HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG

Giáo trình

LÝ THUYẾT XÁC SUẤT THỐNG KÊ

*(Dành cho sinh viên hệ đại học
chuyên ngành Điện tử-Viễn thông-Công nghệ thông tin)*

Biên soạn: PGS.TS. Lê Bá Long

Hà Nội, 2008

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 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.

Giáo trình này được biên soạn cho hệ đại học chuyên ngành Điện tử-Viễn thông và Công nghệ thông tin theo đề cương chi tiết chương trình qui định của Học viện Công nghệ Bưu Chính Viễn Thông. Nội dung của cuốn sách bám sát các giáo trình của các trường đại học khối kỹ thuật 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, các ngành đại học và cao đẳng khối kỹ thuật.

Giáo trình gồm 6 chương tương ứng với 4 đơn vị học trình (60 tiết):

Chương I: Các khái niệm cơ bản về xác suất.

Chương II: Biến ngẫu nhiên và các đặc trưng của chúng.

Chương III: Véc tơ ngẫu nhiên và các đặc trưng của chúng.

Chương IV: Luật số lớn và định lý giới hạn.

Chương V:.Thống kê toán học

Chương VI: Quá trình ngẫu nhiên và chuỗi Markov.

Điều kiện tiên quyết 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 về thời gian của chương trình toán dành cho ngành kỹ thuật, do đó nhiều kết quả và định lý chỉ được phát biểu và minh họa chứ không có điều kiện để chứng minh chi tiết.

Giáo trình được trình bày theo cách thích hợp đối với người tự học. Trước khi nghiên cứu các nội dung chi tiết, người đọc 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 đọc có thể tự đọc và hiểu được cặn kẽ thông qua cách diễn đạt và chỉ dẫn rõ ràng. Đặc biệt bạn đọc nên chú ý đến các nhận xét, bình luận để hiểu sâu hơn hoặc mở rộng tổng quát hơn các kết quả và hướng ứng dụng vào thực tế. Hầu hết các bài toán 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 đọc dễ dàng hơn khi tiếp thu bài học. 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 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 để giải quyết. Vì vậy việc giải các bài tập này giúp học viên nắm chắc hơn lý thuyết và 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 3, mục 3.5 hàm của các biến ngẫu nhiên, mục 3.7 phân bố có điều kiện và kỳ vọng có điều kiện, chương 4 luật số lớn và định lý giới hạn trung tâm, mục 6.3 chương 6 …).

Tuy rằng tác giả đã rất cố gắng, song vì thời gian bị hạn hẹp cùng với yêu cầu cấp bách của Học viện, vì vậy các thiếu sót còn tồn tại trong giáo trình là điều khó tránh khỏi. Tác giả rất mong sự đóng góp ý kiến của bạn bè đồng nghiệp, học viên xa gần và xin cám ơn vì điều đó.

Cuối cùng chúng tôi bày tỏ sự cám ơn đối với Ban Giám đốc Học viện Công nghệ Bưu Chính Viễn Thông 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.

Lê Bá Long

Khoa cơ bản 1

Học Viện CNBCVT

CHƯƠNG I: CÁC KHÁI NIỆM CƠ BẢN VỀ 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.
- Dãy phép thử Bernoulli và xác suất nhị thức

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

Để tính xác suất các biến cố theo phương pháp cổ điển đòi hỏi phải tính số các trường hợp thuận lợi đối với biến cố và số các trường hợp có thể. Vì vậy học viên cần nắm vững các phương pháp đếm - giải tích tổ hợp (đã được học ở lớp 12 và trong chương 1 của toán đại số A2). 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.

Với phép thử gieo con xúc xắc, 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 này đó là sự xuất hiện mặt có số nốt $1, 2, 3, 4, 5, 6$. Ta xem các kết quả này là các 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. Vậy $\Omega = \{1, 2, 3, 4, 5, 6\}$.

Phép thử ngẫu nhiên thường được ký hiệu bởi chữ $\mathscr{C}$.

Ví dụ 1.1:

- Phép thử tung đồng xu có không gian mẫu là $\Omega = \{S, N\}$.
- Phép thử tung đồng thời 2 đồng xu có không gian mẫu là

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

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

1.1.2. Biến cố (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 $\mathscr{C}$.

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.

1.1.3. Quan hệ giữa các biến cố

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

a. 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.

b. 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.

Ví dụ 1.3: Bắn một phát đạn vào bia. Gọi A là biến cố "Bắn trúng bia". Biến cố đối của A là $\overline{A}$ "Bắn trượt bia".

c. Tổng của hai biến cố

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

Ví dụ 1.4: Một mạng điện gồm hai bóng đèn mắc nối tiếp. Gọi A_1 là biến cố "Bóng đèn thứ nhất bị cháy", A_2 là biến cố "Bóng đèn thứ hai bị cháy". Gọi A là biến cố "Mạng mất điện". Ta thấy rằng mạng bị mất điện khi ít nhất một trong hai bóng bị cháy. Vậy $A = A_1 \cup A_2$.

d. 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 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.

Ví dụ 1.5: Một mạng điện gồm hai bóng đèn mắc song song. Gọi A_1 là biến cố "Bóng đèn thứ nhất bị cháy", A_2 là biến cố "Bóng đèn thứ hai bị cháy". Gọi A là biến cố "Mạng mất điện". Ta thấy rằng mạng bị mất điện khi cả hai bóng bị cháy. Vậy $A = A_1 A_2$.

e. 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.

Ví dụ 1.6: Một bình có 3 loại cầu: cầu mầu trắng, mầu đỏ và mầu xanh. Lấy ngẫu nhiên 1 cầu từ bình. Gọi A_t, $A_đ$, A_x lần lượt là biến cố quả cầu rút được là cầu trắng, đỏ, xanh. Các biến cố này xung khắc từng đôi một, vì mỗi quả cầu chỉ có 1 mầu.

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.

f. 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 chắc, 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 1.7 : 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.

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

Định lý 1.2: Nếu A, B độc lập thì các cặp biến cố: $A, \overline{B}$; $\overline{A}, B$; $\overline{A}, \overline{B}$ cũng độc lập.

Ví dụ 1.8: 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.

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

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

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

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

Giải: a. 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.

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

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

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

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

c. Ba biến cố A, B, C độc lập vì biến cố bắn trúng mục tiêu của mỗi xạ thủ là độc lập nhau.

Ba biến cố A, B, C không xung khắc vì có thể cùng bắn trúng mục tiêu.

1.2. ĐỊNH NGHĨA XÁC SUẤT VÀ CÁC TÍNH CHẤT

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 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.

1.2.1. Định nghĩa cổ điển về xác suất

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

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

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

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

Ví dụ 1.9: 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.

1.2.2. Các qui tắc đếm

a. Qui tắc cộng

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

b. 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.

c. 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.

d. 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

$$A_n^k = \frac{n!}{(n-k)!} = n \cdot (n-1) \cdots (n-k) \qquad \textbf{(1.2)}$$

e. Tổ hợp

Một tổ hợp chập k của n phần tử là một tập con k phần tử của tập n phần tử. Cũng có thể xem một tổ hợp chập k của n phần tử là một cách chọn đồng thời k phần tử của tập n phần tử.

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

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

Do đó với mỗi tổ hợp chập k 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

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

Ví dụ 1.10: 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 1.11: Một người gọi điện thoại quên mất hai số cuối của số điện thoại và chỉ nhớ được rằng chúng khác nhau. Tìm xác suất để quay ngẫu nhiên một lần được đúng số cần gọi.

Giải: Gọi A là biến cố "quay ngẫu nhiên một lần được đúng số cần gọi". Số các trường hợp có thể là số các cặp hai chữ số khác nhau từ 10 chữ số từ 0 đến 9. Nó bằng số các chỉnh hợp 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 1.12: Cho các từ mã 6 bit được tạo từ các chuỗi các bit 0 và bit 1 đồng khả năng. Hãy tìm xác suất của các từ có chứa k bit 1, với $k = 0, ..., 6$.

Giải: Số trường hợp có thể $|\Omega| = 2^6$. Đặt A_k là biến cố " từ mã có chứa k bit 1" . Có thể xem mỗi từ mã có chứa k bit 1 là một tổ hợp chập k của 6 phần tử, vậy số trường hợp thuận lợi đối với A_k là số các tổ hợp 6 chập k. Do đó $|A_k| = C_6^k = \frac{6!}{k!(6-k)!}$

Vậy xác suất của các biến cố tương ứng $P(A_k) = \frac{6!}{k!(6-k)!2^6}$, $k = 0, ..., 6$.

Ví dụ 1.13: 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 $P = 14/15$.

1.2.3. 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 $P(A)$.

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

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

Ví dụ 1.14: 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 1.15: 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 n đủ lớn lần 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.

1.2.4. Định nghĩa xác suất theo hình học

Định nghĩa 1.3: *Giả sử không gian mẫu Ω có thể biểu diễn tương ứng với một miền nào đó có diện tích (thể tích, độ dài) hữu hạn và biến cố A tương ứng với một miền con của Ω thì xác suất của biến cố A được định nghĩa:*

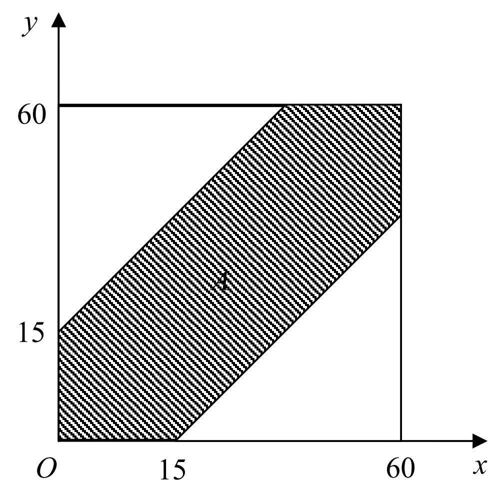
$$P(A)=\frac{\text{diện tích } A}{\text{diện tích } \Omega}.$$

Ví dụ 1.16: Hai người bạn hẹn gặp nhau ở một địa điểm trong khoảng thời gian từ 12h đến 13h. Mỗi người có thể đến điểm hẹn một cách ngẫu nhiên tại một thời điểm trong khoảng thời gian nói trên và họ quy ước rằng ai đến trước thì chỉ đợi người kia trong vòng 15 phút. Tính xác suất để hai người gặp nhau.

Giải: Giả sử x, y là thời điểm người thứ nhất và thứ hai đến điểm hẹn thì:

$$0\le x\le 60,\ 0\le y\le 60.$$

Vậy mỗi cặp thời điểm đến $(x;y)$ là một điểm của hình vuông $\Omega=[0,60]^2$.



Gọi A là biến cố hai người gặp nhau thì

$$A=\{(x;y)\in\Omega \,\big|\, |x-y|\le 15\}$$
$$=\{(x;y)\in\Omega \,\big|\, -15+x\le y\le x+15\}.$$

$$\Rightarrow P(A)=\frac{\text{diện tích } A}{\text{diện tích } \Omega}=1-\frac{45^2}{60^2}=1-\frac{9}{16}=\frac{7}{16}.$$

1.2.5. Các tính chất và định lý xác suất

1.2.5.1. Các tính chất của xác suất

Các định nghĩa trên của xác suất thoả 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.

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

1.2.5.2. Qui tắc cộng xác suất

a. Trường hợp xung khắc

Nếu A, B là hai biến cố xung khắc thì

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

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

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

b. 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

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

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

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

Ví dụ 1.17: 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 $A = A_1 \cup A_2$.

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

1.2.5.3. Quy tắc xác suất của biến cố đối

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

Với mọi biến cố A

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

Ví dụ 1.18: Thống kê cho thấy 12% đàn ông trong vùng nghiện rượu, 23% nghiện thuốc lá và 9% vừa nghiện rượu vừa nghiện thuốc lá. Chọn ngẫu nhiên một người đàn ông trong vùng này, tính xác suất để ông ta không nghiện hai thứ trên.

Giải: Gọi A là biến cố "người đó nghiện rượu", B là biến cố "người đó nghiện thuốc lá".

Gọi H là biến cố "người đó không nghiện hai thứ trên".

Theo giả thiết $P(A)=0,12$, $P(B)=0,23$, $P(AB)=0,09$.

$P\left(\overline{H}\right)=P\left(A\cup B\right)=P(A)+P(B)-P(AB)=0,12+0,23-0,09=0,26$.

Vậy $P(H)=1-P\left(\overline{H}\right)=1-0,26=0,74$.

1.2.6. Nguyên lý xác suất lớn, xác suất nhỏ

Một biến cố không thể có xác suất bằng 0. Tuy nhiên một biến cố có xác suất bằng 0 vẫn có thể xảy ra trong một số lớn các phép thử. 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.

1.3. XÁC SUẤT CÓ ĐIỀU KIỆN

1.3.1. Định nghĩa và các tính chất của 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

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

➢ Khi cố định A với $P(A)>0$ thì xác suất có điều kiện $P\left(B\,|\,A\right)$ có tất cả các tính chất của xác suất thông thường (công thức (1.5)-(1.10)) đối với biến cố B.

Chẳng hạn:

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

Ví dụ 1.19: 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 $P(AB)=\frac{3}{36} \Rightarrow P\left(B|A\right)=\frac{3}{36}\Big/\frac{11}{36}=\frac{3}{11}$.

1.3.2. Quy tắc nhân xác suất

1.3.2.1. 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

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

1.3.2.2. Trường hợp tổng quát:

- $$P(AB)=P(A)P\left(B\,|\,A\right) \qquad \textbf{(1.14)}$$

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

Ví dụ 1.20: Túi I chứa 3 bi trắng, 7 bi đỏ, 15 bi xanh.

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

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

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

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

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

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

$$P\left(A_tB_t \cup A_đB_đ \cup A_xB_x\right) = P\left(A_tB_t\right) + P\left(A_đB_đ\right) + P\left(A_xB_x\right)$$

$$= P\left(A_t\right)P\left(B_t\right) + P\left(A_đ\right)P\left(B_đ\right) + P\left(A_x\right)P\left(B_x\right)$$

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

Ví dụ 1.21: Hai máy bay ném bom 1 mục tiêu, mỗi máy bay ném 1 quả với xác suất trúng mục tiêu tương ứng là $0{,}7$ và $0{,}8$. Tìm xác suất để mục tiêu bị trúng bom.

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

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

Do đó $P(A) = P\left(A_1 \cup A_2\right) = P\left(A_1\right) + P\left(A_2\right) - P\left(A_1\right)P\left(A_2\right) = 0{,}7 + 0{,}8 - 0{,}7.0{,}8 = 0{,}96$.

Ví dụ 1.22: 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

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

1.3.3. 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 của cùng một phép thử, ta có

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

1.3.4. 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 của cùng một phép thử sao cho $P(B) > 0$ ta có :

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

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 1.23: 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 1.24: 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 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. $P(\overline{A}) = P(H)P\left(\overline{A}\,\middle|\,H\right) + P\left(\overline{H}\right)P\left(\overline{A}\,\middle|\,\overline{H}\right) = p(1-\alpha)+(1-p)\beta$.

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

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

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

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

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

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

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

1.4. 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 1.1: Xác suất của biến cố " A xuất hiện ra đúng k lần trong n phép thử":

$$P_n(k\,;p) = C_n^k p^k (1-p)^{n-k};\ k = 0,1,...,n. \qquad \textbf{(1.18)}$$

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:

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

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

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

Ta cần tìm giá trị $k, 0 \le k \le n$ sao cho xác suất $P_n(k;p)$ đạt giá trị lớn nhất.

Định lý 1.2:

(i). $$P_n(k;p) = \frac{(n-k+1)p}{kq} P_n(k-1;p) \quad \textbf{(1.19)}$$

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

$$(n+1)p - 1 \le m \le (n+1)p \quad \textbf{(1.20)}$$

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 $m = (n+1)p$

$$P_{\max} = P_n(m-1;p) = P_n(m;p) \quad \textbf{(1.20)'}$$

Chứng minh:

$$\frac{P_n(k;p)}{P_n(k-1;p)} = \frac{\dfrac{n!}{k!(n-k)!} p^k q^{n-k}}{\dfrac{n!}{(k-1)!(n-k+1)!} p^{k-1} q^{n-k+1}} = \frac{(n-k+1)p}{kq}$$, từ đó có (1.19).

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

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

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

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

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

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

Định nghĩa 1.1: *m xác định bởi công thức (1.20) hoặc (1.20)' được gọi là số lần xuất hiện có khả năng nhất hay giá trị có khả năng xảy ra lớn nhất.*

Ví dụ 1.26: Bắn 7 viên đạn vào bia. Xác suất trúng đích của mỗi viên là $0,6$. Tìm xác suất trong các trường hợp sau:

a. Có đúng 3 viên trúng bia.

b. Có ít nhất 6 viên trúng bia.

c. Có ít nhất 1 viên trúng bia.

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

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

$$P_7(3;0,6)=C_7^3\left(0,6\right)^3\left(0,4\right)^4=0{,}1935\,.$$

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

$$P_7(6;0,6)+P_7(7;0,6)=C_7^6\left(0,6\right)^6\left(0,4\right)+C_7^7\left(0,6\right)^7=0{,}1586\,.$$

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

$$1-P_7(0;0,6)=1-C_7^0\left(0,6\right)^0\left(0,4\right)^7=1-(0,4)^7=0{,}998\,.$$

Ví dụ 1.27: Tín hiệu thông tin được phát đi 3 lần độc lập nhau. Xác suất thu được mỗi lần là 0.4.

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

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

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

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

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

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

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

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

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

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

$$1-\left(0,6\right)^n\geq 0{,}9\Leftrightarrow\left(0,6\right)^n\leq 0{,}1\;\Leftrightarrow n\geq\frac{\lg\left(0,1\right)}{\lg\left(0,6\right)}=\frac{-1}{-1+0{,}778}=4{,}504\,.\text{ Chọn } n=5\,.$$

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 xung khắc thì $P(A \cup B) = P(A) + P(B)$.

Đúng ☐ Sai ☐ .

1.4 Hệ hai biến cố $\{A, \overline{A}\}$ là một hệ đầy đủ.

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 Khi áp dụng công thức xác suất đầy đủ để tính xác suất biến cố B dựa vào hệ đầy đủ $\{A_1, ..., A_n\}$ thì các biến cố B và $A_1, ..., A_n$ phải trong cùng một phép thử.

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% sản phẩm là loại I, 50% sản phẩm là 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 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.

1.18 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.

1.19 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.

1.20 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.

1.21 Một nhà máy ôtô 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.

1.22 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.

1.23 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.

1.24 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 CÁC ĐẶC TRƯNG CỦA CHÚNG

PHẦN 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 biên 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.

Trường hợp 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. Đối với biến ngẫu nhiên nhận giá trị liên tục thì hàm phân bố xác suất có thể được xác định bởi hàm mật độ xác suất.

Ngoài phương pháp sử dụng hàm phân bố xác suất 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.

Một số quy luật phân bố xác suất quan trọng:

- Quy luật nhị thức, quy luật này thường gặp trong dãy phép thử Bernoulli.
- Quy luật Poisson, quy luật này thường gặp trong bài toán về quá trình đếm sự xuất hiện biến cố A nào đó. Quá trình đến của các hệ phục vụ.
- Quy luật phân bố đều một đoạn là quy luật phân bố xác suất của biến ngẫu nhiên liên tục đồng khả năng lấy giá trị trong đoạn 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 phân bố mũ.
- Quy luật phân bố Erlang.
- Quy luật chuẩn.
- Quy luật “khi bình phương”.
- Quy luật Student.

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 ...

Với mỗi quy luật phân bố xác suất ta sẽ khảo sát bảng phân bố xác suất hoặc hàm mật độ xác suất các tính chất và các đặc trưng của nó.

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 định nghĩa 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. KHÁI NIỆM 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, nghĩa là với mọi giá trị thực $x \in \mathbb{R}$ thì $\{X < x\}$ là một biến cố ngẫu nhiên..*

Như vậy 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 một khoảng thời gian nào đó.
- Số cuộc gọi đến một tổng đài trong một khoảng thời gian nào đó.
- Sai số khi đo lường một đại lượng vật lý …

2.1.2. Phân loại

Các biến ngẫu nhiên có thể phân 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 $x_1, x_2, \dots$.

❖ ***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 biến ngẫu nhiên nhận một giá trị cụ thể đều bằng 0 (nghĩa là $P\{X = a\} = 0$ 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 $0, 1, 2, \dots$
- Số cuộc gọi đến một tổng đài là biến ngẫu nhiên rời rạc nhận các giá trị $0, 1, 2, \dots$
- 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.

2.1.3. 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. BIẾN NGẪU NHIÊN RỜI RẠC

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

Với biến ngẫu nhiên rời rạc chúng ta có thể nghiên cứu thông qua bảng ghi các giá trị mà biến ngẫu nhiên nhận với xác suất tương ứng, đó là bảng phân bố xác suất.

Giả sử biến ngẫu nhiên X chỉ nhận các giá trị $x_1, x_2, \dots$ với xác suất tương ứng $p_i = P\{X = x_i\}$ thỏa mãn:

$$p_i > 0 \text{ và } \sum_i p_i = 1. \tag{2.5}$$

Bảng phân bố xác suất của X có dạng sau:

X	x_1	x_2	…
P	p_1	p_2	…

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

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

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

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

$$\{X = x_1\}, \{X = x_2\}, ..., \{X = x_n\} \tag{2.7}$$

lập thành hệ đầy đủ các biến cố.

Hàm phân bố xác suất có dạng:

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

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 xác suất và hàm phân bố xác suất.

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

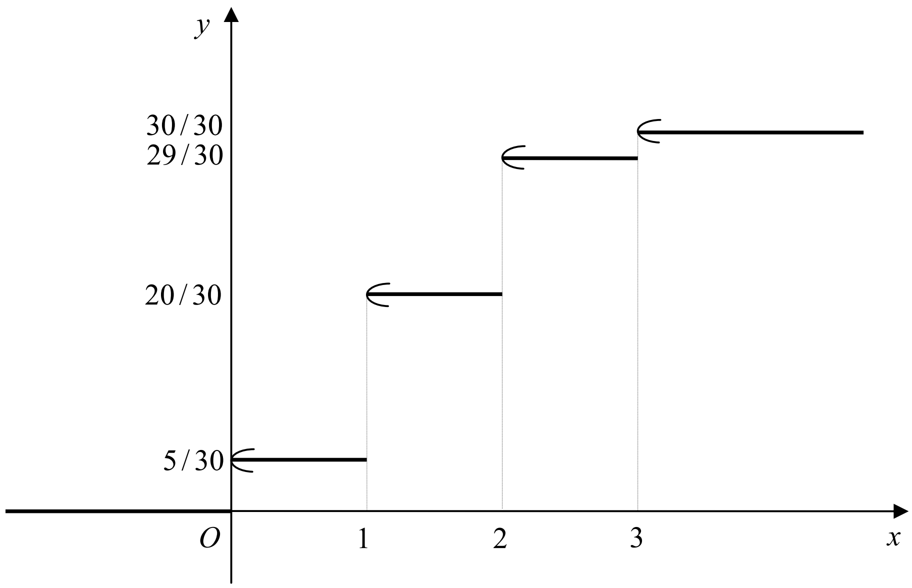
$P\{X=2\} = \dfrac{C_6^1 C_4^2}{C_{10}^3} = \dfrac{9}{30}$, $P\{X=3\} = \dfrac{C_4^3}{C_{10}^3} = \dfrac{1}{30}$

Bảng phân bố xác suất:

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

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

Đồ thị



Ví dụ 2.4: Xác suất để xạ thủ bắn trúng bia là 0,8. Xạ thủ được phát từng viên đạn để bắn cho đến khi trúng bia. Hãy xây dựng bảng phân bố xác suất của số viên đạn được phát.

Giải: Gọi X là "số viên đạn được phát". X là biến ngẫu nhiên rời rạc với các giá trị có thể có là $1,2,...,k,...$

Gọi A_k là biến cố "xạ thủ bắn trúng bia ở viên đạn thứ k". Các biến cố A_k độc lập nhau và $P\left(A_k\right)=0,8$.

$P\{X=1\}=P\left(A_1\right)=0,8$

$P\{X=2\}=P\left(\overline{A_1}A_2\right)=P\left(\overline{A_1}\right)P\left(A_2\right)=0,2.0,8\ \ldots$

$P\{X=k\}=P\left(\overline{A_1}...\overline{A_{k-1}}A_k\right)=P\left(\overline{A_1}\right)...P\left(\overline{A_{k-1}}\right)P\left(A_k\right)=\left(0,2\right)^{k-1}.0,8\ \ldots$

Bảng phân bố xác suất có dạng:

X	1	2	…	k	…
P	0,8	$0,2.0,8$	…	$\left(0,2\right)^{k-1}.0,8$	…

2.2.2. Phân bố nhị thức $\mathscr{B}(n;p)$

Định nghĩa 2.3: *Biến ngẫu nhiên rời rạc X nhận các giá trị $0, 1, ..., n$ với xác suất tương ứng*

$$P\{X=k\}=C_n^k p^k q^{n-k};\ k=0,1,...,n \qquad \textbf{(2.9)}$$

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

Nhận xét:

1. Thực hiện n phép thử Bernoulli với xác suất thành công của biến cố A trong mỗi lần thử là p.

Với mỗi $i=1,2,...,n$; nếu ở lần thử thứ i biến cố A xuất hiện ta cho X_i nhận giá trị 1, nếu biến cố A không xuất hiện ta cho X_i nhận giá trị 0. Như vậy X_i là biến ngẫu nhiên rời rạc có bảng phân bố xác suất:

X_i	0	1
P	$1-p$	p

X_i được gọi là có phân bố không- một $A(p)$. **(2.10)**

Gọi X là số thành công trong n phép thử Bernoulli này thì

$$X=X_1+X_2+\cdots+X_n \sim \mathscr{B}(n;p) \qquad \textbf{(2.11)}$$

2. Từ (2.11) suy ra rằng nếu $X \sim \mathscr{B}(n_1;p)$ và $Y \sim \mathscr{B}(n_2;p)$ thì

$$X+Y \sim \mathscr{B}(n_1+n_2;p) \qquad \textbf{(2.12)}$$

Ví dụ 2.5: Tỉ lệ phế phẩm của lô hàng là 4%. Chọn ngẫu nhiên 20 sản phẩm để kiểm tra. Gọi X là số phế phẩm kiểm tra được.

a. Gọi tên luật phân bố xác suất của X.

b. Tính xác suất có đúng 5 phế phẩm kiểm tra được.

c. Lô hàng được xem là đạt tiêu chuẩn nếu số phế phẩm kiểm tra được không nhiều hơn 2. Tính xác suất để lô hàng đạt tiêu chuẩn.

Giải: Có thể xem kiểm tra chất lượng mỗi sản phẩm và phát hiện ra phế phẩm là thực hiện một phép thử Bernoulli với xác suất thành công của mỗi lần thử là 0,04. Kiểm tra 20 sản phẩm là thực hiện 20 phép thử.

a. Số phế phẩm kiểm tra được là số lần thành công trong 20 phép thử này. Vậy X có phân bố nhị thức $X \sim \mathscr{B}(n;p)$, với $n = 20,\ p = 0,04$.

b. Xác suất để có đúng 5 phế phẩm là $P\{X=5\} = C_{20}^{5}\left(0,04\right)^{5}\left(0,96\right)^{15} = 0,0008$.

c. Xác suất để lô hàng đạt tiêu chuẩn là $P\{X \le 2\} = 0,956$.

2.2.3. Phân bố Poisson

Định nghĩa 2.4: *Biến ngẫu nhiên* X *nhận các giá trị* $k = 0, 1, 2, \ldots$ *với xác suất*

$$P\{X=k\} = e^{-\lambda}\frac{\lambda^{k}}{k!};\ k = 0,1,2,\ldots \tag{2.13}$$

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 là tốc độ trung bình diễn ra trong khoảng thời gian này.

Ví dụ 2.6: 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 thì $X(t)$ có phân bố Poisson tham số $\lambda = 2t$.

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

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

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 sắp hàng, các hệ phục vụ đám đông, các bài toán chuyển mạch trong tổng đài …

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 tham số $\lambda_1 + \lambda_2$.

$$X_1 + X_2 \sim \mathscr{P}(\lambda_1 + \lambda_2) \qquad \textbf{(2.14)}$$

2.3. BIẾN NGẪU NHIÊN LIÊN TỤC

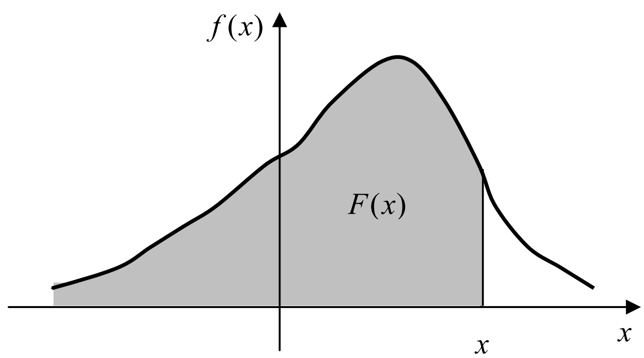
2.3.1. Hàm mật độ xác suất của biến ngẫu nhiên liên tục

Định nghĩa 2.5: *Giả sử X là một biến ngẫu nhiên liên tục có hàm phân bố xác suất $F(x)$. Nếu tồn tại hàm $f(x)$ sao cho với mọi $x \in \mathbb{R}$*

$$F(x) = \int_{-\infty}^{x} f(t)dt \qquad \textbf{(2.15)}$$

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

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



Tính chất của hàm mật độ xác suất

a. $F'(x) = f(x)$ tại các điểm x mà $f(x)$ liên tục. **(2.16)**

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

c. $\int_{-\infty}^{\infty} f(x)dx = 1$, **(2.18)**

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

Ví dụ 2.7 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. Xét tại $x = 1$

$$\begin{cases} F(1) = kx^2\big|_{x=1} = k \\ \lim\limits_{x \to 1^+} F(x) = 1 \end{cases}. \Rightarrow k = 1.$$

b. Theo tính chất (2.16) của hàm mật độ xác suất ta có

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

Ví dụ 2.8 Biến ngẫu nhiên liên tục X với hàm mật độ xác suất có 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 xác suất $F(x)$;

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

b. Từ công thức (2.15) xác định hàm mật độ xác suất ta có

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

c. Từ công thức (2.14) ta có $P\{2<X<3\}=F(3)-F(2)=\frac{2}{3}-\frac{1}{2}=\frac{1}{6}$.

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 $\left(\frac{5}{6}\right)^4\approx 0{,}48$.

2.3.2. Phân bố đều $\mathbf{U}(a,b)$

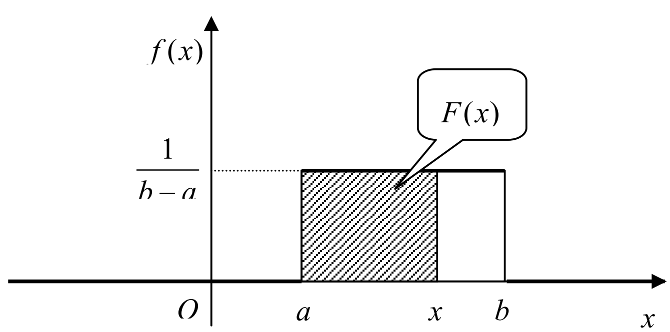
Định nghĩa 2.6: *Biến ngẫu nhiên X được gọi là có phân bố đều trên $[a,b]$ nếu hàm mật độ xác suất của nó xác định bởi:*

$$f(x)=\begin{cases}\dfrac{1}{b-a} & \text{nếu } a\le x\le b\\ 0 & \text{nếu ngược lại}\end{cases} \qquad \textbf{(2.20)}$$

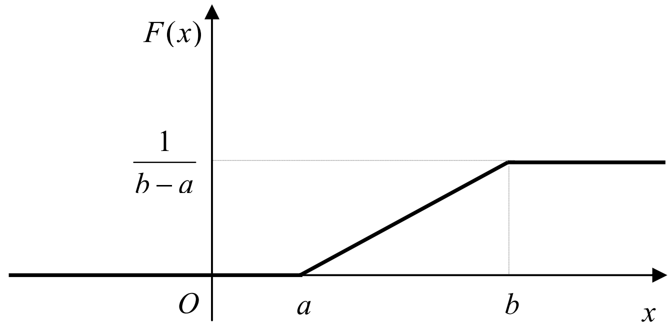
Hàm phân bố xác suất tương ứng

$$F(x)=\int_{-\infty}^{x} f(t)dt=\begin{cases}0 & \text{nếu } x<a\\ \dfrac{x-a}{b-a} & \text{nếu } a\le x\le b\\ 1 & \text{nếu } x>b\end{cases} \qquad \textbf{(2.21)}$$

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 xác suất của phân bố đều $\mathbf{U}(a,b)$



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

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.

2.3.3. Phân bố mũ

Định nghĩa 2.7: *Biến ngẫu nhiên* X *được gọi là có phân bố mũ tham số* $\lambda > 0$ *nếu có hàm mật độ xác suất:*

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{nếu } \; x > 0 \\ 0 & \text{nếu } \; x \le 0 \end{cases} \qquad \textbf{(2.22)}$$

Hàm phân bố xác suất

$$F(x) = \int_{-\infty}^{x} f(t)dt = \begin{cases} 0 & \text{nếu } \; x \le 0 \\ 1 - e^{-\lambda x} & \text{nếu } \; x > 0 \end{cases} \qquad \textbf{(2.23)}$$

Phân bố mũ thường xuất hiện trong các bài toán về thời gian sống của một loài sinh vật, tuổi thọ của thiết bị… hoặc khoảng thời gian giữa hai lần xuất hiện của một biến cố E nào đó mà số lần xuất hiện của E tuân theo luật phân bố Poisson.

Ví dụ 2.8: Tuổi thọ của một mạch điện tử trong máy tính là một biến ngẫu nhiên có phân bố mũ tham số $\lambda > 0$. Giả sử tuổi thọ trung bình của mạch điện tử này là $\frac{1}{\lambda} = 6,25$ (năm). Thời gian bảo hành là 5 năm. Hỏi có bao nhiêu phần trăm mạch điện tử bán ra phải thay thế trong thời gian bảo hành.

Giải: Gọi X là tuổi thọ của mạch điện tử. Xác suất để mạch điện tử bị hỏng trong thời gian bảo hành là:

$$P\{X \le 5\} = 1 - e^{-5\lambda} = 1 - e^{\frac{-5}{6,25}} = 1 - e^{-0,8} = 1 - 0,449 = 0,551.$$

Vậy có khoảng 55% số mạch điện tử bán ra phải thay thế trong thời gian bảo hành.

Biến ngẫu nhiên X *được gọi là không nhớ (memoryless) nếu*

$$P\{X > x+t | X > t\} = P\{X > x\} \quad \forall x,t > 0$$

$$P\{X > x + t\} = P\{X > x\}P\{X > t\} \quad \forall x,t > 0 \qquad \textbf{(2.24)}$$

Gọi $F(x)$ là hàm phân bố xác suất của X, đặt $G(x) = P\{X > x\} = 1 - F(x)$.

Điều kiện (2.23) có thể viết lại

$$G(x+t) = G(x)G(t) \qquad \textbf{(2.25)}$$

Giải phương trình (2.25) (phương trình hàm Cauchy)với điều kiện $G(x) = 1,\ \forall x < 0$ và $G(+\infty) = 0$ ta được $G(x) = e^{-\lambda x}$.

Vậy biến ngẫu nhiên X không nhớ khi và chỉ khi X có phân bố mũ. Chính vì thế phân bố mũ còn được gọi là *phân bố Markov*.

2.3.4. Phân bố Erlang

Định nghĩa 2.8: *Biến ngẫu nhiên X có phân bố Erlang tham số $(k;\lambda)$; k là số tự nhiên và $\lambda > 0$, nếu hàm mật độ xác suất có dạng:*

$$f(x) = \begin{cases} \dfrac{\lambda^k}{(k-1)!} x^{k-1} e^{-\lambda x} & \text{nếu } \ x > 0 \\ 0 & \text{nếu } \ x \le 0 \end{cases} \qquad \textbf{(2.26)}$$

Có thể chứng minh được rằng nếu $X_1, X_2, ..., X_k$ là k biến ngẫu nhiên độc lập cùng có phân bố mũ tham số $\lambda > 0$ thì $X = X_1 + X_2 + \cdots + X_k$ có phân bố Erlang tham số $(k;\lambda)$.

2.3.5. Phân bố chuẩn $\mathbf{N}(\mu;\sigma^2)$

2.3.5.1. Định nghĩa

Định nghĩa 2.9: *Biến ngẫu nhiên liên tục X có phân bố chuẩn $\mathbf{N}(\mu;\sigma^2)$, ký hiệu $X \sim \mathbf{N}(\mu;\sigma^2)$, nếu hàm mật độ xác suất có dạng*

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}; \quad \forall x \in \mathbb{R} \qquad \textbf{(2.27)}$$

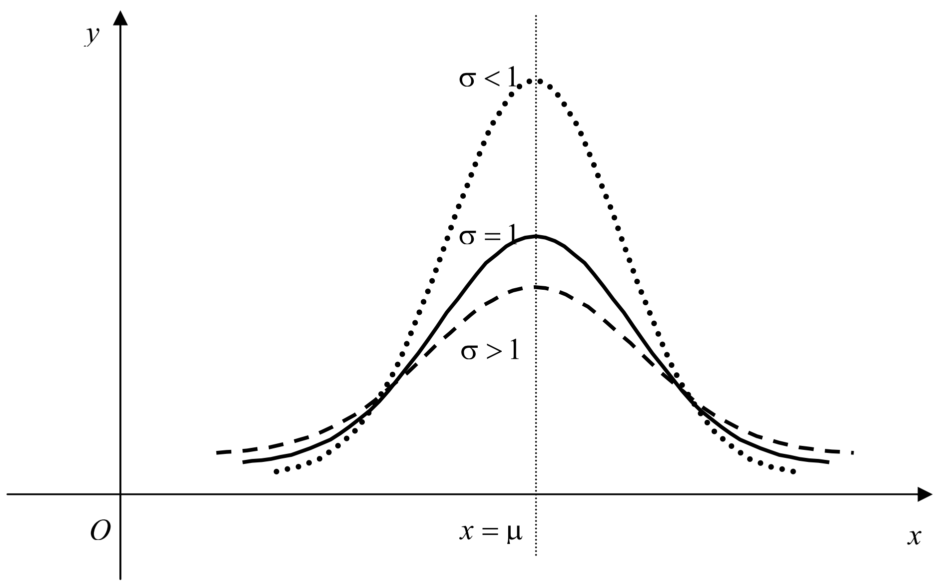
Phân bố chuẩn được Gauss tìm ra năm 1809 nên nó còn được gọi là *phân bố Gauss*. Phân bố chuẩn thường được thấy trong các bài toán về sai số gặp phải khi đo đạc các đại lượng vật lý, thiên văn ...

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 ...

2.3.5.2. Tính chất đồ thị của hàm mật độ xác suất của quy luật chuẩn

Từ công thức xác định hàm mật độ xác suất (2.26) 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 $\dfrac{1}{\sigma\sqrt{2\pi}}$. 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.



Đồ thị của phân bố chuẩn $\mathbf{N}(\mu;\sigma^2)$

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

$$X_1 + X_2 \sim \mathbf{N}(\mu_1 + \mu_2;\sigma_1^2 + \sigma_2^2) \qquad \textbf{(2.28)}$$

2.3.5.3. Phân bố chuẩn tắc

Phân bố chuẩn $\mathbf{N}(\mu;\sigma^2)$ với $\mu = 0$, $\sigma^2 = 1$ gọi là *phân bố chuẩn tắc* $\mathbf{N}(0;1)$.

Hàm mật độ xác suất của phân bố $\mathbf{N}(0;1)$

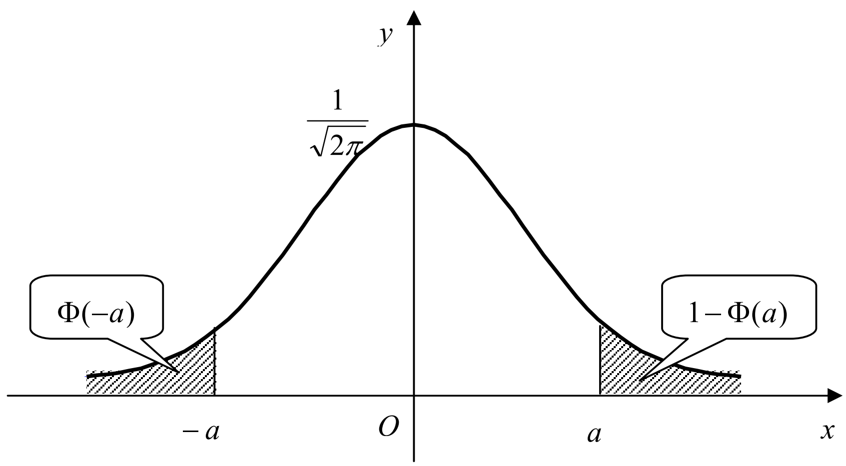
$$\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{\frac{-x^2}{2}} \quad ; \quad \forall x\in\mathbb{R} \qquad \textbf{(2.28)}$$

Hàm phân bố xác suất của $\mathbf{N}(0;1)$

$$\Phi(x)=\int_{-\infty}^{x}\varphi(t)dt=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{\frac{-t^2}{2}}dt \quad ; \quad \forall x\in\mathbb{R} \qquad \textbf{(2.29)}$$

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 xác suất $\varphi(x)$



Đồ thị của phân bố chuẩn tắc $\mathbf{N}(0;1)$

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

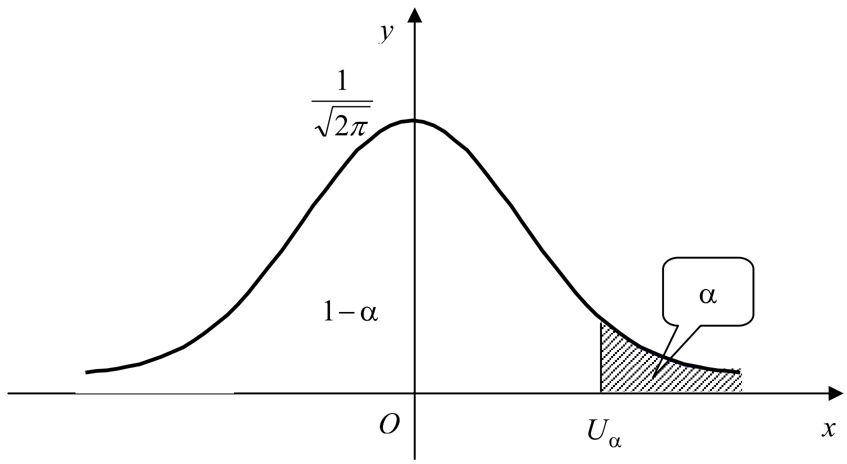
1) $\Phi(x)+\Phi(-x)=1$, $\Phi(-x)=1-\Phi(x)$.

2) Nếu $X\sim\mathbf{N}(0;1)$ thì

$$\forall a>0,\ P\{|X|<a\}=2\Phi(a)-1,\ P\{|X|>a\}=2\big(1-\Phi(a)\big). \qquad \textbf{(2.30)}$$

Định nghĩa 2.10: *Giá trị U_α gọi là giá trị tới hạn mức α của phân bố chuẩn tắc nếu*

$$\Phi^{-1}(1-\alpha)=U_\alpha . \qquad \textbf{(2.31)}$$



Nếu $X \sim \mathbf{N}(0;1)$ thì

$$P\{X > U_{\alpha}\} = \alpha \ ; \ P\left\{|X| > U_{\frac{\alpha}{2}}\right\} = \alpha \ ; \ P\left\{|X| < U_{\frac{\alpha}{2}}\right\} = 1 - \alpha \ . \tag{2.32}$$

Người ta chứng minh được:

$$X \sim \mathbf{N}(\mu;\sigma^2) \ \Rightarrow \ \frac{X-\mu}{\sigma} \sim \mathbf{N}(0;1) \, . \tag{2.33}$$

Từ đó ta có

$$F(x) = P\{X < x\} = P\left\{\frac{X-\mu}{\sigma} < \frac{x-\mu}{\sigma}\right\} = \Phi\left(\frac{x-\mu}{\sigma}\right). \tag{2.34}$$

$$P\{a < X < b\} = P\left\{\frac{a-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{b-\mu}{\sigma}\right\} = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right). \tag{2.35}$$

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

$$P\{|X-\mu| < \varepsilon\} = 2\Phi\left(\frac{\varepsilon}{\sigma}\right) - 1 \tag{2.36}$$

Ví dụ 2.9: Giả sử $X \sim \mathbf{N}(\mu;\sigma^2)$, $\mu = 2100, \ \sigma = 200$. Hãy tìm:

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

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

c) Xác định a để $P\{X > a\} = 0{,}03$.

Giải: Áp dụng công thức (3.23), (3.24), (3.25) ta có:

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

b) $P\{1700 < X < 2200\} = \Phi\left(\dfrac{2200-2100}{200}\right) - \Phi\left(\dfrac{1700-2100}{200}\right)$

$$= \Phi(0,5) - \Phi(-2) = 0,6688 .$$

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

2.3.5.4. Quy tắc hai xích ma và ba xích ma

Nếu trong công thức (2.36) 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

$$P\{\mu - 2\sigma < X < \mu + 2\sigma\} = 0,9544 \tag{2.37}$$

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

$$P\{\mu - 3\sigma < X < \mu + 3\sigma\} = 0,9973 \tag{2.38}$$

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

Nếu X có phân bố chuẩn $\mathbf{N}(\mu;\sigma^2)$ thì có đến 95,44% giá trị của X nằm trong khoảng $(\mu-2\sigma;\, \mu+2\sigma)$ và hầu như toàn bộ giá trị của X nằm trong khoảng $(\mu-3\sigma;\, \mu+3\sigma)$.

2.3.6. Phân bố "khi bình phương"

Định nghĩa 2.11: *Biến ngẫu nhiên liên tục X có phân bố "khi bình phương" n bậc tự do, ký hiệu $X \sim \chi_n^2$ nếu hàm mật độ xác suất có dạng*

$$f(x) = \begin{cases} \dfrac{x^{n/2-1}}{2^{n/2}\Gamma(n/2)} e^{-\frac{x}{2}} & \text{nếu } x > 0 \\ 0 & \text{nếu } x \le 0 \end{cases} \tag{2.39}$$

trong đó $\Gamma(x) = \int\limits_0^{+\infty} t^{x-1} e^{-t} dt$ là hàm Gamma.

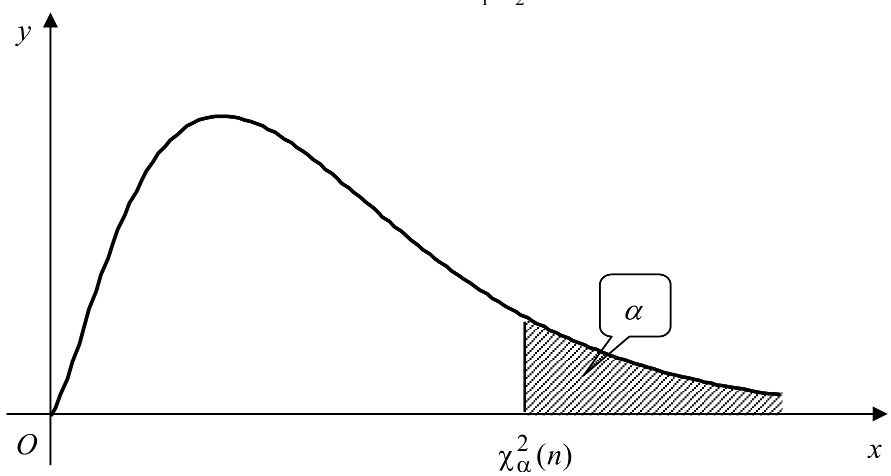
Có thể chứng minh được rằng nếu $X_1, X_2, ..., X_n$ là các biến ngẫu nhiên độc lập có cùng phân bố chuẩn tắc $\mathbf{N}(0,1)$ thì

$$\sum_{i=1}^{n} X_i^2 = X_1^2 + X_2^2 + \cdots + X_n^2 \sim \chi_n^2 \tag{2.40}$$

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

Từ (2.40) 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

$$X_1 + X_2 \sim \chi^2_{n_1+n_2} \tag{2.41}$$



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

$$P\left\{\chi^2 > \chi^2_\alpha(n)\right\} = \alpha\,. \tag{2.42}$$

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

2.3.7. Phân bố Student $\mathbf{T}(n)$

Định nghĩa 2.12: *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 xác suất có dạng:*

$$f(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma(n/2)}\left(1+\frac{t^2}{n}\right)^{\frac{-(n+1)}{2}}, \quad -\infty < t < \infty\,. \tag{2.43}$$

trong đó $\Gamma(x)$ là hàm Gamma.

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

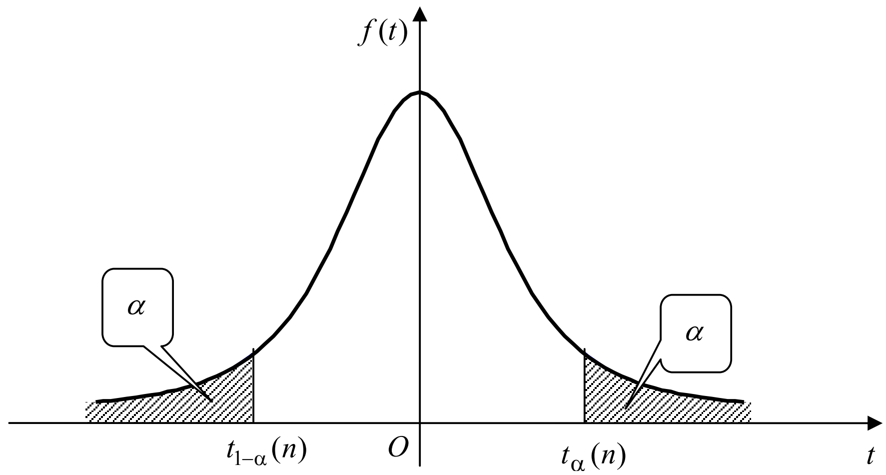
$$T = \frac{Z}{\sqrt{V/n}} \sim \mathbf{T}(n) \tag{2.44}$$

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

$$P\{T > t_\alpha(n)\} = \alpha\,. \tag{2.43}$$

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

Hàm mật độ xác suất (2.43) là hàm chẵn nên đồ thị đối xứng qua trục tung. Khi số bậc tự do tăng lên, phân bố Student hội tụ rất nhanh về phân bố chuẩn tắc $\mathbf{N}(0,1)$. Do đó khi n đủ lớn ($n \ge 30$) 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.



2.4. CÁC THAM SỐ ĐẶC TRƯNG CỦA BIẾN NGẪU NHIÊN

2.4.1. Kỳ vọng toán

2.4.1.1. Định nghĩa

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

$$\mathrm{E}X = \sum_i x_i p_i \qquad \textbf{(2.44)}$$

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

$$\mathrm{E}X = \int_{-\infty}^{\infty} x f(x)dx \qquad \textbf{(2.45)}$$

Kỳ vọng $\mathrm{E}X$ tồn tại nếu chuỗi (2.44) (trường hợp rời rạc) hội tụ tuyệt đối hoặc tích phân (2.45) (trường hợp liên tục) hội tụ tuyệt đối.

Ví dụ 2.10: 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 2.11: 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 (xem ví dụ 1.10). 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 2.12: Tuổi thọ của một loại côn trùng nào đó là một biến ngẫu nhiên X (đơn vị là tháng) với hàm mật độ xác suất như sau:

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

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

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

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

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

2.4.1.2. 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. $\dfrac{r_1x_1 + r_2x_2 + \cdots + r_mx_m}{n}$ là giá trị trung bình của X, trong đó $r_1 + r_2 + \cdots + r_m = n$.

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

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

Trong trường hợp tổng quát thì tần suất f_i được thay bằng xác suất p_i.

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

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

2.4.1.3. Tính chất

1) $\mathrm{E}(C) = C$ với mọi hằng số C. **(2.46)**

2) $\mathrm{E}(CX) = C\mathrm{E}(X)$ với mọi hằng số C. **(2.47)**

3) $\mathrm{E}(X_1 + \cdots + X_n) = \mathrm{E}(X_1) + \cdots + \mathrm{E}(X_n)$ **(2.48)**

4) Cho hàm số $\varphi(x)$, kỳ vọng của biến ngẫu nhiên $Y = \varphi(X)$ được tính theo công thức

$$\mathrm{E}\,Y = \begin{cases} \sum_i \varphi(x_i)p_i & \text{nếu } X \text{ rời rạc có } p_i = P\{X = x_i\} \\ \int_{-\infty}^{\infty} \varphi(x)f(x)dx & \text{nếu } X \text{ liên tục có hàm mật độ } f(x) \end{cases} \quad \textbf{(2.49)}$$

Đặc biệt ta có công thức tính kỳ vọng của X^2:

$$\mathrm{E}\,X^2 = \begin{cases} \sum_i x_i^2 p_i & \text{nếu } X \text{ rời rạc} \\ \int_{-\infty}^{\infty} x^2 f(x)dx & \text{nếu } X \text{ liên tục có hàm mật độ } f(x) \end{cases} \quad \textbf{(2.50)}$$

5) Nếu $X_1, ..., X_n$ độc lập thì $\mathrm{E}(X_1 \cdots X_n) = \mathrm{E}(X_1) \cdots \mathrm{E}(X_n)$. **(2.51)**

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

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

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

Giải: a) Gọi X là số bi trắng trong 3 bi vừa chọn (xem ví dụ 2.3) 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	$5/30$	$15/30$	$9/30$	$1/30$

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

Mặt khác, theo công thức (2.47) và ví dụ 2.10 ta cũng được $\mathrm{E}\,Y = 200\mathrm{E}\,X = 200\times\frac{6}{5} = 240$.

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

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

Ví dụ 2.14: 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 (2.48) 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 $\mathrm{E}\,X_i = \frac{1}{6}\left(1+2+3+4+5+6\right) = \frac{7}{2} \quad \Rightarrow \quad \mathrm{E}\,X = \frac{7}{2}n$.

2.4.2. Phương sai

2.4.2.1. 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 $\mathrm{D}X$ hoặc $\mathrm{D}\left(X\right)$ và định nghĩa như sau:

$$\mathrm{D}X = \mathrm{E}\left(X - \mathrm{E}\,X\right)^2 \qquad \textbf{(2.52)}$$

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

Khai triển vế phải công thức (2.52) và áp dụng các tính chất của kỳ vọng ta có thể tính phương sai theo công thức sau:

$$\mathrm{D}X = \mathrm{E}\,X^2 - \left(\mathrm{E}\,X\right)^2 \qquad \textbf{(2.53)}$$

Theo công thức (2.49) 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

$$\mathrm{D}X=\sum_i\left(x_i-\mathrm{E}X\right)^2p_i=\sum_i x_i^2p_i-\left(\sum_i x_ip_i\right)^2 \qquad \textbf{(2.54)}$$

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

$$\mathrm{D}X=\int_{-\infty}^{\infty}\left(x-\mathrm{E}X\right)^2f(x)dx=\int_{-\infty}^{\infty}x^2f(x)dx-\left(\int_{-\infty}^{\infty}xf(x)dx\right)^2 \qquad \textbf{(2.55)}$$

Ví dụ 2.15: Tính phương sai của biến ngẫu nhiên xét trong ví dụ 2.7.

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

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

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

Ví dụ 2.16: Tính phương sai và độ lệch chuẩn của biến ngẫu nhiên xét trong ví dụ 2.11.

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

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

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 2.11 cho thấy đầu tư bảo hiểm cho những người 25 tuổi là có lãi, nhưng ví dụ 2.15 cho thấy rủi ro của bảo hiểm rất lớn.

2.4.2.2. Tính chất

1) $\mathrm{D}(a)=0$ với mọi hằng số a. **(2.56)**

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

3) $\mathrm{D}(aX+b)=a^2\mathrm{D}(X)$ với mọi hằng số $a,\ b$. **(2.57)**

4) Nếu $X_1,\ldots,X_n$ độc lập và có các phương sai hữu hạn thì

$$\mathrm{D}\left(a_1X_1+\cdots+a_nX_n\right)=a_1^2\mathrm{D}\left(X_1\right)+\cdots+a_n^2\mathrm{D}\left(X_n\right). \qquad \textbf{(2.58)}$$

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

Ví dụ 2.17: Tung con xúc xắc n lần độc lập nhau. Tìm phương sai của tổng số nốt xuất hiện.

Giải: Xét $X=\sum_{i=1}^{n}X_i$ ở ví dụ 2.11. Vì các $X_i\ (i=1,...,n)$ độc lập nhau, do đó theo công thức (2.58) ta có $\mathrm{D}X=\sum_{i=1}^{n}\mathrm{D}X_i$.

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

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

2.4.3. Phân vị, Trung vị (*)

2.4.3.1. Phân vị

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

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

Hay
$$F(v_\alpha)\le\alpha\le F(v_\alpha+0)\qquad\textbf{(2.59)}$$

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

$$v_\alpha=F^{-1}(\alpha)\qquad\textbf{(2.60)}$$

Giá trị tới hạn mức α là phân vị mức $1-\alpha$. **(2.61)**

- Nếu X rời rạc có phân bố:

X	x_1	x_2	…
P	p_1	p_2	…

đặt $P_i=p_1+\cdots+p_i$ thì

$$v_\alpha=\begin{cases} m,\ \ \forall m\in[x_i,x_{i+1}] & \text{nếu } P_i=\alpha<P_{i+1}\\ x_{i+1} & \text{nếu } P_i<\alpha<P_{i+1}\end{cases}\qquad\textbf{(2.62)}$$

2.4.3.2. 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.

2.4.4. Mốt

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

- Mốt của biến ngáu nhiên rời rạc X rời rạc có phân bố:

X	x_1	x_2	…
P	p_1	p_2	…

$$x_{i_0} = \text{Mod}\, X \quad \Leftrightarrow \quad p_{i_0} = \max\{p_1, p_2, ...\} \qquad \textbf{(2.63)}$$

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

$$c = \text{Mod}\, X \quad \Leftrightarrow \quad f(c) = \max\{f(x),\, x \in \mathbb{R}\}. \qquad \textbf{(2.64)}$$

Ví dụ 2.18: Biến ngẫu nhiên X ở ví dụ 2.3 có Mốt và trung vị $\text{Mod}X = \text{Med}X = 1$.

Ví dụ 2.19: 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 2.20: Tìm $\text{Med}\, X$ và $\text{Mod}\, X$ của biến ngẫu nhiên liên tục X xét trong ví dụ 2.5

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

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

Ví dụ 2.21: Tìm $\text{Med}\, X$ và $\text{Mod}\, X$ của biến ngẫu nhiên liên tục X có hàm mật độ xác suất xác định như sau

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

Giải: Hàm phân bố xác suất

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

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

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

- Một của phân bố nhị thức

$$X\sim \mathscr{B}(n;p) \text{ thì } \operatorname{mod} X=\left[(n+1)p\right] \tag{2.65}$$

- Một của phân bố Poisson tham số $\lambda>0$

$$X\sim \mathscr{P}(\lambda) \text{ thì } \operatorname{mod} X=\left[\lambda\right]. \tag{2.66}$$

2.4.5. Moment, hệ số bất đối xứng, hệ số nhọn (*)

1) Moment cấp k $\quad m_k=\mathrm{E}X^k$; $k=1,2,\ldots$ **(2.67)**

2) Moment quy tâm cấp k $\quad \mu_k=\mathrm{E}\left(X-\mathrm{E}X\right)^k$; $k=1,2,\ldots$ **(2.68)**

3) Hệ số bất đối xứng $\quad \alpha_3=\dfrac{\mu_3}{\sigma^3}$ với $\sigma=\sqrt{\mathrm{D}X}$. **(2.69)**

4) Hệ số nhọn $\quad \alpha_4=\dfrac{\mu_4}{\sigma^4}$. **(2.70)**

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 xác suất sẽ lệch về bên trái hơn.

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

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

- Hệ số nhọn α_4 đặc trưng cho độ nhọn của đồ thị hàm mật độ xác suất so với đồ thị hàm mật độ xác suất 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 xác suất sẽ nhọn hơn so với đồ thị hàm mật độ xác suất chuẩn.

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

2.5. HÀM ĐẶC TRƯNG (*)

2.5.1. Định nghĩa

Hàm đặc trưng biến ngẫu nhiên X ký hiệu là $\varphi_X(t)$ và được định nghĩa bởi biểu thức:

$$\varphi_X(t) = \mathrm{E}\left(e^{itX}\right);\ i \text{ là đơn vị ảo thỏa mãn } i^2 = -1. \quad \textbf{(2.71)}$$

- Hàm đặc trưng của biến ngẫu nhiên X rời rạc có bảng phân bố xác suất:

X	x_1	x_2	…
P	p_1	p_2	…

$$\varphi_X(t) = \mathrm{E}\left(e^{itX}\right) = \sum_k e^{itx_k} p_k . \quad \textbf{(2.72)}$$

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

$$\varphi_X(t) = \mathrm{E}\left(e^{itX}\right) = \int_{-\infty}^{\infty} e^{itx} f_X(x)dx \quad \textbf{(2.73)}$$

2.5.2. Tính chất

1) Hàm đặc trưng là biến đổi Fourier ngược của hàm mật độ xác suất $\varphi_X(t) = \mathscr{F}^{-1}\{f_X(x)\}$. Vì vậy hàm đặc trưng có các tính chất như biến đổi Fourier.

2) $|\varphi_X(t)| \le \varphi_X(0) = 1$.

3) $\varphi_X(t)$ xác định không âm.

4) Nếu biến ngẫu nhiên X có môment cấp k thì $\varphi_X(t)$ có đạo hàm đến cấp k và

$$\mathrm{E}\left[X^k\right] = \frac{1}{i^k}\varphi_X^{(k)}(0). \quad \textbf{(2.74)}$$

5) Các biến ngẫu nhiên $X_1, X_2, ..., X_n$ độc lập khi và chỉ khi

$$\varphi_{X_1+X_2+\cdots+X_n}(t) = \varphi_{X_1}(t)\cdots\varphi_{X_n}(t)\,. \qquad \textbf{(2.75)}$$

6) Nếu hàm đặc trưng $\varphi_X(t)$ có biến đổi Fourier (khả tích tuyệt đối và thỏa mãn điều kiện Dirichlet) thì hàm mật độ xác suất được tính theo công thức:

$$f_X(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\varphi_X(t)dt\,. \qquad \textbf{(2.76)}$$

2.5.3. Các đặc trưng của các quy luật phân bố xác suất thường gặp

Phân bố xác suất của X	Kỳ vọng	Phương sai	Hàm đặc trưng
Nhị thức $\mathscr{B}(n;p)$	$\mathrm{E}X = np$	$\mathrm{D}X = npq$	$\left(pe^{it}+q\right)^n$
Poisson $\mathscr{P}(\lambda)$	$\mathrm{E}X = \lambda$	$\mathrm{D}X = \lambda$	$e^{\lambda\left(e^{it}-1\right)}$
Đều $\mathbf{U}(a,b)$	$\mathrm{E}X = \dfrac{a+b}{2}$	$\mathrm{D}X = \dfrac{(b-a)^2}{12}$	$\dfrac{e^{ibt}-e^{iat}}{it(b-a)}$
Phân bố mũ tham số $\lambda > 0$	$\mathrm{E}X = \dfrac{1}{\lambda}$	$\mathrm{D}X = \dfrac{1}{\lambda^2}$	$\dfrac{\lambda}{\lambda - it}$
Phân bố chuẩn $\mathbf{N}(\mu;\sigma^2)$	$\mathrm{E}X = \mu$	$\mathrm{D}X = \sigma^2$	$e^{i\mu t - \frac{1}{2}\sigma^2 t^2}$
Phân bố Erlang tham số $(k;\lambda)$	$\mathrm{E}X = \dfrac{k}{\lambda}$	$\mathrm{D}X = \dfrac{k}{\lambda^2}$	$\dfrac{\lambda^k}{(\lambda - it)^k}$

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 nhận 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 Kỳ vọng của biến ngẫu nhiên rời rạc là giá trị nó lấy thường xuyên nhất.

Đúng ☐ Sai ☐.

2.5 Kỳ vọng của tổng hai biến ngẫu nhiên luôn luôn bằng tổng các kỳ vọng của nó.

Đúng ☐ Sai ☐.

2.6 Hai biến ngẫu nhiên có cùng kỳ vọng sẽ có cùng phương sai.

Đúng ☐ Sai ☐.

2.7 Phương sai của tổng hai biến ngẫu nhiên rời rạc luôn luôn bằng tổng phương sai của nó.

Đúng ☐ Sai ☐.

2.8 Biến ngẫu nhiên tồn tại phương sai thì cũng tồn tại kỳ vọng.

Đúng ☐ Sai ☐.

2.9 Hàm mật độ xác suất $f(x)$ của biến ngẫu nhiên liên tục có tính chất $f(x) \ge 0$.

Đúng ☐ Sai ☐.

2.10 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 ☐.

2.11 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 ☐.

2.12 Nếu X là biến ngẫu nhiên có phân bố Poisson tham số $\lambda > 0$ thì kỳ vọng, phương sai và Mốt của X đều bằng λ.

Đúng ☐ Sai ☐.

2.13 Nếu biến ngẫu nhiên X phân bố theo quy luật chuẩn $\mathbf{N}(\mu;\sigma^2)$ 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 ☐.

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

Đúng ☐ Sai ☐.

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

Đúng ☐ Sai ☐.

2.16 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.

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

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

2.19 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 $\mathrm{E}(\overline{X})$; $\mathrm{D}(\overline{X})$.

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

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

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

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

2.22 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.

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

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

2.24 Tuổi thọ của một loài côn trùng nào đó là một biến ngẫu nhiên X (đơn vị là tháng) với hàm mật độ xác suất 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 $\mathrm{E}X$, $\mathrm{D}X$.

2.25 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 $\mathrm{E}Y$.

2.26 Một xí nghiệp có hai ôtô vận tải hoạt động. Xác suất trong ngày làm việc các ôtô bị hỏng tương ứng bằng 0,1 và 0,2. Gọi X là số ôtô bị hỏng trong thời gian làm việc. Lập bảng phân bố xác suất, tính kỳ vọng EX và phương sai DX của X.

2.27 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 .

2.28 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.

2.29 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.

2.30 Hai kiện tướng bóng bàn ngang sức thi đấu với nhau. Hỏi thắng 2 trong 4 ván dễ hơn hay thắng 3 trong 6 ván dễ hơn.

2.31 Một nữ công nhân quản lý 12 máy dệt. Xác suất để mỗi máy trong khoảng thời gian T cần đến sự chăm sóc của nữ công nhân là 1/3. Tính xác suất:

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

b) Trong khoảng thời gian T có từ 3 đến 6 máy cần đến sự chăm sóc của nữ công nhân.

2.32 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.

2.33 Xác suất để sản phẩm sản xuất ra bị hỏng bằng 0,1.

a) Tìm xác suất để trong 5 sản phẩm sản xuất ra có không quá 2 sản phẩm hỏng.

b) Tìm số sản phẩm hỏng trung bình trong 5 sản phẩm đó.

c) Tìm số sản phẩm hỏng có khả năng xảy ra nhiều nhất.

2.34 Một bài thi trắc nghiệm gồm có 10 câu hỏi, mỗi câu hỏi có 5 phương án trả lời, trong đó chỉ có một phương án đúng. Giả sử mỗi câu trả lời đúng được 4 điểm và câu trả lời sai bị trừ 2 điểm. Một học sinh kém làm bài bằng cách chọn hú hoạ một phương án cho mỗi câu hỏi. Tính xác suất để:

a) Anh ta được 4 điểm.

b) Anh ta bị điểm âm.

2.35 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 suât:

a) Thu được tín hiệu đúng 2 lần.

b) Thu được tín hiệu nhiều nhất 1 lần.

c) Thu được tin.

2.36 Một cầu thủ nổi tiếng về đá phạt đền, xác suất đá vào gôn là 4/5. Có người cho rằng cứ "sút" 5 quả thì chắc chắn rằng có 4 quả vào lưới. Điều khẳng định đó có đúng không? Tìm xác suất để trong 5 lần sút có đúng 4 lần bóng vào lưới.

2.37 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.

2.38 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).

2.39 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).

2.40 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 $\sigma = 100$ gam. Sản phẩm được coi là đạt tiêu chuẩn kỹ thuật nếu trọng lượng của nó đạt từ 98 đến 102gam.

a) Tìm tỷ lệ sản phẩm đạt tiêu chuẩn kỹ thuật của nhà máy.

b) Tìm tỷ lệ phế phẩm của nhà máy.

c) Giải thích bằng đồ thị kết quả tìm được ở phần a).

2.41 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.

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

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

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 III: VÉC TƠ NGẪU NHIÊN

GIỚI THIỆU

Véc tơ ngẫu nhiên là một bộ có thứ tự bao gồm nhiều biến ngẫu nhiên, mỗi biến ngẫu nhiên là một thành phần của véc tơ ngẫu nhiên. Số biến ngẫu nhiên thành phần gọi là chiều của véc tơ ngẫu nhiên.

Tương tự biến ngẫu nhiên, quy luật phân bố xác suất của véc tơ ngẫu nhiên được khảo sát thông qua hàm phân bố xác suất. Trường hợp véc tơ ngẫu nhiên có các biến ngẫu nhiên thành phần rời rạc được gọi là véc tơ ngẫu nhiên rời rạc. Nếu tất cả các biến ngẫu nhiên thành phần liên tục thì véc tơ ngẫu nhiên tương ứng gọi là liên tục. Véc tơ ngẫu nhiên rời rạc được xác định bởi bảng phân bố xác suất đồng thời, còn véc tơ ngẫu nhiên liên tục được xác định bởi hàm mật độ xác suất đồng thời.

Với véc tơ ngẫu nhiên hai chiều rời rạc ta có bảng phân bố xác suất đồng thời, đó là bảng ghi các giá trị của hai biến ngẫu nhiên thành phần theo hàng, theo cột và xác suất tương ứng. 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. Tương tự, từ hàm mật độ xác suất đồng thời ta có thể tìm hàm mật độ xác suất của biến ngẫu nhiên thành phần bằng cách lấy tích phân theo các biến thích hợp.

Ngoài các đặc trưng kỳ vọng, phương sai của các biến ngẫu nhiên thành phần, các véc tơ ngẫu nhiên 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. Dựa vào đặc trưng này ta có thể xây dựng hàm hồi quy tương quan. Hai biến ngẫu nhiên độc lập thì không tương quan.

Tính độc lập của các biến ngẫu nhiên được nhận biết thông qua dấu hiệu của bảng phân bố xác suất, hàm phân bố xác suất hoặc hàm mật độ xác suất.

Đối với biến ngẫu nhiên không độc lập, bằng cách 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.

Từ các biến ngẫu nhiên $X_1, X_2, ..., X_n$ và hàm nhiều biến $\varphi(x_1, ..., x_n)$ ta có thể xây dựng các biến ngẫu nhiên mới $\varphi(X_1, ..., X_n)$ đó là hàm của các biến ngẫu nhiên đã cho. Ta có thể xác định bảng phân bố xác suất hay hàm mật độ xác suất của $\varphi(X_1, ..., X_n)$.

Để học tốt chương này học viên cần nắm vững các tính chất cơ bản của xác suất, xác suất có điều kiện, phân bố xác suất của biến ngẫu nhiên rời rạc; Tích phân suy rộng, hàm nhiều biến.

NỘI DUNG

3.1. KHÁI NIỆM VÉC TƠ NGẪU NHIÊN

3.1.1. Khái niệm

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 véc tơ ngẫu nhiên. Các biến ngẫu nhiên hai chiều, ba chiều, …, n chiều còn được gọi là véc tơ ngẫu nhiên hai chiều, ba chiều, …, n chiều.

Định nghĩa 3.1: *Một véc tơ ngẫu nhiên n chiều là một bộ có thứ tự $(X_1, X_2, ..., X_n)$ với các thành phần $X_1, X_2, ..., X_n$ là các biến ngẫu nhiên. Ta ký hiệu véc tơ 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 3.1: Một nhà máy sản xuất một loại sản phẩm. Nếu kích thước của sản phẩm được đo bằng chiều dài X và chiều rộng Y thì ta có biến ngẫu nhiên hai chiều, còn nếu xét thêm cả chiều cao Z nữa thì ta có biến ngẫu nhiên ba chiều. Nếu ta chỉ quan tâm đến trọng lượng và thể tích của sản phẩm ta cũng được biến ngẫu nhiên hai chiều.

Véc tơ ngẫu nhiên n chiều $(X_1, X_2, ..., X_n)$ là liên tục hay rời rạc nếu tất cả các biến ngẫu nhiên thành phần $X_1, X_2, ..., X_n$ là liên tục hay rời rạc.

3.1.2. Hàm phân bố xác suất

Định nghĩa 3.2: Hàm n biến $F(x_1, x_2, ..., x_n)$ xác định bởi:

$$F(x_1, x_2, ..., x_n) = P\{X_1 < x_1, X_2 < x_2, ..., X_n < x_n\} \quad \textbf{(3.1)}$$

được gọi là hàm phân bố xác suất của véc tơ ngẫu nhiên $X = (X_1, X_2, ..., X_n)$ hay được gọi là hàm phân bố xác suất đồng thời của các biến ngẫu nhiên $X_1, X_2, ..., X_n$.

Hàm phân bố xác suất đồng thời có các tính chất:

1. $0 \le F(x_1, x_2, ..., x_n) \le 1$. **(3.2)**
2. $\lim\limits_{x_k \to -\infty} F(x_1, x_2, ..., x_n) = 0$, với k nào đó thuộc $\{1, ..., n\}$. **(3.3)**
3. $\lim\limits_{(x_1, ..., x_n) \to (\infty, ..., \infty)} F(x_1, x_2, ..., x_n) = 1$. **(3.4)**
4. $F(x_1, x_2, ..., x_n)$ không giảm theo từng biến. **(3.5)**
5. $\lim\limits_{x_k \to \infty} F(x_1, ..., x_k, ..., x_n) = P\{X_1 < x_1, ..., X_{k-1} < x_{k-1}, X_{k+1} < x_{k+1}, ..., X_n < x_n\}$

6. Đặc biệt $F(x,y)$ là hàm phân bố xác suất của véc tơ ngẫu nhiên hai chiều (X,Y) thì:

$$\lim_{y\to\infty} F(x,y) = P\{X < x\} = F_X(x);\ \lim_{x\to\infty} F(x,y) = P\{Y < y\} = F_Y(y) \qquad \textbf{(3.6)}$$

$F_X(x)$, $F_Y(y)$ là các hàm phân bố xác suất của biến ngẫu nhiên X, Y và được gọi là các hàm phân bố xác suất thành phần của véc tơ ngẫu nhiên (X,Y), hay còn gọi là hàm phân bố xác suất biên của hàm phân bố xác suất đồng thời $F(x,y)$.

3.2. BẢNG PHÂN BỐ XÁC SUẤT CỦA VÉC TƠ NGẪU NHIÊN RỜI RẠC HAI CHIỀU

3.2.1. 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 X theo cột, Y theo hàng 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:

$$p(x_i,y_j) = P\{X = x_i, Y = y_j\}, \qquad \textbf{(3.7)}$$

các xác suất này thỏa mãn

$$\begin{cases} p(x_i,y_j) \ge 0,\ \forall i=\overline{1,n},\ j=\overline{1,m} \\ \displaystyle\sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i,y_j) = 1 \end{cases} \qquad \textbf{(3.8)}$$

3.2.2. 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 2.6) ta có:

$$p(y_j)=P\{Y=y_j\}=\sum_{i=1}^{n}P\{X=x_i, Y=y_j\}=\sum_{i=1}^{n}p(x_i,y_j) \;\;;\;\; j=\overline{1,m} \qquad \textbf{(3.9)}$$

$$p(x_i)=P\{X=x_i\}=\sum_{j=1}^{m}P\{X=x_i, Y=y_j\}=\sum_{j=1}^{m}p(x_i,y_j) \;\;;\;\; i=\overline{1,n} \qquad \textbf{(3.10)}$$

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	…	x_i	…	x_n
P	$p(x_1)$	$p(x_2)$	…	$p(x_i)$	…	$p(x_n)$

Y	y_1	y_2	…	y_j	…	y_m
P	$p(y_1)$	$p(y_2)$	…	$p(y_j)$	…	$p(y_m)$

Ví dụ 3.2: 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

Phân bố xác suất của hai biến ngẫu nhiên thành phần:

Cộng các xác suất theo hàng ta được:

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

Cộng các xác suất theo cột ta được:

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

Ví dụ 3.3: 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:

Y \ X	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

Ví dụ 3.4: (Phân bố đa thức) Véc tơ ngẫu nhiên n chiều $X=(X_1,X_2,...,X_n)$ được gọi là có phân bố đa thức với các tham số $N;p_1,...,p_n$ ký hiệu $X\sim \text{MUT}(N;p_1,...,p_n)$ nếu:

$$P\{X_1=k_1,...,X_n=k_n\}=\frac{N!}{k_1!k_2!...k_{n+1}!}p_1^{k_1}p_2^{k_2}\cdots p_n^{kn}q^{k_{n+1}}$$

trong đó $0\le k_i\le N,\ i=1,...,n+1;\ k_{n+1}=N-(k_1+\cdots+k_n);\ q=1-(p_1+\cdots+p_n)$.

Trường hợp $n=1$ ta được phân bố nhị thức.

Xét N phép thử độc lập, thuần nhất mỗi phép thử có $n+1$ kết quả; giả sử xác suất xuất hiện kết quả thứ k là p_k thì $p_1+\cdots+p_n+p_{n+1}=1$.

Gọi X_k là số thành công của kết quả thứ k trong N phép thử thì $X=(X_1,X_2,...,X_n)$ có phân bố đa thức $X\sim \text{MUT}(N;p_1,...,p_n)$.

Gieo xúc xắc cân đối 10 lần. Tính các xác suất:

1) Có đúng 3 lần xuất hiện mặt 5 chấm (biến cố A).

2) Có 2 lần xuất hiện mặt 1 chấm, 4 lần mặt 3 chấm, 1 lần mặt 4 chấm và 3 lần mặt 6 chấm (biến cố B).

Giải: 1) Gọi X là số lần xuất hiện mặt 5 trong 10 lần thử thì X là số lần thành công trong 10 phép thử Bernoulli với xác suất thành công của mỗi lần thử là 1/6. Vậy $X\sim\mathscr{B}(10;1/6)$ và

$$P(A)=P\{X=3\}=C_{10}^3\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^7=0{,}155.$$

2) Gọi X_k là số lần xuất hiện mặt k trong 10 phép thử thì (X_1,X_2,X_3,X_4,X_6) có phân bố đa thức $\text{MUT}(10;1/6,\ 1/6,\ 1/6,\ 1/6,\ 1/6)$.

$$P(B)=P\{X_1=2,X_2=0,X_3=4,X_4=1,X_6=3\}=\frac{10!}{2!0!4!1!3!}\left(\frac{1}{6}\right)^{10}\approx 0{,}0002.$$

3.3. VÉC TƠ NGẪU NHIÊN LIÊN TỤC

3.3.1. Hàm mật độ xác suất của véc tơ ngẫu nhiên liên tục

Định nghĩa 3.3: *Hàm mật độ xác suất của véc tơ ngẫu nhiên liên tục* $X=\left(X_1,X_2,...,X_n\right)$ *là hàm* n *biến* $f(x_1,x_2,...,x_n)\ge 0$ *thoả mãn:*

$$F(x_1,x_2,...,x_n)=P\left\{X_1<x_1,X_2<x_2,...,X_n<x_n\right\}=\int\limits_{-\infty}^{x_1}\cdots\int\limits_{-\infty}^{x_n}f(t_1,t_2,...,t_n)\,dt_1dt_2...\,dt_n \quad \textbf{(3.11)}$$

$f(x_1,x_2,...,x_n)$ *còn được gọi là hàm mật độ xác suất đồng thời của* $X_1,X_2,...,X_n$.

Tính chất: Để đơn giản cho cách biểu diễn ta xét trường hợp véc tơ ngẫu nhiên hai chiều (X,Y) có hàm mật độ xác suất $f(x,y)$.

1) $f(x,y)\ge 0$ với mọi (x,y) và $\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty}f(x,y)dxdy=1$. **(3.12)**

2) $P\left\{(X,Y)\in A\right\}=\iint\limits_{(x,y)\in A}f(x,y)dxdy$ với $A\subset\mathbb{R}^2$. **(3.13)**

3) $f(x,y)=\begin{cases}\dfrac{\partial^2}{\partial x\partial y}F(x,y) & \text{nếu tồn tại đạo hàm tại } (x,y)\\ 0 & \text{nếu ngược lại}\end{cases}$ **(3.14)**

4) $\int\limits_{-\infty}^{\infty}f(x,y)\,dy=f_X(x)$ hàm mật độ xác suất của biến ngẫu nhiên X. **(3.15)**

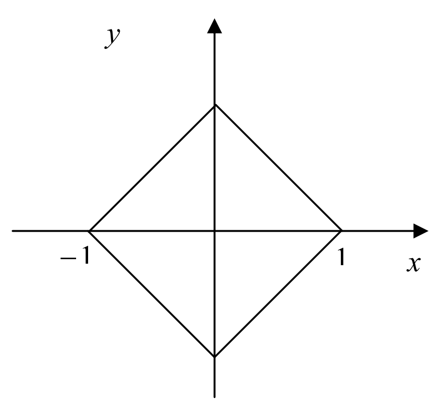
$\int\limits_{-\infty}^{\infty}f(x,y)\,dx=f_Y(y)$ hàm mật độ xác suất của biến ngẫu nhiên Y.

Ví dụ 3.5: Cho véc tơ ngẫu nhiên (X,Y) có hàm mật độ xác suất xác định như sau:

$$f(x,y)=\begin{cases}C & \text{nếu } |x|+|y|\le 1\\ 0 & \text{nếu ngược lại}\end{cases}.$$

a) Tìm C.

b) Tìm các hàm mật độ xác suất của các biến ngẫu nhiên X, Y.



Giải: a) Miền D: $|x|+|y|\le 1$ đối xứng qua hai trục toạ độ Ox, Oy. Phần của D nằm trong góc phần tư thứ nhất là tam giác vuông cân $0\le x, 0\le y$; $x+y\le 1$.

Vậy D là hình vuông có độ dài cạnh bằng $\sqrt{2}$, do đó:

$$1=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy = C\,\mathrm{dt}\,\mathrm{D} = 2C \;\Rightarrow\; C=\frac{1}{2}.$$

b) $|x|+|y|\le 1 \Leftrightarrow |x|-1\le y\le |x|+1 \Leftrightarrow \begin{cases} x-1\le y\le 1-x & \text{nếu } 0\le x\le 1\\ -x-1\le y\le x+1 & \text{nếu } -1\le x\le 0\end{cases}$

$$f_X(x)=\begin{cases}\dfrac{1}{2}\displaystyle\int_{-x-1}^{1+x} dy = 1+x & \text{nếu } -1\le x\le 0\\ 0 & \text{nếu ngược lại}\\ \dfrac{1}{2}\displaystyle\int_{x-1}^{1-x} dy = 1-x & \text{nếu } 0\le x\le 1\end{cases} = \begin{cases}1-|x| & \text{nếu } |x|\le 1\\ 0 & \text{nếu } |x|>1\end{cases}$$

Do tính chất đối xứng của X và Y nên ta cũng có:

$$f_Y(y)=\begin{cases}1-|y| & \text{nếu } |y|\le 1\\ 0 & \text{nếu } |y|>1\end{cases}.$$

3.4. TÍNH ĐỘC LẬP CỦA CÁC BIẾN NGẪU NHIÊN

Định nghĩa 3.4: *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.*

Định lý 3.1: Giả sử $F(x,y)$ là hàm phân bố xác suất của véc tơ ngẫu nhiên (X,Y). Khi đó X,Y độc lập khi và chỉ khi

$$F(x,y)=F_X(x)F_Y(y) \qquad \textbf{(3.16)}$$

trong đó $F_X(x),\ F_Y(y)$ lần lượt là hàm phân bố xác suất của X và Y.

Định lý 3.2: Véc tơ ngẫu nhiên hai chiều rời rạc (X,Y) với phân bố xác suất đồng thời (3.6) là độc lập khi và chỉ khi

$$p(x_i,y_j)=p(x_i)p(y_j)\ \ \forall\, i=\overline{1,n},\ \ j=\overline{1,m}\ . \qquad \textbf{(3.17)}$$

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

- Hai hàng bất kỳ tỉ lệ.
- Hai cột bất kỳ tỉ lệ. **(3.18)**

Ví dụ 3.6: Hai biến ngẫu nhiên X và Y của ví dụ 3.2 không độc lập vì $P\{X=2\}=\dfrac{2}{8}$, $P\{Y=1\}=\dfrac{3}{8}$ nhưng $P\{X=2,Y=1\}=0$.

Hai biến ngẫu nhiên X và Y của ví dụ 3.3 độc lập vì thỏa mãn điều kiện (3.17).

Định lý 3.3: Giả sử véc tơ ngẫu nhiên (X,Y) có hàm mật độ xác suất $f(x,y)$. Khi đó X,Y độc lập khi và chỉ khi

$$f(x,y)=f_X(x)f_Y(y) \qquad \textbf{(3.19)}$$

trong đó $f_X(x)$, $f_Y(y)$ lần lượt là hàm mật độ xác suất của X và Y.

Ví dụ 3.7: Véc tơ ngẫu nhiên (X,Y) của ví dụ 3.5 không độc lập vì hàm mật độ xác suất $f(x,y)\neq f_X(x)f_Y(y)$.

3.5. HÀM CỦA CÁC BIẾN NGẪU NHIÊN (*)

3.5.1. Hàm của một biến ngẫu nhiên

Giả sử X là một biến ngẫu nhiên và $\varphi(x)$ là hàm liên tục một biến số. Khi đó $Y=\varphi(X)$ cũng là một biến ngẫu nhiên gọi là hàm của biến ngẫu nhiên X.

a. Trường hợp rời rạc:

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

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

thì Y cũng rời rạc có bảng phân bố:

Y	$\varphi(x_1)$	$\varphi(x_2)$	$\cdots$
P	p_1	p_2	$\cdots$

b. Trường hợp liên tục:

Nếu X liên tục có hàm phân bố xác suất $F_X(x)$ và hàm mật độ xác suất $f_X(x)$ thì bằng phương pháp giải tích ta có thể tìm được hàm phân bố xác suất và hàm mật độ xác suất của biến ngẫu nhiên $Y=\varphi(X)$.

❖ Trường hợp $\varphi(x)$ là một song ánh. Khi đó $\varphi(x)$ hoặc đơn điệu tăng hoặc đơn điệu giảm:

♦ $\varphi(x)$ đơn điệu tăng

$$F_Y(y)=P\{\varphi(X)<y\}=P\{X<\varphi^{-1}(y)\}=F_X\left(\varphi^{-1}(y)\right) \;\Rightarrow\; F_X(x)=F_Y\left(\varphi(x)\right)$$

$$f_X(x)=f_Y\left(\varphi(x)\right)\varphi'(x) \text{ hay } f_Y(y)=\frac{f_X\left(\varphi^{-1}(y)\right)}{\varphi'\left(\varphi^{-1}(y)\right)}=\frac{f_X\left(\varphi^{-1}(y)\right)}{\left|\varphi'\left(\varphi^{-1}(y)\right)\right|}=f_X(x)\left|\frac{dx}{dy}\right|. \qquad \textbf{(3.20)}$$

♦ $\varphi(x)$ đơn điệu giảm

$$F_Y(y) = P\{\varphi(X) < y\} = P\{X > \varphi^{-1}(y)\} = 1 - F_X\left(\varphi^{-1}(y)\right) \Rightarrow F_X(x) = 1 - F_Y\left(\varphi(x)\right)$$

$$f_X(x) = -f_Y\left(\varphi(x)\right)\varphi'(x) \text{ hay } f_Y(y) = -\frac{f_X\left(\varphi^{-1}(y)\right)}{\varphi'\left(\varphi^{-1}(y)\right)} = \frac{f_X\left(\varphi^{-1}(y)\right)}{\left|\varphi'\left(\varphi^{-1}(y)\right)\right|} = f_X(x)\left|\frac{dx}{dy}\right|. \quad \textbf{(3.21)}$$

❖ Trường hợp $\varphi(x)$ là một hàm liên tục bất kỳ (không đơn điệu) ta tính trực tiếp xác suất $P\{Y < y = \varphi(x)\}$, từ đó suy ra hàm phân bố xác suất và hàm mật độ xác suất của Y.

Ví dụ 3.8: Giả sử $\varphi_1(x) = x^3$, $\varphi_2(x) = x^2$, $\varphi_3(x) = e^{-x}$. X là một biến ngẫu nhiên liên tục có hàm phân bố xác suất $F_X(x)$ và hàm mật độ xác suất $f_X(x)$:

a) Tìm hàm mật độ xác suất của $Y = \varphi_1(X)$.

b) Tìm hàm mật độ xác suất của $Z = \varphi_2(X)$.

c) Tìm hàm mật độ xác suất của $T = \varphi_3(X)$.

d) Tìm hàm mật độ xác suất $f_Y(y)$ và hàm phân bố xác suất $F_Y(y)$ và xác suất biến cố $P\{2,5 < Y < 5\}$ nếu hàm mật độ xác suất $f_X(x)$ của X xác định bởi:

$$f_X(x) = \begin{cases} \dfrac{2}{x^2} & \text{nếu } 1 \le x \le 2 \\ 0 & \text{nếu ngược lại} \end{cases}$$

Giải: a. $y = \varphi_1(x) = x^3$ đơn điệu tăng. $x = \sqrt[3]{y}$, $\dfrac{dx}{dy} = \dfrac{1}{3\sqrt[3]{y^2}}$. Áp dụng công thức (3.20) ta có:

$$f_Y(y) = f_X(x)\frac{dx}{dy} = \frac{1}{3}\frac{f_X(\sqrt[3]{y})}{\sqrt[3]{y^2}}.$$

b. Với $z > 0$ ta có:

$$F_Z(z) = P\{Z < z\} = P\{X^2 < z\} = P\{-\sqrt{z} < X < \sqrt{z}\} = F_X(\sqrt{z}) - F_X(-\sqrt{z}).$$

Với $z \le 0$ thì $P\{Z < z\} = 0$ vì $Z = X^2$ không nhận giá trị âm. Vậy hàm mật độ xác suất của Z.

$$f_Z(z) = \begin{cases} \dfrac{1}{2\sqrt{z}}\left(f_X(\sqrt{z}) + f_X(-\sqrt{z})\right) & \text{nếu } z > 0 \\ 0 & \text{nếu } z \le 0 \end{cases}$$

c. Với $t > 0$ ta có:

$$F_T(t) = P\{T < t\} = P\{e^{-X} < t\} = P\{X > -\ln t\} = 1 - F_X(-\ln t)$$

Khi $t \le 0$ biến cố $\{T < t\}$ là biến cố không thể vì $T = e^{-X}$ chỉ nhận giá trị dương, do đó $P\{T < t\} = 0$. Vậy hàm mật độ xác suất của T xác định như sau:

$$f_T(t) = \begin{cases} \dfrac{f_X(-\ln t)}{t} & \text{nếu } t > 0 \\ 0 & \text{nếu } t \le 0 \end{cases}$$

d. $f_Y(y) = \dfrac{1}{3}\dfrac{f_X(\sqrt[3]{y})}{\sqrt[3]{y^2}} = \begin{cases} \dfrac{2}{3\sqrt[3]{y^4}} & \text{nếu } 1 \le y \le 8 \\ 0 & \text{nếu ngược lại} \end{cases}.$

$$F_Y(y) = \begin{cases} 0 & \text{nếu } y \le 1 \\ 2\left(1 - \dfrac{1}{\sqrt[3]{y}}\right) & \text{nếu } 1 < y \le 8 \\ 1 & \text{nếu } y > 8 \end{cases}.$$

$$P\{2,5 < Y < 5\} = P\left\{2,5 < X^3 < 5\right\} = P\left\{\sqrt[3]{2,5} < X < \sqrt[3]{5}\right\} = \int_{\sqrt[3]{2,5}}^{\sqrt[3]{5}} \frac{2}{x^2}dx = -\frac{2}{x}\bigg|_{\sqrt[3]{2,5}}^{\sqrt[3]{5}} = 2\left(\frac{1}{\sqrt[3]{2,5}} - \frac{1}{\sqrt[3]{5}}\right)$$

hoặc $P\{2,5 < Y < 5\} = F_Y(5) - F_Y(2,5) = 2\left(\dfrac{1}{\sqrt[3]{2,5}} - \dfrac{1}{\sqrt[3]{5}}\right)$.

3.5.2. Hàm của hai biến ngẫu nhiên

Giả sử X, Y là hai biến ngẫu nhiên và $\varphi(x, y)$ là một hàm hai biến liên tục cho trước. Xét biến ngẫu nhiên Z xác định bởi $Z = \varphi(X, Y)$. Khi đó Z là hàm của hai biến ngẫu nhiên X và Y.

a) *Trường hợp rời rạc*:

Nếu X, Y có tập các giá trị $\{x_1, x_2, ..., x_m\}$ và $\{y_1, y_2, ..., y_n\}$ thì $Z = \varphi(X, Y)$ có tập các giá trị là $\left\{\varphi(x_i, y_j) \,\middle|\, i = 1,...m;\, j = 1,..., y_n\right\}$.

Định lý: Giả sử $Z = \varphi(X, Y)$. Khi đó:

$$P\{Z = a\} = \sum_{\{(i,j) | \varphi(x_i, y_j) = a\}} P\left\{X = x_i; Y = y_j\right\}. \qquad \textbf{(3.22)}$$

Ví dụ 3.9: Xét $Z = X + Y$, trong đó véc tơ ngẫu nhiên (X, Y) cho trong ví dụ 3.3.

X	1	1	1	2	2	2	3	3	3
Y	1	2	3	1	2	3	1	2	3
$X+Y$	2	3	4	3	4	5	4	5	6

P	2/36	3/36	1/36	4/36	6/36	2/36	6/36	9/36	3/36

Vậy bảng phân bố xác suất của $Z = X + Y$ xác định như sau

$Z = X + Y$	2	3	4	5	6
P	2/36	7/36	13/36	11/36	3/36

b) *Trường hợp liên tục*:

Giả sử X, Y là hai biến ngẫu nhiên; $u(x,y)$, $v(x,y)$ là hai hàm hai biến cho trước. Khi đó ta có thể thiết lập một cặp biến ngẫu nhiên mới U, V bằng công thức:

$$\begin{cases} U = u(X,Y) \\ V = v(X,Y) \end{cases}$$

Giả sử T là ánh xạ từ $\mathbb{R}^2$ vào $\mathbb{R}^2$ xác định bởi:

$$T(x,y) = \big(u(x,y), v(x,y)\big)$$

Giả thiết T là ánh xạ 1 - 1 (đơn ánh) từ miền $D \subset \mathbb{R}^2$ lên miền $\Delta \subset \mathbb{R}^2$. Khi đó T có ánh xạ ngược T^{-1}: $\Delta \to D$, với mỗi $(u,v) \in \Delta$ tồn tại duy nhất $(x,y) \in D$ sao cho $T(x,y) = (u,v)$, ta ký hiệu $T^{-1}(u,v) = (x,y)$.

Định thức Jacobi của T^{-1} được xác định theo công thức sau:

$$J(u,v) = \frac{D(x,y)}{D(u,v)} = \begin{vmatrix} \dfrac{\partial x}{\partial u} & \dfrac{\partial x}{\partial v} \\ \dfrac{\partial y}{\partial u} & \dfrac{\partial y}{\partial v} \end{vmatrix}$$

và định thức Jacobi của T là: $J(x,y) = \dfrac{D(u,v)}{D(x,y)} = \begin{vmatrix} \dfrac{\partial u}{\partial x} & \dfrac{\partial u}{\partial y} \\ \dfrac{\partial v}{\partial x} & \dfrac{\partial v}{\partial y} \end{vmatrix} = \dfrac{1}{J(u,v)}$.

Định lý 3.4: Giả sử X, Y là hai biến ngẫu nhiên liên tục với hàm mật độ xác suất đồng thời $f_{X,Y}(x,y)$. Ký hiệu:

$$D = \left\{(x,y) \in \mathbb{R}^2 \,\middle|\, f_{X,Y}(x,y) > 0\right\}$$

Giả sử $T(x,y) = \big(u(x,y), v(x,y)\big)$ là ánh xạ 1 - 1 khả vi liên tục từ miền $D \subset \mathbb{R}^2$ lên miền $\Delta \subset \mathbb{R}^2$. Khi đó $U = u(X,Y)$, $V = v(X,Y)$ là hai biến ngẫu nhiên có hàm mật độ xác suất đồng thời

$$f_{U,V}(u,v)=\begin{cases} f_{X,Y}\big(x(u,v),y(u,v)\big)\left|\dfrac{D(x,y)}{D(u,v)}\right| & \text{nếu } (u,v)\in\Delta \\ 0 & \text{nếu ngược lại} \end{cases} \qquad \textbf{(3.23)}$$

Định lý 3.5: Giả sử $V = X + Y$ là tổng của hai biến ngẫu nhiên X , Y .

(i) Nếu X , Y rời rạc thì phân bố xác suất của V :

$$P\{V=v\}=\sum_x P\{X=x;Y=v-x\}=\sum_y P\{Y=y;X=v-y\}. \qquad \textbf{(3.24)}$$

(ii) Nếu X , Y liên tục có hàm mật độ xác suất đồng thời $f_{X,Y}(x,y)$ thì hàm mật độ xác suất của V :

$$f_V(v)=\int_{-\infty}^{\infty} f_{X,Y}(x,v-x)dx. \qquad \textbf{(3.25)}$$

Hệ quả: Giả sử $V = X + Y$ là tổng của hai biến ngẫu nhiên độc lập X , Y .

(i) Nếu X , Y rời rạc thì phân bố xác suất của V :

$$P\{V=v\}=\sum_x P\{X=x\}P\{Y=v-x\}=\sum_y P\{Y=y\}P\{X=v-y\}. \qquad \textbf{(3.26)}$$

(ii) Nếu X , Y liên tục có hàm mật độ xác suất đồng thời $f_X(x)f_Y(y)$ thì hàm mật độ xác suất của V :

$$f_V(v)=\int_{-\infty}^{\infty} f_X(x)f_Y(v-x)dx=f_X(v)*f_Y(v). \qquad \textbf{(3.27)}$$

Như vậy biến ngẫu nhiên là tổng hai biến ngẫu nhiên liên tục X , Y có hàm mật độ xác suất là tích chập của hai hàm mật độ $f_X(v)*f_Y(v)$.

3.6. CÁC THAM SỐ ĐẶC TRƯNG CỦA BIẾN NGẪU NHIÊN

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

Từ bảng phân bố xác suất thành phần (3.9)-(3.10) và hàm mật độ xác suất thành phần (3.15) ta có công thức tính kỳ vọng và phương sai của biến ngẫu nhiên thành phần X , Y của véc tơ ngẫu nhiên (X,Y) :

a. Trường hợp X , Y rời rạc

$$\mathrm{E}X=\sum_{i=1}^{n} x_i p(x_i)=\sum_{i=1}^{n}\sum_{j=1}^{m} x_i p(x_i,y_j) \qquad \textbf{(3.28)}$$

$$\mathrm{E}Y=\sum_{j=1}^{m} y_j p(y_j)=\sum_{j=1}^{m}\sum_{i=1}^{n} y_j p(x_i,y_j) \qquad \textbf{(3.29)}$$

$$\mathrm{D}X = \mathrm{E}X^2 - (\mathrm{E}X)^2 = \sum_{i=1}^{n}\sum_{j=1}^{m} x_i^2 p(x_i, y_j) - (\mathrm{E}X)^2 \qquad \textbf{(3.30)}$$

$$\mathrm{D}Y = \mathrm{E}Y^2 - (\mathrm{E}Y)^2 = \sum_{j=1}^{m}\sum_{i=1}^{n} y_j^2 p(x_i, y_j) - (\mathrm{E}Y)^2 \qquad \textbf{(3.31)}$$

b. Trường hợp X , Y liên tục có hàm mật độ xác suất đồng thời $f_{X,Y}(x,y)$

$$\mathrm{E}X = \int_{-\infty}^{\infty} x f_X(x)dx = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f_{X,Y}(x,y)dxdy \qquad \textbf{(3.32)}$$

$$\mathrm{E}Y = \int_{-\infty}^{\infty} y f_Y(y)dy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y f_{X,Y}(x,y)dxdy \qquad \textbf{(3.33)}$$

$$\mathrm{D}X = \mathrm{E}X^2 - (\mathrm{E}X)^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^2 f_{X,Y}(x,y)dxdy - (\mathrm{E}X)^2 \qquad \textbf{(3.34)}$$

$$\mathrm{D}Y = \mathrm{E}Y^2 - (\mathrm{E}Y)^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y^2 f_{X,Y}(x,y)dxdy - (\mathrm{E}Y)^2 \qquad \textbf{(3.35)}$$

3.6.2. Hiệp phương sai

Định nghĩa 3.5: *Hiệp phương sai (hay còn gọi là Covariance) của hai biến ngẫu nhiên X, Y, ký hiệu $\mathrm{cov}(X,Y)$, là kỳ vọng toán của tích các sai lệch của hai biến ngẫu nhiên đó với kỳ vọng toán của chúng:*

$$\mathrm{cov}(X,Y) = \mathrm{E}\big(X - \mathrm{E}X\big)\big(Y - \mathrm{E}Y\big) = \mathrm{E}\,XY - \mathrm{E}X\,\mathrm{E}Y \qquad \textbf{(3.36)}$$

Nếu X, Y rời rạc có bảng phân bố xác suất đồng thời (3.7) thì

$$\mathrm{cov}(X,Y) = \sum_{j=1}^{m}\sum_{i=1}^{n} x_i y_j p(x_i, y_j) - \mathrm{E}\,X\,\mathrm{E}\,Y \qquad \textbf{(3.37)}$$

Nếu X, Y liên tục có hàm mật độ xác suất đồng thời $f_{X,Y}(x,y)$ thì

$$\mathrm{cov}(X,Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f_{X,Y}(x,y)dxdy - \mathrm{E}\,X\,\mathrm{E}\,Y \qquad \textbf{(3.38)}$$

Tính chất

1) $\mathrm{cov}(X,Y) = \mathrm{cov}(Y,X)$. **(3.39)**

2) $\mathrm{cov}(X,X) = \mathrm{D}X$. **(3.40)**

3) $\mathrm{cov}(aX + c, bY + d) = ab\,\mathrm{cov}(Y,X)$ với mọi hằng số a,b,c,d . **(3.41)**

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

3.6.3. Ma trận hiệp phương sai

Định nghĩa 3.6: Cho véc tơ ngẫu nhiên $X=(X_1,X_2,...,X_n)$.

$$\text{Ma trận } M=\left[C_{ij}\right]_{n\times n} \text{ với } C_{ij}=\operatorname{cov}(X_i,X_j) \quad \textbf{(3.42)}$$

được gọi là ma trận hiệp phương sai (ma trận covariance, ma trận moment) của véc tơ ngẫu nhiên X.

Tính chất

1) Ma trận hiệp phương sai là ma trận đối xứng.

2) Với mọi $t_1,t_2,...,t_n\in\mathbb{R}^n$ luôn có $\sum_j\sum_i C_{ij}t_it_j\ge 0$.

3) Các định thức con chính của M không âm.

Nói cách khác ma trận hiệp phương sai M là ma trận của dạng toàn phương không âm.

3.6.4. Hệ số tương quan

Định nghĩa 3.7: *Hệ số tương quan của hai biến ngẫu nhiên* X,Y *ký hiệu và định nghĩa bởi công thức:*

$$\rho_{X,Y}=\frac{\operatorname{cov}(X,Y)}{\sigma_X\sigma_Y} \quad \textbf{(3.43)}$$

Tính chất

1) $-1\le\rho_{X,Y}\le 1$ với mọi X,Y. **(3.44)**

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

$$\rho_{aX+c,bY+d}=\begin{cases}\rho_{X,Y} & \text{nếu } ab>0\\ -\rho_{X,Y} & \text{nếu } ab<0\end{cases} \quad \textbf{(3.45)}$$

4) $Y=aX+b$, $a\ne 0$ khi và chỉ khi

$$\rho_{X,Y}=\begin{cases}1 & \text{nếu } a>0\\ -1 & \text{nếu } a<0\end{cases} \quad \textbf{(3.46)}$$

Ý nghĩa của hệ số tương quan

Hệ số tương quan đo mức độ phụ thuộc tuyến tính giữa X và Y. Khi $\left|\rho_{X,Y}\right|$ càng gần 1 thì tính chất quan hệ tuyến tính càng chặt, khi $\left|\rho_{X,Y}\right|$ càng gần 0 thì sự phụ thuộc tuyến tính càng ít, càng lỏng lẻo.

Khi $\rho_{X,Y}=0$ ta nói X và Y không tương quan.

3.7. PHÂN BỐ CÓ ĐIỀU KIỆN VÀ KỲ VỌNG CÓ ĐIỀU KIỆN (*)

3.7.1. Phân bố có điều kiện và kỳ vọng có điều kiện của biến ngẫu nhiên X rời rạc với điều kiện B

Định nghĩa 3.8: *Giả sử X biến ngẫu nhiên rời rạc nhận các giá trị $\{x_1, x_2, ..., x_n\}$ và B là một biến cố trong cùng phép thử với X, có xác suất $P(B) > 0$. Khi đó bảng phân bố xác suất có điều kiện của X với điều kiện B.*

$X\|B$	x_1	x_2	…	x_i	…	x_n
P	$p(x_1\|B)$	$p(x_2\|B)$	…	$p(x_i\|B)$	…	$p(x_n\|B)$

trong đó

$$p(x_i|B) = \frac{P(\{X = x_i\}B)}{P(B)} \;;\; i = \overline{1,n}. \tag{3.47}$$

và kỳ vọng của X với điều kiện B:

$$\mathrm{E}\left(X|B\right) = \sum_{i=1}^{n} x_i p(x_i|B) \tag{3.48}$$

3.7.2. Phân bố có điều kiện và kỳ vọng có điều kiện của hai biến ngẫu nhiên X, Y rời rạc

Nếu X, Y có tập các giá trị $\{x_1, x_2, ..., x_n\}$ và $\{y_1, y_2, ..., y_m\}$. Với mỗi y_j, ta có bảng phân bố xác suất có điều kiện của X với điều kiện biến cố $\{Y = y_j\}$:

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

trong đó

$$p(x_i|y_j) = \frac{p(x_i, y_j)}{p(y_j)} \;;\; i = \overline{1,n},\;\; j = \overline{1,m}\;. \tag{3.49}$$

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

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

trong đó

$$p(y_j|x_i)=\frac{p(x_i,y_j)}{p(x_i)}\ ;\ i=\overline{1,n},\ \ j=\overline{1,m}\ . \quad \textbf{(3.50)}$$

Kỳ vọng có điều kiện của X với điều kiện $Y=y_j$ và kỳ vọng có điều kiện của Y với điều kiện $X=x_i$ tương ứng:

$$\mathrm{E}\left(X\middle|Y=y_j\right)=\sum_{i=1}^{n}x_i p(x_i|Y=y_j)\ ;\ \ \mathrm{E}\left(Y\middle|X=x_i\right)=\sum_{j=1}^{m}y_j p(y_j|X=x_i)\,. \quad \textbf{(5.51)}$$

$\mathrm{E}\left(Y\middle|X=x\right)$ *là một hàm của* x*, được gọi là hàm hồi qui của* Y *đối với* X.

$\mathrm{E}\left(X\middle|Y=y\right)$ *là một hàm của* y*, được gọi là hàm hồi qui của* X *đối với* Y.

Ví dụ 3.9: Thực hiện lặp lại nhiều lần phép thử Bernoulli, xác suất xuất hiện của biến cố A trong mỗi lần thử là p. Gọi Y là biến ngẫu nhiên chỉ lần thử đầu tiên xuất hiện biến cố A. Gọi B là biến cố: "Trong n lần thử đầu tiên có duy nhất một lần xuất hiện biến cố A".

a) Tìm bảng phân bố xác suất của Y.

b) Tìm phân bố của Y với điều kiện B.

Giải: a) Ta có bảng phân bố xác suất của Y với $q=1-p$:

Y	1	2	…	k	…
P	p	qp	…	$q^{k-1}p$	…

b) Phân bố của Y với điều kiện B: Khi $P(B)>0$ thì $P\left(\{Y=k\}|B\right)=\dfrac{P\left(\{Y=k\}B\right)}{P(B)}$

- Rõ ràng khi $k>n$ thì biến cố $\{Y=k\}$ kéo theo biến cố trong n phép thử đầu tiên đồng tiền không xuất hiện mặt ngửa. Do đó $P\left(\{Y=k\}B\right)=0$.
- Khi $k\le n$, áp dụng công thức Bernoulli ta có: $P(B)=C_n^1pq^{n-1}=npq^{n-1}$.

Mặt khác $P\left(\{Y=k\}B\right)=P\{\text{chỉ xuất hiện biến cố } A \text{ ở lần thử thứ } k\}=pq^{n-1}$

Vậy

$$P\left(\{Y=k\}|B\right)=\begin{cases}\dfrac{1}{n} & \text{nếu } \ k\le n\\ 0 & \text{nếu } \ k>n\,.\end{cases}$$

Ví dụ 3.10: 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.

$X \backslash Y$	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:

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

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

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.

3.7.3. Phân bố có điều kiện và kỳ vọng có điều kiện của hai biến ngẫu nhiên X, Y liên tục

Định nghĩa 3.10: *Cho hai biến ngẫu nhiên X, Y. Xác suất của biến cố $\{Y < y\}$ với điều kiện $\{X = x\}$ xác định một hàm số của y phụ thuộc tham số x được định nghĩa là hàm phân bố xác suất của Y với điều kiện $\{X = x\}$.*

$$F_{Y|X}(y \mid x) = P\left\{Y < y \middle| X = x\right\}. \tag{3.52}$$

Định lý 3.6: Nếu X, Y liên tục có hàm mật độ xác suất đồng thời là $f(x, y)$ và X có mật độ $f_X(x)$ thì

$$F_{Y|X}(y \mid x) = \int_{-\infty}^{y} \frac{f(x,v)}{f_X(x)} dv \text{, nếu } f_X(x) > 0. \tag{3.53}$$

Đạo hàm của hàm phân bố xác suất có điều kiện $\frac{d}{dy}F_{Y|X}(y\,|\,x)$ được gọi là hàm mật độ xác suất có điều kiện của Y với điều kiện $X = x$ và ký hiệu là $f_{Y|X}(y\,|\,x)$. (Đây là hàm của biến y còn x đóng vai trò là tham số). Từ định lý trên ta có

$$f_{Y|X}(y\,|\,x) = \frac{f(x,y)}{f_X(x)} \text{ với mọi } y \in \mathbb{R} \text{ và } x \text{ thoả mãn } f_X(x) > 0. \qquad \textbf{(3.54)}$$

Tương tự ta có hàm phân bố xác suất và hàm mật độ xác suất có điều kiện của X với $Y = y$

$$F_{X|Y}(x\,|\,y) = \int_{-\infty}^{x} \frac{f(u,y)}{f_Y(y)}du \text{, nếu } f_Y(y) > 0. \qquad \textbf{(3.55)}$$

$$f_{X|Y}(x\,|\,y) = \frac{f(x,y)}{f_Y(y)} \text{ với mọi } x \in \mathbb{R} \text{ và } y \text{ thoả mãn } f_Y(y) > 0. \qquad \textbf{(3.56)}$$

Ví dụ 3.11: Nếu X, Y là hai biến ngẫu nhiên liên tục có hàm mật độ xác suất đồng thời đã cho trong ví dụ 3.5 thì

$$f_{X|Y}(x|y) = \begin{cases} \dfrac{1}{2(1-|y|)} & \text{nếu } |y| \le 1 \text{ và } |x| + |y| \le 1 \\ 0 & \text{nếu ngược lại} \end{cases}.$$

Định nghĩa 3.11: *Giả sử hai biến ngẫu nhiên X, Y có $f_{Y|X}(y\,|\,x)$ là hàm mật độ xác suất có điều kiện của Y với điều kiện $X = x$. Khi đó kỳ vọng của Y với điều kiện $X = x$ được ký hiệu và định nghĩa theo công thức sau*

$$\mathrm{E}\left(Y\middle|X = x\right) = \int_{-\infty}^{\infty} y f_{Y|X}(y|x)dy \qquad \textbf{(3.57)}$$

Định lý sau đây cho công thức liên hệ giũa kỳ vọng và kỳ vọng có điều kiện tương tự công thức (2.49) và công thức xác suất đầy đủ (1.16).

Định lý 3.7: $\mathrm{E}(Y) = \mathrm{E}\left(\mathrm{E}\left(Y\middle|X = x\right)\right)$, cụ thể:

- Nếu X rời rạc với phân bố xác suất $p(x_i) = P\{X = x_i\}$ thì

$$\mathrm{E}(Y) = \sum_i \mathrm{E}\left(Y\middle|X = x_i\right)p(x_i) \qquad \textbf{(3.58)}$$

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

$$\mathrm{E}(Y) = \int_D \mathrm{E}\left(Y\middle|X = x\right)f_X(x)dx, \qquad \textbf{(3.59)}$$

trong đó $D = \{x \in \mathbb{R} \,|\, f_X(x) > 0\}$.

3.8. PHÂN BỐ CHUẨN NHIỀU CHIỀU(*)

3.8.1. Khái niệm phân bố chuẩn n chiều

Định nghĩa 3.12: *Véc tơ ngẫu nhiên $\left(X_1, X_2, ..., X_n\right)$ được gọi là có phân bố chuẩn n chiều nếu có mật độ đồng thời*

$$f(x)=\frac{1}{\sqrt{(2\pi)^n \det M}}\exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{a})M^{-1}(\boldsymbol{x}-\boldsymbol{a})^t\right\} \qquad \textbf{(3.60)}$$

trong đó $\boldsymbol{x}=(x_1,x_2,...,x_n)\in\mathbb{R}^n$; $\boldsymbol{a}=(\mu_1,\mu_2,...,\mu_n)\in\mathbb{R}^n$; $M=\left[C_{ij}\right]_{n\times n}$ là ma trận đối xứng xác định dương; ma trận cột $(\boldsymbol{x}-\boldsymbol{a})^t$ là chuyển vị của ma trận hàng $(\boldsymbol{x}-\boldsymbol{a})$.

Định lý 3.8: Nếu véc tơ ngẫu nhiên $X=\left(X_1, X_2, ..., X_n\right)$ có phân bố chuẩn n chiều với mật độ đồng thời xác định như trong định nghĩa trên thì

i. Hàm đặc trưng: $\varphi_X(t)=\exp\left(i\left\langle \boldsymbol{t},\boldsymbol{a}\right\rangle-\frac{1}{2}\left\langle \boldsymbol{t},M^{-1}\boldsymbol{t}^t\right\rangle\right)$; $\boldsymbol{t}=(t_1,t_2,...,t_n)$. **(3.61)**

trong đó $\left\langle \boldsymbol{t},\boldsymbol{a}\right\rangle=t_1\mu_1+t_2\mu_2+\cdots+t_n\mu_n$ là tích vô hướng của hai véc tơ $\boldsymbol{t}$ và $\boldsymbol{a}$.

ii. Với mọi $i=1,...,n$: $\mu_i=\mathrm{E}X_i$.

iii. M là ma trận hiệp phương sai của véc tơ ngẫu nhiên $X=\left(X_1, X_2, ..., X_n\right)$:

$$C_{ij}=\operatorname{cov}(X_i,X_j). \qquad \textbf{(3.62)}$$

iv. Các biến ngẫu nhiên $X_1, X_2, ..., X_n$ độc lập khi và chỉ khi chúng không tương quan, nghĩa là M là ma trận chéo.

v. Biến ngẫu nhiên thành phần X_i có phân bố chuẩn $\mathbf{N}(\mu_i;C_{ii})$.

$\left(Y_1, Y_2, ..., Y_n\right)=H\left(X_1, X_2, ..., X_n\right)^t$ cũng có phân bố chuẩn, trong đó H là ma trận vuông cấp n bất kỳ có $\det H\neq 0$.

3.8.2. Phân bố chuẩn hai chiều

Định nghĩa 3.13: *Véc tơ ngẫu nhiên (X,Y) được gọi là có phân bố chuẩn hai chiều nếu có mật độ đồng thời:*

$$f(x,y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left(-\frac{1}{2}Q(x,y)\right), \qquad \textbf{(3.63)}$$

$$Q(x,y)=\frac{1}{(1-\rho^2)}\left\{\left(\frac{x-\mu_1}{\sigma_1}\right)^2-2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right)+\left(\frac{y-\mu_2}{\sigma_2}\right)^2\right\}.$$

Định lý 3.9: Giả sử véc tơ ngẫu nhiên (X,Y) có phân bố chuẩn hai chiều với hàm mật độ xác suất đồng thời xác định như trong định nghĩa 3.13. Khi đó

i. X có phân bố chuẩn $\mathbf{N}(\mu_1;\sigma_1^2)$, Y có phân bố chuẩn $\mathbf{N}(\mu_2;\sigma_2^2)$.

ii. Hệ số tương quan $\rho(X,Y)=\rho$.

iii. X, Y không tương quan khi và chỉ khi X, Y độc lập.

iv. Hàm mật độ xác suất của X với điều kiện $Y=y$:

$$f(x|y)=\frac{f(x,y)}{f_Y(y)}=\frac{1}{\sigma_1\sqrt{1-\rho^2}}\varphi\left(\frac{x-\mu_1-(\sigma_1/\sigma_2)\rho(y-\mu_2)}{\sigma_1\sqrt{1-\rho^2}}\right) \quad \textbf{(3.64)}$$

trong đó $\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{\frac{-x^2}{2}}$ là hàm mật độ xác suất của biến ngẫu nhiên chuẩn tắc $\mathbf{N}(0;1)$.

Nhận xét:

1. Ma trận tương quan của (X,Y) là

$$M=\begin{bmatrix}\sigma_1^2 & \rho\sigma_1\sigma_2\\ \rho\sigma_1\sigma_2 & \sigma_2^2\end{bmatrix}\Rightarrow M^{-1}=\frac{1}{1-\rho^2}\begin{bmatrix}1/\sigma_1^2 & -\rho/(\sigma_1\sigma_2)\\ -\rho/(\sigma_1\sigma_2) & 1/\sigma_2^2\end{bmatrix}. \quad \textbf{(3.65)}$$

2. Từ định lý 3.9 suy ra rằng phân bố của X với điều kiện $Y=y$ có phân bố chuẩn với kỳ vọng $\mu_1+\rho\frac{\sigma_1}{\sigma_2}(y-\mu_2)$ và phương sai $\sigma_1^2(1-\rho^2)$, do đó kỳ vọng có điều kiện:

$$\mathrm{E}\left(X\middle|Y=y\right)=\mu_1+\rho\frac{\sigma_1}{\sigma_2}(y-\mu_2). \quad \textbf{(3.66)}$$

là một hàm tuyến tính theo y. Đây là một tính chất rất quan trọng của phân bố chuẩn 2 chiều.

Định lý 3.10: Nếu (X,Y) có phân bố chuẩn hai chiều thì với mọi hằng số $a,b\in\mathbb{R}$ biến ngẫu nhiên $Z=aX+bY$ cũng có phân bố chuẩn.

CÂU HỎI ÔN TẬP VÀ BÀI TẬP

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

Đúng ☐ Sai ☐.

3.2 Bảng phân bố xác suất đồng thời của (X,Y) xác định bảng phân bố xác suất của hai biến ngẫu nhiên thành phần X và Y.

Đúng ☐ Sai ☐.

3.3 Nếu hai biến ngẫu nhiên X, Y độc lập thì bảng phân bố xác suất của X và Y cho phép xác định bảng phân bố xác suất đồng thời của (X,Y).

Đúng ☐ Sai ☐.

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

Đúng ☐ Sai ☐.

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

Đúng ☐ Sai ☐.

3.6 Hiệp phương sai luôn nhận giá trị dương.

Đúng ☐ Sai ☐.

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

Đúng ☐ Sai ☐.

3.8 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 ☐.

3.9 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 ☐.

3.10 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 ☐.

3.11 Hàm mật độ xác suất đồng thời của véc tơ ngẫu nhiên liên tục (X,Y) bằng tích của hai hàm mật độ xác suất thành phần X và Y.

Đúng ☐ Sai ☐.

3.12 Giả sử (X,Y) là véc tơ ngẫu nhiên phân bố chuẩn, khi đó X, Y độc lập khi và chỉ khi X, Y không tương quan.

Đúng ☐ Sai ☐.

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

X \ Y	y_1	y_2	y_3
x_1	0,18	0,22	0,16
x_2	0,08	0,16	0,20

Tìm bảng phân bố xác suất của hai biến ngẫu nhiên thành phần X, Y.

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

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

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

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

X \ 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?

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

X \ Y	1	2	3
1	0,12	0,15	0,03
2	0,28	0,35	0,07

a) Chứng minh rằng X, Y có độc lập.

b) Tìm quy luật phân bố của biến ngẫu nhiên $Z = XY$.

c) Tính các kỳ vọng $\mathrm{E}X, \mathrm{E}Y, \mathrm{E}Z$.

3.17 Cho X, Y là hai biến ngẫu nhiên độc lập có bảng 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 bảng phân bố xác suất đồng thời của X, Y. Tính xác suất $P\{X > Y\}$.

3.18 Cho X, Y là hai biến ngẫu nhiên có bảng 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 $P\{X = 1 | Y = 2\}$.

3.19 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.

3.20 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 $Y = 27$.

3.21 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 $\mathrm{E}X, \mathrm{E}Y$ và phương sai $\mathrm{D}X, \mathrm{D}Y$.

3.22 Cho X, Y là hai biến ngẫu nhiên có bảng phân bố xác suất đồng thời

X \ Y	-1	0	1
-1	4α	α	4α
0	α	2α	α
1	0	2α	0

a) Tìm α. Tính $\mathrm{E}X$, $\mathrm{E}Y$.

b) Tính $\mathrm{cov}(X,Y)$, $\rho(X,Y)$.

c) X và Y có độc lập không.

3.23 Cho véc tơ ngẫu nhiên (X,Y) có hàm mật độ xác suất xác định như sau:

$$f(x,y)=\begin{cases} Cx & \text{nếu } 0<y<x<1 \\ 0 & \text{nếu ngược lại} \end{cases}.$$

a) Tìm C.

b) Tìm các hàm mật độ xác suất của X và của Y.

c) X và Y có độc lập không?

3.24 Hàm phân bố xác suất của véc tơ ngẫu nhiên (X,Y) có dạng

$$F(x,y)=\begin{cases} 1-e^{-x}-e^{-y}+e^{-x-y} & \text{nếu } x>0, y>0 \\ 0 & \text{nếu ngược lại} \end{cases}.$$

Tìm hàm mật độ xác suất đồng thời $f(x,y)$ và hàm mật độ có điều kiện $f(x|y)$.

3.25 Cho véc tơ ngẫu nhiên (X,Y) có hàm mật độ xác suất xác định như sau:

$$f(x,y)=\frac{C}{(1+x^2)(1+y^2)}.$$

a) Tìm C.

b) Tìm hàm phân bố xác suất đồng thời mật độ của X, Y.

c) X và Y có độc lập không?

d) Tính xác suất để véc tơ ngẫu nhiên (X,Y) nhận giá trị nằm trong hình chữ nhật với các đỉnh là $A(1,1); B(\sqrt{3},1); C(1,0)$ và $D(\sqrt{3},0)$.

3.26 Cho véc tơ ngẫu nhiên (X,Y) có hàm mật độ xác suất xác định như sau:

$$f(x,y)=\begin{cases} \dfrac{1}{2x^2y} & \text{nếu } x\ge 1 \text{ và } \dfrac{1}{x}\le y\le x, \\ 0 & \text{nếu ngượclại} \end{cases}$$

a) Tìm hàm mật độ xác suất của X và của Y,

b) Tìm hàm mật độ xác suất có điều kiện $f(x|y)$ và $f(y|x)$.

3.27 Cho X và Y là hai biến ngẫu nhiên có phân bố chuẩn đồng thời với

$$\mathrm{E}X=35,\ \mathrm{E}Y=20,\ \mathrm{D}X=36,\ \mathrm{D}Y=16 \text{ và } \rho_{X,Y}=0{,}8.$$

Tìm kỳ vọng và phương sai của $2X-3Y$.

CHƯƠNG IV: LUẬT SỐ LỚN VÀ ĐỊNH LÝ GIỚI HẠN

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 một nội dung của luật số lớn. Luật số lớn cũng là cơ sở để định nghĩa xác suất của biến cố thông qua tần suất xuất hiện của biến cố đó.

Luật số lớn nghiên cứu sự hội tụ theo xác suất của dãy các biến ngẫu nhiên.

Luật số lớn đầu tiên của James Bernoulli được công bố năm 1713. Về sau, kết quả này được Poisson, Trêbưsép, Markov, Liapunốp mở rộng.

Trong 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 áp dụng cho dãy các biến ngẫu nhiên độc lập có phân bố không – một $A(p)$ công thức (2.9).

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

Định lý giới hạn trung tâm khảo sát sự hội tụ theo phân bố xác suất cúa dãy các biến ngẫu nhiên. Định lý giới hạn trung tâm áp dụng cho dãy các biến ngẫu nhiên độc lập $X_1, X_2, \dots$ có cùng phân bố không – một $A(p)$ ta được định lý Moivre –Laplace.

Áp dụng định lý Moivre –Laplace ta có công thức tính gần đúng phân bố nhị thức.

Luật số lớn và định lý giới hạn trung tâm là cơ sở lý thuyết để giải quyết các bài toán ước lượng và kiểm định giả thiết thống kê.

NỘI DUNG

4.1. CÁC DẠNG HỘI TỤ CỦA DÃY CÁC BIẾN NGẪU NHIÊN

4.1.1. Hội tụ theo xác suất

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

$$\forall \varepsilon > 0 \quad \lim_{n\to\infty} P\left\{\left|X_n - X\right| > \varepsilon\right\} = 0. \tag{4.1}$$

Như vậy dãy các biến ngẫu nhiên $X_1, X_2, ...$ 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.

4.1.2. Hội tụ theo phân bố

Định nghĩa 4.2: *Dãy các biến ngẫu nhiên* $X_1, X_2, ...$ *gọi là hội tụ theo phân bố về biến ngẫu nhiên* X *nếu dãy các hàm phân bố xác suất* $\{F_{X_n}(x)\}_{n=1}^{\infty}$ *hội tụ về hàm phân bố xác suất* $F_X(x)$, *nghĩa là với mọi* $x \in \mathbb{R}$:

$$\lim_{n\to\infty} P\{X_n < x\} = P\{X < x\}. \tag{4.2}$$

Trường hợp dãy các biến ngẫu nhiên rời rạc $\{X_n\}_{n=1}^{\infty}$ và biến ngẫu nhiên rời rạc X có cùng tập giá trị $\mathscr{C} = \{c_1, c_2, ...\}$ thì dãy $\{X_n\}_{n=1}^{\infty}$ hội tụ theo phân bố về biến ngẫu nhiên X khi và chỉ khi với mọi $c_k \in \mathscr{C}$:

$$\lim_{n\to\infty} P\{X_n = c_k\} = P\{X = c_k\}. \tag{4.3}$$

4.2. LUẬT SỐ LỚN

4.2.1. Bất đẳng thức Trêbưsép

Định lý 4.1: Cho Y là biến ngẫu nhiên không âm có kỳ vọng hữu hạn. Khi đó với mọi $a > 0$ ta có:

$$P\{Y > a\} \le \frac{\mathrm{E}Y}{a}. \tag{4.4}$$

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

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

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

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

b) Giả sử Y liên tục có hàm mật độ xác suất $f(x)$. 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 $\quad P\{Y > a\} \le \dfrac{\mathrm{E}Y}{a}$.

Định lý 4.2: Nếu X là biến ngẫu nhiên có kỳ vọng và phương sai hữu hạn thì với mọi $\varepsilon > 0$ ta có:

$$P\{|X - \mathrm{E}X| > \varepsilon\} \le \frac{\mathrm{D}X}{\varepsilon^2}. \qquad \textbf{(4.5)}$$

cũng vậy,

$$P\{|X - \mathrm{E}X| \le \varepsilon\} \ge 1 - \frac{\mathrm{D}X}{\varepsilon^2}. \qquad \textbf{(4.6)}$$

Chứng minh: Áp dụng công thức (4.4) 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 (4.5)-(4.6) 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 4.1: Một cửa hàng muốn ước lượng nhanh chóng sai số của số vải bán ra trong một tháng của mình. Số vải của mỗi khách hàng được làm tròn bởi số nguyên gần nhất (ví dụ trong sổ ghi 195,6 m thì làm tròn là 196m). 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*.

Giải: 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:

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

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

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.

4.2.2. Luật số lớn Trêbưsép

Định lý 4.3: 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

$$\lim_{n\to\infty} P\left(\left|\frac{X_1+\cdots+X_n}{n} - \frac{\mathrm{E}X_1+\cdots+\mathrm{E}X_n}{n}\right| > \varepsilon\right) = 0 \qquad \textbf{(4.7)}$$

Chứng minh: Xét biến ngẫu nhiên $S_n = \dfrac{X_1+\cdots+X_n}{n}$. Từ giả thiết độc lập của dãy các biến ngẫu nhiên $X_1, X_2, ...$ ta suy ra $\mathrm{E}S_n = \dfrac{\mathrm{E}X_1+\cdots+\mathrm{E}X_n}{n}$; $\mathrm{D}S_n = \dfrac{\mathrm{D}X_1+\cdots+\mathrm{D}X_n}{n^2} \le \dfrac{C}{n}$.

Áp dụng bất đẳng thức Trêbưsép (4.5) cho biến ngẫu nhiên S_n ta có:

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

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

$$\frac{X_1+\cdots+X_n}{n} \xrightarrow[n\to\infty]{P} \mu \qquad \textbf{(4.8)}$$

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

$$\frac{X_1+\cdots+X_n}{n} \xrightarrow[n\to\infty]{P} \mu \qquad \textbf{(4.9)}$$

Định lý Trêbưsép chứng tỏ rằng trung bình số học của các biến ngẫu nhiên độc lập hội tụ theo xác suất về trung bình số học của kỳ vọng tương ứng của nó. Nói cách khác nó chứng tỏ sự ổn định của trung bình số học của một số lớn các biến ngẫu nhiên xung quanh trung bình số học của các kỳ vọng của các biến ngẫu nhiên ấy. Như vậy mặc dù từng biến ngẫu nhiên độc lập có thể nhận giá trị khác nhiều so với kỳ vọng của chúng, song trung bình số học của một số lớn các biến ngẫu nhiên lại nhận giá trị gần bằng trung bình số học 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. Giả sử X là số nốt xuất hiện ở mặt trên con xúc xắc. Ta có $\mathrm{E}X = 3,5$. 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.

4.2.3. Luật số lớn Bernoulli

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

Định lý 4.4: Tần suất f_n hội tụ theo xác suất về xác suất p của biến cố A, nghĩa là với mọi $\varepsilon > 0$

$$\lim_{n\to\infty} P\{|f_n - p| < \varepsilon\} = 1 \qquad \textbf{(4.10)}$$

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

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

ta gọi dãy các biến ngẫu nhiên $X_1, X_2, ..., X_n$ độc lập có cùng phân bố không – một $A(p)$ (công thức (2.9)). $\mathrm{E}X_k = p,\ \mathrm{D}X_k = p(1-p)$.

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

Vậy theo hệ quả 2 của định lý 4.2 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 vây ta thấy rằng khi số phép thử tăng lên thì tần suất tương ứng sẽ càng gần 0,5.

4.3. ĐỊNH LÝ GIỚI HẠN TRUNG TÂM

Định lý 4.5: Giả sử $X_1, X_2,...$ là dãy các biến ngẫu nhiên độc lập có cùng phân bố, có kỳ vọng μ và phương sai σ^2. Khi đó dãy biến ngẫu nhiên $S_n = \dfrac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}$ hội tụ theo phân bố về phân bố chuẩn tắc $\mathbf{N}(0;1)$, nghĩa là:

$$\text{Với mọi } x \in \mathbb{R},\ \lim_{n\to\infty} P\left\{S_n < x\right\} = \Phi(x) \qquad \textbf{(4.11)}$$

$\Phi(x)$ là hàm phân bố xác suất của phân bố chuẩn tắc $\mathbf{N}(0;1)$.

Áp dụng định lý giới hạn trung tâm cho dãy các biến ngẫu nhiên độc lập $X_1, X_2,...$ có cùng phân bố không – một $A(p)$ (công thức (2.9)) ta được định lý Moivre –Laplace:

Định lý 4.6 (Moivre –Laplace): Dãy các biến ngẫu nhiên $X_1, X_2,...$ độc lập có cùng phân bố không – một $A(p)$ ta được:

$$\text{Với mọi } x \in \mathbb{R},\ \lim_{n\to\infty} P\left\{\frac{X_1 + \cdots + X_n - np}{\sqrt{npq}} < x\right\} = \Phi(x)\,. \qquad \textbf{(4.12)}$$

4.4. XẤP XỈ PHÂN BỐ NHỊ THỨC

4.4.1. Xấp xỉ phân bố nhị thức bằng phân bố Poisson

Định lý 4.7: Cho $X_1, X_2,...$ là dãy các biến ngẫu nhiên có cùng phân bố nhị thức, ở đó với mỗi n, X_n có phân bố nhị thức $\mathscr{B}(n, p_n)$. Giả sử tồn tại giới hạn $\lim\limits_{n\to\infty} np_n = \lambda > 0$. Thì X_n hội tụ theo phân bố về biến ngẫu nhiên X có phân bố Poisson tham số λ.

Trong thực tế khi $n > 50$ và $p < 0{,}1$ người ta có thể xấp xỉ $\mathscr{B}(n;p)$ với phân bố Poisson $\mathscr{P}(np)$ tham số $\lambda = np$:

$$P\{X_n = k\} \approx e^{-np}\frac{(np)^k}{k!}\,. \qquad \textbf{(4.13)}$$

Ví dụ 3.2: Giả sử xác suất để làm ra mỗi đinh ốc không đúng quy cách là $p = 0{,}015$. Người ta xếp đinh ốc vào từng hộp, mỗi hộp 100 chiếc.

a) Tính tỉ lệ hộp chứa toàn đinh ốc không đúng quy cách.

b) Cần phải xếp ít nhất bao nhiêu đinh ốc trong mỗi hộp để tỉ lệ hộp chứa 100 đinh ốc tốt tối thiểu là 80%.

Giải: a) Nếu gọi X là số đinh ốc không đúng quy cách trong hộp chứa 100 đinh ốc thì $X \sim \mathscr{B}(n;p)$ với $n = 100,\ p = 0{,}015$.

Tính gần đúng: $P\left\{X = 0\right\} \approx e^{-np}\dfrac{(np)^0}{0!} = e^{-1{,}5} = 0{,}2231$

Vậy có khoảng 22,3% số hộp chứa 100 đinh ốc tốt.

b) Giả sử mỗi hộp chứa $100+k$ đinh ốc, k là số tự nhiên. Gọi X là số đinh ốc không đúng quy cách trong mỗi hộp chứa $100+k$ đinh ốc. $X \sim \mathscr{B}(n;p)$ với $n=100+k,\ p=0,015$. Ta phải xác định k nhỏ nhất để

$$P\{X \le k\} = \sum_{i=0}^{k} C_n^i \left(0,015\right)^i \left(0,985\right)^{n-i} \ge 0,8 .$$

Dùng công thức xấp xỉ $P\{X=i\} = C_n^i \left(0,015\right)^i \left(0,985\right)^{n-i} \approx e^{-\lambda} \dfrac{\lambda^i}{i!}$

ở đây $\lambda = np = (100+k)(0,015) \approx 1,5 + 0,015k \approx 1,5$ (vì k nhỏ).

Vậy cần tìm k nhỏ nhất để

$$e^{-1,5}\left\{1+\frac{1,5}{1!}+\frac{1,5^2}{2!}+\cdots+\frac{1,5^k}{k!}\right\} \ge 0,8 \quad \Leftrightarrow \quad 1+\frac{1,5}{1!}+\frac{1,5^2}{2!}+\cdots+\frac{1,5^k}{k!} \ge 0,8e^{1,5} = 3,5853$$

Thử với $k=1,2,...$, ta thấy $k=2$ bất đẳng thức trên được thoả mãn. Như vậy dùng xấp xỉ Poisson ta có thể kết luận mỗi hộp cần đóng 102 chiếc đinh ốc. Khi đó xác suất để có ít nhất 100 đinh ốc tốt trong hộp 102 chiếc là 0,8022.

4.4.2. Xấp xỉ phân bố nhị thức bằng phân bố chuẩn

Giả sử $X_1, X_2, ..., X_n$ độc lập có cùng phân bố không – một A(p). Theo công thức (2.9) và (2.10) ta có $U_n = X_1 + X_2 + \cdots + X_n \sim \mathscr{B}(n,p)$. Công thức (2.8) 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.

Định lý 4.8 (*định lý giới hạn địa phương*): Giả sử X là biến ngẫu nhiên có phân bố nhị thức $\mathscr{B}(n,p)$. Đặt $x_k = \dfrac{k-np}{\sqrt{npq}}$, khi đó

$$P_n(k\,;p) = P\{X=k\} = \varphi\left(\frac{k-np}{\sqrt{npq}}\right)\frac{1}{\sqrt{nqp}}\left(1+\varepsilon_{n,k}\right). \qquad \textbf{(4.14)}$$

trong đó $\left|\varepsilon_{n,k}\right| < \dfrac{C}{\sqrt{n}}$ với C là hằng số.

Như vậy khi n đủ lớn ta có thể xấp xỉ :

$$P\{X=k\} \approx \varphi\left(\frac{k-np}{\sqrt{npq}}\right)\frac{1}{\sqrt{nqp}} \frac{1}{\sqrt{2\pi}} e^{\frac{-(k-np)^2}{2npq}} . \qquad \textbf{(4.15)}$$

Ngoài ra áp dụng định lý Moivre-Laplace và công thức (4.12) ta cũng có công thức xấp xỉ giá trị của hàm phân bố xác suất nhị thức

$$P\{U_n < x\} = P\left\{\frac{U_n - np}{\sqrt{npq}} < \frac{x-np}{\sqrt{npq}}\right\} \approx \Phi\left(\frac{x-np}{\sqrt{npq}}\right) \qquad \textbf{(4.16)}$$

Người ta thấy rằng xấp xỉ là tốt khi np và nq lớn hơn 5 hoặc khi $npq > 20$.

Ví dụ 3.3: 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 $P\{1610 \le X \le 1650\}$.

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

$$P_{3200}(1600; 0,5) = \frac{3200!}{160011600!} 0,5^{3200}.$$

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) $$P\{10 \le X - 1600 \le 50\} = P\left\{\frac{10}{20\sqrt{2}} \le \frac{X-1600}{20\sqrt{2}} \le \frac{50}{20\sqrt{2}}\right\} \approx \Phi(1,7678) - \Phi(0,3536)$$

Tra bảng ta có $\Phi(1,7678) - \Phi(0,3536) \approx 0,9616 - 0,6406 = 0,321$.

CÂU HỎI ÔN TẬP VÀ BÀI TẬP

4.1 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 ☐.

4.2 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 ☐.

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

Đúng ☐ Sai ☐.

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

Đúng ☐ Sai ☐.

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

Đúng ☐ Sai ☐.

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

Đúng ☐ Sai ☐.

4.7 Tổng của dãy các biến ngẫu nhiên độc lập có cùng phân bố với kỳ vọng và phương sai hữu hạn tiệm cận phân bố chuẩn.

Đúng ☐ Sai ☐.

4.8 Luật số lớn xét sự hội tụ theo xác suất còn định lý giới hạn trung tâm xét sự hội tụ theo phân bố của dãy các biến ngẫu nhiên.

Đúng ☐ Sai ☐.

4.9 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

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

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

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

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

4.13 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.

4.14 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.

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

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

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

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

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

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

4.17 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.

4.18 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.

4.19 Một xí nghiệp sản xuất máy tính có xác suất làm ra sản phẩm phế phẩm là 0,02. Chọn ngẫu nhiên 2500 máy tính để kiểm tra. Tính xác suất để:

a) Có đúng hai máy phế phẩm;

b) Có không quá hai máy phế phẩm.

4.20 Một nhà nghỉ có 1000 khách. Nhà ăn phục vụ bữa trưa làm hai đợt liên tiếp. Số chỗ ngồi của nhà ăn phải ít nhất là bao nhiêu để xác suất của biến cố "không đủ chỗ cho người đến ăn" bé hơn 1%?

4.21 Một trường đại học có chỉ tiêu tuyển sinh là 300.

a) Giả sử có 325 người dự thi và xác suất thi đỗ của mỗi người là 90%. Tính xác suất để số người trúng tuyển không vượt quá chỉ tiêu.

b) Cần cho phép tối đa bao nhiêu người dự thi (xác suất đỗ của họ vẫn là 90%) để biến cố "số người trúng tuyển không vượt nhỏ hơn 0,99".

CHƯƠNG V: THỐNG KÊ TOÁN HỌC

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 số liệu thống kê các kết quả quan sát về những 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 những vấn đề cơ bản của lý thuyết thống kê toán học:

- 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.
- Lý thuyết ước lượng.
- Lý thuyết kiểm định giả thiết thống kê.

Đối với mẫu ngẫu nhiên ta xét các vấn đề:

- Các phương pháp mô tả mẫu: bảng phân bố tần số và 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.

Sử dụng phương pháp quy nạp thống kê ta có thể ước lượng các tham số đặc trưng của tổng thể thông qua thống kê của mẫu.

Phương pháp ước lượng điểm chủ trương 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 mômect 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.

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.

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. Kiểm định các tham số đặc trưng: kỳ vọng của biến ngẫu nhiên gốc trong tổng thể, tần suất của tổng thể.

Để học tốt chương này học viên cần nắm vững cơ sở lý thuyết xác suất đã được học trong các chương trước.

NỘI DUNG

5.1. LÝ THUYẾT MẪU

5.1.1. Sự cần thiết phải lấy mẫu

Nhiều bài toán trong thực tế dẫn đến nghiên cứu một hay nhiều dấu hiệu định tính hoặc định lượng 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 mỗi 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. Khi khảo sát một tín hiệu là quá trình ngẫu nhiên người ta tiến hành lấy mẫu tại những thời điểm nào đó và thu được các tín hiệu mẫu.

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.

5.1.2. Tổng thể nghiên cứu, dấu hiệu 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.

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 $\mathscr{C}$. Đối với dấu hiệu định tính có thể mã hóa và xem là biến ngẫu nhiên có phân bố không - một.

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*.

5.1.3. 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 $W=\left(X_1, X_2, ..., X_n\right)$.

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 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 5.1: 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	1/6	1/6	1/6	1/6	1/6	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, $W=\left(X_1, X_2, X_3\right)$.

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.

5.1.4. Các phương pháp mô tả mẫu ngẫu nhiên

5.1.4.1. Bảng phân bố tần số thực nghiệm

Giả sử ta rút được 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

$$x_1 < \cdots < x_k\ ;\ r_1+\cdots+r_k = n. \qquad \textbf{(5.1)}$$

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

X	x_1	x_2	…	x_k
Tần số	r_1	r_2	…	r_k

(5.2)

5.1.4.2. Bảng phân bố tần suất thực nghiệm

Ký hiệu $f_i = \dfrac{r_i}{n}$ gọi là tần suất của x_i.

Ta có bảng phân bố tần suất thực nghiệm của X

X	x_1	x_2	…	x_k
Tần suất	f_1	f_2	…	f_k

(5.3)

Ví dụ 5.2: 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

5.1.4.3. Hàm phân bố thực nghiệm của mẫu

Với mẫu ngẫu nhiên xác định bới công thúc (5.1). Hàm số xác định như sau

$$F_n(x) = \sum_{x_j<x} f_j\ ;\quad -\infty < x < +\infty \qquad \textbf{(5.4)}$$

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.

5.1.4.4. Bảng phân bố ghép lớp

Trong những trường hợp mẫu điều tra có kích thước lớn, hoặc khi các giá trị cụ thể của dấu hiệu X lấy giá trị khác nhau song lại khá gần nhau, người ta thường xác định một số các khoảng $C_1, C_2, ..., C_k$ sao cho mỗi giá trị của dấu hiệu điều tra thuộc vào một khoảng nào đó. Các khoảng này lập thành một phân hoạch của miền giá trị của X.

Việc chọn số khoảng và độ rộng khoảng là tuỳ thuộc vào kinh nghiệm của người nghiên cứu, nhưng nói chung không nên chia quá ít khoảng. Ngoài ra độ rộng các khoảng cũng không nhất thiết phải bằng nhau. Chẳng hạn khi muốn thống kê về tỉ lệ người nghiện thuốc lá thì ta tập trung nhiều vào độ tuổi thanh niên và trung niên.

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 5.3: Một mẫu về chiều cao (cm) của 400 cây con 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

Giá trị $y_i = \dfrac{r_i}{l_i}$ 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 khoảng.

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

5.1.4.5. 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	…	x_k
Tần số	r_1	r_2	…	r_k

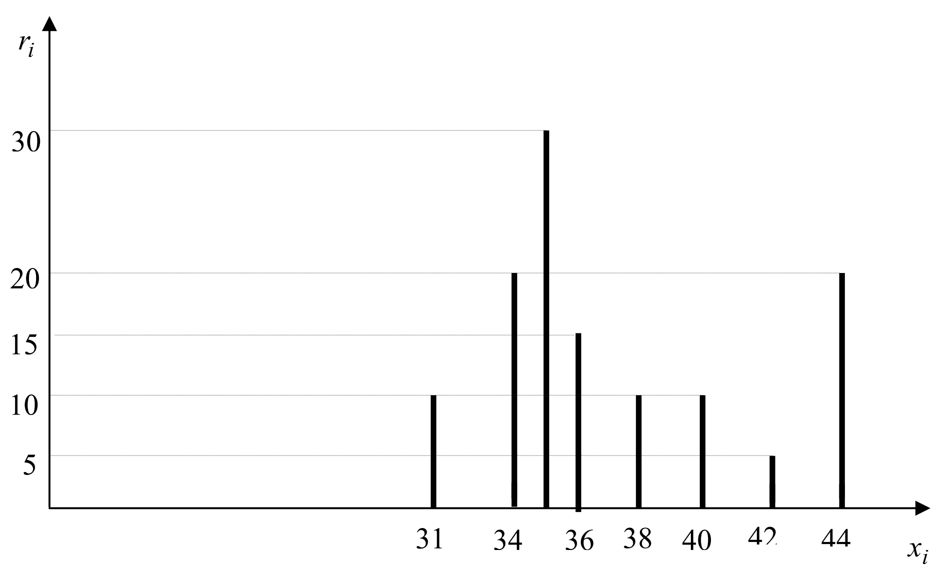
X	x_1	x_2	…	x_k
Tần suất	f_1	f_2	…	f_k

Trong mặt phẳng với hệ trục toạ độ Oxy.

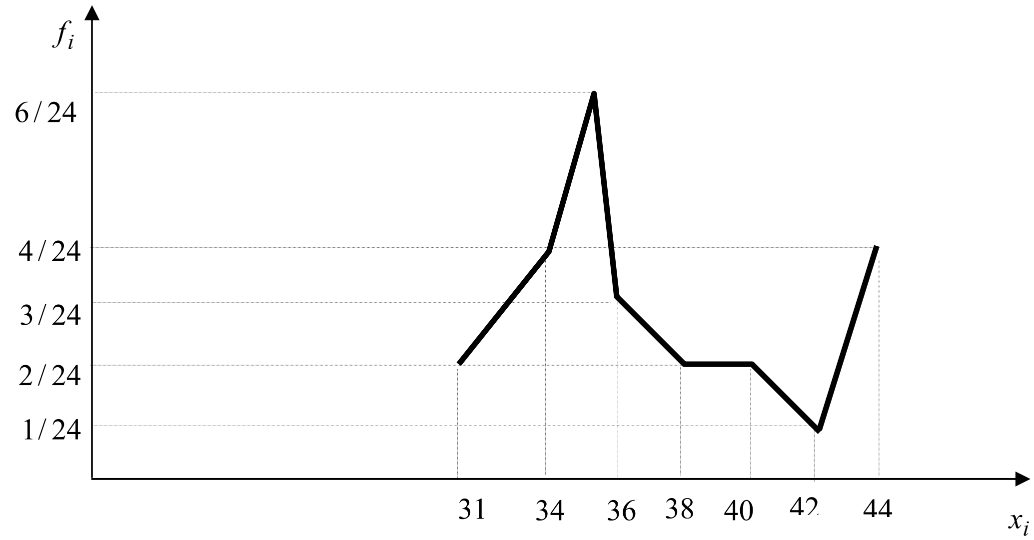
❖ Nối điểm trên trục hoành có toạ độ $(x_i,0)$ với điểm có toạ độ (x_i,r_i); $i=\overline{1,k}$ ta được *biểu đồ tần số hình gậy*.

❖ 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 5.2 có biểu đồ tần số hình gậy và đa giác tần suất



Biểu đồ tần số hình gậy



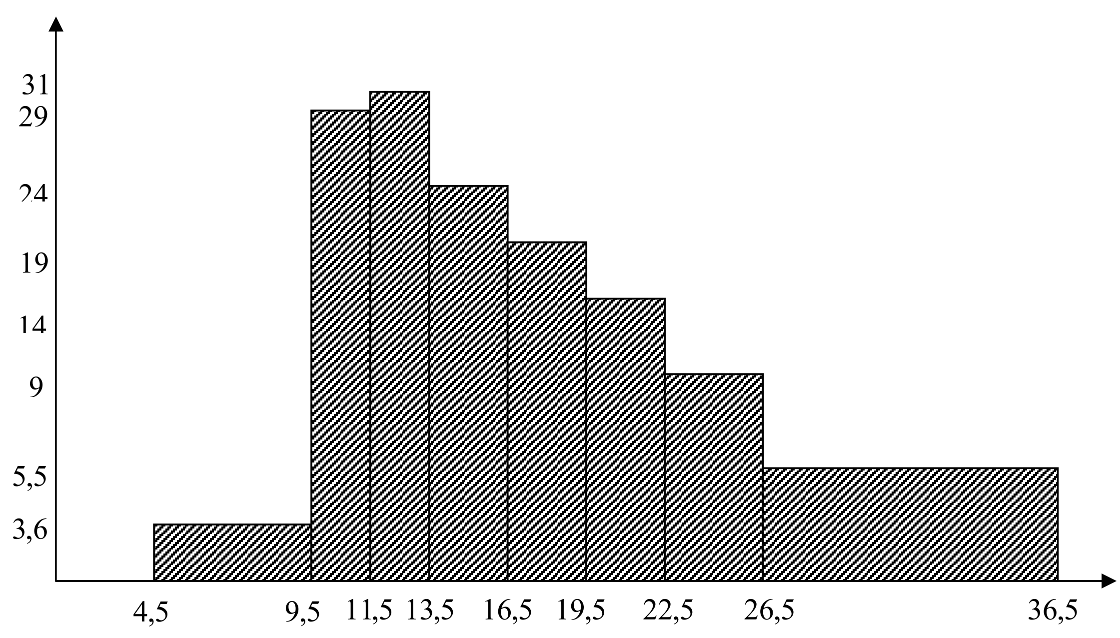
Biểu đồ đa giác tần suất

1.4.6. Tổ chức đồ (histogram)

Đối với bảng phân bố ghép lớp, người ta thường dùng tổ chức đồ để biểu diễn.

Trong mặt phẳng với hệ toạ độ Oxy, trên trục hoành ta chia các khoảng C_i có độ rộng l_i. Với mỗi khoảng C_i ta dựng hình chữ nhật có chiều cao $y_i = \frac{r_i}{l_i}$ (đối với tổ chức đồ tần số), hay $y_i = \frac{f_i}{l_i}$ (đối với tổ chức đồ tần suất).

Tổ chức đồ tần số của mẫu ghép lớp của ví dụ 5.3



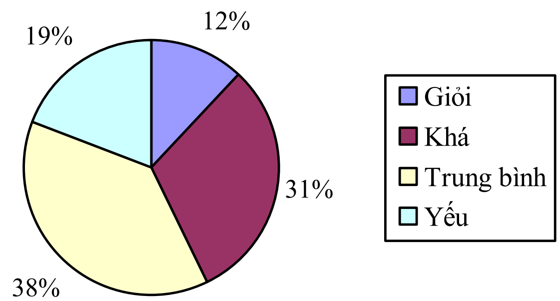
Chú ý rằng diện tích giới hạn bởi tổ chức đồ bằng tần số xuất hiện. Chẳng hạn số cây con nằm trong khoảng (12; 25] chính là diện tích của tổ chức đồ giới hạn bởi đường thẳng $x=12$ và $x=25$.

$$(13{,}5-12)\times 31+(16{,}5-13{,}5)\times 24+(19{,}5-16{,}5)\times 19+(22{,}5-19{,}5)\times 14+(25-22{,}5)\times 9=240$$

Vậy có 240 cây con có chiều cao từ $12\,cm$ đến $25\,cm$.

Khi dấu hiệu điều tra có tính chất 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 5.4: Tổng kết kết quả học tập của sinh viên Học viện ta trong năm 2005 được số liệu sau:



5.1.5. Thống kê và các đặc trưng của mẫu ngẫu nhiên

5.1.5.1. 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:*

$$T=T(X_1,X_2,...,X_n) \quad \textbf{(5.5)}$$

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 $\mathrm{D}T$ …(xem mục 3.5 chương 3). Mặt khác, khi mẫu ngẫu nhiên nhận một giá trị cụ thể $w=(x_1,x_2,...,x_n)$ thì T cũng nhận một giá trị cụ thể còn gọi là giá trị quan sát của thống kê

$$T_{qs}=T(x_1,x_2,...,x_n)$$

Các thống kê 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.

5.1.5.2. Trung bình mẫu

Trung bình mẫu của mẫu ngẫu nhiên $W=(X_1,X_2,...,X_n)$ của biến ngẫu nhiên X được định nghĩa và ký hiệu

$$\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i \quad \textbf{(5.6)}$$

Giá trị trung bình mẫu cụ thể của mẫu ngẫu nhiên cụ thể $w=(x_1,x_2,...,x_n)$ là

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{5.7}$$

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 (2.47), (2.48), (2.58) ta có

$$\mathrm{E}\left(\overline{X}\right) = \mathrm{E}X\ ;\ \ \mathrm{D}\left(\overline{X}\right) = \frac{\mathrm{D}X}{n}. \tag{5.8}$$

5.1.5.3. Phương sai mẫu

- Phương sai mẫu $\widehat{S}^2$:

$$\widehat{S}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^{\,2} - \left(\overline{X}\right)^2 \tag{5.9}$$

- Phương sai mẫu có hiệu chỉnh S^2:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 = \frac{1}{n-1}\sum_{i=1}^{n} X_i^2 - \frac{n}{n-1}\left(\overline{X}\right)^2 \tag{5.9'}$$

- Phương sai mẫu S^{*2} khi dấu hiệu nghiên cứu X của tổng thể có kỳ vọng xác định $\mathrm{E}X = \mu$.

$$S^{*2} = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \mu\right)^2 \tag{5.10}$$

Áp dụng công thức tính kỳ vọng (2.47), (2.48) và (5.8) ta có:

$$\mathrm{E}S^2 = DX \ \text{ và } \ \mathrm{E}S^{*2} = \mathrm{D}X \tag{5.11}$$

5.1.5.4. Độ lệch tiêu chuẩn mẫu

$$S = \sqrt{S^2} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2} \tag{5.12}$$

5.1.5.5. 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 (dấu hiệu định tính). 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 $\mathrm{E}X = p$, $\mathrm{D}X = p(1-p)$ (công thức (2.9)).

Lấy mẫu ngẫu nhiên: $W = \left(X_1, X_2, ..., X_n\right)$, $X_1, X_2, ..., X_n$ 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

$$r = X_1 + X_2 + \cdots + X_n. \tag{5.13}$$

Tần suất mẫu

$$f = \frac{r}{n} \tag{5.14}$$

Tương tự công thức (5.11) ta có công thức tính kỳ vọng và phương sai của tần suất mẫu:

$$\mathrm{E}(f) = p\,;\ \ \mathrm{D}(f) = \frac{p(1-p)}{n} \tag{5.15}$$

5.1.5.6. 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 $r_1, r_2, ..., r_k$ thì giá trị trung bình mẫu và phương sai mẫu cụ thể được tính theo công thức

$$\overline{x} = \frac{\sum_{i=1}^{k} r_i x_i}{n}\,,\ \ \sum_{i=1}^{k} r_i = n \tag{5.16}$$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{k} r_i\left(x_i - \overline{x}\right)^2 = \frac{1}{n-1}\left(\sum_{i=1}^{k} r_i x_i^2 - \frac{\left(\sum_{i=1}^{k} r_i x_i\right)^2}{n}\right) \tag{5.17}$$

2. Nếu giá trị của mẫu cụ thể được cho dưới dạng bảng phân bố ghép lớp với các khoảng $C_1, ..., C_m$ và tần số của C_i là r_i 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:

$$u_i = \frac{x_i - a}{h} \ \Rightarrow\ x_i = hu_i + a \ \Rightarrow\ \overline{x} = h\overline{u} + a;\ \ s_x^2 = h^2 s_u^2 \tag{5.18}$$

$$\overline{u} = \frac{1}{n}\sum_{i=1}^{k} r_i u_i\,,\ s_u^2 = \frac{1}{n-1}\sum_{i=1}^{k} r_i\left(u_i - \overline{u}\right)^2 = \frac{1}{n-1}\left(\sum_{i=1}^{k} r_i u_i^2 - \frac{\left(\sum_{i=1}^{k} r_i u_i\right)^2}{n}\right) \tag{5.18'}$$

Thật vậy: $\overline{x} = \frac{1}{n}\sum_{i=1}^{k} r_i x_i = \frac{1}{n}\sum_{i=1}^{k} r_i\left(hu_i + a\right) = \frac{h}{n}\sum_{i=1}^{k} r_i u_i + \left(\frac{1}{n}\sum_{i=1}^{k} r_i\right)a = h\overline{u} + a$.

$$s_x^2 = \frac{1}{n-1}\sum_{i=1}^{k} r_i\left(x_i - \overline{x}\right)^2 = \frac{1}{n-1}\sum_{i=1}^{k} r_i\left(hu_i + a - h\overline{u} - a\right)^2 = \frac{h^2}{n-1}\sum_{i=1}^{k} r_i\left(u_i - \overline{u}\right)^2 = h^2 s_u^2.$$

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 5.5: Giá trị trung bình mẫu và phương sai mẫu của mẫu ở ví dụ 5.3.

Khoảng	tần số r_i	x_i	$u_i = \dfrac{x_i - 20}{5}$	$r_i u_i$	$r_i u_i^2$
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

$\overline{x} = 5\overline{u} + 20 = 5 \times \dfrac{-177{,}5}{400} + 20 = 17{,}78$. $s_u^2 = \dfrac{1}{399} \times \left(873{,}47 - \dfrac{(-177{,}5)^2}{400} \right) = 1{,}9917$

$s_x^2 = 5^2 \times s_u^2 = 49{,}79 \quad \Rightarrow \quad s = \sqrt{49{,}79} = 7{,}056$.

5.1.6. 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

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 :

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

Các biến ngẫu nhiên thành phần $X_1, X_2, ..., X_n$ độc lập có cùng quy luật phân bố chuẩn $\mathbf{N}(\mu, \sigma^2)$ như X . Theo định lý 3.10 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 $\mathrm{E}(\overline{X}) = \mu$ và phương sai $\mathrm{D}(\overline{X}) = \dfrac{\sigma^2}{n}$, áp dụng công thức (2.33) suy ra thống kê sau có quy luật chuẩn tắc $\mathbf{N}(0;1)$:

$$U = \frac{(\overline{X} - \mu)\sqrt{n}}{\sigma} \sim \mathbf{N}(0;1) \qquad \textbf{(5.19)}$$

2) Thống kê phương sai S^{*2}.

Từ công thức (5.16) ta có: $nS^{*2} = \sum_{i=1}^{n}\left(X_i - \mu\right)^2$ và $\dfrac{nS^{*2}}{\sigma^2} = \sum_{i=1}^{n}\left(\dfrac{X_i-\mu}{\sigma}\right)^2$.

Vì các biến ngẫu nhiên X_i độc lập nên các biến ngẫu nhiên $\dfrac{X_i-\mu}{\sigma}$ cũng độc lập. Mặt khác theo (2.33) $\dfrac{X_i-\mu}{\sigma} \sim \mathbf{N}(0;1)$. Do đó thống kê sau có phân bố khi bình phương n bậc tự do.

$$\chi^2 = \frac{nS^{*2}}{\sigma^2} = \sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 \sim \chi^2(n) \qquad \textbf{(5.20)}$$

3) Thống kê phương sai S^2.

Tương tự thống kê sau có phân bố khi bình phương $n-1$ bậc tự do.

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2} = \sum_{i=1}^{n}\left(\frac{X_i-\overline{X}}{\sigma}\right)^2 \sim \chi^2(n-1) \qquad \textbf{(5.21)}$$

4) Từ (5.19), (5.21) và U, χ^2 độc lập, áp dụng công thức (2.42) thì $T = \dfrac{U}{\sqrt{\chi^2/(n-1)}}$ có phân bố Student $n-1$ bậc tự do.

$$T = \frac{(\overline{X}-\mu)\sqrt{n}}{S} = \frac{\dfrac{(\overline{X}-\mu)\sqrt{n}}{\sigma}}{\sqrt{\dfrac{(n-1)S^2}{\sigma^2(n-1)}}} = \frac{U}{\sqrt{\chi^2/(n-1)}} \sim \mathbf{T}(n-1)\,. \qquad \textbf{(5.22)}$$

Khi n khá lớn thì quy luật Student $\mathbf{T}(n)$ hội tụ khá nhanh về phân bố chuẩn tắc $\mathbf{N}(0;1)$, do đó trong thực tế nếu $n \ge 30$ ta có thể xem thống kê T xấp xỉ $\mathbf{N}(0;1)$.

Quy luật phân bố của tần suất mẫu

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:

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

Từ công thức (5.13), (5.14), (5.15) 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: $\mathrm{E}(f) = p$; $\mathrm{D}(f) = \dfrac{pq}{n}$.

Áp dụng định lý Moivre-Laplace ta có:

$$\text{Với mọi } x \in \mathbb{R},\ \lim_{n\to\infty} P\left\{\frac{(f-p)\sqrt{n}}{\sqrt{pq}} < x\right\} = \Phi(x). \qquad \textbf{(5.23)}$$

Như vậy có thể xấp xỉ thống kê $U = \dfrac{(f-p)\sqrt{n}}{\sqrt{pq}}$ với phân bố chuẩn tắc $\mathbf{N}(0;1)$ khi n đủ lớn. Người ta thấy rằng xấp xỉ là tốt khi $np > 5$ và $nq > 5$ hoặc $npq > 20$.

$$U = \frac{(f-p)\sqrt{n}}{\sqrt{pq}} \sim N(0;1) \text{ khi } \begin{cases} np > 5 \\ nq > 5 \end{cases} \text{ hoặc } npq > 20. \qquad \textbf{(5.24)}$$

5.2. LÝ THUYẾT ƯỚC LƯỢNG

5.2.1. 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 là 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 công thức (5.5):

$$\hat{\theta} = T(X_1, X_2, ..., X_n)$$

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:

5.2.2. 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 5.1: *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*

$$\mathrm{E}(\hat{\theta}) = \theta \qquad \textbf{(5.25)}$$

Nếu $\mathrm{E}(\hat{\theta}) \neq \theta$ thì $\hat{\theta} = T(X_1, X_2, ..., X_n)$ được gọi là một *ước lượng chệch* của θ.

5.2.3. Ước lượng hiệu quả

Điều kiện (5.25) của ước lượng không chệch có nghĩa rằng trung bình các giá trị của $\hat{\theta}$ bằng giá trị θ. Tuy nhiên 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 5.2: *Ước lượng không chệch có phương sai nhỏ nhất so với mọi ước lượng không chệch khác được xây dựng trên cùng một mẫu ngẫu nhiên gọi là ước lượng hiệu quả.*

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 $W = \left(X_1, X_2, ..., X_n\right)$ được lấy từ tổng thể có dấu hiệu nghiên cứu là biến ngẫu nhiên X mà 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 II) và $\hat{\theta}$ là ước lượng không chệch bất kỳ của θ thì

$$\mathrm{D}\left(\hat{\theta}\right) \ge \frac{1}{n\mathrm{E}\left(\dfrac{\partial\left(\ln f(x,\theta)\right)}{\partial\theta}\right)^2} \tag{5.26}$$

Ví dụ 5.6: Dựa vào bất đẳng thức trên ta có thể chứng minh được rằng trung bình mẫu $\overline{X}$ là ước lượng hiệu quả của kỳ vọng μ của dấu hiệu nghiên cứu X của tổng thể có phân bố chuẩn $\mathbf{N}(\mu,\sigma^2)$.

Thật vậy theo công thức (5.8) ta có $\mathrm{D}\left(\overline{X}\right) = \dfrac{\sigma^2}{n}$. Mặt khác theo (2.26) hàm mật độ của X có dạng

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

$$\Rightarrow \ln f(x,\mu) = -\ln\left(\sigma\sqrt{2\pi}\right) - \frac{(x-\mu)^2}{2\sigma^2} \Rightarrow \frac{\partial \ln f(x,\mu)}{\partial\mu} = \frac{x-\mu}{\sigma^2}$$

Vậy $\quad n\mathrm{E}\left(\dfrac{\partial\left(\ln f(X,\theta)\right)}{\partial\theta}\right)^2 = n\mathrm{E}\left(\dfrac{X-\mu}{\sigma^2}\right)^2 = \dfrac{n}{\sigma^4}\mathrm{E}\left(X-\mu\right)^2 = \dfrac{n}{\sigma^4}\mathrm{D}\left(X\right) = \dfrac{n}{\sigma^2}$.

Như vậy $\mathrm{D}\left(\overline{X}\right) = \dfrac{\sigma^2}{n}$ đạt giá trị cực tiểu của bất đẳng thức Cramer-Rao, do đó trung bình mẫu $\overline{X}$ là ước lượng hiệu quả của μ.

5.2.4. Ước lượng vững

Định nghĩa 5.3: *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

$$\lim_{n\to\infty} P\left\{\left|\hat{\theta}-\theta\right|<\varepsilon\right\}=1. \qquad \textbf{(5.27)}$$

Theo hệ quả 2 của luật số lớn Trêbusep, công thức (4.9) chương IV, 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.

Theo luật số lớn Bernoulli ta suy ra 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 S^{*2} (trường hợp μ đã biết) là ước lượng không chệch và vững của phương sai σ^2 của biến ngẫu nhiên gốc của tổng thể.

5.2.5. Phương pháp ước lượng bằng khoảng tin cậy

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 5.4: *Khoảng* $[a;b]$ *có hai đầu mút là hai thống kê*

$$a=a(X_1,X_2,...,X_n),\ b=b(X_1,X_2,...,X_n) \qquad \textbf{(5.28)}$$

phụ thuộc mẫu ngẫu nhiên $W=(X_1,X_2,...,X_n)$ *của biến ngẫu nhiên gốc* X, *gọi là khoảng tin cậy của tham số* θ *với độ tin cậy* β *nếu:*

$$P\{a(X_1,X_2,...,X_n)\le\theta\le b(X_1,X_2,...,X_n)\}=\beta \qquad \textbf{(5.29)}$$

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 $a\le\theta\le b$.

5.2.6. Khoảng tin cậy của kỳ vọng của biến ngẫu nhiên phân bố chuẩn

Giả sử tổng thể biến ngẫu nhiên gốc X có phân bố chuẩn $\mathbf{N}(\mu;\sigma^2)$ nhưng chưa biết tham số μ của nó.

Từ tổng thể rút ra một mẫu ngẫu nhiên kích thước n: $W=(X_1,X_2,...,X_n)$, ta tìm khoảng tin cậy của μ trong các trường hợp sau:

5.2.6.1. Trường hợp phương sai σ^2 đã biết

Định lý 5.1: Khoảng tin cậy của tham số μ với độ tin cậy β có dạng:

$$\left[\overline{X}-U_{\alpha/2}\frac{\sigma}{\sqrt{n}};\ \overline{X}+U_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right] \tag{5.30}$$

trong đó: $\alpha=1-\beta$; $U_{\frac{\alpha}{2}}$ là giá trị tới hạn mức $\frac{\alpha}{2}$ của phân bố chuẩn tắc $\mathbf{N}(0;1)$ (công thức 2.31).

Chứng minh: Theo công thức (5.19) ta có $\dfrac{(\overline{X}-\mu)\sqrt{n}}{\sigma}\sim\mathbf{N}(0;1)$.

Mặt khác: $\overline{X}-U_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}\le\mu\le\overline{X}+U_{\alpha/2}\dfrac{\sigma}{\sqrt{n}} \Leftrightarrow \left|\dfrac{(\overline{X}-\mu)\sqrt{n}}{\sigma}\right|\le U_{\alpha/2}$.

Áp dụng công thức (2.32) 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 5.5: $\varepsilon=U_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}$ 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 cần thiết* là số tự nhiên n nhỏ nhất thỏa mãn:

$$n\ge\frac{\sigma^2U_{\alpha/2}^2}{\varepsilon_0^2} \tag{5.31}$$

Ví dụ 5.7: Trọng lượng của một loại sản phẩm là một biến ngẫu nhiên phân bố chuẩn với độ lệch tiêu chuẩn 1 gram. Cần 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 $\mathbf{N}(\mu;\sigma^2)$ với $\sigma = 1$. Trọng lượng trung bình của sản phẩm là tham số μ. Khoảng tin cậy có dạng (5.30).

Với độ tin cậy $\beta = 0,95 \Rightarrow \dfrac{\alpha}{2} = 0,025 \Rightarrow U_{\frac{\alpha}{2}} = 1,96$.

a. Từ bảng số liệu tìm được trung bình mẫu cụ thể:

$$\overline{x} = \frac{3\cdot 18 + 5\cdot 19 + 15\cdot 20 + 2\cdot 21}{25} = 19,64 .$$

Độ chính xác của ước lượng $\varepsilon = U_{\alpha/2}\dfrac{\sigma}{\sqrt{n}} = 1,96\cdot\dfrac{1}{\sqrt{25}} = 0,392$.

Vậy với độ tin cậy 95% qua mẫu cụ thể này, khoảng tin cậy của tham số μ là:

$$\left[19,64 - 0,392\,;\, 19,64 + 0,392\right]$$

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:

$$n \ge \frac{\sigma^2 U_{\alpha/2}^2}{\varepsilon_0^2} = \frac{1\cdot 1,96^2}{0,3^2} = 42,68 \text{. Chọn } n = 43 .$$

5.2.6.2. 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 (5.12). Mặt khác, theo định lý giới hạn trung tâm thì thống kê $\dfrac{(\overline{X} - \mu)\sqrt{n}}{\sigma}$ có phân bố xấp xỉ phân bố chuẩn, đúng với mọi biến ngẫu nhiên gốc X (không đòi hỏi phân bố chuẩn).

Do đó khoảng tin cậy của tham số μ với độ tin cậy β có thể lấy là

$$\left[\overline{X} - U_{\alpha/2}\frac{S}{\sqrt{n}}; \overline{X} + U_{\alpha/2}\frac{S}{\sqrt{n}}\right] \qquad \textbf{(5.32)}$$

Ví dụ 5.8: 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	r_i	x_i	$u_i = x_i - 8,25$	$r_i u_i$	$r_i u_i^2$
$6,5-7,0$	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 $\beta = 95\%$, $U_{\alpha/2} = 1,96$.

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:

$$7,87 \le \mu \le 8,29\,.$$

5.2.6.3. Trường hợp phương sai σ^2 chưa biết, kích thước mẫu $n < 30$

Trong trường hợp này, theo công thức (5.22) thống kê

$$T = \frac{(\overline{X} - \mu)\sqrt{n}}{S} \qquad \textbf{(5.33)}$$

có phân bố Student $n-1$ bậc tự do. Vì vậy khoảng tin cậy được tính theo kết quả sau:

Định lý 5.2: Khoảng tin cậy của tham số μ với độ tin cậy β có dạng:

$$\left[\overline{X} - t_{\alpha/2}(n-1)\frac{S}{\sqrt{n}}; \overline{X} + t_{\alpha/2}(n-1)\frac{S}{\sqrt{n}}\right], \qquad \textbf{(5.34)}$$

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 (công thức 2.43).

Độ chính xác của ước lượng:

$$\varepsilon = t_{\alpha/2}(n-1)\frac{S}{\sqrt{n}} \quad \textbf{(5.35)}$$

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:

$$n \ge \left(\frac{S\cdot t_{\alpha/2}(n-1)}{\varepsilon_0}\right)^2 \quad \textbf{(5.36)}$$

Ví dụ 5.9: Năng suất của một loại giống mới là một biến ngẫu nhiên có quy luật phân bố chuẩn $\mathbf{N}(\mu;\sigma^2)$. Gieo thử giống hạt 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 hạt giống này là μ thỏa mãn:

$$169,115 \le \mu \le 172,885.$$

5.2.7. Khoảng tin cậy cho tần suất 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 $\mathrm{E}X = p$ và phương sai $\mathrm{D}X = p(1-p)$.

Lấy mẫu ngẫu nhiên $W = (X_1, X_2, ..., X_n)$, trong đó $X_1, X_2, ..., X_n$ là các biến ngẫu nhiên độc lập có cùng phân bố không – một $A(p)$.

$f=\dfrac{X_1+\cdots+X_n}{n}$ là tần suất mẫu. Theo định lý Moivre-Laplace (4.12) thì khi n đủ lớn ta có thể xấp xỉ thống kê $U=\dfrac{(f-p)\sqrt{n}}{\sqrt{p(1-p)}}$ với phân bố chuẩn tắc $\mathbf{N}(0;1)$.

Tuy nhiên vì p chưa biết nên chưa biết $p(1-p)=\mathrm{D}X$.

Mặt khác theo công thức (5.21), (5.24) thì 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 khoảng tin cậy cho tần suất p của tổng thể với độ tin cậy β là:

$$\left[f-U_{\alpha/2}\sqrt{\frac{f(1-f)}{n}}\ ;\ f+U_{\alpha/2}\sqrt{\frac{f(1-f)}{n}}\right] \tag{5.37}$$

Với điều kiện

$$\begin{cases} nf>10 \\ n(1-f)>10 \end{cases} \tag{5.38}$$

trong đó $U_{\alpha/2}$ là giá trị tới hạn mức $\dfrac{\alpha}{2}$ của phân bố chuẩn tắc $\mathbf{N}(0;1)$ với $\alpha=1-\beta$.

Độ chính xác của khoảng tin cậy:

$$\varepsilon=U_{\alpha/2}\sqrt{\frac{f(1-f)}{n}}. \tag{5.39}$$

Với độ tin cậy β và độ chính xác ε_0 cho trước thì *kích thước mẫu cần thiết* là số tự nhiên n nhỏ nhất thỏa mãn:

$$n\ge f(1-f)\left(\frac{U_{\alpha/2}}{\varepsilon_0}\right)^2 \tag{5.40}$$

trong đó f là tần suất mẫu của một mẫu ngẫu nhiên nào đó.

Ví dụ 5.10: 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 (5.37) với điều kiện (5.38).

Từ mẫu cụ thể trên ta có $f=\dfrac{960}{1600}=0,6$ thỏa mãn điều kiện $\begin{cases} nf=960>10 \\ n(1-f)=640>10 \end{cases}$.

Độ chính xác của ước lượng $\varepsilon = U_{\alpha/2}\sqrt{\dfrac{f(1-f)}{n}} = 1{,}96\sqrt{\dfrac{0{,}6\cdot 0{,}4}{1600}} = 0{,}024$.

Khoảng tin cậy: $0{,}576 \le p \le 0{,}624$.

Vậy với độ tin cậy 95% thì tối thiểu có 57,6% cử tri sẽ bỏ phiếu cho ứng cử viên A.

5.3. KIỂM ĐỊNH GIẢ THIẾT THỐNG KÊ

5.3.1. Khái niệm chung

Trong mục 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 mục này ta sẽ nghiên cứu bài toán kiểm định giả thiết về các tham số đặc trưng của tổng thể.

Trước hết ta tìm hiểu các khái niệm sau.

5.3.1.1. Giả thiết thống kê

Giả thiết thống kê là giả thiết về dạng phân bố xác suất, các đặc trưng tham số của biến ngẫu nhiên gốc hoặc giả thiết về sự độc lập của các biến ngẫu nhiên gốc.

Giả thiết đưa ra kiểm nghiệm được ký hiệu là H_0, gọi là *"giả thiết không"*. Đó là giả thiết mà ta nghi ngờ muốn bác bỏ hoặc giả thiết ta muốn bảo vệ. 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 $\mu \ne 1000$, 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:

5.3.1.2. Tiêu chuẩn kiểm định giả thiết thống kê

Từ biến ngẫu nhiên gốc X của tổng thể lập mẫu ngẫu nhiên $W = (X_1, X_2, ..., X_n)$. Chọn thống kê

$$T = T(X_1, X_2, ..., X_n, \theta_0) \tag{5.41}$$

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.*

5.3.1.3. 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:

$$P\{T \in W_\alpha | H_0\} = \alpha \tag{5.42}$$

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.

5.3.1.4. 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 (5.41) ta được giá trị quan sát của tiêu chuẩn kiểm định:

$$T_{qs} = (x_1, x_2, ..., x_n, \theta_0) \tag{5.43}$$

5.3.1.5. 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).

5.3.1.6. 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 giả thiết H_0 trong khi H_0 sai, điều này xảy ra khi giá trị quan sát T_{qs} không thuộc miền bác bỏ W_α trong khi H_1 đúng. Vậy xác suất sai lầm loại II là β xác định như sau:

$$P\left\{T \notin W_\alpha \middle| \mathrm{H}_1\right\} = \beta \tag{5.44}$$

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:

$$P\{T \in W_\alpha | \mathrm{H}_0\} = \alpha \text{ và } P\{T \in W_\alpha | \mathrm{H}_1\} = 1-\beta \to \max$$

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.

5.3.1.7. 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.

5.3.2. Kiểm định giả thiết về kỳ vọng của biến ngẫu nhiên có phân bố chuẩn

Giả sử biến ngẫu nhiên gốc X trong tổng thể có phân bố chuẩn $N(\mu;\sigma^2)$, 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:

5.3.2.1. 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: $W = \left(X_1, X_2, ..., X_n\right)$.

Xét thống kê

$$T = \frac{(\overline{X} - \mu_0)\sqrt{n}}{\sigma} \tag{5.45}$$

Nếu giả thiết H_0 đúng, theo công thức (5.19) 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
$$W_\alpha = \left\{ T = \frac{(\overline{X} - \mu_0)\sqrt{n}}{\sigma}\ ;\ |T| > U_{\alpha/2} \right\}. \tag{5.46}$$

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
$$W_\alpha = \left\{ T = \frac{(\overline{X} - \mu_0)\sqrt{n}}{\sigma}\ ;\ T > U_\alpha \right\}. \tag{5.47}$$

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
$$W_\alpha = \left\{ T = \frac{(\overline{X} - \mu_0)\sqrt{n}}{\sigma}\ ;\ -T > U_\alpha \right\}. \tag{5.48}$$

trong đó $U_{\alpha/2}$, U_{α} lần lượt 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.

5.3.2.2. Trường hợp chưa biết phương sai, kích thước mẫu $n \geq 30$

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 mục 5.2.6.2 ta có thống kê

$$T=\frac{(\overline{X}-\mu_0)\sqrt{n}}{S} \tag{5.49}$$

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 $$W_{\alpha}=\left\{T=\frac{(\overline{X}-\mu_0)\sqrt{n}}{S}\ ;\ |T|>U_{\alpha/2}\right\}. \tag{5.50}$$

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 $$W_{\alpha}=\left\{T=\frac{(\overline{X}-\mu_0)\sqrt{n}}{S}\ ;\ T>U_{\alpha}\right\}. \tag{5.51}$$

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 $$W_{\alpha}=\left\{T=\frac{(\overline{X}-\mu_0)\sqrt{n}}{S}\ ;\ -T>U_{\alpha}\right\}. \tag{5.52}$$

Ví dụ 5.11: 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 α = 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: $T=\dfrac{(\overline{X}-7{,}4)\sqrt{n}}{S}$

$\alpha=0{,}01 \Rightarrow U_{\frac{\alpha}{2}}=2{,}575 \Rightarrow$ Miền bác bỏ: $W_{\alpha}=\left\{T=\dfrac{(\overline{X}-7{,}4)\sqrt{n}}{S}\ ;\ |T|>2{,}575\right\}$.

Từ mẫu cụ thể ta có $T_{qs} = \dfrac{(6,1-7,4)\sqrt{40}}{2,5} = -3,289$.

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 5.12: Một công ti có một hệ thống máy tính có thể xử lí 1200 hóa đơn trong một giờ. Công ti mới nhập một hệ thống máy tính mới. Hệ thống này khi chạy kiểm tra trong 40 giờ cho thấy số hóa đơn được xử lí trung bình trong 1 giờ là 1260 với độ lệch chuẩn 215. Với mức ý nghĩa 5% hãy nhận định xem hệ thống mới có tốt hơn hệ thống cũ hay không?

Giải: Gọi μ là số hóa đơn trung bình mà hệ thống máy tính mới xử lí được trong 1 giờ.

Ta kiểm định: Giả thiết H_0 : $\mu = 1200$; Đối thiết H_1: $\mu > 1200$.

Tiêu chuẩn kiểm định: $T = \dfrac{(\overline{X}-1200)\sqrt{n}}{S}$

$\alpha = 0,05 \Rightarrow U_\alpha = 1,64 \Rightarrow$ Miền bác bỏ: $W_\alpha = \left\{ T = \dfrac{(\overline{X}-1200)\sqrt{n}}{S} \ ;\ T > 1,64 \right\}$.

Thay giá trị cụ thể của mẫu vào công thức (5.49) ta được $T_{qs} = \dfrac{(1260-1200)\sqrt{40}}{215} = 1,76$.

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.

5.3.2.3. Trường hợp chưa biết phương sai, kích thước mẫu $n < 30$

Giả sử giả thiết H_0 đúng, xét thống kê

$$T = \frac{(\overline{X}-\mu_0)\sqrt{n}}{S} \qquad \textbf{(5.53)}$$

theo công thức (5.22) 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: $$W_\alpha = \left\{ T = \frac{(\overline{X}-\mu_0)\sqrt{n}}{S} \ ;\ |T| > t_{\alpha/2}(n-1) \right\}. \qquad \textbf{(5.54)}$$

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: $$W_\alpha = \left\{ T = \frac{(\overline{X}-\mu_0)\sqrt{n}}{S} \ ;\ T > t_{\alpha}(n-1) \right\}. \qquad \textbf{(5.55)}$$

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: $$W_\alpha = \left\{ T = \frac{(\overline{X} - \mu_0)\sqrt{n}}{S} \; ; \; -T > t_\alpha(n-1) \right\}. \qquad \textbf{(5.56)}$$

Ví dụ 5.13: 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 hạt 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 α = 0,05. Biết rằng năng suất giống cây trồng là một biến ngẫu nhiên có phân bố chuẩn $\mathbf{N}(\mu;\sigma^2)$.

Giải: Gọi μ là năng suất trung bình của loại giống mới.

Ta cần kiểm định giả thiết H_0: $\mu = 21{,}5$; Đối thiết H_1: $\mu \neq 21{,}5$.

Tiêu chuẩn kiểm định: $T = \dfrac{(\overline{X} - 21{,}5)\sqrt{n}}{S}$.

Tra bảng ta tính được giá trị tới hạn mức $\dfrac{\alpha}{2} = 0{,}025$ của phân bố Student 15 bậc tự do là 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:

$$\overline{x} = 20{,}406, \quad s = 3{,}038 \;\Rightarrow T_{qs} = \frac{(20{,}406 - 21{,}5)\sqrt{16}}{3{,}038} = -1{,}44.$$

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.

5.3.3. Kiểm định giả thiết về tần suất

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 mục 5.1.5.5 và mục 5.2.7, 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 (5.13)-(5.15)). Xét thống kê

$$T = \frac{(f - p_0)\sqrt{n}}{\sqrt{p_0(1 - p_0)}} \qquad \textbf{(5.57)}$$

Nếu giả thiết H_0 đúng, áp dụng định lý giới hạn trung tâm (công thức (5.24)) thì khi n đủ lớn thống kê trên xấp xỉ phân bố chuẩn tắc $\mathbf{N}(0;1)$. Trong thực tế khi

$$\begin{cases} np_0 > 5 \\ n(1 - p_0) > 5 \end{cases} \qquad \textbf{(5.58)}$$

thì có thể xem thống kê (5.57) có phân bố chuẩn tắc $\mathbf{N}(0;1)$. 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: $p \neq p_0$.

$$W_\alpha = \left\{ T = \frac{(f - p_0)\sqrt{n}}{\sqrt{p_0(1 - p_0)}} \;;\; |T| > U_{\frac{\alpha}{2}} \right\} \qquad \textbf{(5.59)}$$

b) H_0: $p = p_0$; H_1: $p > p_0$.

$$W_\alpha = \left\{ T = \frac{(f - p_0)\sqrt{n}}{\sqrt{p_0(1 - p_0)}} \;;\; T > U_\alpha \right\} \qquad \textbf{(5.60)}$$

c) H_0: $p = p_0$; H_1: $p < p_0$.

$$W_\alpha = \left\{ T = \frac{(f - p_0)\sqrt{n}}{\sqrt{p_0(1 - p_0)}} \;;\; -T > U_\alpha \right\} \qquad \textbf{(5.61)}$$

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 5.14: 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 α = 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 (5.57). 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 $\left|T_{qs}\right| < 1{,}96$. Vậy không có cơ sở để bác bỏ H_0.

CÂU HỎI ÔN TẬP VÀ BÀI TẬP

5.1 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 ☐.

5.2 Một thống kê của mẫu ngẫu nhiên là con số cụ thể về dấu hiệu nghiên cứu.

Đúng ☐ Sai ☐.

5.3 Trung bình mẫu của dấu hiệu nghiên cứu có phân bố chuẩn cũng có phân bố chuẩn.

Đúng ☐ Sai ☐.

5.4 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 ☐.

5.5 Trung bình mẫu là ước lượng vững và hiệu quả của kỳ vọng của biến ngẫu nhiên gốc.

Đúng ☐ Sai ☐.

5.6 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 ☐.

5.7 Tổng của hai ước lượng không chệch là một ước lượng không chệch.

Đúng ☐ Sai ☐.

5.8 Phương sai mẫu hiệu chỉnh S^2 là ước lượng vững không chệch của phương sai của biến ngẫu nhiên gốc.

Đúng ☐ Sai ☐.

5.9 Hai đầu mút của khoảng tin cậy là hai thống kê của mẫu.

Đúng ☐ Sai ☐.

5.10 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 ☐.

5.11 Giả thiết thống kê là giả thiết do nhà thống kê đặt ra cho mẫu ngẫu nhiên.

Đúng ☐ Sai ☐.

5.12 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 ☐.

5.13 Qui tắc kiểm định dựa trên nguyên lý xác suất nhỏ và phép chứng minh phản chứng.

Đúng ☐ Sai ☐.

5.14 Sai lầm loại 1 là sai lầm gặp phải khi thực tế giả thiết đúng nhưng ta bác bỏ.

Đúng ☐ Sai ☐.

5.15 Sai lầm loại 2 luôn luôn lớn hơn sai lầm loại 1.

Đúng ☐ Sai ☐.

5.16 Miền bác bỏ là miền có xác suất rất bé nên ta có thể bỏ qua trong mọi phép kiểm định.

Đúng ☐ Sai ☐.

5.17 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 ☐.

5.18 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 ☐.

5.19 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.

5.20 Giả sử biến ngẫu nhiên gốc có phân bố chuẩn $\mathbf{N}(20;1)$. Chọn mẫu ngẫu nhiên kích thước $n = 100$. Hãy tính xác suất để trung bình mẫu $\overline{X}$ nằm trong khoảng: $19,8 < \overline{X} < 20,2$.

5.21 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.

Tính $\overline{x}$, s^2, s.

5.22 Trong đợt vận động bầu cử tổng thống ở một nước nọ, người ta phỏng vấn ngẫu nhiên 2000 cử tri thì được biết có 1082 người trong số đó sẽ bỏ phiếu cho ứng cử viên A. Với độ tin cậy 98% tối thiểu ứng cử viên A sẽ chiếm được bao nhiêu % số phiếu bầu? Cho biết phân vị mức 0,975 của phân bố chuẩn tắc N(0;1) là 1,96.

5.23 Để xác định sản lượng khai thác điện thoại của đơn vị mình, một đơn vị đã tiến hành thống kê ngẫu nhiên 35 ngày và thu được kết quả sau với đơn vị 100.000 phút/ngày:

0,84 0,96 1,02 1,08 0,88 0,80 0,91 0,97 1,07 0,98 1,04 1,13 0,87 0,82
1,01 0,93 1,03 1,10 0,97 1,05 0,83 0,76 0,95 1,15 1,00 1,05 1,14 0,89
0,81 0,95 1,20 1,16 1,24 0,79 0,77.

Tìm khoảng tin cậy 95% cho sản lượng điện thoại trung bình mỗi ngày.

5.24 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.

5.25 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.

5.26 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 (gram)	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.

5.27 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.

5.28 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 $\alpha = 0,025$ hãy kết luận về ý định nói trên.

5.29 Mức hao phí xăng của một loại ô tô 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 ô tô tương ứng	4	10	9	3	2

Với mức ý nghĩa $\alpha = 0,025$ hãy kết luận về điều nghi ngờ nói trên.

5.30 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?

CHƯƠNG VI: QUÁ TRÌNH NGẪU NHIÊN, CHUỖI MARKOV

GIỚI THIỆU

Hầu hết các hiện tượng xảy ra trong tự nhiên và xã hội đều có tính chất ngẫu nhiên, điều đó phản ánh các mối ràng buộc phức tạp mà ta không biết trước được. Trong các chương trước chúng ta đã tìm hiểu khái niệm biến ngẫu nhiên, véc tơ ngẫu nhiên, đó là các biến nhận các giá trị nào đó phụ thuộc vào các yếu tố ngẫu nhiên. Khi họ các biến ngẫu nhiên phụ thuộc vào thời gian ta có quá trình ngẫu nhiên.

Các tín hiệu truyền dẫn và nhiễu của một hệ thống viễn thông, quá trình sắp hàng ở một tổng đài... là các quá trình ngẫu nhiên. Quá trình ngẫu nhiên có nhiều ứng dụng trong viễn thông là quá trình có tính Markov (memoryless) và quá trình dừng.

Chuỗi Markov là một quá trình Markov có không gian trạng thái rời rạc, thời gian rời rạc và thuần nhất. Chuỗi Markov thường gặp trong bài toán chuyển mạch của hệ thống viễn thông.

Quá trình Poisson là một ví dụ về chuỗi Markov với thời gian liên tục. Quá trình Poisson $X(t)$ mô tả quá trình đếm số lần xuất hiện một biến cố A nào đó cho đến thời điểm t. Quá trình Poisson được ứng dụng nhiều trong viễn thông, liên quan đến bài toán truyền tín hiệu, các hệ phục vụ, bài toán chuyển mạch ...

Tín hiệu viễn thông, nhiễu không có tính Markov. Các quá trình này quá khứ của nó có ảnh hưởng lớn đến sự tiến triển của quá trình trong tương lai. Tuy nhiên hàm trung bình không đổi và hàm tương quan thuần nhất theo thời gian, đó quá trình dừng. Khi các quá trình dừng là các tín hiệu hoặc nhiễu thì biến đổi Fourier của hàm tương quan của quá trình là hàm mật độ phổ công suất của tín hiệu hoặc nhiễu.

Một trong những bài toán quan trọng của lý thuyết chuyển mạch là vấn đề xung đột thông tin, nghẽn mạch hoặc rớt cuộc gọi. Lý thuyết quá trình sắp hàng (queueing process) xác định và tìm các phương án tối ưu để hệ thống phục vụ tốt nhất.

Trong chương này ta chỉ nghiên cứu một cách khái quát khái niệm quá trình ngẫu nhiên và chuỗi Markov thời gian rời rạc thuần nhất.

Quá trình Poisson, quá trình dừng và lý thuyết sắp hàng sẽ được khảo sát trong giáo trình toán chuyên ngành điện tử viễn thông.

Để học tốt chương này học viên cần nắm vững khái niệm xác suất, xác suất có điều kiện, công thức xác suất đầy đủ, biến ngẫu nhiên và kiến thức đại số tuyến tính: ma trận, hệ phương trình tuyến tính.

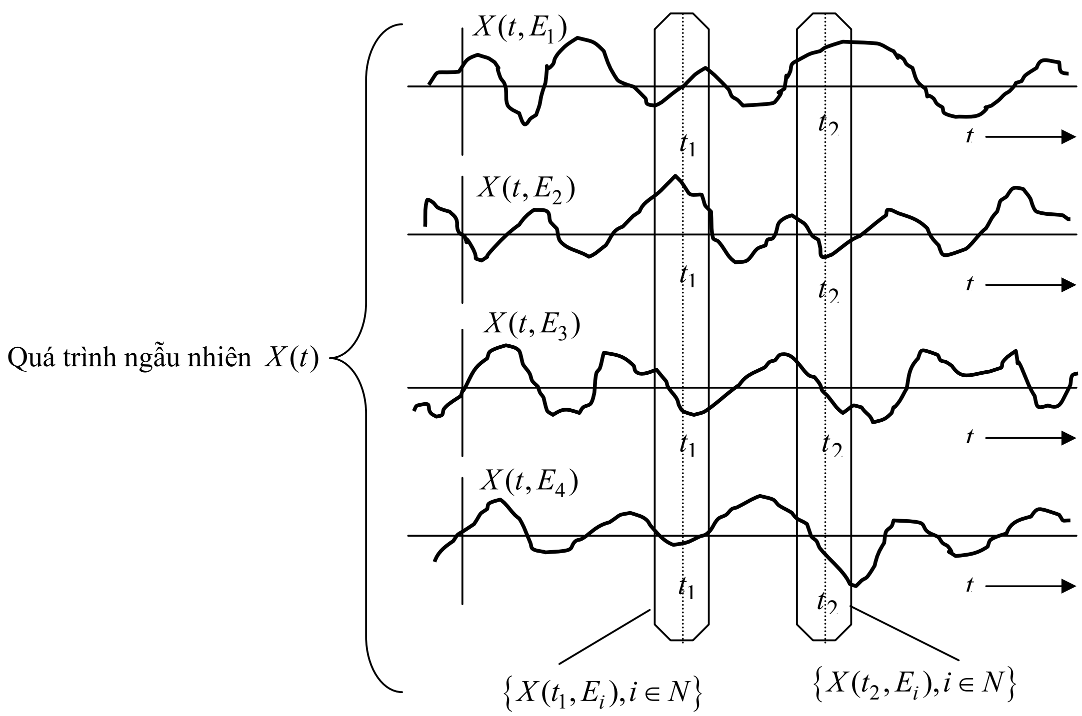
NỘI DUNG

6.1. KHÁI NIỆM VÀ PHÂN LOẠI QUÁ TRÌNH NGẪU NHIÊN

6.1.1. Khái niệm quá trình ngẫu nhiên

Các tín hiệu của các hệ thống thông tin là các tín hiệu ngẫu nhiên vì ngoài thành phần mang tin còn có sự tác động của giao thoa ngẫu nhiên và nhiễu của thiết bị.

Giả sử một tín hiệu nào đó mà tại mỗi thời điểm t chỉ xảy ra ứng với hệ các biến cố $\{E_i, i \in N\}$ của phép thử. Tín hiệu này nhận giá trị là $X(t, E_i)$ tại thời điểm t và khi biến cố E_i xảy ra. Như vậy $X(t, E_i)$ là một mẫu của quá trình ngẫu nhiên $X(t)$. Quá trình ngẫu nhiên $X(t)$ vừa phụ thuộc thời gian t, vừa phụ thuộc yếu tố ngẫu nhiên E_i.



Một cách tổng quát một quá trình ngẫu nhiên là một họ các biến ngẫu nhiên $\{X(t, \omega); t \in I\}$. Các quá trình này vừa phụ thuộc vào thời gian t và khi cố định tham số t thì $X(t, \omega)$ là biến ngẫu nhiên theo ω. Tập chỉ số I thường biểu diễn tham số thời gian.

Để đơn giản trong cách viết người ta ký hiệu quá trình ngẫu nhiên $\{X(t); t \in I\}$ thay cho $\{X(t, \omega); t \in I\}$.

Các tín hiệu video, tín hiệu thoại, dữ liệu máy tính, nhiễu điện trong các thiết bị điện, số khách hàng đến một điểm phục vụ, chỉ số chứng khoán trong thị trường chứng khoán… là các quá trình ngẫu nhiên.

6.1.2. Phân loại quá trình ngẫu nhiên

Các yếu tố chính để phân biệt các quá trình ngẫu nhiên là:

- Không gian trạng thái,
- Tập chỉ số thời gian I ,
- Quan hệ độc lập giữa các biến ngẫu nhiên $X(t)$.

Vì vậy ta có thể phân loại quá trình ngẫu nhiên theo:

6.1.2.1. Tập trạng thái E

Ta ký hiệu E là tập các giá trị của $X(t)$ và gọi là không gian trạng thái của quá trình.

♦ Nếu E là tập đếm được thì $\{X(t); t\in I\}$ gọi là quá trình có trạng thái rời rạc.

♦ Nếu E là 1 khoảng của tập số thực $\mathbb{R}$ thì $\{X(t); t\in I\}$ là quá trình thực.

♦ Nếu E tập con của tập số phức $\mathbb{C}$ thì $\{X(t); t\in I\}$ là quá trình phức.

♦ Nếu $E=\mathbb{R}^k$ thì $\{X(t); t\in I\}$ là quá trình k-véc tơ.

6.1.2.2. Tập các chỉ số I

❖ Nếu $I\subset \mathbb{Z}$ thì quá trình $\{X(t); t\in I\}$ được gọi là quá trình có thời gian rời rạc. Trường hợp này ta ký hiệu $X(n)$ thay cho $X(t)$.

❖ Nếu $I=[0;\infty)$ hoặc $I=\mathbb{R}$ thì $\{X(t); t\in I\}$ được gọi là quá trình có thời gian liên tục.

6.1.2.3. Quan hệ độc lập

Quá trình $\{X(t); t\in I\}$ được gọi là:

a) *Quá trình có gia số độc lập* nếu với mọi cách chọn $t_1<t_2<....<t_n$ thì các biến ngẫu nhiên sau là độc lập

$$X(t_2)-X(t_1), X(t_3)-X(t_2), ..., X(t_n)-X(t_{n-1}). \qquad \textbf{(6.1)}$$

Đặc biệt với quá trình thời gian rời rạc $\{X(n)\}$ thì tính chất gia số độc lập dẫn đến dãy các biến ngẫu nhiên $Z_0=X(0)$, $Z_i=X(i)-X(i-1)$; $i=1,2,...$ là độc lập. Ngoài ra nếu ta biết luật phân bố của từng biến ngẫu nhiên Z_0, Z_1, ... thì ta biết được luật phân bố của mọi $X(i)$, $i=0, 1, ...$. Thật vậy, điều này được suy từ công thức (3.24)-(3.27) và

$$X(i)=Z_0+Z_1+...+Z_i.$$

b) *Quá trình Martingal* nếu với bất kỳ $t_1<t_2<....<t_{n+1}$ và $a_1, a_2, ..., a_n$ thì

$$\mathrm{E}\big(X(t_{n+1})\big|X(t_1)=a_1,...,X(t_n)=a_n\big)=a_n. \qquad \textbf{(6.2)}$$

Martingal có thể xem như là mô hình mô tả trò chơi may rủi, trong đó $X(t)$ là số tiền của người chơi ở thời điểm t. Tính chất Martingal nói rằng *số tiền trung bình* của người chơi sẽ có ở thời điểm t_{n+1} bằng số tiền anh ta có ở thời điểm t_n và *không phụ thuộc vào những gì anh ta có trước đó trong quá khứ.*

Nếu $\{X(t);t\geq 0\}$ là quá trình gia số độc lập với kỳ vọng bằng 0 thì $\{X(t);t\geq 0\}$ là Martingal với thời gian liên tục.

c) Quá trình Markov nếu với mọi thời điểm $t_1<t_2<....<t_n$, với mọi giá trị cho trước $a_1,a_2,....,a_n$, với mọi thời điểm $t>t_n$ và với mọi $a<b$ thì

$$P\left\{a\leq X(t)<b\middle|X(t_1)=a_1,...,X(t_n)=a_n\right\}=P\left\{a\leq X(t)<b\middle|X(t_n)=a_n\right\}. \quad \textbf{(6.3)}$$

Nghĩa là qui luật xác suất trong tương lai chỉ phụ thuộc hiện tại và độc lập với quá khứ. Nói cách khác *quá trình Markov mô tả các hệ không có trí nhớ* (memoryless).

Với mọi $t>s;$ với mọi tập giá trị $A\subset\mathbb{R}$ và giá trị a ta ký hiệu

$$p(s,a;t,A)=P\left\{X(t)\in A\middle|X(s)=a\right\} \quad \textbf{(6.4)}$$

và gọi là *hàm xác suất chuyển từ thời điểm s đến thời điểm t.*

Như vậy công thức (6.3) được viết lại

$$P\left\{a\leq X(t)<b\middle|X(t_1)=a_1,...,X(t_n)=a_n\right\}=p(t_n,a_n;t,A)\text{, trong đó } A=[a;b).$$

d) Quá trình dừng (stationary)

Quá trình $\{X(t);t\in I\}$, $I=\mathbb{R},\mathbb{R}_+,\mathbb{Z},\mathbb{N}$ được gọi là:

- *Dừng theo nghĩa chặt* (strictly stationary) nếu $\forall h>0,\forall t_1,t_2,....,t_n\in I$ thì hàm phân bố đồng thời của $\left(X(t_1+h),X(t_2+h),...,X(t_n+h)\right)$ và $\left(X(t_1),X(t_2),...,X(t_n)\right)$ là như nhau.

  Nói riêng mọi $X(t)$ có cùng phân bố.

- *Dừng theo nghĩa rộng hay dừng hiệp phương sai* (wide sense stationary or covariance stationary) nếu

  i) $\mathrm{E}X(t)=m$=const

  ii) Với mọi t, $\mathrm{E}\left(X(t),X(t+\tau)\right)$ chỉ phụ thuộc τ.

  Đặt

$$K_x(\tau)=\mathrm{E}\left(X(t),X(t+\tau)\right) \quad \textbf{(6.5)}$$

và gọi là *hàm tự tương quan* của quá trình $\{X(t);t\in I\}$.

6.2. CHUỖI MARKOV

Chuỗi Markov là quá trình Markov $\{X(t); t \in I\}$ *có không gian trạng thái* E *đếm được.*

Tuỳ theo tập chỉ số $I = \{0,1,2,...\}$ hoặc $I = (0;\infty)$ ta có tương ứng chuỗi Markov với thời gian rời rạc hoặc liên tục.

Với chuỗi Markov công thức xác suất chuyển (6.4) được viết cụ thể

$$p(s,i;t,j) = P\{X(t) = j | X(s) = i\},\ t > s. \quad \textbf{(6.6)}$$

Nếu xác suất chuyển chỉ phụ thuộc vào $t - s$ nghĩa là

$$p(s,i;t,j) = p(s+h,i;t+h,j) \quad \textbf{(6.7)}$$

với mọi h, thì ta nói *quá trình là thuần nhất theo thời gian*.

6.2.1. Chuỗi Markov với thời gian rời rạc thuần nhất

Định nghĩa 6.1. *Quá trình* $\{X(n),\ n = 0,1,2,...\}$ *với thời gian rời rạc được gọi là chuỗi Markov thời gian rời rạc thuần nhất nếu*

i) Không gian trạng thái E của mọi $X(n)$ *là tập đếm được.*

ii) Hàm xác suất chuyển là thuần nhất theo thời gian, nghĩa là thoả mãn (6.7).

Ta nói tắt chuỗi Markov thay cho chuỗi Markov thời gian rời rạc thuần nhất.

6.2.2. Ma trân xác suất chuyển

Giả sử $\{X(n),\ n = 0,1,2,...\}$ là chuỗi Markov thời gian rời rạc có không gian trạng thái E. Các phần tử của E được ký hiệu $i, j, k ...$

Với mọi $i, j \in E$; đặt

$$p_{ij} = P\{X(n+1) = j | X(n) = i\} = P\{X(1) = j | X(0) = i\} \quad \textbf{(6.8)}$$

không phụ thuộc vào n. Đó là *xác suất để từ trạng thái* i *sau một bước sẽ chuyển thành trạng thái* j.

Đặt

$$p_{ij}^{(k)} = P\{X(n+k) = j | X(n) = i\} = P\{X(k) = j | X(0) = i\}. \quad \textbf{(6.9)}$$

Đó là *xác suất để từ trạng thái* i *sau* k *bước sẽ chuyển thành trạng thái* j.

Định nghĩa 6.1: *Ma trận vuông* $P = \left[p_{ij} \right]$ *gọi là ma trận xác suất chuyển sau 1 bước.*

Ma trận vuông $P^{(k)} = \left[p_{ij}^{(k)} \right]$ *gọi là ma trận xác suất chuyển sau* k *bước.*

Ký hiệu $P^{(0)} = I,\ P^{(1)} = P,\ I$ là ma trận đơn vị.

Chú ý rằng số hàng, số cột của P và $P^{(k)}$ bằng số phần tử của không gian trạng thái E. Vì vậy số hàng, số cột của P và $P^{(k)}$ có thể vô hạn nếu không gian trạng thái E có vô số đếm được các phần tử.

Định lý 6.1: Với mọi $n \ge 0$, ta có phương trình Chapman - Kolmogorov:

$$P^{(n+1)} = PP^{(n)} = P^{(n)}P \tag{6.10}$$

Từ đó suy ra

$$P^{(n)} = P^n. \tag{6.11}$$

Chứng minh: 1) Áp dụng công thức xác xuất đầy đủ (1.16) ta có

$$\begin{aligned} p_{ij}{}^{(n+1)} &= P\left\{X(n+1)=j \,\middle|\, X(0)=i\right\} \\ &= \sum_{k\in E} P\left\{X(n+1)=j \,\middle|\, X(0)=i, X(1)=k\right\} P\left\{X(1)=k \,\middle|\, X(0)=i\right\} \\ &= \sum_{k\in E} P\left\{X(n+1)=j \,\middle|\, X(1)=k\right\} P\left\{X(1)=k \,\middle|\, X(0)=i\right\} = \sum_{k\in E} p_{ik}p_{kj}{}^{(n)} \end{aligned}$$

$\Rightarrow \; P^{(n+1)} = PP^{(n)}$.

Ta cũng có

$$\begin{aligned} p_{ij}{}^{(n+1)} &= P\left\{X(n+1)=j \,\middle|\, X(0)=i\right\} \\ &= \sum_{k\in E} P\left\{X(n+1)=j \,\middle|\, X(0)=i, X(n)=k\right\} P\left\{X(n)=k \,\middle|\, X(0)=i\right\} \\ &= \sum_{k\in E} P\left\{X(n+1)=j \,\middle|\, X(n)=k\right\} P\left\{X(n)=k \,\middle|\, X(0)=i\right\} = \sum_{k\in E} p_{ik}{}^{(n)}p_{kj} \end{aligned}$$

$\Rightarrow \; P^{(n+1)} = P^{(n)}P$.

2) Từ 1) suy ra $P^{(2)} = PP = P^2$, quy nạp ta có $P^{(n)} = P^n$ với mọi $n = 0,1,2,\ldots$.

Đặt

$$p_j^{(n)} = P\left\{X(n)=j\right\},\; n = 0,1,2,\ldots. \tag{6.12}$$

Ma trận hàng $\prod^{(n)} = \left[p_j^{(n)}\right]$ gọi là phân bố của hệ tại thời điểm n.

Khi $n = 0$, ma trận $\prod = \prod^{(0)}$ được gọi là phân bố ban đầu.

Định lý 6.2: Với mọi $n \ge 0$, $m \ge 0$:

$$\prod^{(n)} = \prod P^{(n)}; \tag{6.13}$$

$$\prod^{(n+1)} = \prod^{(n)} P. \tag{6.14}$$

$$\prod^{(n+m)} = \prod^{(n)} P^{(m)}. \tag{6.15}$$

Chứng minh: Từ định lý 6.1 ta suy ra 3 điều trên là tương đương. Vì vậy để chứng minh định lý 6.2 ta chỉ cần chứng minh (6.15).

$$p_j^{(n+m)} = P\{ X(n+m) = j\} = \sum_{k\in E} P\{X(n)=i\} P\{X(n+m)=j \big| X(n)=i\} = \sum_{i\in E} p_i^{(n)} p_{ij}^{(m)}.$$

Vậy chuỗi Markov rời rạc thuần nhất hoàn toàn được xác định bởi bộ ba $(X(n), \prod, P)$ trong đó $X(n)$ là dãy các biến ngẫu nhiên rời rạc.

Ví dụ 6.1: (Mô hình chuỗi Markov về kế hoạch chăm sóc răng miệng).

Để đánh giá về các chương trình chăm sóc răng miệng, Freed và tập thể (1979) đã phân tích các bệnh án của 578 bệnh nhân đến bệnh viện chuyên khoa răng Venice ở Los Angeles từ tháng 7 năm 1971 đến 30 tháng 6 năm 1972. Các trạng thái được xác định như sau: E (episodic care) là nhóm các bệnh nhân phải đến điều trị qua nhiều giai đoạn, I (initial care) là nhóm bệnh nhân đến điều trị một lần và dứt điểm, M (maintenance care) là nhóm các bệnh nhân chỉ cần chăm sóc phòng ngừa và cuối cùng là nhóm N (nonuse care) không cần chăm sóc. Phân tích các số liệu dựa trên kết quả 3 năm nói trên, Freed và đồng nghiệp đã tìm ra ma trận xác suất chuyển 1-năm:

$$P = \begin{array}{c} \\ M \\ I \\ E \\ N \end{array} \begin{array}{c} \begin{array}{cccc} M & \quad I & \quad E & \quad N \end{array} \\ \begin{bmatrix} 0,346 & 0,014 & 0,003 & 0,637 \\ 0,372 & 0,110 & 0,010 & 0,508 \\ 0,191 & 0,152 & 0,095 & 0,562 \\ 0,003 & 0,053 & 0,012 & 0,932 \end{bmatrix} \end{array}$$

Giả sử hiện tại có 14,5% bệnh nhân thuộc nhóm E và 85,5% thuộc nhóm I, sử dụng công thức (6.11), (6.13) ta có thể tính tỉ lệ bệnh nhân thuộc các nhóm trong các năm sau cho trong bảng sau:

Nhóm bệnh nhân	Năm 1	Năm 2	Năm 3	Năm 4	Năm 5
M	34,5	16,8	8,0	5,1	4,2
I	11,5	4,8	5,0	5,4	5,5
E	2,3	1,1	1,1	1,2	1,2
N	51,6	77,3	85,5	88,3	89,0

Ví dụ 6.2: (Mô hình hòa nhập cộng đồng của các bệnh nhân tâm thần được xuất viện).

Các chuyên gia y tế thường tránh chuyển các bệnh nhân tâm thần lâu năm được xuất viện trực tiếp từ bệnh viện đến với cộng đồng. Chẳng hạn ở Billings, Montana, người ta thực hiện như sau: Trước hết người ta chuyển bệnh nhân đến ở tại khu vực được chăm sóc 24/24 giờ. Nếu tình trạng sức khỏe của bệnh nhân tiến triển tốt đáp ứng những tiêu chí đòi hỏi thì được chuyển đến nhóm 40 giờ, tức là được chăm sóc 5 ngày trong 1 tuần và 1 ngày 8 giờ. Nếu tình trạng bệnh nhân tiếp tục tiến triển tốt thì sẽ được đưa đến nhóm có sự tương tác giao tiếp cao hơn, ở đây bệnh nhân được luyện tập tự chủ hành vi của mình. Cuối cùng khi được coi là khỏi bệnh hoàn toàn thì được đưa về hòa nhập với cộng đồng.

Drachman (1981) đã phân tích các dữ liệu thu thập được ở Billings từ 1/1/1978 đến 31/5/1979 và thu được ma trận xác suất chuyển như sau:

$$P = \begin{array}{cc} & \begin{array}{cccccc} H & \quad I & \quad 24 & \quad 40 & \quad A & \quad C \end{array} \\ \begin{array}{c} H \\ I \\ 24 \\ 40 \\ A \\ C \end{array} & \begin{bmatrix} 0,7143 & 0,0714 & 0,0714 & 0,0000 & 0,0000 & 0,1429 \\ 0,1177 & 0,0588 & 0,2941 & 0,1177 & 0,0000 & 0,4118 \\ 0,0109 & 0,0109 & 0,7283 & 0,0652 & 0,0000 & 0,1848 \\ 0,0213 & 0,0213 & 0,0213 & 0,7660 & 0,0426 & 0,1277 \\ 0,0000 & 0,0172 & 0,0172 & 0,0172 & 0,7931 & 0,1552 \\ 0,0136 & 0,0442 & 0,0578 & 0,0034 & 0,0272 & 0,8537 \end{bmatrix} \end{array}$$

trong đó các trạng thái H (ở bệnh viện), I (bắt đầu chuyển khỏi bệnh viện), 24 (nhóm chăm sóc 24/24), 40 (nhóm chăm sóc 40 giờ/1 tuần), A (nhóm được tương tác giao tiếp) và C (nhóm được đưa về cộng đồng). Ở đây 12 tuần được qui tròn thành 3 tháng.

Áp dụng công thức (6.11) ta tính được

$$P^6 = \begin{array}{cc} & \begin{array}{cccccc} H & \quad I & \quad 24 & \quad 40 & \quad A & \quad C \end{array} \\ \begin{array}{c} H \\ I \\ 24 \\ 40 \\ A \\ C \end{array} & \begin{bmatrix} 0,1723 & 0,0424 & 0,2002 & 0,0585 & 0,0372 & 0,4894 \\ 0,0678 & 0,0359 & 0,2032 & 0,1010 & 0,0600 & 0,5323 \\ 0,0454 & 0,0323 & 0,2539 & 0,1167 & 0,0507 & 0,5010 \\ 0,0548 & 0,0330 & 0,1256 & 0,2373 & 0,1046 & 0,4447 \\ 0,0282 & 0,0313 & 0,1180 & 0,0592 & 0,2870 & 0,4762 \\ 0,0489 & 0,0374 & 0,1758 & 0,0548 & 0,0758 & 0,6073 \end{bmatrix} \end{array}.$$

Dữ liệu ban đầu $O_1 = \begin{bmatrix} 1 & 0 & 15 & 8 & 10 & 53 \end{bmatrix}$, áp dụng công thức (6.13) tính được

$$e_7 = O_1 P^6 = \begin{bmatrix} 4,17 & 3,09 & 15,51 & 7,20 & 8,52 & 48,51 \end{bmatrix},$$

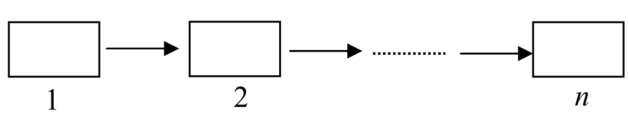
ở đây 17 tháng được làm tròn thành 6 chu kỳ 12-tuần.

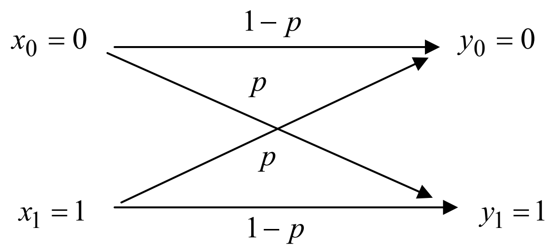
Giá trị thức tế $O_7 = \begin{bmatrix} 5 & 1 & 15 & 7 & 8 & 51 \end{bmatrix}$.

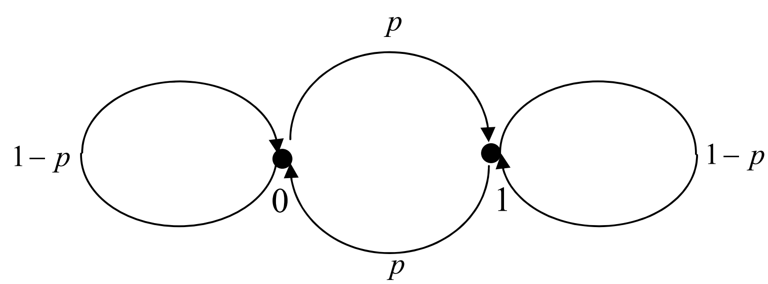
Sử dụng phép thử "khi bình phương" để so sánh, người ta thấy rằng giá trị lý thuyết e_7 phù hợp với giá trị thức tế O_7. Điều này chứng tỏ chuỗi Markov phù hợp với mô hình trên.

Ví dụ 6.3: (Mô hình thu phát tín hiệu kênh đối xứng).

Có n trạm thu phát hai tín hiệu 0, 1 dạng kênh đối xứng nhị phân với xác suất lỗi p:







Đây là 1 mô hình chuỗi Markov có không gian trạng thái $E = \{0,1\}$, thời gian $I = \{1,2,...,n\}$ có ma trận xác suất chuyển

$$P = \begin{bmatrix} 1-p & p \\ p & 1-p \end{bmatrix}$$

6.2.3. Một số mô hình chuỗi Markov quan trọng

6.2.3.1. Mô hình phục vụ đám đông

Xét mô hình phục vụ đám đông (lý thuyết sắp hàng). Khách đến sắp hàng chờ phục vụ theo nguyên tắc FIFO (first in first out) và trong mỗi chu kỳ cửa hàng chỉ phục vụ một khách. Số khách đến trong chu kỳ thứ n là biến ngẫu nhiên ξ_n. Giả sử ξ_1, ξ_2,... là các biến ngẫu nhiên độc lập cùng phân bố với biến ngẫu nhiên ξ có phân bố xác suất.

$$P\{\xi = k\} = a_k \text{ ; } k = 0,1,2,... \text{ ; } a_k > 0; \sum_k a_k = 1 \text{ .} \tag{6.16}$$

Trạng thái của hệ (cửa hàng) là số khách xếp hàng chờ phục vụ tại thời điểm đầu của mỗi chu kỳ. Nếu hiện tại hệ ở trạng thái i và sau 1 chu kỳ hệ rơi vào trạng thái j thì

$$j = \begin{cases} i - 1 + \xi & \text{nếu } i \ge 1, \\ \xi & \text{nếu } i = 0. \end{cases} \tag{6.17}$$

Ký hiệu $X(n)$ là số khách hàng tại thời điểm đầu của chu kỳ thứ n thì

$$X(n+1) = \left(X(n) - 1\right)^+ + \xi_n \text{, trong đó } X^+ = \max(0, X) \text{,}$$

Từ (6.16)-(6.17) suy ra

$$P\left\{X(n+1)=j \middle| X(n)=i\right\}=\begin{cases}P\{\xi=j+1-i\} & \text{nếu } i>0\\ P\{\xi=j\} & \text{nếu } i=0\end{cases}=\begin{cases}0 & \text{nếu } j+1<i\\ a_j & \text{nếu } i=0, j\ge 0\\ a_{j+1-i} & \text{nếu } j+1\ge i>0\end{cases}$$

Vậy $\{X(n); n=0,1,...\}$ là chuỗi Markov thuần nhất với ma trận xác suất chuyển

$$P=\begin{bmatrix}a_0 & a_1 & a_2 & a_3 & \dots\\ a_0 & a_1 & a_2 & a_3 & \dots\\ 0 & a_0 & a_1 & a_2 & \dots\\ 0 & 0 & a_0 & a_1 & \dots\\ \vdots & \vdots & \vdots & \vdots & \ddots\end{bmatrix} \tag{6.18}$$

6.2.3.2. Mô hình kiểm kê (Inventory Model)

Giả thiết phải dự trữ trong kho một loại hàng nào đó để đáp ứng nhu cầu liên tục của khách hàng. Hàng được nhập kho tại cuối mỗi chu kỳ $n=0,1,2,...$ Giả sử tổng số lượng hàng cần phải đáp ứng nhu cầu trong chu kỳ n là biến ngẫu nhiên ξ_n có phân bố độc lập với chu kỳ thời gian, nghĩa là dãy biến ngẫu nhiên $\{\xi_n\}$ độc lập có cùng phân bố với ξ.

$$P\{\xi=k\}=a_k \text{ ; } a_k>0 \text{ và } \sum_k a_k=1. \tag{6.19}$$

Mức hàng dự trữ được kiểm kê tại cuối mỗi chu kỳ. Cách nhập hàng căn cứ vào 2 chỉ số tiêu chuẩn s và S $(s<S)$ như sau: Nếu ở cuối mỗi chu kỳ lượng hàng dự trữ $\le s$ thì ngay tức khắc nhập hàng để có số hàng dự trữ bằng S; Nếu hàng hiện có $>s$ thì không cần nhập hàng.

Ký hiệu $X(n)$ là lượng hàng hiện có tại cuối chu kỳ n và trước khi nhập hàng, như vậy

$$X(n+1)=\begin{cases}X(n)-\xi_{n+1} & \text{nếu } \quad s<X(n)\le S,\\ S-\xi_{n+1} & \text{nếu } \quad X(n)\le s.\end{cases} \tag{6.20}$$

Các trạng thái của quá trình $\{X(n)\}$ là các số lượng hàng dự trữ:

$$S, S-1, ..., 1, 0, -1, -2, ...$$

trong đó giá trị âm là nhu cầu chưa được phục vụ mà sẽ được đáp ứng ngay sau khi nhập hàng

$$p_{ij}=P\left\{X(n+1)=j \middle| X(n)=i\right\}=\begin{cases}P\{\xi_{n+1}=i-j\} & \text{nếu } \quad s<i\le S,\\ P\{\xi_{n+1}=S-j\} & \text{nếu } \quad i\le s.\end{cases} \tag{6.21}$$

Ví dụ 6.4. Xét mô hình kiểm kê phụ tùng thay thế, trong đó yêu cầu có thể là 0, 1 hoặc 2 đơn vị phụ tùng cần thay thế trong một chu kỳ bất kỳ với phân bố xác suất như sau

$$P\{\xi=0\}=0,5;\; P\{\xi=1\}=0,4;\; P\{\xi=2\}=0,1$$

và giả sử $s=0;\ S=2$.

Không gian trạng thái sẽ là $E=\{-1,0,1,2\}$.

Ta có: $p_{-1,-1}=P\{X(n+1)=-1|\,X(n)=-1\}=P(\phi)=0$,

$p_{-1,0}=P\{X(n+1)=0|\,X(n)=-1\}=P(\xi=2)=0{,}1$,

$p_{-1,1}=P\{X(n+1)=1|\,X(n)=-1\}=P(\xi=1)=0{,}4$,

. . . .

Ma trận xác suất chuyển:

$$P=\begin{bmatrix}0,0 & 0,1 & 0,4 & 0,5\\ 0,0 & 0,1 & 0,4 & 0,5\\ 0,1 & 0,4 & 0,5 & 0,0\\ 0,0 & 0,1 & 0,4 & 0,5\end{bmatrix}$$

6.2.4. Phân bố dừng, phân bố giới hạn, phân bố ergodic

Định nghĩa 6.3. $\Pi^*=[\pi_1,\pi_2,...]$ *được gọi là phân bố dừng của chuỗi Markov với ma trân xác suất chuyển* P *nếu thoả mãn 2 điều kiện*:

$$\begin{cases}\sum_j \pi_j=1 & (1)\\ \Pi^*=\Pi^*P & (2)\end{cases} \qquad \textbf{(6.22)}$$

Từ 2) suy ra $\Pi^*=\Pi^*P=\Pi^*P^2=...=\Pi^*P^n$; $\forall\, n$. Do đó nếu lấy Π^* làm phân bố đầu của chuỗi Markov thì $\Pi^{*(n)}=\Pi^*$, $\forall n$. Như vậy nếu chuỗi Markov có phân bố dừng thì hệ sẽ có phân bố xác suất ổn định sau mọi bước chuyển.

Định nghĩa 6.4: *Ta nói rằng chuỗi Markov với ma trân xác suất chuyển* P *có phân bố giới hạn là* $[\pi_1,\pi_2,...]$ *nếu thoả mãn 2 điều kiện:*

1) *Với mọi* j *tồn tại giới hạn* $\lim\limits_{n\to\infty} p_{ij}{}^{(n)}=\pi_j$ *không phụ thuộc* i, **(6.23)**

2) $\sum\limits_j \pi_j=1$, $\pi_j\ge 0$, **(6.24)**

Nếu điều kiện 2) được thay bởi

2') $\sum\limits_j \pi_j=1$, $\pi_j>0$ **(6.25)**

thì chuỗi Markov được gọi là *có tính ergodic còn* $[\pi_1,\pi_2,...]$ *là phân bố ergodic*.

Ví dụ 6.5: Có 3 mạng điện thoại di động A, B, C cùng khai thác thị trường. Tỉ lệ chiếm lĩnh thị trường hiện tại tương ứng là 40%, 30% và 30%. Theo thống kê người ta thấy xác suất thay đổi mạng trong mỗi tháng như sau:

$$P=\begin{array}{c c} & \begin{array}{ccc} A & B & C \end{array} \\ \begin{array}{c} A \\ B \\ C \end{array} & \begin{bmatrix} 0,6 & 0,3 & 0,1 \\ 0,1 & 0,8 & 0,1 \\ 0,1 & 0,2 & 0,7 \end{bmatrix} \end{array}$$

Áp dụng công thức (6.11) và (6.13) ta tính được

$n=0$ $\qquad \pi=\begin{bmatrix}0,4 & 0,3 & 0,3\end{bmatrix}$,

$n=1$ $\qquad P$ $\qquad \pi P=\begin{bmatrix}0,35 & 0,43 & 0,22\end{bmatrix}$,

$$n=6 \qquad P^6=\begin{bmatrix} 0,2125 & 0,5492 & 0,2383 \\ 0,1969 & 0,5648 & 0,2383 \\ 0,1969 & 0,5181 & 0,2853 \end{bmatrix} \qquad \pi P^6=\begin{bmatrix}0,2047 & 0,5476 & 0,2477\end{bmatrix},$$

$$n=12 \qquad P^{12}=\begin{bmatrix} 0,2002 & 0,5503 & 0,2495 \\ 0,2000 & 0,5506 & 0,2495 \\ 0,2000 & 0,5484 & 0,2516 \end{bmatrix} \qquad \pi P^{12}=\begin{bmatrix}0,2001 & 0,550 & 0,2499\end{bmatrix},$$

$$n=18 \qquad P^{18}=\begin{bmatrix} 0,2000 & 0,5500 & 0,2500 \\ 0,2000 & 0,5500 & 0,2500 \\ 0,2000 & 0,5499 & 0,2501 \end{bmatrix} \qquad \pi P^{18}=\begin{bmatrix}0,2000 & 0,550 & 0,25\end{bmatrix}.$$

Ta thấy rằng khi n càng lớn xác suất trên các cột càng gần bằng nhau và phân bố giới hạn khi n tiến đến vô cùng.

Thị trường đạt trạng thái ổn định với tỉ lệ tương ứng 20%, 55% và 25%.

Định lý 6.3: Nếu tồn tại phân bố giới hạn thì đó là phân bố dừng duy nhất.

Chứng minh: Giả sử $[\pi_1, \pi_2, ...]$ là phân bố giới hạn thì với mọi j

$$\pi_j = \lim_{n\to\infty} p_{ij}^{(n+1)} = \lim_{n\to\infty}\left(\sum_k p_{ik}^{(n)} p_{kj}\right) = \sum_k \pi_k p_{kj}$$

$\Rightarrow \qquad [\pi_1, \pi_2, ...] = [\pi_1, \pi_2, ...]P$. Do đó $[\pi_1, \pi_2, ...]$ là một phân bố dừng.

Ngược lại giả sử $[\overline{\pi}_1, \overline{\pi}_2, ...]$ là một phân bố dừng bất kỳ của chuỗi Markov này thì

$$\overline{\pi}_j = \sum_k \overline{\pi}_k p_{kj} = \sum_k \overline{\pi}_k p_{kj}^{(2)} = ... = \sum_k \overline{\pi}_k p_{kj}^{(n)}$$

$$\Rightarrow \quad \overline{\pi}_j = \lim_{n\to\infty}\left(\sum_k \overline{\pi}_k p_{kj}{}^{(n)}\right) = \sum_k \overline{\pi}_k \pi_j = \pi_j .$$

Nghĩa là phân bố giới hạn là phân bố dừng duy nhất.

Định lý 6.4: Nếu chuỗi Markov có không gian trạng thái hữu hạn thì chuỗi này là ergodic khi và chỉ khi tồn tại n_0 sao cho $\min\limits_{i,j} p_{ij}{}^{(n_0)} > 0$.

Chú ý: Từ định lý 6.3 và 6.4 ta thấy rằng nếu chuỗi Markov với ma trận xác suất chuyển $P = \left[p_{ij}\right]$ tồn tại n_0 sao cho $\min\limits_{i,j} p_{ij}{}^{(n_0)} > 0$ thì chuỗi này là ergodic. Phân bố ergodic cũng là phân bố dừng duy nhất, đó là nghiệm của hệ phương trình:

$$\begin{cases} \left[x_1, x_2, \dots\right] = \left[x_1, x_2, \dots\right] P \\ x_j \ge 0, \quad \sum\limits_j x_j = 1. \end{cases} \quad \text{hay} \quad \begin{cases} P^t \begin{bmatrix} x_1 \\ x_2 \\ \vdots \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \end{bmatrix} \\ x_j \ge 0, \quad \sum\limits_j x_j = 1. \end{cases} \tag{6.26}$$

Giải hệ phương trình (6.26) cho trường hợp ví dụ 6.5 ta cũng thu được phân bố dừng tương ứng $\pi = \left[0,20 \quad 0,550 \quad 0,25\right]$.

Ví dụ 6.6: Cho chuỗi Markov có ma trận xác suất chuyển

$$P = \begin{bmatrix} 1-a & a \\ b & 1-b \end{bmatrix}, \ 0 < a, b < 1$$

là chuỗi Markov có tính ergodic với phân bố ergodic là nghiệm của hệ phương trình

$$\begin{bmatrix} 1-a & b \\ a & 1-b \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \Leftrightarrow \begin{cases} -ax_1 + bx_2 = 0 \\ x_1 + x_2 = 1 \end{cases} \Leftrightarrow \begin{cases} x_1 = \dfrac{b}{a+b} \\ x_2 = \dfrac{a}{a+b} \end{cases}$$

$$\Rightarrow \quad \lim_{n\to\infty} P^{(n)} = \begin{bmatrix} \dfrac{b}{a+b} & \dfrac{a}{a+b} \\ \dfrac{b}{a+b} & \dfrac{a}{a+b} \end{bmatrix}$$

Ví dụ 6.7: Trong một bài báo viết năm 1913 A. A. Markov đã chọn 1 dãy gồm 20.000 chữ cái trong trường ca Evghenhi Onheghin của A. X. Puskin và thấy rằng các chữ cái này chuyển đổi liên tiếp theo hai trạng thái nguyên âm (Na) và phụ âm (Pa) với ma trận xác suất chuyển là

$$P = \begin{array}{cc} \begin{bmatrix} 0,128 & 0,872 \\ 0,663 & 0,337 \end{bmatrix} & \begin{array}{c} Na \\ Pa \end{array} \\ \begin{array}{cc} Na & \quad Pa \end{array} & \end{array}$$

Phân bố giới hạn (cũng là phân bố dừng) của chuỗi Markov này là (0,423; 0,568). Vậy có khoảng 42,3% nguyên âm và 56,8% phụ âm trong tác phẩm trên.

6.3. PHÂN LOẠI TRẠNG THÁI CHUỖI MARKOV(*)

Định lý 6.4 cho ta dấu hiệu nhận biết một chuỗi Markov hữu hạn trạng thái tồn tại phân bố ergodic. Trong trường hợp tổng quát, bằng cách phân tích trạng thái của chuỗi Markov ta sẽ tìm điều kiện để tồn tại phân bố giới hạn, phân bố dừng hoặc phân bố ergodic thỏa mãn (6.22)-(6.25).

6.3.1. Các trạng thái liên thông và sự phân lớp

Định nghĩa 6.5: *Ta nói rằng trạng thái j đạt được từ trạng thái i nếu tồn tại $n \ge 0$ sao cho $p_{ij}^{(n)} > 0$ (xác suất để sau n bước chuyển từ trạng thái i sang trạng thái j lớn hơn 0). Ký hiệu $i \to j$.*

Quy ước $p_{ii}^{(0)} = 1$ và $p_{ij}^{(0)} = 0$ khi $i \neq j$.

Hai trạng thái i và j được gọi là liên thông với nhau nếu $i \to j$ và $j \to i$, lúc đó ta ký hiệu $i \leftrightarrow j$.

Có thể chứng minh được rằng $\leftrightarrow$ là một quan hệ tương đương (có tính phản xạ, đối xứng và bắc cầu) trên tập các trạng thái. Do đó ta có thể phân hoạch không gian trạng thái thành các lớp tương đương. Các lớp tương đương này rời nhau, hai trạng thái bất kỳ cùng một lớp thì liên thông với nhau, còn hai trạng thái thuộc hai lớp khác nhau không thể liên thông với nhau.

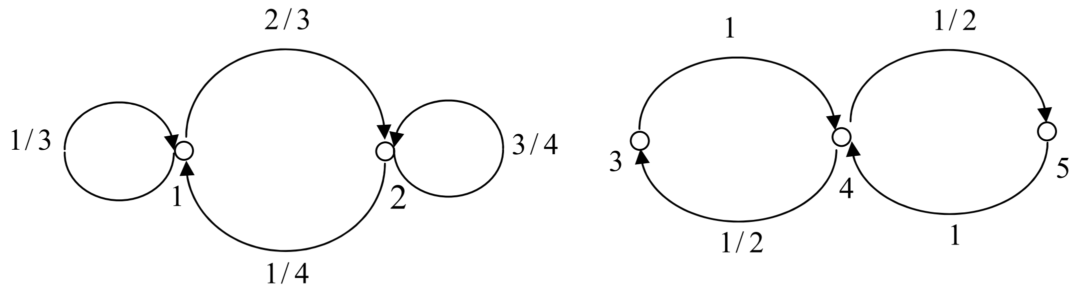
Định nghĩa 6.6: *Chuỗi Markov được gọi là tối giản nếu hai trạng thái bất kỳ của không gian trạng thái liên thông với nhau.*

Như vậy chuỗi Markov tối giản chỉ có một lớp tương đương.

Giả sử không gian trạng thái được tách thành các lớp tương đương $E = E_1 \cup E_2 \cup \cdots$ Trong nhiều trường hợp có thể xem mỗi E_k $(k = 1, 2, ...)$ là không gian trạng thái của chuỗi Markov tối giản. Vì thế $E_1, E_2, ...$ được gọi là các lớp tối giản của chuỗi.

Ví dụ 6.8: Cho chuỗi Markov với ma trận xác suất chuyển

$$P = \begin{bmatrix} 1/3 & 2/3 & 0 & 0 & 0 \\ 1/4 & 3/4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} P_1 & 0 \\ 0 & P_2 \end{bmatrix}$$



Không gian trạng thái $E = \{1,2,3,4,5\}$ phân thành hai lớp $E_1 = \{1,2\}$, $E_2 = \{3,4,5\}$ và có thể xem E_1, E_2 là không gian trạng thái của chuỗi Markov với ma trận xác suất chuyển tương ứng là P_1 và P_2.

6.3.2. Chu kỳ của trạng thái

Định nghĩa 6.7: *Ước chung lớn nhất của tất cả các số tự nhiên $n \ge 1$ thỏa mãn điều kiện $p_{ii}^{(n)} > 0$ được gọi là chu kỳ của trạng thái i ký hiệu $d(i)$. Nếu $p_{ii}^{(n)} = 0$ đối với mọi $n \ge 1$ thì đặt $d(i) = 0$.*

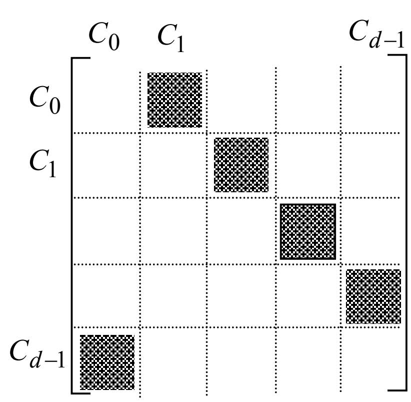
Trong ví dụ 6.8 có $d(1) = d(2) = 1$; $d(3) = d(4) = d(5) = 2$.

Định lý 6.7: Nếu $i \leftrightarrow j$ thì $d(i) = d(j)$. Do đó các trạng thái thuộc cùng một lớp có cùng chu kỳ.

6.3.3. Dạng ma trận xác suất chuyển của chuỗi Markov tối giản

Đối với chuỗi Markov tối giản mọi trạng thái đều có cùng chu kỳ, ta gọi d là chu kỳ chung của mọi trạng thái của chuỗi.

- Nếu $d = 1$ thì ma trận xác suất chuyển P chỉ có 1 khối.
- Nếu $d > 1$ thì tập trạng thái E tách thành d lớp con: $C_0, C_1, ..., C_{d-1}$. Trong trường hợp này sau một bước hệ xuất phát từ C_0 sẽ chuyển sang C_1; xuất phát từ C_1 sẽ chuyển sang C_2; v.v… C_{d-1} sẽ chuyển sang C_0. Ma trận xác suất chuyển P có dạng khối như sau:



(6.27)

Vì vậy mỗi lớp con C_k có thể lấy làm không gian trạng thái của chuỗi Markov mới ma trận xác suất chuyển là $\left[p_{ij}^{(d)}\right]_{i,j\in C_k}$, chuỗi này tối giản có chu kỳ bằng 1. Như vậy ta có thể quy việc nghiên cứu chuỗi Markov tổng quát (đặc biệt là vấn đề tìm $\lim\limits_{n\to\infty} p_{ij}^{(n)}$) về chuỗi Markov tối giản có chu kỳ 1.

6.3.4. Trạng thái hồi quy và trạng thái không hồi quy

Với mỗi trạng thái i ta đặt:

$$f_{ij}^{(n)} = P\{X_n = j, X_1 \neq j, \ldots X_{n-1} \neq j \mid X_0 = i\};\ j \in E\,. \tag{6.28}$$

Như vậy $f_{ij}^{(n)}$ là xác suất để hệ xuất phát từ i lần đầu tiên chuyển sang j tại bước thứ n. Đặc biệt $f_{ii}^{(n)}$ là xác suất để hệ xuất phát từ i lần đầu tiên quay về i tại bước thứ n.

Từ tính Markov và công thức xác suất đầy đủ ta có:

$$p_{ij}^{(n)} = \sum_{k=0}^{n} f_{ij}^{(k)} p_{jj}^{(n-k)}, n \geq 1;\ f_{ij}^{(1)} = p_{ij}\,. \tag{6.29}$$

trong đó ta quy ước $f_{ij}^{(0)} = 0$ với mọi $i,\ j$.

Đặt

$$f_{ij} = \sum_{n=0}^{\infty} f_{ij}^{(n)},\ f_{ii} = \sum_{n=0}^{\infty} f_{ii}^{(n)}\,. \tag{6.30}$$

Như vậy f_{ii} là xác suất để hệ xuất phát từ i trở về i tại thời điểm hữu hạn nào đó.

Do tính chất Markov ta suy ra: " biến cố hệ xuất phát từ i trở lại i ít nhất 2 lần" có xác suất bằng $(f_{ii})^2$, …

$$P = \{\text{hệ xuất phát từ } i \text{ trở lại } i \text{ ít nhất } k \text{ lần}\} = (f_{ii})^k,\ (k = 1, 2, \ldots)\,. \tag{6.31}$$

Định nghĩa 6.7:

- Nếu $f_{ii} = 1$ thì i được gọi hồi *là trạng thái quy* (trạng thái trở về).
- Nếu $f_{ii} < 1$ thì i được gọi là *trạng thái không hồi quy* (trạng thái di chuyển).

Như vậy trạng thái i hồi quy khi và chỉ khi hệ xuất phát từ i, với xác suất 1 hệ lại trở về i tại thời điểm hữu hạn nào đó.

Trường hợp trạng thái i hồi quy $f_{ii} = 1$, theo công thức (6.30) thì $\left(f_{ii}^{(n)}\right)$ lập thành phân bố xác suất. Kết hợp công thức (6.31) ta có thể tính thời gian trung bình hệ trở lại i.

$$\mu_i = \sum_{n=0}^{\infty} n f_{ii}^{(n)} \qquad \textbf{(6.32)}$$

Định nghĩa 6.9: *Giả sử i là trạng thái hồi quy. Ta nói:*

- *i là trạng thái hồi quy dương nếu $\mu_i < \infty$.*
- i là trạng thái hồi quy không nếu $\mu_i = \infty$.

Như vậy không gian trạng thái E được phân loại như sau:

1. Các trạng thái hồi quy:
   - Trạng thái dương,
   - Trạng thái không.
2. Các trạng thái không hồi quy.

6.3.5. Tiêu chuẩn hồi quy và không hồi quy

Định lý 6.8: 1) Trạng thái i là hồi quy khi và chỉ khi $\sum_{n=1}^{\infty} p_{ii}^{(n)} = \infty$.

2) Trạng thái i là không hồi quy khi và chỉ khi $\sum_{n=1}^{\infty} p_{ii}^{(n)} < \infty$.

3) Nếu $i \to j$ và i hồi quy thì $j \to i$ và j cũng hồi quy.

4) Nếu $i \leftrightarrow j$ và j hồi quy thì $f_{ij} = 1$.

6.3.6. Định lý giới hạn cơ bản của chuỗi Markov

Định lý 6.9: Giả sử j là trạng thái hồi quy, chu kỳ $d(j) = 1$. Khi đó:

1. Nếu i và j liên thông thì

$$\lim_{n\to\infty} p_{ij}^{(n)} = \begin{cases} \dfrac{1}{\mu_j} & \text{đối với } j \text{ là trạng thái dương} \\ 0 & \text{đối với } j \text{ là trạng thái không} \end{cases} \qquad \textbf{(6.33)}$$

2. Nếu i và j không liên thông thì

$$\lim_{n\to\infty} p_{ij}^{(n)} = \begin{cases} \dfrac{f_{ij}}{\mu_j} & \text{đối với } j \text{ là trạng thái dương} \\ 0 & \text{đối với } j \text{ là trạng thái không} \end{cases} \qquad \textbf{(6.34)}$$

Định lý 6.10: Giả sử j là trạng thái hồi quy, chu kỳ $d(j) = d > 1$. Khi đó:

1. Nếu i và j liên thông; i thuộc vào lớp con C_r còn j thuộc vào lớp C_{r+a} thì

$$\lim_{n\to\infty} p_{ij}^{(nd+a)} = \frac{d}{\mu_j},\ (a = 0,1,...,d-1) \quad \textbf{(6.35)}$$

2. Nếu i và j không liên thông thì

$$\lim_{n\to\infty} p_{ij}^{(nd+a)} = \left[\sum_{r=0}^{\infty} f_{ij}^{(rd+a)}\right]\frac{d}{\mu_j},\ (a = 0,1,...,d-1) \quad \textbf{(6.36)}$$

6.3.6. Sự tồn tại phân bố dừng và phân bố giới hạn

Định lý 6.11: Giả sử $\{X_n\}$ là chuỗi Markov với không gian trạng thái E. Khi đó phân bố dừng tồn tại và duy nhất khi và chỉ khi trong số các lớp tương đương của không gian trạng thái E có đúng một lớp hồi quy dương.

Giả sử lớp hồi quy dương là C, khi đó phân bố dừng có dạng

$$\pi_j = \begin{cases} \dfrac{1}{\mu_j} & \text{nếu} \quad j \in C \\ 0 & \text{nếu} \quad j \notin C. \end{cases} \quad \textbf{(6.37)}$$

Định lý 6.12: Điều kiện cần và đủ để tồn tại phân bố giới hạn là không gian trạng thái E có đúng một lớp hồi quy dương C, chu kỳ $d(\mathrm{C}) = 1$ sao cho $f_{ij} = 1;\ \forall j \in \mathrm{C},\ \forall i \in E$.

Khi đó phân bố giới hạn cũng là phân bố dừng duy nhất thỏa mãn (6.37).

Định lý 6.13: Giả sử $\{X(n)\}$ là một chuỗi Markov có không gian trạng thái hữu hạn. Khi đó các điều sau là tương đương:

(i) $\{X(n)\}$ tối giản có chu kỳ 1.

(ii) $\{X(n)\}$ tối giản có chu kỳ 1 và tất cả các trạng thái là hồi quy dương.

(iii) $\{X(n)\}$ có tính ergodic, nghĩa là tồn tại phân bố ergodic (thỏa mãn 6.23, 6.25).

(iv) Tồn tại n_0 sao cho $\min\limits_{i,j} p_{ij}^{(n)} > 0$ với mọi $n \ge n_0$ (xem định lý 6.4).

6.4. DI ĐỘNG NGẪU NHIÊN TRÊN ĐƯỜNG THẲNG (*)

6.4.1. Di động ngẫu nhiên trên đường thẳng không có trạng thái hấp thụ

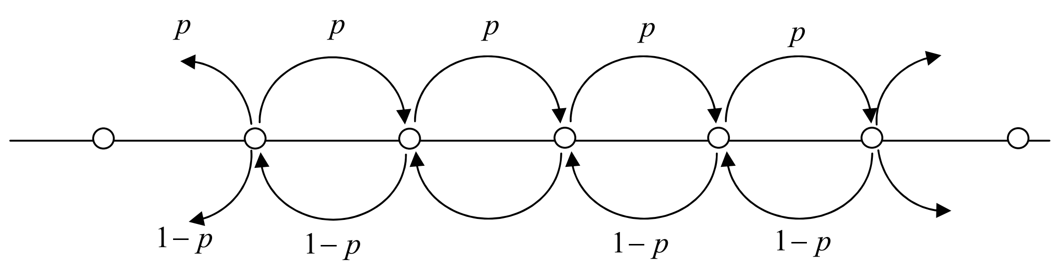
Giả sử $\{\varepsilon_n\}_{n=1}^{\infty}$ là dãy các biến ngẫu nhiên độc lập có cùng phân bố thỏa mãn:

$$P\{\xi_n = 1\} = p,\ P\{\xi_n = -1\} = 1 - p = q;\ 0 \le p \le 1. \quad \textbf{(6.38)}$$

Đặt $X_n = \varepsilon_1 + \varepsilon_2 + \cdots \varepsilon_n$, $(n = 1,2,...)$. Khi đó $\{X_n\}_{n=1}^{\infty}$ lập thành chuỗi Markov với ma trận xác suất chuyển là

$$P = \left[p_{ij}\right] \text{ trong đó } p_{ij} = \begin{cases} p & \text{với } j = i+1 \\ 1-p & \text{với } j = i-1 \\ 0 & \text{với } j \neq i \pm 1 \end{cases} \quad \textbf{(6.39)}$$

Không gian trạng thái của chuỗi này là $E = \{0, \pm 1, \pm 2, ...\}$



Chuỗi này dùng để mô tả *di động ngẫu nhiên trên đường thẳng* của hạt vật chất nào đó: Sau mỗi chu kỳ hạt dịch chuyển sang phải với xác suất p hoặc dịch sang trái với xác suất $1-p$.

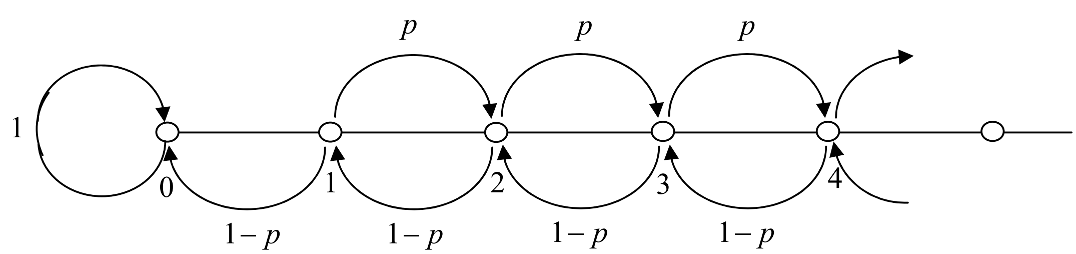
Nếu $p=0$ thì hạt luôn luôn dịch sang trái, còn nếu $p=1$ thì hạt luôn luôn dịch sang phải. Trong những trường hợp như thế vị trí hạt không có tính ngẫu nhiên. Vậy ta luôn giả sử $0 < p < 1$.

Khi $0 < p < 1$ thì di động ngẫu nhiên trên đường thẳng là chuỗi Markov tối giản, có chu kỳ $d = 2$ chuỗi không tồn tại phân bố dừng và không có tính ergodic.

6.4.2. Di động ngẫu nhiên trên đường thẳng có một trạng thái hấp thụ

Đó là di động của hạt vật chất với không gian trạng thái $E = \{0, 1, 2, ...\}$ và ma trận xác suất chuyển là $P = \left[p_{ij}\right]$, trong đó

$$p_{00} = 1,\ p_{ij} = \begin{cases} p & \text{với } j = i+1, i \neq 0, \\ 1-p & \text{với } j = i-1, i \neq 0, \\ 0 & \text{với } j \neq i \pm 1, i \neq 0, \end{cases} ;\ 0 < p < 1 \quad \textbf{(6.40)}$$



Lúc này $\{0\}$ lập thành lớp hồi quy dương duy nhất với chu kỳ $d=1$: $E=\{0\}\cup\{1,2,...\}$. Tất cả các trạng thái $1,2,...$ là không hồi quy (vì nếu 1 hồi quy chẳng hạn thì do $1\to 0$ nên $0\to 1$, điều này không thể xảy ra). Vì vậy theo định lý 6.11 và công thức (6.37) tồn tại phân bố dừng duy nhất, đó là

$$\pi_j=\begin{cases}1 & \text{với } j=0,\\ 0 & \text{với } j\neq 0.\end{cases} \qquad \textbf{(6.41)}$$

Hơn nữa có thể chứng minh được rằng đối với mỗi $i\ge 1$ thì

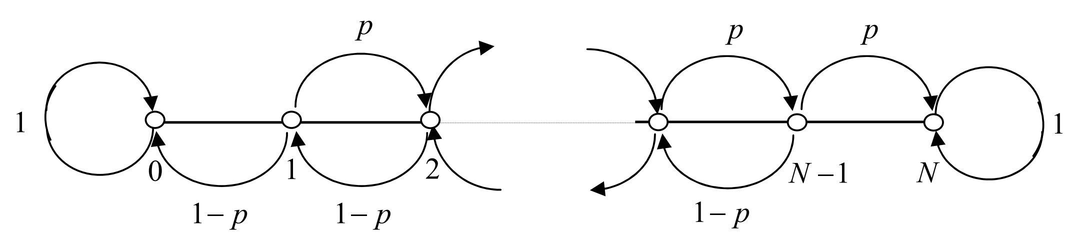
$$\lim_{n\to\infty} p_{i0}^{(n)}=\begin{cases}(q/p)^i & \text{nếu } p>q,\\ 1 & \text{nếu } p\le q;\ q=1-p.\end{cases} \qquad \textbf{(6.42)}$$

Vì vậy:

- Khi $p>q$ thì $\lim\limits_{n\to\infty} p_{i0}^{(n)}=(q/p)^i$ phụ thuộc vào i, do đó không tồn tại phân bố giới hạn.
- Khi $p\le q$ thì tồn tại phân bố giới hạn, đó là $\pi_j=\begin{cases}1 & \text{với } j=0,\\ 0 & \text{với } j\neq 0.\end{cases}$

6.4.3. Di động ngẫu nhiên trên đường thẳng có hai trạng thái hấp thụ

Đó là mô hình di động như hình vẽ



Trong trường hợp này có hai lớp hồi quy dương là $\{0\}$ và $\{N\}$. Các trạng thái còn lại không hồi quy:

$$E=\{0\}\cup\{N\}\cup\{1,2,...,N-1\}.$$

Di động sẽ ngừng lại khi hạt rơi vào trạng thái 0 hoặc trạng thái N.

Do đó tồn tại vô số phân bố dừng $\Pi=[\pi_0,\pi_1,...,\pi_N]$, trong đó

$$\pi_0=a,\ \pi_N=1-a,\ \pi_1=\pi_2=...=\pi_{N-1}=0,\ \text{với } 0\le a\le 1. \qquad \textbf{(6.40)}$$

Không tồn tại phân bố giới hạn. Hơn nữa có thể chứng minh được rằng:

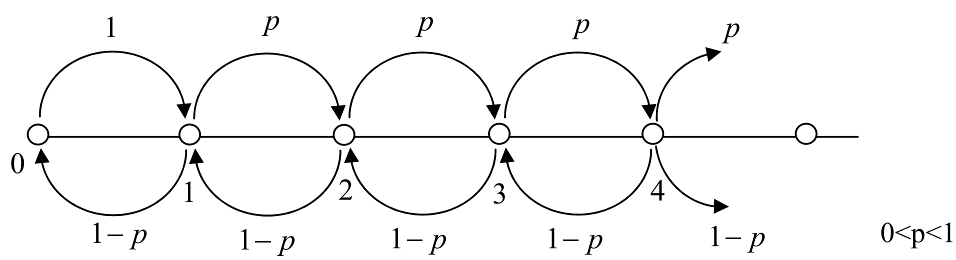
$$\lim_{n\to\infty} p_{i0}^{(n)} = \begin{cases} \dfrac{(q/p)^i - (q/p)^N}{1-(q/p)^N} & \text{nếu } p \neq q, \\ 1 - \dfrac{i}{N} & \text{nếu } p = q = 1/2. \end{cases}$$

$$\lim_{n\to\infty} p_{iN}^{(n)} = 1 - \lim_{n\to\infty} p_{i0}^{(n)},$$

$$\lim_{n\to\infty} p_{ij}^{(n)} = 0 \ (\forall\, j = 1,2,...,N-1).$$

6.4.4. Di động ngẫu nhiên trên đường thẳng có một trạng thái phản hồi

Đó là chuỗi Markov có dạng như hình vẽ



Không gian trạng thái $E = \{0,1,2,...\}$. Chuỗi tối giản, có chu kỳ $d = 2$.

Khi $p > q$, hạt có xu hướng đi sang phải, chuỗi không tồn tại phân bố giới hạn và phân bố dừng.

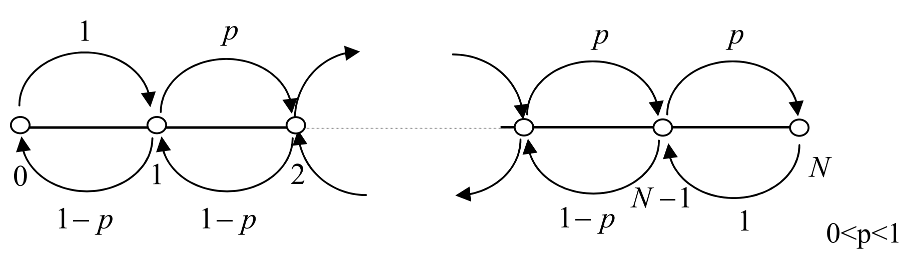
Khi $p = q = 1/2$, tất cả các trạng thái là hồi quy không. Không tồn tại phân bố dừng.

Khi $p < q$, tất cả các trạng thái là hồi quy dương. Tồn tại phân bố dừng duy nhất.

$$\pi_0 = \frac{q-p}{2q}, \pi_1 = \frac{q-p}{2q^2}, ..., \pi_j = \frac{q-p}{2q^2}\left(\frac{p}{q}\right)^{j-1}; j \geq 2. \qquad \textbf{(6.41)}$$

6.4.5. Di động ngẫu nhiên trên đường thẳng có hai trạng thái phản hồi

Đó là chuỗi Markov có dạng như hình vẽ



Chuỗi tối giản, không gian trạng thái hữu hạn $E=\{0,1,2,...,N\}$. Tất cả các trạng thái của chuỗi là hồi quy dương, có chu kỳ $d=2$.

Chuỗi tồn tại phân bố dừng nhưng không tồn tại phân bố giới hạn.

Phân bố dừng là nghiệm duy nhất của hệ phương trình (6.26)

$$\begin{cases} x_j=\sum_{i=0}^{N} x_i p_{ij} \;\; (j=0,1,...,N), \\ x_i \ge 0, \sum_{i=0}^{N} x_i = 1. \end{cases}$$

Từ đó suy ra phân bố dừng

$$\pi_0=\pi_1 q\,,\; \pi_N=\pi_{N-1}p\,;\; \pi_i=\frac{\left(p/q\right)^{i-1}}{1+\sum_{j=1}^{N-1}\left(\frac{p}{q}\right)^{j-1}};(1\le i\le N-1)\,. \tag{6.42}$$

CÂU HỎI ÔN TẬP VÀ BÀI TẬP

6.1 Quá trình ngẫu nhiên $X(t,\omega)$ là một hàm số của hai biến (t,ω).

Đúng ☐ Sai ☐.

6.2 Mọi quá trình có gia số độc lập là quá trình Markov.

Đúng ☐ Sai ☐.

6.3 Chuỗi Markov là quá trình Markov $\{X(t);t\in I\}$ có không gian trạng thái E đếm được.

Đúng ☐ Sai ☐.

6.4 Ma trận xác suất chuyển sau n bước của một chuỗi Markov bằng tích n lần ma trận xác suất chuyển một bước của chuỗi Markov này.

Đúng ☐ Sai ☐.

6.5 Nếu tồn tại phân bố giới hạn thì nó là phân bố dừng duy nhất.

Đúng ☐ Sai ☐.

6.6 Mọi chuỗi Markov có hữu hạn trạng thái luôn tồn tại phân bố dừng duy nhất đó là phân bố ergodic.

Đúng ☐ Sai ☐.

6.7 Cho chuỗi Markov $\{X_n\}_{n=1}^{\infty}$ với không gian trạng thái $E=\{0,1,2\}$ và ma trận xác suất chuyển

$$P=\begin{bmatrix}0,1 & 0,2 & 0,7\\ 0,9 & 0,1 & 0,0\\ 0,1 & 0,8 & 0,1\end{bmatrix}$$

Biết phân bố ban đầu: $p_0=P\{X_0=0\}=3$; $p_1=P\{X_0=1\}=4$; $p_2=P\{X_0=2\}=3$.

Tính $P\{X_0=0,X_1=2,X_2=1\}$.

6.8 Cho chuỗi Markov $\{X_n\}_{n=1}^{\infty}$ với không gian trạng thái $E=\{0,1,2\}$ và ma trận xác suất chuyển

$$P=\begin{bmatrix}0,1 & 0,2 & 0,7\\ 0,2 & 0,2 & 0,6\\ 0,6 & 0,1 & 0,3\end{bmatrix}$$

a) Tính ma trận xác suất chuyển 2 bước.

b) Tính $P\{X_3=1|X_1=0\}$; $P\{X_3=1|X_0=0\}$.

c) Tìm phân bố dừng.

6.9 Xét bài toán truyền một bức điện gồm gồm các tín hiệu 0, 1 thông qua kênh có nhiều trạm và mỗi trạm nhận sai tín hiệu với xác suất không đổi bằng $\alpha\in(0,1)$. Giả sử X_0 là tín hiệu truyền đi và X_n là tín hiệu nhận được tại trạm n. Cho biết $\{X_n;n=0,1,2,...\}$ lập thành chuỗi Markov với ma trận xác suất chuyển

$$P=\begin{bmatrix}\alpha & 1-\alpha\\ 1-\alpha & \alpha\end{bmatrix}.$$

a) Tính $P\{X_0=0,X_1=0,X_2=0\}$.

b) Tính $P\{X_0=0,X_1=0,X_2=0\}+P\{X_0=0,X_1=1,X_2=0\}$.

c) Tính $P\{X_5=0|X_0=0\}$.

6.10 Xét mô hình kiểm kê phụ tùng thay thế với $s=0$ và $S=3$ là các mức căn cứ để nhập hàng cùng với ξ_n là lượng hàng khách yêu cầu trong chu kỳ n. Biết rằng

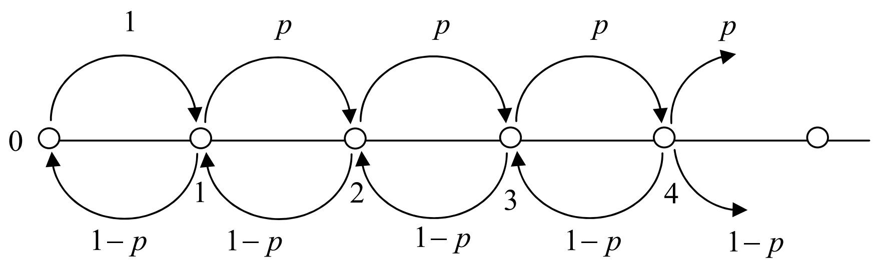
$$P\{\xi_n=0\}=0,4;\ P\{\xi_n=1\}=0,3;\ P\{\xi_n=2\}=0,3.$$

Xác định xác suất chuyển của chuỗi Markov $\{X_n\}$, trong đó X_n là số phụ tùng còn lại tại cuối chu kỳ n.

6.11 Tìm các lớp liên thông trạng thái của chuỗi Markov có không gian trạng thái $E=\{0,1,2,3,4\}$ và ma trận xác suất chuyển

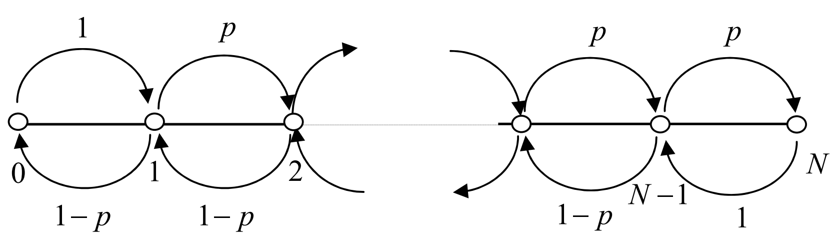
$$P=\begin{bmatrix} 1/2 & 1/2 & 0 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 1/2 & 0 \\ 0 & 0 & 1/2 & 1/2 & 0 \\ 1/4 & 1/4 & 0 & 1/4 & 1/4 \end{bmatrix}$$

6.12 Xét di động ngẫu nhiên trên đường thẳng có một trạng thái phản hồi có dạng như hình vẽ



Với $p=0,4$. Tìm phân bố dừng.

6.13 Xét di động ngẫu nhiên trên đường thẳng có hai trạng thái phản hồi có dạng như hình vẽ



Với $N=10$; $p=0,6$. Tìm phân bố dừng.

.

HƯỚNG DẪN GIẢI 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 $P(A) = \dfrac{1}{216}$.

Lý luận tương tự trên ta có $P_b = \dfrac{6}{216} = \dfrac{1}{36}$; $P_c = \dfrac{A_6^5}{216} = \dfrac{5}{9}$.

1.13 $P = \dfrac{1}{720}$.

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 $P = C_{30}^5 \left(\dfrac{1}{50}\right)^5 \left(\dfrac{49}{50}\right)^{25} = 0{,}00027$.

1.18 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.

$$P(B)=P\left(\overline{A_1}\right)P\left(\overline{A_2}\right)P\left(\overline{A_3}\right)=(1-0,8)(1-0,9)(1-0,99)=0,0002\,.$$

1.19 $P=0,11$.

1.20 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 $P(A)=P(A_1)P(A_2|A_1)P(A_3|A_1A_2)=1\cdot\dfrac{5}{21}\cdot\dfrac{1}{84}=\dfrac{5}{1764}$.

1.21 Gọi A là biến cố sản phẩm kiểm tra là phế phẩm.

Gọi B_i là biến cố sản phẩm lấy ra kiểm tra thuộc phân xưởng thứ i, i=1,2 3.

$P(B_1)=0,36;\ P(B_2)=0,34;\ P(B_3)=0,30$. Hệ $\{B_1,B_2,B_3\}$ đầy đủ

$P(A|B_1)=0,12;\ P(A|B_2)=0,10;\ P(A|B_3)=0,08$.

a. $P(A)=P(B_1)P(A|B_1)+P(B_2)P(A|B_2)+P(B_3)P(A|B_3)=0,1012$

b. $P(B_1|A)=\dfrac{P(B_1)P(A|B_1)}{P(A)}=\dfrac{0,36\times0,12}{0,1012}=0,427$

$$P(B_2|A)=\frac{P(B_2)P(A|B_2)}{P(A)}=\frac{0,34\times0,10}{0,1012}=0,336$$

$$P(B_3|A)=\frac{P(B_3)P(A|B_3)}{P(A)}=\frac{0,30\times0,08}{0,1012}=0,237$$

1.22 Gọi B_i là biến cố xạ thủ được xét thuộc nhóm thứ i, i=1,2,3,4.

Gọi A là biến cố xạ thủ bắn trượt.

Theo đề bài ta có: $P(B_1)=\dfrac{5}{18},\ P(B_2)=\dfrac{7}{18},\ P(B_3)=\dfrac{4}{18},\ P(B_4)=\dfrac{2}{18}$

$$P(A|B_1)=0,2,\ P(A|B_2)=0,3,\ P(A|B_3)=0,4,\ P(A|B_4)=0,5\,.$$

$$\begin{aligned}P(A)&=P(B_1)P(A|B_1)+P(B_2)P(A|B_2)+P(B_3)P(A|B_3)+P(B_4)P(A|B_4)\\&=\frac{5}{18}\times0,2+\frac{7}{18}\times0,3+\frac{4}{18}\times0,4+\frac{2}{18}\times0,5=\frac{57}{180}\end{aligned}$$

Áp dụng công thức Bayer, ta thu được

$$P(B_1|A)=\frac{P(B_1)P(A|B_1)}{P(A)}=\frac{\frac{5}{18}\times0,2}{\frac{57}{180}}=\frac{10}{57},$$

tương tự $P\left(B_2|A\right)=\dfrac{21}{57},\ P\left(B_3|A\right)=\dfrac{16}{57},\ P\left(B_4|A\right)=\dfrac{10}{57}$.

Vậy xạ thủ có khả năng ở nhóm thứ hai nhất.

1.23 Gọi B_1 là biến cố viên đạn thứ nhất trúng mục tiêu, $P\left(B_1\right)=0,7$.

Gọi B_2 là biến cố viên đạn thứ hai trúng mục tiêu, $P\left(B_2\right)=0,4$.

Hai biến cố này độc lập

Xác suất biến cố chỉ có viên đạn thứ nhất trúng mục tiêu

$$P(A)=P\left(B_1\overline{B_2}\cup B_2\overline{B_1}\right)=P\left(B_1\overline{B_2}\right)+P\left(B_2\overline{B_1}\right)$$
$$=0,7\times 0,6+0,4\times 0,3=0,54$$

b. $$P\left(B_1|A\right)=\frac{P\left(B_1A\right)}{P(A)}=\frac{P\left(B_1\left[B_1\overline{B_2}\cup B_2\overline{B_1}\right]\right)}{P(A)}=\frac{P\left(B_1\overline{B_2}\right)}{P(A)}=\frac{0,7\times 0,6}{0,54}=0,778$$

1.24 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.

$P\left(B_T\right)=0,85;\ P\left(B_H\right)=0,15$

Hệ $\left\{B_T, B_H\right\}$ đầy đủ

$P\left(A|B_T\right)=0,9;\ P\left(\overline{A}|B_H\right)=0,95\ \Rightarrow\ P\left(\overline{A}|B_T\right)=0,1;\ P\left(A|B_H\right)=0,05$.

a) $P(A)=P\left(B_T\right)P\left(A|B_T\right)+P\left(B_H\right)P\left(A|B_H\right)=0,85\times 0,9+0,15\times 0,05=0,7725$

b) $P\left(B_H|A\right)=\dfrac{P\left(B_H\right)P\left(A|B_H\right)}{P(A)}=\dfrac{0,15\times 0,05}{0,7725}=0,0097$

c) $P\left(AB_T\cup\overline{A}B_H\right)=P\left(AB_T\right)+P\left(\overline{A}B_H\right)=0,85\times 0,9+0,15\times 0,95=0,9075$.

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	Đúng	Sai	Sai	Đúng	Đúng	Sai
2.11	**2.12**	**2.13**	**2.14**	**2.15**					
Sai	Sai	Đúng	Đúng	Sai					

2.16 $E(X)=0,3$; $D(X)=15,21$.

2.17 $x_3 = 29,1$; $p_3 = 0,2$.

2.18 $\mathrm{E}(X_1)=3,1$; $\mathrm{E}(X_2)=3,4$; $\mathrm{D}(X_1)=1,09$; $\mathrm{D}(X_2)=1,44$.

$\mathrm{E}(X_1+X_2)=6,5$; $\mathrm{D}(X_1+X_2)=2,53$.

2.19 $\mathrm{E}(\overline{X})=0,8$; $\mathrm{D}(\overline{X})=0,12$.

2.20 a) $\mathrm{D}(Z)=61$. b) $\mathrm{D}(Z)=41$.

2.21 $p_1=0,4$; $p_2=0,1$; $p_3=0,5$.

2.22 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	$\frac{1}{243}$	$\frac{10}{243}$	$\frac{40}{243}$	$\frac{80}{243}$	$\frac{80}{243}$	$\frac{32}{243}$

2.23 $\mathrm{E}(X)=0$.

2.24 a) Vì $\int_0^4 x^2(4-x)dx=\frac{64}{3} \Rightarrow k=\frac{3}{64}$.

b) $P\{X<1\}=\int_0^1 \frac{3}{64}x^2(4-x)dx=\frac{13}{256}$.

c) $\mathrm{E}X=\int_0^4 \frac{3}{64}x^3(4-x)dx=\frac{3}{64}\left(x^4-\frac{x^5}{5}\right)\Bigg|_{x=0}^{x=4}=3\left(4-\frac{16}{5}\right)=\frac{12}{5}$

$$\mathrm{E}X^2=\int_0^4 \frac{3}{64}x^4(4-x)dx=\frac{3}{64}\left(\frac{4x^5}{5}-\frac{x^6}{6}\right)\Bigg|_{x=0}^{x=4}=3\times 64\left(\frac{1}{5}-\frac{1}{6}\right)=\frac{32}{5}$$

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

2.25 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) $P\{Y=0\}=0,37$

$P\{Y=1\}=P\{X=1\}+P\{X=-1\}=0,5$

$P\{Y=2\}=P\{X=2\}+P\{X=-2\}=0,13$

$\mathrm{E}Y=0\times 0,37+1\times 0,5+2\times 0,13=0,76$.

2.26 Kí hiệu A_i là biến cố : ”ô tô thứ i bị hỏng”, i = 1, 2.

Dễ thấy $P\left(A_1\right)=0,1$; $P\left(A_2\right)=0,2$

Gọi X là số ôtô bị hỏng trong thời gian làm việc

Từ đó $P\{X=0\}=P\left(\overline{A_1}\right)P\left(\overline{A_2}\right)=0,9\times 0,8=0,72$

$P\{X=1\}=P\left(A_1\right)P\left(\overline{A_2}\right)+P\left(\overline{A_1}\right)P\left(A_2\right)=0,1\times 0,8+0,9\times 0,2=0,26$

$P\{X=2\}=P\left(A_1\right)P\left(A_2\right)=0,02$

X	0	1	2
P	0,72	0,26	0,02

$\mathrm{E}X=0\times 0,72+1\times 0,26+2\times 0,02=0,3$.

$\mathrm{E}X^2=0^2\times 0,72+1^2\times 0,26+2^2\times 0,02=0,34$.

$\mathrm{D}X=\mathrm{E}X^2-\left(\mathrm{E}X\right)^2=0,34-\left(0,3\right)^2=0,25$.

2.27 a) Điều kiện $\begin{cases} p_i>0 \\ \sum\limits_i p_i=1 \end{cases} \Rightarrow \begin{cases} k>0 \\ 10k^2+9k=1 \end{cases} \Rightarrow \begin{cases} k>0 \\ \left[\begin{array}{l} k=-1 \\ k=1/10 \end{array}\right. \end{cases} \Rightarrow k=1/10$.

b) $P\{X\ge 5\}=\frac{1}{10}+\frac{1}{10}=\frac{2}{10}$; $P\{X<3\}=\frac{3}{10}$.

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) $\mathrm{E}X^2=\frac{1}{10}+\frac{8}{10}+\frac{18}{10}+\frac{48}{10}+\frac{25}{100}+\frac{72}{100}+49\left(\frac{7}{100}+\frac{1}{10}\right)=16,8$.

$\mathrm{D}X=\mathrm{E}X^2-\left(\mathrm{E}X\right)^2=16,8-\left(3,66\right)^2=3,404$.

2.28 a) Gọi X là "số phế phẩm gặp phải":

X	0	1
P	0,6	0,4

$\mathrm{E}X=\frac{2}{5}$; $\mathrm{D}X=\frac{6}{25}$.

b) Gọi Ị là "số chính phẩm gặp phải" $\Rightarrow Y=2-X$:

Y	1	2
P	0,4	0,6

$\mathrm{E}Y=\mathrm{E}\left(2-X\right)=2-\frac{2}{5}=\frac{8}{5}$; $\mathrm{D}Y=\frac{6}{25}$.

2.29 Gọi X là "số nữ có trong nhóm được chọn"

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

$$P\{X=2\}=\frac{C_4^2C_6^1}{C_{10}^3}=\frac{9}{30}, \quad P\{X=1\}=\frac{C_4^1C_6^2}{C_{10}^3}=\frac{15}{30}$$

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

$$\mathrm{E}X=0\times\frac{5}{30}+1\times\frac{15}{30}+2\times\frac{9}{30}+3\times\frac{1}{30}=\frac{36}{30}=\frac{6}{5}.$$

2.30 Thắng 2 trong 4 ván dễ hơn.

2.31 a) $P=0{,}238$. b) $P=0{,}751$.

2.32 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) $\mathrm{Mod}X=4$; $P\{X=4\}=0{,}4096$.

2.33 a) $P=0{,}9914$

b) Số sản phẩm hỏng trung bình là 0,5

c) Số sản phẩm hỏng có khả năng xảy ra nhiều nhất là 0.

2.34 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 $P=\sum_{k=0}^{3}C_{10}^4\left(\frac{1}{5}\right)^k\left(\frac{4}{5}\right)^{10-k}=0{,}879$.

2.35 Gọi X là số lần thu được tín hiệu trong 5 lần phát độc lập thì $X\sim B(5;\,0,7)$

a) Xác suất thu được tín hiệu 2 lần $P\{X=2\}=C_5^2\,0,7^2 0,3^3=0{,}132$

b) Xác suất thu được tín hiệu nhiều nhất 1 lần $P\{X\le 1\}=0{,}031$

c) Xác suất thu được tín hiệu $P\{X\ge 1\}=1-P\{X=0\}=1-0{,}002=0{,}998$

2.36 Không đúng; $P=0{,}41$.

2.37 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 $P\{X \geq 1\} = 1 - P\{X = 0\} = 1 - e^{-1/3} = 0,2825$.

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 $P\{Y \leq 3\} = 0,151$.

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 $P\{Z \leq 1\} = 0,406$. Vậy xác suất để trong khoảng thời gian 3 phút liên tiếp mỗi phút có nhiều nhất 1 cuộc gọi là

$$P\{Z \leq 1\}^3 = 0,406^3 = 0,0067.$$

2.38 $P\{8 < X < 12\} = \Phi\left(\frac{12-10}{2}\right) - \Phi\left(\frac{8-10}{2}\right) = 0,6826$.

2.39 $P = 0,3$.

2.40 a) 95,44%; b) 4,56%.

2.41 a) 20,33%; b) $P = 0,9983$.

2.42 $\mathrm{E}(\overline{X}) = \mu$; $\mathrm{D}(\overline{X}) = \frac{\sigma^2}{n} \Rightarrow P\{|\overline{X} - \mu| < \varepsilon\} = 2\Phi\left(\frac{\varepsilon\sqrt{n}}{\sigma}\right) - 1$.

ĐÁP ÁN CHƯƠNG III

3.1	3.2	3.3	3.4	3.5	3.6	3.7	3.8	3.9	3.10
Sai	Đúng	Đúng	Đúng	Sai	Sai	Sai	Đúng	Đúng	Sai
3.11	**3.12**								
Sai	Đúng								

3.13

X	x_1	x_2
P	0,56	0,44

Y	y_1	y_2	y_3
P	0,26	0,38	0,36

3.14 $\mathrm{E}X = -7/15$; $\mathrm{E}Y = 0$;

$\operatorname{cov}(X,Y) = -1/8$; $\rho_{X,Y} = -0,15$.

3.15 $\mathrm{E}X = -1/5$; $\mathrm{E}Y = 0$; $\rho_{X,Y} = 0$. X và Y không độc lập vì

$P\{X=1\}=2/15,\ P\{Y=1\}=5/15$ và $P\{X=1.\,Y=1\}=0$.

3.16 Bảng phân bố xác suất của biến ngẫu nhiên Z

Z	1	2	3	4	6
P	0,12	0,43	0,03	0,35	0,07

$\mathrm{E}X=1{,}7$; $\mathrm{E}Y=1{,}7$; $\mathrm{E}Z=2{,}89$.

3.17 Bảng phân bố xác suất đồng thời của X và Y

X \ Y	0	1	2	3	4
0	0,04	0,12	0,16	0,06	0,02
1	0,03	0,09	0,12	0,045	0,015
2	0,02	0,06	0,08	0,03	0,01
3	0,01	0,03	0,04	0,015	0,005

$P\{X>Y\}=0{,}19$.

3.18 X, Y không độc lập vì

$P\{X=1\}=0{,}5,\ P\{Y=1\}=0{,}45$ và $P\{X=1.\,Y=1\}=0{,}15\neq 0{,}5\cdot 0{,}45$

$P\{X=1|Y=2\}=7/11$.

3.19

Y \ 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 $P\{X=1.\,Y=1\}=0{,}15\neq 0{,}5\cdot 0{,}45$.

3.20

$Y\|X=26$	23	27
P	0,357	0,643

$X\|Y=27$	26	30	41	50
P	0,1268	0,4225	0,1549	0,2958

3.21 $\mathrm{E}\left[Y \mid X=1\right]=5$. $\mathrm{E}X=2{,}93$; $\mathrm{E}Y=4{,}5$; $\mathrm{D}X=4{,}83$; $\mathrm{D}Y=2{,}25$.

3.22 a. $\alpha=15$; $\mathrm{E}X=-0{,}2$; $\mathrm{E}Y=0$.

b. $\operatorname{cov}(X,Y)=0 \Rightarrow \rho(X,Y)=0$.

c. X, Y không độc lập vì $P\{X=1\}=\dfrac{2}{15}$, $P\{Y=1\}=\dfrac{5}{15}$ nhưng $P\{X=1,Y=1\}=0$.

3.23 a) $k=3$;

b) $f_X(x)=\begin{cases} 3x^2 & \text{nếu } 0<x<1 \\ 0 & \text{nếu ngược lại} \end{cases}$; $f_Y(y)=\begin{cases} \dfrac{3}{2}(1-y^2) & \text{nếu } 0<y<1 \\ 0 & \text{nếu ngược lại} \end{cases}$.

c) X và Y không độc lập vì $P\left\{X<\dfrac{1}{2}, Y>\dfrac{1}{2}\right\}=0$ nhưng $P\left\{X<\dfrac{1}{2}\right\}\neq 0$, $P\left\{Y>\dfrac{1}{2}\right\}\neq 0$.

3.24 Áp dụng công thức (3.14) ta được

$$f(x,y)=\frac{\partial^2 F}{\partial x\partial y}=\begin{cases} e^{-x-y} & \text{nếu } x>0, y>0; \\ 0 & \text{nếu ngược lại}. \end{cases}$$

Áp dụng công thức (3.53) ta được

$$f(x|y)=\begin{cases} e^{-x} & \text{nếu } x>0, \\ 0 & \text{nếu } x\le 0. \end{cases}$$

3.25 a) $C=\dfrac{1}{\pi^2}$;

b) $F(x,y)=\left(\dfrac{1}{\pi}\operatorname{arctg} x+\dfrac{1}{2}\right)\left(\dfrac{1}{\pi}\operatorname{arctg} y+\dfrac{1}{2}\right)$;

c) $F_X(x)=\lim\limits_{y\to\infty} F(x,y)=\dfrac{1}{\pi}\operatorname{arctg} x+\dfrac{1}{2}$; $F_Y(y)=\lim\limits_{x\to\infty} F(x,y)=\dfrac{1}{\pi}\operatorname{arctg} y+\dfrac{1}{2}$;

Vì $F(x,y)=F_X(x)F_Y(y)$ nên ta kết luận X và Y độc lập.

d) $P\{1<X<\sqrt{3}, 0<Y<1\}=P\{1<X<\sqrt{3}\}P\{0<Y<1\}=\dfrac{1}{12}\cdot\dfrac{1}{4}=\dfrac{1}{48}$.

3.26 Hàm mật độ của X là $f_X(x)=\begin{cases} \dfrac{\ln x}{x^2} & \text{nếu } x\ge 1, \\ 0 & \text{nếu } x<1. \end{cases}$

Hàm mật độ của Y là $f_Y(y)=\begin{cases}\dfrac{1}{2y^2} & \text{nếu } 1\le y<\infty,\\ \dfrac{1}{2} & \text{nếu } 0\le y\le 1.\end{cases}$

Từ đó hàm mật độ có điều kiện của Y với điều kiện $X=x$ $(x>1)$ là

$$f(y|x)=\frac{f(x,y)}{f_X(x)}=\frac{1}{2y\ln x}\quad \text{nếu}\quad \frac{1}{x}\le y\le x\,;$$

Hàm mật độ có điều kiện của X với điều kiện $Y=y$ $(y>0)$ là

$$f(x|y)=\frac{f(x,y)}{f_Y(y)}=\begin{cases}\dfrac{1}{x^2y} & \text{nếu } 0\le y\le 1, x\ge y\\ \dfrac{y}{x^2} & \text{nếu } y\ge 1, y\le x.\end{cases}$$

3.27 $\mathrm{E}(2X-3Y)=2\mathrm{E}(X)-3\mathrm{E}(Y)=10$;

$\mathrm{D}(2X-3Y)=4\mathrm{D}(X)+9\mathrm{D}(Y)-12\sqrt{\mathrm{D}(X)\mathrm{D}(Y)}\,\rho_{X,Y}=57{,}6$.

ĐÁP ÁN CHƯƠNG IV

4.1	**4.2**	**4.3**	**4.4**	**4.5**	**4.6**	**4.7**	**4.8**
Đúng	Đúng	Sai	Sai	Đúng	Đúng	Đúng	Đúng

4.9 Gọi X là số máy hỏng trong ca. X có phân bố nhị thức $\mathrm{E}X=0{,}5$, $\mathrm{D}X=1$. Áp dụng bất đẳng thức Trêbưsép ta có

$$P\{|X-0{,}05|<2\}\ge 1-\frac{0{,}475}{2^2}=0{,}88\,;\quad P\{|X-0{,}05|\ge 2\}\le\frac{0{,}475}{2^2}=0{,}12\,.$$

4.10 Đặt $S=\sum_{n=1}^{12}X_n$; $\mathrm{E}S=12\cdot 16=192$, $\mathrm{D}S=12$. Theo bất đẳng thức Trêbưsép

$P\{|S-192|\le\varepsilon\}\ge 1-\dfrac{\mathrm{D}S}{\varepsilon^2}\ge 0{,}99$. Chọn $a=157{,}36$; $b=226{,}64$.

4.11 Đặt $S=\sum_{n=1}^{10000}X_n$; $\mathrm{E}S=0$, $\mathrm{D}S=\dfrac{10000}{12}$. Theo bất đẳng thức Trêbưsép

$$P\{|S|\ge 500\}\le\frac{\mathrm{D}S}{500^2}=\frac{1}{300}\,.$$

4.12 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 Trêbưsép

$$P\left\{|S - \mathrm{E}S| < \sqrt{n}\right\} \ge 1 - \frac{\mathrm{D}S}{n} = 1 - \frac{5}{36} = \frac{31}{36} \Leftrightarrow P\left\{\frac{n}{6} - \sqrt{n} < S < \frac{n}{6} + \sqrt{n}\right\} \ge \frac{31}{36}.$$

4.13 Đặt $S = \sum_{n=1}^{12} X_n$. Ta cần tìm M nhỏ nhất để $P\left\{\sum_{n=1}^{12} X_n \le M\right\} \ge 0{,}99$.

Ta có $\mathrm{E}S = 192$, $\mathrm{D}S = 12$. Theo bất đẳng thức Trêbưsép

$P\left\{|S - 192| \le \varepsilon\right\} \ge 1 - \frac{\mathrm{D}S}{\varepsilon^2} \ge 0{,}99 \Rightarrow \varepsilon = 34{,}64$. Vậy M = 192+34,64 = 226,64.

4.14 Thỏa mãn luật số lớn Trêbưsép.

4.15 Thỏa mãn luật số lớn Trêbưsép.

4.16 Thỏa mãn luật số lớn Trêbưsép.

4.17 Áp dụng bất đẳng thức Trêbưsép tính được xác suất $P \ge 0{,}9131$

4.18 Áp dụng bất đẳng thức Trêbưsép cần kiểm tra 23.750 chi tiết.

4.19 Gọi X là số sản phẩm hỏng. Ta có $X \sim \mathscr{B}(250\,;0{,}02)$. X sẽ có xấp xỉ phân bố Poisson với $\lambda = 250 \cdot 0{,}02 = 5$. Từ đó tra bảng ta được:

a) $P\{X = 2\} = 0{,}0842$;

b) $P\{X \le 2\} = 0{,}1247$

4.20 Giả sử X là số người chọn ăn ở đợt 1. Khi đó $1000 - X$ là số người chọn ăn ở đợt 2 . Gọi k là số chỗ ngồi trong nhà ăn. Ta phải chọn k nhỏ nhất để

$P\{X < k,\ 1000 - X < k\} \ge 0{,}99 \Leftrightarrow P\{1000 - k < X < k\} \ge 0{,}99$.

Ta xem X có phân bố chuẩn với $\mu = 500$, $\sigma = \sqrt{250}$. Vậy ta phải có

$$\Phi\left(\frac{k-500}{\sqrt{250}}\right) - \Phi\left(\frac{500-k}{\sqrt{250}}\right) \ge 0{,}99 \Leftrightarrow 2\Phi\left(\frac{k-500}{\sqrt{250}}\right) \ge 1{,}99 \Leftrightarrow \Phi\left(\frac{k-500}{\sqrt{250}}\right) \ge \Phi(2{,}58).$$

Từ đó $k \ge 500 + 2{,}58\sqrt{250} = 540{,}49$. Vậy $k = 541$.

4.21 a) Gọi X là số người trúng tuyển. Ta có $X \sim \mathscr{B}(350\,;0{,}9)$. X có phân bố xấp xỉ chuẩn với $\mu = 292{,}5$, $\sigma = 5{,}4$. Vậy $P\{X \le 300\} \approx \Phi\left(\frac{8}{5{,}4}\right) = \Phi(1{,}48) = 0{,}9306$.

b) Giả sử n là số người được gọi. Phân bố của X xấp xỉ phân bố chuẩn với $\mu = 0{,}9n$, $\sigma = 0{,}3\sqrt{n}$. Vậy

$$P\{X \le 300\} \approx \Phi\left(\frac{300-0{,}9n}{0{,}3\sqrt{n}}\right) \ge 0{,}99 = \Phi(2{,}33) \quad \Leftrightarrow\ 300-0{,}9n \ge (0{,}3)(2{,}33)\sqrt{n}\ .$$

Giải bất phương trình ta được $n \le 319{,}99$. Vậy $n = 319$.

ĐÁP ÁN CHƯƠNG V

5.1	**5.2**	**5.3**	**5.4**	**5.5**	**5.6**	**5.7**	**5.8**	**5.9**	**5.10**
Đúng	Sai	Đúng	Đúng	Đúng	Sai	Sai	Đúng	Đúng	Sai
5.11	**5.12**	**5.13**	**5.14**	**5.15**	**5.16**	**5.17**	**5.18**		
Sai	Sai	Đúng	Đúng	Sai	Sai	Sai	Sai		

5.19 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

$$P\left\{\sum_{i=1}^{10} X_i = 5\right\} = P_{10}(5) = C_{10}^5 (0{,}5)^5 \cdot (0{,}5)^{10-5} = C_{10}^5 (0{,}5)^{10}.$$

5.20 X có phân bố chuẩn $N(\mu;\sigma^2)$ nên $\overline{X}$ có phân bố chuẩn $N(\mu;\frac{\sigma^2}{n})$. Vậy

$P\{|\overline{X}-\mu| < \varepsilon\} = P\{\mu-\varepsilon < \overline{X} < \mu+\varepsilon\} = \Phi\left(\frac{\varepsilon\sqrt{n}}{\sigma}\right) - \Phi\left(\frac{-\varepsilon\sqrt{n}}{\sigma}\right) = 2\Phi\left(\frac{\varepsilon\sqrt{n}}{\sigma}\right)$. Do đó

$$P\{|\overline{X}-20| < 0{,}2\} = 2\Phi\left(\frac{0{,}2\sqrt{100}}{1}\right) = 2\Phi(2) = 0{,}9545.$$

5.21 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

$$F_{10}(x)=\begin{cases}0 & x\le 1\\ 1/5 & 1<x\le 2\\ 3/5 & 2<x\le 3\\ 4/5 & 3<x\le 4\\ 1 & x>4\end{cases}\qquad \bar{x}=6{,}8\,;\ s^2=1{,}15,\quad s=1{,}072\,.$$

5.22 $f=\dfrac{1082}{2000}$; 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.

5.23 $\bar{x}=\dfrac{\sum x_i}{n}=\dfrac{34,15}{35}=0,976$.

$$s^2=\frac{1}{n-1}\left[\sum x_i^{\,2}-\frac{\left(\sum x_i\right)^2}{n}\right]=\frac{1}{34}\left[33,8943-\frac{(34,15)^2}{35}\right]=0,01687$$

$\Rightarrow s=0,1299$; $u_\beta\dfrac{s}{\sqrt{n}}=1,96\times\dfrac{0,1299}{\sqrt{35}}=0,043$. Khoảng tin cậy 95%: $[0,933\,;\,1,019]$.

5.24 Tần suất mẫu $f=\dfrac{53}{400}$, 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 $p=\dfrac{2000}{N}$, trong đó N là số cá trong hồ.

Vậy $0,0993<\dfrac{2000}{N}<0,1657\ \Rightarrow\ \dfrac{2000}{0,1657}<N<\dfrac{2000}{0,0993}\ \Rightarrow\ 12070<N<20141$.

5.25 Đặt $u_i=\dfrac{x_i-18,25}{5}\Rightarrow \bar{x}=5\dfrac{\sum u_i}{n}+18,25=5\dfrac{-1,8}{40}+18,25=18,025$

$$s^2=\frac{5^2}{n-1}\left[\sum u_i^{\,2}-\frac{\left(\sum u_i\right)^2}{n}\right]=\frac{25}{39}\left[0,76-\frac{(-1,8)^2}{40}\right]=0,435$$

$$\Rightarrow s = 0,66;\quad u_\beta \frac{s}{\sqrt{n}} = 1,64 \times \frac{0,66}{\sqrt{40}} = 0,171$$

a) Khoảng tin cậy 90%: $[17,854; 18,196]$.

b) Kích thược mẫu cần thiết $n \geq \dfrac{u^2{}_\beta s^2}{\varepsilon^2} = 116,99$ chọn $n = 117$

5.26 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 $n = 385$

5.27 Gọi μ là trọng lượng trung bình của một bao sản phẩm được đóng gói. Ta kiểm định giả thiết $H_0: \mu = 100$; đối thiết $H_1: \mu < 100$

Tiêu chuẩn kiểm định $T = \dfrac{(100 - \overline{X})\sqrt{n}}{S}$; Miền bác bỏ $W_\alpha = \{T > 2,086\}$.

Đặt $u_i = \dfrac{x_i - 99,25}{5} \Rightarrow \sum r_i u_i = 0,4$; $\sum r_i u_i^2 = 0,42$

$$\Rightarrow \overline{x} = 5 \times \frac{0,4}{29} + 99,25 = 99,319 ;\quad s^2 = 25 \times \frac{1}{28}\left[0,42 - \frac{0,4^2}{29}\right] = 0,37 \Rightarrow s = 0,608$$

$$T_{qs} = \frac{(100 - 99,319)\sqrt{29}}{0,608} = 6,032 \in W_\alpha .$$

Vậy bác bỏ H_0 chấp nhận H_1, nghĩa là sản phẩm bị đóng thiếu.

5.28 Gọi μ là thời gian trung bình hoàn thành một sản phẩm.

Ta kiểm định giả thiết $H_0: \mu = 14$; đối thiết $H_1: \mu \neq 14$

Tiêu chuẩn kiểm định $T = \dfrac{(\overline{X} - 14)\sqrt{n}}{S}$; Miền bác bỏ $W_\alpha = \{|T| > 1,96\}$.

Đặt $u_i = \dfrac{x_i - 15}{2} \Rightarrow \sum r_i u_i = 0$; $\sum r_i u_i^2 = 300$

$$\Rightarrow \overline{x} = 15 ;\quad s^2 = 4 \times \frac{1}{249}\left[300 - \frac{0}{300}\right] = 4,819 \Rightarrow s = 2,195$$

$$\Rightarrow \quad T_{qs} = \frac{(115-14)\sqrt{300}}{2{,}195} = 7{,}89 \in W_\alpha .$$

Vậy bác bỏ H_0 chấp nhận H_1, nghĩa là cần thay đổi định mức.

5.29 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 $W_\alpha = \{T > 2{,}052\}$.

Theo mẫu ta có $\overline{x} = \dfrac{1387,5}{28} = 49,5536$;

$$s^2 = \frac{1}{27}\left(6876375 - \frac{1387,5^2}{28}\right) = \frac{8,1696}{27} = 0,3026 \quad \Rightarrow s = 0,55$$

$$T_{qs} = \frac{(50-49{,}53)\sqrt{30}}{0{,}55} = 4{,}2948 \in W_\alpha .$$

Vậy bác bỏ H_0 chấp nhận H_1, nghĩa là mức hao phí xăng có giảm xuống.

5.30 Gọi μ là số hoá đơn trung bình hệ thống máy tính mới xử lý được trong 1 giờ. Ta kiểm định giả thiết $H_0: \ \mu = 1300$; đối thiết $H_1: \ \mu > 1300$

Tiêu chuẩn kiểm định $T = \dfrac{\left(\overline{X}-1300\right)\sqrt{n}}{S}$; Miền bác bỏ $W_\alpha = \{T > 1{,}96\}$.

Từ mẫu cụ thể ta có $T = \dfrac{\left(1378-1300\right)\sqrt{40}}{215} = 2,294 > 1,96$

Vậy bác bỏ H_0 chấp nhận H_1, nghĩa là hệ thống máy tính mới xử lý tốt hơn.

ĐÁP ÁN CHƯƠNG VI

6.1	**6.2**	**6.3**	**6.4**	**6.5**	**6.6**
Sai	Sai	Đúng	Đúng	Đúng	Sai

6.7 $P\{X_0 = 0, X_1 = 2, X_2 = 1\}$

$$= P\{X_0 = 0\}P\{X_1 = 2 | X_0 = 0\}P\{X_2 = 1 | X_0 = 0, X_1 = 2\}$$

$$= P\left\{X_0=0\right\}P\left\{X_1=2\middle|X_0=0\right\}P\left\{X_2=1\middle|X_1=2\right\}=0,3\cdot 0,7\cdot 0,8=0,168\,.$$

6.8 a) $P^2=\begin{bmatrix}0,47 & 0,13 & 0,40\\ 0,42 & 0,14 & 0,44\\ 0,26 & 0,17 & 0,57\end{bmatrix}$.

b) $P\left\{X_3=1\middle|X_1=0\right\}=P\left\{X_2=1\middle|X_0=0\right\}=0,13$;

$$P\{X_3=1|X_0=0\}=P\{X_3=1,X_2=0|X_0=0\}+P\{X_3=1,X_2=1|X_0=0\}$$

$$+P\{X_3=1,X_2=2|X_0=0\}=P\{X_2=0|X_0=0\}P\{X_3=1|X_0=0,X_2=0\}$$

$$+P\{X_2=1|X_0=0\}P\{X_3=1|X_0=0,X_2=1\}+P\{X_2=2|X_0=0\}P\{X_3=1|X_0=0,X_2=2\}$$

$$=P\{X_2=0|X_0=0\}P\{X_3=1|X_2=0\}+P\{X_2=1|X_0=0\}P\{X_3=1|X_2=1\}$$

$$+P\{X_2=2|X_0=0\}P\{X_3=1|X_2=2\}=0,47\cdot 0,2+0,13\cdot 0,2+0,40\cdot 0,1=0,16\,.$$

c) Phân bố dừng $[x,y,z]$ là nghiệm của hệ phương trình

$$\begin{cases}[x\quad y\quad z]P=[x\quad y\quad z]\\ x,y,z\ge 0;\; x+y+z=1\end{cases}.$$

Như vậy x,y,z là nghiệm không âm của hệ phương trình

$$\begin{cases}-9x+2y+6z=0\\ 2x-8y+z=0\\ x+y+z=1\end{cases}$$

có nghiệm $x=\dfrac{50}{139},\ y=\dfrac{21}{139},\ z=\dfrac{68}{139}$.

6.9 Đặt $p_0=P\left\{X_0=0\right\}$.

a) $P\left\{X_0=0,X_1=0,X_2=0\right\}=P\left\{X_0=0\right\}P\left\{X_1=0,X_2=0\middle|X_0=0\right\}$

$$=P\left\{X_0=0\right\}P\left\{X_1=0\middle|X_0=0\right\}P\left\{X_2=0\middle|X_0=0,X_1=0\right\}=p_0\alpha^2\,.$$

b) $P\left\{X_0=0,X_1=0,X_2=0\right\}+P\left\{X_0=0,X_1=1,X_2=0\right\}=p_0\left(\alpha^2+(1-\alpha)^2\right)$.

c) $P\left\{X_5=0\middle|X_0=0\right\}=16(\alpha-1)^5+40(\alpha-1)^4+40(\alpha-1)^3+20(\alpha-1)^2+5(\alpha-1)+1$.

6.10 Không gian trạng thái sẽ là $E=\left\{-1,0,1,2,3\right\}$.

Theo công thức (6.21) ta có $p_{ij}=P\left\{X(n+1)=j\middle|X(n)=i\right\}=\begin{cases}P\left\{\xi=3-j\right\} & \text{nếu}\quad i\le 0,\\ P\left\{\xi=i-j\right\} & \text{nếu}\quad 0<i\le 3.\end{cases}$

$$p_{-1,-1} = P\left\{X(n+1) = -1 \middle| X(n) = -1\right\} = P(\phi) = 0\,,$$

$$p_{-1,0} = P\left\{X(n+1) = 0 \middle| X(n) = -1\right\} = P(\xi = 3) = P(\phi) = 0\,,$$

$$p_{-1,1} = P\left\{X(n+1) = 1 \middle| X(n) = -1\right\} = P(\xi = 2) = 0,3\,,$$

$$p_{-1,2} = P\left\{X(n+1) = 2 \middle| X(n) = -1\right\} = P(\xi = 1) = 0,3\,,$$

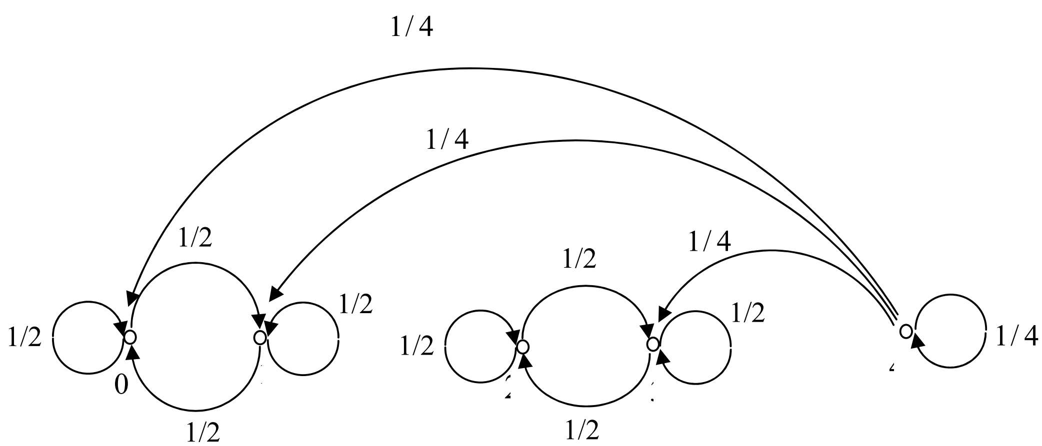
$$p_{-1,3} = P\left\{X(n+1) = 1 \middle| X(n) = -1\right\} = P(\xi = 0) = 0,4\,,$$

. . . .

Ma trận xác suất chuyển:

$$P = \begin{bmatrix} 0 & 0 & 0,3 & 0,3 & 0,4 \\ 0 & 0 & 0,3 & 0,3 & 0,4 \\ 0,3 & 0,3 & 0,4 & 0 & 0 \\ 0 & 0,3 & 0,3 & 0,4 & 0 \\ 0 & 0 & 0,3 & 0,3 & 0,4 \end{bmatrix}$$

6.12 Các trạng thái có chu kỳ 1. Có 3 lớp liên thông là $\{0,1\}$, $\{2,3\}$, $\{4\}$.



PHỤ LỤC

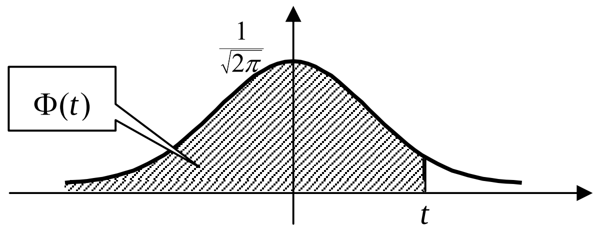
PHỤ LỤC I: GIÁ TRỊ HÀM MẬT ĐỘ XÁC SUẤT PHÂN BỐ CHUẨN TẮC

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

	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

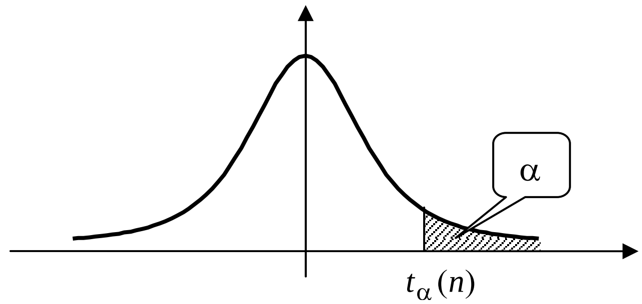
PHỤ LỤC II: GIÁ TRỊ HÀM PHÂN BỐ CHUẨN TẮC

$$\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{x^2}{2}} dx$$



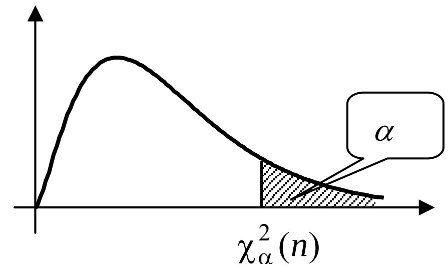
t	0	1	2	3	4	5	6	7	8	9
0,0	0,5000	5040	5080	5120	5160	5199	5239	5279	5319	5359
0,1	5398	5438	5478	5517	5557	5596	5636	5675	5714	5753
0,2	5793	5832	5871	5910	5948	5987	6026	6064	6103	6141
0,3	6179	6217	6255	6293	6331	6368	6406	6443	6480	6517
0,4	6554	6591	6628	6664	6700	6736	6772	6808	6844	6879
0,5	0,6915	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	$\alpha = 0{,}05$	$\alpha = 0{,}025$	$\alpha = 0{,}01$	$\alpha = 0{,}005$	$\alpha = 0{,}001$
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

TÀI LIỆU THAM KHẢO

[1]. Đặng Hùng Thắng, 1997. *Mở đầu về lý thuyết xác suất và các ứng dụng*. NXB GD.

[2]. Đặng Hùng Thắng, *Bài tập xác suất,* NXB Giáo dục – 1998.

[3]. Đặng Hùng Thắng, *Thống kê và ứng dụng,* NXB Giáo dục,1999.

[4]. Nguyễn Cao Văn và Trần Thái Ninh, *Bài giảng xác suất và thống kê toán,* NXB Thống kê, Hà Nội 1999.

[5]. 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.

[6]. Nguyễn Phạm Anh Dũng, 1999. *Các hàm và xác suất ứng dụng trong viễn thông.* Trung Tâm Đào Tạo Bưu Chính Viễn Thông 1.

[7]. Nguyễn Duy Tiến, Vũ Việt Yên, 2000. *Lý thuyết xác suất*. NXB GD.

[8]. Nguyễn Duy Tiến (và tập thể), 2000. *Các mô hình xác suất và ứng dụng*, tập 1, 2, 3. NXB Đại học Quốc gia Hà Nội.

[9]. Tống Đình Quỳ, *Hướng dẫn giải bài tập xác suất thống kê*, NXB Đại Học Quốc Gia Hà Nội, 2004.

[10]. Trần Mạnh Tuấn, *Xác suất và Thống kê, lý thuyết và thực hành tính toán,* NXB Đại Học Quốc Gia Hà Nội, 2004.

[11]. B.V. Gnedenko, *The theory of* probability, Mir publishers, Moscow 1976.

[12]. D. L. (Paul) Minh, *Applied Probability Models*, Duxbury, Thomson Learning, 2001.

[13]. J. L. Doob, 1953. *Stochastic Processes*. Willey and Sons, New York.

[14]. S. Karlin, 1966. *A first Course in Stochastic Processes*. Academic Press, New York and London.

[15]. M. Loeve, 1977. *Probability Theory, I, II. 4th ed*, Springer - Verlag, Berlin and New York.