.

- Khi phân bố của X đối xứng hoặc gần đối xứng thì dùng kỳ vọng để định vị là tốt nhất, song nếu phân bố của X quá lệch thì nên dùng trung vị và mốt để định vị.



Hình 2.7a: Hệ số bất đối xứng $\alpha_3 < 0$



Hình 2.7b: Hệ số bất đối xứng $\alpha_3 > 0$

2.4 MỘT SỐ QUY LUẬT PHÂN BỐ XÁC SUẤT RỜI RẠC THƯỜNG GẶP

2.4.1 Phân bố Bernoulli

2.4.1.1 Định nghĩa

Biến ngẫu nhiên X gọi là có phân bố Bernoulli tham số p (còn được gọi là phân bố Không – một A(p)) nếu X chỉ nhận hai giá trị 0 và 1 với bảng phân bố xác suất như sau:

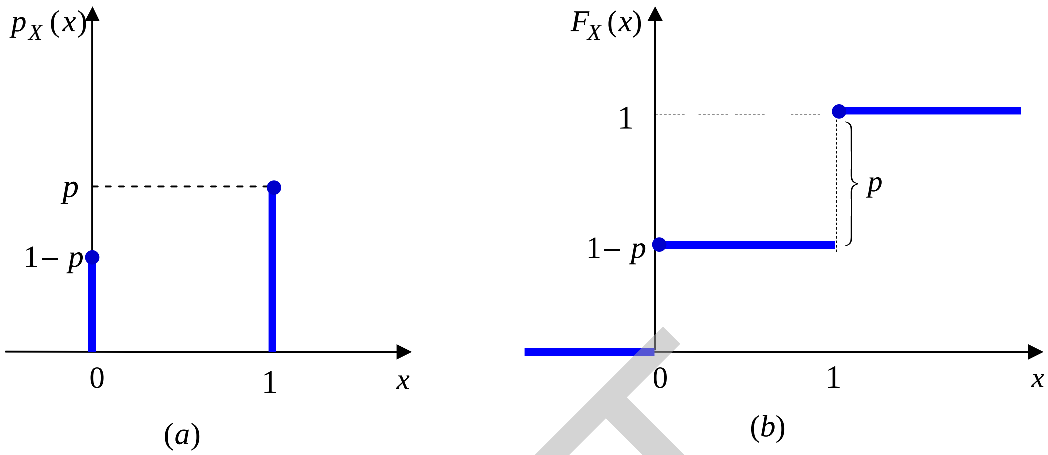
X	0	1
P	q	p

, $q = 1 - p$ **(2.46)**

Phép thử có thể thực hiện lặp lại độc lập và trong mỗi lần thử ta chỉ quan tâm đến sự xuất hiện của biến cố A không đổi với xác suất bằng p được gọi là một phép thử Bernoulli (xem 1.5 chương 1). Gọi X là số lần xuất hiện biến cố A trong một lần thử thì X là biến ngẫu nhiên rời rạc có phân bố Bernoulli tham số p với hàm khối lượng xác suất và hàm phân bố xác suất như sau

$$p_X(x)=\begin{cases}1-p & \text{nÕu } x=0\\ p & \text{nÕu } x=1\\ 0 & \text{nÕu } x\neq 0, x\neq 1\end{cases}\qquad F_X(x)=\begin{cases}0 & \text{nÕu } x<0\\ 1-p & \text{nÕu } 0\le x<1\\ 1 & \text{nÕu } x\ge 1\end{cases}\qquad \textbf{(2.47)}$$

Hình 2.8 minh họa hàm khối lượng xác suất và hàm phân bố xác suất Bernoulli tham số p.



Hình 2.8: *Đồ thị hàm khối lượng xác suất và hàm phân bố Bernoulli tham số* p

2.4.1.2 Các tham số đặc trưng của phân bố Bernoulli tham số *p*

$$\mathrm{E}\,X=0\cdot q+1\cdot p=p\ ;\ \ \mathrm{E}\,X^2=0^2\cdot q+1^2\cdot p=p\ ;\ \ \mathrm{E}(X^2)-(\mathrm{E}\,X)^2=p-p^2=pq\,.$$

Như vậy ta có kỳ vọng, phương sai và độ lệch chuẩn của biến ngẫu nhiên X có phân bố Bernoulli tham số p:

$$\mathrm{E}\,X=p;\ \ \mathrm{D}\,X=pq\ ;\ \ \sigma_X=\sqrt{pq}\qquad \textbf{(2.48)}$$

Trong lý thuyết thống kê biến ngẫu nhiên có phân bố Bernoulli thường được dùng để đặc trưng cho các dấu hiệu nghiên cứu có tính định tính trong đó mỗi cá thể của tổng thể có dấu hiệu này hoặc không có dấu hiệu này.

Chẳng hạn khi muốn nghiên cứu giới tính của khách hàng ta có thể đặc trưng cho giới tính bằng biến ngẫu nhiên với 2 giá trị bằng 0 (Nam) và bằng 1 (Nữ). Trong bài toán bầu cử nếu cử tri nào sẽ bỏ phiếu cho ứng cử viên *A* ta cho nhận giá trị 1, ngược lại ta cho nhận giá trị 0. Để xác định tỷ lệ phế phẩm của lô hàng ta gán cho mỗi sản phẩm một trong hai giá trị 0 và 1, nếu sản phẩm là phế phẩm ta cho nhận giá trị 1 và ngược lại cho nhận giá trị 0 … Đó là các biến ngẫu nhiên có phân bố Bernoulli.

2.4.2 Phân bố nhị thức

2.4.2.1 Định nghĩa

Biến ngẫu nhiên rời rạc X *có miền giá trị* $R_X=\{0, 1, ..., n\}$ *với hàm khối lượng xác suất*

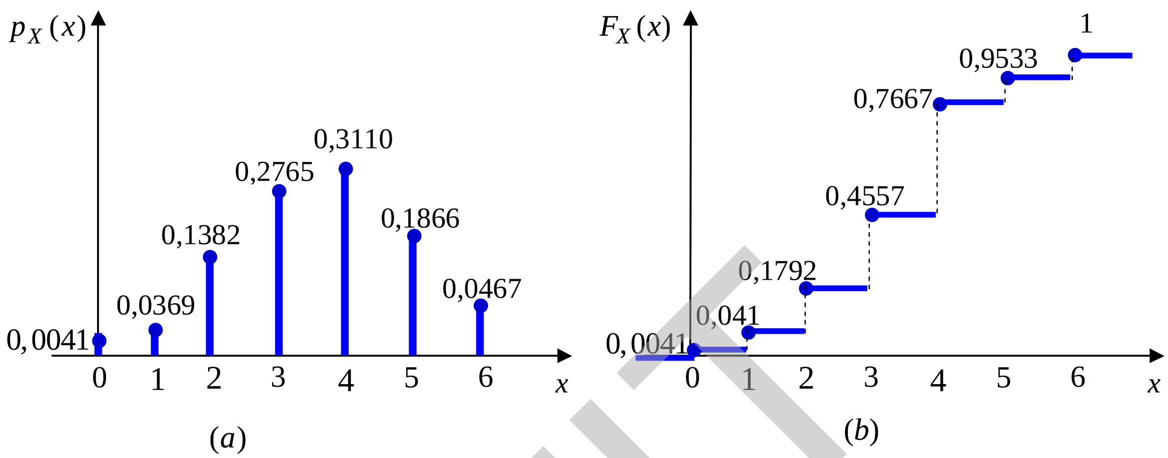
$$p_X(k)=P\{X=k\}=C_n^k p^k q^{n-k}\,,\ k\in R_X\qquad \textbf{(2.49)}$$

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

Bảng phân bố xác suất của biến ngẫu nhiên có phân bố nhị thức $\mathcal{B}(n;p)$

X	0	1	$\dots$	k	$\dots$	n
P	$C_n^0 p^0 q^n$	$C_n^1 p q^{n-1}$	$\dots$	$C_n^k p^k q^{n-k}$	$\dots$	$C_n^n p^n q^0$

Hình 2.9 minh họa phân bố nhị thức với $n=6$ và $p=0,6$.



Hình 2.9: Phân bố nhị thức với $n=6$ *và* $p=0,6$

Hàm phân bố xác suất

$$F_X(x)=\begin{cases} 0 & \text{nÕu } x<0 \\ \sum_{k=0}^{m} C_n^k p^k q^{n-k} & \text{nÕu } m\le x<m+1,\ 0\le m\le n, \\ 1 & \text{nÕu } x\ge n \end{cases} \qquad \textbf{(2.50)}$$

Nếu gọi X là số lần thành công trong dãy n phép thử Bernoulli với xác suất thành công của mỗi lần thử là p thì X có phân bố nhị thức $\mathcal{B}(n;p)$.

Nhận xét 2.4:

1. Phân bố nhị thức $\mathcal{B}(1;p)$ là phân bố Bernoulli tham số p.

2. Thực hiện n lần của cùng một phép thử Bernoulli với xác suất xuất hiện của biến cố A trong mỗi lần thử là p. Gọi $X_1, X_2,\dots, X_n$ lần lượt là số lần xuất hiện của biến cố A trong lần thử thứ $1,2,\dots,n$. Các biến ngẫu nhiên $X_1, X_2,\dots, X_n$ độc lập có cùng phân bố Bernoulli tham số p. Gọi X là số thành công trong n phép thử Bernoulli này thì

$$X = X_1 + X_2 + \cdots + X_n \sim \mathcal{B}(n;p) \qquad \textbf{(2.51)}$$

Nói cách khác tổng của các biến ngẫu nhiên Bernoulli ứng với cùng một dãy phép thử Bernoulli là biến ngẫu nhiên có phân bố nhị thức.

3. Từ (2.51) suy ra rằng nếu $X \sim \mathcal{B}(n_1;p)$, $Y \sim \mathcal{B}(n_2;p)$; X, Y độc lập và cùng phép thử thì

$$X+Y \sim \mathcal{B}(n_1+n_2\,;p) \qquad \textbf{(2.52)}$$

2.4.2.2 Các tham số đặc trưng của biến ngẫu nhiên có phân bố nhị thức

Từ công thức 2.29 và công thức 2.48, 2.51 ta có công thức tính kỳ vọng, phương sai và độ lệch chuẩn của biến ngẫu nhiên $X \sim \mathcal{B}(n;p)$:

$$\mathrm{E}\,X = np;\ \mathrm{D}\,X = npq\ ;\ \ \sigma_X = \sqrt{npq} \qquad \textbf{(2.53)}$$

Từ công thức (1.27) và (1.28) ta có:

- Nếu $(n+1)p$ không nguyên thì

$$\mathrm{Mod}\,X = m \text{ thỏa mãn } (n+1)p-1 \le m \le (n+1)p\,. \qquad \textbf{(2.54)}$$

- Nếu $(n+1)p$ nguyên thì $\mathrm{Mod}\,X$ nhận hai giá trị

$$\mathrm{Mod}\,X = m = np+p \text{ hoặc } \mathrm{Mod}\,X = m-1 = np-q\,. \qquad \textbf{(2.55)}$$

Ví dụ 2.23: Một bưu cục có 10 loại nhật báo khác nhau, xác suất bán hết báo hàng ngày cho mỗi loại là 0,8. Vậy nếu trong một năm với 300 ngày mở cửa thì trung bình có khoảng bao nhiêu ngày bưu cục không bán hết báo.

Giải: Trong mỗi ngày, ta có thể xem việc bán hết mỗi loại nhật báo là một phép thử Bernoulli, gọi X là số loại báo bán hết trong ngày thì $X \sim \mathcal{B}(n;p)$ với $n=10,\ p=0{,}8$. Vậy xác suất để một ngày không bán hết báo là

$$P = P\{X<10\} = 1-P\{X=10\} = 1-0{,}8^{10} = 0{,}8926\,.$$

Tương tự trong một năm với 300 ngày bán hàng tương ứng với 300 phép thử Bernoulli mà kết quả của mỗi lần thử là ngày không bán hết báo, gọi Y là số ngày trong năm bưu cục không bán hết báo thì $Y \sim \mathcal{B}(n;p)$ với $n=300,\ p=0{,}8926$. Vậy số ngày trung bình trong năm mà bưu cục không bán hết báo bằng kỳ vọng toán

$$\mathrm{E}\,Y = np = 300\cdot 0{,}8926 = 267{,}78 \text{ ngày.}$$

2.4.3 Phân bố Poisson

2.4.3.1 Định nghĩa

Biến ngẫu nhiên X có miền giá trị $R_X = \mathbb{N} = \{0,1,\ldots,n,\ldots\}$ với hàm khối lượng xác suất

$$p_X(k) = P\{X=k\} = e^{-\lambda}\frac{\lambda^k}{k!},\ \ k=0,1,2,\ldots \qquad \textbf{(2.56)}$$

gọi là có phân bố Poisson tham số $\lambda>0$, ký hiệu $X \sim \mathcal{P}(\lambda)$.

Hàm phân bố

$$F_X(x) = \begin{cases} 0 & \text{nÕu } x<0 \\ e^{-\lambda}\displaystyle\sum_{k=0}^{n}\frac{\lambda^k}{k!} & \text{nÕu } 0\le n\le x<n+1 \end{cases} \qquad \textbf{(2.57)}$$

Trong thực tế với một số giả thiết thích hợp thì các biến ngẫu nhiên là các quá trình đếm số lần xuất hiện trong một khoảng thời gian xác định nào đó sẽ có phân bố Poisson với tham số

λ:

1) Số cuộc gọi đến một tổng đài.

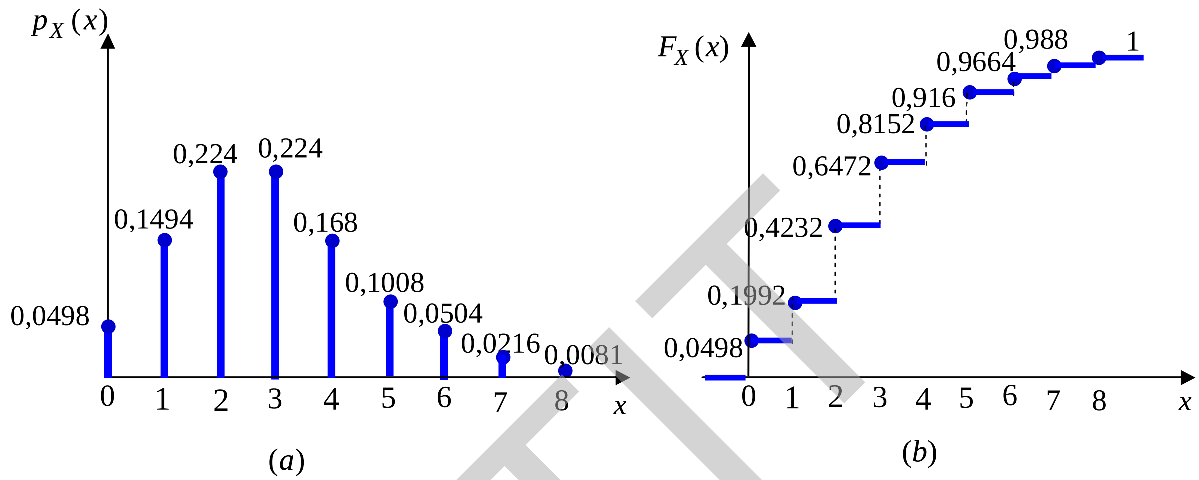
2) Số khách hàng đến 1 điểm phục vụ.

3) Số xe cộ qua 1 ngã tư.

4) Số tai nạn (xe cộ); số các sự cố xảy ra ở một địa điểm …

trong đó λ là tốc độ trung bình của quá trình diễn ra trong khoảng thời gian này.

Hình 2.10 minh họa phân bố Poisson với tham số $\lambda = 3$



Hình 2.10 Phân bố Poisson tham số $\lambda = 3$

Ví dụ 2.24: Ở một tổng đài điện thoại các cuộc gọi đến một cách ngẫu nhiên, độc lập và trung bình có 2 cuộc gọi trong 1 phút. Tìm xác suất để:

a. Có đúng 5 cuộc gọi đến trong 2 phút (biến cố A).

b. Không có một cuộc gọi nào trong 30 giây (biến cố B).

c. Có ít nhất 1 cuộc gọi trong 10 giây (biến cố C).

Giải: Nếu ký hiệu $X(t)$ là số cuộc gọi đến tổng đài trong khoảng thời gian t phút. Theo giả thiết trung bình có 2 cuộc gọi đến tổng đài trong 1 phút, vậy trung bình có $2t$ cuộc gọi trong t phút, do đó $X(t) \sim \mathcal{P}(2t)$.

a. $X(2) \sim \mathcal{P}(4)$, áp dụng công thức 2.56 với $\lambda = 4$, $k = 5$ ta được:

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

b. $X(1/2) \sim \mathcal{P}(1)$, áp dụng công thức 2.56 với $\lambda = 1$, $k = 0$ ta được:

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

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

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

Phân bố Poisson có ứng dụng rộng rãi trong nhiều lĩnh vực thực tế như kiểm tra chất lượng sản phẩm, lý thuyết quản trị dự trữ, lý thuyết xếp hàng … Hầu hết các quá trình đến trong lý thuyết xếp hàng, trong hệ phục vụ đám đông, các bài toán chuyển mạch trong tổng đài …

thường được xét là quá trình đến Poisson.

2.4.3.2 Các tham số đặc trưng của biến ngẫu nhiên có phân bố Poisson

$$\mathrm{E}\,X=\sum_{k=0}^{\infty}ke^{-\lambda}\frac{\lambda^k}{k!}=e^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^k}{(k-1)!}=\lambda e^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!}=\lambda e^{-\lambda}\sum_{m=0}^{\infty}\frac{\lambda^m}{m!}=\lambda,$$

$$\mathrm{E}\,X^2=\sum_{k=0}^{\infty}k^2e^{-\lambda}\frac{\lambda^k}{k!}=e^{-\lambda}\sum_{k=1}^{\infty}k\frac{\lambda^k}{(k-1)!}=\lambda e^{-\lambda}\sum_{k=1}^{\infty}k\frac{\lambda^{k-1}}{(k-1)!}$$

$$=\lambda e^{-\lambda}\left(\sum_{k=1}^{\infty}(k-1)\frac{\lambda^{k-1}}{(k-1)!}+\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!}\right)=\lambda e^{-\lambda}(\lambda e^{\lambda}+e^{\lambda})=\lambda^2+\lambda$$

Như vậy ta có kỳ vọng, phương sai và độ lệch chuẩn của biến ngẫu nhiên X có phân bố Poisson tham số λ:

$$\mathrm{E}\,X=\lambda;\ \mathrm{D}\,X=\lambda;\ \sigma_X=\sqrt{\lambda} \tag{2.58}$$

Từ (2.56) ta có

$$\frac{p_\lambda(k)}{p_\lambda(k+1)}=\frac{k+1}{\lambda}$$

Do đó $p_\lambda(k)<p_\lambda(k+1)\Leftrightarrow k+1<\lambda\Leftrightarrow k<\lambda-1$. Mốt của X là giá trị $m_0=\mathrm{Mod}\,X$ thỏa mãn:

$$\lambda-1\le m_0\le\lambda \tag{2.59}$$

Như vậy

- Nếu λ nguyên thì $\mathrm{Mod}\,X$ nhận 2 giá trị là $\lambda-1$ hoặc λ.
- Nếu λ không nguyên thì $\mathrm{Mod}\,X$ bằng phần nguyên của λ.

Chẳng hạn Hình 2.9 cho ta thấy $X\sim\mathcal{P}(3)$ có $\mathrm{Mod}\,X=2$ và $\mathrm{Mod}\,X=3$ với xác suất $P_{\max}=0,224$

Ta sẽ chứng minh: nếu X_1,X_2 là hai biến ngẫu nhiên độc lập trong cùng phép thử có phân bố Poisson tham số lần lượt λ_1, λ_2 thì X_1+X_2 cũng có phân bố Poisson tham số $\lambda_1+\lambda_2$

$$X_1+X_2\sim\mathcal{P}(\lambda_1+\lambda_2) \tag{2.60}$$

Thật vậy, áp dụng công thức xác suất đầy đủ, chú ý đến điều kiện miền giá trị là tập số tự nhiên và tính chất độc lập của X_1,X_2 ta có

$$P\{X_1+X_2=k\}=\sum_{m=0}^{k}P\{X_1=m\}P\{X_1+X_2=k\,|\,X_1=m\}=\sum_{m=0}^{k}P\{X_1=m\}P\{X_2=k-m\}$$

$$=\sum_{m=0}^{k}e^{-\lambda_1}\frac{\lambda_1^m}{m!}\cdot e^{-\lambda_2}\frac{\lambda_2^{k-m}}{(k-m)!}=e^{-(\lambda_1+\lambda_2)}\sum_{m=0}^{k}\frac{\lambda_1^m}{m!}\cdot\frac{\lambda_2^{k-m}}{(k-m)!}$$

$$=e^{-(\lambda_1+\lambda_2)}\frac{1}{k!}\sum_{m=0}^{k}C_k^m\lambda_1^m\lambda_2^{k-m}=e^{-(\lambda_1+\lambda_2)}\frac{(\lambda_1+\lambda_2)^k}{k!}.$$

2.5 MỘT SỐ QUY LUẬT PHÂN BỐ XÁC SUẤT LIÊN TỤC THƯỜNG GẶP

2.5.1 Phân bố đều

2.5.1.1 Định nghĩa: *Biến ngẫu nhiên* X *được gọi là có phân bố đều trên* (a,b), ký hiệu $X \sim \mathrm{U}(a;b)$ nếu hàm mật độ của X xác định bởi:

$$f_X(x) = \begin{cases} \dfrac{1}{b-a} & \text{nÕu} \quad a \le x \le b \\ 0 & \text{nếu ngược lại} \end{cases} \qquad \textbf{(2.61)}$$

Trường hợp $x \le a$, $P\{X \le x\} = 0$.

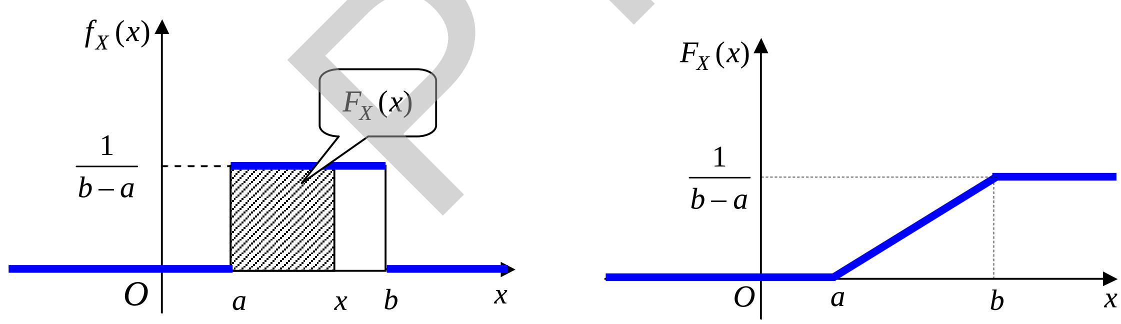
Trường hợp $a \le x \le b$, $P\{X \le x\}$ bằng diện tích hình chữ nhật có chiều rộng bằng $x-a$ và chiều cao bằng $\dfrac{1}{b-a}$ (xem hình 2.10), do đó $P\{X \le x\} = \dfrac{x-a}{b-a}$.

Trường hợp $x \ge b$, $P\{X \le x\}$ bằng diện tích hình chữ nhật có chiều rộng bằng $b-a$ và chiều cao bằng $\dfrac{1}{b-a}$ (xem hình 2.10), do đó $P\{X \le x\} = 1$.

Do đó hàm phân bố xác suất xác định như sau

$$F_X(x) = \int_{-\infty}^{x} f(t)dt = \begin{cases} 0 & \text{nÕu} \quad x < a \\ \dfrac{x-a}{b-a} & \text{nÕu} \quad a \le x \le b \\ 1 & \text{nÕu} \quad x > b \end{cases} \qquad \textbf{(2.62)}$$

Vậy X có khả năng nhận giá trị trong khoảng (a,b) là "đều nhau" và không nhận giá trị ngoài khoảng (a,b).



Hình 2.11 Phân bố đều $\mathbf{U}(a,b)$

2.5.1.2 Các tham số đặc trưng của biến ngẫu nhiên có phân bố đều

Bằng cách tính tích phân xác định ta nhận được công thức kỳ vọng và phương sai của biến ngẫu nhiên $X \sim \mathbf{U}(a,b)$:

$$\mathrm{E}\,X = \frac{a+b}{2}\,;\ \mathrm{D}\,X = \frac{(b-a)^2}{12} \qquad \textbf{(2.63)}$$

Từ đồ thị của hàm mật độ (2.5.15) ta suy ra rằng:

$$\mathrm{Med}\,X = \mathrm{E}\,X = \frac{a+b}{2} \text{ và } \mathrm{Mod}\,X = m,\ \forall m \in (a;b).$$

Phân bố đều có nhiều ứng dụng trong lý thuyết thống kê, trong bài toán mô phỏng thống kê, đặc biệt trong phương pháp phi tham số. Trong một số lý thuyết kết luận thống kê người ta thường xuất phát từ quy tắc sau đây: Nếu ta không biết gì về giá trị của tham số cần ước lượng thì mỗi giá trị có thể có của tham số đó là đồng khả năng, điều đó dẫn đến việc quan niệm tham số cần ước lượng như một biến ngẫu nhiên có phân bố đều.

2.5.2 Phân bố chuẩn

2.5.2.1 Định nghĩa

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

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}; \quad \forall x \in \mathbb{R} \tag{2.64}$$

Phân bố chuẩn được Gauss tìm ra năm 1809 nên nó còn được gọi là *phân bố Gauss*. Phân bố chuẩn thường được thấy trong các bài toán về sai số gặp phải khi đo đạc các đại lượng trong vật lý, thiên văn ...

Trong thực tế, nhiều biến ngẫu nhiên có phân bố chuẩn hoặc tiệm cận chuẩn (Định lý giới hạn trung tâm). Chẳng hạn: trọng lượng, chiều cao của một nhóm người nào đó, điểm thi của thí sinh, năng suất cây trồng, mức lãi suất của một công ty, nhu cầu tiêu thụ của một mặt hàng nào đó ...

2.5.2.2 Các tham số đặc trưng của phân bố chuẩn

Từ công thức xác định hàm mật độ (2.64) ta suy ra các tính chất sau của đồ thị:

- Nhận trục $x = \mu$ làm trục đối xứng.
- Tiệm cận với trục hoành khi $x \to \pm\infty$.
- Diện tích giới hạn bởi đồ thị và trục hoành bằng 1.
- Đạt cực đại tại $x = \mu$ và có giá trị cực đại bằng $\frac{1}{\sigma\sqrt{2\pi}}$.
- Có 2 điểm uốn tại $x = \mu \pm \sigma$.

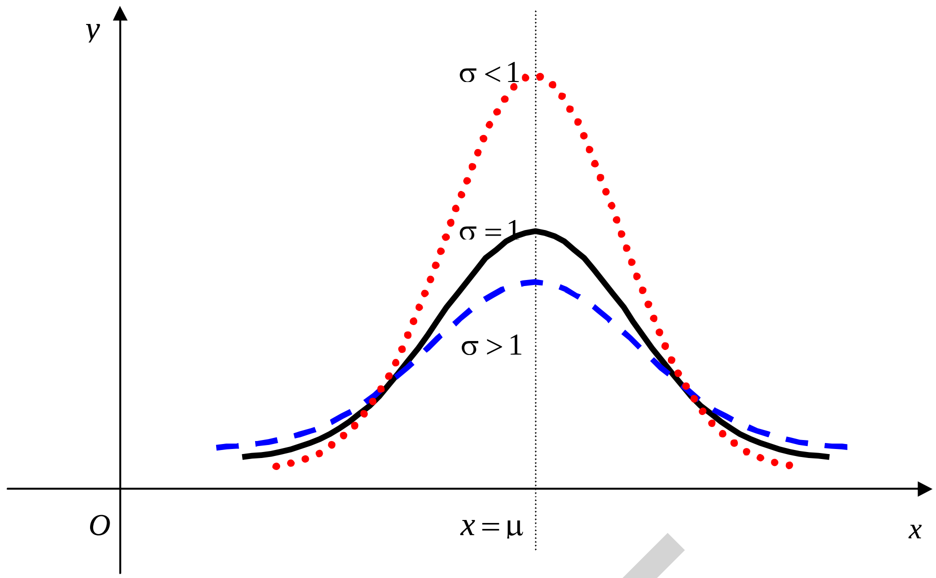
Do đó khi μ tăng lên thì đồ thị dịch sang phải, còn khi μ giảm đồ thị dịch sang trái.

Khi σ tăng lên thì đồ thị sẽ thấp xuống, còn khi σ giảm đồ thị cao lên và nhọn hơn.

$$\text{Mod}\, X = \text{Med}\, X = \mu$$

Bằng cách tính các tích phân suy rộng ta có thể tính được kỳ vọng và phương sai của biến ngẫu nhiên liên tục $X \sim \mathbf{N}(\mu;\sigma^2)$:

$$\text{E}\, X = \mu\,; \ \text{D}\, X = \sigma^2\,; \ \sigma_X = \sqrt{\text{D}X} = \sigma \tag{2.65}$$



Hình 2.12: *Đồ thị hàm mật độ của phân bố chuẩn* $\mathbf{N}(\mu;\sigma^2)$

Có thể chứng minh được, nếu X_1, X_2 là hai biến ngẫu nhiên độc lập lần lượt có phân bố chuẩn $X_1 \sim \mathbf{N}(\mu_1;\sigma_1^2)$ và $X_2 \sim \mathbf{N}(\mu_2;\sigma_2^2)$ thì tổ hợp tuyến tính bất kỳ của X_1, X_2 cũng có phân bố chuẩn. Cụ thể, với mọi a, b khác không thì

$$aX_1 + bX_2 \sim \mathbf{N}(a\mu_1 + b\mu_2; a^2\sigma_1^2 + b^2\sigma_2^2) \quad \textbf{(2.66)}$$

2.5.2.3 Phân bố chuẩn tắc

Phân bố chuẩn $\mathbf{N}(0;1)$ với kỳ vọng bằng 0, phương sai bằng 1 gọi là *phân bố chuẩn tắc.*

Hàm mật độ của phân bố chuẩn tắc $\mathbf{N}(0;1)$

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \; ; \; \forall x \in \mathbb{R} \quad \textbf{(2.67)}$$

Hàm phân bố của $\mathbf{N}(0;1)$

$$\Phi(x) = \int_{-\infty}^{x} \varphi(t)dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt \; ; \; \forall x \in \mathbb{R} \quad \textbf{(2.68)}$$

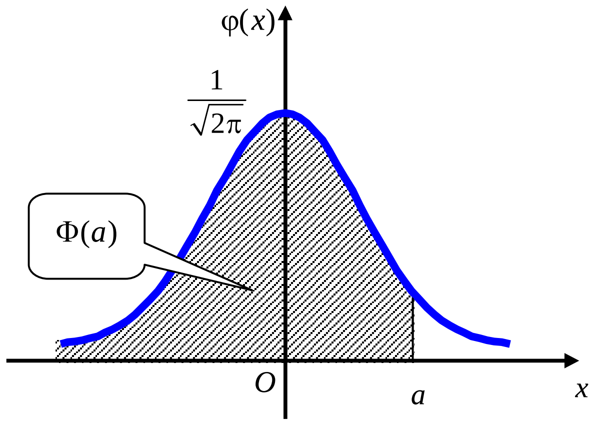
Có bảng tính sẵn các giá trị của $\varphi(x)$ và $\Phi(x)$ (xem Phụ lục I và Phụ lục II).

Cần chú ý rằng một số tài liệu cho bảng tính $\Phi_0(x) = \int_0^x \varphi(t)dt = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt$ với $x \geq 0$.

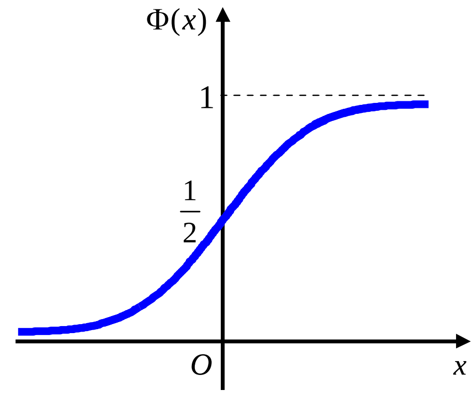
Công thức liên hệ là

$$\Phi(x) = \Phi_0(x) + 0{,}5 \text{ với mọi } x \geq 0 .$$

Đồ thị của hàm mật độ xác suất $\varphi(x)$ và hàm phân bố xác suất $\Phi(x)$



Hình 2.13a: Đồ thị hàm mật độ của phân bố chuẩn tắc $\mathbf{N}(0;1)$



Hình 2.13b: Đồ thị hàm phân bố của phân bố chuẩn tắc $\mathbf{N}(0;1)$

Từ đồ thị của hàm mật độ $\varphi(x)$ (hình 2.12a) ta có:

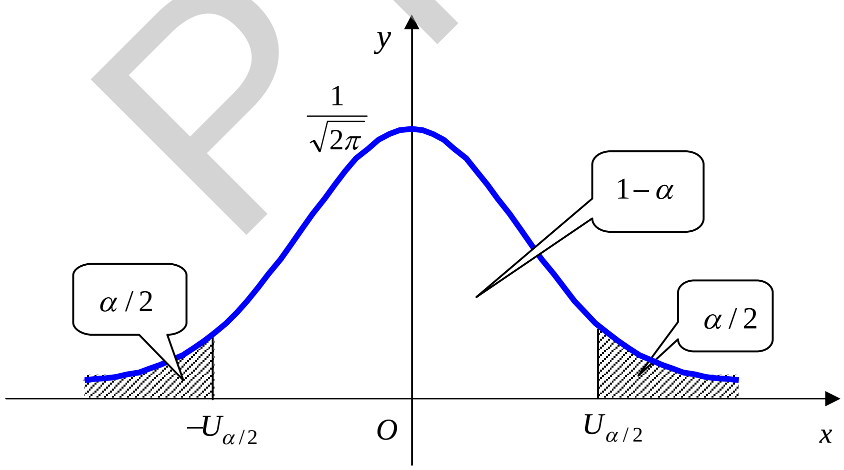
1) $\Phi(x)+\Phi(-x)=1$, $\Phi(-x)=1-\Phi(x)$.

2) Nếu $X \sim \mathbf{N}(0;1)$ thì

$$\forall a>0,\ P\{|X|>a\}=2(1-\Phi(a)),\ P\{|X|<a\}=2\Phi(a)-1. \qquad \textbf{(2.69)}$$

Định nghĩa 2.5: *Giá trị* U_α *gọi là giá trị tới hạn của phân bố chuẩn tắc mức* α *nếu*

$$\Phi^{-1}(1-\alpha)=U_\alpha . \qquad \textbf{(2.70)}$$



Hình 2.14: *Giá trị tới hạn của phân bố chuẩn tắc* $\mathbf{N}(0;1)$

Từ định nghĩa 2.5 và từ hình 2.14 ta có: Nếu $X \sim \mathbf{N}(0;1)$ thì

$$\forall \alpha: 0<\alpha<1; P\{X>U_\alpha\}=\alpha\ ;\ P\left\{|X|>U_{\frac{\alpha}{2}}\right\}=\alpha\ ;\ P\left\{|X|<U_{\frac{\alpha}{2}}\right\}=1-\alpha \qquad \textbf{(2.71)}$$

3) Có thể chứng minh được:

$$\text{Nếu } X \sim \mathbf{N}(\mu;\sigma^2) \text{ thì } \frac{X-\mu}{\sigma} \sim \mathbf{N}(0;1). \tag{2.72}$$

Do đó

$$F_X(x) = P\{X \le x\} = P\left\{\frac{X-\mu}{\sigma} \le \frac{x-\mu}{\sigma}\right\} = \Phi\left(\frac{x-\mu}{\sigma}\right). \tag{2.73}$$

$$P\{a < X < b\} = P\left\{\frac{a-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{b-\mu}{\sigma}\right\} = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right). \tag{2.74}$$

Công thức 2.73-2.74 vẫn còn đúng khi thay dấu “<’ thành dấu “≤” (xem công thức 2.19).

Từ công thức (2.69) và (2.72) ta có công thức tính xác suất của sự sai lệch giữa biến ngẫu nhiên có phân bố chuẩn $X \sim \mathbf{N}(\mu;\sigma^2)$ và kỳ vọng của nó theo công thức:

$$P\{|X-\mu| < \varepsilon\} = P\left\{\left|\frac{X-\mu}{\sigma}\right| < \frac{\varepsilon}{\sigma}\right\} = 2\Phi\left(\frac{\varepsilon}{\sigma}\right) - 1 \tag{2.75}$$

Ví dụ 2.25: Giả sử $X \sim \mathbf{N}(\mu;\sigma^2)$, $\mu = 2100$, $\sigma = 200$. Hãy tìm:

a. $P\{X \le 2400\}$.

b. $P\{1700 < X < 2200\}$.

c. Xác định a để $P\{X > a\} = 0{,}03$.

Giải: Áp dụng các công thức (2.73) ta được

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

b. Áp dụng các công thức (2.74) ta được

$$P\{1700 < X < 2200\} = \Phi\left(\frac{2200-2100}{200}\right) - \Phi\left(\frac{1700-2100}{200}\right) = \Phi(0{,}5) - \Phi(-2) = 0{,}6688$$

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

Tra bảng ta được $0{,}97 = \Phi(1{,}881) \Rightarrow \frac{a-2100}{200} = 1{,}881 \Rightarrow a = 2476{,}2$.

2.5.2.4 Quy tắc hai xích ma và ba xích ma

Nếu trong công thức (2.75) ta đặt $\varepsilon = 2\sigma$ tức là bằng hai lần độ lệch chuẩn của X thì

$$P\{|X-\mu| < 2\sigma\} = 2\Phi(2) - 1 = 0{,}9544$$

$$P\{\mu - 2\sigma < X < \mu + 2\sigma\} = 0{,}9544 \tag{2.76}$$

Tương tự thay $\varepsilon = 3\sigma$ ta được

$$P\{\mu - 3\sigma < X < \mu + 3\sigma\} = 0{,}9973 \tag{2.77}$$

Hai công thức (2.76)-(2.77) là cơ sở của quy tắc hai xích ma và ba xích ma:

Nếu X có phân bố chuẩn $\mathbf{N}(\mu;\sigma^2)$ thì có đến 95,44% giá trị của X nằm trong khoảng

$(\mu - 2\sigma\, ;\, \mu + 2\sigma)$ *và hầu như toàn bộ giá trị của* X *nằm trong khoảng* $(\mu - 3\sigma\, ;\, \mu + 3\sigma)$.

2.5.3 Tính gần đúng phân bố nhị thức

Giả sử $X_1, X_2, ..., X_n$ độc lập có cùng phân bố Bernoulli tham số p. Theo công thức (2.51) ta có

$$U_n = X_1 + X_2 + \cdots + X_n \sim \mathrm{B}(n; p).$$

Biến ngẫu nhiên U_n nhận các giá trị $k = 0,1,...,n$ với xác suất

$$P\{U_n = k\} = C_n^k p^k q^{n-k} = \frac{n!}{k!(n-k)!} p^k q^{n-k} \text{ (công thức 2.49).}$$

Tuy nhiên khi n khá lớn thì $n!$ rất lớn, do đó ta không thể áp dụng công thức này để tính mà cần đến công thức xấp xỉ.

Có thể xấp xỉ phân bố nhị thức theo phân bố chuẩn hoặc phân bố Poisson.

2.5.3.1 Tính gần đúng phân bố nhị thức bằng phân bố chuẩn

Người ta chứng minh được rằng (*Định lý Moivre –Laplace,* chương 3):

Với mọi $x \in \mathbb{R}$: $$\lim_{n\to\infty} P\left\{\frac{U_n - np}{\sqrt{npq}} \le x\right\} = \Phi(x)$$

Định lý Moivre-Laplace cho phép xấp xỉ phân bố nhị thức $\mathrm{B}(n; p)$ với phân bố chuẩn $\mathbf{N}(np; npq)$ khi n đủ lớn (công thức 2.78). Người ta thấy rằng xấp xỉ là tốt khi

$$np \text{ và } nq \text{ lớn hơn 5 hoặc khi } npq > 20. \tag{2.78}$$

Do đó khi n đủ lớn

$$P\{U_n \le x\} = P\left\{\frac{U_n - np}{\sqrt{npq}} \le \frac{x - np}{\sqrt{npq}}\right\} \approx \Phi\left(\frac{x - np}{\sqrt{npq}}\right) \tag{2.79}$$

$$P\{a < U_n \le b\} \approx \Phi\left(\frac{b - np}{\sqrt{npq}}\right) - \Phi\left(\frac{a - np}{\sqrt{npq}}\right) \tag{2.80}$$

Hơn nữa áp dụng định lý giới hạn địa phương ta có thể tính xấp xỉ xác suất tại từng điểm khi n đủ lớn:

$$P\{U_n = k\} = C_n^k p^k q^{n-k} \approx \frac{1}{\sqrt{npq}} \varphi\left(\frac{k - np}{\sqrt{npq}}\right) = \frac{1}{\sqrt{npq}} \frac{1}{\sqrt{2\pi}} e^{-\frac{(k-np)^2}{2npq}}. \tag{2.81}$$

Công thức (2.80), (2.81) tính xấp xỉ phân bố của biến ngẫu nhiên rời rạc bởi phân bố của biến ngẫu nhiên liên tục, do đó công thức tính xấp xỉ tốt hơn như sau

$$P\{U_n \le x\} \approx \Phi\left(\frac{x + 0,5 - np}{\sqrt{npq}}\right) \tag{2.82}$$

$$P\left\{a < U_n \le b\right\} \approx \Phi\left(\frac{b+0,5-np}{\sqrt{npq}}\right) - \Phi\left(\frac{a-0,5-np}{\sqrt{npq}}\right) \quad \textbf{(2.83)}$$

Ví dụ 2.26: Gieo 3200 lần một đồng xu cân đối và đồng chất. Gọi X là số lần xuất hiện mặt sấp trong 3200 lần gieo đó.

a. Tìm số lần xuất hiện mặt sấp có khả năng nhất. Tính xác suất tương ứng.

b. Tính xác suất $P\left\{5\sqrt{2}+1600 \le X \le 10\sqrt{2}+1600\right\}$.

Giải: **a.** Ta có: $n = 3200$, $p = 0,5 \Rightarrow (n+1)p = 1600,5$. Vậy số lần xuất hiện mặt sấp có khả năng nhất là 1600 với xác suất tương ứng

$$P_{3200}(1600;0,5) = \frac{3200!}{1600!1600!}0,5^{3200}.$$

Ta thấy rằng khó tính được con số cụ thể của $\dfrac{3200!}{1600!1600!}0,5^{3200}$.

Tuy nhiên ta có thể tính gần đúng như sau: $x_m = \dfrac{1600-3200\cdot 0,5}{\sqrt{3200\cdot 0,5\cdot 0,5}} = 0$. Do đó

$$P_{3200}(1600;0,5) \approx \frac{1}{\sqrt{3200\cdot 0,5\cdot 0,5}}\varphi(0) = \frac{1}{40\sqrt{\pi}} \approx 0,014.$$

b. Ta có $\sqrt{npq} = \sqrt{3200\cdot 0,5\cdot 0,5} = \sqrt{800} = 20\sqrt{2}$, áp dụng công thức (2.83) ta được:

$$P\left\{5\sqrt{2} \le X - 1600 \le 10\sqrt{2}\right\} = P\left\{0,25 \le \frac{X-1600}{20\sqrt{2}} \le 0,5\right\} \approx \Phi(0,5) - \Phi(0,25) = 0,09.$$

2.5.3.2 Tính gần đúng phân bố nhị thức bằng phân bố Poisson

Công thức (2.81)-(2.83) cho phép xấp xỉ $\mathcal{B}(n;p)$ bởi $\mathbf{N}(np;npq)$ khi npq đủ lớn, khi npq bé thì xấp xỉ này không tốt. Tuy nhiên trong trường hợp

$$n > 50 \text{ và } p < 0,1 \quad \textbf{(2.84)}$$

người ta chứng minh được rằng có thể xấp xỉ $\mathcal{B}(n;p)$ với phân bố Poisson $\mathcal{P}(\lambda)$ tham số $\lambda = np$.

Ví dụ 2.27: Một nhà phân tích kinh tế tiên đoán 3,5% số công ti nhỏ trong 1 vùng địa phương A sẽ bị thanh lý phá sản trong năm tới. Với 1 mẫu ngẫu nhiên gồm 100 công ti được chọn ra. Tìm xác suất để ít nhất 3 công ti trong mẫu ấy bị thanh lý phá sản trong năm tới (giả sử rằng tiên đoán của nhà phân tích trên là đúng).

Giải: Gọi X là số công ti bị phá sản trong năm tới của 100 công ti được lấy mẫu thì $X \sim \mathcal{B}(n;p)$ với $n = 100$, $p = 0,035$ thỏa mãn điều kiện (2.84).

Ta có thể xấp xỉ $\quad X \sim \mathcal{P}(np) = \mathcal{P}(3,5)$.

Xác suất cần tính là:

$$P\left\{X \ge 3\right\} = 1 - \left(P\left\{X=0\right\} + P\left\{X=1\right\} + P\left\{X=2\right\}\right) \approx 1 - e^{-3,5}\left\{1 + \frac{3,5}{1!} + \frac{3,5^2}{2!}\right\}$$

$$\approx 1-\left(0,0302+0,1057+0,1850\right)=0,697=69,7\% .$$

2.5.4 Phân bố "Khi bình phương"

2.5.4.1 Định nghĩa

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

$$f_X(x)=\begin{cases} \dfrac{x^{n/2-1}}{2^{n/2}\Gamma(n/2)}e^{-\frac{x}{2}} & \text{nÕu } x>0 \\ 0 & \text{nÕu } x\le 0 \end{cases} \qquad \textbf{(2.85)}$$

trong đó $\Gamma(x)=\int\limits_0^{+\infty} t^{x-1}e^{-t}dt$ là hàm Gamma.

Có thể chứng minh được rằng nếu $X_1, X_2, ..., X_n$ là các biến ngẫu nhiên độc lập có cùng phân bố chuẩn tắc $\mathbf{N}(0,1)$ thì

$$\sum_{i=1}^{n} X_i^2 = X_1^2+X_2^2+\cdots+X_n^2 \sim \chi_n^2 \qquad \textbf{(2.86)}$$

Phân bố "khi bình phương" do Karl Pearson đưa ra vào năm 1900.

Nhận xét 2.5:

1. Trong thực tế ta thường sử dụng phân bố "khi bình phương" dưới dạng tổng của bình phương của các biến ngẫu nhiên độc lập cùng có phân bố chuẩn tắc theo công thức (2.86) mà ít xét đến hàm mật độ xác suất (2.85).
2. Từ (2.86) suy ra rằng nếu $X_1, X_2, ..., X_k$ là các biến ngẫu nhiên độc lập có phân bố "khi bình phương" với bậc tự do lần lượt $n_1,\ n_2,\ \ldots,\ n_k$ thì $X_1+X_2+\cdots+X_k$ là biến ngẫu nhiên có phân bố "khi bình phương" $n_1+n_2+\cdots+n_k$ bậc tự do
$$X_1+X_2+\cdots+X_k \sim \chi^2_{n_1+n_2+\cdots+n_k} \qquad \textbf{(2.87)}$$
3. Có thể chứng minh được: Nếu X_1, X_2 là hai biến ngẫu nhiên độc lập, X_1 có phân bố "khi bình phương" với n_1 bậc tự do và X_1+X_2 có phân bố "khi bình phương" với n bậc tự do, $n>n_1$ thì X_2 là biến ngẫu nhiên có phân bố "khi bình phương" với $n-n_1$ bậc tự do.

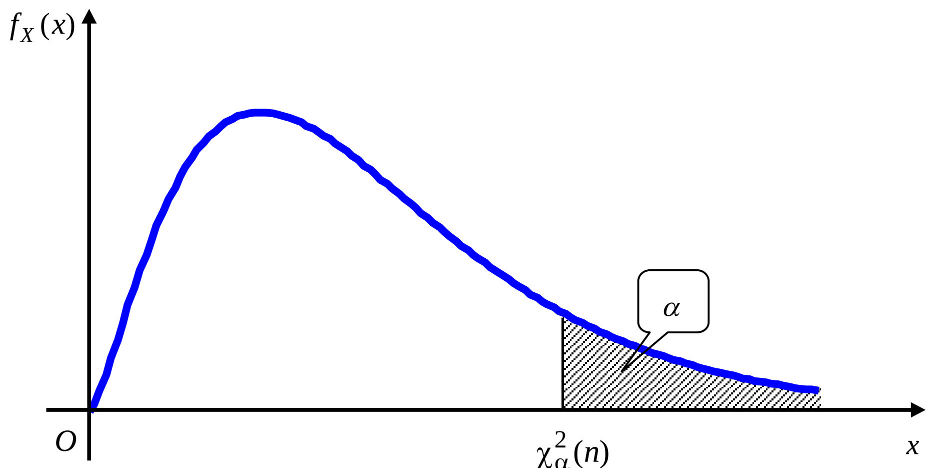
2.5.4.2 Các tham số đặc trưng của phân bố "khi bình phương" n bậc tự do

Có thể chứng minh được rằng:

Nếu $X_i \sim \mathrm{N}(0;1)$ thì $\mathrm{E}(X_i^2)=\mathrm{D}\,X_i=1$ và $\mathrm{E}(X_i^4)=3 \Rightarrow \mathrm{D}(X_i^2)=2$.

Áp dụng công thức 2.24, 2.34 và 2.86 ta nhận được công thức kỳ vọng và phương sai của biến ngẫu nhiên "khi bình phương" n bậc tự do $\chi^2(n)$

$$\mathrm{E}\chi^2(n)=n;\ \mathrm{D}\chi^2(n)=2n \qquad \textbf{(2.88)}$$



Hình 2.15: Giá trị tới hạn của phân bố "khi bình phương"

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

$$P\left\{\chi^2 > \chi^2_\alpha(n)\right\} = \alpha\,. \tag{2.89}$$

Bảng các giá trị tới hạn $\chi^2_\alpha(n)$ được tính sẵn trong bảng ở Phụ lục IV.

2.5.5 Phân bố Student

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

$$f_T(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma(n/2)}\left(1+\frac{t^2}{n}\right)^{\frac{-(n+1)}{2}}, \quad -\infty < t < \infty\,. \tag{2.90}$$

trong đó $\Gamma(x)$ là hàm Gamma.

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

$$T = \frac{Z}{\sqrt{V/n}} \sim \mathbf{T}(n) \tag{2.91}$$

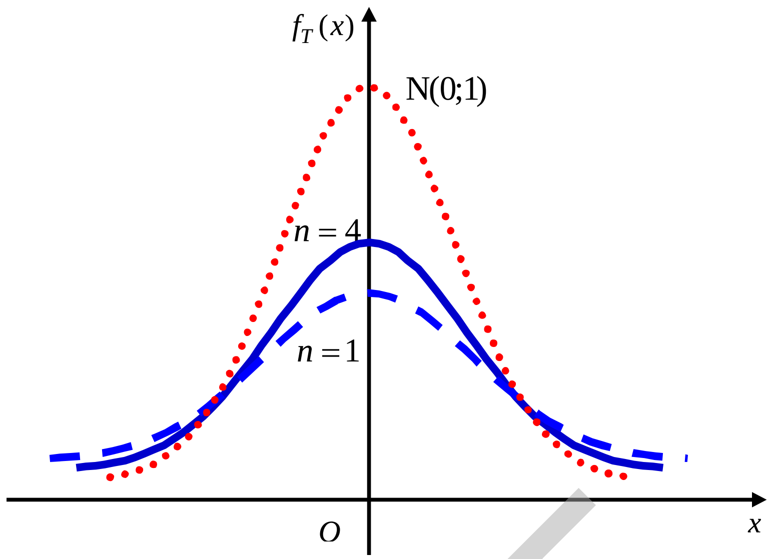
Kỳ vọng và phương sai của biến ngẫu nhiên T có phân bố Student n bậc tự do:

$$\mathrm{E}T = 0,\ \ \mathrm{D}T = \frac{n}{n-2} \tag{2.92}$$

Giá trị tới hạn mức α của phân bố Student n bậc tự do, ký hiệu $t_\alpha(n)$, thỏa mãn:

$$P\left\{T > t_\alpha(n)\right\} = \alpha \tag{2.93}$$

Bảng tính các giá trị tới hạn $t_\alpha(n)$ cho trong Phụ lục III.



Hình 2.16: Đồ thị hàm mật độ của phân bố Student

Hàm mật độ (2.90) là hàm chẵn nên đồ thị đối xứng qua trục tung. Khi số bậc tự do tăng lên, phân bố Student hội tụ rất nhanh về phân bố chuẩn tắc $\mathbf{N}(0;1)$. Do đó khi n đủ lớn ($n \geq 30$) có thể dùng phân bố chuẩn tắc thay cho phân bố Student. Tuy nhiên khi n nhỏ ($n < 30$) việc thay thế như trên sẽ gặp sai số lớn.

CÂU HỎI ÔN TẬP VÀ BÀI TẬP CHƯƠNG 2

2.1 Biến ngẫu nhiên luôn luôn nhận giá trị dương.

Đúng ☐ Sai ☐.

2.2 Biến ngẫu nhiên rời rạc chỉ nhận một số hữu hạn các giá trị.

Đúng ☐ Sai ☐.

2.3 Nếu biến ngẫu nhiên X rời rạc có miền giá trị $R_X = \{x_1, ..., x_n\}$ thì hệ các biến cố $\{X = x_1\}$, …, $\{X = x_n\}$ lập thành một hệ đầy đủ.

Đúng ☐ Sai ☐.

2.4 Hàm phân bố xác suất là một hàm liên tục.

Đúng ☐ Sai ☐.

2.5 Kỳ vọng của biến ngẫu nhiên rời rạc là giá trị nó lấy thường xuyên nhất.

Đúng ☐ Sai ☐.

2.6 Kỳ vọng của tổng hai biến ngẫu nhiên của cùng một phép thử luôn luôn bằng tổng các kỳ vọng của nó.

Đúng ☐ Sai ☐.

2.7 Hai biến ngẫu nhiên có cùng kỳ vọng sẽ có cùng phương sai.

Đúng ☐ Sai ☐.

2.8 Phương sai của tổng hai biến ngẫu nhiên rời rạc luôn luôn bằng tổng phương sai của nó.

Đúng ☐ Sai ☐.

2.9 Biến ngẫu nhiên tồn tại phương sai thì cũng tồn tại kỳ vọng.

Đúng ☐ Sai ☐.

2.10 Hàm mật độ $f_X(x)$ của biến ngẫu nhiên liên tục X có tính chất $f_X(x) \ge 0$.

Đúng ☐ Sai ☐.

2.11 Tổng của n biến ngẫu nhiên độc lập phân bố Bernoulli tham số p của cùng một phép thử là một biến ngẫu nhiên có phân bố nhị thức $\mathrm{B}(n;p)$.

Đúng ☐ Sai ☐.

2.12 Tổng của hai biến ngẫu nhiên có phân bố nhị thức bất kỳ luôn luôn là một biến ngẫu nhiên có phân bố nhị thức.

Đúng ☐ Sai ☐.

2.13 Biến ngẫu nhiên có phân bố Poisson là biến ngẫu nhiên rời rạc nên chỉ nhận một số hữu hạn các giá trị.

Đúng ☐ Sai ☐.

2.14 Nếu X là biến ngẫu nhiên có phân bố nhị thức $\mathrm{B}(n;p)$ thì X chỉ nhận duy nhất một giá trị $\mathrm{Mod}\, X = m$ thỏa mãn $(n+1)p-1 \le m \le (n+1)p$.

Đúng ☐ Sai ☐.

2.15 Nếu X là biến ngẫu nhiên có phân bố Poisson tham số $\lambda > 0$ thì kỳ vọng, phương sai và mốt của X đều bằng λ.

Đúng ☐ Sai ☐.

2.16 Nếu biến ngẫu nhiên X có phân bố chuẩn $\mathbf{N}(\mu;\sigma^2)$ thì xác suất sai lệch giữa X và kỳ vọng của nó thỏa mãn $P\left\{|X-\mu|<\varepsilon\right\} = 2\Phi\left(\dfrac{\varepsilon}{\sigma}\right)-1$.

Đúng ☐ Sai ☐.

2.17 Nếu biến ngẫu nhiên X có phân bố chuẩn $\mathbf{N}(\mu;\sigma^2)$ thì $\dfrac{X-\mu}{\sigma}$ có phân bố chuẩn tắc $\mathbf{N}(0;1)$

Đúng ☐ Sai ☐.

2.18 Ta có thể tính gần đúng xác suất của biến ngẫu nhiên có phân bố nhị thức $\mathrm{B}(n;p)$ bằng phân bố chuẩn nếu n đủ lớn và $p > 0{,}1$ hoặc xấp xỉ bằng phân bố Poisson khi $n > 50$ và $p < 0{,}1$

Đúng ☐ Sai ☐.

2.19 Biến ngẫu nhiên có phân bố Student chỉ nhận những giá trị dương.

Đúng ☐ Sai ☐.

2.20 Có thể xấp xỉ phân bố “khi bình phương” n bậc tự do bởi phân bố chuẩn tắc $\mathbf{N}(0;1)$ khi $n \ge 30$.

Đúng ☐ Sai ☐ .

2.21 Biến ngẫu nhiên X có bảng phân bố

X	-5	2	3	4
P	0,4	0,3	0,1	0,2

Tính kỳ vọng EX và phương sai DX.

2.22 Biến ngẫu nhiên rời rạc X nhận ba giá trị có thể có là x_1, x_2, x_3. Biết $x_1 = 0,6$; $x_2 = 4$ với xác suất tương ứng $p_1 = 0,3$, $p_2 = 0,5$ và có kỳ vọng $\mathrm{E}\,X = 8$. Tìm x_3 và p_3.

2.23 Cho X_1 và X_2 là hai biến ngẫu nhiên độc lập có bảng phân bố xác suất như sau:

X_1	2	3	5
P	0,3	0,5	0,2

X_2	1	4
P	0,2	0,8

a. Tính $\mathrm{E}\,X_1$; $\mathrm{E}\,X_2$; $\mathrm{D}\,X_1$; $\mathrm{D}\,X_2$.

b. Tính $\mathrm{E}(X_1 + X_2)$ và $\mathrm{D}(X_1 + X_2)$.

2.24 Cho X_1, X_2, X_3 là ba biến ngẫu nhiên độc lập có bảng phân bố xác suất như sau:

X_1	0	2
P	0,6	0,4

X_2	1	2
P	0,4	0.6

X_3	0	2
P	0,8	0.2

Lập bảng phân bố xác suất của $\overline{X} = \dfrac{X_1 + X_2 + X_3}{3}$. Tính $\mathrm{E}(\overline{X})$; $\mathrm{D}(\overline{X})$.

2.25 Hai biến ngẫu nhiên X, Y độc lập. Tính $\mathrm{D}(Z)$ với:

a. $Z = 2X + 3Y$. **b.** $Z = -3X + Y$.

Cho biết $\mathrm{D}(X) = 4$, $\mathrm{D}(Y) = 5$.

2.26 Biến ngẫu nhiên rời rạc X nhận các giá trị có thể có là $x_1 = -1$; $x_2 = 0$; $x_3 = 1$. Tìm các xác suất tương ứng p_1; p_2; p_3 biết rằng $\mathrm{E}(X) = 0,1$ và $\mathrm{D}(X) = 0,89$.

2.27 Xếp ngẫu nhiên 5 hành khách lên 3 toa tầu I, II, III. Gọi X là số khách lên toa I và Y là số khách lên toa II và III.

a. Tính xác suất để cả 3 toa đều có khách.

b. Lập bảng phân bố xác suất của biến ngẫu nhiên X và biến ngẫu nhiên Y.

2.28 Tính kỳ vọng và phương sai của biến ngẫu nhiên liên tục X có hàm mật độ xác suất

$$f_X(x) = \begin{cases} \dfrac{\cos x}{2} & \text{nÕu } x \in (-\pi/2; \pi/2) \\ 0 & \text{nÕu } x \notin (-\pi/2; \pi/2) \end{cases}$$

2.29. Cho biến ngẫu nhiên X liên tục với hàm mật độ như sau

$$f_X(x)=\begin{cases} kx & \text{nÕu } \ 0\le x\le 1 \\ k & \text{nÕu } \ 1\le x\le 4 \\ 0 & \text{nÕu tr¸i l¹i} \end{cases}$$

a. Tìm k và hàm phân bố $F_X(x)$.

b. Tính kỳ vọng $\mathrm{E}X$ và phương sai $\mathrm{D}X$.

2.30 Tuổi thọ của một loài côn trùng nào đó là một biến ngẫu nhiên X (đơn vị là tháng) với hàm mật độ như sau

$$f_X(x)=\begin{cases} kx^2(2-x) & \text{nÕu } \ 0\le x\le 2 \\ 0 & \text{nÕu tr¸i l¹i} \end{cases}$$

a. Tìm k.

b. Tính xác suất để côn trùng chết trước khi nó được một tháng tuổi.

c. Tìm $\mathrm{E}X, \mathrm{D}X$.

2.31 Hai xạ thủ A và B tập bắn. Mỗi người bắn hai phát. Xác suất bắn trúng đích của A trong mỗi lần bắn là 0,4; còn của B là 0,5.

a. Gọi X là số phát bắn trúng của **A** trừ đi số phát bắn trúng của **B**. Tìm phân bố xác suất của X, kỳ vọng $\mathrm{E}X$ và phương sai $\mathrm{D}X$.

b. Tìm phân bố xác suất của $Y=|X|$ và kỳ vọng $\mathrm{E}Y$.

2.32 Một xí nghiệp có hai ô tô vận tải hoạt động. Xác suất trong ngày làm việc các ô tô bị hỏng tương ứng bằng 0,1 và 0,2. Gọi X là số ô tô bị hỏng trong thời gian làm việc. Lập bảng phân bố xác suất, tính kỳ vọng $\mathrm{E}X$ và phương sai $\mathrm{D}X$ của X.

2.33 Cho biến ngẫu nhiên liên tục X có hàm mật độ

$$f_X(x)=\begin{cases} kx^2 & \text{nÕu } \ 0\le x\le 3 \\ 0 & \text{víi } x \text{ cßn l¹i} \end{cases}$$

a. Tìm k.

b. Tính xác suất $P\{X>2\}$.

c. Tìm hàm phân bố của X.

d. Tìm α để $P\{X<\alpha\}=\dfrac{3}{4}$.

2.34 Cho biến ngẫu nhiên rời rạc X có bảng phân bố xác suất

X	1	2	3	4	5	6	7
P	k	$2k$	$2k$	$3k$	k^2	$2k^2$	$7k^2+k$

a. Xác định k.

b. Tính xác suất $P\{X\ge 5\}$ và $P\{X<3\}$.

c. Tính kỳ vọng $\mathrm{E}X$.

d. Tính phương sai $\mathrm{D}X$.

2.35 Có 5 sản phẩm trong đó có 4 chính phẩm và 1 phế phẩm. Người ta lấy ra lần lượt 2 sản phẩm (lấy không hoàn lại).

a. Gọi X là "số phế phẩm có thể gặp phải". Lập bảng phân bố xác suất của X.

Tính kỳ vọng $\mathrm{E}X$ và phương sai $\mathrm{D}X$.

b. Gọi Y là "số chính phẩm có thể nhận được". Lập hệ thức cho biết mối quan hệ giữa Y và X. Tính kỳ vọng $\mathrm{E}Y$ và phương sai $\mathrm{D}Y$.

2.36 Một nhóm có 10 người trong đó có 6 nam và 4 nữ. Chọn ngẫu nhiên ra 3 người. Gọi X là số nữ có trong nhóm được chọn. Lập bảng phân bố xác suất của X. Tính kỳ vọng $\mathrm{E}X$.

2.37 Cho biến ngẫu nhiên liên tục X có hàm mật độ

$$f_X(x) = \begin{cases} kx^2(1-x) & \text{nÕu } 0 \le x \le 1 \\ 0 & \text{víi } x \text{ cßn l¹i} \end{cases}$$

a. Tìm k.

b. Tính $P\{0,4 < X < 0,6\}$.

c. Tính kỳ vọng $\mathrm{E}X$.

2.38 Để kinh doanh có lãi một cửa hàng rau tươi cần nhập mỗi ngày bao nhiêu kg? Biết rằng 1 kg rau bán được lãi 50 đồng, không bán được thì lỗ 20 đồng. Theo dự kiến trong 100 ngày thì sẽ có 10 ngày không bán được kg nào, 30 ngày bán được 1000kg/ngày, 45 ngày bán được 2000kg/ngày và15 ngày bán được 3000kg/ ngày.

2.39 Hai kiện tướng bóng bàn ngang sức thi đấu với nhau. Hỏi thắng 2 trong 4 ván dễ hơn hay thắng 3 trong 6 ván dễ hơn.

2.40 Một nữ công nhân quản lý 12 máy dệt. Xác suất để mỗi máy trong khoảng thời gian T cần đến sự chăm sóc của nữ công nhân là 1/3. Tính xác suất:

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

b. Trong khoảng thời gian T có từ 3 đến 6 máy cần đến sự chăm sóc của nữ công nhân.

2.41 Trong một lô hàng có 800 sản phẩm loại 1 và 200 sản phẩm loại 2. Lấy ngẫu nhiên ra 5 sản phẩm theo phương thức có hoàn lại. Gọi X là số sản phẩm loại 1 lấy được.

a. X có phân bố phân bố gì? Viết biểu thức tổng quát của quy luật.

b. Tìm kỳ vọng và phương sai của X.

c. Tìm mốt của X và tính khả năng để xảy ra điều đó.

2.42 Xác suất để sản phẩm sản xuất ra bị hỏng bằng 0,1.

a. Tìm xác suất để trong 5 sản phẩm sản xuất ra có không quá 2 sản phẩm hỏng.

b. Tìm số sản phẩm hỏng trung bình trong 5 sản phẩm đó.

c. Tìm số sản phẩm hỏng có khả năng xảy ra nhiều nhất.

2.43 Một bài thi trắc nghiệm gồm có 10 câu hỏi, mỗi câu hỏi có 5 phương án trả lời, trong đó chỉ có một phương án đúng. Giả sử mỗi câu trả lời đúng được 4 điểm và câu trả lời sai bị trừ 2 điểm. Một học sinh kém làm bài bằng cách chọn hú hoạ một phương án cho mỗi câu hỏi. Tính xác suất để:

a. Anh ta được 4 điểm.

b. Anh ta bị điểm âm.

2.44 Tín hiệu thông tin được phát đi 5 lần độc lập nhau. Xác suất thu được tin của mỗi lần phát là 0,7. Tính xác suất:

a. Thu được tín hiệu đúng 2 lần ;

b. Thu được tín hiệu nhiều nhất 1 lần ;

c. Thu được tin.

2.45 Trong một cuộc thi bắn, mỗi xạ thủ được bắn 5 phát vào bia. Xác suất bắn trúng bia của xạ thủ A là 0,8. Tinh xác suất:

a. A bắn trúng bia đúng 2 lần ;

b. A bắn trúng bia nhiều nhất 2 lần ;

c. A bắn trúng bia.

2.46 Một cầu thủ nổi tiếng về đá phạt đền, xác suất đá vào gôn là 4/5. Có người cho rằng cứ "sút" 5 quả thì chắc chắn rằng có 4 quả vào lưới. Điều khẳng định đó có đúng không? Tìm xác suất để trong 5 lần sút có đúng 4 lần bóng vào lưới.

2.47 Ở một tổng đài điện thoại các cuộc điện thoại gọi đến xuất hiện một cách ngẫu nhiên, độc lập với nhau và trung bình có 2 cuộc gọi trong một phút. Tính xác suất để:

a. Có ít nhất một cuộc gọi trong khoảng thời gian 10 giây.

b. Trong khoảng thời gian 3 phút có nhiều nhất ba cuộc gọi.

c. Trong khoảng thời gian 3 phút liên tiếp mỗi phút có nhiều nhất một cuộc gọi.

2.48 Số cuộc gọi điện thoại đến một trạm điện thoại A trong một phút là một đại lượng ngẫu nhiên (biến ngẫu nhiên) X có phân bố Poisson với tham số $\lambda = 1,5$. Tính xác suất để trong một phút:

a. Trạm điện thoại A không nhận được cuộc gọi nào.

b. Trạm điện thoại A nhận được nhiều nhất 2 cuộc gọi.

c. Trạm điện thoại A nhận được ít nhất 4 cuộc gọi.

2.49 Số khách hàng vào một siêu thị trong một giờ là biến ngẫu nhiên có phân bố Poisson với tham số $\lambda = 8$(8 là số khách hàng đến trung bình trong một giờ). Tìm xác suất để trong một giờ nào đó có hơn 4 khách vào.

2.50 Cho biến ngẫu nhiên $X \sim \mathrm{N}(0;1)$. Tính các xác suất sau:

a. $P\{0 < X < 1,2\}$ **b.** $P\{-0,68 < X < 0\}$ **c.** $P\{-0,46 < X < 2,21\}$

d. $P\{0,81 < X < 1,94\}$ **e.** $P\{X > -1,28\}$

2.51 Biến ngẫu nhiên X có phân bố chuẩn với kỳ vọng $\mu = 10$ và phương sai $\sigma^2 = 4$. Tính xác suất để X nhận giá trị trong khoảng (8; 12).

2.52 Biến ngẫu nhiên X có phân bố chuẩn với kỳ vọng $\mu = 10$. Xác suất để X nhận giá trị trong khoảng (10; 20) là 0,3. Tìm xác suất để X nhận giá trị trong khoảng (0; 10).

2.53 Trọng lượng sản phẩm X do một máy tự động sản xuất là một biến ngẫu nhiên có phân bố chuẩn với $\mu = 100$ gam và độ lệch chuẩn $\sigma = 1$ gam. Sản phẩm được coi là đạt tiêu chuẩn kỹ thuật nếu trọng lượng của nó đạt từ 98 đến 102gam.

a. Tìm tỷ lệ sản phẩm đạt tiêu chuẩn kỹ thuật của nhà máy.

b. Tìm tỷ lệ phế phẩm của nhà máy.

c. Giải thích bằng đồ thị kết quả tìm được ở phần a).

2.54 Chiều cao của nam giới khi trưởng thành ở một vùng dân cư là biến ngẫu nhiên phân bố chuẩn với kỳ vọng $\mu = 160$ cm và độ lệch chuẩn $\sigma = 6$ cm. Một thanh niên bị coi là lùn nếu có chiều cao nhỏ hơn 155 cm.

a. Tìm tỷ lệ thanh niên lùn ở vùng đó.

b. Tìm xác suất để lấy ngẫu nhiên 4 người thì có ít nhất 1 người không bị lùn.

CHƯƠNG 3: VÉC TƠ NGẪU NHIÊN

Khái niệm biến ngẫu nhiên hai chiều hay còn gọi là véc tơ ngẫu nhiên hai chiều là trường hợp riêng của biến ngẫu nhiên nhiều chiều bao gồm nhiều biến ngẫu nhiên lập thành một bộ có thứ tự. Mỗi biến ngẫu nhiên là một thành phần của nó.

Tương tự biến ngẫu nhiên, quy luật phân bố xác suất của biến ngẫu nhiên nhiều chiều được khảo sát thông qua hàm phân bố. Biến ngẫu nhiên nhiều chiều có các biến ngẫu nhiên thành phần rời rạc gọi là biến ngẫu nhiên nhiều chiều rời rạc. Nếu tất cả các biến ngẫu nhiên thành phần liên tục thì biến ngẫu nhiên nhiều chiều tương ứng gọi là liên tục. Biến ngẫu nhiên nhiều chiều rời rạc được xác định bởi bảng phân bố xác suất đồng thời hoặc hàm khối lượng xác suất đồng thời, còn biến ngẫu nhiên liên tục được xác định bởi hàm mật độ xác suất đồng thời.

Chương trình toán cao cấp của hệ đại học ngành kinh tế không xét đến khái niệm tích phân bội, vì vậy trong chương này ta chỉ có thể xét biến ngẫu nhiên hai chiều rời rạc.

Từ hàm phân bố xác suất đồng thời có thể tính được hàm phân bố xác suất của các biến ngẫu nhiên thành phần. Cũng vậy, từ bảng phân bố xác suất đồng thời của biến ngẫu nhiên hai chiều rời rạc có thể tìm được bảng phân bố xác suất của hai biến ngẫu nhiên thành phần.

Áp dụng công thức tính xác suất có điều kiện suy ra quy luật phân bố xác suất có điều kiện của các biến ngẫu nhiên thành phần.

Ngoài các đặc trưng kỳ vọng, phương sai của hai biến ngẫu nhiên thành phần, biến ngẫu nhiên hai chiều còn được đặc trưng bởi hiệp phương sai và hệ số tương quan. Hệ số tương quan đo mức độ phụ thuộc tuyến tính của hai biến ngẫu nhiên thành phần, khi hệ số tương quan càng gần 1 thì mức độ phụ thuộc tuyến tính càng chặt. Hai biến ngẫu nhiên thành phần không tương quan thì hệ số tương quan bằng 0.

Áp dụng công thức xác suất có điều kiện ta xây dựng phân bố xác suất có điều kiện. Từ đó có thể tính kỳ vọng có điều kiện của biến ngẫu nhiên thành phần này với điều kiện biến ngẫu nhiên thành phần kia nhận giá trị cụ thể nào đó và xây dựng hàm hồi quy tương quan.

Để học tốt chương này học viên cần nắm vững các tính chất cơ bản của xác suất, xác suất có điều kiện và phân bố xác suất của biến ngẫu nhiên rời rạc.

3.1. KHÁI NIỆM VÉC TƠ NGẪU NHIÊN

3.1.1 Khái niệm và phân loại véc tơ ngẫu nhiên

Trong các chương trước ta xét các biến ngẫu nhiên mà giá trị chúng nhận được có thể biểu diễn bằng một số, đó là các biến ngẫu nhiên một chiều. Tuy nhiên trong thực tế có thể gặp các đại lượng ngẫu nhiên mà giá trị nhận được là một bộ gồm hai, ba, …, n số. Những đại lượng này được gọi một cách tương ứng là biến ngẫu nhiên hai chiều, ba chiều, …, n chiều và được gọi chung là biến ngẫu nhiên nhiều chiều hoặc véc tơ ngẫu nhiên. Các biến ngẫu nhiên hai chiều, ba chiều, …, n chiều còn được gọi là véc tơ ngẫu nhiên hai chiều, ba chiều, …, n chiều.

Một biến ngẫu nhiên n chiều là một bộ có n thành phần $(X_1, X_2, ..., X_n)$, trong đó các thành phần $X_1, X_2, ..., X_n$ là các biến ngẫu nhiên. Trường hợp biến ngẫu nhiên hai chiều ta ký hiệu là (X,Y), trong đó X là biến ngẫu nhiên thành phần thứ nhất và Y là biến ngẫu nhiên thành phần thứ hai.

Biến ngẫu nhiên n chiều $(X_1, X_2, ..., X_n)$ là liên tục hay rời rạc nếu tất cả các biến ngẫu nhiên thành phần $X_1, X_2, ..., X_n$ là liên tục hay rời rạc.

Trong chương này ta chỉ xét biến ngẫu nhiên hai chiều (X,Y) rời rạc.

Ví dụ 3.1: Một nhà máy sản xuất một loại sản phẩm. Nếu kích thước của sản phẩm được đo bằng chiều dài X và chiều rộng Y thì ta có biến ngẫu nhiên hai chiều, còn nếu xét thêm cả chiều cao Z nữa thì ta có biến ngẫu nhiên ba chiều. Nếu ta chỉ quan tâm đến trọng lượng và thể tích của sản phẩm ta cũng được biến ngẫu nhiên hai chiều.

3.1.2 Hàm phân bố xác suất đồng thời và hàm phân bố xác suất biên

Định nghĩa 3.2: Hàm n biến $F_{X_1,..., X_n}(x_1, x_2, ..., x_n)$ xác định bởi:

$$F_{X_1,..., X_n}(x_1, x_2, ..., x_n) = P\left\{X_1 \le x_1, X_2 \le x_2, ..., X_n \le x_n\right\} \qquad \textbf{(3.1)}$$

trong đó $\left\{X_1 \le x_1, X_2 \le x_2, ..., X_n \le x_n\right\}$ *là biến cố tích* $\left\{X_1 \le x_1\right\} \cap \left\{X_2 \le x_2\right\} \cap ... \cap \left\{X_n \le x_n\right\}$, *được gọi là hàm phân bố xác suất của véc tơ ngẫu nhiên* $X = (X_1, X_2, ..., X_n)$ *hoặc hàm phân bố xác suất đồng thời của các biến ngẫu nhiên* $X_1, X_2, ..., X_n$.

Hàm phân bố xác suất đồng thời có các tính chất:

1. $0 \le F_{X_1,..., X_n}(x_1, ..., x_n) \le 1$. **(3.2)**
2. $\lim\limits_{x_k \to -\infty} F_{X_1,..., X_n}(x_1, ..., x_n) = 0$, với k nào đó thuộc $\{1, ..., n\}$. **(3.3)**
3. $\lim\limits_{(x_1,...,x_n) \to (\infty,...,\infty)} F_{X_1,..., X_n}(x_1, ..., x_n) = 1$. **(3.4)**
4. $F_{X_1,..., X_n}(x_1, ..., x_n)$ không giảm theo từng biến. **(3.5)**
5. $\lim\limits_{x_1 \to \infty} F_{X_1,..., X_n}(x_1, ..., x_n) = F_{X_2,..., X_n}(x_2, ..., x_n)$.

   Như vậy nếu lấy giới hạn của hàm phân bố xác suất đồng thời của n biến ngẫu nhiên $X_1, X_2, ..., X_n$ khi biến x_1 tiến đến vô cùng thì được hàm phân bố xác suất đồng thời của $n-1$ biến ngẫu nhiên còn lại $X_2, ..., X_n$.

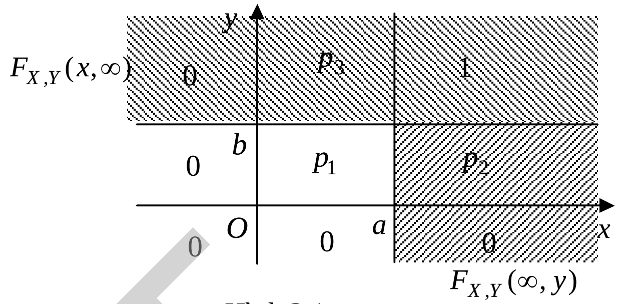
   Tương tự nếu lấy giới hạn của hàm phân bố xác suất đồng thời của $X_1, X_2, ..., X_n$ khi biến x_k tiến đến vô cùng, với k nào đó thuộc $\{1, 2, ..., n\}$, thì được hàm phân bố xác suất đồng thời của $n-1$ biến ngẫu nhiên còn lại $X_1, ..., X_{k-1}, X_{k+1}, ..., X_n$.
6. Đặc biệt nếu $F_{X,Y}(x, y)$ là hàm phân bố xác suất của véc tơ ngẫu nhiên hai chiều (X,Y) thì:

$$\lim_{y\to\infty} F_{X,Y}(x,y) = P\left\{X \le x\right\} = F_X(x);\ \lim_{x\to\infty} F_{X,Y}(x,y) = P\left\{Y \le y\right\} = F_Y(y) \qquad \textbf{(3.6)}$$

trong đó $F_X(x)$, $F_Y(y)$ là các hàm phân bố xác suất của biến ngẫu nhiên X, Y và được gọi là các hàm phân bố xác suất thành phần của véc tơ ngẫu nhiên (X,Y), hay còn gọi là ***hàm phân bố xác suất biên*** của hàm phân bố xác suất đồng thời $F_{X,Y}(x,y)$.

Ví dụ 3.2: Hàm phân bố xác suất của véc tơ ngẫu nhiên (X,Y) xác định như sau

$$F_{X,Y}(x,y) = \begin{cases} 0 & x<0 \text{ hoặc } y<0 \\ p_1 & 0\le x<a,\ 0\le y<b \\ p_2 & x\ge a,\quad 0\le y<b \\ p_3 & 0\le x<a,\ y\ge b \\ 1 & x\ge a,\ y\ge b \end{cases}$$



Hình 3.1

Có hai hàm phân bố xác suất biên

$$F_X(x) = F_{X,Y}(x,\infty) = \begin{cases} 0 & x<0 \\ p_3 & 0\le x<a \\ 1 & x\ge a \end{cases}.\quad F_Y(y) = F_{X,Y}(\infty,y) = \begin{cases} 0 & y<0 \\ p_2 & 0\le y<b \\ 1 & y\ge b \end{cases}$$

3.2. HÀM KHỐI LƯỢNG XÁC SUẤT VÀ BẢNG PHÂN BỐ XÁC SUẤT

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

3.2.1 Hàm khối lượng xác suất đồng thời và bảng phân bố xác suất đồng thời

Hàm khối lượng xác suất đồng thời của biến ngẫu nhiên hai chiều (X,Y) ký hiệu và xác định bởi:

$$p_{X,Y}(x_i,y_j) = P\left\{X = x_i, Y = y_j\right\} = P\left(\left\{X = x_i\right\}\cap\left\{Y = y_j\right\}\right), \qquad \textbf{(3.7)}$$

Hàm khối lượng xác suất đồng thời thỏa mãn điều kiện

$$\begin{cases} p_{X,Y}(x_i,y_j) \ge 0,\ \forall\, i=1,\dots,n,\ j=1,\dots,m \\ \displaystyle\sum_{i=1}^{n}\sum_{j=1}^{m} p_{X,Y}(x_i,y_j) = 1 \end{cases} \qquad \textbf{(3.8)}$$

Để trực quan hơn chúng ta biểu diễn hàm khối lượng xác suất đồng thời của biến ngẫu nhiên rời rạc hai chiều thông qua bảng phân bố xác suất đồng thời.

Bảng phân bố xác suất đồng thời của biến ngẫu nhiên rời rạc hai chiều (X,Y) là bảng liệt kê tất cả các giá trị của X theo cột, các giá trị của Y theo hàng và các xác suất tương ứng có dạng sau, trong đó x_i ($i=1,...,n$) là các giá trị có thể có của thành phần X; y_j ($j=1,...,m$) là các giá trị có thể có của thành phần Y.

$R_{X,Y}=\left\{(x_i,y_j)\,\middle|\,i=1,...,n;\, j=1,...,m\right\}$ gọi là ***miền giá trị của véc tơ ngẫu nhiên*** (X,Y).

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

Hàm phân bố xác suất đồng thời được xác định từ hàm khối lượng xác suất đồng thời

$$F_{X,Y}(x,y)=\sum_{x_i\le x}\sum_{y_j\le y} p_{X,Y}(x_i,y_j);\ (x_i,y_j)\in R_{X,Y} \qquad \textbf{(3.9)}$$

3.2.2 Bảng phân bố xác suất biên

Áp dụng công thức xác suất đầy đủ (1.239) cho hệ $\{X=x_1\}, \{X=x_2\}, ..., \{X=x_n\}$ (xem công thức 2.13) ta có:

$$p_Y(y_j)=P\{Y=y_j\}=\sum_{i=1}^{n}P\{X=x_i,Y=y_j\}=\sum_{i=1}^{n}p_{X,Y}(x_i,y_j)\ ;\ j=1,...,m \qquad \textbf{(3.10)}$$

$$p_X(x_i)=P\{X=x_i\}=\sum_{j=1}^{m}P\{X=x_i,Y=y_j\}=\sum_{j=1}^{m}p_{X,Y}(x_i,y_j)\ ;\ i=1,...,n \qquad \textbf{(3.11)}$$

Nhận xét 3.1:

1. Từ công thức (3.10) và (3.11), ta thấy rằng nếu ta cộng các xác suất của bảng phân bố xác suất đồng thời của (X,Y) theo cột thì ta được các xác suất tương ứng với các giá trị của Y, nếu ta cộng các xác suất theo hàng ta được các xác suất tương ứng với giá trị của X. Từ đó nhận được phân bố xác suất của biến ngẫu nhiên thành phần Y và biến ngẫu nhiên thành phần X.

2. Hàm khối lượng xác suất của véc tơ ngẫu nhiên rời rạc n chiều $(X_1,X_2,...,X_n)$

$$p_{X_1,X_2,...,X_n}(x_1,x_2,...,x_n)=P\{X_1=x_1,X_2=x_2,...,X_n=x_n\}. \qquad \textbf{(3.12)}$$

Hàm phân bố xác suất đồng thời có thể xác lập từ hàm khối lượng xác suất đồng thời

$$F_{X_1,X_2,...,X_n}(y_1,y_2,...,y_n)=\sum_{x_1\le y_1}...\sum_{x_n\le y_n} p_{X_1,X_2,...,X_n}(x_1,x_2,...,x_n)\ ;\ (x_1,x_2,...,x_n)\in R_{X_1,X_2,...,X_n}\ . \quad \textbf{(3.13)}$$

$X \backslash Y$	y_1	y_2	…	y_j	…	y_m	$\sum_j$
x_1	$p_{XY}(x_1,y_1)$	$p_{XY}(x_1,y_2)$	…	$p_{XY}(x_1,y_j)$	…	$p_{XY}(x_1,y_m)$	$p_X(x_1)$
x_2	$p_{XY}(x_2,y_1)$	$p_{XY}(x_2,y_2)$	…	$p_{XY}(x_2,y_j)$	…	$p_{XY}(x_2,y_m)$	$p_X(x_2)$
$\vdots$	$\vdots$	$\vdots$	…	$\vdots$	…	$\vdots$	$\vdots$
x_i	$p_{XY}(x_i,y_1)$	$p_{XY}(x_i,y_2)$	…	$p_{XY}(x_i,y_j)$	…	$p_{XY}(x_i,y_m)$	$p_X(x_i)$
$\vdots$	$\vdots$	$\vdots$	…	$\vdots$	…	$\vdots$	$\vdots$
x_n	$p_{XY}(x_n,y_1)$	$p_{XY}(x_n,y_2)$	…	$p_{XY}(x_n,y_j)$	…	$p_{XY}(x_n,y_m)$	$p_X(x_n)$
$\sum_i$	$p_Y(y_1)$	$p_Y(y_2)$	…	$p_Y(y_j)$	…	$p_Y(y_m)$	1

X	x_1	x_2	…	x_i	…	x_n
P	$p_X(x_1)$	$p_X(x_2)$	…	$p_X(x_i)$	…	$p_X(x_n)$

Y	y_1	y_2	…	y_j	…	y_m
P	$p_Y(y_1)$	$p_Y(y_2)$	…	$p_Y(y_j)$	…	$p_Y(y_m)$

Ví dụ 3.3: Gieo 3 đồng tiền cân đối A, B, C. Gọi X là số mặt ngửa xuất hiện của 2 đồng tiền A, B và Y là số mặt ngửa xuất hiện của cả 3 đồng tiền A, B, C. Hãy lập bảng phân bố xác suất đồng thời của X,Y.

Giải: Chúng ta có bảng liệt kê 8 kết quả đồng khả năng khi gieo 3 đồng tiền cân đối và tính các giá trị của X,Y tương ứng, trong đó N là ký hiệu mặt ngửa xuất hiện còn S là mặt sấp.

A	N	N	N	N	S	S	S	S
B	N	N	S	S	N	N	S	S
C	N	S	N	S	N	S	N	S
X	2	2	1	1	1	1	0	0
Y	3	2	2	1	2	1	1	0

Sử dụng công thức tính xác suất cổ điển (1.1) ta có:

$$P\{X=2,Y=3\}=\frac{1}{8}\ ;\ P\{X=2,Y=2\}=\frac{1}{8}\ ;\ P\{X=1,Y=2\}=\frac{2}{8}\ \ldots$$

Vậy bảng phân bố xác suất đồng thời của X và Y là

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

Cộng các xác suất theo hàng và theo cột ta được:

$X \backslash Y$	0	1	2	3	$\sum$
0	1/8	1/8	0	0	2/8
1	0	2/8	2/8	0	4/8
2	0	0	1/8	1/8	2/8
$\sum$	1/8	3/8	3/8	1/8	1

Vậy phân bố xác suất của hai biến ngẫu nhiên thành phần:

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

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

Ví dụ 3.4: Có hai hộp, mỗi hộp đựng 6 bi.

Hộp I có: 1 bi mang số 1, 2 bi mang số 2, 3 bi mang số 3.

Hộp II có: 2 bi mang số 1, 3 bi mang số 2, 1 bi mang số 3.

Rút ngẫu nhiên từ mỗi hộp 1 bi. Gọi X,Y lần lượt là số ghi trên bi rút được từ hộp I và hộp II. Hãy lập bảng phân bố xác suất đồng thời của X,Y.

Giải: Mỗi hộp có 6 bi cho nên số các trường hợp có thể có của phép thử là $6\cdot 6=36$, trong đó có 2 trường hợp $(1,1)$, 3 trường hợp $(1,2)$, 4 trường hợp $(2,1)$, …

Vậy bảng phân bố xác suất đồng thời của X,Y như sau:

$X \backslash Y$	1	2	3
1	2/36	3/36	1/36
2	4/36	6/36	2/36
3	6/36	9/36	3/36

Ví dụ 3.5: (Phân bố đa thức, multinomial) Véc tơ ngẫu nhiên k chiều $X=\left(X_1, X_2, \ldots, X_k\right)$ được gọi là có phân bố đa thức với các tham số $\left(n; p_1, \ldots, p_k\right)$ ký hiệu $X \sim \mathrm{MUT}\left(n; p_1, \ldots, p_k\right)$ nếu hàm khối lượng xác suất đồng thời có dạng:

$$p_{X_1 X_2 \ldots X_k}\left(m_1, m_2, \ldots, m_k\right)=P\left\{X_1=m_1, \ldots, X_k=m_k\right\}=\frac{n!}{m_1!m_2!\ldots m_k!} p_1^{m_1} p_2^{m_2} \cdots p_k^{m_k} \qquad \textbf{(3.14)}$$

trong đó $0 \le m_i \le n,\ i=1,\ldots,n;\ m_1+\cdots+m_k=n;\ p_i>0,\ i=1,\ldots,n;\ p_1+\cdots+p_k=1$

Trường hợp $k=2$: $\left(X_1, X_2\right) \sim \mathrm{MUT}(n; p, 1-p)$ thì $X_1 \sim \mathcal{B}(n; p)$ và $X_2 \sim \mathcal{B}(n; 1-p)$.

Xét phép thử độc lập, lặp lại và mỗi lần thử có k kết quả ngẫu nhiên $A_1, \ldots, A_k$ tạo thành hệ đầy đủ biến cố; giả sử xác suất xuất hiện của biến cố A_i là p_i thỏa mãn $p_1+\cdots+p_k=1$.

Thực hiện n phép thử, gọi X_i là số lần xuất hiện của biến cố A_i trong n phép thử thì $X=\left(X_1, X_2, \ldots, X_k\right)$ có phân bố đa thức $X \sim \mathrm{MUT}\left(n; p_1, \ldots, p_k\right)$.

Chẳng hạn, gieo một con xúc xắc cân đối 10 lần. Tính các xác suất:

1) Có đúng 3 lần xuất hiện mặt 5 chấm (biến cố A).
2) Có 2 lần xuất hiện mặt 1 chấm, 4 lần mặt 3 chấm, 1 lần mặt 4 chấm và 3 lần mặt 6 chấm (biến cố B).

Giải: 1) Xét phép thử Bernoulli với thành công của mỗi lần thử là xuất hiện mặt có 5 chấm, vậy xác suất thành công mỗi lần thử là 1/6. Gọi X là số lần xuất hiện mặt 5 trong 10 lần thử thì X có phân bố nhị thức tham số $(10; 1/6)$, do đó

$$P(A)=P\{X=3\}=C_{10}^{3}\left(\frac{1}{6}\right)^{3}\left(\frac{5}{6}\right)^{7}=0{,}155.$$

2) Gọi X_i là số lần xuất hiện mặt i chấm trong 10 phép thử thì

$\left(X_1, X_2, X_3, X_4, X_5, X_6\right)$ có phân bố đa thức $\mathrm{MUT}(10; 1/6, 1/6, 1/6, 1/6, 1/6, 1/6)$.

$$P(B)=P\left\{X_1=2, X_2=0, X_3=4, X_4=1, X_5=0, X_6=3\right\}=\frac{10!}{2!0!4!1!0!3!}\left(\frac{1}{6}\right)^{10} \approx 0{,}0002.$$

3.2.3 Quy luật phân bố xác suất có điều kiện

Định nghĩa 3.7: *Giả sử* X *biến ngẫu nhiên rời rạc có miền giá trị* $R_X=\left\{x_1, x_2, \ldots, x_n\right\}$, B *là một biến cố trong cùng phép thử với* X *và có xác suất* $P(B)>0$. *Khi đó bảng phân bố xác suất* *của X với điều kiện B được xác định như sau*

$X\|B$	x_1	x_2	$\dots$	x_i	$\dots$	x_n
P	$p_{X\|B}(x_1\|B)$	$p_{X\|B}(x_2\|B)$	$\dots$	$p_{X\|B}(x_i\|B)$	$\dots$	$p_{X\|B}(x_n\|B)$

trong đó

$$p_{X|B}(x_i|B) = \frac{P\left(\{X = x_i\} \cap B\right)}{P(B)} \; ; \; i = 1, \dots, n \; . \tag{3.15}$$

Hàm $p_{X|B}(x_i|B)$ xác định bởi công thức (3.15) được gọi là ***hàm khối lượng xác suất của biến ngẫu nhiên X với điều kiện B.***

Định nghĩa 3.8: *Giả sử X, Y có tập các giá trị $R_X = \{x_1, x_2, \dots, x_n\}$ và $R_Y = \{y_1, y_2, \dots, y_m\}$. Với mỗi $y_j \in R_Y$ ta có bảng phân bố xác suất có điều kiện của X với điều kiện biến cố $\{Y = y_j\}$:*

$X\|Y = y_j$	x_1	x_2	$\dots$	x_i	$\dots$	x_n
P	$p_{X\|Y}(x_1\|y_j)$	$p_{X\|Y}(x_2\|y_j)$	$\dots$	$p_{X\|Y}(x_i\|y_j)$	$\dots$	$p_{X\|Y}(x_n\|y_j)$

trong đó

$$p_{X|Y}(x_i \mid y_j) = \frac{p_{X,Y}(x_i, y_j)}{p_Y(y_j)} \; ; \; i = 1, \dots, n \, . \tag{3.16}$$

Tương tự với mỗi $x_i \in R_X$ ta có bảng phân bố xác suất có điều kiện của Y với điều kiện $\{X = x_i\}$

$Y\|X = x_i$	y_1	y_2	$\dots$	y_j	$\dots$	y_m
P	$p_{Y\|X}(y_1\|x_i)$	$p_{Y\|X}(y_2\|x_i)$	$\dots$	$p_{Y\|X}(y_j\|x_i)$	$\dots$	$p_{Y\|X}(y_m\|x_i)$

trong đó

$$p_{Y|X}(y_j|x_i) = \frac{p_{X,Y}(x_i, y_j)}{p_X(x_i)} \; ; \; j = 1, \dots, m \, . \tag{3.17}$$

Ngược lại, từ công thức xác suất có điều kiện (3.16)-(3.17) ta có công thức tính xác suất đồng thời:

$$p_{X,Y}(x_i, y_j) = p_X(x_i) p_{Y|X}(y_j|x_i) = p_Y(y_j) p_{X|Y}(x_i|y_j) ; \; i = 1, \dots, n \; ; \; j = 1, \dots, m \, . \tag{3.18}$$

Tính chất 3.1:

a. $0 \le p_{X|Y}(x_i \mid y_j) \le 1$; $\forall i = 1, \dots, n$, $j = 1, \dots, m$.

b. Với mỗi j: $\sum_{i=1}^{n} p_{X|Y}(x_i \mid y_j) = 1$.

c. Nếu X, Y độc lập thì $p_{X|Y}(x_i \mid y_j) = p_X(x_i)$ và $p_{Y|X}(y_j \mid x_i) = p_Y(y_j)$.

Ví dụ 3.6: Cho X , Y là hai biến ngẫu nhiên rời rạc có phân bố xác suất đồng thời

$X \backslash Y$	0,5	2	4	$\sum$
0,1	0,07	0,12	0,08	0,27
0,2	0,11	0,18	0,19	0,48
0,4	0,05	0,13	0,07	0,25

a. Tìm bảng phân bố xác suất của các thành phần X và Y.

b. Tìm bảng phân bố xác suất của Y với điều kiện $X = 0,2$.

Giải: **a.** Cộng xác suất của bảng phân bố xác suất đồng thời ta được bảng phân bố xác suất thành phần của X và Y

X	0,1	0,2	0,4
P	0,27	0,48	0,25

Y	0,5	2	4
P	0,23	0,43	0,34

b. Bảng phân bố xác suất của Y với điều kiện $X = 0,2$

$Y \mid X = 0,2$	0,5	2	4
P	11/48	18/48	19/48

Ví dụ 3.7: Thực hiện lặp lại cùng một phép thử Bernoulli với xác suất xuất hiện của biến cố A trong mỗi lần thử là p, $0 < p < 1$. Gọi Y là biến ngẫu nhiên chỉ lần thử đầu tiên xuất hiện biến cố A. Gọi B là biến cố: "Trong n lần thử đầu tiên có duy nhất một lần xuất hiện biến cố A".

a. Tìm bảng phân bố xác suất của Y.

b. Tìm phân bố của Y với điều kiện B.

Giải: **a.** Ta có bảng phân bố xác suất của Y (xem ví dụ 2.9)

Y	1	2	…	k	…
P	p	qp	…	$q^{k-1}p$	…

với $q = 1 - p$.

b. Phân bố của Y với điều kiện B:

Khi $P(B) > 0$ thì $P\left(\{Y = k\} \middle| B\right) = \dfrac{P\left(\{Y = k\} \cap B\right)}{P(B)}$.

♦ Rõ ràng khi $k > n$ thì biến cố $\{Y = k\}$ kéo theo trong n phép thử đầu tiên biến cố A không xuất hiện. Do đó $\{Y = k\} \cap B = \varnothing$, vậy $P\left(\{Y = k\} \cap B\right) = 0$.

- Khi $k \le n$, áp dụng công thức Bernoulli ta có: $P(B) = C_n^1 pq^{n-1} = npq^{n-1}$.

Mặt khác $P\left(\left\{Y=k\right\} \cap B\right) = P\left\{\text{chØ xuÊt hiÖn biÕn cè } A \text{ ë lÇn thö thø } k\right\} = pq^{n-1}$

Vậy hàm khối lượng xác suất của Y với điều kiện B có dạng

$$P_{X|B}(k\mid B) = P\left(\left\{Y=k\right\}\middle|B\right) = \begin{cases} \dfrac{1}{n} & \text{nÕu} \quad 1 \le k \le n \\ 0 & \text{nÕu} \quad k > n. \end{cases}$$

3.2.4 Tính độc lập của các biến ngẫu nhiên

Hai biến ngẫu nhiên X và Y gọi là độc lập nếu mỗi biến ngẫu nhiên nhận giá trị này hay giá trị khác không ảnh hưởng gì đến phân bố xác suất của biến ngẫu nhiên kia (xem định nghĩa 2.2)

Xét véc tơ ngẫu nhiên rời rạc hai chiều (X,Y) có miền giá trị $R_X = \left\{x_1, x_2, ..., x_n\right\}$ và $R_Y = \left\{y_1, y_2, ..., y_m\right\}$ với hàm khối lượng xác suất đồng thời xác định theo công thức (3.7). Khi đó X và Y là độc lập khi và chỉ khi

$$p_{X,Y}(x_i, y_j) = p_X(x_i)\, p_Y(y_j) \quad \forall i = 1,...,n\ ;\ \forall j = 1,...,m. \tag{3.19}$$

Một dấu hiệu để nhận biết hai biến ngẫu nhiên rời rạc độc lập là bảng phân bố xác suất đồng thời có tính chất:

- Hai hàng bất kỳ tỉ lệ với nhau.
- Hai cột bất kỳ tỉ lệ với nhau. **(3.20)**

Ví dụ 3.8: Y là số mặt ngửa xuất hiện của cả 3 đồng tiền A, B, C phụ thuộc số mặt ngữa X của 2 đồng tiền A, B (xét trong ví dụ 3.3), do đó X và Y không độc lập. Mặt khác ta cũng thấy các hàng của bảng phân bố xác suất đồng thời không tỉ lệ hoặc có thể kiểm tra như sau:

$$P\left\{X=2\right\} = \frac{2}{8},\ P\left\{Y=1\right\} = \frac{3}{8} \text{ nhưng } P\left\{X=2, Y=1\right\} = 0 \ne P\left\{X=2\right\} P\left\{Y=1\right\}.$$

Hai biến ngẫu nhiên X và Y của ví dụ 3.4 độc lập. Ta cũng thấy bảng phân bố xác suất đồng thời thỏa mãn điều kiện (3.19), (3.20). Cụ thể các hàng của bảng phân bố xác suất đồng thời tỉ lệ nhau theo tỉ lệ 1:2:3.

3.3 CÁC THAM SỐ ĐẶC TRƯNG CỦA BIẾN NGẪU NHIÊN RỜI RẠC HAI CHIỀU

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

Từ bảng phân bố xác suất thành phần (3.10), (3.11) ta có công thức tính kỳ vọng và phương sai của biến ngẫu nhiên thành phần X, Y của véc tơ ngẫu nhiên (X,Y):

$$\mathrm{E}\,X = \sum_{i=1}^{n} x_i\, p_X(x_i) = \sum_{i=1}^{n}\sum_{j=1}^{m} x_i\, p_{X,Y}(x_i, y_j) \tag{3.21}$$

$$\mathrm{E}\,Y=\sum_{j=1}^{m}y_j p_Y(y_j)=\sum_{j=1}^{m}\sum_{i=1}^{n}y_j p_{X,Y}(x_i,y_j) \qquad \textbf{(3.22)}$$

$$\mathrm{E}\,X^2=\sum_{i=1}^{n}x_i^2 p_X(x_i);\ \mathrm{D}\,X=\mathrm{E}(X^2)-\left(\mathrm{E}\,X\right)^2 \qquad \textbf{(3.23)}$$

$$\mathrm{E}\,Y^2=\sum_{j=1}^{m}y_j^2 p_Y(y_j);\ \mathrm{D}\,Y=\mathrm{E}(Y^2)-\left(\mathrm{E}\,Y\right)^2 \qquad \textbf{(3.24)}$$

3.3.2 Hiệp phương sai

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

$$\mathrm{cov}(X,Y)=\mathrm{E}\left[\left(X-\mathrm{E}\,X\right)\left(Y-\mathrm{E}\,Y\right)\right] \qquad \textbf{(3.25)}$$

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

$$\mathrm{cov}(X,Y)=\mathrm{E}(XY)-(\mathrm{E}\,X)(\mathrm{E}\,Y) \qquad \textbf{(3.26)}$$

Nếu X,Y có hàm khối lượng xác suất đồng thời theo công thức (3.7) thì

$$\mathrm{E}\,(XY)=\sum_{j=1}^{m}\sum_{i=1}^{n}x_i y_j p_{X,Y}(x_i,y_j). \qquad \textbf{(3.27)}$$

Tính chất 3.2

1) $\mathrm{cov}(X,Y)=\mathrm{cov}(Y,X)$.

2) $\mathrm{cov}(X,X)=\mathrm{D}X$.

3) $\mathrm{cov}(aX+c,bY+d)=ab\,\mathrm{cov}(X,Y)$ với mọi hằng số a,b,c,d.

4) Từ công thức (3.26) và (2.21) suy ra rằng, nếu X,Y độc lập thì $\mathrm{cov}(X,Y)=0$. Tuy nhiên điều ngược lại không đúng, nghĩa là tồn tại hai biến ngẫu nhiên X, Y không độc lập nhưng $\mathrm{cov}(X,Y)=0$.

3.3.3 Hệ số tương quan

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

$$\rho_{X,Y}=\frac{\mathrm{cov}(X,Y)}{\sqrt{\mathrm{D}\,X}\sqrt{\mathrm{D}\,Y}}. \qquad \textbf{(3.28)}$$

Tính chất 3.3

1) $-1\le\rho_{X,Y}\le 1$ với mọi X,Y.

2) Nếu X,Y độc lập thì $\rho_{X,Y}=0$, nhưng điều ngược lại chưa chắc đúng.

3) Với mọi hằng số $a,b,c,d;\ a\neq 0$

$$\rho_{aX+c,bY+d}=\begin{cases}\rho_{X,Y} & \text{nÕu } ab>0\\ -\rho_{X,Y} & \text{nÕu } ab<0\end{cases} \tag{3.29}$$

4) $Y=aX+b$, $a\neq 0$ khi và chỉ khi

$$\rho_{X,Y}=\begin{cases}1 & \text{nÕu } a>0\\ -1 & \text{nÕu } a<0\end{cases} \tag{3.30}$$

Ý nghĩa của hệ số tương quan

Hệ số tương quan đo mức độ phụ thuộc tuyến tính giữa X và Y (Tính chất 3.2.4). Khi $\left|\rho_{X,Y}\right|$ càng gần 1 thì tính chất tương quan tuyến tính càng chặt, khi $\left|\rho_{X,Y}\right|$ càng gần 0 thì sự phụ thuộc tuyến tính càng ít, càng lỏng lẻo.

Khi $\rho_{X,Y}=0$ ta nói X và Y không tương quan.

Như vậy hai biến ngẫu nhiên độc lập thì không tương quan, nhưng ngược lại chưa chắc đúng (xem ví dụ 3.9).

Ví dụ 3.9: Xét biến ngẫu nhiên hai chiều (X,Y) ở ví dụ 3.3

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

Có bảng phân bố xác suất biên

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

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

$$\mathrm{E}\,X=0.\frac{2}{8}+1.\frac{4}{8}+2.\frac{2}{8}=1;\ \mathrm{E}\,X^2=0^2.\frac{2}{8}+1^2.\frac{4}{8}+2^2.\frac{2}{8}=\frac{3}{2}\Rightarrow \mathrm{D}\,X=\mathrm{E}\,X^2-(\mathrm{E}\,X)^2=\frac{1}{2}$$

$$\mathrm{E}\,Y=0.\frac{1}{8}+1.\frac{3}{8}+2.\frac{3}{8}+3.\frac{1}{8}=\frac{3}{2};\ \mathrm{E}\,Y^2=3\Rightarrow \mathrm{D}\,Y=\frac{3}{4}.$$

$$\mathrm{E}(XY)=0.0.\frac{1}{8}+0.1.\frac{1}{8}+0.2.0+0.3.0+1.0.0+1.1.\frac{2}{8}+1.2.\frac{2}{8}+1.3.0+2.0.0+2.1.0+2.2.\frac{1}{8}+2.3.\frac{1}{8}=2$$

Hiệp phương sai $\quad \operatorname{cov}(X,Y)=\mathrm{E}(XY)-(\mathrm{E}\,X)(\mathrm{E}\,Y)=2-1.\frac{3}{2}=\frac{1}{2}.$

Hệ số tương quan $\rho_{X,Y}=\dfrac{\operatorname{cov}(X,Y)}{\sqrt{\mathrm{D}\,X}\sqrt{\mathrm{D}\,Y}}=\dfrac{1/2}{\sqrt{(1/2)(3/4)}}=\sqrt{\dfrac{2}{3}}$.

Ví dụ 3.10: Xét biến ngẫu nhiên hai chiều (X,Y) ở ví dụ 3.4 với bảng phân bố xác suất đồng thời:

$X \backslash Y$	1	2	3
1	2/36	3/36	1/36
2	4/36	6/36	2/36
3	6/36	9/36	3/36

Có bảng phân bố xác suất biên

X	1	2	3
P	1/6	2/6	3/6

Y	1	2	3
P	2/6	3/6	1/6

$\mathrm{E}\,X=1.\dfrac{1}{6}+2.\dfrac{2}{6}+3.\dfrac{3}{6}=\dfrac{7}{3}$; $\mathrm{E}\,X^2=1^2.\dfrac{1}{6}+2^2.\dfrac{2}{6}+3^2.\dfrac{3}{6}=6 \Rightarrow \mathrm{D}\,X=\mathrm{E}\,X^2-(\mathrm{E}\,X)^2=\dfrac{5}{9}$

$\mathrm{E}\,Y=1.\dfrac{2}{6}+2.\dfrac{3}{6}+3.\dfrac{1}{6}=\dfrac{11}{6}$; $\mathrm{E}\,Y^2=1^2.\dfrac{2}{6}+2^2.\dfrac{3}{6}+3^2.\dfrac{1}{6}=\dfrac{23}{6} \Rightarrow \mathrm{D}\,Y=\dfrac{17}{36}$.

$$\mathrm{E}(XY)=1.1.\frac{2}{36}+1.2.\frac{3}{36}+1.3.\frac{1}{36}+2.1.\frac{4}{36}+2.2.\frac{6}{36}+2.3.\frac{2}{36}+3.1.\frac{6}{36}+3.2.\frac{9}{36}+3.3.\frac{3}{36}=\frac{77}{18}$$

Hiệp phương sai $\operatorname{cov}(X,Y)=\mathrm{E}(XY)-\mathrm{E}\,X\,\mathrm{E}\,Y=\dfrac{77}{18}-\dfrac{7}{3}\cdot\dfrac{11}{6}=0$.

Hệ số tương quan $\rho_{X,Y}=\dfrac{\operatorname{cov}(X,Y)}{\sqrt{\mathrm{D}\,X}\sqrt{\mathrm{D}\,Y}}=\dfrac{0}{\sqrt{(5/9)(17/36)}}=0$.

Điều này phù hợp với tính chất độc lập của hai biến ngẫu nhiên X, Y.

Ví dụ 3.11: Xét véc tơ ngẫu nhiên (X,Y) có bảng phân bố xác suất đồng thời

$X \backslash Y$	-1	0	1
-1	4/15	1/15	4/15
0	1/15	2/15	1/15
1	0	2/15	0

Có bảng phân bố xác suất biên

X	-1	0	1
P	9/15	4/15	2/15

Y	-1	0	1
P	1/3	1/3	1/3

$$\mathrm{E}\,X = -1.\frac{9}{15} + 0.\frac{4}{15} + 1.\frac{2}{15} = -\frac{7}{15};$$

$$\mathrm{E}\,X = (-1)^2.\frac{9}{15} + 0^2.\frac{4}{15} + 1^2.\frac{2}{15} = \frac{11}{15} \Rightarrow \mathrm{D}\,X = \mathrm{E}\,X^2 - (\mathrm{E}\,X)^2 = \frac{116}{225}$$

$$\mathrm{E}\,Y = (-1).\frac{1}{3} + 0.\frac{1}{3} + 1.\frac{1}{3} = 0;\ \mathrm{E}\,Y^2 = \frac{2}{3} \Rightarrow \mathrm{D}\,Y = \frac{2}{3}.$$

$$\mathrm{E}(XY) = (-1).(-1).\frac{4}{15} + (-1).0.\frac{1}{15} + (-1).1.\frac{4}{15}$$

$$+0.(-1).\frac{1}{15} + 0.0.\frac{2}{15} + 0.1.\frac{1}{15} + 1.(-1).0 + 1.0.\frac{2}{15} + 1.1.0 = 0$$

Hiệp phương sai $\quad \mathrm{cov}(X,Y) = \mathrm{E}(XY) - (\mathrm{E}\,X)(\mathrm{E}\,Y) = 0 - \frac{-7}{15}.0 = 0.$

Hệ số tương quan $\quad \rho_{X,Y} = \frac{\mathrm{cov}(X,Y)}{\sqrt{\mathrm{D}\,X}\sqrt{\mathrm{D}Y}} = \frac{0}{\sqrt{(116/225)(2/3)}} = 0.$

Ma trận hiệp phương sai $\quad M = \begin{bmatrix} 116/225 & 0 \\ 0 & 2/3 \end{bmatrix}.$

Vì các hàng của bảng phân bố xác suất đồng thời không tỉ lệ nên hai biến ngẫu nhiên X, Y không độc lập, mặc dù hiệp phương sai $\mathrm{cov}(X,Y) = 0$.

3.3.4 Kỳ vọng có điều kiện, hàm hồi quy

Từ luật phân bố điều kiện (3.15)- (3.17) ta có thể định nghĩa và ký hiệu kỳ vọng

Định nghĩa 3.11: *Kỳ vọng của X với điều kiện B được định nghĩa:*

$$\mathrm{E}\left[X\middle|B\right] = \sum_{i=1}^{n} x_i\, p_{X|B}(x_i|B) \qquad \textbf{(3.31)}$$

Định nghĩa 3.12: *Kỳ vọng có điều kiện của Y với điều kiện $X = x_i$, $x_i \in R_X$* được tính theo công thức sau

$$\mathrm{E}\left[Y\middle|X = x_i\right] = \sum_{j=1}^{m} y_j\, p_{Y|X}(y_j \mid x_i) \qquad \textbf{(3.32)}$$

Khi cho giá trị x_i thay đổi thì kỳ vọng có điều kiện của Y phụ thuộc vào giá trị của X gọi là hàm hồi quy của Y đối với X.

$$f(x)=\begin{cases}\mathrm{E}\left[Y\middle|X=x_i\right] & \text{nÕu } x=x_i\in R_X\\ 0 & \text{nÕu } x\notin R_X\end{cases}\tag{3.33}$$

Tương tự ta có định nghĩa kỳ vọng có điều kiện của X với điều kiện $\{Y=y_j\}$, $y_j\in R_Y$ và hàm hồi quy của X đối với Y:

$$\mathrm{E}\left[X\middle|Y=y_j\right]=\sum_{i=1}^{n}x_i\,p_{X|Y}(x_i\,|\,y_j)\tag{3.34}$$

$$g(y)=\begin{cases}\mathrm{E}\left[X\middle|Y=y_j\right] & \text{nÕu } y=y_j\in R_Y\\ 0 & \text{nÕu } y\notin R_Y\end{cases}\tag{3.35}$$

Ví dụ 3.12: Thống kê dân cư của một thành phố nọ ở độ tuổi trưởng thành về thu nhập hàng tháng X và lứa tuổi Y thu được kết quả trong bảng sau.

X \ Y	30	45	70
2	0,01	0,02	0,05
4	0,03	0,06	0,10
6	0,18	0,21	0,15
8	0,07	0,08	0,04

trong đó $X=2,4,6,8$ tương ứng chỉ thu nhập triệu đồng /tháng.

$Y=30,\ 45,\ 70$ chỉ độ tuổi của người dân trong khoảng: 25-35, 35-55, 55-85.

Tìm thu nhập trung bình theo lứa tuổi.

Giải: Thu nhập trung bình theo lứa tuổi là kỳ vọng có điều kiện của X theo Y.

Với $Y=30$ bảng phân bố xác suất điều kiện tương ứng:

$X\|Y=30$	2	4	6	8
P	$\frac{0{,}01}{0{,}29}$	$\frac{0{,}03}{0{,}29}$	$\frac{0{,}18}{0{,}29}$	$\frac{0{,}07}{0{,}29}$

Từ đó $\mathrm{E}\left[X\middle|Y=30\right]=2\cdot\frac{1}{29}+4\cdot\frac{3}{29}+6\cdot\frac{18}{29}+8\cdot\frac{7}{29}=\frac{178}{29}=6{,}138$.

Tương tự $\mathrm{E}\left[X\middle|Y=45\right]=5{,}892$; $\mathrm{E}\left[X\middle|Y=70\right]=5{,}058$.

Vậy thu nhập trung bình:

độ tuổi 30 là 6.138.000đ/tháng,

độ tuổi 45 là 5.892.000đ/tháng

và độ tuổi 70 là 5.058.000 đ/tháng.

3.4 LUẬT SỐ LỚN VÀ ĐỊNH LÝ GIỚI HẠN TRUNG TÂM

Lý thuyết xác suất nghiên cứu khả năng xuất hiện của các biến cố ngẫu nhiên. Khi tung một đồng xu ta sẽ không biết mặt sấp hay mặt ngửa xuất hiện nhưng nếu tung nhiều lần thì ta thấy rằng số lần mặt sấp và mặt ngửa xuất hiện là xấp xỉ gần bằng nhau. Như vậy khi thực hiện nhiều lần phép thử ngẫu nhiên ta sẽ tìm được quy luật xuất hiện của biến cố ngẫu nhiên, đấy là nội dung của luật số lớn. Luật số lớn cũng là cơ sở để định nghĩa xác suất của biến cố thông qua tần suất xuất hiện của biến cố đó.

Luật số lớn nghiên cứu sự hội tụ theo xác suất của dãy các biến ngẫu nhiên.

Luật số lớn đầu tiên của James Bernoulli được công bố năm 1713. Về sau, kết quả này được Poisson, Trêbưsép, Markov, Liapunốp mở rộng.

Trong mục này ta xét hai định lý về luật số lớn. Định lý Trêbưsép là dạng tổng quát của luật số lớn và định lý Bernoulli là trường hợp đơn giản nhất của luất số lớn áp dụng cho các biến ngẫu nhiên có phân bố Bernoulli. Để chứng minh định lý Trêbưsép ta sử dụng bất đẳng thức Trêbưsép.

3.4.1 Bất đẳng thức Markov và bất đẳng thức Trêbưsép

Định lý 3.1: Cho Y là biến ngẫu nhiên không âm có kỳ vọng hữu hạn. Khi đó với mọi $a>0$ ta có:

$$P\{Y\ge a\}\le\frac{\mathrm{E}Y}{a}. \tag{3.36}$$

Chứng minh:

a. Trường hợp Y rời rạc có tập giá trị $R_Y=\{y_1,y_2,...\}$.

Đặt $R_1=\{y_i\in R_Y, y_i<a\}$; $R_2=\{y_i\in R_Y, y_i\ge a\}$.

$$\mathrm{E}Y=\sum_{y_i\in R_Y} y_iP\{Y=y_i\}\ (*)$$

- Trường hợp tổng (*) có hữu hạn các số hạng thì đương nhiên có thể thay đổi thứ tự lấy tổng
- Trường hợp tổng (*) có vô hạn số hạng thì đây là tổng của một chuỗi số dương do đó có thể thay đổi thứ tự, vì vậy có thể viết lại:

$$\mathrm{E}Y=\sum_{y_i\in R_Y} y_iP\{Y=y_i\}=\sum_{y_i\in R_1} y_iP\{Y=y_i\}+\sum_{y_i\in R_2} y_iP\{Y=y_i\}$$

$$\ge\sum_{y_i\in R_2} y_iP\{Y=y_i\}\ge a\sum_{y_i\in R_2} P\{Y=y_i\}=aP\{Y\ge a\}.$$

Suy ra $\quad P\{Y\ge a\}\le\dfrac{\mathrm{E}Y}{a}$.

b. Giả sử Y liên tục có hàm mật độ xác suất $f_Y(y)$. Y là biến ngẫu nhiên không âm do đó $f_Y(y)=0, \forall y\le 0$, vì vậy

$$\mathrm{E}Y=\int_0^{+\infty} yf_Y(y)dy=\int_0^{a} yf_Y(y)dy+\int_a^{+\infty} yf_Y(y)dy\ge\int_a^{+\infty} yf_Y(y)dy\ge a\int_a^{+\infty} f(y)dy=aP\{Y\ge a\}.$$

Suy ra $\quad P\{Y \ge a\} \le \dfrac{\mathrm{E}Y}{a}$.

Định lý 3.2: Giả sử X là biến ngẫu nhiên có kỳ vọng và phương sai hữu hạn, khi đó với mọi $\varepsilon > 0$ ta có:

$$P\{|X - \mathrm{E}\,X| > \varepsilon\} \le \frac{\mathrm{D}X}{\varepsilon^2}. \qquad \textbf{(3.37)}$$

Sử dụng công thức xác suất biến cố đối ta cũng có

$$P\{|X - \mathrm{E}\,X| \le \varepsilon\} \ge 1 - \frac{\mathrm{D}X}{\varepsilon^2}. \qquad \textbf{(3.38)}$$

Chứng minh: Áp dụng công thức (3.36) cho biến ngẫu nhiên $Y = \left(X - \mathrm{E}\,X\right)^2$ và $a = \varepsilon^2$ ta có:

$$P\{|X - \mathrm{E}\,X| > \varepsilon\} = P\left\{Y > \varepsilon^2\right\} \le \frac{\mathrm{E}\,Y}{\varepsilon^2} = \frac{\mathrm{E}\left(X - \mathrm{E}\,X\right)^2}{\varepsilon^2} = \frac{\mathrm{D}\,X}{\varepsilon^2}.$$

Bất đẳng thức (3.37)-(3.38) được gọi là bất đẳng thức Trêbưsép.

Bất đẳng thức Trêbưsép có nhiều ứng dụng. Trước hết nó cho phép ta đánh giá cận trên hoặc cận dưới xác suất để biến ngẫu nhiên X nhận giá trị sai lệch so với kỳ vọng $\mathrm{E}\,X$ không quá ε. Bất đẳng thức Trêbưsép có ý nghĩa to lớn về mặt lý thuyết, nó được sử dụng để chứng minh các định lý của luật số lớn.

3.4.2 Hội tụ theo xác suất

Định nghĩa 3.13: Dãy các biến ngẫu nhiên trong cùng một phép thử $X_1, X_2, \dots$ gọi là hội tụ theo xác suất về biến ngẫu nhiên X (của cùng phép thử), ký hiệu $X_n \xrightarrow[n\to\infty]{P} X$, nếu:

$$\forall \varepsilon > 0:\ \lim_{n\to\infty} P\{|X_n - X| > \varepsilon\} = 0. \qquad \textbf{(3.39)}$$

Như vậy dãy các biến ngẫu nhiên $X_1, X_2, \dots$ hội tụ theo xác suất về biến ngẫu nhiên X thì với n đủ lớn, thực tế gần như chắc chắn ta có thể coi rằng X_n không khác mấy so với X.

3.4.3 Luật số lớn Trêbưsép

Định lý 3.3: Giả sử $X_1, X_2, \dots$ là dãy các biến ngẫu nhiên độc lập trong cùng một phép thử, có các kỳ vọng hữu hạn và phương sai đều bị chặn trên bởi hằng số C ($\mathrm{D}\,X_k \le C;\ \forall k = 1, 2, \dots$). Khi đó với mọi $\varepsilon > 0$:

$$\lim_{n\to\infty} P\left(\left|\frac{X_1 + \cdots + X_n}{n} - \frac{\mathrm{E}\,X_1 + \cdots + \mathrm{E}\,X_n}{n}\right| > \varepsilon\right) = 0 \qquad \textbf{(3.40)}$$

Chứng minh: Xét biến ngẫu nhiên $S_n = \dfrac{X_1 + \cdots + X_n}{n}$. Từ giả thiết độc lập của dãy các biến ngẫu nhiên $X_1, X_2, \dots$ ta suy ra :

$$\operatorname{E} S_n = \frac{\operatorname{E} X_1 + \cdots + \operatorname{E} X_n}{n};\ \operatorname{D} S_n = \frac{\operatorname{D} X_1 + \cdots + \operatorname{D} X_n}{n^2} \le \frac{C}{n}.$$

Áp dụng bất đẳng thức Trêbưsép (3.37) cho biến ngẫu nhiên S_n ta có:

$$P\left(\left|\frac{X_1 + \cdots + X_n}{n} - \frac{\operatorname{E} X_1 + \cdots + \operatorname{E} X_n}{n}\right| > \varepsilon\right) \le \frac{C}{n\varepsilon^2} \xrightarrow[n\to\infty]{} 0.$$

Hệ quả 3.1: Giả sử $X_1, X_2, \ldots$ là dãy các biến ngẫu nhiên độc lập trong cùng một phép thử có cùng có kỳ vọng μ và phương sai đều bị chặn trên bởi hằng số C ($\operatorname{D} X_k \le C;\ \forall k = 1, 2, \ldots$). Khi đó

$$\frac{X_1 + \cdots + X_n}{n} \xrightarrow[n\to\infty]{P} \mu \tag{3.41}$$

Hệ quả 3.2: Giả sử $X_1, X_2, \ldots$ là dãy các biến ngẫu nhiên độc lập trong cùng một phép thử có cùng phân bố, có kỳ vọng μ và phương sai σ^2. Khi đó

$$\frac{X_1 + \cdots + X_n}{n} \xrightarrow[n\to\infty]{P} \mu \tag{3.42}$$

Định lý Trêbưsép chứng tỏ rằng trung bình số học của các biến ngẫu nhiên độc lập hội tụ theo xác suất về trung bình số học của kỳ vọng tương ứng của nó. Nói cách khác nó chứng tỏ sự ổn định của trung bình số học của một số lớn các biến ngẫu nhiên xung quanh trung bình số học của các kỳ vọng của các biến ngẫu nhiên ấy. Như vậy mặc dù từng biến ngẫu nhiên độc lập có thể nhận giá trị khác nhiều so với kỳ vọng của chúng, song trung bình số học của một số lớn các biến ngẫu nhiên lại nhận giá trị gần bằng trung bình số học các kỳ vọng của chúng với xác suất rất lớn. Điều đó cho phép dự đoán giá trị trung bình số học của các biến ngẫu nhiên. Chẳng hạn, gieo một con xúc xắc cân đối, gọi X là số chấm xuất hiện ở mặt trên con xúc xắc. Theo lý thuyết ta tính được $\operatorname{E} X = 3,5$. Một nhà thống kê đã gieo một con xúc xắc cân đối 1 triệu lần (nhờ sự trợ giúp của máy vi tính) và ghi lại số chấm xuất hiện ở mặt trên con xúc xắc. Số trung bình của 1 triệu lần gieo được tìm thấy là

$$\frac{x_1 + \cdots + x_{10^6}}{10^6} \approx 3,500867 \approx 3,5.$$

Định lý Trêbưsép có ứng dụng rộng rãi trong nhiều lĩnh vực, chẳng hạn nó chính là cơ sở cho phương pháp đo lường trong vật lý. Để xác định giá trị của một đại lượng vật lý nào đó người ta thường tiến hành đo n lần độc lập và lấy trung bình số học của các kết quả đo làm giá trị thực của đại lượng cần đo. Thật vậy, giả sử xem kết quả của n lần đo là các biến ngẫu nhiên $X_1, X_2, \ldots, X_n$. Ta thấy rằng các biến ngẫu nhiên này độc lập, có cùng kỳ vọng bằng chính giá trị thực của đại lượng vật lý (giả sử không có sai số hệ thống), các phương sai của chúng đều bị chặn trên bởi bình phương của độ chính xác của thiết bị đo. Do đó theo định lý Trêbưsép ta có thể cho rằng trung bình số học của các kết quả đo sẽ sai lệch rất ít so với giá trị thực của đại lượng vật lý với xác suất gần như bằng một.

Định lý Trêbưsép còn là cơ sở cho phương pháp mẫu ứng dụng trong thống kê.

3.4.4 Luật số lớn Bernoulli

Xét phép thử ngẫu nhiên C và A là một biến cố liên quan đến phép thử C. Tiến hành n lần độc lập phép thử C và gọi k_n là tần số xuất hiện biến cố A trong n phép thử đó. $f_n = \dfrac{k_n}{n}$ được gọi là tần suất xuất hiện của A trong n phép thử.

Định lý 3.4: (Định lý Bernoulli) Tần suất f_n hội tụ theo xác suất về xác suất p của biến cố A, nghĩa là với mọi $\varepsilon > 0$

$$\lim_{n\to\infty} P\left\{\left| f_n - p\right| \le \varepsilon\right\} = 1 \qquad \textbf{(3.43)}$$

Chứng minh: Xét dãy các biến ngẫu nhiên $X_1, X_2, ..., X_n, ...$ xác định như sau:

$$X_k = \begin{cases} 1 & \text{nÕu } A \text{ x¶y ra ë phÐp thö thø } k \\ 0 & \text{nÕu } A \text{ kh«ng x¶y ra ë phÐp thö thø } k \end{cases}$$

thì dãy các biến ngẫu nhiên $X_1, X_2, ..., X_n, ...$ độc lập có cùng phân bố Bernoulli tham số p. Có kỳ vọng là $\mathrm{E}\, X_k = p$ và phương sai $\mathrm{D}\, X_k = p(1-p) < 1$ với mọi $k = 1, 2, ...$

Ta có $$\frac{X_1 + \cdots + X_n}{n} = \frac{k_n}{n} = f_n .$$

Vậy theo hệ quả 3.2 của định lý 3.3 suy ra f_n hội tụ theo xác suất về p.

Định lý Bernoulli chỉ ra rằng tần suất xuất hiện của biến cố trong n phép thử độc lập sẽ hội tụ theo xác suất về xác suất của biến cố đó khi số lần thử tăng lên vô hạn. Chính vì vậy định lý Bernoulli là cơ sở lý thuyết của định nghĩa thống kê về xác suất.

Ở thế kỷ 18, nhà toán học Pháp Buffon gieo một đồng tiền 4040 lần và ghi được 2048 lần xuất hiện mặt ngửa, tần suất là 0,507. Một nhà thống kê người Anh gieo đồng tiền 12000 lần và thu được 6019 lần xuất hiện mặt ngửa, tần suất tương ứng 0,5016. Trong một thí nghiệm khác, ông ta gieo 24000 lần và thu được 12012 lần xuất hiện mặt ngửa, tần suất tương ứng là 0,5005. Như vây ta thấy rằng khi số phép thử tăng lên thì tần suất tương ứng sẽ càng gần 0,5.

3.4.5 Định lý giới hạn trung tâm

Giả sử $X_1, X_2, ...$ là dãy các biến ngẫu nhiên độc lập có cùng phân bố, có kỳ vọng μ và phương sai σ^2. Theo công thức (2.24), (2.34) ta có

$$\mathrm{E}\left[X_1 + \cdots + X_n\right] = n\mu \quad \text{và} \quad \mathrm{D}\left[X_1 + \cdots + X_n\right] = n\sigma^2$$

Do đó

$$S_n = \frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} = \frac{X_1 + \cdots + X_n - \mathrm{E}\left[X_1 + \cdots + X_n\right]}{\sqrt{\mathrm{D}\left[X_1 + \cdots + X_n\right]}} \text{ có } \mathrm{E}\, S_n = 0 \text{ và } \mathrm{D}\, S_n = 1.$$

Định lý 3.5: Giả sử $X_1, X_2, ...$ là dãy các biến ngẫu nhiên độc lập có cùng phân bố, có kỳ vọng μ và phương sai σ^2. Khi đó dãy biến ngẫu nhiên $S_n = \dfrac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}$ hội tụ theo phân bố về phân bố chuẩn tắc $\mathbf{N}(0;1)$, tức là:

$$\text{Với mọi } x \in \mathbb{R},\quad \lim_{n\to\infty} F_{S_n}(x) = \lim_{n\to\infty} P\{S_n \le x\} = \Phi(x) \qquad \textbf{(3.44)}$$

$\Phi(x)$ là hàm phân bố xác suất của phân bố chuẩn tắc $\mathbf{N}(0;1)$.

Áp dụng định lý giới hạn trung tâm cho dãy các biến ngẫu nhiên độc lập $X_1, X_2, \ldots$ có cùng phân bố Bernoulli tham số p (công thức (2.46)) ta được định lý Moivre –Laplace:

Định lý 3.6 (Moivre –Laplace): Đối với dãy các biến ngẫu nhiên $X_1, X_2, \ldots$ độc lập có cùng phân bố Bernoulli tham số p thì:

$$\text{Với mọi } x \in \mathbb{R},\ \lim_{n\to\infty} P\left\{\frac{X_1 + \cdots + X_n - np}{\sqrt{npq}} \le x\right\} = \Phi(x). \qquad \textbf{(3.45)}$$

Định lý Moivre-Laplace cho phép xấp xỉ phân bố nhị thức $\mathsf{B}(n;p)$ với phân bố chuẩn $\mathbf{N}(np;npq)$ khi n đủ lớn (công thức 2.78). Người ta thấy rằng xấp xỉ là tốt khi np và nq lớn hơn 5 hoặc khi $npq > 20$.

CÂU HỎI ÔN TẬP VÀ BÀI TẬP CHƯƠNG 3

3.1 Từ bảng phân bố xác suất của thành phần X và Y luôn xác định phân bố xác suất đồng thời của (X,Y).

Đúng ☐ Sai ☐.

3.2 Bảng phân bố xác suất đồng thời của (X,Y) xác định phân bố xác suất của từng biến ngẫu nhiên thành phần X và Y.

Đúng ☐ Sai ☐.

3.3 Nếu hai biến ngẫu nhiên X, Y độc lập thì bảng phân bố xác suất của X và Y cho phép xác định phân bố xác suất đồng thời của (X,Y).

Đúng ☐ Sai ☐.

3.4 Hai biến ngẫu nhiên độc lập có hiệp phương sai bằng 0.

Đúng ☐ Sai ☐.

3.5 Hai biến ngẫu nhiên có hiệp phương sai bằng 0 thì độc lập.

Đúng ☐ Sai ☐.

3.6 Hiệp phương sai luôn nhận giá trị dương.

Đúng ☐ Sai ☐.

3.7 Nếu $Y = aX + b$, $a \ne 0$ thì hệ số tương quan $\rho_{X,Y}$ luôn luôn bằng 1.

Đúng ☐ Sai ☐.

3.8 Với mọi biến ngẫu nhiên X, Y, Z trong cùng một phép thử và với mọi hằng số α, β ta có: $\mathrm{E}\left[\alpha Y + \beta Z \middle| X = x_i\right] = \alpha\,\mathrm{E}\left[Y \middle| X = x_i\right] + \beta\,\mathrm{E}\left[Z \middle| X = x_i\right]$.

Đúng ☐ Sai ☐.

3.9 Nếu hai biến ngẫu nhiên X, Y độc lập thì hàm hồi quy $f(x) = \mathrm{E}\left[Y \middle| X = x\right]$ của Y đối với X và hàm hồi quy $g(y) = \mathrm{E}\left[X \middle| Y = y\right]$ của X đối với Y là hai hàm hằng trong miền giá trị R_X và R_Y.

Đúng ☐ Sai ☐.

3.10 Nếu hiệp phương sai của hai biến ngẫu nhiên bằng 0 thì hai kỳ vọng của chúng bằng nhau.

Đúng ☐ Sai ☐.

3.11 Luật số lớn kết luận về sự hội tụ theo xác suất của trung bình cộng các biến ngẫu nhiên độc lập về trung bình cộng của kỳ vọng của chúng nếu các phương sai của các biến ngẫu nhiên này bị chặn.

Đúng ☐ Sai ☐.

3.12 Giả sử $\{X_n\}$ là dãy các biến ngẫu nhiên độc lập có kỳ vọng bằng nhau và phương sai dần tới 0, khi đó dãy sẽ hội tụ theo xác suất đến kỳ vọng chung của dãy biến ngẫu nhiên trên.

Đúng ☐ Sai ☐.

3.13 Bất đẳng thức Trêbưsép chỉ đúng đối với các biến ngẫu nhiên rời rạc.

Đúng ☐ Sai ☐.

3.14 Bất đẳng thức Trêbưsép chỉ đúng đối với các biến ngẫu nhiên nhận giá trị dương.

Đúng ☐ Sai ☐.

3.15 Luật số lớn Bernoulli là một trường hợp đặc biệt của luật số lớn Trêbưsép khi dãy các biến ngẫu nhiên được xét có cùng phân bố Bernoulli tham số p.

Đúng ☐ Sai ☐.

3.16 Luật số lớn Bernoulli là cơ sở lý thuyết của định nghĩa thống kê về xác suất.

Đúng ☐ Sai ☐.

3.17 Cho X, Y là hai biến ngẫu nhiên có phân bố xác suất đồng thời như sau

X \ Y	y_1	y_2	y_3
x_1	0,18	0,22	0,16
x_2	0,08	0,16	0,20

Tìm phân bố xác suất của hai biến ngẫu nhiên thành phần X, Y.

3.18 Cho X, Y là hai biến ngẫu nhiên có phân bố xác suất đồng thời như sau

X \ Y	-1	1
-1	1/6	1/4
0	1/6	1/8
1	1/6	1/8

Hãy tính $\mathrm{E}X, \mathrm{E}Y$, $\mathrm{cov}(X,Y)$ và $\rho_{X,Y}$.

3.19 Cho X, Y là hai biến ngẫu nhiên có phân bố xác suất đồng thời như sau

$X \backslash Y$	-1	0	1
-1	4/15	1/15	4/15
0	1/15	2/15	1/15
1	0	2/15	0

Hãy tính $\mathrm{E}X, \mathrm{E}Y$; $\mathrm{cov}(X,Y)$ và $\rho_{X,Y}$. X, Y có độc lập không?

3.20 Cho X, Y là hai biến ngẫu nhiên có phân bố xác suất đồng thời như sau

$X \backslash Y$	1	2	3
1	0,12	0,15	0,03
2	0,28	0,35	0,07

a. Chứng minh rằng X, Y có độc lập.

b. Tìm quy luật phân bố của biến ngẫu nhiên $Z = XY$.

c. Tính các kỳ vọng $\mathrm{E}X, \mathrm{E}Y, \mathrm{E}Z$.

3.21 Cho X, Y là hai biến ngẫu nhiên độc lập có phân bố xác suất như sau:

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

Y	0	1	2	3	4
P	0,1	0,3	0,4	0,15	0,05

Tìm phân bố xác suất đồng thời của X, Y. Tính xác suất $P\{X > Y\}$.

3.22 Cho X, Y là hai biến ngẫu nhiên có phân bố xác suất đồng thời như sau

$X \backslash Y$	1	2	3
1	0,15	0,20	0,10
2	0,35	0,05	0,15

Hai biến ngẫu nhiên X, Y có độc lập không. Tính xác suất $P\{X = 1 | Y = 2\}$.

3.23 Gieo đồng thời một con xúc xắc và một đồng tiền. Gọi X là biến ngẫu nhiên chỉ số chấm của con xúc xắc và Y là biến ngẫu nhiên chỉ mặt sấp (1) hay mặt ngửa (0) của đồng tiền. Lập bảng phân bố xác suất đồng thời của X và Y. X và Y có độc lập không.

3.24 Cho bảng phân bố xác suất đồng thời của X, Y

$X \backslash Y$	26	30	41	50
23	0,05	0,08	0,12	0,04
27	0,09	0,30	0,11	0,21

Tìm bảng phân bố xác suất điều kiện của Y khi $X = 27$ và của X khi $Y = 26$.

3.25 Cho bảng phân bố xác suất đồng thời của X, Y

$X \backslash Y$	1	3	4	8
3	0,15	0,06	0,25	0,04
6	0,30	0,10	0,03	0,07

a. Tìm kỳ vọng có điều kiện của X khi $Y = 1$.

b. Tìm các kỳ vọng $\mathrm{E}X, \mathrm{E}Y$ và phương sai $\mathrm{D}X, \mathrm{D}Y$.

3.26 Gieo hai con xúc sắc cân đối. Đặt $X = 0$ hoặc 1 tương ứng với con xúc sắc thứ nhất ra mặt chẵn hay lẻ. Tương tự $Y = 0$ hoặc 1 tương ứng với con xúc sắc thứ hai ra mặt chẵn hay lẻ. Tìm bảng phân bố xác suất đồng thời của X và Y.

3.27 Cho X, Y là hai biến ngẫu nhiên có phân bố xác suất đồng thời

$X \backslash Y$	-1	0	1
-1	4α	α	4α
0	α	2α	α
1	0	2α	0

a. Tìm α. Tính $\mathrm{E}X$, $\mathrm{E}Y$.

b. Tính $\mathrm{cov}(X,Y)$, $\rho(X,Y)$.

c. X và Y có độc lập không.

3.28 Lấy ngẫu nhiên 3 quả cầu từ một túi đựng 2 quả cầu mầu đỏ, 3 quả cầu mầu trắng và 4 quả cầu mầu xanh. Xét véc tơ ngẫu nhiên hai chiều (X,Y), trong đó X là số bi mầu đỏ và Y là số bi mầu trắng chọn được.

a. Tìm bảng phân bố xác suất đồng thời của X và Y.

b. Tìm bảng phân bố xác suất thành phần X và thành phần Y.

c. X và Y có độc lập không.

3.29 Tung hai đồng xu 3 lần. Đồng xu A cân đối nhưng đồng xu B không cân đối với xác suất xuất hiện mặt sấp là $1/4$ và mặt ngửa là $3/4$. Xét véc tơ ngẫu nhiên hai chiều (X,Y), trong đó

X là lần mặt sấp của A xuất hiện và Y là lần mặt sấp của B xuất hiện.

a. Tìm bảng phân bố xác suất đồng thời của X và Y.

b. Tính các xác suất: $P\{X=Y\}$, $P\{X>Y\}$ và $P\{X+Y\le 4\}$.

3.30 Có 10 máy hoạt động độc lập với nhau. Xác suất để trong ca làm việc mỗi máy bị hỏng là 0,05. Dựa vào bất đẳng thức Trêbưsép hãy đánh giá xác suất của sự sai lệch giữa số máy hỏng và số máy hỏng trung bình.

a. Nhỏ hơn 2.

b. Lớn hơn 2

3.31 Cho $X_1, X_2, ..., X_{12}$ là các biến ngẫu nhiên độc lập với $\mathrm{E}X_i = 16$, $\mathrm{D}X_i = 1$ ($i = 1,...,12$). Sử dụng bất đẳng thức Trêbưsép để tìm hai hằng số a, b sao cho

$$P\left\{a \le \sum_{i=1}^{12} X_i \le b\right\} \ge 0{,}99.$$

3.32 Cho $X_1, X_2, ..., X_{10000}$ là các biến ngẫu nhiên độc lập có phân bố đều trong đoạn $\left[-\frac{1}{2}, \frac{1}{2}\right]$

Chứng minh rằng $P\left\{\left|\sum_{i=1}^{10000} X_i\right| \ge 500\right\} \ge \frac{1}{300}$.

3.33 Gieo một con xúc xắc cân đối n lần một cách độc lập. Gọi S là số lần xuất hiện mặt lục. Chứng minh rằng $P\left\{\frac{n}{6} - \sqrt{n} < S < \frac{n}{6} + \sqrt{n}\right\} \ge \frac{31}{36}$.

3.34 Giả sử tiền điện của một gia đình phải trả trong 1 tháng là một biến ngẫu nhiên với trung bình 16 USD và độ lệch tiêu chuẩn 1 USD. Sử dụng bất đẳng thức Trêbưsép, hãy xác định số M nhỏ nhất để với xác suất 0,99 số tiền điện phải trả trong 1 năm (12 tháng) không vượt quá M.

3.35 Xác suất chậm tầu của mỗi hành khách là 0,007. Dùng bất đẳng thức Trêbưsép hãy đánh giá xác suất để trong 20.000 hành khách có từ 100 đến 180 người chậm tầu.

3.36 Phải kiểm tra bao nhiêu chi tiết để với xác suất không nhỏ hơn 0,98 có thể hy vọng rằng sai lệch giữa tần suất xuất hiện chi tiết tốt và xác suất để chi tiết là tốt bằng 0,95 sẽ không vượt quá 0,01.

CHƯƠNG 4: CƠ SỞ LÝ THUYẾT MẪU

Thống kê toán là bộ môn toán học nghiên cứu qui luật của các hiện tượng ngẫu nhiên có tính chất số lớn trên cơ sở thu thập và xử lý các số liệu thống kê các kết quả quan sát về các hiện tượng ngẫu nhiên này.

Nếu ta thu thập được tất cả các số liệu liên quan đến đối tượng cần nghiên cứu thì ta có thể biết được đối tượng này (phương pháp toàn bộ). Tuy nhiên trong thực tế điều đó không thể thực hiện được vì quy mô của đối tượng nghiên cứu quá lớn hoặc trong quá trình nghiên cứu đối tượng nghiên cứu bị phá hủy. Lý thuyết mẫu cung cấp phương pháp nghiên cứu tổng thể thông qua mẫu.

Phương pháp mẫu là một trong những phương pháp quan trọng của lý thuyết thống kê.

Trong chương này chúng ta tìm hiểu cơ sở của lý thuyết mẫu. Các phương pháp chọn mẫu: mẫu ngẫu nhiên đơn, mẫu ngẫu nhiên hệ thống, mẫu chùm, mẫu phân tổ, mẫu nhiều cấp.

Đối với mẫu ngẫu nhiên ta xét các vấn đề:

- Biểu diễn các giá trị quan sát của mẫu ngẫu nhiên theo bảng và theo biểu đồ.
- Thống kê của mẫu ngẫu nhiên.
- Các đặc trưng của thống kê mẫu ngẫu nhiên.
- Quy luật phân bố xác suất của một số thống kê đặc trưng mẫu.

Chương này là cơ sở cho chương tiếp.

4.1 SỰ CẦN THIẾT PHẢI LẤY MẪU

Nhiều bài toán trong thực tế dẫn đến nghiên cứu một hay nhiều *dấu hiệu định tính* hoặc *định lượng* liên quan đến các phần tử của một tập hợp nào đó. Chẳng hạn nếu muốn điều tra thu nhập bình quân của các gia đình ở Hà Nội thì tập hợp cần nghiên cứu là các hộ gia đình ở Hà Nội với dấu hiệu nghiên cứu có tính chất định lượng là thu nhập của từng hộ gia đình. Một doanh nghiệp có thể nghiên cứu các khách hàng của mình với dấu hiệu định tính là sự hài lòng của khách hàng đối với sản phẩm hoặc dịch vụ của doanh nghiệp, còn dấu hiệu định lượng là số lượng sản phẩm của doanh nghiệp mà khách hàng có nhu cầu được đáp ứng.

Dấu hiệu định lượng được thể hiện qua các đơn vị đo của đại lượng như trọng lượng (gram, kg …), độ dài (cm, m …), thời gian (giây, giờ …), áp suất (atmosphere…), chiếc, tá … Trái lại, dấu hiệu định tính như giới tính (nam, nữ), sở thích (yêu, ghét, thích loại sản phẩm nào đó), loại phương tiện (di động, cố định, internet)… không có đơn vị. Dù rằng có thể mã hóa những dấu hiệu này ví dụ nam ứng với 1, nữ ứng với 0, song không có đơn vị đo cho dấu hiệu này.

Để xử lý dấu hiệu cần nghiên cứu đôi khi người ta sử dung phương pháp nghiên cứu toàn bộ, đó là điều tra toàn bộ các phần tử của tập hợp theo dấu hiệu cần nghiên cứu để rút ra các kết luận cần thiết. Tuy nhiên trong thực tế việc áp dụng phương pháp này gặp phải những khó khăn sau:

- Do qui mô của tập hợp cần nghiên cứu quá lớn nên việc nghiên cứu toàn bộ sẽ đòi hỏi nhiều chi phí về vật chất và thời gian, có thể không kiểm soát được dẫn đến bị chồng chéo hoặc bỏ sót.
- Trong nhiều trường hợp không thể nắm được toàn bộ các phần tử của tập hợp cần nghiên cứu, do đó không thể tiến hành toàn bộ được. Chẳng hạn doanh nghiệp không thể xác định toàn bộ các khách hàng của mình.
- Có thể trong quá trình điều tra sẽ phá hủy đối tượng nghiên cứu. Chẳng hạn để xác định tuổi thọ một thiết bị mà nhà máy sản xuất, nếu sử dụng phương pháp nghiên cứu toàn bộ thì sản xuất ra bao nhiêu phải thử hết bấy nhiêu!

Vì thế trong thực tế phương pháp nghiên cứu toàn bộ thường chỉ áp dụng đối với các tập hợp có qui mô nhỏ, còn chủ yếu người ta sử dụng phương pháp không toàn bộ mà đặc biệt là phương pháp nghiên cứu chọn mẫu.

4.2 MẪU NGẪU NHIÊN

4.2.1 Khái niệm mẫu ngẫu nhiên

Toàn bộ tập hợp các phần tử đồng nhất theo một dấu hiệu nghiên cứu định tính hay định lượng nào đó được gọi là ***tổng thể***, ký hiệu $\mathcal{C}$.

Số lượng các phần tử của tổng thể được gọi là ***kích thước của tổng thể***, ký hiệu N. Thường thì kích thước N của tổng thể là hữu hạn, song nếu tổng thể quá lớn hoặc không thể nắm được toàn bộ tổng thể ta có thể giả thiết rằng kích thước của tổng thể là vô hạn. Điều giả thiết này dựa trên cơ sở là khi tăng kích thước của tổng thể lên khá lớn thì thực tế không ảnh hưởng gì đến kết quả tính toán trên số liệu của từng bộ phận rút ra từ tổng thể đó.

Mỗi phần tử của tổng thể được gọi là ***cá thể***.

Các cá thể của tổng thể được nghiên cứu thông qua các dấu hiệu nghiên cứu. Dấu hiệu nghiên cứu này có thể được định tính hoặc định lượng. Nếu dấu hiệu nghiên cứu có tính định lượng, nghĩa là được thể hiện bằng cách cho tương ứng mỗi cá thể của tổng thể $\mathcal{C}$ nhận một giá trị thực nào đó thì dấu hiệu này được gọi là một ***biến lượng,*** ký hiệu X. Có thể xem biến lượng X là một biến ngẫu nhiên chung của tổng thể.

Với dấu hiệu định tính ta chỉ xét trường hợp các dấu hiệu có thể mã hóa thành biến ngẫu nhiên chỉ nhận hai giá trị 0 và 1, như vậy dấu hiệu định tính X có thể xem là biến ngẫu nhiên có phân bố Bernoulli. Chẳng hạn một doanh nghiệp muốn nghiên cứu các khách hàng của mình về dấu hiệu định tính là sự hài lòng của khách hàng đối với sản phẩm hoặc dịch vụ của doanh nghiệp, câu trả lời của các khách hàng chỉ chọn theo một trong hai giá trị sau: có hài lòng (giá trị 1) hoặc không hài lòng (giá trị 0) và không xét các ý kiến khác.

Việc chọn n cá thể nào đó từ tổng thể được gọi là ***phép lấy mẫu***. Ta gọi các cá thể chọn được này là *một mẫu*, n là kích thước mẫu. Nếu cá thể chọn xong không trả lại tổng thể để chọn tiếp thì mẫu được gọi là *không hoàn lại*. Nếu chọn xong trả lại tổng thể để chọn tiếp thì mẫu được gọi *có hoàn lại*.

Ta nói rằng một mẫu là ***mẫu ngẫu nhiên*** nếu trong phép lấy mẫu đó mỗi cá thể của tổng thể được chọn một cách độc lập và có xác suất được chọn như nhau.

Trường hợp kích thước của tổng thể vô hạn hoặc không xác định thì mẫu ngẫu nhiên có thể chọn theo cách hoàn lại hoặc không hoàn lại, nhưng khi kích thước của tổng thể không lớn thì mẫu ngẫu nhiên phải thực hiện theo cách hoàn lại.

4.2.2 Một vài phương pháp chọn mẫu ngẫu nhiên

Một trong những nhiệm vụ quan trọng nhất của phương pháp thống kê là xây dựng các phương pháp cho phép ta có thể rút ra các kết luận, ra quyết định, lập các dự báo về toàn bộ tổng thể trên cơ sở xử lý các thông tin thu được trên mẫu. Vì vậy vấn đề lấy mẫu rất quan trọng.

Tùy theo đặc điểm của tổng thể mà mẫu có thể được chọn theo nhiều phương pháp khác nhau để đảm bảo yêu cầu về tính đại diện của mẫu:

a. Mẫu ngẫu nhiên đơn: 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 để rút ra n phần tử của tổng thể theo một bảng số ngẫu nhiên nào đó.

Phương pháp này có ưu điểm là cho phép thu được mẫu có tính đại diện cao, cho phép suy rộng các kết quả của mẫu cho tổng thể với một sai số xác định, song để sử dụng phương pháp này cần phải có toàn bộ danh sách của tổng thể nghiên cứu. Mặt khác chi phí chọn mẫu sẽ khá lớn.

b. Mẫu ngẫu nhiên hệ thống: là loại mẫu ngẫu nhiên đã được đơn giản hóa trong cách chọn, trong đó chỉ có phần tử đầu tiên được lựa chọn một cách ngẫu nhiên, sau đó dựa trên danh sách đã được đánh số của tổng thể các phần tử còn lại của mẫu được chọn theo một thủ tục hay quy luật nào đó.

Nhược điểm chính của phương pháp này là dễ mắc sai số hệ thống khi danh sách của tổng thể không được sắp xếp một cách ngẫu nhiên mà theo một trật tự chủ quan nào đó. Tuy vậy do cách thức đơn giản của nó, mẫu ngẫu nhiên hệ thống hay được dùng trong trường hợp tổng thể tương đối thuần nhất.

c. Mẫu chùm: Trong một số trường hợp, để tiện cho việc nghiên cứu người ta muốn qui diện nghiên cứu gọn về từng chùm chứ không để cho các phần tử của mẫu phân tán quá rộng. Chẳng hạn muốn điều tra về chi tiêu hàng tháng thì ta tiến hành điều tra với từng hộ gia đình chứ không xét từng người riêng lẻ. Mỗi hộ gia đình là một chùm.

d. Mẫu phân tổ: Để chọn mẫu phân tổ, trước hết người ta phân chia tổng thể ra thành các tổ có độ thuần nhất cao để chọn ra các phần tử đại diện cho từng tổ. Việc phân tổ có hiệu quả khi tổng thể nghiên cứu không thuần nhất theo dấu hiệu nghiên cứu. Sau khi đã phân tổ thì kích thước mẫu được phân bố cho mỗi tổ theo một qui tắc nào đó, chẳng hạn tỷ lệ thuận với kích thước mỗi tổ.

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

4.2.3 Mô hình hóa mẫu ngẫu nhiên

Giả sử các cá thể của tổng thể được nghiên cứu thông qua dấu hiệu X. Với mỗi mẫu ta chỉ cần quan tâm dấu hiệu nghiên cứu X của mỗi cá thể của mẫu.

Chẳng hạn, khi cần nghiên cứu chiều cao trung bình của thanh niên trong một vùng nào đó thì với cá thể A được chọn làm mẫu ta chỉ quan tâm về chiều cao của A, tức là dấu hiệu chiều cao X_A *của A, mà không quan tâm đến các đặc trưng khác của cá thể này.*

Vì vậy, mỗi cá thể được chọn khi lấy mẫu có thể đồng nhất với dấu hiệu nghiên cứu X của cá thể đó. Bằng cách đồng nhất mỗi cá thể của mẫu ngẫu nhiên với các dấu hiệu nghiên cứu tương ứng của cá thể ta có thể xác định mẫu ngẫu nhiên như sau:

Mẫu ngẫu nhiên kích thước n *là một dãy gồm* n *biến ngẫu nhiên:* $X_1, X_2, ..., X_n$ *độc lập cùng phân bố với* X *, ký hiệu* $W = (X_1, X_2, ..., X_n)$*, trong đó* X_i là dấu hiệu X của cá thể thứ i của mẫu ($i = 1,...,n$).

Thực hiện một phép thử đối với mẫu ngẫu nhiên W chính là thực hiện một phép thử đối với mỗi thành phần của mẫu. Giả sử X_i nhận giá trị x_i ($i = 1,...,n$), khi đó các giá trị $x_1, x_2, ..., x_n$ tạo thành một *giá trị cụ thể* của mẫu ngẫu nhiên, hay còn gọi là một thể hiện của mẫu ngẫu nhiên, ký hiệu $w = (x_1, x_2, ..., x_n)$.

Ví dụ 4.1: Gọi X là số chấm xuất hiện khi tung con xúc xắc cân đối thì X là biến ngẫu nhiên có bảng phân bố xác suất sau

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

Tung con xúc xắc 3 lần và gọi X_i là số chấm xuất hiện trong lần tung thứ i ($i = 1, 2, 3$) thì ta có 3 biến ngẫu nhiên độc lập có cùng quy luật phân bố xác suất với X . Vậy ta có mẫu ngẫu nhiên kích thước 3, $W = (X_1, X_2, X_3)$.

Thực hiện một phép thử đối với mẫu ngẫu nhiên này tức là tung con xúc xắc 3 lần. Giả sử lần thứ nhất được mặt có 2 chấm, lần thứ hai được 5 chấm lần ba được 3 chấm thì $w = (2,5,3)$ là một mẫu cụ thể của mẫu ngẫu nhiên W .

4.2.4 Biểu diễn giá trị cụ thể của mẫu ngẫu nhiên theo bảng và theo biểu đồ

4.2.4.1 Bảng phân bố tần số thực nghiệm

Từ một mẫu cụ thể của mẫu ngẫu nhiên kích thước n của X ta sắp xếp các giá trị của mẫu cụ thể theo thứ tự tăng dần, giả sử giá trị x_i xuất hiện với tần số r_i, $i = 1,...,k$

$$x_1 < \cdots < x_k \,;\; r_1 + \cdots + r_k = n. \qquad \textbf{(4.1)}$$

Khi đó ta có thể biểu diễn giá trị quan sát của mẫu ngẫu nhiên trên qua bảng phân bố tần số thực nghiệm

X	x_1	x_2	…	x_k
Tần số	r_1	r_2	…	r_k

(4.2)

Ở đây x_1 là giá trị cụ thể nhỏ nhất, chưa chắc là giá trị cụ thể của X_1.

4.2.4.2 Bảng phân bố tần suất thực nghiệm

Ký hiệu $f_i = \dfrac{r_i}{n}$ gọi là tần suất của x_i.

Ta có bảng phân bố tần suất thực nghiệm của X

X	x_1	x_2	…	x_k
Tần suất	f_1	f_2	…	f_k

(4.3)

Ví dụ 4.2: Giá trị quan sát của một mẫu ngẫu nhiên kích thước 120 được biểu diễn theo bảng phân bố tần số thực nghiệm

X	31	34	35	36	38	40	42	44	$\sum$
Tần số	10	20	30	15	10	10	5	20	120

Bảng phân bố tần suất thực nghiệm tương ứng

X	31	34	35	36	38	40	42	44	$\sum$
Tần suất	$2/24$	$4/24$	$6/24$	$3/24$	$2/24$	$2/24$	$1/24$	$4/24$	1

4.2.4.3 Hàm phân bố thực nghiệm của mẫu

Tương tự công thức (2.11) xác định hàm phân bố của biến ngẫu nhiên rời rạc từ hàm khối lượng xác suất, với mẫu cụ thể của mẫu ngẫu nhiên xác định bởi công thúc (4.1) ta có ***hàm phân bố thực nghiệm của mẫu*** xác định như sau

$$F_n(x) = \sum_{x_j \le x} f_j \; ; \quad -\infty < x < +\infty \tag{4.4}$$

Định lý Glivenco chỉ ra rằng hàm phân bố thực nghiệm $F_n(x)$ xấp xỉ với phân bố lý thuyết $F_X(x) = P\{X \le x\}$ khi n đủ lớn, trong đó X là biến ngẫu nhiên gốc của tổng thể.

4.2.4.4 Bảng phân bố ghép lớp

Trong những trường hợp mẫu điều tra có kích thước lớn, hoặc khi các giá trị cụ thể của dấu hiệu X lấy giá trị khác nhau song lại khá gần nhau, người ta thường xác định một số các khoảng $C_1, C_2, ..., C_k$ sao cho mỗi giá trị của dấu hiệu điều tra thuộc vào một khoảng nào đó. Các khoảng này lập thành một phân hoạch của miền giá trị của X.

Việc chọn số khoảng và độ rộng khoảng là tuỳ thuộc vào kinh nghiệm của người nghiên cứu, nhưng nói chung không nên chia quá ít khoảng. Ngoài ra độ rộng các khoảng cũng không nhất thiết phải bằng nhau. Chẳng hạn khi muốn thống kê về tỉ lệ người nghiện thuốc lá thì ta tập trung nhiều vào độ tuổi thanh niên và trung niên, vì vậy các khoảng trong độ tuổi này sẽ nhỏ hơn.

Ví dụ 4.3: Một mẫu cụ thể về chiều cao (đơn vị pound) của 40 nam sinh viên đại học có số đo được làm tròn

138	164	150	132	144	125	149	157	146	158
140	147	136	148	152	144	168	126	138	176
163	119	154	165	146	173	142	147	135	153
140	135	161	145	135	142	150	156	145	128

trọng lượng lớn nhất 176 và nhỏ nhất 119, vậy khoảng cách tối đa $176 - 119 = 57$.

Nếu chia với độ rộng khoảng 5 thi có xấp xỉ $57/5 \simeq 11$ khoảng

Nếu chia với độ rộng khoảng 20 thi có xấp xỉ $57/20 \simeq 3$ khoảng.

Ta chọn độ rộng khoảng là 5, vì chọn độ rộng khoảng là 20 thì chỉ có 3 khoảng và sai số sẽ lớn.

Ta có hai bảng phân bố ghép lớp với độ rộng khoảng bằng 5 có 11 khoảng và độ rộng khoảng bằng 9 có 7 khoảng

Trọng lượng (pound)	Tần số
118-127	3
127-136	5
136-145	9
145-154	12
154-163	5
163-172	4
172-180	2
Tổng số	40

Độ rộng khoảng bằng 9

Trọng lượng (pound)	Tần số
118-123	1
123-128	2
128-133	2
133-138	4
138-143	6
143-148	8
148-153	5
153-158	4
158-163	2
163-168	3
168-173	1
173-178	2
Tổng số	40

Độ rộng khoảng bằng 5

Nhận xét 4.1:

1) Ví dụ trên cho thấy ta nên chia khoảng các giá trị, vì trong thực tế các số đo nhận được là các giá trị xấp xỉ quy tròn.

2) Người ta quy ước đầu mút bên phải của mỗi khoảng thuộc vào khoảng đó mà không thuộc khoảng tiếp theo khi tính tần số của mỗi khoảng.

3) Một trong những gợi ý để chọn số khoảng k tối ưu là hãy chọn k nguyên nhỏ nhất sao cho $2^k \geq n$ như sau:

n: kích thước mẫu	33 – 64	65 – 127	129 – 256	257 – 512	513 – 1024
k: số khoảng	6	7	8	9	10

4) Độ rộng các khoảng không đòi hỏi bằng nhau.

Ví dụ 4.4: Một mẫu về chiều cao của 400 cây con trong một vườn ươm được trình bày trong bảng phân bố ghép lớp sau:

Kho¶ng	TÇn sè r_i	TÇn suÊt f_i	§ é réng kho¶ng l_i	$y_i = r_i / l_i$
4,5 – 9,5	18	0,045	5	3,6
9,5 – 11,5	58	0,145	2	29
11,5 – 13,5	62	0,155	2	31
13,5 – 16,5	72	0,180	3	24
16,5 – 19,5	57	0,1425	3	19
19,5 – 22,5	42	0,105	3	14
22,5 – 26,5	36	0,090	4	9
26,5 – 36,5	55	0,1375	10	5,5

Chiều cao $y_i = \dfrac{r_i}{l_i}$ là tần số xuất hiện trong một đơn vị khoảng của khoảng có độ dài l_i.

4.2.4.5 Biểu diễn bằng biểu đồ

Để có hình ảnh trực quan ta có thể biểu diễn các giá trị của mẫu bằng biểu đồ.

Giả sử dấu hiệu điều tra X có bảng phân bố tần số thực nghiệm

X	x_1	x_2	…	x_k
Tần số	r_1	r_2	…	r_k

và bảng phân bố tần suất thực nghiệm

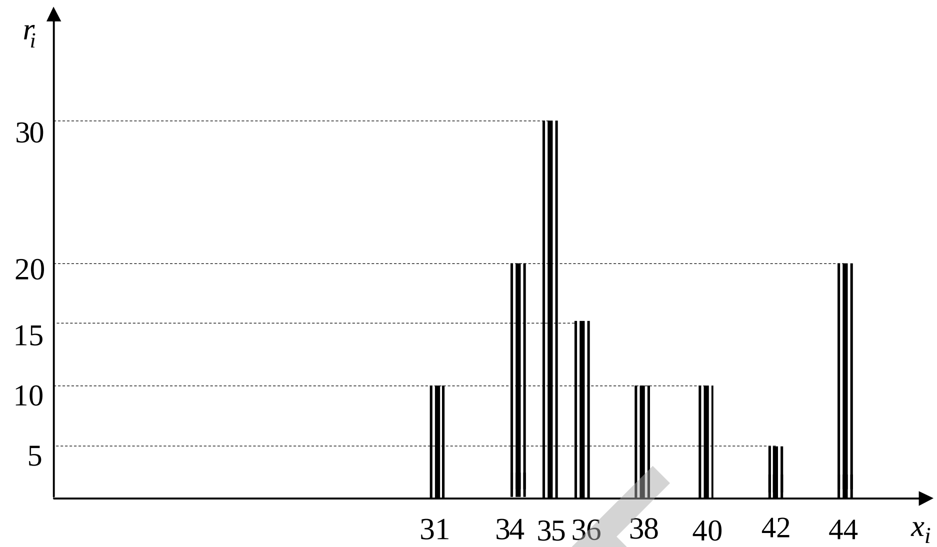
X	x_1	x_2	…	x_k
Tần suất	f_1	f_2	…	f_k

Trong mặt phẳng với hệ trục toạ độ Oxy.

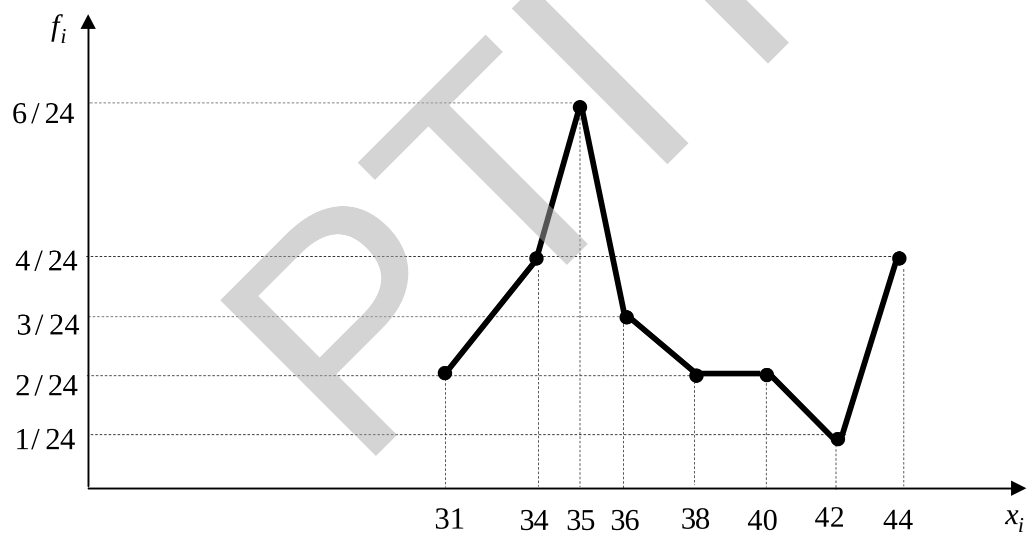
❖ Nối điểm trên trục hoành có toạ độ $(x_i, 0)$ với điểm có toạ độ (x_i, r_i) với mọi $i = 1, 2, \ldots, k$ ta được ***biểu đồ tần số hình gậy***.

❖ Nối lần lượt điểm có toạ độ (x_i, f_i) với điểm có toạ độ (x_{i+1}, f_{i+1}) với mọi $i = 1, 2, \ldots, k-1$ ta được ***biểu đồ đa giác tần suất***.

Bảng phân bố tần số và tần suất thực nghiệm trong ví dụ 4.2 có biểu đồ tần số hình gậy và biểu đồ đa giác tần suất



Hình 4.1: *Biểu đồ tần số hình gậy*



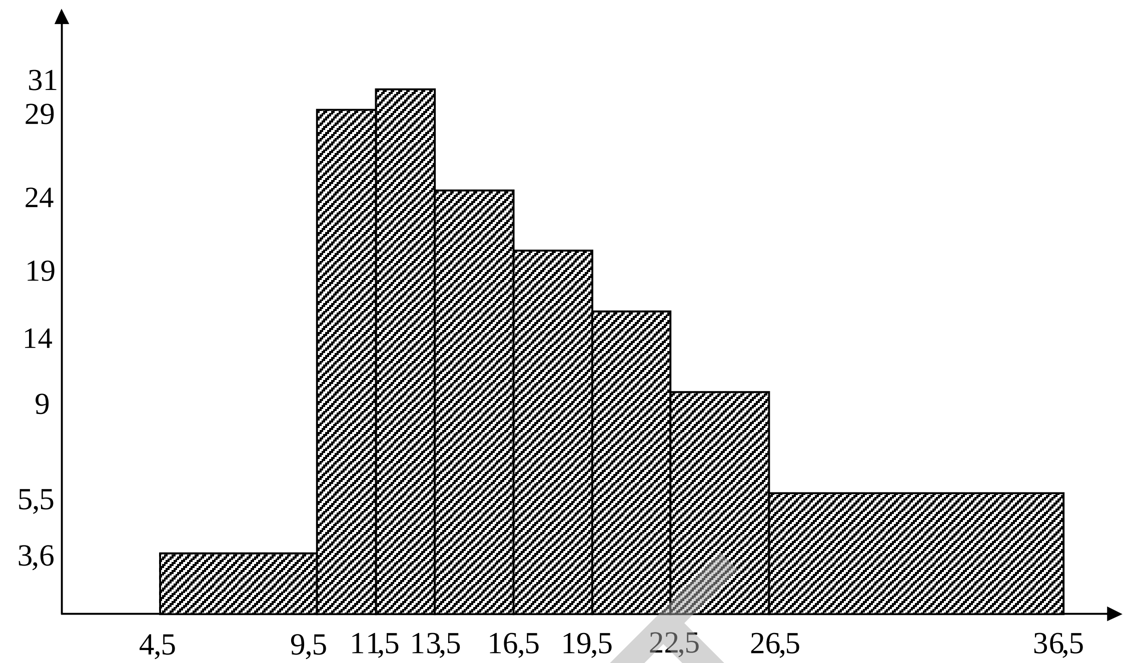
Hình 4.2*: Biểu đồ đa giác tần suất*

4.2.4.6 Tổ chức đồ (histogram)

Đối với bảng phân bố ghép lớp, người ta thường dùng tổ chức đồ để biểu diễn.

Trong mặt phẳng với hệ toạ độ Oxy, trên trục hoành ta chia các khoảng C_i có độ rộng l_i. Với mỗi khoảng C_i ta dựng hình chữ nhật có chiều cao $y_i = \dfrac{r_i}{l_i}$ (đối với tổ chức đồ tần số), hay $y_i = \dfrac{f_i}{l_i}$ (đối với tổ chức đồ tần suất).

Tổ chức đồ tần số của mẫu ghép lớp của ví dụ 4.3



Hình 4.3: *Tổ chức đồ*

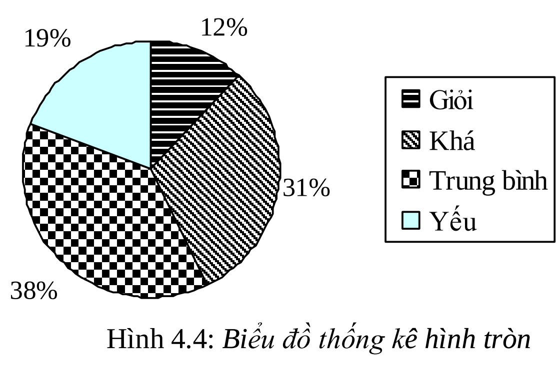
Ý nghĩa hình học của tổ chức đồ là ta có thể tính tần số giá trị cụ thể của mẫu thông qua diện tích giới hạn bởi tổ chức đồ. Chẳng hạn số cây nằm trong khoảng (12; 25] chính là diện tích của tổ chức đồ giới hạn bởi đường thẳng $x=12$ và $x=25$. Cụ thể

$$(13,5-12)\cdot 31+(16,5-13,5)\cdot 24+(19,5-16,5)\cdot 19+(22,5-19,5)\cdot 14+(25-22,5)\cdot 9=240$$

Vậy có 240 cây có chiều cao từ $12m$ đến $25m$.

Khi dấu hiệu điều tra của tổng thể có thể biểu diễn dưới dạng các tần số tỷ lệ người ta thường mô tả các số liệu mẫu bằng biểu đồ hình bánh xe. Đó là hình tròn được chia thành những góc có diện tích tỷ lệ với các tần số tương ứng của mẫu.

Ví dụ 4.5: Tổng kết kết quả học tập của sinh viên Học viện ta trong năm 2005 được số liệu sau:



Hình 4.4: *Biểu đồ thống kê hình tròn*

4.3 THỐNG KÊ VÀ CÁC ĐẶC TRƯNG CỦA MẪU NGẪU NHIÊN

4.3.1 Định nghĩa thống kê

Một thống kê của mẫu là một hàm phụ thuộc các biến ngẫu nhiên thành phần của mẫu.

Thống kê của mẫu ngẫu nhiên $W=(X_1,X_2,...,X_n)$ có dạng:

$$T=T(X_1,X_2,...,X_n) \qquad \textbf{(4.5)}$$

Như vậy thống kê T cũng là một biến ngẫu nhiên tuân theo một quy luật phân bố xác suất nhất định và có các tham số đặc trưng như kỳ vọng $\mathrm{E}T$ phương sai $\mathrm{D}T$ … Mặt khác, khi mẫu ngẫu nhiên nhận một giá trị cụ thể $w=(x_1,x_2,...,x_n)$ thì T cũng nhận một giá trị cụ thể còn gọi là giá trị quan sát của thống kê

$$T_{qs}=T(x_1,x_2,...,x_n)$$

Các thống kê mẫu cùng với quy luật phân bố xác suất của chúng là cơ sở để khảo sát dấu hiệu nghiên cứu của tổng thể từ các thông tin của mẫu.

Có hai nhóm thống kê mẫu quan trọng đặc trưng cho biến ngẫu nhiên của tổng thể:

❖ Các số đặc trưng cho ta hình ảnh về vị trí trung tâm của mẫu, tức là xu thế các số liệu trong mẫu tụ tập xung quanh những con số nào đó. Chẳng hạn trung bình mẫu, trung vị mẫu, mốt mẫu...

❖ Các số đặc trưng cho sự phân tán của các số liệu: biên độ, độ lệch trung bình, độ lệch tiêu chuẩn và phương sai mẫu.

Ta sẽ xem xét một số thống kê đặc trưng mẫu quan trọng sau:

4.3.2 Trung bình mẫu

Trung bình mẫu của mẫu ngẫu nhiên $W=(X_1,X_2,...,X_n)$ của biến ngẫu nhiên X được định nghĩa và ký hiệu

$$\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i \qquad \textbf{(4.6)}$$

Giá trị trung bình mẫu cụ thể của mẫu ngẫu nhiên cụ thể $w=(x_1,x_2,...,x_n)$ là

$$\overline{x}=\frac{1}{n}\sum_{i=1}^{n}x_i \qquad \textbf{(4.7)}$$

Giả sử dấu hiệu nghiên cứu X có kỳ vọng và phương sai hữu hạn, áp dụng các công thức tính kỳ vọng và phương sai của tổng các biến ngẫu nhiên độc lập (công thức (2.23), (2.24), (2.34) và ví dụ 2.18) ta có

$$\mathrm{E}\left(\overline{X}\right)=\mathrm{E}\,X\ ;\ \ \mathrm{D}\left(\overline{X}\right)=\frac{\mathrm{D}\,X}{n}. \qquad \textbf{(4.8)}$$

Độ lệch chuẩn của trung bình mẫu $\sigma_{\overline{X}}$ thường được dùng để phản ảnh sai số ước lượng, do đó người ta còn gọi là sai số chuẩn Se của trung bình mẫu:

$$\sigma_{\overline{X}}=\frac{\sqrt{\mathrm{D}\,X}}{\sqrt{n}}. \qquad \textbf{(4.9)}$$

4.3.3 Phương sai mẫu, Độ lệch chuẩn mẫu

Một cách tương tự trung bình mẫu, ta định nghĩa phương sai mẫu là trung bình cọng của độ lệch bình phương các thành phần của mẫu với trung bình mẫu và ký hiệu

▪ Phương sai mẫu $\widehat{S}^2$:

$$\widehat{S}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 = \frac{1}{n}\sum_{i=1}^{n}X_i^2 - \left(\overline{X}\right)^2 \tag{4.10}$$

Phương sai mẫu $\widehat{S}^2$ cũng là một biến ngẫu nhiên, sử dụng các tính chất kỳ vọng ta có:

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

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

$$= \frac{1}{n}\mathrm{E}\left[\left(\sum_{i=1}^{n}\left(X_i - \mu\right)^2\right) - n\left(\overline{X} - \mu\right)^2\right] = \frac{n-1}{n}\mathrm{D}\,X\,.$$

Vậy

$$\mathrm{E}\,\widehat{S}^2 = \frac{n-1}{n}\mathrm{D}\,X \tag{4.11}$$

Để kỳ vọng của phương sai mẫu trùng với phương sai $\mathrm{D}\,X$ của biến ngẫu nhiên gốc ta cần hiệu chỉnh như sau.

▪ Phương sai mẫu có hiệu chỉnh S^2:

$$S^2 = \frac{n}{n-1}\widehat{S}^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 = \frac{1}{n-1}\sum_{i=1}^{n}X_i^2 - \frac{n}{n-1}\left(\overline{X}\right)^2 \tag{4.12}$$

Từ (4.11), (4.12) ta được

$$\mathrm{E}\,S^2 = \mathrm{D}\,X \tag{4.13}$$

Trường hợp biến ngẫu nhiên gốc X của tổng thể có kỳ vọng xác định $\mathrm{E}\,X = \mu$ ta có thể thay phương sai mẫu có dạng như sau:

▪ Phương sai mẫu trường hợp đã biết kỳ vọng của biến ngẫu nhiên gốc S^{*2}:

$$S^{*2} = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \mu\right)^2 \tag{4.14}$$

Áp dụng công thức tính kỳ vọng (2.23), (2.24) và (2.28) ta có:

$$\mathrm{E}\,S^{*2} = \mathrm{D}\,X \tag{4.15}$$

Độ lệch chuẩn mẫu

$$S = \sqrt{S^2} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2} \tag{4.16}$$

4.3.4 Tần suất mẫu

Trường hợp cần nghiên cứu một dấu hiệu định tính A nào đó mà mỗi cá thể của tổng thể có thể có hoặc không, giả sử p là tần suất có dấu hiệu A của tổng thể.

Nếu cá thể có dấu hiệu A ta cho nhận giá trị 1, trường hợp ngược lại ta cho nhận giá trị 0. Lúc đó dấu hiệu nghiên cứu có thể xem là biến ngẫu nhiên X có phân bố Bernoulli tham số p có kỳ vọng $\mathrm{E}X = p$ và phương sai $\mathrm{D}X = p(1-p)$ (công thức (2.48)).

Lấy mẫu ngẫu nhiên kích thước n: $W = (X_1, X_2, ..., X_n)$, trong đó $X_1, X_2, ..., X_n$ là dãy các biến ngẫu nhiên độc lập có cùng phân bố Bernoulli tham số p. Tần số xuất hiện dấu hiệu A của mẫu là

$$r = X_1 + X_2 + \cdots + X_n . \quad \textbf{(4.17)}$$

Tần suất mẫu

$$f = \frac{r}{n} = \overline{X} \quad \textbf{(4.18)}$$

Như vậy tần suất mẫu là trung bình mẫu của biến ngẫu nhiên X có phân bố Bernoulli tham số p.

Tương tự ta có công thức tính kỳ vọng và phương sai của tần suất mẫu:

$$\mathrm{E}(f) = p\,;\quad \mathrm{D}(f) = \frac{p(1-p)}{n} \quad \textbf{(4.19)}$$

Sai số chuẩn của tần suất mẫu

$$\sigma_f = \frac{\sqrt{p(1-p)}}{\sqrt{n}} \quad \textbf{(4.20)}$$

4.3.5 Cách tính giá trị cụ thể của trung bình mẫu $\overline{x}$ và phương sai mẫu có hiệu chỉnh s^2

1) Nếu mẫu chỉ nhận các giá trị $x_1, x_2, ..., x_k$ với tần số tương ứng $r_1, r_2, ..., r_k$ thì giá trị trung bình mẫu và phương sai mẫu cụ thể được tính theo công thức

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{k} r_i x_i \text{ , trong đó } \sum_{i=1}^{k} r_i = n \quad \textbf{(4.21)}$$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{k} r_i \left(x_i - \overline{x}\right)^2 = \frac{1}{n-1}\left(\sum_{i=1}^{k} r_i x_i^2 - \frac{\left(\sum_{i=1}^{k} r_i x_i\right)^2}{n}\right) \quad \textbf{(4.22)}$$

2) Nếu giá trị của mẫu cụ thể được cho dưới dạng bảng phân bố ghép lớp với các khoảng $C_1, ..., C_m$ và tần số của C_i là r_i, khi đó giá trị trung bình mẫu và phương sai mẫu được tính theo công thức trên, trong đó x_i là trung điểm của khoảng C_i.

3) Mẫu thu gọn: Nếu các giá trị của mẫu cụ thể x_i không gọn (quá lớn hoặc quá bé hoặc phân tán) ta có thể thu gọn mẫu bằng cách đổi biến:

$$u_i = \frac{x_i - a}{h} \text{ hoặc } x_i = hu_i + a \quad \textbf{(4.23)}$$

Khi đó

$$\overline{x} = h\overline{u} + a\,;\quad s^2 = h^2 s_u^2 \quad \textbf{(4.24)}$$

Trong đó

$$\overline{u}=\frac{1}{n}\sum_{i=1}^{k}r_iu_i\,;\quad s_u^2=\frac{1}{n-1}\sum_{i=1}^{k}r_i\left(u_i-\overline{u}\right)^2=\frac{1}{n-1}\left(\sum_{i=1}^{k}r_iu_i^2-\frac{\left(\sum_{i=1}^{k}r_iu_i\right)^2}{n}\right)\qquad \textbf{(4.25)}$$

Thật vậy: $\overline{x}=\frac{1}{n}\sum_{i=1}^{k}r_ix_i=\frac{1}{n}\sum_{i=1}^{k}r_i\left(hu_i+a\right)=h\left(\frac{1}{n}\sum_{i=1}^{k}r_iu_i\right)+\left(\frac{1}{n}\sum_{i=1}^{k}r_i\right)a=h\overline{u}+a$.

$$s^2=\frac{1}{n-1}\sum_{i=1}^{k}r_i\left(x_i-\overline{x}\right)^2=\frac{1}{n-1}\sum_{i=1}^{k}r_i\left(hu_i+a-h\overline{u}-a\right)^2=\frac{h^2}{n-1}\sum_{i=1}^{k}r_i\left(u_i-\overline{u}\right)^2=h^2s_u^2.$$

Các số a và h được chọn phù hợp sao cho $\overline{u}$, s_u^2 tính dễ dàng hơn.

Thông thường ta chọn a là điểm giữa của các giá trị x_i và h được chọn sao cho phép chia thực hiện dễ dàng.

Ví dụ 4.6: Giá trị trung bình mẫu, phương sai mẫu và độ lệch chuẩn mẫu của mẫu cụ thể cho trong ví dụ 4.5.

Khoảng	tần số r_i	x_i	$u_i=\frac{x_i-20}{5}$	r_iu_i	$r_iu_i^2$
4,5 – 9,5	18	7	– 2,6	– 46,8	121,68
9,5 – 11,5	58	10,5	– 1,9	– 110,2	209,38
11,5 – 13,5	62	12,5	– 1,5	– 93	139,5
13,5 – 16,5	72	15	– 1	– 72	72
16,9 – 19,5	57	18	– 0,4	– 22,8	9,12
19,5 – 22,5	42	21	0,2	8,4	1,68
22,5 – 26,5	36	24,5	0,9	32,4	29,16
26,5 – 36,5	55	31,5	2,3	126,5	290,95
$\sum$	400			– 177,5	873,47

$$\overline{x}=5\overline{u}+20=5\cdot\frac{-177,5}{400}+20=17,78.\ \ s_u^2=\frac{1}{399}\cdot\left(873,47-\frac{(-177,5)^2}{400}\right)=1,9917$$

$$\Rightarrow s^2=5^2\cdot s_u^2=49,79\ \Rightarrow s=\sqrt{49,79}=7,056.$$

4.4 MẪU NGẪU NHIÊN HAI CHIỀU

4.4.1 Khái niệm mẫu ngẫu nhiên hai chiều

Ta xét đồng thời hai dấu hiệu nghiên cứu, trong đó dấu hiệu nghiên cứu thứ nhất có thể xem là biến ngẫu nhiên X, còn dấu hiệu nghiên cứu thứ hai là biến ngẫu nhiên Y. Lúc đó việc nghiên cứu đồng thời hai dấu hiệu của tổng thể tương đương với việc nghiên cứu biến ngẫu nhiên hai chiều (X,Y).

Mẫu ngẫu nhiên hai chiều kích thước n của dấu hiệu nghiên cứu (X,Y) là một dãy gồm n biến ngẫu nhiên hai chiều (X_1,Y_1), (X_2,Y_2), …, (X_n,Y_n) độc lập và có cùng quy luật phân bố xác suất với (X,Y).

Mẫu ngẫu nhiên hai chiều được ký hiệu là $W = \left[(X_1,Y_1), (X_2,Y_2), ..., (X_n,Y_n)\right]$

Giả sử thành phần (X_i,Y_i) nhận giá trị (x_i, y_i) ($i=1,...,n$) ta thu được mẫu cụ thể:

$$w = \left[(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)\right].$$

Các giá trị x_i ($i=1,...,n$) gọi là giá trị cụ thể của thành phần X của mẫu, còn các giá trị y_i ($i=1,...,n$) gọi là giá trị cụ thể của thành phần Y của mẫu.

4.4.2 Biểu diễn giá trị cụ thể của mẫu ngẫu nhiên hai chiều

Thực hiện phép thử từ tổng thể ta nhận được giá trị cụ thể (giá trị quan sát) của mẫu ngẫu nhiên kích thước n, ta có thể sắp xếp các giá trị cụ thể thành phần X và Y của mẫu theo thứ tự tăng dần:

$$x_1 < x_2 < ... < x_i < ... < x_h \quad \text{và} \quad y_1 < y_2 < ... < y_j < ... < y_k.$$

Ký hiệu r_{ij} là tần số của cặp giá trị (x_i, y_j) của mẫu w, rõ ràng các tần số r_{ij} thỏa mãn hệ thức $\sum_{j=1}^{k}\sum_{i=1}^{h} r_{ij} = n$. Lúc đó giá trị cụ thể mẫu có thể biểu diễn dưới dạng bảng phân bố tần số thực nghiệm sau:

$X \backslash Y$	y_1	y_2	…	y_j	…	y_k	r_{i*}
x_1	r_{11}	r_{12}	…	r_{1j}	…	r_{1k}	r_{1*}
x_2	r_{21}	r_{22}	…	r_{2j}	…	r_{2k}	r_{2*}
⋮	⋮	⋮	…	⋮	…	⋮	⋮
x_i	r_{i1}	r_{i2}	…	r_{ij}	…	r_{ik}	r_{i*}
⋮	⋮	⋮	…	⋮	…	⋮	⋮
x_h	r_{h1}	r_{h2}	…	r_{hj}	…	r_{hk}	r_{h*}
r_{*j}	r_{*1}	r_{*2}	…	r_{*j}	…	r_{*k}	$\sum = n$

trong đó r_{i*} là tần số của giá trị x_i ($i=1,...,h$) của thành phần X,

r_{*j} là tần số của giá trị y_j ($j=1,...,k$) của thành phần Y.

4.4.3 Giá trị quan sát của một số thống kê đặc trưng mẫu ngẫu nhiên hai chiều

Từ bảng phân bố thực nghiệm của mẫu ngẫu nhiên hai chiều ta có thể tính được các giá trị quan sát của các đặc trưng mẫu sau

1) Bảng phân bố thực nghiệm của thành phần X

X	x_1	x_2	...	x_i	...	x_h
Tần số	r_{1*}	r_{2*}	...	r_{i*}	...	r_{h*}

Giá trị quan sát của trung bình mẫu và phương sai mẫu của X :

$$\overline{x}=\frac{1}{n}\sum_{i=1}^{h}r_{i*}x_i\ ;\quad s_x^2=\frac{1}{n-1}\sum_{i=1}^{h}r_{i*}\left(x_i-\overline{x}\right)^2 \tag{4.26}$$

2) Bảng phân bố thực nghiệm của thành phần Y

Y	y_1	y_2	...	y_j	...	y_k
Tần số	r_{*1}	r_{*2}	...	r_{*j}	...	r_{*k}

Giá trị quan sát của trung bình mẫu và phương sai mẫu của Y:

$$\overline{y}=\frac{1}{n}\sum_{j=1}^{k}r_{*j}y_j\ ;\quad s_y^2=\frac{1}{n-1}\sum_{j=1}^{k}r_{*j}\left(y_j-\overline{y}\right)^2 \tag{4.27}$$

3) Giá trị quan sát của hệ số tương quan mẫu

$$r=\frac{\sum_{i=1}^{h}\sum_{j=1}^{k}r_{ij}(x_i-\overline{x})(y_j-\overline{y})}{\sqrt{\left(\sum_{i=1}^{h}r_{i*}\left(x_i-\overline{x}\right)^2\right)\left(\sum_{j=1}^{k}r_{*j}\left(y_j-\overline{y}\right)^2\right)}} \tag{4.28}$$

4.5 PHÂN BỐ XÁC SUẤT CỦA MỘT SỐ THỐNG KÊ ĐẶC TRƯNG MẪU

4.5.1 Trường hợp biến ngẫu nhiên gốc có phân bố chuẩn

Giả sử dấu hiệu nghiên cứu trong tổng thể có thể xem như một biến ngẫu nhiên X có phân bố chuẩn với kỳ vọng $\mathrm{E}X=\mu$ và phương sai $\mathrm{D}X=\sigma^2$. Các tham số này có thể đã biết hoặc chưa biết. Từ tổng thể rút ra một mẫu ngẫu nhiên kích thước n:

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

Các biến ngẫu nhiên thành phần $X_1,X_2,...,X_n$ độc lập có cùng quy luật phân bố chuẩn $\mathbf{N}(\mu,\sigma^2)$ của X. Mọi tổ hợp tuyến tính của các biến ngẫu nhiên có phân bố chuẩn là biến ngẫu nhiên có phân bố chuẩn (công thức 2.66). Vì vậy ta có các kết quả sau:

4.5.1.1 Phân bố của thống kê trung bình mẫu

Trung bình mẫu $\overline{X}$ có phân bố chuẩn với kỳ vọng $\mathrm{E}(\overline{X})=\mu$ và phương sai $\mathrm{D}\left(\overline{X}\right)=\dfrac{\sigma^2}{n}$.

Áp dụng công thức (2.72) suy ra thống kê sau có phân bố chuẩn tắc $\mathbf{N}(0;1)$:

$$U=\frac{(\overline{X}-\mu)\sqrt{n}}{\sigma}\sim \mathbf{N}(0;1) \tag{4.29}$$

Trường hợp biến ngẫu nhiên gốc có phân bố tùy ý, áp dụng định lý giới hạn trung tâm (định lý 3.5) ta cũng có thể xấp xỉ phân bố của thống kê $U=\frac{(\overline{X}-\mu)\sqrt{n}}{\sigma}$ với phân bố chuẩn tắc $\mathbf{N}(0;1)$ khi *n* đủ lớn, trong thực tế với $n\geq 30$ ta có thể xấp xỉ $U=\frac{(\overline{X}-\mu)\sqrt{n}}{\sigma}\approx \mathbf{N}(0;1)$.

Ví dụ 4.7: Chiều cao *X* của các nam sinh viên đại học là biến ngẫu nhiên có phân bố chuẩn với trung bình 163cm và độ lệch chuẩn 3cm. Lấy 80 mẫu của mẫu ngẫu nhiên 25 sinh viên

a. Tìm kỳ vọng và phương sai của trung bình mẫu.

b. Có bao nhiêu mẫu trong số 80 mẫu lấy giá trị trung bình trong khoảng từ 161,8 cm đến 163,3 cm.

c. Có bao nhiêu mẫu trong số 80 mẫu lấy giá trị trung bình nhỏ hơn 161,4 cm.

Giải: **a.** $\mathrm{E}(\overline{X})=\mu=\mathrm{E}(X)=163\,\mathrm{cm}$, $\mathrm{D}\left(\overline{X}\right)=\frac{\sigma^2}{n}=\frac{3^2}{25}=0,36$.

b. Áp dụng công thức 4.22 ta được $U=\frac{(\overline{X}-163)}{0,6}\sim \mathbf{N}(0;1)$

$P\left\{161,86<\overline{X}<163,3\right\}=\Phi(0,5)-\Phi(-1,9)=\Phi(0,5)+\Phi(1,9)-1=0,6627$.

Vậy số mẫu thỏa mãn điều kiện cần tìm là 80.0,6627 hoặc xấp xỉ 53 mẫu.

c. $P\left\{\overline{X}<161,4\right\}=\Phi(-2,67)=1-\Phi(2,67)=0,0038$. Đây là biến cố có xác suất bé, vì vậy không có mẫu nào trong số 80 mẫu có số đo trung bình nhỏ hơn 161,4 cm.

Thật vậy $80.0,0038=0,304<<1$.

4.5.1.2 Phân bố của thống kê phương sai mẫu S^{*2}

Từ công thức (4.14) ta có: $nS^{*2}=\sum_{i=1}^{n}\left(X_i-\mu\right)^2$ và $\frac{nS^{*2}}{\sigma^2}=\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2$.

Vì các biến ngẫu nhiên X_i độc lập nên các biến ngẫu nhiên $\frac{X_i-\mu}{\sigma}$ cũng độc lập.

Mặt khác theo (2.72) ta có $\frac{X_i-\mu}{\sigma}\sim \mathbf{N}(0;1)$. Do đó thống kê $\chi^2=\frac{nS^{*2}}{\sigma^2}$ có phân bố "khi bình phương" n bậc tự do (công thức 2.86).

$$\chi^2=\frac{nS^{*2}}{\sigma^2}=\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2\sim \chi^2(n) \tag{4.30}$$

4.5.1.3 Phân bố của thống kê phương sai mẫu S^2

Ta có $$\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2=\sum_{i=1}^{n}\left(\frac{X_i-\overline{X}}{\sigma}\right)^2+\frac{\left(\overline{X}-\mu\right)^2}{\sigma^2/n}$$

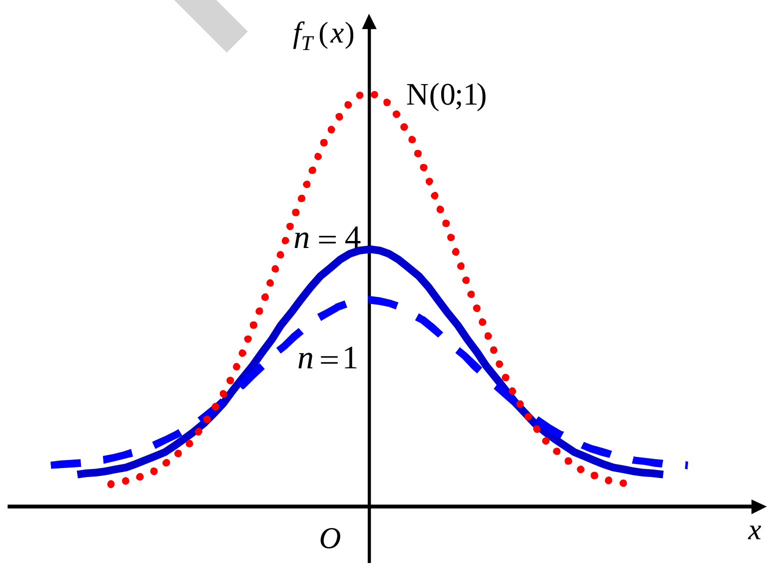
Thống kê $\dfrac{\left(\overline{X}-\mu\right)^2}{\sigma^2/n}$ có phân bố "khi bình phương" 1 bậc tự do và theo công thức (4.30) ta có $\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2\sim\chi^2(n)$. Sử dụng tính chất của phân bố "khi bình phương" (Nhận xét 2.5-3 chương 2) ta suy ra thống kê $\dfrac{(n-1)S^2}{\sigma^2}$ có phân bố "khi bình phương" $n-1$ bậc tự do

$$\frac{(n-1)S^2}{\sigma^2}=\sum_{i=1}^{n}\left(\frac{X_i-\overline{X}}{\sigma}\right)^2\sim\chi^2(n-1) \qquad \textbf{(4.31)}$$

Áp dụng công thức (2.91) với biến ngẫu nhiên U từ công thức (4.29) và $\chi^2=\dfrac{(n-1)S^2}{\sigma^2}$ từ công thức (4.31), thì $T=\dfrac{U}{\sqrt{\chi^2/(n-1)}}$ có phân bố Student $n-1$ bậc tự do.

Vậy

$$T=\frac{(\overline{X}-\mu)\sqrt{n}}{S}=\frac{\dfrac{(\overline{X}-\mu)\sqrt{n}}{\sigma}}{\sqrt{\dfrac{(n-1)S^2}{\sigma^2(n-1)}}}=\frac{U}{\sqrt{\chi^2/(n-1)}}\sim\mathbf{T}(n-1) \qquad \textbf{(4.32)}$$



Hình 4.4: Đồ thị hàm mật độ của phân bố Student

Phân bố Student $\mathbf{T}(n)$ hội tụ khá nhanh về phân bố chuẩn tắc $\mathbf{N}(0;1)$, do đó trong thực tế khi $n \ge 30$ ta có thể xem thống kê T xấp xỉ $\mathbf{N}(0;1)$.

$$T = \frac{(\overline{X} - \mu)\sqrt{n}}{S} \approx \mathbf{N}(0;1) \text{ , khi } n \ge 30 \qquad \textbf{(4.33)}$$

4.5.2 Trường hợp biến ngẫu nhiên gốc hai chiều cùng có phân bố chuẩn

Giả sử ta xét cùng một lúc hai tổng thể. Ở tổng thể thứ nhất dấu hiệu nghiên cứu được xem như biến ngẫu nhiên X_1 có phân bố chuẩn $\mathbf{N}(\mu_1;\sigma_1^2)$. Ở tổng thể thứ hai dấu hiệu nghiên cứu được xem như biến ngẫu nhiên X_2 có phân bố chuẩn $\mathbf{N}(\mu_2;\sigma_2^2)$.

Từ hai tổng thể nói trên rút ra hai mẫu ngẫu nhiên độc lập kích thước tương ứng là n_1 và n_2

$$W_1 = \left(X_{11}, X_{12}, ..., X_{1n_1}\right) \ ; \quad W_2 = \left(X_{21}, X_{22}, ..., X_{2n_2}\right) .$$

4.5.2.1 Phân bố thống kê hiệu của hai trung bình mẫu $\overline{X}_1 - \overline{X}_2$.

Theo tính chất của phân bố chuẩn (mọi tổ hợp tuyến tính của các phân bố chuẩn cũng có phân bố chuẩn) do đó $\overline{X}_1$, $\overline{X}_2$ và $\overline{X}_1 - \overline{X}_2$ cũng quy luật phân bố chuẩn.

Mặt khác kỳ vọng và phương sai của $\overline{X}_1 - \overline{X}_2$

$$\mathrm{E}\left(\overline{X}_1 - \overline{X}_2\right) = \mathrm{E}\left(\overline{X}_1\right) - \mathrm{E}\left(\overline{X}_2\right) = \mu_1 - \mu_2 .$$

$$\mathrm{D}\left(\overline{X}_1 - \overline{X}_2\right) = \mathrm{D}\left(\overline{X}_1\right) + \mathrm{D}\left(\overline{X}_2\right) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} .$$

Do đó thống kê sau có phân bố chuẩn tắc

$$U = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \sim \mathbf{N}(0;1) . \qquad \textbf{(4.34)}$$

4.5.2.2 Phân bố hai thống kê phương sai thành phần:

$$\chi_1^2 = \frac{(n_1 - 1)S_1^2}{\sigma_1^2} \sim \chi^2(n_1 - 1); \quad \chi_2^2 = \frac{(n_2 - 1)S_2^2}{\sigma_2^2} \sim \chi^2(n_2 - 1) \text{ và độc lập.} \qquad \textbf{(4.35)}$$

Do đó thống kê:

$$\chi^2 = \frac{(n_1 - 1)S_1^2}{\sigma_1^2} + \frac{(n_2 - 1)S_2^2}{\sigma_2^2} \sim \chi^2(n_1 + n_2 - 2) . \qquad \textbf{(4.36)}$$

Từ (4.34)-(4.36) suy ra:

- Trường hợp $\sigma_1^2 = \sigma_2^2$ ta có

$$T=\frac{(\overline{X}_1-\overline{X}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim \mathbf{T}(n_1+n_2-2)\,. \qquad \textbf{(4.37)}$$

có phân bố Student n_1+n_2-2 bậc tự do.

♦ Trường hợp $\sigma_1^2\neq\sigma_2^2$ ta có

$$T=\frac{(\overline{X}_1-\overline{X}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}}\sim \mathbf{T}(k)\,. \qquad \textbf{(4.38)}$$

có phân bố Student k bậc tự do, trong đó:

$$k\approx\frac{(n_1-1)(n_2-1)}{(n_2-1)C^2+(n_1-1)(1-C)^2}\ ;\ C=\frac{\frac{S_1^2}{n_1}}{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}\,.$$

4.5.3 Trường hợp biến ngẫu nhiên gốc có phân bố Bernoulli

Giả sử trong tổng thể dấu hiệu nghiên cứu có thể xem như biến ngẫu nhiên phân bố theo quy luật Bernoulli. Từ tổng thể rút ra một mẫu ngẫu nhiên kích thước n:

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

Từ công thức (4.17), (4.18), (4.19) ta đã biết tần số mẫu $r=X_1+\cdots+X_n\sim \mathcal{B}(n;p)$ có phân bố nhị thức với các tham số *n, p*. Theo định lý Moivre-Laplace (công thức 3.45):

$$\text{Với mọi } x\in\mathbb{R},\ \lim_{n\to\infty}P\left\{\frac{X_1+\cdots+X_n-np}{\sqrt{npq}}\le x\right\}=\Phi(x)\,. \qquad \textbf{(4.39)}$$

Như vậy định lý Moivre-Laplace cho phép xấp xỉ phân bố của thống kê

$$U=\frac{X_1+\cdots+X_n-np}{\sqrt{npq}}=\frac{\left(\frac{X_1+\cdots+X_n}{n}-p\right)n}{\sqrt{npq}}=\frac{\left(\frac{X_1+\cdots+X_n}{n}-p\right)\sqrt{n}}{\sqrt{pq}}=\frac{(f-p)\sqrt{n}}{\sqrt{pq}}$$

với phân bố chuẩn tắc $\mathbf{N}(0;1)$ khi n đủ lớn. Người ta thấy rằng xấp xỉ là tốt khi $np>5$ và $nq>5$ hoặc $npq>20$.

$$U=\frac{(f-p)\sqrt{n}}{\sqrt{pq}}\sim\mathbf{N}(0;1)\ \text{khi}\ \begin{cases}np>5\\ nq>5\end{cases}\ \text{hoặc}\ npq>20\,. \qquad \textbf{(4.40)}$$

Hoặc tính theo tần số mẫu $r=X_1+\cdots+X_n\sim\mathcal{B}(n;p)$, theo công thức (3.73) ta cũng có

$$U=\frac{(r-np)}{\sqrt{npq}}\sim\mathbf{N}(0;1)\ \text{khi}\ \begin{cases}np>5\\ nq>5\end{cases}\ \text{hoặc}\ npq>20\,. \qquad \textbf{(4.41)}$$

Ví dụ 4.8: Gieo 120 lần đồng xu cân đối đồng chất.

a. Tính xác suất có khoảng 40% đến 60% lần số mặt sấp xuất hiện.

b. Tính xác suất tỷ lệ mặt sấp xuất hiện lớn hơn hoặc bằng $\frac{5}{8}$.

c. Một nhóm 500 người, mỗi người gieo 120 lần đồng xu cân đối đồng chất. Có bao nhiêu người có kết quả mặt sấp xuất hiện trong khoảng 40% đến 60%.

Giải: Có thể xem mỗi lần gieo đồng xu là thực hiện phép thử Bernoulli với sự thành công của phép thử là sự xuất hiện mặt sấp, từ giả thiết ta có xác suất thành công của phép thử là $0,5$ Như vậy biến ngẫu nhiên gốc X có phân bố Bernoulli tham số $0,5$.

Gieo 120 lần là lấy mẫu ngẫu nhiên với kích thước 120 của biến ngẫu nhiên gốc, do đó tần suất mẫu $f = \dfrac{X_1 + \cdots + X_{120}}{120}$.

Ta có $np = nq = 120.0,5 = 60$, $\sqrt{npq} = 5,48$ thỏa mãn điều kiện kích thước đủ lớn.

a. 40% và 60% của 120 bằng 48 và 72. Áp dụng công thức (4.41) và (2.83) ta có

$$P\{48 < r < 42\} \approx \Phi\left(\frac{72+0,5-60}{5,48}\right) - \Phi\left(\frac{48-0,5-60}{5,48}\right)$$

$$= \Phi(2,28) - \Phi(-2,28) = 2\Phi(2,28) - 1 = 0,9774.$$

b. $\frac{5}{8}.120 = 75$, vậy xác suất tỷ lệ mặt sấp xuất hiện lớn hơn hoặc bằng $\frac{5}{8}$ là

$$P\{r > 75 - 0,5\} \approx 1 - \Phi\left(\frac{74,5-60}{5,48}\right) = 1 - \Phi\left(2,65\right) = 1 - 0,9960 = 0,0040.$$

c. Theo ý a) xác suất gieo 120 lần đồng xu (mẫu ngẫu nhiên kích thước 120) với 40% đến 60% lần mặt sấp xuất hiện là 0,9774. Vậy 500 người thực hiện 120 lần gieo đồng xu (500 quan sát cụ thể của mẫu ngẫu nhiên kích thước 120) thì số người có kết quả gieo với số mặt sấp xuất hiện trong khoảng 40% đến 60% là $500.0,9774 = 488,7 \approx 489$.

4.5.4 Trường hợp biến ngẫu nhiên gốc hai chiều cùng có phân bố Bernoulli

Giả sử ta xét cùng một lúc hai tổng thể trong đó dấu hiệu nghiên cứu trong hai tổng được xem như biến ngẫu nhiên có phân bố Bernoulli với các tham số lần lượt là p_1 và p_2.

Từ hai tổng thể nói trên rút ra hai mẫu ngẫu nhiên độc lập kích thước tương ứng là n_1 và n_2

$$W_1 = \left(X_{11}, X_{12}, ..., X_{1n_1}\right) \; ; \; W_2 = \left(X_{21}, X_{22}, ..., X_{2n_2}\right).$$

Xét thống kê $f_1 - f_2$ là hiệu hai tần suất mẫu. Lúc đó khi $n_1 > 30$ và $n_2 > 30$ ta có thể xấp xỉ phân bố của $f_1 - f_2$ theo phân bố chuẩn (áp dụng định lý giới hạn trung tâm) với các tham số đặc trưng là:

$$\mathrm{E}\left(f_1 - f_2\right) = p_1 - p_2 \text{ và } \mathrm{D}\left(f_1 - f_2\right) = \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}.$$

Do đó thống kê sau xấp xỉ phân bố chuẩn tắc:

$$U=\frac{(f_1-f_2)-(p_1-p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1}+\frac{p_2(1-p_2)}{n_2}}}\sim \mathbf{N}(0;1)\,. \qquad \mathbf{(4.42)}$$

CÂU HỎI ÔN TẬP VÀ BÀI TẬP CHƯƠNG 4

4.1 Mẫu ngẫu nhiên kích thước n về dấu hiệu nghiên cứu X có thể biểu diễn dưới dạng một dãy gồm n biến ngẫu nhiên $X_1, X_2, ..., X_n$ độc lập cùng phân bố với X.

Đúng ☐ Sai ☐ .

4.2 Tổ chức đồ dùng để biểu diễn mẫu ngẫu nhiên cho dưới dạng bảng phân bố ghép lớp.

Đúng ☐ Sai ☐ .

4.3 Một thống kê của mẫu ngẫu nhiên là con số cụ thể về dấu hiệu nghiên cứu.

Đúng ☐ Sai ☐ .

4.4 Nếu biến ngẫu nhiên gốc X có phân bố chuẩn thì trung bình mẫu của X cũng có phân bố chuẩn.

Đúng ☐ Sai ☐ .

4.5 Một thống kê của mẫu là một hàm của các biến ngẫu nhiên thành phần của mẫu do đó cũng là một biến ngẫu nhiên .

Đúng ☐ Sai ☐ .

4.6 Nếu biến ngẫu nhiên gốc X có phân bố chuẩn thì phương sai mẫu của X cũng có phân bố chuẩn.

Đúng ☐ Sai ☐ .

4.7 Nếu hai biến ngẫu nhiên gốc X_1, X_2 có phân bố chuẩn thì thống kê hiệu của hai trung bình mẫu $\overline{X}_1-\overline{X}_2$ cũng có phân bố chuẩn.

Đúng ☐ Sai ☐ .

4.8 Khi bậc tự do của phân bố Student lớn hơn 30, phân bố Student tiệm cận phân bố chuẩn.

Đúng ☐ Sai ☐ .

4.9 Khi kích thước mẫu $n_1>30$ và $n_2>30$ thì thống kê hiệu hai tần suất mẫu f_1-f_2 của hai dấu hiệu nghiên cứu có phân bố Bernoulli sẽ xấp xỉ phân bố chuẩn.

Đúng ☐ Sai ☐ .

4.10 Nếu biến ngẫu nhiên gốc có phân bố chuẩn thì phương sai mẫu có phân bố χ^2.

Đúng ☐ Sai ☐ .

4.11 Cho ví dụ về một biến ngẫu nhiên, một mẫu ngẫu nhiên kích thước $n=10$ và một giá trị của mẫu ngẫu nhiên xây dựng từ biến ngẫu nhiên ấy.

4.12 Cho ví dụ về mẫu ngẫu nhiên kích thước n được xây dựng từ biến ngẫu nhiên có phân bố Bernoulli, phân bố nghị thức, phân bố chuẩn.

4.13 Hãy tính giá trị trung bình mẫu $\overline{x}$ và phương sai mẫu s^2 của mẫu cụ thể có bảng phân bố tần số thực nghiệm sau

x_i	21	24	25	26	28	32	34
r_i	10	20	30	15	10	10	5

4.14 Hãy tính giá trị của trung bình mẫu $\overline{x}$ và độ lệch chuẩn mẫu S của mẫu cụ thể có bảng phân bố tần số thực nghiệm sau

x_i	4	7	8	12
r_i	5	2	3	10

4.15 Để nghiên cứu tuổi thọ X của một loại bóng đèn, người ta thắp thử 100 bóng và có số liệu sau:

x_i	1010-1030	1030-1050	1050-1070	1070-1090	1090-1110	1110-1130	1130-1150	1150-1170	1170-1190	1190-1210
r_i	2	3	8	13	25	20	12	10	6	1

Tính tuổi thọ trung bình $\overline{x}$ và độ lệch chuẩn mẫu s.

4.16 Giả sử biến ngẫu nhiên gốc có phân bố Bernoulli. Chọn mẫu ngẫu nhiên kích thước $n = 10$. Hãy tính kỳ vọng và phương sai của trung bình mẫu.

4.17 Giả sử biến ngẫu nhiên gốc có phân bố Bernoulli tham số $p = 0,5$. Hãy lập mẫu ngẫu nhiên kích thước $n = 10$, tính xác suất để trung bình mẫu của mẫu ngẫu nhiên này nhận giá trị 0,5.

4.18 Giả sử biến ngẫu nhiên gốc có phân bố chuẩn $\mathbf{N}(20;1)$. Chọn mẫu ngẫu nhiên kích thước $n = 100$. Hãy tính xác suất để trung bình mẫu $\overline{X}$ nằm trong khoảng:

$$19,8 < \overline{X} < 20,2 .$$

4.19 Một mẫu cụ thể của biến ngẫu nhiên X như sau:

$$2\ ;\ 3\ ;\ 2\ ;\ 4\ ;\ 1\ ;\ 4\ ;\ 2\ ;\ 2\ ;\ 3\ ;\ 1\ \ (n = 10).$$

a. Lập bảng phân bố tần suất.

b. Xây dựng hàm phân bố thực nghiệm.

c. Tính $\overline{x}$, s^2, s.

4.20 Theo dõi thời gian và số người hoàn thành một sản phẩm ở hai nhóm công nhân ta có bảng sau:

Nhóm 1

X (phút)	42	44	50	58	60	64
r_i (số người)	4	5	20	10	8	3

Nhóm 2

X (phút)	46	48	51
r_i (số người)	2	40	8

Tính $\overline{x}$ và độ lệch chuẩn mẫu s của hai mẫu cụ thể trên. Cho nhận xét.

CHƯƠNG 5: ƯỚC LƯỢNG CÁC THAM SỐ CỦA BIẾN NGẪU NHIÊN

Trong thực tế, có nhiều trường hợp dấu hiệu nghiên cứu của tổng thể có thể xác định được dạng phân bố phụ thuộc các tham số nào đó bằng lý thuyết. Lúc đó bài toán đưa về ước lượng các tham số này.

Chương này giới thiệu hai phương pháp sử dụng thống kê để ước lượng tham số: đó là ước lượng điểm và ước lượng bằng khoảng tin cậy.

Phương pháp ước lượng điểm dùng giá trị cụ thể của thống kê để thay cho một tham số nào đó theo các tiêu chuẩn: vững, không chệch, hiệu quả. Hai phương pháp tìm ước lượng điểm thường được sử dụng là phương pháp mômen và phương pháp hợp lý cực đại.

Khoảng tin cậy là khoảng mà tham số của dấu hiệu nghiên cứu của tổng thể rơi vào khoảng này với xác suất bằng độ tin cậy.

Trong chương này ta sẽ xây dựng ước lượng cho kỳ vọng, phương sai của dấu hiệu nghiên cứu có phân bố chuẩn và ước lượng cho tần suất của tổng thể.

Để học tốt học viên cần nắm vững chương 4 về lý thuyết mẫu.

5.1. PHƯƠNG PHÁP ƯỚC LƯỢNG ĐIỂM

Bằng cách áp dụng nguyên lý xác suất lớn, phương pháp ước lượng điểm chủ trương dùng một giá trị để thay cho giá trị của tham số θ chưa biết của tổng thể. Thông thường giá trị được chọn này giá trị cụ thể của một thống kê $\widehat{\Theta}$ nào đó của mẫu ngẫu nhiên.

Với mẫu ngẫu nhiên $W=(X_1, X_2,..., X_n)$, thống kê ước lượng cho tham số θ có dạng:

$$\widehat{\Theta}=T(X_1, X_2,..., X_n)$$

Khi đó với giá trị cụ thể của mẫu $w=(x_1, x_2,..., x_n)$ giá trị quan sát $\hat{\theta}=T_{qs}(x_1, x_2,..., x_n)$ của thống kê là ước lượng của tham số θ.

Cùng với một mẫu ngẫu nhiên có thể xây dựng nhiều thống kê $\widehat{\Theta}$ khác nhau để ước lượng cho tham số θ. Vì vậy cần đặt các tiêu chuẩn cho việc ước lượng tốt.

5.1.1 Ước lượng không chệch (unbiased estimator)

Thống kê $\widehat{\Theta}=T(X_1, X_2,..., X_n)$ là một hàm của các biến ngẫu nhiên $X_1, X_2,..., X_n$ nên cũng là một biến ngẫu nhiên. Do đó ta có thể xét các đặc trưng của thống kê này.

Định nghĩa 5.1: *Thống kê $\widehat{\Theta}=T(X_1, X_2,..., X_n)$ được gọi là ước lượng không chệch của tham số θ của tổng thể nếu*

$$\mathrm{E}\left[\widehat{\Theta}(X_1, X_2,..., X_n)\right]=\theta \tag{5.1}$$

Nếu $\mathrm{E}(\widehat{\Theta})\neq\theta$ thì $\widehat{\Theta}=T(X_1, X_2,..., X_n)$ được gọi là ước lượng chệch của θ.

Như vậy ước lượng không chệch có trung bình bằng tham số cần ước lượng.

Ví dụ 5.1: Dựa vào các công thức (4.8), (4.18), (4.19) của lý thuyết mẫu ta có các kết quả sau:

- Trung bình mẫu $\overline{X}$ là ước lượng không chệch của kỳ vọng μ của biến ngẫu nhiên gốc của tổng thể.
- Phương sai mẫu S^2 và S^{*2} là ước lượng không chệch cho phương sai σ^2 của biến ngẫu nhiên gốc của tổng thể.
- Tần suất mẫu f là ước lượng không chệch của tần suất p của tổng thể.

5.1.2 Ước lượng hiệu quả (efficient estimator)

Điều kiện (5.1) của ước lượng không chệch có nghĩa rằng trung bình các giá trị của $\widehat{\Theta}$ bằng giá trị θ. Tuy nhiên từng giá trị của $\widehat{\Theta}$ có thể sai lệch rất lớn so với θ. Vì vậy ta tìm ước lượng không chệch sao cho độ sai lệch trung bình là bé nhất.

Định nghĩa 5.2: *Ước lượng không chệch có phương sai nhỏ nhất so với mọi ước lượng không chệch khác được xây dựng trên cùng một mẫu ngẫu nhiên gọi là ước lượng hiệu quả (hay ước lượng phương sai bé nhất).*

Như vậy, để xét xem ước lượng không chệch $\widehat{\Theta}$ có phải là ước lượng hiệu quả của θ hay không ta cần phải tìm một cận dưới của phương sai của các ước lượng không chệch và so sánh phương sai của $\widehat{\Theta}$ với cận dưới này. Điều này được giải quyết bằng bất đẳng thức Cramer-Rao phát biểu như sau:

Cho mẫu ngẫu nhiên $W=(X_1,X_2,...,X_n)$ được lấy từ tổng thể với dấu hiệu nghiên cứu là biến ngẫu nhiên X có hàm mật độ xác suất (hay hàm khối lượng xác suất) $f(x,\theta)$ thỏa mãn một số điều kiện nhất định (thường được thỏa mãn trong thực tế, ít ra là các phân bố xác suất đã xét trong chương 2) và $\widehat{\Theta}$ là ước lượng không chệch bất kỳ của θ. Khi đó

$$\mathrm{D}\left(\widehat{\Theta}\right)\ge \frac{1}{n\,\mathrm{E}\left(\dfrac{\partial\left(\ln f(X,\theta)\right)}{\partial\theta}\right)^2} \qquad \textbf{(5.2)}$$

Ví dụ 5.2: Dựa vào bất đẳng thức trên ta có thể chứng minh được rằng trung bình mẫu $\overline{X}$ là ước lượng hiệu quả của kỳ vọng μ của dấu hiệu nghiên cứu X của tổng thể có phân bố chuẩn $\mathbf{N}(\mu,\sigma^2)$.

Giải: Thật vậy theo công thức (4.8) ta có $\mathrm{D}\left(\overline{X}\right)=\dfrac{\sigma^2}{n}$. Mặt khác hàm mật độ của $X\sim\mathbf{N}(\mu,\sigma^2)$ có dạng

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

$$\Rightarrow\ \ln f(x,\mu,\sigma)=-\ln\left(\sigma\sqrt{2\pi}\right)-\frac{(x-\mu)^2}{2\sigma^2}\ \Rightarrow\ \frac{\partial\ln f(x,\mu,\sigma)}{\partial\mu}=\frac{x-\mu}{\sigma^2}$$

Vậy $n\mathrm{E}\left(\dfrac{\partial\left(\ln f(X,\mu,\sigma)\right)}{\partial\mu}\right)^2 = n\mathrm{E}\left(\dfrac{X-\mu}{\sigma^2}\right)^2 = \dfrac{n}{\sigma^4}\mathrm{E}\left(X-\mu\right)^2 = \dfrac{n}{\sigma^4}\mathrm{D}\left(X\right) = \dfrac{n}{\sigma^2}$.

Như vậy $\mathrm{D}\left(\overline{X}\right) = \dfrac{\sigma^2}{n}$ đạt giá trị cực tiểu của bất đẳng thức Cramer-Rao, do đó trung bình mẫu $\overline{X}$ là ước lượng hiệu quả của μ .

5.1.3 Ước lượng vững (consistent estimator)

Định nghĩa 5.3: *Thống kê $\widehat{\Theta} = T\left(X_1, X_2,..., X_n\right)$ được gọi là ước lượng vững của tham số θ của biến ngẫu nhiên gốc X nếu $\widehat{\Theta} = T\left(X_1, X_2,..., X_n\right)$ hội tụ theo xác suất đến θ khi $n \to \infty$*, nghĩa là với mọi $\varepsilon > 0$, luôn có

$$\lim_{n\to\infty} P\left\{\left|\widehat{\Theta}\left(X_1, X_2,..., X_n\right) - \theta\right| < \varepsilon\right\} = 1 \qquad \textbf{(5.3)}$$

Hoặc một cách tương đương

$$\lim_{n\to\infty} P\left\{\left|\widehat{\Theta}\left(X_1, X_2,..., X_n\right) - \theta\right| \ge \varepsilon\right\} = 0 . \qquad \textbf{(5.4)}$$

Theo hệ quả 3.2 của luật số lớn Trêbusep ta có trung bình mẫu $\overline{X}$ là ước lượng vững của kỳ vọng μ ; S^2, $\widehat{S}^2$ và S^{*2} (trường hợp kỳ vọng μ đã biết) là ước lượng vững của phương sai σ^2 của biến ngẫu nhiên gốc X của tổng thể.

Theo luật số lớn Bernoulli (công thức 3.43) tần suất mẫu f là ước lượng vững của tần suất p của tổng thể. Tần suất mẫu f là trung bình mẫu của biến ngẫu nhiên của tổng thể có phân bố Bernoulli, do đó là ước lượng không chệch. Bằng cách sử dụng bất đẳng thức Cramer-Rao ta cũng chứng minh được tần suất mẫu f là ước lượng hiệu quả của tần suất tổng thể.

Tóm lại ta có kết quả sau:

- Trung bình mẫu $\overline{X}$ là ước lượng không chệch, hiệu quả và vững của kỳ vọng μ của biến ngẫu nhiên gốc của tổng thể.
- Tần suất mẫu f là ước lượng không chệch, hiệu quả và vững của xác suất p của tổng thể.
- Phương sai mẫu S^2 và S^{*2} (trường hợp μ đã biết) là ước lượng không chệch và vững của phương sai σ^2 của biến ngẫu nhiên gốc của tổng thể.

5.2 PHƯƠNG PHÁP ƯỚC LƯỢNG BẰNG KHOẢNG TIN CẬY

Phương pháp ước lượng điểm nói trên có nhược điểm là khi kích thước mẫu bé thì ước lượng điểm có thể sai lệch rất nhiều so với giá trị của tham số cần ước lượng. Mặt khác phương pháp trên cũng không thể đánh giá được khả năng mắc sai lầm khi ước lượng là bao nhiêu. Do đó khi kích thước mẫu bé người ta thường dùng phương pháp ước lượng khoảng tin cậy. Theo phương pháp này từ mẫu ngẫu nhiên ta có thể tìm được khoảng [a; b] chứa tham số θ với xác suất

β đủ lớn cho trước (β được gọi là độ tin cậy và thường được chọn trong khoảng 0,95 đến 0,99).

5.2.1 Khái niệm khoảng tin cậy

Định nghĩa 5.4: *Khoảng* $[a;b]$ *có hai đầu mút là hai thống kê*

$$a = a(X_1, X_2, ..., X_n),\ b = b(X_1, X_2, ..., X_n) \qquad \textbf{(5.6)}$$

phụ thuộc mẫu ngẫu nhiên $W = (X_1, X_2, ..., X_n)$ *của biến ngẫu nhiên gốc* X *, gọi là khoảng tin cậy của tham số* θ *với độ tin cậy* β *nếu:*

$$P\{a(X_1, X_2, ..., X_n) \le \theta \le b(X_1, X_2, ..., X_n)\} = \beta \qquad \textbf{(5.7)}$$

Trong thực tế thường yêu cầu độ tin cậy β khá lớn, khi đó theo nguyên lý xác suất lớn biến cố $\{a(X_1, X_2, ..., X_n) \le \theta \le b(X_1, X_2, ..., X_n)\}$ hầu như chắc chắn sẽ xảy ra trong một phép thử.

Tiến hành một phép thử với mẫu ngẫu nhiên $W = (X_1, X_2, ..., X_n)$ ta thu được giá trị cụ thể của mẫu $w = (x_1, x_2, ..., x_n)$, tính được giá trị quan sát $a = a_{qs}(x_1, x_2, ..., x_n)$, $b = b_{qs}(x_1, x_2, ..., x_n)$.

Lúc đó có thể kết luận là: Qua mẫu cụ thể với độ tin cậy β tham số θ của biến ngẫu nhiên gốc X sẽ nằm trong khoảng $[a;b]$, tức là $a \le \theta \le b$.

5.2.2 Khoảng tin cậy của kỳ vọng của biến ngẫu nhiên có phân bố chuẩn

Hầu hết các biến ngẫu nhiên của tổng thể có phân bố chuẩn. Chẳng hạn trọng lượng, chiều cao, năng suất, điểm thi … là các dấu hiệu của tổng thể có phân bố chuẩn.

Ngoài ra theo định lý giới hạn trung tâm khi kích thước mẫu lớn thì trung bình mẫu của biến ngẫu nhiên tổng thể có phân bố tùy ý sẽ xấp xỉ phân bố chuẩn.

Giả sử biến ngẫu nhiên gốc X có phân bố chuẩn $\mathbf{N}(\mu;\sigma^2)$ nhưng chưa biết tham số μ của nó. Từ tổng thể rút ra một mẫu ngẫu nhiên kích thước n: $W = (X_1, X_2, ..., X_n)$, ta tìm khoảng tin cậy của μ.

5.2.2.1 Trường hợp phương sai σ^2 đã biết

Định lý 5.1: Khoảng tin cậy của tham số μ với độ tin cậy β có dạng:

$$\left[\overline{X} - U_{\alpha/2}\frac{\sigma}{\sqrt{n}};\ \overline{X} + U_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right] \qquad \textbf{(5.8)}$$

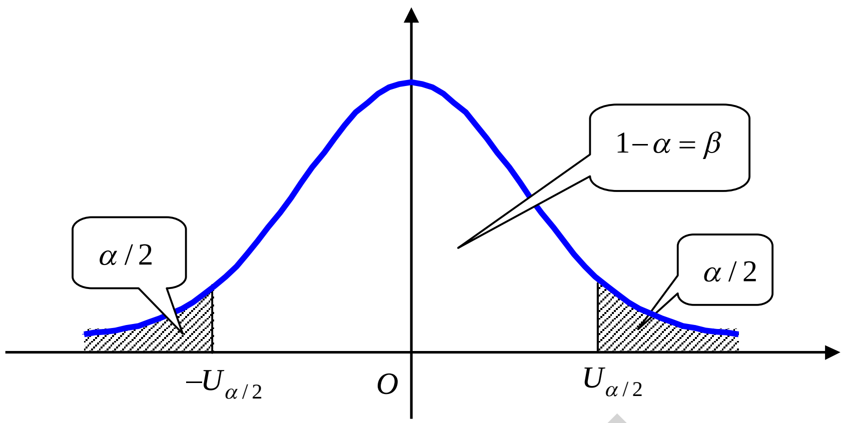
trong đó: $\alpha = 1-\beta$; $U_{\alpha/2}$ là giá trị tới hạn mức $\frac{\alpha}{2}$ của phân bố chuẩn tắc $\mathbf{N}(0;1)$ (công thức 2.42).

Chứng minh: Theo công thức (4.29) ta có $\frac{(\overline{X}-\mu)\sqrt{n}}{\sigma} \sim \mathbf{N}(0;1)$.

Mặt khác: $\overline{X} - U_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \mu \le \overline{X} + U_{\alpha/2}\frac{\sigma}{\sqrt{n}} \Leftrightarrow \left|\frac{(\overline{X}-\mu)\sqrt{n}}{\sigma}\right| \le U_{\alpha/2}$.

Áp dụng công thức (2.71) ta có

$$P\left[\overline{X}-U_{\alpha/2}\frac{\sigma}{\sqrt{n}}\le\mu\le\overline{X}+U_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]=P\left\{\left|\frac{(\overline{X}-\mu)\sqrt{n}}{\sigma}\right|\le U_{\alpha/2}\right\}=1-\alpha=\beta\,.$$



Hình 5.1: *Khoảng tin cậy của kỳ vọng phân bố chuẩn*

Định nghĩa 5.5: $\varepsilon=U_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}$ được gọi là *độ chính xác của ước lượng*.

Với phương sai đã biết (σ^2 không đổi) và độ tin cậy β không đổi thì giá trị tới hạn $U_{\alpha/2}$ không đổi, khi đó độ chính xác chỉ phụ thuộc vào kích thước mẫu n. Khi n càng lớn thì ε càng bé, do đó khoảng ước lượng càng chính xác. Nói cách khác độ chính xác phụ thuộc kích thước mẫu. Khoảng ước lượng có độ chính xác càng bé càng có ý nghĩa, vì vậy trong thực tế người ta mong muốn độ chính xác ở mức ý nghĩa cho trước.

Nếu muốn ước lượng với độ chính xác ε_0 và độ tin cậy β cho trước, *kích thước mẫu cần thiết* là số tự nhiên n nhỏ nhất thỏa mãn:

$$n\ge\frac{\sigma^2U_{\alpha/2}^2}{\varepsilon_0^2}\,. \qquad \textbf{(5.9)}$$

Ví dụ 5.5: Trọng lượng của một loại sản phẩm là một biến ngẫu nhiên phân bố theo quy luật chuẩn với độ lệch tiêu chuẩn 1 gram. Cần thử 25 sản phẩm loại này ta thu được kết quả:

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

Với độ tin cậy 95%

a. Hãy tìm khoảng tin cậy của trọng lượng trung bình của loại sản phẩm trên.

b. Nếu muốn độ chính xác của ước lượng không vượt quá 0,3 thì cần cân thử ít nhất bao nhiêu sản phẩm.

Giải: Gọi X là trọng lượng sản phẩm, theo giả thiết X có phân bố chuẩn $\mathbf{N}(\mu;\sigma^2)$ với $\sigma=1$.

Trọng lượng trung bình của sản phẩm là tham số μ. Khoảng tin cậy có dạng (5.7).

Với độ tin cậy $\beta = 0,95 \Rightarrow \dfrac{\alpha}{2} = 0,025 \Rightarrow U_{\frac{\alpha}{2}} = 1,96$.

a. Từ bảng số liệu tìm được trung bình mẫu cụ thể:

$$\overline{x} = \frac{3\cdot 18 + 5\cdot 19 + 15\cdot 20 + 2\cdot 21}{25} = 19,64.$$

Độ chính xác của ước lượng $\varepsilon = U_{\alpha/2}\dfrac{\sigma}{\sqrt{n}} = 1,96\cdot\dfrac{1}{\sqrt{25}} = 0,392$.

Vậy với độ tin cậy 95% qua mẫu cụ thể này, khoảng tin cậy của tham số μ là:

$$[19,64 - 0,392;\ 19,64 + 0,392]$$

$$19,248 \le \mu \le 20,032.$$

b. Nếu muốn độ chính xác của ước lượng không vượt quá 0,3 thì cần cân thử ít nhất n sản phẩm sao cho:

$$n \ge \frac{\sigma^2 U_{\alpha/2}^2}{\varepsilon_0^2} = \frac{1\cdot 1,96^2}{0,3^2} = 42,68.$$

Chọn $n = 43$.

5.2.2.2 Trường hợp phương sai σ^2 chưa biết, kích thước mẫu $n \ge 30$

Trong nhiều bài toán thực tế, ta không biết phương sai σ^2 của biến ngẫu nhiên gốc X của tổng thể. Nhưng khi kích thước mẫu n đủ lớn ($n \ge 30$) ta có thể xấp xỉ độ lệch chuẩn σ bởi độ lệch chuẩn mẫu S (vì S^2 là ước lượng vững không chệch của σ^2), S được xác định bởi công thức (4.16). Mặt khác, theo định lý giới hạn trung tâm thì thống kê $\dfrac{(\overline{X} - \mu)\sqrt{n}}{\sigma}$ có phân bố xấp xỉ phân bố chuẩn, đúng với mọi biến ngẫu nhiên gốc X của tổng thể (không đòi hỏi phân bố chuẩn).

Bằng cách chứng minh tương tự định lý 5.1 ta được khoảng tin cậy của tham số μ với độ tin cậy β có thể lấy là

$$\left[\overline{X} - U_{\alpha/2}\frac{S}{\sqrt{n}};\ \overline{X} + U_{\alpha/2}\frac{S}{\sqrt{n}}\right] \qquad \textbf{(5.10)}$$

Độ chính xác của ước lượng $\varepsilon = U_{\alpha/2}\dfrac{S}{\sqrt{n}}$.

Ví dụ 5.6: Để xác định chiều cao trung bình của các cây bạch đàn trong khu rừng rộng trồng bạch đàn, ta tiến hành đo ngẫu nhiên 35 cây và có kết quả cho trong bảng sau:

Khoảng	n_i	x_i	$u_i = x_i - 8,25$	$n_i u_i$	$n_i u_i^2$
$6,5 - 7,0$	2	6,75	$-1,5$	-3	4,5
$7,0 - 7,5$	4	7,25	$-1,0$	-4	4

$7,5-8,0$	10	7,75	$-0,5$	-5	2,5
$8,0-8,5$	11	8,25	0	0	0
$8,5-9,0$	5	8,75	0,5	2,5	1,25
$9,0-9,5$	3	9,25	1,0	3	3
$\sum$	35			$-6,5$	15,25

$\overline{u} = \dfrac{-6,5}{35} = -0,1857 \;\Rightarrow\; \overline{x} = 8,25 - 0,1857 \approx 8,06$.

$s^2 = s_u^2 = \dfrac{1}{34}\left(15,25 - \dfrac{(-6,5)^2}{35}\right) = 0,413 \;\Rightarrow\; s = 0,64$.

Với độ tin cậy $\beta = 95\%$, $U_{\alpha/2} = 1,96$.

Độ chính xác của ước lượng $\varepsilon = U_{\alpha/2}\dfrac{s}{\sqrt{n}} = 1,96 \cdot \dfrac{0,64}{\sqrt{35}} = 0,21$.

Vậy với độ tin cậy 95% có thể xem chiều cao trung bình μ của các cây bạch đàn thỏa mãn:

$$7,85 \le \mu \le 8,27 .$$

5.2.2.3 Trường hợp phương sai σ^2 chưa biết, kích thước mẫu $n < 30$

Trong trường hợp này, theo công thức (4.32) thống kê

$$T = \frac{(\overline{X}-\mu)\sqrt{n}}{S} \sim \mathbf{T}(n-1) \qquad \textbf{(5.11)}$$

có phân bố Student $n-1$ bậc tự do. Vì vậy khoảng tin cậy được tính theo kết quả sau:

Định lý 5.2: Khoảng tin cậy của tham số μ với độ tin cậy β có dạng:

$$\left[\overline{X} - t_{\alpha/2}(n-1)\frac{S}{\sqrt{n}};\; \overline{X} + t_{\alpha/2}(n-1)\frac{S}{\sqrt{n}}\right] \qquad \textbf{(5.12)}$$

trong đó $t_{\alpha/2}(n-1)$ là giá trị tới hạn mức $\alpha/2$ của phân bố Student $n-1$ bậc tự do (công thức 2.57).

Độ chính xác của ước lượng:

$$\varepsilon = t_{\alpha/2}(n-1)\frac{S}{\sqrt{n}} \qquad \textbf{(5.13)}$$

Với độ tin cậy β và độ chính xác ε_0 cho trước thì kích thước mẫu cần thiết là số tự nhiên n nhỏ nhất thỏa mãn:

$$n \ge \left(\frac{S \cdot t_{\alpha/2}(n-1)}{\varepsilon_0}\right)^2 \qquad \textbf{(5.14)}$$

Ví dụ 5.7: Năng suất của một loại giống mới là một biến ngẫu nhiên có quy luật phân bố chuẩn

$\mathbf{N}(\mu;\sigma^2)$. Gieo thử giống này trên 16 mảnh vườn thí nghiệm thu được như sau (đơn vị kg/ha):

172, 173, 173,174, 174, 175, 176, 166, 166, 167, 165, 173, 171, 170, 171, 170.

Hãy tìm khoảng tin cậy cho năng suất trung bình của loại giống này với độ tin cậy $\beta = 95\%$.

Giải: Năng suất trung bình của hạt giống là tham số μ .

Từ các số liệu trên ta tính được: $\overline{x} = 171;\ s = 3,4254$. $\alpha = 0,05;\ \dfrac{\alpha}{2} = 0,025$.

Tra bảng phân bố Student với 15 bậc tự do ta tìm được $t_{\alpha/2}(n-1) = t_{0,025}(15) = 2,131$.

Độ chính xác $\varepsilon = t_{\alpha/2}(n-1)\dfrac{s}{\sqrt{n}} = 2,131 \cdot \dfrac{3,4254}{\sqrt{16}} = 1,885$.

Vậy với độ tin cậy 95% có thể xem năng suất trung bình của loại hạt giống này là μ thỏa mãn:

$$169,115 \le \mu \le 172,885.$$

5.2.3 Khoảng tin cậy cho tham số *p* của biến ngẫu nhiên gốc có phân bố Bernoulli

Ta cần nghiên cứu một dấu hiệu định tính A nào đó mà mỗi cá thể của tổng thể có thể có hoặc không. Nếu cá thể có dấu hiệu A ta cho nhận giá trị 1, trường hợp ngược lại ta cho nhận giá trị 0. Lúc đó dấu hiệu nghiên cứu có thể xem là biến ngẫu nhiên X có phân bố Bernoulli với tham số p . Kỳ vọng $\mathrm{E}X = p$ và phương sai $\mathrm{D}X = p(1-p)$.

Lấy mẫu ngẫu nhiên $W = (X_1, X_2, ..., X_n)$ của biến ngẫu nhiên gốc X có phân bố Bernoulli với tham số p , khi đó $X_1, X_2, ..., X_n$ là các biến ngẫu nhiên độc lập có cùng phân bố Bernoulli tham số p .

Tần suất mẫu $f = \dfrac{X_1 + \cdots + X_n}{n}$.

Theo định lý Moivre-Laplace và công thức (4.40) ta có thể xấp xỉ phân bố xác suất của thống kê $U = \dfrac{(f-p)\sqrt{n}}{\sqrt{p(1-p)}}$ với phân bố chuẩn tắc $\mathbf{N}(0;1)$ khi n đủ lớn.

Tuy nhiên vì p chưa biết nên chưa biết $p(1-p) = \mathrm{D}X$.

Mặt khác tần suất mẫu f là ước lượng vững, không chệch và hiệu quả của xác suất p tổng thể. Vì vậy khi n đủ lớn ta có thể thay p bằng f .

Do đó khoảng tin cậy cho xác suất p của tổng thể với độ tin cậy β là:

$$\left[f - U_{\alpha/2}\sqrt{\frac{f(1-f)}{n}}\ ;\ f + U_{\alpha/2}\sqrt{\frac{f(1-f)}{n}} \right] \quad \textbf{(5.15)}$$

Với điều kiện n đủ lớn:

$$\begin{cases} nf > 10 \\ n(1-f) > 10 \end{cases} \quad \textbf{(5.16)}$$

trong đó $U_{\alpha/2}$ là giá trị tới hạn mức $\alpha/2$ của phân bố chuẩn tắc $\mathbf{N}(0;1)$ với $\alpha=1-\beta$.

Độ chính xác của khoảng tin cậy:

$$\varepsilon=U_{\alpha/2}\sqrt{\frac{f(1-f)}{n}}\,.$$

Với độ tin cậy β và độ chính xác ε_0 cho trước thì *kích thước mẫu cần thiết* là số tự nhiên n nhỏ nhất thỏa mãn:

$$n\ge f(1-f)\left(\frac{U_{\alpha/2}}{\varepsilon_0}\right)^2 \qquad \textbf{(5.17)}$$

trong đó f là tần suất mẫu của một mẫu ngẫu nhiên nào đó.

Ví dụ 5.8: Trong đợt vận động bầu cử tổng thống ở một nước nọ, người ta phỏng vấn ngẫu nhiên 1600 cử tri, được biết có 960 người trong số đó sẽ bỏ phiếu cho ứng cử viên A.

a. Với độ tin cậy $\beta=95\%$ thì ứng cử viên A sẽ chiếm được tối thiểu bao nhiêu % số phiếu bầu.

b. Với độ tin cậy $\beta=95\%$, nếu muốn độ chính xác của ước lượng không vượt quá 0,02 thì cần phỏng vấn tối thiểu bao nhiêu cử tri.

Giải: Gọi p là tỉ lệ số phiếu sẽ bầu cho ứng cử viên A. Tổng thể nghiên cứu là tập hợp tất cả các cử tri. Dấu hiệu nghiên cứu là cử tri sẽ bỏ phiếu cho A, có thể xem là biến ngẫu nhiên có phân bố Bernoulli tham số p. Khoảng tin cậy cho p có dạng (5.14) với điều kiện (5.15).

a. Từ mẫu cụ thể trên ta có $f=\dfrac{960}{1600}=0,6$ thỏa mãn điều kiện $\begin{cases} nf=960>10 \\ n(1-f)=640>10 \end{cases}$.

Độ chính xác của ước lượng $\varepsilon=U_{\alpha/2}\sqrt{\dfrac{f(1-f)}{n}}=1,96\sqrt{\dfrac{0,6\cdot 0,4}{1600}}=0,024$.

Khoảng tin cậy: $0,576\le p\le 0,624$.

Vậy với độ tin cậy 95% thì tối thiểu có 57,6% cử tri sẽ bỏ phiếu cho ứng cử viên A.

b. Theo công thức 5.17 ta có

$$n\ge 0,6.0,4\left(\frac{1,96}{0,02}\right)^2=2304,96\,;\ \text{chọn } n=2.305\,.$$

5.2.4 Ước lượng phương sai của biến ngẫu nhiên có phân bố chuẩn

Giả sử biến ngẫu nhiên gốc X của tổng thể có phân bố chuẩn $\mathbf{N}(\mu;\sigma^2)$, phương sai σ^2 chưa biết và ta cần ước lượng. Từ tổng thể rút ra một mẫu ngẫu nhiên kích thước n:

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

Ta sẽ chọn thống kê thích hợp để ước lượng cho tham số σ^2 phụ thuộc điều kiện kỳ vọng μ đã biết hoặc chưa biết.

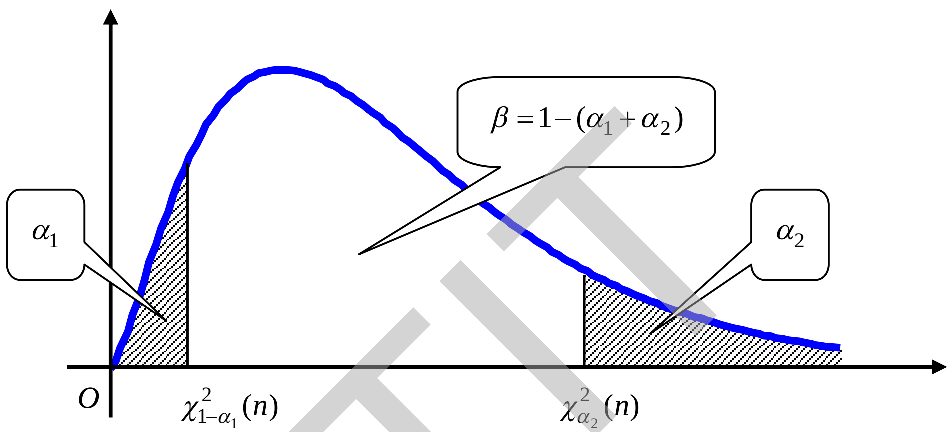
5.2.4.1 Trường hợp kỳ vọng μ đã biết

Chọn thống kê

$$T=\frac{nS^{*2}}{\sigma^2}=\frac{\sum_{i=1}^{n}\left(X_i-\mu\right)^2}{\sigma^2} \qquad \textbf{(5.18)}$$

Theo công thức (4.30) thống kê T có phân bố "khi bình phương" n bậc tự do. Do đó với độ tin cậy β cho trước, với cặp số $\alpha_1,\ \alpha_2$ sao cho $\alpha_1+\alpha_2=1-\beta$ có thể tìm hai giá trị tới hạn của T mức $\alpha_1,\ \alpha_2$ là $\chi^2_{1-\alpha_1}(n)$, $\chi^2_{\alpha_2}(n)$ xác định bởi:

$$P\left\{T>\chi^2_{1-\alpha_1}(n)\right\}=1-\alpha_1 \ \text{ và } \ P\left\{T>\chi^2_{\alpha_2}(n)\right\}=\alpha_2. \qquad \textbf{(5.19)}$$



Hình 5.2: Khoảng tin cậy của phương sai

Ta được

$$P\left\{\chi^2_{1-\alpha_1}(n)<T<\chi^2_{\alpha_2}(n)\right\}=1-(\alpha_1+\alpha_2)=\beta. \qquad \textbf{(5.20)}$$

Thay thống kê T từ công thức (5.18) vào biến cố ở công thức (5.20), ta được:

$$\left\{\chi^2_{1-\alpha_1}(n)<T<\chi^2_{\alpha_2}(n)\right\}=\left\{\chi^2_{1-\alpha_1}(n)<\frac{nS^{*2}}{\sigma^2}<\chi^2_{\alpha_2}(n)\right\}=\left\{\frac{nS^{*2}}{\chi^2_{\alpha_2}(n)}<\sigma^2<\frac{nS^{*2}}{\chi^2_{1-\alpha_1}(n)}\right\}$$

Do đó

$$P\left\{\frac{nS^{*2}}{\chi^2_{\alpha_2}(n)}<\sigma^2<\frac{nS^{*2}}{\chi^2_{1-\alpha_1}(n)}\right\}=\beta. \qquad \textbf{(5.21)}$$

Như vậy, với độ tin cậy β khoảng tin cậy của phương sai σ^2 có dạng:

$$\left(\frac{nS^{*2}}{\chi^2_{\alpha_2}(n)};\frac{nS^{*2}}{\chi^2_{1-\alpha_1}(n)}\right) \qquad \textbf{(5.22)}$$

Tùy theo cách chọn mức $\alpha_1,\ \alpha_2$ thỏa mãn $\alpha_1+\alpha_2=1-\beta$ ta nhận được các khoảng tin cậy của phương sai σ^2 với độ tin cậy β :

- Trường hợp $\alpha_1 = \alpha_2 = \dfrac{\alpha}{2}$, theo công thức (5.22) khoảng tin cậy đối xứng có dạng:

$$\left(\frac{nS^{*2}}{\chi^2_{\alpha/2}(n)} ; \frac{nS^{*2}}{\chi^2_{1-\alpha/2}(n)} \right) \tag{5.23}$$

- Trường hợp $\alpha_1 = 0; \alpha_2 = \alpha$

Công thức (5.20) trở thành $P\left\{T < \chi^2_\alpha(n)\right\} = 1 - \alpha = \beta$

$$\left\{T < \chi^2_\alpha(n)\right\} = \left\{ \frac{nS^{*2}}{\sigma^2} < \chi^2_\alpha(n) \right\} = \left\{ \frac{nS^{*2}}{\chi^2_\alpha(n)} < \sigma^2 \right\} \Rightarrow P\left\{ \frac{nS^{*2}}{\chi^2_\alpha(n)} < \sigma^2 \right\} = 1 - \alpha = \beta$$

Vậy khoảng tin cậy bên phải của σ^2 có dạng:

$$\left(\frac{nS^{*2}}{\chi^2_\alpha(n)} ; +\infty \right) \tag{5.24}$$

- Trường hợp $\alpha_2 = 0; \alpha_1 = \alpha$

Công thức (5.20) trở thành $P\left\{\chi^2_{1-\alpha}(n) < T\right\} = 1 - \alpha = \beta$

$$\left\{\chi^2_{1-\alpha}(n) < T\right\} = \left\{ \chi^2_{1-\alpha}(n) < \frac{nS^{*2}}{\sigma^2} \right\} = \left\{ \sigma^2 < \frac{nS^{*2}}{\chi^2_{1-\alpha}(n)} \right\} \Rightarrow P\left\{ \sigma^2 < \frac{nS^{*2}}{\chi^2_{1-\alpha}(n)} \right\} = 1 - \alpha = \beta$$

Vậy khoảng tin cậy bên trái của σ^2 có dạng:

$$\left(0 ; \frac{nS^{*2}}{\chi^2_{1-\alpha}(n)} \right) \tag{5.25}$$

Nếu không nói rõ tìm khoảng tin cậy bên phải hay bên trái ta thì ta quy ước là cần tìm khoảng tin cậy đối xứng.

Ví dụ 5.9: Mức hao phí nguyên liệu cho 1 đơn vị sản phẩm là biến ngẫu nhiên có phân bố 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 sau:

Hao phí nguyên liệu (gam)	19,5	20,0	20,5
Số sản phẩm tương ứng	5	18	2

Với độ tin cậy $\beta = 90\%$ hãy tìm khoảng tin cậy của σ^2 nếu $\alpha_1 = \alpha_2 = \dfrac{\alpha}{2} = 0,05$.

Giải: Gọi X là mức hao phí nguyên liệu cho 1 đơn vị sản phẩm. X có phân bố chuẩn với kỳ vọng đã biết $\mu = 20$. Đây là ước lượng phương sai σ^2 của phân bố chuẩn $\mathbf{N}(\mu;\sigma^2)$ khi đã biết μ. Khoảng tin cậy đối xứng theo công thức (5.23).

Tra bảng $\chi^2(n)$ ta có:

$$\chi^2_{\alpha/2}(n) = \chi^2_{0,05}(25) = 37,65;\ \chi^2_{1-\alpha/2}(n) = \chi^2_{0,95}(25) = 14,61.$$

Để tìm s^{*2} ta lập bảng sau:

x_i	r_i	$x_i-\mu$	$(x_i-\mu)^2$	$r_i(x_i-\mu)^2$
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
$\sum$	25			1,75

$s^{*2}=\dfrac{1,75}{25}=0,07$.

Vậy với độ tin cậy 90%, qua mẫu cụ thể này, khoảng tin cậy của σ^2 là:

$$\left(\frac{25\cdot 0,07}{37,65};\frac{25\cdot 0,07}{14,61}\right)\ \text{hay}\quad 0,0464<\sigma^2<0,1198.$$

5.2.4.2. Chưa biết kỳ vọng μ

Chọn thống kê

$$T=\frac{(n-1)S^2}{\sigma^2}=\sum_{i=1}^{n}\frac{\left(X_i-\overline{X}\right)^2}{\sigma^2} \qquad \textbf{(5.26)}$$

Theo công thức (4.31) thống kê T có phân bố khi bình phương $n-1$ bậc tự do: $\chi^2(n-1)$. Tương tự trường hợp trên, với độ tin cậy β cho trước, với cặp số α_1, α_2 sao cho $\alpha_1+\alpha_2=1-\beta$ ta có thể tìm hai giá trị tới hạn của T mức α_1, α_2 là $\chi^2_{1-\alpha_1}(n-1)$, $\chi^2_{\alpha_2}(n-1)$ xác định bởi:

$$P\left\{T>\chi^2_{1-\alpha_1}(n-1)\right\}=1-\alpha_1 \text{ và } P\left\{T>\chi^2_{\alpha_2}(n-1)\right\}=\alpha_2 \qquad \textbf{(5.27)}$$

Do đó

$$P\left\{\chi^2_{1-\alpha_1}(n-1)<T<\chi^2_{\alpha_2}(n-1)\right\}=1-(\alpha_1+\alpha_2)=\beta \qquad \textbf{(5.28)}$$

Thay thống kê T từ công thức (5.26) vào biến cố trong công thức (5.28)

$$\left\{\chi^2_{1-\alpha_1}(n-1)<T<\chi^2_{\alpha_2}(n-1)\right\}=\left\{\chi^2_{1-\alpha_1}(n-1)<\frac{(n-1)S^2}{\sigma^2}<\chi^2_{\alpha_2}(n-1)\right\}$$

$$=\left\{\frac{(n-1)S^2}{\chi^2_{\alpha_2}(n-1)}<\sigma^2<\frac{(n-1)S^2}{\chi^2_{1-\alpha_1}(n-1)}\right\}$$

Theo công thức (5.28) ta được

$$P\left\{\frac{(n-1)S^2}{\chi^2_{\alpha_2}(n-1)}<\sigma^2<\frac{(n-1)S^2}{\chi^2_{1-\alpha_1}(n-1)}\right\}=\beta \qquad \textbf{(5.29)}$$

Như vậy, với độ tin cậy β khoảng tin cậy của σ^2 có dạng:

$$\left(\frac{(n-1)S^2}{\chi^2_{\alpha_2}(n-1)};\frac{(n-1)S^2}{\chi^2_{1-\alpha_1}(n-1)}\right) \qquad \textbf{(5.30)}$$

Tùy theo cách chọn mức α_1, α_2 thỏa mãn $\alpha_1+\alpha_2=1-\beta$ ta nhận được các khoảng tin cậy của phương sai σ^2 với độ tin cậy β :

- Trường hợp $\alpha_1=\alpha_2=\frac{\alpha}{2}$ khoảng tin cậy đối xứng có dạng:

$$\left(\frac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)};\frac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)}\right) \qquad \textbf{(5.31)}$$

- Trường hợp $\alpha_1=0;\ \alpha_2=\alpha$, tương tự công thức (5.24) ta có khoảng tin cậy bên phải của σ^2 dạng:

$$\left(\frac{(n-1)S^2}{\chi^2_{\alpha}(n-1)};+\infty\right) \qquad \textbf{(5.32)}$$

- Trường hợp $\alpha_2=0;\ \alpha_1=\alpha$ khoảng tin cậy bên trái của σ^2 có dạng:

$$\left(0;\frac{(n-1)S^2}{\chi^2_{1-\alpha}(n-1)}\right) \qquad \textbf{(5.33)}$$

CÂU HỎI ÔN TẬP VÀ BÀI TẬP CHƯƠNG 5

5.1 Trung bình mẫu là ước lượng vững và hiệu quả của kỳ vọng của biến ngẫu nhiên gốc.

Đúng ☐ Sai ☐ .

5.2 Có thể tìm được ước lượng không chệch của θ có phương sai nhỏ hơn đại lượng $\dfrac{1}{n\mathrm{E}\left(\dfrac{\partial\left(\ln f(X,\theta)\right)}{\partial\theta}\right)^2}$.

Đúng ☐ Sai ☐.

5.3 Trung bình cộng của hai ước lượng không chệch là một ước lượng không chệch.

Đúng ☐ Sai ☐.

5.4 Phương sai mẫu hiệu chỉnh S^2 là ước lượng vững không chệch của phương sai của biến ngẫu nhiên gốc.

Đúng ☐ Sai ☐ .

5.5 Việc tìm điểm cực đại của hàm hợp lý $L(x_1,x_2,...,x_n,\theta)$ tương đương với việc tìm điểm cực đại của hàm $\ln L(x_1,x_2,...,x_n,\theta)$.

Đúng ☐ Sai ☐.

5.6 Mọi ước lượng vững là ước lượng hiệu quả.

Đúng $\square$ Sai $\square$.

5.7 Hai đầu mút của khoảng tin cậy là hai thống kê của mẫu.

Đúng $\square$ Sai $\square$.

5.8 Muốn tìm khoảng tin cậy cho tham số μ của biến ngẫu nhiên gốc có phân bố chuẩn $\mathbf{N}(\mu;\sigma^2)$ thì kích thước mẫu n phải lớn hơn 30.

Đúng $\square$ Sai $\square$.

5.9 Để tìm khoảng tin cậy cho tham số p của biến ngẫu nhiên gốc có phân bố Bernoulli ta xấp xỉ tần suất mẫu với quy luật chuẩn tắc $\mathbf{N}(0;1)$ nếu n thỏa mãn điều kiện $\begin{cases} nf > 10 \\ n(1-f) > 10 \end{cases}$.

Đúng $\square$ Sai $\square$.

5.10 Có thể tìm kích thước mẫu cần thiết để khoảng tin cậy cho tham số p của biến ngẫu nhiên gốc có phân bố Bernoulli thỏa mãn độ tin cậy và độ chính xác cho trước.

Đúng $\square$ Sai $\square$.

5.11 Cho mẫu ngẫu nhiên $W=(X_1, X_2,..., X_n)$ của biến ngẫu nhiên gốc X có phân bố mũ tham số $\lambda>0$. Chứng minh rằng $\frac{n-1}{n}\overline{X}$ là ước lượng không chệch của λ.

5.12 Một nghiên cứu trên 50 em bé 6 tuổi cho thấy số giờ xem tivi trung bình trong một tuần của nhóm này là 38 giờ với độ lệch tiêu chuẩn là 6,4 giờ. Tìm khoảng tin cậy 99% cho thời gian xem tivi trung bình trong một tuần của các em nhỏ 6 tuổi.

5.13 Trong đợt vận động bầu cử tổng thống ở một nước nọ, người ta phỏng vấn ngẫu nhiên 2000 cử tri thì được biết có 1082 người trong số đó sẽ bỏ phiếu cho ứng cử viên A. Với độ tin cậy 98% tối thiểu ứng cử viên A sẽ chiếm được bao nhiêu % số phiếu bầu?

5.14 Để xác định sản lượng khai thác điện thoại của đơn vị mình, một đơn vị đã tiến hành thống kê ngẫu nhiên 35 ngày và thu được kết quả sau với đơn vị 100.000 phút/ngày:

0,84 0,96 1,02 1,08 0,88 0,80 0,91 0,97 1,07 0,98 1,04 1,13 0,87 0,82 1,01
0,93 1,03 1,10 0,97 1,05 0,83 0,76 0,95 1,15 1,00 1,05 1,14 0,89 0,81
0,95 1,20 1,16 1,24 0,79 0,77.

Tìm khoảng tin cậy 95% cho sản lượng điện thoại trung bình mỗi ngày.

5.15 Muốn ước lượng số cá trong hồ, người ta bắt 2000 con cá trong hồ đánh dấu rồi thả lại xuống hồ. Sau đó bắt lại 400 con và thấy có 53 con có dấu. Hãy ước lượng số cá trong hồ với độ tin cậy là 95%.

5.16 Hao phí nguyên liệu cho một đơn vị sản phẩm là một biến ngẫu nhiên có phân bố chuẩn với độ lệch chuẩn $\sigma=0,03$. Sản xuất thử 36 sản phẩm và thu được số liệu sau:

Mức hao phí nguyên liệu	19,5 – 19,7	19,7 –19,9	19,9 –20,1	90,1 –20,3
Số sản phẩm	8	8	18	2

Với độ tin cậy 95% hãy ước lượng hao phí nguyên liệu trung bình cho 1 đơn vị sản phẩm.

5.17 Để xác định chiều cao trung bình của các cây con trong một vườn ươm người ta tiến hành đo ngẫu nhiên 40 cây. Kết quả đo được như sau:

Chiều cao (cm)	16,5 –17	17 –17,5	17,5 –18	18 –18,5	18,5 –19	19 –19,5
Số cây tương ứng	3	5	11	12	6	3

a. Tìm khoảng tin cậy 90% cho chiều cao trung bình của vườn cây con.

b. Nếu muốn khoảng ước lượng có độ chính xác $\varepsilon = 0,1$ thì cần lấy mẫu bao nhiêu cây.

5.18 Để ước lượng trọng lượng trung bình của một loại sản phẩm, người ta cân ngẫu nhiên 100 sản phẩm và thu được số liệu sau:

Trọng lượng (kg)	40 – 42	42 – 44	44 – 46	46 – 48	48 – 50	50 – 52
Số sản phẩm	7	13	25	35	15	5

Giả sử biến ngẫu nhiên chỉ trọng lượng X có phân bố chuẩn.

a. Tìm khoảng tin cậy 95% cho trọng lượng trung bình của sản phẩm.

b. Nếu muốn khoảng ước lượng có độ chính xác $\varepsilon = 0,4$ thì cần lấy mẫu gồm bao nhiêu sản phẩm.

5.19 Trọng lượng của một loại sản phẩm A là một biến ngẫu nhiên có phân bố theo quy luật chuẩn với độ lệch chuẩn là 1 gam. Cân thử 27 bao loại này ta thu được kết quả:

Trọng lượng (gam)	47,5 – 48,5	48,5 – 49,5	49,5 – 50,5	50,5 – 51,5	51,5 – 52,5
Số sản phẩm	3	6	15	2	1

a. Tìm khoảng tin cậy 95% của trọng lượng trung bình của loại sản phẩm trên.

b. Nếu muốn độ chính xác $\varepsilon = 0,1$ thì kích thước mẫu cần thiết là bao nhiêu.

5.20 Để xác định chiều cao trung bình của trẻ em 8 tuổi, người ta tiến hành ngẫu nhiên đo chiều cao của 100 em học sinh lớp 3 (8 tuổi) ở một số trường tiểu học và được kết quả:

Khoảng chiều cao	110-112	112-114	114-116	116-118	118-120	120-122	122-124	124-126	126-128
Số học sinh	3	8	14	17	20	16	10	6	4

a. Tìm khoảng tin cậy 95% cho chiều cao trung bình của trẻ em 8 tuổi.

b. Nếu muốn khoảng ước lượng có độ chính xác $\varepsilon = 0,5$ cm thì cần phải lấy mẫu kích thước bao nhiêu.

5.21 Để ước lượng tỷ lệ phần trăm phế phẩm của một lô hàng người ta tiến hành kiểm tra ngẫu nhiên 400 sản phẩm và nhận thấy có 16 phế phẩm. Với mức tin cậy 95% hãy ước lượng tỷ lệ phế phẩm tối đa của lô hàng.

5.22 Để xác định giá trung bình đối với một loại hàng hoá trên thị trường, người ta điều tra ngẫu nhiên tại 100 cửa hàng thu được số liệu sau đây

Giá X (nghìn đồng)	83	85	87	89	91	93	95	97	99	101
Số cửa hàng	6	7	12	15	30	10	8	6	4	2

Với độ tin cậy 95% hãy tìm khoảng tin cậy cho giá trung bình của loại hàng hoá nói trên.

5.23 Người ta đo một đại lượng không đổi 25 lần bằng một dụng cụ đo không có sai số hệ thống và sai số đo trung bình bằng 0. Giả sử sai số của phép đo là một biến ngẫu nhiên có phân bố chuẩn và phương sai mẫu đo được bằng 0,5. Hãy xác định khoảng tin cậy 95% cho phương sai của sai số đo.

CHƯƠNG 6: KIỂM ĐỊNH GIẢ THIẾT THỐNG KÊ

Một dạng khác của quy nạp thống kê là kiểm định giả thiết thống kê. Đây là một phương pháp quan trọng cho phép giải quyết nhiều bài toán trong thực tế. Nội dung của kiểm định giả thiết thống kê là dựa vào mẫu cụ thể và các quy tắc hay thủ tục quyết định dẫn đến bác bỏ hay chấp nhận giả thiết của tổng thể.

Giả thiết thống kê là giả thiết về dạng phân bố xác suất của biến ngẫu nhiên gốc của tổng thể hoặc các tham số đặc trưng hoặc tính chất của biến ngẫu nhiên này. Giả thiết thống kê là những điều ta muốn bảo vệ hoặc ta nghi ngờ muốn bác bỏ, được phát biểu dưới dạng H_0 (Null hypothesis). Cạnh tranh với giả thiết này là đối thiết H_1 (Alternative hypothesis), theo nghĩa rằng nếu bác bỏ H_0 thì chấp nhận H_1 và ngược lại.

Phép kiểm định giả thiết thống kê dựa vào hai nguyên lý: Phép chứng minh phản chứng và nguyên lý xác suất nhỏ. Để kiểm định giả thiết H_0, dựa vào hai nguyên lý này ta giả sử rằng H_0 đúng từ đó xây dựng một biến cố W_α có xác suất bé (bằng mức ý nghĩa của phép kiểm định). Theo nguyên lý xác suất nhỏ thì trong một lần thử biến cố W_α không xảy ra. Vì vậy nếu với một mẫu cụ thể nào đó mà biến cố W_α xảy ra thì giả thiết cho rằng H_0 đúng là vô lý do đó ta bác bỏ H_0, còn nếu W_α không xảy ra thì ta chưa có cơ sở để bác bỏ H_0. Biến cố W_α được gọi là miền bác bỏ.

Lý thuyết kiểm định giả thiết thống kê có nhiều ứng dụng trong thực tế, giúp các nhà quản lý kiểm tra tính đúng đắn của các quyết định.

Để học tốt chương này học viên cần nắm vững chương 4 về lý thuyết mẫu.

6.1 KHÁI NIỆM CHUNG VỀ KIỂM ĐỊNH GIẢ THIẾT THỐNG KÊ

Trong chương trước ta giải quyết các bài toán về ước lượng tham số đặc trưng của dấu hiệu nghiên cứu của tổng thể bằng cách đưa về ước lượng các tham số đặc trưng của các biến ngẫu nhiên gốc. Trong chương này ta sẽ nghiên cứu bài toán kiểm định giả thiết về các tham số đặc trưng của tổng thể.

Phương pháp kiểm định giả thiết thống kê là dựa vào mẫu cụ thể và các quy tắc hay thủ tục quyết định dẫn đến bác bỏ hay chấp nhận giả thiết của tổng thể. Để giải quyết bài toán này ta cần tìm hiểu các khái niệm sau: giả thiết thống kê, các nguyên tắc để xây dựng quy tắc kiểm định, miền bác bỏ, sai lầm khi kiểm định …

6.1.1 Giả thiết thống kê

Vì các dấu hiệu nghiên cứu có thể xem là các biến ngẫu nhiên gốc do đó giả thiết thống kê có thể là giả thiết về dạng phân bố xác suất. Chẳng hạn số khách hàng đến điểm phục vụ có theo quy luật phân bố Poisson hay không? Nhu cầu của thị trường đối với sản phẩm nào đó có theo quy luật chuẩn không?…

Khi đã biết dạng phân bố của biến ngẫu nhiên gốc nhưng chưa xác định được các tham số đặc trưng của biến ngẫu nhiên gốc (như giá trị trung bình, phương sai, tham số p của phân bố

Bernoulli …), thì giả thiết thống kê là giả thiết về tham số của phân bố đó.

Đối với bài toán có hai dấu hiệu nghiên cứu thì giả thiết thống kê có thể là giả thiết về sự độc lập của chúng hoặc so sánh các tham số đặc trưng của chúng.

Giả thiết đưa ra kiểm nghiệm được ký hiệu là H_0, gọi là "giả thiết không". Đó là giả thiết mà ta muốn bảo vệ hoặc ta nghi ngờ muốn bác bỏ. Ngoài giả thiết H_0 ra, ta còn phải định ra một giả thiết cạnh tranh với H_0 gọi là đối thiết, ký hiệu H_1. Đối thiết H_1 sẽ được chấp nhận khi H_0 bị bác bỏ.

Cần chú ý rằng đối thiết H_1 không nhất thiết là phủ định của giả thiết H_0. Chẳng hạn giả thiết H_0: nhu cầu thị trường về loại hàng hóa này là $\mu = 1000$ đơn vị/tháng. Nếu ta nghi ngờ rằng nhu cầu này không đúng thì đối thiết H_1 là $\mu \neq 1000$, nhưng nếu do tiếp thị tốt, do chính sách hậu mãi tốt người ta nghĩ rằng nhu cầu về mặt hàng này tăng lên thì đối thiết H_1 là $\mu > 1000$.

Qui tắc kiểm định dựa trên hai nguyên lý sau:

* ***Nguyên lý xác suất nhỏ***: "Nếu một biến cố có xác rất nhỏ thì trong một phép thử biến cố đó coi như không xảy ra".
* ***Phương pháp phản chứng***: "Nếu từ giả thiết H_0 đúng dẫn đến một điều vô lý thì ta bác bỏ H_0 "

Dựa vào hai nguyên lý này ta đưa ra phương pháp chung để kiểm định một giả thiết thống kê như sau: Để kiểm định H_0 trước hết giả sử H_0 đúng từ đó ta tìm được biến cố A mà xác suất xuất hiện biến cố W_α là rất bé và ta có thể xem W_α không thể xảy ra trong một phép thử. Thực hiện phép thử, nếu với mẫu cụ thể quan sát được mà biến cố W_α xảy ra thì điều này trái với nguyên lý xác suất nhỏ. Vậy H_0 sai và bác bỏ nó. Còn nếu W_α không xảy ra thì ta chưa có cơ sở để bác bỏ H_0.

Ta thực hiện phương pháp trên bằng các bước cụ thể sau:

6.1.2 Tiêu chuẩn kiểm định giả thiết thống kê

Miền bác bỏ W_α được xây dựng từ thống kê T của mẫu gọi là tiêu chuẩn kiểm định và được xác định như sau:

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

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

Chọn thống kê

$$T = T(X_1, X_2, ..., X_n, \theta) \tag{6.1}$$

trong đó θ là tham số liên quan đến giả thiết cần kiểm định.

Nếu H_0 đúng thì thống kê T có quy luật phân bố xác suất xác định, từ đó có thể xây dựng miền bác bỏ W_α.

6.1.3 Miền bác bỏ giả thiết

Sau khi đã chọn tiêu chuẩn kiểm định T, với α bé cho trước (thường α được lấy bằng 0,05 hoặc 0,01) và với điều kiện H_0 đúng ta có thể tìm được miền W_α sao cho T nhận giá trị trong miền W_α với xác suất bằng α :

$$P\left\{T \in W_\alpha \middle| \mathrm{H}_0\right\} = \alpha \qquad \textbf{(6.2)}$$

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

6.1.4 Giá trị quan sát của tiêu chuẩn kiểm định

Thực hiện phép thử với mẫu ngẫu nhiên $W = \left(X_1, X_2, ..., X_n\right)$ thu được mẫu cụ thể $w = (x_1, x_2, ..., x_n)$, thay giá trị này vào thống kê (6.1) ta được giá trị quan sát của tiêu chuẩn kiểm định:

$$T_{qs} = (x_1, x_2, ..., x_n, \theta_0) \qquad \textbf{(6.3)}$$

6.1.5 Quy tắc kiểm định giả thiết thống kê

So sánh giá trị quan sát của tiêu chuẩn kiểm định với miền bác bỏ W_α và kết luận theo quy tắc sau:

1. Nếu giá trị quan sát T_{qs} rơi vào miền bác bỏ W_α , theo nguyên tắc kiểm định thì H_0 sai, do đó ta bác bỏ H_0 thừa nhận H_1.
2. Nếu giá trị quan sát T_{qs} không rơi vào miền bác bỏ W_α thì điều này chưa khẳng định rằng H_0 đúng mà chỉ có nghĩa là qua mẫu cụ thể đó chưa khẳng định được là H_0 sai. Vì vậy ta chỉ có thể nói rằng qua mẫu cụ thể này chưa có cơ sở để bác bỏ H_0 (trên thực tế là thừa nhận H_0).

6.1.6 Sai lầm loại một và sai lầm loại hai

Với quy tắc kiểm định như trên có thể mắc hai loại sai lầm sau:

1. Sai lầm loại I: Đó là sai lầm mắc phải khi bác bỏ giả thiết H_0 trong khi H_0 đúng. Ta thấy xác suất mắc sai lầm loại I đúng bằng mức ý nghĩa α . Thật vậy, xác suất ta bác bỏ H_0 bằng xác suất biến cố $\left\{T \in W_\alpha\right\}$, do đó khi H_0 đúng thì xác suất này là xác suất có điều kiện $P\left\{T \in W_\alpha \middle| \mathrm{H}_0\right\} = \alpha$.

   Sai lầm loại I sinh ra do kích thước mẫu quá nhỏ, do phương pháp lấy mẫu v.v…

Thực tế / Quyết định	H_0 đúng	H_0 sai
Bác bỏ H_0	Sai lầm loại I Xác suất $= \alpha$	Quyết định đúng Xác suất $= 1 - \beta$
Không bác bỏ H_0	Quyết định đúng Xác suất $= 1 - \alpha$	Sai lầm loại II Xác suất $= \beta$

2. Sai lầm loại II: Đó là sai lầm mắc phải khi thừa nhận giả thiết H_0 trong khi H_0 sai, điều này xảy ra khi giá trị quan sát T_{qs} không thuộc miền bác bỏ W_α trong khi H_1 đúng.

Vậy xác suất sai lầm loại II là β xác định như sau:

$$P\left\{T \notin W_\alpha \middle| \mathrm{H}_1\right\} = \beta \qquad \textbf{(6.4)}$$

Xác suất của biến cố đối của sai lầm loại II: $P\left\{T \in W_\alpha \middle| H_1\right\} = 1-\beta$ gọi là ***lực lượng kiểm định.***

Ta muốn tìm một qui tắc kiểm định mà cả hai loại sai lầm trên là cực tiểu. Nhưng không tồn tại kiểm định lý tưởng như vậy, vì nói chung khi giảm sai lầm loại I thì sai lầm loại II tăng và ngược lại. Chẳng hạn nếu lấy $\alpha = 0$ thì sẽ không bác bỏ bất kỳ giả thiết nào, kể cả giả thiết sai, vậy β sẽ đạt cực đại. Mặt khác trong bài toán kiểm định thì giả thiết H_0 là giả thiết quan trọng, do đó sai lầm về nó càng nhỏ càng tốt. Vì vậy các nhà thống kê đưa ra phương pháp sau:

Sau khi ta chọn sai lầm loại I nhỏ ở mức ý nghĩa α, với mẫu kích thước n xác định, ta chọn ra miền bác bỏ W_α sao cho xác suất sai lầm loại II là nhỏ nhất hay lực lượng kiểm định là lớn nhất. Nghĩa là cần tìm miền bác bỏ W_α thỏa mãn điều kiện:

$$P\left\{T \in W_\alpha \middle| \mathrm{H}_0\right\} = \alpha \text{ và } P\left\{T \in W_\alpha \middle| \mathrm{H}_1\right\} = 1-\beta \to \max$$

Định lý Neymann - Pearson chỉ ra rằng nhiều bài toán quan trọng trong thực tiễn có thể tìm được miền bác bỏ W_α thỏa mãn điều kiện trên.

Việc chọn mức ý nghĩa α bằng bao nhiêu tùy thuộc vào từng trường hợp cụ thể, tùy thuộc vào ý nghĩa của bài toán.

6.1.7 Thủ tục kiểm định giả thiết thống kê

Qua nội dung trình bày ở trên ta có thể xây dựng một thủ tục kiểm định giả thiết thống kê bao gồm các bước sau:

a. Phát biểu giả thiết H_0 và đối thiết H_1.

b. Từ tổng thể nghiên cứu lập mẫu ngẫu nhiên kích thước n.

c. Chọn tiêu chuẩn kiểm định T và xác định quy luật phân bố xác suất của T với điều kiện giả thiết H_0 đúng.

d. Với mức ý nghĩa α, xác định miền bác bỏ W_α tốt nhất tùy thuộc vào đối thiết H_1.

e. Từ mẫu cụ thể tính giá trị quan sát của tiêu chuẩn kiểm định T_{qs}.

f. So sánh giá trị quan sát của tiêu chuẩn kiểm định T_{qs} với miền bác bỏ W_α và kết luận.

6.2 KIỂM ĐỊNH THAM SỐ

Trong mục này ta xét bài toán kiểm định tham số của biến ngẫu nhiên gốc có phân bố chuẩn (dấu hiệu định lượng) và biến ngẫu nhiên gốc có phân bố Bernoulli (dấu hiệu định tính).

6.2.1 Kiểm định giả thiết về kỳ vọng của biến ngẫu nhiên có phân bố chuẩn

Giả sử biến ngẫu nhiên gốc X trong tổng thể có phân bố chuẩn $\mathbf{N}(\mu;\sigma^2)$, cần kiểm định về giá trị của kỳ vọng μ. Nếu có cơ sở để giả thiết rằng kỳ vọng μ bằng giá trị μ_0 thì ta có thể đưa ra giả thiết thống kê

$$\mathrm{H}_0:\ \mu = \mu_0.$$

Bài toán phụ thuộc điều kiện phương sai đã biết hoặc chưa biết.

6.2.1.1 Trường hợp đã biết phương sai

Giả sử phương sai σ^2 của biến ngẫu nhiên gốc X trong tổng thể có phân bố chuẩn $\mathbf{N}(\mu;\sigma^2)$ đã biết, từ tổng thể rút ra một mẫu ngẫu nhiên kích thước n:

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

Xét thống kê

$$T=\frac{(\overline{X}-\mu)\sqrt{n}}{\sigma} \tag{6.5}$$

Theo công thức (4.29) chương 4 thống kê $T=\dfrac{(\overline{X}-\mu)\sqrt{n}}{\sigma}$ có phân bố chuẩn tắc $\mathbf{N}(0;1)$.

Mặt khác nếu giả thiết H_0 đúng thì thống kê T trở thành $T=\dfrac{(\overline{X}-\mu_0)\sqrt{n}}{\sigma}$, do đó

$$\{T\in W_\alpha \mid H_0\}=\left\{\frac{(\overline{X}-\mu_0)\sqrt{n}}{\sigma}\in W_\alpha\right\}$$

là biến cố thống kê T có giá trị rơi vào miền W_α với điều kiện giả thiết H_0 đúng.

Áp dụng công thức 2.71 chương 2 ta có thể xây dựng các miền bác bỏ dựa vào đối thiết H_1 như sau:

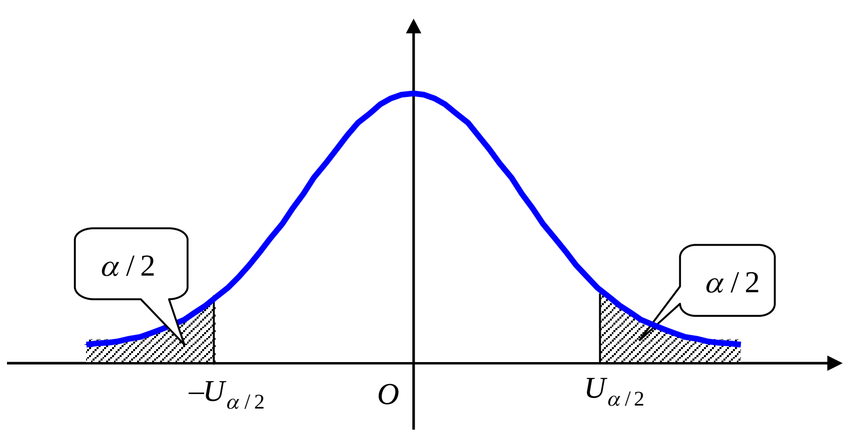
a. <u>Bài toán 1</u>: H_0: $\mu=\mu_0$; H_1: $\mu\neq\mu_0$. Ta nói đây là bài toán kiểm định hai phía.

Miền bác bỏ

$$W_\alpha=\left\{T=\frac{(\overline{X}-\mu_0)\sqrt{n}}{\sigma}\ ;\ |T|>U_{\alpha/2}\right\}. \tag{6.6}$$

trong đó $U_{\alpha/2}$, U_α là giá trị tới hạn mức $\alpha/2$ và mức α của phân bố chuẩn tắc $\mathbf{N}(0;1)$.

$$T=\frac{(\overline{X}-\mu_0)\sqrt{n}}{\sigma}\sim\mathbf{N}(0;1)\Rightarrow P(W_\alpha)=P\left\{\left|\frac{(\overline{X}-\mu_0)\sqrt{n}}{\sigma}\right|>U_{\alpha/2}\right\}=\alpha$$



Hình 6.1: *Miền bác bỏ của kỳ vọng phân bố chuẩn, bài toán 1*

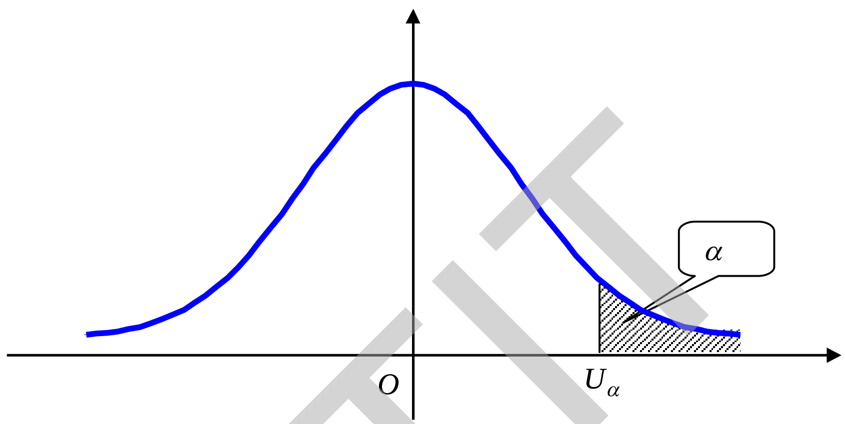
Như vậy miền bác bỏ W_α là biến cố thống kê $T = \dfrac{(\overline{X} - \mu_0)\sqrt{n}}{\sigma}$ có giá trị thỏa mãn $|T| > U_{\alpha/2}$.

b. <u>Bài toán 2</u>: H$_0$: $\mu = \mu_0$; H$_1$: $\mu > \mu_0$. Đây là bài toán kiểm định một phía.

Miền bác bỏ

$$W_\alpha = \left\{ T = \frac{(\overline{X} - \mu_0)\sqrt{n}}{\sigma}\ ;\ T > U_\alpha \right\}. \quad \textbf{(6.7)}$$

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



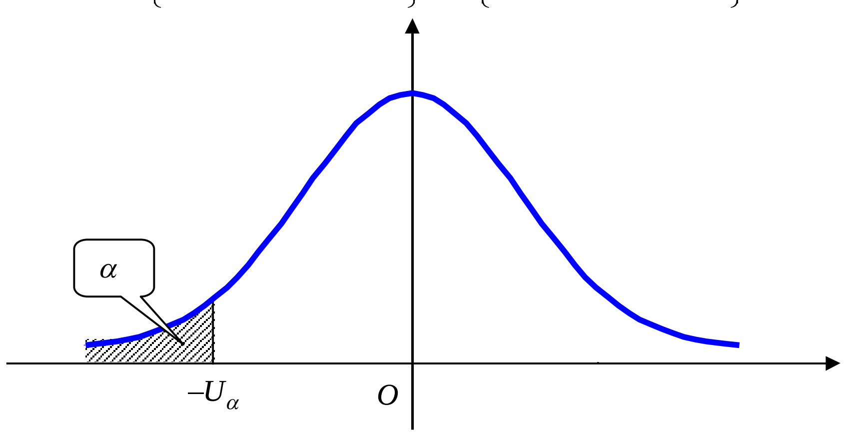
Hình 6.2: *Miền bác bỏ của kỳ vọng phân bố chuẩn, bài toán* 2

c. <u>Bài toán 3</u>: H$_0$: $\mu = \mu_0$; H$_1$: $\mu < \mu_0$. Đây là bài toán kiểm định một phía.

Miền bác bỏ

$$W_\alpha = \left\{ T = \frac{(\overline{X} - \mu_0)\sqrt{n}}{\sigma}\ ;\ -T > U_\alpha \right\} \quad \textbf{(6.8)}$$

$$P(W_\alpha) = P\left\{ -\frac{(\overline{X} - \mu_0)\sqrt{n}}{\sigma} > U_\alpha \right\} = P\left\{ \frac{(\overline{X} - \mu_0)\sqrt{n}}{\sigma} < -U_\alpha \right\} = \alpha$$



Hình 6.3: *Miền bác bỏ của kỳ vọng phân bố chuẩn, bài toán* 3

Từ giá trị của mẫu cụ thể $w=(x_1,x_2,...,x_n)$, tính giá trị quan sát của tiêu chuẩn kiểm định $T_{qs}=\dfrac{(\overline{x}-\mu_0)\sqrt{n}}{\sigma}$ và so sánh với miền bác bỏ W_α để kết luận.

6.2.1.2 Trường hợp chưa biết phương sai, kích thước mẫu $n \ge 30$

Trường hợp phương sai σ^2 chưa biết, ta thay độ lệch chuẩn σ bằng độ lệch chuẩn mẫu S (công thức 4.16). Khi kích thước n đủ lớn ($n \ge 30$) thống kê

$$T=\frac{(\overline{X}-\mu)\sqrt{n}}{S} \tag{6.9}$$

xấp xỉ phân bố chuẩn tắc $\mathbf{N}(0;1)$ (điều này cũng đúng ngay cả khi biến ngẫu nhiên gốc không có phân bố chuẩn, xem công thức 4.33 và tương tự mục 5.2.2.2 chương 5).

Khi giả thiết H_0 đúng thì thống kê T trở thành $T=\dfrac{(\overline{X}-\mu_0)\sqrt{n}}{S}$

Ta xây dựng các miền bác bỏ dựa vào đối thiết H_1 như sau:

a. Bài toán 1: H_0: $\mu=\mu_0$; H_1: $\mu\ne\mu_0$. Ta nói đây là bài toán kiểm định hai phía.

Miền bác bỏ

$$W_\alpha=\left\{T=\frac{(\overline{X}-\mu_0)\sqrt{n}}{S}\,;\ |T|>U_{\alpha/2}\right\}. \tag{6.10}$$

b. Bài toán 2: H_0: $\mu=\mu_0$; H_1: $\mu>\mu_0$. Đây là bài toán kiểm định một phía.

Miền bác bỏ

$$W_\alpha=\left\{T=\frac{(\overline{X}-\mu_0)\sqrt{n}}{S}\,;\ T>U_\alpha\right\}. \tag{6.11}$$

c. Bài toán 3: H_0: $\mu=\mu_0$; H_1: $\mu<\mu_0$. Đây là bài toán kiểm định một phía.

Miền bác bỏ

$$W_\alpha=\left\{T=\frac{(\overline{X}-\mu_0)\sqrt{n}}{S}\,;\ -T>U_\alpha\right\}. \tag{6.12}$$

Ví dụ 6.1: Một hãng buôn muốn biết xem phải chăng có sự không ổn định trung bình về lượng hàng bán được trung bình trên một nhân viên bán hàng so với các năm trước (lượng đó bằng 7,4). Một mẫu ngẫu nhiên gồm 40 nhân viên bán hàng được lựa chọn và tìm thấy lượng hàng trung bình của họ là $\overline{x}=6{,}1$ với độ lệch chuẩn là $s=2{,}5$. Với mức ý nghĩa $\alpha=1\%$ có thể nói rằng lượng hàng bán được trung bình trên mỗi đầu người có sự thay đổi không?

Giải: Gọi μ là lượng hàng bán được trung bình trên mỗi nhân viên bán hàng của hãng buôn.

Ta kiểm định: Giả thiết H_0: $\mu=7{,}4$; Đối thiết H_1: $\mu\ne 7{,}4$.

Tiêu chuẩn kiểm định: $T = \dfrac{(\overline{X}-\mu)\sqrt{n}}{S}$

Mức ý nghĩa: $\alpha = 0,01 \Rightarrow U_{\frac{\alpha}{2}} = 2,575$

Miền bác bỏ: $W_\alpha = \left\{ T = \dfrac{(\overline{X}-7,4)\sqrt{n}}{S} \;;\; |T| > 2,575 \right\}.$

Giá trị quan sát của tiêu chuẩn kiểm định:

$$T_{qs} = \frac{(6,1-7,4)\sqrt{40}}{2,5} = -3,289\,.$$

Với mẫu cụ thể này giá trị quan sát của tiêu chuẩn kiểm định rơi vào miền bác bỏ, vậy ta có thể kết luận rằng số lượng hàng bán được trung bình của mỗi nhân viên bán hàng là có thay đổi.

Ví dụ 6.2: Một công ti có một hệ thống máy tính có thể xử lí 1200 hóa đơn trong một giờ. Công ti mới nhập một hệ thống máy tính mới. Hệ thống này khi chạy kiểm tra trong 40 giờ cho thấy số hóa đơn được xử lí trung bình trong 1 giờ là 1260 với độ lệch chuẩn 215. Với mức ý nghĩa 5% hãy nhận định xem hệ thống mới có tốt hơn hệ thống cũ hay không?

Giải: Gọi μ là số hóa đơn trung bình mà hệ thống máy tính mới xử lí được trong 1 giờ.

Ta kiểm định: Giả thiết $H_0:\ \mu = 1200$; Đối thiết $H_1:\ \mu > 1200$.

Tiêu chuẩn kiểm định: $T = \dfrac{(\overline{X}-\mu)\sqrt{n}}{S}$

Mức ý nghĩa: $\alpha = 0,05 \Rightarrow U_\alpha = 1,64$

Miền bác bỏ: $W_\alpha = \left\{ T = \dfrac{(\overline{X}-1200)\sqrt{n}}{S} \;;\; T > 1,64 \right\}.$

Thay giá trị cụ thể của mẫu vào công thức (6.9) ta được

$$T_{qs} = \frac{(1260-1200)\sqrt{40}}{215} = 1,76\,.$$

Với mẫu cụ thể này giá trị quan sát của tiêu chuẩn kiểm định rơi vào miền bác bỏ, vậy ta có thể kết luận hệ thống máy tính mới tốt hơn hệ thống cũ.

6.2.1.3 Trường hợp chưa biết phương sai, kích thước mẫu $n < 30$

Xét thống kê

$$T = \frac{(\overline{X}-\mu)\sqrt{n}}{S} \qquad \textbf{(6.13)}$$

theo công thức (4.32) thống kê T có phân bố Student $n-1$ bậc tự do.

Khi giả thiết H_0 đúng thống kê T trở thành

$$T = \frac{(\overline{X}-\mu_0)\sqrt{n}}{S}$$

Miền bác bỏ phụ thuộc đối thiết H_1 xác định như sau:

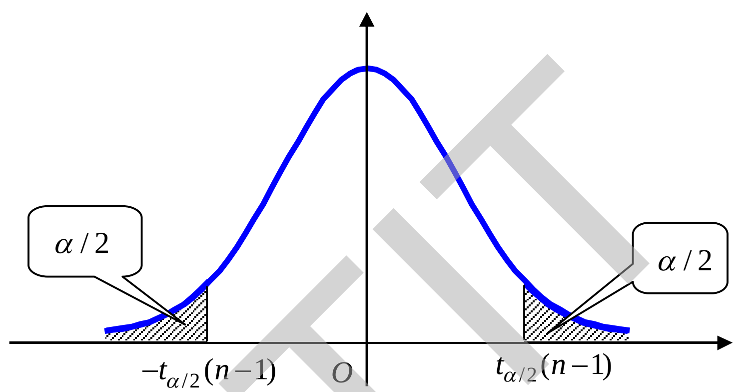
a. Bài toán 1: H_0: $\mu = \mu_0$; H_1: $\mu \neq \mu_0$. Bài toán kiểm định hai phía.

Miền bác bỏ:

$$W_\alpha = \left\{ T = \frac{(\overline{X} - \mu_0)\sqrt{n}}{S} \; ; \; |T| > t_{\alpha/2}(n-1) \right\}. \tag{6.14}$$

trong đó $t_{\alpha/2}(n-1)$ là giá trị tới hạn mức $\frac{\alpha}{2}$ và $t_\alpha(n-1)$ là giá trị tới hạn mức α của phân bố Student $n-1$ bậc tự do (xem phụ lục III).

$$T = \frac{(\overline{X} - \mu_0)\sqrt{n}}{\sigma} \sim \mathbf{T}(n-1) \Rightarrow P(W_\alpha) = P\left\{ \left| \frac{(\overline{X} - \mu_0)\sqrt{n}}{\sigma} \right| > t_{\alpha/2}(n-1) \right\} = \alpha$$



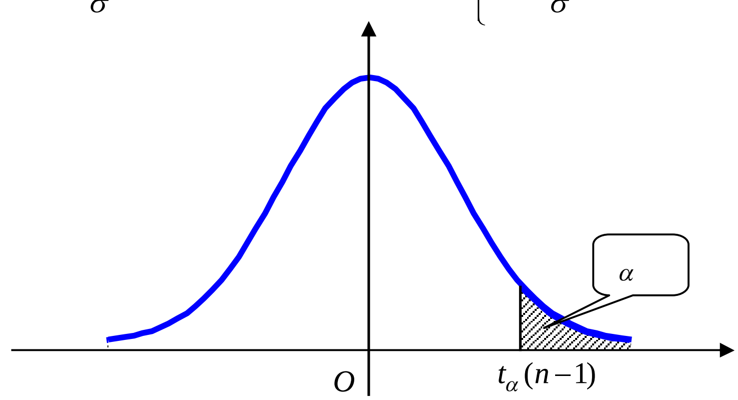
Hình 6.4: *Miền bác bỏ của kỳ vọng với* σ^2 *chưa biết* $n < 30$, *bài toán* 1

b. Bài toán 2: H_0: $\mu = \mu_0$; H_1: $\mu > \mu_0$.

Miền bác bỏ:

$$W_\alpha = \left\{ T = \frac{(\overline{X} - \mu_0)\sqrt{n}}{S} \; ; \; T > t_\alpha(n-1) \right\}. \tag{6.15}$$

$$T = \frac{(\overline{X} - \mu_0)\sqrt{n}}{\sigma} \sim \mathbf{T}(n-1) \Rightarrow P(W_\alpha) = P\left\{ \frac{(\overline{X} - \mu_0)\sqrt{n}}{\sigma} > t_\alpha(n-1) \right\} = \alpha$$



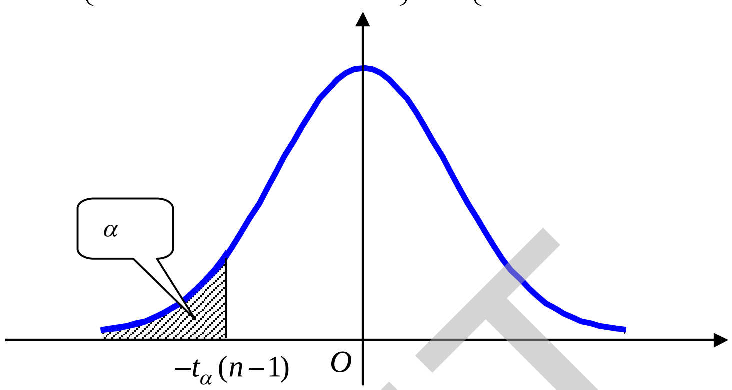
Hình 6.5: *Miền bác bỏ của kỳ vọng với* σ^2 *chưa biết* $n < 30$, *bài toán 2*

c. Bài toán 3: H_0: $\mu = \mu_0$; H_1: $\mu < \mu_0$.

Miền bác bỏ:

$$W_\alpha = \left\{ T = \frac{(\overline{X} - \mu_0)\sqrt{n}}{S} \; ; \; -T > t_\alpha(n-1) \right\}. \qquad \textbf{(6.16)}$$

$$P(W_\alpha) = P\left\{ -\frac{(\overline{X} - \mu_0)\sqrt{n}}{\sigma} > t_\alpha(n-1) \right\} = P\left\{ \frac{(\overline{X} - \mu_0)\sqrt{n}}{\sigma} < -t_\alpha(n-1) \right\} = \alpha$$



Hình 6.6: *Miền bác bỏ của kỳ vọng với* σ^2 *chưa biết* $n < 30$, *bài toán 3*

Ví dụ 6.3: Một công ty sản xuất hạt giống tuyên bố rằng một loại giống mới của họ có năng suất trung bình là 21,5 tạ/ha. Gieo thử giống mới này tại 16 vườn thí nghiệm và thu được kết quả:

19,2; 18,7; 22,4; 20,3; 16,8; 25,1; 17,0; 15,8; 21,0; 18,6; 23,7; 24,1; 23,4; 19,8; 21,7; 18,9.

Dựa vào kết quả này hãy xác nhận xem quảng cáo của công ty có đúng không. Mức ý nghĩa được lựa chọn là $\alpha = 0,05$. Biết rằng năng suất giống cây trồng là một biến ngẫu nhiên có phân bố chuẩn $\mathbf{N}(\mu;\sigma^2)$.

Giải: Gọi μ là năng suất trung bình của loại giống mới.

Ta cần kiểm định giả thiết H_0: $\mu = 21,5$; Đối thiết H_1: $\mu \neq 21,5$.

Tiêu chuẩn kiểm định: $T = \dfrac{(\overline{X} - \mu)\sqrt{n}}{S}$.

Tra bảng ta tính được giá trị tới hạn mức $\dfrac{\alpha}{2} = 0,025$ của phân bố Student 15 bậc tự do là $t_{0,025}(15) = 2,131$.

Do đó miền bác bỏ: $W_\alpha = \left\{ T = \dfrac{(\overline{X} - 21,5)\sqrt{n}}{S} \; ; \; |T| > 2,131 \right\}$.

Từ mẫu cụ thể trên tính được:

$$\overline{x} = 20,406\,, \; s = 3,038 \;\Rightarrow T_{qs} = \frac{(20,406 - 21,5)\sqrt{16}}{3,038} = -1,44\,.$$

Vì $\left|T_{qs}\right| = 1,44 < 2,131$ nên chưa có cơ sở để bác bỏ H$_0$. Nghĩa là với số liệu này thì có thể chấp nhận lời quảng cáo của công ty.

6.2.2 Kiểm định giả thiết về phương sai của biến ngẫu nhiên có phân bố chuẩn

Giả sử trong tổng thể biến ngẫu nhiên gốc có phân bố chuẩn $\mathbf{N}(\mu;\sigma^2)$, phương sai σ^2 chưa biết song có cơ sở để giả thiết rằng giá trị của nó bằng σ_0^2.

Ta kiểm định giả thiết H$_0$: $\quad \sigma^2 = \sigma_0^2$.

Từ tổng thể rút ra một mẫu ngẫu nhiên kích thước n: $W = (X_1, X_2, ..., X_n)$.

Xét thống kê

$$T = \frac{(n-1)S^2}{\sigma^2} \tag{6.17}$$

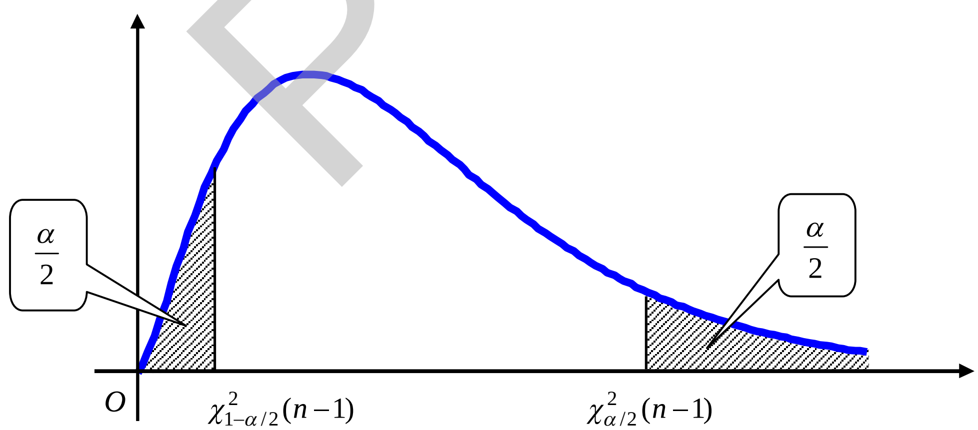
Theo công thức (3.31) thống kê T có phân bố "khi bình phương" $n-1$ bậc tự do

Khi giả thiết H$_0$ đúng thống kê T trở thành $T = \dfrac{(n-1)S^2}{\sigma_0^2}$.

Với mức ý nghĩa α, ta xây dựng các miền bác bỏ tùy thuộc vào đối thiết H$_1$ như sau:

a. <u>Bài toán 1</u>: H$_0$: $\sigma^2 = \sigma_0^2$; H$_1$: $\sigma^2 \neq \sigma_0^2$. (Bài toán kiểm định hai phía).

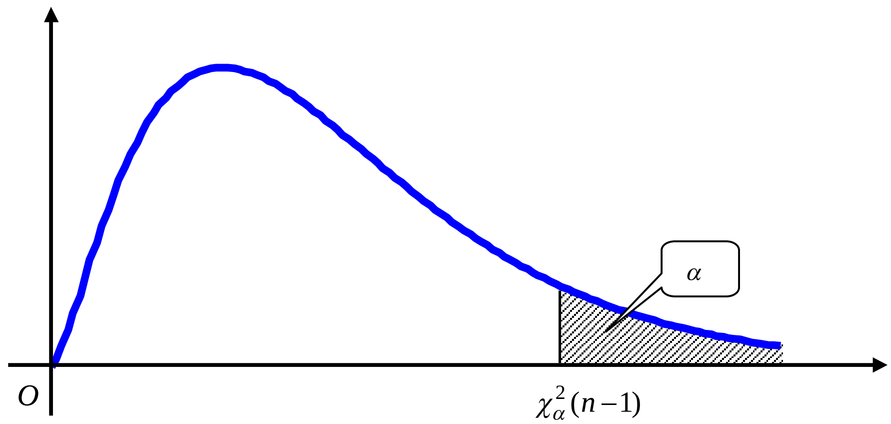
$$W_\alpha = \left\{ T = \frac{(n-1)S^2}{\sigma_0^2} \; ; \; T < \chi^2_{1-\frac{\alpha}{2}}(n-1) \text{ hoặc } T > \chi^2_{\frac{\alpha}{2}}(n-1) \right\} \tag{6.18}$$



Hình 6.7: *Miền bác bỏ bài toán 1 kiểm định phương sai*

b. <u>Bài toán 2</u>: H$_0$: $\sigma^2 = \sigma_0^2$; H$_1$: $\sigma^2 > \sigma_0^2$.

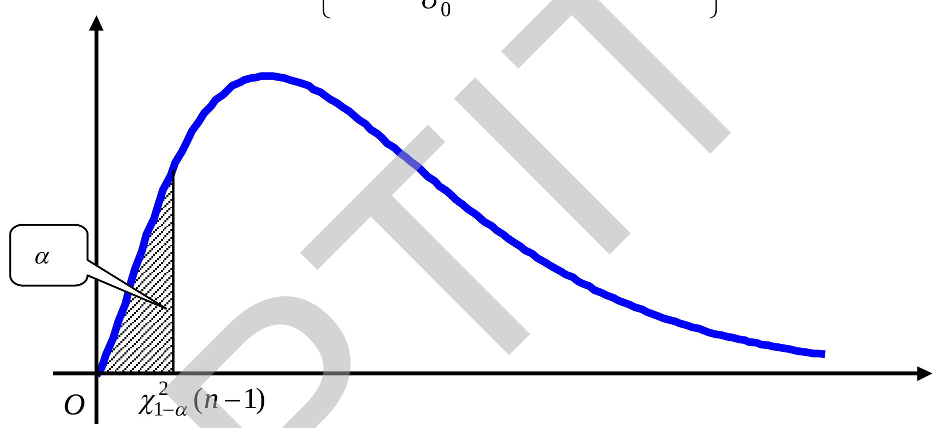
$$W_\alpha = \left\{ T = \frac{(n-1)S^2}{\sigma_0^2} \; ; \; T > \chi^2_\alpha(n-1) \right\} \tag{6.19}$$



Hình 6.8: *Miền bác bỏ bài toán 2 kiểm định phương sai*

c. Bài toán 3: H_0: $\sigma^2=\sigma_0^2$; H_1: $\sigma^2<\sigma_0^2$.

$$W_{\alpha}=\left\{T=\frac{(n-1)S^2}{\sigma_0^2}\ ;\ T<\chi_{1-\alpha}^{2}(n-1)\right\} \qquad \textbf{(6.20)}$$



Hình 6.9: *Miền bác bỏ bài toán 3 kiểm định phương sai*

Ví dụ 6.4: Để kiểm tra độ chính xác của một 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 về hoạt động của máy, biết rằng kích thước chi tiết do máy đó sản xuất ra là biến ngẫu nhiên có phân bố chuẩn có dung sai theo thiết kế là $\sigma^2=12$.

Giải: Gọi X là kích thước chi tiết, theo giả thiết X có phân bố chuẩn $\mathbf{N}(\mu;\sigma^2)$. Vậy đây là bài toán kiểm định giả thiết:

$$H_0:\ \sigma^2=12\text{; đối thiết } H_1:\ \sigma^2\neq 12.$$

Với mức ý nghĩa $\alpha=0{,}01$, tra bảng khi bình phương 14 bậc tự do ta được:

$$\chi_{\frac{\alpha}{2}}^{2}(n-1)=\chi_{0,005}^{2}(14)=3{,}57\ ;\ \ \chi_{1-\frac{\alpha}{2}}^{2}(n-1)=\chi_{0,995}^{2}(14)=28{,}3.$$

Thay giá trị mẫu cụ thể vào công thức (6.17) ta tính được giá trị quan sát của tiêu chuẩn kiểm định:

$$T_{qs} = \frac{14 \cdot 14{,}6}{12} = 17{,}03 .$$

Giá trị quan sát của tiêu chuẩn kiểm định không rơi vào miền bác bỏ, vậy chưa có cơ sở để bác bỏ H_0. Nói cách khác có thể coi máy móc vẫn làm việc bình thường.

6.2.3 Kiểm định giả thiết về tần suất p của tổng thể

Ta để ý đến dấu hiệu định tính là đặc trưng A nào đó mà mỗi cá thể của tổng thể có thể có tính chất này hoặc không. Gọi p là tần suất có đặc trưng A của tổng thể. Như đã thấy trong chương 4, có thể xem dấu hiệu nghiên cứu này là một biến ngẫu nhiên X có phân bố Bernoulli với tham số p. Ta cần kiểm định giả thiết cho rằng giá trị p bằng p_0.

Giả thiết không $\qquad$ H_0: $p = p_0$.

Từ tổng thể rút ra một mẫu ngẫu nhiên kích thước n, gọi f là tần suất mẫu (công thức (4.17)-(4.18)). Xét thống kê

$$T = \frac{(f - p)\sqrt{n}}{\sqrt{p(1-p)}} \qquad \textbf{(6.21)}$$

Nếu giả thiết H_0 đúng thống kê trên trở thành $T = \dfrac{(f - p_0)\sqrt{n}}{\sqrt{p_0(1-p_0)}}$.

Áp dụng định lý giới hạn trung tâm (công thức (3.45)) ta có thể xem phân bố của thống kê T xấp xỉ phân bố chuẩn tắc $\mathbf{N}(0;1)$ khi n đủ lớn.

Trong thực tế khi khi n thỏa mãn điều kiện sau ta xem là đủ lớn

$$\begin{cases} np_0 > 5 \\ n(1-p_0) > 5 \end{cases} \qquad \textbf{(6.22)}$$

Với mức ý nghĩa α cho trước, các miền bác bỏ của phép kiểm định phụ thuộc đối thiết H_1 xác định như sau.

a. <u>Bài toán 1</u>: H_0: $p = p_0$; H_1: $p \neq p_0$.

$$W_\alpha = \left\{ T = \frac{(f - p_0)\sqrt{n}}{\sqrt{p_0(1-p_0)}} \; ; \; |T| > U_{\frac{\alpha}{2}} \right\} \qquad \textbf{(6.23)}$$

b. <u>Bài toán 2</u>: H_0: $p = p_0$; H_1: $p > p_0$.

$$W_\alpha = \left\{ T = \frac{(f - p_0)\sqrt{n}}{\sqrt{p_0(1-p_0)}} \; ; \; T > U_\alpha \right\} \qquad \textbf{(6.24)}$$

c. <u>Bài toán 3</u>: H_0: $p = p_0$; H_1: $p < p_0$.

$$W_\alpha = \left\{ T = \frac{(f - p_0)\sqrt{n}}{\sqrt{p_0(1-p_0)}} \; ; \; -T > U_\alpha \right\} \qquad \textbf{(6.25)}$$

Với mẫu cụ thể tính giá trị quan sát của tiêu chuẩn kiểm định T_{qs}, so sánh với W_α và kết luận.

Ví dụ 6.5: Một đảng chính trị trong một cuộc bầu cử tổng thống ở nước nọ tuyên bố rằng có ít nhất 45% cử tri sẽ bỏ phiếu cho ứng cử viên A của đảng họ.

Chọn ngẫu nhiên 2000 cử tri để thăm dò ý kiến và cho thấy có 924 cử tri tuyên bố sẽ bỏ phiếu cho A.

Với mức $\alpha = 2,5\%$, hãy kiểm định xem dự đoán của đảng trên có đúng không.

Giải: Gọi p là tỉ lệ cử tri sẽ bỏ phiếu cho ứng cử viên A.

Ta cần kiểm định: Giả thiết H_0: $p = 0,45$; Đối thiết H_1: $p \geq 0,45$.

Điều kiện $\begin{cases} np_0 = 2000 \cdot 0,45 = 900 > 5 \\ n(1-p_0) = 2000 \cdot 0,55 = 1100 > 5 \end{cases}$ thỏa mãn, do đó tiêu chuẩn kiểm định có thể xấp xỉ phân bố chuẩn tắc.

Miền bác bỏ

$$\alpha = 0,025 \Rightarrow U_\alpha = 1,96 \Rightarrow W_\alpha = \{\, T > 1,96 \}$$

Từ mẫu cụ thể ta có tần suất mẫu $f = \dfrac{924}{2000} = 0,462$,

do đó giá trị quan sát của tiêu chuẩn kiểm định

$$T_{qs} = \frac{(0,462 - 0,45)\sqrt{2000}}{\sqrt{0,45 \cdot 0,55}} = 1,079 .$$

Ta thấy $T_{qs} < 1,96$, giá trị quan sát của tiêu chuẩn kiểm định không rơi vào miền bác bỏ. Vậy với mẫu trên ta chưa có cơ sở để bác bỏ H_0.

6.2.4 Kiểm định giả thiết về hai kỳ vọng của hai biến ngẫu nhiên có phân bố chuẩn

Giả sử ta xét cùng một lúc hai dấu hiệu nghiên cứu, dấu hiệu nghiên cứu thứ nhất được xem như biến ngẫu nhiên X_1 có phân bố chuẩn $\mathbf{N}(\mu_1;\sigma_1^2)$, dấu hiệu nghiên cứu thứ hai được xem như biến ngẫu nhiên X_2 có phân bố chuẩn $\mathbf{N}(\mu_2;\sigma_2^2)$. Nếu μ_1 và μ_2 chưa biết song có cơ sở để giả thiết rằng giá trị của chúng bằng nhau, nghĩa là giả thiết không có dạng:

$$H_0:\ \mu_1 = \mu_2$$

Ta kiểm định giả thiết trên theo các trường hợp sau:

6.2.4.1 Trường hợp hai phương sai σ_1^2 và σ_2^2 của hai biến ngẫu nhiên gốc đã biết

Lấy hai mẫu ngẫu nhiên độc lập kích thước tương ứng là n_1, n_2 của dấu hiệu nghiên cứu X_1 và X_2.

$$W_1 = \left(X_{11}, X_{12}, \ldots, X_{1n_1}\right) ;\ W_2 = \left(X_{21}, X_{22}, \ldots, X_{2n_2}\right) .$$

Chọn thống kê sau làm tiêu chuẩn kiểm định

$$T=\frac{\left(\overline{X_1}-\overline{X_2}\right)-\left(\mu_1-\mu_2\right)}{\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}} \qquad \textbf{(6.26)}$$

Theo công thức (4.34) chương 4, thống kê T có phân bố chuẩn tắc $\mathbf{N}(0;1)$. Nếu giả thiết H_0 đúng thì thống kê T có dạng

$$T=\frac{\overline{X_1}-\overline{X_2}}{\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}} \sim \mathbf{N}(0;1) \qquad \textbf{(6.27)}$$

Vì vậy với mức ý nghĩa α cho trước và tùy thuộc vào dạng của đối thiết H_1, tương tự trường hợp ở mục 6.2.1.1 và 6.2.1.2 ta có thể xây dựng các miền bác bỏ tương ứng như sau:

a. H_0: $\mu_1=\mu_2$; H_1: $\mu_1\neq\mu_2$ (kiểm định hai phía).

$$W_\alpha=\left\{T=\frac{\overline{X_1}-\overline{X_2}}{\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}}\ ;\ |T|>U_{\frac{\alpha}{2}}\right\} \qquad \textbf{(6.28)}$$

b. H_0: $\mu_1=\mu_2$; H_1: $\mu_1>\mu_2$ (kiểm định một phía).

$$W_\alpha=\left\{T=\frac{\overline{X_1}-\overline{X_2}}{\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}}\ ;\ T>U_\alpha\right\} \qquad \textbf{(6.29)}$$

c. H_0: $\mu_1=\mu_2$; H_1: $\mu_1<\mu_2$ (kiểm định một phía).

$$W_\alpha=\left\{T=\frac{\overline{X_1}-\overline{X_2}}{\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}}\ ;\ -T>U_\alpha\right\} \qquad \textbf{(6.30)}$$

Lập hai mẫu cụ thể từ X_1 và X_2 và tính giá trị quan sát của tiêu chuẩn kiểm định T_{qs}, so sánh với miền bác bỏ W_α và kết luận.

Ví dụ 6.6: Tại một xí nghiệp người ta xây dựng hai phương án gia công cùng 1 loại chi tiết. Để đánh giá xem chi phí trung bình về nguyên liệu theo hai phương án ấy có khác nhau hay không người ta tiến hành sản xuất thử và thu được các kết quả sau:

Phương án 1:	2,5	3,2	3,5	3,8	3,5	
Phương án 2:	2,0	2,7	2,5	2,9	2,3	2,6

Với mức ý nghĩa $\alpha=0,05$, hãy kết luận về vấn đề trên biết rằng chi phí nguyên liệu theo cả hai phương án gia công đều là các biến ngẫu nhiên có phân bố chuẩn với phương sai

$\sigma_1^2 = \sigma_2^2 = 0,16$.

Giải: Gọi X_1 và X_2 tương ứng là chi phí nguyên liệu theo hai phương án gia công trên. Theo giả thiết X_1 và X_2 có phân bố chuẩn. Vậy chi phí trung bình theo các phương án đó là μ_1 và μ_2. Đây là bài toán kiểm định: Giả thiết H$_0$: $\mu_1 = \mu_2$; Đối thiết H$_1$: $\mu_1 \neq \mu_2$.

Giả sử H$_0$ đung, theo công thức (6.27) tiêu chuẩn kiểm định trở thành

$$T = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{\frac{0,16}{5} + \frac{0,16}{6}}}.$$

Do $\alpha = 0,05 \Rightarrow U_{\alpha/2} = 1,96$.

Vậy miền bác bỏ $W_\alpha = \{|T| > 1,96\}$.

Từ mẫu cụ thể ta tính được: $\overline{x_1} = 3,3$; $\overline{x_2} = 2,5$,

thay vào ta có giá trị quan sát của tiêu chuẩn kiểm định

$$T_{qs} = \frac{3,3 - 2,5}{\sqrt{\frac{0,16}{5} + \frac{0,16}{6}}} = 3,33 > 1,96.$$

Vậy giá trị quan sát của tiêu chuẩn kiểm định rơi vào miền bác bỏ. Bác bỏ H$_0$ thừa nhận H$_1$, tức là chi phí nguyên liệu theo hai phương án gia công trên thực sự khác nhau.

6.2.4.2 Trường hợp hai phương sai σ_1^2 và σ_2^2 của hai biến ngẫu nhiên gốc chưa biết nhưng biết rằng chúng bằng nhau

Lấy hai mẫu ngẫu nhiên độc lập kích thước tương ứng là n_1, n_2 của dấu hiệu nghiên cứu X_1 và X_2.

$$W_1 = (X_{11}, X_{12}, ..., X_{1n_1}); \quad W_2 = (X_{21}, X_{22}, ..., X_{2n_2}).$$

Chọn thống kê sau làm tiêu chuẩn kiểm định

$$T = \frac{(\overline{X_1} - \overline{X_2}) - (\mu_1 - \mu_2)}{Sp\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad \textbf{(6.31)}$$

$$Sp = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}$$

Theo công thức (4.37) chương 4, thống kê T có phân bố Student với $n_1 + n_2 - 2$ bậc tự do. Nếu giả thiết H$_0$ đúng thì thống kê T trở thành

$$T = \frac{(\overline{X_1} - \overline{X_2})}{Sp\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad \textbf{(6.32)}$$

Vì vậy với mức ý nghĩa α , ta xây dựng các miền bác bỏ tương ứng:

a. H₀: $\mu_1 = \mu_2$; H₁: $\mu_1 \neq \mu_2$ (kiểm định hai phía).

$$W_\alpha = \left\{ T = \frac{\left(\overline{X_1} - \overline{X_2}\right)}{Sp\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} \; ; \; |T| > t_{\alpha/2}(n_1 + n_2 - 2) \right\} \qquad \textbf{(6.33)}$$

b. H₀: $\mu_1 = \mu_2$; H₁: $\mu_1 > \mu_2$ (kiểm định một phía).

$$W_\alpha = \left\{ T = \frac{\left(\overline{X_1} - \overline{X_2}\right)}{Sp\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} \; ; \; T > t_{\alpha}(n_1 + n_2 - 2) \right\} \qquad \textbf{(6.34)}$$

c. H₀: $\mu_1 = \mu_2$; H₁: $\mu_1 < \mu_2$ (kiểm định một phía).

$$W_\alpha = \left\{ T = \frac{\left(\overline{X_1} - \overline{X_2}\right)}{Sp\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} \; ; \; -T > t_{\alpha}(n_1 + n_2 - 2) \right\} \qquad \textbf{(6.35)}$$

Lập hai mẫu cụ thể từ X_1 và X_2, tính giá trị quan sát của tiêu chuẩn kiểm định T_{qs}, so sánh với miền bác bỏ W_α và kết luận.

Ví dụ 6.7: Một nghiên cứu được thực hiện đối với 20 người của thành phố A và 19 người thành phố B xem thu nhập hàng năm (tính bằng triệu đồng) của dân cư hai thành phố đó có thực sự khác nhau hay không. Các số liệu mẫu thu được như sau:

$$n_1 = 20 \; ; \; \overline{x_1} = 18{,}27 \; ; \; s_1^2 = 8{,}74 \, .$$

$$n_2 = 19 \; ; \; \overline{x_2} = 16{,}78 \; ; \; s_2^2 = 6{,}53 \, .$$

Vậy với mức ý nghĩa $\alpha = 0{,}05$ có thể cho rằng thu nhập bình quân của dân cư ở hai thành phố này là khác nhau hay không. Giả sử thu nhập hàng năm của dân cư hai thành phố này cùng phân bố chuẩn với phương sai như nhau.

Giải: Gọi X_1 và X_2 tương ứng là thu nhập hàng năm của dân cư hai thành phố đó. Theo giả thiết X_1 và X_2 có phân bố chuẩn với các phương sai bằng nhau. Do đó để kiểm định giả thiết

$$\mathrm{H}_0: \; \mu_1 = \mu_2 \, ; \quad \mathrm{H}_1: \; \mu_1 \neq \mu_2 \, .$$

Giả sử giả thiết H₀ đúng, ta sử dụng tiêu chuẩn kiểm định (công thức 6.32) và miền bác bỏ W_α (công thức 6.33).

$$\alpha = 0{,}05 \Rightarrow t_{\alpha/2}(n_1 + n_2 - 2) = t_{0,025}(37) = 2{,}021 \;\; \Rightarrow W_\alpha = \left\{ \, |T| > 2{,}021 \right\}.$$

Từ mẫu cụ thể ta tính được

$$Sp = \sqrt{\frac{19 \cdot 8,74 + 18 \cdot 6,53}{20 + 19 - 2}} = 2,773.$$

Do đó giá trị quan sát của tiêu chuẩn kiểm định

$$T_{qs} = \frac{18,27 - 16,78}{2,773\sqrt{\frac{1}{20} + \frac{1}{19}}} = 0,172.$$

Vì giá trị quan sát T_{qs} không rơi vào miền bác bỏ W_α, do đó với hai mẫu cụ thể này thì chưa có cơ sở để bác bỏ H_0, tức là vẫn chấp nhận giả thiết cho rằng thu nhập trung bình hàng năm của dân hai thành phố đó là như nhau.

6.2.4.3 Trường hợp hai phương sai σ_1^2 và σ_2^2 của hai biến ngẫu nhiên gốc chưa biết và không thể cho rằng chúng bằng nhau ($\sigma_1^2 \neq \sigma_2^2$)

Chọn thống kê sau làm tiêu chuẩn kiểm định

$$T = \frac{\left(\overline{X_1} - \overline{X_2}\right) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \qquad \textbf{(6.36)}$$

Theo công thức (4.38) chương 4, thống kê T có phân bố Student k bậc tự do, trong đó:

$$k \approx \frac{(n_1 - 1)(n_2 - 1)}{(n_2 - 1)C^2 + (n_1 - 1)(1 - C)^2} \; ; \quad C = \frac{\frac{S_1^2}{n_1}}{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}.$$

Nếu giả thiết H_0 đúng thì thống kê T có dạng

$$T = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \qquad \textbf{(6.37)}$$

Với mức ý nghĩa α, miền bác bỏ phụ thuộc đối thiết và được xác định như sau:

a. H_0: $\mu_1 = \mu_2$; H_1: $\mu_1 \neq \mu_2$ (kiểm định hai phía).

$$W_\alpha = \left\{ T = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \; ; \; |T| > t_{\alpha/2}(k) \right\} \qquad \textbf{(6.38)}$$

b. H_0: $\mu_1 = \mu_2$; H_1: $\mu_1 > \mu_2$ (kiểm định một phía).

$$W_\alpha = \left\{ T = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}} \; ; \; T > t_\alpha(k) \right\} \qquad \textbf{(6.39)}$$

c. H0: $\mu_1 = \mu_2$; H1: $\mu_1 < \mu_2$ (kiểm định một phía).

$$W_\alpha = \left\{ T = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}} \; ; \; -T > t_\alpha(k) \right\} \qquad \textbf{(6.40)}$$

Lập hai mẫu cụ thể từ X_1 và X_2, tính giá trị quan sát của tiêu chuẩn kiểm định T_{qs}, so sánh với miền bác bỏ W_α và kết luận.

Ví dụ 6.8: Để kiểm nghiệm hiệu quả của một sáng kiến mới nhằm tăng năng suất sản xuất, người ta chọn ngẫu nhiên 7 ngày làm việc của từng nhóm: Nhóm I có áp dụng phương pháp mới và Nhóm II không áp dụng phương pháp mới. Kết quả có được về năng suất của từng nhóm như sau:

Nhóm I:	40	54	26	63	21	37	39
Nhóm II:	18	43	28	50	16	32	13

Với mức ý nghĩa $\alpha = 0,05$ hãy kết luận xem phương pháp mới nói trên có thực sự hiệu quả hay không. Giả thiết rằng năng suất của mỗi loại đều có phân bố chuẩn.

Giải: Gọi X_1 và X_2 tương ứng là số sản phẩm sản xuất được trong mỗi ngày thuộc hai nhóm trên. Theo giả thiết X_1 và X_2 có phân bố chuẩn với các phương sai σ_1^2, σ_2^2 chưa biết và không thể cho rằng chúng bằng nhau. Số sản phẩm trung bình sản xuất được tương ứng là μ_1 và μ_2. Ta cần kiểm định giả thiết

$$\text{H}_0: \; \mu_1 = \mu_2; \qquad \text{H}_1: \; \mu_1 > \mu_2 .$$

Giả sử giả thiết H0 đúng, ta sử dụng tiêu chuẩn kiểm định (6.37) và miền bác bỏ W_α (6.39).

Từ hai mẫu cụ thể ta tính được:

$$n_1 = 7 \; ; \; \overline{x_1} = 40,00 \; ; \; s_1^2 = 215,33.$$

$$n_2 = 7 \; ; \; \overline{x_2} = 28,57 \; ; \; s_2^2 = 198,62 .$$

Ta có $$C = \frac{\dfrac{s_1^2}{n_1}}{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}} = 0,520 \;\Rightarrow\; k = \frac{6 \cdot 6}{6 \cdot 0,4798^2 + 6 \cdot 0,5202^2} \approx 12 .$$

Miền bác bỏ

$$\alpha = 0,05 \Rightarrow t_{\alpha}(k) = t_{0,05}(12) = 1,782 \Rightarrow W_{\alpha} = \{ T > 1,782 \}.$$

Giá trị quan sát của tiêu chuẩn kiểm định

$$T_{qs} = \frac{40 - 28,52}{\sqrt{\frac{198,62}{7} + \frac{215,33}{19}}} = 1,49.$$

Vì giá trị quan sát T_{qs} không rơi vào miền bác bỏ W_{α}, do đó với hai mẫu cụ thể này chưa có cơ sở để bác bỏ H_0, tức là chưa thể nói phương pháp mới có hiệu quả hơn.

Nhận xét 6.1: Theo định lý giới hạn trung tâm, khi $n_1 \ge 30$ phân bố của $\overline{X_1}$ xấp xỉ phân bố chuẩn $\mathbf{N}\left(\mu_1; \frac{S_1^2}{n_1}\right)$ và $n_2 \ge 30$ thì phân bố của $\overline{X_2}$ xấp xỉ phân bố chuẩn $\mathbf{N}\left(\mu_2; \frac{S_2^2}{n_2}\right)$. Do đó nếu giả thiết H_0 đúng thì $T \approx \mathbf{N}(0;1)$.

6.2.5 Kiểm định giả thiết về sự bằng nhau của hai tần suất tương ứng với hai tổng thể

Giả sử ta xét cùng một lúc hai tổng thể. Ở tổng thể thứ nhất tần suất xuất hiện của dấu hiệu A là p_1, ở tổng thể thứ hai tần suất xuất hiện của dấu hiệu A là p_2. Bằng cách mã hóa ta có thể xem dấu hiệu nghiên cứu của tổng thể thứ nhất là biến ngẫu nhiên X_1 có phân bố Bernoulli tham số p_1 và dấu hiệu nghiên cứu của tổng thể thứ hai là biến ngẫu nhiên X_2 có phân bố Bernoulli tham số p_2. Nếu hai tham số p_1 và p_2 chưa biết song có cơ sở để giả thiết rằng giá trị của chúng bằng nhau, vậy giả thiết thống kê là:

$$H_0: \quad p_1 = p_2$$

Để kiểm định giả thiết H_0 chúng ta rút ra hai mẫu ngẫu nhiên độc lập kích thước tương ứng là n_1 và n_2 lần lượt từ hai tổng thể nói trên:

$$W_1 = (X_{11}, X_{12}, ..., X_{1n_1}); \quad W_2 = (X_{21}, X_{22}, ..., X_{2n_2}).$$

Chọn thống kê sau làm tiêu chuẩn kiểm định

$$T = \frac{(f_1 - f_2) - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}}. \quad \textbf{(6.41)}$$

Theo công thức (4.42) chương 4, thống kê T có phân bố xấp xỉ chuẩn tắc $\mathbf{N}(0;1)$ khi $n_1 > 30$ và $n_2 > 30$.

Nếu giả thiết H_0 đúng ($p_1 = p_2 = p$) thì tiêu chuẩn kiểm định trở thành

$$T = \frac{f_1 - f_2}{\sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}.$$

Thông thường p chưa biết nên được thay bằng ước lượng (trung bình cộng) của nó là

$$\overline{f} = \frac{n_1 f_1 + n_2 f_2}{n_1 + n_2}.$$

Như vậy ta có tiêu chuẩn kiểm định

$$T = \frac{f_1 - f_2}{\sqrt{\overline{f}(1-\overline{f})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}. \quad \textbf{(6.42)}$$

Thống kê T có phân bố xấp xỉ chuẩn tắc $\mathbf{N}(0;1)$ nếu $n_1 > 30$ và $n_2 > 30$. Do đó với mức ý nghĩa α cho trước, miền bác bỏ của phép kiểm định phụ thuộc đối thiết H_1 có dạng sau.

a. H_0: $p_1 = p_2$; H_1: $p_1 \neq p_2$ (kiểm định hai phía).

$$W_\alpha = \left\{ T = \frac{f_1 - f_2}{\sqrt{\overline{f}(1-\overline{f})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \;;\; |T| > U_{\frac{\alpha}{2}} \right\} \quad \textbf{(6.43)}$$

b. H_0: $p_1 = p_2$; H_1: $p_1 > p_2$ (kiểm định một phía).

$$W_\alpha = \left\{ T = \frac{f_1 - f_2}{\sqrt{\overline{f}(1-\overline{f})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \;;\; T > U_\alpha \right\} \quad \textbf{(6.44)}$$

c. H_0: $p_1 = p_2$; H_1: $p_1 < p_2$ (kiểm định một phía).

$$W_\alpha = \left\{ T = \frac{f_1 - f_2}{\sqrt{\overline{f}(1-\overline{f})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \;;\; -T > U_\alpha \right\} \quad \textbf{(6.45)}$$

Lập hai mẫu cụ thể từ X_1 và X_2, tính giá trị quan sát của tiêu chuẩn kiểm định T_{qs}, so sánh với miền bác bỏ W_α và kết luận.

Ví dụ 6.9: Kiểm tra ngẫu nhiên các sản phẩm cùng loại do hai nhà máy sản xuất thu được các số liệu sau:

Nhà máy	Số sản phẩm được kiểm tra	Số phế phẩm
A	$n_1 = 1100$	$x_1 = 22$
B	$n_2 = 900$	$x_2 = 31$

Với mức ý nghĩa $\alpha = 0,05$ có thể coi tỷ lệ phế phẩm của hai nhà máy là như nhau hay không?

Giải: Gọi p_1 và p_2 tương ứng là tỷ lệ phế phẩm của hai nhà máy A và B. Như vậy đây là bài toán kiểm định cặp giả thiết, đối thiết:

$$H_0:\ p_1 = p_2;\ H_1:\ p_1 \neq p_2.$$

Giả sử giả thiết H_0 đúng chọn tiêu chuẩn kiểm định (6.42) và khi n_1 và n_2 đủ lớn thì miền bác bỏ dạng (6.43).

Miền bác bỏ:

$$\alpha = 0,05 \Rightarrow U_{\alpha/2} = 1,96 \Rightarrow W_\alpha = \{|T| > 1,96\}.$$

Với hai mẫu cụ thể trên ta tính được

$$f_1 = \frac{22}{1100} = 0,02;\quad f_2 = \frac{31}{900} = 0,0344;\quad \overline{f} = \frac{22+31}{1100+900} = 0,0265.$$

Giá trị quan sát của tiêu chuẩn kiểm định:

$$T_{qs} = \frac{0,02 - 0,0344}{\sqrt{0,0265 \cdot 0,9735\left(\frac{1}{1100} + \frac{1}{900}\right)}} = -1,995.$$

Vì giá trị quan sát T_{qs} rơi vào miền bác bỏ W_α, do đó ta có thể bác bỏ giả thiết H_0. Vậy có thể cho rằng tỷ lệ phế phẩm của hai nhà máy khác nhau.

CÂU HỎI ÔN TẬP VÀ BÀI TẬP CHƯƠNG 6

6.1 Giả thiết thống kê là giả thiết do nhà thống kê đặt ra cho mẫu ngẫu nhiên.

Đúng ☐ Sai ☐.

6.2 Bác bỏ giả thiết dẫn đến chấp nhận đối thiết và ngược lại do đó đối thiết luôn luôn là phủ định của giả thiết.

Đúng ☐ Sai ☐.

6.3 Qui tắc kiểm định dựa trên nguyên lý xác suất nhỏ và phép chứng minh phản chứng.

Đúng ☐ Sai ☐.

6.4 Sai lầm loại 1 là sai lầm gặp phải khi thực tế giả thiết đúng nhưng ta bác bỏ.

Đúng ☐ Sai ☐.

6.5 Sai lầm loại 2 luôn luôn lớn hơn sai lầm loại 1.

Đúng ☐ Sai ☐.

6.6 Định lý Neymann - Pearson chỉ ra rằng có thể tìm được tiêu chuẩn kiểm định có sai lầm loại 1 và loại 2 bé tùy ý cho trước.

Đúng ☐ Sai ☐.

6.7 Miền bác bỏ là miền có xác suất rất bé nên ta có thể bỏ qua trong mọi phép kiểm định.

Đúng ☐ Sai ☐.

6.8 Miền bác bỏ của tiêu chuẩn kiểm định giả thiết hai phía về giá trị trung bình của biến ngẫu nhiên có phân bố chuẩn luôn luôn là miền đối xứng.

☐ ☐

Đúng Sai .

6.9 Khi xây dựng tiêu chuẩn kiểm định T ta luôn giả sử rằng giả thiết H_0 sai vì giả thiết H_0 là điều ta nghi ngờ muốn bác bỏ.

Đúng ☐ Sai ☐.

6.10 Kiểm định hai phía là kiểm định đối với những tham số có thể nhận giá trị âm dương bất kỳ, còn kiểm định một phía khi tham số cần kiểm định chỉ nhận giá trị dương hoặc âm.

Đúng ☐ Sai ☐.

6.11 Trọng lượng đóng bao của một loại sản phẩm X là biến ngẫu nhiên có phân bố theo quy luật chuẩn với trọng lượng trung bình theo quy định là 100kg. Nghi ngờ sản phẩm bị đóng thiếu, người ta cân thử 29 bao loại này ta thu được kết quả:

Trọng lượng (kg)	98,0-98,5	98,5-99,0	99,0-99,5	99,5-100	100-100,5	100,5-101
Số bao tương ứng	2	6	10	7	3	1

Với mức ý nghĩa $\alpha = 0,025$ hãy kết luận về điều nghi ngờ nói trên.

6.12 Định mức thời gian hoàn thành sản phẩm là 14 phút. Liệu có cần thay đổi định mức không, nếu theo dõi thời gian hoàn thành sản phẩm ở 250 công nhân ta thu được kết quả như sau:

Thời gian hoàn thành sản phẩm (phút)	10 - 12	12 - 14	14 - 16	16 - 18	18 - 20
Số công nhân tương ứng	20	60	100	40	30

Với mức ý nghĩa $\alpha = 0,05$ hãy kết luận về ý định nói trên.

6.13 Mức hao phí xăng của một loại ôtô chạy từ A đến B là một biến ngẫu nhiên có phân bố chuẩn với kỳ vọng 50 lít. Đoạn đường được sửa chữa lại. Người ta cho rằng mức hao phí xăng trung bình giảm xuống. Quan sát 28 ô tô cùng loại thu được

Mức hao phí xăng (lít)	48,5 - 49,0	49,0 - 49,5	49,5 - 50,0	50,0 - 50,5	50,5-51
Số ôtô tương ứng	4	10	9	3	2

Với mức ý nghĩa $\alpha = 0,025$ hãy kết luận về điều nghi ngờ nói trên.

6.14 Trọng lượng X của sản phẩm do một nhà máy sản xuất ra là một biến ngẫu nhiên có phân bố chuẩn $\mathbf{N}(\mu, \sigma^2)$ với độ lệch $\sigma = 2kg$ và trọng lượng trung bình $\mu = 20kg$. Nghi ngờ trọng lượng đóng gói thay đổi, người ta cân thử 100 sản phẩm và thu được kết quả sau:

Trọng lượng sản phẩm (kg)	18	19	20	21	22
Số sản phẩm	5	25	40	20	10

Hãy kiểm định điều nghi ngờ trên với mức ý nghĩa $\alpha = 0,05$.

6.15 Một công ty có một hệ thống máy tính có thể xử lý 1300 hoá đơn trong 1 giờ. Công ty mới nhập một hệ thống máy tính mới, hệ thống này chạy kiểm tra trong 40 giờ cho thấy số hoá đơn xử lý trung bình trong 1 giờ là 1378 với độ lệch tiêu chuẩn 215. Với mức ý nghĩa 2,5%

hãy nhận định xem hệ thống mới có tốt hơn hệ thống cũ hay không?

6.16 Tỉ lệ phế phẩm do một máy tự động sản xuất là 5%. Kiểm tra ngẫu nhiên 300 sản phẩm thấy có 24 sản phẩm là phế phẩm. Từ đó có ý kiến cho rằng tỷ lệ phế phẩm do máy đó sản xuất có chiều hướng tăng lên. Hãy kết luận ý kiến trên với mức ý nghĩa $\alpha = 0,05$.

6.17 Nếu máy móc hoạt động bình thường thì trọng lượng sản phẩm là một biến ngẫu nhiên có phân bố chuẩn với độ lệch tiêu chuẩn 1 kg. Có thể coi máy móc còn hoạt động tốt hay không nếu cân thử 30 sản phẩm ta thấy độ lệch tiêu chuẩn mẫu tăng lên tới 1,1 kg. Hãy kiểm định giả thiết trên với mức ý nghĩa $\alpha = 0,01$.

6.18 Có hai loại bi bằng thép, loại I và loại II. Loại I có đường kính là biến ngẫu nhiên X có phân bố chuẩn $\mathbf{N}(\mu_1;\sigma^2)$, Loại II có đường kính là biến ngẫu nhiên Y có phân bố chuẩn $\mathbf{N}(\mu_2;\sigma^2)$. Người ta cho rằng đường kính trung bình của hai loại vòng bi này bằng nhau. Lấy ngẫu nhiên 10 phần tử loại I và 10 phần tử loại II. Ta thu được kết quả sau:

X	2,066	2,068	2,063	2,060	2,067	2,059	2,065	2,062	2,061	2,064
Y	2,063	2,060	2,057	2,056	2,059	2,058	2,062	2,059	2,062	2,059

Hãy kiểm định giả thiết trên với mức ý nghĩa $\alpha = 0,05$.

6.19 Để so sánh trọng lượng trung bình của trẻ sơ sinh thành thị và nông thôn, người ta theo dõi 1000 cháu và thu được bảng sau

Kết quả	n số cháu	Trung bình mẫu $\overline{x}$	Độ lệch mẫu s
Nông thôn	500	3,0 kg	0,4 kg
Thành thị	500	3,2 kg	0,3 kg

Với mức ý nghĩa $\alpha = 0,05$ có thể coi trọng lượng trẻ sơ sinh ở thành thị cao hơn nông thôn được không?

6.20 Sau khi theo dõi thời gian hoàn thành sản phẩm của hai công nhân A và B ta có kết quả sau:

Kết quả	số sản phẩm n	Trung bình mẫu $\overline{x}$	Độ lệch mẫu s
Công nhân A	50	32 phút	4 phút
Công nhân B	60	30 phút	3 phút

Với mức ý nghĩa $\alpha = 0,05$ có thể coi công nhân B hoàn thành sản phẩm nhanh hơn công nhân A không. Giả sử thời gian hoàn thành sản phẩm của hai công nhân là hai biến ngẫu nhiên phân bố chuẩn.

HƯỚNG DẪN GIẢI VÀ ĐÁP ÁN BÀI TẬP

HƯỚNG DẪN GIẢI VÀ ĐÁP ÁN CHƯƠNG 1

1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	1.10
Đúng	Sai	Đúng	Sai	Sai	Đúng	Sai	Sai	Đúng	Đúng

1.11 Có 5! cách đi.

1.12 **a.** $P = 0{,}246$ **b.** $P = 0{,}495$.

1.13 **A.** 1/3 **b.** 3/5 **c.** 2/5 **d.** 4/5

1.14 **a.** $P = \dfrac{C_3^2}{C_5^2} = \dfrac{3}{10}$ **b.** $P = \dfrac{C_2^2}{C_5^2} = \dfrac{1}{10}$ **c.** $P = \dfrac{C_2^1 C_3^1}{C_5^2} = \dfrac{3}{5}$.

1.15 Mỗi khách đều có 6 khả năng để ra ở 6 tầng còn lại của tòa nhà. Do đó số kết cục đồng khả năng có thể $N = 6^3 = 216$.

Gọi A là biến cố tất cả cùng ra ở tầng bốn, biến cố này chỉ có 1 trường hợp thuận lợi. Do đó $P(A) = \dfrac{1}{216}$.

Lý luận tương tự trên ta có $P_b = \dfrac{6}{216} = \dfrac{1}{36}$; $P_c = \dfrac{A_6^3}{216} = \dfrac{5}{9}$.

1.16 $P = \dfrac{1}{720}$.

1.17 **a.** $A_1 \cap A_2 \cap \ldots . \cap A_{10}$ **b.** $A_1 \cup A_2 \cup \ldots \cup A_{10}$

c. $\overline{A_1} \cap \overline{A_2} \cap \ldots . \cap \overline{A_6} \cap A_7 \cap A_8 \cap A_9 \cap A_{10}$ **d.** $A_1 \cap A_2 \cap \ldots \cap A_6$.

1.18 Gọi A_1 và A_2 tương ứng là biến cố người thứ nhất và thứ hai bắn trúng mục tiêu, A là biến cố chỉ có một người bắn trúng mục tiêu. $A = \left(A_1 \cap \overline{A}_2\right) \cup \left(\overline{A}_1 \cap A_2\right)$. Sử dụng qui tắc cộng xác suất trường hợp xung khắc và qui tắc nhân trường hợp độc lập ta có:

$$P(A) = P(A_1 \cap \overline{A}_2) + P(\overline{A}_1 \cap A_2) = P(A_1)P(\overline{A}_2) + P(\overline{A}_1)P(A_2) = 0{,}8 \cdot 0{,}1 + 0{,}2 \cdot 0{,}9 = 0{,}26$$

Tương tự ta có: $P_b = 0{,}98$; $P_c = 0{,}02$.

1.19 Gọi A_1 là biến cố sản phẩm lấy ra thuộc loại 1.

Gọi A_2 là biến cố sản phẩm lấy ra thuộc loại 2.

Gọi A là biến cố sản phẩm lấy ra thuộc loại 1 hoặc loại 2: $A = A_1 \cup A_2$

Vì A_1, A_2 xung khắc do đó

$$P(A)=P(A_1\cup A_2)=P(A_1)+P(A_2)=0,4+0,5=0,9\,.$$

1.20 $P=1-P_0(3;0,6)=1-0,6^3=0,784$.

1.21 $P=\dfrac{C_{980}^{25}C_{20}^{5}}{C_{1000}^{30}}\approx 0,00027$.

1.22 **a.** $\dfrac{C_4^4C_{48}^1}{C_{52}^5}=\dfrac{1}{54.145}$ **b.** $\dfrac{C_4^4C_4^1}{C_{52}^5}=\dfrac{1}{649.740}$ **c.** $\dfrac{C_4^3C_4^2}{C_{52}^5}=\dfrac{1}{108.290}$

d. $\dfrac{C_4^1C_4^1C_4^1C_4^1C_4^1}{C_{52}^5}=\dfrac{64}{162.435}=0,000394$ **e.** $\dfrac{4C_{13}^5}{C_{52}^5}=\dfrac{1}{500}=0,002$

f. $\dfrac{4}{C_{52}^5}=\dfrac{1}{649.740}$ **g.** $1-\dfrac{C_{48}^5}{C_{52}^5}=1-\dfrac{35673}{54145}=\dfrac{18472}{54145}=0,3412$

1.23 **a.** $P=\dfrac{4^3}{52^3}=\dfrac{1}{2197}$ **b.** $P=\dfrac{4}{52}\dfrac{3}{51}\dfrac{2}{50}=\dfrac{1}{5525}$

1.24 Số phế phẩm trung bình trong một ngày sản xuất là $12000.3\%=360$ và số sản phẩm không phải phế phẩm là 11.640. Vậy xác suất cần tìm là $P=\dfrac{C_{360}^{12}C_{11640}^{588}}{C_{12000}^{600}}$.

1.25 Gọi A_i là biến cố sản phẩm đã qua kiểm tra chất lượng ở vòng thứ i, $i=1,2,3$.

Gọi B là biến cố phế phẩm được nhập kho.

$$P(B)=P(\overline{A_1})P(\overline{A_2})P(\overline{A_3})=(1-0,8)(1-0,9)(1-0,99)=0,0002\,.$$

1.26 P = 0,11.

1.27 Gọi A_i là biến cố lần thứ i lấy ra 3 sản phẩm mới để kiểm tra, ($i=1,2,3$). Gọi A là biến cố sau 3 lần kiểm tra tất cả các sản phẩm đều được kiểm tra $A=A_1\cap A_2\cap A_3$. Vì các biến cố phụ thuộc nên $P(A)=P(A_1)P(A_2|A_1)P(A_3|A_1\cap A_2)=1\cdot\dfrac{5}{21}\cdot\dfrac{1}{84}=\dfrac{5}{1764}$.

1.28 $P(A)+P(B)-P(A)P(B)=0,65\ \Rightarrow P(B)=0,5$.

1.29 **a.** $P(A|B)=3/4$ **b.** $P(B|A)=1/2$ **c.** $P(A\cup B)=\dfrac{1}{2}+\dfrac{1}{3}-\dfrac{1}{3}=\dfrac{7}{12}$

1.30 Gọi A_i là biến cố lần thứ nhất chọn được là số i, $i=0,1,...,9$. Gọi B là biến cố lần thứ hai chọn được số 4.

$$P(A_i)=\frac{1}{10};\ P(B\,|\,A_i)=\frac{1}{9},\ \forall i\neq 4;\ P(B\,|\,A_4)=0\,.$$

Vậy $P(B)=\displaystyle\sum_{i=1}^{10}P(A_i)P(B\,|\,A_i)=9\left(\frac{1}{9}\right)\left(\frac{1}{10}\right)=\frac{1}{10}$.

1.31 Gọi A_1, A_2 lần lượt là biến cố người thứ nhất và người thứ hai chọn được sản phẩm loại I.

$$P(A_2)=P(A_1)P(A_2\mid A_1)+P(\overline{A_1})P(A_2\mid \overline{A_1})=\frac{4}{9}\frac{3}{8}+\frac{5}{9}\frac{4}{8}=\frac{4}{9}$$

$$P(A_1\mid A_2)=\frac{P(A_1\cap A_2)}{P(A_2)}=\frac{P(A_1)P(A_2\mid A_1)}{P(A_2)}=\frac{(4/9)(3/8)}{(4/9)}=\frac{3}{8}$$

1.32 Gọi A là biến cố sản phẩm kiểm tra là phế phẩm.

Gọi B_i là biến cố sản phẩm lấy ra kiểm tra thuộc phân xưởng thứ i, i=1,2 3.

$P(B_1)=0,36;\ P(B_2)=0,34;\ P(B_3)=0,30$. Hệ $\{B_1,B_2,B_3\}$ đầy đủ

$P(A|B_1)=0,12;\ P(A|B_2)=0,10;\ P(A|B_3)=0,08$.

a. $P(A)=P(B_1)P(A|B_1)+P(B_2)P(A|B_2)+P(B_3)P(A|B_3)=0,1012$

b. $P(B_1|A)=\dfrac{P(B_1)P(A|B_1)}{P(A)}=\dfrac{0,36\times 0,12}{0,1012}=0,427$

$$P(B_2|A)=\frac{P(B_2)P(A|B_2)}{P(A)}=\frac{0,34\times 0,10}{0,1012}=0,336$$

$$P(B_3|A)=\frac{P(B_3)P(A|B_3)}{P(A)}=\frac{0,30\times 0,08}{0,1012}=0,237$$

1.33 Gọi B_i là biến cố xạ thủ được xét thuộc nhóm thứ *i, i*=1,2,3,4.

Gọi A là biến cố xạ thủ bắn trượt.

Theo đề bài ta có: $P(B_1)=\dfrac{5}{18},\ P(B_2)=\dfrac{7}{18},\ P(B_3)=\dfrac{4}{18},\ P(B_4)=\dfrac{2}{18}$

$$P(A|B_1)=0,2,\ P(A|B_2)=0,3,\ P(A|B_3)=0,4,\ P(A|B_4)=0,5.$$

$$P(A)=P(B_1)P(A|B_1)+P(B_2)P(A|B_2)+P(B_3)P(A|B_3)+P(B_4)P(A|B_4)$$
$$=\frac{5}{18}.0,2+\frac{7}{18}.0,3+\frac{4}{18}.\,0,4+\frac{2}{18}.0,5=\frac{57}{180}$$

Áp dụng công thức Bayes, ta thu được

$$P(B_1|A)=\frac{P(B_1)P(A|B_1)}{P(A)}=\frac{\frac{5}{18}.0,2}{\frac{57}{180}}=\frac{10}{57},$$

$$P(B_2|A)=\frac{21}{57},\ P(B_3|A)=\frac{16}{57},\ P(B_4|A)=\frac{10}{57}.$$

Vậy xạ thủ có khả năng ở nhóm thứ hai nhất.

1.34 Gọi B_1 là biến cố viên đạn thứ nhất trúng mục tiêu, $P(B_1)=0,7$.

Gọi B_2 là biến cố viên đạn thứ hai trúng mục tiêu, $P(B_2)=0,4$.

Hai biến cố này độc lập

Xác suất biến cố chỉ có viên đạn thứ nhất trúng mục tiêu

$$P(A)=P\left(\left(B_1\cap\overline{B_2}\right)\cup\left(B_2\cap\overline{B_1}\right)\right)=P\left(B_1\cap\overline{B_2}\right)+P\left(B_2\cap\overline{B_1}\right)$$
$$=0,7.0,6+0,4.0,3=0,54$$

b. $P\left(B_1|A\right)=\dfrac{P\left(B_1\cap A\right)}{P(A)}=\dfrac{P\left(B_1\cap\left[\left(B_1\cap\overline{B_2}\right)\cup\left(B_2\cap\overline{B_1}\right)\right]\right)}{P(A)}=\dfrac{P\left(B_1\cap\overline{B_2}\right)}{P(A)}=\dfrac{0,7.0,6}{0,54}=0,778$

1.35 Gọi H_A và H_B là biến cố tín hiệu A và B tương ứng đã được phát.

Ta có $P(H_A)=0,85;\ P(H_B)=0,15$, $\{H_A,H_B\}$ là hệ đầy đủ.

Gọi luôn A là biến cố thu được tín hiệu A.

$$P\left(A|H_A\right)=1-\frac{1}{7}=\frac{6}{7};\ P\left(A|H_B\right)=\frac{1}{8}$$

a. $P(A)=P(H_A)P\left(A|H_A\right)+P(H_B)P\left(A|H_B\right)=0,85\times\dfrac{6}{7}+0,15\times\dfrac{1}{8}=0,747$

b. $P\left(H_A|A\right)=\dfrac{P(H_A)P\left(A|H_A\right)}{P(A)}=\dfrac{0,85\times\frac{6}{7}}{0,747}=0,975$

1.36 A_1, A_2 lần lượt là biến cố sản phẩm được chọn từ lô hàng thứ nhất, thứ hai. Gọi B là biến cố chọn được sản phẩm loại I.

$$P(A_1)=P(A_2)=\frac{1}{2};\ P(B\,|\,A_1)=\frac{3}{5},\ P(B\,|\,A_2)=\frac{2}{7}\ \Rightarrow P(B)=\frac{1}{2}\left(\frac{3}{5}+\frac{2}{7}\right)$$

$$\Rightarrow P\left(A_1|B\right)=\frac{P(A_1)P(B\,|\,A_1)}{P(B)}=\frac{21}{31}.$$

1.37 Gọi A là biến cố sản phẩm kiểm tra có kết luận đạt tiêu chuẩn chất lượng.

Gọi B_T là biến cố sản phẩm đạt tiêu chuẩn chất lượng.

Gọi B_H là biến cố sản phẩm không đạt tiêu chuẩn chất lượng.

$$P(B_T)=0,85;\ P(B_H)=0,15$$

Hệ $\{B_T,B_H\}$ đầy đủ

$$P\left(A|B_T\right)=0,9;\ P\left(\overline{A}|B_H\right)=0,95\ \Rightarrow\ P\left(\overline{A}|B_T\right)=0,1;\ P\left(A|B_H\right)=0,05.$$

a. $P(A)=P(B_T)P\left(A|B_T\right)+P(B_H)P\left(A|B_H\right)=0,85\times0,9+0,15\times0,05=0,7725$

b. $P\left(B_H|A\right)=\dfrac{P(B_H)P\left(A|B_H\right)}{P(A)}=\dfrac{0,15\times0,05}{0,7725}=0,0097$

c. $P\left(\left(A\cap B_T\right)\cup\left(\overline{A}\cap B_H\right)\right)=P\left(A\cap B_T\right)+P\left(\overline{A}\cap B_H\right)=0,85\times0,9+0,15\times0,95=0,9075$.

HƯỚNG DẪN GIẢI VÀ ĐÁP ÁN CHƯƠNG 2

2.1	**2.2**	**2.3**	**2.4**	**2.5**	**2.6**	**2.7**	**2.8**	**2.9**	**2.10**
Sai	Sai	Đúng	Sai	Sai	Đúng	Sai	Sai	Đúng	Đúng

2.11	**2.12**	**2.13**	**2.14**	**2.15**	**2.16**	**2.17**	**2.18**	**2.19**	**2.20**
Đúng	Sai	Sai	Sai	Sai	Đúng	Đúng	Đúng	Sai	Sai

2.14 $\mathrm{E}(X) = -0,3;\ \mathrm{D}(X) = 15,21$.

2.15 $x_3 = 29,1;\ p_3 = 0,2$.

2.16 $\mathrm{E}(X_1) = 3,1;\ \mathrm{E}(X_2) = 3,4;$

$\mathrm{D}(X_1) = 1,09;\ \mathrm{D}(X_2) = 1,44$.

$\mathrm{E}(X_1 + X_2) = 6,5;\ \mathrm{D}(X_1 + X_2) = 2,53$.

2.17 $\mathrm{E}(\overline{X}) = 0,8;\ \ \mathrm{D}(\overline{X}) = 0,12$.

2.18 a) $\mathrm{D}(Z) = 61$. b) $\mathrm{D}(Z) = 41$.

2.19 $p_1 = 0,4;\ p_2 = 0,1;\ p_3 = 0,5$.

2.20 a) Gọi A_i là biến cố toa i có người ngồi ($i = \overline{1,3}$). Gọi A là biến cố cả 3 toa đều có người ngồi. Khi đó: $A = A_1A_2A_3 \Rightarrow \overline{A} = \overline{A_1} + \overline{A_2} + \overline{A_3}$.

$$P(\overline{A_i}) = \frac{2^5}{3^5} = \frac{32}{243};\ P(\overline{A_i}\,\overline{A_j}) = \frac{1}{243};\ P(\overline{A_1}\,\overline{A_2}\,\overline{A_3}) = 0$$

$$P(\overline{A}) = P(\overline{A_1}) + P(\overline{A_2}) + P(\overline{A_3}) - P(\overline{A_1}\,\overline{A_2}) - P(\overline{A_1}\,\overline{A_3}) - P(\overline{A_2}\,\overline{A_3}) + P(\overline{A_1}\,\overline{A_2}\,\overline{A_3}) = \frac{31}{81}$$

$$\Rightarrow\ P(A) = \frac{50}{81}.$$

b) $P\{X = 0\} = P(\overline{A_1}) = \dfrac{2^5}{3^5} = \dfrac{32}{243};\ P\{X = k\} = \dfrac{C_5^k\, 2^{5-k}}{3^5}$

X	0	1	2	3	4	5
P	$\frac{32}{243}$	$\frac{80}{243}$	$\frac{80}{243}$	$\frac{40}{243}$	$\frac{10}{243}$	$\frac{1}{243}$

Y	0	1	2	3	4	5
P	$\frac{1}{243}$	$\frac{10}{243}$	$\frac{40}{243}$	$\frac{80}{243}$	$\frac{80}{243}$	$\frac{32}{243}$

2.21 $\mathrm{E}(X)=0$. $\mathrm{E}X^2=\dfrac{\pi^2}{4}-2 \Rightarrow \mathrm{D}X=\dfrac{\pi^2}{4}-2$.

2.22 a) $\displaystyle\int_{-\infty}^{\infty} f_X(x)dx = k\int_0^1 xdx+3k=1 \Rightarrow k=\frac{2}{7}$

$$F_X(x)=\begin{cases} 0 & \text{nÕu } x\le 0\\ \dfrac{x^2}{7} & \text{nÕu } 0<x\le 1\\ \dfrac{1}{7}+\dfrac{2}{7}(x-1) & \text{nÕu } 1<x\le 4\\ 1 & \text{nÕu } 4<x\end{cases}$$

b. $\displaystyle \mathrm{E}X=\frac{2}{7}\int_0^1 x^2dx+\frac{2}{7}\int_1^4 xdx=\frac{47}{21}=2,238$

$$\mathrm{E}X^2=\frac{2}{7}\int_0^1 x^3dx+\frac{2}{7}\int_1^4 x^2dx=\frac{85}{14} \Rightarrow \mathrm{D}X=\mathrm{E}X^2-(\mathrm{E}X)^2=\frac{937}{882}=1,062\,.$$

2.23 a) Vì $\displaystyle\int_0^4 x^2(4-x)dx=\frac{64}{3} \Rightarrow k=\frac{3}{64}$.

b) $\displaystyle P\{X<1\}=\int_0^1 \frac{3}{64}x^2(4-x)dx=\frac{13}{256}$.

c) $\displaystyle \mathrm{E}X=\int_0^4 \frac{3}{64}x^3(4-x)dx=\frac{3}{64}\left(x^4-\frac{x^5}{5}\right)\Bigg|_{x=0}^{x=4}=3\left(4-\frac{16}{5}\right)=\frac{12}{5}$

$$\mathrm{E}X^2=\int_0^4 \frac{3}{64}x^4(4-x)dx=\frac{3}{64}\left(\frac{4x^5}{5}-\frac{x^6}{6}\right)\Bigg|_{x=0}^{x=4}=3\times 64\left(\frac{1}{5}-\frac{1}{6}\right)=\frac{32}{5}$$

$$\Rightarrow \mathrm{D}X=\mathrm{E}X^2-(\mathrm{E}X)^2=\frac{32}{5}-\left(\frac{12}{5}\right)^2=\frac{16}{25}.$$

2.24 a) Kí hiệu A_i là biến cố : "A bắn trúng i viên",

B_i là biến cố : "B bắn trúng i viên"; i = 0, 1 2. Dễ thấy

$P(A_0)=0,36$; $P(A_1)=0,48$; $P(A_2)=0,16$;

$P(B_0)=0,25$; $P(B_1)=0,5$; $P(B_2)=0,25$.

Từ đó $P\{X=-2\}=P(A_0)P(B_2)=0,09$

$P\{X=-1\}=P(A_0)P(B_1)+P(A_1)P(B_2)=0,18+0,12=0,3$

$P\{X=0\}=P(A_0)P(B_0)+P(A_1)P(B_1)+P(A_2)P(B_2)=0,37$

$P\{X=1\}=P(A_1)P(B_0)+P(A_2)P(B_1)=0,2$

$P\{X=2\}=P(A_2)P(B_0)=0,04$

Vậy bảng phân bố xác suất của X

X	-2	-1	0	1	2
P	0,09	0,3	0,37	0,2	0,04

$\mathrm{E}X=(-2)\times 0,09+(-1)\times 0,3+0\times 0,37+1\times 0,2+2\times 0,04=-0,2$

$\mathrm{E}X^2=(-2)^2\times 0,09+(-1)^2\times 0,3+0^2\times 0,37+1^2\times 0,2+2^2\times 0,04=1,02$

$\mathrm{D}X=\mathrm{E}X^2-(\mathrm{E}X)^2=1,02-(-0,2)^2=0,98$

b) $P\{Y=0\}=0,37$

$P\{Y=1\}=P\{X=1\}+P\{X=-1\}=0,5$

$P\{Y=2\}=P\{X=2\}+P\{X=-2\}=0,13$

$\mathrm{E}Y=0\times 0,37+1\times 0,5+2\times 0,13=0,76$.

2.25 Kí hiệu A_i là biến cố : "ô tô thứ i bị hỏng", i = 1, 2.

Dễ thấy $P(A_1)=0,1$; $P(A_2)=0,2$

Gọi X là số ô tô bị hỏng trong thời gian làm việc

Từ đó $P\{X=0\}=P(\overline{A_1})P(\overline{A_2})=0,9\times 0,8=0,72$

$P\{X=1\}=P(A_1)P(\overline{A_2})+P(\overline{A_1})P(A_2)=0,1\times 0,8+0,9\times 0,2=0,26$

$P\{X=2\}=P(A_1)P(A_2)=0,02$

X	0	1	2
P	0,72	0,26	0,02

$\mathrm{E}X=0\times 0,72+1\times 0,26+2\times 0,02=0,3$.

$\mathrm{E}X^2=0^2\times 0,72+1^2\times 0,26+2^2\times 0,02=0,34$.

$\mathrm{D}X=\mathrm{E}X^2-(\mathrm{E}X)^2=0,34-(0,3)^2=0,25$.

2.26 a) $1=\int\limits_{-\infty}^{\infty} f_X(x)dx=k\int\limits_0^3 x^2dx=9k \Rightarrow k=\frac{1}{9}$

b) $P\{X>2\}=\frac{1}{9}\int\limits_2^3 x^2dx=\frac{19}{27}$.

c) Hàm phân bố $F_X(x)=\begin{cases} 0 & \text{nÕu} \quad x<0 \\ \dfrac{x^3}{27} & \text{nÕu} \quad 0\le x\le 3 \\ 1 & \text{nÕu} \quad x>3 \end{cases}$

d) $F_X(\alpha)=\dfrac{3}{4} \Rightarrow \dfrac{\alpha^3}{27}=\dfrac{3}{4} \Rightarrow \alpha=2,726$.

2.27 a) Điều kiện $\begin{cases} p_i>0 \\ \sum\limits_i p_i=1 \end{cases} \Rightarrow \begin{cases} k>0 \\ 10k^2+9k=1 \end{cases} \Rightarrow \begin{cases} k>0 \\ \left[\begin{array}{l} k=-1 \\ k=1/10 \end{array}\right. \end{cases} \Rightarrow k=1/10$.

b) $P\{X\ge 5\}=\dfrac{1}{10}+\dfrac{1}{10}=\dfrac{2}{10}$; $P\{X<3\}=\dfrac{3}{10}$.

c) $\mathrm{E}X=\dfrac{1}{10}+\dfrac{4}{10}+\dfrac{6}{10}+\dfrac{12}{10}+\dfrac{5}{100}+\dfrac{12}{100}+7\left(\dfrac{7}{100}+\dfrac{1}{10}\right)=3,66$.

d) $\mathrm{E}X^2=\dfrac{1}{10}+\dfrac{8}{10}+\dfrac{18}{10}+\dfrac{48}{10}+\dfrac{25}{100}+\dfrac{72}{100}+49\left(\dfrac{7}{100}+\dfrac{1}{10}\right)=16,8$.

$\mathrm{D}X=\mathrm{E}X^2-(\mathrm{E}X)^2=16,8-(3,66)^2=3,404$.

2.28 a) Gọi X là "số phế phẩm gặp phải":

X	0	1
P	0,6	0,4

$\mathrm{E}X=\dfrac{2}{5}$; $\mathrm{D}X=\dfrac{6}{25}$.

b) Gọi Y là "số chính phẩm nhận được"

$\Rightarrow Y=2-X$:

Y	1	2
P	0,4	0,6

$\mathrm{E}Y=\mathrm{E}(2-X)=2-\dfrac{2}{5}=\dfrac{8}{5}$; $\mathrm{D}Y=\dfrac{6}{25}$.

2.29 Gọi X là "số nữ có trong nhóm được chọn"

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

$P\{X=2\}=\dfrac{C_4^2C_6^1}{C_{10}^3}=\dfrac{9}{30}$, $P\{X=1\}=\dfrac{C_4^1C_6^2}{C_{10}^3}=\dfrac{15}{30}$

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

$\mathrm{E}X=0\times\dfrac{5}{30}+1\times\dfrac{15}{30}+2\times\dfrac{9}{30}+3\times\dfrac{1}{30}=\dfrac{36}{30}=\dfrac{6}{5}$.

2.30 a) $1=\int\limits_{-\infty}^{\infty} f(x)dx=k\int\limits_0^1 x^2(1-x)dx=\dfrac{k}{12} \Rightarrow k=12$.

$$P\{0,4<X<0,6\}=12\int_{0,4}^{0,6}x^2(1-x)dx=0,296.$$

b) $\mathrm{E}X=12\int_0^1 x^3(1-x)dx=\dfrac{3}{5}$.

2.31 Gọi X là biến ngẫu nhiên chỉ số rau bán được trong một ngày.

X	0	1000	2000	3000
P	0,1	0,30	0,45	0,15

Ta xác định được lượng rau bán trung bình trong một ngày:

$$\mathrm{E}X=0\cdot 0,1+1000\cdot 0,3+2000\cdot 0,45+3000\cdot 0,15=1650\,\text{kg/ngày}.$$

Gọi Y là biến ngẫu nhiên chỉ số tiền lãi trung bình trong 1 ngày. Ta thấy

$$Y=\begin{cases}1650\cdot 50-(X-1650)\cdot 20 & \text{nÕu}\quad X>1650\\ X\cdot 50 & \text{nÕu}\quad X\le 1650\end{cases}$$

Do đó $Y(1000)=50.000$; $Y(2000)=75.500$; $Y(3000)=55.500$.

Vậy mỗi ngày nhập 2000kg rau thì có lãi nhiều nhất.

2.32 Thắng 2 trong 4 ván dễ hơn.

2.33 a) $P=0,238$. b) $P=0,751$.

2.34 a) X tuân theo quy luật nhị thức $\mathrm{B}(n;p)$ với $n=5$ và $p=0,8$.

b) $\mathrm{E}X=4$; $\mathrm{D}X=0,8$.

c) $\mathrm{Mod}X=4$; $P\{X=4\}=0,4096$.

2.35 a) $P=0,9914$

b) Số sản phẩm hỏng trung bình là 0,5

c) Số sản phẩm hỏng có khả năng xảy ra nhiều nhất là 0.

2.36 Gọi x là số câu hỏi học sinh trả lời đúng. Số điểm anh ta nhận được là

$$4x+(10-x)(-2)=6x-20$$

a) Anh ta được 4 điểm khi trả lời đúng: $6x-20=4\Rightarrow x=4$.

Vậy xác suất để anh ta được điểm 4 là $P=C_{10}^4\left(\frac{1}{5}\right)^4\left(\frac{4}{5}\right)^6=0,088$.

b) Anh ta được điểm âm khi trả lời đúng: $6x-20<0\Rightarrow x=0,1,2,3$.

Vậy xác suất để anh ta được điểm âm là $P=C_{10}^4\left(\frac{1}{5}\right)^4\left(\frac{4}{5}\right)^6=0,088$.

2.37 Gọi X là số lần thu được tín hiệu trong 5 lần phát độc lập thì $X\sim B(5;\,0,7)$

a) Xác suất thu được tín hiệu 2 lần $P\{X=2\}=C_5^2\,0,7^2\,0,3^3=0,132$

b) Xác suất thu được tín hiệu nhiều nhất 1 lần $P\{X \le 1\} = 0,031$

c) Xác suất thu được tín hiệu $P\{X \ge 1\} = 1 - P\{X = 0\} = 1 - 0,002 = 0,998$

2.38 Gọi X là số lần xạ thủ A bắn trúng bia trong 5 lần bắn độc lập thì $X \sim B(5;\, 0,8)$

a) Xác suất A bắn trúng bia 2 lần $P\{X = 2\} = C_5^2 0,8^2 0,2^3 = 0,051$

b) Xác suất A bắn trúng bia nhiều nhất 2 lần $P\{X \le 2\} = 0,058$

c) Xác suất A bắn trúng bia $P\{X \ge 1\} = 1 - P\{X = 0\} = 0,9996$

2.39 Không đúng; $P = 0,41$.

2.40 a) Gọi X là số cuộc gọi trong khoảng thời gian 10 giây thì X có phân bố Poisson tham số $\lambda = 1/3$. Vậy xác suất có ít nhất một cuộc gọi trong khoảng thời gian 10 giây là $P\{X \ge 1\} = 1 - P\{X = 0\} = 1 - e^{-1/3} = 0,2825$.

b) Gọi Y là số cuộc gọi trong khoảng thời gian 3 phút thì Ị có phân bố Poisson tham số $\lambda = 6$. Vậy xác suất có nhiều nhất ba cuộc gọi trong khoảng thời gian 3 phút là $P\{Y \le 3\} = 0,151$.

c) Gọi Z là số cuộc gọi trong khoảng thời gian 1 phút thì Z có phân bố Poisson tham số $\lambda = 2$. Xác suất có nhiều nhất 1 cuộc gọi trong khoảng thời gian 1 phút là $P\{Z \le 1\} = 0,406$. Vậy xác suất để trong khoảng thời gian 3 phút liên tiếp mỗi phút có nhiều nhất 1 cuộc gọi là

$$P\{Z \le 1\}^3 = 0,406^3 = 0,0067.$$

2.41 Gọi X là số cuộc gọi điện thoại đến trạm điện thoại A trong khoảng thời gian 1 phút thì X có phân bố Poisson tham số $\lambda = 1,5$.

a) Xác suất trạm điện thoại A không nhận được cuộc gọi nào: $P\{X = 0\} = 0,223$.

b) Xác suất trạm điện thoại A nhận được nhiều nhất 2 cuộc gọi: $P\{X \le 2\} = 0,809$.

c) Xác suất trạm điện thoại A nhận được ít nhất 4 cuộc gọi

$$P\{X \ge 4\} = 1 - P\{X \le 3\} = 1 - 0,934 = 0,066.$$

2.42 $P = 0,9$.

2.43 Tra bảng phụ lục II ta được

a) 0,3849 b) 0,2517 c) 0,6636 d) 0,1828 e) 0,8997

2.44 $P\{8 < X < 12\} = \Phi\left(\frac{12-10}{2}\right) - \Phi\left(\frac{8-10}{2}\right) = 0,6826$.

2.45 $P = 0,3$.

2.46 a) 95,44%; b) 4,56%.

2.47 a) 20,33%; b) $P = 0,9983$.

HƯỚNG DẪN GIẢI VÀ ĐÁP ÁN CHƯƠNG 3

3.1	**3.2**	**3.3**	**3.4**	**3.5**	**3.6**	**3.7**	**3.8**	**3.9**	**3.10**
Sai	Đúng	Đúng	Đúng	Sai	Sai	Sai	Đúng	Đúng	Sai

3.11	**3.12**	**3.13**	**3.14**	**3.15**	**3.16**
Đúng	Đúng	Sai	Sai	Đúng	Đúng

3.17

X	x_1	x_2
P	0,56	0,44

Y	y_1	y_2	y_3
P	0,26	0,38	0,36

3.18 $\mathrm{E}X = -7/15$; $\mathrm{E}Y = 0$;

$\mathrm{cov}(X,Y) = -1/8$; $\rho_{X,Y} = -0{,}15$.

3.19 $\mathrm{E}X = -1/5$; $\mathrm{E}Y = 0$; $\rho_{X,Y} = 0$. X và Y không độc lập vì

$P\{X=1\} = 2/15$, $P\{Y=1\} = 5/15$ và $P\{X=1.\,Y=1\} = 0$.

3.20 Bảng phân bố xác suất của biến ngẫu nhiên Z

Z	1	2	3	4	6
P	0,12	0,43	0,03	0,35	0,07

$\mathrm{E}X = 1{,}7$; $\mathrm{E}Y = 1{,}7$; $\mathrm{E}Z = 2{,}89$.

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

X \ Y	0	1	2	3	4
0	0,04	0,12	0,16	0,06	0,02
1	0,03	0,09	0,12	0,045	0,015
2	0,02	0,06	0,08	0,03	0,01
3	0,01	0,03	0,04	0,015	0,005

$P\{X > Y\} = 0{,}19$.

3.22 X, Y không độc lập vì

$P\{X=1\} = 0{,}5$, $P\{Y=1\} = 0{,}45$ và $P\{X=1.\,Y=1\} = 0{,}15 \neq 0{,}5 \cdot 0{,}45$

$P\{X=1|Y=2\} = 7/11$.

3.23

$$P\{X=1\} = 0{,}5,\ P\{Y=1\} = 0{,}45 \text{ và } P\{X=1.\,Y=1\} = 0{,}15 \neq 0{,}5 \cdot 0{,}45.$$

$Y \backslash X$	1	2	3	4	5	6
0	1/12	1/12	1/12	1/12	1/12	1/12
1	1/12	1/12	1/12	1/12	1/12	1/12

3.24

$X\|Y=26$	23	27
P	0,357	0,643

$Y\|X=27$	26	30	41	50
P	0,1268	0,4225	0,1549	0,2958

3.25 $\mathrm{E}\left[X\middle|Y=1\right]=5$. $\mathrm{E}X=4{,}5$; $\mathrm{E}Y=2{,}93$; $\mathrm{D}X=2{,}25$; $\mathrm{D}Y=4{,}83$.

3.26 $P\{X=0\}=P\{X=1\}=\dfrac{3}{6}=\dfrac{1}{2}$; $P\{Y=0\}=P\{Y=1\}=\dfrac{3}{6}=\dfrac{1}{2}$.

X và Y độc lập do đó $P\{X=i;Y=j\}=P\{X=i\}P\{Y=j\}=\dfrac{1}{4}$; $\forall i,j=0,1$.

3.27 a. $\alpha=15$; $\mathrm{E}X=-0{,}2$; $\mathrm{E}Y=0$. b. $\operatorname{cov}(X,Y)=0 \Rightarrow \rho(X,Y)=0$.

c. X , Y không độc lập vì $P\{X=1\}=\dfrac{2}{15}$, $P\{Y=1\}=\dfrac{5}{15}$ nhưng $P\{X=1,Y=1\}=0$.

3.28 $p_{XY}(0,0)=\dfrac{C_4^3}{C_9^3}=\dfrac{4}{84}$ $\quad p_{XY}(0,1)=\dfrac{C_3^1C_4^2}{C_9^3}=\dfrac{18}{84}$ $\quad p_{XY}(0,2)=\dfrac{C_3^2C_4^1}{C_9^3}=\dfrac{12}{84}$

$p_{XY}(0,3)=\dfrac{C_3^3}{C_9^3}=\dfrac{1}{84}$ $\quad p_{XY}(1,0)=\dfrac{C_2^1C_4^2}{C_9^3}=\dfrac{12}{84}$ $\quad p_{XY}(1,1)=\dfrac{C_2^1C_3^1C_4^1}{C_9^3}=\dfrac{24}{84}$

$p_{XY}(1,2)=\dfrac{C_2^1C_3^2}{C_9^3}=\dfrac{6}{84}$ $\quad p_{XY}(2,0)=\dfrac{C_2^2C_4^1}{C_9^3}=\dfrac{4}{84}$ $\quad p_{XY}(2,1)=\dfrac{C_2^2C_3^1}{C_9^3}=\dfrac{3}{84}$.

Bảng phân bố xác suất đồng thời

$X \backslash Y$	0	1	2	3
0	4/84	18/84	12/84	1/84
1	12/84	24/84	6/84	0
2	4/84	3/84	0	0

X và Y không độc lập.

3.29 $p_X(0)=C_3^0\left(\frac{1}{2}\right)^3=\frac{1}{8}$ $\quad p_X(1)=C_3^1\left(\frac{1}{2}\right)^3=\frac{3}{8}$ $\quad p_X(2)=C_3^2\left(\frac{1}{2}\right)^3=\frac{3}{8}$

$p_X(3)=C_3^3\left(\frac{1}{2}\right)^3=\frac{1}{8}$ $\quad p_Y(0)=C_3^0\left(\frac{1}{4}\right)^0\left(\frac{3}{4}\right)^3=\frac{27}{64}$ $\quad p_Y(1)=C_3^1\left(\frac{1}{4}\right)\left(\frac{3}{4}\right)^2=\frac{27}{64}$

$p_Y(2)=C_3^2\left(\frac{1}{4}\right)^2\left(\frac{3}{4}\right)=\frac{9}{64}$ $\quad p_Y(3)=C_3^3\left(\frac{1}{4}\right)^3\left(\frac{3}{4}\right)^0=\frac{1}{64}$.

X và Y độc lập do đó ta có bảng phân bố xác suất đồng thời

X \ Y	0	1	2	3
0	27/512	27/512	9/512	1/512
1	81/512	81/512	27/512	3/512
2	81/512	81/512	27/512	3/512
3	27/512	27/512	9/512	1/512

$P\{X=Y\}=\frac{1}{512}(27+81+27+1)=\frac{126}{512}$; $P\{X>Y\}=\frac{1}{512}(81+81+27+81+27+9)=\frac{306}{512}$

$P\{X+Y>4\}=\frac{1}{512}(3+9+1)=\frac{13}{512}\Rightarrow P\{X+Y\le 4\}=1-P\{X+Y>4\}=1-\frac{13}{512}=\frac{499}{512}$

3.30 Gọi X là số máy hỏng trong ca. X có phân bố nhị thức $\mathrm{E}X=0{,}5$, $\mathrm{D}X=1$. Áp dụng bất đẳng thức Trêbưsép ta có

$$P\{|X-0{,}05|<2\}\ge 1-\frac{0{,}475}{2^2}=0{,}88;\quad P\{|X-0{,}05|\ge 2\}\le\frac{0{,}475}{2^2}=0{,}12.$$

3.31 Đặt $S=\sum_{n=1}^{12}X_n$; $\mathrm{E}S=12\cdot 16=192$, $\mathrm{D}S=12$. Theo bất đẳng thức Trêbưsép

$P\{|S-192|\le\varepsilon\}\ge 1-\frac{\mathrm{D}S}{\varepsilon^2}\ge 0{,}99$. Chọn $a=157{,}36$; $b=226{,}64$.

3.32 Đặt $S=\sum_{n=1}^{10000}X_n$; $\mathrm{E}S=0$, $\mathrm{D}S=\frac{10000}{12}$. Theo bất đẳng thức Trêbưsép

$$P\{|S|\ge 500\}\le\frac{\mathrm{D}S}{500^2}=\frac{1}{300}.$$

3.33 Ta biết rằng S là biến ngẫu nhiên có phân bố nhị thức tham số $p=\frac{1}{6}$.

$$\mathrm{E}S=\frac{n}{6}\text{ và }\mathrm{D}S=\frac{5n}{36}.$$

Theo bất đẳng thức Trêbưsép

$$P\left\{\left|S-\mathrm{E}S\right|<\sqrt{n}\right\}\ge 1-\frac{\mathrm{D}S}{n}=1-\frac{5}{36}=\frac{31}{36}\Leftrightarrow P\left\{\frac{n}{6}-\sqrt{n}<S<\frac{n}{6}+\sqrt{n}\right\}\ge\frac{31}{36}.$$

3.34 Đặt $S=\sum_{n=1}^{12}X_n$. Ta cần tìm M nhỏ nhất để $P\left\{\sum_{n=1}^{12}X_n\le M\right\}\ge 0{,}99$.

$$\mathrm{E}S=192,\quad \mathrm{D}S=12.$$

Theo bất đẳng thức Trêbưsép

$P\left\{\left|S-192\right|\le\varepsilon\right\}\ge 1-\dfrac{\mathrm{D}S}{\varepsilon^2}\ge 0{,}99\Rightarrow\ \varepsilon=34{,}64$. Vậy M = 192+34,64 = 226,64.

3.35 Áp dụng bất đẳng thức Trêbưsép tính được xác suất $P\ge 0{,}9131$

3.36 Áp dụng bất đẳng thức Trêbưsép cần kiểm tra 23.750 chi tiết.

HƯỚNG DẪN GIẢI VÀ ĐÁP ÁN CHƯƠNG 4

4.1	4.2	4.3	4.4	4.5	4.6	4.7	4.8	4.9	4.10
Đúng	Đúng	Sai	Đúng	Đúng	Sai	Đúng	Đúng	Đúng	Sai

4.13 Đặt $u_i=x_i-26\Rightarrow\overline{x}=\overline{u}+26=26$;

$$s^2=\frac{1}{99}\left(1080-\frac{0^2}{100}\right)=10{,}909$$

4.14 $\overline{x}=8{,}9$; $s=2{,}18$

4.15 $\overline{x}=1111{,}1$ (giờ); $s=37$.

4.16 Ta có $\mathrm{E}X=p$; $\mathrm{D}X=p(1-p)$. Do đó $\mathrm{E}\overline{X}=p$; $\mathrm{D}\overline{X}=\dfrac{p(1-p)}{10}$.

4.17 Mẫu ngẫu nhiên có kích thước 10: $W=(X_1,X_2,...,X_{10})$.

$$P\left\{\overline{X}=\frac{1}{2}\right\}=P\left\{\frac{1}{10}\sum_{i=1}^{10}X_i=\frac{1}{2}\right\}=P\left\{\sum_{i=1}^{10}X_i=5\right\}.$$

Vì X có phân bố nhị thức nên

$$P\left\{\sum_{i=1}^{10}X_i=5\right\}=P_{10}(5)=C_{10}^5(0{,}5)^5\cdot(0{,}5)^{10-5}=C_{10}^5(0{,}5)^{10}.$$

4.18 X có phân bố chuẩn $\mathbf{N}(\mu;\sigma^2)$ nên $\overline{X}$ có phân bố chuẩn $\mathbf{N}(\mu;\dfrac{\sigma^2}{n})$. Vậy

$$P\left\{\left|\overline{X}-\mu\right|<\varepsilon\right\}=P\left\{\mu-\varepsilon<\overline{X}<\mu+\varepsilon\right\}=\Phi\left(\frac{\varepsilon\sqrt{n}}{\sigma}\right)-\Phi\left(\frac{-\varepsilon\sqrt{n}}{\sigma}\right)=2\Phi\left(\frac{\varepsilon\sqrt{n}}{\sigma}\right).$$

Do đó $P\left\{\left|\overline{X}-20\right|<0,2\right\}=2\Phi\left(\dfrac{0,2\sqrt{100}}{1}\right)=2\Phi(2)=0,9545$.

4.19 Bảng phân bố tần số

X	1	2	3	4
Tần số	2	4	2	2

Bảng phân bố tần suất

X	1	2	3	4
Tần suất	1/5	2/5	1/5	1/5

Hàm phân bố thực nghiệm

$$F_{10}(x)=\begin{cases}0 & x\le 1\\ 1/5 & 1<x\le 2\\ 3/5 & 2<x\le 3\\ 4/5 & 3<x\le 4\\ 1 & x>4\end{cases}\qquad \overline{x}=6,8;\ s^2=1,15,\ \ s=1,072.$$

4.20 Nhóm 1: $\overline{x}=51,08$; $s^2=220,13$, $s=14,83$.

Nhóm 2: $\overline{x}=45,36$; $s^2=287\ \Rightarrow\ s=16,9$.

Thời gian trung bình của nhóm 2 ít hơn, nhưng độ sai lệch lớn hơn.

HƯỚNG DẪN GIẢI VÀ ĐÁP ÁN CHƯƠNG 5

5.1	**5.2**	**5.3**	**5.4**	**5.5**	**5.6**	**5.7**	**5.8**	**5.9**	**5.10**
Đúng	Sai	Sai	Đúng	Đúng	Sai	Đúng	Sai	Đúng	Đúng

5.12 $\overline{x}=38$; $\dfrac{\sigma}{\sqrt{n}}U_{\alpha/2}=\dfrac{6,4}{\sqrt{50}}.2,575=2,331$. Khoảng tin cậy $[35,669;40,331]$.

5.13 $f=\dfrac{1082}{2000}$; Điều kiện $\begin{cases}nf=1082>10\\ n(1-f)=918>10\end{cases}$

$$f-U_{\alpha/2}\frac{\sqrt{f(1-f)}}{\sqrt{n}}=\frac{1082}{2000}-2,33\frac{\sqrt{918\times1082}}{\sqrt{2000}}=0,515$$

Vậy tối thiểu có 51,5% số phiếu bầu cho ứng cử viên A.

5.14 $\overline{x}=\dfrac{\sum x_i}{n}=\dfrac{34,15}{35}=0,976$.

$$s^2=\frac{1}{n-1}\left[\sum x_i{}^2-\frac{\left(\sum x_i\right)^2}{n}\right]=\frac{1}{34}\left[33,8943-\frac{(34,15)^2}{35}\right]=0,01687\,.$$

$$\Rightarrow s=0,1299\,;\;\; U_{\alpha/2}\frac{s}{\sqrt{n}}=1,96\times\frac{0,1299}{\sqrt{35}}=0,043\,.$$

Khoảng tin cậy 95%: $[0,933\,;\,1,019]$.

5.15 Tần suất mẫu $f=\frac{53}{400}$, điều kiện $\begin{cases} nf=53>10 \\ n(1-f)=347>10 \end{cases}$ thỏa mãn.

Gọi p là xác suất bắt được con cá có đánh dấu, khoảng tin cậy 95% của p :

$$U_{\alpha/2}\frac{\sqrt{f(1-f)}}{\sqrt{n}}=1,96\frac{\sqrt{53\times 347}}{400\sqrt{400}}=0,0332$$

Khoảng ước lượng $[0,0993\,;\,0,1657]$

Mặt khác $p=\frac{2000}{N}$, trong đó N là số cá trong hồ.

Vậy $0,0993<\frac{2000}{N}<0,1657$

$$\Rightarrow\;\frac{2000}{0,1657}<N<\frac{2000}{0,0993}\;\Rightarrow\; 12070<N<20141$$

5.16 Khoảng tin cậy 95% của hao phí nguyên liệu trung bình cho 1 đơn vị sản phẩm là $\left[\overline{x}-U_{\alpha/2}\frac{\sigma}{\sqrt{n}}\,;\,\overline{x}+U_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$, trong đó $U_{\alpha/2}=1,96$ là giá trị tới hạn mức 0,975 của phân bố chuẩn tắc N(0;1).

Theo mẫu ta tính được $\overline{x}=19,87$; $\sigma=0,03$ $\Rightarrow U_{\alpha/2}\frac{\sigma}{\sqrt{n}}=1,96\frac{0,03}{\sqrt{36}}=0,0098$

Khoảng tin cậy $[19,8602\,;\,19,8798]$.

5.17 Đặt $u_i=\frac{x_i-18,25}{5}\Rightarrow \overline{x}=5\frac{\sum u_i}{n}+18,25=5\frac{-1,8}{40}+18,25=18,025\,.$

$$s^2=\frac{5^2}{n-1}\left[\sum u_i{}^2-\frac{\left(\sum u_i\right)^2}{n}\right]=\frac{25}{39}\left[0,76-\frac{(-1,8)^2}{40}\right]=0,435\,.$$

$$\Rightarrow s=0,66\,;\, U_{\alpha/2}\frac{s}{\sqrt{n}}=1,64\times\frac{0,66}{\sqrt{40}}=0,171\,.$$

a) Khoảng tin cậy 90%: $[17,854;\,18,196]$.

b) Kích thược mẫu cần thiết $n\geq\frac{U^2_{\alpha/2}s^2}{\varepsilon^2}=116,99$ chọn $n=117$

5.18 Đặt $u_i = \dfrac{x_i - 47}{2} \Rightarrow \bar{x} = 2\dfrac{\sum u_i}{n} + 47 = 2\dfrac{-47}{100} + 47 = 46,06$.

$$s^2 = \frac{2^2}{n-1}\left[\sum {u_i}^2 - \frac{\left(\sum u_i\right)^2}{n}\right] = \frac{4}{99}\left[175 - \frac{(-47)^2}{100}\right] = 6,178.$$

$$\Rightarrow s = 2,486;\ U_{\alpha/2}\frac{s}{\sqrt{n}} = 1,96 \times \frac{2,486}{\sqrt{100}} = 0,487.$$

a) Khoảng tin cậy 95%: $[45,573;\ 46,547]$.

b) Kích thược mẫu cần thiết $n \ge \dfrac{U^2_{\alpha/2} s^2}{\varepsilon^2} = 148,33$ chọn $n = 149$

5.19 Đặt $u_i = x_i - 50 \Rightarrow \bar{x} = \dfrac{\sum u_i}{n} + 50 = \dfrac{-8}{27} + 50 = 49,704$

$$U_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 1,96 \times \frac{1}{\sqrt{27}} = 0,377.$$

a) Khoảng tin cậy 95%: $[49,327\ ;\ 50,081]$.

b) Kích thược mẫu cần thiết $n \ge \dfrac{U^2_{\alpha/2}\sigma^2}{\varepsilon^2} = 384,16$ chọn $n = 385$

5.20 Đặt $u_i = \dfrac{x_i - 119}{2} \Rightarrow \bar{x} = 2\dfrac{\sum u_i}{n} + 119 = 2\dfrac{-19}{100} + 119 = 118,62$.

$$s^2 = \frac{2^2}{n-1}\left[\sum {u_i}^2 - \frac{\left(\sum u_i\right)^2}{n}\right] = \frac{4}{99}\left[399 - \frac{(-19)^2}{100}\right] = 15,9752.$$

$$\Rightarrow s = 3,9969;\ U_{\alpha/2}\frac{s}{\sqrt{n}} = 1,96 \times \frac{3,9969}{\sqrt{100}} = 0,783.$$

a) Khoảng tin cậy 95%: $[117,837\ ;\ 119,403]$.

b) Kích thược mẫu cần thiết $n \ge \dfrac{U^2_{\alpha/2} s^2}{\varepsilon^2} = 245,48$ chọn $n = 246$.

5.21 a) Tần suất mẫu $f = \dfrac{16}{400} = 0,04$, điều kiện $\begin{cases} nf = 16 > 10 \\ n(1-f) = 384 > 10 \end{cases}$

Gọi p là xác suất phế phẩm của lô hàng, khoảng tin cậy 95% của p :

$$U_{\alpha/2}\frac{\sqrt{f(1-f)}}{\sqrt{n}} = 1,96\frac{\sqrt{16 \times 384}}{400\sqrt{400}} = 0,0192$$

Khoảng ước lượng $[0,0208\ ;\ 0,0592]$

Vậy tỷ lệ phế phẩm tối đa của lô hàng là $5,9\%$.

5.22 $\overline{x}=2\times\dfrac{-14}{100}+91=90,72$; $s^2=4\times\dfrac{1}{99}\left(432-\dfrac{14^2}{100}\right)=17,375 \Rightarrow s=4,168$.

$U_{\alpha/2}\dfrac{S}{\sqrt{n}}=0,817$. Vậy khoảng tin cậy 95% của giá trung bình của loại hàng hoá trên là $[89,903\ ;\ 91,537]$.

5.23 Khoảng tin cậy 95% của phương sai được tính theo công thức (7.22).

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

Tra bảng χ^2 với 15 bậc tự do và với giả thiết $\hat{S}^2=0,5$ ta tìm được khoảng tin cậy:

$$\left(\frac{25\cdot 0,5}{40,646};\frac{25\cdot 0,5}{13,120}\right)=(0,3075\ ;\ 0,9520).$$

HƯỚNG DẪN GIẢI VÀ ĐÁP ÁN CHƯƠNG 6

6.1	6.2	6.3	6.4	6.5	6.6	6.7	6.8	6.9	6.10
Sai	Sai	Đúng	Đúng	Sai	Đúng	Sai	Đúng	Sai	Sai

6.11 Gọi μ là trọng lượng trung bình của một bao sản phẩm được đóng gói. Ta kiểm định giả thiết $H_0:\ \mu=100$; đối thiết $H_1:\ \mu<100$

Tiêu chuẩn kiểm định $T=\dfrac{(100-\overline{X})\sqrt{n}}{S}$; Miền bác bỏ $W_\alpha=\{T>2,086\}$.

Đặt $u_i=\dfrac{x_i-99,25}{5} \quad \Rightarrow \sum r_iu_i=0,4$; $\sum r_iu_i^2=0,42$

$\Rightarrow\ \overline{x}=5\times\dfrac{0,4}{29}+99,25=99,319$;

$s^2=25\times\dfrac{1}{28}\left[0,42-\dfrac{0,4^2}{29}\right]=0,37 \Rightarrow s=0,608$

$T_{qs}=\dfrac{(100-99,319)\sqrt{29}}{0,608}=6,032\in W_\alpha$.

Vậy bác bỏ H_0 chấp nhận H_1, nghĩa là sản phẩm bị đóng thiếu.

6.12 Gọi μ là thời gian trung bình hoàn thành một sản phẩm.

Ta kiểm định giả thiết $H_0:\ \mu=14$; đối thiết $H_1:\ \mu\neq 14$

Tiêu chuẩn kiểm định $T=\dfrac{(\overline{X}-14)\sqrt{n}}{S}$;

Miền bác bỏ $W_\alpha = \{|T| > 1{,}96\}$.

Đặt $u_i = \dfrac{x_i - 15}{2}$ $\Rightarrow \sum r_i u_i = 0$; $\sum r_i {u_i}^2 = 300$

$$\Rightarrow \bar{x} = 15\ ; \quad s^2 = 4 \times \frac{1}{249}\left[300 - \frac{0}{300}\right] = 4{,}819 \Rightarrow s = 2{,}195$$

$$\Rightarrow T_{qs} = \frac{(115-14)\sqrt{300}}{2{,}195} = 7{,}89 \in W_\alpha .$$

Vậy bác bỏ H_0 chấp nhận H_1, nghĩa là cần thay đổi định mức.

6.13 Gọi μ là mức hao phí xăng trung bình của ô tô chạy từ A đến B. Ta kiểm định giả thiết $H_0:\ \mu = 50$; đối thiết $H_1:\ \mu < 50$

Tiêu chuẩn kiểm định $T = \dfrac{\left(50 - \overline{X}\right)\sqrt{n}}{S}$;

Miền bác bỏ $W_\alpha = \{T > 2{,}052\}$.

Theo mẫu ta có $\bar{x} = \dfrac{1387{,}5}{28} = 49{,}5536$;

$$s^2 = \frac{1}{27}\left(6876375 - \frac{1387{,}5^2}{28}\right) = \frac{8{,}1696}{27} = 0{,}3026 \Rightarrow s = 0{,}55$$

$$T_{qs} = \frac{(50-49{,}53)\sqrt{30}}{0{,}55} = 4{,}2948 \in W_\alpha .$$

Vậy bác bỏ H_0 chấp nhận H_1, nghĩa là mức hao phí xăng có giảm xuống.

6.14 Gọi μ là trọng lượng đóng bao trung bình sản phẩm của nhà máy. Ta kiểm định giả thiết $H_0:\ \mu = 20$; đối thiết $H_1:\ \mu \neq 20$

Tiêu chuẩn kiểm định $T = \dfrac{\left(\overline{X} - 20\right)\sqrt{n}}{\sigma}$;

Miền bác bỏ $W_\alpha = \{|T| > 1{,}96\}$.

Theo mẫu ta có $\bar{x} = \bar{u} + 20 = \dfrac{5}{100} + 20 = 20{,}05$.

$$T_{qs} = \frac{(20{,}05-20)\sqrt{100}}{2} = 0{,}25 \notin W_\alpha .$$

Vậy chưa có cơ sở để bác bỏ H_0.

6.15 Gọi μ là số hoá đơn trung bình hệ thống máy tính mới xử lý được trong 1 giờ. Ta kiểm định giả thiết $H_0:\ \mu = 1300$; đối thiết $H_1:\ \mu > 1300$

Tiêu chuẩn kiểm định $T=\dfrac{\left(\overline{X}-1300\right)\sqrt{n}}{S}$;

Miền bác bỏ $W_\alpha=\{T>1,96\}$.

Từ mẫu cụ thể ta có $T_{qs}=\dfrac{(1378-1300)\sqrt{40}}{215}=2,294>1,96$

Vậy bác bỏ H_0 chấp nhận H_1, nghĩa là hệ thống máy tính mới xử lý tốt hơn.

6.16 Gọi p là tỉ lệ phế phẩm do nhà máy sản xuất.

Ta kiểm định giả thiết $H_0: \; p=0,05$; đối thiết $H_1: \; p>0,05$

Tiêu chuẩn kiểm định $T=\dfrac{(f-0,05)\sqrt{n}}{\sqrt{0,05(1-0,05)}}$.

Miền bác bỏ $W_\alpha=\{T>1,64\}$.

Từ mẫu cụ thể ta có $f=0,08$ thoả mãn điều kiện $\begin{cases} nf=24>5 \\ n(1-f)=276>5 \end{cases}$

$$T_{qs}=\frac{(0,08-0,05)\sqrt{300}}{\sqrt{0,05(1-0,05)}}=2,384\in W_\alpha .$$

Vậy bác bỏ H_0 chấp nhận H_1, nghĩa là tỉ lệ phế phẩm của nhà máy có xu hướng tăng lên.

6.17 Ta kiểm định giả thiết H$_0$: $\sigma^2=\sigma_0^2=1$; đối thiết H$_1$: $\sigma^2>1$.

Tiêu chuẩn kiểm định $T=\dfrac{(n-1)S^2}{\sigma_0^2}\sim\chi^2(n-1)$;

Miền bác bỏ $W_\alpha=\{T>49,6\}$.

Từ mẫu cụ thể ta có $T_{qs}=35,09\notin W_\alpha$.

Vậy chưa có cơ sở để bác bỏ H_0.

6.18 Ta kiểm định giả thiết H$_0$: $\mu_1=\mu_2$; đối thiết H$_1$: $\mu_1\neq\mu_2$.

Tiêu chuẩn kiểm định $T=\dfrac{\overline{X}-\overline{Y}}{\sqrt{\dfrac{S_X^2}{n}+\dfrac{S_Y^2}{m}}}\sim\mathbf{N}(0;1)$;

Miền bác bỏ $W_\alpha=\{|T|>1,96\}$.

Từ mẫu cụ thể ta có $T_{qs}\notin W_\alpha$.

Vậy chưa có cơ sở để bác bỏ H_0.

6.19 Ta kiểm định giả thiết H$_0$: $\mu_1=\mu_2$; đối thiết H$_1$: $\mu_1>\mu_2$.

Tiêu chuẩn kiểm định $T = \dfrac{\overline{X} - \overline{Y}}{\sqrt{\dfrac{S_X^2}{n} + \dfrac{S_Y^2}{m}}} \sim \mathbf{N}(0;1)$;

Miền bác bỏ $W_\alpha = \{T > 1{,}64\}$.

Từ mẫu cụ thể ta có $T_{qs} \in W_\alpha$.

Vậy bác bỏ H_0 chấp nhận H_1.

6.20 Gọi X là thời gian hoàn thành sản phẩm của công nhân A,

Y là thời gian hoàn thành sản phẩm của công nhân B.

$$X \sim N(\mu_1, \sigma_1^2);\ Y \sim N(\mu_2, \sigma_2^2)$$

Ta kiểm định giả thiết H_0: $\mu_1 = \mu_2$; đối thiết H_1: $\mu_1 < \mu_2$.

Tiêu chuẩn kiểm định $T = \dfrac{\overline{X_1} - \overline{X_2}}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}}$;

$$k = \frac{(n_1 - 1)(n_2 - 1)}{(n_2 - 1)C^2 + (n_1 - 1)(1 - C)^2};\ C = \frac{\dfrac{S_1^2}{n_1}}{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}.$$

T có phân bố Student k bậc tự do.

Miền bác bỏ $W_\alpha = \{-T > t_\alpha(k)\}$.

Từ mẫu cụ thể ta có $T_{qs} \notin W_\alpha$.

Vậy chưa có cơ sở để bác bỏ H_0.

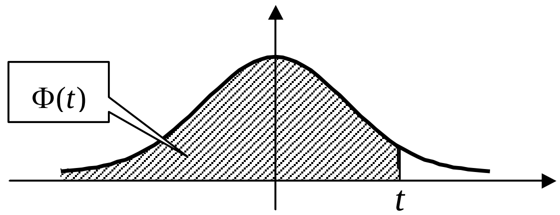
PHỤ LỤC I: GIÁ TRỊ HÀM MẬT ĐỘ

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

	0	1	2	3	4	5	6	7	8	9
0,0	0,3989	3989	3989	3988	3986	3984	3982	3980	3977	3973
0,1	3970	3965	3961	3956	3951	3945	3939	3932	3925	3918
0,2	3910	3902	3894	3885	3876	3867	3857	3847	3836	3825
0,3	3814	3802	3790	3778	3765	3752	3739	3726	3712	3697
0,4	3683	3668	3653	3637	3621	3605	3589	3572	3555	3538
0,5	3521	3503	3485	3467	3448	3429	3410	3391	3372	3352
0,6	3332	3312	3292	3271	3251	3230	3209	3187	3166	3144
0,7	3123	3101	3079	3056	3034	3011	2989	2966	2943	2920
0,8	2897	2874	2850	2827	2803	2780	2756	2732	2709	2685
0,9	2661	2637	2613	2589	2565	2541	2516	2492	2468	2444
1,0	0,2420	2396	2370	2347	2320	2299	2275	2251	2227	2203
1,1	2179	2155	2131	2107	2083	2059	2036	2012	1989	1965
1,2	1942	1919	1895	1872	1849	1826	1804	1781	1758	1736
1,3	1714	1691	1669	1647	1626	1604	1582	1561	1539	1518
1,4	1497	1476	1456	1435	1415	1394	1374	1354	1334	1315
1,5	1295	1276	1257	1238	1219	1200	1182	1163	1145	1127
1,6	1109	1092	1074	1057	1040	1023	1006	0989	0973	0957
1,7	0940	0925	0909	0893	0878	0863	0848	0833	0818	0804
1,8	0790	0775	0761	0748	0734	0721	0707	0694	0681	0669
1,9	0656	0644	0632	0620	0608	0596	0584	0573	0562	0551
2,0	0,0540	0529	0519	0508	0498	0488	0478	0468	0459	0449
2,1	0440	0431	0422	0413	0404	0396	0387	0379	0371	0363
2,2	0355	0347	0339	0332	0325	0317	0310	0303	0297	0290
2,3	0283	0277	0270	0264	0258	0252	0246	0241	0235	0229
2,4	0224	0219	0213	0208	0203	0198	0194	0189	0184	0180
2,5	0175	0171	0167	0163	0158	0154	0151	0147	0143	0139
2,6	0136	0132	0129	0126	0122	0119	0116	0113	0110	0107
2,7	0104	0101	0099	0096	0093	0091	0088	0086	0084	0081
2,8	0079	0077	0075	0073	0071	0069	0067	000065	0063	0061
2,9	0060	0058	0056	0055	0053	0051	0050	0048	0047	0046
3,0	0,0044	0043	0042	0040	0039	0038	0037	0036	0035	0034
3,1	0033	0032	0031	0030	0029	0028	0027	0026	0025	0025
3,2	0024	0023	0022	0022	0021	0020	0020	0019	0018	0018
3,3	0017	0017	0016	0016	0015	0015	0014	0014	0013	0013
3,4	0012	0012	0012	0011	0011	0010	0010	0010	0009	0009
3,5	0009	0008	0008	00080	0008	0007	0007	0007	0007	0006
3,6	0005	0005	0005	0005	0005	0005	0005	0005	0005	0004
3,7	0004	0004	0004	0004	0004	0004	0003	0003	0003	0003
3,8	0003	0003	0003	0003	0003	0002	0002	0002	0002	0002
3,9	0002	0002	0002	0002	0002	0002	0002	0002	0001	0001

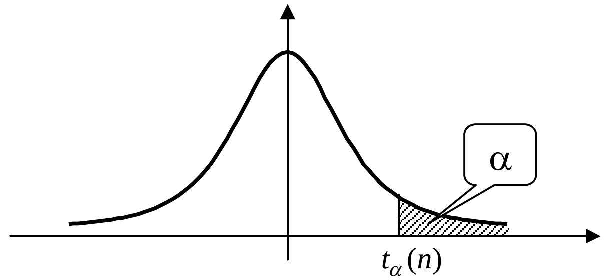
PHỤ LỤC II: GIÁ TRỊ HÀM PHÂN BỐ CHUẨN TẮC

$$\Phi(t)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t} e^{-\frac{x^2}{2}}\,dx$$



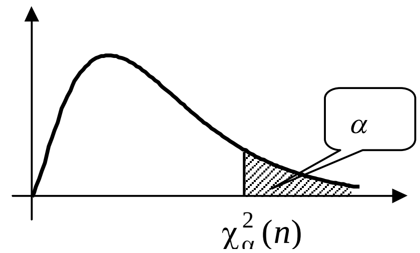
t	0	1	2	3	4	5	6	7	8	9
0,0	0,5000	5040	5080	5120	5160	5199	5239	5279	5319	5359
0,1	5398	5438	5478	5517	5557	5596	5636	5675	5714	5753
0,2	5793	5832	5871	5910	5948	5987	6026	6064	6103	6141
0,3	6179	6217	6255	6293	6331	6368	6406	6443	6480	6517
0,4	6554	6591	6628	6664	6700	6736	6772	6808	6844	6879
0,5	0,6915	6950	6985	7019	7054	7088	7123	7156	7190	7224
0,6	7257	7291	7324	7357	7389	7422	7454	7486	7517	7549
0,7	7580	7611	7642	7673	7703	7734	7764	7794	7823	7852
0,8	7881	7910	7939	7967	7995	8023	8051	8078	8106	8132
0,9	8159	8186	8212	8238	8264	8289	8315	8340	8365	8389
1,0	0,8413	8438	8461	8485	8508	8531	8554	8577	8599	8621
1,1	8643	8665	8686	8708	8729	8749	8770	8790	8810	8830
1,2	8849	8869	8888	8907	8925	8944	8962	8980	8997	9015
1,3	9032	9049	9066	9082	9099	9115	9131	9147	9162	9177
1,4	9192	9207	9222	9236	9251	9265	9279	9292	9306	9319
1,5	0,9332	9345	9357	9370	9382	9394	9406	9418	9429	9441
1,6	9452	9463	9474	9484	9495	9505	9515	9525	9535	9545
1,7	9554	9564	9573	9582	9591	9599	9608	9616	9625	9633
1,8	9641	9649	9656	9664	9671	9678	9686	9693	9699	9706
1,9	9712	9719	9726	9732	9738	9744	9750	9756	9761	9767
2,0	0,9773	9778	9783	9788	9793	9798	9803	9808	9812	9817
2,1	9821	9826	9830	9834	9838	9842	9846	9850	9854	9857
2,2	9861	9864	9868	9871	9875	9878	9881	9884	9887	9890
2,3	9893	9896	9898	9901	9904	9906	9909	9911	9913	9916
2,4	9918	9920	9922	9925	9927	9929	9931	9932	9934	9936
2,5	0,9938	9940	9941	9943	9945	9946	9948	9949	9951	9952
2,6	9953	9955	9956	9957	9959	9960	9961	9962	9963	9964
2,7	9965	9966	9967	9968	9969	9970	9971	9972	9973	9974
2,8	9974	9975	9976	9977	9977	9978	9979	9979	9980	9981
2,9	9981	9982	9982	9983	9984	9984	9985	9985	9986	9986
t	3,0	3,1	3,2	3,3	3,4	3,5	3,6	3,7	3,8	3,9
$\Phi(t)$	0,9987	9990	9993	9995	9996	9997	9998	9999	9999	9999

PHỤ LỤC III: GIÁ TRỊ TỚI HẠN CỦA PHÂN BỐ STUDENT



Bậc tự do	α = 0,05	α = 0,025	α = 0,01	α = 0,005	α = 0,001
1	6,314	12,706	31,821	63,657	318,309
2	2,920	4,303	6,965	9,925	22,327
3	2,353	3,128	4,541	5,841	10,215
4	2,132	2,776	3,747	4,604	7,173
5	2,015	2,571	3,365	4,032	5,893
6	1,943	2,447	3,143	3,707	5,208
7	1,895	2,365	2,998	3,499	4,705
8	1,860	2,306	2,896	3,355	4,501
9	1,833	2,262	2,821	3,250	4,297
10	1,812	2,228	2,764	3,169	4,144
11	1,796	2,201	2,718	3,106	4,025
12	1,782	2,179	2,681	3,055	3,930
13	1,771	2,160	2,650	3,012	3,852
14	1,761	2,145	2,624	2,977	3,787
15	1,753	2,131	2,606	2,947	3,733
16	1,746	2,120	2,583	2,921	3,686
17	1,740	2,110	2,567	2,898	3,646
18	1,734	2,101	2,552	2,878	3,610
19	1,729	2,093	2,539	2,861	3,579
20	1,725	2,086	2,58	2,845	3,552
21	1,721	2,080	2,518	2,831	3,527
22	1,717	2,074	2,508	2,819	3,505
23	1,714	2,069	2,500	2,807	3,485
24	1,711	2,064	2,492	2,797	3,467
25	1,708	2,060	2,485	2,787	3,450
26	1,796	2,056	2,479	2,779	3,435
27	1,703	2,052	2,473	2,771	3,421
28	1,701	2,048	2,467	2,763	3,408
29	1,699	2,045	2,462	2,756	3,396
inf	1,645	1,960	2,326	2,576	3,090

PHỤ LỤC IV: GIÁ TRỊ TỚI HẠN CỦA PHÂN BỐ "KHI BÌNH PHƯƠNG"



Bậc tự do	$\chi^2_{0,995}$	$\chi^2_{0,99}$	$\chi^2_{0,97}$	$\chi^2_{0,95}$	$\chi^2_{0,05}$	$\chi^2_{0,025}$	$\chi^2_{0,01}$	$\chi^2_{0,005}$
1	0,000	0,000	0,001	0,004	3,841	5,024	6,635	7,879
2	0,010	0,020	0,051	0,103	5,991	7,378	9,210	10,597
3	0,072	0,115	0,216	0,352	7,815	9,348	11,345	12,838
4	0,207	0,297	0,484	0,711	9,488	11,143	13,277	14,860
5	0,412	0,554	0,831	1,145	11,070	12,832	15,086	16,750
6	0,676	0,872	1,237	1,635	12,592	14,449	16,812	18,548
7	0,989	1,239	1,690	2,167	14,067	16,013	18,475	20,278
8	1,344	1,646	2,180	2,733	15,507	17,535	20,090	21,955
9	1,735	2,088	2,700	3,325	16,919	19,023	21,666	23,589
10	2,156	2,558	3,247	3,940	18,307	20,483	23,209	25,188
11	2,603	3,053	3,816	4,575	19,675	21,920	24,725	26,757
12	3,074	3,571	4,404	5,226	21,026	23,337	26,217	28,300
13	3,565	4,107	5,009	5,982	22,362	24,736	27,688	28,819
14	4,075	4,660	5,629	6,571	23,685	26,119	29,141	31,319
15	5,001	5,229	6,262	7,261	24,996	27,488	30,578	32,801
16	5,142	5,812	6,908	7,962	26,296	28,845	32,000	34,267
17	5,697	6,408	7,564	8,672	27,587	30,191	33,409	35,718
18	6,265	7,015	8,231	9,390	28,869	31,524	34,805	37,156
19	6,844	7,633	8,907	10,117	30,144	32,852	36,191	38,582
20	7,343	8,260	9,591	10,851	31,410	34,170	37,566	39,997
21	8,034	8,897	10,283	11,591	32,671	35,479	38,932	41,401
22	8,543	9,542	10,982	12,388	33,924	36,781	30,289	42,796
23	9,260	10,196	11,689	13,091	35,172	38,076	41,638	44,181
24	9,886	10,856	12,401	13,848	36,415	39,364	42,980	45,558
25	10,520	11,524	13,120	14,611	37,625	40,646	44,314	46,928
26	11,160	12,198	13,844	15,379	38,885	41,923	45,642	48,290
27	11,808	12,879	14,573	16,151	40,113	43,194	46,993	46,645
28	12,461	13,565	15,308	16,928	41,337	44,461	48,278	50,993
29	13,121	14,256	16,047	17,708	42,557	45,722	49,588	52,336
30	13,787	14,930	16,791	18,493	43,773	46,979	50,892	53,672

PHỤ LỤC V: GIÁ TRỊ HÀM KHỐI LƯỢNG XÁC SUẤT POISSON

$$P\left\{X=k\right\}=\frac{e^{-\lambda}\lambda^{k}}{k!}$$

k \ λ	0,1	0,2	0,3	0,4	0,5	0,6
0	0,904837	0,818731	0,740818	0,670320	0,606531	0,548812
1	0,090484	0,163746	0,222245	0,268128	0,303265	0,329287
2	0,004524	0,016375	0,033337	0,053626	0,075817	0,098786
3	0,000151	0,001091	0,003334	0,007150	0,012636	0,019757
4	0,000004	0,000055	0,000250	0,000715	0,001580	0,002964
5		0,000002	0,000015	0,000057	0,000158	0,000356
6			0,000001	0,000004	0,000013	0,000035
7					0,000001	0,000003

k \ λ	0,7	0,8	0,9	1,0	2,0	3,0
0	0,496585	0,449329	0,406570	0,367877	0,135335	0,049787
1	0,347610	0,359463	0,365913	0,367789	0,270671	0,149361
2	0,121663	0,143785	0,164661	0,183940	0,270671	0,224042
3	0,028388	0,038343	0,049398	0,061313	0,180447	0,224042
4	0,004968	0,007669	0,011115	0,015328	0,090224	0,168031
5	0,000695	0,001227	0,002001	0,003066	0,036089	0,100819
6	0,000081	0,000164	0,000300	0,000511	0,012030	0,050409
7	0,000008	0,000019	0,000039	0,000073	0,003437	0,021604
8		0,000002	0,000004	0,000009	0,000859	0,008101
9				0,000001	0,000191	0,002701
10					0,000038	0,000810
11					0,000007	0,000221
12					0,000001	0,000055
13						0,000013
14						0,000003
15						0,000001

$k \backslash \lambda$	4,0	5,0	6,0	7,0	8,0	9,0
0	0,018316	0,006738	0,002479	0,000912	0,000335	0,000123
1	0,073263	0,033690	0,014873	0,006383	0,002684	0,001111
2	0,146525	0,084224	0,044618	0,022341	0,010735	0,004998
3	0,195367	0,140374	0,089235	0,052129	0,028626	0,014994
4	0,195367	0,175467	0,133853	0,191226	0,057252	0,033737
5	0,156293	0,175467	0,160623	0,127717	0,091604	0,060727
6	0,104194	0,146223	0,160623	0,149003	0,122138	0,091090
7	0,059540	0,104445	0,137677	0,149003	0,139587	0,117116
8	0,029770	0,065278	0,103258	0,130377	0,139587	0,131756
9	0,013231	0,036266	0,068838	0,011405	0,120477	0,131756
10	0,005292	0,018133	0,041303	0,070983	0,099262	0,118580
11	0,001925	0,008242	0,022529	0,045171	0,072190	0,097020
12	0,000642	0,003434	0,011262	0,026350	0,048127	0,072765
13	0,000197	0,001321	0,005199	0,014188	0,029616	0,050376
14	0,000056	0,000472	0,002228	0,007094	0,013924	0,032384
15	0,000015	0,000157	0,000891	0,003311	0,009026	0,019431
16	0,000004	0,000049	0,000334	0,001448	0,004513	0,010930
17	0,000001	0,000014	0,000118	0,000596	0,002124	0,005786
18		0,000004	0,000039	0,000232	0,000944	0,002893
19		0,000001	0,000012	0,000085	0,000397	0,001370
20			0,000004	0,000030	0,000159	0,000617
21			0,000001	0,000010	0,000061	0,000264
22				0,000003	0,000022	0,000108
23				0,000001	0,000008	0,000042
24					0,000003	0,000016
25					0,000001	0,000006
26						0,000002
27						0,000001

PHỤ LỤC VI: GIÁ TRỊ HÀM PHÂN BỐ POISSON

$$P\{X \le k\} = \sum_{i=0}^{k} \frac{e^{-\lambda}\lambda^{i}}{i!}$$

k \ λ	0,1	0,2	0,3	0,4	0,5	0,6
0	0,904837	0,818731	0,740818	0,670320	0,606531	0,548812
1	0,995321	0,982477	0,963063	0,938448	0,909796	0,878099
2	0,999845	0,998853	0,996400	0,992074	0,985612	0,976885
3	0,999996	0,999943	0,999734	0,999224	0,998248	0,996642
4	1,000000	0,999998	0,999984	0,999939	0,999828	0,999606
5		1,000000	0,999999	0,999996	0,999986	0,999962
6				1,000000	0,999999	0,999997
7						1,000000

k \ λ	0,7	0,8	0,9	1,0	2,0	3,0
0	0,496585	0,449329	0,406570	0,367877	0,135335	0,049787
1	0,844195	0,808792	0,772483	0,735759	0,406006	0,199148
2	0,965858	0,952577	0,937144	0,919699	0,676677	0,423190
3	0,994246	0,990920	0,986542	0,981012	0,857124	0,647232
4	0,999214	0,998589	0,997657	0,996340	0,947348	0,815263
5	0,999909	0,999816	0,999658	0,999403	0,983437	0,916082
6	0,999990	0,999980	0,999958	0,999917	0,995467	0,966491
7	0,999999	0,999998	0,999997	0,999990	0,998904	0,988095
8			1,000000	0,999999	0,999753	0,996196
9				1,000000	0,999954	0,998897
10					0,999992	0,999707
11					0,999999	0,999928
12					1,000000	0,999983
13						0,999996
14						0,999999
15						1,000000

k \ λ	4,0	5,0	6,0	7,0	8,0	9,0
0	0,018316	0,006738	0,002479	0,000912	0,000335	0,000123
1	0,091579	0,040428	0,017352	0,007295	0,003019	0,001234
2	0,238105	0,124652	0,061970	0,029636	0,013754	0,006232
3	0,433472	0,265026	0,151205	0,081765	0,042380	0,021228
4	0,785132	0,615960	0,445681	0,300708	0,191236	0,115690
5	0,889326	0,762183	0,606304	0,449711	0,313374	0,206780
6	0,948866	0,866628	0,743981	0,598711	0,452961	0,323896
7	0,978636	0,931806	0,847239	0,729091	0,592548	0,455652
8	0,991867	0,968172	0,916077	0,830496	0,716625	0,587408
9	0,997159	0,986305	0,957380	0,901479	0,815887	0,705988
10	0,999084	0,984547	0,979909	0,946650	0,888077	0,803008
11	0,999726	0,997981	0,991173	0,973000	0,936204	0,875773
12	0,999923	0,999202	0,996372	0,987188	0,965820	0,926149
13	0,999979	0,999774	0,998600	0,994282	0,982744	0,958533
14	0,999994	0,999931	0,999491	0,997593	0,991770	0,977964
15	0,999998	0,999980	0,999825	0,999041	0,996283	0,988894
16	0,999999	0,999994	0,999943	0,999637	0,998407	0,994680
17	0,999999	0,999998	0,999982	0,999869	0,999351	0,997573
18	0,999999	0,999999	0,999994	0,999955	0,999748	0,998943
19	1,000000	0,999999	0,999998	0,999985	0,999907	0,999560
20		1,000000	0,999999	0,999995	0,999967	0,999824
21			0,999999	0,999998	0,999989	0,999932
22			1,000000	0,999999	0,999997	0,999974
23				0,999999	0,999998	0,999990
24				1,000000	0,999999	0,999996
25					1,000000	0,999998
26						0,999999
27						1,000000

BẢNG CHỈ DẪN THUẬT NGỮ

TÀI LIỆU THAM KHẢO

[1]. Lê Bá Long, *Giáo trình Xác suất và thống kê*, NXB Thông tin và truyền thông, 2009.

[2]. Đào Hữu Hồ, *Xác suất Thống kê*, NXB Đại Học Quốc Gia Hà Nội, 1999.

[3]. Nguyễn Cao Văn và Trần Thái Ninh, *Bài giảng xác suất và thống kê toán*, NXB Thống kê, Hà Nội 1999.

[4]. Nguyễn Cao Văn, Trần Thái Ninh và Nguyễn Thế Hệ, *Bài tập lý thuyết xác suất và thống kê toán*, NXB Giáo dục, Hà Nội 2002.

[5]. Nguyễn Văn Phấn, Lương Hữu Thanh, *Bài tập xác suất và thống kê*, Đại Học Giao Thông Vận Tải, 1996.

[6]. Tống Đình Quỳ, *Hướng dẫn giải bài tập xác suất thống kê*, NXB Đại Học Quốc Gia Hà Nội, 2004.

[7]. Đặng Hùng Thắng, *Mở đầu về lý thuyết xác suất và các ứng dụng*, NXB Giáo dục, 1997.

[8]. Đặng Hùng Thắng, *Thống kê và ứng dụng*, NXB Giáo dục,1999.

[9]. Nguyễn Duy Tiến, Vũ Việt Yên, *Lý thuyết xác suất*, NXB Giáo dục, 2000.

[10]. Trần Mạnh Tuấn, *Xác suất và Thống kê, lý thuyết và thực hành tính toán*, NXB Đại Học Quốc Gia Hà Nội, 2004.

[11]. Nguyễn Bác Văn, *Xác suất và xử lí số liệu thống kê*, NXB Giáo dục,1996.

[12]. Harald Cramer, *Phương pháp toán học trong thống kê*, NXB Khoa học và Kỉ thuật, Hà Nội 1970.

[13]. Prasad Chalasani & Somesh Jha, *Stochastic Calculus and Finance*, Steven E. Shreve, 1996.

[14]. George G. Roussas, *A Course in Mathematical Statistics*, ACADEMIC PRESS USA , 1997.

[15]. Murray R. Spiegel, John Schiller, R. Alu Srinivasan; *Probability and Slatistics*, Schaum's outline Series. Mc Graw Hill, 2000.