HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG

SÁCH HƯỚNG DẪN HỌC TẬP

LÝ THUYẾT XÁC SUẤT VÀ THỐNG KÊ TOÁN

Biên soạn : Ts. LÊ BÁ LONG

Lưu hành nội bộ

HÀ NỘI - 2006

LỜI NÓI ĐẦU

Lý thuyết xác suất thống kê là một bộ phận của toán học, nghiên cứu các hiện tượng ngẫu nhiên và ứng dụng chúng vào thực tế. Ta có thể hiểu hiện tượng ngẫu nhiên là hiện tượng không thể nói trước nó xảy ra hay không xảy ra khi thực hiện một lần quan sát. 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 cũng là cơ sở để nghiên cứu Thống kê - môn học nghiên cứu các các phương pháp thu thập thông tin chọn mẫu, xử lý thông tin, nhằm rút ra các kết luận hoặc 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.

Có nhiều sách giáo khoa và tài liệu chuyên khảo viết về lý thuyết xác suất thống kê. Tuy nhiên, với phương thức đào tạo từ xa có những đặc thù riêng, đòi hỏi học viên phải làm việc độc lập nhiều hơn, vì vậy cần phải có tài liệu hướng dẫn học tập của từng môn học thích hợp cho đối tượng này. Tập tài liệu "Hướng dẫn học môn toán xác suất thống kê" này được biên soạn cũng nhằm mục đích trên.

Tập tài liệu "Hướng dẫn học môn Lý thuyết xác suất và thống kê toán" được biên soạn theo chương trình qui định của Học viện Công nghệ Bưu Chính Viễn Thông dành cho hệ đại học chuyên ngành Quản trị kinh doanh. 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.

Giáo trình gồm 8 chương tương ứng với 4 đơn vị học trình (60 tiết):

Chương I: Biến cố ngẫu nhiên và xác suất.

Chương II: Biến ngẫu nhiên và quy luật phân bố xác suất.

Chương III: Một số quy luật phân bố xác suất quan trọng.

Chương IV: Biến ngẫu nhiên hai chiều.

Chương V: Luật số lớn.

Chương VI: Cơ sở lý thuyết mẫu.

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

Chương VIII: Kiểm định giả thiết thống kê.

Năm 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 đại số và giải tích trong chương trình toán đại cương. 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, chứ không có điều kiện để chứng minh chi tiết.

Giáo trình này được trình bày theo phương pháp phù hợp đối với người tự học, đặc biệt phục vụ đắc lực cho công tác đào tạo từ xa. 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.

Hầu hết các bài toán trong giáo trình được xây dựng theo lược đồ: đặt bài toán, chứng minh sự tồn tại lời giải bằng lý thuyết và cuối cùng nêu thuật toán giải quyết bài toán này. 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 các chương có phần tóm tắt các nội dung chính, và cuối cùng là các câu hỏi luyện tập. 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.

Giáo trình được viết theo đúng đề cương chi tiết môn học đã được Học Viện ban hành. Các kiến thức được trang bị tương đối đầy đủ, có hệ thống. Tuy nhiên, nếu người học không có điều kiện đọc kỹ toàn bộ giáo trình thì các nội dung có đánh dấu (*) được coi là phần tham khảo thêm (chẳng hạn: chương 5 luật số lớn và định lý giới hạn trung tâm (*), mục 6.6 chương 6 …).

Tuy tác giả đã rất cố gắng, song do thời gian bị hạn hẹp, nên các thiếu sót còn tồn tại trong giáo trình 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 TS Tô Văn Ban, CN Nguyễn Đình Thực, đã đọc bản thảo và cho những ý kiến phản biện quý giá và đặc biệt tới KS Nguyễn Chí Thành người đã giúp tôi biên tập hoàn chỉnh cuốn tài liệu.

Cuối cùng, tác giả xin bày tỏ sự cám ơn đối với Ban Giám đốc Học viện Công nghệ Bưu Chính Viễn Thông, Trung tâm Đào tạo Bưu Chính Viễn Thông 1 và 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, đầu năm 2006.
TÁC GIẢ

CHƯƠNG I: BIẾN CỐ NGẪU NHIÊN VÀ XÁC SUẤT

GIỚI THIỆU

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 ta biết chắc chắn rằng lông của quạ có mầu đen, một vật được thả từ trên cao chắc chắn sẽ rơi xuống đất... 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 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… 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 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 nắm vững các kiến thức về đại số tập hợp như: hợp, giao tập hợp, 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 có thể. Vì vậy học viên cần nắm vững các phương pháp đếm - giải tích tổ hợp (đã được học ở lớp 12). Tuy nhiên để thuận lợi cho người học chúng tôi sẽ nhắc lại các kết quả chính trong mục 3.

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.

NỘI DUNG

1.1. PHÉP THỬ VÀ BIẾN CỐ

1.1.1. Phép thử (Experiment)

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 $\mathscr{C}$. Tuy không biết kết quả sẽ xảy ra như thế nào, nhưng ta có thể liệt kê được hoặc biểu diễn tất cả các kết quả của phép thử $\mathscr{C}$. Mỗi kết quả của phép thử $\mathscr{C}$ được gọi là một biến cố sơ cấp. Tập hợp tất cả các biến cố sơ cấp của phép thử được gọi là không gian mẫu, ký hiệu Ω.

Ví dụ 1.1:

▪ Phép thử tung đồng xu có không gian mẫu là $\Omega = \{S, N\}$.

▪ Với phép thử tung xúc xắc, các biến cố sơ cấp có thể xem là số các nốt trên mỗi mặt xuất hiện. Vậy $\Omega = \{1, 2, 3, 4, 5, 6\}$.

▪ 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 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.

1.2.1. Biến cố (Event)

Với phép thử $\mathscr{C}$ ta thường 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 $\mathscr{C}$. Các biến cố ngẫu nhiên được ký hiệu bằng các chữ in hoa A, B, C, … Mỗi kết quả ω của $\mathscr{C}$ được gọi là kết quả thuận lợi cho biến cố A nếu A xảy ra khi kết quả của $\mathscr{C}$ là ω.

Ví dụ 1.2: Nếu gọi A là biến cố số nốt xuất hiện là chẵn trong phép thử tung xúc xắc ở ví dụ 1.1 thì A có các kết quả thuận lợi là 2, 4, 6.

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); (N,S).

Như vậy mỗi biến cố A được đồng nhất với một tập con của không gian mẫu Ω bao gồm các kết quả thuận lợi đối với A.

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, biến cố này trùng với không gian mẫu Ω.

- *Biến cố không thể*: là biến cố nhất định không xảy ra khi thực hiện phép thử. Biến cố không thể được ký hiệu ϕ.

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

1.2. ĐỊNH NGHĨA XÁC SUẤT

Việc 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.

Dựa vào bản chất của phép thử (đồng khả năng) ta có thể suy luận về khả năng xuất hiện của biến cố, với cách tiếp cận này ta có *định nghĩa xác suất theo phương pháp cổ điển*.

Khi thực hiện nhiều lần lặp lại độc lập một phép thử ta có thể tính được tần suất xuất hiện (số lần xuất hiện) của một biến cố nào đó. Tần suất thể hiện khả năng xuất hiện của biến cố, với cách tiếp cận này ta có *định nghĩa xác suất theo thống kê*.

1.3. ĐỊNH NGHĨA CỔ ĐIỂN VỀ XÁC SUẤT

1.3.1. Định nghĩa và ví dụ

Giả sử phép thử $\mathscr{C}$ 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 ta định nghĩa xác suất của biến cố A là

$$\mathrm{P(A)} = \frac{\text{số trường hợp thuận lợi đối với A}}{\text{số trường hợp có thể}} \tag{1.1}$$

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

$$P(A) = \frac{\text{số phần tử của } A}{\text{số phần tử của } \Omega} = \frac{|A|}{|\Omega|} \tag{1.1'}$$

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

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

1.3.2. Các qui tắc đếm

1.3.2.1. 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.

1.3.2.2. Qui tắc nhân

Giả sử công việc H gồm nhiều công đoạn liên tiếp H_1, H_2, ..., H_k và mỗi công đoạn H_i có n_i cách thực hiện thì có tất cả $n_1 \times n_2 \times \cdots \times n_k$ cách thực hiện công việc H.

1.3.2.3. Hoán vị

Mỗi phép đổi chỗ của n phần tử 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.

1.3.2.4. Chỉnh hợp

Chọn lần lượt k 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:

$$A_n^k = \frac{n!}{(n-k)!} \tag{1.2}$$

1.3.2.5. 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 n chập k 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 của n phần tử 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

$$C_n^k = \frac{A_n^k}{k!} = \frac{n!}{k!(n-k)!} \tag{1.3}$$

Ví dụ 1.4: Tung một con xúc xắc hai lần. Tìm xác suất để trong đó có 1 lần ra 6 nốt.

Giải: Số các trường hợp có thể là 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, nghĩa là 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 xác suất để chỉ có một lần ra mặt 6 khi tung xúc xắc 2 lần là $\frac{10}{36}$.

Ví dụ 1.5: 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 10 chập 2. 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. Do đó $P(A) = \frac{1}{90}$.

Ví dụ 1.6: 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 $|\Omega| = C_6^2 = 15$.

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

b. Có $C_4^2 = 6$ cách chọn 2 trong 4 nữ, vậy xác suất tương ứng $P = 6/15$.

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 $P = 14/15$.

1.4. ĐỊNH NGHĨA THỐNG KÊ VỀ XÁC SUẤT

Định nghĩa xác suất theo cổ điển trực quan, dễ hiểu. Tuy nhiên khi số các kết quả có thể vô hạn hoặc không đồng khả năng thì cách tính xác suất cổ điển không áp dụng được.

Giả sử phép thử $\mathscr{C}$ 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 $\mathscr{C}$, biến cố A xuất hiện $k_n(A)$ lần 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) 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 $P(A)$.

$$P(A)=\lim_{n\to\infty} f_n(A) \quad \textbf{(1.4)}$$

Trên thực tế $P(A)$ được tính xấp xỉ bởi tần suất $f_n(A)$ khi n đủ lớn.

Ví dụ 1.7: Một công ty bảo hiểm muốn xác định xác suất để một người Mỹ 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 0,008.

Ví dụ 1.8: 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: Định nghĩa xác suất theo thống kê khắc phục được hạn chế của định nghĩa cổ điển, nó hoàn toàn dựa trên các thí nghiệm 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.

1.5. QUAN HỆ GIỮA CÁC BIẾN CỐ

Trong lý thuyết xác suất người ta xét các quan hệ sau đây cho các biến cố.

1.5.1. Quan hệ kéo theo

Biến cố A kéo theo biến cố B, ký hiệu $A\subset B$, nếu A xảy ra thì B xảy ra.

1.5.2. Quan hệ biến cố đối

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

1.5.3. Tổng của hai biến cố

Tổng của hai biến cố A, B là biến cố được ký hiệu $A\cup B$ (hoặc $A+B$). Biến cố $A\cup B$ 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 $\bigcup_{i=1}^{n} A_i$. Biến cố này xảy ra khi có ít nhất một trong các biến cố A_i xảy ra.

1.5.4. Tích của hai biến cố

Tích của hai biến cố A, B là biến cố được ký hiệu AB. Biến cố AB xảy ra khi và chỉ khi cả hai biến cố A, B cùng xảy ra.

Tích của một dãy các biến cố $\{A_1, A_2, ..., A_n\}$ là biến cố $\prod_{i=1}^{n} A_i$. Biến cố này xảy ra khi tất cả các biến cố A_i cùng xảy ra.

1.5.5. Biến cố xung khắc

Hai biến số A, B gọi là xung khắc nếu biến cố tích AB là biến cố không thể. Nghĩa là hai biến cố này không thể đồng thời xảy ra.

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 tích các biến cố có tính giao hoán, kết hợp, tổng phân bố đối với tích, tích phân bố đối với tổng, luật De Morgan …

1.5.6. 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:

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

(ii) Tổng của chúng là biến cố chắc chắn, nghĩa là $\bigcup_{i=1}^{n} A_i = \Omega$.

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

Ví dụ 1.9: 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.

1.5.7. 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.

Tổng quát hơn 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 $1 \leq k \leq n$, không làm ảnh hưởng tới việc xảy ra hay không xảy ra của các biến cố còn lại.

Ví dụ 1.10: Ba xạ thủ $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$ mỗi người bắn một viên đạn vào mục tiêu. Gọi A, B, C lần lượt là biến cố $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$ bắn trúng mục tiêu.

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

ii. 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 $\mathcal{C}$ bắn trúng.
- G: Chỉ có 1 xạ thủ bắn trúng.

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

Giải:

i. ABC : cả 3 đều bắn trúng. $\overline{A}\,\overline{B}\,\overline{C}$: cả 3 đều bắn trượt. $A \cup B \cup C$: có ít nhất 1 người bắn trúng.

ii. $D = AB \cup BC \cup CA$.

Có nhiều nhất một xạ thủ bắn trúng có nghĩa là có ít nhất hai xạ thủ bắn trượt, vậy:

$$E = \overline{A}\,\overline{B} \cup \overline{B}\,\overline{C} \cup \overline{C}\,\overline{A}.$$

$$F = \overline{A}\,\overline{B}C. \quad G = A\overline{B}\,\overline{C} \cup \overline{A}B\overline{C} \cup \overline{A}\,\overline{B}C.$$

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

1.6. CÁC TÍNH CHẤT VÀ ĐỊNH LÝ XÁC SUẤT

1.6.1. Các tính chất của xác suất

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

1. Với mọi biến cố A:

$$0 \le P(A) \le 1. \tag{1.5}$$

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

$$P(\phi) = 0,\ P(\Omega) = 1 \tag{1.6}$$

1.6.2. Qui tắc cộng

1.6.2.1. Trường hợp xung khắc

Nếu A, B là hai biến cố xung khắc thì

$$P(A \cup B) = P(A) + P(B). \tag{1.7}$$

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

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i)\,. \qquad \textbf{(1.7)'}$$

Từ công thức (1.6) và (1.7)' ta có hệ quả: Nếu $\{A_1, A_2, ..., A_n\}$ là một hệ đầy đủ thì

$$\sum_{i=1}^{n} P(A_i) = 1 \qquad \textbf{(1.8)}$$

1.6.2.2. Trường hợp tổng quát

- Nếu A, B là hai biến cố bất kỳ thì

$$P(A \cup B) = P(A) + P(B) - P(AB) \qquad \textbf{(1.9)}$$

- Nếu A, B, C là ba biến cố bất kỳ thì

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(AB) - P(BC) - P(CA) + P(ABC) \quad \textbf{(1.9)'}$$

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

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) - \sum_{i<j} P(A_i A_j) + \sum_{i<j<k} P(A_i A_j A_k) - \cdots + (-1)^{n-1} P(A_1 A_2 ... A_n)\,. \quad \textbf{1.9)''}$$

Ví dụ 11: 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. Vậy $A = A_1 \cup A_2$.

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

Áp dụng công thức (1.8) cho hệ đầy đủ $\{A, \overline{A}\}$ ta được quy tắc xác suất biến cố đối

1.6.3. Quy tắc xác suất của biến cố đối

Với mọi biến cố A:

$$P(\overline{A}) = 1 - P(A)\,. \qquad \textbf{(1.10)}$$

1.6.4. 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

- Nếu $P(A)>0$ thì:

$$P(B\,|\,A)=\frac{P(AB)}{P(A)}. \qquad \textbf{(1.11)}$$

- 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 (1.5)-(1.10)") đối với biến cố B.

Chẳng hạn:

$$P(\overline{B}\,|\,A)=1-P(B\,|\,A),\ P(B_1\cup B_2\,|\,A)=P(B_1|\,A)+P(B_2|\,A)-P(B_1B_2\,|\,A).$$

Ví dụ 12: Gieo đồng thời hai con xúc xắc cân đối. Tính xác suất để tổng số nốt xuất hiện trên hai con xúc xắc ≥ 10 biết rằng ít nhất một con đã ra nốt 5.

Giải: Gọi A là biến cố " ít nhất một con ra nốt 5".

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

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

Biến cố AB có 3 kết quả thuận lợi là (5,6), (6,5), (5,5).

Vậy $P(AB)=\dfrac{3}{36} \Rightarrow P\left(B|A\right)=\dfrac{3}{36}\Big/\dfrac{11}{36}=\dfrac{3}{11}$.

1.6.5. Quy tắc nhân

1.6.5.1. Trường hợp độc lập

- Nếu A, B là hai biến cố độc lập thì

$$P(AB)=P(A)P(B). \qquad \textbf{(1.12)}$$

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

$$P(A_1A_2...A_n)=P(A_1)P(A_2)...P(A_n). \qquad \textbf{(1.13)}$$

1.6.5.2. Trường hợp tổng quát

- $P(AB)=P(A)P(B\,|\,A)$ **(1.14)**

- $P\left(A_1A_2...A_n\right)=P\left(A_1\right)P\left(A_2|\,A_1\right)P\left(A_3|\,A_1A_2\right)...P\left(A_n|\,A_1A_2...A_{n-1}\right).$ **(1.15)**

Ví dụ 1.14:

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 độc lập với các biến cố B_t, $B_đ$, B_x. Vậy xác suất để 2 bi được rút cùng mầu là:

$$P\left(A_tB_t \cup A_đB_đ \cup A_xB_x\right) = P\left(A_tB_t\right) + P\left(A_đB_đ\right) + P\left(A_xB_x\right) \text{(do xung khắc)}$$

$$= P\left(A_t\right)P\left(B_t\right) + P\left(A_đ\right)P\left(B_đ\right) + P\left(A_x\right)P\left(B_x\right) \text{(do độc lập)}$$

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

Ví dụ 1.15: 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 mở được kho ở lần thứ ba.

Giải: Ký hiệu A_i là biến cố "thử đúng chìa ở lần thứ i". Vậy xác suất cần tìm là

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

1.6.6. Công thức xác suất đầy đủ

Định lý 1.3: Nếu $\{A_1, A_2, ..., A_n\}$ là một hệ đầy đủ các biến cố. Với mọi biến cố B (trong cùng 1 phép thử) ta có

$$P(B) = \sum_{i=1}^{n} P(A_i)P\left(B \mid A_i\right) \qquad \textbf{(1.16)}$$

1.6.7. Công thức Bayes

Định lý 1.4: Nếu $\{A_1, A_2, ..., A_n\}$ là một hệ đầy đủ các biến cố. Với mọi biến cố B (trong cùng 1 phép thử) sao cho $P(B) > 0$ ta có :

$$P\left(A_k \mid B\right) = \frac{P(A_kB)}{P(B)} = \frac{P(A_k)P\left(B \mid A_k\right)}{\sum_{i=1}^{n} P(A_i)P\left(B \mid A_i\right)}. \qquad \textbf{(1.17)}$$

Giải thích: 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.

Ví dụ 1.16: 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.

Giải: Gọi là A biến cố "phát tín hiệu A" và B là biến cố "phát tín hiệu B". Khi đó $\{A, B\}$ là hệ đầy đủ. Gọi là T_A biến cố "thu được tín hiệu A" và là T_B biến cố "thu được tín hiệu B".

$$P(A) = 0{,}85,\; P(B) = 0{,}15;\; P(T_B|A) = \frac{1}{7},\; P(T_A|B) = \frac{1}{8}.$$

a. Áp dụng công thức xác suất đầy đủ ta có xác suất thu được tín hiệu A:

$$P(T_A) = P(A)P(T_A|A) + P(B)P(T_A|B) = 0{,}85 \times \frac{6}{7} + 0{,}15 \times \frac{1}{8} = 0{,}7473$$

b. Áp dụng công thức Bayes ta có

$$P(A|T_A) = \frac{P(A)P(T_A|A)}{P(T_A)} = \frac{0{,}85 \times \frac{6}{7}}{0{,}7473} = 0{,}975.$$

Ví dụ 1.17: 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 (biến cố A).

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 H là biến cố “sản phẩm được chọn là phế phẩm”. Theo giả thiết ta có:

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

a. Áp dụng công thức đầy đủ cho hệ đầy đủ $\{H, \overline{H}\}$ ta có:

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

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

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

Ví dụ 1.18: 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 A là biến cố "người được phỏng vấn sẽ mua sản phẩm".

Gọi H_1, H_2, H_3 lần lượt là 3 biến cố tương ứng với 3 cách trả lời của khách hàng được phỏng vấn:

H_1 - người đó trả lời "sẽ mua"

H_2 - người đó trả lời "có thể mua"

H_3 - người đó trả lời "không mua"

H_1, H_2, H_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 $P\left(A\middle|H_1\right)=0{,}7$; $P\left(A\middle|H_2\right)=0{,}3$; $P\left(A\middle|H_3\right)=0{,}01$.

a. Theo công thức xác suất đầy đủ

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

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

b. Theo công thức Bayes

$$P\left(H_1\middle|A\right)=\frac{P(H_1)P\left(A\middle|H_1\right)}{P(A)}=\frac{0{,}17\cdot 0{,}7}{0{,}268}=0{,}444=44{,}4\%.$$

1.7. NGUYÊN LÝ XÁC SUẤT LỚN, XÁC SUẤT NHỎ

Một biến cố không thể có xác suất bằng 0. 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.*

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 tuyên bố 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.

TÓM TẮT

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. Mỗi kết quả của phép thử $\mathscr{C}$ được gọi là một biến cố sơ cấp. Tập hợp tất cả các biến cố sơ cấp của phép thử được gọi là không gian mẫu, ký hiệu Ω.

Biến cố

Mỗi biến cố A được đồng nhất với một tập con của không gian mẫu Ω bao gồm các kết quả thuận lợi đối với A.

Xác suất

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.

Định nghĩa cổ điển về xác suất

Xác suất của biến cố A là $P(A) = \dfrac{\text{số trường hợp thuận lợi đối với } A}{\text{số trường hợp có thể}}$

Định nghĩa thống kê về xác suất

Xác suất của biến cố A là $P(A) \approx f_n(A) = \dfrac{k_n(A)}{n}$

trong đó $k_n(A)$ số lần xuất hiện biến cố A trong n phép thử.

Quan hệ kéo theo

Biến cố A kéo theo biến cố B, ký hiệu $A \subset B$, nếu A xảy ra thì B xảy ra.

Quan hệ biến cố đối

$\overline{A}$ là biến cố đối của A. A xảy ra khi và chỉ khi $\overline{A}$ không xảy ra.

Tổng của hai biến cố

Biến cố $A \cup B$ tổng ($A + B$) của hai biến cố A, B xảy ra khi và chỉ khi có ít nhất A hoặc B xảy ra. Biến cố tổng $\bigcup_{i=1}^{n} A_i$ của một dãy các biến cố $\{A_1, A_2, ..., A_n\}$ xảy ra khi có ít nhất một trong các biến cố A_i xảy ra.

Tích của hai biến cố

Biến cố AB của hai biến cố A, B xảy ra khi và chỉ khi cả hai biến cố A, B cùng xảy ra.

Biến cố tích $\prod_{i=1}^{n} A_i$ của dãy các biến cố $\{A_1, A_2, ..., A_n\}$ xảy ra khi tất cả các biến cố A_i cùng xảy ra.

Biến cố xung khắc

Hai biến số A, B gọi là xung khắc nếu AB là biến cố không thể.

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 chúng xung khắc từng đôi một và tổng của chúng là biến cố chắc chắc.

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.

Tổng quát 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 các biến cố còn lại.

Qui tắc cộng

Trường hợp xung khắc: $P(A \cup B) = P(A) + P(B)$.

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

Trường hợp tổng quát: $P(A \cup B) = P(A) + P(B) - P(AB)$

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(AB) - P(BC) - P(CA) + P(ABC)$$

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

Quy tắc xác suất của biến cố đối

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

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 $P(B \mid A)$.

Quy tắc nhân

Trường hợp độc lập:

$$P(AB) = P(A)P(B); \quad P(A_1 A_2 \dots A_n) = P(A_1)P(A_2) \dots P(A_n).$$

Trường hợp không độc lập:

$$P(AB) = P(A)P(B \mid A);$$

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

Công thức xác suất đầy đủ

Giả sử $\{A_1, A_2, ..., A_n\}$ là một hệ đầy đủ. Với mọi biến cố B ta có:

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

Công thức Bayes

Nếu $\{A_1, A_2, ..., A_n\}$ là một hệ đầy đủ và với mọi biến cố B sao cho $P(B) > 0$ ta có :

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

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.

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.

CÂU HỎI ÔN TẬP VÀ BÀI TẬP

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

Đúng ☐ Sai ☐

1.2 Các biến cố A và $\overline{A}\cup\overline{B}$ là 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 Thông tin liên quan đến việc xuất hiện biến cố B làm tăng xác suất của biến cố A, tức là $P(A|B)\ge P(A)$?

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 2 xác suất.

Đúng ☐ Sai ☐

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

Đúng ☐ Sai ☐

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

Đúng ☐ Sai ☐

1.11 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.

1.12 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.

1.13 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.

1.14 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 ($k = \overline{1,10}$) là biến cố chỉ sản phẩm kiểm tra thứ k thuộc loại xấu. Biểu diễn các biến cố sau theo A_k:

a) Cả 10 sản phẩm đều xấu.

b) Có ít nhất một sản phẩm xấu.

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 xấu.

1.15 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.

1.16 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 là phế phẩm.

1.17 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.

1.18 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.

1.19 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.

1.20 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 được thử thì không thử lại. Tính xác suất anh ta mở được cửa ở lần thử thứ 4.

1.21 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.

1.22 Một nhà máy ô tô có ba phân xưởng I, II, III cùng sản xuất ra một loại pít-tông. 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.

1.23 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.

1.24 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.

1.25 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.

1.26 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 II: BIẾN NGẪU NHIÊN VÀ LUẬT PHÂN BỐ XÁC SUẤT

GIỚI THIỆU

Trong chương này ta khảo sát các biến cố gắn với các giá trị nào đó, khi các giá trị này thay đổi ta được các 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 nào đó với xác suất bao nhiêu. Nói cách khác biến ngẫu nhiên X có thể được khảo sát thông qua hàm phân bố xác suất của nó $F(x)=P\{X<x\}$. Như vậy khi ta biết qui luật phân bố xác suất của một biến ngẫu nhiên thì ta đã nắm được toàn bộ thông tin về biến ngẫu nhiên này.

Khi biến ngẫu nhiên chỉ nhận các giá trị rời rạc thì hàm phân bố xác suất hoàn toàn được xác định bởi bảng phân bố xác suất, đó là bảng ghi các giá trị mà biến ngẫu nhiên nhận với xác suất tương ứng. Khi biến ngẫu nhiên nhận giá trị liên tục thì hàm phân bố xác suất được xác định bởi hàm mật độ xác suất.

Các biến ngẫu nhiên đặc biệt thường gặp sẽ được xét trong chương sau.

Ngoài phương pháp sử dụng hàm phân bố để xác định biến ngẫu nhiên, trong nhiều trường hợp bài toán chỉ đòi hỏi cần 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 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.

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 tích phân xác định.

NỘI DUNG

2.1. ĐỊNH NGHĨA VÀ PHÂN LOẠI BIẾN NGẪU NHIÊN

2.1.1. Khái niệm 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.*

Người ta thường ký hiệu các biến ngẫu nhiên bằng các chữ in hoa $X, Y, Z, \ldots$ và các chữ thường ký hiệu các trị số của chúng. Vì vậy với biến ngẫu nhiên X và với mọi giá trị thực $x \in \mathbb{R}$ thì $\{X < x\}$ là một biến cố ngẫu nhiên.

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

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

- Số nốt xuất hiện khi gieo một con xúc xắc.
- Tuổi thọ của một thiết bị đang hoạt động.
- Số khách hàng vào một điểm phục vụ trong 1 đơn vị thời gian.
- Số cuộc gọi đến một tổng đài.
- Sai số khi đo lường một đại lượng vật lý …

2.1.2. Phân loại

Người ta phân các biến ngẫu nhiên thành hai loại:

❖ ***Biến ngẫu nhiên rời rạc*** nếu nó chỉ nhận một số hữu hạn hoặc vô hạn đếm được các giá trị. Nghĩa là có thể liệt kê các giá trị thành một dãy $x_1, x_2, \ldots$.

❖ ***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 và xác suất $P\{X = a\}$ bằng không với mọi a.

Ví dụ 2.2:

• Gọi X là số nốt xuất hiện khi gieo một con xúc xắc 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 1 đơn vị thời gian, Z là biến ngẫu nhiên rời rạc nhận các giá trị $0, 1, 2, \ldots$

• Số cuộc gọi đến một tổng đài là biến ngẫu nhiên rời rạc nhận các giá trị $0, 1, 2, \ldots$

• 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. QUY LUẬT PHÂN BỐ XÁC SUẤT CỦA BIẾN NGẪU NHIÊN

Biến ngẫu nhiên nhận các giá trị nào đó phụ thuộc vào yếu tố ngẫu nhiên vì vậy có thể sử dụng các phương pháp sau để xác định luật phân bố xác suất của biến ngẫu nhiên.

2.2.1. Hàm phân bố xác suất

Định nghĩa 2.2: Hàm phân bố xác suất (cumulative distribution function, viết tắt CDF) của biến ngẫu nhiên X là hàm số $F(x)$ xác định với mọi $x \in \mathbb{R}$ bởi công thức:

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

Hàm phân bố có các tính chất sau:

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

b. $F(x)$ là hàm không giảm, liên tục bên trái. Nếu X là biến ngẫu nhiên liên tục thì $F(x)$ là hàm liên tục.

c. $F(-\infty) = \lim\limits_{x \to -\infty} F(x) = 0$; $F(+\infty) = \lim\limits_{x \to +\infty} F(x) = 1$, **(2.3)**

d. $P\{a \le X < b\} = F(b) - F(a)$. **(2.4)**

2.2.2. Bảng phân bố xác suất của biến ngẫu nhiên rời rạc

Giả sử biến ngẫu nhiên X chỉ nhận các giá trị $x_1, x_2, \ldots$ với xác suất tương ứng $p_i = P\{X = x_i\}$. $p_i > 0$ và $\sum\limits_i p_i = 1$.

Bảng phân bố xác suất của X có dạng sau:

$$\begin{array}{c|ccc} X & x_1 & x_2 & \cdots \\ \hline P & p_1 & p_2 & \cdots \end{array} \qquad \textbf{(2.5)}$$

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

$$F(x) = \begin{cases} 0 & \text{nếu} \quad x \le x_1 \\ p_1 + p_2 + \cdots + p_{k-1} & \text{nếu} \quad x_{k-1} < x \le x_k,\ \forall k > 1 \end{cases} \qquad \textbf{(2.6)}$$

Đồ thị của $F(x)$ là hàm bậc thang có bước nhảy tại $x_1, x_2, \ldots$

- Nếu X chỉ nhận các giá trị $x_1, x_2, \ldots, x_n$ thì các biến cố

$$\{X = x_1\}, \{X = x_2\}, \ldots, \{X = x_n\} \qquad \textbf{(2.7)}$$

lập thành hệ đầy đủ các biến cố.

Hàm phân bố có dạng:

$$F(x)=\begin{cases} 0 & \text{nếu} \quad x \le x_1 \\ p_1+p_2+\cdots+p_{k-1} & \text{nếu} \quad x_{k-1} < x \le x_k \;, 1<k\le n \\ 1 & \text{nếu} \quad x > x_n \end{cases} \qquad \textbf{(2.8)}$$

Ví dụ 2.3: 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 và hàm phân bố của biến ngẫu nhiên X.

Giải: $P\{X=0\}=\dfrac{C_6^3}{C_{10}^3}=\dfrac{5}{30}$, $P\{X=1\}=\dfrac{C_6^2C_4^1}{C_{10}^3}=\dfrac{15}{30}$, $P\{X=2\}=\dfrac{C_6^1C_4^2}{C_{10}^3}=\dfrac{9}{30}$,

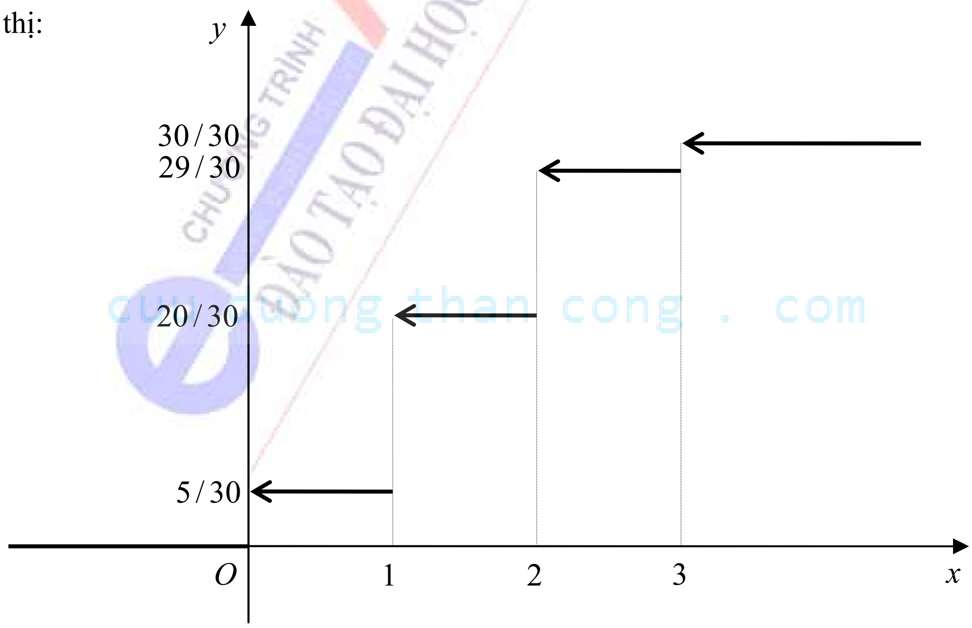
$P\{X=3\}=\dfrac{C_4^3}{C_{10}^3}=\dfrac{1}{30}$.

Bảng phân bố xác suất:

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

Hàm phân bố:
$$F(x)=\begin{cases} 0 & \text{nếu} \quad x\le 0 \\ 5/30 & \text{nếu} \quad 0<x\le 1 \\ 20/30 & \text{nếu} \quad 1<x\le 2 \\ 29/30 & \text{nếu} \quad 2<x\le 3 \\ 1 & \text{nếu} \quad x>3 \end{cases}$$

Đồ thị:



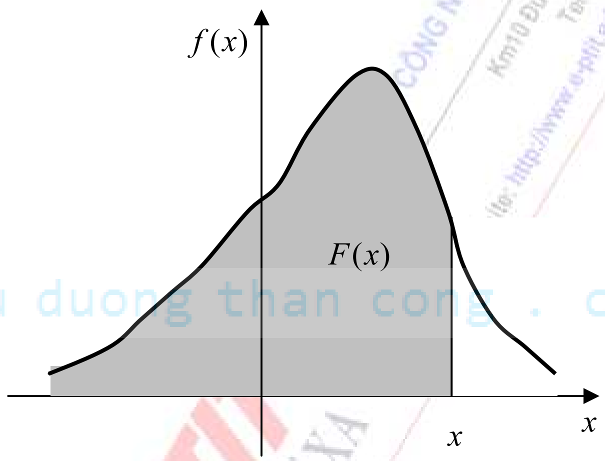
2.2.3. Hàm mật độ phân bố xác suất của biến ngẫu nhiên liên tục

Định nghĩa 2.3: *Giả sử X là một biến ngẫu nhiên liên tục có hàm phân bố $F(x)$. Nếu tồn tại hàm $f(x)$ sao cho với mọi $x \in \mathbb{R}$*

$$F(x) = \int_{-\infty}^{x} f(t)dt \qquad \textbf{(2.9)}$$

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

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



Tính chất của hàm mật độ

a. $F'(x) = f(x)$ tại các điểm x mà $f(x)$ liên tục. **(2.10)**

b. $f(x) \ge 0$ với mọi $x \in \mathbb{R}$, **(2.11)**

c. $\int_{-\infty}^{\infty} f(x)dx = 1$, **(2.12)**

d. $P\{a < X < b\} = P\{a \le X \le b\} = P\{a < X \le b\} = P\{a \le X < b\} = \int_a^b f(x)dx$. **(2.13)**

Ví dụ 2.4: Giả sử hàm phân bố xác suất của biến ngẫu nhiên liên tục X có dạng:

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

a) Xác định hệ số k;

b) Tìm hàm mật độ xác suất $f(x)$.

Giải:

a) Vì hàm phân bố xác suất $F(x)$ liên tục, do đó tại $x=1 \Rightarrow \quad 1=F(1)=kx^2\Big|_{x=1}=k$.

b) Theo tính chất (2.10) của hàm mật độ xác suất ta có

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

Ví dụ 2.5: Biến ngẫu nhiên liên tục X có hàm mật độ dạng

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

Hãy xác định:

a) Hệ số k;

b) Hàm phân bố $F(x)$;

c) Xác suất $P\{2<X<3\}$;

d) Xác suất để trong 4 phép thử độc lập biến ngẫu nhiên X đều không lấy giá trị trong khoảng $(2,3)$.

Giải:

a) Dựa vào tính chất (2.12) ta có:

$$1=\int_{-\infty}^{\infty} f(x)dx=\int_{1}^{\infty}\frac{k}{x^2}dx=-\lim_{a\to\infty}\left(\frac{k}{x}\Big|_1^a\right)=k \text{, từ đó } k=1.$$

b) Từ công thức (2.9) xác định hàm mật độ ta có

$$F(x)=\int_{-\infty}^{x} f(t)dt=\begin{cases} 0 & \text{với} \quad x<1 \\ \dfrac{x-1}{x} & \text{với} \quad x\ge 1 \end{cases}$$

c) Từ công thức (2.13) ta có $P\{2<X<3\}=F(3)-F(2)=\dfrac{2}{3}-\dfrac{1}{2}=\dfrac{1}{6}$.

d) 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

2.3.1. Kỳ vọng toán

2.3.1.1. Định nghĩa

Kỳ vọng hoặc giá trị trung bình (average, mean value, expected value) của biến ngẫu nhiên X ký hiệu là $\mathrm{E}X$ và được xác định như sau:

(i) Nếu X rời rạc nhận các giá trị x_i với xác suất tương ứng $p_i = P\{X = x_i\}$ thì

$$\mathrm{E}X = \sum_i x_i p_i \qquad \textbf{(2.14)}$$

(ii) Nếu X liên tục có hàm mật độ $f(x)$ thì

$$\mathrm{E}X = \int_{-\infty}^{\infty} x f(x)dx \qquad \textbf{(2.15)}$$

Kỳ vọng $\mathrm{E}X$ tồn tại nếu chuỗi (2.14) (trường hợp rời rạc) hội tụ tuyệt đối hoặc tích phân (2.15) (trường hợp liên tục) hội tụ tuyệt đối.

Ví dụ 2.6: Tính kỳ vọng của biến ngẫu nhiên X cho ở ví dụ 2.3.

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

Ví dụ 2.7: Theo thống kê, việc một người Mỹ 25 tuổi sẽ sống thêm trên một năm có xác suất là 0,992, còn xác suất để người đó chết trong vòng một năm tới là 0,008. Một chương trình bảo hiểm đề nghị người đó bảo hiểm sinh mạng cho 1 năm với số tiền chi trả 1000 đô la, còn tiền đóng là 10 đô la. Hỏi lợi nhuận của công ty bảo hiểm nhận được là bao nhiêu?

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

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

Do đó kỳ vọng $\mathrm{E}X = (-990)\cdot 0{,}008 + 10\cdot 0{,}992 = 2$. Ta thấy lợi nhuận trung bình là một số dương vì vậy công ty bảo hiểm có thể làm ăn có lãi.

Ví dụ 2.8: Tuổi thọ của một loại côn trùng nào đó là một biến ngẫu nhiên X (đơn vị là tháng) với hàm mật độ như sau:

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

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

Giải: Vì $\int_0^4 x^2(4-x)dx = \frac{64}{3} \ \Rightarrow \ k = \frac{3}{64}$. Hàm phân bố xác suất

$$F(x) = \int_{-\infty}^{x} f(t)dt = \begin{cases} 0 & \text{nếu } \ x \le 0 \\ \frac{3x^3}{64}\left(\frac{4}{3} - \frac{x}{4}\right) & \text{nếu } \ 0 < x \le 4 \\ 1 & \text{nếu } \ x > 4 \end{cases}$$

Tuổi thọ trung bình $\mathrm{E}\,X = \frac{3}{64}\int_0^4 x^3(4-x)dx = \frac{3}{64}\left(x^4 - \frac{x^5}{5}\right)\Bigg|_0^4 = \frac{12}{5}$ (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. Giả sử biến ngẫu nhiên X nhận các giá trị $x_1, x_2, ..., x_m$ với các tần số tương ứng $r_1, r_2, ..., r_m$.

$r_i x_i$ là tổng giá trị X nhận được với cùng giá trị x_i. Do đó $r_1x_1 + r_2x_2 + \cdots + r_mx_m$ là tổng tất cả các giá trị X nhận được.

Vậy $\frac{r_1x_1 + r_2x_2 + \cdots + r_mx_m}{n} = f_1x_1 + f_2x_2 + \cdots + f_mx_m$ là giá trị trung bình của X , trong đó $f_i = \frac{r_i}{n}$ là tần suất nhận giá trị x_i của X. Trong trường hợp tổng quát thì tần suất f_i được thay bằng xác suất p_i và ta có công thức (2.14).

Trường hợp biến ngẫu nhiên liên tục phép tính tổng của giá trị trung bình được thay bằng phép tính tích phân xác định, công thức (2.15).

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

Ví dụ 2.9: 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	0,3	0,25	0,18	0,14	0,1	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 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:

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

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

Nhu cầu

Lượng Hàng Nhập

i \ P_j, j	0,3 20	0,25 21	0,18 22	0,14 23	0,10 24	0,03 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

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

Vậy cửa hàng nên nhập 21 cuốn.

2.3.1.3. Tính chất

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

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

3) $\mathrm{E}(X_1+\cdots+X_n)=\mathrm{E}(X_1)+\cdots+\mathrm{E}(X_n)$ **(2.16)**

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

$$\mathrm{E}\,Y=\begin{cases}\sum_i \varphi(x_i)p_i & \text{nếu } X \text{ rời rạc có } p_i=P\{X=x_i\}\\ \int_{-\infty}^{\infty}\varphi(x)f(x)dx & \text{nếu } X \text{ liên tục có hàm mật độ } f(x)\end{cases} \qquad \textbf{(2.17)}$$

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

$$\mathrm{E}\,X^2=\begin{cases}\sum_i x_i^2 p_i & \text{nếu } X \text{ rời rạc}\\ \int_{-\infty}^{\infty}x^2 f(x)dx & \text{nếu } X \text{ liên tục có hàm mật độ } f(x)\end{cases} \qquad \textbf{(2.18)}$$

5) Nếu $X_1, ..., X_n$ độc lập thì $\mathrm{E}(X_1\cdots X_n)=\mathrm{E}(X_1)\cdots\mathrm{E}(X_n)$. **(2.19)**

Ví dụ 2.10: 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 2.3) 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	$5/30$	$15/30$	$9/30$	$1/30$

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

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

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

Ví dụ 2.11: Tung con xúc xắc n lần. Tìm kỳ vọng của tổng số nốt thu được.

Giải: Gọi $X_i\ (i=1,...,n)$ là số nốt thu được ở lần tung thứ i, gọi X là tổng số nốt thu được trong n lần tung. Như vậy $X=\sum_{i=1}^{n}X_i$.

Theo công thức (3.5) 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 $\mathrm{E}X_i = \frac{1}{6}(1+2+3+4+5+6) = \frac{7}{2} \quad \Rightarrow \quad \mathrm{E}X = \frac{7}{2}n$.

2.3.2. Phương sai

2.3.2.1. Định nghĩa

Phương sai (variance) hay độ lệch 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 $\mathrm{E}X$.

Phương sai của X được ký hiệu là $\mathrm{D}X$ hay $\mathrm{var}\,X$ và định nghĩa như sau:

$$\mathrm{D}X = \mathrm{E}(X - \mathrm{E}X)^2 \tag{2.20}$$

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

Khai triển vế phải công thức (2.20) và áp dụng các tính chất của kỳ vọng ta có thể tính phương sai theo công thức sau:

$$\mathrm{D}X = \mathrm{E}X^2 - (\mathrm{E}X)^2 \tag{2.21}$$

Từ công thức (2.17) thì phương sai có thể tính theo công thức sau:

(i) Nếu X rời rạc nhận các giá trị với xác suất tương ứng $p_i = P\{X = x_i\}$ thì

$$\mathrm{D}X = \sum_i (x_i - \mathrm{E}X)^2 p_i = \sum_i x_i^2 p_i - (\mathrm{E}X)^2 \tag{2.22}$$

(ii) Nếu X liên tục có hàm mật độ $f(x)$ thì

$$\mathrm{D}X = \int_{-\infty}^{\infty} (x - \mathrm{E}X)^2 f(x)dx = \int_{-\infty}^{\infty} x^2 f(x)dx - (\mathrm{E}X)^2 \tag{2.23}$$

Ví dụ 2.12: Tính phương sai của biến ngẫu nhiên xét trong ví dụ 2.7.

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 2.13: Tính phương sai của biến ngẫu nhiên xét trong ví dụ 2.8.

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

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

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 2.8 cho thấy đầu tư bảo hiểm cho những người 25 tuổi là có lãi, nhưng ví dụ 2.12 cho thấy rủi ro của bảo hiểm rất lớn.

Ví dụ 2.14: 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

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

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

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

1) $\mathrm{D}(aX) = a^2\mathrm{D}(X)$ với mọi hằng số a. **(2.24)**

2) $\mathrm{D}(aX+b) = a^2\mathrm{D}(X)$ với mọi hằng số a, b. **(2.25)**

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

$$\mathrm{D}\left(a_1X_1+\cdots+a_nX_n\right)=a_1^2\mathrm{D}\left(X_1\right)+\cdots+a_n^2\mathrm{D}\left(X_n\right). \quad \textbf{(2.26)}$$

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 2.15: Tung con xúc xắc n lần độc lập nhau. Tìm phương sai của tổng số nốt xuất hiện.

Giải: Xét $X=\sum_{i=1}^{n}X_i$ ở ví dụ 3.4. Vì các $X_i\,(i=1,...,n)$ độc lập nhau, theo công thức (2.26) ta có $\mathrm{D}X=\sum_{i=1}^{n}\mathrm{D}X_i$.

Mặt khác $\mathrm{E}X_i=\frac{7}{2}$; $\mathrm{E}X_i^2=\frac{1}{6}\left(1^2+2^2+3^2+4^2+5^2+6^2\right)=\frac{91}{6}$,

do đó $\mathrm{D}X_i=\frac{91}{6}-\frac{7^2}{2^2}=\frac{35}{12}$. Vậy $\mathrm{D}X=\frac{35}{12}n$.

2.3.3. Phân vị, Trung vị

2.3.3.1. Phân vị

Phân vị mức α của biến ngẫu nhiên X có hàm phân bố $F(x)$ là giá trị v_α thỏa mãn

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

Hay
$$F(v_\alpha)\le\alpha\le F(v_\alpha+0) \quad \textbf{(2.27)}$$

- Nếu $F(x)$ liên tục tăng chặt thì phân vị v_α là nghiệm duy nhất của phương trình $F(x)=\alpha$, nghĩa là

$$v_\alpha=F^{-1}(\alpha) \quad \textbf{(2.27)'}$$

- Nếu X rời rạc có phân bố: $\begin{array}{c|ccc} X & x_1 & x_2 & \cdots \\ \hline P & p_1 & p_2 & \cdots \end{array}$ đặt $P_i=p_1+\cdots+p_i$ thì

$$v_\alpha=\begin{cases} m,\ \ \forall m\in\left[x_i,x_{i+1}\right] & \text{nếu } P_i=\alpha<P_{i+1} \\ x_{i+1} & \text{nếu } P_i<\alpha<P_{i+1}\end{cases} \quad \textbf{(2.27)}$$

2.3.3.2.Trung vị

Phân vị mức 1/2 được gọi là *median* hay *trung vị* của X, ký hiệu $\mathrm{Med}X$. 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 theo nghĩa sau:

- Nếu X rời rạc có phân bố: $\begin{array}{c|ccc} X & x_1 & x_2 & \cdots \\ \hline P & p_1 & p_2 & \cdots \end{array}$ thì

$$x_{i_0} = \text{Mod}\,X \iff p_{i_0} = \max\{p_1, p_2, ...\} \quad \textbf{(2.28)}$$

- Nếu X liên tục có hàm mật độ $f(x)$

$$c = \text{Mod}\,X \iff f(c) = \max\{f(x),\ x \in \mathbb{R}\}. \quad \textbf{(2.29)}$$

Ví dụ 2.16: biến ngẫu nhiên X ở ví dụ 2.3 có Mốt, trung vị và $\text{Mod}X = \text{Med}X = 1$.

Ví dụ 2.17: 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 $\text{Mod}X = 20$.

Hàm phân bố xác suất của X

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

Từ đó suy ra $\text{Med}X = 21$.

Ví dụ 2.18: Tìm $\text{Med}X$ và $\text{Mod}X$ của biến ngẫu nhiên liên tục X xét trong ví dụ 2.4

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

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

Ví dụ 2.19: Tìm $\text{Med}X$ và $\text{Mod}X$ của biến ngẫu nhiên liên tục X có hàm mật độ xác định như sau:

$$f(x)=\begin{cases}\frac{3}{4}x(2-x) & \text{với} \quad 0\le x\le 2\\ 0 & \text{nếu trái lại}\end{cases}$$

Giải: Hàm phân bố xác suất:

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

$\mathrm{Med}X$ là nghiệm của phương trình $F(x)=\frac{1}{2} \Leftrightarrow \begin{cases}x^3-3x^2+2=0\\ 0<x\le 2\end{cases}$. Từ đó $\mathrm{Med}X=1$.

Hàm mật độ $f(x)$ có đạo hàm $f'(x)=\begin{cases}\frac{3}{2}(1-x) & \text{với} \quad 0<x<2\\ 0 & \text{nếu trái lại}\end{cases}$ đổi dấu từ dương sang âm khi đi qua $x=1$, do đó đạt cực đại tại điểm này. Vậy $\mathrm{Mod}X=1$.

2.3.5. Moment, hệ số bất đối xứng, hệ số nhọn

1) Moment gốc cấp k $\quad m_k=\mathrm{E}X^k$; $k=1,2,\ldots$ **(2.8)**

2) Moment quy tâm cấp k $\quad \mu_k=\mathrm{E}(X-\mathrm{E}X)^k$; $k=1,2,\ldots$ **(2.9)**

3) Hệ số bất đối xứng $\quad \alpha_3=\frac{\mu_3}{\sigma^3}$ với $\sigma=\sqrt{\mathrm{D}X}$. **(2.10)**

4) Hệ số nhọn $\quad \alpha_4=\frac{\mu_4}{\sigma^4}$. **(2.11)**

Nhận xét:

- $m_1=\mathrm{E}X$, $\mu_1=0$, $\mu_2=\mathrm{D}X$.
- 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 Median và Mode để định vị.

TÓM TẮT

Biến ngẫu nhiên

Biến ngẫu nhiên X là đại lượng nhận các giá trị nào đó phụ thuộc vào các yếu tố ngẫu nhiên, nghĩa là với mọi giá trị thực $x \in \mathbb{R}$ thì $\{X < x\}$ là một biến cố ngẫu nhiên.

Người ta phân các biến ngẫu nhiên thành hai loại:

❖ ***Biến ngẫu nhiên rời rạc*** nếu nó chỉ nhận một số hữu hạn hoặc vô hạn đếm được các giá trị. Nghĩa là có thể liệt kê các giá trị thành một dãy $x_1, x_2, \ldots$.

❖ ***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 và xác suất $P\{X = a\} = 0$ với mọi a.

Hàm phân bố xác suất

Hàm phân bố xác suất của biến ngẫu nhiên X là hàm số $F(x)$ xác định với mọi $x \in \mathbb{R}$ bởi công thức: $F(x) = P\{X < x\};\ -\infty < x < \infty$.

Bảng phân bố xác suất của biến ngẫu nhiên rời rạc

Giả sử biến ngẫu nhiên X chỉ nhận các giá trị $x_1, x_2, \ldots$ với xác suất tương ứng $p_i = P\{X = x_i\}$. $p_i > 0$ và $\sum_i p_i = 1$. Bảng phân bố xác suất của X có dạng sau:

X	x_1	x_2	$\cdots$
P	p_1	p_2	$\cdots$

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

Giả sử X là một biến ngẫu nhiên liên tục có hàm phân bố $F(x)$. Hàm mật độ của biến ngẫu nhiên X là hàm $f(x)$ sao cho với mọi $x \in \mathbb{R}$, $F(x) = \int_{-\infty}^{x} f(t)dt$.

Kỳ vọng

Kỳ vọng hoặc giá trị trung bình của biến ngẫu nhiên X được định nghĩa và ký hiệu như sau:

- Nếu X rời rạc nhận các giá trị với xác suất tương ứng $p_i = P\{X = x_i\}$ thì $\mathrm{E}\,X = \sum_i x_i p_i$

- Nếu X liên tục có hàm mật độ $f(x)$ thì $\mathrm{E}\,X = \int_{-\infty}^{\infty} xf(x)dx$.

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

Phương sai

Phương sai hay độ lệch 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 $\mathrm{E}X$. Phương sai của X được ký hiệu là $\mathrm{D}X$ hay $\mathrm{var}\,X$ và định nghĩa như sau: $\mathrm{D}X = \mathrm{E}\left(X - \mathrm{E}\,X\right)^2$.

Độ lệch tiêu chuẩn

$\sigma_X = \sqrt{\mathrm{D}X}$ được gọi là độ lệch tiêu chuẩn của X.

Phân vị

Phân vị mức α của biến ngẫu nhiên X có hàm phân bố $F(x)$ là giá trị v_α thỏa mãn:

$$P\{X < v_\alpha\} \le \alpha \le P\{X \le v_\alpha\} \text{ hay } F(v_\alpha) \le \alpha \le F(v_\alpha + 0)$$

Trung vị

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

Mốt

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

- Nếu X rời rạc có phân bố:

X	x_1	x_2	$\dots$
P	p_1	p_2	$\dots$

thì $x_{i_0} = \mathrm{Mod}\,X \Leftrightarrow p_{i_0} = \max\{p_1, p_2, ...\}$

- Nếu X liên tục có hàm mật độ $f(x)$ thì:

$$c = \mathrm{Mod}\,X \Leftrightarrow f(c) = \max\{f(x),\ x \in \mathbb{R}\}.$$

Moment, hệ số bất đối xứng, hệ số nhọn

Moment gốc cấp k: $m_k = \mathrm{E}X^k$; $k = 1, 2, ...$

Moment quy tâm cấp k : $\mu_k = \mathrm{E}(X - \mathrm{E}X)^k$; $k = 1, 2, ...$

Hệ số bất đối xứng: $\alpha_3 = \dfrac{\mu_3}{\sigma^3}$ với $\sigma = \sqrt{\mathrm{D}X}$.

Hệ số nhọn: $\alpha_4 = \dfrac{\mu_4}{\sigma^4}$.

- Hệ số bất đối xứng α_3 đo mức độ bất đối xứng của luật phân bố.
- Hệ số nhọn α_4 cho phép bổ sung thêm thông tin về phương sai của phân bố.

CÂU HỎI ÔN TẬP VÀ BÀI TẬP

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 chỉ nhân các giá trị $x_1, ..., x_n$ thì hệ các biến cố $\{X = x_1\}$, …, $\{X = x_n\}$ lập thành một hệ đầy đủ.

Đúng ☐ Sai ☐

2.4 Một biến ngẫu nhiên có thể có hai hàm phân bố khác nhau.

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 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 $f(x)$ của biến ngẫu nhiên liên tục có tính chất $f(x) \geq 0$.

Đúng ☐ Sai ☐

2.11 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.

2.12 Biến ngẫu nhiên rời rạc X nhận ba giá trị có thể có là x_1, x_2, x_3. Biết $x_1 = 4, x_2 = 0,6$ với xác suất tương ứng $p_1 = 0,5$, $p_2 = 0,3$ và có kỳ vọng EX = 8. Tìm x_3 và p_3.

2.13 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 $\mathrm{D}(X_1 + X_2)$.

2.14 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 $\overline{X} = \dfrac{X_1 + X_2 + X_3}{3}$. Tính $\mathrm{E}(\overline{X})$; $\mathrm{D}(\overline{X})$.

2.15 Hai biến ngẫu nhiên X, Y độc lập. Tính $\mathrm{D}(Z)$ với:

a) $Z = 2X + 3Y$. b) $Z = -3X + Y$.

Cho biết $\mathrm{D}(X) = 4$, $\mathrm{D}(Y) = 5$.

2.16 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 $\mathrm{D}(X) = 0,89$.

2.17 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.

2.18 Tính kỳ vọng của biến ngẫu nhiên liên tục X có hàm mật độ xác suất:

$$f(x)=\begin{cases}\dfrac{\cos x}{2} & \text{nếu } x\in(-\pi/2;\pi/2)\\ 0 & \text{nếu } x\notin(-\pi/2;\pi/2)\end{cases}$$

2.19. Cho biến ngẫu nhiên X liên tục với hàm mật độ như sau:

$$f(x)=\begin{cases}kx & \text{nếu } 0\le x\le 1\\ k & \text{nếu } 1\le x\le 4\\ 0 & \text{nếu trái lại}\end{cases}$$

a) Tìm k và hàm phân bố F(x).

b) Tính kỳ vọng $\mathrm{E}X$ và phương sai $\mathrm{D}X$.

2.20 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:

$$f(x)=\begin{cases}kx^2(2-x) & \text{nếu } 0\le x\le 2\\ 0 & \text{nếu trái lại}\end{cases}$$

a) Tìm k.

b) Tính xác suất để côn trùng chết trước khi nó được một tháng tuổi.

c) Tìm $\mathrm{E}X$, $\mathrm{D}X$.

2.21 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 DX.

b) Tìm phân bố xác suất của $Y=|X|$ và kỳ vọng $\mathrm{E}Y$.

2.22 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 EX và phương sai DX của X.

2.23 Cho biến ngẫu nhiên liên tục X có hàm mật độ

$$f(x)=\begin{cases}kx^2 & \text{nếu } 0\le x\le 3\\ 0 & \text{với } x \text{ còn lại}\end{cases}$$

a) Tìm k.

b) Tính kỳ $P\{X>2\}$. .

c) Tìm hàm phân bố của X.

d) Tìm α để $P\{X < \alpha\} = \dfrac{3}{4}$

2.24 Cho biến ngẫu nhiên rời rạc X có bảng phân bố xác suất

X	0	1	2	3	4	5	6	7
P	0	k	$2k$	$2k$	$3k$	k^2	k^2	k^2+k

a) Xác định k.

b) Tính xác suất $P\{X \ge 5\}$ và $P\{X < 3\}$.

c) Tính kỳ vọng EX.

d) Tính phương sai DX .

2.25 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 EX và phương sai DX.

b) Gọi Y là "số chính phẩm có thể gặp phải". Lập hệ thức cho biết mối quan hệ giữa Y và X. Tính kỳ vọng EY và phương sai DY.

2.26 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 EX.

2.27 Cho biến ngẫu nhiên liên tục X có hàm mật độ:

$$f(x) = \begin{cases} kx^2(1-x) & \text{nếu } \quad 0 \le x \le 1 \\ 0 & \text{với } x \text{ còn lại} \end{cases}$$

a) Tìm k. Tính $P\{0,4 < X < 0,6\}$.

b) Tính kỳ vọng EX.

2.28 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.

CHƯƠNG III: MỘT SỐ QUY LUẬT PHÂN BỐ XÁC SUẤT QUAN TRỌNG

GIỚI THIỆU

Trong chương này chúng ta sẽ khảo sát một số quy luật phân bố xác suất quan trọng:

- **Quy luật không – một A(p).**

Quy luật 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 xét biến ngẫu nhiên này dưới dạng tần suất xuất hiện biến cố A 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 …

- **Quy luật nhị thức**

Quy luật này thường gặp trong dãy phép thử Bernoulli, tổng của các biến ngẫu nhiên độc lập có quy luật không – một A(p).

- **Quy luật Poisson**

Quy luật này thường gặp trong bài toán về quá trình đếm sự xuất hiện biến cố A nào đó: 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.

- **Quy luật phân bố đều**

Quy luật phân bố đều trên một đoạn là quy luật phân bố xác suất của biến ngẫu nhiên liên tục đồng khả năng lấy giá trị trong khoảng đó. Quy luật 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.

- **Quy luật chuẩn**
- **Quy luật khi bình phương**
- **Quy luật Student**
- **Quy luật Fisher – Snedecor.**

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 quy luật 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 ...

Trong thống kê quy luật chuẩn và quy luật Student được ứng dụng để giải quyết các bài toán về giá trị trung bình và quy luật Fisher-Snedecor giải quyết các bài toán về phương sai của biến ngẫu nhiên có quy luật chuẩn. Quy luật khi bình phương được ứng dụng để kiểm định các giả thiết về xác suất của hệ đầy đủ.

Với mỗi quy luật xác suất ta sẽ khảo sát bảng phân bố xác suất hoặc hàm mật độ các tính chất và các đặc trưng của nó. Vì vậy để học tốt chương này học viên cần nắm chắc các khái niệm và kết quả trong chương 1 và chương 2.

NỘI DUNG

3.1. QUY LUẬT KHÔNG - MỘT A(p)

3.1.1. Định nghĩa

Biến ngẫu nhiên X gọi là có phân bố xác suất theo quy luật không – một tham số p nếu X chỉ nhận hai giá trị 0 và 1 với bảng phân bố xác suất như sau:

$$\begin{array}{c|cc} X & 0 & 1 \\ \hline P & q & p \end{array} \; ; \quad q = 1 - p \tag{3.1}$$

Quy luật không – một được ký hiệu $A(p)$.

Giả sử ta tiến hành một phép thử, trong đó xác suất xuất hiện của biến cố A bằng p. Gọi X là số lần xuất hiện biến cố A trong phép thử đó thì X là biến ngẫu nhiên rời rạc có quy luật không – một $A(p)$.

3.1.2. Các tham số đặc trưng của quy luật không – một A(p)

$$\mathrm{E}X = 0 \cdot q + 1 \cdot p = p \; ; \; \mathrm{E}X^2 = 0^2 \cdot q + 1^2 \cdot p = p$$

Như vậy kỳ vọng, phương sai và độ lệch chuẩn của biến ngẫu nhiên X có phân bố không – một tham số p được cho bởi

$$\mathrm{E}X = p; \; \mathrm{D}X = pq \; ; \; \sigma_X = \sqrt{pq} \tag{3.2}$$

Trong thực tế quy luật không – một 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 xét biến ngẫu nhiên có quy luật không - một, trong đó sản phẩm là phế phẩm ta cho nhận giá trị 1 và ngược lại cho nhận giá trị 0.

3.2. QUI LUẬT NHỊ THỨC $\mathscr{B}(n;p)$

3.2.1. Định nghĩa

Biến ngẫu nhiên rời rạc X nhận các giá trị $0, 1, ..., n$ với xác suất tương ứng

$$P\{X=k\}=C_n^k p^k q^{n-k} \tag{3.3}$$

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 $X \sim \mathscr{B}(n;p)$.

Bảng phân bố xác suất của biến ngẫu nhiên có quy luật nhị thức $\mathscr{B}(n;p)$

X	0	1	…	k	…	n
P	$C_n^0 p^0 q^n$	$C_n^1 pq^{n-1}$	…	$C_n^k p^k q^{n-k}$	…	$C_n^n p^n q^0$

3.2.2. Dãy phép thử Bernoulli

Dãy các phép thử lặp lại, độc lập, trong mỗi phép thử chỉ có 2 kết cục: A, $\overline{A}$ và xác suất xuất hiện của biến cố A không đổi $P(A)=p$, $(0<p<1)$ được gọi là dãy phép thử Bernoulli. p là xác suất thành công trong mỗi lần thử.

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 3.1: $P_n(k;p)=C_n^k p^k (1-p)^{n-k}$; $k=0,1,...,n$. **(3.4)**

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:

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

Mỗi biến cố này có xác suất $P(\underbrace{A...A}_{k \text{ lần}}\underbrace{\overline{A}...\overline{A}}_{n-k \text{ lần}})=p^k(1-p)^{n-k}$.

Vậy $P_n(k;p)=C_n^k p^k (1-p)^{n-k}$.

Định nghĩa: *Nếu $P_n(m;p)=\max\limits_{0\le k\le n} P_n(k;p)$ thì m được gọi là giá trị chắc chắn nhất của số thành công hay giá trị có khả năng xảy ra lớn nhất của phân bố nhị thức.*

Định lý 3.2: 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:

$$(n+1)p-1 \le m \le (n+1)p \quad \textbf{(3.5)}$$

Như vậy:

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

$$P_{\max} = P_n(m-1;p) = P_n(m;p) \quad \textbf{(3.6)}$$

Hệ quả: Nếu gọi X là số thành công trong n phép thử Bernoulli với xác suất thành công của mỗi lần thử là p thì X có phân bố nhị thức $\mathscr{B}(n;p)$ và $\text{Mod}X = m$.

Ví dụ 3.1: 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) Tìm giá trị chắc chắn nhất của số thành công.

d) 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

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

b) Xác suất để nguồn thu nhận được thông tin là $P = 1-(0{,}6)^3 = 0{,}784$.

c) $(n+1)p = (3+1)0,4 = 1,6 \Rightarrow m = 1$.

d) Xác suất để nguồn thu nhận được thông tin khi phát n lần là $P = 1-(0{,}6)^n$.

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

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

Nhận xét:

1. Phân bố nhị thức $\mathscr{B}(1;p)$ là phân bố không – một $A(p)$.

2. Thực hiện n phép thử Bernoulli với xác suất thành công của biến cố A trong mỗi lần thử là p. Gọi $X_1, X_2, ..., X_n$ lần lượt là số lần xuất hiện của biến cố A trong lần thử thứ $1, 2, ..., n$. Các biến ngẫu nhiên $X_1, X_2, ..., X_n$ độc lập có cùng phân bố không - một $\mathcal{A}(p)$. Gọi X là số thành công trong n phép thử Bernoulli này thì

$$X = X_1 + X_2 + \cdots + X_n \sim \mathcal{B}(n; p) \quad \textbf{(3.7)}$$

3. Từ (3.7) suy ra rằng nếu $X \sim \mathcal{B}(n_1; p)$ và $Y \sim \mathcal{B}(n_2; p)$ thì

$$X + Y \sim \mathcal{B}(n_1 + n_2; p) \quad \textbf{(3.8)}$$

3.2.3. Các tham số đặc trưng của quy luật nhị thức

Từ công thức (3.2) và (3.7) ta có công thức tính kỳ vọng, phương sai và độ lệch chuẩn của biến ngẫu nhiên $X \sim \mathcal{B}(n; p)$:

$$\mathrm{E}X = np; \ \mathrm{D}X = npq; \ \sigma_X = \sqrt{npq} \quad \textbf{(3.9)}$$

Từ công thức (3.5) và (3.6) ta có:

- Nếu $(n+1)p$ không nguyên thì

$$\mathrm{Mod}X = m \text{ thỏa mãn } (n+1)p - 1 \le m \le (n+1)p. \quad \textbf{(3.10)}$$

- Nếu $(n+1)p$ nguyên thì $\mathrm{Mod}X$ nhận hai giá trị

$$\mathrm{Mod}X = m = np + p \text{ hoặc } \mathrm{Mod}X = m - 1 = np - q. \quad \textbf{(3.10)'}$$

Ví dụ 3.2: 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 có khoảng bao nhiêu ngày bưu cục không bán hết báo.

Giải: Trong một ngày nào đó 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:

$$\mathrm{E}X = np = 300 \cdot 0{,}8926 = 267{,}78 \text{ ngày}.$$

3.3. PHÂN BỐ POISSON

3.3.1. Định nghĩa

Biến ngẫu nhiên X nhận các giá trị $k = 0, 1, 2, ...$ với xác suất

$$P\{X=k\}=e^{-\lambda}\frac{\lambda^k}{k!} \qquad \textbf{(3.11)}$$

gọi là có phân bố Poisson tham số $\lambda>0$*, ký hiệu* $X\sim\mathscr{P}(\lambda)$.

Trong thực tế với một số giả thiết thích hợp thì các biến ngẫu nhiên là các quá trình đếm sau:

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 một khoảng thời gian xác định nào đó sẽ có phân bố Poisson với tham số λ.

λ là tốc độ trung bình của quá trình diễn ra trong khoảng thời gian này.

Ví dụ 3.3: 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 $X(t)\sim\mathscr{P}(2t)$.

a) $X(2)\sim\mathscr{P}(4)$, do đó $P(A)=P\{X(2)=5\}=e^{-4}\dfrac{4^5}{5!}\approx 0{,}156$.

b) $X(1/2)\sim\mathscr{P}(1)$, do đó $P(B)=P\{X(1/2)=0\}=e^{-1}\approx 0{,}3679$.

c) $X(1/6)\sim\mathscr{P}(1/3)$, do đó

$$P(C)=P\{X(1/6)\ge 1\}=1-P\{X(1/6)=0\}=1-e^{-1/3}\approx 0{,}2835.$$

Quy luật 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.

3.3.2. Các tham số đặc trưng của quy luật Poisson

Kỳ vọng và phương sai của biến ngẫu nhiên X có quy luật Poisson tham số λ đều bằng λ:

$$\mathrm{E}X=\lambda;\ \ \mathrm{D}X=\lambda;\ \ \sigma_X=\sqrt{\lambda} \qquad \textbf{(3.12)}$$

Mốt của X là giá trị $m_0 = \text{Mod}X$ thỏa mãn:

$$\lambda - 1 \le m_0 \le \lambda \tag{3.13}$$

Như vậy

- Nếu λ nguyên thì $\text{Mod}X$ nhận 2 giá trị là $\lambda - 1$ hoặc λ.
- Nếu λ không nguyên thì $\text{Mod}X$ bằng phần nguyên của λ.

Nếu X_1, X_2 là hai biến ngẫu nhiên độc lập có phân bố Poisson tham số lần lượt λ_1, λ_2 thì $X_1 + X_2$ cũng có phân bố Poisson:

$$X_1 + X_2 \sim \mathscr{P}(\lambda_1 + \lambda_2) \tag{3.14}$$

Các quy luật phân bố xét ở trên là các quy luật phân bố quan trọng của biến ngẫu nhiên rời rạc. Sau đây ta xét một vài quy luật phân bố của biến ngẫu nhiên liên tục.

3.4. QUY LUẬT PHÂN BỐ ĐỀU $\mathbf{U}(a,b)$

3.4.1. Định nghĩa

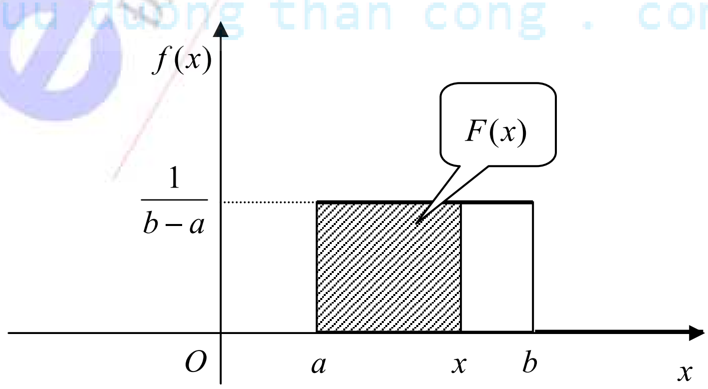
Biến ngẫu nhiên X được gọi là có phân bố đều trên $[a,b]$, ký hiệu $X \sim U(a;b)$ nếu hàm mật độ của X xác định bởi:

$$f(x) = \begin{cases} \dfrac{1}{b-a} & \text{nếu } \ a \le x \le b \\ 0 & \text{nếu ngược lại} \end{cases} \tag{3.15}$$

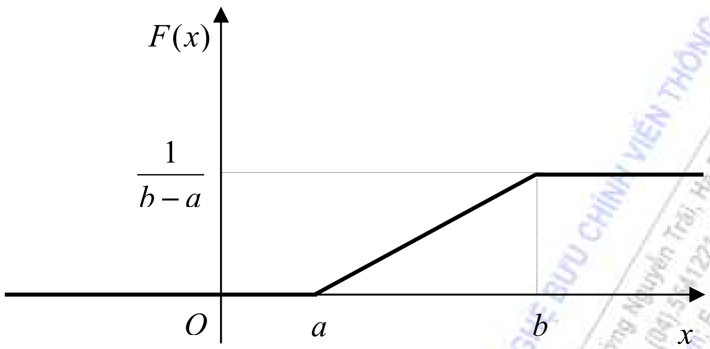
Hàm phân bố:

$$F(x) = \int_{-\infty}^{x} f(t)dt = \begin{cases} 0 & \text{nếu } \ x < a \\ \dfrac{x-a}{b-a} & \text{nếu } \ a \le x \le b \\ 1 & \text{nếu } \ x > b \end{cases} \tag{3.16}$$

Vậy X có khả năng nhận giá trị trong khoảng $[a,b]$ là "đều nhau" và không nhận giá trị ngoài $[a,b]$.



Đồ thị hàm mật độ của phân bố đều $\mathbf{U}(a,b)$



Đồ thị của hàm phân bố đều $\mathbf{U}(a,b)$

3.4.2. Các tham số đặc trưng của quy luật phân bố đều $\mathrm{U}(a,b)$

Bằng cách tính tích phân xác định ta nhận được công thức kỳ vọng và phương sai của biến ngẫu nhiên $X \sim \mathbf{U}(a,b)$:

$$\mathrm{E}X = \frac{a+b}{2}\ ;\quad \mathrm{D}X = \frac{(b-a)^2}{12} \tag{3.17}$$

Từ đồ thị của hàm mật độ (3.15) ta suy ra rằng: $\mathrm{Med}X = \mathrm{E}X = \dfrac{a+b}{2}$ và $\mathrm{Mod}X = m$, $\forall m \in [a\ b]$.

Quy luật phân bố đều có nhiều ứng dụng trong thống kê toán như 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 quy luật phân bố đều.

3.5. QUY LUẬT CHUẨN $\mathrm{N}(\mu;\sigma^2)$

3.5.1. Định nghĩa

Biến ngẫu nhiên liên tục X *có phân bố chuẩn* $N(\mu;\sigma^2)$, *ký hiệu* $X \sim N(\mu;\sigma^2)$, *nếu hàm mật độ có dạng:*

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{\frac{-(x-\mu)^2}{2\sigma^2}}\;;\quad \forall x\in\mathbb{R} \tag{3.18}$$

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 tuân theo quy luật 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 ...

3.5.2. Các tham số đặc trưng của quy luật chuẩn

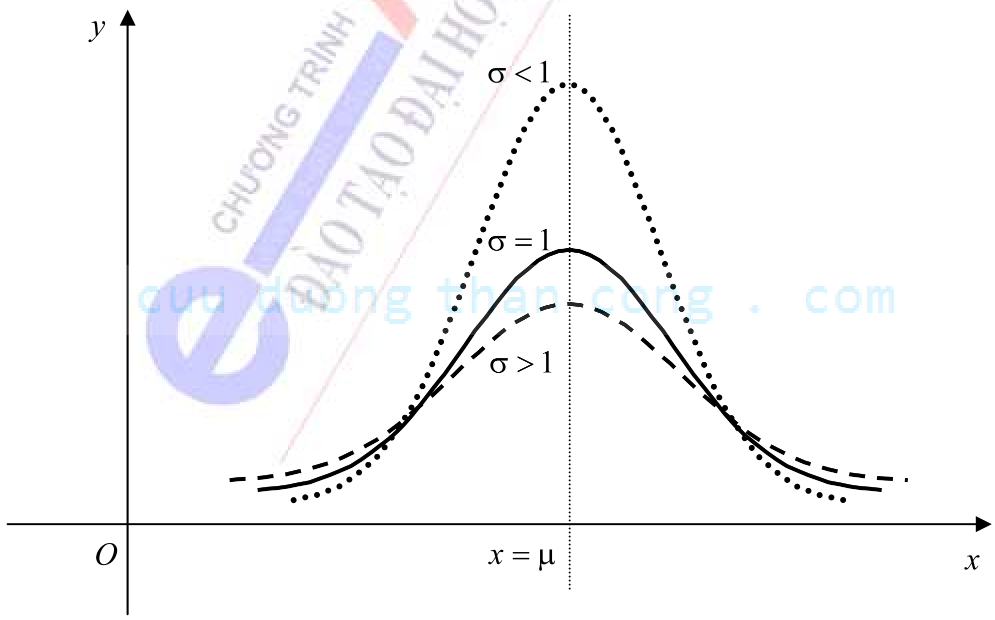
Từ công thức xác định hàm mật độ (3.18) ta suy ra các tính chất sau của đồ thị:

- 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 $X \sim N(\mu;\sigma^2)$:

$$\mathrm{E}X = \mu\ ; \quad \mathrm{D}X = \sigma^2\ ; \quad \sigma_X = \sqrt{\mathrm{D}X} = \sigma \tag{3.19}$$

Nếu X_1, X_2 là hai biến ngẫu nhiên độc lập có phân bố chuẩn $X_1 \sim N(\mu_1;\sigma_1^2)$ và $X_2 \sim N(\mu_2;\sigma_2^2)$ thì tổ hợp tuyến tính bất kỳ của X_1, X_2 cũng có phân bố chuẩn. Đặc biệt:

$$X_1 + X_2 \sim N(\mu_1 + \mu_2;\sigma_1^2 + \sigma_2^2) \tag{3.20}$$

3.5.3. Phân bố chuẩn tắc

Phân bố chuẩn $N(0;1)$ với kỳ vọng bằng 0, phương sai bằng 1 gọi là *phân bố chuẩn tắc.*

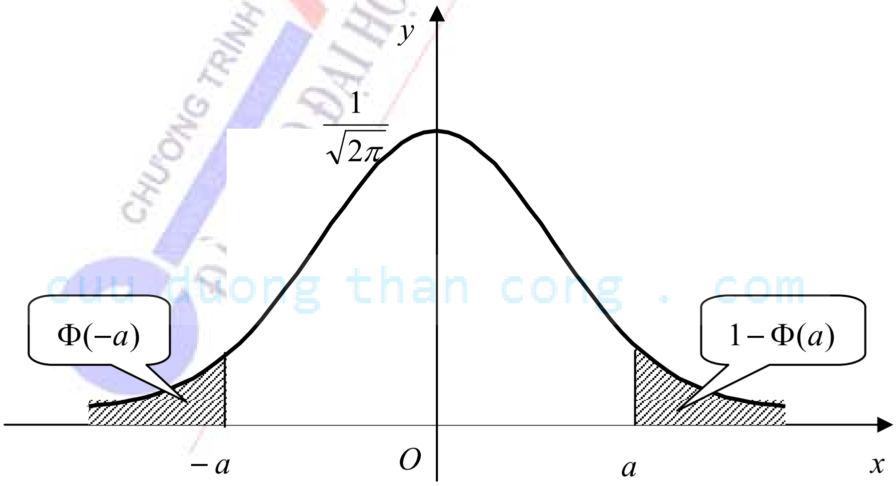
Hàm mật độ của $N(0;1)$:

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}\ ; \quad \forall x \in \mathbb{R} \tag{3.21}$$

Hàm phân bố của $N(0;1)$:

$$\Phi(x) = \int_{-\infty}^{x} \varphi(t)dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{\frac{-t^2}{2}} dt\ ; \quad \forall x \in \mathbb{R} \tag{3.22}$$

Có bảng tính sẵn các giá trị của $\varphi(x)$ và $\Phi(x)$ (xem Phụ lục I và Phụ lục II).



Đồ thị của hàm mật độ $\varphi(x)$

Các tính chất của hàm phân bố $\Phi(x)$

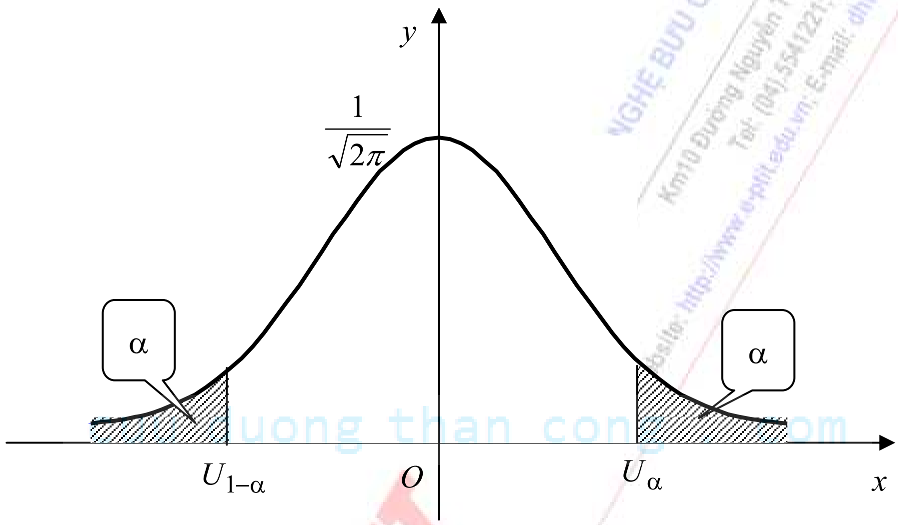
1) $\Phi(x)+\Phi(-x)=1$, $\Phi(-x)=1-\Phi(x)$.

2) Nếu $X \sim N(0;1)$ thì

$$\forall\, a>0,\ P\{|X|<a\}=2\Phi(a)-1,\ P\{|X|>a\}=2\big(1-\Phi(a)\big). \tag{3.23}$$

Định nghĩa: *Giá trị U_α gọi là giá trị tới hạn của phân bố chuẩn tắc mức α nếu*

$$\Phi^{-1}(1-\alpha)=U_\alpha. \tag{3.24}$$



Nếu $X \sim N(0;1)$ thì

$$P\{X>U_\alpha\}=\alpha\ ;\ P\left\{|X|>U_{\frac{\alpha}{2}}\right\}=\alpha\ ;\ P\left\{|X|<U_{\frac{\alpha}{2}}\right\}=1-\alpha \tag{3.25}$$

3) Người ta chứng minh được nếu $X \sim N(\mu;\sigma^2)$ thì $\dfrac{X-\mu}{\sigma} \sim N(0;1)$.

Từ đó ta có $F(x)=P\{X<x\}=P\left\{\dfrac{X-\mu}{\sigma}<\dfrac{x-\mu}{\sigma}\right\}=\Phi\left(\dfrac{x-\mu}{\sigma}\right)$. **(3.26)**

$$P\{a<X<b\}=P\left\{\frac{a-\mu}{\sigma}<\frac{X-\mu}{\sigma}<\frac{b-\mu}{\sigma}\right\}=\Phi\left(\frac{b-\mu}{\sigma}\right)-\Phi\left(\frac{a-\mu}{\sigma}\right). \tag{3.26'}$$

Xác suất của sự sai lệch giữa biến ngẫu nhiên có quy luật chuẩn $X \sim N(\mu;\sigma^2)$ và kỳ vọng của nó được tính theo công thức:

$$P\{|X-\mu|<\varepsilon\}=2\Phi\left(\frac{\varepsilon}{\sigma}\right)-1 \tag{3.27}$$

Ví dụ 3.4: Giả sử $X \sim N(\mu;\sigma^2)$, $\mu = 2100,\ \sigma = 200$. Hãy tìm:

a) $P\{X > 2400\}$.

b) $P\{1700 < X < 2200\}$.

c) Xác định a để $P\{X > a\} = 0,03$.

Giải:

a) $P\{X > 2400\} = 1 - \Phi\left(\dfrac{2400-2100}{200}\right) = 1 - \Phi(1,5) = 1 - 0,9332 = 0,0668$.

b) Áp dụng công thức (3.23):

$$P\{1700 < X < 2200\} = \Phi\left(\frac{2200-2100}{200}\right) - \Phi\left(\frac{1700-2100}{200}\right) = \Phi(0,5) - \Phi(-2) = 0,6688$$

c) $P\{X > a\} = 1 - \Phi\left(\dfrac{a-2100}{200}\right) = 0,03 \ \Rightarrow\ \Phi\left(\dfrac{a-2100}{200}\right) = 0,97$.

Tra bảng ta được $0,97 = \Phi(1,881) \ \Rightarrow\ \dfrac{a-2100}{200} = 1,881 \ \Rightarrow\ a = 2476,2$.

3.5.4. Quy tắc hai xích ma và ba xích ma

Nếu trong công thức (3.27) ta đặt $\varepsilon = 2\sigma$ tức là bằng hai lần độ lệch chuẩn của X thì $P\{|X-\mu| < 2\sigma\} = 2\Phi(2) - 1 = 0,9544$. Vậy

$$P\{\mu - 2\sigma < X < \mu + 2\sigma\} = 0,9544 \qquad \textbf{(3.28)}$$

Tương tự thay $\varepsilon = 3\sigma$ ta được

$$P\{\mu - 3\sigma < X < \mu + 3\sigma\} = 0,9973 \qquad \textbf{(3.29)}$$

Hai công thức trên là cơ sở của quy tắc hai xích ma và ba xích ma:

Nếu X có phân bố chuẩn $N(\mu;\sigma^2)$ 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 $(\mu - 3\sigma;\ \mu + 3\sigma)$.

3.5.5. Tính gần đúng phân bố nhị thức

Giả sử $X_1, X_2, ..., X_n$ độc lập có cùng phân bố không – một $A(p)$. Theo công thức (3.7) ta có $U_n = X_1 + X_2 + \cdots + X_n \sim \mathscr{B}(n;p)$. Công thức (3.3) cho phép tính xác suất $P\{U_n = k\} = C_n^k p^k q^{n-k}$.

Tuy nhiên khi n khá lớn ta không thể áp dụng công thức này để tính mà cần đến công thức xấp xỉ.

3.5.5.1. Tính gần đúng phân bố nhị thức bằng phân bố chuẩn

Định lý Moivre –Laplace:

Với mọi $x \in \mathbb{R}$: $\lim\limits_{n\to\infty} P\left\{\dfrac{U_n - np}{\sqrt{npq}} < x\right\} = \Phi(x)$

Do đó khi n đủ lớn

$$P\{U_n < x\} = P\left\{\frac{U_n - np}{\sqrt{npq}} < \frac{x-np}{\sqrt{npq}}\right\} \approx \Phi\left(\frac{x-np}{\sqrt{npq}}\right) \qquad \textbf{(3.30)}$$

Định lý Moivre-Laplace cho phép xấp xỉ phân bố nhị thức $\mathscr{B}(n;p)$ với phân bố chuẩn $N(np;npq)$ khi n đủ lớn (công thức 3.30). Người ta thấy rằng xấp xỉ là tốt khi np và nq lớn hơn 5 hoặc khi $npq > 20$.

Nếu $X \sim \mathscr{B}(n;p)$ thì khi n đủ lớn ta có công thức tính xấp xỉ:

$$P_n(k;p) = P\{X = k\} \approx \frac{1}{\sqrt{npq}}\varphi\left(\frac{k-np}{\sqrt{npq}}\right) = \frac{1}{\sqrt{npq}}\frac{1}{\sqrt{2\pi}}e^{\frac{-(k-np)^2}{2npq}}. \qquad \textbf{(3.31)}$$

$$P\{a < X < b\} \approx \Phi\left(\frac{b-np}{\sqrt{npq}}\right) - \Phi\left(\frac{a-np}{\sqrt{npq}}\right) \qquad \textbf{(3.32)}$$

Ví dụ 3.5: 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:

$$P_{3200}(1600;0,5) = \frac{3200!}{1600!1600!}0,5^{3200}.$$

Mặt khác nếu tính gần đúng ta có: $x_m = \dfrac{1600 - 3200\cdot 0,5}{\sqrt{3200\cdot 0,5\cdot 0,5}} = 0$. 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) $P\left\{5\sqrt{2} \le X - 1600 \le 10\sqrt{2}\right\} = P\left\{0,25 \le \dfrac{X-1600}{20\sqrt{2}} \le 0,5\right\} \approx \Phi(0,5) - \Phi(0,25) = 0,09$.

3.5.5.2. Tính gần đúng phân bố nhị thức bằng phân bố Poisson

Khi npq bé thì xấp xỉ $\mathscr{B}(n;p)$ bởi $N(np;npq)$ không tốt. Tuy nhiên trong trường hợp $n > 50$ và $p < 0,1$ người ta chứng minh được rằng có thể xấp xỉ $\mathscr{B}(n;p)$ với phân bố Poisson $\mathscr{P}(np)$ tham số $\lambda = np$.

Ví dụ 3.6: Một nhà phân tích kinh tế tiên đoán 3,5% số công ti nhỏ trong 1 miền 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 $X \sim \mathscr{B}(n;p)$ với $n = 100,\ p = 0,035$. Ta có thể xấp xỉ $X \sim \mathscr{P}(np) = \mathscr{P}(3,5)$.

Ta cần tính:

$$P\{X \ge 3\} = 1 - \left[P\{X=0\} + P\{X=1\} + P\{X=2\}\right] = 1 - e^{-3,5}\left\{1 + \frac{3,5}{1!} + \frac{3,5^2}{2!}\right\}$$

$$= 1 - \left[0,0302 + 0,1057 + 0,1850\right] = 0,697 \approx 69,7\% .$$

3.6. QUY LUẬT KHI BÌNH PHƯƠNG

3.6.1 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:

$$f(x) = \begin{cases} \dfrac{x^{n/2-1}}{2^{n/2}\Gamma(n/2)} e^{\frac{-x}{2}} & \text{nếu } \ x > 0 \\ 0 & \text{nếu } \ x \le 0 \end{cases} \qquad \textbf{(3.33)}$$

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 $N(0,1)$ thì:

$$\sum_{i=1}^{n} X_i^2 = X_1^2 + X_2^2 + \cdots + X_n^2 \sim \chi_n^2 \qquad \textbf{(3.34)}$$

Phân bố χ^2 do Karl Pearson đưa ra vào năm 1900.

Từ (3.34) suy ra rằng nếu X_1, X_2 là hai biến ngẫu nhiên độc lập có phân bố khi bình phương lần lượt n_1 và n_2 bậc tự do thì $X_1 + X_2$ là biến ngẫu nhiên có phân bố khi bình phương $n_1 + n_2$ bậc tự do:

$$X_1 + X_2 \sim \chi^2(n_1 + n_2) \qquad \textbf{(3.35)}$$

3.6.2. Các tham số đặc trưng của phân bố "Khi bình phương" n bậc tự do

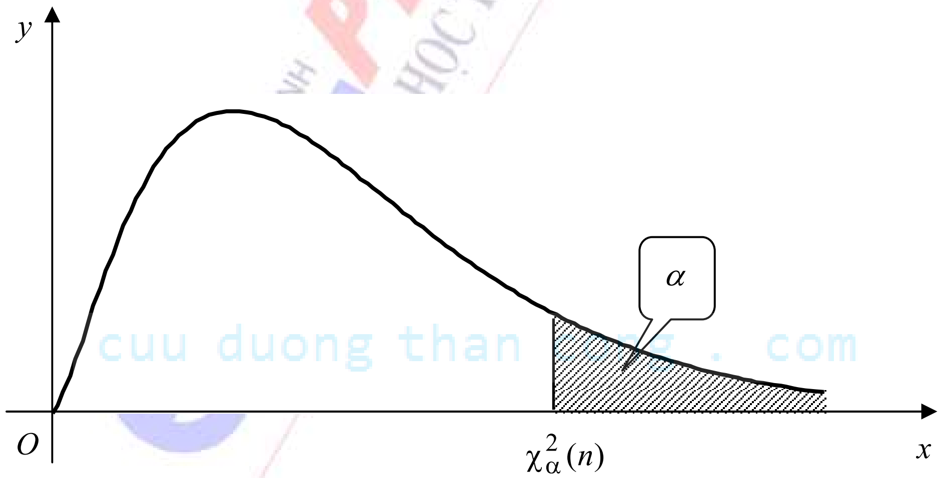
Có thể chứng minh được rằng, kỳ vọng và phương sai của biến ngẫu nhiên "Khi bình phương" n bậc tự do $\chi^2(n)$ có công thức:

$$\mathrm{E}\chi^2(n) = n;\ \mathrm{D}\chi^2(n) = 2n \qquad \textbf{(3.36)}$$

Giá trị tới hạn khi bình phương n bậc tự do mức α, ký hiệu $\chi_\alpha^2(n)$, được định nghĩa như sau, với $\chi^2 \sim \chi^2(n)$:

$$P\{\chi^2 > \chi_\alpha^2(n)\} = \alpha \qquad \textbf{(3.37)}$$

Bảng các giá trị tới hạn $\chi_\alpha^2(n)$ được tính sẵn trong bảng ở Phụ lục III.



3.7. QUY LUẬT STUDENT $\mathbf{T}(n)$

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:

$$f(t)=\frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma(n/2)}\left(1+\frac{t^2}{n}\right)^{\frac{-(n+1)}{2}}, \quad -\infty<t<\infty. \tag{3.38}$$

trong đó $\Gamma(x)$ là hàm Gamma.

Người ta chứng minh được rằng nếu $Z \sim N(0;1)$, $V \sim \chi^2(n)$; Z và V độc lập thì

$$T=\frac{Z}{\sqrt{V/n}} \sim \mathbf{T}(n) \tag{3.39}$$

Kỳ vọng và phương sai của biến ngẫu nhiên T có phân bố Student n bậc tự do:

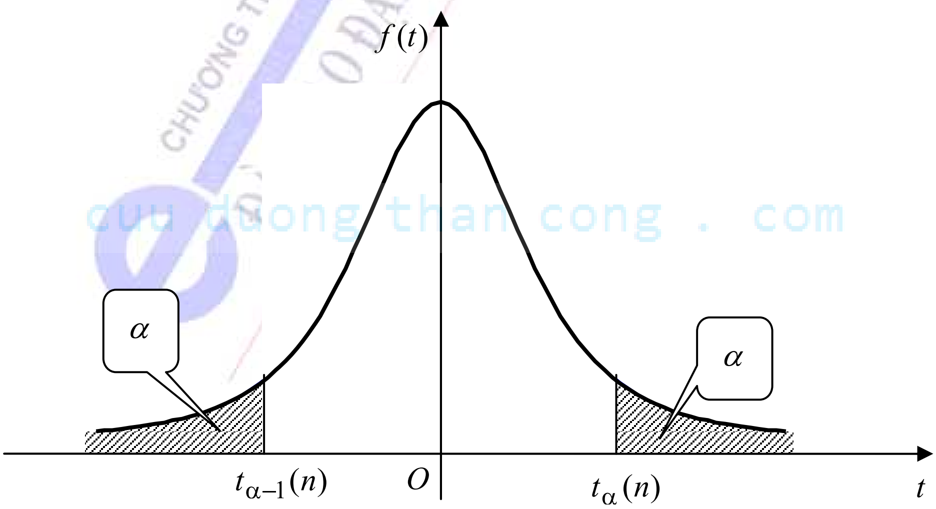
$$\mathrm{E}T=0,\ \mathrm{D}T=\frac{n}{n-2} \tag{3.40}$$

Giá trị tới hạn mức α của phân bố Student n bậc tự do, ký hiệu $t_\alpha(n)$, thỏa mãn:

$$P\{T>t_\alpha(n)\}=\alpha \tag{3.41}$$

Bảng tính các giá trị tới hạn $t_\alpha(n)$ cho trong Phụ lục IV.

Hàm mật độ (3.38) 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 $N(0;1)$. Do đó khi n đủ lớn ($n \ge 30$) 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.



3.8. QUY LUẬT FISHER-SNEDECOR $\mathbf{F}(n_1, n_2)$

Biến ngẫu nhiên liên tục F có phân bố Fisher-Snedecor với n_1 và n_2 bậc tự do, ký hiệu $F \sim \mathbf{F}(n_1, n_2)$, nếu hàm mật độ có dạng:

$$f(x) = \begin{cases} 0 & \text{nếu } x \le 0 \\ C\dfrac{x^{\frac{n_1-2}{2}}}{(n_2+n_1x)^{\frac{n_1+n_2}{2}}} & \text{nếu } x > 0 \end{cases} \quad \textbf{(3.42)}$$

trong đó $C = \dfrac{\Gamma\left(\dfrac{n_1+n_2}{2}\right) n_1^{\frac{n_1}{2}} n_2^{\frac{n_2}{2}}}{\Gamma\left(\dfrac{n_1}{2}\right)\Gamma\left(\dfrac{n_2}{2}\right)}$.

Người ta chứng minh được rằng nếu $U \sim \chi^2(n_1)$, $V \sim \chi^2(n_2)$; U và V độc lập thì:

$$F = \frac{U/n_1}{V/n_2} \sim \mathbf{F}(n_1, n_2) \quad \textbf{(3.43)}$$

Kỳ vọng và phương sai của biến ngẫu nhiên F có phân bố Fisher-Snedecor với n_1 và n_2 bậc tự do:

$$\mathrm{E}F = \frac{n_2}{n_2-2}, \; \mathrm{D}F = \frac{2n_2^2(n_1+n_2-2)}{n_1^2(n_2-2)^2(n_2-4)} \quad \textbf{(3.44)}$$

Giá trị tới hạn mức α của phân bố Fisher-Snedecor với n_1 và n_2 bậc tự do, ký hiệu $f_\alpha(n_1, n_2)$, thỏa mãn:

$$P\{F > f_\alpha(n_1, n_2)\} = \alpha \quad \textbf{(3.45)}$$

Bảng tính các giá trị tới hạn $f_\alpha(n_1, n_2)$ cho trong Phụ lục Va, Vb, Vc, Vd.

TÓM TẮT

Quy luật không – một A(p)

Biến ngẫu nhiên X gọi là có phân bố xác suất theo quy luật không – một tham số p nếu X chỉ nhận hai giá trị 0 và 1 với bảng phân bố xác suất như sau:

$$\begin{array}{c|cc} X & 0 & 1 \\ \hline P & q & p \end{array} \; ; \quad q = 1 - p$$

$$\mathrm{E}X = p;\ \mathrm{D}X = pq\ ;\ \ \sigma_X = \sqrt{pq}$$

Qui luật nhị thức $\mathscr{B}(n;p)$

Biến ngẫu nhiên rời rạc X nhận các giá trị 0, 1, ..., n với xác suất tương ứng:

$$P\{X = k\} = C_n^k p^k q^{n-k}$$

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 $X \sim \mathscr{B}(n;p)$.

$$\mathrm{E}X = np;\ \mathrm{D}X = npq\ ;\ \ \sigma_X = \sqrt{npq}$$

Dãy phép thử Bernoulli

Dãy các phép thử lặp lại, độc lập và trong mỗi phép thử sự xuất hiện của biến cố A có xác suất $P(A) = p$, $(0 < p < 1)$ không đổi được gọi là dãy phép thử Bernoulli.

Khi $m = [(n+1)p]$ thì $P_n(m;p) = C_n^m p^m (1-p)^{n-m}$ đạt giá trị lớn nhất. Gọi m là giá trị có khả năng xảy ra lớn nhất của dãy phép thử Bernoulli.

Nếu gọi X là số thành công trong n phép thử Bernoulli với xác suất thành công của mỗi lần thử là p, theo định lý 3.1 thì X có phân bố nhị thức $\mathscr{B}(n;p)$.

Nếu $X_1, X_2, ..., X_n$ là các biến ngẫu nhiên độc lập có cùng phân bố không – một A(p) thì

$$X = X_1, + X_2 + \cdots + X_n \sim \mathscr{B}(n;p).$$

Phân bố Poisson

Biến ngẫu nhiên X nhận các giá trị $k = 0, 1, 2, ...$ với xác suất $P\{X = k\} = e^{-\lambda} \dfrac{\lambda^k}{k!}$ gọi là có phân bố Poisson tham số $\lambda > 0$, ký hiệu $X \sim \mathscr{P}(\lambda)$.

$$EX = \lambda;\ \mathrm{D}X = \lambda\ ;\ \ \sigma_X = \sqrt{\lambda}\ ;\ \lambda - 1 \le \mathrm{Mod}X \le \lambda.$$

Quy luật phân bố đều $\mathbf{U}(a,b)$

Biến ngẫu nhiên X được gọi là có phân bố đều trên $[a, b]$ nếu hàm mật độ và hàm phân bố

$$f(x)=\begin{cases}\dfrac{1}{b-a} & \text{nếu } a\le x\le b\\ 0 & \text{nếu ngược lại}\end{cases}\qquad F(x)=\int_{-\infty}^{x} f(t)dt=\begin{cases}0 & \text{nếu } x<a\\ \dfrac{x-a}{b-a} & \text{nếu } a\le x\le b\\ 1 & \text{nếu } x>b\end{cases}$$

$$\mathrm{E}X=\frac{a+b}{2}\ ;\quad \mathrm{D}X=\frac{(b-a)^2}{12}\ ;\ \mathrm{Med}X=\mathrm{E}X=\frac{a+b}{2}\ \text{ và } \mathrm{Mod}X=m,\ \forall m\in[a\ b].$$

Quy luật chuẩn $\mathbf{N}(\mu;\sigma^2)$

Biến ngẫu nhiên liên tục X có phân bố chuẩn $N(\mu;\sigma^2)$, ký hiệu $X\sim N(\mu;\sigma^2)$, nếu hàm mật độ có dạng $f(x)=\dfrac{1}{\sigma\sqrt{2\pi}}e^{\frac{-(x-\mu)^2}{2\sigma^2}}$; $-\infty<x<\infty$.

$$\mathrm{Mod}X=\mathrm{Med}X=\mathrm{E}X=\mu\ ;\quad \mathrm{D}X=\sigma^2\ ;\ \ \sigma_X=\sqrt{\mathrm{D}X}=\sigma.$$

Phân bố chuẩn tắc

Phân bố chuẩn $N(0;1)$ với kỳ vọng bằng 0, phương sai bằng 1 gọi là *phân bố chuẩn tắc.*

Hàm mật độ: $\varphi(x)=\dfrac{1}{\sqrt{2\pi}}e^{\frac{-x^2}{2}}$; $\forall x\in\mathbb{R}$

Hàm phân bố: $\Phi(x)=\displaystyle\int_{-\infty}^{x}\varphi(t)dt=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{\frac{-t^2}{2}}dt$; $\forall x\in\mathbb{R}$.

Giá trị U_α thỏa mãn $\Phi^{-1}(1-\alpha)=U_\alpha$ gọi là giá trị tới hạn của phân bố chuẩn mức α .

Nếu $X\sim N(0;1)$ thì $P\{X>U_\alpha\}=\alpha$; $P\left\{|X|>U_{\frac{\alpha}{2}}\right\}=\alpha$.

$X\sim N(\mu;\sigma^2)$ thì $\dfrac{X-\mu}{\sigma}\sim N(0;1)$.

Từ đó $P\{X<x\}=P\left\{\dfrac{X-\mu}{\sigma}<\dfrac{x-\mu}{\sigma}\right\}=\Phi\left(\dfrac{x-\mu}{\sigma}\right)$.

Quy luật khi bình phương

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

$$f(x)=\begin{cases}\dfrac{x^{n/2-1}}{2^{n/2}\Gamma(n/2)}e^{-\frac{x}{2}} & \text{nếu } x>0\\ 0 & \text{nếu } x\le 0\end{cases}$$

Nếu $X_1, X_2, ..., X_n$ là các biến ngẫu nhiên độc lập có cùng phân bố chuẩn tắc $N(0,1)$ thì

$$\sum_{i=1}^{n} X_i^2 = X_1^2 + X_2^2 + \cdots + X_n^2 \sim \chi_n^2$$

$\mathrm{E}\chi^2 = n$; $\mathrm{D}\chi^2 = 2n$. Giá trị tới hạn $\chi_\alpha^2(n)$ được định nghĩa: $P\{\chi^2 > \chi_\alpha^2(n)\} = \alpha$.

Quy luật Student $\mathbf{T}(n)$

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: $f(t)=\dfrac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma(n/2)}\left(1+\dfrac{t^2}{n}\right)^{\frac{-(n+1)}{2}}$, $-\infty < t < \infty$..

Nếu $Z \sim N(0;1)$, $V \sim \chi^2(n)$; Z và V độc lập thì $T = \dfrac{Z}{\sqrt{V/n}} \sim \mathbf{T}(n)$

$\mathrm{E}T = 0$, $\mathrm{D}T = \dfrac{n}{n-2}$. Giá trị tới hạn mức α của phân bố Student n bậc tự do ký hiệu $t_\alpha(n)$ thỏa mãn: $P\{T > t_\alpha(n)\} = \alpha$.

Quy luật Fisher-Snedecor $\mathbf{F}(n_1, n_2)$

Biến ngẫu nhiên liên tục F có phân bố Fisher-Snedecor với n_1 và n_2 bậc tự do, ký hiệu $F \sim \mathbf{F}(n_1, n_2)$, nếu hàm mật độ có dạng:

$$f(x)=\begin{cases}0 & \text{nếu } x\le 0\\ C\dfrac{x^{\frac{n_1-2}{2}}}{(n_2+n_1x)^{\frac{n_1+n_2}{2}}} & \text{nếu } x>0\end{cases} \quad \text{trong đó } C=\frac{\Gamma\left(\frac{n_1+n_2}{2}\right)n_1^{\frac{n_1}{2}}n_2^{\frac{n_2}{2}}}{\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)}.$$

Nếu $U \sim \chi_{n_1}^2$, $V \sim \chi_{n_2}^2$; U và V độc lập thì $F = \dfrac{U/n_1}{V/n_2} \sim \mathbf{F}(n_1, n_2)$.

$$\mathrm{E}F = \frac{n_2}{n_2 - 2}, \quad \mathrm{D}F = \frac{2n_2^2(n_1 + n_2 - 2)}{n_1^2(n_2 - 2)^2(n_2 - 4)}.$$

Giá trị tới hạn mức α của phân bố Fisher-Snedecor với n_1 và n_2 bậc tự do ký hiệu $f_\alpha(n_1, n_2)$ thỏa mãn: $P\{F > f_\alpha(n_1, n_2)\} = \alpha$.

CÂU HỎI ÔN TẬP VÀ BÀI TẬP

3.1 Tổng của n biến ngẫu nhiên độc lập phân bố theo quy luật không – một A(p) là một biến ngẫu nhiên phân bố theo quy luật nhị thức $\mathscr{B}(n;p)$.

Đúng ☐ Sai ☐

3.2 Tổng của hai biến ngẫu nhiên phân bố theo quy luật nhị thức bất kỳ luôn luôn là một biến ngẫu nhiên phân bố theo quy luật nhị thức.

Đúng ☐ Sai ☐

3.3 Biến ngẫu nhiên phân bố theo quy luật Poisson là biến ngẫu nhiên rời rạc nên chỉ nhận một số hữu hạn các giá trị.

Đúng ☐ Sai ☐

3.4 Nếu X là biến ngẫu nhiên phân bố theo quy luật nhị thức $\mathscr{B}(n;p)$ thì X chỉ nhận một giá trị $\mathrm{Mod}X = m$ thỏa mãn $(n+1)p - 1 \le m \le (n+1)p$.

Đúng ☐ Sai ☐

3.5 Nếu X là biến ngẫu nhiên phân bố theo quy luật Poisson tham số λ>0 thì kỳ vọng, phương sai và mốt của X đều bằng λ.

Đúng ☐ Sai ☐

3.6 Nếu biến ngẫu nhiên X phân bố theo quy luật chuẩn $N(\mu;\sigma^2)$ thì xác suất sai lệch giữa X và kỳ vọng của nó thỏa mãn $P\{|X - \mu| < \varepsilon\} = 2\Phi\left(\frac{\varepsilon}{\sigma}\right) - 1$.

Đúng ☐ Sai ☐

3.7 Nếu biến ngẫu nhiên X phân bố theo quy luật chuẩn $N(\mu;\sigma^2)$ thì $\dfrac{X-\mu}{\sigma}$ phân bố theo quy luật chuẩn tắc $N(0;1)$.

Đúng ☐ Sai ☐

3.8 Ta có thể tính gần đúng quy luật nhị thức $\mathscr{B}(n;p)$ bằng phân bố chuẩn nếu n đủ lớn và p > 0,1. Khi $n > 50$ và $p < 0,1$ thì xấp xỉ bằng phân bố Poisson.

☐ ☐

Đúng Sai .

3.9 Biến ngẫu nhiên phân bố theo quy luật Student chỉ nhận những giá trị dương.

Đúng ☐ Sai ☐

3.10 Biến ngẫu nhiên liên tục $F \sim \mathbf{F}(n_1, n_2)$ phân bố theo quy luật Fisher-Snedecor với n_1 và n_2 bậc tự do có tính chất đối xứng đối với n_1 và n_2.

Đúng ☐ Sai ☐

3.11 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.

3.12 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.

3.13 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 tuân theo quy luật 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.

3.14 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.

3.15 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.

3.16 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.

3.17 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.

3.18 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.

3.19 Ở một tổng đài bưu điện 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.

3.20 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 Poát-xông 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.

3.21 Số khách hàng vào một siêu thị trong một giờ là biến ngẫu nhiên tuân theo quy luật 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.

3.22 Biến ngẫu nhiên X tuân theo quy luật chuẩn với kỳ vọng $\mu = 10$ và phương sai $\sigma^2 = 4$. Tính xác suất để X nhận giá trị trong khoảng (8; 12).

3.23 Biến ngẫu nhiên X tuân theo quy luật 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).

3.24 Trọng lượng sản phẩm X do một máy tự động sản xuất là một biến ngẫu nhiên tuân theo quy luật 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).

3.25 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.

3.26 Cho X_i $(i = \overline{1,n})$là các biến ngẫu nhiên độc lập, cùng tuân theo quy luật chuẩn với

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

Lập công thức tính $P\left\{\left|\overline{X} - \mu\right| < \varepsilon\right\}$ biết rằng $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ và cũng tuân theo quy luật chuẩn, $\varepsilon > 0$ tùy ý.

CHƯƠNG IV: BIẾN NGẪU NHIÊN HAI CHIỀU

GIỚI THIỆU

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 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, còn biến ngẫu nhiên liên tục được xác định bởi hàm mật độ xác suất.

Trong chương này ta chỉ xét biến ngẫu nhiên hai chiều rời rạc.

Từ bảng phân bố xác suất đồng thời của biến ngẫu nhiên hai chiều rời rạc, dựa vào công thức cộng xác suất đầy đủ ta 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 rời rạc 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, 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 đối với biến ngẫu nhiên thành phần kia 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.

NỘI DUNG

1. KHÁI NIỆM BIẾN NGẪU NHIÊN NHIỀU CHIỀU

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. 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 $\left(X_1, X_2, ..., X_n\right)$ với các thành phần $X_1, X_2, ..., X_n$ là các biến ngẫu nhiên. Ta ký hiệu biến ngẫu nhiên hai chiều là (X, Y), trong đó X là biến ngẫu nhiên thành phần thứ nhất và Y là biến ngẫu nhiên thành phần thứ hai.

Ví dụ 4.1: 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.

Biến ngẫu nhiên n chiều $\left(X_1, X_2, ..., X_n\right)$ 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.

2. BẢNG PHÂN BỐ XÁC SUẤT CỦA BIẾN NGẪU NHIÊN RỜI RẠC HAI CHIỀU

2.1. 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 nó và các xác suất tương ứng.

$X \backslash Y$	y_1	y_2	…	y_j	…	y_m	$\sum_j$
x_1	$p(x_1, y_1)$	$p(x_1, y_2)$	…	$p(x_1, y_j)$	…	$p(x_1, y_m)$	$p(x_1)$
x_2	$p(x_2, y_1)$	$p(x_2, y_2)$	…	$p(x_2, y_j)$	…	$p(x_2, y_m)$	$p(x_2)$
$\vdots$	$\vdots$	$\vdots$	…	$\vdots$	…	$\vdots$	$\vdots$
x_i	$p(x_i, y_1)$	$p(x_i, y_2)$	…	$p(x_i, y_j)$	…	$p(x_i, y_m)$	$p(x_i)$
$\vdots$	$\vdots$	$\vdots$	…	$\vdots$	…	$\vdots$	$\vdots$
x_n	$p(x_n, y_1)$	$p(x_n, y_2)$	…	$p(x_n, y_j)$	…	$p(x_n, y_m)$	$p(x_n)$
$\sum_i$	$p(y_1)$	$p(y_2)$	…	$p(y_j)$	…	$p(y_m)$	1

Trong đó x_i $(i = \overline{1, n})$ là các giá trị có thể có của thành phần X; y_j $(j = \overline{1, m})$ là các giá trị có thể có của thành phần Y. $p(x_i, y_j)$ là xác suất đồng thời của biến ngẫu nhiên hai chiều (X, Y) nhận giá trị (x_i, y_j), nghĩa là:

$$p(x_i, y_j) = P\left\{X = x_i, Y = y_j\right\},$$

các xác suất này thỏa mãn:

$$\begin{cases} p(x_i, y_j) \ge 0,\ \forall i = \overline{1,n},\ j = \overline{1,m} \\ \sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i, y_j) = 1 \end{cases} \quad \textbf{(4.1)}$$

2.2. Bảng phân bố xác suất biên

Áp dụng công thức xác suất đầy đủ (1.16) cho hệ $\{X = x_1\}, \{X = x_2\}, ..., \{X = x_n\}$ (xem công thức 2.7) ta có:

$$p(y_j) = P\left\{Y = y_j\right\} = \sum_{i=1}^{n} P\left\{X = x_i, Y = y_j\right\} = \sum_{i=1}^{n} p(x_i, y_j)\ ;\ \ j = \overline{1,m} \quad \textbf{(4.2)}$$

$$p(x_i) = P\left\{X = x_i\right\} = \sum_{j=1}^{m} P\left\{X = x_i, Y = y_j\right\} = \sum_{j=1}^{m} p(x_i, y_j)\ ;\ \ i = \overline{1,n} \quad \textbf{(4.3)}$$

Như vậy từ bảng phân bố xác suất đồng thời của (X,Y), nếu ta cộng các xác suất 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.

X	x_1	x_2	$\dots$	x_i	$\dots$	x_n
P	$p(x_1)$	$p(x_2)$	$\dots$	$p(x_i)$	$\dots$	$p(x_n)$

Y	y_1	y_2	$\dots$	y_j	$\dots$	y_m
P	$p(y_1)$	$p(y_2)$	$\dots$	$p(y_j)$	$\dots$	$p(y_m)$

Ví dụ 4.2: 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 8 kết quả đồng khả năng của việc 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	**B**	**C**	X	Y
N	N	N	2	3
N	N	S	2	2
N	S	N	1	2
N	S	S	1	1
S	N	N	1	1
S	N	S	1	2
S	S	N	0	1
S	S	S	0	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 \ 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

Cộng các cột ta được phân bố xác suất của biến ngẫu nhiên thành phần X :

X	0	1	2
P	2/8	4/8	2/8

Cộng các hàng ta được phân bố xác suất của biến ngẫu nhiên thành phần Y :

Y	0	1	2	3
P	1/8	3/8	3/8	1/8

Ví dụ 4.3: 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	$\sum$
1	2/36	3/36	1/36	1/6
2	4/36	6/36	2/36	2/6
3	6/36	9/36	3/36	3/6
$\sum$	2/6	3/6	1/6	1

2.3. Quy luật phân bố xác suất có điều kiện của các thành phần

Áp dụng công thức tính xác suất có điều kiện (1.11) và (1.11)' ta suy ra rằng biến ngẫu nhiên thành phần X với điều kiện biến cố $\{Y = y_j\}$ xảy ra nhận các giá trị $x_1, x_2, ..., x_n$ với bảng phân bố xác suất điều kiện tương ứng:

$X \mid Y = y_j$	x_1	x_2	…	x_i	…	x_n
P	$p(x_1 \mid y_j)$	$p(x_2 \mid y_j)$	…	$p(x_i \mid y_j)$	…	$p(x_n \mid y_j)$

trong đó:

$$p(x_i \mid y_j) = \frac{p(x_i, y_j)}{p(y_j)}\ ;\ i = \overline{1,n},\ \ j = \overline{1,m}\ . \tag{4.4}$$

Tương tự ta cũng có bảng phân bố xác suất có điều kiện của Y với điều kiện $\{X = x_i\}$

$Y \mid X = x_i$	y_1	y_2	…	y_j	…	y_m
P	$p(y_1 \mid x_i)$	$p(y_2 \mid x_i)$	…	$p(y_j \mid x_i)$	…	$p(y_m \mid x_i)$

trong đó:

$$p(y_j|x_i) = \frac{p(x_i, y_j)}{p(x_i)} \; ; \; i = \overline{1,n}, \;\; j = \overline{1,m} \; . \qquad \textbf{(4.5)}$$

Từ công thức xác suất có điều kiện (4.4)-(4.5) ta có công thức tính xác suất đồng thời:

$$p(x_i, y_j) = p(x_i)p(y_j|x_i) = p(y_j)p(x_i|y_j) \; ; \; i = \overline{1,n}, \;\; j = \overline{1,m} \; . \qquad \textbf{(4.6)}$$

Ví dụ 4.4: Phân bố xác suất có điều kiện của biến ngẫu nhiên Y với điều kiện $\{X = 1\}$ của ví dụ 4.2.

$Y\|X = 1$	0	1	2	3
P	0	2/4	2/4	0

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.

Vậy biến ngẫu nhiên hai chiều rời rạc (X,Y) với phân bố xác suất (4.1) là độc lập khi và chỉ khi:

$$p(x_i, y_j) = p(x_i)p(y_j) \;\; \forall i = \overline{1,n}, \;\; j = \overline{1,m} \; . \qquad \textbf{(4.7)}$$

Một dấu hiệu để nhận biết hai biến ngẫu nhiên độc lập là bảng phân bố xác suất đồng thời có tính chất:

- Hai hàng bất kỳ tỉ lệ.
- Hai cột bất kỳ tỉ lệ.

Ví dụ 4.5: Hai biến ngẫu nhiên X và Y của ví dụ 4.2 không độc lập vì $P\{X = 2\} = \frac{2}{8}$, $P\{Y = 1\} = \frac{3}{8}$ nhưng $P\{X = 2, Y = 1\} = 0$.

Hai biến ngẫu nhiên X và Y của ví dụ 4.3 độc lập vì thỏa mãn điều kiện (4.7).

3. CÁC THAM SỐ ĐẶC TRƯNG CỦA BIẾN NGẪU NHIÊN HAI CHIỀU

3.1. Kỳ vọng và phương sai của các biến ngẫu nhiên thành phần

Từ bảng phân bố xác suất thành phần (4.2)-(4.3) ta có công thức tính kỳ vọng và phương sai:

$$\mathrm{E}X=\sum_{i=1}^{n}x_i p(x_i)=\sum_{i=1}^{n}\sum_{j=1}^{m}x_i p(x_i,y_j) \quad \textbf{(4.8)}$$

$$\mathrm{E}Y=\sum_{j=1}^{m}y_j p(y_j)=\sum_{j=1}^{m}\sum_{i=1}^{n}y_j p(x_i,y_j) \quad \textbf{(4.9)}$$

$$\mathrm{D}X=\mathrm{E}X^2-(\mathrm{E}X)^2=\sum_{i=1}^{n}\sum_{j=1}^{m}x_i^2 p(x_i,y_j)-(\mathrm{E}X)^2 \quad \textbf{(4.10)}$$

$$\mathrm{D}Y=\mathrm{E}Y^2-(\mathrm{E}Y)^2=\sum_{j=1}^{m}\sum_{i=1}^{n}y_j^2 p(x_i,y_j)-(\mathrm{E}Y)^2 \quad \textbf{(4.11)}$$

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 các biến ngẫu nhiên đó với kỳ vọng toán của chúng:

$$\mathrm{cov}(X,Y)=\mathrm{E}\big(X-\mathrm{E}X\big)\big(Y-\mathrm{E}Y\big) \quad \textbf{(4.12)}$$

Nếu X,Y rời rạc có bảng phân bố đồng thời (4.1) thì:

$$\mathrm{cov}(X,Y)=\sum_{j=1}^{m}\sum_{i=1}^{n}x_i y_j p(x_i,y_j)-\mathrm{E}X\,\mathrm{E}Y \quad \textbf{(4.13)}$$

Tính chất

1) $\mathrm{cov}(X,Y)=\mathrm{E}XY-\mathrm{E}X\,\mathrm{E}Y$.

2) $\mathrm{cov}(X,Y)=\mathrm{cov}(Y,X)$.

3) $\mathrm{cov}(X,X)=\mathrm{D}X$.

4) $\mathrm{cov}(aX+c,bY+d)=ab\,\mathrm{cov}(Y,X)$ với mọi hằng số a,b,c,d.

5) Nếu X,Y độc lập thì $\mathrm{cov}(X,Y)=0$ nhưng ngược lại chưa chắc đúng.

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:

$$\rho_{X,Y}=\frac{\mathrm{cov}(X,Y)}{\sigma_X\sigma_Y} \quad \textbf{(4.14)}$$

Tính chất

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

$$\rho_{aX+c,bY+d} = \begin{cases} \rho_{X,Y} & \text{nếu } ab > 0 \\ -\rho_{X,Y} & \text{nếu } ab < 0 \end{cases} \quad \textbf{(4.15)}$$

4) $Y = aX + b$, $a \neq 0$ khi và chỉ khi

$$\rho_{X,Y} = \begin{cases} 1 & \text{nếu } a > 0 \\ -1 & \text{nếu } a < 0 \end{cases} \quad \textbf{(4.16)}$$

Ý nghĩa của hệ số tương quan:

Hệ số tương quan đo mức độ phụ thuộc tuyến tính giữa X và Y. Khi $|\rho_{X,Y}|$ càng gần 1 thì tính chất quan hệ tuyến tính càng chặt, khi $|\rho_{X,Y}|$ 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.

3.4. Kỳ vọng có điều kiện, hàm hồi quy

Từ luật phân bố điều kiện (4.5), kỳ vọng của biến ngẫu nhiên Y với điều kiện $\{X = x_i\}$ được ký hiệu và tính theo công thức sau:

$$\mathrm{E}(Y|X = x_i) = \sum_{j=1}^{m} y_j p(y_j|x_i) \quad \textbf{(4.17)}$$

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.

$$f(x) = \mathrm{E}(Y|X = x) \quad \textbf{(4.18)}$$

Tương tự ta có định nghĩa kỳ vọng có điều kiện của X với điều kiện $\{Y = y\}$ và hàm hồi quy của X đối với Y:

$$\mathrm{E}(X|Y = y) = \sum_{i=1}^{n} x_i p(x_i|y) \quad \textbf{(4.19)}$$

$$g(y) = \mathrm{E}(X|Y = y) \quad \textbf{(4.20)}$$

Hàm hồi qui (4.18) có miền xác định là miền giá trị của X, đó là tập $\{x_1, x_2, ..., x_n\}$. Tương tự, hàm hồi qui (4.20) có miền xác định là miền giá trị của Y, đó là tập $\{y_1, y_2, ..., y_m\}$. Khi X, Y liên tục thì miền xác định của hàm hồi qui là một khoảng nào đó.

Ví dụ 4.6: 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.

Tìm thu nhập trung bình theo lứa tuổi.

X \ Y	30	45	70
1	0,01	0,02	0,05
2	0,03	0,06	0,10
3	0,18	0,21	0,15
4	0,07	0,08	0,04

trong đó $X = 1, 2, 3, 4$ 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.

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:

$X \vert Y = 30$	0	1	2	3
P	$\frac{0,01}{0,29}$	$\frac{0,03}{0,29}$	$\frac{0,18}{0,29}$	$\frac{0,07}{0,29}$

Từ đó $\quad \mathrm{E}(X|Y=30) = 0\cdot\frac{1}{29} + 1\cdot\frac{3}{29} + 2\cdot\frac{18}{29} + 3\cdot\frac{7}{29} = 2{,}069$.

Tương tự $\mathrm{E}(X|Y=45) = 1{,}946$; $\mathrm{E}(X|Y=70) = 1{,}529$.

Vậy thu nhập trung bình ở độ tuổi 30 là 2.069.000đ/tháng, độ tuổi 45 là 1.946.000đ/tháng và độ tuổi 70 là 1.529.000 đ/tháng.

TÓM TẮT

Biến ngẫu nhiên nhiều chiều

Một biến ngẫu nhiên n chiều là một bộ $(X_1, X_2, ..., X_n)$ với các thành phần $X_1, X_2, ..., X_n$ là các biến ngẫu nhiên. Biến ngẫu nhiên n chiều 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.

Ta ký hiệu biến ngẫu nhiên hai chiều là (X, Y), trong đó X là biến ngẫu nhiên thành phần thứ nhất và Y là biến ngẫu nhiên thành phần thứ hai.

Bảng phân bố xác suất đồng thời của biến ngẫu nhiên rời rạc hai chiều

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 nó và các xác suất tương ứng, trong đó x_i $(i = \overline{1, n})$ là các giá trị có thể có của thành phần X; y_j $(j = \overline{1, m})$ là các giá trị có thể có của thành phần Y: $p(x_i, y_j) = P\{X = x_i, Y = y_j\}$

Bảng phân bố xác suất biên

$$p(y_j) = P\{Y = y_j\} = \sum_{i=1}^{n} P\{X = x_i, Y = y_j\} = \sum_{i=1}^{n} p(x_i, y_j) \; ; \quad j = \overline{1, m}$$

$$p(x_i) = P\{X = x_i\} = \sum_{j=1}^{m} P\{X = x_i, Y = y_j\} = \sum_{j=1}^{m} p(x_i, y_j) \; ; \quad i = \overline{1, n}$$

Quy luật phân bố xác suất có điều kiện của các thành phần

$X \mid Y = y_j$	x_1	x_2	…	x_i	…	x_n
P	$p(x_1 \mid y_j)$	$p(x_2 \mid y_j)$	…	$p(x_i \mid y_j)$	…	$p(x_n \mid y_j)$

$$p(x_i \mid y_j) = \frac{p(x_i, y_j)}{p(y_j)} \; ; \; i = \overline{1, n}, \; j = \overline{1, m} \, .$$

$Y \mid X = x_i$	y_1	y_2	…	y_j	…	y_m
P	$p(y_1 \mid x_i)$	$p(y_2 \mid x_i)$	…	$p(y_j \mid x_i)$	…	$p(y_m \mid x_i)$

$$p(y_j \mid x_i) = \frac{p(x_i, y_j)}{p(x_i)} \; ; \; i = \overline{1, n}, \; j = \overline{1, m} \, .$$

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.

Trường hợp hai biến ngẫu nhiên X và Y rời rạc nhận các giá trị $x_i, y_j; \forall i = \overline{1,n}, j = \overline{1,m}$ thì tính chất độc lập tương đương với:

$$p(x_i, y_j) = p(x_i)p(y_j) \ \ \forall i = \overline{1,n}, \ \ j = \overline{1,m} \ .$$

Kỳ vọng và phương sai của các biến ngẫu nhiên thành phần

$$\mathrm{E}X = \sum_{i=1}^{n} x_i p(x_i) = \sum_{i=1}^{n}\sum_{j=1}^{m} x_i p(x_i, y_j)$$

$$\mathrm{E}Y = \sum_{j=1}^{m} y_j p(y_j) = \sum_{j=1}^{m}\sum_{i=1}^{n} y_j p(x_i, y_j)$$

$$\mathrm{D}X = \mathrm{E}X^2 - (\mathrm{E}X)^2 = \sum_{i=1}^{n}\sum_{j=1}^{m} x_i^2 p(x_i, y_j) - (\mathrm{E}X)^2$$

$$\mathrm{D}Y = \mathrm{E}Y^2 - (\mathrm{E}Y)^2 = \sum_{j=1}^{m}\sum_{i=1}^{n} y_j^2 p(x_i, y_j) - (\mathrm{E}Y)^2$$

Hiệp phương sai

$$\mathrm{cov}(X,Y) = \mathrm{E}\big(X - \mathrm{E}X\big)\big(Y - \mathrm{E}Y\big) = \sum_{j=1}^{m}\sum_{i=1}^{n} x_i y_j p(x_i, y_j) - \mathrm{E}\,X\,\mathrm{E}\,Y \,.$$

Hệ số tương quan

Hệ số tương quan của hai biến ngẫu nhiên X, Y ký hiệu và định nghĩa bởi công thức:

$$\rho_{X,Y} = \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y}$$

Kỳ vọng có điều kiện

Kỳ vọng có điều kiện của Y với điều kiện $\{X = x_i\}$: $\mathrm{E}\big(Y \big| X = x_i\big) = \sum_{j=1}^{m} y_j p\big(y_j \big| x_i\big)$.

Kỳ vọng có điều kiện của X với điều kiện $\{Y = y\}$: $\mathrm{E}\big(X \big| Y = y\big) = \sum_{i=1}^{n} x_i p\big(x_i \big| y\big)$.

Hàm hồi quy

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: $f(x) = \mathrm{E}(Y|X = x)$.

Hàm hồi quy của X đối với Y: $g(y) = \mathrm{E}(X|Y = y)$.

CÂU HỎI ÔN TẬP VÀ BÀI TẬP

4.1 Bảng phân bố xác suất của X và Y cho phép xác định phân bố xác suất đồng thời của (X,Y).

Đúng ☐ Sai ☐

4.2 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 ☐

4.3 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 ☐

4.4 Hai biến ngẫu nhiên độc lập có hiệp phương sai bằng 0.

Đúng ☐ Sai ☐

4.5 Hai biến ngẫu nhiên có hiệp phương sai bằng 0 thì độc lập.

Đúng ☐ Sai ☐

4.6 Hiệp phương sai luôn nhận giá trị dương.

Đúng ☐ Sai ☐

4.7 Nếu $Y = aX + b$, $a \neq 0$ thì hệ số tương quan $\rho_{X,Y}$ luôn luôn bằng 1.

Đúng ☐ Sai ☐

4.8 Nếu $\{x_1, x_2, ..., x_n\}$ là tập giá trị của X thì $\{f(x_1), f(x_2), ..., f(x_n)\}$ là tập giá trị của hàm hồi quy $f(x) = \mathrm{E}(Y|X = x)$ của Y đối với X.

Đúng ☐ Sai ☐

4.9 Nếu hai biến ngẫu nhiên X, Y độc lập thì hàm hồi quy $f(x) = \mathrm{E}(Y|X = x)$ của Y đối với X và hàm hồi quy $g(y) = \mathrm{E}(X|Y = y)$ của X đối với Y là hai hàm hằng.

Đúng ☐ Sai ☐

4.10 Nếu hiệp phương sai của hai biến ngẫu nhiên bằng 0 thì hai kỳ vọng của chúng bằng nhau ($\mathrm{cov}(X,Y) = 0$ thì $\mathrm{E}X = \mathrm{E}Y$).

Đúng ☐ Sai ☐

4.11 Cho X, Y là hai biến ngẫu nhiên có phân bố xác suất đồng thời như sau

$X \backslash Y$	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.

4.12 Cho X, Y là hai biến ngẫu nhiên có phân bố xác suất đồng thời như sau

$X \backslash Y$	-1	1
-1	1/6	1/4
0	1/6	1/8
1	1/6	1/8

Hãy xác định $\mathrm{E}X, \mathrm{E}Y$, $\mathrm{cov}(X,Y)=0$ và $\rho_{X,Y}$.

4.13 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 xác định $\mathrm{E}X, \mathrm{E}Y$, $\mathrm{cov}(X,Y)=0$ và $\rho_{X,Y}$. X, Y có độc lập không?

4.14 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 $\mathrm{E}X, \mathrm{E}Y, \mathrm{E}Z$.

4.15 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 $P\{X > Y\}$.

4.16 Cho X, Y là hai biến ngẫu nhiên có phân bố xác suất đồng thời như sau:

X \ 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 $P\{X = 1 | Y = 2\}$.

4.17 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.

4.18 Cho bảng phân bố xác suất đồng thời của X, Y:

Y \ X	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 $X = 26$ và của X khi $Y = 27$.

4.19 Cho bảng phân bố xác suất đồng thời của X, Y:

Y \ X	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 Y khi $X = 1$.

b) Tìm các kỳ vọng EX, EY và phương sai DX, DY.

4.20 Cho X, Y là hai ĐLNN 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 $\operatorname{cov}(X,Y)$, $\rho(X,Y)$.

c) X và Y có độc lập không.

CHƯƠNG V: LUẬT SỐ LỚN (*)

GIỚI THIỆU

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, Máckốp, Liapunốp mở rộng.

Trong chương này ta xét hai định lý về luật số lớn. Định lý Trêbưsép là dạng tổng quát của luật số lớn còn định lý Bernoulli là trường hợp đơn giản nhất cho các biến ngẫu nhiên có phân bố không – một.

Để chứng minh định lý Trêbưsép ta sử dụng bất đẳng thức Trêbưsép.

NỘI DUNG

5.1. BẤT ĐẲNG THỨC TRÊBƯSÉP

Định lý 5.1: Cho Y là biến ngẫu nhiên không âm có kỳ vọng hữu hạn. Khi đó với mọi $a>0$ ta có:

$$P\{Y>a\}\le\frac{\mathrm{E}Y}{a}. \qquad \textbf{(5.1)}$$

Chứng minh:

a) Trường hợp Y rời rạc có tập giá trị $V=\{y_1,y_2,...\}$.

Đặt $V_1=\{y_i\in V, y_i\le a\}$; $V_2=\{y_i\in V, y_i>a\}$. Khi đó

$$\mathrm{E}Y=\sum_{y_i\in V}y_iP\{Y=y_i\}=\sum_{y_i\in V_1}y_iP\{Y=y_i\}+\sum_{y_i\in V_2}y_iP\{Y=y_i\}$$

$$\ge\sum_{y_i\in V_2}y_iP\{Y=y_i\}\ge a\sum_{y_i\in V_2}P\{Y=y_i\}=aP\{Y>a\}.$$

Suy ra $\quad P\{Y > a\} \le \dfrac{\mathrm{E}Y}{a}$.

b) Giả sử Y liên tục có hàm mật độ $f(x)$. Ta có

$$\mathrm{E}Y = \int_0^{+\infty} xf(x)dx = \int_0^{a} xf(x)dx + \int_a^{+\infty} xf(x)dx \ge \int_a^{+\infty} xf(x)dx \ge a\int_a^{+\infty} f(x)dx = aP\{Y > a\}.$$

Suy ra $\quad P\{Y > a\} \le \dfrac{\mathrm{E}Y}{a}$.

Hệ quả: Nếu X là biến ngẫu nhiên có kỳ vọng và phương sai hữu hạn thì với mọi $\varepsilon > 0$ ta có:

$$P\{|X - \mathrm{E}X| > \varepsilon\} \le \frac{\mathrm{D}X}{\varepsilon^2}. \qquad \textbf{(5.2)}$$

cũng vậy:

$$P\{|X - \mathrm{E}X| \le \varepsilon\} \ge 1 - \frac{\mathrm{D}X}{\varepsilon^2}. \qquad \textbf{(5.3)}$$

Chứng minh: Áp dụng công thức (5.1) cho biến ngẫu nhiên $Y = (X - \mathrm{E}X)^2$ và $a = \varepsilon^2$ ta có:

$$P\{|X - \mathrm{E}X| > \varepsilon\} = P\{Y > \varepsilon^2\} \le \frac{\mathrm{E}Y}{\varepsilon^2} = \frac{\mathrm{E}(X - \mathrm{E}X)^2}{\varepsilon^2} = \frac{\mathrm{D}X}{\varepsilon^2}.$$

Bất đẳng thức (5.2)-(5.3) 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.

Ví dụ 5.1: Một cửa hàng muốn ước lượng nhanh chóng sai số của số vải bán ra trong một tháng của mình. Biết rằng số vải của mỗi khách hàng được làm tròn bởi số nguyên M gần nhất (ví dụ trong sổ ghi 195,6 *m* thì làm tròn là 196*m*).

Giải: Ký hiệu X_i là sai số giữa số mét vải thực bán và số mét vải đã làm tròn của khách hàng thứ *i*. Các sai số $X_1, X_2, ..., X_n$ là các biến ngẫu nhiên độc lập có phân bố đều trên đoạn $[-0,5; 0,5]$. Khi đó $\mathrm{E}X_i = 0, \mathrm{D}X_i = \dfrac{1}{12}$. Sai số tổng cộng trong cả tháng là $S = X_1 + \cdots + X_n$ (trong đó n là số khách hàng mua hàng trong tháng). Ta có:

$$\mathrm{E}S = \sum_{i=1}^{n} \mathrm{E}X_i = 0,\ \mathrm{D}S = \sum_{i=1}^{n} \mathrm{D}X_i = \frac{n}{12}.$$

Theo bất đẳng thức Trêbưsép, xác suất để sai số vượt quá ε mét sẽ được đánh giá bởi:

$$P\{|S| > \varepsilon\} \le \frac{DS}{\varepsilon^2} = \frac{n}{12\varepsilon^2}.$$

Giả sử có $n = 10^4$ khách hàng trong tháng. Để xác suất $P\{|S| > \varepsilon\}$ bé hơn 0,01 ta phải có $\frac{n}{12\varepsilon^2} \le 0,01$ hay $\varepsilon \ge \sqrt{\frac{n}{12 \cdot 0,01}} = 288,67$.

Vậy ta có thể kết luận: Với xác suất 0,99 sai số giữa số vải thực bán với số vải đã tính tròn không vượt quá 289 *m*, nếu số khách hàng là 1 vạn.

5.2. LUẬT SỐ LỚN

5.2.1. Hội tụ theo xác suất

Định nghĩa 5.1: Dãy các biến ngẫu nhiên $X_1, X_2, \ldots$ gọi là hội tụ theo xác suất về biến ngẫu nhiên X, ký hiệu $X_n \xrightarrow[n\to\infty]{P} X$, nếu:

$$\forall \varepsilon > 0 \quad \lim_{n\to\infty} P\{|X_n - X| > \varepsilon\} = 0. \qquad \textbf{(5.4)}$$

Như vậy dãy các biến ngẫu nhiên $X_1, X_2, \ldots$ hội tụ theo xác suất về biến ngẫu nhiên X thì với n đủ lớn, thực tế ta có thể coi rằng X_n không khác mấy so với X.

5.2.2. Luật số lớn Trêbưsép

Định lý 5.2: Giả sử $X_1, X_2, \ldots$ là dãy các biến ngẫu nhiên độc lập, có các kỳ vọng hữu hạn và phương sai đều bị chặn trên bởi hằng số C ($DX_i \le C;\ \forall i = 1, 2, \ldots$). Khi đó:

$$\lim_{n\to\infty} P\left(\left|\frac{X_1 + \cdots + X_n}{n} - \frac{EX_1 + \cdots + EX_n}{n}\right| > \varepsilon\right) = 0 \qquad \textbf{(5.5)}$$

Chứng minh: Xét biến ngẫu nhiên $S_n = \frac{X_1 + \cdots + X_n}{n}$. Từ giả thiết độc lập của dãy các biến ngẫu nhiên $X_1, X_2, \ldots$ ta suy ra:

$$ES_n = \frac{EX_1 + \cdots + EX_n}{n};\ DS_n = \frac{DX_1 + \cdots + DX_n}{n^2} \le \frac{C}{n}.$$

Áp dụng bất đẳng thức Trêbưsép (5.2) cho biến ngẫu nhiên S_n ta có:

$$P\left(\left|\frac{X_1 + \cdots + X_n}{n} - \frac{EX_1 + \cdots + EX_n}{n}\right| > \varepsilon\right) \le \frac{C}{n\varepsilon^2} \xrightarrow[n\to\infty]{} 0.$$

Hệ quả 1: Giả sử $X_1, X_2, ...$ là dãy các biến ngẫu nhiên độc lập có cùng có kỳ vọng μ và phương sai đều bị chặn trên bởi hằng số C ($\mathrm{D}X_i \le C$; $\forall i = 1, 2, ...$). Khi đó:

$$\frac{X_1 + \cdots + X_n}{n} \xrightarrow[n\to\infty]{P} \mu \tag{5.6}$$

Hệ quả 2: 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:

$$\frac{X_1 + \cdots + X_n}{n} \xrightarrow[n\to\infty]{P} \mu \tag{5.7}$$

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 của một số lớn các biến ngẫu nhiên lại nhận giá trị gần bằng trung bình số học 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 nốt xuất hiện ở mặt trên con xúc xắc. Theo lý thuyết ta tính được $\mathrm{E}X = 3,5$. 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 nốt 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 $X_1, X_2, ..., X_n$. 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.

5.3. LUẬT SỐ LỚN BERNOULLI

Xét phép thử ngẫu nhiên $\mathscr{C}$ và A là một biến cố liên quan đến phép thử $\mathscr{C}$. Tiến hành phép thử $\mathscr{C}$ n lần độc lập và gọi k_n là tần số xuất hiện biến cố A trong n phép thử đó. $f_n = \frac{k_n}{n}$ được gọi là tần suất xuất hiện của A trong n phép thử.

Định lý 5.3: Tần suất f_n hội tụ theo xác suất về xác suất p của biến cố A, nghĩa là với mọi $\varepsilon > 0$:

$$\lim_{n\to\infty} P\{|f_n - p| > \varepsilon\} = 0 \qquad \textbf{(5.8)}$$

Chứng minh: Xét dãy các biến ngẫu nhiên $X_1, X_2, ..., X_n$ xác định như sau:

$$X_k = \begin{cases} 1 & \text{nếu } A \text{ xảy ra ở phép thử thứ } k \\ 0 & \text{nếu } A \text{ không xảy ra ở phép thử thứ } k \end{cases}$$

thì dãy các biến ngẫu nhiên $X_1, X_2, ..., X_n$ độc lập có cùng phân bố không – một $A(p)$.

Ta có $\dfrac{X_1 + \cdots + X_n}{n} = \dfrac{k_n}{n} = f_n$.

Vậy theo hệ quả 2 của định lý 5.2 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 vậy ta thấy rằng khi số phép thử tăng lên thì tần suất tương ứng sẽ càng gần 0,5.

TÓM TẮT

Bất đẳng thức Trêbưsép

Cho Y là biến ngẫu nhiên không âm có kỳ vọng hữu hạn. Khi đó với mọi $a > 0$ ta có: $P\{Y > a\} \le \dfrac{\mathrm{E}Y}{a}$.

Nếu X là biến ngẫu nhiên có kỳ vọng và phương sai hữu hạn thì với mọi $\varepsilon > 0$ ta có:

$$P\{|X - \mathrm{E}X| > \varepsilon\} \le \frac{\mathrm{D}X}{\varepsilon^2} \text{ và } P\{|X - \mathrm{E}X| \le \varepsilon\} \ge 1 - \frac{\mathrm{D}X}{\varepsilon^2}.$$

Luật số lớn Trêbưsép

Giả sử $X_1, X_2, ...$ là dãy các biến ngẫu nhiên độc lập, có các kỳ vọng hữu hạn và phương sai đều bị chặn trên bởi hằng số C ($\mathrm{D}X_i \le C;\ \forall i = 1, 2, ...$). Khi đó:

$$\lim_{n\to\infty} P\left(\left|\frac{X_1+\cdots+X_n}{n}-\frac{\mathrm{E}X_1+\cdots+\mathrm{E}X_n}{n}\right|>\varepsilon\right)=0.$$

Giả sử $X_1, X_2,...$ là dãy các biến ngẫu nhiên độc lập có cùng có kỳ vọng μ và phương sai đều bị chặn trên bởi hằng số C ($\mathrm{D}X_i \le C;\ \forall i=1,2,...$) hoặc độc lập có cùng phân bố, có kỳ vọng μ và phương sai σ^2. Khi đó $\frac{X_1+\cdots+X_n}{n} \xrightarrow[n\to\infty]{P} \mu$.

Luật số lớn Bernoulli

Tần suất xuất hiện biến cố A trong n phép thử f_n hội tụ theo xác suất về xác suất p của biến cố A, nghĩa là với mọi $\varepsilon > 0$ $\lim_{n\to\infty} P\{|f_n - p| > \varepsilon\} = 0$.

CÂU HỎI ÔN TẬP VÀ BÀI TẬP

5.1 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 ☐

5.2 Giả sử $\{X_n\}$ là dãy các biến ngẫu nhiên 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 ☐

5.3 Bất đẳng thức Trêbưsép chỉ đúng đối với các biến ngẫu nhiên rời rạc.

Đúng ☐ Sai ☐

5.4 Bất đẳng thức Trêbưsép chỉ đúng đối với các biến ngẫu nhiên nhận giá trị dương.

Đúng ☐ Sai ☐

5.5 Luật số lớn Bernoulli là một trường hợp đặc biết của luật số lớn Trêbưsép khi dãy các biến ngẫu nhiên được có cùng phân bố không – một A(p) .

Đúng ☐ Sai ☐

5.6 Luật số lớn Bernoulli là cơ sở lý thuyết của định nghĩa thống kê về xác suất.

Đúng ☐ Sai ☐

5.7 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

5.8 Cho $X_1, X_2, ..., X_{12}$ là các biến ngẫu nhiên độc lập với $EX_i = 16$, $DX_i = 1$ ($i = \overline{1,12}$). Sử dụng bất đẳng thức Trêbưsép để tìm hai hằng số a, b sao cho:

$$P\left\{a \le \sum_{i=1}^{12} X_i \le b\right\} \ge 0{,}99 .$$

5.9 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 $P\left\{\left|\sum_{i=1}^{10000} X_i\right| \ge 500\right\} \ge \frac{1}{300}$.

5.10 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 $P\left\{\frac{n}{6} - \sqrt{n} < S < \frac{n}{6} + \sqrt{n}\right\} \ge \frac{31}{36}$.

5.11 Giả sử tiền điện của một gia đình phải trả trong 1 tháng là một biến ngẫu nhiên với trung bình 16USD và độ lệch tiêu chuẩn 1USD. 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.

5.12 Cho dãy các biến ngẫu nhiên độc lập $\{X_n\}$ xác định như sau:

X_n	-2^n	0	2^n
P	$2^{-(2n+1)}$	$1-2^{-2n}$	$2^{-(2n+1)}$

Chứng minh rằng dãy $\{X_n\}$ thỏa mãn luật số lớn Trêbưsép.

5.13 Cho dãy các biến ngẫu nhiên độc lập $\{X_n\}$ xác định như sau:

X_n	$-na$	0	na
P	$\frac{1}{2n^2}$	$1-\frac{1}{n^2}$	$\frac{1}{2n^2}$

trong đó a là một hằng số. Dãy $\{X_n\}$ thỏa mãn luật số lớn Trêbưsép không?

5.14 Cho dãy các biến ngẫu nhiên độc lập $\{X_n\}$ xác định như sau:

X_n	$-a$	a
P	$\frac{n+1}{2n+1}$	$\frac{n}{2n+1}$

trong đó a là một hằng số. Dãy $\{X_n\}$ thỏa mãn luật số lớn Trêbưsép không?

5.15 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.

5.16 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 VI: CƠ SỞ LÝ THUYẾT MẪU

GIỚI THIỆU

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. 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. Vì vậy cần lấy mẫu để nghiên cứu.

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:

- Các phương pháp mô tả mẫu:
  - Bảng phân bố tần số thực nghiệm, bảng phân bố tần suất thực nghiệm, bảng phân bố ghép lớp.
  - Biểu đồ tần số hình gậy, đa giác tần suất và tổ chức 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 hai chương tiếp.

NỘI DUNG

6.1. KHÁI NIỆM VỀ PHƯƠNG PHÁP MẪU

Nhiều bài toán trong thực tế dẫn đến nghiên cứu một hay nhiều dấu hiệu định tính hoặc định lượng đặc trưng cho 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, dấu hiệu nghiên cứu là thu nhập của từng hộ gia đình. Một doanh nghiệp muốn nghiên cứu các khách hàng của mình về dấu hiệu định tính có thể là mức độ 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, độ dài, thời gian, áp suất, 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.
- Có thể trong quá trình điều tra sẽ phá hủy đối tượng nghiên cứu …

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.

6.2. TỔNG THỂ NGHIÊN CỨU

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 $\mathscr{C}$.

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 phân 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 $\mathscr{C}$ nhận một giá trị thực nào đó thì dấu hiệu này được gọi là một *biến lượng*, ký hiệu X. Bằng cách mô hình hóa ta có thể xem biến lượng X là một biến ngẫu nhiên xác định trên tổng thể $\mathscr{C}$.

Việc chọn ra từ tổng thể một tập con nào đó gọi là *phép lấy mẫu*. Tập hợp con này được gọi là *một mẫu*.

6.3. MẪU NGẪU NHIÊN

6.3.1. Các phương pháp chọn mẫu

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.

6.3.2. Định nghĩa mẫu ngẫu nhiên

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.

Giả sử các cá thể của tổng thể được nghiên cứu thông qua dấu hiệu X .

Với mẫu ngẫu nhiên kích thước n, gọi X_i là dấu hiệu X của phần tử thứ i của mẫu ($i = \overline{1,n}$). Bằng cách đồng nhất mẫu ngẫu nhiên với các dấu hiệu nghiên cứu của mẫu ta có định nghĩa về 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 $W = (X_1, X_2, ..., X_n)$.

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 ($i = \overline{1,n}$), 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 6.1: Gọi X là số nốt xuất hiện khi tung con xúc xắc cân đối, X là biến ngẫu nhiên có bảng phân bố xác suất sau:

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

Nếu tung con xúc xắc 3 lần và gọi X_i là số nốt xuất hiện trong lần tung thứ i ($i = \overline{1,3}$) 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 2 nốt, lần thứ hai được 5 nốt lần ba được 3 nốt thì $w = (2,5,3)$ là một mẫu cụ thể của mẫu ngẫu nhiên W.

6.3.3. Các phương pháp mô tả mẫu ngẫu nhiên

6.3.3.1. Bảng phân bố tần số thực nghiệm

Nếu một mẫu ngẫu nhiên kích thước n của X nhận giá trị x_i với tần số xuất hiện r_i, $i = 1, ..., k$ trong đó:

$$x_1 < \cdots < x_k \; ; \; r_1 + \cdots + r_k = n . \tag{6.1}$$

Khi đó ta có thể mô tả mẫu ngẫu nhiên trên qua bảng phân bố tần số thực nghiệm:

$$\begin{array}{c|cccc} X & x_1 & x_2 & \cdots & x_k \\ \hline \text{Tần số} & r_1 & r_2 & \cdots & r_k \end{array} \tag{6.2}$$

6.3.3.2. 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:

$$\begin{array}{c|cccc} X & x_1 & x_2 & \cdots & x_k \\ \hline \text{Tần suất} & f_1 & f_2 & \cdots & f_k \end{array} \tag{6.3}$$

Ví dụ 6.2: Lấy một mẫu ngẫu nhiên kích thược 120 ta có bảng phân bố thực nghiệm tần số và tần suất:

X	31	34	35	36	38	40	42	44	$\sum$
Tần số	10	20	30	15	10	10	5	20	120
Tần suất	2/24	4/24	6/24	3/24	2/24	2/24	1/24	4/24	1

6.3.3.3. Hàm phân bố thực nghiệm của mẫu

Với mẫu ngẫu nhiên xác định bới công thúc (6.1). Hàm số xác định như sau:

$$F_n(x) = \sum_{x_j < x} \frac{r_j}{n} \; ; \; -\infty < x < +\infty \tag{6.4}$$

gọi là hàm phân bố thực nghiệm của mẫu đã cho.

Định lý Glivenco chỉ ra rằng hàm phân bố thực nghiệm $F_n(x)$ xấp xỉ với phân bố lý thuyết $F(x) = P\{X < x\}$ khi n đủ lớn.

6.3.3.4. 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 tùy 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.

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	9 − 16	17 − 32	33 − 64	65 − 127	129 − 256	257 − 512
k : số khoảng	4	5	6	7	8	9

Ví dụ 6.3: Một mẫu về chiều cao của 400 cây con trong một vườn ươm được trình bày trong bảng phân bố ghép lớp sau:

Khoảng	Tần số r_i	Tần suất f_i	Độ rộng khoảng 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.

Qui ước: Đầu mút bên phải của mỗi khoảng thuộc vào khoảng đó mà không thuộc khoảng tiếp theo khi tính tần số của mỗi lớp.

Trong ví dụ trên ta có các khoảng $[4,5;9,5]$, $(9,5;11,5]$, $(11,5;13,5]$,...

6.3.3.5. Biểu diễn bằng biểu đồ

Giả sử dấu hiệu điều tra X có bảng phân bố tần số và tần suất thực nghiệm:

X	x_1	x_2	$\cdots$	x_k
Tần số	r_1	r_2	$\cdots$	r_k

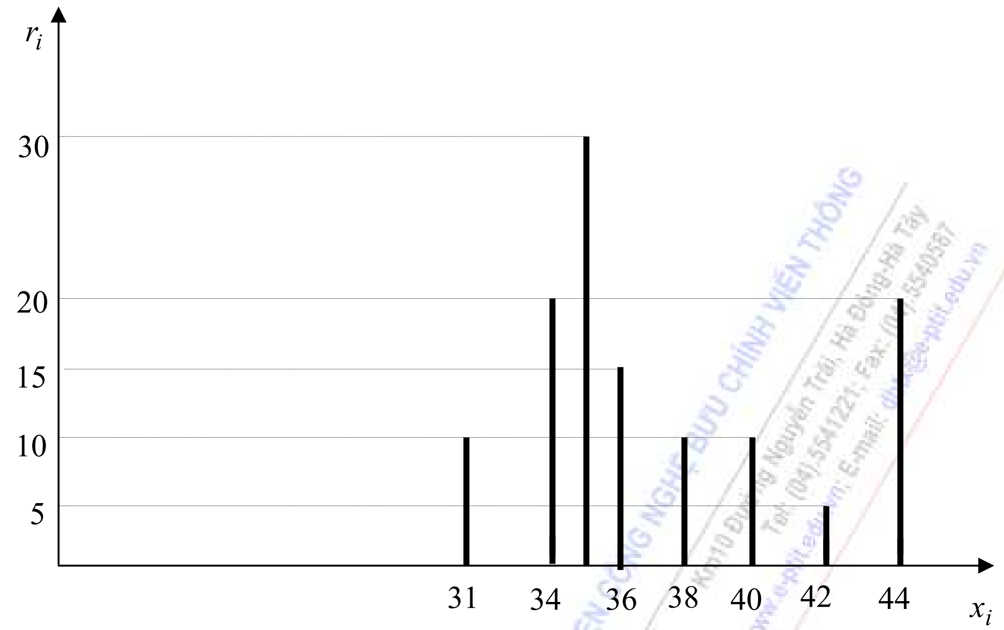
X	x_1	x_2	$\cdots$	x_k
Tần suất	f_1	f_2	$\cdots$	f_k

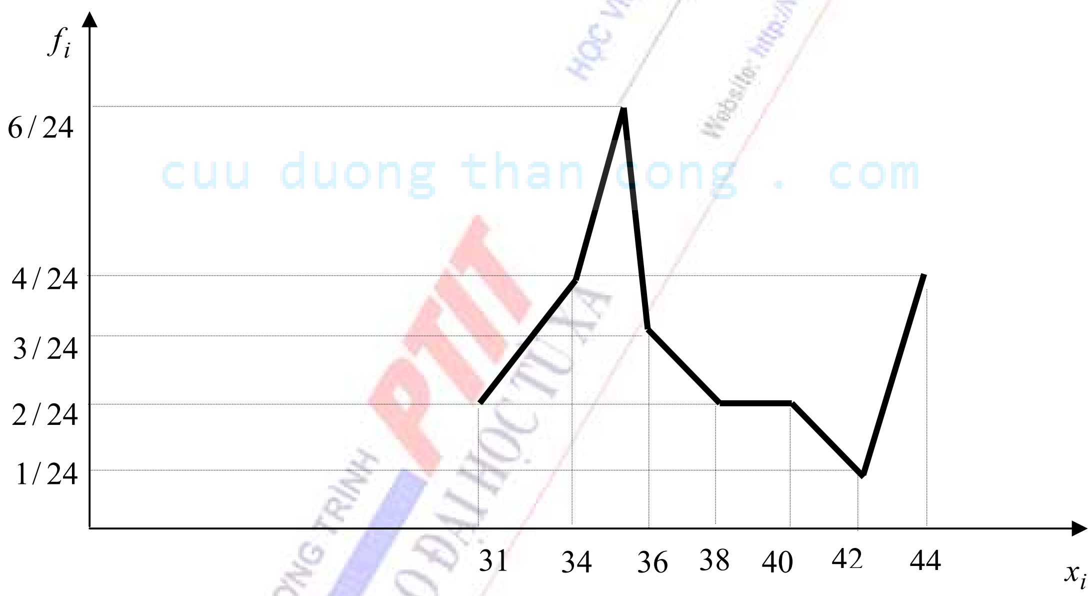
Trong mặt phẳng với hệ trục toạ độ Oxy.

❖ Nối điểm trên trục hoành có toạ độ $(x_i, 0)$ với điểm có toạ độ (x_i, r_i); $i = \overline{1,k}$ 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}); $i = \overline{1, k-1}$ ta được *biểu đồ đa giác tần suất*.

Bảng phân bố tần số và tần suất thực nghiệm trong ví dụ 6.2 có biểu đồ tần số hình gậy và đa giác tần suất



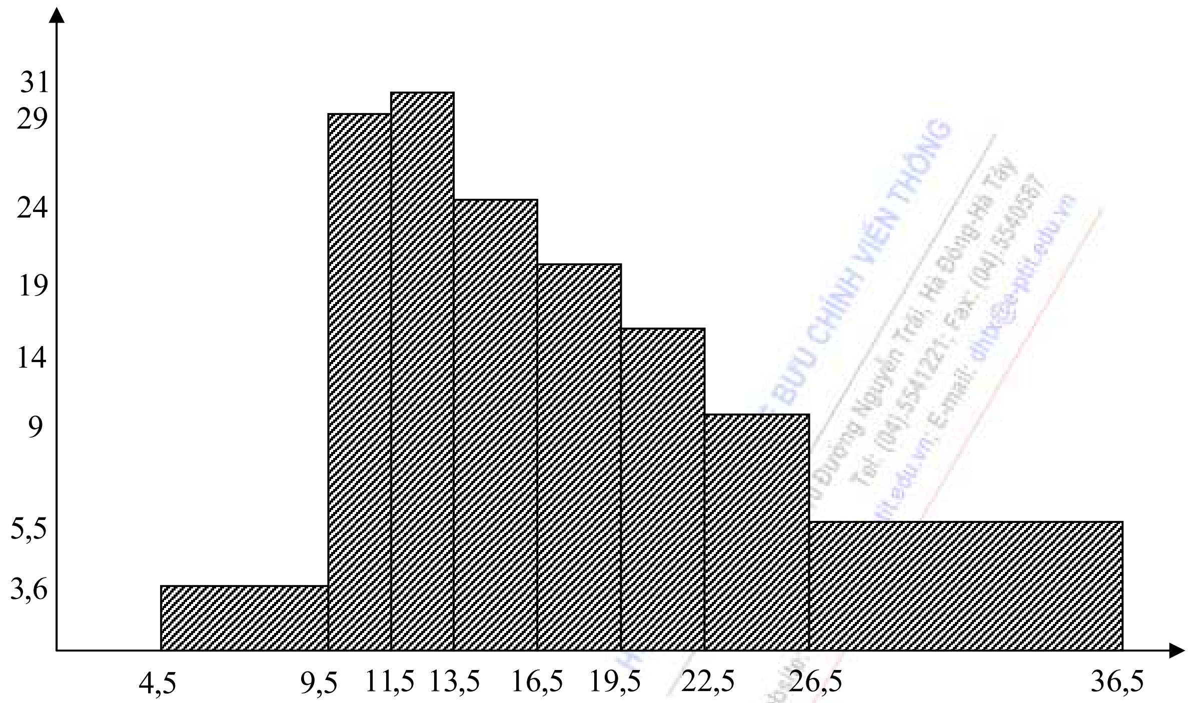


6.3.3.6. Tổ chức đồ (histogram)

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 6.3



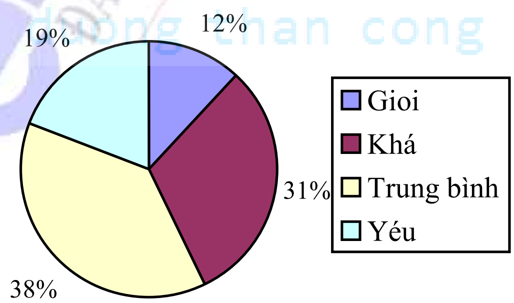
Chú ý rằng diện tích giới hạn bởi tổ chức đồ bằng tần số xuất hiện. Chẳng hạn số cây nằm trong khoảng (12; 25] chính là diện tích của tổ chức đồ giới hạn bởi đường thẳng $x=12$ và $x=25$.

$$(13,5-12)\times 31+(16,5-13,5)\times 24+(19,5-16,5)\times 19+(22,5-19,5)\times 14+(25-22,5)\times 9=240$$

Vậy có 240 cây có chiều cao từ $12m$ đến $25m$.

Với các dấu hiệu điều tra là định tính thì 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 6.4: Tổng kết kết quả học tập của sinh viên Học viện ta trong năm 2005 được số liệu sau:



6.4. THỐNG KÊ VÀ CÁC ĐẶC TRƯNG CỦA MẪU NGẪU NHIÊN

6.4.1. Định nghĩa thống kê

Một thống kê của mẫu là một hàm của các biến ngẫu nhiên thành phần của mẫu. Thống kê của mẫu ngẫu nhiên $W=(X_1,X_2,...,X_n)$ có dạng:

$$T=T(X_1,X_2,...,X_n) \tag{6.5}$$

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 gọi là giá trị quan sát được của thống kê:

$$T_{qs}=T(x_1,x_2,...,x_n)$$

Các thống kê cùng với quy luật phân bố xác suất của chúng là cơ sở để suy rộng các thông tin của mẫu cho dấu hiệu nghiên cứu của tổng thể.

6.4.2. Một số thống kê đặc trưng của mẫu ngẫu nhiên

Có hai nhóm chứa các số đặc trưng mẫu:

❖ 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.

Ta sẽ xem xét một số thống kê đặc trưng mẫu quan trọng sau:

6.4.2.1. 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:

$$\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i \tag{6.6}$$

Giá trị trung bình mẫu cụ thể của mẫu ngẫu nhiên cụ thể $w=(x_1,x_2,...,x_n)$ là:

$$\overline{x}=\frac{1}{n}\sum_{i=1}^{n}x_i \tag{6.7}$$

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 (2.18), (2.19), (2.26) ta có:

$$\mathrm{E}(\overline{X}) = \mathrm{E}X \; ; \; \mathrm{D}(\overline{X}) = \frac{\mathrm{D}X}{n} . \qquad \textbf{(6.8)}$$

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:

$$\mathrm{Se}(\overline{X}) = \sigma_{\overline{X}} = \frac{\sqrt{\mathrm{D}X}}{\sqrt{\mathrm{n}}} . \qquad \textbf{(6.9)}$$

6.4.2.2. Tổng bình phương các sai lệch và trung bình của tổng bình phương các sai lệch (*)

Tổng bình phương các sai lệch giữa các giá trị của mẫu và trung bình mẫu của mẫu ngẫu nhiên $W = (X_1, X_2, ..., X_n)$ được ký hiệu và xác định bởi công thức sau:

$$\mathrm{SS} = \sum_{i=1}^{n} (X_i - \overline{X})^2 \qquad \textbf{(6.10)}$$

Nếu trong mẫu các giá trị X_i xuất hiện với tần số n_i $(i = \overline{1,k})$ và $\sum_{i=1}^{k} n_i = n$ thì:

$$\mathrm{SS} = \sum_{i=1}^{k} n_i (X_i - \overline{X})^2 \qquad \textbf{(6.11)}$$

Trung bình của tổng bình phương các sai lệch giữa các giá trị của mẫu và trung bình mẫu của mẫu ngẫu nhiên $W = (X_1, X_2, ..., X_n)$ được ký hiệu và xác định bởi công thức sau:

$$\mathrm{MS} = \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})^2 \text{ hoặc } \mathrm{MS} = \frac{1}{n} \sum_{i=1}^{k} n_i (X_i - \overline{X})^2 \qquad \textbf{(6.12)}$$

Ta cũng có thể chứng minh được:

$$\mathrm{MS} = \frac{1}{n} \sum_{i=1}^{k} n_i X_i - (\overline{X})^2 \qquad \textbf{(6.13)}$$

Nếu dấu hiệu nghiên cứu X có kỳ vọng và phương sai hữu hạn thì:

$$\mathrm{E(MS)} = \frac{n-1}{n} \mathrm{D}X . \qquad \textbf{(6.14)}$$

6.4.2.3. Phương sai mẫu

- Phương sai mẫu $\hat{S}^2$:

$$\hat{S}^2 = \mathrm{MS} = \frac{1}{n} \sum_{i=1}^{k} n_i (X_i - \overline{X})^2 = \frac{1}{n} \sum_{i=1}^{k} n_i X_i^2 - (\overline{X})^2 \qquad \textbf{(6.15)}$$

- Phương sai mẫu có hiệu chỉnh S^2:

$$S^2 = \frac{n}{n-1}\text{MS} = \frac{1}{n-1}\sum_{i=1}^{k} n_i\left(X_i - \overline{X}\right)^2 = \frac{1}{n-1}\sum_{i=1}^{k} n_i X_i^2 - \frac{n}{n-1}\left(\overline{X}\right)^2 \qquad \textbf{(6.15)}$$

- Phương sai mẫu S^{*2}: khi dấu hiệu nghiên cứu X của tổng thể có kỳ vọng xác định $\mathrm{E}X = \mu$.

$$S^{*2} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2 \qquad \textbf{(6.16)}$$

Áp dụng công thức tính kỳ vọng (2.18), (2.19) và (6.15) ta có:

$$\mathrm{E}S^2 = DX \text{ và } \mathrm{E}S^{*2} = DX \qquad \textbf{(6.17)}$$

6.4.2.4. Độ lệch tiêu chuẩn mẫu

$$S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{k} n_i\left(X_i - \overline{X}\right)^2} \qquad \textbf{(6.18)}$$

6.4.2.5. Tần suất mẫu

Ta cần nghiên cứu một dấu hiệu nghiên cứu A nào đó mà mỗi cá thể của tổng thể có thể có hoặc không có dấu hiệu A. 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 quy luật phân bố xác suất không – một $A(p)$ có kỳ vọng và phương sai $\mathrm{E}X = p$; $\mathrm{D}X = p(1-p)$.

Lấy mẫu ngẫu nhiên kích thước n: $W = \left(X_1, X_2, ..., X_n\right)$, trong đó $X_1, X_2, ..., X_n$ là các biến ngẫu nhiên độc lập có cùng phân bố không – một $A(p)$. Tần số xuất hiện dấu hiệu A của mẫu là:

$$r = X_1 + X_2 + \cdots + X_n . \qquad \textbf{(6.19)}$$

Tần suất mẫu:

$$f = \frac{r}{n} \qquad \textbf{(6.20)}$$

Tương tự ta có công thức tính kỳ vọng và phương sai của tần suất mẫu:

$$\mathrm{E}(f) = p\,;\ \mathrm{D}(f) = \frac{p(1-p)}{n} \qquad \textbf{(6.21)}$$

Sai số chuẩn của tần suất mẫu:

$$\mathrm{Se}(f) = \frac{\sqrt{p(1-p)}}{\sqrt{n}} \qquad \textbf{(6.22)}$$

6.4.2.6. Cách tính giá trị cụ thể của trung bình mẫu và phương sai mẫu $\overline{x}$, s^2

1. Nếu mẫu chỉ nhận các giá trị $x_1, x_2, ..., x_k$ với tần số tương ứng $n_1, n_2, ..., n_k$ thì giá trị trung bình mẫu và phương sai mẫu cụ thể được tính theo công thức:

$$\overline{x} = \frac{\sum_{i=1}^{k} n_i x_i}{n}, \quad \sum_{i=1}^{k} n_i = n \tag{6.23}$$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{k} n_i \left(x_i - \overline{x}\right)^2 = \frac{1}{n-1}\left(\sum_{i=1}^{k} n_i x_i^2 - \frac{\left(\sum_{i=1}^{k} n_i x_i\right)^2}{n}\right) \tag{6.24}$$

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 n_i thì giá trị trung bình mẫu và phương sai mẫu được tính như 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:

$$u_i = \frac{x_i - a}{h} \Rightarrow x_i = hu_i + a \Rightarrow \overline{x} = h\overline{u} + a; \quad s_x^2 = h^2 s_u^2 \tag{6.25}$$

Thật vậy: $\overline{x} = \sum_{i=1}^{k} n_i x_i = \sum_{i=1}^{k} n_i \left(hu_i + a\right) = h\sum_{i=1}^{k} n_i u_i + a = h\overline{u} + a$.

$$s_x^2 = \frac{1}{n-1}\sum_{i=1}^{k} n_i \left(x_i - \overline{x}\right)^2 = \frac{1}{n-1}\sum_{i=1}^{k} n_i \left(hu_i + a - h\overline{u} - a\right)^2 = \frac{h^2}{n-1}\sum_{i=1}^{k} n_i \left(u_i - \overline{u}\right)^2 = h^2 s_u^2$$

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.

Ví dụ 6.5: Giá trị trung bình mẫu và phương sai mẫu của mẫu ở ví dụ 6.3.

Khoảng	tần số n_i	x_i	$u_i = \dfrac{x_i - 20}{5}$	$n_i u_i$	$n_i u_i^2$
$4{,}5 - 9{,}5$	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

$$\overline{x} = 5\overline{u} + 20 = 5 \times \frac{-177,5}{400} + 20 = 17,78 .$$

$$s_u^2 = \frac{1}{399} \times \left(873,47 - \frac{(-177,5)^2}{400} \right) = 1,9917 \Rightarrow$$

$$s_x^2 = 5^2 \times s_u^2 = 49,79 \Rightarrow s = \sqrt{49,79} = 7,056 .$$

6.5. MẪU NGẪU NHIÊN HAI CHIỀU

6.5.1. Khái niệm mẫu ngẫu nhiên hai chiều

Giả sử trên cùng một tổng thể phải nghiên cứu 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:

$$W = [(X_1, Y_1), (X_2, Y_2), ..., (X_n, Y_n)]$$

Giả sử thành phần (X_i, Y_i) nhận giá trị (x_i, y_i) $(i = \overline{1, n})$ ta thu được mẫu cụ thể:

$$w = [(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)].$$

Các giá trị x_i $(i = \overline{1, n})$ gọi là thành phần X của mẫu, còn các giá trị y_i $(i = \overline{1, n})$ gọi là thành phần Y của mẫu.

6.5.2. Phương pháp mô tả mẫu ngẫu nhiên hai chiều

Giả sử từ tổng thể ta rút ra được mẫu ngẫu nhiên cụ thể w kích thước n. Ta hãy sắp xếp các giá trị thành phần X và Y của mẫu theo thứ tự tăng dần:

$$x_1 < x_2 < \ldots < x_i < \ldots < x_h \text{ và } y_1 < y_2 < \ldots < y_j < \ldots < y_k .$$

Ký hiệu n_{ij} là tần số của cặp giá trị (x_i, y_j) của mẫu w, rõ ràng các tần số n_{ij} thỏa mãn hệ thức $\sum_{j=1}^{k}\sum_{i=1}^{h} n_{ij} = 1$. Lúc đó mẫu cụ thể w được mô tả bằng bảng phân bố xác suất tần số thực nghiệm sau:

X \ Y	y_1	y_2	$\ldots$	y_j	$\ldots$	y_k	n_{i*}
x_1	n_{11}	n_{12}	$\ldots$	n_{1j}	$\ldots$	n_{1k}	n_{1*}
x_2	n_{21}	n_{22}	$\ldots$	n_{2j}	$\ldots$	n_{2k}	n_{2*}
$\vdots$	$\vdots$	$\vdots$	$\ldots$	$\vdots$	$\ldots$	$\vdots$	$\vdots$
x_i	n_{i1}	n_{i2}	$\ldots$	n_{ij}	$\ldots$	n_{ik}	n_{i*}
$\vdots$	$\vdots$	$\vdots$	$\ldots$	$\vdots$	$\ldots$	$\vdots$	$\vdots$
x_h	n_{h1}	n_{h2}	$\ldots$	n_{hj}	$\ldots$	n_{hk}	n_{h*}
n_{*j}	n_{*1}	n_{*2}	$\ldots$	n_{*j}	$\ldots$	n_{*k}	$\sum = n$

trong đó n_{i*} là tần số của giá trị $x_i\,(i = \overline{1,h})$ của thành phần X, n_{*j} là tần số của giá trị $y_j\,(j = \overline{1,k})$ của thành phần Y.

6.5.3. Một số thống kê đặc trưng của mẫu ngẫu nhiên hai chiều

Từ bảng phân bố thực nghiệm của mẫu ngẫu nhiên hai chiều ta có thể rút ra:

1. Bảng phân bố thực nghiệm của thành phần X:

X	x_1	x_2	$\ldots$	x_i	$\ldots$	x_h
Tần số	n_{1*}	n_{2*}	$\ldots$	n_{i*}	$\ldots$	n_{h*}

Các thống kê đặc trưng của X:

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{h} n_{i*}X_i \;;\; S_X^2 = \frac{1}{n-1}\sum_{i=1}^{h} n_{i*}\left(X_i - \overline{X}\right)^2 \quad \textbf{(6.26)}$$

2. Bảng phân bố thực nghiệm của thành phần Y

Y	y_1	y_2	$\dots$	y_j	$\dots$	y_k
Tần số	n_{*1}	n_{*2}	$\dots$	n_{*j}	$\dots$	n_{*k}

Các thống kê đặc trưng của Y:

$$\overline{Y} = \frac{1}{n}\sum_{j=1}^{k} n_{*j} Y_j \; ; \quad S_Y^2 = \frac{1}{n-1}\sum_{j=1}^{k} n_{*j}\left(Y_j - \overline{Y}\right)^2 \tag{6.27}$$

3. Hệ số tương quan mẫu:

$$r = \frac{1}{n}\cdot\frac{\sum_{i=1}^{h}\sum_{j=1}^{k} n_{ij}(X_i - \overline{X})(Y_j - \overline{Y})}{\sqrt{\mathrm{MS}_X \mathrm{MS}_Y}} \tag{6.28}$$

6.6. QUY LUẬT PHÂN BỐ XÁC SUẤT CỦA MỘT SỐ THỐNG KÊ ĐẶC TRƯNG MẪU (*)

6.6.1. Trường hợp biến ngẫu nhiên gốc tuân theo quy luật phân bố chuẩn

Giả sử dấu hiệu nghiên cứu trong tổng thể có thể xem như một biến ngẫu nhiên X tuân theo quy luật 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:

$$W = \left(X_1, X_2, \dots, X_n\right)$$

Các biến ngẫu nhiên thành phần $X_1, X_2, \dots, X_n$ độc lập có cùng quy luật phân bố chuẩn $N(\mu, \sigma^2)$ như X. Theo công thức (3.20) mọi tổ hợp tuyến tính của các biến ngẫu nhiên có phân bố chuẩn là biến ngẫu nhiên có phân bố chuẩn. Vì vậy ta có các kết quả sau:

1. Thống kê trung bình mẫu:

Trung bình mẫu $\overline{X}$ có quy luật phân bố chuẩn với kỳ vọng $\mathrm{E}\left(\overline{X}\right) = \mu$ và phương sai $\mathrm{D}\left(\overline{X}\right) = \frac{\sigma^2}{n}$. Theo công thức (3.26), biến ngẫu nhiên sau có quy luật chuẩn tắc $N(0;1)$

$$U = \frac{(\overline{X} - \mu)\sqrt{n}}{\sigma} \sim N(0;1) \tag{6.29}$$

2. Thống kê phương sai $\hat{S}^2$.

Từ công thức (6.16) ta có: $n\hat{S}^2 = \sum_{i=1}^{n}(X_i - \mu)^2$ và $\frac{n\hat{S}^2}{\sigma^2} = \sum_{i=1}^{n}\left(\frac{X_i - \mu}{\sigma}\right)^2$.

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 (3.26) $\frac{X_i - \mu}{\sigma} \sim N(0;1)$. Do đó biến ngẫu nhiên sau có phân bố "Khi bình phương" n bậc tự do.

$$\chi^2 = \frac{n\hat{S}^2}{\sigma^2} = \sum_{i=1}^{n}\left(\frac{X_i - \mu}{\sigma}\right)^2 \sim \chi^2(n) \quad \textbf{(6.30)}$$

3. Thống kê phương sai S^2.

Có thể chứng minh được thống kê sau có phân bố "Khi bình phương" $n-1$ bậc tự do.

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2} = \sum_{i=1}^{n}\left(\frac{X_i - \overline{X}}{\sigma}\right)^2 \sim \chi^2(n-1) \quad \textbf{(6.31)}$$

4. Hơn nữa, từ (6.29), (6.31) và U, χ^2 độc lập, áp dụng công thức (3.39) thì:

$T = \frac{U}{\sqrt{\chi^2/(n-1)}}$ có phân bố Student $n-1$ bậc tự do.

$$T = \frac{U}{\sqrt{\chi^2/(n-1)}} = \frac{\dfrac{(\overline{X}-\mu)\sqrt{n}}{\sigma}}{\sqrt{\dfrac{(n-1)S^2}{\sigma^2(n-1)}}} = \frac{(\overline{X}-\mu)\sqrt{n}}{S} \sim \mathbf{T}(n-1) \quad \textbf{(6.32)}$$

Khi n khá lớn thì quy luật Student $\mathbf{T}(n)$ hội tụ khá nhanh về phân bố chuẩn tắc $N(0;1)$, do đó trong thực tế nếu $n > 30$ ta có thể xem thống kê T xấp xỉ $N(0;1)$.

6.6.2. Trường hợp có hai biến ngẫu nhiên gốc cùng phân bố theo quy luật 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 $N(\mu_1;\sigma_1^2)$. Ở tổng thể thứ hai dấu hiệu nghiên cứu được xem như biến ngẫu nhiên X_2 có phân bố chuẩn $N(\mu_2;\sigma_2^2)$.

Từ hai tổng thể nói trên rút ra hai mẫu ngẫu nhiên độc lập kích thước tương ứng là n_1 và n_2

$$W_1 = (X_{11}, X_{12}, ..., X_{1n_1})\ ; \qquad W_2 = (X_{21}, X_{22}, ..., X_{2n_2})\ .$$

1. Thống kê hiệu của hai trung bình mẫu $\overline{X}_1 - \overline{X}_2$.

Đây là một tổ hợp tuyến tính của các phân bố chuẩn do đó cũng có quy luật phân bố chuẩn. Mặt khác kỳ vọng và phương sai

$$\mathrm{E}\left(\overline{X}_1 - \overline{X}_2\right) = \mathrm{E}\left(\overline{X}_1\right) - \mathrm{E}\left(\overline{X}_2\right) = \mu_1 - \mu_2 .$$

$$\mathrm{D}\left(\overline{X}_1 - \overline{X}_2\right) = \mathrm{D}\left(\overline{X}_1\right) + \mathrm{D}\left(\overline{X}_2\right) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} .$$

Do đó thống kê sau có phân bố chuẩn tắc:

$$U = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \sim N(0;1) . \qquad \textbf{(6.33)}$$

2. Hai thống kê phương sai thành phần:

$$\chi_1^2 = \frac{(n_1 - 1)S_1^2}{\sigma_1^2} \sim \chi^2(n_1 - 1) \, ; \; \chi_2^2 = \frac{(n_2 - 1)S_2^2}{\sigma_2^2} \sim \chi^2(n_2 - 1) \text{ và độc lập.} \qquad \textbf{(6.34)}$$

Do đó thống kê:

$$\chi^2 = \frac{(n_1 - 1)S_1^2}{\sigma_1^2} + \frac{(n_2 - 1)S_2^2}{\sigma_2^2} \sim \chi^2(n_1 + n_2 - 2) . \qquad \textbf{(6.35)}$$

Từ (6.33)-(6.35) suy ra:

♦ Nếu $\sigma_1^2 = \sigma_2^2$ thì thống kê:

$$T = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}} \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} \sim \mathbf{T}(n_1 + n_2 - 2) . \qquad \textbf{(6.36)}$$

có phân bố Student $n_1 + n_2 - 2$ bậc tự do.

♦ Nếu $\sigma_1^2 \neq \sigma_2^2$ thì thống kê:

$$T = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}} \sim \mathbf{T}(k) . \qquad \textbf{(6.37)}$$

có phân bố Student k bậc tự do, trong đó:

$$k=\frac{(n_1-1)(n_2-1)}{(n_2-1)C^2+(n_1-1)(1-C)^2}\ ;\ C=\frac{\frac{S_1^2}{n_1}}{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}.$$

Từ (6.34) và (3.43) suy ra:

$$F=\frac{\frac{\chi_1^2}{n_1-1}}{\frac{\chi_2^2}{n_2-1}}=\frac{S_1^2\sigma_2^2}{S_2^2\sigma_1^2}\sim F[(n_1-1);(n_2-1)]. \qquad \textbf{(6.38)}$$

có phân bố Fisher-Snedecor với n_1-1 và n_2-1 bậc tự do.

6.6.3. Trường hợp biến ngẫu nhiên gốc phân bố theo quy luật không – một

Định lý: (***Định lý giới hạn trung tâm***) Giả sử $X_1, X_2, ...$ là dãy các biến ngẫu nhiên độc lập có cùng phân bố, với 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 $N(0;1)$, nghĩa là: Với mọi $x\in\mathbb{R}$

$$\lim_{n\to\infty}P\{S_n<x\}=\Phi(x) \qquad \textbf{(6.38)}$$

$\Phi(x)$ là hàm phân bố của phân bố chuẩn tắc $N(0;1)$.

Áp dụng định lý giới hạn trung tâm cho dãy các ĐLNN độc lập $X_1, X_2, ...$ có cùng phân bố không – một $A(p)$ ta được định lý Moivre –Laplace:

$$\text{Với mọi } x\in\mathbb{R},\ \lim_{n\to\infty}P\left\{\frac{X_1+\cdots+X_n-np}{\sqrt{npq}}<x\right\}=\Phi(x). \qquad \textbf{(6.39)}$$

Giả sử trong tổng thể dấu hiệu nghiên cứu có thể xem như biến ngẫu nhiên phân bố theo quy luật không – một. Từ tổng thể rút ra một mẫu ngẫu nhiên kích thước n:

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

Từ công thức (6.19), (6.20), (6.21) ta đã biết tần suất mẫu $f=\dfrac{X_1+\cdots+X_n}{n}$ có quy luật nhị thức với các tham số đặc trưng là: $\mathrm{E}(f)=p$; $\mathrm{D}(f)=\dfrac{pq}{n}$. Do đó công thức (6.39) trở thành:

$$\text{Với mọi } x \in \mathbb{R}, \ \lim_{n\to\infty} P\left\{\frac{(f-p)\sqrt{n}}{\sqrt{pq}} < x\right\} = \Phi(x). \qquad \textbf{(6.40)}$$

Định lý Moivre-Laplace cho phép xấp xỉ thống kê $U = \dfrac{(f-p)\sqrt{n}}{\sqrt{pq}}$ với phân bố chuẩn tắc $N(0;1)$ khi n đủ lớn. Người ta thấy rằng xấp xỉ là tốt khi $np > 5$ và $nq > 5$ hoặc $npq > 20$.

$$U = \frac{(f-p)\sqrt{n}}{\sqrt{pq}} \sim N(0;1) \text{ khi } \begin{cases} np > 5 \\ nq > 5 \end{cases} \text{ hoặc } npq > 20. \qquad \textbf{(6.41)}$$

6.6.4. Trường hợp có hai biến ngẫu nhiên gốc phân bố theo quy luật không - một

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 không – một 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:

$$W_1 = \left(X_{11}, X_{12}, \dots, X_{1n_1}\right); \ W_2 = \left(X_{21}, X_{22}, \dots, X_{2n_2}\right).$$

Xét thống kê $f_1 - f_2$ là hiệu hai tần suất mẫu. Lúc đó nếu $n_1 > 30$ và $n_2 > 30$ thì $f_1 - f_2$ sẽ có phân bố xấp xỉ chuẩn theo 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:

$$U = \frac{\left(f_1 - f_2\right) - (p_1 - p_2)}{\sqrt{\dfrac{p_1(1-p_1)}{n_1} + \dfrac{p_2(1-p_2)}{n_2}}} \sim N(0;1). \qquad \textbf{(6.42)}$$

TÓM TẮT

Tổng thể

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. Việc chọn ra từ tổng thể một tập con nào đó gọi là *phép lấy mẫu*. Tập hợp con này được gọi là *một mẫu*.

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.

Mẫu ngẫu nhiên của dấu hiệu nghiên cứu X

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

Bảng phân bố tần số thực nghiệm và tần suất thực nghiệm

Nếu một mẫu ngẫu nhiên kích thước n của X nhận giá trị x_i với tần số xuất hiện r_i, $i = 1, ..., k$: $x_1 < \cdots < x_k$; $r_1 + \cdots + r_k = n$. $f_i = \dfrac{r_i}{n}$ gọi là tần suất của x_i.

Khi đó ta có thể mô tả mẫu ngẫu nhiên trên qua bảng phân bố tần số thực nghiệm hoặc tần suất thực nghiệm của X

X	x_1	x_2	$\cdots$	x_k
Tần số	r_1	r_2	$\cdots$	r_k

X	x_1	x_2	$\cdots$	x_k
Tần suất	f_1	f_2	$\cdots$	f_k

Hàm phân bố thực nghiệm của mẫu

Hàm số xác định như sau: $F_n(x) = \sum_{x_j < x} \dfrac{r_j}{n}$; $-\infty < x < +\infty$.

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.

Biểu diễn bằng biểu đồ

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); $i = \overline{1,k}$ 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}); $i = \overline{1, k-1}$ ta được *biểu đồ đa giác tần suất*.

❖ Đối với bảng phân bố ghép lớp, người ta thường dùng tổ chức đồ để biểu diễn: 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).

Thống kê

Một thống kê của mẫu là một hàm của các biến ngẫu nhiên thành phần của mẫu. Thống kê của mẫu ngẫu nhiên $W=(X_1,X_2,...,X_n)$ có dạng:

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

Trung bình mẫu

Trung bình mẫu của mẫu ngẫu nhiên $W=(X_1,X_2,...,X_n)$ là: $\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$.

Trung bình của tổng bình phương các sai lệch giữa các giá trị của mẫu và trung bình mẫu

$$\mathrm{MS}=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2$$

Phương sai mẫu

- Phương sai mẫu có hiệu chỉnh S^2:

$$S^2=\frac{n}{n-1}\mathrm{MS}=\frac{1}{n-1}\sum_{i=1}^{k}n_i\left(X_i-\overline{X}\right)^2=\frac{1}{n-1}\sum_{i=1}^{k}n_iX_i^2-\frac{n}{n-1}\left(\overline{X}\right)^2$$

- Phương sai mẫu S^{*2} được xét khi dấu hiệu nghiên cứu X của tổng thể có kỳ vọng xác định $\mathrm{E}X=\mu$.

$$S^{*2}=\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2.$$

Độ lệch tiêu chuẩn mẫu

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

Tần suất mẫu

Xét biến ngẫu nhiên gốc X có phân bố không – một $A(p)$. Lấy mẫu ngẫu nhiên kích thước n: $W=(X_1,X_2,...,X_n)$, $X_1,X_2,...,X_n$ là các biến ngẫu nhiên độc lập có cùng phân bố với X. Tần số xuất hiện dấu hiệu A của mẫu là: $r=X_1+X_2+\cdots+X_n$. Tần suất mẫu $f=\frac{r}{n}$.

Cách tính giá trị cụ thể của trung bình mẫu và phương sai mẫu $\overline{x}$, s^2

Nếu mẫu chỉ nhận các giá trị $x_1,x_2,...,x_k$ với tần số tương ứng $n_1,n_2,...,n_k$ thì giá trị trung bình mẫu và phương sai mẫu cụ thể được tính theo công thức:

$$\overline{x} = \frac{\sum_{i=1}^{k} n_i x_i}{n}, \quad \sum_{i=1}^{k} n_i = n; \; s^2 = \frac{1}{n-1}\sum_{i=1}^{k} n_i \left(x_i - \overline{x}\right)^2 = \frac{1}{n-1}\left(\sum_{i=1}^{k} n_i x_i^2 - \frac{\left(\sum_{i=1}^{k} n_i x_i\right)^2}{n}\right)$$

Mẫu ngẫu nhiên hai chiều

Giả sử trên cùng một tổng thể phải nghiên cứu đồng thời hai dấu hiệu nghiên cứu, 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 được ký hiệu là:

$$W = [(X_1, Y_1), (X_2, Y_2), ..., (X_n, Y_n)]$$

Giả sử thành phần (X_i, Y_i) nhận giá trị (x_i, y_i) $(i = \overline{1, n})$ ta thu được mẫu cụ thể:

$$w = [(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)].$$

Phương pháp mô tả mẫu ngẫu nhiên hai chiều

Giả sử từ tổng thể rút ra được mẫu ngẫu nhiên cụ thể w kích thước n. Ta sắp xếp các giá trị thành phần X và Y của mẫu theo thứ tự tăng dần:

$$x_1 < x_2 < ... < x_i < ... < x_h \text{ và } y_1 < y_2 < ... < y_j < ... < y_k .$$

Ký hiệu n_{ij} là tần số của cặp giá trị (x_i, y_j) của mẫu w, $\sum_{j=1}^{k}\sum_{i=1}^{h} n_{ij} = 1$.

Lúc đó mẫu cụ thể w được mô tả bằng bảng phân bố xác suất tần số thực nghiệm sau:

$X \backslash Y$	y_1	y_2	…	y_j	…	y_k	n_{i*}
x_1	n_{11}	n_{12}	…	n_{1j}	…	n_{1k}	n_{1*}
x_2	n_{21}	n_{22}	…	n_{2j}	…	n_{2k}	n_{2*}
$\vdots$	$\vdots$	$\vdots$	…	$\vdots$	…	$\vdots$	$\vdots$
x_i	n_{i1}	n_{i2}	…	n_{ij}	…	n_{ik}	n_{i*}
$\vdots$	$\vdots$	$\vdots$	…	$\vdots$	…	$\vdots$	$\vdots$
x_h	n_{h1}	n_{h2}	…	n_{hj}	…	n_{hk}	n_{h*}
n_{*j}	n_{*1}	n_{*2}	…	n_{*j}	…	n_{*k}	$\sum = n$

trong đó n_{i*} là tần số của giá trị $x_i\,(i=\overline{1,h})$ của thành phần X, n_{*j} là tần số của giá trị $y_j\,(j=\overline{1,k})$ của thành phần Y.

Một số thống kê đặc trưng của mẫu ngẫu nhiên hai chiều

Bảng phân bố thực nghiệm của thành phần X:

X	x_1	x_2	...	x_i	...	x_h
Tần số	n_{1*}	n_{2*}	...	n_{i*}	...	n_{h*}

Các thống kê đặc trưng của X: $\overline{X}=\dfrac{1}{n}\sum_{i=1}^{h} n_{i*}X_i$; $S_X^2=\dfrac{1}{n-1}\sum_{i=1}^{h} n_{i*}\left(X_i-\overline{X}\right)^2$

Bảng phân bố thực nghiệm của thành phần Y:

Y	y_1	y_2	...	y_j	...	y_k
Tần số	n_{*1}	n_{*2}	...	n_{*j}	...	n_{*k}

Các thống kê đặc trưng của Y: $\overline{Y}=\dfrac{1}{n}\sum_{j=1}^{k} n_{*j}Y_j$; $S_Y^2=\dfrac{1}{n-1}\sum_{j=1}^{k} n_{*j}\left(Y_j-\overline{Y}\right)^2$

Hệ số tương quan mẫu: $r=\dfrac{\sum_{i=1}^{h}\sum_{j=1}^{k} n_{ij}(X_i-\overline{X})(Y_j-\overline{Y})}{n\sqrt{\mathrm{MS}_X\mathrm{MS}_Y}}$.

Quy luật phân bố xác suất của một số thống kê đặc trưng mẫu

Trường hợp biến ngẫu nhiên gốc tuân theo quy luật phân bố chuẩn:

$$U=\frac{(\overline{X}-\mu)\sqrt{n}}{\sigma}\sim N(0;1).$$

$$\chi^2=\frac{n\hat{S}^2}{\sigma^2}=\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2\sim\chi^2(n).$$

$$\chi^2=\frac{(n-1)S^2}{\sigma^2}=\sum_{i=1}^{n}\left(\frac{X_i-\overline{X}}{\sigma}\right)^2\sim\chi^2(n-1).$$

$$T=\frac{U}{\sqrt{\chi^2/(n-1)}}=\frac{\frac{(\overline{X}-\mu)\sqrt{n}}{\sigma}}{\sqrt{\frac{(n-1)S^2}{\sigma^2(n-1)}}}=\frac{(\overline{X}-\mu)\sqrt{n}}{S}\sim\mathbf{T}(n-1).$$

Trường hợp có hai biến ngẫu nhiên gốc cùng phân bố theo quy luật chuẩn:

$X_1\sim N(\mu_1;\sigma_1^2)$, $X_2\sim N(\mu_2;\sigma_2^2)$

$$U=\frac{(\overline{X}_1-\overline{X}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}\sim N(0;1).$$

$$\chi^2=\frac{(n_1-1)S_1^2}{\sigma_1^2}+\frac{(n_2-1)S_2^2}{\sigma_2^2}\sim\chi^2(n_1+n_2-2).$$

$$T=\frac{(\overline{X}_1-\overline{X}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim\mathbf{T}(n_1+n_2-2).$$

$$F=\frac{\frac{\chi_1^2}{n_1-1}}{\frac{\chi_2^2}{n_2-1}}=\frac{S_1^2\sigma_2^2}{S_2^2\sigma_1^2}\sim F[(n_1-1);(n_2-1)].$$

Trường hợp biến ngẫu nhiên gốc có phân bố theo quy luật không – một:

$$U=\frac{(f-p)\sqrt{n}}{\sqrt{pq}}\sim N(0;1)\ \text{khi}\ \begin{cases}np>5\\nq>5\end{cases}\ \text{hoặc}\ npq>20.$$

Trường hợp có hai biến ngẫu nhiên gốc có phân bố theo quy luật không – một:

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 không – một với các tham số lần lượt là p_1 và p_2. Xét thống kê f_1-f_2 là hiệu hai tần suất mẫu. Lúc đó nếu $n_1>30$ và $n_2>30$ thì:

$$U=\frac{(f_1-f_2)-(p_1-p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1}+\frac{p_2(1-p_2)}{n_2}}}\sim N(0;1).$$

CÂU HỎI ÔN TẬP VÀ BÀI TẬP

6.1 Mẫu ngẫu nhiên kích thước n về dấu hiệu nghiên cứu X là một dãy gồm n biến ngẫu nhiên $X_1, X_2, ..., X_n$ độc lập cùng phân bố với X .

Đúng ☐ Sai ☐

6.2 Tổ chức đồ dùng để biểu diễn mẫu ngẫu nhiên cho dưới dạng bảng phân bố ghép lớp.

Đúng ☐ Sai ☐

6.3 Một thống kê của mẫu ngẫu nhiên là con số cụ thể về dấu hiệu nghiên cứu.

Đúng ☐ Sai ☐

6.4 Trung bình mẫu của dấu hiệu nghiên cứu có phân bố chuẩn cũng có phân bố chuẩn.

Đúng ☐ Sai ☐

6.5 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 ☐

6.6 Phương sai mẫu của dấu hiệu nghiên cứu có phân bố chuẩn cũng có phân bố chuẩn.

Đúng ☐ Sai ☐

6.7 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 ☐

6.8 Khi bậc tự do của phân bố Student lớn hơn 30, phân bố Student tiệm cận phân bố chuẩn.

Đúng ☐ Sai ☐

6.9 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 quy luật không – một sẽ xấp xỉ phân bố chuẩn.

Đúng ☐ Sai ☐

6.10 Nếu biến ngẫu nhiên gốc có phân bố chuẩn thì phương sai mẫu có phân bố χ^2 .

Đúng ☐ Sai ☐

6.11 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.

6.12 Cho ví dụ về mẫu ngẫu nhiên kích thước n được xây dựng từ biến ngẫu nhiên có phân bố không – một, phân bố nghị thức, phân bố chuẩn.

6.13 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
n_i	10	20	30	15	10	10	5

6.14 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
n_i	5	2	3	10

6.15 Để nghiên cứu tuổi thọ (X) của một loại bóng đèn, người ta thắp thử 100 bóng và có số liệu sau:

Tuổi thọ X(giờ)	Số bóng tương ứng (n_i)
$1010-1030$	2
$1030-1050$	3
$1050-1070$	8
$1070-1090$	13
$1090-1110$	25
$1110-1130$	20
$1130-1150$	12
$1150-1170$	10
$1170-1190$	6
$1190-1210$	1
	$n=100$

Tính tuổi thọ trung bình $\overline{x}$ và độ lệch chuẩn mẫu s .

6.16 Từ tổng thể có dấu hiệu nghiên cứu X có bảng phân bố xác suất sau

X	0	1
P	0,5	0,5

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.

6.17 Giả sử biến ngẫu nhiên gốc có phân bố không - một A(p). Chọn mẫu ngẫu nhiên kích thước $n=10$. Hãy tính kỳ vọng và phương sai của trung bình mẫu.

6.18 Giả sử biến ngẫu nhiên gốc có phân bố chuẩn $N(20;1)$. Chọn mẫu ngẫu nhiên kích thước $n=100$. Hãy tính xác suất để trung bình mẫu $\overline{X}$ nằm trong khoảng:

$$19{,}8 < \overline{X} < 20{,}2\,.$$

6.19 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.

6.20 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
n_i (số người)	4	5	20	10	8	3

Nhóm 2

X (phút)	46	48	51
n_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 VII: ƯỚC LƯỢNG CÁC THAM SỐ CỦA BIẾN NGẪU NHIÊN

GIỚI THIỆU

Quy luật phân bố xác suất của các thống kê đặc trưng mẫu phản ánh mối liên hệ chặt chẽ giữa các tham số của mẫu với các tham số của dấu hiệu nghiên cứu tương ứng của tổng thể. Lý thuyết thống kê sử dụng hai phương pháp sau

- Suy diễn thống kê:

Nếu đã biết qui luật phân bố xác suất cũng như các tham số đặc trưng của tổng thể thì có thể sử dụng các kết luận trên để suy đoán về tính chất của một mẫu ngẫu nhiên rút ra từ tổng thể đó. Chẳng hạn, nếu biết dấu hiệu nghiên cứu X có phân bố chuẩn $N(\mu,\sigma^2)$ thì thống kê $\dfrac{(\overline{X}-\mu)\sqrt{n}}{\sigma^2}$ có phân bố chuẩn tắc $N(0;1)$.

- Quy nạp thống kê:

Sử dụng các phương pháp thống kê để từ các đặc trưng mẫu suy ra các đặc trưng của tổng thể.

Chính vì vậy các phương pháp thống kê giải quyết được nhiều bài toán của thực tế, có thể giúp cho các nhà nghiên cứu tìm ra quy luật của tổng thể, giúp các nhà hoạch định chính sách dự đoán sự phát triển trong tương lai, để ra các quyết định chấp nhận hoặc bác bỏ các giả thiết nào đó.

Trong chương này ta sử dụng phương pháp quy nạp thống kê để ước lượng các tham số đặc trưng của tổng thể thông qua thống kê của mẫu.

Nếu dấu hiệu nghiên cứu của tổng thể có thể xem như một biến ngẫu nhiên và giả sử bằng lý thuyết đã xác định được dạng phân bố xác suất của nó thì vấn đề xác định các tham số đặc trưng của tổng thể sẽ quy về bài toán xác định các tham số đặc trưng của quy luật phân bố xác suất của biến ngẫu nhiên gốc X. Chẳng hạn nếu đã biết dấu hiệu nghiên cứu của tổng thể có phân bố chuẩn $N(\mu,\sigma^2)$ thì bài toán đặt ra là phải ước lượng các tham số là kỳ vọng μ và phương sai σ^2, hai tham số này cũng chính là trung bình và phương sai của tổng thể.

Có hai phương pháp sử dụng thống kê để ước lượng tham số là ước lượng điểm và khoảng tin cậy:

Ước lượng điểm là dùng thống kê để ước lượng một tham số nào đó theo các tiêu chuẩn: vững, không chệch, hiệu quả. Có hai phương pháp ước lượng điểm 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 chương này học viên cần nắm vững chương 6 về lý thuyết mẫu.

NỘI DUNG

7.1. PHƯƠNG PHÁP ƯỚC LƯỢNG ĐIỂM

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 $\hat{\theta}$ 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 (6.5):

$$\hat{\theta}=T(X_1,X_2,...,X_n)$$

Khi đó với mẫu cụ thể $w=(x_1,x_2,...,x_n)$ giá trị cụ thể của thống kê $\hat{\theta}=T(x_1,x_2,...,x_n)$ là ước lượng của θ.

Cùng với một mẫu ngẫu nhiên có thể xây dựng nhiều thống kê $\hat{\theta}$ khác nhau để ước lượng cho tham số θ. Vì vậy ta cần lựa chọn thống kê tốt nhất để ước lượng cho tham số θ dựa vào các tiêu chuẩn sau:

7.1.1. Ước lượng không chệch

Thống kê $\hat{\theta}=T(X_1,X_2,...,X_n)$ 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 7.1: *Thống kê $\hat{\theta}=T(X_1,X_2,...,X_n)$ được gọi là ước lượng không chệch của θ nếu với mọi giá trị của tham số θ,*

$$\mathrm{E}(\hat{\theta})=\theta \tag{7.1}$$

Nếu $\mathrm{E}(\hat{\theta})\neq\theta$ thì $\hat{\theta}=T(X_1,X_2,...,X_n)$ được gọi là ước lượng chệch của θ.

Ví dụ 7.1: Dựa vào các công thức (6.8), (6.17), (6.21) 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.
- Phương sai mẫu S^2 và $\hat{S}^2$ là ước lượng không chệch cho phương sai σ^2 của biến ngẫu nhiên gốc.
- Tần suất mẫu f là ước lượng không chệch của tần suất p của tổng thể.

7.1.2. Ước lượng hiệu quả

Điều kiện (7.1) của ước lượng không chệch có nghĩa rằng trung bình các giá trị của $\hat{\theta}$ bằng giá trị θ. Từng giá trị của $\hat{\theta}$ có thể sai lệch rất lớn so với θ. Vì vậy ta tìm ước lượng không chệch sao cho độ sai lệch trên bé nhất.

Định nghĩa 7.2: *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.*

Như vậy, để xét xem ước lượng không chệch $\hat{\theta}$ 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 $\hat{\theta}$ 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 có dấu hiệu nghiên cứu là biến ngẫu nhiên X có hàm mật độ xác suất (hay biểu thức 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 III) và $\hat{\theta}$ là ước lượng không chệch bất kỳ của θ thì

$$\mathrm{D}(\hat{\theta}) \ge \frac{1}{n\mathrm{E}\left(\dfrac{\partial(\ln f(x,\theta))}{\partial\theta}\right)^2} \qquad \textbf{(7.2)}$$

Ví dụ 7.2: 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 $N(\mu,\sigma^2)$.

Thật vậy theo công thức (6.8) ta có $\mathrm{D}(\overline{X})=\dfrac{\sigma^2}{n}$. Mặt khác theo (3.18) hàm mật độ của X có dạng:

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

$$\Rightarrow \ln f(x,\mu)=-\ln\left(\sigma\sqrt{2\pi}\right)-\frac{(x-\mu)^2}{2\sigma^2} \Rightarrow \frac{\partial \ln f(x,\mu)}{\partial\mu}=\frac{x-\mu}{\sigma^2}$$

Vậy: $n\mathrm{E}\left(\dfrac{\partial(\ln f(X,\theta))}{\partial\theta}\right)^2=n\mathrm{E}\left(\dfrac{X-\mu}{\sigma^2}\right)^2=\dfrac{n}{\sigma^4}\mathrm{E}(X-\mu)^2=\dfrac{n}{\sigma^4}\mathrm{D}(X)=\dfrac{n}{\sigma^2}$.

Như vậy $\mathrm{D}(\overline{X})=\dfrac{\sigma^2}{n}$ 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.

7.1.3. Ước lượng vững

Định nghĩa 7.3: *Thống kê* $\hat{\theta} = T(X_1, X_2, ..., X_n)$ *được gọi là ước lượng vững của tham số* θ *của biến ngẫu nhiên gốc* X *nếu* $\hat{\theta} = T(X_1, X_2, ..., X_n)$ *hội tụ theo xác suất đến* θ *khi* $n \to \infty$, nghĩa là với mọi $\varepsilon > 0$, luôn có:

$$\lim_{n\to\infty} P\left\{\left|\hat{\theta} - \theta\right| < \varepsilon\right\} = 1. \quad \textbf{(7.3)}$$

Theo luật số lớn hệ quả 2, công thức (5.7) chương V, ta có trung bình mẫu $\overline{X}$ là ước lượng vững của kỳ vọng μ, S^2 và $\hat{S}^2$ là ước lượng vững của phương sai σ^2 của biến ngẫu nhiên gốc X của tổng thể. Tần suất mẫu f là ước lượng vững của tần suất p của 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 tần suất p của tổng thể.
- Phương sai mẫu S^2 và $\hat{S}^2$ là ước lượng không chệch và vững của phương sai σ^2 của biến ngẫu nhiên gốc của tổng thể.

7.1.4. Ước lượng hợp lý cực đại

Giả sử đã biết quy luật phân bố xác suất của dấu hiệu nghiên cứu biến ngẫu nhiên gốc X có hàm mật độ $f(x,\theta)$ (hoặc có thể là biểu thức xác suất nếu X là biến ngẫu nhiên rời rạc). Cần phải ước lượng tham số θ nào đó của X. Từ tổng thể rút ra một mẫu ngẫu nhiên kích thước n:

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

Vì các biến ngẫu nhiên $X_1, X_2, ..., X_n$ độc lập nên hàm mật độ đồng thời có dạng:

$$L(x_1, x_2, ..., x_n, \theta) = f(x_1, \theta) \cdot f(x_2, \theta) \cdots f(x_n, \theta). \quad \textbf{(7.4)}$$

Hàm $L(x_1, x_2, ..., x_n, \theta)$ được gọi là *hàm hợp lý của tham số* θ.

Phương pháp hợp lý cực đại chủ trương rằng $\hat{\theta} = g(x_1, x_2, ..., x_n)$ ước lượng tốt nhất cho tham số θ nếu: Khi ta xem $x_1, ..., x_n$ là tham số còn θ là biến thì hàm hợp lý $L(x_1, ..., x_n, \theta)$ đạt cực đại tại $\hat{\theta} = g(x_1, x_2, ..., x_n)$.

Thống kê $\hat{\theta} = g(X_1, X_2, ..., X_n)$ *được gọi là ước lượng hợp lý cực đại của* θ.

Mặt khác vì hàm logarit đồng biến nên hàm $L(x_1,x_2,...,x_n,\theta)$ và $\ln L(x_1,x_2,...,x_n,\theta)$ đạt cực đại tại cùng một giá trị của θ, do đó để tìm ước lượng hợp lý cực đại ta cần tìm θ thỏa mãn các điều kiện sau:

$$\begin{cases} \dfrac{\partial \ln L}{\partial\theta} = 0 \\ \dfrac{\partial^2 \ln L}{\partial\theta^2} < 0 \end{cases} \tag{7.5}$$

Ví dụ 7.3: Tìm ước lượng hợp lý cực đại của tần suất tổng thể p.

Bài toán tương đương với tìm ước lượng hợp lý cực đại của tham số p của biến ngẫu nhiên gốc có quy luật phân bố không – một $A(p)$. Biểu thức xác suất có dạng $P_x = p^x(1-p)^{1-x}; x = 0,1$

Hàm hợp lý $L(x_1,x_2,...,x_n,p) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i}$. Từ đó:

$$\ln L(x_1,x_2,...,x_n,p) = \sum_{i=1}^{n}\left(x_i \ln p + (1-x_i)\ln(1-p)\right)$$

$$\frac{\partial \ln L}{\partial p} = \frac{1}{p}\sum_{i=1}^{n} x_i + \frac{1}{1-p}\sum_{i=1}^{n}(x_i - 1)\,; \quad \frac{\partial \ln L}{\partial p} = 0 \Rightarrow p = \overline{x}\,.$$

Hơn nữa $\dfrac{\partial^2 \ln L}{\partial p^2} = -\dfrac{n}{p(1-p)} < 0$, do đó ước lượng hợp lý cực đại của p là tần suất mẫu $\overline{X} = f$ (công thức (6.20), (6.21)).

Ví dụ 7.4: Tìm ước lượng hợp lý cực đại của hai tham số μ và σ^2 của biến ngẫu nhiên X có phân bố chuẩn $N(\mu;\sigma^2)$.

Hàm hợp lý có dạng $L(x_1,x_2,...,x_n,\mu,\sigma) = \dfrac{1}{(\sigma\sqrt{2\pi})^n} e^{\frac{-1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}$.

Các đạo hàm riêng của $\ln L$ theo μ và σ^2 như sau:

$$\frac{\partial \ln L}{\partial \mu} = -\frac{n\overline{x} - n\mu}{\sigma^2}\;;\quad \frac{\partial \ln L}{\partial \sigma^2} = -\frac{n}{\sigma} + \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{\sigma^3}\,.$$

Giải hệ phương trình (7.5) thu được nghiệm $\mu = \overline{x}$ và $\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2$. Tại điểm này hàm hợp lý $L(x_1, x_2, ..., x_n, \mu, \sigma)$ đạt cực đại. Vậy ước lượng hợp lý cực đại của kỳ vọng μ là trung bình mẫu $\overline{X}$, còn của σ^2 là MS.

7.2. PHƯƠNG PHÁP ƯỚC LƯỢNG BẰNG KHOẢNG TIN CẬY

7.2.1. Khái niệm

Các 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. Nghĩa là từ mẫu ngẫu nhiên tìm 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 là 0,95 hay 0,99).

Định nghĩa 7.4: *Khoảng $[a;b]$ có hai đầu mút là hai thống kê:*

$$a = a(X_1, X_2, ..., X_n),\ b = b(X_1, X_2, ..., X_n) \qquad \textbf{(7.6)}$$

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:

$$P\{a \le \theta \le b\} = \beta \qquad \textbf{(7.7)}$$

Trong thực tế thường yêu cầu độ tin cậy β khá lớn, khi đó theo nguyên lý xác suất lớn biến cố $\{a \le \theta \le b\}$ 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 một mẫu cụ thể $w = (x_1, x_2, ..., x_n)$, tính được giá trị cụ thể $a = a(x_1, x_2, ..., x_n)$, $b = b(x_1, x_2, ..., x_n)$.

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 $a \le \theta \le b$.

7.2.2. Khoảng tin cậy của kỳ vọng của biến ngẫu nhiên phân bố theo quy luật chuẩn

Giả sử tổng thể biến ngẫu nhiên gốc X có phân bố chuẩn $N(\mu;\sigma^2)$ 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 trong các trường hợp sau.

7.2.2.1. Trường hợp phương sai σ^2 đã biết

Định lý 7.1: Khoảng tin cậy của tham số μ với độ tin cậy β có dạng:

$$\left[\overline{X}-U_{\alpha/2}\frac{\sigma}{\sqrt{n}};\ \overline{X}+U_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right] \tag{7.8}$$

trong đó: $\alpha=1-\beta$; $U_{\frac{\alpha}{2}}$ là giá trị tới hạn mức $\frac{\alpha}{2}$ của phân bố chuẩn tắc $N(0;1)$ (công thức 3.24).

Chứng minh: Theo công thức (6.29) ta có $\frac{(\overline{X}-\mu)\sqrt{n}}{\sigma}\sim N(0;1)$.

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 (3.25) ta có

$$P\left[\overline{X}-U_{\alpha/2}\frac{\sigma}{\sqrt{n}}\le\mu\le\overline{X}+U_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]=P\left\{\left|\frac{(\overline{X}-\mu)\sqrt{n}}{\sigma}\right|\le U_{\alpha/2}\right\}=1-\alpha=\beta.$$

Định nghĩa 7.5: $\varepsilon=U_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ được gọi là *độ chính xác của ước lượng*.

Với độ chính xác ε_0 và độ tin cậy β cho trước, thì *kích thước mẫu tối thiểu* là số tự nhiên n nhỏ nhất thỏa mãn:

$$n\ge\frac{\sigma^2U_{\alpha/2}^2}{\varepsilon_0^2} \tag{7.9}$$

Ví dụ 7.5: 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. Cân thử 25 sản phẩm loại này ta thu được kết quả:

Trọng lượng (gram)	18	19	20	21
Số SP 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 $N(\mu;\sigma^2)$ với $\sigma=1$. Trọng lượng trung bình của sản phẩm là tham số μ. Khoảng tin cậy có dạng (7.8).

Với độ tin cậy $\beta = 0,95 \Rightarrow \frac{\alpha}{2} = 0,025 \Rightarrow U_{\frac{\alpha}{2}} = 1,96$.

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 $\varepsilon = U_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 1,96\cdot\frac{1}{\sqrt{25}} = 0,392$.

Vậy với độ tin cậy 95% qua mẫu cụ thể này, khoảng tin cậy của tham số μ là:

$$[19,64 - 0,392; 19,64 + 0,392]$$

hay $$19,248 \le \mu \le 20,032.$$

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 $n = 43$.

7.2.2.2. 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 nếu 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 (6.18). Mặt khác, theo định lý giới hạn trung tâm thì thống kê $\frac{(\overline{X}-\mu)\sqrt{n}}{\sigma}$ xấp xỉ chuẩn, đúng với mọi biến ngẫu nhiên gốc X (không đòi hỏi phân bố chuẩn).

Do đó khoảng tin cậy của tham số μ với độ tin cậy β có thể lấy là:

$$\left[\overline{X} - U_{\alpha/2}\frac{S}{\sqrt{n}}; \overline{X} + U_{\alpha/2}\frac{S}{\sqrt{n}}\right] \quad \textbf{(7.10)}$$

Ví dụ 7.6: 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			$-6,5$	$15,25$

$$\overline{u}=\frac{-6,5}{35}=-0,1857 \;\Rightarrow\; \overline{x}=8,25-0,1857\approx 8,06\,.$$

$$s_X^2=s_U^2=\frac{1}{34}\left(15,25-\frac{(-6,5)^2}{35}\right)=0,413 \;\Rightarrow\; s=0,64\,.$$

Với độ tin cậy $\beta=95\%$, $U_{\frac{\alpha}{2}}=1,96$.

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 khoảng tin cậy cho chiều cao trung bình μ của các cây bạch đàn là:

$7,87\le\mu\le 8,29$.

7.2.2.3. Trường hợp phương sai σ^2 chưa biết, kích thước mẫu $n<30$

Trong trường hợp này, theo công thức (6.32) thống kê

$$T=\frac{(\overline{X}-\mu)\sqrt{n}}{S} \tag{7.11}$$

có phân bố Student $n-1$ bậc tự do. Vì vậy khoảng tin cậy được tính theo kết quả sau:

Định lý 7.2: Khoảng tin cậy của tham số μ với độ tin cậy β có dạng:

$$\left[\overline{X}-t_{\alpha/2}(n-1)\frac{S}{\sqrt{n}}\,;\;\overline{X}+t_{\alpha/2}(n-1)\frac{S}{\sqrt{n}}\right] \tag{7.12}$$

trong đó $t_{\alpha/2}(n-1)$ là giá trị tới hạn mức $\dfrac{\alpha}{2}$ của phân bố Student $n-1$ bậc tự do (công thức 3.41).

Độ chính xác của ước lượng:

$$\varepsilon=t_{\alpha/2}(n-1)\frac{S}{\sqrt{n}} \tag{7.13}$$

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:

$$n \ge \left(\frac{S \cdot t_{\alpha/2}(n-1)}{\varepsilon_0}\right)^2 \tag{7.14}$$

Ví dụ 7.7: Năng suất của một loại giống mới là một biến ngẫu nhiên có quy luật phân bố chuẩn $N(\mu;\sigma^2)$. 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 hạt 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: $\overline{x} = 171$; $s = 3,4254$. $\alpha = 0,05$; $\frac{\alpha}{2} = 0,025$.

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 $\varepsilon = t_{\alpha/2}(n-1)\frac{S}{\sqrt{n}} = 2,131 \cdot \frac{3,4254}{\sqrt{16}} = 1,885$.

Vậy khoảng tin cậy cho năng suất trung bình của loại giống này thỏa mãn:

$$169,115 \le \mu \le 172,885.$$

7.2.3. Khoảng tin cậy cho xác suất p của biến ngẫu nhiên có phân bố theo quy luật không – một $A(p)$

Ta cần nghiên cứu một dấu hiệu nghiên cứu A nào đó mà mỗi cá thể của tổng thể có thể có hoặc không có dấu hiệu A. 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 quy luật phân bố xác suất không – một $A(p)$ có kỳ vọng và phương sai $\mathrm{E}X = p$; $\mathrm{D}X = p(1-p)$.

Lấy mẫu ngẫu nhiên kích thước n: $W = (X_1, X_2, ..., X_n)$, trong đó $X_1, X_2, ..., X_n$ là các biến ngẫu nhiên độc lập có cùng phân bố không – một $A(p)$.

Theo định lý Moivre-Laplace (6.40) thì khi n đủ lớn ta có thể xấp xỉ thống kê $U = \frac{(f-p)\sqrt{n}}{\sqrt{p(1-p)}}$ với phân bố chuẩn tắc $N(0;1)$.

Tuy nhiên vì p chưa biết nên chưa biết $p(1-p) = \mathrm{D}X$.

Mặt khác theo công thức (6.19), (6.20) và mục 2. ta có tần suất mẫu f là ước lượng vững, không chệch và hiệu quả của tần suất p tổng thể. Vì vậy khi n đủ lớn ta có thể thay p bằng f.

Do đó ta có khoảng tin cậy cho tần suất p của tổng thể với độ tin cậy β là:

$$\left[f - U_{\alpha/2}\sqrt{\frac{f(1-f)}{n}}\ ;\ f + U_{\alpha/2}\sqrt{\frac{f(1-f)}{n}} \right] \qquad \textbf{(7.15)}$$

Với điều kiện:

$$\begin{cases} nf > 10 \\ n(1-f) > 10 \end{cases} \qquad \textbf{(7.16)}$$

trong đó $\alpha = 1-\beta$; $U_{\frac{\alpha}{2}}$ là giá trị tới hạn mức $\frac{\alpha}{2}$ của phân bố chuẩn tắc $N(0;1)$.

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:

$$n \geq f(1-f)\left(\frac{U_{\alpha/2}}{\varepsilon_0}\right)^2 \qquad \textbf{(7.17)}$$

trong đó f là tần suất mẫu của một mẫu ngẫu nhiên nào đó.

Ví dụ 7.8: 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. Với độ tin cậy $\beta = 95\%$ tối thiểu ứng cử viên A sẽ chiếm được bao nhiêu % số phiếu bầu.

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, là biến ngẫu nhiên có quy luật phân bố không – một $A(p)$. Khoảng tin cậy cho p có dạng (7.15) với điều kiện (7.16).

Từ mẫu cụ thể trên ta có $f = \frac{960}{1600} = 0,6$.

Điều kiện $\begin{cases} nf = 960 > 10 \\ n(1-f) = 640 > 10 \end{cases}$ được thỏa mãn.

Độ chính xác của ước lượng $\varepsilon = U_{\alpha/2}\sqrt{\frac{f(1-f)}{n}} = 1,96\sqrt{\frac{0,6\cdot 0,4}{1600}} = 0,024$.

Khoảng tin cậy: $0,576 \leq p \leq 0,624$.

Vậy với độ tin cậy 95% thì tối thiểu có 57,6% cử tri sẽ bỏ phiếu cho ứng cử viên A.

7.2.4. Ước lượng phương sai của biến ngẫu nhiên có phân bố theo quy luật chuẩn $N(\mu;\sigma^2)$

Giả sử trong tổng thể biến ngẫu nhiên gốc X có phân bố chuẩn $N(\mu;\sigma^2)$ nhưng chưa biết phương sai σ^2 của nó. Từ tổng thể rút ra một mẫu ngẫu nhiên kích thước n: $W = (X_1, X_2, ..., X_n)$.

Ta sẽ chọn thống kê thích hợp để ước lượng cho tham số σ^2 trong các trường hợp sau:

7.2.4.1. Đã biết kỳ vọng μ

Chọn thống kê
$$T = \frac{n\hat{S}^2}{\sigma^2} \qquad \textbf{(7.18)}$$

Theo công thức (6.30) thống kê T có phân bố khi bình phương n bậc tự do: $\chi^2(n)$. Do đó với độ tin cậy β cho trước, với cặp số α_1, α_2 sao cho $\alpha_1 + \alpha_2 = 1-\beta$ có thể tìm hai giá trị tới hạn của T mức α_1, α_2 là $\chi^2_{1-\alpha_1}(n)$, $\chi^2_{\alpha_2}(n)$:

$$P\left\{T > \chi^2_{1-\alpha_1}(n)\right\} = 1-\alpha_1 \text{ và } P\left\{T > \chi^2_{\alpha_2}(n)\right\} = \alpha_2 .$$

Do đó
$$P\left\{\chi^2_{1-\alpha_1}(n) < T < \chi^2_{\alpha_2}(n)\right\} = 1-(\alpha_1+\alpha_2) = \beta . \qquad \textbf{(7.19)}$$

Thay thống kê T từ công thức (7.18) vào (7.19) và giải theo σ^2, ta được:

$$P\left\{\frac{n\hat{S}^2}{\chi^2_{\alpha_2}(n)} < \sigma^2 < \frac{n\hat{S}^2}{\chi^2_{1-\alpha_1}(n)}\right\} = \beta . \qquad \textbf{(7.20)}$$

Như vậy, với độ tin cậy β khoảng tin cậy của σ^2 có dạng:

$$\left(\frac{n\hat{S}^2}{\chi^2_{\alpha_2}(n)} ; \frac{n\hat{S}^2}{\chi^2_{1-\alpha_1}(n)}\right) \qquad \textbf{(7.21)}$$

- Nếu $\alpha_1 = \alpha_2 = \dfrac{\alpha}{2}$ khoảng tin cậy có dạng:

$$\left(\frac{n\hat{S}^2}{\chi^2_{\alpha/2}(n)} ; \frac{n\hat{S}^2}{\chi^2_{1-\alpha/2}(n)}\right) \qquad \textbf{(7.22)}$$

- Nếu $\alpha_1 = 0$; $\alpha_2 = \alpha$ khoảng tin cậy bên phải của σ^2 có dạng:

$$\left(\frac{n\hat{S}^2}{\chi^2_{\alpha}(n)} \, ; +\infty \right) \tag{7.23}$$

- Nếu $\alpha_2 = 0; \alpha_1 = \alpha$ khoảng tin cậy bên trái của σ^2 có dạng:

$$\left(0 \, ; \frac{n\hat{S}^2}{\chi^2_{1-\alpha}(n)} \right) \tag{7.24}$$

Ví dụ 7.9: 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 $N(\mu;\sigma^2)$ khi đã biết μ. Khoảng tin cậy theo công thức (2.22).

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 $\hat{s}^2$ ta lập bảng sau:

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

$$\hat{s}^2 = \frac{1,75}{25} = 0,07.$$

Vậy với độ tin cậy 90%, qua mẫu cụ thể này, khoảng tin cậy của σ^2 là:

$$\left(\frac{25\cdot 0,07}{37,65};\ \frac{25\cdot 0,07}{14,61}\right) \text{ hay } 0,0464 < \sigma^2 < 0,1198.$$

7.2.4.2. Chưa biết kỳ vọng μ

Chọn thống kê:

$$T = \frac{(n-1)S^2}{\sigma^2} \tag{7.25}$$

Theo công thức (6.31) thống kê T có phân bố khi bình phương $n-1$ bậc tự do: $\chi^2(n-1)$. Do đó với độ tin cậy β cho trước, với cặp số α_1, α_2 sao cho $\alpha_1+\alpha_2=1-\beta$ có thể tìm hai giá trị tới hạn của T mức α_1, α_2 là $\chi^2_{1-\alpha_1}(n-1)$, $\chi^2_{\alpha_2}(n-1)$:

$$P\left\{T > \chi^2_{1-\alpha_1}(n-1)\right\} = 1-\alpha_1 \text{ và } P\left\{T > \chi^2_{\alpha_2}(n-1)\right\} = \alpha_2.$$

Do đó:

$$P\left\{\chi^2_{1-\alpha_1}(n-1) < T < \chi^2_{\alpha_2}(n-1)\right\} = 1-(\alpha_1+\alpha_2) = \beta. \tag{7.26}$$

Thay thống kê T từ công thức (7.25) vào (7.26) và giải theo σ^2, ta được:

$$P\left\{\frac{(n-1)S^2}{\chi^2_{\alpha_2}(n-1)} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{1-\alpha_1}(n-1)}\right\} = \beta. \tag{7.27}$$

Như vậy, với độ tin cậy β khoảng tin cậy của σ^2 có dạng:

$$\left(\frac{(n-1)S^2}{\chi^2_{\alpha_2}(n-1)};\frac{(n-1)S^2}{\chi^2_{1-\alpha_1}(n-1)}\right) \tag{7.28}$$

- Nếu $\alpha_1=\alpha_2=\dfrac{\alpha}{2}$ khoảng tin cậy có dạng:

$$\left(\frac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)};\frac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)}\right) \tag{7.29}$$

- Nếu $\alpha_1=0$; $\alpha_2=\alpha$ khoảng tin cậy bên phải của σ^2 có dạng:

$$\left(\frac{(n-1)S^2}{\chi^2_{\alpha}(n-1)};+\infty\right) \tag{7.30}$$

• Nếu $\alpha_2 = 0$; $\alpha_1 = \alpha$ khoảng tin cậy bên trái của σ^2 có dạng:

$$\left(0\,;\frac{(n-1)S^2}{\chi^2_{1-\alpha}(n-1)}\right) \tag{7.31}$$

TÓM TẮT

Ước lượng không chệch

Thống kê $\hat{\theta} = T(X_1, X_2, ..., X_n)$ được gọi là ước lượng không chệch của θ nếu: $\mathrm{E}(\hat{\theta}) = \theta$.

Ước lượng hiệu quả

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.

Ước lượng vững

Thống kê $\hat{\theta} = T(X_1, X_2, ..., X_n)$ được gọi là ước lượng vững của tham số θ của biến ngẫu nhiên gốc X nếu $\hat{\theta} = T(X_1, X_2, ..., X_n)$ hội tụ theo xác suất đến θ khi $n \to \infty$.

Ước lượng hợp lý cực đại

Giả sử đã biết quy luật phân bố xác suất của dấu hiệu nghiên cứu biến ngẫu nhiên gốc X có hàm mật độ $f(x,\theta)$ (hoặc có thể là biểu thức xác suất nếu X là biến ngẫu nhiên rời rạc). Cần phải ước lượng tham số θ nào đó của X. Từ tổng thể rút ra một mẫu ngẫu nhiên kích thước n, hàm mật độ đồng thời có dạng của mẫu ngẫu nhiên có dạng

$$L(x_1, x_2, ..., x_n, \theta) = f(x_1, \theta) \cdot f(x_2, \theta) \cdots f(x_n, \theta).$$

Hàm $L(x_1, x_2, ..., x_n, \theta)$ được gọi là hàm hợp lý của tham số θ.

Khi ta xem $x_1, ..., x_n$ là tham số còn θ là biến và giả sử hàm hợp lý $L(x_1, ..., x_n, \theta)$ đạt cực đại tại $\hat{\theta} = g(x_1, x_2, ..., x_n)$.

Thống kê $\hat{\theta} = g(X_1, X_2, ..., X_n)$ được gọi là ước lượng hợp lý cực đại của θ.

Khoảng tin cậy

Khoảng $[a;b]$ có hai đầu mút là hai thống kê $a = a(X_1, X_2, ..., X_n)$, $b = b(X_1, X_2, ..., X_n)$ 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: $P\{a \le \theta \le b\} = \beta$.

Khoảng tin cậy của kỳ vọng của biến ngẫu nhiên phân bố theo quy luật chuẩn

- Trường hợp phương sai σ^2 đã biết:

$\left[\overline{X} - U_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}; \overline{X} + U_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}\right]$, $U_{\alpha/2}$ là giá trị tới hạn mức $\alpha/2$ của phân bố chuẩn tắc $N(0;1)$; $\alpha = 1-\beta$.

- Trường hợp phương sai σ^2 chưa biết, $n \ge 30$:

$$\left[\overline{X} - U_{\alpha/2}\frac{S}{\sqrt{n}}; \overline{X} + U_{\alpha/2}\frac{S}{\sqrt{n}}\right].$$

- Trường hợp phương sai σ^2 chưa biết, $n < 30$:

$$\left[\overline{X} - t_{\alpha/2}(n-1)\frac{S}{\sqrt{n}}; \overline{X} + t_{\alpha/2}(n-1)\frac{S}{\sqrt{n}}\right],$$

trong đó $t_{\alpha/2}(n-1)$ là giá trị tới hạn mức $\dfrac{\alpha}{2}$ của phân bố Student $n-1$ bậc tự do.

Khoảng tin cậy cho xác suất p của biến ngẫu nhiên có phân bố theo quy luật không – một $A(p)$:

$\left[f - U_{\alpha/2}\sqrt{\dfrac{f(1-f)}{n}}\ ; f + U_{\alpha/2}\sqrt{\dfrac{f(1-f)}{n}}\right]$ Với điều kiện $\begin{cases} nf > 10 \\ n(1-f) > 10 \end{cases}$

trong đó $\alpha = 1-\beta$; $U_{\frac{\alpha}{2}}$ là giá trị tới hạn mức $\dfrac{\alpha}{2}$ của phân bố chuẩn tắc $N(0;1)$.

Ước lượng phương sai của biến ngẫu nhiên có phân bố theo quy luật chuẩn $N(\mu;\sigma^2)$

- Trường hợp đã biết kỳ vọng μ : .

Khoảng tin cậy của σ^2 với độ tin cậy β có dạng: $\left(\dfrac{n\hat{S}^2}{\chi^2_{\alpha_2}(n)}; \dfrac{n\hat{S}^2}{\chi^2_{1-\alpha_1}(n)}\right)$, trong đó cặp số α_1, α_2 sao cho $\alpha_1 + \alpha_2 = 1-\beta$. $\chi^2_{1-\alpha_1}(n)$, $\chi^2_{\alpha_2}(n)$ lần lượt là giá trị tới hạn mức $1-\alpha_1$ và α_2 của phân bố "Khi bình phương" n bậc tự do.

- Trường hợp chưa biết kỳ vọng μ :

Khoảng tin cậy của σ^2 với độ tin cậy β có dạng: $\left(\dfrac{(n-1)S^2}{\chi^2_{\alpha_2}(n-1)};\dfrac{(n-1)S^2}{\chi^2_{1-\alpha_1}(n-1)}\right)$, trong đó cặp số α_1, α_2 sao cho $\alpha_1+\alpha_2=1-\beta$. $\chi^2_{\alpha_2}(n-1)$, $\chi^2_{1-\alpha_1}(n-1)$ lần lượt là giá trị tới hạn mức α_2 và $1-\alpha_1$ của phân bố "Khi bình phương" $n-1$ bậc tự do.

CÂU HỎI ÔN TẬP VÀ BÀI TẬP

7.1 Trung bình mẫu là ước lượng vững và hiệu quả của kỳ vọng của biến ngẫu nhiên gốc.

Đúng ☐ Sai ☐

7.2 Có thể tìm được ước lượng không chệch của θ có phương sai nhỏ hơn đại lượng $\dfrac{1}{n\mathrm{E}\left(\dfrac{\partial(\ln f(x,\theta))}{\partial\theta}\right)^2}$.

Đúng ☐ Sai ☐

7.3 Tổng của hai ước lượng không chệch là một ước lượng không chệch.

Đúng ☐ Sai ☐

7.4 Phương sai mẫu hiệu chỉnh S^2 là ước lượng vững không chệch của phương sai của biến ngẫu nhiên gốc.

Đúng ☐ Sai ☐

7.5 Việc tìm giá trị cực đại của hàm hợp lý $L(x_1,x_2,...,x_n,\theta)$ tương đương với việc tìm giá trị cực đại của hàm $\ln L(x_1,x_2,...,x_n,\theta)$.

Đúng ☐ Sai ☐

7.6 Có thể tìm khoảng tin cậy chứa một tham số với xác suất bằng 1.

Đúng ☐ Sai ☐

7.7 Hai đầu mút của khoảng tin cậy là hai thống kê của mẫu.

Đúng ☐ Sai ☐

7.8 Muốn tìm khoảng tin cậy cho tham số μ của biến ngẫu nhiên gốc có phân bố chuẩn $N(\mu;\sigma^2)$ thì kích thước mẫu n phải lớn hơn 30.

Đúng ☐ Sai ☐

7.9 Để tìm khoảng tin cậy cho tham số p của biến ngẫu nhiên gốc có phân bố không – một A(p) ta xấp xỉ tần suất mẫu với quy luật chuẩn nếu n đủ lớn: điều kiện $\begin{cases} nf > 10 \\ n(1-f) > 10 \end{cases}$.

Đúng ☐ Sai ☐

7.10 Có thể tìm kích thước mẫu cần thiết để khoảng tin cậy cho tham số p của biến ngẫu nhiên gốc có phân bố không – một A(p) thỏa mãn độ tin cậy và độ chính xác cho trước.

Đúng ☐ Sai ☐

7.11 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.

7.12 Bằng phương pháp hợp lý cực đại hãy ước lượng tham số của quy luật phân bố mũ.

7.13 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.

7.14 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 thiếu ứng cử viên A sẽ chiếm được bao nhiêu % số phiếu bầu? Cho biết phân vị mức 0,975 của phân bố chuẩn tắc N(0;1) là 1,96.

7.15 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.

7.16 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 0,95.

7.17 Hao phí nguyên liệu cho một đơn vị sản phẩm là một biến ngẫu nhiên tuân theo luật phân bố chuẩn với độ lệch chuẩn $\sigma = 0,03$. Sản xuất thử 36 sản phẩm và thu được số liệu sau:

$X(gram)$	19,5-19,7	19,7-19,9	19,9-20,1	20,1-20,3
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.

7.18 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:

Khoảng 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.

7.19 Để ước lượng năng suất trung bình của một giống lúa mới, người ta gặt ngẫu nhiên 100 thửa ruộng trồng thí nghiệm và thu được số liệu sau:

X (tạ/ha)	40 - 42	42 - 44	44 - 46	46 - 48	48 - 50	50 - 52
Số thửa ruộng	7	13	25	35	15	5

Giả sử biến ngẫu nhiên chỉ năng suất X tuân theo quy luật chuẩn.

a) Tìm khoảng tin cậy 95% cho năng suất trung bình của giống lúa mới.

b) Nếu muốn khoảng ước lượng có độ chính xác $\varepsilon = 0,4$ thì cần lấy mẫu gồm bao nhiêu thửa ruộng.

7.20 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 bao tương ứng	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.

7.21 Để xác định chiều cao trung bình của trẻ em 8 tuổi ở thành phố, người ta tiến hành ngẫu nhiên đo chiều cao của 100 em học sinh lớp 3 (8 tuổi) ở một 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 em tương ứng	5	8	14	17	20	16	10	6	4

a) Tìm khoảng tin cậy 95% cho chiều cao trung bình của trẻ em 8 tuổi ở thành phố.

b) Nếu muốn khoảng ước lượng có độ chính xác $\varepsilon = 0,5\ cm$ thì cần phải lấy mẫu kích thước bao nhiêu.

7.22 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.

7.23 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

Gía X (nghìn đồng)	83	85	87	89	91	93	95	97	99	101
Số cửa hàng n_i	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.

7.24 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 tuân theo quy luật 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 VIII: KIỂM ĐỊNH GIẢ THIẾT THỐNG KÊ

GIỚI THIỆU

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, 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. Cạnh tranh với giả thiết này là đối thiết H_1, 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 A có xác suất bé (bằng mức ý nghĩa của phép kiểm định). Theo nguyên lý xác suất nhỏ thì trong một vài lần thử biến cố A không xảy ra. Vì vậy nếu với một mẫu cụ thể nào đó mà A xảy ra thì giả thiết rằng H_0 đúng là vô lý do đó ta bác bỏ H_0, còn nếu A không xảy ra thì ta chưa có cơ sở để bác bỏ H_0. Biến cố A được gọi là miền bác bỏ.

Hai dạng bài toán kiểm định xét trong chương này là: kiểm định tham số và kiểm định phi tham số.

- Kiểm định tham số là kiểm định tham số đặc trưng của biến ngẫu nhiên gốc trong tổng thể: kỳ vọng, phương sai, tần suất của tổng thể. Kiểm định về sự bằng nhau của hai kỳ vọng, hai phương sai hoặc hai tần suất của tổng thể.
- Kiểm định phi tham số bao gồm các bài toán về quy luật phân bố xác suất của biến ngẫu nhiên gốc của tổng thể, về tính độc lập của hai dấu hiệu nghiên cứu định tính, về tính thuần nhất của tổng thể nghiên cứu.

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 VI về lý thuyết mẫu.

NỘI DUNG

8.1. KHÁI NIỆM CHUNG

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 các quy tắc hay thủ tục quyết định dẫn đến bác bỏ hay chấp nhận giả thiết của tổng thể. Trong mục này ta tìm hiểu các khái niệm: giả thiết thống kê, các nguyên tắc để xây dựng quy tắc kiểm định, miền bác bỏ, sai lầm khi kiểm định …

8.1.1. 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 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?…

Nếu phân bố của biến ngẫu nhiên gốc được đặc trưng bởi các tham số (như giá trị trung bình, phương sai, tham số p của quy luật không – một …), 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 hay vài phép thử thì biến cố đó coi như không xảy ra".
* Phương pháp phản chứng: "Để bác bỏ A ta giả sử A đúng thì dẫn đến một điều vô lý".

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 A là rất bé và ta có thể xem A không thể xảy ra trong một phép thử về biến cố này. Lúc đó nếu trên một mẫu cụ thể quan sát được mà biến cố A xuất hiện thì điều này trái với nguyên lý xác suất nhỏ. Vậy H_0 sai và bác bỏ nó. Còn nếu A 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:

8.1.2. Tiêu chuẩn kiểm định giả thiết thống kê

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

$$T=T(X_1,X_2,...,X_n,\theta_0) \tag{8.1}$$

trong đó θ_0 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. Thống kê T được gọi là tiêu chuẩn kiểm định.

8.1.3. 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:

$$P\{T\in W_\alpha | H_0\}=\alpha \tag{8.2}$$

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.

8.1.4. Giá trị quan sát của tiêu chuẩn kiểm định

Thực hiện phép thử với mẫu ngẫu nhiên $W=(X_1,X_2,...,X_n)$ thu được mẫu cụ thể $w=(x_1,x_2,...,x_n)$, thay giá trị này vào thống kê (8.1) ta được giá trị quan sát của tiêu chuẩn kiểm định:

$$T_{qs}=(x_1,x_2,...,x_n,\theta_0) \tag{8.3}$$

8.1.5. 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 $T_{qs}\in W_\alpha$, theo nguyên tắc kiểm định thì H_0 sai, do đó ta bác bỏ H_0 thừa nhận H_1.

2. Nếu $T_{qs}\notin W_\alpha$ thì điều này chưa khẳng định rằng H_0 đúng mà chỉ có nghĩa là qua mẫu cụ thể này chưa khẳng định được là H_0 sai. Do đó 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).

8.1.6. 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 $\{T \in W_\alpha\}$, do đó khi H_0 đúng thì xác suất này bằng $P\{T \in W_\alpha | H_0\} = \alpha$. Sai lầm loại I sinh ra do kích thước mẫu quá nhỏ, do phương pháp lấy mẫu v.v…

2. Sai lầm loại II: Đó là sai lầm mắc phải khi thừa nhận giải 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:

$$P\{T \notin W_\alpha | H_1\} = \beta \tag{8.4}$$

Xác suất của biến cố đối của sai lầm loại II: $P\{T \in W_\alpha | H_1\} = 1 - \beta$ gọi là *lực lượng kiểm định*.

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 $= \beta$

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:

$$P\{T \in W_\alpha | H_0\} = \alpha \text{ và } P\{T \in W_\alpha | H_1\} = 1 - \beta \to \max$$

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.

8.1.7. 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.

8.2. KIỂM ĐỊNH THAM SỐ

8.2.1. Kiểm định giả thiết về kỳ vọng của biến ngẫu nhiên có phân bố theo quy luật chuẩn

Giả sử biến ngẫu nhiên gốc X trong tổng thể có phân bố chuẩn $N(\mu;\sigma^2)$, cần kiểm định kỳ vọng μ. Nếu có cơ sở để giả thiết rằng kỳ vọng μ bằng giá trị μ_0 ta đưa ra giả thiết thống kê

$$H_0:\ \mu = \mu_0 .$$

Ta xét các trường hợp sau:

8.2.1.1. 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 $N(\mu;\sigma^2)$ đã biết. Từ tổng thể rút ra một mẫu ngẫu nhiên kích thước n:

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

Xét thống kê

$$T = \frac{(\overline{X} - \mu_0)\sqrt{n}}{\sigma} \quad \textbf{(8.5)}$$

Nếu giả thiết H_0 đúng (theo công thức (6.29) chương VI) thì thống kê T có phân bố chuẩn tắc $N(0;1)$. Ta xây dựng các miền bác bỏ dựa vào đối thiết H_1.

a) Bài toán 1: H_0: $\mu = \mu_0$; H_1: $\mu \neq \mu_0$. Ta nói đây là bài toán kiểm định hai phía.

Miền bác bỏ
$$W_\alpha = \left\{ T = \frac{(\overline{X} - \mu_0)\sqrt{n}}{\sigma}\ ;\ |T| > U_{\alpha/2} \right\}. \quad \textbf{(8.6)}$$

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
$$W_\alpha = \left\{ T = \frac{(\overline{X} - \mu_0)\sqrt{n}}{\sigma}\ ;\ T > U_\alpha \right\}. \quad \textbf{(8.7)}$$

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 $$W_\alpha = \left\{ T = \frac{(\overline{X} - \mu_0)\sqrt{n}}{\sigma} \; ; \; -T > U_\alpha \right\} \tag{8.8}$$

trong đó $U_{\alpha/2}$, U_α là giá trị tới hạn mức $\frac{\alpha}{2}$ và mức α của phân bố chuẩn tắc $N(0;1)$.

Lập mẫu cụ thể $w = (x_1, x_2, ..., x_n)$ và 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.

8.2.1.2. Trường hợp chưa biết phương sai, kích thước mẫu $n \geq 30$ (có thể không cần phân bố chuẩn)

Trường hợp phương sai σ^2 chưa biết: Với kích thước n đủ lớn ($n \geq 30$) và giả thiết H$_0$ đúng (tương tự (2.2.2) chương 7) thì thống kê:

$$T = \frac{(\overline{X} - \mu_0)\sqrt{n}}{S} \tag{8.9}$$

xấp xỉ phân bố chuẩn tắc $N(0;1)$. Ta xây dựng các miền bác bỏ dựa vào đối thiết H$_1$.

a) Bài toán 1: H$_0$: $\mu = \mu_0$; H$_1$: $\mu \neq \mu_0$. Ta nói đây là bài toán kiểm định hai phía.

Miền bác bỏ $$W_\alpha = \left\{ T = \frac{(\overline{X} - \mu_0)\sqrt{n}}{S} \; ; \; |T| > U_{\alpha/2} \right\}. \tag{8.10}$$

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 $$W_\alpha = \left\{ T = \frac{(\overline{X} - \mu_0)\sqrt{n}}{S} \; ; \; T > U_\alpha \right\}. \tag{8.11}$$

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 $$W_\alpha = \left\{ T = \frac{(\overline{X} - \mu_0)\sqrt{n}}{S} ; -T > U_\alpha \right\}. \tag{8.12}$$

Ví dụ 8.1: 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: $\mu \neq 7,4$.

Tiêu chuẩn kiểm định: $T = \dfrac{(\overline{X} - 7,4)\sqrt{n}}{S}$

Miền bác bỏ: $\alpha = 0,01$

$$\Rightarrow U_{\frac{\alpha}{2}} = 2,575 \Rightarrow W_\alpha = \left\{ T = \frac{(\overline{X} - 7,4)\sqrt{n}}{S} \; ; \; |T| > 2,575 \right\}.$$

$T_{qs} = \dfrac{(6,1 - 7,4)\sqrt{40}}{2,5} = -3,289$. 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 8.2: Một công tyi 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: $T = \dfrac{(\overline{X} - 1200)\sqrt{n}}{S}$

Miền bác bỏ: $\alpha = 0,05 \Rightarrow U_\alpha = 1,64 \Rightarrow W_\alpha = \left\{ T = \dfrac{(\overline{X} - 1200)\sqrt{n}}{S} \; ; \; T > 1,64 \right\}$.

Thay giá trị cụ thể của mẫu vào công thức (8.9) 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.

8.2.1.3. Trường hợp chưa biết phương sai, kích thước mẫu $n < 30$

Giả sử giả thiết H_0 đúng, xét thống kê

$$T = \frac{(\overline{X} - \mu_0)\sqrt{n}}{S} \qquad \textbf{(8.13)}$$

theo công thức (6.32) thống kê T có phân bố Student $n-1$ bậc tự do. Ta xây dựng các miền bác bỏ dựa vào đối thiết H_1.

a) Bài toán 1: H_0: $\mu = \mu_0$; H_1: $\mu \neq \mu_0$.

Miền bác bỏ: $$W_\alpha = \left\{ T = \frac{(\overline{X}-\mu_0)\sqrt{n}}{S} \; ; \; |T| > t_{\alpha/2}(n-1) \right\}. \qquad \textbf{(8.14)}$$

trong đó $t_{\alpha/2}(n-1)$ là giá trị tới hạn mức $\frac{\alpha}{2}$ của phân bố Student $n-1$ bậc tự do.

b) Bài toán 2: H_0: $\mu = \mu_0$; H_1: $\mu > \mu_0$.

Miền bác bỏ: $$W_\alpha = \left\{ T = \frac{(\overline{X}-\mu_0)\sqrt{n}}{S} \; ; \; T > t_{\alpha}(n-1) \right\}. \qquad \textbf{(8.15)}$$

trong đó $t_{\alpha}(n-1)$ là giá trị tới hạn mức α của phân bố Student $n-1$ bậc tự do.

c) Bài toán 3: H_0: $\mu = \mu_0$; H_1: $\mu < \mu_0$.

Miền bác bỏ: $$W_\alpha = \left\{ T = \frac{(\overline{X}-\mu_0)\sqrt{n}}{S} \; ; \; -T > t_{\alpha}(n-1) \right\}. \qquad \textbf{(8.16)}$$

Ví dụ 8.3: 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 sất giống cây trồng là một biến ngẫu nhiên có phân bố chuẩn $N(\mu;\sigma^2)$.

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: $T = \dfrac{(\overline{X}-21{,}5)\sqrt{n}}{S}$.

Tra bảng ta tính được giá trị tới hạn mức $\frac{\alpha}{2} = 0{,}025$ của phân bố Student 15 bậc tự do là 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 $|T_{qs}| = 1,44 < 2,131$ nên chưa có cơ sở để bác bỏ H$_0$. Có nghĩa là với số liệu này thì có thể chấp nhận lời quảng cáo của công ty.

8.2.2. 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 $N(\mu;\sigma^2)$, 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$_0$: $\sigma^2 = \sigma_0^2$.

Từ tổng thể rút ra một mẫu ngẫu nhiên kích thước n: $W = (X_1, X_2, ..., X_n)$.

Nếu giả thiết H$_0$ đúng, theo công thức (6.31) thì thống kê sau có phân bố khi bình phương $n-1$ bậc tự do

$$T = \frac{(n-1)S^2}{\sigma_0^2} \tag{8.17}$$

Với mức ý nghĩa α, ta xây dựng các miền bác bỏ tùy thuộc vào đối thiết H$_1$ như sau:

a. H$_0$: $\sigma^2 = \sigma_0^2$; H$_1$: $\sigma^2 \neq \sigma_0^2$.

$$W_\alpha = \left\{ T = \frac{(n-1)S^2}{\sigma_0^2} \; ; \; T < \chi^2_{1-\frac{\alpha}{2}}(n-1) \quad \text{hoặc} \quad T > \chi^2_{\frac{\alpha}{2}}(n-1) \right\} \tag{8.18}$$

b. H$_0$: $\sigma^2 = \sigma_0^2$; H$_1$: $\sigma^2 > \sigma_0^2$.

$$W_\alpha = \left\{ T = \frac{(n-1)S^2}{\sigma_0^2} \; ; \; T > \chi^2_{\alpha}(n-1) \right\} \tag{8.19}$$

c. H$_0$: $\sigma^2 = \sigma_0^2$; H$_1$: $\sigma^2 < \sigma_0^2$.

$$W_\alpha = \left\{ T = \frac{(n-1)S^2}{\sigma_0^2} \; ; \; T < \chi^2_{1-\alpha}(n-1) \right\} \tag{8.20}$$

Ví dụ 8.4: 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 α =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 σ^2 =12.

Giải: Gọi X là kích thước chi tiết, theo giả thiết X có phân bố chuẩn $N(\mu;\sigma^2)$. Vậy đây là bài toán kiểm định giả thiết:

H$_0$: $\sigma^2 = 12$; H$_1$: $\sigma^2 \neq 12$.

Do $\alpha = 0{,}01$, tra bảng khi bình phương 14 bậc tự do ta được:

$$\chi^2_{\frac{\alpha}{2}}(n-1) = \chi^2_{0,005}(14) = 3{,}57\,;\ \chi^2_{1-\frac{\alpha}{2}}(n-1) = \chi^2_{0,995}(14) = 28{,}3\,.$$

Thay giá trị mẫu cụ thể vào công thức (8.17) ta tính được giá trị quan sát của tiêu chuẩn kiểm định:

$$T_{qs} = \frac{14\cdot 14{,}6}{12} = 17{,}03\,.$$

Giá trị này không rơi vào miền bác bỏ, vậy chưa có cơ sở để bác bỏ H_0 hay có thể nói máy móc vẫn làm việc bình thường.

8.2.3. Kiểm định giả thiết về kỳ vọng của biến ngẫu nhiên có phân bố không – một A(p)

Giả sử ta để ý đến một 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 VII dấu hiệu nghiên cứu này là một biến ngẫu nhiên X có phân bố không – một $A(p)$ với kỳ vọng bằng p. Nếu p chưa biết, song có cơ sở để giả thiết rằng giá trị này bằng p_0. Ta kiểm định giả thiết 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 (5.19)-(6.21)). Xét thống kê

$$T = \frac{(f-p_0)\sqrt{n}}{\sqrt{p_0(1-p_0)}} \tag{8.21}$$

Nếu giả thiết H_0 đúng, áp dụng định lý giới hạn trung tâm (công thức (6.41)) thì khi n đủ lớn thống kê trên xấp xỉ phân bố chuẩn tắc $N(0;1)$.

Trong thực tế khi

$$\begin{cases} np_0 > 5 \\ n(1-p_0) > 5 \end{cases} \tag{8.22}$$

thì có thể xem thống kê (8.21) có phân bố chuẩn tắc $N(0;1)$. Do đó với mức ý nghĩa α và tùy thuộc đối thiết H_1 ta có thể xây dựng các miền bác bỏ tương ứng.

a) H_0: $p = p_0$; H_1: $p \neq p_0$.

$$W_\alpha = \left\{ T = \frac{(f-p_0)\sqrt{n}}{\sqrt{p_0(1-p_0)}}\ ;\ |T| > U_{\frac{\alpha}{2}} \right\} \tag{8.23}$$

b) H_0: $p = p_0$; H_1: $p > p_0$.

$$W_\alpha = \left\{ T = \frac{(f-p_0)\sqrt{n}}{\sqrt{p_0(1-p_0)}}\ ;\ T > U_\alpha \right\} \tag{8.24}$$

c) H₀: $p = p_0$; H₁: $p < p_0$.

$$W_\alpha = \left\{ T = \frac{(f - p_0)\sqrt{n}}{\sqrt{p_0(1-p_0)}} \; ; \; -T > U_\alpha \right\} \quad \textbf{(8.25)}$$

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 8.5: Một đảng chính trị trong một cuộc bầu cử tổng thống ở nước nọ tuyên bố rằng có 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 862 cử tri tuyên bố sẽ bỏ phiếu cho A.

Với mức $\alpha = 5\%$, 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: $p \neq 0{,}45$.

(Bởi vì ta không có cơ sở nào để cho rằng dự đoán của đảng trên là cao hơn 0,45 hay thấp hơn 0,45).

Vì rằng 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 nên có thể chọn tiêu chuẩn kiểm định theo công thức (8.21). Thay mẫu cụ thể với $f = \dfrac{862}{2000} = 0{,}431$ ta được giá trị quan sát của tiêu chuẩn kiểm định $T_{qs} = \dfrac{(0{,}431 - 0{,}45)\sqrt{2000}}{\sqrt{0{,}45 \cdot 0{,}55}} = -1{,}708$.

Với mức ý nghĩa $\alpha = 0{,}05 \Rightarrow U_{\frac{\alpha}{2}} = 1{,}96$. Ta thấy $|T_{qs}| < 1{,}96$. Vậy không có cơ sở để bác bỏ H_0.

8.2.4. 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 tổng thể. Ở tổng thể thứ nhất dấu hiệu nghiên cứu được xem như biến ngẫu nhiên X_1 có phân bố chuẩn $N(\mu_1;\sigma_1^2)$. Ở tổng thể thứ hai dấu hiệu nghiên cứu được xem như biến ngẫu nhiên X_2 có phân bố chuẩn $N(\mu_2;\sigma_2^2)$. Nếu μ_1 và μ_2 chưa biết song có cơ sở để giả thiết rằng giá trị của chúng bằng nhau, người ta đưa ra giả thiết thống kê: H_0: $\mu_1 = \mu_2$

Ta kiểm định giả thiết trên theo các trường hợp sau:

8.2.4.1. Trường hợp hai phương sai σ_1^2 và σ_2^2 của hai biến ngẫu nhiên gốc đã biết

Từ hai tổng thể nói trên rút ra hai mẫu ngẫu nhiên độc lập kích thước tương ứng là n_1 và n_2

$$W_1 = (X_{11}, X_{12}, ..., X_{1n1}) ; \qquad W_2 = (X_{21}, X_{22}, ..., X_{2n2}) .$$

Chọn thống kê sau làm tiêu chuẩn kiểm định:

$$T = \frac{(\overline{X_1} - \overline{X_2}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \tag{8.26}$$

Theo công thức (6.33) chương VI, thống kê T có phân bố chuẩn tắc $N(0;1)$. Nếu giả thiết H_0 đúng thì thống kê T có dạng:

$$T = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0;1) \tag{8.27}$$

Vì vậy với mức ý nghĩa α cho trước và tùy thuộc vào dạng của đối thiết H_1, ta xây dựng các miền bác bỏ tương ứng:

a) H_0: $\mu_1 = \mu_2$; H_1: $\mu_1 \neq \mu_2$ (kiểm định hai phía).

$$W_\alpha = \left\{ T = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \ ; \ |T| > U_{\frac{\alpha}{2}} \right\} \tag{8.28}$$

b) H_0: $\mu_1 = \mu_2$; H_1: $\mu_1 > \mu_2$ (kiểm định một phía).

$$W_\alpha = \left\{ T = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \ ; \ T > U_\alpha \right\} \tag{8.29}$$

c) H_0: $\mu_1 = \mu_2$; H_1: $\mu_1 < \mu_2$ (kiểm định một phía).

$$W_\alpha = \left\{ T = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \ ; \ -T > U_\alpha \right\} \tag{8.30}$$

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 8.6: 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 $\sigma_1^2 = \sigma_2^2 = 0,16$.

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: $\mu_1 \neq \mu_2$.

Theo công thức (8.27) tiêu chuẩn kiểm định là $T = \dfrac{\overline{X_1} - \overline{X_2}}{\sqrt{\dfrac{0,16}{5} + \dfrac{0,16}{6}}}$.

Do $\alpha = 0,05 \Rightarrow U_{\frac{\alpha}{2}} = 1,96$. 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 $T_{qs} = \dfrac{3,3 - 2,5}{\sqrt{\dfrac{0,16}{5} + \dfrac{0,16}{6}}} = 3,33 > 1,96$. Vậy $T_{qs} \in W_\alpha$, 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.

8.2.4.2. Trường hợp hai phương sai σ_1^2 và σ_2^2 của hai biến ngẫu nhiên gốc chưa biết nhưng $\sigma_1^2 = \sigma_2^2$

Từ hai tổng thể nói trên rút ra hai mẫu ngẫu nhiên độc lập kích thước tương ứng là n_1 và n_2

$$W_1 = (X_{11}, X_{12}, \dots, X_{1n_1}); \quad W_2 = (X_{21}, X_{22}, \dots, X_{2n_2}).$$

Chọn thống kê sau làm tiêu chuẩn kiểm định:

$$T = \frac{(\overline{X_1} - \overline{X_2}) - (\mu_1 - \mu_2)}{Sp\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} \quad \textbf{(8.31)}$$

Với $Sp = \sqrt{\dfrac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}$

Theo công thức (6.36) chương VI, 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 có dạng :

$$T = \frac{\left(\overline{X_1} - \overline{X_2}\right)}{Sp\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \tag{8.32}$$

Vì vậy với mức ý nghĩa α , ta xây dựng các miền bác bỏ tương ứng:

a) H_0: $\mu_1 = \mu_2$; H_1: $\mu_1 \neq \mu_2$ (kiểm định hai phía).

$$W_\alpha = \left\{ T = \frac{\left(\overline{X_1} - \overline{X_2}\right)}{Sp\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \;;\; |T| > t_{\alpha/2}(n_1 + n_2 - 2) \right\} \tag{8.33}$$

b) H_0: $\mu_1 = \mu_2$; H_1: $\mu_1 > \mu_2$ (kiểm định một phía).

$$W_\alpha = \left\{ T = \frac{\left(\overline{X_1} - \overline{X_2}\right)}{Sp\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \;;\; T > t_{\alpha}(n_1 + n_2 - 2) \right\} \tag{8.34}$$

c) H_0: $\mu_1 = \mu_2$; H_1: $\mu_1 < \mu_2$ (kiểm định một phía).

$$W_\alpha = \left\{ T = \frac{\left(\overline{X_1} - \overline{X_2}\right)}{Sp\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \;;\; -T > t_{\alpha}(n_1 + n_2 - 2) \right\} \tag{8.35}$$

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 8.7: 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

$$\text{H}_0\text{: } \mu_1 = \mu_2 \text{ ; } \text{H}_1\text{: } \mu_1 \neq \mu_2 .$$

Ta sử dụng tiêu chuẩn kiểm định (8.32) và miền bác bỏ W_α (8.33).

Với $\alpha = 0,05 \Rightarrow t_{\alpha/2}(n_1 + n_2 - 2) = t_{0,025}(37) = 2,021$.

Từ mẫu cụ thể ta tính được $Sp = \sqrt{\dfrac{19 \cdot 8,74 + 18 \cdot 6,58}{20 + 19 - 2}} = 2,773$.

Do đó giá trị quan sát của tiêu chuẩn kiểm định $T_{qs} = \dfrac{18,27 - 16,78}{0,773\sqrt{\dfrac{1}{20} + \dfrac{1}{19}}} = 1,677$.

Vì $T_{qs} \notin W_\alpha$ do đó với hai mẫu cụ thể này thì chưa có cơ sở để bác bỏ H₀, tức là có thể xem thu nhập trung bình hàng năm của dân hai thành phố đó là như nhau.

8.2.4.3. 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:

$$T = \frac{\left(\overline{X_1} - \overline{X_2}\right) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}} \quad \textbf{(8.36)}$$

Theo công thức (6.37) chương VI, thống kê T có phân bố Student k bậc tự do, trong đó:

$$k = \frac{(n_1 - 1)(n_2 - 1)}{(n_2 - 1)C^2 + (n_1 - 1)(1 - C)^2} \;;\; C = \frac{\dfrac{S_1^2}{n_1}}{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}.$$

Nếu giả thiết H₀ đúng thì thống kê T có dạng:

$$T = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}} \quad \textbf{(8.37)}$$

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).

$$W_\alpha = \left\{ T = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \; ; \; |T| > t_{\alpha/2}(k) \right\} \qquad \textbf{(8.38)}$$

b) H₀: $\mu_1 = \mu_2$; H₁: $\mu_1 > \mu_2$ (kiểm định một phía).

$$W_\alpha = \left\{ T = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \; ; \; T > t_\alpha(k) \right\} \qquad \textbf{(8.39)}$$

c) H₀: $\mu_1 = \mu_2$; H₁: $\mu_1 < \mu_2$ (kiểm định một phía).

$$W_\alpha = \left\{ T = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \; ; \; -T > t_\alpha(k) \right\} \qquad \textbf{(8.40)}$$

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 8.8: Để kiểm nghiệm hiệu quả của một loại thuốc tẩy giun cho lợn, người ta bắt ngẫu nhiên 14 con lợn từ một trại chăn nuôi và chia thành 2 nhóm: Nhóm I cho uống thuốc tẩy giun và Nhóm II không cho uống thuốc tẩy giun.

Sau thời gian dùng thuốc, khi giết thịt hai nhóm lợn trên cho kết quả sau về số giun có trong những con lợn thuộc hai nhóm trên.

Nhóm I:	18	43	28	50	16	32	13
Nhóm II:	40	54	26	63	21	37	39

Với mức ý nghĩa $\alpha = 0{,}05$ hãy kết luận xem loại thuốc tẩy giun nói trên có thực sự hiệu quả hay không. Giả thiết rằng số lượng giun có phân bố chuẩn.

Giải: Gọi X_1 và X_2 tương ứng là số giun trong mỗi con lợn 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 giun trung bình là μ_1 và μ_2. Để kiểm định giả thiết:

H₀: $\mu_1 = \mu_2$; H₁: $\mu_1 < \mu_2$.

Ta sử dụng tiêu chuẩn kiểm định (8.37) và miền bác bỏ W_α (8.40).

Từ hai mẫu cụ thể ta tính được: $n_1 = 7$; $\overline{x_1} = 28,57$; $s_1^2 = 198,62$.

$$n_2 = 7 \ ; \ \overline{x_2} = 40,00 \ ; \ s_2^2 = 215,33 .$$

Ta có $C = \dfrac{\frac{S_1^2}{n_1}}{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} = 0,4798 \Rightarrow k = \dfrac{6 \cdot 6}{6 \cdot 0,4798^2 + 6 \cdot 0,5202^2} \approx 12$.

Với $\alpha = 0,05 \Rightarrow t_\alpha(k) = t_{0,05}(12) = 1,782$.

Giá trị quan sát của tiêu chuẩn kiểm định $T_{qs} = \dfrac{28,52 - 40}{\sqrt{\dfrac{198,62}{7} + \dfrac{215,33}{19}}} = -1,49$.

Vì $T_{qs} \notin W_\alpha$ do đó với mức ý nghĩa 0,05 và từ hai mẫu cụ thể này thì chưa có cơ sở để bác bỏ H_0, tức là chưa thể nói loại thuốc tẩy giun được thử nghiệm là có hiệu quả.

8.2.5. Kiểm định giả thiết về sự bằng nhau của hai phương sai của hai biến ngẫu nhiên có phân bố chuẩn (*)

Xét hai biến ngẫu nhiên gốc của hai tổng thể cùng có phân bố chuẩn. biến ngẫu nhiên X_1 có phân bố chuẩn $N(\mu_1; \sigma_1^2)$. biến ngẫu nhiên X_2 có phân bố chuẩn $N(\mu_2; \sigma_2^2)$. Khác với trường hợp 2.4, ở đây ta cần kiểm định sự bằng nhau của hai phương sai σ_1^2 và σ_2^2:

$$H_0: \sigma_1^2 = \sigma_2^2$$

Để kiểm định giả thiết trên từ hai tổng thể nói trên rút ra hai mẫu ngẫu nhiên độc lập kích thước tương ứng là n_1 và n_2:

$$W_1 = (X_{11}, X_{12}, ..., X_{1n_1}) \ ; \ W_2 = (X_{21}, X_{22}, ..., X_{2n_2}) .$$

Chọn thống kê sau làm tiêu chuẩn kiểm định:

$$T = \frac{S_1^2 \sigma_2^2}{S_2^2 \sigma_1^2} . \qquad \textbf{(8.41)}$$

Giả sử $S_1^2 > S_2^2$, từ công thức (6.38) chương VI ta biết thống kê T có phân bố Fisher-Snedecor với $n_1 - 1$ và $n_2 - 1$ bậc tự do.

Nếu giả thiết H_0 đúng thì tiêu chuẩn kiểm định có dạng:

$$T = \frac{S_1^2}{S_2^2}\ ;\ \ (S_1^2 > S_2^2)\,. \tag{8.42}$$

và vẫn có phân bố Fisher-Snedecor với $n_1 - 1$ và $n_2 - 1$ bậc tự do. Do đó với mức ý nghĩa α cho trước và tùy thuộc vào dạng của đối thiết H_1, ta xây dựng các miền bác bỏ tương ứng:

a) H_0: $\sigma_1^2 = \sigma_2^2$; H_1: $\sigma_1^2 \neq \sigma_2^2$ (kiểm định hai phía).

$$W_\alpha = \left\{ T = \frac{S_1^2}{S_2^2}\ ;\ \ T < f_{1-\frac{\alpha}{2}}(n_1 - 1, n_2 - 1) \ \text{ hoặc } \ T > f_{\alpha/2}(n_1 - 1, n_2 - 1) \right\} \quad \textbf{(8..43)}$$

b) H_0: $\sigma_1^2 = \sigma_2^2$; H_1: $\sigma_1^2 > \sigma_2^2$ (kiểm định một phía).

$$W_\alpha = \left\{ T = \frac{S_1^2}{S_2^2}\ ;\ \ T > f_\alpha(n_1 - 1, n_2 - 1) \right\} \tag{8.44}$$

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 8.9: Có hai giống lúa có năng suất trung bình xấp xỉ như nhau song mức độ phân tán về năng suất có thể khác nhau. Để kiểm tra điều đó người ta gặt mẫu tại 2 vùng trồng hai giống lúa đó và thu được kết quả sau:

Giống lúa	Số điểm gặt	Phương sai
A	$n_1 = 41$	$s_1^2 = 11,41$
B	$n_2 = 30$	$s_2^2 = 6,52$

Với mức ý nghĩa $\alpha = 0,05$ hãy kết luận về vấn đề trên, biết năng suất lúa là biến ngẫu nhiên có phân bố chuẩn.

Giải: Gọi X_1 và X_2 tương ứng là năng suất của hai giống lúa A và B. Theo giả thiết X_1 và X_2 có phân bố chuẩn. Đây là bài toán kiểm định cặp giả thiết:

$$H_0:\ \sigma_1^2 = \sigma_2^2;\quad H_1:\ \sigma_1^2 \neq \sigma_2^2.$$

Ta sử dụng tiêu chuẩn kiểm định (8.42) và miền bác bỏ W_α (8.43).

Với $\alpha = 0,05 \Rightarrow \begin{cases} f_{\frac{\alpha}{2}}(n_1 - 1, n_2 - 1) = f_{0,025}(40, 29) = 2,03 \\ f_{1-\frac{\alpha}{2}}(n_1 - 1, n_2 - 1) = \dfrac{1}{f_{\frac{\alpha}{2}}(n_1 - 1, n_2 - 1)} = 0,52 \end{cases}$.

Vậy miền bác bỏ là $W_\alpha = \left\{ T = \dfrac{S_1^2}{S_2^2} \;;\; T < 0,52 \text{ hoặc } T > 2,03 \right\}$.

Giá trị quan sát của tiêu chuẩn kiểm định $T_{qs} = \dfrac{11,41}{6,52} = 1,75$.

Vì $T_{qs} \notin W_\alpha$ do đó với mức ý nghĩa 0,05 và từ hai mẫu cụ thể này thì chưa có cơ sở để bác bỏ H_0, tức là độ phân tán của năng suất hai giống lúa trên là như nhau.

8.2.6. Kiểm định giả thiết về sự bằng nhau của hai kỳ vọng của hai biến ngẫu nhiên có phân bố $A(p)$

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 không – một $A(p_1)$. Ở tổng thể thứ hai dấu hiệu nghiên cứu được xem như biến ngẫu nhiên X_2 có phân bố không – một $A(p_2)$. 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, ta đưa ra giả thiết thống kê:

$$H_0:\; p_1 = p_2 .$$

Để kiểm định giả thiết trên từ hai tổng thể nói trên rút ra hai mẫu ngẫu nhiên độc lập kích thước tương ứng là n_1 và n_2:

$$W_1 = (X_{11}, X_{12}, ..., X_{1n_1})\;;\; W_2 = (X_{21}, X_{22}, ..., X_{2n_2}) .$$

Chọn thống kê sau làm tiêu chuẩn kiểm định

$$T = \frac{(f_1 - f_2) - (p_1 - p_2)}{\sqrt{\dfrac{p_1(1-p_1)}{n_1} + \dfrac{p_2(1-p_2)}{n_2}}} . \qquad \textbf{(8.45)}$$

Theo công thức (6.42) chương VI, thống kê T có phân bố xấp xỉ chuẩn tắc $N(0;1)$ 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:

$$T = \frac{f_1 - f_2}{\sqrt{p(1-p)\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}} .$$

Thông thường p chưa biết nên được thay bằng ước lượng 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:

$$T = \frac{f_1 - f_2}{\sqrt{\overline{f}(1-\overline{f})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}. \quad \textbf{(8.46)}$$

Thống kê T có phân bố xấp xỉ chuẩn tắc $N(0;1)$ nếu $n_1 > 30$ và $n_2 > 30$. Do đó tùy theo mức ý nghĩa α và đối thiết H₁ ta xây dựng các miền bác bỏ tương ứng:

a) H$_0$: $p_1 = p_2$; H$_1$: $p_1 \neq p_2$ (kiểm định hai phía).

$$W_\alpha = \left\{ T = \frac{f_1 - f_2}{\sqrt{\overline{f}(1-\overline{f})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \;;\; |T| > U_{\frac{\alpha}{2}} \right\} \quad \textbf{(8.47)}$$

b) H$_0$: $p_1 = p_2$; H$_1$: $p_1 > p_2$ (kiểm định một phía).

$$W_\alpha = \left\{ T = \frac{f_1 - f_2}{\sqrt{\overline{f}(1-\overline{f})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \;;\; T > U_\alpha \right\} \quad \textbf{(8.48)}$$

c) H$_0$: $p_1 = p_2$; H$_1$: $p_1 < p_2$ (kiểm định một phía).

$$W_\alpha = \left\{ T = \frac{f_1 - f_2}{\sqrt{\overline{f}(1-\overline{f})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \;;\; -T > U_\alpha \right\} \quad \textbf{(8.49)}$$

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 8.10: 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 B	$n_1 = 1100$ $n_2 = 950$	$x_1 = 20$ $x_2 = 31$

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$_1$: $p_1 \neq p_2$.

Chọn tiêu chuẩn kiểm định (8.46) và khi n_1 và n_2 đủ lớn thì miền bác bỏ dạng (8.47).

Do $\alpha = 0,05 \Rightarrow U_{\frac{\alpha}{2}} = 1,96$. Miền bác bỏ $W_\alpha = \{|T| > 1,96\}$.

Với hai mẫu cụ thể trên ta tính được

$$f_1 = \frac{22}{1100} = 0,02; \quad f_2 = \frac{31}{900} = 0,0344 \; ; \quad \overline{f} = \frac{22+31}{1100+900} = 0,0265;$$

$$T_{qs} = \frac{0,02 - 0,0344}{\sqrt{0,0265 \cdot 0,9735\left(\frac{1}{1100} + \frac{1}{900}\right)}} = -1,995.$$

Vì $T_{qs} \in W_\alpha$ do đó ta có thể bác bỏ giả thiết H$_0$, vậy tỷ lệ phế phẩm của hai nhà máy khác nhau.

8.3. KIỂM ĐỊNH PHI THAM SỐ

Trong mục 2 ta kiểm định các giả thiết về các dấu hiệu nghiên cứu của tổng thể có chứa các tham số và bài toán đưa về kiểm định giả thiết tham số. Tuy nhiên trong thực tế ta cũng cần kiểm định giả thiết không chứa các tham số, chẳng hạn kiểm định về quy luật phân bố xác suất của dấu hiệu nghiên cứu, kiểm định tính độc lập của hai dấu hiệu nghiên cứu …, đó là bài toán kiểm định phi tham số.

8.3.1. Kiểm định giả thiết về quy luật phân bố xác suất của biến ngẫu nhiên

Giả sử chưa biết quy luật phân bố xác suất của biến ngẫu nhiên gốc X trong tổng thể song có cơ sở để giả thiết rằng X có phân bố theo quy luật A nào đó. Ta có thể kiểm định:

Giả thiết H$_0$: X có phân bố xác suất theo quy luật A;

Đối thiết H$_1$: X có phân bố xác suất không theo quy luật A.

Sau đây ta sử dụng *tiêu chuẩn phù hợp Pearson* để kiểm định giả thiết trên.

Phân hoạch miền giá trị của X thành k khoảng rời nhau $C_1, C_2, ..., C_k$. Từ tổng thể rút ra mẫu ngẫu nhiên kích thước n và gọi $n_1, n_2, ..., n_k$ là các tần số tương ứng của mẫu nằm trong k khoảng này (gọi là tần số thực nghiệm).

Nếu giả thiết H_0 đúng ta có thể tính được các xác suất:

$$p_i = P\{X \in C_i\};\ i = \overline{1,k}\ .$$

Từ đó ta có tần số lý thuyết:

$$\hat{n}_i = np_i\ ;\ i = \overline{1,k}\ .$$

Tiêu chuẩn kiểm định được chọn là thống kê:

$$T = \sum_{i=1}^{k} \frac{(n_i - \hat{n}_i)^2}{\hat{n}_i} \qquad \textbf{(8.50)}$$

Người ta chứng minh được rằng không phụ thuộc vào quy luật phân bố xác suất của biến ngẫu nhiên gốc X trong tổng thể, khi $n \to \infty$ thống kê T sẽ hội tụ về quy luật phân bố "Khi bình phương" với $k-r-1$ bậc tự do, trong đó r là số tham số cần ước lượng của quy luật cần kiểm định. Các tham số này được ước lượng bằng phương pháp hợp lý cực đại. Chẳng hạn nếu quy luật cần kiểm định là quy luật phân bố chuẩn thì có hai tham số cần ước lượng là μ và σ^2 nên $r = 2$, còn nếu quy luật cần kiểm định là quy luật Poisson thì $r = 1$.

Với mức ý nghĩa α ta xây dựng miền bác bỏ:

$$W_\alpha = \left\{ T = \sum_{i=1}^{k} \frac{(n_i - \hat{n}_i)^2}{\hat{n}_i}\ ;\ T > \chi_\alpha(k-r-1) \right\} \qquad \textbf{(8.51)}$$

Chú ý rằng tiêu chuẩn phù hợp của Pearson chỉ áp dụng được khi kích thước mẫu đủ lớn ($n > 50$) và các tần số thực nghiệm cũng đủ lớn ($n_i \ge 5$). Khi các tần số quá nhỏ ($n_i < 5$) thì ta phải ghép các giá trị hay các khoảng giới hạn của mẫu lại để tăng giá trị tần số thực nghiệm lên.

Ví dụ 8.11: Số lời gọi đến một trạm điện thoại X trong một phút được cho trong bảng sau:

Số lời gọi (x_i)	0	1	2	3	4	5	≥ 6
Số phút gọi tương ứng (n_i)	18	22	25	19	11	3	2

Với mức ý nghĩa $\alpha = 0{,}01$ có thể coi X có phân bố theo quy luật Poisson được không?

Giải: Kiểm định giả thiết H_0: X phân bố theo quy luật Poisson;

Đối thiết H_1: X không phân bố theo quy luật Poisson.

Từ mẫu tính được $\overline{x} = 2$ là ước lượng của tham số λ trong phân bố Poisson. Để áp dụng tiêu chuẩn kiểm định (8.50) và miền bác bỏ (8.51) ta phải nhập lại để thỏa mãn điều kiện $n_i \geq 5$.

Số lời gọi (x_i)	0	1	2	3	4	≥ 5
Số phút gọi tương ứng (n_i)	18	22	25	19	11	5

Để tính giá trị quan sát của tiêu chuẩn kiểm định T_{qs} ta lập bảng sau:

x_i	n_i	$p_i = \dfrac{e^{-2}2^{x_i}}{x_i!}$	$\hat{n}_i = np_i$	$n_i - \hat{n}_i$	$\dfrac{(n_i - \hat{n}_i)^2}{\hat{n}_i}$
0	18	0,1353	13,53	4,47	1,477
1	22	0,2707	27,07	-5,07	0,95
2	25	0,2707	27,07	-2,07	0,158
3	19	0,1804	18,04	0,96	0,051
4	11	0,0902	9,02	1,98	0,435
≥ 5	5	0,0527	5,27	-0,27	0,014
$\sum$	$n = 100$	1,0000	100		3,085

Do $\alpha = 0{,}01 \Rightarrow \chi^2_\alpha(k - r - 1) = \chi^2_{0,01}(6 - 1 - 1) = 13{,}277$. Vậy $T_{qs} \notin W_\alpha$ do đó chưa có cơ sở để bác bỏ H$_0$ hay có thể coi X phân bố theo quy luật Poisson.

8.3.2. Kiểm định giả thiết về tính độc lập của hai dấu hiệu định tính(*)

Giả sử cần nghiên cứu đồng thời hai dấu hiệu định tính A và B. Ta chia dấu hiệu A thành h mức độ $A_1, A_2, \ldots, A_h$ và dấu hiệu B thành k mức độ $B_1, B_2, \ldots, B_k$.

Ta kiểm định giả thiết H$_0$: A và B độc lập;

Đối thiết H$_1$: A và B phụ thuộc.

Để kiểm định giả thiết trên từ tổng thể lập mẫu ngẫu nhiên kích thước n và trình bày dưới dạng sau:

A \ B	B_1	B_2	…	B_k	Tổng
A_1	n_{11}	n_{12}	…	n_{1k}	n_{1*}
A_2	n_{21}	n_{22}	…	n_{2k}	n_{2*}
…	…	…	…	…	…
A_h	n_{h1}	n_{h2}	…	n_{hk}	n_{h*}
Tổng	n_{*1}	n_{*2}	…	n_{*k}	n

trong đó: n_{i*} là tổng số các tần số ứng với dấu hiệu thành phần A_i;

n_{*j} là tổng số các tần số ứng với dấu hiệu thành phần B_j;

n_{ij} là tần số ứng với các phần tử đồng thời mang dấu hiệu A_i và B_j.

Với n đủ lớn thì theo định nghĩa thống kê về xác suất ta có:

$$P(A_iB_j) \approx \frac{n_{ij}}{n},\ i = \overline{1,h}\ ;\quad j = \overline{1,k}\ .$$

$$P(A_i) \approx \frac{n_{i*}}{n}\ ,\ i = \overline{1,h}\ .\quad P(B_j) \approx \frac{n_{*j}}{n}\ ,\ j = \overline{1,k}\ .$$

Nếu giả thiết H_0 đúng tức là A và B độc lập thì với mọi $i = \overline{1,h}\ ; j = \overline{1,k}$:

$$P(A_iB_j) = P(A_i)P(B_j) \text{ tức là } \frac{n_{ij}}{n} = \frac{n_{i*}}{n} \cdot \frac{n_{*j}}{n}\ .$$

Xét thống kê: $$T = n\sum_{i=1}^{h}\sum_{j=1}^{k}\frac{\left(\dfrac{n_{ij}}{n} - \dfrac{n_{i*}}{n}\cdot\dfrac{n_{*j}}{n}\right)^2}{\dfrac{n_{i*}}{n}\cdot\dfrac{n_{*j}}{n}}$$

Khai triển vế phải công thức trên và rút gọn ta được

$$T = n\left(\sum_{i=1}^{h}\sum_{j=1}^{k}\frac{n_{ij}^2}{n_{i*}\cdot n_{*j}} - 1\right) \quad \textbf{(8.52)}$$

Có thể chứng minh được rằng khi $n \to \infty$ phân bố xác suất của T dần tới phân bố χ^2 với $(h-1)(k-1)$ bậc tự do. Vì vậy khi n đủ lớn, với mức ý nghĩa α ta xây dựng miền bác bỏ:

$$W_\alpha = \left\{ T = n\left(\sum_{i=1}^{h} \sum_{j=1}^{k} \frac{n_{ij}^2}{n_{i*} \cdot n_{*j}} - 1 \right);\ T > \chi_\alpha^2 (h-1)(k-1) \right\} \qquad \textbf{(8.53)}$$

Ví dụ 8.12: Bảng sau cho ta tần số sinh viên ứng với 3 mức của A (khả năng học toán: thấp, trung bình, cao) và 3 mức của B (sự quan tâm đến môn xác suất thống kê). Với mức ý nghĩa $\alpha = 0{,}01$ hãy kiểm định giả thiết H₀: A và B độc lập với nhau.

B \ A	Thấp	Trung bình	Cao	Tổng
Ít	43	24	15	82
Trung bình	31	46	37	114
Nhiều	12	20	49	81
Tổng	86	90	101	277

Với $\alpha = 0{,}01$ ta có $\chi_\alpha^2 (h-1)(k-1) = \chi_{0,01}^2 (4) = 13{,}28$. Từ mẫu cụ thể trên ta tính được giá trị quan sát của tiêu chuẩn kiểm định $T_{qs} = 33{,}533 > 13{,}28$ nên H_0 bị bác bỏ. Vậy các đặc tính A và B không độc lập, tức năng lực toán học của sinh viên có quan hệ với việc quan tâm hay thích thú môn xác suất thống kê.

8.3.3. Kiểm định giả thiết về tần số của nhiều tổng thể (*)

Ta xét k tổng thể $\mathscr{H}_1, \mathscr{H}_2, \dots, \mathscr{H}_k$. Mỗi cá thể của chúng có thể mang hay không mang dấu hiệu A . Gọi p_i là tần suất có dấu hiệu A của tổng thể $\mathscr{H}_i$, ($i = \overline{1,k}$).

Ta kiểm định giả thiết H₀: $p_1 = p_2 = \dots = p_k$.

Đối thiết H₁: Các tần suất p_i ($i = \overline{1,k}$) không đồng thời bằng nhau.

Từ mỗi tổng thể $\mathscr{H}_i$ rút ra một mẫu ngẫu nhiên có kích thước n_i trong đó có m_i cá thể có dấu hiệu A và $l_i = n_i - m_i$ cá thể không có dấu hiệu A. Các mẫu ngẫu nhiên này được trình bày trong bảng sau:

Tổng thể	1	2	…	k	Tổng
Có A	m_1	m_2	…	m_k	m
Không A	l_1	l_2	…	l_k	l
Tổng	n_1	n_2	…	n_k	$N = m + l = \sum_{i=1}^{k} n_i$

Có thể chứng minh được rằng nếu giả thiết H_0 đúng và khi kích thước mẫu đủ lớn (tần số thực nghiệm $m_i \geq 5$, $l_i \geq 5$) thì thống kê sau xấp xỉ phân bố χ^2 với $k-1$ bậc tự do:

$$T = \frac{N^2}{ml}\sum_{i=1}^{k}\frac{m_i^2}{n_i} - N\frac{m}{l} \qquad \textbf{(8.54)}$$

Ta chọn thống kê trên làm tiêu chuẩn kiểm định của giả thiết H_0.

Miền bác bỏ $$W_\alpha = \left\{T = \frac{N^2}{ml}\sum_{i=1}^{k}\frac{m_i^2}{n_i} - N\frac{m}{l};\ T > \chi_\alpha^2(k-1)\right\} \qquad \textbf{(8.55)}$$

Ví dụ 8.13: So sánh tác dụng của 6 mẫu thuốc thử nghiệm trên 6 lô chuột, kết quả thu được như sau:

Mẫu thuốc	1	2	3	4	5	6	Tổng
Số sống	79	82	77	83	76	81	478
Số chết	21	18	23	17	24	19	122
Tổng	100	100	100	100	100	100	600

Với mức ý nghĩa $\alpha = 5\%$ hãy kiểm định giả thiết H_0: Tỉ lệ chết trong 6 mẫu thuốc là như nhau; Đối thiết H_1: Tỉ lệ chết trong 6 mẫu thuốc khác nhau.

Giải: Tra bảng phân bố χ^2 với 5 bậc tự do ta có $\chi_\alpha^2(k-1) = \chi_{0,05}^2(5) = 11,07$.

Từ mẫu trên ta có

$$T_{qs} = \frac{600^2}{478\cdot 122}\left(\frac{79^2}{100} + \frac{82^2}{100} + \cdots + \frac{81^2}{100}\right) - 600\cdot\frac{478}{122} = 2,42 < 11,07.$$

Vậy $T_{qs} \notin W_\alpha$ do đó chưa có cơ sở để bác bỏ H_0 hay có thể cho rằng tỉ lệ chết trong 6 mẫu thuốc là như nhau.

TÓM TẮT

Giả thiết thống kê

Giả thiết thống kê là giả thiết về dạng phân bố xác suất của biến ngẫu nhiên gốc của tổng thể, các tham số đặc trưng hoặc tính chất của các biến ngẫu nhiên này. Giả thiết thống kê là những điều ta nghi ngờ muốn bác bỏ, được phát biểu dưới dạng H_0. Cạnh tranh với giả thiết này là đối thiết H_1, theo nghĩa rằng nếu bác bỏ H_0 thì chấp nhận H_1 và ngược lại.

Thủ tục kiểm định giả thiết thống kê

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.

Kiểm định giả thiết về kỳ vọng của biến ngẫu nhiên có phân bố theo quy luật chuẩn

1) ***Trường hợp đã biết phương sai***

Tiêu chuẩn kiểm định: $T = \dfrac{(\overline{X} - \mu_0)\sqrt{n}}{\sigma}$.

a. Bài toán 1: H_0: $\mu = \mu_0$; H_1: $\mu \neq \mu_0$. Miền bác bỏ $W_\alpha = \left\{|T| > U_{\alpha/2}\right\}$.

b. Bài toán 2: H_0: $\mu = \mu_0$; H_1: $\mu > \mu_0$. Miền bác bỏ $W_\alpha = \left\{ T > U_\alpha \right\}$.

c. Bài toán 3: H_0: $\mu = \mu_0$; H_1: $\mu < \mu_0$. Miền bác bỏ $W_\alpha = \left\{ -T > U_\alpha \right\}$.

2) ***Trường hợp chưa biết phương sai*** $n \geq 30$

Tiêu chuẩn kiểm định: $T = \dfrac{(\overline{X} - \mu_0)\sqrt{n}}{S}$.

a. Bài toán 1: H_0: $\mu = \mu_0$; H_1: $\mu \neq \mu_0$. Miền bác bỏ $W_\alpha = \left\{|T| > U_{\alpha/2}\right\}$.

b. Bài toán 2: H_0: $\mu = \mu_0$; H_1: $\mu > \mu_0$. Miền bác bỏ $W_\alpha = \left\{ T > U_\alpha \right\}$.

c. Bài toán 3: H_0: $\mu = \mu_0$; H_1: $\mu < \mu_0$. Miền bác bỏ $W_\alpha = \left\{ -T > U_\alpha \right\}$.

3) ***Trường hợp chưa biết phương sai*** $n < 30$

Tiêu chuẩn kiểm định: $T = \dfrac{(\overline{X} - \mu_0)\sqrt{n}}{S}$.

a. Bài toán 1: H_0: $\mu = \mu_0$; H_1: $\mu \neq \mu_0$. Miền bác bỏ $W_\alpha = \left\{|T| > t_{\alpha/2}(n-1)\right\}$.

b. Bài toán 2: H_0: $\mu = \mu_0$; H_1: $\mu > \mu_0$. Miền bác bỏ $W_\alpha = \left\{ T > t_\alpha(n-1) \right\}$.

c. Bài toán 3: H_0: $\mu = \mu_0$; H_1: $\mu < \mu_0$. Miền bác bỏ $W_\alpha = \left\{ -T > t_\alpha(n-1) \right\}$.

Kiểm định giả thiết về phương sai của biến ngẫu nhiên có phân bố chuẩn

Tiêu chuẩn kiểm định: $T = \dfrac{(n-1)S^2}{\sigma_0^2}$.

a. H_0: $\sigma^2 = \sigma_0^2$; H_1: $\sigma^2 \neq \sigma_0^2$:

$$W_\alpha = \left\{ T < \chi^2_{1-\frac{\alpha}{2}}(n-1) \quad \text{hoặc} \quad T > \chi^2_{\frac{\alpha}{2}}(n-1) \right\}.$$

b. H_0: $\sigma^2 = \sigma_0^2$; H_1: $\sigma^2 > \sigma_0^2$: $W_\alpha = \left\{ T > \chi^2_\alpha(n-1) \right\}$.

c. H_0: $\sigma^2 = \sigma_0^2$; H_1: $\sigma^2 < \sigma_0^2$: $W_\alpha = \left\{ T < \chi^2_{1-\alpha}(n-1) \right\}$.

Kiểm định giả thiết về kỳ vọng của biến ngẫu nhiên có phân bố không – một A(p)

Tiêu chuẩn kiểm định: $T = \dfrac{(f - p_0)\sqrt{n}}{\sqrt{p_0(1-p_0)}}$ với điều kiện $\begin{cases} np_0 > 5 \\ n(1-p_0) > 5 \end{cases}$.

f là tần suất mẫu.

a. H_0: $p = p_0$; H_1: $p \neq p_0$. $W_\alpha = \left\{ |T| > U_{\alpha/2} \right\}$.

b. H_0: $p = p_0$; H_1: $p > p_0$. $W_\alpha = \left\{ T > U_\alpha \right\}$.

c. H_0: $p = p_0$; H_1: $p < p_0$. $W_\alpha = \left\{ -T > U_\alpha \right\}$.

Kiểm định giả thiết về hai kỳ vọng của hai biến ngẫu nhiên có phân bố chuẩn

1) Trường hợp hai phương sai σ_1^2 và σ_2^2 của hai biến ngẫu nhiên gốc đã biết

Tiêu chuẩn kiểm định: $T = \dfrac{\left(\overline{X_1} - \overline{X_2}\right) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$

a) H_0: $\mu_1 = \mu_2$; H_1: $\mu_1 \neq \mu_2$: $W_\alpha = \left\{ |T| > U_{\alpha/2} \right\}$.

b) H_0: $\mu_1 = \mu_2$; H_1: $\mu_1 > \mu_2$: $W_\alpha = \left\{ T > U_\alpha \right\}$.

c) H_0: $\mu_1 = \mu_2$; H_1: $\mu_1 < \mu_2$: $W_\alpha = \left\{ -T > U_\alpha \right\}$.

2) Trường hợp hai phương sai σ_1^2 và σ_2^2 của hai biến ngẫu nhiên gốc chưa biết nhưng $\sigma_1^2 = \sigma_2^2$

Tiêu chuẩn kiểm định: $T = \dfrac{\left(\overline{X_1} - \overline{X_2}\right)}{Sp\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$,

với $Sp = \sqrt{\dfrac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}$

a) H$_0$: $\mu_1 = \mu_2$; H$_1$: $\mu_1 \neq \mu_2$: $W_\alpha = \left\{ |T| > t_{\alpha/2}(n_1 + n_2 - 2) \right\}$.

b) H$_0$: $\mu_1 = \mu_2$; H$_1$: $\mu_1 > \mu_2$: $W_\alpha = \left\{ T > t_\alpha(n_1 + n_2 - 2) \right\}$.

c) H$_0$: $\mu_1 = \mu_2$; H$_1$: $\mu_1 < \mu_2$: $W_\alpha = \left\{ -T > t_\alpha(n_1 + n_2 - 2) \right\}$.

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 ($\sigma_1^2 \neq \sigma_2^2$)

Tiêu chuẩn kiểm định: $T = \dfrac{\overline{X_1} - \overline{X_2}}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}}$; $k = \dfrac{(n_1 - 1)(n_2 - 1)}{(n_2 - 1)C^2 + (n_1 - 1)(1 - C)^2}$;

$C = \dfrac{\dfrac{S_1^2}{n_1}}{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}$. T có phân bố Student k bậc tự do.

a) H$_0$: $\mu_1 = \mu_2$; H$_1$: $\mu_1 \neq \mu_2$: $W_\alpha = \left\{ |T| > t_{\alpha/2}(k) \right\}$.

b) H$_0$: $\mu_1 = \mu_2$; H$_1$: $\mu_1 > \mu_2$: $W_\alpha = \left\{ T > t_\alpha(k) \right\}$.

c) H$_0$: $\mu_1 = \mu_2$; H$_1$: $\mu_1 < \mu_2$: $W_\alpha = \left\{ -T > t_\alpha(k) \right\}$.

Kiểm định giả thiết về sự bằng nhau của hai phương sai của hai biến ngẫu nhiên có phân bố chuẩn

X_1 có phân bố chuẩn $N(\mu_1; \sigma_1^2)$. X_2 có phân bố chuẩn $N(\mu_2; \sigma_2^2)$. H$_0$: $\sigma_1^2 = \sigma_2^2$

Tiêu chuẩn kiểm định: $T = \dfrac{S_1^2}{S_2^2}$; $(S_1^2 > S_2^2)$ có phân bố Fisher-Snedecor với $n_1 - 1$ và $n_2 - 1$ bậc tự do.

a) H$_1$: $\sigma_1^2 \neq \sigma_2^2$: $W_\alpha = \left\{ T < f_{1-\frac{\alpha}{2}}(n_1 - 1, n_2 - 1) \text{ hoặc } T > f_{\alpha/2}(n_1 - 1, n_2 - 1) \right\}$.

b) H₁: $\sigma_1^2 > \sigma_2^2$: $W_\alpha = \left\{ T = \dfrac{S_1^2}{S_2^2} \;;\; T > f_\alpha(n_1 - 1, n_2 - 1) \right\}$.

Kiểm định giả thiết về sự bằng nhau của hai kỳ vọng của hai biến ngẫu nhiên có phân bố $A(p)$

Biến ngẫu nhiên X_1 của tổng thể thứ nhất có phân bố không – một $A(p_1)$. biến ngẫu nhiên X_2 của tổng thể thứ hai có phân bố không – một $A(p_2)$. Kiểm định H_0: $p_1 = p_2$.

Tiêu chuẩn kiểm định: $T = \dfrac{f_1 - f_2}{\sqrt{\overline{f}(1-\overline{f})\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$.

$$\overline{f} = \frac{n_1 f_1 + n_2 f_2}{n_1 + n_2}.\; n_1 > 30 \text{ và } n_2 > 30.$$

f_1, f_2 là tần suất mẫu, tương ứng với kích thước mẫu n_1, n_2.

a) H_0: $p_1 = p_2$; H_1: $p_1 \neq p_2$: $W_\alpha = \left\{ |T| > U_{\alpha/2} \right\}$.

b) H_0: $p_1 = p_2$; H_1: $p_1 > p_2$: $W_\alpha = \left\{ T > U_\alpha \right\}$.

c) H_0: $p_1 = p_2$; H_1: $p_1 < p_2$: $W_\alpha = \left\{ -T > U_\alpha \right\}$.

Kiểm định giả thiết về quy luật phân bố xác suất của biến ngẫu nhiên

Giả thiết H_0: X có phân bố xác suất theo quy luật A;

Đối thiết H_1: X có phân bố xác suất không theo quy luật A.

Phân hoạch miền giá trị của X thành k khoảng rời nhau $C_1, C_2, \ldots, C_k$. Từ tổng thể rút ra mẫu ngẫu nhiên kích thước n và gọi $n_1, n_2, \ldots, n_k$ là các tần số tương ứng của mẫu nằm trong k khoảng này (gọi là tần số thực nghiệm). Nếu giả thiết H_0 đúng ta có thể tính được các xác suất: $p_i = P\{X \in C_i\}$; $i = \overline{1,k}$. Tần số lý thuyết: $\hat{n}_i = np_i$; $i = \overline{1,k}$.

Tiêu chuẩn kiểm định: $T = \sum\limits_{i=1}^{k} \dfrac{(n_i - \hat{n}_i)^2}{\hat{n}_i}$ có phân bố χ^2 với $k - r - 1$ bậc tự do, trong đó r là số tham số cần ước lượng của quy luật cần kiểm định.

Miền bác bỏ $W_\alpha = \left\{ T = \sum\limits_{i=1}^{k} \dfrac{(n_i - \hat{n}_i)^2}{\hat{n}_i} \;;\; T > \chi_\alpha(k - r - 1) \right\}$.

Kiểm định giả thiết về tính độc lập của hai dấu hiệu định tính

Giả sử cần nghiên cứu đồng thời hai dấu hiệu định tính A và B. Ta chia dấu hiệu A thành h mức độ $A_1, A_2, \ldots, A_h$ và dấu hiệu B thành k mức độ $B_1, B_2, \ldots, B_k$.

Ta kiểm định giả thiết H_0: A và B độc lập; Đối thiết H_1: A và B phụ thuộc.

n là kích thức mẫu; n_{i*}, n_{*j} là tổng số các tần số ứng với dấu hiệu thành phần A_i và B_j;

n_{ij} là tần số ứng với các phần tử đồng thời mang dấu hiệu A_i và B_j.

Tiêu chuẩn kiểm định $T = n\left(\sum_{i=1}^{h}\sum_{j=1}^{k}\frac{n_{ij}^2}{n_{i*}\cdot n_{*j}} - 1\right)$.

Miền bác bỏ $W_\alpha = \left\{T > \chi_\alpha^2(h-1)(k-1)\right\}$.

Kiểm định giả thiết về tần số của nhiều tổng thể

Ta xét k tổng thể $\mathscr{H}_1, \mathscr{H}_2, \ldots, \mathscr{H}_k$. Mỗi cá thể của chúng có thể mang hay không mang dấu hiệu A . Gọi p_i là tần suất có dấu hiệu A của tổng thể $\mathscr{H}_i$, $(i = \overline{1,k})$. Ta kiểm định giả thiết H_0: $p_1 = p_2 = \ldots = p_k$. H_1: Các tần suất p_i $(i = \overline{1,k})$ không đồng thời bằng nhau. Từ mỗi tổng thể $\mathscr{H}_i$ rút ra một mẫu ngẫu nhiên có kích thước n_i trong đó có m_i cá thể có dấu hiệu A và $l_i = n_i - m_i$ cá thể không có dấu hiệu A. $m = \sum_{i=1}^{k} m_i$; $l = \sum_{i=1}^{k} l_i$; $N = m + l = \sum_{i=1}^{k} n_i$.

Tiêu chuẩn kiểm định: $T = \frac{N^2}{ml}\sum_{i=1}^{k}\frac{m_i^2}{n_i} - N\frac{m}{l}$. Miền bác bỏ $W_\alpha = \left\{T > \chi_\alpha^2(k-1)\right\}$.

CÂU HỎI ÔN TẬP VÀ BÀI TẬP

8.1 Giả thiết thống kê là giả thiết do nhà thống kê đặt ra cho mẫu ngẫu nhiên.

Đúng ☐ Sai ☐

8.2 Bác bỏ giả thiết dẫn đến chấp nhận đối thiết và ngược lại do đó đối thiết là phủ định của giả thiết.

Đúng ☐ Sai ☐

8.3 Qui tắc kiểm định dựa trên nguyên lý xác suất nhỏ và phép chứng minh phản chứng.

Đúng ☐ Sai ☐

8.4 Sai lầm loại 1 là sai lầm gặp phải khi thực tế giả thiết đúng nhưng ta bác bỏ.

Đúng ☐ Sai ☐

8.5 Sai lầm loại 2 luôn luôn lớn hơn sai lầm loại 1.

Đúng ☐ Sai ☐

8.6 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 ☐

8.7 Miền bác bỏ là miền có xác suất rất bé nên ta có thể bỏ qua trong mọi phép kiểm định.

Đúng ☐ Sai ☐

8.8 Miền bác bỏ của tiêu chuẩn kiểm định giả thiết hai phía luôn luôn là miền đối xứng.

Đúng ☐ Sai ☐

8.9 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 ☐

8.10 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 ☐

8.11 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.

8.12 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:

X (phút)	10 - 12	12 - 14	14 - 16	16 - 18	18 - 20
Số công nhân	20	60	100	40	30

Với mức ý nghĩa $\alpha = 0,05$ hãy kết luận về ý định nói trên.

8.13 Mức hao phí xăng của một loại Ôtô chạy từ A đến B là một biến ngẫu nhiên có quy luật 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

X hao phí (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.

8.14 Trọng lượng X của sản phẩm do một nhà máy sản xuất ra là một biến ngẫu nhiên có phân bố chuẩn $N(\mu,\sigma^2)$ 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:

X (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 $\alpha = 0,05$.

8.15 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?

8.16 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 $\alpha = 0,05$.

8.17 Nếu máy móc hoạt động bình thường thì trọng lượng sản phẩm tuân theo quy luật 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 điều chỉnh mẫu tăng lên tới 1,1 kg. Kiểm định giả thiết trên với mức ý nghĩa $\alpha = 0,01$.

8.18 Có hai loại bi bằng thép: loại I và loại II. Loại I có đường kính là biến ngẫu nhiên X tuân theo quy luật phân bố chuẩn $N(\mu_1;\sigma^2)$, Loại II có đường kính là biến ngẫu nhiên Y tuân theo quy luật phân bố chuẩn $N(\mu_2;\sigma^2)$. 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 $\alpha = 0,05$.

8.19 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 $\bar{x}$	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?

8.20 Sau khi theo dõi thời gian hoàn thành sản phẩm của hai công nhân A và B ta có kết quả sau:

Kết quả	n số sản phẩm	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 theo quy luật chuẩn.

8.21 Điều tra ngẫu nhiên phân bố thực nghiệm của 4000 gia đình có ba con, ta có bảng sau:

Số con trai	0	1	2	3
Số gia đình	450	1460	1530	560

Với mức ý nghĩa $\alpha = 0{,}05$ có thể xem X là biến ngẫu nhiên chỉ số con trai trong gia đình ba con tuân theo luật phân bố nhị thức không?

8.22 Số con (X) của 2000 phụ nữ thủ đô dưới 25 tuổi cho ở bảng sau

X : Số con	0	1	2	3	4
Số phụ nữ	1090	650	220	30	10

Với mức ý nghĩa $\alpha = 0{,}05$ có thể xem X tuân theo quy luật Poisson được không?

8.23 Một máy tự động làm ra sản phẩm là một loại trục có đường kính (X) quan sát được trên 200 trục ta nhận được số liệu sau

Giá trị X	0,3	0,5	0,7	0,9	1,1	1,3	1,5	1,7	1,9	2.1	2.3
Số lần n_i	6	9	26	25	30	26	21	24	20	8	5

Với mức ý nghĩa $\alpha = 0{,}05$ hãy kiểm định giả thiết H_0: “X có phân bố chuẩn”, đối thiết H_1: “X không có phân bố chuẩn”.

8.24 Nghiên cứu ảnh hưởng của thành phần thức ăn của bố mẹ (X) đối với giới tính (Y) của con cái ta có kết quả sau

X -Thành phần thức ăn / Y -Giới tính con	Không có Vitamin	Có Vitamin	Tổng cộng
Nam	123	145	268
Nữ	153	150	303
Tổng cộng	276	295	571

Với mức ý nghĩa $\alpha = 0{,}05$ có thể xem X và Y độc lập với nhau được không?

HƯỚNG DẪN GIẢI VÀ ĐÁP ÁN BÀI TẬP

ĐÁP ÁN CHƯƠNG I

1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	1.10
Đúng	Sai	Đúng	Đúng	Sai	Đúng	Sai	Sai	Đúng	Đúng

1.11 a) $P = 0{,}246$ b) $P = 0{,}495$.

1.12 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 $N = A_6^3 = 216$. Gọi A là biến cố tất cả cùng ra ở tầng bốn, biến cố này chỉ có 1 trường hợp thuận lợi. Do đó $P(A) = \frac{1}{216}$.

Lý luận tương tự trên ta có $P_b = \frac{6}{216} = \frac{1}{36}$; $P_c = \frac{A_6^5}{216} = \frac{5}{9}$.

1.13 $P = \frac{1}{720}$.

1.15 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. $A = A_1\overline{A}_2 + \overline{A}_1 A_2$. Sử dụng qui tắc cộng xác suất trường hợp xung khắc và qui tắc nhân trường hợp độc lập ta có:

$$P(A) = P(A_1\overline{A}_2) + P(\overline{A}_1 A_2) = P(A_1)P(\overline{A}_2) + P(\overline{A}_1)P(A_2) = 0{,}8 \cdot 0{,}1 + 0{,}2 \cdot 0{,}9 = 0{,}26$$

Tương tự ta có: $P_b = 0{,}98$; $P_c = 0{,}02$.

1.16 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: $A = A_1 + A_2$

Vì A_1, A_2 xung khắc do đó:

$$P(A) = P(A_1 + A_2) = P(A_1) + P(A_2) = 0{,}4 + 0{,}5 = 0{,}9$$

1.17 $P = 1 - P_0(3; 0,6) = 1 - 0{,}6^3 = 0{,}784$.

1.18 $P = C_{30}^{5}\left(\frac{1}{50}\right)^{5}\left(\frac{49}{50}\right)^{25} = 0,00027$.

1.19 Gọi A_i là biến cố sản phẩm đã qua kiểm tra chất lượng ở phòng thứ i, i=1,2,3.

Gọi B là biến cố phế phẩm được nhập kho.

$P(B) = P(\overline{A_1})P(\overline{A_2})P(\overline{A_3}) = (1-0,8)(1-0,9)(1-0,99) = 0,0002$.

1.20 $P = 0,11$.

1.21 Gọi A_i là biến cố lần thứ i lấy ra 3 sản phẩm mới để kiểm tra, $(i=\overline{1,3})$. Gọi A là biến cố sau 3 lần kiểm tra tất cả các sản phẩm đều được kiểm tra $A = A_1A_2A_3$. Vì các biến cố phụ thuộc nên $P(A) = P(A_1)P(A_2|A_1)P(A_3|A_1A_2) = 1\cdot\frac{5}{21}\cdot\frac{1}{84} = \frac{5}{1764}$.

1.22 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. $P(B_1|A) = \frac{P(B_1)P(A|B_1)}{P(A)} = \frac{0,36\times 0,12}{0,1012} = 0,427$

$$P(B_2|A) = \frac{P(B_2)P(A|B_2)}{P(A)} = \frac{0,34\times 0,10}{0,1012} = 0,336$$

$$P(B_3|A) = \frac{P(B_3)P(A|B_3)}{P(A)} = \frac{0,30\times 0,08}{0,1012} = 0,237$$

1.23 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: $P(B_1) = \frac{5}{18},\ P(B_2) = \frac{7}{18},\ P(B_3) = \frac{4}{18},\ P(B_4) = \frac{2}{18}$

$P(A|B_1) = 0,2,\ P(A|B_2) = 0,3,\ P(A|B_3) = 0,4,\ P(A|B_4) = 0,5$.

$$P(A)=P(B_1)P(A|B_1)+P(B_2)P(A|B_2)+P(B_3)P(A|B_3)+P(B_4)P(A|B_4)$$
$$=\frac{5}{18}\times 0,2+\frac{7}{18}\times 0,3+\frac{4}{18}\times 0,4+\frac{2}{18}\times 0,5=\frac{57}{180}$$

Áp dụng công thức Bayer, ta thu được:

$$P(B_1|A)=\frac{P(B_1)P(A|B_1)}{P(A)}=\frac{\frac{5}{18}\times 0,2}{\frac{57}{180}}=\frac{10}{57},$$

$$P(B_2|A)=\frac{21}{57},\ P(B_3|A)=\frac{16}{57},\ P(B_4|A)=\frac{10}{57}.$$

Vậy xạ thủ có khả năng ở nhóm thứ hai nhất.

1.24 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:

$$P(A)=P(B_1\overline{B_2}\cup B_2\overline{B_1})=P(B_1\overline{B_2})+P(B_2\overline{B_1})$$
$$=0,7\times 0,6+0,4\times 0,3=0,54$$

b. $$P(B_1|A)=\frac{P(B_1A)}{P(A)}=\frac{P\left(B_1\left[B_1\overline{B_2}\cup B_2\overline{B_1}\right]\right)}{P(A)}=\frac{P(B_1\overline{B_2})}{P(A)}=\frac{0,7\times 0,6}{0,54}=0,778$$

1.25 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.

$$P(A|H_A)=1-\frac{1}{7}=\frac{6}{7};\ P(A|H_B)=\frac{1}{8}$$

a. $P(A)=P(H_A)P(A|H_A)+P(H_B)P(A|H_B)=0,85\times\frac{6}{7}+0,15\times\frac{1}{8}=0,747$

b. $P(H_A|A)=\dfrac{P(H_A)P(A|H_A)}{P(A)}=\dfrac{0,85\times\frac{6}{7}}{0,747}=0,975$

1.26 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

$P(A|B_T)=0,9;\ P(\overline{A}|B_H)=0,95 \Rightarrow P(\overline{A}|B_T)=0,1;\ P(A|B_H)=0,05$.

a) $P(A)=P(B_T)P(A|B_T)+P(B_H)P(A|B_H)=0,85\times0,9+0,15\times0,05=0,7725$

b) $P(B_H|A)=\dfrac{P(B_H)P(A|B_H)}{P(A)}=\dfrac{0,15\times0,05}{0,7725}=0,0097$

c) $P(AB_T\cup\overline{A}B_H)=P(AB_T)+P(\overline{A}B_H)=0,85\times0,9+0,15\times0,95=0,9075$.

ĐÁP ÁN CHƯƠNG II

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 $E(X)=0,3;\ D(X)=15,21$.

2.12 $x_3=29,1;\ p_3=0,2$.

2.13 $\mathrm{E}(X_1)=3,1;\ \mathrm{E}(X_2)=3,4;\ \mathrm{D}(X_1)=1,09;\ \mathrm{D}(X_2)=1,44$.

$\mathrm{E}(X_1+X_2)=6,5;\ \mathrm{D}(X_1+X_2)=2,53$.

2.14 $\mathrm{E}(\overline{X})=0,8;\ \mathrm{D}(\overline{X})=0,12$.

2.15 a) $\mathrm{D}(Z)=61$. b) $\mathrm{D}(Z)=41$.

2.16 $p_1=0,4;\ p_2=0,1;\ p_3=0,5$.

2.17 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: $A=A_1A_2A_3 \Rightarrow \overline{A}=\overline{A_1}+\overline{A_2}+\overline{A_3}$.

$$P(\overline{A})=P(\overline{A}_1)+P(\overline{A}_2)+P(\overline{A}_3)-P(\overline{A}_1\overline{A}_2)-P(\overline{A}_1\overline{A}_3)-P(\overline{A}_2\overline{A}_3)+P(\overline{A}_1\overline{A}_2\overline{A}_3)=\frac{93}{243}$$

$$\Rightarrow\ P(A)=\frac{150}{243}.$$

b)

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	$\frac{1}{243}$	$\frac{10}{243}$	$\frac{40}{243}$	$\frac{80}{243}$	$\frac{80}{243}$	$\frac{32}{243}$

2.18 $\mathrm{E}(X)=0$.

2.19 a) $\int\limits_{-\infty}^{\infty} f(x)dx = k\int\limits_0^1 xdx + 3k = 1 \Rightarrow k=\frac{2}{7}$

$$F(x)=\begin{cases} 0 & \text{nếu } x\le 0 \\ \dfrac{x^2}{7} & \text{nếu } 0<x\le 1 \\ \dfrac{1}{7}+\dfrac{2}{7}(x-1) & \text{nếu } 1<x\le 4 \\ 1 & \text{nếu } 4<x \end{cases}$$

b. $EX=\frac{2}{7}\int\limits_0^1 x^2dx+\frac{2}{7}\int\limits_1^4 xdx=\frac{47}{21}=2,238$

$EX^2=\frac{2}{7}\int\limits_0^1 x^3dx+\frac{2}{7}\int\limits_1^4 x^2dx=\frac{85}{14} \quad \Rightarrow DX=EX^2-(EX)^2=\frac{937}{882}=1,062$.

2.20 a) Vì $\int\limits_0^4 x^2(4-x)dx=\frac{64}{3} \Rightarrow k=\frac{3}{64}$.

b) $P\{X<1\}=\int\limits_0^1 \frac{3}{64}x^2(4-x)dx=\frac{13}{256}$.

c) $\mathrm{E}X=\int\limits_0^4 \frac{3}{64}x^3(4-x)dx=\frac{3}{64}\left(x^4-\frac{x^5}{5}\right)\Bigg|_{x=0}^{x=4}=3\left(4-\frac{16}{5}\right)=\frac{12}{5}$

$$\mathrm{E}X^2 = \int_0^4 \frac{3}{64} x^4 (4-x)\,dx = \frac{3}{64}\left(\frac{4x^5}{5} - \frac{x^6}{6}\right)\Bigg|_{x=0}^{x=4} = 3\times 64\left(\frac{1}{5}-\frac{1}{6}\right) = \frac{32}{5}$$

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

2.21 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) $P\{Y=0\} = 0,37$

$$P\{Y=1\} = P\{X=1\} + P\{X=-1\} = 0,5$$

$$P\{Y=2\} = P\{X=2\} + P\{X=-2\} = 0,13$$

$$\mathrm{E}Y = 0\times 0,37 + 1\times 0,5 + 2\times 0,13 = 0,76.$$

2.22 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 ôtô bị hỏng trong thời gian làm việc

Từ đó $P\{X=0\}=P\left(\overline{A_1}\right)P\left(\overline{A_2}\right)=0,9\times 0,8=0,72$

$P\{X=1\}=P(A_1)P\left(\overline{A_2}\right)+P\left(\overline{A_1}\right)P(A_2)=0,1\times 0,8+0,9\times 0,2=0,26$

$P\{X=2\}=P(A_1)P(A_2)=0,02$

X	0	1	2
P	0,72	0,26	0,02

$\mathrm{E}X=0\times 0,72+1\times 0,26+2\times 0,02=0,3$.

$\mathrm{E}X^2=0^2\times 0,72+1^2\times 0,26+2^2\times 0,02=0,34$.

$\mathrm{D}X=\mathrm{E}X^2-(\mathrm{E}X)^2=0,34-(0,3)^2=0,25$.

2.23 a) $1=\int\limits_{-\infty}^{\infty} f(x)dx=k\int\limits_0^3 x^2dx=9k \Rightarrow k=\frac{1}{9}$

b) $P\{X>2\}=\frac{1}{9}\int\limits_2^3 x^2dx=\frac{19}{27}$.

c) Hàm phân bố $F(x)=\begin{cases}0 & \text{nếu } x<0\\ \dfrac{x^3}{27} & \text{nếu } 0\le x\le 3\\ 1 & \text{nếu } x>3\end{cases}$

d) $F(\alpha)=\frac{3}{4} \Rightarrow \frac{\alpha^3}{27}=\frac{3}{4} \Rightarrow \alpha=2,726$.

2.24 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) $P\{X\ge 5\}=\frac{1}{10}+\frac{1}{10}=\frac{2}{10}$; $P\{X<3\}=\frac{3}{10}$.

c) $\mathrm{E}X=\frac{1}{10}+\frac{4}{10}+\frac{6}{10}+\frac{12}{10}+\frac{5}{100}+\frac{12}{100}+7\left(\frac{7}{100}+\frac{1}{10}\right)=3,66$.

d) $\mathrm{E}X^2=\frac{1}{10}+\frac{8}{10}+\frac{18}{10}+\frac{48}{10}+\frac{25}{100}+\frac{72}{100}+49\left(\frac{7}{100}+\frac{1}{10}\right)=16,8$.

$\mathrm{D}X = \mathrm{E}X^2 - (\mathrm{E}X)^2 = 16{,}8 - (3{,}66)^2 = 3{,}404$.

2.25 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 Ị là "số chính phẩm gặp phải" $\Rightarrow Y = 2 - X$:

Y	1	2
P	0,4	0,6

$\mathrm{E}Y = \mathrm{E}(2 - X) = 2 - \dfrac{2}{5} = \dfrac{8}{5}$; $\mathrm{D}Y = \dfrac{6}{25}$.

2.26 Gọi X là "số nữ có trong nhóm được chọn"

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

$$P\{X=2\} = \frac{C_4^2 C_6^1}{C_{10}^3} = \frac{9}{30}, \quad P\{X=1\} = \frac{C_4^1 C_6^2}{C_{10}^3} = \frac{15}{30}$$

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

$$\mathrm{E}X = 0\times\frac{5}{30} + 1\times\frac{15}{30} + 2\times\frac{9}{30} + 3\times\frac{1}{30} = \frac{36}{30} = \frac{6}{5}.$$

2.27 a) $1 = \displaystyle\int_{-\infty}^{\infty} f(x)dx = k\int_0^1 x^2(1-x)dx = \frac{k}{12} \Rightarrow k = 12$.

$$P\{0{,}4 < X < 0{,}6\} = 12\int_{0{,}4}^{0{,}6} x^2(1-x)dx = 0{,}296.$$

b) $\mathrm{E}X = 12\displaystyle\int_0^1 x^3(1-x)dx = \frac{3}{5}$.

2.28 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

$$Y = \begin{cases} 1650\cdot 50 - (X-1650)\cdot 20 & \text{nếu} \quad X > 1650 \\ X\cdot 50 & \text{nếu} \quad X \le 1650 \end{cases}$$

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.

ĐÁP ÁN CHƯƠNG III

3.1	3.2	3.3	3.4	3.5	3.6	3.7	3.8	3.9	3.10
Đúng	Sai	Sai	Sai	Sai	Đúng	Đúng	Đúng	Sai	Sai

3.11 Thắng 2 trong 4 ván dễ hơn.

3.12 a) $P = 0{,}238$. b) $P = 0{,}751$.

3.13 a) X tuân theo quy luật nhị thức $\mathscr{B}(n;p)$ với $n=5$ và $p=0{,}8$.

b) $\mathrm{E}X = 4$; $\mathrm{D}X = 0{,}8$. c) $\mathrm{Mod}X = 4$; $P\{X=4\} = 0{,}4096$.

3.14 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.

3.15 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 $P = \sum_{k=0}^{3} C_{10}^4\left(\frac{1}{5}\right)^k\left(\frac{4}{5}\right)^{10-k} = 0{,}879$.

3.16 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 $P\{X \ge 1\} = 1 - P\{X = 0\} = 1 - 0,002 = 0,998$

3.17 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 $P\{X \ge 1\} = 1 - P\{X = 0\} = 0,9996$

3.18 Không đúng; $P = 0,41$.

3.19 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

$$P\{Z \le 1\}^3 = 0,406^3 = 0,0067.$$

3.20 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

$$P\{X \ge 4\} = 1 - P\{X \le 3\} = 1 - 0,934 = 0,066.$$

3.21 $P = 0,9$.

3.22 $P\{8 < X < 12\} = \Phi\left(\frac{12-10}{2}\right) - \Phi\left(\frac{8-10}{2}\right) = 0,6826$.

3.23 $P = 0,3$.

3.24 a) 95,44%; b) 4,56%.

3.25 a) 20,33%; b) $P = 0{,}9983$.

3.26 $E(\overline{X}) = \mu$; $D(\overline{X}) = \dfrac{\sigma^2}{n}$ $\Rightarrow$ $P\{|\overline{X} - \mu| < \varepsilon\} = 2\Phi\left(\dfrac{\varepsilon\sqrt{n}}{\sigma}\right) - 1$.

ĐÁP ÁN CHƯƠNG IV

4.1	4.2	4.3	4.4	4.5	4.6	4.7	4.8	4.9	4.10
Sai	Đúng	Đúng	Đúng	Sai	Sai	Sai	Đúng	Đúng	Sai

4.11

X	x_1	x_2
P	0,56	0,44

Y	y_1	y_2	y_3
P	0,26	0,38	0,36

4.12 $EX = -7/15$; $EY = 0$;

$\text{cov}(X,Y) = -1/8$; $\rho_{X,Y} = -0{,}15$.

4.13 $EX = -1/5$; $EY = 0$; $\rho_{X,Y} = 0$. X và Y không độc lập vì

$P\{X = 1\} = 2/15$, $P\{Y = 1\} = 5/15$ và $P\{X = 1.\ Y = 1\} = 0$.

4.14 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

$EX = 1{,}7$; $EY = 1{,}7$; $EZ = 2{,}89$.

4.15 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

$P\{X > Y\} = 0{,}19$.

4.16 X, Y không độc lập vì

$P\{X=1\}=0{,}5,\ P\{Y=1\}=0{,}45$ và $P\{X=1.\,Y=1\}=0{,}15\neq 0{,}5\cdot 0{,}45$

$P\{X=1|Y=2\}=7/11$.

4.17

$Y \backslash X$	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

$P\{X=1\}=0{,}5,\ P\{Y=1\}=0{,}45$ và $P\{X=1.\,Y=1\}=0{,}15\neq 0{,}5\cdot 0{,}45$.

4.18

$Y\|X=26$	23	27
P	0,357	0,643

$X\|Y=27$	26	30	41	50
P	0,1268	0,4225	0,1549	0,2958

4.19 $\mathrm{E}[Y|X=1]=5$. $\mathrm{E}X=2{,}93$; $\mathrm{E}Y=4{,}5$; $\mathrm{D}X=4{,}83$; $\mathrm{D}Y=2{,}25$.

4.20 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 $P\{X=1,Y=1\}=0$.

ĐÁP ÁN CHƯƠNG V

5.1	5.2	5.3	5.4	5.5	5.6
Đúng	Đúng	Sai	Sai	Đúng	Đúng

5.7 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

$$P\left\{\left|X-0,05\right|<2\right\}\geq 1-\frac{0,475}{2^2}=0,88\,;\; P\left\{\left|X-0,05\right|\geq 2\right\}\leq\frac{0,475}{2^2}=0,12\,.$$

5.8 Đặt $S=\sum_{n=1}^{12}X_n$; $\mathrm{E}S=12\cdot 16=192$, $\mathrm{D}S=12$. Theo bất đẳng thức Trêbusép

$$P\left\{\left|S-192\right|\leq\varepsilon\right\}\geq 1-\frac{\mathrm{D}S}{\varepsilon^2}\geq 0,99.$$ Chọn $a=157,36$; $b=226,64$.

5.9 Đặt $S=\sum_{n=1}^{10000}X_n$; $\mathrm{E}S=0$, $\mathrm{D}S=\frac{10000}{12}$. Theo bất đẳng thức Trêbusép

$$P\left\{\left|S\right|\geq 500\right\}\leq\frac{\mathrm{D}S}{500^2}=\frac{1}{300}\,.$$

5.10 Ta biết rằng S là biến ngẫu nhiên có phân bố nhị thức tham số $p=\frac{1}{6}$. $\mathrm{E}S=\frac{n}{6}$ và $\mathrm{D}S=\frac{5n}{36}$. Theo bất đẳng thức Trêbusép

$$P\left\{\left|S-\mathrm{E}S\right|<\sqrt{n}\right\}\geq 1-\frac{\mathrm{D}S}{n}=1-\frac{5}{36}=\frac{31}{36}\Leftrightarrow P\left\{\frac{n}{6}-\sqrt{n}<S<\frac{n}{6}+\sqrt{n}\right\}\geq\frac{31}{36}.$$

5.11 Đặt $S=\sum_{n=1}^{12}X_n$. Ta cần tìm M nhỏ nhất để $P\left\{\sum_{n=1}^{12}X_n\leq M\right\}\geq 0,99$.

Ta có $\mathrm{E}S=192$, $\mathrm{D}S=12$. Theo bất đẳng thức Trêbusép

$$P\left\{\left|S-192\right|\leq\varepsilon\right\}\geq 1-\frac{\mathrm{D}S}{\varepsilon^2}\geq 0,99\Rightarrow\varepsilon=34,64.$$ Vậy M = 192+34,64 = 226,64.

5.12 Thỏa mãn luật số lớn Trêbusép.

5.13 Thỏa mãn luật số lớn Trêbusép.

5.14 Thỏa mãn luật số lớn Trêbusép.

5.15 Áp dụng bất đẳng thức Trêbusép tính được xác suất $P\geq 0,9131$

5.16 Áp dụng bất đẳng thức Trêbusép cần kiểm tra 23.750 chi tiết.

ĐÁP ÁN CHƯƠNG VI

6.1	6.2	6.3	6.4	6.5	6.6	6.7	6.8	6.9	6.10
Đúng	Đúng	Sai	Đúng	Đúng	Đúng	Đúng	Đúng	Đúng	Sai

6.13 Đặt $u_i = x_i - 26 \Rightarrow \overline{x} = \overline{u} + 26 = 26$;

$$s^2 = \frac{1}{99}\left(1080 - \frac{0^2}{100}\right) = 10,909$$

6.14 $\overline{x} = 8,9$; $s = 2,18$

6.15 $\overline{x} = 1111,1$ (giờ); $s = 37$.

6.16 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

$$P\left\{\sum_{i=1}^{10} X_i = 5\right\} = P_{10}(5) = C_{10}^5 (0,5)^5 \cdot (0,5)^{10-5} = C_{10}^5 (0,5)^{10}.$$

6.17 Ta có $\mathrm{E}X = p$; $\mathrm{D}X = p(1-p)$. Do đó $\mathrm{E}\overline{X} = p$; $\mathrm{D}\overline{X} = \frac{p(1-p)}{10}$

6.18 X có phân bố chuẩn $N(\mu;\sigma^2)$ nên $\overline{X}$ có phân bố chuẩn $N(\mu;\frac{\sigma^2}{n})$. 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

$$P\left\{\left|\overline{X} - 20\right| < 0,2\right\} = 2\Phi\left(\frac{0,2\sqrt{100}}{1}\right) = 2\Phi(2) = 0,9545.$$

6.19 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

$$F_{10}(x)=\begin{cases}0 & x\le 1\\ 1/5 & 1<x\le 2\\ 3/5 & 2<x\le 3\\ 4/5 & 3<x\le 4\\ 1 & x>4\end{cases} \qquad \overline{x}=6{,}8;\ s^2=1{,}15,\ \ s=1{,}072.$$

6.20 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.

ĐÁP ÁN CHƯƠNG VII

7.1	7.2	7.3	7.4	7.5	7.6	7.7	7.8	7.9	7.10
Đúng	Sai	Sai	Đúng	Đúng	Sai	Đúng	Sai	Đúng	Đúng

7.13 $\overline{x}=38$; $\dfrac{\sigma}{\sqrt{n}}u_\beta=\dfrac{6{,}4}{\sqrt{50}}.2{,}575=2{,}331$. Khoảng tin cậy $[35{,}669;40{,}331]$.

7.14 $f=\dfrac{1082}{2000}$; Điều kiện $\begin{cases}nf=1082>10\\ n(1-f)=918>10\end{cases}$

$$f-u_\beta\frac{\sqrt{f(1-f)}}{\sqrt{n}}=\frac{1082}{2000}-2{,}33\frac{\sqrt{918\times 1082}}{\sqrt{2000}}=0{,}515$$

Vậy tối thiểu có 51,5% số phiếu bầu cho ứng cử viên A.

7.15 $\overline{x}=\dfrac{\sum x_i}{n}=\dfrac{34{,}15}{35}=0{,}976$.

$$s^2=\frac{1}{n-1}\left[\sum x_i{}^2-\frac{\left(\sum x_i\right)^2}{n}\right]=\frac{1}{34}\left[33{,}8943-\frac{(34{,}15)^2}{35}\right]=0{,}01687.$$

$$\Rightarrow s=0{,}1299;\quad u_\beta\frac{s}{\sqrt{n}}=1{,}96\times\frac{0{,}1299}{\sqrt{35}}=0{,}043.$$

Khoảng tin cậy 95%: $[0,933 ; 1,019]$.

7.16 Tần suất mẫu $f=\frac{53}{400}$, điều kiện $\begin{cases} nf=53>10 \\ n(1-f)=347>10 \end{cases}$

Gọi p là xác suất bắt được con cá có đánh dấu, khoảng tin cậy 95% của p:

$$u_\beta \frac{\sqrt{f(1-f)}}{\sqrt{n}}=1,96\frac{\sqrt{53\times 347}}{400\sqrt{400}}=0,0332$$

Khoảng ước lượng $[0,0993 ; 0,1657]$

Mặt khác $p=\frac{2000}{N}$, trong đó N là số cá trong hồ.

Vậy $0,0993<\frac{2000}{N}<0,1657$

$$\Rightarrow \frac{2000}{0,1657}<N<\frac{2000}{0,0993} \Rightarrow 12070<N<20141$$

7.17 Khoảng tin cậy 95% của hao phí nguyên liệu trung bình cho 1 đơn vị sản phẩm là $\left[\overline{x}-u_\beta\frac{\sigma}{\sqrt{n}} ; \overline{x}+u_\beta\frac{\sigma}{\sqrt{n}}\right]$, trong đó $u_\beta=1,96$ là phân vị mức 0,975 của phân bố chuẩn tắc N(0;1).

Theo mẫu ta tính được $\overline{x}=19,87$; $\sigma=0,03$ $\Rightarrow u_\beta\frac{\sigma}{\sqrt{n}}=1,96\frac{0,03}{\sqrt{36}}=0,0098$

Khoảng tin cậy $[19,8602 ; 19,8798]$.

7.18 Đặt $u_i=\frac{x_i-18,25}{5} \Rightarrow \overline{x}=5\frac{\sum u_i}{n}+18,25=5\frac{-1,8}{40}+18,25=18,025$.

$$s^2=\frac{5^2}{n-1}\left[\sum u_i{}^2-\frac{\left(\sum u_i\right)^2}{n}\right]=\frac{25}{39}\left[0,76-\frac{(-1,8)^2}{40}\right]=0,435.$$

$$\Rightarrow s=0,66;\ u_\beta\frac{s}{\sqrt{n}}=1,64\times\frac{0,66}{\sqrt{40}}=0,171.$$

a) Khoảng tin cậy 90%: $[17,854; 18,196]$.

b) Kích thược mẫu cần thiết $n\ge\frac{u^2{}_\beta s^2}{\varepsilon^2}=116,99$ chọn $n=117$

7.19 Đặt $u_i = \dfrac{x_i - 47}{2} \Rightarrow \overline{x} = 2\dfrac{\sum u_i}{n} + 47 = 2\dfrac{-47}{100} + 47 = 46,06$.

$$s^2 = \frac{2^2}{n-1}\left[\sum u_i{}^2 - \frac{\left(\sum u_i\right)^2}{n}\right] = \frac{4}{99}\left[175 - \frac{(-47)^2}{100}\right] = 6,178.$$

$$\Rightarrow s = 2,486;\quad u_\beta \frac{s}{\sqrt{n}} = 1,96 \times \frac{2,486}{\sqrt{100}} = 0,487.$$

a) Khoảng tin cậy 95%: $[45,573;\ 46,547]$.

b) Kích thược mẫu cần thiết $n \geq \dfrac{u^2{}_\beta s^2}{\varepsilon^2} = 148,33$ chọn $n = 149$

7.20 Đặt $u_i = x_i - 50 \Rightarrow \overline{x} = \dfrac{\sum u_i}{n} + 50 = \dfrac{-8}{27} + 50 = 49,704$

$$u_\beta \frac{\sigma}{\sqrt{n}} = 1,96 \times \frac{1}{\sqrt{27}} = 0,377.$$

a) Khoảng tin cậy 95%: $[49,327\ ;\ 50,081]$.

b) Kích thược mẫu cần thiết $n \geq \dfrac{u^2{}_\beta \sigma^2}{\varepsilon^2} = 384,16$ chọn $n = 385$

7.21 Đặt $u_i = \dfrac{x_i - 119}{2} \Rightarrow \overline{x} = 2\dfrac{\sum u_i}{n} + 119 = 2\dfrac{-19}{100} + 119 = 118,62$.

$$s^2 = \frac{2^2}{n-1}\left[\sum u_i{}^2 - \frac{\left(\sum u_i\right)^2}{n}\right] = \frac{4}{99}\left[399 - \frac{(-19)^2}{100}\right] = 15,9752.$$

$$\Rightarrow s = 3,9969;\quad u_\beta \frac{s}{\sqrt{n}} = 1,96 \times \frac{3,9969}{\sqrt{100}} = 0,783.$$

a) Khoảng tin cậy 95%: $[117,837\ ;\ 119,403]$.

b) Kích thược mẫu cần thiết $n \geq \dfrac{u^2{}_\beta s^2}{\varepsilon^2} = 245,48$ chọn $n = 246$.

7.22 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:

$$u_\beta \frac{\sqrt{f(1-f)}}{\sqrt{n}} = 1{,}96\frac{\sqrt{16\times 384}}{400\sqrt{400}} = 0{,}0192$$

Khoảng ước lượng $[0{,}0208\ ;\ 0{,}0592]$

Vậy tỷ lệ phế phẩm tối đa của lô hàng là $5{,}9\%$.

7.23 $\overline{x} = 2\times\frac{-14}{100} + 91 = 90{,}72$; $s^2 = 4\times\frac{1}{99}\left(432 - \frac{14^2}{100}\right) = 17{,}375 \Rightarrow s = 4{,}168$.

$u_\beta \frac{S}{\sqrt{n}} = 0{,}817$. Vậy khoảng tin cậy 95% của giá trung bình của loại hàng hoá trên là $[89{,}903\ ;\ 91{,}537]$.

7.24 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:

$$\left(\frac{25\cdot 0{,}5}{40{,}646}; \frac{25\cdot 0{,}5}{13{,}120}\right) = (0{,}3075\ ;\ 0{,}9520).$$

ĐÁP ÁN CHƯƠNG VIII

8.1	8.2	8.3	8.4	8.5	8.6	8.7	8.8	8.9	8.10
Sai	Sai	Đúng	Đúng	Sai	Đúng	Sai	Đúng	Sai	Sai

8.11 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 $T = \frac{(100-\overline{X})\sqrt{n}}{S}$; Miền bác bỏ $W_\alpha = \{T > 2{,}086\}$.

Đặt $u_i = \frac{x_i - 99{,}25}{5} \quad \Rightarrow \quad \sum r_i u_i = 0{,}4\ ;\ \sum r_i u_i^2 = 0{,}42$

$$\Rightarrow\ \overline{x} = 5\times\frac{0,4}{29} + 99,25 = 99,319\ ;\quad s^2 = 25\times\frac{1}{28}\left[0,42 - \frac{0,4^2}{29}\right] = 0,37 \Rightarrow s = 0,608$$

$$T_{qs} = \frac{(100-99,319)\sqrt{29}}{0,608} = 6,032 \in W_\alpha .$$

Vậy bác bỏ H_0 chấp nhận H_1, nghĩa là sản phẩm bị đóng thiếu.

8.12 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 $T = \dfrac{\left(\overline{X}-14\right)\sqrt{n}}{S}$; Miền bác bỏ $W_\alpha = \left\{|T| > 1,96\right\}$.

Đặt $u_i = \dfrac{x_i - 15}{2} \Rightarrow \sum r_i u_i = 0;\ \sum r_i u_i^2 = 300$

$$\Rightarrow\ \overline{x} = 15\ ;\quad s^2 = 4\times\frac{1}{249}\left[300 - \frac{0}{300}\right] = 4,819 \Rightarrow s = 2,195$$

$$\Rightarrow\ T_{qs} = \frac{(115-14)\sqrt{300}}{2,195} = 7,89 \in W_\alpha .$$

Vậy bác bỏ H_0 chấp nhận H_1, nghĩa là cần thay đổi định mức.

8.13 Gọi μ là mức hao phí xăng trung bình của ôtô 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 $W_\alpha = \left\{T > 2,052\right\}$.

Theo mẫu ta có $\overline{x} = \dfrac{1387,5}{28} = 49,5536$;

$$s^2 = \frac{1}{27}\left(6876375 - \frac{1387,5^2}{28}\right) = \frac{8,1696}{27} = 0,3026 \Rightarrow s = 0,55$$

$$T_{qs} = \frac{(50-49,53)\sqrt{30}}{0,55} = 4,2948 \in W_\alpha .$$

Vậy bác bỏ H_0 chấp nhận H_1, nghĩa là mức hao phí xăng có giảm xuống.

8.14 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 $T = \dfrac{(\bar{X} - 20)\sqrt{n}}{\sigma}$; 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.

8.15 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 $T = \dfrac{(\bar{X} - 1300)\sqrt{n}}{S}$; Miền bác bỏ $W_\alpha = \{T > 1{,}96\}$.

Từ mẫu cụ thể ta có $T = \dfrac{(1378 - 1300)\sqrt{40}}{215} = 2{,}294 > 1{,}96$

Vậy bác bỏ H_0 chấp nhận H_1, nghĩa là hệ thống máy tính mới xử lý tốt hơn.

8.16 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.

8.17 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.

8.18 Ta kiểm định giả thiết H0: $\mu_1 = \mu_2$; đối thiết H1: $\mu_1 \neq \mu_2$.

Tiêu chuẩn kiểm định $T = \dfrac{\overline{X} - \overline{Y}}{\sqrt{\dfrac{S_X^2}{n} + \dfrac{S_Y^2}{m}}} \sim N(0;1)$; 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.

8.19 Ta kiểm định giả thiết H0: $\mu_1 = \mu_2$; đối thiết H1: $\mu_1 > \mu_2$.

Tiêu chuẩn kiểm định $T = \dfrac{\overline{X} - \overline{Y}}{\sqrt{\dfrac{S_X^2}{n} + \dfrac{S_Y^2}{m}}} \sim N(0;1)$; 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.

8.20 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 H0: $\mu_1 = \mu_2$; đối thiết H1: $\mu_1 < \mu_2$.

Tiêu chuẩn kiểm định $T = \dfrac{\overline{X_1} - \overline{X_2}}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}}$; $k = \dfrac{(n_1 - 1)(n_2 - 1)}{(n_2 - 1)C^2 + (n_1 - 1)(1 - C)^2}$;

$$C = \frac{\dfrac{S_1^2}{n_1}}{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}.$$

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.

8.21 Gọi X là biến ngẫu nhiên chỉ số con trai trong gia đình 3 con. Kiểm định giả thiết về luật phân bố xác suất, giả thiết H0: X tuân theo luật nhị thức, đối thiết H1: X không tuân theo luật nhị thức.

Giả sử X là biến ngẫu nhiên nhị thức với $n = 3$ và $p = 0{,}5$. Ta có bảng phân bố lý thuyết

X	0	1	2	3
P	0.125	0.375	0.375	0.125

Từ bảng điều tra 4000 gia đình ta có

X	n_i	p_i	np_i	$(n_i - np_i)^2$	$\frac{(n_i - np_i)^2}{np_i}$
0	450	0,125	500	2500	5
1	1460	0,375	1500	1600	1,06
2	1530	0,375	1500	900	0,6
3	560	0,125	500	3600	7,2
	$n = 4000$				$\sum = 13{,}86$

Tiêu chuẩn kiểm định $T = \sum_{i=0}^{3} \frac{(n_i - np_i)^2}{np_i}$. Miền bác bỏ $W_\alpha = \left\{T > \chi^2_\alpha(k-r-1)\right\}$. Ở đây $k = 4$, $r = 1$; $\alpha = 0{,}05$. Tra bảng ta được $W_\alpha = \{T > 6\}$

Từ mẫu cụ thể ta có $T_{qs} = 13{,}36 \in W_\alpha$. Vậy bác bỏ H_0 chấp nhận H_1.

8.22 Tiêu chuẩn kiểm định $T = \sum_{i=0}^{3} \frac{(n_i - np_i)^2}{np_i}$. Miền bác bỏ $W_\alpha = \left\{T > \chi^2_\alpha(k-r-1)\right\}$. Ở đây $k = 4$, $r = 1$; $\alpha = 0{,}05$. Tra bảng ta được $W_\alpha = \{T > 6\}$

Từ mẫu cụ thể ta có $T_{qs} = 0{,}318 \notin W_\alpha$. Chưa có cơ sở để bác bỏ H_0. Vậy số con của phụ nữ dưới 25 tuổi tuân theo quy luật Poisson.

8.23 Ta kiểm định giả thiết H₀: "X có phân bố chuẩn", đối thiết H₁: "X không có phân bố chuẩn".

Từ mẫu ta có ước lượng $\overline{x} = 1{,}26$, $s = 0{,}490$.

Ta tính xác suất xấp xỉ $p_i \approx \frac{h}{S}\varphi\left(\frac{x_i - \overline{X}}{S}\right)$; $h = x_i - x_{i-1} = 0{,}2$; trong đó $\varphi(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}$ (xem phụ lục I).

Tiêu chuẩn kiểm định $T = \sum_{i=1}^{10} \frac{(n_i - np_i)^2}{np_i}$. Miền bác bỏ $W_\alpha = \left\{T > \chi^2_\alpha(k-r-1)\right\}$. Ở đây $k = 10$, $r = 2$; $\alpha = 0{,}05$. Tra bảng ta được $W_\alpha = \{T > 15{,}507\}$.

Từ mẫu cụ thể ta có $T_{qs} = 12{,}425 \notin W_\alpha$. Chưa có cơ sở để bác bỏ H_0. Vậy có thể xem X tuân theo quy luật phân bố chuẩn.

8.24 Tiêu chuẩn kiểm định $T = n\left(\sum_{i=1}^{h}\sum_{j=1}^{k}\frac{n_{ij}^2}{n_{i*}\cdot n_{*j}} - 1\right)$.

Miền bác bỏ $W_\alpha = \left\{T > \chi_\alpha^2(h-1)(k-1)\right\}$.

$k = 2, h = 2,\ \alpha = 0{,}01 \Rightarrow \chi_\alpha^2(h-1)(k-1) = 6{,}6$.

Từ mẫu cụ thể trên ta có $T_{qs} = 1{,}2 \notin W_\alpha$. Chưa có cơ sở để bác bỏ H_0. Vậy thành phần thức ăn của bố mẹ không ảnh hưởng đến giới tính của con cái.

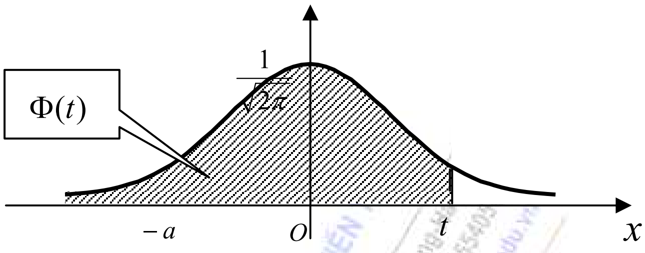
PHỤ LỤC

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

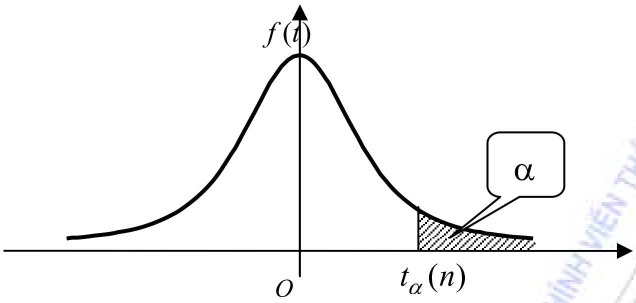
PHỤ LỤC II: GIÁ TRỊ HÀM PHÂN BỐ CHUẨN TẮC

$$\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{x^2}{2}} dx$$



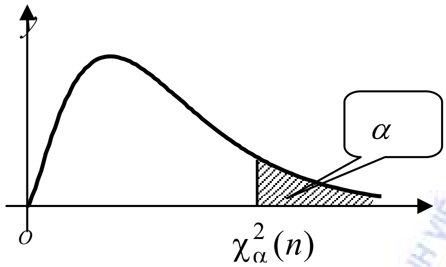
t	0	1	2	3	4	5	6	7	8	9
0,0	0,5000	5040	5080	5120	5160	5199	5239	5279	5319	5359
0,1	5398	5438	5478	5517	5557	5596	5636	5675	5714	5753
0,2	5793	5832	5871	5910	5948	5987	6026	6064	6103	6141
0,3	6179	6217	6255	6293	6331	6368	6406	6443	6480	6517
0,4	6554	6591	6628	6664	6700	6736	6772	6808	6844	6879
0,5	0,6915	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



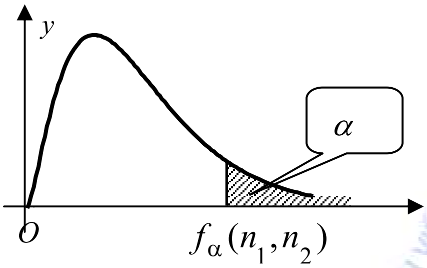
Bậc tự do	α = 0,05	α = 0,025	α = 0,01	α = 0,005	α = 0,001
1	6,314	12,706	31,821	63,657	318,309
2	2,920	4,303	6,965	9,925	22,327
3	2,353	3,128	4,541	5,841	10,215
4	2,132	2,776	3,747	4,604	7,173
5	2,015	2,571	3,365	4,032	5,893
6	1,943	2,447	3,143	3,707	5,208
7	1,895	2,365	2,998	3,499	4,705
8	1,860	2,306	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 χ^2



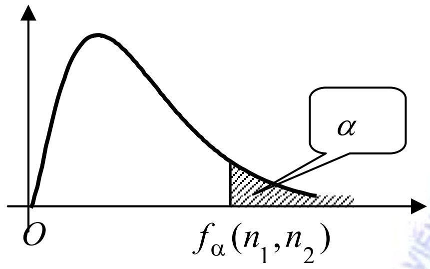
Bậc tự do	$\chi^2_{0,995}$	$\chi^2_{0,99}$	$\chi^2_{0,97}$	$\chi^2_{0,95}$	$\chi^2_{0,05}$	$\chi^2_{0,025}$	$\chi^2_{0,01}$	$\chi^2_{0,005}$
1	0,000	0,000	0,001	0,004	3,841	5,024	6,635	7,879
2	0,010	0,020	0,051	0,103	5,991	7,378	9,210	10,597
3	0,072	0,115	0,216	0,352	7,815	9,348	11,345	12,838
4	0,207	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 Va: GIÁ TRỊ TỚI HẠN CỦA PHÂN BỐ FISHER $F(n_1,n_2)$ MỨC $\alpha = 0{,}05$



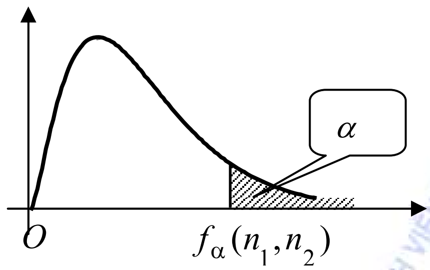
Bậc tự do n_2	Bậc tự do n_1									
	1	2	3	4	5	6	7	8	9	10
1	161	200	216	225	230	234	237	239	241	242
2	18.5	19,0	19,2	19,2	19,3	19,3	19,4	19,4	19,4	19,4
3	10,1	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74
6	5,99	5,74	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,32
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14
10	4,96	4,10	3,71	3,84	3,33	3,22	3,14	3,07	3,02	2,98
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49
17	4,45	3,69	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,81	2,46	2,41
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38
20	4,35	3,19	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32	2,27
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,24
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16
40	4,08	3,23	2,86	2,61	2,45	2,34	2,25	2,18	2,12	2,08
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99
120	3,92	3,07	26,8	2,45	2,29	2,18	2,09	2,02	1,96	1,91
∞	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,88	1,83

PHỤ LỤC Vb: GIÁ TRỊ TỚI HẠN CỦA PHÂN BỐ FISHER $F(n_1,n_2)$ MỨC $\alpha = 0,05$



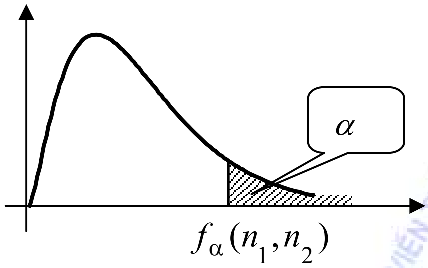
Bậc tự do n_2	Bậc tự do n_1								
	12	15	20	24	30	40	60	120	∞
1	244	246	248	249	250	251	252	3,70	3,67
2	19,4	19,4	19,4	19,5	19,5	19,5	19,5	3,27	3,23
3	8,74	8,70	8,66	8,64	8,62	8,59	8,57	2,97	2,93
4	5,91	5,86	5,80	5,77	5,75	5,72	5,69	2,75	2,71
5	4,68	4,62	4,56	4,53	4,50	4,46	4,43	2,58	2,54
6	4,00	3,94	3,87	3,84	3,81	3,77	3,74	2,45	2,40
7	3,57	3,51	3,44	3,41	3,38	3,34	3,30	2,34	2,30
8	3,28	3,22	3,15	3,12	3,08	3,04	3,01	2,25	2,21
9	3,07	3,01	2,94	2,90	2,86	2,83	2,79	2,18	2,13
10	2,91	2,85	2,77	2,74	2,70	2,66	2,62	2,11	2,07
11	2,79	2,72	2,65	2,61	2,57	2,53	2,49	2,06	2,40
12	2,69	2,62	2,54	2,51	2,47	2,43	2,38	2,01	2,30
13	2,60	2,53	2,46	2,42	2,38	2,34	2,30	1,79	2,21
14	2,53	2,46	2,39	2,35	2,31	2,27	2,22	1,73	2,13
15	2,48	2,40	2,33	2,29	2,25	2,20	2,16	1,90	2,07
16	2,42	2,35	2,28	2,24	2,19	2,15	2,11	1,87	2,01
17	2,38	2,31	2,23	2,19	2,15	2,10	2,06	1,84	1,96
18	2,34	2,27	2,19	2,15	2,11	2,06	2,02	1,81	1,92
19	2,31	2,23	2,16	2,11	2,07	2,03	1,98	1,79	1,88
20	2,28	2,20	2,12	2,08	2,04	1,99	1,95	1,77	1,84
21	2,25	2,18	2,10	2,05	2,01	1,96	1,74	1,68	1,81
22	2,23	2,15	2,07	2,03	1,98	1,94	1,64	1,58	1,78
23	2,20	2,13	2,05	2,01	1,96	1,91	1,53	1,47	1,76
24	2,18	2,11	2,03	1,98	1,94	1,89	1,43	1,35	1,73
25	2,15	2,09	2,01	1,96	1,92	1,78	1,32	1,22	1,71
30	2,09	2,01	1,93	1,89	1,84	1,79	2,53	2,54	1,62
40	2,00	1,92	1,84	1,79	1,74	1,69	29,5	19,5	1,51
60	1,92	1,84	1,75	1,70	1,65	1,59	8,55	8,53	1,39
120	1,83	1,75	1,66	1,61	1,55	1,59	5,66	5,63	1,25
∞	1,75	1,67	1,57	1,52	1,46	1,39	4,40	4,37	1,00

PHỤ LỤC Vc: GIÁ TRỊ TỚI HẠN CỦA PHÂN BỐ FISHER $F(n_1, n_2)$ MỨC $\alpha = 0{,}01$



Bậc tự do n_2	Bậc tự do n_1									
	1	2	3	4	5	6	7	8	9	10
1	4,052	5,000	5,403	5,625	5,764	5,859	5,928	5,982	6,023	6,056
2	98,5	99,0	99,2	99,2	99,3	99,3	99,4	99,4	99,4	99,4
3	34,1	30,8	29,5	28,7	28,2	27,9	27,7	27,5	27,3	27,2
4	21,2	18,0	16,7	16,0	15,5	15,2	15,0	14,8	14,7	14,5
5	16,3	13,3	12,1	11,4	11,0	10,7	10,5	10,3	10,2	10,1
6	13,7	10,9	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87
7	12,2	9,55	8,45	7,85	7,46	7,19	6,99	6,84	6,72	6,62
8	11,3	8,65	7,59	7,01	6,63	6,37	6,18	6,03	5,91	5,81
9	10,6	8,02	6,99	6,42	6,06	5,80	5,61	5,47	5,35	5,26
10	10,0	7,56	6,55	5,99	5,64	5,39	5,20	5,06	4,94	4,85
11	9,65	7,21	6,22	5,67	5,32	5,07	4,89	4,74	4,63	4,54
12	9,33	6,93	5,95	5,41	5,06	4,82	4,64	4,50	4,39	4,30
13	9,07	6,70	5,74	5,21	4,86	4,62	4,44	4,30	4,19	4,10
14	8,86	6,51	5,56	5,04	4,70	4,46	4,28	4,14	4,03	3,94
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80
16	8,53	5,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78	3,69
17	8,40	6,71	5,19	4,67	4,34	4,10	3,93	3,79	3,68	3,59
18	8,29	6,01	5,09	4,58	4,25	4,01	3,84	3,71	3,60	3,51
19	8,19	5,93	5,01	4,50	4,17	3,94	3,77	3,63	3,52	3,43
20	8,10	5,85	4,94	4,43	4,10	3,87	3,70	3,56	3,46	3,37
21	8,02	5,78	4,87	4,37	4,04	3,81	3,64	3,51	3,40	3,31
22	7,95	5,72	4,82	4,31	3,99	3,76	3,59	3,45	3,34	3,26
23	7,88	5,66	4,76	4,26	3,94	3,71	3,54	3,41	3,30	3,21
24	7,82	5,61	4,72	4,22	3,90	3,67	3,50	3,36	3,26	3,17
25	7,77	5,57	4,68	4,78	3,86	3,63	3,46	3,32	3,22	3,13
30	7,56	5,39	4,51	4,02	3,70	3,47	3,30	3,17	3,07	2,98
40	7,31	5,18	4,31	3,83	3,51	3,29	3,12	2,99	2,89	2,80
60	7,08	4,98	4,13	3,65	3,34	3,12	2,95	2,82	2,72	2,63
120	6,85	4,79	3,95	3,48	3,17	2,96	2,79	2,66	2,56	2,47
∞	6,63	4,61	3,78	3,32	3,02	2,80	2,64	2,51	2,41	2,32

PHỤ LỤC Vd: GIÁ TRỊ TỚI HẠN CỦA PHÂN BỐ FISHER $F(n_1, n_2)$ MỨC $\alpha = 0{,}01$



Bậc tự do n_2	Bậc tự do n_1								
	12	15	20	24	30	40	60	120	∞
1	6,106	6,157	6,209	6,235	6,261	6,287	6,313	6,339	6,366
2	99,4	99,4	99,4	99,5	99,5	99,5	99,5	99,5	99,5
3	27,1	26,9	26,7	26,6	26,5	26,4	26,3	26,2	26,1
4	14,4	14,2	14,2	13,9	13,8	13,7	13,7	13,6	13,5
5	9,89	9,72	9,55	9,47	9,38	9,29	9,20	9,11	9,02
6	7,72	7,56	7,40	7,31	7,23	7,14	7,06	6,97	6,88
7	6,47	6,31	6,16	6,07	5,99	5,91	5,82	5,74	5,65
8	5,67	5,52	5,36	5,28	5,20	5,12	5,03	4,95	4,86
9	5,11	4,96	4,81	4,73	4,65	4,57	4,48	4,40	4,31
10	4,71	4,56	4,41	4,33	4,25	4,17	4,08	4,00	3,91
11	4,40	4,25	4,10	4,02	3,94	3,86	3,78	3,69	3,60
12	4,16	4,01	3,86	3,78	3,70	3,62	3,54	3,45	3,36
13	3,96	3,82	3,66	3,59	3,51	3,43	3,34	3,25	3,77
14	3,80	3,66	3,51	3,43	3,35	3,27	3,18	3,09	3,00
15	3,67	3,52	3,37	3,29	3,21	3,13	3,05	2,96	2,87
16	3,55	3,41	3,26	3,18	3,10	3,02	2,93	2,84	2,75
17	3,46	3,31	3,16	3,08	3,00	2,92	2,83	2,75	2,65
18	3,37	3,23	3,08	3,00	2,92	2,84	2,75	2,66	2,57
19	3,30	3,15	3,00	2,92	2,84	2,76	2,67	2,56	2,49
20	3,23	3,09	2,94	2,86	2,78	2,69	2,61	2,52	2,42
21	3,17	3,03	2,88	2,80	2,72	2,64	2,55	2,46	2,36
22	3,12	2,98	2,83	2,75	2,67	2,58	2,50	2,40	2,31
23	3.07	2.93	2.78	2.70	2,62	2,54	2,45	2,35	2,36
24	3,03	2,89	2,74	2,66	2,58	2,49	2,40	2,30	2,21
25	2,99	2,85	2,70	2,62	2,53	2,45	2,36	2,27	2,17
30	2,84	2,70	2,55	2,47	2,39	2,30	2,21	2,11	2,01
40	2,66	2,52	2,37	2,29	2,20	2,11	2,02	1,92	1,80
60	2,50	2,35	2,20	2,12	2,03	1,94	1,84	1,73	1,60
120	2,34	2,10	2,03	1,95	1,86	1,76	1,66	1,53	1,38
∞	2,18	2,04	1,88	1,79	1,70	1,59	1,47	1,32	1,00

TÀI LIỆU THAM KHẢO

[1]. Đào Hữu Hồ, *Xác suất Thống kê,* NXB Đại Học Quốc Gia Hà Nội, 1999.

[2]. Nguyễn Văn Phấn, Lương Hữu Thanh, *Bài tập xác suất và thống kê,* Đại Học Giao Thông Vận Tải, 1996.

[3]. Tống Đình Quỳ, *Hướng dẫn giải bài tập xác suất thống kê*, NXB Đại Học Quốc Gia Hà Nội, 2004.

[4]. Đặng Hùng Thắng, *Mở đầu về lý thuyết xác suất và các ứng dụng,* NXB Giáo dục, 1997.

[5]. Đặng Hùng Thắng, *Bài tập xác suất,* NXB Giáo dục - 1998.

[6]. Đặng Hùng Thắng, *Thống kê và ứng dụng,* NXB Giáo dục,1999.

[7]. Nguyễn Duy Tiến, Vũ Việt Yên, *Lý thuyết xác suất,* NXB Giáo dục, 2000.

[8]. Nguyễn Duy Tiến (và tập thể), *Các mô hình xác suất và ứng dụng*, tập 1, 2, 3. NXB Đại Học Quốc Gia Hà Nội, 2000.

[9]. Trần Mạnh Tuấn, *Xác suất và Thống kê, lý thuyết và thực hành tính toán,* NXB Đại Học Quốc Gia Hà Nội, 2004.

[10]. Nguyễn Bác Văn, *Xác suất và xử lí số liệu thống kê,* NXB Giáo dục,1996.

[11]. Nguyễn Cao Văn và Trần Thái Ninh, *Bài giảng xác suất và thống kê toán,* NXB Thống kê, Hà Nội 1999.

[12]. 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.

[13]. D. L. (Paul) Minh, *Applied Probability Models*, Duxbury, Thomson Learning, 2001.

[14]. B.V. Gnedenko, *The theory of probability,* Mir publishers, Moscow 1976.

[15]. Harald Cramer, *Phương pháp toán học trong thống kê*, NXB Khoa học và Kỹ thuật, Hà Nội 1970.

LÝ THUYẾT
XÁC SUẤT VÀ THỐNG KÊ TOÁN

Mã số: 497XSU210

Chịu trách nhiệm bản thảo

TRUNG TÂM ĐÀO TẠO BƯU CHÍNH VIỄN THÔNG 1

(Tài liệu này được ban hành theo Quyết định số: 375/QĐ-TTĐT1 ngày 22/05/2006 của Giám đốc Học viện Công nghệ Bưu chính Viễn thông)

In tại : *Công ty cổ phần In Bưu điện*
Số lượng : *2000 cuốn, khổ 19 x 26 cm*
Ngày hoàn thành : *01/06/2006.*