

HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG
PGS.TS. Lê Bá Long

BÀI GIẢNG
XÁC SUẤT VÀ THỐNG KÊ

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

Hà Nội, 2013

LỜI NÓI ĐẦU

Tập bài giảng Xác suất và Thông kê dành cho sinh viên hệ đại học chuyên ngành Điện tử-Viễn thông, Công nghệ thông tin và An toàn thông tin được biên soạn lại trên cơ sở giáo trình Xác suất và Thống kê của cùng tác giả xuất bản năm 2009, nhằm đáp ứng yêu cầu đào tạo theo hình thức tín chỉ và phù hợp với đề cương chi tiết môn học do Học viện Công nghệ Bưu Chính Viễn Thông ban hành năm 2012 theo hình thức đào tạo tín chỉ.

Nội dung của cuốn sách cũng được hoàn thiện từ các bài giảng trong nhiều năm của tác giả theo định hướng ứng dụng trong các ngành kỹ thuật. Chính vì thế, tập bài giảng này có thể dùng làm tài liệu học tập, tài liệu tham khảo cho sinh viên của các trường đại học và cao đẳng khối kỹ thuật.

Giáo trình gồm 5 chương tương ứng với 2 tín chỉ:

Chương 1: Các khái niệm cơ bản về xác suất.

Chương 2: Biến ngẫu nhiên và các đặc trưng của chúng.

Chương 3: Véc tơ ngẫu nhiên và các đặc trưng của chúng.

Chương 4: Lý thuyết mẫu

Chương 5: Lý thuyết ước lượng và kiểm định giả thiết thống kê.

Điều kiện tiên quyết cho môn học xác suất và thống kê là môn đại số và giải tích 1, giải tích 2 trong chương trình toán đại cương.

Giáo trình được viết cho đối tượng là sinh viên các trường đại học khối kỹ thuật, vì vậy tác giả cung cấp nhiều ví dụ minh họa tương ứng với từng phần lý thuyết và có nhiều ví dụ ứng dụng vào lĩnh vực chuyên ngành Điện tử Viễn thông và Công nghệ thông tin. Ngoài ra tác giả cũng có ý thức trình bày thích hợp đối với người tự học. Trước khi nghiên cứu các nội dung chi tiết, người 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. Trong mỗi nội dung tác giả luôn có ý thức cung cấp nhiều ví dụ để minh họa 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 mỗi chương có các câu hỏi luyện tập và bài tập. Có khoảng từ 30 đến 40 bài tập cho mỗi chương, tương ứng với 8 -10 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 tự kiểm tra được mức độ tiếp thu lý thuyết của mình.

Với thời lượng ứng với 2 tín chỉ của môn học giảng viên khó có đủ thời gian để trình bày hết các nội dung của tập bài giảng ở trên lớp. Vì vậy tác giả đánh dấu (*) cho các nội dung dành cho sinh viên tự học.

Tác giả xin chân thành cảm ơn PGS.TS. Phạm Ngọc Anh, PGS. TS. Tô Văn Ban, PGS. TS. Nguyễn Năng Anh, TS. Nguyễn Hắc Hải, GVC. Ths. Lê Bá Cầu, Ths. Trần Việt Anh đã cho những ý kiến đóng góp quý giá.

Mặc dù tác giả đã rất cố gắng, song do yêu cầu cấp bách cần có tài liệu phục vụ việc giảng dạy và học tập của Học viện theo hình thức tín chỉ, thời gian biên soạn bị hạn hẹp vì vậy các thiếu sót còn tồn tại trong giáo trình là điều khó tránh khỏi. Tác giả rất mong nhận được sự đóng góp ý kiến của bạn đọc xa gần.

Cuối cùng tác giả bày tỏ sự cảm 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 tác giả hoàn thành giáo trình này.

Lê Bá Long

Khoa cơ bản 1

Học Viện CNBCVT

MỤC LỤC

LỜI NÓI ĐẦU	3
MỤC LỤC	5
CHƯƠNG I: CÁC KHÁI NIỆM CƠ BẢN VỀ XÁC SUẤT	9
1.1 PHÉP THỬ VÀ BIẾN CỐ	9
1.1.1 Phép thử	9
1.1.2 Biến cố	10
1.1.3 Quan hệ giữa các biến cố	10
1.2 CÁC ĐỊNH NGHĨA VÀ TÍNH CHẤT CỦA XÁC SUẤT	13
1.2.1 Định nghĩa cổ điển về xác suất	13
1.2.2 Các qui tắc đếm	15
1.2.3 Định nghĩa xác suất theo thống kê	21
1.2.4 Định nghĩa xác suất theo hình học	21
1.2.5 Các tính chất và định lý xác suất	23
1.2.6 Nguyên lý xác suất lớn, xác suất nhỏ	26
1.3 XÁC SUẤT CÓ ĐIỀU KIỆN	27
1.3.1 Định nghĩa và các tính chất của xác suất có điều kiện	27
1.3.2 Quy tắc nhân xác suất	29
1.3.3 Công thức xác suất đầy đủ	32
1.3.4 Công thức Bayes	34
1.4 DÃY PHÉP THỬ BERNOULLI	38
CÂU HỎI ÔN TẬP VÀ BÀI TẬP CHƯƠNG 1	40
CHƯƠNG 2: BIẾN NGẪU NHIÊN VÀ CÁC ĐẶC TRƯNG CỦA CHÚNG	45
2.1 ĐỊNH NGHĨA VÀ PHÂN LOẠI BIẾN NGẪU NHIÊN	45
2.1.1 Định nghĩa biến ngẫu nhiên	46
2.1.2 Hàm phân bố xác suất	46
2.1.3 Phân loại	50
2.2 BIẾN NGẪU NHIÊN RỜI RẠC	51
2.2.1 Hàm khối lượng xác suất và bảng phân bố xác suất của biến ngẫu nhiên rời rạc	51
2.2.2 Các phân bố rời rạc thường gặp	54
2.3 BIẾN NGẪU NHIÊN LIÊN TỤC	59
2.3.1 Hàm mật độ xác suất của biến ngẫu nhiên liên tục	59
2.3.2 Các phân bố liên tục thường gặp	61
2.4 CÁC THAM SỐ ĐẶC TRƯNG CỦA BIẾN NGẪU NHIÊN	70
2.4.1 Kỳ vọng toán	70
2.4.2 Phương sai	74
2.4.3 Phân vị, Trung vị	76
2.4.4 Môđ	77
2.4.5 Moment, hệ số bất đối xứng, hệ số nhọn (*)	78
2.4.6 Kỳ vọng và phương sai của các biến ngẫu nhiên có phân bố xác suất thường gặp	79
TÓM TẮT	80

CÂU HỎI ÔN TẬP VÀ BÀI TẬP CHƯƠNG 2	81
CHƯƠNG 3: VÉC TƠ NGẪU NHIÊN VÀ CÁC ĐẶC TRƯNG CỦA CHÚNG	87
3.1 KHÁI NIỆM VÉC TƠ NGẪU NHIÊN	87
3.1.1 Khái niệm và phân loại véc tơ ngẫu nhiên	87
3.1.2 Hàm phân bố xác suất đồng thời và hàm phân bố xác suất biên	88
3.2 VÉC TƠ NGẪU NHIÊN RỜI RẠC	90
3.2.1 Hàm khối lượng xác suất đồng thời và bảng phân bố xác suất đồng thời.....	90
3.2.2 Bảng phân bố xác suất biên	91
3.3 VÉC TƠ NGẪU NHIÊN LIÊN TỤC	94
3.3.1 Hàm mật độ xác suất đồng thời	94
3.3.2 Hàm mật độ xác suất biên	95
3.4 TÍNH ĐỘC LẬP CỦA CÁC BIẾN NGẪU NHIÊN.....	97
3.5 CÁC THAM SỐ ĐẶC TRƯNG CỦA VÉC TƠ NGẪU NHIÊN	98
3.5.1 Kỳ vọng và phương sai của các biến ngẫu nhiên thành phần.....	98
3.5.2 Hiệp phương sai.....	99
3.5.3 Ma trận hiệp phương sai.....	99
3.5.4 Hệ số tương quan	100
3.6 PHÂN BỐ CÓ ĐIỀU KIỆN VÀ KỲ VỌNG CÓ ĐIỀU KIỆN.....	102
3.6.1 Phân bố có điều kiện và kỳ vọng có điều kiện của biến ngẫu nhiên rời rạc.....	102
3.6.2 Phân bố có điều kiện và kỳ vọng có điều kiện của biến ngẫu nhiên liên tục	104
3.6.3 Kỳ vọng có điều kiện	106
3.7 LUẬT SỐ LỚN VÀ ĐỊNH LÝ GIỚI HẠN	107
3.7.1 Hội tụ theo xác suất và hội tụ theo phân bố của dãy biến ngẫu nhiên	108
3.7.2 Luật số lớn.....	108
3.7.3 Định lý giới hạn trung tâm	113
3.7.4 Xấp xỉ phân bố nhị thức	113
TÓM TẮT.....	116
CÂU HỎI ÔN TẬP VÀ BÀI TẬP CHƯƠNG 3	117
CHƯƠNG 4: LÝ THUYẾT MẪU	124
4.1 SỰ CẦN THIẾT PHẢI LẤY MẪU	124
4.2 MẪU NGẪU NHIÊN.....	125
4.2.1 Khái niệm mẫu ngẫu nhiên.....	125
4.2.2 Mô hình hóa mẫu ngẫu nhiên	125
4.2.3 Biểu diễn giá trị cụ thể của mẫu ngẫu nhiên theo bảng và theo biểu đồ	126
4.3 THỐNG KÊ VÀ CÁC ĐẶC TRƯNG CỦA MẪU NGẪU NHIÊN	131
4.3.1 Định nghĩa thống kê.....	131
4.3.2 Trung bình mẫu.....	131
4.3.3 Phương sai mẫu, Độ lệch chuẩn mẫu.....	132
4.3.4 Tần suất mẫu.....	133
4.3.5 Cách tính giá trị cụ thể của trung bình mẫu \bar{x} và phương sai mẫu có hiệu chỉnh s^2	133
4.4 PHÂN BỐ XÁC SUẤT CỦA MỘT SỐ THỐNG KÊ ĐẶC TRƯNG MẪU.....	135

4.4.1 Trường hợp biến ngẫu nhiên gốc có phân bố chuẩn	135
4.4.2 Trường hợp biến ngẫu nhiên gốc có phân bố Bernoulli	137
CÂU HỎI ÔN TẬP VÀ BÀI TẬP CHƯƠNG 4.....	139
CHƯƠNG 5: ƯỚC LƯỢNG THAM SỐ VÀ KIỂM ĐỊNH GIẢ THIẾT THỐNG KÊ.....	142
5.1 PHƯƠNG PHÁP ƯỚC LƯỢNG ĐIỂM.....	142
5.1.1 Khái niệm ước lượng điểm	142
5.1.2 Ước lượng không chệch (unbiased estimator)	142
5.1.3 Ước lượng hiệu quả (efficient estimator)	143
5.1.4 Ước lượng vững (consistent estimator)	144
5.2 PHƯƠNG PHÁP ƯỚC LƯỢNG BẰNG KHOẢNG TIN CẬY	144
5.2.1 Khái niệm khoảng tin cậy	145
5.2.2 Khoảng tin cậy của kỳ vọng của biến ngẫu nhiên phân bố chuẩn	145
5.2.2 Khoảng tin cậy cho tần suất của tổng thể	149
5.3 KHÁI NIỆM CHUNG KIỂM ĐỊNH GIẢ THIẾT THỐNG KÊ	150
5.3.1 Giả thiết thống kê	150
5.3.2 Tiêu chuẩn kiểm định giả thiết thống kê	151
5.3.3 Miền bác bỏ giả thiết	151
5.3.4 Giá trị quan sát của tiêu chuẩn kiểm định.....	151
5.3.5 Quy tắc kiểm định giả thiết thống kê.....	151
5.3.6 Sai lầm loại một và sai lầm loại hai.....	152
5.3.7 Thủ tục kiểm định giả thiết thống kê.....	153
5.4 KIỂM ĐỊNH THAM SỐ.....	153
5.4.1 Kiểm định giả thiết về kỳ vọng của biến ngẫu nhiên có phân bố chuẩn	153
5.4.2 Kiểm định tham số của biến ngẫu nhiên phân bố Bernoulli	159
TÓM TẮT	160
CÂU HỎI ÔN TẬP VÀ BÀI TẬP CHƯƠNG 5.....	161
HƯỚNG DẪN VÀ ĐÁP ÁN	165
HƯỚNG DẪN VÀ ĐÁP ÁN CHƯƠNG 1	165
HƯỚNG DẪN VÀ ĐÁP ÁN CHƯƠNG 2	167
HƯỚNG DẪN VÀ ĐÁP ÁN CHƯƠNG 3	173
HƯỚNG DẪN VÀ ĐÁP ÁN CHƯƠNG 4	179
HƯỚNG DẪN VÀ ĐÁP ÁN CHƯƠNG 5	180
PHỤ LỤC 1: GIÁ TRỊ HÀM MẬT ĐỘ XÁC SUẤT PHÂN BỐ CHUẨN TẮC	185
PHỤ LỤC 2: GIÁ TRỊ HÀM PHÂN BỐ CHUẨN TẮC	186
PHỤ LỤC 3: GIÁ TRỊ TỚI HẠN CỦA PHÂN BỐ STUDENT	187
PHỤ LỤC 4: GIÁ TRỊ TỚI HẠN CỦA PHÂN BỐ “KHI BÌNH PHƯƠNG”	188
PHỤ LỤC 5: GIÁ TRỊ HÀM PHÂN BỐ POISSON.....	189
BẢNG CHỈ DẪN THUẬT NGỮ	191
TÀI LIỆU THAM KHẢO	194

PTT

CHƯƠNG I: CÁC KHÁI NIỆM CƠ BẢN VỀ XÁC SUẤT

Các hiện tượng trong tự nhiên hay xã hội xảy ra một cách ngẫu nhiên (không biết trước kết quả) hoặc tất định (biết trước kết quả sẽ xảy ra). Chẳng hạn một vật nặng được thả từ trên cao chắc chắn sẽ rơi xuống đất, trong điều kiện bình thường nước sôi ở 100°C ... Đó là những hiện tượng diễn ra có tính quy luật, tất nhiên. Trái lại khi tung đồng xu ta không biết mặt sấp hay mặt ngửa sẽ xuất hiện. Ta không thể biết trước có bao nhiêu cuộc gọi đến tổng đài, có bao nhiêu khách hàng đến điểm phục vụ trong khoảng thời gian nào đó. Ta không thể xác định trước chỉ số chứng khoán trên thị trường chứng khoán ở một thời điểm khớp lệnh trong tương lai... Đó là những hiện tượng ngẫu nhiên. Tuy nhiên, nếu tiến hành quan sát khá nhiều lần một hiện tượng ngẫu nhiên trong những hoàn cảnh như nhau, thì trong nhiều trường hợp ta có thể rút ra những kết luận có tính quy luật về những hiện tượng này. Lý thuyết xác suất nghiên cứu các quy 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

1.1 PHÉP THỬ VÀ BIẾN CỐ

1.1.1 Phép thử

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

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

Chẳng hạn, với phép thử gieo con xúc xắc (6 mặt), 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 chấm 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 . Không gian mẫu của phép thử gieo con xúc xắc là $\Omega = \{1, 2, 3, 4, 5, 6\}$.

Ví dụ 1.1:

- Phép thử tung đồng xu có hai khả năng xảy ra là mặt sấp, ký hiệu S , hoặc mặt ngửa, ký hiệu N . Ta gọi S, N là các biến cố sơ cấp. Không gian mẫu của phép thử 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

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

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

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

Nhận xét 1.1:

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

2. Mỗi biến cố chỉ có thể xảy ra khi một phép thử được thực hiện, nghĩa là gắn với không gian mẫu nào đó.

Có hai biến cố đặc biệt sau:

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

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

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

1.1.3 Quan hệ giữa các biến cố

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

A) Quan hệ biến cố đối

Với mỗi biến cố A , luôn luôn có biến cố gọi là biến cố đối của A , ký hiệu \bar{A} và được xác định như sau: Biến cố \bar{A} xảy ra khi và chỉ khi biến cố 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 \bar{A} : “bắn trượt bia”.

B) Tổng của hai biến cố

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

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

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

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

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 $A = A_1 \cup A_2$.

C) Tích của hai biến cố

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

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

Tích của một dãy các biến cố $\{A_1, A_2, \dots, A_n\}$ là biến cố $A_1 \cap A_2 \cap \dots \cap A_n$ hoặc $\bigcap_{i=1}^n A_i$. Biến cố tích xảy ra khi **tất cả các biến cố A_i đồng thời cùng xảy ra**, với mọi $i = 1, \dots, n$.

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 $A = A_1 \cap A_2$.

Ví dụ 1.6: Hai xạ thủ A và B mỗi người bắn một viên đạn vào bia. Gọi A là biến cố “ A bắn trúng bia”, B là biến cố “ B bắn trúng bia”. Khi đó $A \cup B$ là biến cố “có ít nhất một người bắn trúng bia” và $A \cap B$ là biến cố “cả hai người cùng bắn trúng bia”.

D) Biến cố xung khắc

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

Nói cách khác biến cố tích $A \cap B$ là biến cố không thể, nghĩa là $A \cap B = \emptyset$.

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

Ví dụ 1.7: Một bình có 3 loại cầu: cầu màu trắng, màu đỏ và màu xanh. Lấy ngẫu nhiên 1 cầu từ bình. Gọi A_t, A_d, 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 màu.

E) Hệ đầy đủ các biến cố

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

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

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

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

Ví dụ 1.8: Một nhà máy có ba phân xưởng sản xuất ra cùng một loại sản phẩm. Giả sử rằng mỗi sản phẩm của nhà máy chỉ do một trong ba phân xưởng này sản xuất. Chọn ngẫu nhiên một sản phẩm, gọi A_1, A_2, A_3 lần lượt là biến cố sản phẩm được chọn do phân xưởng thứ nhất, thứ hai, thứ ba sản xuất. Khi đó hệ ba biến cố $\{A_1, A_2, A_3\}$ là hệ đầy đủ.

Hệ ba biến cố $\{A_t, A_d, A_x\}$ trong ví dụ 1.7 cũng là đầy đủ.

F) Tính độc lập của các biến cố

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

Tổng quát hơn, các biến cố A_1, A_2, \dots, 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 một nhóm nào đó các biến cố còn lại.

Ví dụ 1.9: Ba xạ thủ A, B, C mỗi người bắn một viên đạn vào mục tiêu. Gọi A, B, C lần lượt là biến cố A, B, C bắn trúng mục tiêu.

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

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

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

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

Giải: a. $A \cap B \cap C$: cả 3 đều bắn trúng. $\bar{A} \cap \bar{B} \cap \bar{C}$: cả 3 đều bắn trượt. $A \cup B \cup C$: có ít nhất 1 người bắn trúng.

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

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

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

$$F = \bar{A} \cap \bar{B} \cap C.$$

$$G = (A \cap \bar{B} \cap \bar{C}) \cup (\bar{A} \cap B \cap \bar{C}) \cup (\bar{A} \cap \bar{B} \cap C).$$

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.

Nhận xét 1.2:

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

toán tích các biến cố có tính giao hoán, kết hợp, tổng phân bố đối với tích, tích phân bố đối với tổng, thỏa mãn luật De Morgan ...

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

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

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

1.2 CÁC ĐỊNH NGHĨA VÀ TÍNH CHẤT CỦA XÁC SUẤT

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

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

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

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

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

1.2.1 Định nghĩa cổ điển về xác suất

Định nghĩa 1.1: Giả sử phép thử C thỏa 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

$$P(A) = \frac{\text{số trường hợp thuận lợi đối với } A}{\text{số trường hợp cơ bản}} \quad (1.1a)$$

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

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

Ví dụ 1.10: Biến cố A xuất hiện mặt chẵn trong phép thử gieo con xúc xắc ở ví dụ 1.2 có 3 trường hợp thuận lợi ($|A| = 3$) và 6 trường hợp có thể ($|\Omega| = 6$). Vậy $P(A) = \frac{3}{6} = \frac{1}{2}$.

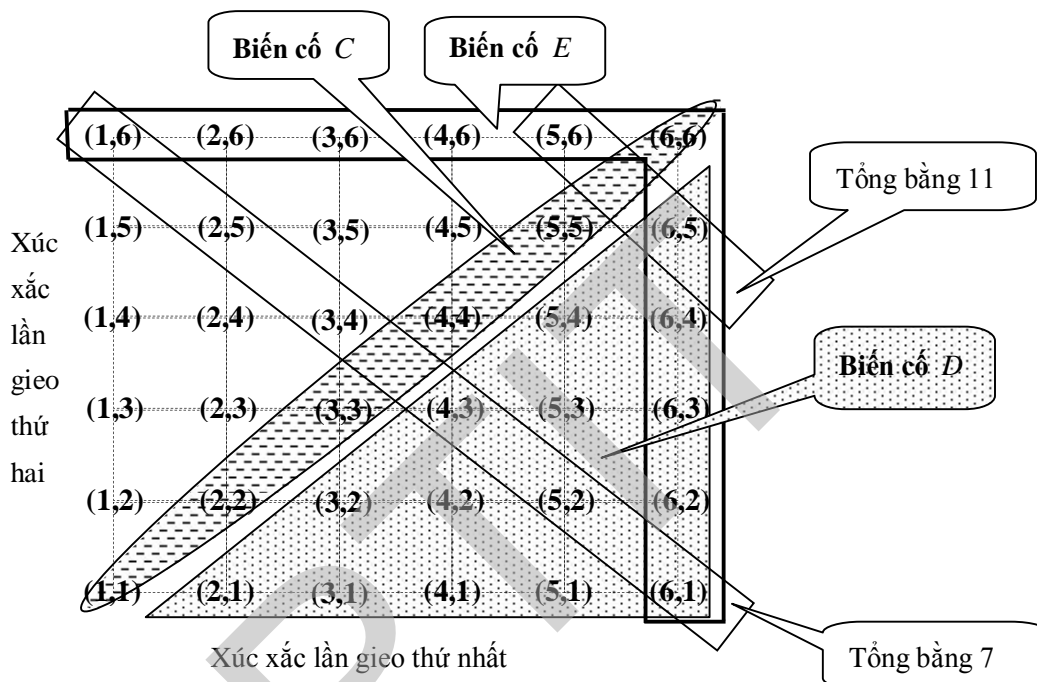
Biến cố xuất hiện một mặt sấp và một mặt ngửa khi gieo đồng thời hai đồng xu có 2 kết quả thuận lợi và 4 kết quả đồng khả năng có thể, vậy có xác suất xuất hiện của biến cố đó là $\frac{1}{2}$.

Ví dụ 1.11: Xét phép thử gieo liên tiếp 2 lần con xúc xắc. Tính xác suất của các biến cố sau:

- a. Tổng số chấm xuất hiện là chẵn (biến cố A).

- b. Tổng số chấm xuất hiện bằng 7 hoặc 11 (biến cố B).
- c. Số chấm xuất hiện của hai con xúc xắc bằng nhau (biến cố C).
- d. Số chấm của xúc xắc thứ nhất lớn hơn xúc xắc thứ hai (biến cố D).
- e. Ít nhất một xúc xắc xuất hiện mặt 6 chấm (biến cố E).

Giải: Để có hình ảnh trực quan ta có thể biểu diễn không gian mẫu của phép thử và các biến cố tương ứng dưới dạng biểu đồ. Các biến cố sơ cấp được biểu diễn các cặp số tương tự tọa độ của các điểm. Không gian mẫu tương ứng với 36 điểm.



Hình 1.1: Phép thử gieo 2 xúc xắc

Mỗi hàng có 3 biến cố sơ cấp thuận lợi đối với biến cố A , chẳng hạn hàng dưới cùng có (1,1), (1,3), (1,5) hàng tiếp (2,2), (2,4), (2,6) ... như vậy biến cố A có 18 kết quả thuận lợi.

Các điểm thuộc đường chéo thứ hai (hoặc song song đường chéo thứ hai) có tổng hai thành phần bằng nhau: $6+1=5+2=4+3=\dots=1+6=7$.

Biến cố C là các điểm thuộc đường chéo.

Biến cố D là các điểm phía dưới đường chéo.

Theo định nghĩa xác suất (1.1a) ta có:

$$\text{a. } P(A) = \frac{18}{36} = \frac{1}{2}.$$

$$\text{b. } P(B) = \frac{8}{36} = \frac{2}{9}.$$

$$\text{c. } P(C) = \frac{6}{36} = \frac{1}{6}.$$

$$\text{d. } P(D) = \frac{15}{36} = \frac{5}{12}.$$

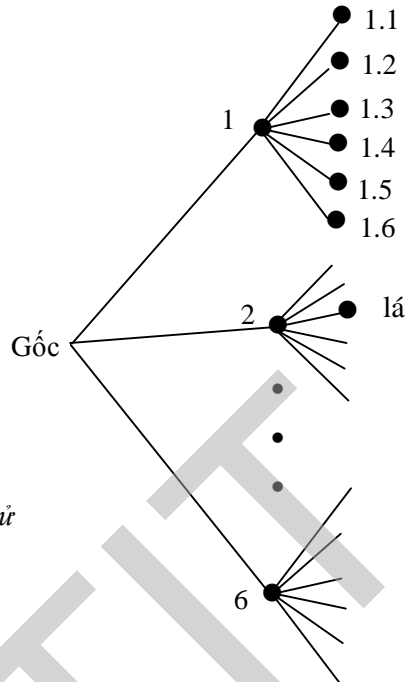
$$\text{e. } P(E) = \frac{11}{36}.$$

Ví dụ 1.12: Sơ đồ cây

Nhiều phép thử có tính chất nối tiếp lập thành dãy, chẳng hạn phép thử tung liên tiếp đồng xu ba lần, quan sát chỉ số chứng khoán trong năm ngày liên tiếp, hoặc tám ký số liên tiếp

nhận được của một bộ nhận thông tin ... Trong trường hợp này ta có thể biểu diễn không gian mẫu và các biến cố tương ứng dưới dạng sơ đồ cây.

Không gian mẫu và biến cố B của ví dụ 1.11 được biểu diễn dạng sơ đồ cây như sau



Hình 1.2: Sơ đồ cây của phép thử gieo 2 xúc xắc

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 $m_1 + m_2 + \dots + m_n$ cách chọn một trong các đối tượng đã cho.

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

B. Qui tắc nhân

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

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

Ví dụ 1.13: Một nhân viên có 4 chiếc áo sơ mi và 3 quần dài đồng phục, thì anh ta có $4 \cdot 3 = 12$ cách chọn áo sơ mi và quần đồng phục.

Ví dụ 1.14: Tung một con xúc xắc (6 mặt) hai lần. Tìm xác suất để trong đó có 1 lần ra 6 chấm.

Giải: Theo quy tắc nhân ta có số các trường hợp có thể khi tung con xúc xắc 2 lần là $6 \cdot 6 = 36$.

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

Ví dụ 1.15:

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

Giải:

- a. Có 9 cách chọn chữ số đầu tiên (vì chữ số đầu tiên khác 0) và các chữ số còn lại có 10 cách chọn cho từng chữ số. Vậy có $9 \cdot 10 \cdot 10 \cdot 10 = 9000$ số cần tìm.
- b. Có 9 cách chọn chữ số đầu tiên (vì chữ số đầu tiên khác 0), 9 cách chọn chữ số thứ hai, 8 cách chọn chữ số thứ ba và 7 cách chọn chữ số thứ tư. Vậy có $9 \cdot 9 \cdot 8 \cdot 7 = 4536$ số cần tìm.
- c. Vì chữ số thứ tư là số 0 và các chữ số này khác nhau do đó có 9 cách chọn chữ số đầu tiên, 8 cách chọn chữ số thứ hai, 7 cách chọn chữ số thứ ba. Vậy có $9 \cdot 8 \cdot 7 = 504$ số cần tìm.

C. Hoán vị

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

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

Ví dụ 1.16:

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

Giải:

- a. Số cách bố trí 9 SV (gồm 5 nam SV và 4 nữ SV) theo một hàng là $9! = 362880$.
- b. Có $5!$ cách bố trí nam SV, ứng với mỗi cách bố trí nam SV có $4!$ cách bố trí nữ SV vào vị trí chẵn tương ứng. Vậy có $5!4! = 2880$ cách bố trí theo yêu cầu.

Ví dụ 1.17: (Hoán vị vòng tròn) Có n người ($n \geq 3$), trong đó có hai người là anh em.

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

Giải:

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

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

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

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

Ví dụ 1.18: Xếp ngẫu nhiên 6 cuốn sách toán và 4 sách lý vào 1 giá sách. Tính xác suất 3 cuốn sách toán đứng cạnh nhau.

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

Ta xem 3 cuốn sách toán đứng cạnh nhau như là một cuốn sách lớn. Như vậy ta cần sắp xếp 8 cuốn sách vào giá sách (có $8!$ cách), ngoài ra 3 cuốn sách toán đứng cạnh nhau có

$3!$ cách sắp xếp. Do đó số các trường hợp thuận lợi là $8!3!$. Vậy $P = \frac{8!3!}{10!} = \frac{1}{15}$.

D. Chính hợp

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

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

Ví dụ 1.19: Có $A_{10}^4 = 10 \cdot 9 \cdot 8 \cdot 7 = 5040$ cách bố trí 10 người ngồi vào 4 chỗ.

Ví dụ 1.20: 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 chính hợp chập 2 của 10 phần tử. Vậy số các trường hợp có thể là $A_{10}^2 = 10 \cdot 9 = 90$.

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

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

E. Tổ hợp

Một tổ hợp chập k của n phần tử là một cách chọn đồng thời k phần tử từ một tập có n phần tử ($1 \leq k \leq n$). Cũng có thể xem một tập con k phần tử của tập n phần tử là một tổ hợp chập k của n phần tử.

Hai chính hợp chập k của n phần tử là khác nhau nếu thỏa mãn một trong hai điều kiện sau:

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

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

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

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

Một vài trường hợp cụ thể

$$C_n^0 = 1; C_n^1 = n; C_n^2 = \frac{n(n-1)}{2}; C_n^3 = \frac{n(n-1)(n-2)}{6}; C_n^k = C_n^{n-k}. \quad (1.4)$$

Ví dụ 1.21: 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 1 nữ trúng tuyển.

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

- a. Chỉ có 1 trường hợp cả 2 nam đều trúng tuyển do đó xác suất tương ứng là $P = \frac{1}{15}$.
- b. Có $C_4^2 = \frac{4 \cdot 3}{2} = 6$ cách chọn 2 trong 4 nữ, vậy xác suất tương ứng $P = \frac{6}{15}$.
- c. Trong 15 trường hợp có thể chỉ có 1 trường hợp cả 2 nam được chọn, vậy có 14 trường hợp ít nhất 1 nữ được chọn. Do đó xác suất tương ứng $P = \frac{14}{15}$.

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

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

- Có 2 nữ được chọn: Có 6 cách
- Có 1 nữ và 1 nam được chọn: Có $4 \cdot 2$ cách chọn

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

Ví dụ 1.22: Một hộp có 8 bi màu đỏ, 3 bi trắng và 9 bi màu xanh. Lấy ngẫu nhiên 3 bi từ hộp. Tính xác suất trong các trường hợp sau:

- a. 3 bi lấy được cùng màu đỏ
- b. 2 đỏ và 1 trắng
- c. Ít nhất 1 trắng
- d. Mỗi màu 1 bi
- e. Nếu lấy lần lượt không hoàn lại 3 bi, tính xác suất lấy được mỗi màu 1 bi.

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

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

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

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

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

Ví dụ 1.23: 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 các trường hợp $k = 0, \dots, 6$.

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 chập k của 6 phần tử. Do đó $|A_k| = C_6^k = \frac{6!}{k!(6-k)!}$

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

Tương tự xác suất của các từ có chứa k bit 0 cũng bằng $\frac{6!}{k!(6-k)!2^6}$ (điều này có thể suy ra từ tính chất $C_n^k = C_n^{n-k}$).

Nhận xét 1.3:

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

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

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

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

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

Ta có thể mở rộng kết quả này như sau.

Công thức tổ hợp mở rộng

Số tổ hợp chập k của n phần tử bằng số tổ hợp chập $n - k$ của n phần tử:

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

Chúng ta thấy rằng: số tổ hợp chập k của n phần tử (số cách chọn đồng thời k phần tử của tập n phần tử) bằng số cách sắp xếp n vật theo một hàng, trong đó có k vật giống nhau và $n - k$ vật còn lại cũng giống nhau. Ta có thể mở rộng kết quả này như sau.

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

$$\frac{n!}{n_1! n_2! \dots n_k!} \quad (1.5)$$

Công thức này có thể giải thích như sau:

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

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

Ví dụ 1.24: Cần sắp xếp 4 cuốn sách toán, 6 sách lý và 2 sách hóa khác nhau trên cùng một giá sách. Có bao nhiêu cách sắp xếp trong mỗi trường hợp sau:

- a. Các cuốn sách cùng môn học phải đứng cạnh nhau.
- b. Chỉ cần các sách toán đứng cạnh nhau.
- c. Nếu các cuốn sách trong mỗi môn học giống nhau thì có bao nhiêu cách sắp xếp.

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

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

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

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

c. Vì các cuốn sách cùng loại không phân biệt do đó có thể áp dụng công thức (1.5) và số cách sắp xếp là $\frac{12!}{4!6!2!}=13.860$.

1.2.3 Định nghĩa xác suất theo thống kê

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

Giả sử phép thử 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 C biến cố A xuất hiện $k_n(A)$ lần (gọi là **tần số** xuất hiện) thì tỉ số:

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

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

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

$$P(A) = \lim_{n \rightarrow \infty} f_n(A)$$

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

$$P(A) \approx f_n(A) \quad (1.6)$$

Ví dụ 1.25: Một công ty bảo hiểm muốn xác định xác suất để một thanh niên 25 tuổi sẽ bị chết trong năm tới, người ta theo dõi 100.000 thanh niên và thấy rằng có 798 người bị chết trong vòng 1 năm sau đó. Theo công thức (1.6) ta có thể tính xấp xỉ xác suất cần tìm bằng $\frac{798}{100.000} \approx 0,008$.

Ví dụ 1.26: Thống kê cho thấy tần suất sinh con trai xấp xỉ 0,513. Vậy xác suất để bé trai ra đời lớn hơn bé gái.

Nhận xét 1.4: 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 1.2: 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:

$$P(A) = \frac{\text{diện tích } A}{\text{diện tích } \Omega}. \quad (1.7)$$

Ví dụ 1.27: Hai người bạn X, Y 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 lần lượt là thời điểm X và Y đến điểm hẹn thì:

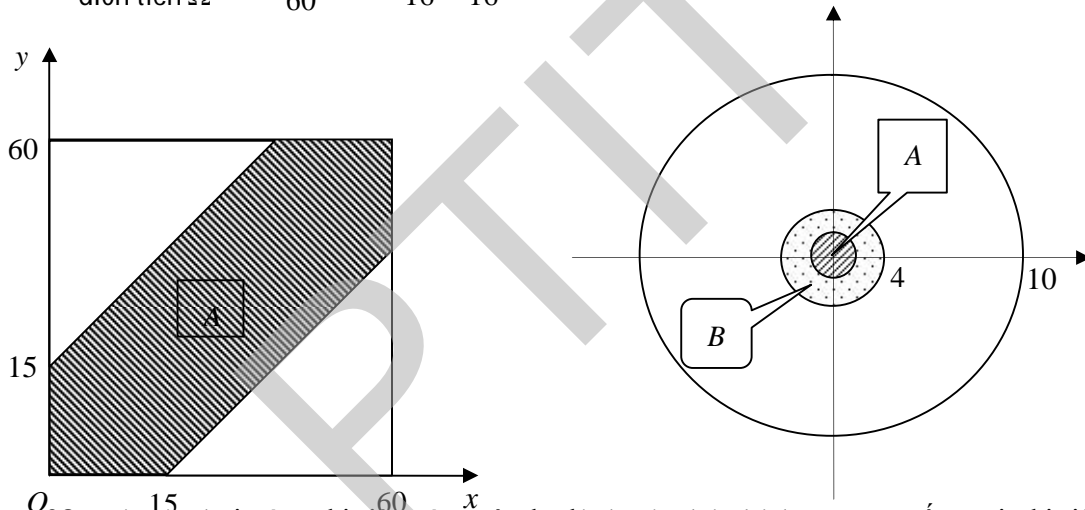
$$0 \leq x \leq 60, 0 \leq y \leq 60.$$

Vậy mỗi cặp thời điểm đến $(x; y)$ là một điểm của hình vuông $\Omega = [0, 60]^2$ (Hình 1.3).

Gọi A là biến cố hai người gặp nhau thì

$$A = \{(x; y) \in \Omega \mid |x - y| \leq 15\} = \{(x; y) \in \Omega \mid -15 + x \leq y \leq x + 15\}.$$

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



Ví dụ 1.28: Xét trò chơi ném phi tiêu vào một đĩa hình tròn bán kính 10 cm. Nếu mũi phi tiêu cắm vào ≤ 2 cm thì được giải nhất, nếu mũi phi tiêu ở trong khoảng 2 cm đến 4 cm thì được giải thứ hai. Giả sử mũi phi tiêu luôn cắm vào trong đĩa và đồng khả năng. Tính xác suất để người chơi được giải nhất, được giải nhì.

Giải: Gọi A, B lần lượt là biến cố người chơi nhận được giải nhất, giải nhì.

Có thể biểu diễn không gian mẫu Ω là hình tròn bán kính 10 (Hình 1.4). Khi đó biến cố A là hình tròn cùng tâm có bán kính 2 và biến cố B là hình vành khăn bán kính đường tròn trong bằng 2 và bán kính đường tròn ngoài bằng 4. Vậy xác suất để người chơi được giải nhất, được giải nhì lần lượt là:

$$P(A) = \frac{\text{diện tích } A}{\text{diện tích } \Omega} = \frac{\pi \cdot 2^2}{\pi \cdot 10^2} = \frac{2}{50},$$

$$P(B) = \frac{\pi \cdot (4^2 - 2^2)}{\pi \cdot 10^2} = \frac{7}{50}.$$

Ta đã có ba cách tiếp cận khác nhau về xác suất một biến cố, tất cả các định nghĩa này cùng có các tính chất sau.

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 :

$$0 \leq P(A) \leq 1. \quad (1.8)$$

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

$$P(\emptyset) = 0, P(\Omega) = 1 \quad (1.9)$$

1.2.5.2 Qui tắc cộng xác suất

A. Trường hợp xung khắc

Nếu A, B là hai biến cố xung khắc thì

$$P(A \cup B) = P(A) + P(B). \quad (1.10)$$

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

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

Từ công thức (1.9) và (1.11) ta có hệ quả: Nếu $\{A_1, A_2, \dots, A_n\}$ là một hệ đầy đủ thì

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

B. Trường hợp không xung khắc

- Nếu A, B là hai biến cố bất kỳ thì

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \quad (1.13)$$

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

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(B \cap C) - P(C \cap A) + P(A \cap B \cap C) \quad (1.14)$$

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

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

Ví dụ 1.29: 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.

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

Chương 1: Các khái niệm cơ bản về xác suất

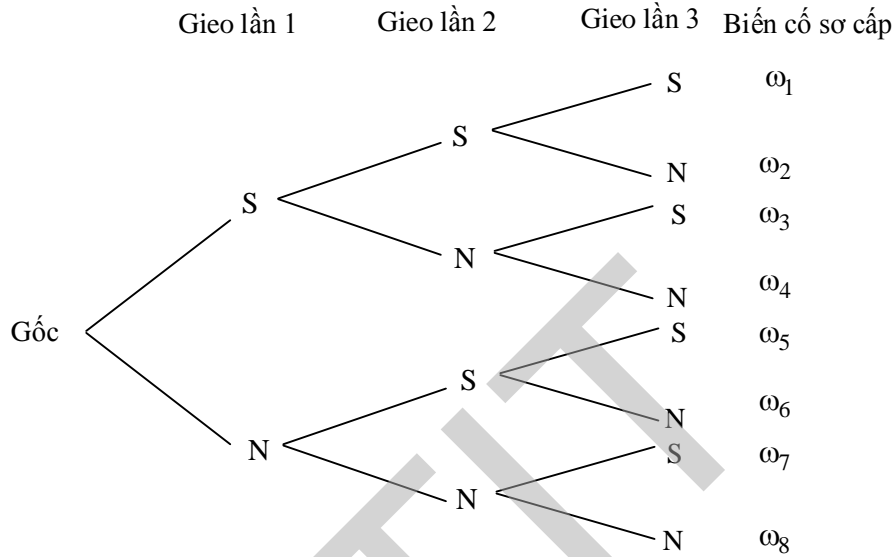
Gọi A là biến cố sản phẩm được chọn đạt tiêu chuẩn chất lượng, ta có $A = A_1 \cup A_2$.

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

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

Ví dụ 1.30: Gieo liên tiếp một đồng xu 3 lần.

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



Hình 1.5: Sơ đồ cây của phép thử gieo đồng xu liên tiếp 3 lần

Từ sơ đồ ta có $P(A) = P(B) = \frac{1}{2}$

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

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

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

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

Ví dụ 1.31: Xét hai biến cố A, B trong cùng một phép thử có xác suất $P(A) = 0,7$, $P(B) = 0,6$.

- Hai biến cố A, B có xung khắc không?
- Giả sử $A \cup B$ là biến cố chắc chắn, tìm $P(A \cap B)$.

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

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

Vậy hai biến cố A, B không xung khắc.

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

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

1.2.5.3 Quy tắc tính xác suất của biến cố đối

Áp dụng công thức (1.13) cho hệ đầy đủ $\{A, \bar{A}\}$ ta được quy tắc tính xác suất biến cố đối:

Với mọi biến cố A

$$P(\bar{A}) = 1 - P(A); P(A) = 1 - P(\bar{A}). \quad (1.16)$$

Ví dụ 1.32: Trong phòng có n người ($n < 365$).

- Tính xác suất có ít nhất hai người có cùng ngày sinh?
- Tính xác suất này khi $n = 10$.

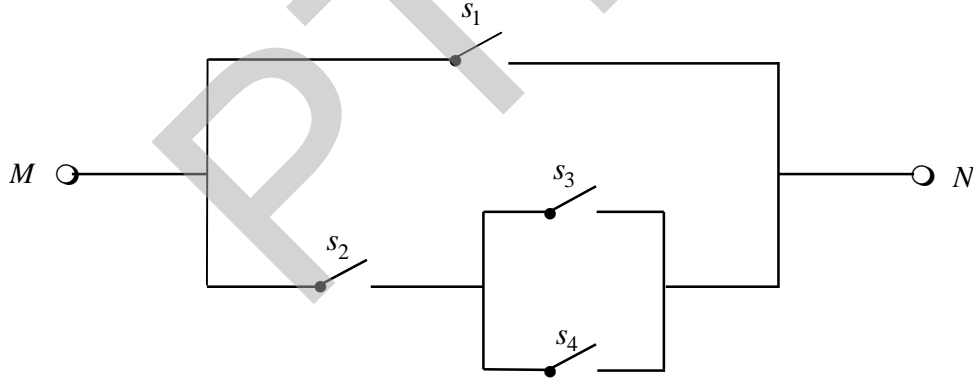
Giải : a. Gọi A là biến cố có ít nhất hai người trong phòng có cùng ngày sinh. Biến cố đối \bar{A} là biến cố mọi người không trùng ngày sinh. Ngày sinh của mỗi người đồng khả năng xảy ra tại 1 trong 365 ngày của năm.

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

b. Khi $n = 10$ thì

$$P(\bar{A}) = \frac{A_{365}^{10}}{365^{10}} = 0,883, P(A) = 1 - 0,883 = 0,117.$$

Ví dụ 1.33: Xét mạng gồm 4 chuyển mạch cho trong hình 1.6. Mỗi vị trí chuyển mạch đều có hai trạng thái đóng hoặc mở đồng khả năng. Tính xác suất đoạn mạch giữa M và N ở trạng thái đóng.



Hình 1.6

Giải: Đặt A_k là biến cố “chuyển mạch s_k ở trạng thái đóng”. Gọi A là biến cố “đoạn mạch giữa M và N ở trạng thái đóng”. Từ nhận xét 1.2 ta có

$$A = A_1 \cup [A_2 \cap (A_3 \cup A_4)] = A_1 \cup (A_2 \cap A_3) \cup (A_2 \cap A_4).$$

Áp dụng công thức (1.15) ta có

$$\begin{aligned} P(A) &= P[A_1 \cup (A_2 \cap A_3) \cup (A_2 \cap A_4)] = P(A_1) + P(A_2 \cap A_3) + P(A_2 \cap A_4) - P[A_1 \cap (A_2 \cap A_3)] \\ &\quad - P[A_1 \cap (A_2 \cap A_4)] - P[(A_2 \cap A_3) \cap (A_2 \cap A_4)] + P(A_1 \cap A_2 \cap A_3 \cap A_4). \end{aligned}$$

Chương 1: Các khái niệm cơ bản về xác suất

Mỗi chuyển mạch s_k có 2 trạng thái, vậy đoạn mạch giữa M và N có 16 trạng thái đồng khả năng. Nếu chuyển mạch ở trạng thái đóng ta ký hiệu 1 và ở trạng thái mở ta ký hiệu 0. Ta có thể liệt kê tất cả các trường hợp có thể và sự xuất hiện các biến cố theo bảng sau:

s_1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0
s_2	1	1	1	1	0	0	0	0	1	1	1	1	0	0	0	0
s_3	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0
s_4	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0

Do đó

$$P(A_1) = \frac{8}{16}, P(A_2 \cap A_3) = P(A_2 \cap A_4) = \frac{4}{16},$$

$$P(A_1 \cap A_2 \cap A_3) = P(A_1 \cap A_2 \cap A_4) = P(A_2 \cap A_3 \cap A_4) = \frac{2}{16}, P(A_1 \cap A_2 \cap A_3 \cap A_4) = \frac{1}{16}.$$

$$\text{Vậy } P(A) = \frac{8}{16} + 2 \cdot \frac{4}{16} - 3 \cdot \frac{2}{16} + \frac{1}{16} = \frac{11}{16} \approx 0,688.$$

Ví dụ 1.34: Giả sử phép thử C có không gian mẫu $\Omega = \{a, b, c, d\}$ với xác suất

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

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

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

Giải: $P(A) = P(a) + P(b) = 0,2 + 0,3 = 0,5$;

$$P(B) = P(b) + P(c) + P(d) = 0,3 + 0,4 + 0,1 = 0,8$$

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

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

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

1.2.6 Nguyên lý xác suất lớn, xác suất nhỏ

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

Khi tung đồng xu, ngoài khả năng mặt sấp hay mặt ngửa xuất hiện còn có khả năng đồng xu ở trạng thái đứng. Tuy nhiên khả năng thứ ba rất khó xảy ra, vì vậy thực tế ta luôn công nhận chỉ có hai khả năng mặt sấp và mặt ngửa xuất hiện.

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

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

Mức xác suất nhỏ này được gọi là *mức ý nghĩa*. Nếu α là mức ý nghĩa thì số $\beta = 1 - \alpha$ gọi là *độ tin cậy*. Khi dựa trên nguyên lý xác suất nhỏ ta khẳng định rằng: “Biến cố A có xác suất nhỏ (tức là $P(A) \leq \alpha$) 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 biến cố A xảy ra được gọi là xác suất của B với điều kiện A . Ký hiệu $P(B|A)$.

Tính chất

- Nếu $P(A) > 0$ thì

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

- Khi cố định A với $P(A) > 0$ thì xác suất có điều kiện $P(B|A)$ có tất cả các tính chất của xác suất thông thường (công thức (1.7)-(1.15)) đối với biến cố B .

Chẳng hạn:

$$P(\bar{B}|A) = 1 - P(B|A), \quad P(B_1 \cup B_2|A) = P(B_1|A) + P(B_2|A) - P(B_1 \cap B_2|A) \dots \quad (1.18)$$

Nhận xét 1.5: Ta có thể tính xác suất có điều kiện $P(B|A)$ bằng cách áp dụng công thức (1.17) hoặc tính trực tiếp.

Ví dụ 1.35: Gieo đồng thời hai con xúc xắc (6 mặt) cân đối. Tính xác suất để tổng số chấm xuất hiện trên hai con xúc xắc ≥ 10 biết rằng ít nhất một con đã ra chấm 5.

Giải: Gọi A là biến cố “ít nhất một con ra chấm 5”.

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

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

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

$$\text{Vậy } P(A \cap B) = \frac{3}{36} \Rightarrow P(B|A) = \frac{3}{36} / \frac{11}{36} = \frac{3}{11}.$$

Ta cũng có thể tính trực tiếp như sau.

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

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

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

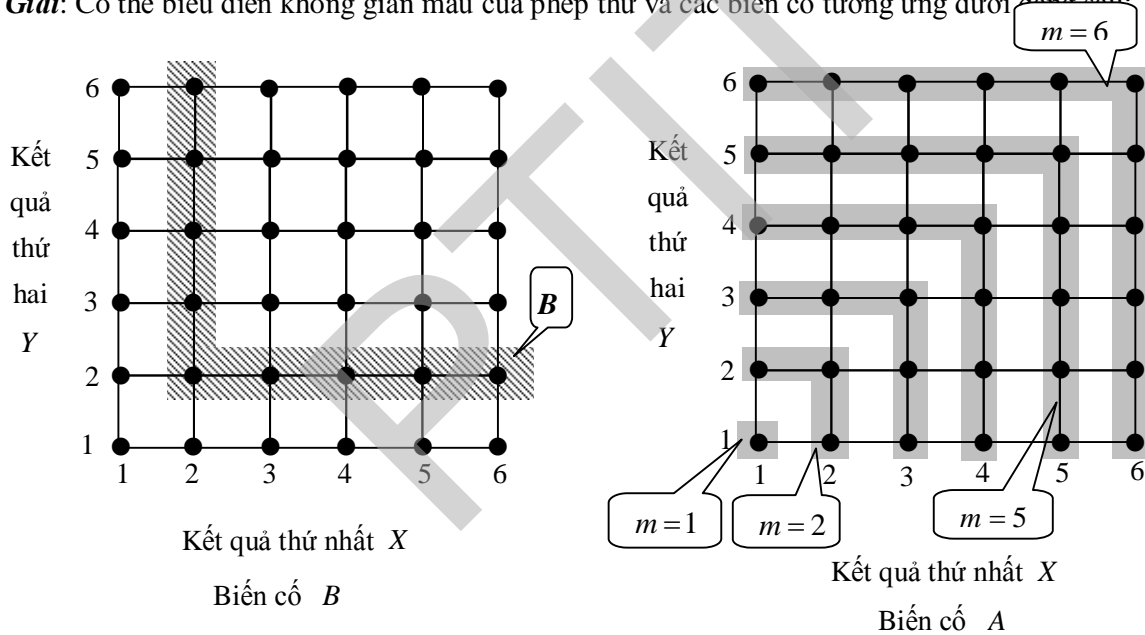
$$\text{Vậy } P(B|A) = \frac{3}{11}$$

Ví dụ 1.36: Xét phép thử gieo liên tiếp 2 lần con xúc xắc 4 mặt trong ví dụ 1.4. Gọi X, Y lần lượt là số chấm xuất hiện khi gieo lần thứ nhất và lần thứ hai. Ta tính xác suất có điều kiện $P(B|A)$ trong đó

$$A = \{\max(X, Y) = m\}, \quad B = \{\min(X, Y) = 2\}$$

Và m nhận một trong các giá trị 1, 2, 3, 4, 5, 6.

Giải: Có thể biểu diễn không gian mẫu của phép thử và các biến cố tương ứng dưới dạng sau:



Hình 1.7: Phép thử gieo liên tiếp 2 lần xúc xắc

Từ hình 1.7 ta được:

$$P(B) = \frac{9}{36}; \quad P(A \cap B) = \begin{cases} 8/36 & \text{nếu } m = 3, 4, 5, 6 \\ 1/36 & \text{nếu } m = 2 \\ 0 & \text{nếu } m = 1 \end{cases} \Rightarrow P(A|B) = \begin{cases} 8/9 & \text{nếu } m = 3, 4, 5, 6 \\ 1/9 & \text{nếu } m = 2 \\ 0 & \text{nếu } m = 1 \end{cases}.$$

Ví dụ 1.37: Có hai phân xưởng của nhà máy sản xuất cùng một loại sản phẩm. Phân xưởng I sản xuất được 1000 sản phẩm trong đó có 100 phế phẩm. Phân xưởng II sản xuất được 2000 sản phẩm trong đó có 150 phế phẩm. Lấy ngẫu nhiên một sản phẩm để kiểm tra và đó là phế phẩm. Tính xác suất phế phẩm này do phân xưởng thứ I sản xuất.

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

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

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

Áp dụng công thức (1.17) ta được $P(A|B) = \frac{1/30}{1/12} = \frac{2}{5} = 0,4$.

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

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

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

1.3.2 Quy tắc nhân xác suất

1.3.2.1 Trường hợp độc lập:

▪ Nếu A, B là hai biến cố độc lập thì xác suất của biến cố B không phụ thuộc vào A có xảy ra hay không (xem mục 1.1.3-f), nghĩa là $P(B|A) = P(B)$. Theo (1.17) ta có

$$P(A \cap B) = P(A)P(B). \quad (1.19)$$

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

$$P(A_1 \cap A_2 \cap \dots \cap A_n) = P(A_1)P(A_2) \dots P(A_n). \quad (1.20)$$

Thông thường tính độc lập của các biến cố được suy ra từ ý nghĩa thực tế. Chẳng hạn nếu A là biến cố xạ thủ thứ nhất bắn trúng mục tiêu và B là biến cố xạ thủ thứ hai bắn trúng mục tiêu (xem ví dụ 1.16) thì A, B là hai biến cố độc lập.

1.3.2.2 Trường hợp không độc lập:

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

$$P(A \cap B) = P(A)P(B|A) \quad (1.21)$$

▪ Với n biến cố bất kỳ A_1, A_2, \dots, A_n :

$$P(A_1 \cap A_2 \cap \dots \cap A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \dots P(A_n|A_1 \cap A_2 \cap \dots \cap A_{n-1}) \quad (1.22)$$

Ví dụ 1.38: 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_d, A_x lần lượt là biến cố bi được rút từ túi I là trắng, đỏ, xanh.

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

Các biến cố A_t, A_d, A_x xung khắc, B_t, B_d, B_x xung khắc;

Các biến cố A_t, A_d, A_x độc lập với các biến cố B_t, B_d, B_x .

Biến cố 2 bi được rút cùng màu là $(A_t \cap B_t) \cup (A_d \cap B_d) \cup (A_x \cap B_x)$

Vậy xác suất cần tìm:

$$P((A_t \cap B_t) \cup (A_d \cap B_d) \cup (A_x \cap B_x)) = P(A_t \cap B_t) + P(A_d \cap B_d) + P(A_x \cap B_x)$$

(theo công thức 1.10)

$$= P(A_t)P(B_t) + P(A_d)P(B_d) + P(A_x)P(B_x) \quad (\text{theo công thức 1.19})$$

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

Ví dụ 1.39: Một hộp đựng 100 con chip bán dẫn trong đó có 20 chip là phế phẩm. Lấy ngẫu nhiên không hoàn lại 2 chip bán dẫn ở trong hộp.

- Tính xác suất con chip lấy được lần đầu là phế phẩm.
- Tính xác suất con chip lấy được lần thứ hai là phế phẩm biết rằng con chip lấy lần đầu cũng là phế phẩm.
- Tính xác suất cả hai con chip lấy được đều là phế phẩm.

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

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

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

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

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

Ví dụ 1.40: Một tủ kho có một chùm chìa khóa gồm 9 chiếc, bề ngoài chúng giống hệt nhau nhưng trong đó chỉ có đúng 2 chiếc mở được kho. Anh ta thử ngẫu nhiên từng chìa (chìa nào không trúng thì bỏ ra). Tính xác suất để đến lần thử thứ ba mới mở được kho.

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

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

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

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

$$P(\overline{A_1} \cap \overline{A_2} \cap A_3) = P(\overline{A_1})P(\overline{A_2} | \overline{A_1})P(A_3 | \overline{A_1} \cap \overline{A_2}).$$

Từ giả thiết ta có thể tính được

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

Do đó

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

Ví dụ 1.41: Rút lần lượt ngẫu nhiên không hoàn lại 3 quân bài từ cỗ bài tứ lơ khơ. Tính xác suất trong các trường hợp sau:

- Cả 3 quân bài rút được không phải là quân bích.
- Lần thứ nhất rút được không phải quân bích và lần thứ hai rút được quân bích.
- Hai lần đầu rút được không phải quân bích và lần thứ ba rút được quân bích.

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

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

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

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

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

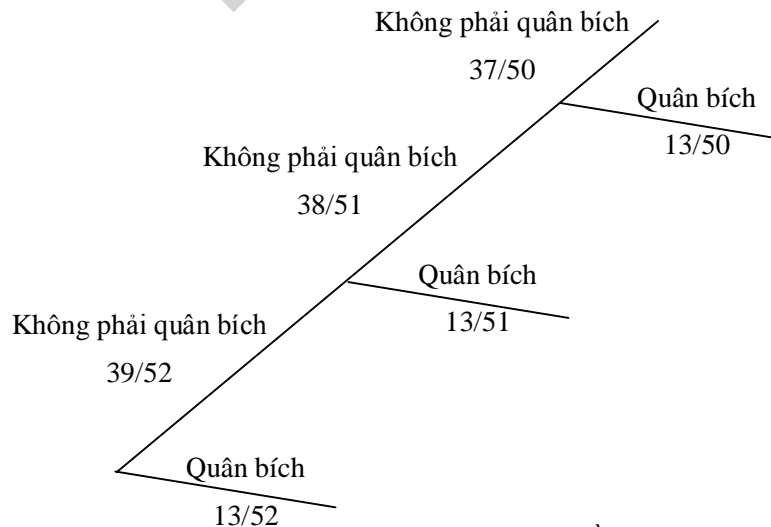
- Xác suất lần thứ nhất rút được không phải quân bích và lần thứ hai rút được quân bích là

$$P(A_1 \cap \overline{A_2}) = P(A_1)P(\overline{A_2} | A_1) = \frac{39}{52} \cdot \frac{13}{51}.$$

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

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

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



Hình 1.8: Sơ đồ cây rút liên tiếp 3 quân bài

Ví dụ 1.42: Cô thư ký xếp ngẫu nhiên n bức thư (với địa chỉ người nhận khác nhau) vào n phong bì đã có sẵn địa chỉ. Tính xác suất ít nhất một bức thư xếp đúng địa chỉ cần gửi.

Giải: Gọi A_1, A_2, \dots, A_n lần lượt là biến cố bức thư thứ nhất, hai, ..., thứ n xếp đúng địa chỉ cần gửi. Vậy biến cố ít nhất một bức thư xếp đúng địa chỉ cần gửi là $B = A_1 \cup A_2 \cup \dots \cup A_n$.

Áp dụng công thức (1.14):

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

Ta có $P(A_i) = \frac{1}{n}$ (vì có n kết quả có thể và 1 kết quả thuận lợi đối với biến cố)

$$\forall i < j : P(A_i \cap A_j) = P(A_i)P(A_j | A_i) = \frac{1}{n} \cdot \frac{1}{n-1} = \frac{(n-2)!}{n!}$$

$$\forall i < j < k : P(A_i \cap A_j \cap A_k) = P(A_i)P(A_j | A_i)P(A_k | A_i \cap A_j) = \frac{1}{n} \cdot \frac{1}{n-1} \cdot \frac{1}{n-2} = \frac{(n-3)!}{n!}$$

...

$$\begin{aligned} P(A_1 \cap A_2 \cap \dots \cap A_n) &= P(A_1)P(A_2 | A_1)P(A_3 | A_1 \cap A_2) \dots P(A_n | A_1 \cap A_2 \cap \dots \cap A_{n-1}) \\ &= \frac{1}{n} \cdot \frac{1}{n-1} \cdot \frac{1}{n-2} \dots \frac{1}{1} = \frac{1}{n!} \end{aligned}$$

Tổng $\sum_{1 \leq i < j \leq n} P(A_i \cap A_j)$ có $C_n^2 = \frac{n!}{2!(n-2)!}$ số hạng bằng nhau và bằng $\frac{(n-2)!}{n!}$,

$$\sum_{1 \leq i < j < k \leq n} P(A_i \cap A_j \cap A_k) \text{ có } C_n^3 = \frac{n!}{3!(n-3)!} \text{ số hạng bằng nhau và bằng } \frac{(n-3)!}{n!}, \dots$$

Thay vào công thức trên ta được

$$P(B) = \sum_{i=1}^n \frac{1}{n} - \sum_{i < j} \frac{(n-2)!}{n!} + \sum_{i < j < k} \frac{(n-3)!}{n!} - \dots + (-1)^{n-1} \frac{1}{n!} = 1 - \frac{1}{2!} + \frac{1}{3!} - \dots + (-1)^{n-1} \frac{1}{n!}.$$

Từ công thức khai triển Mc. Laurin hàm mũ ta có

$$e^x = 1 + x + \frac{x^2}{2!} + \dots + \frac{x^n}{n!} + \dots$$

$$\Rightarrow e^{-1} = 1 - 1 + \frac{1}{2!} + \dots + \frac{(-1)^n}{n!} + \dots = 1 - \left(1 - \frac{1}{2!} + \frac{1}{3!} + \dots + (-1)^{n-1} \frac{1}{n!} + \dots\right)$$

Chuỗi số trên hội tụ khá nhanh, vì vậy khi n đủ lớn ($n > 10$) ta có giá trị xấp xỉ

$$P(B) \approx 1 - e^{-1} = 0,6321.$$

1.3.3 Công thức xác suất đầy đủ

Định lý 1.2: Giả sử $\{A_1, A_2, \dots, A_n\}$ là một hệ đầy đủ các biến cố. Khi đó, với mọi biến cố B của cùng một phép thử ta có

$$P(B) = \sum_{i=1}^n P(A_i \cap B) = \sum_{i=1}^n P(A_i)P(B|A_i) \quad (1.23)$$

Ví dụ 1.43: Một người tham gia thi đấu cờ vua với một nhóm các đấu thủ chia làm ba loại: loại I chiếm $1/2$ số đấu thủ, loại II chiếm $1/4$ số đấu thủ và loại III chiếm $1/4$ số đấu thủ còn lại. Xác suất anh ta thắng đấu thủ loại I là 0,3, thắng đấu thủ loại II là 0,4 và thắng đấu thủ loại III là 0,5. Anh ta thi đấu ngẫu nhiên với một trong các đấu thủ loại I, loại II hoặc loại III. Tính xác suất anh ta thắng cuộc.

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

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

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

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

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

$$\begin{aligned} P(B) &= P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + P(A_3)P(B|A_3) \\ &= 0,5 \cdot 0,3 + 0,25 \cdot 0,4 + 0,25 \cdot 0,5 = 0,375. \end{aligned}$$

Ví dụ 1.44: 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.

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

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

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

Từ đó ta tính được các xác suất có điều kiện

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

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

Ví dụ 1.45: Gieo xúc xắc. Nếu mặt 1 chấm hoặc 2 chấm xuất hiện ta gieo tiếp lần nữa và ngược lại. Tính xác suất để tổng số chấm xuất hiện ít nhất là 5.

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

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

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

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

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

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

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

Áp dụng công thức xác suất đầy đủ ta được

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

1.3.4 Công thức Bayes

Từ công thức (1.21) ta có

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

Dựa vào đẳng thức này ta được kết quả sau.

Định lý 1.3: Giả sử $\{A_1, A_2, \dots, A_n\}$ là một hệ đầy đủ các biến cố. Khi đó, với mọi biến cố B của cùng một phép thử và $P(B) > 0$ ta có:

$$P(A_k | B) = \frac{P(A_k)P(B | A_k)}{P(B)} = \frac{P(A_k)P(B | A_k)}{\sum_{i=1}^n P(A_i)P(B | A_i)}; \quad k = 1, 2, \dots, n. \quad (1.24)$$

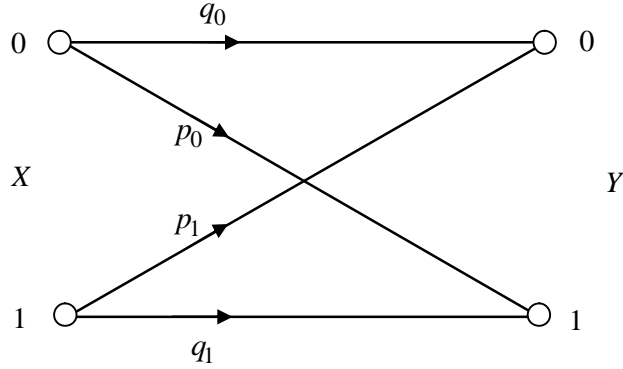
Nhận xét 1.7: Trong thực tế các xác suất $\{P(A_1), P(A_2), \dots, P(A_n)\}$ đã biết và được gọi là các *xác suất tiên nghiệm*. Sau khi quan sát biết được biến cố B xảy ra, các xác suất của A_k được tính trên thông tin này (xác suất có điều kiện $P(A_k | B)$) được gọi là *xác suất hậu nghiệm*. Vì vậy công thức Bayes còn được gọi là công thức xác suất hậu nghiệm.

Ví dụ 1.46: Xét kênh viễn thông nhị phân được biểu diễn như sơ đồ Hình 1.9.

Đầu vào của kênh ký hiệu là X và giả thiết rằng chỉ có hai trạng thái 0 và 1, tương tự đầu ra ký hiệu là Y và cũng chỉ có hai trạng thái 0 và 1. Do bị nhiễu kênh nên đầu vào 0 có thể chuyển thành đầu ra là 1 và ngược lại.

Gọi là X_0 biến cố “ X có trạng thái 0” và X_1 là biến cố “ X có trạng thái 1”.

Gọi là Y_0 biến cố “đầu ra Y có trạng thái 0” và là Y_1 biến cố “đầu ra Y có trạng thái 1”.
 Khi đó $\{X_0, X_1\}$ và $\{Y_0, Y_1\}$ là hai hệ đầy đủ.



Hình 1.9

Kênh được đặc trưng bởi các xác suất chuyển p_0, q_0, p_1 và q_1 , trong đó

$$p_0 = P(Y_1|X_0) \text{ và } p_1 = P(Y_0|X_1)$$

$$q_0 = P(Y_0|X_0) \text{ và } q_1 = P(Y_1|X_1)$$

$$p_0 + q_0 = 1 = p_1 + q_1.$$

p_0, p_1 được gọi là xác suất lỗi

Giả sử $P(X_0) = 0,5$ (hai tín hiệu 0, 1 đầu vào đồng khả năng), $p_0 = 0,1$ và $p_1 = 0,2$.

- Tìm xác suất đầu ra của kênh là 0 và xác suất đầu ra của kênh là 1.
- Giả sử đầu ra của kênh nhận được là 0. Tìm xác suất nhận đúng tín hiệu đầu vào.
- Tính xác suất lỗi P_e

Giải: $P(X_1) = 1 - P(X_0) = 0,5$; $q_0 = 1 - p_0 = 1 - 0,1 = 0,9$; $q_1 = 1 - p_1 = 1 - 0,2 = 0,8$

- Áp dụng công thức xác suất đầy đủ với hệ đầy đủ $\{X_0, X_1\}$ ta được:

$$P(Y_0) = P(X_0)P(Y_0|X_0) + P(X_1)P(Y_0|X_1) = 0,5 \times 0,9 + 0,5 \times 0,2 = 0,55$$

$$P(Y_1) = P(X_0)P(Y_1|X_0) + P(X_1)P(Y_1|X_1) = 0,5 \times 0,1 + 0,5 \times 0,8 = 0,45.$$

- Áp dụng công thức Bayes ta có

$$P(X_0|Y_0) = \frac{P(X_0)P(Y_0|X_0)}{P(Y_0)} = \frac{0,5 \times 0,9}{0,55} = 0,818.$$

- Xác suất lỗi là xác suất của biến cố đầu vào 0 và đầu ra 1 hoặc biến cố đầu vào 1 và đầu ra 0. Vậy

$$P_e = P((X_0 \cap Y_1) \cup (X_1 \cap Y_0)) = P(X_0)P(Y_1|X_0) + P(X_1)P(Y_0|X_1) = 0,5 \times 0,1 + 0,5 \times 0,2 = 0,15$$

Ví dụ 1.47: Một nhà máy có ba phân xưởng I, II, III cùng sản xuất ra một loại sản phẩm. Phân xưởng I, II, III sản xuất tương ứng 36%, 34%, 30% sản lượng của nhà máy, với tỷ lệ phế phẩm tương ứng là 0,12; 0,1; 0,08.

- a. Tìm tỷ lệ phế phẩm chung của nhà máy.
- b. Lấy ngẫu nhiên một sản phẩm kiểm tra và đó là phế phẩm. Tính xác suất để phế phẩm đó là do phân xưởng I sản xuất.

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

Gọi A_1, A_2, A_3 lần lượt là biến cố sản phẩm lấy ra kiểm tra do phân xưởng I, II, III sản xuất. Theo giả thiết ta có: hệ 3 biến cố $\{A_1, A_2, A_3\}$ đầy đủ (xem ví dụ 1.8).

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

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

- a. Xác suất của biến cố B cũng là tỉ lệ phế phẩm chung của nhà máy. Áp dụng công thức xác suất đầy đủ (1.23) ta có

$$P(B) = P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + P(A_3)P(B|A_3) = 0,1012$$

- b. Áp dụng công thức Bayes ta được

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

Ví dụ 1.48: 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:

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

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

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

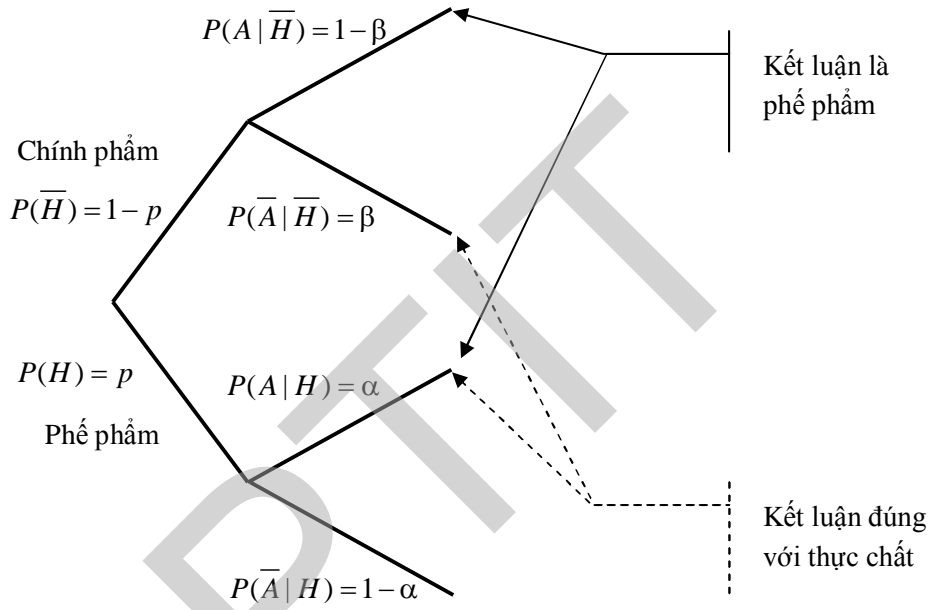
b. Biến cố sản phẩm kiểm tra được kết luận đạt chất lượng nhưng là phế phẩm là biến cố H với điều kiện \bar{A} . Áp dụng công thức Bayes ta được

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

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

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

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



Hình 1.10: Sơ đồ cây xác suất đầy đủ

Ví dụ 1.49: Giả sử hai biến cố A , B có xác suất $P(A) = 2/5$, $P(B) = 1/3$ và $P(AB) = 1/6$.

Hãy tính

a. $P(A | B)$

b. $P(A \cup B)$

c. $P(A \cap \bar{B})$

d. $P(\bar{B} | \bar{A})$.

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

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

c. $P(B | A) = \frac{P(B \cap A)}{P(A)} = \frac{1/6}{2/5} = \frac{5}{12} \Rightarrow P(A \cap \bar{B}) = P(A)P(\bar{B} | A) = \frac{2}{5} \cdot \left(1 - \frac{5}{12}\right) = \frac{7}{30}$.

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

1.4 DÃY PHÉP THỬ BERNOULLI

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

p là xác suất thành công trong mỗi lần thử.

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

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

Đặt $P_n(k; p) = P(H_k)$.

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

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

Chứng minh: H_k là tổng của C_n^k các biến cố xung khắc từng đôi nhận được bằng cách hoán vị các chữ A và \bar{A} trong biến cố tích sau (xem nhận xét 1.2):

$$\underbrace{A \cap \dots \cap A}_{k \text{ lần}} \cap \underbrace{\bar{A} \cap \dots \cap \bar{A}}_{n-k \text{ lần}}$$

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

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

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

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

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

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

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

$$(n+1)p - 1 \leq m \leq (n+1)p \quad (1.27)$$

Như vậy, $P_{\max} = P_n(m; p)$

- 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 $m = (n+1)p$

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

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

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

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

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

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

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

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

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

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

Ví dụ 1.50: Bắn 7 viên đạn vào bia. Xác suất trúng đích của mỗi viên là 0,6. Tìm xác suất trong các trường hợp sau:

- Có đúng 3 viên trúng bia.
- Có ít nhất 6 viên trúng bia.
- Có ít nhất 1 viên trúng bia.
- Tìm số viên đạn trúng bia có khả năng lớn nhất.

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

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

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

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

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

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

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

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

Ví dụ 1.49: Tín hiệu thông tin được phát đi 3 lần độc lập nhau. Xác suất thu được mỗi lần là 0.4.

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

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

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

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

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

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

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

TÓM TẮT

Trong chương này ta xét đến phép thử, biến cố và xác suất của biến cố.

Có thể xem biến cố của một phép thử là tập con của không gian mẫu của phép thử này. Do đó ta có các quan hệ giữa các biến cố tương tự với các phép toán giữa các tập hợp, đó là phép toán hợp, giao và lấy phần bù của tập hợp.

Để tính xác suất của biến cố trường hợp đồng khả năng ta sử dụng phương pháp xác suất cổ điển (công thức 1.1a) và các quy tắc đếm.

Trường hợp đã biết xác suất các biến cố nào đó và cần tính xác suất của các biến cố mới có liên quan ta sử dụng các quy tắc tính xác suất, trong đó có các công thức sau:

- Công thức cộng xác suất (1.10-1.13)
- Công thức xác suất biến cố đối (1.14)
- Công thức nhân xác suất (1.17-1.20)
- Công thức xác suất đầy đủ, công thức Bayes (1.23-1.24)
- Công thức xác suất của dãy phép thử Bernoulli (1.25).

CÂU HỎI ÔN TẬP VÀ BÀI TẬP CHƯƠNG 1

- 1.1 Ta có thể có hai không gian mẫu Ω các biến cố sơ cấp cho cùng một phép thử C ?

Đúng ☐ Sai ☐.

- 1.2 Các biến cố A và $\overline{A \cup B}$ 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 $\{A_1, \dots, A_n\}$ thì các biến cố B và A_1, \dots, 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 Một hộp có 10 bi màu đỏ, 30 trắng, 20 xanh và 15 vàng. Lấy ngẫu nhiên từ hộp 1 bi, tính xác suất bi lấy được trong các trường hợp sau:

- a. màu vàng hoặc đỏ
- b. không phải màu và không phải màu xanh
- c. màu trắng
- d. màu đỏ hoặc trắng hoặc xanh

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

- a. cả hai cùng màu xanh
- b. cả hai cùng màu đỏ
- c. 1 bi màu đỏ và 1 bi màu xanh

1.14 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.

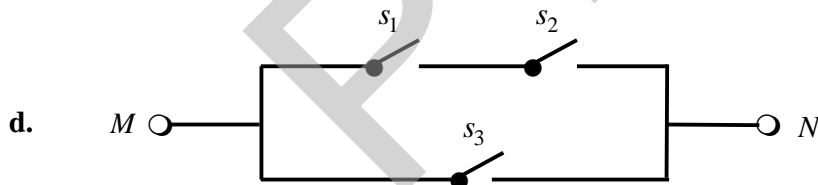
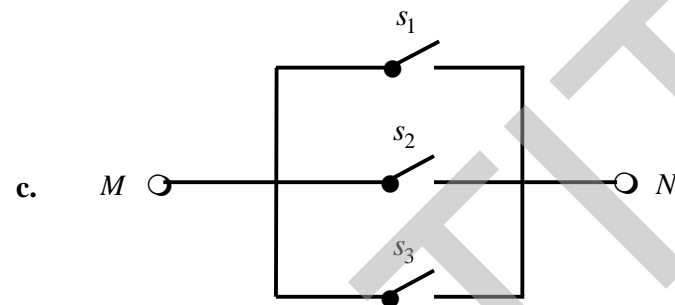
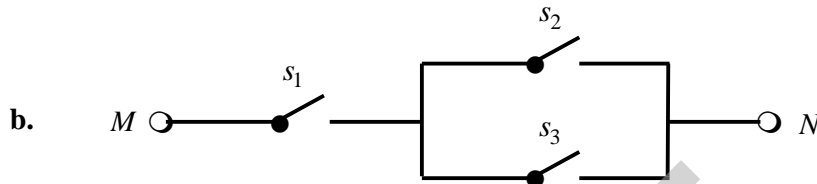
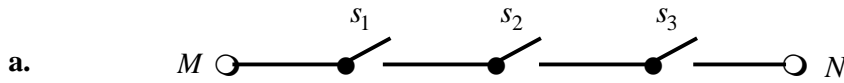
1.15 Một người gọi điện thoại cho bạn nhưng lại quên mất 3 chữ số cuối và chỉ nhớ rằng chúng khác nhau. Tìm xác suất để người đó quay số một lần được đúng số điện thoại của bạn.

1.16 Ta kiểm tra theo thứ tự một lô hàng có 10 sản phẩm. Mỗi sản phẩm thuộc một trong hai loại: Tốt hoặc Xấu. Ký hiệu A_k ($k = 1, \dots, 10$) 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.

1.17 Xét các mạng với ba vị trí chuyển mạch tương ứng trong các sơ đồ sau. Gọi A_k là biến cố chuyển mạch s_k ($k = 1, 2, 3$) ở trạng thái đóng. Gọi A là biến cố mạng ở trạng thái đóng từ M đến N . Hãy biểu diễn A theo các biến cố A_k ($k = 1, 2, 3$).



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

- Chỉ có một người bắn trúng mục tiêu.
- Có người bắn trúng mục tiêu.
- Cả hai người bắn trượt.

1.19 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.

1.20 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.

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

- Có 4 quân Át
- 4 quân Át và 1 quân K

- c. 3 quân 10 và 2 quân J
e. Đồng chất
g. Có ít nhất 1 quân Át.
- d. 10, J, Q, K và Át
f. 10, J, Q, K, Át và đồng chất
- 1.22** Tính xác suất rút lần lượt được 3 quân Át từ cỗ bài tứ lơ khơ trong hai trường hợp sau:
a. Rút có hoàn lại
b. Rút không hoàn lại.
- 1.23** 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.
- 1.24** 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.
- 1.25** Hai biến cố A , B có xác suất $P(A) = 0,3$, $P(A \cup B) = 0,65$. Giả sử A , B độc lập nhưng không xung khắc. Tính $P(B)$.
- 1.26** Giả sử hai biến cố A , B có xác suất $P(A) = 1/2$, $P(B) = 1/3$ và $P(A \cap B) = 1/4$. Hãy tính
a. $P(A|B)$
b. $P(B|A)$
c. $P(A \cup B)$
d. $P(A \cap \bar{B})$
e. $P(\bar{A} \cap B)$
f. $P(\bar{B}|\bar{A})$
g. $P(\bar{A}|\bar{B})$
h. $P(\bar{A} \cap \bar{B})$.
- 1.27** Chọn ngẫu nhiên lần lượt không hoàn lại 2 số từ các số $\{0,1,\dots,9\}$. Tính xác suất số thứ hai chọn được là số 4.
- 1.28** Một nhà máy ô tô có ba phân xưởng I, II, III cùng sản xuất ra một loại pít-tông. Phân xưởng I, II, III sản xuất tương ứng 36%, 34%, 30% sản lượng của nhà máy, với tỷ lệ phế phẩm tương ứng là 0,12; 0,1; 0,08.
a. Tìm tỷ lệ phế phẩm chung của nhà máy.
b. Lấy ngẫu nhiên một sản phẩm kiểm tra và được sản phẩm là phế phẩm. Tính xác suất để phế phẩm đó là do phân xưởng I, II, III sản xuất.
- 1.29** 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.
- 1.30** 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.
- 1.31** Hộp thứ nhất có 2 viên bi màu trắng và 3 bi đen; hộp thứ hai có 4 trắng và 1 đen; hộp thứ ba có 3 trắng và 4 đen. Chọn ngẫu nhiên một hộp trong ba hộp và lấy ngẫu nhiên một viên bi. Biết rằng viên bi được lấy ra màu trắng, tính xác suất viên bi này được lấy từ hộp thứ nhất.

1.32 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.

1.33 Chứng minh rằng nếu $P(A|B) > P(A)$ thì $P(B|A) > P(B)$.

CHƯƠNG 2: BIẾN NGẪU NHIÊN VÀ CÁC ĐẶC TRƯNG CỦA CHÚNG

Trong chương này ta khảo sát các biến cố của đại lượng nhận các giá trị nào đó, khi các giá trị này thay đổi ta được 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 với xác suất bao nhiêu. Các biến ngẫu nhiên trong các phép thử khác nhau có thể có các phân bố xác suất như nhau, nghĩa là cùng quy luật phân bố xác suất. Phân bố xác suất của biến ngẫu nhiên X có thể được khảo sát thông qua hàm phân bố xác suất $F_X(x) = P\{X \leq x\}$. Khi ta biết qui luật phân bố xác suất của một biến ngẫu nhiên thì ta có thể tính các xác suất liên quan đến biến ngẫu nhiên này.

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

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

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

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

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

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

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

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

2.1 ĐỊNH NGHĨA VÀ PHÂN LOẠI BIẾN NGẪU NHIÊN

Ta xét ví dụ đơn giản sau về biến ngẫu nhiên:

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

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

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

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

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

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

2.1.1 Định nghĩa biến ngẫu nhiên

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

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

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

Ví dụ 2.1: Gieo đồng thời hai con xúc xắc. Ký hiệu A_k , $k = 2, 3, \dots, 12$ là biến cố tổng số chấm xuất hiện của hai con xúc xắc là k .

Nếu gọi X là tổng số chấm xuất hiện khi gieo hai con xúc xắc thì X là một biến ngẫu nhiên có miền giá trị $R_X = \{2, 3, \dots, 12\}$ và $\{X = k\} = A_k$ với $k = 2, 3, \dots, 12$.

Ví dụ 2.2: Các đại lượng sau là biến ngẫu nhiên

- Tuổi thọ của một thiết bị đang hoạt động.
- Số khách hàng vào một điểm phục vụ trong 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 ...

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

$$\{X \leq x\}, \{Y \leq y\}.$$

Trong chương 3 ta sẽ đưa ra các tiêu chuẩn để nhận biết tính chất độc lập của hai biến ngẫu nhiên.

2.1.2 Hàm phân bố xác suất

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

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

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

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

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

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

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

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

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

$$F_X(x) = P\{X \leq x\}; \quad -\infty < x < \infty \quad (2.1)$$

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

Hàm phân bố có các tính chất sau:

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

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

$$\text{Với mọi } a, b \in \mathbb{R}: a < b \Rightarrow F_X(a) \leq F_X(b)$$

Với mọi $a \in \mathbb{R}$, ta có

$$F_X(a^+) = F_X(a) \text{ với } F_X(a^+) = \lim_{x \rightarrow a, x > a} F_X(x) \quad (2.3)$$

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

c. $F_X(-\infty) = \lim_{x \rightarrow -\infty} F_X(x) = 0$; $F_X(+\infty) = \lim_{x \rightarrow +\infty} F_X(x) = 1$, (2.4)

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

e. $P\{X > a\} = 1 - F_X(a)$; $P\{X < a\} = F_X(a^-)$ với $F_X(a^-) = \lim_{x \rightarrow a, x < a} F_X(x)$ (2.6)

Nhận xét 2.1: Một số tài liệu coi

$$G_X(x) = P\{X < x\}; \quad -\infty < x < \infty$$

là hàm phân bố của biến ngẫu nhiên X .

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

$$P\{a \leq X < b\} = G_X(b) - G_X(a).$$

Mỗi cách định nghĩa có thuận lợi riêng, tuy nhiên trong giáo trình này ta sử dụng hàm phân bố $F_X(x)$ theo công thức (2.1).

Ví dụ 2.3: Một nguồn thông tin sinh ra các ký hiệu ngẫu nhiên từ bốn ký tự $\{a, b, c, d\}$ với xác suất $P(a) = 1/2$, $P(b) = 1/4$ và $P(c) = P(d) = 1/8$. Mã hóa các ký hiệu này theo các mã nhị phân sau

a	0
b	10
c	110
d	111

Đặt X là biến ngẫu nhiên ký hiệu độ dài của mã, đó là số các bit.

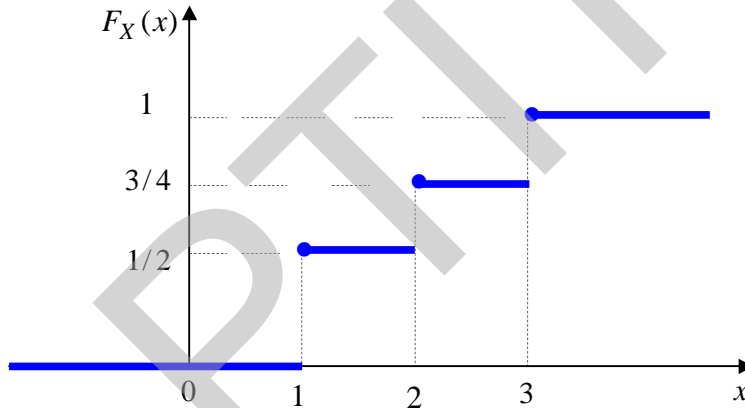
- Tìm miền giá trị của X .
- Giả sử các ký hiệu được sinh độc lập. Tính các xác suất $P\{X=1\}$, $P\{X=2\}$ và $P\{X=3\}$.
- Tìm hàm phân bố $F_X(x)$ và vẽ đồ thị của $F_X(x)$.

Giải: a. Miền giá trị $R_X = \{1, 2, 3\}$.

- $P\{X=1\} = P(a) = 1/2$, $P\{X=2\} = P(b) = 1/4$,
 $P\{X=3\} = P\{c, d\} = P(c) + P(d) = 1/4$.

- $$F_X(x) = \begin{cases} 0 & x < 1 \\ 1/2 & 1 \leq x < 2 \\ 3/4 & 2 \leq x < 3 \\ 1 & x \geq 3 \end{cases}$$

Đồ thị của $F_X(x)$ có dạng bậc thang.



Hình 2.1

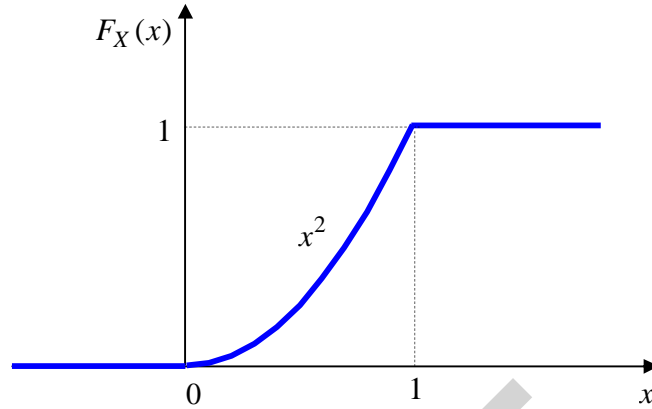
Ví dụ 2.4: Xét phép thử ném phi tiêu vào một đĩa tròn có bán kính bằng 1 (xem ví dụ 1.18). Ký hiệu X là biến ngẫu nhiên đo khoảng cách từ điểm mũi phi tiêu cắm vào đĩa đến tâm của đĩa. Giả sử mũi phi tiêu luôn cắm vào đĩa và đồng khả năng tại mọi điểm của đĩa.

- Tìm miền giá trị của X .
- Tìm hàm phân bố $F_X(x)$ và vẽ đồ thị của $F_X(x)$.

Giải: a. Miền giá trị của X là $R_X = \{x \in \mathbb{R} | 0 \leq x < 1\}$.

- $P\{X \leq x\} = \frac{\pi \cdot x^2}{\pi \cdot 1^2} = x^2$ nếu $0 \leq x \leq 1$.

Vậy hàm phân bố và đồ thị $F_X(x) = \begin{cases} 0 & x \leq 0 \\ x^2 & 0 < x \leq 1 \\ 1 & x > 1 \end{cases}$



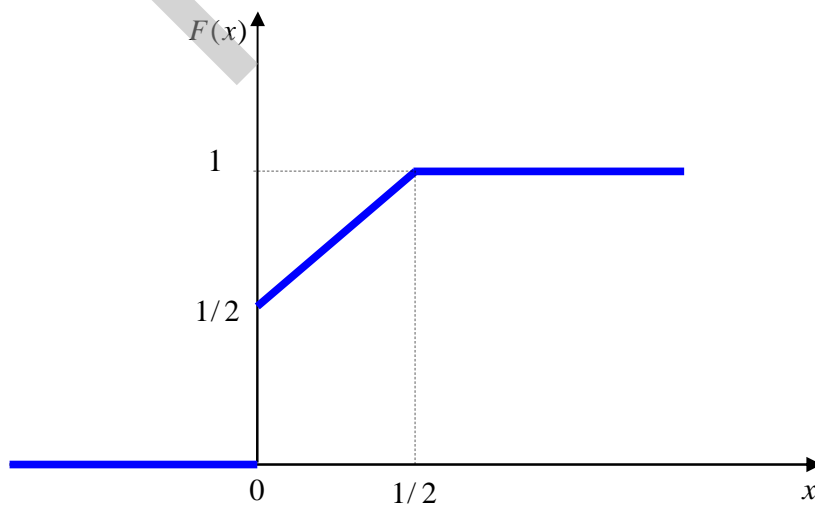
Hình 2.2

Ví dụ 2.5: Xét hàm số $F(x) = \begin{cases} 0 & x < 0 \\ x + 1/2 & 0 \leq x < 1/2 \\ 1 & x \geq 1/2 \end{cases}$

- Vẽ đồ thị của $F(x)$ và chứng minh rằng $F(x)$ thỏa mãn các tính chất (2.2)-(2.4).
- Nếu X là biến ngẫu nhiên có hàm phân bố $F(x)$. Tính

$$P\{X \leq 1/4\}, P\{0 < X \leq 1/4\}, P\{X = 0\} \text{ và } P\{0 \leq X \leq 1/4\}.$$

Giải: a. Từ đồ thị của $F(x)$ (hình 2.3) có thể suy ra các tính chất (2.2)-(2.4) của hàm phân bố.



Hình 2.3

b.
$$P\left\{X \leq \frac{1}{4}\right\} = F\left(\frac{1}{4}\right) = \frac{1}{4} + \frac{1}{2} = \frac{3}{4};$$

$$P\left\{0 < X \leq \frac{1}{4}\right\} = F\left(\frac{1}{4}\right) - F(0) = \frac{3}{4} - \frac{1}{2} = \frac{1}{4}$$

Áp dụng công thức (2.6) ta được

$$P\{X = 0\} = P\{X \leq 0\} - P\{X < 0\} = F(0) - F(0^-) = \frac{1}{2} - 0 = \frac{1}{2}$$

$$P\{0 \leq X \leq 1/4\} = P\{X = 0\} + P\{0 < X \leq 1/4\} = \frac{1}{2} + \frac{3}{4} - \frac{1}{2} = \frac{3}{4}.$$

2.1.3 Phân loại

Hàm phân bố của biến ngẫu nhiên là hàm số không giảm và bị chặn do đó hàm phân bố chỉ có thể gián đoạn tại một số không quá đếm được các điểm. Dựa vào dạng hàm phân bố của biến ngẫu nhiên ta có thể chia biến ngẫu nhiên thành ba loại: Biến ngẫu nhiên rời rạc, biến ngẫu nhiên liên tục và loại hỗn hợp.

❖ *Biến ngẫu nhiên rời rạc*

Biến ngẫu nhiên X là rời rạc nếu miền giá trị gồm một số hữu hạn hoặc vô hạn đếm được các giá trị, nghĩa là có thể liệt kê các giá trị của miền giá trị R_X thành một dãy có dạng x_1, x_2, \dots . Do đó hàm phân bố của biến ngẫu nhiên rời rạc có đồ thị dạng hình thang (Hình 2.1, ví dụ 2.3).

❖ *Biến ngẫu nhiên liên tục*

Biến ngẫu nhiên X là biến ngẫu nhiên liên tục nếu các giá trị của nó có thể lấp đầy một hoặc một số các khoảng hữu hạn hoặc vô hạn (như vậy miền giá trị R_X là một khoảng hoặc hợp của một số khoảng hữu hạn hoặc vô hạn) và xác suất X nhận giá trị tại từng điểm đều bằng 0 (nghĩa là $P\{X = a\} = 0$ với mọi a). Do đó hàm phân bố là hàm số liên tục (Hình 2.2, ví dụ 2.4).

Biến ngẫu nhiên có hàm phân bố ở ví dụ 2.5 thuộc loại hỗn hợp, không rời rạc vì miền giá trị chứa khoảng $[0; 1/2]$ và không liên tục vì $P\{X = 0\} = 1/2 \neq 0$.

Tuy nhiên chúng ta thường chỉ gặp hai loại biến ngẫu nhiên rời rạc hoặc liên tục. Tập bài giảng này cũng chỉ xét hai loại đó.

Ví dụ 2.6:

- Gọi X là số chấm 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, ...
- Số cuộc gọi đến một tổng đài là biến ngẫu nhiên rời rạc nhận các giá trị 0, 1, 2, ...
- Sai số Y khi đo lường một đại lượng vật lý nào đó là biến ngẫu nhiên liên tục nhận giá trị trong một khoảng.

2.2 BIẾN NGẪU NHIÊN RỜI RẠC

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

Biến ngẫu nhiên rời rạc có miền giá trị là một tập hữu hạn hoặc vô hạn đếm được. Phân bố xác suất chỉ tập trung tại các giá trị này.

Giả sử biến ngẫu nhiên rời rạc X có đồ thị của $F_X(x)$ là hàm bậc thang có bước nhảy tại x_1, x_2, \dots thì

$$F_X(x_k) - F_X(x_k^-) = P\{X \leq x_k\} - P\{X < x_k\} = P\{X = x_k\} \quad (2.7)$$

Đặt

$$p_X(x) = P\{X = x\} \quad (2.8)$$

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

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

$$1. \quad p_X(x_k) > 0, \text{ với mọi } x_k \in R_X \quad (2.9)$$

$$2. \quad \sum_{x_k \in R_X} p_X(x_k) = 1 \quad (2.10)$$

$$3. \quad p_X(x) = 0 \text{ với mọi } x \notin R_X \quad (2.11)$$

Hàm phân bố của X có thể xác định từ hàm khối lượng xác suất

$$F_X(x) = P\{X \leq x\} = \sum_{x_k \leq x; x_k \in R_X} p_X(x_k) \quad (2.12)$$

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

$$F_X(x) = \begin{cases} 0 & \text{nếu } x < x_1 \\ p_X(x_1) + \dots + p_X(x_{k-1}) & \text{nếu } x_{k-1} \leq x < x_k, \forall k > 1 \end{cases} \quad (2.13)$$

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

$$\{X = x_1\}, \{X = x_2\}, \dots, \{X = x_n\} \quad (2.14)$$

lập thành hệ đầy đủ các biến cố.

Hàm phân bố xác suất có dạng:

$$F_X(x) = \begin{cases} 0 & \text{nếu } x < x_1 \\ p_X(x_1) + \dots + p_X(x_{k-1}) & \text{nếu } x_{k-1} \leq x < x_k \\ 1 & \text{nếu } x \geq x_n \end{cases} \quad (2.15)$$

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

Bảng phân bố xác suất của X có dạng sau:

X	x_1	x_2	...
P	$p_X(x_1)$	$p_X(x_2)$...

Ví dụ 2.7: Xét phép thử tung đồng thời 2 đồng xu (Ví dụ 1.1). Không gian mẫu của phép thử là $\Omega = \{(S, S), (S, N), (N, S), (N, N)\}$ gồm 4 kết quả đồng khả năng. Gọi X là số mặt sấp xuất hiện, khi đó X là một biến ngẫu nhiên rời rạc.

Bảng phân bố xác suất

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

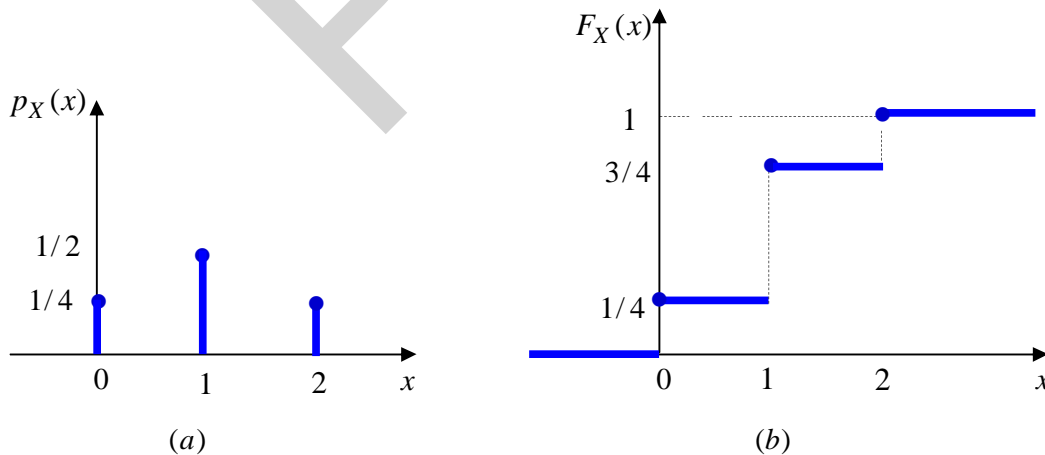
Hàm khối lượng xác suất

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

Hàm phân bố xác suất

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

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



Hình 2.4

Ví dụ 2.8: Chọn ngẫu nhiên 3 bi từ một túi có 6 bi đen, 4 bi trắng. Gọi X là số bi trắng trong 3 bi vừa chọn thì X là một biến ngẫu nhiên rời rạc. Tìm bảng phân bố xác suất và hàm phân bố xác suất.

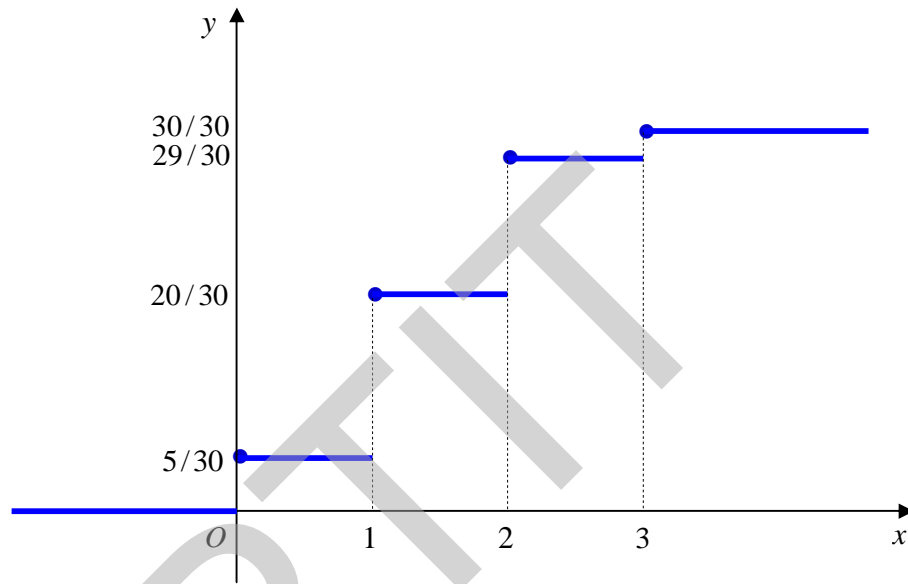
Giải: $P\{X=0\} = \frac{C_6^3}{C_{10}^3} = \frac{5}{30}, P\{X=1\} = \frac{C_6^2 C_4^1}{C_{10}^3} = \frac{15}{30},$

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

Hàm khối lượng xác suất

$$p_X(0) = \frac{5}{30}, p_X(1) = \frac{15}{30}, p_X(2) = \frac{9}{30}, p_X(3) = \frac{1}{30}; p_X(x) = 0 \text{ với mọi } x \text{ khác } 0, 1, 2, 3.$$

Đồ thị



Hình 2.5

Bảng phân bố xác suất

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

Hàm phân bố xác suất

$$F_X(x) = \begin{cases} 0 & \text{nếu } x < 0 \\ 5/30 & \text{nếu } 0 \leq x < 1 \\ 20/30 & \text{nếu } 1 \leq x < 2 \\ 29/30 & \text{nếu } 2 \leq x < 3 \\ 1 & \text{nếu } x \geq 3 \end{cases}$$

Ví dụ 2.9: Xét một dãy phép thử Bernoulli với xác suất thành công mỗi lần thử là p , $0 < p < 1$. Phép thử được thực hiện đến khi sự thành công xuất hiện. Gọi X là số lần thử.

a. Tìm hàm khối lượng xác suất $p_X(x)$.

b. Tìm hàm phân bố xác suất $F_X(x)$.

c. Khi $p = \frac{1}{2}$, tính $P\{1 < X \leq 4\}$ và $P\{X > 4\}$

Giải: X là biến ngẫu nhiên rời rạc với các giá trị có thể có là $1, 2, \dots, k, \dots$

Gọi A_k là biến cố “phép thử thành công ở lần thử thứ k ”. Các biến cố A_k độc lập nhau và $P(A_k) = p$.

$$P\{X = 1\} = P(A_1) = p$$

$$P\{X = 2\} = P(\overline{A_1} \cap A_2) = P(\overline{A_1})P(A_2) = qp \text{ với } q = 1 - p$$

Tổng quát $P\{X = k\} = P(\overline{A_1} \cap \dots \cap \overline{A_{k-1}} \cap A_k) = P(\overline{A_1}) \dots P(\overline{A_{k-1}})P(A_k) = q^{k-1}p \dots$

a. Vậy hàm khối lượng xác suất

$$p_X(x) = \begin{cases} q^{k-1}p & x = k \\ 0 & x \neq k \end{cases} ; k = 1, 2, \dots; 0 < p < 1 \quad (2.16)$$

Bảng phân bố xác suất có dạng:

X	1	2	...	k	...
P	p	qp	...	$q^{k-1}p$...

Biến ngẫu nhiên với phân bố rời rạc có dạng trên được gọi là **biến ngẫu nhiên phân bố cấp số nhân** (geometric random variable), hàm khối lượng xác suất lập thành cấp số nhân có công bội là q .

$$P\{X > k\} = \sum_{n=k+1}^{\infty} q^{n-1}p = \sum_{m=k}^{\infty} q^m p = q^k p \sum_{m=0}^{\infty} q^m = \frac{q^k p}{1-q} = q^k$$

$$\Rightarrow P\{X \leq k\} = 1 - P\{X > k\} = 1 - q^k$$

b. Hàm phân bố xác suất

$$F_X(x) = \begin{cases} 0 & \text{nếu } x < 0 \\ 1 - q^k & \text{nếu } k \leq x < k+1 \end{cases}$$

c. Khi $p = \frac{1}{2}$: $P\{1 < X \leq 4\} = F_X(4) - F_X(1) = \frac{15}{16} - \frac{1}{2} = \frac{7}{16}$

$$P\{X > 4\} = 1 - P\{X \leq 4\} = 1 - F_X(4) = 1 - \frac{15}{16} = \frac{1}{16}.$$

2.2.2 Các phân bố rời rạc thường gặp

2.2.2.1 Phân bố Bernoulli

Định nghĩa 2.4: Biến ngẫu nhiên rời rạc X nhận hai giá trị 0, 1 với xác suất tương ứng

$$p_X(k) = P\{X = k\} = p^k q^{1-k}; k = 0, 1 \quad (2.17)$$

trong đó $0 < p < 1$, $q = 1 - p$, được gọi là có phân bố Bernoulli tham số p .

Bảng phân bố xác suất

X	0	1
P	q	p

, $q = 1 - p$

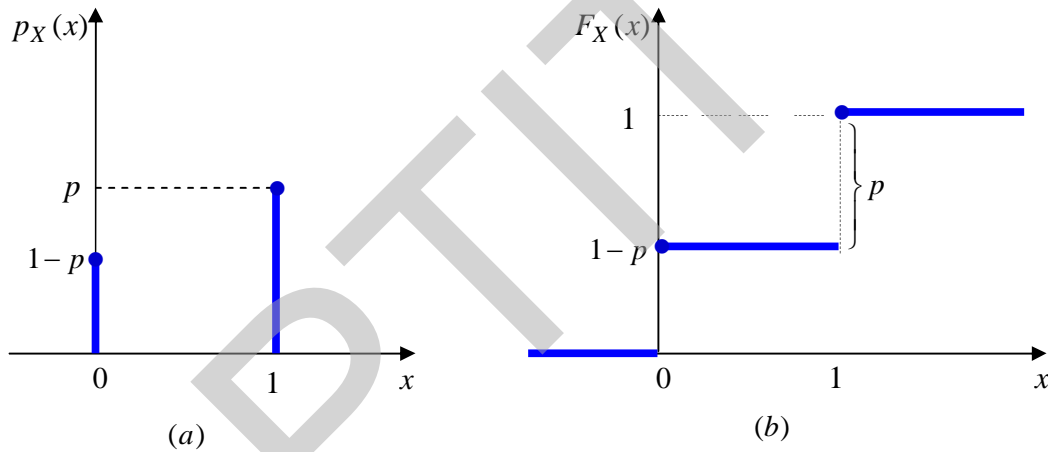
Hàm khối lượng xác suất và hàm phân bố xác suất như sau

$$p_X(x) = \begin{cases} 1-p & \text{nếu } x=0 \\ p & \text{nếu } x=1 \\ 0 & \text{nếu } x \neq 0, x \neq 1 \end{cases} \quad F_X(x) = \begin{cases} 0 & \text{nếu } x < 0 \\ 1-p & \text{nếu } 0 \leq x < 1 \\ 1 & \text{nếu } x \geq 1 \end{cases}$$

Xét phép thử Bernoulli với sự thành công của phép thử là sự xuất hiện của biến cố A và giả sử xác suất xuất hiện của A trong mỗi lần thử là p (xem 1.5 chương 1). Gọi X là số lần thành công trong một lần thử thì X là biến ngẫu nhiên rời rạc có phân bố Bernoulli tham số p .

Biến ngẫu nhiên X còn được gọi là phân bố không – một $A(p)$.

Hình 2.5 minh họa phân bố Bernoulli tham số p .



Hình 2.6: Đồ thị hàm khối lượng xác suất và hàm phân bố Bernoulli tham số p

Trong lý thuyết thống kê biến ngẫu nhiên có phân bố Bernoulli thường được dùng để đặc trưng cho các dấu hiệu nghiên cứu có tính định tính trong đó mỗi cá thể của tổng thể có dấu hiệu này hoặc không có dấu hiệu này (xem chương 4).

Chẳng hạn khi muốn nghiên cứu giới tính của khách hàng ta có thể đặc trưng cho giới tính bằng biến ngẫu nhiên với 2 giá trị bằng 0 (Nam) và bằng 1 (Nữ). Trong bài toán bầu cử nếu cử tri nào sẽ bỏ phiếu cho ứng cử viên A ta cho nhận giá trị 1, ngược lại ta cho nhận giá trị 0. Để xác định tỷ lệ phế phẩm của lô hàng ta gán cho mỗi sản phẩm một trong hai giá trị 0 và 1, nếu sản phẩm là phế phẩm ta cho nhận giá trị 1 và ngược lại cho nhận giá trị 0 ... Đó là các biến ngẫu nhiên có phân bố Bernoulli.

2.2.2.2 Phân bố nhị thức

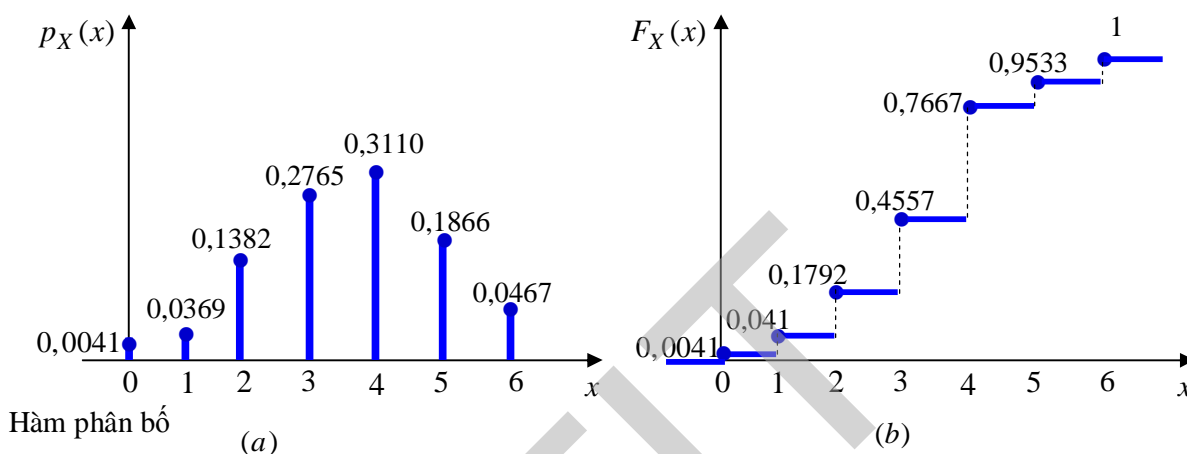
Định nghĩa 2.5: Biến ngẫu nhiên rời rạc X nhận các giá trị $0, 1, \dots, n$ với xác suất tương ứng

$$p_X(k) = P\{X = k\} = C_n^k p^k q^{n-k}; \quad k = 0, 1, \dots, n \quad (2.18)$$

trong đó n là số tự nhiên và $0 < p < 1$, được gọi là có phân bố nhị thức tham số n, p , ký hiệu $X \sim B(n; p)$.

Bảng phân bố xác suất của biến ngẫu nhiên có phân bố nhị thức $B(n; p)$

	X	0	1	...	k	...	n
Trong đó $P_q = 1 - p$		$C_n^0 p^0 q^n$	$C_n^1 p^1 q^{n-1}$...	$C_n^k p^k q^{n-k}$...	$C_n^n p^n q^0$



Hình 2.7: phân bố nhị thức với $n = 6$ và $p = 0,6$

$$F_X(x) = \begin{cases} 0 & \text{nếu } x < 0 \\ \sum_{k=0}^m C_n^k p^k q^{n-k} & \text{nếu } m \leq x < m+1, 0 \leq m \leq n-1, \\ 1 & \text{nếu } x \geq n \end{cases} \quad (2.19)$$

Nhận xét 2.2:

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

$$X = X_1 + X_2 + \dots + X_n \sim B(n; p) \quad (2.20)$$

Nói cách khác tổng của các biến ngẫu nhiên Bernoulli ứng với cùng một dãy phép thử Bernoulli là biến ngẫu nhiên có phân bố nhị thức.

3. Từ (2.20) suy ra rằng nếu $X \sim B(n_1; p)$, $Y \sim B(n_2; p)$; X, Y độc lập và cùng phép thử thì

$$X + Y \sim B(n_1 + n_2; p) \quad (2.21)$$

Ví dụ 2.10: Tỷ lệ phế phẩm của lô hàng là 4%. Chọn ngẫu nhiên 20 sản phẩm để kiểm tra. Gọi X là số phế phẩm phát hiện được.

- a. Gọi tên luật phân bố xác suất của X .
- b. Tính xác suất có đúng 5 phế phẩm phát hiện được.

c. Lô hàng được xem là đạt tiêu chuẩn nếu số phế phẩm phát hiện 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 là thực hiện một phép thử Bernoulli với sự thành công của phép thử là phát hiện ra phế phẩm. Theo giả thiết 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 phát hiện được là số lần thành công trong 20 phép thử này. Vậy X có phân bố nhị thức $X \sim B(n; p)$, với $n = 20$, $p = 0,04$.

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

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

Ví dụ 2.11: Một nguồn nhị phân phát ra hai ký số (digit) 1 và 0 một cách ngẫu nhiên với xác suất tương ứng 0,6 và 0,4.

a. Tính xác suất có đúng hai ký số 1 và ba ký số 0 trong dãy có năm ký số.

b. Tính xác suất có ít nhất ba ký số 1 trong dãy có năm ký số.

Giải: Gọi X là số các ký số 1 trong dãy có năm ký số. Theo giả thiết chỉ có hai khả năng đầu vào có thể là 1 và 0, xác suất xuất hiện của 1 không đổi và bằng 0,6. Vậy có thể xem mỗi ký số xuất hiện trong dãy tương ứng với một phép thử Bernoulli mà sự thành công của mỗi lần thử là xuất hiện ký số 1. Mỗi dãy có năm ký số tương ứng với 5 phép thử Bernoulli, vậy X có phân bố nhị thức $X \sim B(n; p)$, với $n = 5$, $p = 0,6$.

a. Xác suất để dãy có đúng hai ký số 1 và ba ký số 0 trong dãy có năm ký số là

$$P\{X = 2\} = C_5^2 (0,6)^2 (0,4)^3 = 0,23.$$

b. Xác suất có ít nhất ba ký số 1 trong dãy có năm ký số là

$$P\{X \geq 3\} = 1 - P\{X \leq 2\}$$

Trong đó
$$P\{X \leq 2\} = \sum_{k=0}^2 C_5^k (0,6)^k (0,4)^{5-k} = 0,317$$

Vậy
$$P\{X \geq 3\} = 1 - 0,317 = 0,683.$$

2.2.2.3 Phân bố Poisson

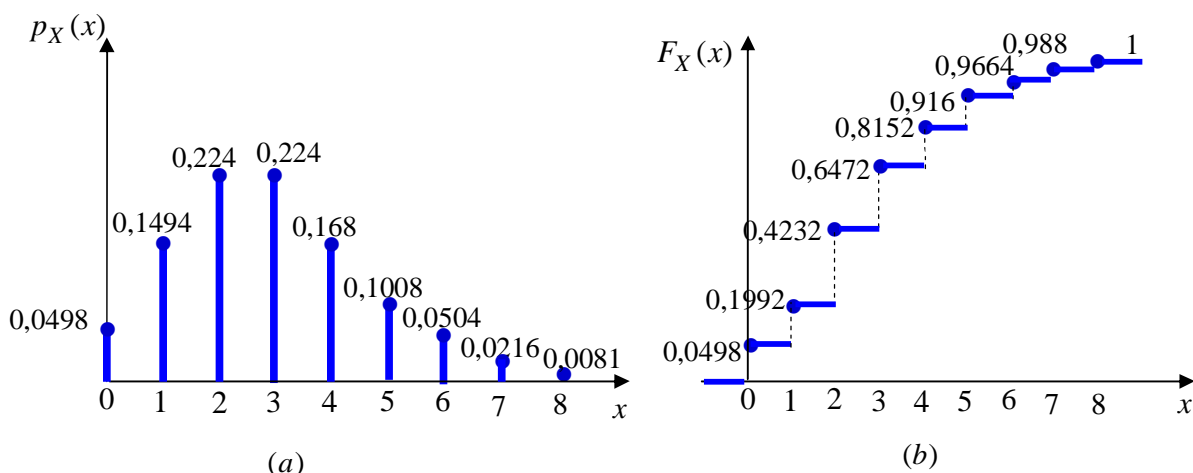
Định nghĩa 2.6: Biến ngẫu nhiên X có miền giá trị $R_X = \mathbb{N} = \{0, 1, \dots, n, \dots\}$ với hàm khối lượng xác suất

$$p_X(k) = P\{X = k\} = e^{-\lambda} \frac{\lambda^k}{k!}; \lambda > 0; k = 0, 1, 2, \dots \quad (2.22)$$

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

Hàm phân bố

$$F_X(x) = e^{-\lambda} \sum_{k=0}^n \frac{\lambda^k}{k!}, \quad n \leq x < n+1 \quad (2.23)$$



Hình 2.8 Phân bố Poisson tham số $\lambda = 3$

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

Ví dụ 2.12: Ở một tổng đài điện thoại các cuộc gọi đến một cách ngẫu nhiên, độc lập và trung bình có 2 cuộc gọi trong 1 phút. Tìm xác suất để:

- a. Có đúng 5 cuộc gọi đến trong 2 phút (biến cố A).
- b. Không có một cuộc gọi nào trong 30 giây (biến cố B).
- c. Có ít nhất 1 cuộc gọi trong 10 giây (biến cố C).

Giải: Nếu ký hiệu $X(t)$ là số cuộc gọi đến tổng đài trong khoảng thời gian t phút. Theo giả thiết trung bình có 2 cuộc gọi đến tổng đài trong 1 phút, vậy trung bình có $2t$ cuộc gọi trong t phút, do đó $X(t) \sim P(2t)$.

- a. $X(2) \sim P(4)$, áp dụng công thức 2.22 với $\lambda = 4$, $k = 5$ ta được:

$$P(A) = P\{X(2) = 5\} = e^{-4} \frac{4^5}{5!} \approx 0,156.$$

- b. $X(1/2) \sim P(1)$, áp dụng công thức 2.22 với $\lambda = 1$, $k = 0$ ta được:

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

- c. $X(1/6) \sim P(1/3)$, do đó

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

Phân bố Poisson có ứng dụng rộng rãi trong nhiều lĩnh vực thực tế như kiểm tra chất lượng sản phẩm, lý thuyết quản trị dự trữ, lý thuyết xếp hàng ... Hầu hết các quá trình đến trong lý thuyết xếp hàng, trong hệ phục vụ đám đông, các bài toán chuyển mạch trong tổng đài ... thường được xét là quá trình đến Poisson.

Ta sẽ chứng minh: nếu X_1, X_2 là hai biến ngẫu nhiên độc lập trong cùng phép thử có phân bố Poisson tham số lần lượt λ_1, λ_2 thì $X_1 + X_2$ cũng có phân bố Poisson tham số $\lambda_1 + \lambda_2$

$$X_1 + X_2 \sim P(\lambda_1 + \lambda_2). \quad (2.24)$$

Thật vậy, áp dụng công thức xác suất đầy đủ, chú ý đến điều kiện miền giá trị là tập số tự nhiên và tính chất độc lập của X_1, X_2 ta có

$$\begin{aligned} P\{X_1 + X_2 = k\} &= \sum_{m=0}^k P\{X_1 = m\} P\{X_1 + X_2 = k | X_1 = m\} = \sum_{m=0}^k P\{X_1 = m\} P\{X_2 = k - m\} \\ &= \sum_{m=0}^k e^{-\lambda_1} \frac{\lambda_1^m}{m!} \cdot e^{-\lambda_2} \frac{\lambda_2^{k-m}}{(k-m)!} = e^{-(\lambda_1 + \lambda_2)} \sum_{m=0}^k \frac{\lambda_1^m}{m!} \cdot \frac{\lambda_2^{k-m}}{(k-m)!} \\ &= e^{-(\lambda_1 + \lambda_2)} \frac{1}{k!} \sum_{m=0}^k C_k^m \lambda_1^m \lambda_2^{k-m} = e^{-(\lambda_1 + \lambda_2)} \frac{(\lambda_1 + \lambda_2)^k}{k!}. \end{aligned}$$

2.3 BIẾN NGẪU NHIÊN LIÊN TỤC

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

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

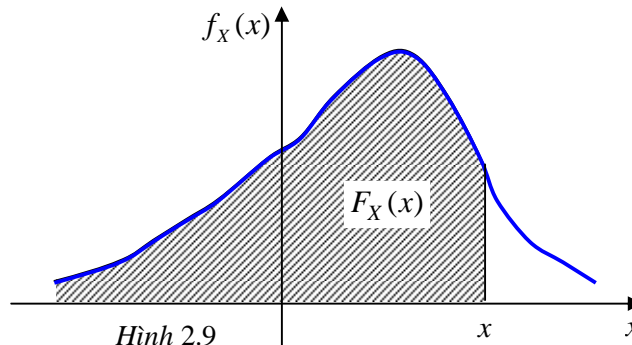
Định nghĩa 2.7: Giả sử X là một biến ngẫu nhiên liên tục có hàm phân bố xác suất $F_X(x)$, nếu tồn tại hàm $f_X(x)$ sao cho

$$F_X(x) = \int_{-\infty}^x f_X(t) dt, \text{ với mọi } x \in \mathbb{R} \quad (2.25)$$

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

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

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



Hình 2.9

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

$$1. \quad \frac{d}{dx} F_X(x) = f_X(x) \text{ tại các điểm } x \text{ mà } f_X(x) \text{ liên tục.} \quad (2.26)$$

$$2. \quad f_X(x) \geq 0 \text{ với mọi } x \in \mathbb{R}, \quad (2.27)$$

$$3. \quad \int_{-\infty}^{\infty} f_X(x) dx = 1, \quad (2.28)$$

Vì xác suất tại từng điểm của biến ngẫu nhiên liên tục bằng 0, do đó ta có

$$4. \quad P\{a < X < b\} = P\{a \leq X \leq b\} = P\{a \leq X < b\} = P\{a < X \leq b\} = \int_a^b f_X(x) dx. \quad (2.29)$$

Nhận xét 2.3: Công thức 2.25 chỉ ra rằng hàm phân bố $F_X(x)$ là một nguyên hàm của hàm mật độ $f_X(x)$, vì vậy khi biết hàm mật độ ta có thể tìm hàm phân bố bằng cách tính tích phân. Ngược lại công thức 2.26 cho phép tìm hàm mật độ xác suất từ hàm phân bố xác suất bằng cách lấy đạo hàm của hàm phân bố xác suất. Các ví dụ sau minh họa điều này.

Ví dụ 2.13: Hàm phân bố xác suất của biến ngẫu nhiên liên tục X có dạng

$$F_X(x) = \begin{cases} 0 & \text{vì } x < 0 \\ kx^2 & \text{vì } 0 \leq x < 1 \\ 1 & \text{vì } x \geq 1 \end{cases}$$

- a. Xác định hệ số k ;
- b. Tìm hàm mật độ xác suất $f_X(x)$.

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

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

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

b. Theo tính chất (2.26) của hàm mật độ xác suất ta có

$$f_X(x) = \frac{d}{dx} F_X(x) = \begin{cases} 0 & \text{vì } x \leq 0 \\ 2x & \text{vì } 0 < x < 1 \\ 0 & \text{vì } x \geq 1 \end{cases}$$

Ví dụ 2.14: Biến ngẫu nhiên liên tục X với hàm mật độ xác suất có dạng

$$f_X(x) = \begin{cases} 0 & \text{vì } x < 1 \\ \frac{k}{x^2} & \text{vì } x \geq 1 \end{cases}$$

Hãy xác định:

- a. Hệ số k ;

- b.** Hàm phân bố xác suất $F_X(x)$;
- c.** Tính xác suất $P\{2 < X < 3\}$;
- d.** Thực hiện 4 lần phép thử độc lập biến ngẫu nhiên X , tính xác suất biến ngẫu nhiên X không lấy giá trị trong khoảng $(2; 3)$.

Giải: **a.** Theo tính chất (2.18) ta có $1 = \int_{-\infty}^{\infty} f_X(x)dx = \int_1^{\infty} \frac{k}{x^2} dx = -\lim_{a \rightarrow \infty} \left(\frac{k}{x} \right) \Big|_1^a = k$, từ đó $k = 1$.

b. Từ công thức (2.15) ta có

$$F_X(x) = \int_{-\infty}^x f_X(t)dt = \begin{cases} 0 & \text{vì } x < 1 \\ 1 - \frac{1}{x} & \text{vì } x \geq 1 \end{cases}$$

c. Từ công thức (2.19) và (2.5) ta có

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

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

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

2.3.2 Các phân bố liên tục thường gặp

2.3.2.1 Phân bố đều $U(a, b)$

Định nghĩa 2.8: Biến ngẫu nhiên X được gọi là có phân bố đều trên (a, b) , ký hiệu $X \sim U(a; b)$ nếu hàm mật độ của X xác định bởi:

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{nếu } a < x < b \\ 0 & \text{nếu ngược lại} \end{cases} \quad (2.30)$$

Trường hợp $x \leq a$, $P\{X \leq x\} = 0$.

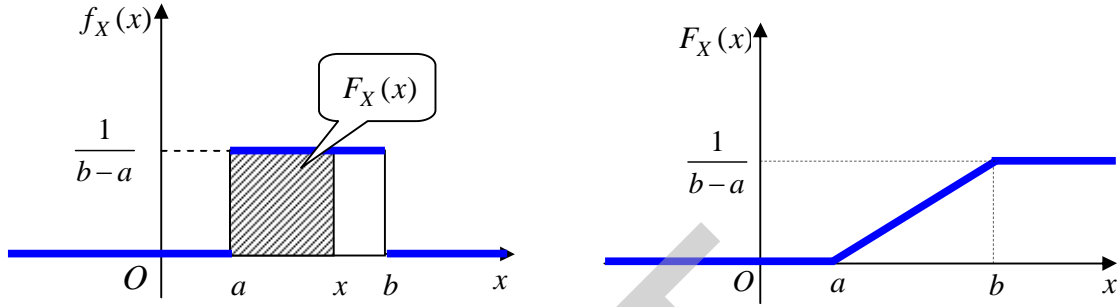
Trường hợp $a \leq x \leq b$, $P\{X \leq x\}$ bằng diện tích hình chữ nhật có chiều rộng bằng $x - a$ và chiều cao bằng $\frac{1}{b-a}$ (xem hình 2.10), do đó $P\{X \leq x\} = \frac{x-a}{b-a}$.

Trường hợp $x \geq b$, $P\{X \leq x\}$ bằng diện tích hình chữ nhật có chiều rộng bằng $b - a$ và chiều cao bằng $\frac{1}{b-a}$ (xem hình 2.10), do đó $P\{X \leq x\} = 1$.

Hàm phân bố xác suất tương ứng

$$F_X(x) = \int_{-\infty}^x f_X(t)dt = \begin{cases} 0 & \text{nếu } x \leq a \\ \frac{x-a}{b-a} & \text{nếu } a < x < b \\ 1 & \text{nếu } x \geq b \end{cases} \quad (2.31)$$

Vậy X có khả năng nhận giá trị trong khoảng $(a;b)$ là “đều nhau” và không nhận giá trị ngoài khoảng $(a;b)$.



Hình 2.10 Phân bố đều $U(a,b)$

Phân bố đều có nhiều ứng dụng trong bài toán mô phỏng thống kê, đặc biệt trong phương pháp phi tham số. Trong một số lý thuyết kết luận thống kê người ta thường xuất phát từ quy tắc sau đây: Nếu ta không biết gì về giá trị của tham số cần ước lượng thì mỗi giá trị có thể có của tham số đó là đồng khả năng, điều đó dẫn đến việc quan niệm tham số cần ước lượng như một biến ngẫu nhiên có phân bố đều.

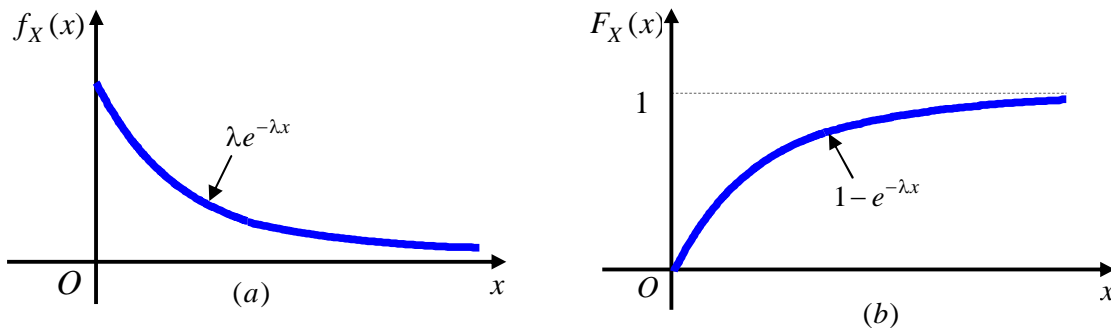
2.3.2.2 Phân bố mũ

Định nghĩa 2.9: Biến ngẫu nhiên liên tục X được gọi là có phân bố mũ tham số $\lambda > 0$ nếu hàm mật độ xác suất xác định như sau:

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{nếu } x > 0 \\ 0 & \text{nếu } x \leq 0 \end{cases} \quad (2.32)$$

Hàm phân bố xác suất

$$F_X(x) = \int_{-\infty}^x f_X(t)dt = \begin{cases} 1 - e^{-\lambda x} & \text{nếu } x > 0 \\ 0 & \text{nếu } x \leq 0 \end{cases} \quad (2.33)$$



Hình 2.11 Phân bố mũ tham số $\lambda = 1/7$

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 2.15: 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) (xem kỳ vọng của phân bố mũ ở mục 2.5.3). Thời gian bảo hành là 2 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:

$$P\{X \leq 2\} = 1 - e^{-2\lambda} = 1 - e^{-\frac{2}{6,25}} = 1 - e^{-0,32} = 1 - 0,726 = 0,274.$$

Vậy có khoảng 27,4% số mạch điện tử bán ra phải thay thế trong thời gian bảo hành.

Nhận xét 2.4: Tính chất thú vị nhất của phân bố mũ là tính “không nhớ”.

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

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

Điều này tương đương với

$$P\{X > x+t\} = P\{X > x\}P\{X > t\}; \quad \forall x, t > 0 \quad (2.34)$$

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

Điều kiện (2.34) có thể viết lại

$$G(x+t) = G(x)G(t) \quad (2.35)$$

Giải phương trình (2.35) (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 là phân bố có tính chất “không nhớ” và được gọi là *phân bố Markov*.

Ví dụ 2.16: Giả sử thời gian mỗi cuộc gọi điện thoại (tính theo phút) là một biến ngẫu nhiên X với phân bố mũ tham số $\lambda = 1/10$. Một bộ điện thoại chỉ phục vụ từng người và giả sử A vào bộ điện thoại trước khi B đến. Tính xác suất B phải chờ đến lượt mình trong khoảng thời gian:

- a. Ít hơn 5 phút b. Trong khoảng từ 5 đến 10 phút.

Giải: a. Vì tính chất “không nhớ” của phân bố mũ do đó thời gian chờ của B bằng thời gian A tiếp tục hoàn thành cuộc gọi tính từ lúc B đến và không phụ thuộc A đã gọi trong thời gian bao lâu. Theo công thức (2.33) ta được xác suất B phải chờ ít hơn 5 phút là

$$P\{X < 5\} = P\{X \leq 5\} = F_X(5) = 1 - e^{-0,5} \approx 0,394.$$

- b. Lập luận tương tự trên ta được xác suất B phải chờ trong khoảng từ 5 đến 10 phút là

$$P\{5 < X < 10\} = F_X(10) - F_X(5) = (1 - e^{-1}) - (1 - e^{-0,5}) = e^{-0,5} - e^{-1} \approx 0,239.$$

2.3.2.3 Phân bố Erlang

Định nghĩa 2.10: 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:

$$f_X(x) = \begin{cases} \frac{(\lambda x)^{k-1}}{(k-1)!} \lambda e^{-\lambda x} & \text{nếu } x > 0 \\ 0 & \text{nếu } x \leq 0 \end{cases} \quad (2.36)$$

Có thể chứng minh được rằng nếu X_1, X_2, \dots, X_k là k biến ngẫu nhiên độc lập cùng có phân bố mũ tham số $\lambda > 0$ thì

$$X = X_1 + X_2 + \dots + X_k$$

có phân bố Erlang tham số $(k; \lambda)$.

Phân bố Erlang tham số $(1; \lambda)$ là phân bố mũ tham số λ .

2.3.2.4 Phân bố chuẩn

A. Định nghĩa

Định nghĩa 2.11: Biến ngẫu nhiên liên tục X có phân bố chuẩn $N(\mu; \sigma^2)$, ký hiệu $X \sim N(\mu; \sigma^2)$, nếu hàm mật độ xác suất có dạng

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}; \quad \forall x \in \mathbb{R} \quad (2.37)$$

Phân bố chuẩn được Gauss tìm ra năm 1809 nên nó còn được gọi là *phân bố Gauss*. Phân bố chuẩn thường được sử dụng trong các bài toán đo đạc các đại lượng vật lý, thiên văn ...

Trong thực tế, nhiều biến ngẫu nhiên tuân theo phân bố chuẩn hoặc tiệm cận chuẩn (Định lý giới hạn trung tâm). Chẳng hạn: trọng lượng, chiều cao của một nhóm người nào đó, điểm thi của thí sinh, năng suất cây trồng, mức lãi suất của một công ty, nhu cầu tiêu thụ của một mặt hàng nào đó, nhiễu trắng trên các kênh thông tin ... là các biến ngẫu nhiên có phân bố chuẩn.

Có thể chứng minh được, nếu X_1, X_2 là hai biến ngẫu nhiên độc lập lần lượt có phân bố chuẩn $X_1 \sim N(\mu_1; \sigma_1^2)$ và $X_2 \sim N(\mu_2; \sigma_2^2)$ thì tổ hợp tuyến tính bất kỳ của X_1, X_2 cũng có phân bố chuẩn. Cụ thể, với mọi a, b khác không thì

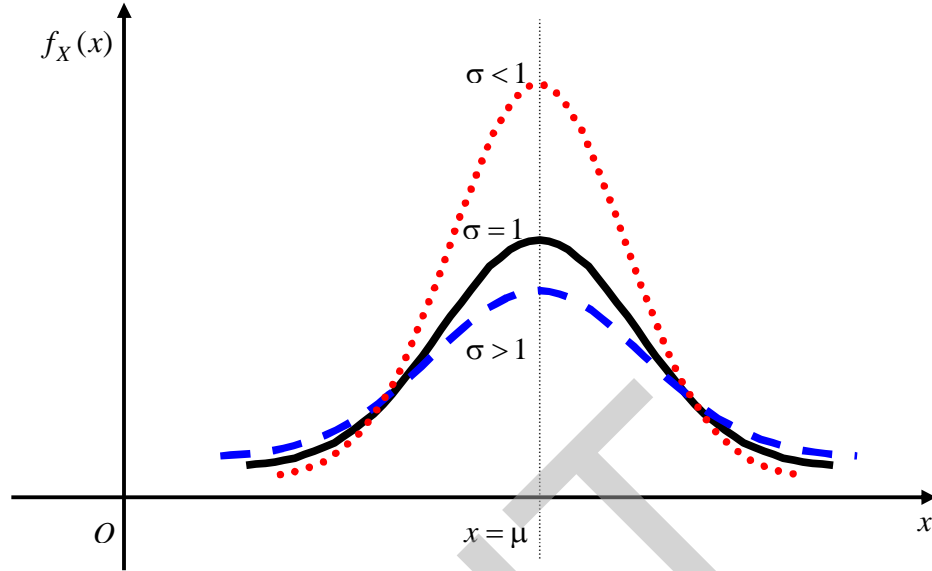
$$aX_1 + bX_2 \sim N(a\mu_1 + b\mu_2; a^2\sigma_1^2 + b^2\sigma_2^2) \quad (2.38)$$

B. Tính chất đồ thị của hàm mật độ xác suất của phân bố chuẩn

Từ công thức xác định hàm mật độ xác suất (2.37) ta suy ra các tính chất sau của đồ thị:

- Nhận trục $x = \mu$ làm trục đối xứng.
- Tiệm cận với trục hoành khi $x \rightarrow \pm\infty$.
- Diện tích giới hạn bởi đồ thị và trục hoành bằng 1.
- Đạt cực đại tại $x = \mu$ và có giá trị cực đại bằng $\frac{1}{\sigma\sqrt{2\pi}}$.
- Có 2 điểm uốn tại $x = \mu \pm \sigma$.

- Do đó khi μ tăng lên thì đồ thị dịch sang phải, còn khi μ giảm đồ thị dịch sang trái.
- Khi σ tăng lên thì đồ thị sẽ thấp xuống, còn khi σ giảm đồ thị cao lên và nhọn hơn.



Hình 2.12: Đồ thị hàm mật độ của phân bố chuẩn $N(\mu; \sigma^2)$

C. Phân bố chuẩn tắc

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

Hàm mật độ xác suất của phân bố chuẩn tắc $N(0; 1)$

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}; \forall x \in \mathbb{R} \quad (2.39)$$

Hàm phân bố xác suất của $N(0; 1)$

$$\Phi(x) = \int_{-\infty}^x \varphi(t) dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt; \forall x \in \mathbb{R} \quad (2.40)$$

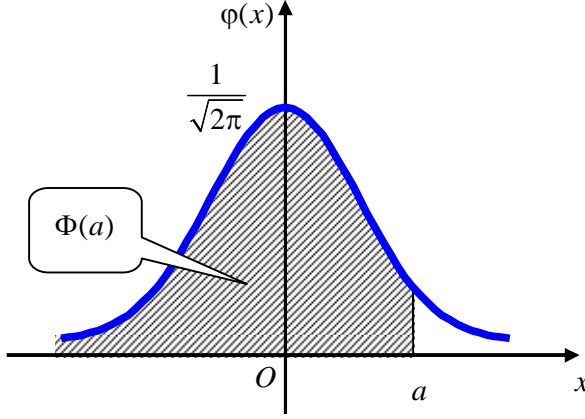
Có bảng tính sẵn các giá trị của $\varphi(x)$ và $\Phi(x)$ (xem Phụ lục I và Phụ lục II).

Cần chú ý rằng một số tài liệu cho bảng tính $\Phi_0(x) = \int_0^x \varphi(t) dt = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt$ với $x \geq 0$.

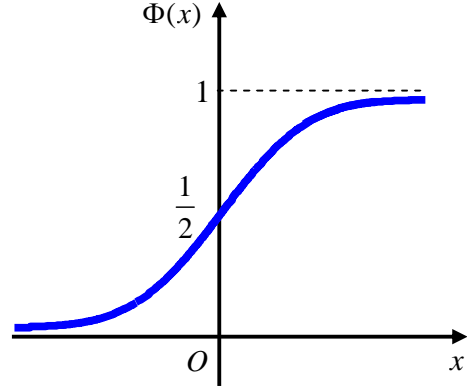
Công thức liên hệ là

$$\Phi(x) = \Phi_0(x) + 0,5 \text{ với mọi } x \geq 0.$$

Đồ thị của hàm mật độ xác suất $\varphi(x)$ và hàm phân bố xác suất $\Phi(x)$



Hình 2.13a: Đồ thị hàm mật độ của phân bố chuẩn tắc $N(0;1)$



Hình 2.13b: Đồ thị hàm phân bố của phân bố chuẩn tắc $N(0;1)$

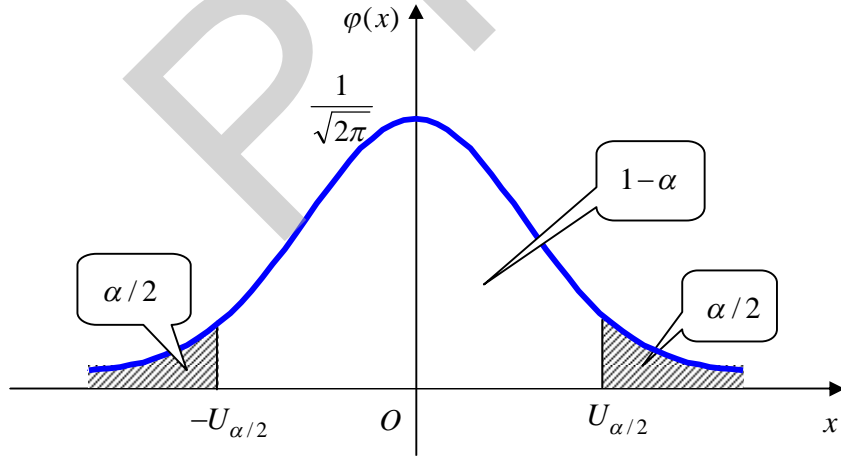
Từ đồ thị của hàm mật độ $\varphi(x)$ (hình 2.13a) ta có:

- 1) $\Phi(x) + \Phi(-x) = 1$, $\Phi(-x) = 1 - \Phi(x)$.
- 2) Nếu $X \sim N(0;1)$ thì

$$\forall a > 0, P\{|X| > a\} = 2(1 - \Phi(a)), P\{|X| < a\} = 2\Phi(a) - 1. \quad (2.41)$$

Định nghĩa 2.12: Giá trị U_α gọi là giá trị tới hạn của phân bố chuẩn tắc mức α nếu

$$\Phi^{-1}(1 - \alpha) = U_\alpha. \quad (2.42)$$



Hình 2.14: Giá trị tới hạn của phân bố chuẩn tắc $N(0;1)$

Từ định nghĩa 2.12 và từ hình 2.14 ta có: Nếu $X \sim N(0;1)$ thì

$$\forall \alpha : 0 < \alpha < 1; P\{X > U_\alpha\} = \alpha; P\{|X| > U_{\frac{\alpha}{2}}\} = \alpha; P\{|X| < U_{\frac{\alpha}{2}}\} = 1 - \alpha \quad (2.43)$$

- 3) Ta có thể chứng minh được:

$$\text{Nếu } X \sim \mathbf{N}(\mu; \sigma^2) \text{ thì } \frac{X - \mu}{\sigma} \sim \mathbf{N}(0; 1). \quad (2.44)$$

$$\text{Từ đó ta có } F_X(x) = P\{X \leq x\} = P\left\{\frac{X - \mu}{\sigma} \leq \frac{x - \mu}{\sigma}\right\} = \Phi\left(\frac{x - \mu}{\sigma}\right). \quad (2.45)$$

$$P\{a < X < b\} = P\left\{\frac{a - \mu}{\sigma} < \frac{X - \mu}{\sigma} < \frac{b - \mu}{\sigma}\right\} = \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right). \quad (2.46)$$

Công thức 2.45-2.46 vẫn còn đúng khi thay dấu “<” thành dấu “≤” và ngược lại (xem công thức 2.29).

Ví dụ 2.17: Giả sử $X \sim \mathbf{N}(\mu; \sigma^2)$, $\mu = 2100$, $\sigma = 200$. Hãy tìm:

- $P\{X \leq 2400\}$.
- $P\{1700 < X < 2200\}$.
- Xác định a để $P\{X > a\} = 0,03$.

Giải: Áp dụng công thức (2.44), (2.45), (2.46) ta có:

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

b. Áp dụng các công thức (2.46) ta được

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

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

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

D. Quy tắc hai xích ma và ba xích ma

Áp dụng công thức (2.41), (2.44) ta có thể tính xác suất của sự sai lệch giữa biến ngẫu nhiên có phân bố chuẩn $X \sim \mathbf{N}(\mu; \sigma^2)$ và tham số μ của nó theo công thức

$$P\{|X - \mu| < \varepsilon\} = 2\Phi\left(\frac{\varepsilon}{\sigma}\right) - 1 \quad (2.47)$$

Nếu trong công thức (2.47) ta đặt $\varepsilon = 2\sigma$ tức là bằng hai lần độ lệch chuẩn của X (xem mục 2.42) thì $P\{|X - \mu| < 2\sigma\} = 2\Phi(2) - 1 = 0,9546$. Vậy

$$P\{\mu - 2\sigma < X < \mu + 2\sigma\} = 0,9546 \quad (2.48)$$

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

$$P\{\mu - 3\sigma < X < \mu + 3\sigma\} = 0,9974 \quad (2.49)$$

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

Nếu X có phân bố chuẩn $N(\mu; \sigma^2)$ thì có đến 95,46% 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 $(\mu - 3\sigma; \mu + 3\sigma)$.

2.3.2.5 Phân bố “khi bình phương”

Định nghĩa 2.13: Biến ngẫu nhiên liên tục X có phân bố “Khi bình phương” n bậc tự do, ký hiệu $X \sim \chi^2(n)$ nếu hàm mật độ có dạng

$$f_X(x) = \begin{cases} \frac{\left(\frac{x}{2}\right)^{n/2-1}}{\Gamma(n/2)} \frac{1}{2} e^{-\frac{x}{2}} & \text{nếu } x > 0 \\ 0 & \text{nếu } x \leq 0 \end{cases} \quad (2.50)$$

trong đó $\Gamma(x) = \int_0^{+\infty} t^{x-1} e^{-t} dt$, $x > 0$ là hàm Gamma.

Phân bố “khi bình phương” do Karl Pearson đưa ra vào năm 1900.

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

$$\sum_{i=1}^n X_i^2 = X_1^2 + X_2^2 + \dots + X_n^2 \sim \chi_n^2 \quad (2.51)$$

Nhận xét 2.5:

1. Trong thực tế ta thường sử dụng phân bố “khi bình phương” dưới dạng tổng của bình phương của các biến ngẫu nhiên độc lập cùng có phân bố chuẩn tắc theo công thức (2.51) mà ít xét đến hàm mật độ xác suất (2.50).
2. Từ (2.51) suy ra rằng nếu X_1, X_2, \dots, X_k là các biến ngẫu nhiên độc lập có phân bố “khi bình phương” với bậc tự do lần lượt n_1, n_2, \dots, n_k thì $X_1 + X_2 + \dots + X_k$ là biến ngẫu nhiên có phân bố “khi bình phương” $n_1 + n_2 + \dots + n_k$ bậc tự do

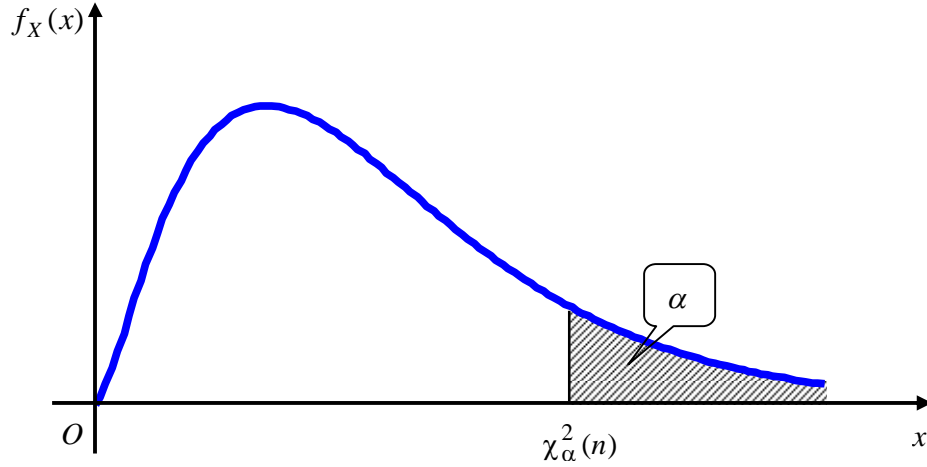
$$X_1 + X_2 + \dots + X_k \sim \chi_{n_1+n_2+\dots+n_k}^2 \quad (2.52)$$

3. Có thể chứng minh được: Nếu X_1, X_2 là hai biến ngẫu nhiên độc lập, X_1 có phân bố “khi bình phương” với n_1 bậc tự do và $X_1 + X_2$ có phân bố “khi bình phương” với n bậc tự do, $n > n_1$ thì X_2 là biến ngẫu nhiên có phân bố “khi bình phương” với $n - n_1$ bậc tự do.

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

$$P\{\chi^2 > \chi_\alpha^2(n)\} = \alpha. \quad (2.53)$$

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



Hình 2.15: Giá trị tới hạn của phân bố “khi bình phương”

Nhận xét 2.6: Biến ngẫu nhiên X có phân bố Gamma tham số (α, λ) , ký hiệu $X \sim \Gamma(\alpha; \lambda)$, nếu hàm mật độ có dạng

$$f_X(x) = \begin{cases} \frac{(\lambda x)^{\alpha-1}}{\Gamma(\alpha)} \lambda e^{-\lambda x} & \text{nếu } x > 0 \\ 0 & \text{nếu } x \leq 0 \end{cases} \quad (2.54)$$

Trong đó $\alpha, \lambda > 0$.

- Khi $\alpha = k$ là một số nguyên dương thì phân bố $\Gamma(k; \lambda)$ là phân bố Erlang tham số $(k; \lambda)$ có hàm mật độ xác định bởi công thức (2.36).
- Khi $\alpha = 1$ thì $\Gamma(1; \lambda)$ là phân bố mũ tham số λ .
- Khi $\lambda = 1/2, \alpha = n/2$ thì phân bố $\Gamma(n/2, 1/2)$ là phân bố “khi bình phương” n bậc tự do.

2.3.2.6 Phân bố Student

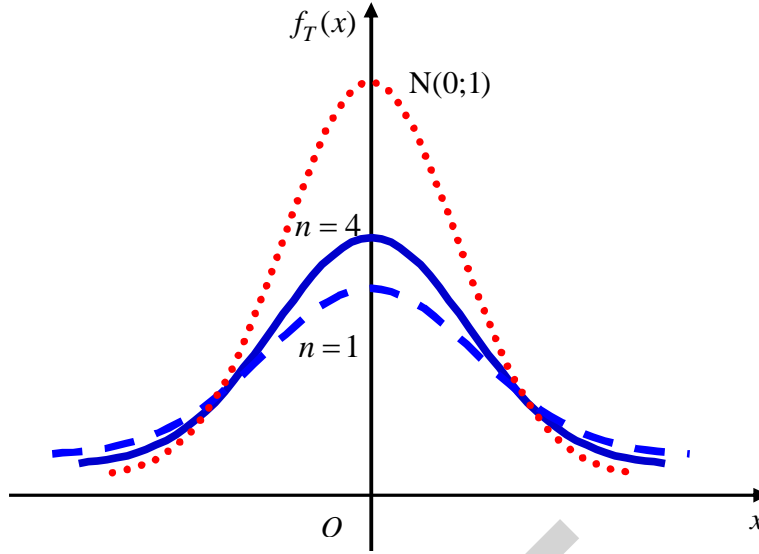
Định nghĩa 2.14: Biến ngẫu nhiên liên tục T có phân bố Student n bậc tự do, ký hiệu $T \sim \mathbf{T}(n)$, nếu hàm mật độ xác suất có dạng:

$$f_T(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma(n/2)} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}, \quad -\infty < x < \infty. \quad (2.55)$$

trong đó $\Gamma(x)$ là hàm Gamma.

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

$$T = \frac{Z}{\sqrt{V/n}} \sim \mathbf{T}(n) \quad (2.56)$$



Hình 2.16: Đồ thị hàm mật độ của phân bố Student

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

$$P\{T > t_\alpha(n)\} = \alpha. \quad (2.57)$$

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

Hàm mật độ xác suất (2.54) là hàm chẵn nên đồ thị đối xứng qua trục tung. Khi số bậc tự do tăng lên, phân bố Student hội tụ rất nhanh về phân bố chuẩn tắc $N(0,1)$. Do đó khi n đủ lớn ($n \geq 30$) có thể dùng phân bố chuẩn tắc thay cho phân bố Student. Tuy nhiên khi n nhỏ ($n < 30$) việc thay thế như trên sẽ gặp sai số lớn.

2.4 CÁC THAM SỐ ĐẶC TRƯNG CỦA BIẾN NGẪU NHIÊN

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

2.4.1 Kỳ vọng toán

2.4.1.1 Định nghĩa

Với mọi biến ngẫu nhiên X ta ký hiệu EX hoặc $E(X)$ hoặc $E[X]$ và xác định như sau:

- (i) Trường hợp biến ngẫu nhiên X rời rạc có miền giá trị R_X với hàm khối lượng xác suất $p_X(x_i)$

$$EX = \sum_{x_i \in R_X} x_i p_X(x_i) \quad (2.58)$$

- (ii) Trường hợp biến ngẫu nhiên X liên tục có hàm mật độ xác suất $f_X(x)$

$$EX = \int_{-\infty}^{\infty} x f_X(x) dx \quad (2.59)$$

Nếu chuỗi (2.58) hội tụ tuyệt đối (trường hợp X rời rạc) hoặc tích phân (2.59) hội tụ tuyệt đối (trường hợp X liên tục) thì ta gọi EX là **kỳ vọng của biến ngẫu nhiên** X (expected value). Trường hợp ngược lại ta nói X không tồn tại kỳ vọng.

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

Ví dụ 2.18: Tính kỳ vọng của biến ngẫu nhiên X cho ở ví dụ 2.8.

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

Ví dụ 2.19: Theo thống kê việc một thanh niên 25 tuổi sẽ sống thêm trên một năm có xác suất là 0,992, xác suất để người đó chết trong vòng một năm tới là 0,008 (xem ví dụ 1.18). Một chương trình bảo hiểm kinh doanh bảo hiểm sinh mạng trong 1 năm cho thanh niên độ tuổi 25 với số tiền chi trả 1000 đô la, tiền mua bảo hiểm là 10 đô la. Hỏi lợi nhuận trung bình của công ty bảo hiểm nhận được trên mỗi khách hàng là bao nhiêu?

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

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

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

Ví dụ 2.20: Tuổi thọ của một loại côn trùng nào đó là một biến ngẫu nhiên X (đơn vị là tháng) với hàm mật độ xác suất như sau:

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

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

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

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

$$\text{Tuổi thọ trung bình } EX = \frac{3}{64} \int_0^4 x^3(4-x)dx = \frac{3}{64} \left(x^4 - \frac{x^5}{5} \right) \Big|_0^4 = \frac{12}{5} \text{ (tháng)}.$$

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. Ta minh họa điều này trong trường hợp rời rạc sau.

Chương 2: Biến ngẫu nhiên và các đặc trưng của chúng

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

$r_i x_i$ là tổng giá trị X nhận được ứng với giá trị x_i .

$r_1 x_1 + r_2 x_2 + \dots + r_m x_m$ là tổng tất cả các giá trị X nhận được.

Vậy $\frac{r_1 x_1 + r_2 x_2 + \dots + r_m x_m}{n}$ là giá trị trung bình của X , trong đó $r_1 + r_2 + \dots + r_m = n$.

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

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

$$\frac{r_1 x_1 + r_2 x_2 + \dots + r_m x_m}{n} = f_1 x_1 + f_2 x_2 + \dots + f_m x_m.$$

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

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

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

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

Vậy tổng thu nhập cả năm

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

Thu nhập bình quân một tháng

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

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

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

Ví dụ 2.21: (Bài toán đố). Ví dụ này khi mở rộng một cách thích hợp sẽ là mô hình về sự lựa chọn thứ tự tối ưu thỏa mãn điều kiện cho trước.

Xét trò chơi trả lời 2 câu hỏi A và B; người chơi có quyền chọn câu hỏi nào để trả lời đầu tiên. Câu hỏi A được trả lời đúng với xác suất 0,8 và khi đó người chơi sẽ được thưởng \$100, câu hỏi B được trả lời đúng với xác suất 0,6 và người chơi được thưởng \$200. Nếu không trả lời đúng lần thứ nhất sẽ không được trả lời tiếp. Vậy người chơi nên chọn câu hỏi nào trả lời đầu tiên để tiền thưởng trung bình nhận được cao hơn.

Gọi X là số tiền thưởng nhận được khi người chơi chọn câu hỏi A trả lời đầu tiên thì

X	\$0	\$100	\$300
P	0,2	0,8.0,4	0,8.0,6

Chương 2: Biến ngẫu nhiên và các đặc trưng của chúng

Do đó $EX = 0,8.0 + 4.100 + 0,8.0 + 6.300 = \176

Gọi Y là số tiền thưởng nhận được khi người chơi chọn câu hỏi B trả lời đầu tiên thì

Y	\$0	\$200	\$300
P	0,4	0,6.0,2	0,6.0,8

Do đó $EY = 0,6.0 + 2.200 + 0,6.0 + 8.300 = \168

Vậy nên chọn câu hỏi A để trả lời đầu tiên để có khả năng nhận thưởng cao hơn.

2.4.1.3 Tính chất

$$1) E(C) = C \text{ với mọi hằng số } C. \quad (2.60)$$

$$2) E(CX) = CE(X) \text{ với mọi hằng số } C. \quad (2.61)$$

$$3) E(X_1 + \dots + X_n) = E(X_1) + \dots + E(X_n) \quad (2.62)$$

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

$$EY = \begin{cases} \sum_{x_i \in R_X} g(x_i) p_X(x_i) & \text{nếu } X \text{ rời rạc với } p_X(x_i) = P\{X = x_i\} \\ \int_{-\infty}^{\infty} g(x) f_X(x) dx & \text{nếu } X \text{ liên tục có hàm mật độ } f_X(x) \end{cases} \quad (2.63)$$

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

$$EX^2 = \begin{cases} \sum_{x_i \in R_X} x_i^2 p_X(x_i) & \text{nếu } X \text{ rời rạc} \\ \int_{-\infty}^{\infty} x^2 f_X(x) dx & \text{nếu } X \text{ liên tục có hàm mật độ } f_X(x) \end{cases} \quad (2.64)$$

$$5) \text{ Nếu } X_1, \dots, X_n \text{ độc lập thì } E(X_1 \dots X_n) = E(X_1) \dots E(X_n). \quad (2.65)$$

Các tính chất (2.62)-(2.65) được chứng minh trong chương 3.

Ví dụ 2.22: Chọn ngẫu nhiên 3 bi từ một túi có 6 bi đen, 4 bi trắng. Xét hai bài toán sau:

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 2.8) thì $Y = g(X) = 200X$ là một biến ngẫu nhiên rời rạc có bảng phân bố sau:

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

$$EY = 0 \cdot \frac{5}{30} + 200 \cdot \frac{15}{30} + 400 \cdot \frac{9}{30} + 600 \cdot \frac{1}{30} = 240.$$

Mặt khác, theo công thức (2.61) và ví dụ 2.18 ta cũng được $EY = 200EX = 200 \cdot \frac{6}{5} = 240$.

$$\text{b. } Z = 200X + 300(3 - X) = 900 - 100X$$

$$\Rightarrow EZ = E(900 - 100X) = 900 - 100EX = 900 - 100 \cdot \frac{6}{5} = 780\$.$$

2.4.2 Phương sai

2.4.2.1 Định nghĩa

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

Phương sai của X được ký hiệu DX hoặc $\text{Var } X$. Vậy

$$DX = E(X - EX)^2 \quad (2.66)$$

Độ lệch chuẩn của X là $\sigma_X = \sqrt{DX}$.

$$\text{Ta có} \quad (X - EX)^2 = X^2 - (2EX)X + (EX)^2$$

Áp dụng các công thức 2.60, 2.61, 2.62 của kỳ vọng ta có thể tính phương sai theo công thức sau:

$$DX = EX^2 - (EX)^2 \quad (2.67)$$

Sử dụng công thức (2.64), (2.67) ta suy ra công thức thường dùng để tính phương sai của biến ngẫu nhiên rời rạc và phương sai của biến ngẫu nhiên liên tục:

(i). Nếu X rời rạc nhận các giá trị x_i với hàm khối lượng xác suất $p_X(x_i)$ thì

$$EX^2 = \sum_{x_i \in R_X} x_i^2 p_X(x_i); \quad DX = EX^2 - (EX)^2 = \sum_{x_i \in R_X} x_i^2 p_X(x_i) - \left(\sum_{x_i \in R_X} x_i p_X(x_i) \right)^2 \quad (2.68)$$

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

$$EX^2 = \int_{-\infty}^{\infty} x^2 f_X(x) dx; \quad DX = EX^2 - (EX)^2 = \int_{-\infty}^{\infty} x^2 f_X(x) dx - \left(\int_{-\infty}^{\infty} x f_X(x) dx \right)^2 \quad (2.69)$$

Ví dụ 2.23: Tính phương sai của biến ngẫu nhiên xét trong ví dụ 2.19.

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

$$\Rightarrow DX = EX^2 - (EX)^2 = 7940 - 4 = 7936 \Rightarrow \sigma_X = \sqrt{DX} = \sqrt{7936} \approx 89,08.$$

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

Ví dụ 2.24: Tính phương sai và độ lệch chuẩn của biến ngẫu nhiên xét trong ví dụ 2.20.

$$\text{Giải: } EX^2 = \frac{3}{64} \int_0^4 x^4 (4-x) dx = \frac{3}{64} \left(\frac{4x^5}{5} - \frac{x^6}{6} \right) \Big|_0^4 = \frac{32}{5}$$

$$\Rightarrow DX = EX^2 - (EX)^2 = \frac{32}{5} - \left(\frac{12}{5}\right)^2 = \frac{16}{25} \Rightarrow \sigma_X = \frac{4}{5}.$$

Phương sai của biến ngẫu nhiên X là độ lệch bình phương trung bình quanh giá trị trung bình EX . Trong kỹ thuật phương sai đặc trưng cho mức độ phân tán của các chỉ 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 2.19 cho thấy đầu tư bảo hiểm cho những người 25 tuổi là có lãi, nhưng ví dụ 2.24 cho thấy rủi ro của bảo hiểm rất lớn.

2.4.2.2 Tính chất

$$1) D(a) = 0 \text{ với mọi hằng số } a. \quad (2.70)$$

$$2) D(aX + b) = a^2 D(X) \text{ với mọi hằng số } a, b. \quad (2.71)$$

$$3) \text{ Nếu } X_1, \dots, X_n \text{ độc lập và có các phương sai hữu hạn thì}$$

$$D(a_1 X_1 + \dots + a_n X_n) = a_1^2 D(X_1) + \dots + a_n^2 D(X_n). \quad (2.72)$$

Nói riêng: Nếu X, Y độc lập và DX, DY hữu hạn thì $D(X \pm Y) = DX + DY$.

Ví dụ 2.25: Tung con xúc xắc n lần. Tìm kỳ vọng của tổng số chấm thu được.

Giải: Gọi $X_i (i=1, \dots, n)$ là số chấm thu được ở lần tung thứ i , gọi X là tổng số chấm thu được trong n lần tung.

Như vậy $X = \sum_{i=1}^n X_i$. Theo công thức (2.62) ta có $EX = \sum_{i=1}^n EX_i$.

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

$$EX_i^2 = \frac{1}{6}(1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2) = \frac{91}{6} \Rightarrow DX_i = \frac{91}{6} - \frac{7^2}{6} = \frac{35}{12}.$$

$$\text{Vậy } DX = \sum_{i=1}^n DX_i = \frac{35}{12}n.$$

Ví dụ 2.26: Cho n biến ngẫu nhiên $X_i (i=1, \dots, n)$ độc lập trong cùng một phép thử, có các kỳ vọng bằng nhau và các phương sai bằng nhau

$$EX_1 = EX_2 = \dots = EX_n = \mu; \quad DX_1 = DX_2 = \dots = DX_n = \sigma^2$$

$$\text{Tìm kỳ vọng và phương sai của } \bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}.$$

Giải: $E \bar{X} = E \left(\frac{X_1 + X_2 + \dots + X_n}{n} \right) = \frac{E X_1 + E X_2 + \dots + E X_n}{n} = \frac{n\mu}{n} = \mu.$

$$D \bar{X} = D \left(\frac{X_1 + X_2 + \dots + X_n}{n} \right) = \frac{D X_1 + D X_2 + \dots + D X_n}{n^2} = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}.$$

2.4.3 Phân vị, Trung vị

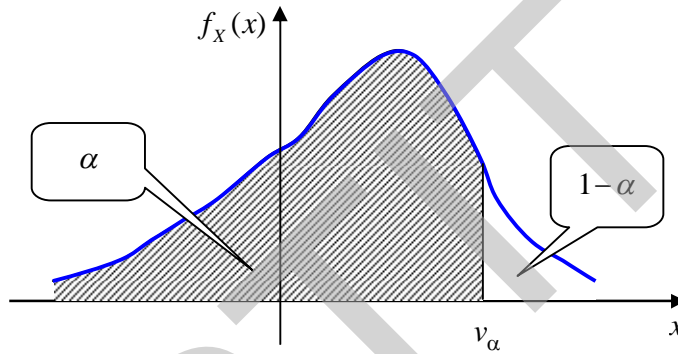
2.4.3.1 Phân vị

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

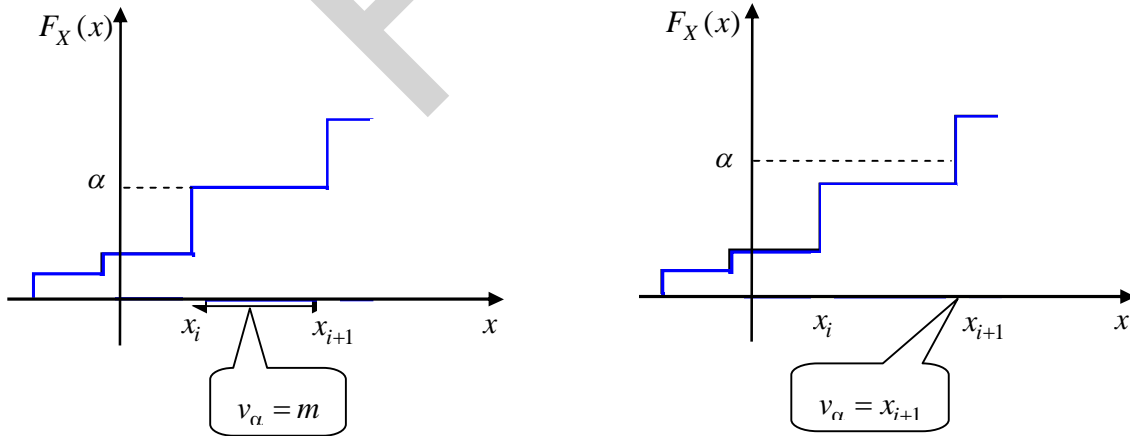
$$P\{X < v_\alpha\} \leq \alpha \leq P\{X \leq v_\alpha\}$$

Nghĩa là

$$F_X(v_\alpha -) \leq \alpha \leq F_X(v_\alpha) \quad (2.73)$$



Hình 2.17: Phân vị mức α của biến ngẫu nhiên liên tục



Hình 2.18: Phân vị mức α của biến ngẫu nhiên rời rạc

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

$$v_\alpha = F_X^{-1}(\alpha) \quad (2.74)$$

Giá trị tới hạn mức α là phân vị mức $1 - \alpha$ (hình 2.17). (2.75)

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

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

$$v_\alpha = \begin{cases} m, & \forall m \in [x_i, x_{i+1}] & \text{nếu } \alpha = F_X(x_i) \\ x_{i+1} & & \text{nếu } F_X(x_i) < \alpha < F_X(x_{i+1}) \end{cases} \quad (2.76)$$

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

2.4.3.2 Trung vị

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

2.4.4 Một

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

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

X	x_1	x_2	...
P	$p_X(x_1)$	$p_X(x_2)$...

được xác định như sau:

$$x_{i_0} = \text{Mod } X \Leftrightarrow p_X(x_{i_0}) = \max\{p_X(x_1), p_X(x_2), \dots\} \quad (2.77)$$

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

$$c = \text{Mod } X \Leftrightarrow f_X(c) = \max\{f_X(x), x \in \mathbb{R}\}. \quad (2.78)$$

Ví dụ 2.27: Biến ngẫu nhiên X ở ví dụ 2.8 có Một và trung vị $\text{Mod } X = \text{Med } X = 1$.

Ví dụ 2.28: Tìm trung vị và Một 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

$$F_X(x) = \begin{cases} 0 & \text{nếu } x < 20 \\ 0,3 & \text{nếu } 20 \leq x < 21 \\ 0,55 & \text{nếu } 21 \leq x < 22 \\ 0,73 & \text{nếu } 22 \leq x < 23 \\ 0,87 & \text{nếu } 23 \leq x < 24 \\ 1 & \text{nếu } x \geq 24 \end{cases}$$

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

Ví dụ 2.29: Tìm Med X và Mod X của biến ngẫu nhiên liên tục X xét trong ví dụ 2.6

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

$$\text{Hàm mật độ xác suất } f_X(x) = \begin{cases} 0 & \text{vì } x \leq 0 \\ 2x & \text{vì } 0 < x \leq 1 \text{ đạt cực đại tại } x = 1, \text{ vậy } \text{Mod } X = 1 \\ 0 & \text{vì } x > 1 \end{cases}$$

Ví dụ 2.30: Tìm Med X và Mod X của biến ngẫu nhiên liên tục X có hàm mật độ xác suất xác định như sau

$$f_X(x) = \begin{cases} \frac{3}{4}x(2-x) & \text{vì } 0 \leq x \leq 2 \\ 0 & \text{nếu } x < 0 \text{ hoặc } x > 2 \end{cases}$$

Giải: Hàm phân bố xác suất

$$F_X(x) = \begin{cases} 0 & \text{vì } x < 0 \\ \frac{3}{4} \left(x^2 - \frac{x^3}{3} \right) & \text{vì } 0 \leq x < 2 \\ 1 & \text{vì } x \geq 2 \end{cases}$$

Med X là nghiệm của phương trình

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

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

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

$$\frac{d}{dx} f_X(x) = \begin{cases} \frac{3}{2}(1-x) & \text{vì } 0 < x < 2 \\ 0 & \text{nếu } x < 0 \text{ hoặc } x > 2 \end{cases}$$

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

Vậy Mod $X = 1$.

- Mốt của phân bố nhị thức (xem công thức (1.20))

$$X \sim B(n; p) \text{ thì } \text{mod } X = [(n+1)p] \quad (2.79)$$

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

$$X \sim P(\lambda) \text{ thì } \text{mod } X = [\lambda]. \quad (2.80)$$

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

$$1) \text{ Moment cấp } k \quad m_k = EX^k; \quad k = 1, 2, \dots \quad (2.81)$$

$$2) \text{ Moment quy tâm cấp } k \quad \mu_k = E(X - EX)^k; \quad k = 1, 2, \dots \quad (2.82)$$

$$3) \quad \text{Hệ số bất đối xứng} \quad \alpha_3 = \frac{\mu_3}{\sigma^3} \quad \text{với } \sigma = \sqrt{DX}. \quad (2.83)$$

$$4) \quad \text{Hệ số nhọn} \quad \alpha_4 = \frac{\mu_4}{\sigma^4}. \quad (2.84)$$

Nhận xét 2.7:

- $m_1 = EX$, $\mu_1 = 0$, $\mu_2 = DX$.
- α_3 đo mức độ bất đối xứng của luật phân bố :

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

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

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

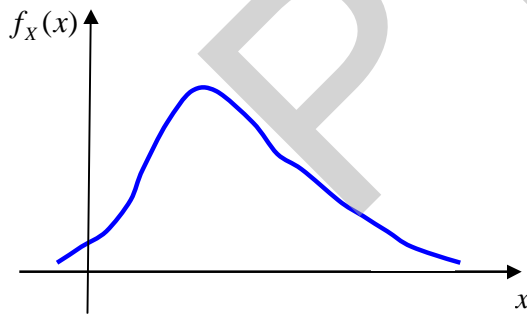
- Hệ số nhọn α_4 đặc trưng cho độ nhọn của đồ thị hàm mật độ so với đồ thị hàm mật độ của phân bố chuẩn.

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

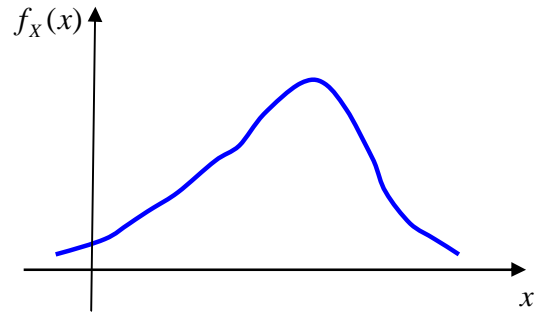
$\alpha_4 > 3$ thì đồ thị hàm mật độ sẽ nhọn hơn so với đồ thị hàm mật độ chuẩn.

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

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



Hình 2.19a: Hệ số bất đối xứng $\alpha_3 < 0$



Hình 2.19b: Hệ số bất đối xứng $\alpha_3 > 0$

2.4.6 Kỳ vọng và phương sai của các biến ngẫu nhiên có phân bố xác suất thường gặp

Phân bố xác suất của X	Hàm khối lượng $p_X(x_k)$ hoặc mật độ $f_X(x)$	Kỳ vọng EX	Phương sai DX
Bernoulli tham số p	$p_X(k) = p^k q^{1-k}; k = 0, 1$	p	pq
Nhị thức $B(n; p)$	$p_X(k) = C_n^k p^k q^{n-k}; k = 0, 1, \dots, n$	np	npq
Poisson $P(\lambda)$	$p_X(k) = e^{-\lambda} \frac{\lambda^k}{k!}; \lambda > 0; k = 0, 1, 2, \dots$	λ	λ

Chương 2: Biến ngẫu nhiên và các đặc trưng của chúng

Đều $U(a, b)$	$\frac{1}{b-a}, a < x < b$	$(a+b)/2$	$(b-a)^2/12$
Phân bố mũ tham số $\lambda > 0$	$\lambda e^{-\lambda x}, x > 0$	$1/\lambda$	$1/\lambda^2$
Phân bố chuẩn $N(\mu; \sigma^2)$	$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}; \forall x \in \mathbb{R}$	μ	σ^2
Phân bố Erlang tham số $(k; \lambda)$	$\frac{(\lambda x)^{k-1}}{(k-1)!} \lambda e^{-\lambda x}, x > 0$	k/λ	$\frac{k}{\lambda^2}$
“Khi bình phương” n bậc tự do	$\frac{(x/2)^{n/2-1}}{\Gamma(n/2)} \frac{1}{2} e^{-(x/2)}, x > 0$	n	$2n$
Student n bậc tự do, $n > 2$	$\frac{\Gamma((n+1)/2)}{\sqrt{n\pi}\Gamma(n/2)} \left(1 + \frac{x^2}{n}\right)^{-(n+1)/2}, x \in \mathbb{R}$	0	$\frac{n}{n-2}$
Gamma $\Gamma(\alpha, \lambda)$	$\frac{(\lambda x)^{\alpha-1}}{\Gamma(\alpha)} \lambda e^{-\lambda x}, x > 0$	$\frac{\alpha}{\lambda}$	$\frac{\alpha}{\lambda^2}$

TÓM TẮT

Biến ngẫu nhiên X được nghiên cứu qua hàm phân bố xác suất $F_X(x) = P\{X \leq x\}$.

- Nếu biến ngẫu nhiên X rời rạc với hàm khối lượng xác suất $p_X(x_k), x_k \in R_X$

Hàm phân bố được xác định theo công thức $F_X(x) = \sum_{x_k \leq x; x_k \in R_X} p_X(x_k)$

Ngược lại $p_X(x_k) = F_X(x_k) - F_X(x_{k-1}^-), x_k \in R_X$.

- Nếu biến ngẫu nhiên X liên tục với hàm mật độ $f_X(x)$ thì

$$F_X(x) = \int_{-\infty}^x f_X(x) dx \text{ và } f_X(x) = \frac{d}{dx} F_X(x).$$

Kỳ vọng $E X = \sum_{x_k \in R_X} x_k p_X(x_k)$ (trường hợp rời rạc)

$$E X = \int_{-\infty}^{\infty} x f_X(x) dx \quad (\text{trường hợp liên tục})$$

Phương sai $D X = E(X - E X)^2 = E X^2 - (E X)^2$, trong đó

$$E X^2 = \sum_{x_k \in R_X} x_k^2 p_X(x_k) \quad (\text{trường hợp rời rạc})$$

$$E X^2 = \int_{-\infty}^{\infty} x^2 f_X(x) dx \quad (\text{trường hợp liên tục})$$

Phân vị mức α của biến ngẫu nhiên X , ký hiệu v_α , là giá trị thỏa mãn $F_X(v_\alpha -) \leq \alpha \leq F_X(v_\alpha)$.

$$x_{i_0} = \text{Mod } X \Leftrightarrow p_X(x_{i_0}) = \max \{p_X(x_1), p_X(x_2), \dots\} \text{ (trường hợp rời rạc)}$$

$$c = \text{Mod } X \Leftrightarrow f_X(c) = \max \{f_X(x), x \in \mathbb{R}\} \text{ (trường hợp liên tục)}$$

CÂU HỎI ÔN TẬP VÀ BÀI TẬP CHƯƠNG 2

2.1 Biến ngẫu nhiên luôn luôn nhận giá trị dương.

Đúng ☐ Sai ☐.

2.2 Biến ngẫu nhiên rời rạc chỉ nhận một số hữu hạn các giá trị.

Đúng ☐ Sai ☐.

2.3 Nếu biến ngẫu nhiên X rời rạc chỉ nhận các giá trị x_1, \dots, x_n thì hệ các biến cố $\{X = x_1\}, \dots, \{X = x_n\}$ 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 $f(x) \geq 0$.

Đúng ☐ Sai ☐.

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

Đúng ☐ Sai ☐.

2.11 Biến ngẫu nhiên phân bố theo phân bố 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 Một của X đều bằng λ .

Đúng ☐ Sai ☐.

2.13 Nếu biến ngẫu nhiên X phân bố theo phân bố chuẩn $N(\mu; \sigma^2)$ thì xác suất sai lệch giữa

$$X \text{ và kỳ vọng của nó thỏa mãn } P\{|X - \mu| < \varepsilon\} = 2\Phi\left(\frac{\varepsilon}{\sigma}\right) - 1.$$

Đúng ☐ Sai ☐.

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

Đúng ☐ Sai ☐.

2.15 Biến ngẫu nhiên có phân bố Student chỉ nhận những giá trị dương.

Đúng ☐ Sai ☐.

2.16 Xác định các hằng số a, b sao cho hàm số sau là hàm phân bố của biến ngẫu nhiên X nào đó

$$F(x) = \begin{cases} 1 - ae^{-x/b} & \text{nếu } x \geq 0 \\ 0 & \text{nếu } x < 0 \end{cases}$$

2.17 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 .

2.18 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 .

2.19 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 $EX_1; EX_2; DX_1; DX_2$.

b. Tính $E(X_1 - X_2)$ và $D(X_1 - X_2)$.

2.20 Cho X_1, X_2, X_3 là ba biến ngẫu nhiên độc lập có bảng phân bố xác suất như sau:

X_1	0	2
P	0,6	0,4

X_2	1	2
P	0,4	0,6

X_3	0	2
P	0,8	0,2

Lập bảng phân bố xác suất của $\bar{X} = \frac{X_1 + X_2 + X_3}{3}$. Tính $E(\bar{X}); D(\bar{X})$.

2.21 Hai biến ngẫu nhiên X, Y độc lập. Tính $D(Z)$ với:

a. $Z = 2X + 3Y$.

b. $Z = -3X + Y$.

Cho biết $D(X) = 4, D(Y) = 5$.

2.22 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 $E(X) = 0,1$ và $D(X) = 0,89$.

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

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

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

$$f_X(x) = \begin{cases} kx^2(2-x) & \text{nếu } 0 \leq x \leq 2 \\ 0 & \text{nếu } x < 0 \text{ hoặc } x > 2 \end{cases}$$

- a. Tìm k ; Tính xác suất để côn trùng chết trước khi nó được một tháng tuổi;
- b. Tìm EX, DX .

2.25 Cho biến ngẫu nhiên liên tục X với hàm mật độ xác suất như sau

$$f_X(x) = \begin{cases} ke^{-2x} & \text{nếu } x > 0 \\ 0 & \text{nếu } x \leq 0 \end{cases}$$

- a. Tìm k ; Tính hàm phân bố xác suất $F_X(x)$;
- b. Tìm EX, DX .

2.26 Cho biến ngẫu nhiên X liên tục với hàm mật độ như sau

$$f_X(x) = \begin{cases} kx^2 & \text{nếu } 1 \leq x < 2 \\ kx & \text{nếu } 2 \leq x < 3 \\ 0 & \text{nếu } x < 1 \text{ hoặc } x \geq 3 \end{cases}$$

- a. Tìm k . Tính $P\{X > 2\}, P\{0,5 < X < 1,5\}$.
- b. Tìm hàm phân bố $F_X(x)$.

2.27 Một công ty xổ số phát hành 10.000 vé, trong đó có 200 giải 5\$, 20 giải 25\$ và 5 giải 100\$. Để công ty không bị lỗ cần phải bán với giá vé tối thiểu là bao nhiêu?

2.28 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 EX và phương sai DX .
- b. Tìm phân bố xác suất của $Y = |X|$ và kỳ vọng EY .

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

2.30 Gieo ngẫu nhiên hai con xúc xắc cân đối đồng chất, gọi X là tổng số chấm trên mặt xuất hiện của hai con xúc xắc. Tìm hàm khối lượng xác suất, hàm phân bố xác suất của X .

2.31 Xếp ngẫu nhiên 5 hành khách lên 3 toa tàu 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 .

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

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

- a. Xác định k .
- b. Tính xác suất $P\{X \geq 5\}$ và $P\{X < 3\}$.
- c. Tính kỳ vọng EX .
- d. Tính phương sai DX .
- 2.33** 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 nhận được". Lập hệ thức cho biết mối quan hệ giữa Y và X . Tính kỳ vọng EY và phương sai DY .
- 2.34** 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 .
- 2.35** 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.
- 2.36** Trong một lô hàng có 800 sản phẩm loại 1 và 200 sản phẩm loại 2. Lấy ngẫu nhiên ra 5 sản phẩm theo phương thức có hoàn lại. Gọi X là số sản phẩm loại 1 lấy được.
- a. X có phân bố gì? Viết biểu thức tổng quát của phân bố.
- b. Tìm kỳ vọng và phương sai của X .
- c. Tìm một của X và tính khả năng để xảy ra điều đó.
- 2.37** 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.
- 2.38** 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 hù họa một phương án cho mỗi câu hỏi. Tính xác suất để:
- a. Anh ta được 4 điểm.
- b. Anh ta bị điểm âm.
- 2.39** Tín hiệu thông tin được phát đi 5 lần độc lập nhau. Xác suất thu được tin của mỗi lần phát là 0,7. Tính xác suất:
- a. Thu được tín hiệu đúng 2 lần.
- a) Thu được tín hiệu nhiều nhất 1 lần.

- b) Thu được tin.
- 2.40** 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.
- 2.41** 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:
- Có ít nhất một cuộc gọi trong khoảng thời gian 10 giây.
 - Trong khoảng thời gian 3 phút có nhiều nhất ba cuộc gọi.
 - Trong khoảng thời gian 3 phút liên tiếp mỗi phút có nhiều nhất một cuộc gọi.
- 2.42** Giả sử X là biến ngẫu nhiên rời rạc với phân bố cấp số nhân tham số p có hàm khối lượng xác suất $p_X(x)$ thỏa mãn công thức (2.16).
- Nghiệm lại công thức $\sum_{k=1}^{\infty} p_X(k) = 1$. Tìm hàm phân bố $F_X(x)$.
 - Tính kỳ vọng EX , phương sai DX .
- 2.43** Cho biến ngẫu nhiên $X \sim N(0;1)$. Tính các xác suất sau:
- $P\{0 < X < 1,2\}$
 - $P\{-0,68 < X < 0\}$
 - $P\{-0,46 < X < 2,21\}$
 - $P\{0,81 < X < 1,94\}$
 - $P\{X > -1,28\}$
- 2.44** Cho biến ngẫu nhiên $X \sim N(0;1)$. Tính các giá trị a thỏa mãn các điều kiện tương ứng sau
- $P\{0 < X < a\} = 0,377$
 - $P\{X < a\} = 0,8621$
 - $P\{-1,5 < X < a\} = 0,0217$
 - $P\{a < X < -1,5\} = 0,0217$
- 2.45** Biến ngẫu nhiên X tuân theo phân bố chuẩn với kỳ vọng $\mu = 10$ và phương sai $\sigma^2 = 4$. Tính xác suất để X nhận giá trị trong khoảng $(8; 12)$.
- 2.46** Biến ngẫu nhiên X tuân theo phân bố chuẩn với kỳ vọng $\mu = 10$. Xác suất để X nhận giá trị trong khoảng $(10; 20)$ là 0,3. Tìm xác suất để X nhận giá trị trong khoảng $(0; 10)$.
- 2.47** Trọng lượng sản phẩm X do một máy tự động sản xuất là một biến ngẫu nhiên tuân theo phân bố chuẩn với $\mu = 100$ gam và độ lệch chuẩn $\sigma = 1$ gam. Sản phẩm được coi là đạt tiêu chuẩn kỹ thuật nếu trọng lượng của nó đạt từ 98 đến 102gam.
- Tìm tỷ lệ sản phẩm đạt tiêu chuẩn kỹ thuật của nhà máy.
 - Tìm tỷ lệ phế phẩm của nhà máy.
 - Giải thích bằng đồ thị kết quả tìm được ở phần a.
- 2.48** Cho X_i ($i = \overline{1, n}$) là các biến ngẫu nhiên độc lập, cùng tuân theo phân bố chuẩn với

$$E(X_1) = E(X_2) = \dots = E(X_n) = \mu; \quad D(X_1) = D(X_2) = \dots = D(X_n) = \sigma^2$$

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

2.49 Cho biến ngẫu nhiên $X \sim N(0;1)$. Tìm hàm mật độ xác suất của X^2 , từ đó có thể kết luận X^2 có phân bố “khi bình phương” với 1 bậc tự do.

2.50 Biến ngẫu nhiên liên tục X được gọi là **biến ngẫu nhiên Rayleigh** tham số σ nếu hàm mật độ có dạng

$$f_X(x) = \begin{cases} \frac{x}{\sigma^2} e^{-x^2/(2\sigma^2)} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

a. Xác định hàm phân bố $F_X(x)$

b. Tính kỳ vọng EX , phương sai DX .

2.51 Một túi chứa 4 quả cầu trắng, 3 quả cầu đen. Hai người A và B lần lượt rút một quả cầu trong túi (rút xong không trả lại). Trò chơi kết thúc khi có người rút được quả cầu đen người đó xem như thua cuộc và trả cho người kia số tiền bằng số quả cầu rút ra nhân với 5 USD. Giả sử A là người rút trước và X là tiền A thu được

a. Lập bảng phân bố xác suất của X . Tính EX .

b. Nếu chơi 150 ván thì trung bình A được bao nhiêu?

CHƯƠNG 3: VÉC TƠ NGẪU NHIÊN VÀ CÁC ĐẶC TRƯNG CỦA CHÚNG

Khái niệm biến ngẫu nhiên nhiều chiều hay còn gọi véc tơ ngẫu nhiên là một bộ có thứ tự các biến ngẫu nhiên. Mỗi biến ngẫu nhiên là một thành phần của nó.

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

Từ hàm phân bố xác suất đồng thời có thể tính được hàm phân bố xác suất của các biến ngẫu nhiên thành phần. Cũng vậy, từ bảng phân bố xác suất đồng thời của biến ngẫu nhiên hai chiều rời rạc có thể tìm được bảng phân bố xác suất của hai biến ngẫu nhiên thành phần và từ hàm mật độ xác suất đồng thời của véc tơ ngẫu nhiên liên tục có thể tìm được hàm mật độ xác suất của các biến ngẫu nhiên thành phần.

Ngoài các đặc trưng kỳ vọng, phương sai của hai biến ngẫu nhiên thành phần, biến ngẫu nhiên nhiều chiều còn được đặc trưng bởi hiệp phương sai và hệ số tương quan. Hệ số tương quan đo mức độ phụ thuộc tuyến tính của hai biến ngẫu nhiên thành phần, khi hệ số tương quan càng gần 1 thì mức độ phụ thuộc tuyến tính càng chặt. Hai biến ngẫu nhiên thành phần không tương quan thì hệ số tương quan bằng 0.

Áp dụng công thức xác suất có điều kiện ta xây dựng phân bố xác suất có điều kiện. Từ đó có thể tính kỳ vọng có điều kiện của biến ngẫu nhiên thành phần này với điều kiện biến ngẫu nhiên thành phần kia nhận giá trị cụ thể nào đó và xây dựng hàm hồi quy tương quan.

Để học tốt chương này học viên cần nắm vững các tính chất cơ bản của xác suất, xác suất có điều kiện và phân bố xác suất của biến ngẫu nhiên.

3.1 KHÁI NIỆM VÉC TƠ NGẪU NHIÊN

3.1.1 Khái niệm và phân loại véc tơ ngẫu nhiên

Trong chương trước ta xét các biến ngẫu nhiên mà giá trị chúng nhận được có thể biểu diễn bằng giá trị số, đó là các biến ngẫu nhiên một chiều. Tuy nhiên trong thực tế có thể gặp các biến ngẫu nhiên mà giá trị nhận được là các 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.

Định nghĩa 3.1: Một véc tơ ngẫu nhiên n chiều là một bộ có thứ tự (X_1, X_2, \dots, X_n) với các thành phần X_1, X_2, \dots, X_n là các biến ngẫu nhiên xác định trong cùng một phép thử.

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 thì ta có biến ngẫu nhiên ba chiều. Ngoài ra nếu quan tâm thêm trọng lượng của sản phẩm thì ta được biến ngẫu nhiên bốn chiều ...

Véc tơ ngẫu nhiên n chiều (X_1, X_2, \dots, X_n) là **liên tục** hay **rời rạc** nếu tất cả các biến ngẫu nhiên thành phần X_1, X_2, \dots, X_n là liên tục hay rời rạc. Mặc dù tồn tại những véc tơ ngẫu nhiên có một thành phần rời rạc và một số thành phần liên tục, tuy nhiên giáo trình này không xét trường hợp đó.

3.1.2 Hàm phân bố xác suất đồng thời và hàm phân bố xác suất biên

Định nghĩa 3.2: Hàm n biến $F_{X_1, \dots, X_n}(x_1, x_2, \dots, x_n)$ xác định bởi:

$$F_{X_1, \dots, X_n}(x_1, x_2, \dots, x_n) = P\{X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n\} \quad (3.1)$$

trong đó $\{X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n\}$ là biến cố tích $\{X_1 \leq x_1\} \cap \{X_2 \leq x_2\} \cap \dots \cap \{X_n \leq x_n\}$,

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

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

$$1. \quad 0 \leq F_{X_1, \dots, X_n}(x_1, \dots, x_n) \leq 1. \quad (3.2)$$

$$2. \quad \lim_{x_k \rightarrow -\infty} F_{X_1, \dots, X_n}(x_1, \dots, x_n) = 0, \text{ với } k \text{ nào đó thuộc } \{1, \dots, n\}. \quad (3.3)$$

$$3. \quad \lim_{(x_1, \dots, x_n) \rightarrow (\infty, \dots, \infty)} F_{X_1, \dots, X_n}(x_1, \dots, x_n) = 1. \quad (3.4)$$

$$4. \quad F_{X_1, \dots, X_n}(x_1, \dots, x_n) \text{ không giảm theo từng biến.} \quad (3.5)$$

$$5. \quad \lim_{x_1 \rightarrow \infty} F_{X_1, \dots, X_n}(x_1, \dots, x_n) = F_{X_2, \dots, X_n}(x_2, \dots, x_n).$$

Như vậy nếu lấy giới hạn của hàm phân bố xác suất đồng thời của n biến ngẫu nhiên X_1, X_2, \dots, X_n khi biến x_1 tiến đến vô cùng thì được hàm phân bố xác suất đồng thời của $n-1$ biến ngẫu nhiên X_2, \dots, X_n .

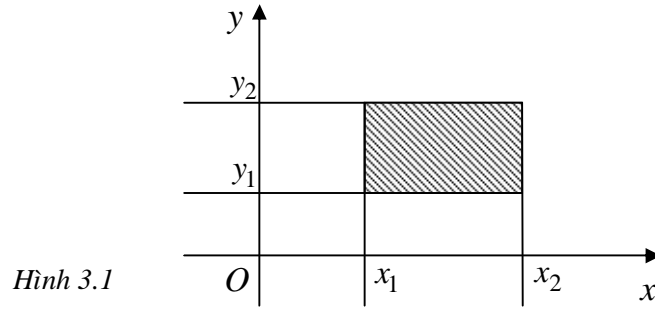
Tương tự nếu lấy giới hạn của hàm phân bố xác suất đồng thời của X_1, X_2, \dots, X_n khi biến x_k tiến đến vô cùng, với k nào đó thuộc $\{1, 2, \dots, n\}$, thì được hàm phân bố xác suất đồng thời của $n-1$ biến ngẫu nhiên còn lại $X_1, \dots, X_{k-1}, X_{k+1}, \dots, X_n$.

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

$$\lim_{y \rightarrow \infty} F_{X,Y}(x, y) = P\{X \leq x\} = F_X(x); \quad \lim_{x \rightarrow \infty} F_{X,Y}(x, y) = P\{Y \leq y\} = F_Y(y) \quad (3.6)$$

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

$$7. \quad P\{x_1 < X \leq x_2; y_1 < Y \leq y_2\} = F_{X,Y}(x_2, y_2) - F_{X,Y}(x_1, y_2) - F_{X,Y}(x_2, y_1) + F_{X,Y}(x_1, y_1) \quad (3.7)$$



Hình 3.1

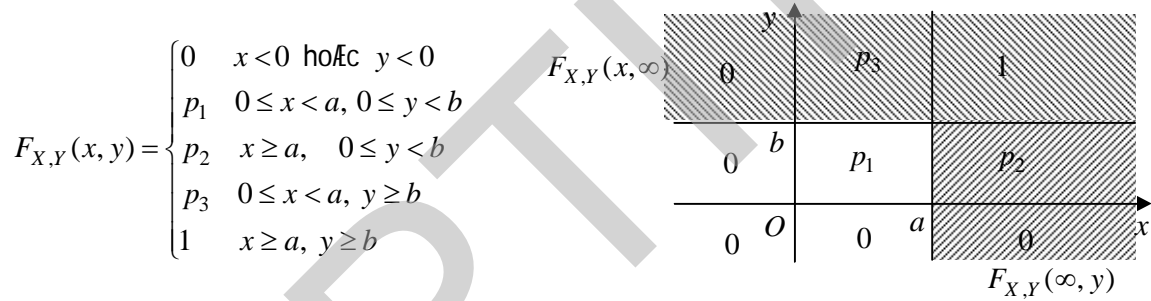
Ví dụ 3.2: Hàm số $F(x, y) = \begin{cases} 1 - e^{-(x+y)} & 0 \leq x < \infty, 0 \leq y < \infty \\ 0 & \text{nếu ngược lại} \end{cases}$

Xét $x_2 = y_2 = 2$ và $x_1 = y_1 = 1$, ta có

$$\begin{aligned} F(2, 2) - F(1, 2) - F(2, 1) + F(1, 1) &= (1 - e^{-4}) - (1 - e^{-3}) - (1 - e^{-3}) + (1 - e^{-2}) \\ &= -e^{-4} + 2e^{-3} - e^{-2} = -(e^{-2} - e^{-1})^2 < 0. \end{aligned}$$

Vậy $F(x, y)$ không thể là hàm phân bố của véc tơ ngẫu nhiên vì không thỏa mãn (3.7).

Ví dụ 3.3: Hàm phân bố xác suất của véc tơ ngẫu nhiên (X, Y) xác định như sau



Hình 3.2

Có hai hàm phân bố xác suất biên

$$F_X(x) = F_{X,Y}(x, \infty) = \begin{cases} 0 & x < 0 \\ p_3 & 0 \leq x < a \\ 1 & x \geq a \end{cases} \quad F_Y(y) = F_{X,Y}(\infty, y) = \begin{cases} 0 & y < 0 \\ p_2 & 0 \leq y < b \\ 1 & y \geq b \end{cases}$$

Ví dụ 3.4: Cho hàm phân bố xác suất của véc tơ ngẫu nhiên (X, Y) xác định như sau

$$F_{X,Y}(x, y) = \begin{cases} (1 - e^{-\alpha x})(1 - e^{-\beta y}) & x \geq 0, y \geq 0; \alpha, \beta > 0 \\ 0 & \text{nếu ngược lại} \end{cases}$$

a. Tìm hai hàm phân bố xác suất biên $F_X(x)$, $F_Y(y)$.

b. Tính $P\{X \leq 1, Y \leq 1\}$, $P\{X \leq 1\}$, $P\{Y > 1\}$ và $P\{X > x, Y > y\}$.

Giải: a. $F_X(x) = F_{X,Y}(x, \infty) = \begin{cases} 1 - e^{-\alpha x} & x \geq 0 \\ 0 & x < 0 \end{cases}$; $F_Y(y) = F_{X,Y}(\infty, y) = \begin{cases} 1 - e^{-\beta y} & y \geq 0 \\ 0 & y < 0 \end{cases}$.

b. $P\{X \leq 1, Y \leq 1\} = F_{X,Y}(1, 1) = (1 - e^{-\alpha})(1 - e^{-\beta})$;

$$P\{X \leq 1\} = F_X(1) = 1 - e^{-\alpha}; P\{Y > 1\} = 1 - P\{Y \leq 1\} = 1 - F_Y(1) = e^{-\beta};$$

Áp dụng luật De Morgan ta có

$$\overline{\{X > x\} \cap \{Y > y\}} = \overline{\{X > x\}} \cup \overline{\{Y > y\}} = \{X \leq x\} \cup \{Y \leq y\}$$

Do đó

$$\begin{aligned} P(\overline{\{X > x\} \cap \{Y > y\}}) &= P(\{X \leq x\} \cup \{Y \leq y\}) = P\{X \leq x\} + P\{Y \leq y\} - P\{X \leq x; Y \leq y\} \\ &= F_X(x) + F_Y(y) - F_{X,Y}(x, y) \\ &= (1 - e^{-\alpha x}) + (1 - e^{-\beta y}) - (1 - e^{-\alpha x})(1 - e^{-\beta y}) = 1 - e^{-\alpha x} e^{-\beta y}. \end{aligned}$$

$$\text{Vậy } P\{X > x, Y > y\} = 1 - P(\overline{\{X > x\} \cap \{Y > y\}}) = e^{-\alpha x} e^{-\beta y}.$$

3.2 VÉC TƠ NGẪU NHIÊN RỜI RẠC

Tương tự trường hợp biến ngẫu nhiên rời rạc, quy luật phân bố xác suất của véc tơ ngẫu nhiên rời rạc được xác định thông qua hàm khối lượng xác suất đồng thời. Đặc biệt quy luật phân bố xác suất của véc tơ ngẫu nhiên rời rạc 2 chiều có thể được xác định thông qua bảng phân bố xác suất đồng thời hoặc hàm khối lượng xác suất đồng thời.

3.2.1 Hàm khối lượng xác suất đồng thời và bảng phân bố xác suất đồng thời

Hàm khối lượng xác suất đồng thời của biến ngẫu nhiên hai chiều (X, Y) ký hiệu và xác định bởi:

$$p_{X,Y}(x_i, y_j) = P\{X = x_i, Y = y_j\} = P(\{X = x_i\} \cap \{Y = y_j\}), \quad (3.8)$$

Hàm khối lượng xác suất đồng thời thỏa mãn điều kiện

$$\begin{cases} p_{X,Y}(x_i, y_j) \geq 0, \forall i = 1, \dots, n, j = 1, \dots, m \\ \sum_{i=1}^n \sum_{j=1}^m p_{X,Y}(x_i, y_j) = 1 \end{cases} \quad (3.9)$$

Hàm phân bố xác suất đồng thời có thể xác lập từ hàm khối lượng xác suất đồng thời

$$F_{X,Y}(x, y) = \sum_{x_i \leq x} \sum_{y_j \leq y} p_{X,Y}(x_i, y_j); (x_i, y_j) \in R_{X,Y} \quad (3.10)$$

Để trực quan hơn chúng ta biểu diễn hàm khối lượng xác suất đồng thời của biến ngẫu nhiên rời rạc hai chiều thông qua bảng phân bố xác suất đồng thời.

Bảng phân bố xác suất đồng thời của biến ngẫu nhiên rời rạc hai chiều (X, Y) là bảng liệt kê tất cả các giá trị của X theo cột, các giá trị của Y theo hàng và các xác suất tương ứng có dạng sau, trong đó x_i ($i = 1, \dots, n$) là các giá trị có thể có của thành phần X ; y_j ($j = 1, \dots, m$) là các giá trị có thể có của thành phần Y .

$$R_{X,Y} = \{(x_i, y_j) | i = 1, \dots, n; j = 1, \dots, m\} \text{ gọi là } \textbf{miền giá trị của véc tơ ngẫu nhiên } (X, Y).$$

$X \backslash Y$	y_1	y_2	...	y_j	...	y_m
x_1	$p_{XY}(x_1, y_1)$	$p_{XY}(x_1, y_2)$...	$p_{XY}(x_1, y_j)$...	$p_{XY}(x_1, y_m)$
x_2	$p_{XY}(x_2, y_1)$	$p_{XY}(x_2, y_2)$...	$p_{XY}(x_2, y_j)$...	$p_{XY}(x_2, y_m)$
\vdots	\vdots	\vdots	...	\vdots	...	\vdots
x_i	$p_{XY}(x_i, y_1)$	$p_{XY}(x_i, y_2)$...	$p_{XY}(x_i, y_j)$...	$p_{XY}(x_i, y_m)$
\vdots	\vdots	\vdots	...	\vdots	...	\vdots
x_n	$p_{XY}(x_n, y_1)$	$p_{XY}(x_n, y_2)$...	$p_{XY}(x_n, y_j)$...	$p_{XY}(x_n, y_m)$

3.2.2 Bảng phân bố xác suất biên

Áp dụng công thức xác suất đầy đủ (1.19) cho hệ $\{X = x_1\}, \{X = x_2\}, \dots, \{X = x_n\}$ (xem công thức 2.14) ta có:

$$p_Y(y_j) = P\{Y = y_j\} = \sum_{i=1}^n P\{X = x_i, Y = y_j\} = \sum_{i=1}^n p_{X,Y}(x_i, y_j) ; \quad j=1, \dots, m \quad (3.11)$$

$$p_X(x_i) = P\{X = x_i\} = \sum_{j=1}^m P\{X = x_i, Y = y_j\} = \sum_{j=1}^m p_{X,Y}(x_i, y_j) ; \quad i=1, \dots, n \quad (3.12)$$

Nhận xét 3.1:

1. Từ công thức (3.11) và (3.12), ta thấy rằng nếu ta cộng các xác suất của bảng phân bố xác suất đồng thời của (X, Y) theo cột thì ta được các xác suất tương ứng với các giá trị của Y , 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 .

$X \backslash Y$	y_1	y_2	...	y_j	...	y_m	\sum_j
x_1	$p_{XY}(x_1, y_1)$	$p_{XY}(x_1, y_2)$...	$p_{XY}(x_1, y_j)$...	$p_{XY}(x_1, y_m)$	$p_X(x_1)$
x_2	$p_{XY}(x_2, y_1)$	$p_{XY}(x_2, y_2)$...	$p_{XY}(x_2, y_j)$...	$p_{XY}(x_2, y_m)$	$p_X(x_2)$
\vdots	\vdots	\vdots	...	\vdots	...	\vdots	\vdots
x_i	$p_{XY}(x_i, y_1)$	$p_{XY}(x_i, y_2)$...	$p_{XY}(x_i, y_j)$...	$p_{XY}(x_i, y_m)$	$p_X(x_i)$
\vdots	\vdots	\vdots	...	\vdots	...	\vdots	\vdots
x_n	$p_{XY}(x_n, y_1)$	$p_{XY}(x_n, y_2)$...	$p_{XY}(x_n, y_j)$...	$p_{XY}(x_n, y_m)$	$p_X(x_n)$
\sum_i	$p_Y(y_1)$	$p_Y(y_2)$...	$p_Y(y_j)$...	$p_Y(y_m)$	1

Chương 3: Véc tơ ngẫu nhiên và các đặc trưng của chúng

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

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

2. Hàm khối lượng xác suất của véc tơ ngẫu nhiên rời rạc n chiều (X_1, X_2, \dots, X_n)

$$p_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) = P\{X_1 = x_1, X_2 = x_2, \dots, X_n = x_n\}. \quad (3.13)$$

Hàm phân bố xác suất đồng thời có thể xác lập từ hàm khối lượng xác suất đồng thời

$$F_{X_1, X_2, \dots, X_n}(y_1, y_2, \dots, y_n) = \sum_{x_1 \leq y_1} \dots \sum_{x_n \leq y_n} p_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n); (x_1, x_2, \dots, x_n) \in R_{X_1, X_2, \dots, X_n}. \quad (3.14)$$

Ví dụ 3.5: Gieo 3 đồng tiền cân đối A, B, C. Gọi X là số mặt ngửa xuất hiện của 2 đồng tiền A, B và Y là số mặt ngửa xuất hiện của cả 3 đồng tiền A, B, C. Hãy lập bảng phân bố xác suất đồng thời của X, Y .

Giải: Chúng ta có bảng liệt kê 8 kết quả đồng khả năng khi gieo 3 đồng tiền cân đối và tính các giá trị của X, Y tương ứng, trong đó N là ký hiệu mặt ngửa xuất hiện còn S là mặt sấp.

A	N	N	N	N	S	S	S	S
B	N	N	S	S	N	N	S	S
C	N	S	N	S	N	S	N	S
X	2	2	1	1	1	1	0	0
Y	3	2	2	1	2	1	1	0

Sử dụng công thức tính xác suất cổ điển (1.1a) ta có:

$$P\{X = 2, Y = 3\} = \frac{1}{8}; \quad P\{X = 2, Y = 2\} = \frac{1}{8}; \quad P\{X = 1, Y = 2\} = \frac{2}{8} \dots$$

Vậy bảng phân bố xác suất đồng thời của X và Y là

$Y \backslash X$	0	1	2	3
0	1/8	1/8	0	0
1	0	2/8	2/8	0
2	0	0	1/8	1/8

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 3.6: 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 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:

$Y \backslash X$	1	2	3
1	2/36	3/36	1/36
2	4/36	6/36	2/36
3	6/36	9/36	3/36

Ví dụ 3.7: (Phân bố đa thức, multinomial) Véc tơ ngẫu nhiên k chiều $X = (X_1, X_2, \dots, X_k)$ được gọi là có phân bố đa thức với các tham số $(n; p_1, \dots, p_k)$ ký hiệu $X \sim \text{MUT}(n; p_1, \dots, p_k)$ nếu hàm khối lượng xác suất đồng thời có dạng:

$$P_{X_1 X_2 \dots X_k}(m_1, m_2, \dots, m_k) = P\{X_1 = m_1, \dots, X_k = m_k\} = \frac{n!}{m_1! m_2! \dots m_k!} p_1^{m_1} p_2^{m_2} \dots p_k^{m_k} \quad (3.15)$$

trong đó $0 \leq m_i \leq n, i = 1, \dots, k; m_1 + \dots + m_k = n; p_i > 0, i = 1, \dots, k; p_1 + \dots + p_k = 1$

Trường hợp $k = 2$: $(X_1, X_2) \sim \text{MUT}(n; p, 1-p)$ thì $X_1 \sim \text{B}(n; p)$ và $X_2 \sim \text{B}(n; 1-p)$.

Xét n phép thử độc lập, thuần nhất và mỗi phép thử có k kết quả ngẫu nhiên A_1, \dots, A_k tạo thành hệ đầy đủ biến cố; giả sử xác suất xuất hiện kết quả thứ A_i là $p_i; p_1 + \dots + p_k = 1$.

Gọi X_i là số lần xuất hiện của biến cố A_i trong n phép thử thì $X = (X_1, X_2, \dots, X_k)$ có phân bố đa thức $X \sim \text{MUT}(n; p_1, \dots, p_k)$.

Chẳng hạn, gieo một con xúc xắc cân đối 10 lần. Tính các xác suất:

- 1) Có đúng 3 lần xuất hiện mặt 5 chấm (biến cố A).
- 2) Có 2 lần xuất hiện mặt 1 chấm, 4 lần mặt 3 chấm, 1 lần mặt 4 chấm và 3 lần mặt 6 chấm (biến cố B).

Giải: 1) Xét phép thử Bernoulli với thành công của mỗi lần thử là xuất hiện mặt có 5 chấm, vậy xác suất thành công mỗi lần thử là $1/6$. Gọi X là số lần xuất hiện mặt 5 trong 10 lần thử thì X có phân bố nhị thức tham số $(10; 1/6)$, do đó

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

- 2) Gọi X_i là số lần xuất hiện mặt i chấm trong 10 phép thử thì $(X_1, X_2, X_3, X_4, X_5, X_6)$ có phân bố đa thức $\text{MUT}(10; 1/6, 1/6, 1/6, 1/6, 1/6, 1/6)$.

$$P(B) = P\{X_1 = 2, X_2 = 0, X_3 = 4, X_4 = 1, X_5 = 0, X_6 = 3\} = \frac{10!}{2!0!4!1!0!3!} \left(\frac{1}{6}\right)^{10} \approx 0,0002.$$

3.3 VÉC TƠ NGẪU NHIÊN LIÊN TỤC

3.3.1 Hàm mật độ xác suất đồng thời

Hàm phân bố xác suất của véc tơ ngẫu nhiên rời rạc có thể được xác định qua hàm khối lượng xác suất đồng thời, công thức (3.10), (3.14). Tương tự hàm phân bố xác suất của véc tơ ngẫu nhiên liên tục có thể được xác định qua hàm mật độ xác suất đồng thời; trong đó hàm khối lượng xác suất đồng thời của véc tơ ngẫu nhiên rời rạc được thay bằng hàm mật độ xác suất đồng thời của véc tơ ngẫu nhiên liên tục, dấu tổng của công thức 3.14- xác định hàm phân bố xác suất đồng thời chuyển thành dấu tích phân.

Định nghĩa 3.3: Hàm mật độ xác suất của véc tơ ngẫu nhiên liên tục (X_1, X_2, \dots, X_n) là hàm n biến $f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) \geq 0$ thỏa mãn:

$$F_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} f_{X_1, X_2, \dots, X_n}(t_1, t_2, \dots, t_n) dt_1 dt_2 \dots dt_n \quad (3.16)$$

$f_{X_1, \dots, X_n}(x_1, \dots, x_n)$ còn được gọi là hàm mật độ xác suất đồng thời của các biến ngẫu nhiên X_1, X_2, \dots, X_n .

Tính chất 3.1: 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 $f_{XY}(x, y)$.

$$1) \quad f_{X,Y}(x, y) \geq 0 \quad \text{với mọi } (x, y) \quad \text{và} \quad \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x, y) dx dy = 1. \quad (3.17)$$

$$2) \quad P\{(X, Y) \in A\} = \iint_{(x,y) \in A \cap R_{X,Y}} f_{X,Y}(x, y) dx dy \quad \text{với mọi } A \subset \mathbb{R}^2, \quad (3.18)$$

Trong đó $R_{X,Y}$ là miền giá trị của (X,Y) .

$$3) \quad f_{X,Y}(x,y) = \begin{cases} \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x,y) & \text{nếu tần suất đồng thời } (x,y) \\ 0 & \text{nếu ngược lại} \end{cases} \quad (3.19)$$

3.3.2 Hàm mật độ xác suất biên

Từ bảng phân bố xác suất đồng thời của véc tơ ngẫu nhiên rời rạc (X,Y) , nếu ta cộng giá trị của hàm khối lượng xác suất theo cột tức là cộng tất cả các giá trị của hàm khối lượng xác suất theo mọi giá trị của X thì ta được giá trị của hàm khối lượng xác suất với giá trị tương ứng của Y , tương tự nếu ta cộng các giá trị của hàm khối lượng xác suất theo hàng ta được giá trị của hàm khối lượng xác suất với giá trị tương ứng của X .

Trường hợp véc tơ ngẫu nhiên (X,Y) liên tục thay vì cộng ta thực hiện lấy tích phân của hàm mật độ đồng thời $f_{X,Y}(x,y)$ theo biến y ta được hàm mật độ $f_X(x)$ của biến ngẫu nhiên thành phần X và lấy tích phân của hàm mật độ đồng thời $f_{X,Y}(x,y)$ theo biến x ta được hàm mật độ $f_Y(y)$ của biến ngẫu nhiên thành phần Y , cụ thể

$$\int_{-\infty}^{\infty} f_{X,Y}(x,y) dy = f_X(x) \text{ hàm mật độ xác suất của biến ngẫu nhiên } X. \quad (3.20)$$

$$\int_{-\infty}^{\infty} f_{X,Y}(x,y) dx = f_Y(y) \text{ hàm mật độ xác suất của biến ngẫu nhiên } Y. \quad (3.21)$$

Ví dụ 3.8: Cho véc tơ ngẫu nhiên (X,Y) có hàm mật độ xác suất xác định như sau:

$$f_{X,Y}(x,y) = \begin{cases} k & \text{nếu } 0 < y \leq x < 1 \\ 0 & \text{nếu ngược lại} \end{cases}.$$

- Tìm k .
- Tìm các hàm mật độ xác suất của các biến ngẫu nhiên thành phần X, Y .
- Tìm xác suất $P\{0 < X \leq 1/2; 0 < Y \leq 1/2\}$.

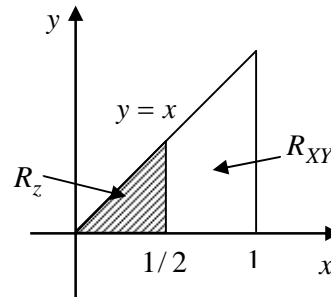
Giải: a. Miền giá trị của X, Y là tam giác R_{XY} (xem hình 3.1).

$$\text{Do đó } 1 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,y) dx dy = k dt(R_{X,Y}) = k \frac{1}{2} \Rightarrow k = 2$$

Trong đó $dt(R_{X,Y})$ là diện tích của miền giá trị $R_{X,Y}$

$$\text{b. } f_X(x) = \begin{cases} \int_0^x 2 dy = 2x & \text{nếu } 0 < x < 1 \\ 0 & \text{nếu ngược lại} \end{cases}$$

$$f_Y(y) = \begin{cases} \int_y^1 2 dx = 2(1-y) & \text{nếu } 0 < y < 1 \\ 0 & \text{nếu ngược lại} \end{cases}$$



Hình 3.2

$$c. P\{0 < X \leq 1/2; 0 < Y \leq 1/2\} = P\{0 < X \leq 1/2; 0 < Y < X\} = \iint_{R_z} f_{X,Y}(x, y) dx dy = 2 \left(\frac{1}{8} \right) = \frac{1}{4}.$$

Ví dụ 3.9: Cho véc tơ ngẫu nhiên (X, Y) có hàm mật độ xác suất xác định như sau:

$$f_{X,Y}(x, y) = \begin{cases} k & \text{nếu } |x| + |y| \leq 1 \\ 0 & \text{nếu ngược lại} \end{cases}$$

a. Tìm k .

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

Giải: a. Miền $D: |x| + |y| \leq 1$ đối xứng qua hai trục tọa độ Ox, Oy .

Phần của D nằm trong góc phần tư thứ nhất là tam giác vuông cân $0 \leq x, 0 \leq y; x + y \leq 1$.

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

$$1 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x, y) dx dy = k \cdot \text{kt} D = 2k \Rightarrow k = \frac{1}{2}.$$

$$b. |x| + |y| \leq 1 \Leftrightarrow |x| - 1 \leq y \leq 1 - |x|$$

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy = \begin{cases} \frac{1}{2} \int_{|x|-1}^{1-|x|} dy & \text{nếu } -1 < x < 1 \\ 0 & \text{nếu ngược lại} \end{cases} = \begin{cases} 1 - |x| & \text{nếu } |x| < 1 \\ 0 & \text{nếu } |x| \geq 1 \end{cases}$$

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

$$f_Y(y) = \begin{cases} 1 - |y| & \text{nếu } |y| < 1 \\ 0 & \text{nếu } |y| \geq 1 \end{cases}$$

Ví dụ 3.10: Chọn ngẫu nhiên một điểm từ hình tròn tâm O bán kính R . Đặt X, Y là hoành độ và tung độ của điểm được chọn thì (X, Y) là véc tơ ngẫu nhiên liên tục có hàm mật độ

$$f_{X,Y}(x, y) = \begin{cases} k & x^2 + y^2 \leq R^2 \\ 0 & x^2 + y^2 > R^2 \end{cases}$$

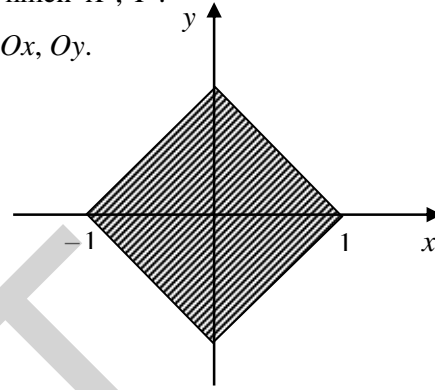
a. Xác định k .

b. Tìm các hàm mật độ thành phần $f_X(x), f_Y(y)$.

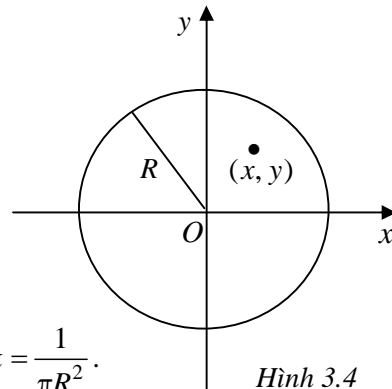
c. Tìm xác suất để khoảng cách từ gốc O đến điểm được chọn không lớn hơn a , với $0 \leq a \leq R$.

$$\text{Giải: a. } 1 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x, y) dx dy = k \iint_{x^2 + y^2 \leq R^2} dx dy = k \pi R^2 \Rightarrow k = \frac{1}{\pi R^2}.$$

$$b. \text{ Khi } x^2 \leq R^2: f_X(x) = \frac{1}{\pi R^2} \int_{-\sqrt{R^2 - x^2}}^{\sqrt{R^2 - x^2}} dy = \frac{2}{\pi R^2} \sqrt{R^2 - x^2}.$$



Hình 3.3



Hình 3.4

$$\text{Vậy} \quad f_X(x) = \begin{cases} \frac{2}{\pi R^2} \sqrt{R^2 - x^2} & |x| \leq R \\ 0 & |x| > R \end{cases}$$

$$\text{Tương tự} \quad f_Y(y) = \begin{cases} \frac{2}{\pi R^2} \sqrt{R^2 - y^2} & |y| \leq R \\ 0 & |y| > R \end{cases}.$$

c. Với $0 \leq a \leq R$:

$$P\{X^2 + Y^2 \leq a^2\} = \iint_{x^2 + y^2 \leq a^2} f_{X,Y}(x, y) dx dy = \frac{1}{\pi R^2} \iint_{x^2 + y^2 \leq a^2} dx dy = \frac{\pi a^2}{\pi R^2} = \frac{a^2}{R^2}.$$

3.4 TÍNH ĐỘC LẬP CỦA CÁC BIẾN NGẪU NHIÊN

Từ định nghĩa 2.2 về tính độc lập của hai biến ngẫu nhiên ta có thể xây dựng các dấu hiệu sau đây để nhận biết tính độc lập của chúng thông qua tính chất của hàm phân bố xác suất đồng thời hoặc hàm khối lượng xác suất đồng thời đối với hai biến ngẫu nhiên rời rạc hoặc hàm mật độ xác suất đồng thời đối với hai biến ngẫu nhiên liên tục.

Giả sử X, Y là hai biến ngẫu nhiên độc lập, khi đó

$$F_{X,Y}(x, y) = P(\{X \leq x\} \cap \{Y \leq y\}) = P\{X \leq x\} P\{Y \leq y\} = F_X(x) F_Y(y).$$

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

$$F_{X,Y}(x, y) = F_X(x) F_Y(y) \quad (3.22)$$

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

Ví dụ 3.11: Hàm phân bố xác suất đồng thời của (X, Y) trong ví dụ 3.3 thỏa mãn (3.22) do đó độc lập.

Trường hợp véc tơ ngẫu nhiên (X, Y) rời rạc thì điều kiện (3.22) của hàm phân bố được thay bởi điều kiện của hàm khối lượng xác suất như sau.

Định lý 3.2: Véc tơ ngẫu nhiên rời rạc hai chiều (X, Y) có miền giá trị $R_X = \{x_1, x_2, \dots, x_n\}$ và $R_Y = \{y_1, y_2, \dots, y_m\}$ với hàm khối lượng xác suất đồng thời xác định theo công thức (3.8). Khi đó X và Y là độc lập khi và chỉ khi

$$p_{X,Y}(x_i, y_j) = p_X(x_i) p_Y(y_j); \quad \forall i = 1, \dots, n, j = 1, \dots, m. \quad (3.23)$$

Tuy nhiên kiểm tra trực tiếp điều kiện (3.23) khá phức tạp và người ta thường sử dụng dấu hiệu sau để nhận biết hai biến ngẫu nhiên rời rạc độc lập: đó là bảng phân bố xác suất đồng thời có tính chất:

- Hai hàng bất kỳ (các khối lượng xác suất trong bảng) tỉ lệ với nhau.
- Hai cột bất kỳ tỉ lệ với nhau.

(3.24)

Ví dụ 3.12: Y là số mặt ngửa xuất hiện của cả 3 đồng tiền A, B, C phụ thuộc số mặt ngửa X của 2 đồng tiền A, B (xét trong ví dụ 3.4), do đó X và Y không độc lập. Mặt khác ta cũng thấy các hàng của bảng phân bố xác suất đồng thời không tỉ lệ hoặc có thể kiểm tra như sau:

$$P\{X = 2\} = \frac{2}{8}, P\{Y = 1\} = \frac{3}{8} \text{ nhưng } P\{X = 2, Y = 1\} = 0 \neq P\{X = 2\} P\{Y = 1\}.$$

Từ ý nghĩa thực tế của phép thử suy ra hai biến ngẫu nhiên X và Y của ví dụ 3.5 độc lập. Ta cũng thấy bảng phân bố xác suất đồng thời thỏa mãn điều kiện (3.23), (3.24). Cụ thể các hàng của bảng phân bố xác suất đồng thời tỉ lệ nhau theo tỉ lệ 1:2:3.

Trường hợp véc tơ ngẫu nhiên (X, Y) liên tục, bằng cách lấy đạo hàm hai vế của (3.22) (xem công thức 3.19) ta được:

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

$$f_{X,Y}(x, y) = f_X(x)f_Y(y) \quad (3.25)$$

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

Ví dụ 3.13: Véc tơ ngẫu nhiên (X, Y) của ví dụ 3.7, ví dụ 3.8, ví dụ 3.9 không độc lập vì hàm mật độ xác suất $f_{X,Y}(x, y) \neq f_X(x)f_Y(y)$.

Ví dụ 3.14: Véc tơ ngẫu nhiên (X, Y) có hàm mật độ:

$$f_{X,Y}(x, y) = \begin{cases} 4xy & \text{nếu } 0 < x < 1, 0 < y < 1 \\ 0 & \text{nếu ngược lại} \end{cases}$$

Có hai hàm mật độ thành phần

$$f_X(x) = \begin{cases} 2x & \text{nếu } 0 < x < 1 \\ 0 & \text{nếu ngược lại} \end{cases}; f_Y(y) = \begin{cases} 2y & \text{nếu } 0 < y < 1 \\ 0 & \text{nếu ngược lại} \end{cases}$$

Ta có $f_{X,Y}(x, y) = f_X(x)f_Y(y)$. Vậy X và Y độc lập.

3.5 CÁC THAM SỐ ĐẶC TRƯNG CỦA VÉC TƠ NGẪU NHIÊN

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

Từ hàm khối lượng xác suất thành phần (3.11), (3.12) hoặc hàm mật độ xác suất thành phần (3.20), (3.21) 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

$$E X = \sum_{i=1}^n x_i p_X(x_i) = \sum_{i=1}^n \sum_{j=1}^m x_i p_{X,Y}(x_i, y_j) \quad (3.26)$$

$$E Y = \sum_{j=1}^m y_j p_Y(y_j) = \sum_{j=1}^m \sum_{i=1}^n y_j p_{X,Y}(x_i, y_j) \quad (3.27)$$

$$E X^2 = \sum_{i=1}^n x_i^2 p_X(x_i); D X = E(X^2) - (E X)^2 \quad (3.28)$$

$$E Y^2 = \sum_{j=1}^m y_j^2 p_Y(y_j); D Y = E(Y^2) - (E Y)^2 \quad (3.29)$$

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

$$E X = \int_{-\infty}^{\infty} x f_X(x) dx = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f_{X,Y}(x, y) dx dy \quad (3.30)$$

$$E Y = \int_{-\infty}^{\infty} y f_Y(y) dy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y f_{X,Y}(x, y) dx dy \quad (3.31)$$

$$E X^2 = \int_{-\infty}^{\infty} x^2 f_X(x) dx = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^2 f_{X,Y}(x, y) dx dy ; D X = E(X^2) - (E X)^2 \quad (3.32)$$

$$E Y^2 = \int_{-\infty}^{\infty} y^2 f_Y(y) dy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y^2 f_{X,Y}(x, y) dy dx ; D Y = E(Y^2) - (E Y)^2 \quad (3.33)$$

3.5.2 Hiệp phương sai

Định nghĩa 3.4: Hiệp phương sai (hay còn gọi là Covariance) của hai biến ngẫu nhiên X, Y , ký hiệu $\text{cov}(X, Y)$, là kỳ vọng toán của tích các sai lệch của hai biến ngẫu nhiên đó với kỳ vọng toán của chúng:

$$\text{cov}(X, Y) = E[(X - E X)(Y - E Y)] \quad (3.34)$$

Khai triển về phải và áp dụng tính chất của kỳ vọng ta được

$$\text{cov}(X, Y) = E(XY) - (E X)(E Y) \quad (3.35)$$

Nếu X, Y rời rạc có hàm khối lượng xác suất đồng thời theo công thức (3.8) thì

$$E(XY) = \sum_{j=1}^m \sum_{i=1}^n x_i y_j p_{X,Y}(x_i, y_j) \quad (3.36)$$

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

$$E(XY) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f_{X,Y}(x, y) dx dy \quad (3.37)$$

Tính chất 3.2: Từ định nghĩa ta có thể chứng minh được các tính chất sau

$$1) \text{cov}(X, Y) = \text{cov}(Y, X). \quad (3.38)$$

$$2) \text{cov}(X, X) = D X. \quad (3.39)$$

$$3) \text{cov}(aX + c, bY + d) = ab \text{cov}(X, Y) \text{ với mọi hằng số } a, b, c, d. \quad (3.40)$$

4) Từ công thức (3.35) và (2.65) suy ra rằng, nếu X, Y độc lập thì $\text{cov}(X, Y) = 0$. Tuy nhiên điều ngược lại không đúng, nghĩa là tồn tại hai biến ngẫu nhiên X, Y không độc lập nhưng $\text{cov}(X, Y) = 0$.

3.5.3 Ma trận hiệp phương sai

Định nghĩa 3.5: Cho véc tơ ngẫu nhiên $X = (X_1, X_2, \dots, X_n)$.

$$\text{Ma trận } M = [C_{ij}]_{n \times n} \text{ với } C_{ij} = \text{cov}(X_i, X_j) \quad (3.41)$$

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

Tính chất 3.3:

- 1) Ma trận hiệp phương sai là ma trận đối xứng.
- 2) Với mọi $t_1, t_2, \dots, t_n \in \mathbb{R}^n$ luôn có $\sum_j \sum_i C_{ij} t_i t_j \geq 0$.

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.

3.5.4 Hệ số tương quan

Định nghĩa 3.6: Hệ số tương quan của hai biến ngẫu nhiên X, Y với phương sai khác 0 được ký hiệu và định nghĩa bởi công thức:

$$\rho_{X,Y} = \frac{\text{cov}(X,Y)}{\sqrt{D X} \sqrt{D Y}}. \quad (3.42)$$

Tính chất 3.4:

$$1) -1 \leq \rho_{X,Y} \leq 1 \text{ với mọi } X, Y. \quad (3.43)$$

2) Nếu X, Y độc lập thì $\rho_{X,Y} = 0$, nhưng điều ngược lại chưa chắc đúng.

3) Từ công thức (3.37) ta suy ra: với mọi hằng số a, b, c, d

$$\rho_{aX+c, bY+d} = \begin{cases} \rho_{X,Y} & \text{nếu } ab > 0 \\ -\rho_{X,Y} & \text{nếu } ab < 0 \end{cases} \quad (3.44)$$

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

$$\rho_{X,Y} = \begin{cases} 1 & \text{nếu } a > 0 \\ -1 & \text{nếu } a < 0 \end{cases} \quad (3.45)$$

Nhận xét 3.2: Ý nghĩa của hệ số tương quan

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

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

Như vậy hai biến ngẫu nhiên độc lập thì không tương quan, nhưng ngược lại chưa chắc đúng (xem tính chất 3.2-4).

Ví dụ sau đây minh họa điều đó.

Ví dụ 3.15: Xét véc tơ ngẫu nhiên (X, Y) có bảng phân bố xác suất đồng thời

$X \backslash Y$	-1	0	1
-1	4/15	1/15	4/15
0	1/15	2/15	1/15
1	0	2/15	0

Có bảng phân bố xác suất biên

X	-1	0	1
P	9/15	4/15	2/15

Y	-1	0	1
P	1/3	1/3	1/3

$$E X = -1 \cdot \frac{9}{15} + 0 \cdot \frac{4}{15} + 1 \cdot \frac{2}{15} = -\frac{7}{15};$$

$$E X^2 = (-1)^2 \cdot \frac{9}{15} + 0^2 \cdot \frac{4}{15} + 1^2 \cdot \frac{2}{15} = \frac{11}{15} \Rightarrow D X = E X^2 - (E X)^2 = \frac{116}{225}$$

$$E Y = (-1) \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} = 0; E Y^2 = \frac{2}{3} \Rightarrow D Y = \frac{2}{3}.$$

$$E(XY) = (-1) \cdot (-1) \cdot \frac{4}{15} + (-1) \cdot 0 \cdot \frac{1}{15} + (-1) \cdot 1 \cdot \frac{4}{15} \\ + 0 \cdot (-1) \cdot \frac{1}{15} + 0 \cdot 0 \cdot \frac{2}{15} + 0 \cdot 1 \cdot \frac{1}{15} + 1 \cdot (-1) \cdot 0 + 1 \cdot 0 \cdot \frac{2}{15} + 1 \cdot 1 \cdot 0 = 0$$

Hiệp phương sai $\text{cov}(X, Y) = E(XY) - (E X)(E Y) = 0 - \frac{-7}{15} \cdot 0 = 0.$

Hệ số tương quan $\rho_{X,Y} = \frac{\text{cov}(X,Y)}{\sqrt{D X} \sqrt{D Y}} = \frac{0}{\sqrt{(116/225)(2/3)}} = 0.$

Ma trận hiệp phương sai $M = \begin{bmatrix} 116/225 & 0 \\ 0 & 2/3 \end{bmatrix}.$

Vì các hàng của bảng phân bố xác suất đồng thời không tỉ lệ nên hai biến ngẫu nhiên X, Y không độc lập, mặc dù hiệp phương sai $\text{cov}(X, Y) = 0$ và $\rho_{X,Y} = 0.$

Ví dụ 3.16: Xét kênh viễn thông nhị phân được cho trong ví dụ 1.31 chương 1. Ký hiệu X là đầu vào và Y đầu ra của kênh. (X, Y) là véc tơ ngẫu nhiên thỏa mãn

$$P\{X = 0\} = 0,5; P\{Y = 1 | X = 0\} = 0,1 \text{ và } P\{Y = 0 | X = 1\} = 0,2.$$

- Tìm bảng phân bố xác suất đồng thời của (X, Y) .
- Tìm bảng phân bố xác suất thành phần X và Y . X và Y có độc lập không?
- Tính $E X$, $E Y$; $D X$, $D Y$
- Tìm hiệp phương sai, hệ số tương quan và ma trận hiệp phương sai.

Giải: a.

$Y \backslash X$	0	1	Σ
0	0,45	0,05	0,5
1	0,10	0,40	0,5
Σ	0,55	0,45	1

b. Bảng phân bố xác suất thành phần X và Y

X	0	1
P	0,5	0,5

Y	0	1
P	0,55	0,45

Hai hàng của bảng phân bố xác suất đồng thời của (X, Y) không tỉ lệ, do đó X, Y không độc lập.

Ta cũng thấy rằng $P\{X=0; Y=0\} = 0,45 \neq P\{X=0\}P\{Y=0\} = 0,55 \cdot 0,5 = 0,275$.

c. Áp dụng các công thức (3.26)-(3.29) ta có

$$EX = 0,5, EX^2 = 0,5 \Rightarrow DX = 0,25$$

$$EY = 0,45, EY^2 = 0,45 \Rightarrow DY = 0,2475.$$

d. Áp dụng công thức (3.35), (3.36), (3.41), (3.42) ta được

Hiệp phương sai $E(XY) = 0,4 \Rightarrow \text{cov}(X, Y) = 0,4 - 0,5 \times 0,45 = 0,175.$

Hệ số tương quan $\rho_{X,Y} = \frac{\text{cov}(X,Y)}{\sqrt{DX}\sqrt{DY}} = \frac{0,175}{\sqrt{0,25 \times 0,2475}} = 0,704.$

Ma trận hiệp phương sai $M = \begin{bmatrix} 0,25 & 0,175 \\ 0,175 & 0,2475 \end{bmatrix}.$

Ta thấy giá trị $\rho_{X,Y} = 0,704$ khá xa 1, do đó Y không phụ thuộc tuyến tính đối với X .

3.6 PHÂN BỐ CÓ ĐIỀU KIỆN VÀ KỲ VỌNG CÓ ĐIỀU KIỆN

3.6.1 Phân bố có điều kiện và kỳ vọng có điều kiện của biến ngẫu nhiên rời rạc

Định nghĩa 3.7: Giả sử X biến ngẫu nhiên rời rạc có miền giá trị $R_X = \{x_1, x_2, \dots, x_n\}$, B là một biến cố trong cùng phép thử với X và có xác suất $P(B) > 0$. Khi đó bảng phân bố xác suất của X với điều kiện B được xác định như sau

$X B$	x_1	x_2	...	x_i	...	x_n
P	$p_{X B}(x_1 B)$	$p_{X B}(x_2 B)$...	$p_{X B}(x_i B)$...	$p_{X B}(x_n B)$

trong đó

$$p_{X|B}(x_i|B) = \frac{P(\{X = x_i\} \cap B)}{P(B)}; i = 1, \dots, n. \quad (3.46)$$

Hàm $p_{X|B}(x_i|B)$ xác định bởi công thức (3.46) được gọi là **hàm khối lượng xác suất của biến ngẫu nhiên X với điều kiện B** .

Định nghĩa 3.8: Giả sử X, Y có tập các giá trị $R_X = \{x_1, x_2, \dots, x_n\}$ và $R_Y = \{y_1, y_2, \dots, y_m\}$. Với mỗi $y_j \in R_Y$ ta có bảng phân bố xác suất có điều kiện của X với điều kiện biến cố $\{Y = y_j\}$:

Chương 3: Véc tơ ngẫu nhiên và các đặc trưng của chúng

$X Y = y_j$	x_1	x_2	...	x_i	...	x_n
P	$p_{X Y}(x_1 y_j)$	$p_{X Y}(x_2 y_j)$...	$p_{X Y}(x_i y_j)$...	$p_{X Y}(x_n y_j)$

trong đó

$$p_{X|Y}(x_i|y_j) = \frac{p_{X,Y}(x_i, y_j)}{p_Y(y_j)}; i = 1, \dots, n. \quad (3.47)$$

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

$Y X = x_i$	y_1	y_2	...	y_j	...	y_m
P	$p_{Y X}(y_1 x_i)$	$p_{Y X}(y_2 x_i)$...	$p_{Y X}(y_j x_i)$...	$p_{Y X}(y_m x_i)$

trong đó

$$p_{Y|X}(y_j|x_i) = \frac{p_{X,Y}(x_i, y_j)}{p_X(x_i)}; j = 1, \dots, m. \quad (3.48)$$

Tính chất 3.5:

- $0 \leq p_{X|Y}(x_i|y_j) \leq 1; \forall i = 1, \dots, n, j = 1, \dots, m.$
- Với mỗi $j: \sum_{i=1}^n p_{X|Y}(x_i|y_j) = 1.$
- Nếu X, Y độc lập thì $p_{X|Y}(x_i|y_j) = p_X(x_i)$ và $p_{Y|X}(y_j|x_i) = p_Y(y_j).$

Ví dụ 3.17: Cho X, Y là hai biến ngẫu nhiên rời rạc có phân bố xác suất đồng thời

$\begin{matrix} & Y \\ X \end{matrix}$	0,5	2	4	Σ
0,1	0,07	0,12	0,08	0,27
0,2	0,11	0,18	0,19	0,48
0,4	0,05	0,13	0,07	0,25

- Tìm bảng phân bố xác suất của các thành phần X và Y .
- Tìm bảng phân bố xác suất của Y với điều kiện $X = 0,2$.

Giải: a. Cộng xác suất của bảng phân bố xác suất đồng thời ta được bảng phân bố xác suất thành phần của X và Y

X	0,1	0,2	0,4
P	0,27	0,48	0,25

Y	0,5	2	4
P	0,23	0,43	0,34

- Bảng phân bố xác suất của Y với điều kiện $X = 0,2$

$Y X = 0, 2$	0,5	2	4
P	11/48	18/48	19/48

Ví dụ 3.18: Thực hiện lặp lại cùng một phép thử Bernoulli với xác suất xuất hiện của biến cố A trong mỗi lần thử là p , $0 < p < 1$. Gọi Y là biến ngẫu nhiên chỉ lần thử đầu tiên xuất hiện biến cố A . Gọi B là biến cố: “Trong n lần thử đầu tiên có duy nhất một lần xuất hiện biến cố A ”.

- Tìm bảng phân bố xác suất của Y .
- Tìm phân bố của Y với điều kiện B .

Giải: a. Ta có bảng phân bố xác suất của Y (xem ví dụ 2.9)

Y	1	2	...	k	...
P	p	qp	...	$q^{k-1}p$...

với $q = 1 - p$.

- Phân bố của Y với điều kiện B :

$$\text{Khi } P(B) > 0 \text{ thì } P(\{Y = k\} | B) = \frac{P(\{Y = k\} \cap B)}{P(B)}.$$

♦ Rõ ràng khi $k > n$ thì biến cố $\{Y = k\}$ kéo theo trong n phép thử đầu tiên biến cố A không xuất hiện. Do đó $\{Y = k\} \cap B = \emptyset$, vậy $P(\{Y = k\} \cap B) = 0$.

♦ Khi $k \leq n$, áp dụng công thức Bernoulli ta có: $P(B) = C_n^1 pq^{n-1} = npq^{n-1}$.

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

Vậy hàm khối lượng xác suất của Y với điều kiện B có dạng

$$P_{X|B}(k | B) = P(\{Y = k\} | B) = \begin{cases} \frac{1}{n} & \text{nếu } 1 \leq k \leq n \\ 0 & \text{nếu } k > n. \end{cases}$$

3.6.2 Phân bố có điều kiện và kỳ vọng có điều kiện của biến ngẫu nhiên liên tục

Định nghĩa 3.9: Xét biến ngẫu nhiên X và biến cố B trong cùng một phép thử, thỏa mãn điều kiện $P(B) > 0$. Hàm phân bố xác suất của X với điều kiện B được định nghĩa và ký hiệu như sau

$$F_{X|B}(x | B) = P(\{X \leq x\} | B) = \frac{P(\{X \leq x\} \cap B)}{P(B)}. \quad (3.49)$$

Hàm mật độ xác suất của X với điều kiện B được định nghĩa và ký hiệu như sau

$$f_{X|B}(x | B) = \frac{dF_{X|B}(x | B)}{dx}. \quad (3.50)$$

Định nghĩa 3.10: Giả sử $f_{X,Y}(x, y)$ là hàm mật độ xác suất đồng thời của hai biến ngẫu nhiên liên tục X, Y và $f_X(x)$ là hàm mật độ của biến ngẫu nhiên thành phần X . Hàm phân bố xác suất của Y với điều kiện $\{X = x\}$ được định nghĩa và ký hiệu như sau:

$$F_{Y|X}(y|x) = \int_{-\infty}^y \frac{f_{X,Y}(x, v)}{f_X(x)} dv, \text{ nếu } f_X(x) > 0. \quad (3.51)$$

Biến ngẫu nhiên Y với điều kiện biến cố $\{X = x\}$ đôi khi được viết dưới dạng “với điều kiện $X = x$ ”.

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).

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x, y)}{f_X(x)} \text{ với mọi } (x, y) \in R_{X,Y} \text{ và thỏa mãn } f_X(x) > 0. \quad (3.52)$$

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

$$F_{X|Y}(x|y) = \int_{-\infty}^x \frac{f_{X,Y}(u, y)}{f_Y(y)} du, \text{ nếu } f_Y(y) > 0. \quad (3.53)$$

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)} \text{ với mọi } (x, y) \in R_{X,Y} \text{ và thỏa mãn } f_Y(y) > 0. \quad (3.54)$$

Tính chất 3.6:

- a. $f_{X|Y}(x|y) \geq 0$.
- b. Với mỗi y thỏa mãn $f_Y(y) > 0$ thì có $\int_{-\infty}^{\infty} f_{X|Y}(x|y) dx = 1$.
- c. Trường hợp X, Y độc lập ta có $f_{X|Y}(x|y) = f_X(x)$ và $f_{Y|X}(y|x) = f_Y(y)$.

Ví dụ 3.19: Giả sử X, Y là hai biến ngẫu nhiên liên tục có hàm mật độ xác suất đồng thời đã cho trong ví dụ 3.9.

Hàm mật độ xác suất đồng thời:

$$f_{X,Y}(x, y) = \begin{cases} \frac{1}{2} & \text{nếu } |x| + |y| \leq 1 \\ 0 & \text{nếu ngược lại} \end{cases}.$$

Hàm mật độ xác suất thành phần Y

$$f_Y(y) = \begin{cases} 1 - |y| & \text{nếu } |y| < 1 \\ 0 & \text{nếu } |y| \geq 1 \end{cases}$$

Theo công thức (3.54) suy ra hàm mật độ xác suất có điều kiện của X với $\{Y = y\}$

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

3.6.3 Kỳ vọng có điều kiện

Định nghĩa 3.11: Kỳ vọng của X với điều kiện B được định nghĩa:

$$E[X|B] = \sum_{i=1}^n x_i p_{X|B}(x_i|B) \quad (3.55)$$

Định nghĩa 3.12: 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 được tính theo công thức sau

$$E[X|Y = y_j] = \sum_{i=1}^n x_i p_{X|Y}(x_i|y_j); \quad E[Y|X = x_i] = \sum_{j=1}^m y_j p_{Y|X}(y_j|x_i). \quad (3.56)$$

Ví dụ 3.20: Thống kê dân cư của một thành phố nọ ở độ tuổi trưởng thành về thu nhập hàng tháng X và lứa tuổi Y thu được kết quả trong bảng sau.

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

Tìm thu nhập trung bình theo lứa tuổi.

Giải: Thu nhập trung bình theo lứa tuổi là kỳ vọng có điều kiện của X theo Y .

Với $Y = 30$ bảng phân bố xác suất điều kiện tương ứng:

$X Y = 30$	1	2	3	4
P	$\frac{0,02}{0,37}$	$\frac{0,06}{0,37}$	$\frac{0,21}{0,37}$	$\frac{0,08}{0,37}$

$$\text{Từ đó } E[X|Y = 30] = 1 \cdot \frac{2}{37} + 2 \cdot \frac{6}{37} + 3 \cdot \frac{21}{37} + 4 \cdot \frac{8}{37} = \frac{109}{37} = 2,9459.$$

$$\text{Tương tự } E[X|Y = 45] = \frac{89}{29} = 3,069. \quad E[X|Y = 70] = \frac{86}{34} = 2,5294.$$

Vậy thu nhập trung bình ở độ tuổi 30 là 2.945.900đ/tháng, độ tuổi 45 là 3.069.000đ/tháng và độ tuổi 70 là 2.529.400 đ/tháng.

Định nghĩa 3.13: 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

$$E[Y|X = x] = \int_{-\infty}^{\infty} y f_{Y|X}(y|x) dy \quad (3.57)$$

Kỳ vọng của X với điều kiện $\{Y = y\}$ được ký hiệu và định nghĩa

$$E[X|Y = y] = \int_{-\infty}^{\infty} x f_{X|Y}(x|y) dy \quad (3.58)$$

$E[Y|X = x]$ xác định ở công thức (3.57), (3.58) là một hàm của x , được gọi là **hàm hồi qui của Y đối với X** .

$E[X|Y = y]$ là một hàm của y , được gọi là **hàm hồi qui của X đối với Y** .

Ví dụ 3.21: Giả sử X, Y là hai biến ngẫu nhiên liên tục có hàm mật độ xác suất đồng thời đã cho trong ví dụ 3.7

$$f_{X,Y}(x,y) = \begin{cases} 2 & \text{nếu } 0 < y \leq x < 1 \\ 0 & \text{nếu ngược lại} \end{cases}$$

Tìm hàm mật độ có điều kiện $f_{X|Y}(x|y)$, $f_{Y|X}(y|x)$ và kỳ vọng điều kiện.

Giải: Theo kết quả của ví dụ 3.7 ta có

$$f_X(x) = 2x \quad \text{nếu } 0 < x < 1$$

$$f_Y(y) = 2(1-y) \quad \text{nếu } 0 < y < 1$$

Áp dụng công thức (3.57), (3.58) ta được

$$f_{Y|X}(y|x) = \frac{1}{x} \quad 0 < y \leq x < 1, 0 < x < 1$$

$$f_{X|Y}(x|y) = \frac{1}{1-y} \quad 0 < y \leq x < 1, 0 < y < 1.$$

Kỳ vọng của Y với điều kiện $\{X = x\}$

$$E[Y|X = x] = \int_{-\infty}^{\infty} y f_{Y|X}(y|x) dy = \int_0^x y \left(\frac{1}{x} \right) dy = \frac{y^2}{2x} \Big|_0^x = \frac{x}{2} \quad \text{nếu } 0 < x < 1.$$

Kỳ vọng của X với điều kiện $\{Y = y\}$

$$E[X|Y = y] = \int_{-\infty}^{\infty} x f_{X|Y}(x|y) dx = \int_y^1 x \left(\frac{1}{1-y} \right) dx = \frac{x^2}{2(1-y)} \Big|_{x=y}^{x=1} = \frac{1+y}{2} \quad \text{nếu } 0 < y < 1.$$

3.7 LUẬT SỐ LỚN VÀ ĐỊNH LÝ GIỚI HẠN

Trong mục này ta nghiên cứu sự hội tụ của dãy các biến ngẫu nhiên theo xác suất và sự hội tụ của dãy các biến ngẫu nhiên theo phân bố xác suất

Với sự hội tụ theo xác suất ta có luật số lớn và với sự hội tụ theo phân bố ta có định lý giới hạn trung tâm.

3.7.1 Hội tụ theo xác suất và hội tụ theo phân bố của dãy biến ngẫu nhiên

Định nghĩa 3.14: Xét dãy biến ngẫu nhiên $\{X_n\}_{n=1}^{\infty}$ và biến ngẫu nhiên X trong cùng một phép thử. Ta nói rằng dãy các biến ngẫu nhiên $\{X_n\}_{n=1}^{\infty}$ hội tụ theo xác suất về biến ngẫu nhiên X , ký hiệu $X_n \xrightarrow[n \rightarrow \infty]{P} X$, nếu:

$$\forall \varepsilon > 0 \quad \lim_{n \rightarrow \infty} P\{|X_n - X| > \varepsilon\} = 0. \quad (3.59)$$

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

Định nghĩa 3.15: Dãy các biến ngẫu nhiên $\{X_n\}_{n=1}^{\infty}$ được 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 $F_X(x)$. Tức là với mọi $x \in \mathbb{R}$:

$$\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x). \quad (3.60)$$

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

$$\lim_{n \rightarrow \infty} P\{X_n = c_k\} = P\{X = c_k\}. \quad (3.61)$$

3.7.2 Luật số lớn

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, Trêbusép, Markov, Liapunốp mở rộng.

Trong mục này ta xét hai định lý về luật số lớn. Định lý Trêbusép là dạng tổng quát của luật số lớn và định lý Bernoulli là trường hợp đơn giản nhất của luật số lớn áp dụng cho các biến ngẫu nhiên có phân bố Bernoulli. Để chứng minh định lý Trêbusép ta sử dụng bất đẳng thức Trêbusép.

3.7.2.1 Bất đẳng thức Markov

Định lý 3.4: Cho Y là biến ngẫu nhiên không âm có kỳ vọng hữu hạn. Khi đó với mọi $a > 0$ ta có bất đẳng thức:

$$P\{Y \geq a\} \leq \frac{EY}{a}. \quad (3.62)$$

Chứng minh:

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

Đặt $R_1 = \{y_i \in R_Y, y_i < a\}$; $R_2 = \{y_i \in R_Y, y_i \geq a\}$.

$$EY = \sum_{y_i \in R_Y} y_i P\{Y = y_i\} \quad (*)$$

- Trường hợp tổng (*) có hữu hạn các số hạng thì đương nhiên có thể thay đổi thứ tự lấy tổng
- Trường hợp tổng (*) có vô hạn số hạng thì đây là tổng của một chuỗi số dương do đó cũng có thể thay đổi thứ tự, vì vậy có thể viết lại:

$$\begin{aligned} EY &= \sum_{y_i \in R_Y} y_i P\{Y = y_i\} = \sum_{y_i \in R_1} y_i P\{Y = y_i\} + \sum_{y_i \in R_2} y_i P\{Y = y_i\} \\ &\geq \sum_{y_i \in R_2} y_i P\{Y = y_i\} \geq a \sum_{y_i \in R_2} P\{Y = y_i\} = aP\{Y \geq a\}. \end{aligned}$$

Suy ra $P\{Y \geq a\} \leq \frac{EY}{a}$.

b) Giả sử Y liên tục có hàm mật độ xác suất $f_Y(y)$. Y là biến ngẫu nhiên không âm do đó $f_Y(y) = 0, \forall y \leq 0$, vì vậy

$$EY = \int_0^{+\infty} y f_Y(y) dy = \int_0^a y f_Y(y) dy + \int_a^{+\infty} y f_Y(y) dy \geq \int_a^{+\infty} y f_Y(y) dy \geq a \int_a^{+\infty} f_Y(y) dy = aP\{Y \geq a\}.$$

Suy ra $P\{Y \geq a\} \leq \frac{EY}{a}$.

3.7.2.2 Bất đẳng thức Trêbusép

Định lý 3.5: Giả sử X là biến ngẫu nhiên có kỳ vọng và phương sai hữu hạn, khi đó với mọi $\varepsilon > 0$ ta có:

$$P\{|X - EX| > \varepsilon\} \leq \frac{DX}{\varepsilon^2}. \quad (3.63)$$

Sử dụng công thức xác suất biến cố đối ta cũng có

$$P\{|X - EX| \leq \varepsilon\} \geq 1 - \frac{DX}{\varepsilon^2}. \quad (3.64)$$

Chứng minh: Áp dụng công thức (3.62) cho biến ngẫu nhiên $Y = (X - EX)^2$ và $a = \varepsilon^2$ ta có:

$$P\{|X - EX| > \varepsilon\} = P\{Y > \varepsilon^2\} \leq \frac{EY}{\varepsilon^2} = \frac{E(X - EX)^2}{\varepsilon^2} = \frac{DX}{\varepsilon^2}.$$

Từ (3.63), áp dụng quy tắc xác suất biến cố đối ta được bất đẳng thức (3.64).

Bất đẳng thức (3.63)-(3.64) được gọi là bất đẳng thức Trêbusép.

Bất đẳng thức Trêbusép có nhiều ứng dụng. Trước hết nó cho phép ta đánh giá cận trên hoặc cận dưới của xác suất để biến ngẫu nhiên X nhận giá trị sai lệch so với kỳ vọng EX không quá ε . Bất đẳng thức Trêbusép 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 3.22: 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 số 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, \dots, X_n là các biến ngẫu nhiên độc lập có phân bố đều trên đoạn $[-0,5; 0,5]$. Khi đó $EX_i = 0, DX_i = \frac{1}{12}$.

Sai số tổng cộng trong cả tháng là $S = X_1 + \dots + X_n$ (trong đó n là số khách hàng mua hàng trong tháng). Ta có:

$$ES = \sum_{i=1}^n EX_i = 0, DS = \sum_{i=1}^n DX_i = \frac{n}{12}.$$

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

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

Giả sử có $n = 10^4$ khách hàng trong tháng. Để xác suất $P\{|S| > \varepsilon\}$ bé hơn 0,01 ta phải có

$$\frac{n}{12\varepsilon^2} \leq 0,01 \text{ hay } \varepsilon \geq \sqrt{\frac{n}{12 \cdot 0,01}} = 288,67.$$

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

3.7.2.3 Luật số lớn Trêbusép

Định lý 3.6: Giả sử X_1, X_2, \dots là dãy các biến ngẫu nhiên độc lập trong cùng một phép thử, có các kỳ vọng hữu hạn và phương sai đều bị chặn trên bởi hằng số C ($DX_k \leq C; \forall k = 1, 2, \dots$). Khi đó với mọi $\varepsilon > 0$:

$$\lim_{n \rightarrow \infty} P\left(\left|\frac{X_1 + \dots + X_n}{n} - \frac{EX_1 + \dots + EX_n}{n}\right| > \varepsilon\right) = 0 \quad (3.65)$$

Chứng minh: Xét biến ngẫu nhiên $S_n = \frac{X_1 + \dots + X_n}{n}$. Từ giả thiết độc lập của dãy các biến ngẫu

nhien X_1, X_2, \dots ta suy ra $ES_n = \frac{EX_1 + \dots + EX_n}{n}$; $DS_n = \frac{DX_1 + \dots + DX_n}{n^2} \leq \frac{C}{n}$.

Áp dụng bất đẳng thức Trêbusép (3.65) cho biến ngẫu nhiên S_n ta có:

$$P\left(\left|\frac{X_1 + \dots + X_n}{n} - \frac{EX_1 + \dots + EX_n}{n}\right| > \varepsilon\right) \leq \frac{C}{n\varepsilon^2} \xrightarrow{n \rightarrow \infty} 0.$$

Hệ quả 3.1: Giả sử X_1, X_2, \dots là dãy các biến ngẫu nhiên độc lập trong cùng một phép thử có cùng kỳ vọng μ và phương sai đều bị chặn trên bởi hằng số C ($DX_k \leq C; \forall k = 1, 2, \dots$).

Khi đó

$$\frac{X_1 + \dots + X_n}{n} \xrightarrow[n \rightarrow \infty]{P} \mu \quad (3.66)$$

Hệ quả 3.2: Giả sử X_1, X_2, \dots là dãy các biến ngẫu nhiên độc lập trong cùng một phép thử có cùng phân bố, có kỳ vọng μ và phương sai σ^2 . Khi đó

$$\frac{X_1 + \dots + X_n}{n} \xrightarrow[n \rightarrow \infty]{P} \mu \quad (3.67)$$

Định lý Trêbusép 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 có sự ổn định của trung bình số học của một số lớn các biến ngẫu nhiên xung quanh trung bình số học của các kỳ vọng của các biến ngẫu nhiên ấy. Như vậy mặc dù từng biến ngẫu nhiên độc lập có thể nhận giá trị khác nhiều so với kỳ vọng của chúng, song trung bình số học của một số lớn các biến ngẫu nhiên lại nhận giá trị gần bằng trung bình số học các kỳ vọng của chúng với xác suất rất lớn. Điều đó cho phép dự đoán giá trị trung bình số học của các biến ngẫu nhiên. Chẳng hạn, gieo một con xúc xắc cân đối. Giả sử X là số nốt xuất hiện ở mặt trên con xúc xắc. Ta có $E X = 3,5$. 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

$$\frac{x_1 + \dots + x_{10^6}}{10^6} \approx 3,500867 \approx 3,5.$$

Định lý Trêbusép 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, \dots, 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 Trêbusép 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 Trêbusép còn là cơ sở cho lý thuyết mẫu ứng dụng trong thống kê.

3.7.2.4 Luật số lớn Bernoulli

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

Định lý 3.7: (Định lý Bernoulli) Tần suất f_n hội tụ theo xác suất về xác suất p của biến cố A , nghĩa là với mọi $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} P\{|f_n - p| \leq \varepsilon\} = 1 \quad (3.68)$$

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

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

Như vậy đây các biến ngẫu nhiên X_1, X_2, \dots, X_n độc lập có cùng phân bố Bernoulli tham số p (công thức (2.9)). $EX_k = p, DX_k = p(1-p)$.

Ta có
$$\frac{X_1 + \dots + X_n}{n} = \frac{k_n}{n} = f_n.$$

Vậy theo hệ quả 3-2 của Định lý 3.6 suy ra f_n hội tụ theo xác suất về p .

Định lý Bernoulli chỉ ra rằng tần suất xuất hiện của biến cố trong n phép thử độc lập sẽ hội tụ theo xác suất về xác suất của biến cố đó khi số lần thử tăng lên vô hạn. Chính vì vậy định lý Bernoulli là cơ sở lý thuyết của định nghĩa thống kê về xác suất.

Ở thế kỷ 18, nhà toán học Pháp Buffon gieo một đồng tiền 4040 lần và ghi được 2048 lần xuất hiện mặt ngửa, tần suất là 0,507. Một nhà thống kê người Anh gieo đồng tiền 12000 lần và thu được 6019 lần xuất hiện mặt ngửa, tần suất tương ứng 0,5016. Trong một thí nghiệm khác, ông ta gieo 24000 lần và thu được 12012 lần xuất hiện mặt ngửa, tần suất tương ứng là 0,5005. Như vậy ta thấy rằng khi số phép thử tăng lên thì tần suất tương ứng sẽ càng gần 0,5.

Ví dụ 3.23: Giả sử p là tỉ lệ cử tri sẽ bầu cho ứng cử viên A. Để ước lượng trước tỉ lệ này người ta phỏng vấn ngẫu nhiên n cử tri. Có thể coi kết quả bầu của cử tri thứ i là biến ngẫu nhiên X_i có phân bố Bernoulli tham số p (X_i nhận giá trị 1 nếu cử tri thứ i bầu A và nhận giá trị 0 trong trường hợp ngược lại). Các biến ngẫu nhiên $X_i, i = 1, \dots, n$ độc lập và có cùng phân bố Bernoulli tham số p với kỳ vọng $EX_i = p$ phương sai $DX_i = p(1-p)$.

Tần số
$$k_n = X_1 + \dots + X_n$$

Áp dụng bất đẳng thức Tchebusev ta có

$$P\{|k_n - p| \geq \varepsilon\} \leq \frac{p(1-p)}{n\varepsilon^2}.$$

Tham số p chưa biết nhưng theo bất đẳng thức Cauchy ta có $p(1-p) \leq 1/4$. Vậy

$$P\{|k_n - p| \geq \varepsilon\} \leq \frac{1}{4n\varepsilon^2} \text{ cũng có nghĩa là } P\{|k_n - p| < \varepsilon\} \geq 1 - \frac{1}{4n\varepsilon^2}$$

Chẳng hạn, nếu $\varepsilon = 0,1$ và $n = 100$ thì $P\{|k_{100} - p| \geq 0,1\} \leq \frac{1}{4 \cdot 100 \cdot (0,1)^2} = 0,25$

Nói cách khác, nếu phỏng vấn 100 cử tri và lấy kết quả này để ước lượng cho tỉ lệ cử tri sẽ bỏ phiếu cho ứng cử viên A thì sai số vượt quá 0,1 có xác suất nhỏ hơn 0,25.

Nếu muốn ước lượng tin cậy hơn (chẳng hạn xác suất lớn hơn 95%) và chính xác hơn (với sai số 0,01) thì

$$P\{|k_n - p| \geq 0,01\} \leq \frac{1}{4n(0,01)^2} \text{ cũng có nghĩa là } P\{|k_n - p| < 0,01\} \geq 1 - \frac{1}{4n(0,01)^2}$$

Vậy số cử tri phải phỏng vấn thỏa mãn

$$1 - \frac{1}{4n(0,01)^2} \geq 0,95 \Rightarrow \frac{1}{4n(0,01)^2} \leq 1 - 0,95 = 0,05 \Rightarrow n \geq 50.000.$$

Như vậy để ước lượng với độ tin cậy cao và độ chính xác cao thì cần phải lấy mẫu với số lượng lớn. Tuy nhiên ở đây ta chỉ dựa vào bất đẳng thức Trêbusép để giải quyết bài toán, trong chương 6 ta sẽ nghiên cứu về bài toán ước lượng này và bằng phương pháp khác ta sẽ chỉ ra số cỡ tri phải phòng vẫn nhỏ hơn kết quả trên.

3.7.3 Định lý giới hạn trung tâm

Giả sử X_1, X_2, \dots là dãy các biến ngẫu nhiên độc lập có cùng phân bố, có kỳ vọng μ và phương sai σ^2 . Theo công thức (2.61), (2.71) ta có

$$E[X_1 + \dots + X_n] = n\mu \text{ và } D[X_1 + \dots + X_n] = n\sigma^2$$

Do đó

$$S_n = \frac{X_1 + \dots + X_n - n\mu}{\sigma\sqrt{n}} = \frac{X_1 + \dots + X_n - E[X_1 + \dots + X_n]}{\sqrt{D[X_1 + \dots + X_n]}} \text{ có } ES_n = 0 \text{ và } DS_n = 1.$$

Định lý 3.8: Giả sử X_1, X_2, \dots 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 $S_n = \frac{X_1 + \dots + X_n - n\mu}{\sigma\sqrt{n}}$ hội tụ theo phân bố về phân bố chuẩn tắc $N(0;1)$, tức là:

$$\text{Với mọi } x \in \mathbb{R}, \lim_{n \rightarrow \infty} F_{S_n}(x) = \lim_{n \rightarrow \infty} P\{S_n \leq x\} = \Phi(x) \quad (3.69)$$

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

Áp dụng định lý giới hạn trung tâm cho dãy các biến ngẫu nhiên độc lập X_1, X_2, \dots có cùng phân bố Bernoulli tham số p (công thức (2.9)) ta được định lý Moivre – Laplace:

Định lý 3.9 (Moivre – Laplace): Đối với dãy các biến ngẫu nhiên X_1, X_2, \dots độc lập có cùng phân bố Bernoulli tham số p thì:

$$\text{Với mọi } x \in \mathbb{R}, \lim_{n \rightarrow \infty} P\left\{ \frac{X_1 + \dots + X_n - np}{\sqrt{npq}} \leq x \right\} = \Phi(x). \quad (3.70)$$

Định lý Moivre-Laplace cho phép xấp xỉ phân bố nhị thức $B(n; p)$ với phân bố chuẩn $N(np; npq)$ khi n đủ lớn (công thức 2.78). Người ta thấy rằng xấp xỉ là tốt khi np và nq lớn hơn 5 hoặc khi $npq > 20$.

3.7.4 Xấp xỉ phân bố nhị thức

3.7.4.1 Xấp xỉ phân bố nhị thức bằng phân bố chuẩn

Giả sử X_1, X_2, \dots, X_n độc lập có cùng phân bố Bernoulli tham số p . Theo công thức (2.17), (2.20) ta có

$$U_n = X_1 + X_2 + \dots + X_n \sim B(n; p).$$

Biến ngẫu nhiên U_n nhận các giá trị $k = 0, 1, \dots, n$ với xác suất

$$P\{U_n = k\} = C_n^k p^k q^{n-k} = \frac{n!}{k!(n-k)!} p^k q^{n-k} \text{ (công thức 2.21).}$$

Tuy nhiên khi n khá lớn thì $n!$ rất lớn, do đó ta không thể áp dụng công thức này để tính mà cần đến công thức xấp xỉ.

Sử dụng các định lý giới hạn ta có thể tính xấp xỉ các xác suất này.

Định lý 3.10 (*định lý giới hạn địa phương*): Giả sử X là biến ngẫu nhiên có phân bố nhị thức $B(n, p)$. Đặt $x_k = \frac{k - np}{\sqrt{npq}}$, khi đó

$$P_n(k; p) = P\{X = k\} = \varphi\left(\frac{k - np}{\sqrt{npq}}\right) \frac{1}{\sqrt{npq}} (1 + \varepsilon_{n,k}). \quad (3.71)$$

trong đó $|\varepsilon_{n,k}| < \frac{C}{\sqrt{n}}$ với C là hằng số.

Vì vậy khi n đủ lớn ta có thể xấp xỉ:

$$P\{X = k\} \approx \varphi\left(\frac{k - np}{\sqrt{npq}}\right) \frac{1}{\sqrt{npq}} = \frac{1}{\sqrt{npq}} \frac{1}{\sqrt{2\pi}} e^{-\frac{(k - np)^2}{2npq}}. \quad (3.72)$$

Để tính xấp xỉ giá trị của hàm phân bố xác suất nhị thức ta có thể áp dụng định lý Moivre-Laplace - công thức (3.72) như sau

$$P\{U_n \leq x\} = P\left\{\frac{U_n - np}{\sqrt{npq}} \leq \frac{x - np}{\sqrt{npq}}\right\} \approx \Phi\left(\frac{x - np}{\sqrt{npq}}\right) \quad (3.73)$$

$$P\{a \leq U_n \leq b\} \approx \Phi\left(\frac{b - np}{\sqrt{npq}}\right) - \Phi\left(\frac{a - np}{\sqrt{npq}}\right) \quad (3.74)$$

Người ta thấy rằng xấp xỉ là tốt khi np và nq lớn hơn 5 hoặc khi $npq > 20$.

Khi $a = b = k$, $0 \leq k \leq n$, vế trái của công thức (3.74) sẽ là $P\{U_n = k\} \neq 0$, trong khi đó vế phải bằng 0. Điều này xảy ra vì ta đã dùng hàm phân bố liên tục để xấp xỉ phân bố rời rạc, vì vậy để xấp xỉ tốt hơn người ta thường sử dụng công thức có dạng sau

$$P\{U_n \leq x\} \approx \Phi\left(\frac{x + 0,5 - np}{\sqrt{npq}}\right); P\{a \leq U_n \leq b\} \approx \Phi\left(\frac{b + 0,5 - np}{\sqrt{npq}}\right) - \Phi\left(\frac{a - 0,5 - np}{\sqrt{npq}}\right) \quad (3.75)$$

Ví dụ 3.24: Giả sử U_n là một biến ngẫu nhiên có phân bố nhị thức với tham số $n = 36$ và $p = 0,5$. Xác suất U_n nhận giá trị ≤ 21 được tính chính xác theo công thức là

$$P\{U_n \leq 21\} = \sum_{k=0}^{21} \frac{36!}{k!(36-k)!} (0,5)^{36} = 0,8785$$

Áp dụng công thức (3.73) ta được

$$P\{U_n \leq 21\} \approx \Phi\left(\frac{21 - np}{\sqrt{npq}}\right) = \Phi\left(\frac{21 - 18}{3}\right) = \Phi(1) = 0,8413.$$

Áp dụng công thức (3.75) ta được

$$P\{U_n \leq 21\} \approx \Phi\left(\frac{21,5 - np}{\sqrt{npq}}\right) = \Phi\left(\frac{21,5 - 18}{3}\right) = \Phi(1,17) = 0,879.$$

Như vậy sử dụng công thức 3-17 có kết quả gần với giá trị chính xác.

$$P\{U_n = 19\} \approx \Phi\left(\frac{19,5 - 18}{3}\right) - \Phi\left(\frac{18,5 - 18}{3}\right) = 0,6915 - 0,5675 = 0,124.$$

Giá trị này cũng khá gần với giá trị chính xác

$$\frac{36!}{19!(36-19)!} (0,5)^{36} = 0,1251.$$

Ví dụ 3.25: 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 $P\{1610 \leq X \leq 1650\}$.

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

$$P_{3200}(1600; 0,5) = \frac{3200!}{1600!1600!} 0,5^{3200} \text{ (không thể tính được vì tràn bộ nhớ).}$$

Mặt khác nếu tính gần đúng ta có: $x_m = \frac{1600 - 3200 \cdot 0,5}{\sqrt{3200 \cdot 0,5 \cdot 0,5}} = 0$. Do đó

$$P_{3200}(1600; 0,5) \approx \frac{1}{\sqrt{3200 \cdot 0,5 \cdot 0,5}} \varphi(0) = \frac{1}{40\sqrt{\pi}} \approx 0,014.$$

$$\text{a. } P\{10 \leq X - 1600 \leq 50\} = P\left\{\frac{10}{20\sqrt{2}} \leq \frac{X - 1600}{20\sqrt{2}} \leq \frac{50}{20\sqrt{2}}\right\} \approx \Phi(1,7678) - \Phi(0,3536)$$

Tra bảng ta có $\Phi(1,7678) - \Phi(0,3536) \approx 0,9616 - 0,6406 = 0,321$.

3.7.4.2 Xấp xỉ phân bố nhị thức bằng phân bố Poisson

Khi điều kiện xấp xỉ phân bố nhị thức bằng phân bố chuẩn không thỏa mãn (điều kiện: np và nq lớn hơn 5 hoặc khi $npq > 20$ không thỏa mãn), ta có thể xấp xỉ phân bố nhị thức bằng phân bố Poisson.

Định lý 3.11: Cho X_1, X_2, \dots là dãy các biến ngẫu nhiên có cùng phân bố nhị thức, ở đó với mỗi n , X_n có phân bố nhị thức $B(n, p_n)$. Nếu tồn tại giới hạn $\lim_{n \rightarrow \infty} np_n = \lambda > 0$ 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 $B(n; p)$ với phân bố Poisson $P(np)$ tham số $\lambda = np$:

$$P\{X_n = k\} \approx e^{-np} \frac{(np)^k}{k!}. \quad (3.76)$$

Ví dụ 3.25: 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 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 $X \sim B(n; p)$ với $n = 100, p = 0,015$.

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

Vậy có khoảng 22,31% 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. $X \sim B(n; p)$ với $n = 100 + k, p = 0,015$. Ta phải xác định k nhỏ nhất để

$$P\{X \leq k\} = \sum_{i=0}^k C_n^i (0,015)^i (0,985)^{n-i} \geq 0,8.$$

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

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

Vậy cần tìm k nhỏ nhất để

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

Thử với $k = 1, 2, \dots$, 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.

TÓM TẮT

Tương tự biến ngẫu nhiên, véc tơ ngẫu nhiên cũng được nghiên cứu theo hàm phân bố xác suất đồng thời (3.1).

Trường hợp véc tơ ngẫu nhiên rời rạc hai chiều (X, Y) , hàm phân bố xác suất đồng thời được xác định từ hàm khối lượng xác suất đồng thời (3.8)-(3.10). Hàm phân bố xác suất đồng thời của véc tơ ngẫu nhiên liên tục được xác định từ hàm mật độ xác suất đồng thời (3.14).

Từ bảng phân bố xác suất đồng thời của véc tơ ngẫu nhiên rời rạc hai chiều (X, Y) ta có thể xác định được bảng phân bố xác suất thành phần X và thành phần Y (Nhận xét 3.1).

Công thức (3.18)-(3.19) cho phép tìm hàm mật độ thành phần X và thành phần Y từ hàm mật độ xác suất đồng thời của véc tơ ngẫu nhiên liên tục (X, Y) .

Các tiêu chuẩn để nhận biết hai biến ngẫu nhiên X, Y độc lập (công thức 3.20–3.23).

Ngoài hai đặc trưng kỳ vọng và phương sai đã được xét ở chương 2, trong chương này còn xét các đặc trưng hiệp phương sai $\text{cov}(X, Y)$ (công thức 3.32a–3.34), hệ số tương quan ρ_{XY} (công thức 3.39) và ma trận hiệp phương sai (công thức 3.38).

Hàm khối lượng xác suất có điều kiện (công thức 3.65, 3.67, 3.68). Hàm mật độ xác suất có điều kiện (công thức 3.70 – 3.75).

Kỳ vọng có điều kiện (công thức 3.66, 3.69, 3.76).

Luật số lớn Trêbusép xét sự hội tụ theo xác suất của trung bình cộng của dãy các biến ngẫu nhiên độc lập có kỳ vọng và phương sai hữu hạn dựa vào bất đẳng thức Trêbusép. Khi các biến ngẫu nhiên của dãy cùng có phân bố Bernoulli thì ta có luật số lớn Bernoulli.

Định lý giới hạn trung tâm xét sự hội tụ theo phân bố của tổng các số hạng của dãy các biến ngẫu nhiên về phân bố chuẩn tắc. Trường hợp các biến ngẫu nhiên của dãy cùng có phân bố Bernoulli thì ta có định lý Moivre-Laplace.

Áp dụng định lý giới hạn trung tâm ta có thể tính gần đúng xác suất của phân bố nhị thức khi tham số n khá lớn, công thức (3.13) - (3.17). Khi tham số n lớn nhưng p bé thì xấp xỉ phân bố nhị thức $B(n; p)$ với phân bố Poisson $P(np)$ tham số $\lambda = np$, công thức (3.18).

CÂU HỎI ÔN TẬP VÀ BÀI TẬP CHƯƠNG 3

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

Đúng ☐ Sai ☐.

3.9 Nếu hai biến ngẫu nhiên X, Y độc lập thì hàm hồi quy $f(x) = E(Y|X = x)$ của Y đối với X và hàm hồi quy $g(y) = E(X|Y = y)$ 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 ($\text{cov}(X, Y) = 0$ thì $EX = EY$).

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 Luật số lớn kết luận về sự hội tụ theo xác suất của trung bình cộng các biến ngẫu nhiên độc lập về trung bình cộng của kỳ vọng của chúng nếu các phương sai của các biến ngẫu nhiên này bị chặn.

Đúng ☐ Sai ☐.

3.14 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 ☐.

3.15 Bất đẳng thức Trêbusép chỉ đúng đối với các biến ngẫu nhiên rời rạc.

Đúng ☐ Sai ☐.

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

Đúng ☐ Sai ☐.

3.17 Luật số lớn Bernoulli là một trường hợp đặc biệt của luật số lớn Trêbusép khi dãy các biến ngẫu nhiên được có cùng phân bố Bernoulli $A(p)$.

Đúng ☐ Sai ☐.

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

Đúng ☐ Sai ☐.

3.19 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 ☐.

3.20 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 ☐.

3.21 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 .

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

Chương 3: Véc tơ ngẫu nhiên và các đặc trưng của chúng

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

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

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

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

- Chứng minh rằng X, Y có độc lập.
- Tìm quy luật phân bố của biến ngẫu nhiên $Z = XY$.
- Tính các kỳ vọng EX, EY, EZ .

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

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

$X \backslash Y$	26	30	41	50
23	0,05	0,08	0,12	0,04
27	0,09	0,30	0,11	0,21

Tìm bảng phân bố xác suất điều kiện của Y khi $X = 26$ và của X khi $Y = 27$.

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

$X \backslash Y$	1	3	4	8
3	0,15	0,06	0,25	0,04
6	0,30	0,10	0,03	0,07

- a. Tìm kỳ vọng có điều kiện của Y khi $X = 1$.
- b. Tìm các kỳ vọng EX, EY và phương sai DX, DY .

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

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

- a. Tìm α . Tính EX, EY .
- b. Tính $\text{cov}(X, Y), \rho_{X, Y}$.
- c. X và Y có độc lập không.

3.28 Giả sử X là biến ngẫu nhiên rời rạc có phân bố Poisson tham số λ . Tìm hàm khối lượng xác suất với điều kiện $B = \{X \leq n\}$.

3.29 Cho véc tơ ngẫu nhiên rời rạc (X, Y) có hàm khối lượng xác suất xác định như sau:

$$f_{X,Y}(x, y) = \begin{cases} k(2x + y) & \text{nếu } x = 0, 1, 2; y = 0, 1, 2, 3 \\ 0 & \text{nếu ngược lại} \end{cases}.$$

- a. Tìm k . Tính $P\{X = 2, Y = 1\}, P\{X \geq 1, Y \leq 2\}$
- b. Tìm hàm khối lượng điều kiện và kỳ vọng điều kiện của Y khi $X = 2$.

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

$$f_{X,Y}(x, y) = \begin{cases} kx & \text{nếu } 0 < y < x < 1 \\ 0 & \text{nếu ngược lại} \end{cases}.$$

- a. Tìm k .
- b. Tìm các hàm mật độ xác suất của X và của Y .
- c. X và Y có độc lập không?

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

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

Tìm hàm mật độ xác suất đồng thời $f_{X,Y}(x, y)$ và hàm mật độ có điều kiện $f_{X|Y}(x|y)$.

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

$$f_{X,Y}(x, y) = \frac{k}{(1+x^2)(1+y^2)}.$$

- a. Tìm k . Tìm hàm phân bố xác suất đồng thời của X, Y .
- b. X và Y có độc lập không?

c. 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 $D(\sqrt{3},0)$.

3.33 Chứng minh rằng hàm phân bố điều kiện và hàm mật độ điều kiện của biến ngẫu nhiên X với điều kiện $B = \{a < X \leq b\}$ có dạng

$$F_{X|B}(x|B) = \begin{cases} 0 & x \leq a \\ \frac{F_X(x) - F_X(a)}{F_X(b) - F_X(a)} & a < x \leq b; \\ 1 & x > b \end{cases}$$

$$f_{X|B}(x|B) = \begin{cases} 0 & x \leq a \\ f_X(x) / \int_a^b f_X(y) dy & a < x \leq b \\ 0 & x > b. \end{cases}$$

3.34 Giả sử biến ngẫu nhiên $X \sim N(0; \sigma^2)$. Tìm hàm mật độ điều kiện, hàm phân bố điều kiện và kỳ vọng điều kiện với điều kiện $B = \{X > 0\}$.

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

$$f_{X,Y}(x, y) = \begin{cases} k(x+y) & \text{nếu } 0 < x < 2, 0 < y < 2 \\ 0 & \text{nếu ngược lại} \end{cases}$$

a. Tìm k . Tìm hàm mật độ xác suất của X và của Y . X, Y có độc lập không?

b. Tìm hàm mật độ xác suất có điều kiện $f_{X|Y}(x|y)$ và $f_{Y|X}(y|x)$.

c. Tính $P\{0 < Y < 1/2 | X = 1\}$.

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

$$f_{X,Y}(x, y) = \begin{cases} kxy & \text{nếu } 1 < x < 4, 1 < y < 5 \\ 0 & \text{nếu ngược lại} \end{cases}$$

a. Tìm k . Tìm hàm mật độ xác suất của X và của Y . X, Y có độc lập không?

b. Tính $P\{1 < X < 2, 2 < Y < 3\}$; $P\{X \geq 3; Y \leq 2\}$.

c. Tính $P\{X + Y < 3\}$.

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

$$f_{X,Y,Z}(x, y, z) = \begin{cases} 24xy^2z^3 & \text{nếu } 0 < x < 1, 0 < y < 1, 0 < z < 1 \\ 0 & \text{nếu ngược lại} \end{cases}$$

a. Tính $P\{X > 0,5; Y < 0,5; Z > 0,5\}$.

b. Tính $P\{Z < X + Y\}$.

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

$EX = 35, EY = 20, DX = 36, DY = 16$ và $\rho_{X,Y} = 0,8$.

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

3.39 Cho hai biến ngẫu nhiên X, Y .

a. Chứng minh bất đẳng thức Cauchy-Schwarz: $(E[XY])^2 \leq (EX^2)(EY^2)$.

b. Chứng minh hệ số tương quan thỏa mãn: $|\rho_{X,Y}| \leq 1$.

3.40 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 Trêbusép hãy đánh giá xác suất của sự sai lệch giữa số máy hỏng và số máy hỏng trung bình.

a. Nhỏ hơn 2.

b. Lớn hơn 2

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

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

3.42 Cho $X_1, X_2, \dots, X_{10000}$ là các biến ngẫu nhiên độc lập có phân bố đều trong đoạn $\left[-\frac{1}{2}, \frac{1}{2}\right]$.

Chứng minh rằng $P\left\{\left|\sum_{i=1}^{10000} X_i\right| \geq 500\right\} \leq \frac{1}{300}$.

3.43 Xét biến ngẫu nhiên liên tục X có hàm mật độ $f_X(x)$ và giả sử X_1, X_2, \dots, X_n là các biến ngẫu nhiên độc lập có cùng phân bố với X . Ta gọi (X_1, X_2, \dots, X_n) là một mẫu ngẫu nhiên kích thước n của X . Trung bình mẫu được định nghĩa là $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$. Ký hiệu $EX = \mu, DX = \sigma^2$.

Cần phải lấy mẫu kích thước bao nhiêu để $P\left\{\left|\bar{X} - \mu\right| \leq \frac{\sigma}{10}\right\} \geq 0,95$.

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

3.45 Giả sử tiền điện của một gia đình phải trả trong 1 tháng là một biến ngẫu nhiên với trung bình 16 USD và độ lệch tiêu chuẩn 1 USD. Sử dụng bất đẳng thức Trêbusép, hãy xác định số M nhỏ nhất để với xác suất 0,99 số tiền điện phải trả trong 1 năm (12 tháng) không vượt quá M .

3.46 Xác suất chậm tàu của mỗi hành khách là 0,007. Dùng bất đẳng thức Trêbusép hãy đánh giá xác suất để trong 20.000 hành khách có từ 100 đến 180 người chậm tàu.

3.47 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.

3.48 Một xí nghiệp sản xuất máy tính có xác suất làm ra sản phẩm phế phẩm là 0,02. Chọn ngẫu nhiên 250 máy tính để kiểm tra. Tính xác suất để:

- a. Có đúng hai máy phế phẩm;
- b. Có không quá hai máy phế phẩm.

3.49 Một nhà nghỉ có 1000 khách. Nhà ăn phục vụ bữa trưa làm hai đợt liên tiếp. Áp dụng định lý Moivre-Laplace hãy tìm 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%?

3.50 Một kênh truyền có nhiễu với xác suất lỗi trên mỗi ký số là 0,01.

- a. Tính xác suất có nhiều hơn một lỗi trong 10 ký số nhận được;
- b. Tính kết quả của ý a) bằng cách sử dụng công thức xấp xỉ Poisson.

3.51 Giả sử X là biến ngẫu nhiên có phân bố Poisson tham số λ . Áp dụng định lý giới hạn trung tâm hãy xây dựng công thức xấp xỉ

$$P\{X \leq x\} \approx \Phi\left(\frac{x-\lambda}{\sqrt{\lambda}}\right).$$

CHƯƠNG 4: LÝ THUYẾT MẪU

Thống kê toán là bộ môn toán học nghiên cứu qui luật của các hiện tượng ngẫu nhiên có tính chất số lớn trên cơ sở thu thập và xử lý số liệu thống kê các kết quả quan sát về những hiện tượng ngẫu nhiên này. Nếu ta thu thập được các số liệu liên quan đến tất cả đối tượng cần nghiên cứu thì ta có thể biết được đối tượng này (phương pháp toàn bộ). Tuy nhiên trong thực tế điều đó không thể thực hiện được vì quy mô của các đối tượng cần nghiên cứu quá lớn hoặc trong quá trình nghiên cứu đối tượng nghiên cứu bị phá hủy. Lý thuyết mẫu cung cấp phương pháp nghiên cứu tổng thể thông qua mẫu.

Chương này giới thiệu về phương pháp lấy mẫu ngẫu nhiên, xác định các thống kê thường gặp của mẫu ngẫu nhiên và các đặc trưng của chúng.

4.1 SỰ CẦN THIẾT PHẢI LẤY MẪU

Nhiều bài toán trong thực tế dẫn đến nghiên cứu một hay nhiều dấu hiệu định tính hoặc định lượng 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 gia đình (dấu hiệu định lượng). Một doanh nghiệp muốn nghiên cứu các khách hàng của mình về dấu hiệu định tính có thể là sự hài lòng của khách hàng đối với sản phẩm hoặc dịch vụ của doanh nghiệp, còn dấu hiệu định lượng là số lượng sản phẩm của doanh nghiệp mà khách hàng có nhu cầu được đáp ứng.

Dấu hiệu định lượng được thể hiện qua các đơn vị đo của đại lượng như trọng lượng (gram, kg ...), độ dài (cm, m ...), thời gian (giờ, giây ...), áp suất (atmosphere...), chiếc, tá ... Trái lại, dấu hiệu định tính như giới tính (nam, nữ), sở thích (yêu, ghét, thích loại sản phẩm nào đó), loại phương tiện (di động, cố định, internet)... không có đơn vị. Dù rằng có thể mã hóa những dấu hiệu này ví dụ nam ứng với 1, nữ ứng với 0, song không có đơn vị đo cho dấu hiệu này.

Ngoài thành phần mang tin của các tín hiệu trên đường truyền còn có sự tác động của nhiễu vì vậy các tín hiệu theo thời gian là các quá trình ngẫu nhiên. Tín hiệu thu được là các tín hiệu mẫu của quá trình ngẫu nhiên đó.

Để xử lý dấu hiệu cần nghiên cứu đôi khi người ta sử dụng 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 chông chéo hoặc bỏ sót.
- Trong nhiều trường hợp không thể nắm được toàn bộ các phần tử của tập hợp cần nghiên cứu, do đó không thể tiến hành toàn bộ được. Chẳng hạn doanh nghiệp không thể xác định toàn bộ các khách hàng của mình.

- Có thể trong quá trình điều tra sẽ phá hủy đối tượng nghiên cứu ... Nhà máy sản xuất diêm không thể sử dụng phương pháp toàn bộ để xác định tỷ lệ các que diêm cháy do nhà máy sản xuất.

Vì thế trong thực tế phương pháp nghiên cứu toàn bộ thường chỉ áp dụng đối với các tập hợp có qui mô nhỏ, còn chủ yếu người ta sử dụng phương pháp không toàn bộ mà đặc biệt là phương pháp nghiên cứu chọn mẫu.

4.2 MẪU NGẪU NHIÊN

4.2.1 Khái niệm mẫu ngẫu nhiên

Toàn bộ tập hợp các phần tử đồng nhất theo một dấu hiệu nghiên cứu định tính hay định lượng nào đó được gọi là tổng thể, ký hiệu 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 phần tử của tổng thể được gọi là *cá thể*.

Các cá thể của tổng thể được nghiên cứu thông qua các dấu hiệu nghiên cứu. Dấu hiệu nghiên cứu này có thể được định tính hoặc định lượng. Nếu dấu hiệu nghiên cứu có tính định lượng, nghĩa là được thể hiện bằng cách cho tương ứng mỗi cá thể của tổng thể C nhận một giá trị thực nào đó thì dấu hiệu này được gọi là một *biến lượng*, ký hiệu X . Có thể xem biến lượng X là một biến ngẫu nhiên chung của tổng thể.

Với dấu hiệu định tính ta chỉ xét trường hợp các dấu hiệu có thể mã hóa thành biến ngẫu nhiên chỉ nhận hai giá trị 0 và 1, như vậy dấu hiệu định tính X có thể xem là biến ngẫu nhiên có phân bố Bernoulli. Chẳng hạn một doanh nghiệp muốn nghiên cứu các khách hàng của mình về dấu hiệu định tính là sự hài lòng của khách hàng đối với sản phẩm hoặc dịch vụ của doanh nghiệp, câu trả lời của các khách hàng chỉ chọn theo một trong hai giá trị sau: có hài lòng (giá trị 1) hoặc không hài lòng (giá trị 0) và không xét các ý kiến khác.

Việc chọn n cá thể nào đó từ tổng thể được gọi là *phép lấy mẫu*. Ta gọi các cá thể chọn được này là *một mẫu*, n là kích thước mẫu. Nếu cá thể chọn xong không trả lại tổng thể để chọn tiếp thì mẫu được gọi là *không hoàn lại*. Nếu chọn xong trả lại tổng thể để chọn tiếp thì mẫu được gọi *có hoàn lại*.

Ta nói rằng một mẫu là *mẫu ngẫu nhiên* nếu trong phép lấy mẫu đó mỗi cá thể của tổng thể được chọn một cách độc lập và có xác suất được chọn như nhau.

Trường hợp kích thước của tổng thể vô hạn hoặc không xác định thì mẫu ngẫu nhiên có thể chọn theo cách hoàn lại hoặc không hoàn lại, nhưng khi kích thước của tổng thể không lớn thì mẫu ngẫu nhiên phải thực hiện theo cách hoàn lại.

Có nhiều cách chọn mẫu ngẫu nhiên, khi kích thước tổng thể hữu hạn ta có thể chọn được mẫu ngẫu nhiên từ tổng thể bằng cách sử dụng các số ngẫu nhiên.

4.2.2 Mô hình hóa mẫu ngẫu nhiên

Giả sử các cá thể của tổng thể được nghiên cứu thông qua dấu hiệu X . Với mỗi mẫu ta chỉ cần quan tâm dấu hiệu nghiên cứu X của mỗi cá thể của mẫu.

Chẳng hạn, khi cần nghiên cứu chiều cao trung bình của thanh niên trong một vùng nào đó thì với cá thể A được chọn làm mẫu ta chỉ quan tâm về chiều cao của A , tức là dấu hiệu chiều cao X_A của A , mà không quan tâm đến các đặc trưng khác của cá thể này.

Vì vậy, mỗi cá thể được chọn khi lấy mẫu có thể đồng nhất với dấu hiệu nghiên cứu X của cá thể đó. Bằng cách đồng nhất mỗi cá thể của mẫu ngẫu nhiên với các dấu hiệu nghiên cứu tương ứng của cá thể ta có thể xác định mẫu ngẫu nhiên như sau:

Mẫu ngẫu nhiên kích thước n là một dãy gồm n biến ngẫu nhiên: X_1, X_2, \dots, X_n độc lập cùng phân bố với X , ký hiệu $W = (X_1, X_2, \dots, X_n)$, trong đó X_i là dấu hiệu X của cá thể thứ i của mẫu ($i = 1, \dots, n$).

Thực hiện một phép thử đối với mẫu ngẫu nhiên W chính là thực hiện một phép thử đối với mỗi thành phần của mẫu. Giả sử X_i nhận giá trị x_i ($i = 1, \dots, n$), khi đó các giá trị x_1, x_2, \dots, x_n tạo thành một giá trị cụ thể của mẫu ngẫu nhiên, hay còn gọi là một thể hiện của mẫu ngẫu nhiên, ký hiệu $w = (x_1, x_2, \dots, x_n)$.

Ví dụ 4.1: Gọi X là số chấm của mặt xuất hiện khi gieo con xúc xắc cân đối, X là biến ngẫu nhiên rời rạc 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

Giả sử tung con xúc xắc 3 lần, gọi X_i là số chấm xuất hiện trong lần tung thứ i ($i = 1, 2, 3$) thì ta có 3 biến ngẫu nhiên độc lập có cùng quy luật phân bố xác suất với X . Vậy ta có mẫu ngẫu nhiên kích thước 3, $W = (X_1, X_2, X_3)$.

Thực hiện một phép thử đối với mẫu ngẫu nhiên này tức là tung con xúc xắc 3 lần. Giả sử lần thứ nhất được 2 chấm, lần thứ hai được 5 chấm, lần ba được 3 chấm thì $w = (2, 5, 3)$ là một mẫu cụ thể của mẫu ngẫu nhiên W .

4.2.3 Biểu diễn giá trị cụ thể của mẫu ngẫu nhiên theo bảng và theo biểu đồ

4.2.3.1 Bảng phân bố tần số thực nghiệm

Từ một mẫu cụ thể của mẫu ngẫu nhiên kích thước n của X ta sắp xếp các giá trị của mẫu cụ thể theo thứ tự tăng dần, giả sử giá trị x_i xuất hiện với tần số r_i , $i = 1, \dots, k$

$$x_1 < \dots < x_k; r_1 + \dots + r_k = n. \quad (4.1)$$

Khi đó ta có thể biểu diễn giá trị quan sát của mẫu ngẫu nhiên trên qua bảng phân bố tần số thực nghiệm

X	x_1	x_2	...	x_k
Tần số	r_1	r_2	...	r_k

(4.2)

Ở đây x_1 là giá trị cụ thể nhỏ nhất, chưa chắc là giá trị cụ thể của X_1 .

4.2.3.2 Bảng phân bố tần suất thực nghiệm

Ký hiệu $f_i = \frac{r_i}{n}$ và gọi là tần suất của giá trị x_i .

Ta có bảng phân bố tần suất thực nghiệm tương ứng của X

X	x_1	x_2	...	x_k
Tần suất	f_1	f_2	...	f_k

(4.3)

Ví dụ 4.2: Giá trị quan sát của một mẫu ngẫu nhiên kích thước 120 được biểu diễn theo bảng phân bố tần số thực nghiệm

X	31	34	35	36	38	40	42	44	Σ
Tần số	10	20	30	15	10	10	5	20	120

Bảng phân bố tần suất thực nghiệm tương ứng

X	31	34	35	36	38	40	42	44	Σ
Tần suất	2/24	4/24	6/24	3/24	2/24	2/24	1/24	4/24	1

4.2.3.3 Hàm phân bố thực nghiệm của mẫu

Tương tự công thức (2.12) xác định hàm phân bố của biến ngẫu nhiên rời rạc từ hàm khối lượng xác suất, với mẫu cụ thể của mẫu ngẫu nhiên xác định bởi công thức (4.1) ta có **hàm phân bố thực nghiệm của mẫu** xác định như sau

$$F_n(x) = \sum_{x_j \leq x} f_j; \quad -\infty < x < +\infty \quad (4.4)$$

Định lý Glivenco chỉ ra rằng hàm phân bố thực nghiệm $F_n(x)$ xấp xỉ với phân bố lý thuyết $F_X(x) = P\{X \leq x\}$ khi n đủ lớn, trong đó X là biến ngẫu nhiên gốc của tổng thể.

4.2.3.4 Bảng phân bố ghép lớp

Trong những trường hợp mẫu điều tra có kích thước lớn, hoặc khi các giá trị cụ thể của dấu hiệu X lấy giá trị khác nhau song lại khá gần nhau, người ta thường xác định một số các khoảng C_1, C_2, \dots, C_k sao cho mỗi giá trị của dấu hiệu điều tra thuộc vào một khoảng nào đó. Các khoảng này lập thành một phân hoạch của miền giá trị của X .

Việc chọn số khoảng và độ rộng khoảng là tùy thuộc vào kinh nghiệm của người nghiên cứu, nhưng nói chung không nên chia quá ít khoảng. Ngoài ra độ rộng các khoảng cũng không nhất thiết phải bằng nhau. Chẳng hạn khi muốn thống kê về tỉ lệ người nghiện thuốc lá thì ta tập trung nhiều vào độ tuổi thanh niên và trung niên, vì vậy các khoảng trong độ tuổi này sẽ nhỏ hơn.

Ví dụ 4.3: Một mẫu cụ thể về chiều cao (đơn vị pound) của 40 nam sinh viên đại học có số đo được làm tròn

138	164	150	132	144	125	149	157	146	158
140	147	136	148	152	144	168	126	138	176

Chương 4: Lý thuyết mẫu

163	119	154	165	146	173	142	147	135	153
140	135	161	145	135	142	150	156	145	128

trọng lượng lớn nhất 176 và nhỏ nhất 119, vậy khoảng cách tối đa $176-119=57$.

Nếu chia với độ rộng khoảng 5 thì có xấp xỉ $57/5 \approx 11$ khoảng

Nếu chia với độ rộng khoảng 20 thì có xấp xỉ $57/20 \approx 3$ khoảng.

Ta chọn độ rộng khoảng là 5, vì chọn độ rộng khoảng là 20 thì chỉ có 3 khoảng và sai số sẽ lớn.

Ta có hai bảng phân bố ghép lớp với độ rộng khoảng bằng 5 có 11 khoảng và độ rộng khoảng bằng 9 có 7 khoảng

Trọng lượng (pound)	Tần số
118-127	3
127-136	5
136-145	9
145-154	12
154-163	5
163-172	4
172-180	2
Tổng số	40

Độ rộng khoảng bằng 9

Trọng lượng (pound)	Tần số
118-123	1
123-128	2
128-133	2
133-138	4
138-143	6
143-148	8
148-153	5
153-158	4
158-163	2
163-168	3
168-173	1
173-178	2
Tổng số	40

Độ rộng khoảng bằng 5

Nhận xét 4.1:

1) Ví dụ trên cho thấy ta nên chia khoảng các giá trị, vì trong thực tế các số đo nhận được là các giá trị xấp xỉ quy tròn.

2) Người ta quy ước đầu mút bên phải của mỗi khoảng thuộc vào khoảng đó mà không thuộc khoảng tiếp theo khi tính tần số của mỗi khoảng.

3) Một trong những gợi ý để chọn số khoảng k tối ưu là hãy chọn k nguyên nhỏ nhất sao cho $2^k \geq n$ như sau:

n : kích thước mẫu	33 –64	65 –127	129 –256	257 –512	513 –1024
k : số khoảng	6	7	8	9	10

4) Độ rộng các khoảng không đòi hỏi bằng nhau.

Ví dụ 4.4: Một mẫu về chiều cao (cm) của 400 cây con trong vườn ươm được trình bày trong bảng phân bố ghép lớp sau

Chương 4: Lý thuyết mẫu

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

4.2.3.5 Biểu diễn bằng biểu đồ

Để có hình ảnh trực quan ta có thể biểu diễn các giá trị của mẫu bằng biểu đồ.

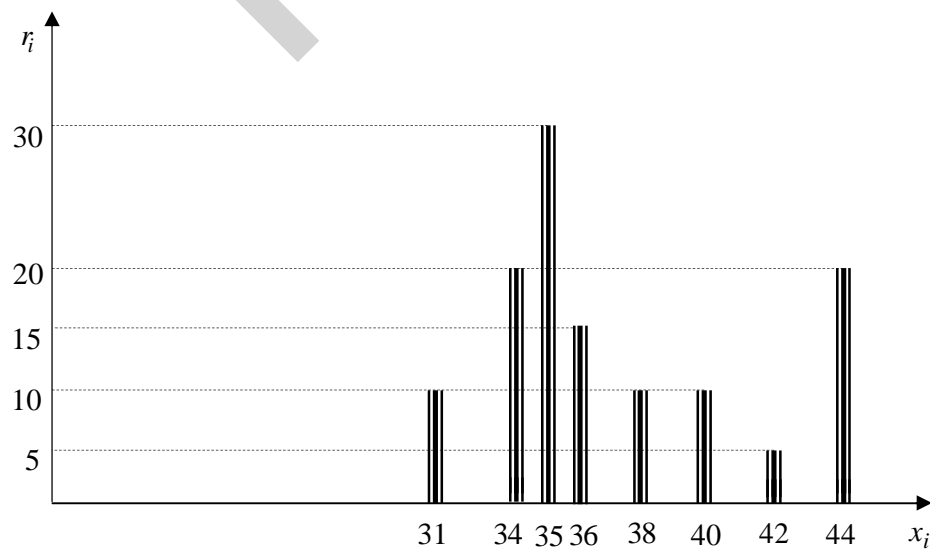
Giả sử dấu hiệu điều tra X có bảng phân bố tần số thực nghiệm và bảng phân bố tần suất thực nghiệm

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 tọa độ Oxy .

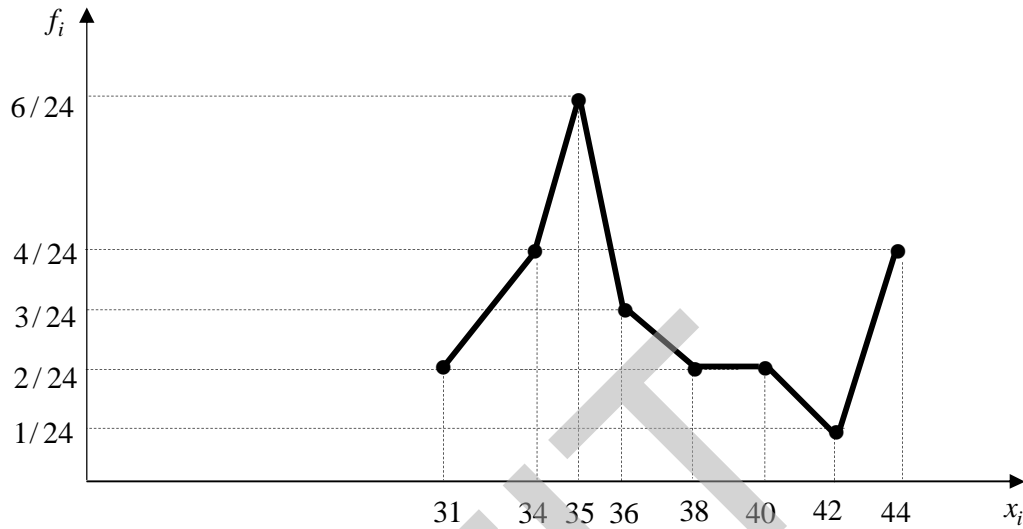
❖ Nối điểm trên trục hoành có tọa độ $(x_i, 0)$ với điểm có tọa độ (x_i, r_i) với mọi $i = 1, 2, \dots, k$ ta được **biểu đồ tần số hình gậy**.



Hình 4.1: Biểu đồ tần số hình gậy

❖ Nối lần lượt điểm có tọa độ (x_i, f_i) với điểm có tọa độ (x_{i+1}, f_{i+1}) với mọi $i = 1, 2, \dots, k-1$ ta được **biểu đồ đa giác tần suất**.

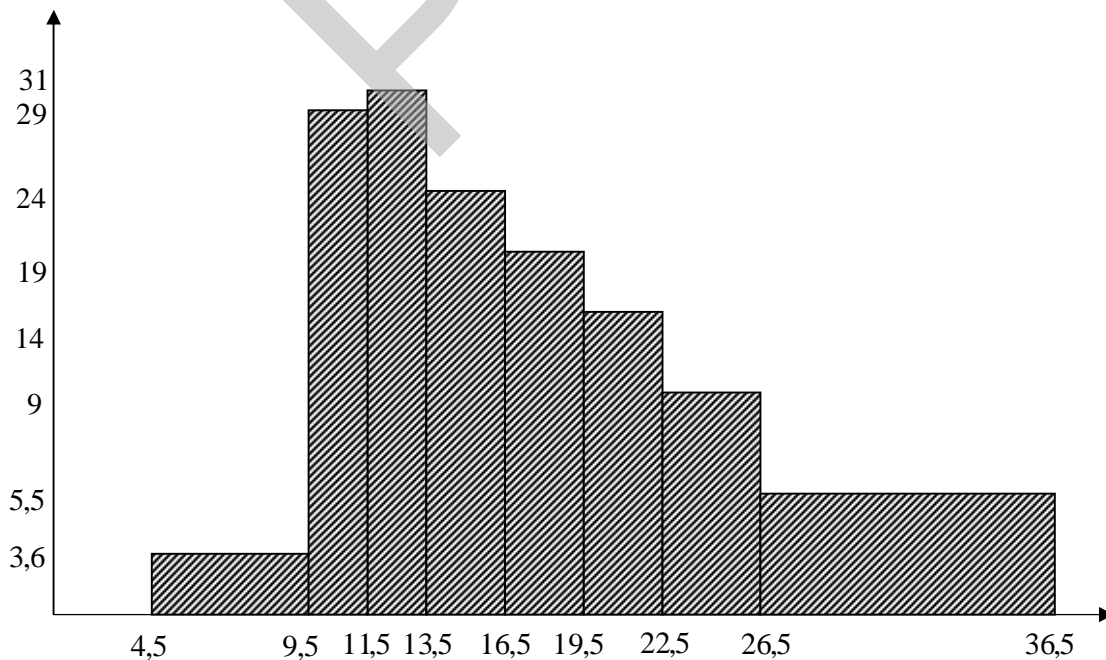
Bảng phân bố tần số và tần suất thực nghiệm trong ví dụ 4.2 có biểu đồ tần số hình gậy và biểu đồ đa giác tần suất



Hình 4.2: Biểu đồ đa giác tần suất

4.2.3.6 Tổ chức đồ (histogram)

Tổ chức đồ tần số của mẫu ghép lớp của ví dụ 4.3



Hình 4.3: Tổ chức đồ

Chương 4: Lý thuyết mẫu

Đối với bảng phân bố ghép lớp, người ta thường dùng tổ chức đồ để biểu diễn.

Trong mặt phẳng với hệ tọa 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).

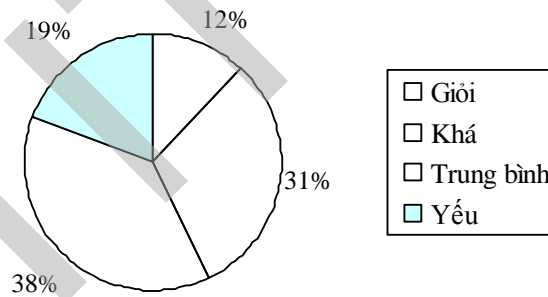
Ý nghĩa hình học của tổ chức đồ là ta có thể tính tần số giá trị cụ thể của mẫu thông qua diện tích giới hạn bởi tổ chức đồ. Chẳng hạn số cây nằm trong khoảng $(12; 25]$ chính là diện tích của tổ chức đồ giới hạn bởi đường thẳng $x = 12$ và $x = 25$. Cụ thể

$$(13,5 - 12) \cdot 31 + (16,5 - 13,5) \cdot 24 + (19,5 - 16,5) \cdot 19 + (22,5 - 19,5) \cdot 14 + (25 - 22,5) \cdot 9 = 240$$

Vậy có 240 cây con có chiều cao từ 12 cm đến 25 cm.

Khi dấu hiệu điều tra của tổng thể có thể biểu diễn dưới dạng các tần số tỷ lệ người ta thường mô tả các số liệu mẫu bằng biểu đồ hình bánh xe. Đó là hình tròn được chia thành những góc có diện tích tỷ lệ với các tần số tương ứng của mẫu.

Ví dụ 5.4: Tổng kết kết quả học tập của sinh viên Học viện CNBCVT trong năm 2005 được số liệu sau:



4.3 THỐNG KÊ VÀ CÁC ĐẶC TRƯNG CỦA MẪU NGẪU NHIÊN

4.3.1 Định nghĩa thống kê

Một thống kê của mẫu là một hàm của các biến ngẫu nhiên thành phần của mẫu. Thống kê của mẫu ngẫu nhiên $W = (X_1, X_2, \dots, X_n)$ có dạng:

$$T = T(X_1, X_2, \dots, X_n) \quad (4.5)$$

Như vậy thống kê T cũng là một biến ngẫu nhiên tuân theo một quy luật phân bố xác suất nhất định và có các tham số đặc trưng như kỳ vọng ET phương sai $DT \dots$ (xem mục 3.5 chương 3). Mặt khác, khi mẫu ngẫu nhiên nhận một giá trị cụ thể $w = (x_1, x_2, \dots, x_n)$ thì T cũng nhận một giá trị cụ thể còn gọi là giá trị quan sát của thống kê

$$T_{qs} = T(x_1, x_2, \dots, x_n)$$

Các thống kê mẫu cùng với quy luật phân bố xác suất của chúng là cơ sở để khảo sát dấu hiệu nghiên cứu của tổng thể từ các thông tin của mẫu.

Chúng ta xét một số thống kê mẫu quan trọng sau.

4.3.2 Trung bình mẫu

Trung bình mẫu của mẫu ngẫu nhiên $W = (X_1, X_2, \dots, X_n)$ của biến ngẫu nhiên gốc X là trung bình cộng của các thành phần của mẫu và ký hiệu

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad (4.6)$$

Giá trị trung bình mẫu cụ thể của mẫu ngẫu nhiên cụ thể $w = (x_1, x_2, \dots, x_n)$ là

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (4.7)$$

Giả sử dấu hiệu nghiên cứu X có kỳ vọng và phương sai hữu hạn, áp dụng các công thức tính kỳ vọng và phương sai của tổng các biến ngẫu nhiên độc lập (2.47), (2.48), (2.58) và ví dụ 2.26) ta có

$$E(\bar{X}) = E X = \mu; \quad D(\bar{X}) = \frac{D X}{n} = \frac{\sigma^2}{n}. \quad (4.8)$$

4.3.3 Phương sai mẫu, Độ lệch chuẩn mẫu

Một cách tương tự trung bình mẫu, ta định nghĩa phương sai mẫu là trung bình cộng của độ lệch bình phương các thành phần của mẫu với trung bình mẫu và ký hiệu

- Phương sai mẫu \hat{S}^2 :

$$\hat{S}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - (\bar{X})^2 \quad (4.9)$$

Phương sai mẫu \hat{S}^2 cũng là một biến ngẫu nhiên, sử dụng các tính chất kỳ vọng ta có:

$$\begin{aligned} E[\hat{S}^2] &= E\left[\frac{1}{n} \sum_{i=1}^n ((X_i - \mu) + (\mu - \bar{X}))^2\right] = \frac{1}{n} E\left[\sum_{i=1}^n ((X_i - \mu)^2 + (\bar{X} - \mu)^2 - 2(X_i - \mu)(\bar{X} - \mu))\right] \\ &= \frac{1}{n} E\left[\left(\sum_{i=1}^n (X_i - \mu)^2\right) + n(\bar{X} - \mu)^2 - 2n(\bar{X} - \mu)(\bar{X} - \mu)\right] \\ &= \frac{1}{n} E\left[\left(\sum_{i=1}^n (X_i - \mu)^2\right) - n(\bar{X} - \mu)^2\right] = \frac{1}{n} \left(n D X - n \frac{D X}{n}\right) = \frac{n-1}{n} D X. \end{aligned}$$

Đề kỳ vọng của phương sai mẫu trùng với phương sai của biến ngẫu nhiên gốc ta cần hiệu chỉnh như sau.

- Phương sai mẫu có hiệu chỉnh S^2 :

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n-1} \left(\sum_{i=1}^n X_i^2\right) - \frac{n}{n-1} (\bar{X})^2 \quad (4.10)$$

- Trường hợp biến ngẫu nhiên gốc X có kỳ vọng $E X = \mu$ đã xác định thì phương sai mẫu được chọn là S^{*2} xác định như sau

$$S^{*2} = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2 \quad (4.11)$$

Áp dụng công thức tính kỳ vọng (2.61), (2.62) và (2.66) ta có:

$$E S^2 = D X \quad \text{và} \quad E S^{*2} = D X \quad (4.12)$$

Độ lệch chuẩn mẫu

$$S = \sqrt{S^2} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2} \quad (4.13)$$

4.3.4 Tần suất mẫu

Trường hợp cần nghiên cứu một dấu hiệu định tính A nào đó mà mỗi cá thể của tổng thể có thể có hoặc không, giả sử p là tần suất có dấu hiệu A của tổng thể.

Nếu cá thể có dấu hiệu A ta cho nhận giá trị 1, trường hợp ngược lại ta cho nhận giá trị 0. Lúc đó dấu hiệu nghiên cứu có thể xem là biến ngẫu nhiên X có phân bố Bernoulli tham số p có kỳ vọng $E X = p$ và phương sai $DX = p(1-p)$ (công thức (2.17)).

Lấy mẫu ngẫu nhiên: $W = (X_1, X_2, \dots, X_n)$, X_1, X_2, \dots, X_n là các biến ngẫu nhiên độc lập có cùng phân bố Bernoulli với tham số p . Tần số xuất hiện dấu hiệu A của mẫu là

$$r = X_1 + X_2 + \dots + X_n. \quad (4.14)$$

Tần suất mẫu

$$f = \frac{r}{n} = \bar{X} \quad (4.15)$$

Như vậy tần suất mẫu là trung bình mẫu của biến ngẫu nhiên X có phân bố Bernoulli tham số p .

Tương tự ta có công thức tính kỳ vọng và phương sai của tần suất mẫu:

$$E(f) = p; \quad D(f) = \frac{p(1-p)}{n} \quad (4.16)$$

4.3.5 Cách tính giá trị cụ thể của trung bình mẫu \bar{x} và phương sai mẫu có hiệu chỉnh s^2

1. Nếu mẫu chỉ nhận các giá trị x_1, x_2, \dots, x_k với tần số tương ứng r_1, r_2, \dots, r_k thì giá trị trung bình mẫu và phương sai mẫu cụ thể được tính theo công thức

$$\bar{x} = \frac{1}{n} \sum_{i=1}^k r_i x_i, \quad \sum_{i=1}^k r_i = n \quad (4.17)$$

$$s^2 = \frac{1}{n-1} \sum_{i=1}^k r_i (x_i - \bar{x})^2 = \frac{1}{n-1} \left(\sum_{i=1}^k r_i x_i^2 - \frac{\left(\sum_{i=1}^k r_i x_i \right)^2}{n} \right) \quad (4.18)$$

2. Nếu giá trị của mẫu cụ thể được cho dưới dạng bảng phân bố ghép lớp với các khoảng C_1, \dots, C_m và tần số của C_i là r_i thì giá trị trung bình mẫu và phương sai mẫu được tính theo công thức trên, trong đó

$$x_i \text{ là trung điểm của khoảng } C_i. \quad (4.19)$$

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:

$$u_i = \frac{x_i - a}{h} \Rightarrow x_i = h u_i + a \Rightarrow \bar{x} = h \bar{u} + a; \quad s^2 = h^2 s_u^2 \quad (4.20)$$

trong đó

$$\bar{u} = \frac{1}{n} \sum_{i=1}^k r_i u_i, \quad s_u^2 = \frac{1}{n-1} \sum_{i=1}^k r_i (u_i - \bar{u})^2 = \frac{1}{n-1} \left(\sum_{i=1}^k r_i u_i^2 - \frac{\left(\sum_{i=1}^k r_i u_i \right)^2}{n} \right) \quad (4.21)$$

Thật vậy:

$$\begin{aligned} \bar{x} &= \frac{1}{n} \sum_{i=1}^k r_i x_i = \frac{1}{n} \sum_{i=1}^k r_i (h u_i + a) = \frac{h}{n} \sum_{i=1}^k r_i u_i + \left(\frac{1}{n} \sum_{i=1}^k r_i \right) a = h \bar{u} + a. \\ s^2 &= \frac{1}{n-1} \sum_{i=1}^k r_i (x_i - \bar{x})^2 = \frac{1}{n-1} \sum_{i=1}^k r_i (h u_i + a - h \bar{u} - a)^2 = \frac{h^2}{n-1} \sum_{i=1}^k r_i (u_i - \bar{u})^2 = h^2 s_u^2. \end{aligned}$$

Các số a và h được chọn phù hợp sao cho \bar{u} , s_u^2 tính dễ dàng hơn.

Thông thường ta chọn a là điểm giữa của các giá trị x_i và h được chọn sao cho phép chia thực hiện dễ dàng.

Ví dụ 4.4: Giá trị trung bình mẫu, phương sai mẫu và độ lệch chuẩn mẫu của mẫu cụ thể cho trong ví dụ 4.4.

Khoảng	tần số r_i	x_i	$u_i = \frac{x_i - 20}{5}$	$r_i u_i$	$r_i u_i^2$
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
Σ	400			– 177,5	873,47

$$\bar{x} = 5\bar{u} + 20 = 5 \cdot \frac{-177,5}{400} + 20 = 17,78. \quad s_u^2 = \frac{1}{399} \cdot \left(873,47 - \frac{(-177,5)^2}{400} \right) = 1,9917$$

$$\Rightarrow s^2 = 5^2 \cdot s_u^2 = 49,79 \Rightarrow s = \sqrt{49,79} = 7,056.$$

4.4 PHÂN BỐ XÁC SUẤT CỦA MỘT SỐ THỐNG KÊ ĐẶC TRƯNG MẪU

4.4.1 Trường hợp biến ngẫu nhiên gốc có phân bố chuẩn

Giả sử dấu hiệu nghiên cứu trong tổng thể có thể xem như một biến ngẫu nhiên X có phân bố chuẩn với kỳ vọng $E X = \mu$ và phương sai $D X = \sigma^2$. Các tham số này có thể đã biết hoặc chưa biết. Từ tổng thể rút ra một mẫu ngẫu nhiên kích thước n :

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

Các biến ngẫu nhiên thành phần X_1, X_2, \dots, X_n độc lập có cùng quy luật phân bố chuẩn $N(\mu, \sigma^2)$ của X . Mọi tổ hợp tuyến tính của các biến ngẫu nhiên có phân bố chuẩn là biến ngẫu nhiên có phân bố chuẩn (công thức 2.38). Vì vậy ta có các kết quả sau:

4.4.1.1 Phân bố của thống kê trung bình mẫu

Trung bình mẫu \bar{X} có phân bố chuẩn với kỳ vọng $E(\bar{X}) = \mu$ và phương sai $D(\bar{X}) = \frac{\sigma^2}{n}$. Áp dụng công thức (2.44) suy ra thống kê sau có phân bố chuẩn tắc $N(0;1)$:

$$U = \frac{(\bar{X} - \mu)\sqrt{n}}{\sigma} \sim N(0;1) \quad (4.22)$$

Trường hợp biến ngẫu nhiên gốc có phân bố tùy ý, áp dụng định lý giới hạn trung tâm (định lý 3.8 chương 3) ta có thể xấp xỉ phân bố của thống kê $U = \frac{(\bar{X} - \mu)\sqrt{n}}{\sigma}$ với phân bố chuẩn tắc $N(0;1)$ khi n đủ lớn, trong thực tế với $n \geq 30$ ta có thể xấp xỉ $U = \frac{(\bar{X} - \mu)\sqrt{n}}{\sigma} \approx N(0;1)$.

Ví dụ 4.4: Chiều cao X của các nam sinh viên đại học là biến ngẫu nhiên có phân bố chuẩn với trung bình 163cm và độ lệch chuẩn 3cm. Lấy 80 mẫu của mẫu ngẫu nhiên 25 sinh viên

- Tìm kỳ vọng và phương sai của trung bình mẫu.
- Có bao nhiêu mẫu trong số 80 mẫu lấy giá trị trung bình trong khoảng từ 161,8 cm đến 163,3 cm.
- Có bao nhiêu mẫu trong số 80 mẫu lấy giá trị trung bình nhỏ hơn 161,4 cm.

Giải: a. $E(\bar{X}) = \mu = E(X) = 163\text{cm}$, $D(\bar{X}) = \frac{\sigma^2}{n} = \frac{3^2}{25} = 0,36$.

b. Áp dụng công thức 4.22 ta được $U = \frac{(\bar{X} - 163)}{0,6} \sim N(0;1)$

$$P\{161,86 < \bar{X} < 163,3\} = \Phi(0,5) - \Phi(-1,9) = \Phi(0,5) + \Phi(1,9) - 1 = 0,6627.$$

Vậy số mẫu thỏa mãn điều kiện cần tìm là 80.0,6627 hoặc xấp xỉ 53 mẫu.

c. $P\{\bar{X} < 161,4\} = \Phi(-2,67) = 1 - \Phi(2,67) = 0,0038$. Đây là biến cố có xác suất bé, vì vậy không có mẫu nào trong số 80 mẫu có số đo trung bình nhỏ hơn 161,4 cm.

Thật vậy $80.0,0038 = 0,304 < 1$.

4.4.1.2 Phân bố của thống kê phương sai mẫu S^{*2} .

Từ công thức (4.18) ta có: $nS^{*2} = \sum_{i=1}^n (X_i - \mu)^2$ và $\frac{nS^{*2}}{\sigma^2} = \sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma} \right)^2$.

Vì các biến ngẫu nhiên X_i độc lập nên các biến ngẫu nhiên $\frac{X_i - \mu}{\sigma}$ cũng độc lập. Mặt khác theo công thức (2.44) thì $\frac{X_i - \mu}{\sigma} \sim N(0;1)$. Do đó thống kê $\chi^2 = \frac{nS^{*2}}{\sigma^2}$ có phân bố “khi bình phương” n bậc tự do (công thức 2.51)

$$\chi^2 = \frac{nS^{*2}}{\sigma^2} = \sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma} \right)^2 \sim \chi^2(n) \quad (4.23)$$

4.4.1.3 Phân bố của thống kê phương sai mẫu S^2 .

Ta có
$$\sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma} \right)^2 = \sum_{i=1}^n \left(\frac{X_i - \bar{X}}{\sigma} \right)^2 + \frac{(\bar{X} - \mu)^2}{\sigma^2/n}$$

Thống kê $\frac{(\bar{X} - \mu)^2}{\sigma^2/n}$ có phân bố “khi bình phương” 1 bậc tự do và theo công thức (4.23) ta có $\sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma} \right)^2 \sim \chi^2(n)$. Sử dụng tính chất của phân bố “khi bình phương” (mục 2.3.2.5 chương 2) ta suy ra thống kê $\frac{(n-1)S^2}{\sigma^2}$ có phân bố “khi bình phương” $n-1$ bậc tự do

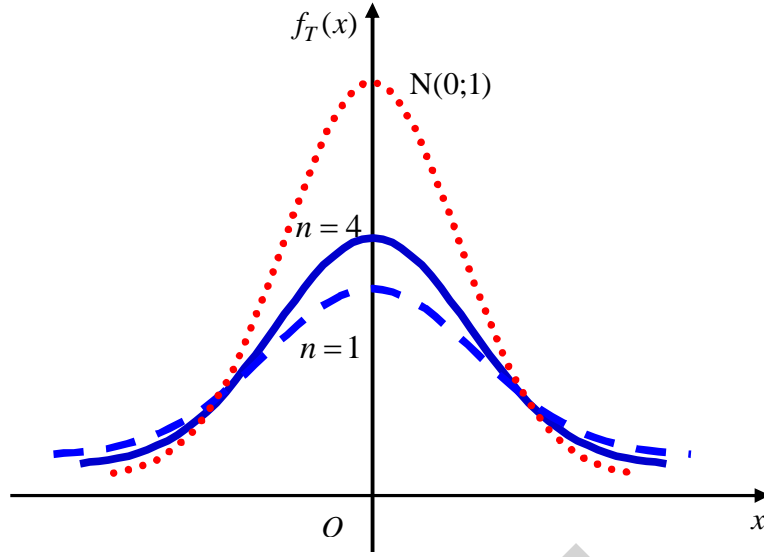
$$\chi^2 = \frac{(n-1)S^2}{\sigma^2} = \sum_{i=1}^n \left(\frac{X_i - \bar{X}}{\sigma} \right)^2 \sim \chi^2(n-1) \quad (4.24)$$

Áp dụng công thức (2.55) với biến ngẫu nhiên U từ công thức (4.22) và $\chi^2 = \frac{(n-1)S^2}{\sigma^2}$ từ công thức (4.24), thì $T = \frac{U}{\sqrt{\chi^2/(n-1)}}$ có phân bố Student $n-1$ bậc tự do.

Vậy

$$T = \frac{(\bar{X} - \mu)\sqrt{n}}{S} = \frac{\frac{(\bar{X} - \mu)\sqrt{n}}{\sigma}}{\sqrt{\frac{1}{n-1} \cdot \frac{(n-1)S^2}{\sigma^2}}} = \frac{U}{\sqrt{\frac{\chi^2}{n-1}}} \sim T(n-1). \quad (4.25)$$

Phân bố Student $T(n)$ hội tụ khá nhanh về phân bố chuẩn tắc $N(0;1)$, do đó trong thực tế khi $n \geq 30$ ta có thể xem thống kê $T(n)$ xấp xỉ $N(0;1)$.



Hình 4.4: Đồ thị hàm mật độ của phân bố Student

4.4.2 Trường hợp biến ngẫu nhiên gốc có phân bố Bernoulli

Giả sử trong tổng thể dấu hiệu nghiên cứu có thể xem như biến ngẫu nhiên có phân bố Bernoulli tham số p . Từ tổng thể rút ra một mẫu ngẫu nhiên kích thước n :

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

Từ công thức (4.14), (4.15), (4.16) ta biết rằng tần suất mẫu $f = \frac{X_1 + \dots + X_n}{n}$ là biến

ngẫu nhiên có kỳ vọng và phương sai: $E(f) = p$; $D(f) = \frac{pq}{n}$.

Áp dụng định lý 4.6 (Định lý Moivre-Laplace) và công thức 4.12 ta có:

$$\text{Với mọi } x \in \mathbb{R}, \lim_{n \rightarrow \infty} P\left\{\frac{(f-p)\sqrt{n}}{\sqrt{pq}} < x\right\} = \Phi(x). \quad (4.26)$$

Như vậy có thể xấp xỉ thống kê $U = \frac{(f-p)\sqrt{n}}{\sqrt{pq}}$ với phân bố chuẩn tắc $\mathbf{N}(0;1)$ khi n đủ lớn. Người ta thấy rằng xấp xỉ là tốt khi $np > 5$ và $nq > 5$ hoặc $npq > 20$.

$$U = \frac{(f-p)\sqrt{n}}{\sqrt{pq}} \sim \mathbf{N}(0;1) \text{ khi } \begin{cases} np > 5 \\ nq > 5 \end{cases} \text{ hoặc } npq > 20. \quad (4.27)$$

Hoặc tính theo tần số $r = X_1 + \dots + X_n \sim \mathbf{B}(n; p)$, theo công thức (3.73) ta cũng có

$$U = \frac{(r-np)}{\sqrt{npq}} \sim \mathbf{N}(0;1) \text{ khi } \begin{cases} np > 5 \\ nq > 5 \end{cases} \text{ hoặc } npq > 20. \quad (4.28)$$

Ví dụ 4.6: Gieo 120 lần đồng xu cân đối đồng chất.

- Tính xác suất có khoảng 40% đến 60% lần số mặt sấp xuất hiện.
- Tính xác suất tỷ lệ mặt sấp xuất hiện lớn hơn hoặc bằng $\frac{5}{8}$.

- c) Một nhóm 500 người, mỗi người gieo 120 lần đồng xu cân đối đồng chất. Có bao nhiêu người có kết quả mặt sấp xuất hiện trong khoảng 40% đến 60%.

Giải: Có thể xem mỗi lần gieo đồng xu là thực hiện phép thử Bernoulli với sự thành công của phép thử là sự xuất hiện mặt sấp, từ giả thiết ta có xác suất thành công của phép thử là 0,5. Như vậy biến ngẫu nhiên gốc X có phân bố Bernoulli tham số 0,5.

Gieo 120 lần là lấy mẫu ngẫu nhiên với kích thước 120 của biến ngẫu nhiên gốc, do đó tần suất mẫu $f = \frac{X_1 + \dots + X_{120}}{120}$.

Ta có $np = nq = 120 \cdot 0,5 = 60$, $\sqrt{npq} = 5,48$ thỏa mãn điều kiện kích thước đủ lớn.

- a) 40% và 60% của 120 bằng 48 và 72. Áp dụng công thức (4.41) và (2.83) ta có

$$P\{48 < r < 72\} \approx \Phi\left(\frac{72 + 0,5 - 60}{5,48}\right) - \Phi\left(\frac{48 - 0,5 - 60}{5,48}\right) \\ = \Phi(2,28) - \Phi(-2,28) = 2\Phi(2,28) - 1 = 0,9774.$$

- b) $\frac{5}{8} \cdot 120 = 75$, vậy xác suất tỷ lệ mặt sấp xuất hiện lớn hơn hoặc bằng $\frac{5}{8}$ là

$$P\{r > 75 - 0,5\} \approx 1 - \Phi\left(\frac{74,5 - 60}{5,48}\right) = 1 - \Phi(2,65) = 1 - 0,9960 = 0,0040.$$

- c) Theo ý a) xác suất gieo 120 lần đồng xu (mẫu ngẫu nhiên kích thước 120) với 40% đến 60% lần mặt sấp xuất hiện là 0,9774. Vậy 500 người thực hiện 120 lần gieo đồng xu (500 quan sát cụ thể của mẫu ngẫu nhiên kích thước 120) thì số người có kết quả gieo với số mặt sấp xuất hiện trong khoảng 40% đến 60% là $500 \cdot 0,9774 = 488,7 \approx 489$.

TÓM TẮT

Mẫu ngẫu nhiên kích thước n của biến ngẫu nhiên gốc X là véc tơ ngẫu nhiên $W = (X_1, \dots, X_n)$, trong đó các biến ngẫu nhiên thành phần X_1, \dots, X_n độc lập và có cùng phân bố với biến ngẫu nhiên gốc X . Mỗi thống kê là một hàm của mẫu $T = T(X_1, \dots, X_n)$.

Các thống kê thường gặp: trung bình mẫu \bar{X} , phương sai mẫu có hiệu chỉnh S^2 và tần suất mẫu $f = \bar{X}$ của dấu hiệu định tính được đặc trưng bởi biến ngẫu nhiên gốc X có phân bố Bernoulli.

Áp dụng công thức (4.17), (4.18) để tính giá trị cụ thể của trung bình mẫu và phương sai mẫu. Trường hợp giá trị cụ thể của không gợn ta sử dụng công thức (4.20) tính theo mẫu rút gọn.

Nếu biến ngẫu nhiên gốc X có phân bố chuẩn $N(\mu, \sigma^2)$ thì trung bình mẫu \bar{X} cũng có phân bố chuẩn $N(\mu, \sigma^2/n)$, vậy $U = \frac{(\bar{X} - \mu)\sqrt{n}}{\sigma} \sim N(0;1)$.

Thống kê $\frac{(n-1)S^2}{\sigma^2}$ có phân bố “khi bình phương” $n-1$ bậc tự do, trong đó S^2 là phương sai mẫu có hiệu chỉnh, do đó $\frac{(\bar{X} - \mu)\sqrt{n}}{S} \sim T(n-1)$.

Áp dụng định lý Moivre-Laplace cho tần suất mẫu ta có thể xấp xỉ thống kê của tần suất mẫu với phân bố chuẩn tắc $U = \frac{(f - p)\sqrt{n}}{\sqrt{pq}} \sim N(0;1)$.

CÂU HỎI ÔN TẬP VÀ BÀI TẬP CHƯƠNG 4

4.1 Mẫu ngẫu nhiên kích thước n về dấu hiệu nghiên cứu X là một dãy gồm n biến ngẫu nhiên: X_1, X_2, \dots, X_n độc lập cùng phân bố với X .

Đúng ☐ Sai ☐.

4.2 Một thống kê của mẫu ngẫu nhiên là con số cụ thể về dấu hiệu nghiên cứu.

Đúng ☐ Sai ☐.

4.3 Trung bình mẫu của dấu hiệu nghiên cứu có phân bố chuẩn cũng có phân bố chuẩn.

Đúng ☐ Sai ☐.

4.4 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 ☐.

4.5 Tổ chức đồ dùng để biểu diễn mẫu ngẫu nhiên cho dưới dạng bảng phân bố ghép lớp.

Đúng ☐ Sai ☐.

4.6 Nếu biến ngẫu nhiên gốc X có phân bố chuẩn thì phương sai mẫu của X cũng có phân bố chuẩn.

Đúng ☐ Sai ☐.

4.7 Khi bậc tự do của phân bố Student lớn hơn 30, phân bố Student tiệm cận phân bố chuẩn.

Đúng ☐ Sai ☐.

4.8 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.

4.9 Giả sử biến ngẫu nhiên gốc có phân bố chuẩn $N(20;1)$. Chọn mẫu ngẫu nhiên kích thước $n = 100$. Hãy tính xác suất để trung bình mẫu \bar{X} nằm trong khoảng: $19,8 < \bar{X} < 20,2$.

4.10 Một mẫu cụ thể của biến ngẫu nhiên X như sau:

2 ; 3 ; 2 ; 4 ; 1 ; 4 ; 2 ; 2 ; 3 ; 1 ($n = 10$).

- Lập bảng phân bố tần suất.
- Xây dựng hàm phân bố thực nghiệm.

Chương 4: Lý thuyết mẫu

Tính \bar{x} , s^2 , s .

- 4.11** Hãy tính giá trị trung bình mẫu \bar{x} và phương sai mẫu s^2 của mẫu cụ thể có bảng phân bố tần số thực nghiệm sau

x_i	21	24	25	26	28	32	34
r_i	10	20	30	15	10	10	5

- 4.12** Hãy tính giá trị trung bình mẫu \bar{x} và phương sai mẫu s^2 của mẫu cụ thể có bảng phân bố tần số thực nghiệm sau

x_i	18,6	19,0	19,4	19,8	20,2	20,6
r_i	4	6	30	40	18	2

- 4.13** Giá trị quan sát của một mẫu ngẫu nhiên về chiều cao của 100 sinh viên đại học được cho trong bảng phân bố ghép lớp sau:

Chiều cao (inche)	59,5-62,5	62,5-65,5	65,5-68,5	68,5-71,5	71,5-74,5
Tần số	5	18	42	27	8

- a. Vẽ tổ chức đồ
 - b. Tính giá trị quan sát của trung bình mẫu, phương sai mẫu và độ lệch chuẩn mẫu.
- 4.14** Một cỗ máy sản xuất với tỉ lệ phế phẩm 2%. Lấy mẫu ngẫu nhiên 400 sản phẩm, tính xác suất
- a. Có ít nhất 3% phế phẩm.
 - b. Có nhiều nhất 2% phế phẩm.
- 4.15** Trọng lượng sản phẩm do nhà máy sản xuất là biến ngẫu nhiên X có phân bố chuẩn với trọng lượng trung bình 22,40 oz và độ lệch chuẩn 0,048 oz. Lấy mẫu ngẫu nhiên kích thước 36 các sản phẩm.
- a. Tính kỳ vọng và độ lệch chuẩn của thống kê trung bình mẫu \bar{X} .
 - b. Tính $P\{22,39 < \bar{X} < 22,41\}$, $P\{\bar{X} > 22,42\}$, $P(\{\bar{X} < 22,38\} \cup \{\bar{X} > 22,41\})$.
- 4.16** Với 300 mẫu cụ thể của mẫu ngẫu nhiên kích thước 36 ở câu trên, có bao nhiêu mẫu có giá trị trung bình thỏa mãn:
- a. Nằm trong khoảng 22,39 oz đến 22,41 oz.
 - b. Lớn hơn 22,42 oz.
 - c. Nhỏ hơn 22,38oz hoặc lớn hơn 22,41oz.
- 4.17** Tuổi thọ của bóng đèn do nhà máy sản xuất với giá trị trung bình $\mu = 800$ giờ và độ lệch chuẩn $\sigma = 56$ giờ. Lấy mẫu ngẫu nhiên kích thước 64, tính xác suất của trung bình mẫu \bar{X} có giá trị:
- a. Trong khoảng từ 790 giờ đến 810 giờ.
 - b. Nhỏ hơn 795 giờ.

- c. Lớn hơn 820 giờ.
- d. Tìm ε thỏa mãn $P\{|\bar{X} - \mu| \leq \varepsilon\} \geq 0,95$.

4.18 Giả sử đứa trẻ ra đời là trai hay gái đồng khả năng. Lấy mẫu ngẫu nhiên 200 đứa trẻ sắp chào đời. Tính xác suất các trường hợp sau:

- a. Số bé trai ít hơn 40%.
- b. Số bé gái trong khoảng từ 43% đến 57%.
- c. Số bé trai nhiều hơn 54%.

4.19 Một bảng phân bố ghép lớp tải trọng tối đa của các dây cáp do một nhà máy sản xuất như sau

Tải trọng tối đa (tấn)	9,3- 9,8	9,8- 10,3	10,3- 10,8	10,8- 11,3	11,3- 11,8	11,8- 12,3	12,3- 12,8	12,8- 13,3
Tần số	2	5	12	17	14	6	3	1

- a. Vẽ tổ chức đồ
- b. Tính giá trị cụ thể của trung bình mẫu và độ lệch chuẩn mẫu.

4.20 Một mẫu cụ thể về trọng lượng sản phẩm do nhà máy sản xuất được cho trong bảng sau.

Trọng lượng (gram)	48	51	54	57	60	63	66	69	72	75	78	81
Tần số	2	6	8	15	42	68	49	25	18	12	4	1

- a. Tính giá trị cụ thể của trung bình mẫu và độ lệch chuẩn mẫu.
- b. Tính tỷ lệ phần trăm số sản phẩm có trọng lượng thuộc khoảng $(\bar{x} \pm s)$, $(\bar{x} \pm 2s)$, $(\bar{x} \pm 3s)$
- c. Giả sử trọng lượng sản phẩm có phân bố chuẩn, so sánh kết quả b) với cách tính dựa vào xác suất của thống kê mẫu tương ứng.

CHƯƠNG 5: ƯỚC LƯỢNG THAM SỐ VÀ KIỂM ĐỊNH GIẢ THIẾT THỐNG KÊ

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 giá trị quan sát của một thống kê để ước lượng một tham số (véc tơ tham số) nào đó theo các tiêu chuẩn: vững, không chệch, hiệu quả. Có nhiều phương pháp khác nhau để ước lượng điểm.

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 cho trường hợp một tham số. Khoảng tin cậy là khoảng mà tham số của dấu hiệu nghiên cứu của tổng thể rơi vào khoảng này với xác suất bằng độ tin cậy.

Trong chương này ta sẽ xây dựng ước lượng điểm và khoảng tin cậy cho kỳ vọng của dấu hiệu nghiên cứu có phân bố chuẩn và tần suất của tổng thể.

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.

5.1 PHƯƠNG PHÁP ƯỚC LƯỢNG ĐIỂM

5.1.1 Khái niệm 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 $W = (X_1, X_2, \dots, X_n)$, thống kê ước lượng cho tham số θ được ký hiệu:

$$\hat{\theta} = T(X_1, X_2, \dots, X_n)$$

Khi đó với mẫu cụ thể $w = (x_1, x_2, \dots, x_n)$ giá trị cụ thể của thống kê $\hat{\theta} = T(x_1, x_2, \dots, x_n)$ là ước lượng của tham số θ của tổng thể.

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.

5.1.2 Ước lượng không chệch (unbiased estimator)

Thống kê $\hat{\theta} = T(X_1, X_2, \dots, X_n)$ là một hàm của các biến ngẫu nhiên X_1, X_2, \dots, X_n , do đó thống kê cũng là một biến ngẫu nhiên. Vì vậy ta có thể xét hàm phân bố và các đặc trưng của thống kê này.

Định nghĩa 5.1: Thống kê $\hat{\theta} = T(X_1, X_2, \dots, X_n)$ được gọi là ước lượng không chệch của tham số θ của tổng thể nếu

$$E[\hat{\Theta}(X_1, X_2, \dots, X_n)] = \theta \quad (5.1)$$

Trường hợp $E(\hat{\Theta}) \neq \theta$ thì $\hat{\Theta} = T(X_1, X_2, \dots, X_n)$ được gọi là một *ước lượng chệch* của θ .

Như vậy ước lượng không chệch có trung bình bằng tham số cần ước lượng.

Ví dụ 5.1: Dựa vào các công thức (4.8), (4.13), (4.17) của lý thuyết mẫu ta có các kết quả sau:

- Trung bình mẫu \bar{X} là ước lượng không chệch của kỳ vọng μ của biến ngẫu nhiên gốc của tổng thể.
- Phương sai mẫu S^2 và S^{*2} là ước lượng không chệch cho phương sai σ^2 của biến ngẫu nhiên gốc của tổng thể.
- Tần suất mẫu f là ước lượng không chệch của tần suất p của tổng thể.

5.1.3 Ước lượng hiệu quả (efficient estimator)

Điều kiện (5.1) 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 trung bình là bé nhất.

Định nghĩa 5.2: *Ước lượng không chệch có phương sai nhỏ nhất so với mọi ước lượng không chệch khác được xây dựng trên cùng một mẫu ngẫu nhiên gọi là ước lượng hiệu quả (hay ước lượng phương sai bé nhất).*

Như vậy, để xét xem ước lượng không chệch $\hat{\Theta}$ có phải là ước lượng hiệu quả của tham số θ hay không ta cần phải tìm một cận dưới các phương sai của các ước lượng không chệch và so sánh phương sai của thống kê $\hat{\Theta}$ với cận dưới này. Điều này được giải quyết bằng bất đẳng thức Cramer-Rao phát biểu như sau:

Cho mẫu ngẫu nhiên $W = (X_1, X_2, \dots, X_n)$ được lấy từ tổng thể có dấu hiệu nghiên cứu là biến ngẫu nhiên X mà hàm mật độ xác suất (hay hàm khối lượng xác suất) $f(x, \theta)$ thỏa mãn một số điều kiện nhất định (thường được thỏa mãn trong thực tế, ít ra là các phân bố xác suất đã xét trong chương 2) và $\hat{\Theta}$ là ước lượng không chệch bất kỳ của tham số θ thì

$$D(\hat{\Theta}) \geq \frac{1}{nE\left(\frac{\partial(\ln f(X, \theta))}{\partial \theta}\right)^2} \quad (5.2)$$

Ví dụ 5.2: Dựa vào bất đẳng thức Cramer-Rao ta có thể chứng minh được rằng trung bình mẫu \bar{X} là ước lượng hiệu quả của kỳ vọng μ của dấu hiệu nghiên cứu X của tổng thể có phân bố chuẩn $N(\mu, \sigma^2)$.

Giải: Thật vậy theo công thức (4.8) ta có $D(\bar{X}) = \frac{\sigma^2}{n}$. Mặt khác hàm mật độ của $X \sim N(\mu, \sigma^2)$ có dạng

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

$$\Rightarrow \ln f(x, \mu) = -\ln(\sigma\sqrt{2\pi}) - \frac{(x-\mu)^2}{2\sigma^2} \Rightarrow \frac{\partial \ln f(x, \mu)}{\partial \mu} = \frac{x-\mu}{\sigma^2}$$

$$\text{Vậy } nE\left(\frac{\partial(\ln f(X, \mu))}{\partial \mu}\right)^2 = nE\left(\frac{X-\mu}{\sigma^2}\right)^2 = \frac{n}{\sigma^4}E(X-\mu)^2 = \frac{n}{\sigma^4}D(X) = \frac{n}{\sigma^2}.$$

Như vậy $D(\bar{X}) = \frac{\sigma^2}{n}$ đạt giá trị cực tiểu của bất đẳng thức Cramer-Rao, do đó trung bình mẫu \bar{X} là ước lượng hiệu quả của μ .

5.1.4 Ước lượng vững (consistent estimator)

Định nghĩa 5.3: Thống kê $\hat{\Theta} = T(X_1, X_2, \dots, X_n)$ được gọi là ước lượng vững của tham số θ của tổng thể nếu $\hat{\Theta} = T(X_1, X_2, \dots, X_n)$ hội tụ theo xác suất đến θ khi $n \rightarrow \infty$.

Nghĩa là với mọi $\varepsilon > 0$, luôn có

$$\lim_{n \rightarrow \infty} P\left\{\left|\hat{\Theta}(X_1, X_2, \dots, X_n) - \theta\right| < \varepsilon\right\} = 1 \quad (5.3)$$

Hoặc một cách tương đương

$$\lim_{n \rightarrow \infty} P\left\{\left|\hat{\Theta}(X_1, X_2, \dots, X_n) - \theta\right| \geq \varepsilon\right\} = 0. \quad (5.4)$$

Theo hệ quả 3.2 của luật số lớn Trêbusep và công thức (3.67) chương 3, ta có trung bình mẫu \bar{X} là ước lượng vững của kỳ vọng μ ; S^2 , \hat{S}^2 và S^{*2} (trường hợp kỳ vọng μ đã biết) là ước lượng vững của phương sai σ^2 của biến ngẫu nhiên gốc X của tổng thể.

Theo luật số lớn Bernoulli (công thức 3.68) tần suất mẫu f là ước lượng vững của tần suất p của tổng thể. Tần suất mẫu f là trung bình mẫu của biến ngẫu nhiên của tổng thể có phân bố Bernoulli, do đó là ước lượng không chệch. Bằng cách sử dụng bất đẳng thức Cramer-Rao ta cũng chứng minh được tần suất mẫu f là ước lượng hiệu quả của tần suất tổng thể.

Tóm lại ta có kết quả sau:

- Trung bình mẫu \bar{X} là ước lượng không chệch, hiệu quả và vững của kỳ vọng μ của biến ngẫu nhiên gốc của tổng thể.
- Tần suất mẫu f là ước lượng không chệch, hiệu quả và vững của xác suất p của tổng thể.
- Phương sai mẫu S^2 và S^{*2} (trường hợp μ đã biết) là ước lượng không chệch và vững của phương sai σ^2 của biến ngẫu nhiên gốc của tổng thể.

5.2 PHƯƠNG PHÁP ƯỚC LƯỢNG BẰNG KHOẢNG TIN CẬY

Phương pháp ước lượng điểm nói trên có nhược điểm là khi kích thước mẫu bé thì ước lượng điểm có thể sai lệch rất nhiều so với giá trị của tham số cần ước lượng. Mặt khác phương pháp trên cũng không thể đánh giá được khả năng mắc sai lầm khi ước lượng là bao nhiêu. Do đó khi kích thước mẫu bé người ta thường dùng phương pháp ước lượng khoảng tin cậy. Theo phương pháp này từ mẫu ngẫu nhiên ta có thể tìm được khoảng $[a; b]$ chứa tham số θ với xác

suất β đủ lớn cho trước (β được gọi là độ tin cậy và thường được chọn trong khoảng 0,95 đến 0,99).

5.2.1 Khái niệm khoảng tin cậy

Định nghĩa 5.4: Khoảng $[a; b]$ có hai đầu mút là hai thống kê

$$a = a(X_1, X_2, \dots, X_n), \quad b = b(X_1, X_2, \dots, X_n) \quad (5.5)$$

phụ thuộc mẫu ngẫu nhiên $W = (X_1, X_2, \dots, X_n)$ của biến ngẫu nhiên gốc X , gọi là khoảng tin cậy của tham số θ với độ tin cậy β nếu:

$$P\{a(X_1, X_2, \dots, X_n) \leq \theta \leq b(X_1, X_2, \dots, X_n)\} = \beta \quad (5.6)$$

Trong thực tế thường yêu cầu độ tin cậy β khá lớn, khi đó theo nguyên lý xác suất lớn biến cố $\{a(X_1, X_2, \dots, X_n) \leq \theta \leq b(X_1, X_2, \dots, X_n)\}$ hầu như chắc chắn sẽ xảy ra trong một phép thử.

Tiến hành một phép thử với mẫu ngẫu nhiên $W = (X_1, X_2, \dots, X_n)$ ta thu được giá trị cụ thể của mẫu $w = (x_1, x_2, \dots, x_n)$, tính được giá trị quan sát $a = a_{qs}(x_1, x_2, \dots, x_n)$, $b = b_{qs}(x_1, x_2, \dots, x_n)$.

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 $a \leq \theta \leq b$.

5.2.2 Khoảng tin cậy của kỳ vọng của biến ngẫu nhiên phân bố chuẩn

Hầu hết các biến ngẫu nhiên của tổng thể có phân bố chuẩn. Chẳng hạn trọng lượng, chiều cao, năng suất, điểm thi ... là các dấu hiệu của tổng thể có phân bố chuẩn.

Ngoài ra theo định lý giới hạn trung tâm khi kích thước mẫu lớn thì trung bình mẫu của biến ngẫu nhiên tổng thể có phân bố tùy ý sẽ xấp xỉ phân bố chuẩn.

Giả sử biến ngẫu nhiên gốc X có phân bố chuẩn $N(\mu; \sigma^2)$ nhưng chưa biết tham số μ của nó. Từ tổng thể rút ra một mẫu ngẫu nhiên kích thước n : $W = (X_1, X_2, \dots, X_n)$, ta tìm khoảng tin cậy của μ trong các trường hợp sau:

5.2.1.1 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:

$$\left[\bar{X} - U_{\alpha/2} \frac{\sigma}{\sqrt{n}}; \bar{X} + U_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right] \quad (5.7)$$

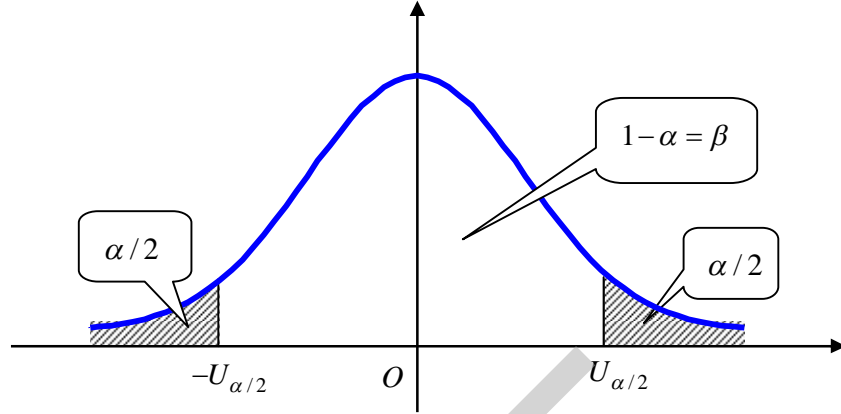
trong đó: $\alpha = 1 - \beta$; $U_{\alpha/2}$ là giá trị tới hạn mức $\frac{\alpha}{2}$ của phân bố chuẩn tắc $N(0;1)$ (công thức 2.42).

Chứng minh: Theo công thức (4.22) ta có $\frac{(\bar{X} - \mu)\sqrt{n}}{\sigma} \sim N(0;1)$.

$$\text{Mặt khác: } \bar{X} - U_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + U_{\alpha/2} \frac{\sigma}{\sqrt{n}} \Leftrightarrow \left| \frac{(\bar{X} - \mu)\sqrt{n}}{\sigma} \right| \leq U_{\alpha/2}.$$

Áp dụng công thức (2.43) ta có

$$P\left[\bar{X} - U_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + U_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right] = P\left\{\left|\frac{(\bar{X} - \mu)\sqrt{n}}{\sigma}\right| \leq U_{\alpha/2}\right\} = 1 - \alpha = \beta.$$



Hình 5.1: Khoảng tin cậy của kỳ vọng phân bố chuẩn

Định nghĩa 5.5: $\varepsilon = U_{\alpha/2} \frac{\sigma}{\sqrt{n}}$ được gọi là *độ chính xác của ước lượng*.

Với phương sai đã biết σ^2 không đổi và độ tin cậy β không đổi thì giá trị tới hạn $U_{\alpha/2}$ không đổi, khi đó độ chính xác chỉ phụ thuộc vào kích thước mẫu n . Khi n càng lớn thì ε càng bé, do đó khoảng ước lượng càng chính xác. Nói cách khác độ chính xác phụ thuộc kích thước mẫu. Khoảng ước lượng có độ chính xác càng bé càng có ý nghĩa, vì vậy trong thực tế người ta mong muốn độ chính xác ở mức ý nghĩa cho trước.

Nếu muốn ước lượng với độ chính xác ε_0 và độ tin cậy β cho trước, *kích thước mẫu cần thiết* là số tự nhiên n nhỏ nhất thỏa mãn:

$$n \geq \frac{\sigma^2 U_{\alpha/2}^2}{\varepsilon_0^2} \quad (5.8)$$

Ví dụ 5.5: 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%

- Hãy tìm khoảng tin cậy của trọng lượng trung bình của loại sản phẩm trên.
- Nếu muốn độ chính xác của ước lượng không vượt quá 0,3 gram thì cần cân thử ít nhất bao nhiêu sản phẩm.

Giải: Gọi X là trọng lượng sản phẩm, theo giả thiết X có phân bố chuẩn $N(\mu; \sigma^2)$ với $\sigma = 1$. Trọng lượng trung bình của sản phẩm là tham số μ . Khoảng tin cậy có dạng (5.7).

Với độ tin cậy $\beta = 0,95 \Rightarrow \frac{\alpha}{2} = 0,025 \Rightarrow U_{\frac{\alpha}{2}} = 1,96$.

a) Từ bảng số liệu tìm được trung bình mẫu cụ thể:

$$\bar{x} = \frac{3 \cdot 18 + 5 \cdot 19 + 15 \cdot 20 + 2 \cdot 21}{25} = 19,64.$$

Độ chính xác của ước lượng $\varepsilon = U_{\alpha/2} \frac{\sigma}{\sqrt{n}} = 1,96 \cdot \frac{1}{\sqrt{25}} = 0,392$.

Vậy với độ tin cậy 95% qua mẫu cụ thể này, khoảng tin cậy của tham số μ là:

$$[19,64 - 0,392; 19,64 + 0,392]$$

hay $19,248 \leq \mu \leq 20,032$.

b) Nếu muốn độ chính xác của ước lượng không vượt quá 0,3 thì cần cần thử ít nhất n sản phẩm sao cho:

$$n \geq \frac{\sigma^2 U_{\alpha/2}^2}{\varepsilon_0^2} = \frac{1 \cdot 1,96^2}{0,3^2} = 42,68. \text{ Chọn } n = 43.$$

5.2.1.2 Trường hợp phương sai σ^2 chưa biết, kích thước mẫu $n \geq 30$

Trong nhiều bài toán thực tế, ta không biết phương sai σ^2 của biến ngẫu nhiên gốc X của tổng thể. Nhưng khi kích thước mẫu n đủ lớn ($n \geq 30$) ta có thể xấp xỉ độ lệch chuẩn σ bởi độ lệch chuẩn mẫu S (vì S^2 là ước lượng vững không chệch của σ^2), S được xác định bởi công thức (4.13). Mặt khác, theo định lý giới hạn trung tâm thì thống kê $\frac{(\bar{X} - \mu)\sqrt{n}}{\sigma}$ có phân bố xấp xỉ phân bố chuẩn, đúng với mọi biến ngẫu nhiên gốc X của tổng thể (không đòi hỏi phân bố chuẩn).

Do đó khoảng tin cậy của tham số μ với độ tin cậy β có thể lấy là

$$\left[\bar{X} - U_{\alpha/2} \frac{S}{\sqrt{n}}; \bar{X} + U_{\alpha/2} \frac{S}{\sqrt{n}} \right] \quad (5.9)$$

Ví dụ 5.6: Để xác định chiều cao trung bình của các cây bạch đàn trong khu rừng rộng trồng bạch đàn, ta tiến hành đo ngẫu nhiên 35 cây và có kết quả cho trong bảng sau:

Khoảng	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

Σ	35			-6,5	15,25
----------	----	--	--	------	-------

$$\bar{u} = \frac{-6,5}{35} = -0,1857 \Rightarrow \bar{x} = 8,25 - 0,1857 \approx 8,06.$$

$$s^2 = s_u^2 = \frac{1}{34} \left(15,25 - \frac{(-6,5)^2}{35} \right) = 0,413 \Rightarrow s = 0,64.$$

Với độ tin cậy $\beta = 95\%$, $U_{\alpha/2} = 1,96$.

Độ chính xác của ước lượng $\varepsilon = U_{\alpha/2} \frac{s}{\sqrt{n}} = 1,96 \cdot \frac{0,64}{\sqrt{35}} = 0,21$.

Vậy khoảng tin cậy cho chiều cao trung bình μ của các cây bạch đàn là:

$$7,85 \leq \mu \leq 8,27.$$

5.2.1.3 Trường hợp phương sai σ^2 chưa biết, kích thước mẫu $n < 30$

Trong trường hợp này, theo công thức (4.25) thống kê

$$T = \frac{(\bar{X} - \mu)\sqrt{n}}{S} \sim T(n-1) \quad (5.10)$$

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:

$$\left[\bar{X} - t_{\alpha/2}(n-1) \frac{S}{\sqrt{n}}; \bar{X} + t_{\alpha/2}(n-1) \frac{S}{\sqrt{n}} \right] \quad (5.11)$$

trong đó $t_{\alpha/2}(n-1)$ là giá trị tới hạn mức $\alpha/2$ của phân bố Student $n-1$ bậc tự do (công thức 2.57).

Độ chính xác của ước lượng:

$$\varepsilon = t_{\alpha/2}(n-1) \frac{S}{\sqrt{n}} \quad (5.12)$$

Với độ tin cậy β và độ chính xác ε_0 cho trước thì kích thước mẫu cần thiết là số tự nhiên n nhỏ nhất thỏa mãn:

$$n \geq \left(\frac{S \cdot t_{\alpha/2}(n-1)}{\varepsilon_0} \right)^2 \quad (5.13)$$

Ví dụ 5.7: Năng suất của một loại giống mới là một biến ngẫu nhiên có quy luật phân bố chuẩn $N(\mu; \sigma^2)$. Gieo thử giống này trên 16 mảnh vườn thí nghiệm thu được như sau (đơn vị kg/ha):

172, 173, 173, 174, 174, 175, 176, 166, 166, 167, 165, 173, 171, 170, 171, 170.

Hãy tìm khoảng tin cậy cho năng suất trung bình của loại 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: $\bar{x} = 171$; $s = 3,4254$. $\alpha = 0,05$; $\frac{\alpha}{2} = 0,025$.

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:

$$169,115 \leq \mu \leq 172,885.$$

5.2.2 Khoảng tin cậy cho tần suất của tổng thể

Ta cần nghiên cứu một dấu hiệu định tính A nào đó mà mỗi cá thể của tổng thể có thể có hoặc không. Nếu cá thể có dấu hiệu A ta cho nhận giá trị 1, trường hợp ngược lại ta cho nhận giá trị 0. Lúc đó dấu hiệu nghiên cứu có thể xem là biến ngẫu nhiên X có phân bố Bernoulli với tham số p . Kỳ vọng $EX = p$ và phương sai $DX = p(1-p)$.

Lấy mẫu ngẫu nhiên $W = (X_1, X_2, \dots, X_n)$ của biến ngẫu nhiên gốc X có phân bố Bernoulli với tham số p , khi đó X_1, X_2, \dots, X_n là các biến ngẫu nhiên độc lập có cùng phân bố Bernoulli tham số p .

Tần suất mẫu $f = \frac{X_1 + \dots + X_n}{n}$.

Theo định lý Moivre-Laplace và công thức (4.27) ta có thể xấp xỉ thống kê $U = \frac{(f-p)\sqrt{n}}{\sqrt{p(1-p)}}$ với phân bố chuẩn tắc $N(0;1)$ khi n đủ lớn.

Tuy nhiên vì p chưa biết nên chưa biết $p(1-p) = DX$.

Mặt khác theo công thức (5.6), (5.17), (5.18) thì tần suất mẫu f là ước lượng vững, không chệch và hiệu quả của xác suất p tổng thể. Vì vậy khi n đủ lớn ta có thể thay p bằng f .

Do đó khoảng tin cậy cho xác suất p của tổng thể với độ tin cậy β là:

$$\left[f - U_{\alpha/2} \sqrt{\frac{f(1-f)}{n}} ; f + U_{\alpha/2} \sqrt{\frac{f(1-f)}{n}} \right] \quad (5.14)$$

Với điều kiện n đủ lớn thỏa mãn

$$\begin{cases} nf > 10 \\ n(1-f) > 10 \end{cases} \quad (5.15)$$

trong đó $U_{\alpha/2}$ là giá trị tới hạn mức $\alpha/2$ của phân bố chuẩn tắc $N(0;1)$ với $\alpha = 1 - \beta$.

Độ chính xác của khoảng tin cậy:

$$\varepsilon = U_{\alpha/2} \sqrt{\frac{f(1-f)}{n}}. \quad (5.16)$$

Với độ tin cậy β và độ chính xác ε_0 cho trước thì kích thước mẫu cần thiết là số tự nhiên n nhỏ nhất thỏa mãn:

$$n \geq f(1-f) \left(\frac{U_{\alpha/2}}{\varepsilon_0} \right)^2 \quad (5.17)$$

trong đó f là tần suất mẫu của một mẫu ngẫu nhiên nào đó.

Ví dụ 5.8: Trong đợt vận động bầu cử tổng thống ở một nước nọ, người ta phỏng vấn ngẫu nhiên 1600 cử tri, được biết có 960 người trong số đó sẽ bỏ phiếu cho ứng cử viên A.

a) Với độ tin cậy $\beta = 95\%$ thì ứng cử viên A sẽ chiếm được tối thiểu bao nhiêu % số phiếu bầu.

b) Với độ tin cậy $\beta = 95\%$, nếu muốn độ chính xác của ước lượng không vượt quá 0,02 thì cần phỏng vấn tối thiểu bao nhiêu cử tri.

Giải: Gọi p là tỉ lệ số phiếu sẽ bầu cho ứng cử viên A. Tổng thể nghiên cứu là tập hợp tất cả các cử tri. Dấu hiệu nghiên cứu là cử tri sẽ bỏ phiếu cho A, có thể xem là biến ngẫu nhiên có phân bố Bernoulli tham số p . Khoảng tin cậy cho p có dạng (5.14) với điều kiện (5.15).

a) Từ mẫu cụ thể trên ta có $f = \frac{960}{1600} = 0,6$ thỏa mãn điều kiện $\begin{cases} nf = 960 > 10 \\ n(1-f) = 640 > 10 \end{cases}$.

Độ chính xác của ước lượng $\varepsilon = U_{\alpha/2} \sqrt{\frac{f(1-f)}{n}} = 1,96 \sqrt{\frac{0,6 \cdot 0,4}{1600}} = 0,024$.

Khoảng tin cậy: $0,576 \leq p \leq 0,624$.

Vậy với độ tin cậy 95% thì tối thiểu có 57,6% cử tri sẽ bỏ phiếu cho ứng cử viên A.

b) Theo công thức 5.26 ta có

$$n \geq 0,6 \cdot 0,4 \left(\frac{1,96}{0,02} \right)^2 = 2304,96 \text{ chọn } n = 2.305.$$

5.3 KHÁI NIỆM CHUNG KIỂM ĐỊNH GIẢ THIẾT THỐNG KÊ

Trong 5.1, 5.2 ta đã giải quyết bài toán ước lượng các tham số đặc trưng của biến ngẫu nhiên gốc của tổng thể. Trong mục này ta sẽ sử dụng phương pháp kiểm định giả thiết thống kê để kiểm định các tham số đặc trưng của tổng thể.

Trước hết ta tìm hiểu các khái niệm sau.

5.3.1 Giả thiết thống kê

Giả thiết thống kê là giả thiết về biến ngẫu nhiên gốc của tổng thể, bao gồm: 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*” (Null hypothesis). 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 có tính chất đối chứng, cạnh tranh với H_0 gọi là *đối thiết* (Alternative hypothesis), ký hiệu H_1 . Đối thiết H_1 sẽ được chấp nhận khi H_0 bị bác bỏ và ngược lại.

Cần chú ý rằng đối thiết H_1 không nhất thiết là phủ định của giả thiết H_0 . Chẳng hạn giả thiết H_0 : nhu cầu thị trường về loại hàng hóa này là $\mu = 1000$ đơn vị/tháng. Nếu ta nghi ngờ

rằng nhu cầu này không đúng thì đối thiết H_1 là $\mu \neq 1000$, nhưng nếu do tiếp thị tốt, do chính sách hậu mãi tốt người ta nghĩ rằng nhu cầu về mặt hàng này tăng lên thì đối thiết H_1 là $\mu > 1000$.

5.3.2 Tiêu chuẩn kiểm định giả thiết thống kê

Quy tắc kiểm định dựa trên hai nguyên lý sau:

- Nguyên lý xác suất nhỏ: "Nếu một biến cố có xác suất rất nhỏ thì với một phép thử biến cố đó coi như không xảy ra".
- Phương pháp phản chứng: "Đề bác bỏ A ta giả sử A đúng; nếu A đúng dẫn đến một điều vô lý thì ta bác bỏ A ".

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 với 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:

Từ biến ngẫu nhiên gốc X của tổng thể lập mẫu ngẫu nhiên $W = (X_1, X_2, \dots, X_n)$. Chọn thống kê

$$T = T(X_1, X_2, \dots, X_n, \theta) \quad (5.18)$$

trong đó θ là tham số liên quan đến giả thiết cần kiểm định. Nếu H_0 đúng thì thống kê T có quy luật phân bố xác suất xác định. Thống kê T được gọi là *tiêu chuẩn kiểm định*.

5.3.3 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 giả thiết 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 :

$$P\{T \in W_\alpha | H_0\} = \alpha \quad (5.19)$$

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

5.3.4 Giá trị quan sát của tiêu chuẩn kiểm định

Thực hiện phép thử với mẫu ngẫu nhiên $W = (X_1, X_2, \dots, X_n)$ thu được mẫu cụ thể $w = (x_1, x_2, \dots, x_n)$, thay giá trị này vào thống kê (5.18) ta được giá trị quan sát của tiêu chuẩn kiểm định:

$$T_{qs} = (x_1, x_2, \dots, x_n, \theta_0) \quad (5.20)$$

5.3.5 Quy tắc kiểm định giả thiết thống kê

So sánh giá trị quan sát của tiêu chuẩn kiểm định với miền bác bỏ W_α và kết luận theo quy tắc sau:

1. Nếu giá trị quan sát T_{qs} rơi vào miền bác bỏ W_α , theo nguyên tắc kiểm định thì H_0 sai, do đó ta bác bỏ H_0 thừa nhận H_1 .

2. Nếu giá trị quan sát T_{qs} không rơi vào miền bác bỏ W_α thì điều này chưa khẳng định rằng H_0 đúng mà chỉ có nghĩa là qua mẫu cụ thể 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).

5.3.6 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 phát sinh 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:

$$P\{T \notin W_\alpha | H_1\} = \beta \quad (5.21)$$

Xác suất của biến cố đối của sai lầm loại II: $P\{T \in W_\alpha | H_1\} = 1 - \beta$ gọi là *lực lượng của 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 mong muốn tìm một qui tắc kiểm định mà cả hai loại sai lầm trên là cực tiểu. Nhưng không tồn tại kiểm định lý tưởng như vậy, vì nói chung khi giảm sai lầm loại I thì sai lầm loại II tăng và ngược lại. Chẳng hạn nếu lấy $\alpha = 0$ thì sẽ không bác bỏ bất kỳ giả thiết nào, kể cả giả thiết sai, vậy β sẽ đạt cực đại. Mặt khác trong bài toán kiểm định thì giả thiết H_0 là giả thiết quan trọng, do đó sai lầm về nó càng nhỏ càng tốt. Vì vậy các nhà thống kê đưa ra phương pháp sau:

Sau khi ta chọn sai lầm loại I nhỏ ở mức ý nghĩa α , với mẫu kích thước n xác định, ta chọn ra miền bác bỏ W_α sao cho xác suất sai lầm loại II là nhỏ nhất hay lực lượng kiểm định là lớn nhất. Nghĩa là cần tìm miền bác bỏ W_α thỏa mãn điều kiện:

$$P\{T \in W_\alpha | H_0\} = \alpha \quad \text{và} \quad P\{T \in W_\alpha | H_1\} = 1 - \beta \rightarrow \max$$

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.

5.3.7 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:

- Phát biểu giả thiết H_0 và đối thiết H_1 .
- Từ tổng thể nghiên cứu lập mẫu ngẫu nhiên kích thước n .
- 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.
- Với mức ý nghĩa α , xác định miền bác bỏ W_α tốt nhất tùy thuộc vào đối thiết H_1 .
- Từ mẫu cụ thể tính giá trị quan sát của tiêu chuẩn kiểm định T_{qs} .
- So sánh giá trị quan sát của tiêu chuẩn kiểm định T_{qs} với miền bác bỏ W_α và kết luận.

5.4 KIỂM ĐỊNH THAM SỐ

Trong mục này ta xét bài toán kiểm định tham số của biến ngẫu nhiên gốc có phân bố chuẩn (dấu hiệu định lượng) và biến ngẫu nhiên gốc có phân bố Bernoulli (dấu hiệu định tính).

5.4.1 Kiểm định giả thiết về kỳ vọng của biến ngẫu nhiên có phân bố chuẩn

Giả sử biến ngẫu nhiên gốc X trong tổng thể có phân bố chuẩn $N(\mu; \sigma^2)$, 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

$$H_0: \mu = \mu_0.$$

Tiêu chuẩn kiểm định và miền bác bỏ phụ thuộc các trường hợp sau:

5.4.1.1 Trường hợp đã biết phương sai

Giả sử phương sai σ^2 của biến ngẫu nhiên gốc X trong tổng thể có phân bố chuẩn $N(\mu; \sigma^2)$ đã biết. Từ tổng thể rút ra một mẫu ngẫu nhiên kích thước n : $W = (X_1, X_2, \dots, X_n)$.

Tiêu chuẩn kiểm định

$$T = \frac{(\bar{X} - \mu)\sqrt{n}}{\sigma} \quad (5.22)$$

Theo công thức (4.29) chương 4 thống kê $T = \frac{(\bar{X} - \mu)\sqrt{n}}{\sigma}$ có phân bố chuẩn tắc $N(0;1)$.

Mặt khác nếu giả thiết H_0 đúng thì thống kê T trở thành $T = \frac{(\bar{X} - \mu_0)\sqrt{n}}{\sigma}$, do đó

$$\{T \in W_\alpha | H_0\} = \left\{ \frac{(\bar{X} - \mu_0)\sqrt{n}}{\sigma} \in W_\alpha \right\}$$

là biến cố thống kê T có giá trị rơi vào miền W_α với điều kiện giả thiết H_0 đúng.

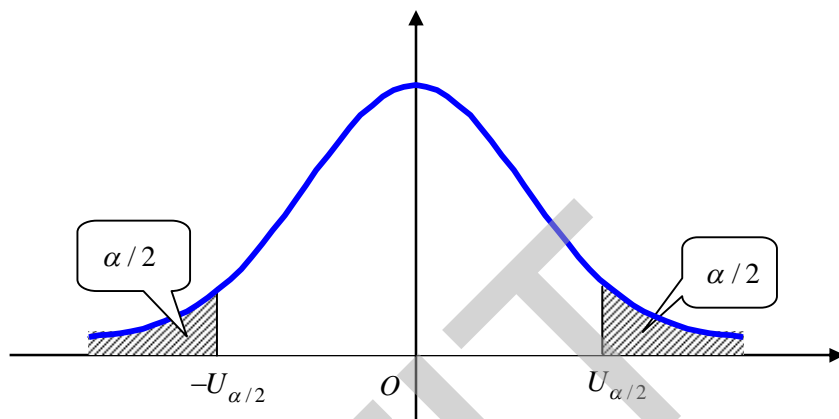
Áp dụng công thức 3.24-3.25 chương 3 ta có thể xây dựng các miền bác bỏ dựa vào đối thiết H_1 như sau:

a. Bài toán 1: $H_0: \mu = \mu_0$; $H_1: \mu \neq \mu_0$. Ta nói đây là bài toán kiểm định hai phía.

$$\text{Miền bác bỏ} \quad W_\alpha = \left\{ T = \frac{(\bar{X} - \mu_0)\sqrt{n}}{\sigma}; |T| > U_{\alpha/2} \right\}. \quad (5.23)$$

trong đó $U_{\alpha/2}$, U_α là giá trị tới hạn mức $\alpha/2$ và mức α của phân bố chuẩn tắc $N(0;1)$.

$$T = \frac{(\bar{X} - \mu_0)\sqrt{n}}{\sigma} \sim N(0;1) \Rightarrow P(W_\alpha) = P\left\{ \left| \frac{(\bar{X} - \mu_0)\sqrt{n}}{\sigma} \right| > U_{\alpha/2} \right\} = \alpha$$



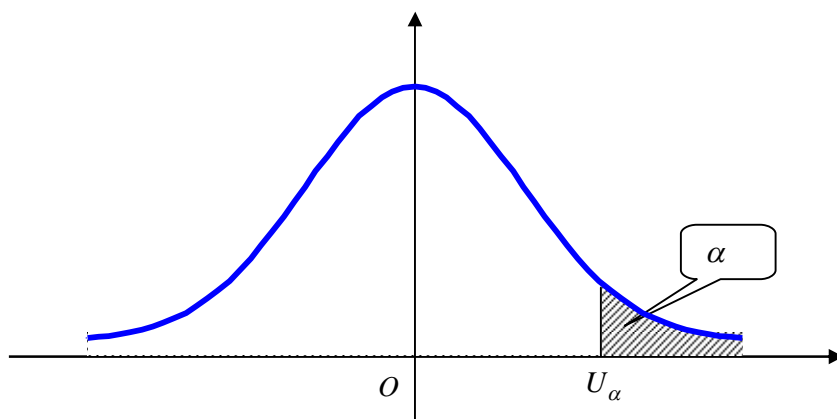
Hình 5.1: Miền bác bỏ của kỳ vọng phân bố chuẩn, bài toán 1

Như vậy miền bác bỏ W_α là biến cố thống kê $T = \frac{(\bar{X} - \mu_0)\sqrt{n}}{\sigma}$ có giá trị thỏa mãn $|T| > U_{\alpha/2}$.

b. Bài toán 2: $H_0: \mu = \mu_0$; $H_1: \mu > \mu_0$. Đây là bài toán kiểm định một phía.

$$\text{Miền bác bỏ} \quad W_\alpha = \left\{ T = \frac{(\bar{X} - \mu_0)\sqrt{n}}{\sigma}; T > U_\alpha \right\}. \quad (5.24)$$

$$P(W_\alpha) = P\left\{ \frac{(\bar{X} - \mu_0)\sqrt{n}}{\sigma} > U_\alpha \right\} = \alpha$$



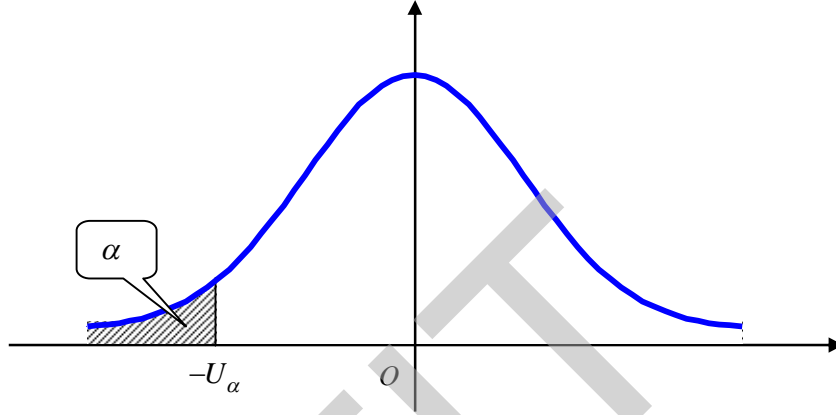
Hình 5.2: Miền bác bỏ của kỳ vọng phân bố chuẩn, bài toán 2

c. Bài toán 3: $H_0: \mu = \mu_0$; $H_1: \mu < \mu_0$. Đây là bài toán kiểm định một phía.

$$\text{Miền bác bỏ} \quad W_\alpha = \left\{ T = \frac{(\bar{X} - \mu_0)\sqrt{n}}{\sigma}; -T > U_\alpha \right\}. \quad (5.25)$$

trong đó $U_{\alpha/2}$, U_α là giá trị tới hạn mức $\alpha/2$ và mức α của phân bố chuẩn tắc $N(0;1)$.

$$P(W_\alpha) = P\left\{ -\frac{(\bar{X} - \mu_0)\sqrt{n}}{\sigma} > U_\alpha \right\} = P\left\{ \frac{(\bar{X} - \mu_0)\sqrt{n}}{\sigma} < -U_\alpha \right\} = \alpha$$



Hình 5.3: Miền bác bỏ của kỳ vọng phân bố chuẩn, bài toán 3

Từ giá trị của mẫu cụ thể $w = (x_1, x_2, \dots, x_n)$, tính giá trị quan sát của tiêu chuẩn kiểm định

$$T_{qs} = \frac{(\bar{x} - \mu_0)\sqrt{n}}{\sigma} \text{ và so sánh với miền bác bỏ } W_\alpha \text{ để kết luận.}$$

5.4.1.2 Trường hợp chưa biết phương sai, kích thước mẫu $n \geq 30$

Trường hợp phương sai σ^2 chưa biết, ta thay độ lệch chuẩn σ bằng độ lệch chuẩn mẫu S (công thức 4.16). Khi kích thước n đủ lớn ($n \geq 30$) thống kê

$$T = \frac{(\bar{X} - \mu)\sqrt{n}}{S} \quad (5.26)$$

xấp xỉ phân bố chuẩn tắc $N(0;1)$ (điều này cũng đúng ngay cả khi biến ngẫu nhiên gốc không có phân bố chuẩn, xem công thức (3.69)).

$$\text{Khi giả thiết } H_0 \text{ đúng thì thống kê } T \text{ trở thành } T = \frac{(\bar{X} - \mu_0)\sqrt{n}}{S}$$

Ta xây dựng các miền bác bỏ dựa vào đối thiết H_1 như sau:

a. Bài toán 1: $H_0: \mu = \mu_0$; $H_1: \mu \neq \mu_0$. Ta nói đây là bài toán kiểm định hai phía.

$$\text{Miền bác bỏ} \quad W_\alpha = \left\{ T = \frac{(\bar{X} - \mu_0)\sqrt{n}}{S}; |T| > U_{\alpha/2} \right\}. \quad (5.27)$$

b. Bài toán 2: $H_0: \mu = \mu_0$; $H_1: \mu > \mu_0$. Đây là bài toán kiểm định một phía.

$$\text{Miền bác bỏ} \quad W_{\alpha} = \left\{ T = \frac{(\bar{X} - \mu_0)\sqrt{n}}{S}; T > U_{\alpha} \right\}. \quad (5.28)$$

c. Bài toán 3: $H_0: \mu = \mu_0$; $H_1: \mu < \mu_0$. Đây là bài toán kiểm định một phía.

$$\text{Miền bác bỏ} \quad W_{\alpha} = \left\{ T = \frac{(\bar{X} - \mu_0)\sqrt{n}}{S}; -T > U_{\alpha} \right\}. \quad (5.29)$$

Ví dụ 5.14: 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 $\bar{x} = 6,1$ với độ lệch chuẩn là $s = 2,5$. Với mức ý nghĩa $\alpha = 1\%$ có thể nói rằng lượng hàng bán được trung bình trên mỗi đầu người có sự thay đổi không?

Giải: Gọi μ là lượng hàng bán được trung bình trên mỗi nhân viên bán hàng của hãng buôn.

Ta kiểm định giả thiết $H_0: \mu = 7,4$; đối thiết $H_1: \mu \neq 7,4$.

$$\text{Tiêu chuẩn kiểm định:} \quad T = \frac{(\bar{X} - 7,4)\sqrt{n}}{S}$$

$$\text{Miền bác bỏ:} \quad \alpha = 0,01 \Rightarrow U_{\frac{\alpha}{2}} = 2,575; W_{\alpha} = \left\{ T = \frac{(\bar{X} - 7,4)\sqrt{n}}{S}; |T| > 2,575 \right\}.$$

$$\text{Từ mẫu cụ thể ta có} \quad T_{qs} = \frac{(6,1 - 7,4)\sqrt{40}}{2,5} = -3,289.$$

Với mẫu cụ thể này giá trị quan sát của tiêu chuẩn kiểm định rơi vào miền bác bỏ, vậy ta có thể kết luận rằng số lượng hàng bán được trung bình của mỗi nhân viên bán hàng là có thay đổi.

Ví dụ 5.15: Một công ti có một hệ thống máy tính có thể xử lý 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 lý 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 lý được trong 1 giờ.

Ta kiểm định: Giả thiết $H_0: \mu = 1200$; đối thiết $H_1: \mu > 1200$.

$$\text{Tiêu chuẩn kiểm định} \quad T = \frac{(\bar{X} - 1200)\sqrt{n}}{S}$$

$$\text{Miền bác bỏ:} \quad \alpha = 0,05 \Rightarrow U_{\alpha} = 1,64; W_{\alpha} = \left\{ T = \frac{(\bar{X} - 1200)\sqrt{n}}{S}; T > 1,64 \right\}.$$

$$\text{Thay giá trị cụ thể của mẫu vào công thức (5.26) ta được } T_{qs} = \frac{(1260 - 1200)\sqrt{40}}{215} = 1,76.$$

Với mẫu cụ thể này giá trị quan sát của tiêu chuẩn kiểm định rơi vào miền bác bỏ, vậy ta có thể kết luận hệ thống máy tính mới tốt hơn hệ thống cũ.

5.4.1.3 Trường hợp chưa biết phương sai, kích thước mẫu $n < 30$

Xét thống kê

$$T = \frac{(\bar{X} - \mu)\sqrt{n}}{S} \quad (5.30)$$

theo công thức (4.32) thống kê T có phân bố Student $n-1$ bậc tự do.

Khi giả thiết H_0 đúng thống kê T trở thành

$$T = \frac{(\bar{X} - \mu_0)\sqrt{n}}{S}$$

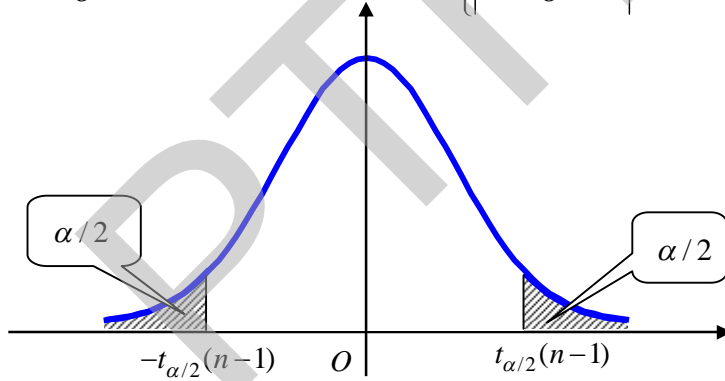
Miền bác bỏ phụ thuộc đối thiết H_1 xác định như sau:

a. Bài toán 1: $H_0: \mu = \mu_0; H_1: \mu \neq \mu_0$.

$$\text{Miền bác bỏ: } W_\alpha = \left\{ T = \frac{(\bar{X} - \mu_0)\sqrt{n}}{S}; |T| > t_{\alpha/2}(n-1) \right\}. \quad (5.31)$$

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.

$$T = \frac{(\bar{X} - \mu_0)\sqrt{n}}{\sigma} \sim \mathbf{T}(n-1) \Rightarrow P(W_\alpha) = P\left\{ \left| \frac{(\bar{X} - \mu_0)\sqrt{n}}{\sigma} \right| > t_{\alpha/2}(n-1) \right\} = \alpha$$



Hình 5.4: Miền bác bỏ của kỳ vọng với σ^2 chưa biết $n < 30$, bài toán 1

b. Bài toán 2: $H_0: \mu = \mu_0; H_1: \mu > \mu_0$.

$$\text{Miền bác bỏ: } W_\alpha = \left\{ T = \frac{(\bar{X} - \mu_0)\sqrt{n}}{S}; T > t_\alpha(n-1) \right\}. \quad (5.32)$$

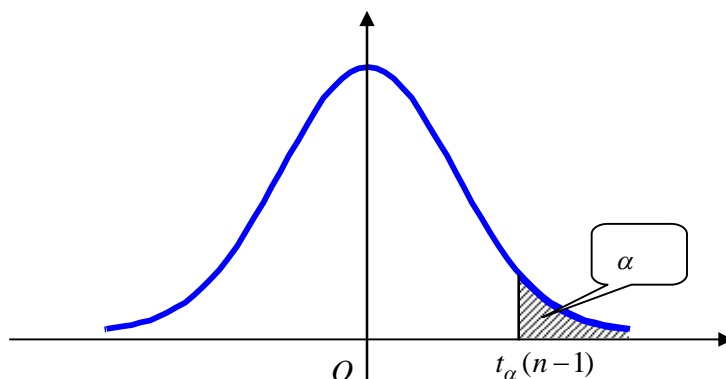
trong đó $t_\alpha(n-1)$ là giá trị tới hạn mức α của phân bố Student $n-1$ bậc tự do.

$$T = \frac{(\bar{X} - \mu_0)\sqrt{n}}{\sigma} \sim \mathbf{T}(n-1) \Rightarrow P(W_\alpha) = P\left\{ \frac{(\bar{X} - \mu_0)\sqrt{n}}{\sigma} > t_\alpha(n-1) \right\} = \alpha \quad (\text{Hình 5.5}).$$

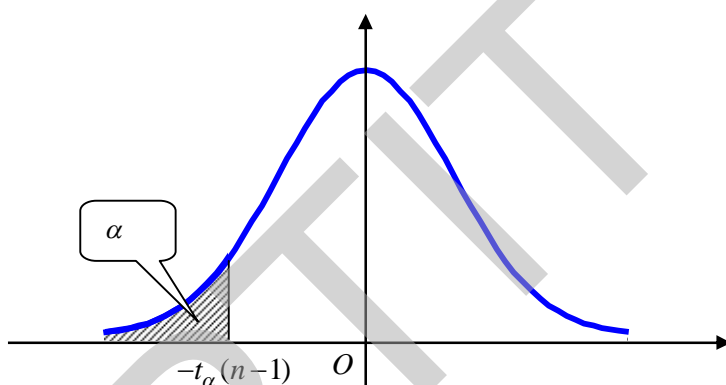
c. Bài toán 3: $H_0: \mu = \mu_0; H_1: \mu < \mu_0$.

$$\text{Miền bác bỏ: } W_\alpha = \left\{ T = \frac{(\bar{X} - \mu_0)\sqrt{n}}{S}; -T > t_\alpha(n-1) \right\}. \quad (5.33)$$

$$P(W_\alpha) = P\left\{-\frac{(\bar{X} - \mu_0)\sqrt{n}}{\sigma} > t_\alpha(n-1)\right\} = P\left\{\frac{(\bar{X} - \mu_0)\sqrt{n}}{\sigma} < -t_\alpha(n-1)\right\} = \alpha \quad (\text{Hình 5.6}).$$



Hình 5.5: Miền bác bỏ của kỳ vọng với σ^2 chưa biết $n < 30$, bài toán 2



Hình 5.6: Miền bác bỏ của kỳ vọng với σ^2 chưa biết $n < 30$, bài toán 3

Ví dụ 5.16: 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 $\alpha = 0,05$. Biết rằng năng suất giống cây trồng là một biến ngẫu nhiên có phân bố chuẩn $N(\mu; \sigma^2)$.

Giải: Gọi μ là năng suất trung bình của loại giống mới.

Ta cần kiểm định giả thiết $H_0: \mu = 21,5$; đối thiết $H_1: \mu \neq 21,5$.

Tiêu chuẩn kiểm định
$$T = \frac{(\bar{X} - 21,5)\sqrt{n}}{S}.$$

Tra bảng ta tính được giá trị tới hạn mức $\frac{\alpha}{2} = 0,025$ của phân bố Student 15 bậc tự do là 2,131.

Do đó miền bác bỏ
$$W_\alpha = \left\{ T = \frac{(\bar{X} - 21,5)\sqrt{n}}{S} ; |T| > 2,131 \right\}.$$

Từ mẫu cụ thể trên tính được:

$$\bar{x} = 20,406, s = 3,038 \Rightarrow T_{qs} = \frac{(20,406 - 21,5)\sqrt{16}}{3,038} = -1,44.$$

Vì $|T_{qs}| = 1,44 < 2,131$ nên chưa có cơ sở để bác bỏ H_0 . Có nghĩa là với số liệu này thì có thể chấp nhận lời quảng cáo của công ty.

5.4.2 Kiểm định tham số của biến ngẫu nhiên phân bố Bernoulli

Ta đề ý đến dấu hiệu định tính là đặc trưng A nào đó mà mỗi cá thể của tổng thể có thể có tính chất này hoặc không. Gọi p là tần suất có đặc trưng A của tổng thể. Như đã thấy trong chương 4, có thể xem dấu hiệu nghiên cứu này là một biến ngẫu nhiên X có phân bố Bernoulli với tham số p . Ta cần kiểm định giả thiết cho rằng giá trị p bằng p_0 .

Giả thiết không $H_0: p = p_0$.

Từ tổng thể rút ra một mẫu ngẫu nhiên kích thước n , gọi f là tần suất mẫu (công thức (4.17)-(4.18)). Xét thống kê

$$T = \frac{(f - p)\sqrt{n}}{\sqrt{p(1 - p)}} \quad (5.34)$$

Nếu giả thiết H_0 đúng thống kê trên trở thành $T = \frac{(f - p_0)\sqrt{n}}{\sqrt{p_0(1 - p_0)}}$.

Trong thực tế khi

$$\begin{cases} np_0 > 5 \\ n(1 - p_0) > 5 \end{cases} \quad (5.35)$$

ta có thể xem thống kê (5.34) có phân bố chuẩn tắc $N(0;1)$. Do đó với mức ý nghĩa α và tùy thuộc đối thiết H_1 ta có thể xây dựng các miền bác bỏ tương ứng.

a. $H_0: p = p_0; H_1: p \neq p_0$.

Miền bác bỏ
$$W_\alpha = \left\{ T = \frac{(f - p_0)\sqrt{n}}{\sqrt{p_0(1 - p_0)}} ; |T| > U_{\frac{\alpha}{2}} \right\} \quad (5.36)$$

b. $H_0: p = p_0; H_1: p > p_0$.

Miền bác bỏ
$$W_\alpha = \left\{ T = \frac{(f - p_0)\sqrt{n}}{\sqrt{p_0(1 - p_0)}} ; T > U_\alpha \right\} \quad (5.37)$$

c. $H_0: p = p_0; H_1: p < p_0$.

Miền bác bỏ
$$W_\alpha = \left\{ T = \frac{(f - p_0)\sqrt{n}}{\sqrt{p_0(1 - p_0)}} ; -T > U_\alpha \right\} \quad (5.38)$$

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 5.17: 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 đó.

Chọn ngẫu nhiên 2000 cử tri để thăm dò ý kiến và cho thấy có 862 cử tri tuyên bố sẽ bỏ phiếu cho A.

Với mức $\alpha = 5\%$, hãy kiểm định xem dự đoán của đảng trên có đúng không.

Giải: Gọi p là tỉ lệ cử tri sẽ bỏ phiếu cho ứng cử viên A.

Ta cần kiểm định: Giả thiết $H_0: p = 0,45$; Đối thiết $H_1: p \neq 0,45$.

(Bởi vì ta không có cơ sở nào để cho rằng dự đoán của đảng trên là cao hơn 0,45 hay thấp hơn 0,45).

Vì rằng điều kiện $\begin{cases} np_0 = 2000 \cdot 0,45 = 900 > 5 \\ n(1 - p_0) = 2000 \cdot 0,55 = 1100 > 5 \end{cases}$ thỏa mãn điều kiện (5.35) nên tiêu chuẩn kiểm định theo công thức (5.34) có xấp xỉ phân bố chuẩn tắc.

Thay mẫu cụ thể với $f = \frac{862}{2000} = 0,431$ ta được giá trị quan sát của tiêu chuẩn kiểm định

$$T_{qs} = \frac{(0,431 - 0,45)\sqrt{2000}}{\sqrt{0,45 \cdot 0,55}} = -1,708.$$

Với mức ý nghĩa $\alpha = 0,05 \Rightarrow U_{\frac{\alpha}{2}} = 1,96$. Ta thấy $|T_{qs}| < 1,96$. Vậy chưa có cơ sở để bác bỏ H_0 .

TÓM TẮT

Trung bình mẫu \bar{X} là ước lượng không chệch, hiệu quả và vững của kỳ vọng μ của biến ngẫu nhiên gốc của tổng thể.

Tần suất mẫu f là ước lượng không chệch, hiệu quả và vững của xác suất p của tổng thể.

Phương sai mẫu S^2 và S^{*2} (trường hợp μ đã biết) là ước lượng không chệch và vững của phương sai σ^2 của biến ngẫu nhiên gốc của tổng thể.

Khoảng tin cậy của kỳ vọng của biến ngẫu nhiên gốc có phân bố chuẩn:

- Trường hợp phương sai σ^2 đã biết: công thức (5.16).
- Trường hợp phương sai σ^2 chưa biết, $n \geq 30$: công thức (5.18).
- Trường hợp phương sai σ^2 chưa biết, $n < 30$: công thức (5.20).

Khoảng tin cậy cho xác suất p : công thức (5.23).

Khoảng tin cậy của phương sai của biến ngẫu nhiên gốc có phân bố chuẩn:

- Trường hợp kỳ vọng μ đã biết: công thức (5.30)-(5.33).
- Trường hợp kỳ vọng μ chưa biết: công thức (5.37)-(5.40).

Một thủ tục kiểm định giả thiết thống kê bao gồm các bước sau:

1. Phát biểu giả thiết H_0 và đối thiết H_1 .
2. Từ tổng thể nghiên cứu lập mẫu ngẫu nhiên kích thước n .
3. 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.
4. Với mức ý nghĩa α , xác định miền bác bỏ W_α tốt nhất tùy thuộc vào đối thiết H_1 .
5. Từ mẫu cụ thể tính giá trị quan sát của tiêu chuẩn kiểm định T_{qs} .
6. So sánh giá trị quan sát của tiêu chuẩn kiểm định T_{qs} với miền bác bỏ W_α và kết luận.

Kiểm định kỳ vọng của biến ngẫu nhiên gốc có phân bố chuẩn $N(\mu; \sigma^2)$.

Giả thiết $H_0: \mu = \mu_0$

	Tiêu chuẩn kiểm định	Miền bác bỏ tương ứng với đối thiết H_1		
		$H_1: \mu \neq \mu_0$	$H_1: \mu > \mu_0$	$H_1: \mu < \mu_0$
σ^2 đã biết	$T = \frac{(\bar{X} - \mu_0)\sqrt{n}}{\sigma}$	$W_\alpha = \{ T > U_{\alpha/2}\}$	$W_\alpha = \{T > U_\alpha\}$	$W_\alpha = \{-T > U_\alpha\}$
σ^2 chưa biết, $n \geq 30$	$T = \frac{(\bar{X} - \mu_0)\sqrt{n}}{S}$	$W_\alpha = \{ T > U_{\alpha/2}\}$	$W_\alpha = \{T > U_\alpha\}$	$W_\alpha = \{-T > U_\alpha\}$
σ^2 chưa biết, $n < 30$	$T = \frac{(\bar{X} - \mu_0)\sqrt{n}}{S}$	$W_\alpha = \{ T > t_{\alpha/2}(n-1)\}$	$W_\alpha = \{T > t_\alpha(n-1)\}$	$W_\alpha = \{-T > t_\alpha(n-1)\}$

Kiểm định giả thiết về xác suất. Giả thiết $H_0: p = p_0$

Tiêu chuẩn kiểm định $T = \frac{(f - p_0)\sqrt{n}}{\sqrt{p_0(1 - p_0)}}$, với điều kiện $\begin{cases} np_0 > 5 \\ n(1 - p_0) > 5 \end{cases}$.

Đối thiết H_1 :	$p \neq p_0$	$p > p_0$	$p < p_0$
Miền bác bỏ	$W_\alpha = \{ T > U_{\alpha/2}\}$	$W_\alpha = \{T > U_\alpha\}$	$W_\alpha = \{-T > U_\alpha\}$

CÂU HỎI ÔN TẬP VÀ BÀI TẬP CHƯƠNG 5

5.1 Trung bình mẫu là ước lượng vững và hiệu quả của kỳ vọng của biến ngẫu nhiên gốc.

Đúng ☐ Sai ☐.

5.2 Có thể tìm được ước lượng không chệch của θ có phương sai nhỏ hơn đại lượng

$$\frac{1}{nE\left(\frac{\partial(\ln f(X, \theta))}{\partial \theta}\right)^2}.$$

Đúng ☐ Sai ☐.

5.3 Tổng của hai ước lượng không chệch là một ước lượng không chệch.

Đúng ☐ Sai ☐.

5.4 Phương sai mẫu hiệu chỉnh S^2 là ước lượng vững không chệch của phương sai của biến ngẫu nhiên gốc.

Đúng ☐ Sai ☐.

5.5 Hai đầu mút của khoảng tin cậy là hai thống kê của mẫu.

Đúng ☐ Sai ☐.

5.6 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 ☐.

5.7 Giả thiết thống kê là giả thiết do nhà thống kê đặt ra cho mẫu ngẫu nhiên.

Đúng ☐ Sai ☐.

5.8 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 ☐.

5.9 Quy tắc kiểm định dựa trên nguyên lý xác suất nhỏ và phép chứng minh phản chứng.

Đúng ☐ Sai ☐.

5.10 Sai lầm loại 1 là sai lầm gặp phải khi thực tế giả thiết đúng nhưng ta bác bỏ.

Đúng ☐ Sai ☐.

5.11 Sai lầm loại 2 luôn luôn lớn hơn sai lầm loại 1.

Đúng ☐ Sai ☐.

5.12 Miền bác bỏ là miền có xác suất rất bé nên ta có thể bỏ qua trong mọi phép kiểm định.

Đúng ☐ Sai ☐.

5.13 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 ☐.

5.14 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 ☐.

5.15 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.

5.16 Giả sử biến ngẫu nhiên gốc có phân bố chuẩn $N(20;1)$. Chọn mẫu ngẫu nhiên kích thước $n = 100$. Hãy tính xác suất để trung bình mẫu \bar{X} nằm trong khoảng: $19,8 < \bar{X} < 20,2$.

5.17 Một mẫu cụ thể của biến ngẫu nhiên X như sau:

2 ; 3 ; 2 ; 4 ; 1 ; 4 ; 2 ; 2 ; 3 ; 1 ($n = 10$).

c. Lập bảng phân bố tần suất.

d. Xây dựng hàm phân bố thực nghiệm. Tính \bar{x} , s^2 , s .

5.18 Giả sử $\widehat{\Theta}_n = T(X_1, X_2, \dots, X_n)$ là một ước lượng của θ thỏa mãn:

$$\lim_{n \rightarrow \infty} E(\widehat{\Theta}_n) = \theta \text{ và } \lim_{n \rightarrow \infty} D(\widehat{\Theta}_n) = 0.$$

Chứng minh $\widehat{\Theta}_n$ là một ước lượng vững của θ .

5.19 Trong đợt vận động bầu cử tổng thống ở một nước nọ, người ta phỏng vấn ngẫu nhiên 2000 cử tri thì được biết có 1082 người trong số đó sẽ bỏ phiếu cho ứng cử viên A. Với độ tin cậy 98% tối thiểu ứng cử viên A sẽ chiếm được bao nhiêu % số phiếu bầu?

5.20 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.

5.21 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.

5.22 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.

5.23 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.

5.24 Người ta đo một đại lượng không đổi 25 lần bằng một dụng cụ đo không có sai số hệ thống và sai số đo trung bình bằng 0. Giả sử sai số của phép đo tuân theo quy luật phân bố chuẩn và phương sai mẫu đo được bằng 0,5. Hãy xác định khoảng tin cậy 95% cho phương sai của sai số đo.

5.25 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:

Chương 5: Ước lượng tham số và kiểm định giả thiết thống kê

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.

5.26 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.

5.27 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.

5.28 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?

HƯỚNG DẪN VÀ ĐÁP ÁN

HƯỚNG DẪN VÀ ĐÁP ÁN CHƯƠNG 1

1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	1.10
Đúng	Sai	Đúng	Đúng	Sai	Đúng	Sai	Sai	Đúng	Đúng

1.11 a) $P = 0,246$ b) $P = 0,495$.

1.12 a) $1/3$ b) $3/5$ c) $2/5$ d) $4/5$

1.13 a) $P = \frac{C_3^2}{C_5^2} = \frac{3}{10}$ b) $P = \frac{C_2^2}{C_5^2} = \frac{1}{10}$ c) $P = \frac{C_2^1 C_3^1}{C_5^2} = \frac{3}{5}$.

1.14 Mỗi khách đều có 6 khả năng để ra ở 6 tầng còn lại của tòa nhà. Do đó số kết cục đồng khả năng có thể $N = 6^3 = 216$. Gọi A là biến cố tất cả cùng ra ở tầng bốn, biến cố này chỉ có 1 trường hợp thuận lợi. Do đó $P(A) = \frac{1}{216}$.

Lý luận tương tự trên ta có $P(B) = \frac{6}{216} = \frac{1}{36}$; $P(C) = \frac{A_6^3}{216} = \frac{5}{9}$.

1.15 $P = \frac{1}{720}$.

1.16 a) $A_1 \cap A_2 \cap \dots \cap A_{10}$ b) $A_1 \cup A_2 \cup \dots \cup A_{10}$
 c) $\overline{A_1} \cap \overline{A_2} \cap \dots \cap \overline{A_6} \cap A_7 \cap A_8 \cap A_9 \cap A_{10}$ d) $A_1 \cap A_2 \cap \dots \cap A_6$.

1.17 a) $A = A_1 \cap A_2 \cap A_3$ b) $A = A_1 \cup A_2 \cup A_3$
 c) $A = A_1 \cap (A_2 \cup A_3)$ d) $A = (A_1 \cap A_2) \cup A_3$.

1.18 Gọi A_1 và A_2 tương ứng là biến cố người thứ nhất và thứ hai bắn trúng mục tiêu, A là biến cố chỉ có một người bắn trúng mục tiêu. $A = (A_1 \cap \overline{A_2}) \cup (\overline{A_1} \cap A_2)$. Sử dụng qui tắc cộng xác suất trường hợp xung khắc và qui tắc nhân trường hợp độc lập ta có:

$$P(A) = P(A_1 \cap \overline{A_2}) + P(\overline{A_1} \cap A_2) = P(A_1)P(\overline{A_2}) + P(\overline{A_1})P(A_2) = 0,8 \cdot 0,1 + 0,2 \cdot 0,9 = 0,26.$$

Tương tự ta có: $P_b = 0,98$; $P_c = 0,02$.

1.19 Gọi A_1 là biến cố sản phẩm lấy ra thuộc loại 1.

Gọi A_2 là biến cố sản phẩm lấy ra thuộc loại 2.

Gọi A là biến cố sản phẩm lấy ra thuộc loại 1 hoặc loại 2: $A = A_1 \cup A_2$

Vì A_1, A_2 xung khắc do đó $P(A) = P(A_1 \cup A_2) = P(A_1) + P(A_2) = 0,4 + 0,5 = 0,9$

1.20 $P = \frac{C_{20}^5 C_{980}^{25}}{C_{1000}^{30}} = 0,00027$.

$$1.21 \quad a) \frac{C_4^4 C_{48}^1}{C_{52}^5} = \frac{1}{54.145} \quad b) \frac{C_4^4 C_4^1}{C_{52}^5} = \frac{1}{649.740} \quad c) \frac{C_4^3 C_4^2}{C_{52}^5} = \frac{1}{108.290}$$

$$d) \frac{C_4^1 C_4^1 C_4^1 C_4^1 C_4^1}{C_{52}^5} = \frac{64}{162.435} = 0,000394 \quad e) \frac{4C_{13}^5}{C_{52}^5} = \frac{1}{500} = 0,002$$

$$f) \frac{4}{C_{52}^5} = \frac{1}{649.740} \quad g) 1 - \frac{C_{48}^5}{C_{52}^5} = 1 - \frac{35673}{54145} = \frac{18472}{54145} = 0,3412$$

$$1.22 \quad a) P = \frac{4^3}{52^3} = \frac{1}{2197} \quad b) P = \frac{4}{52} \frac{3}{51} \frac{2}{50} = \frac{1}{5525}$$

1.23 Gọi A_i là biến cố sản phẩm đã qua kiểm tra chất lượng ở vòng thứ i , $i = 1, 2, 3$.

Gọi B là biến cố phế phẩm được nhập kho.

$$P(B) = P(\overline{A_1})P(\overline{A_2})P(\overline{A_3}) = (1-0,8)(1-0,9)(1-0,99) = 0,0002.$$

$$1.24 \quad P = 0,11.$$

$$1.25 \quad P(A) + P(B) - P(A)P(B) = 0,65 \Rightarrow P(B) = 0,5.$$

$$1.26 \quad a) P(A|B) = 3/4 \quad b) P(B|A) = 1/2 \quad c) P(A \cup B) = \frac{1}{2} + \frac{1}{3} - \frac{1}{3} = \frac{7}{12}$$

1.27 Gọi A_i là biến cố lần thứ nhất chọn được là số i , $i = 0, 1, \dots, 9$. Gọi B là biến cố lần thứ hai chọn được số 4.

$$P(A_i) = \frac{1}{10}; P(B|A_i) = \frac{1}{9}, \forall i \neq 4; P(B|A_4) = 0.$$

$$\text{Vậy } P(B) = \sum_{i=1}^{10} P(A_i)P(B|A_i) = \frac{1}{10}.$$

1.28 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 , $i=1, 2, 3$.

$$P(B_1) = 0,36; P(B_2) = 0,34; P(B_3) = 0,30. \text{ Hệ } \{B_1, B_2, B_3\} \text{ đầy đủ}$$

$$P(A|B_1) = 0,12; P(A|B_2) = 0,10; P(A|B_3) = 0,08.$$

$$a. P(A) = 0,1012$$

$$b. P(B_1|A) = 0,427; P(B_2|A) = 0,336; P(B_3|A) = 0,237.$$

1.29 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.

$$\text{Theo đề bài ta có: } P(B_1) = \frac{5}{18}, P(B_2) = \frac{7}{18}, P(B_3) = \frac{4}{18}, P(B_4) = \frac{2}{18}$$

$$P(A|B_1) = 0,2, P(A|B_2) = 0,3, P(A|B_3) = 0,4, P(A|B_4) = 0,5.$$

$$P(A) = \frac{57}{180}$$

Áp dụng công thức Bayes, ta thu được

$$P(B_1|A) = \frac{10}{57}, \quad P(B_2|A) = \frac{21}{57}, \quad P(B_3|A) = \frac{16}{57}, \quad P(B_4|A) = \frac{10}{57}.$$

Vậy xạ thủ có khả năng ở nhóm thứ hai nhất.

1.30 Gọi B_1 là biến cố viên đạn thứ nhất trúng mục tiêu, $P(B_1) = 0,7$.

Gọi B_2 là biến cố viên đạn thứ hai trúng mục tiêu, $P(B_2) = 0,4$.

Hai biến cố này độc lập

Xác suất biến cố chỉ có viên đạn thứ nhất trúng mục tiêu $P(A) = 0,54$

b. $P(B_1|A) = 0,778$.

1.31 A_1, A_2, A_3 lần lượt là biến cố chọn từ hộp thứ nhất, thứ hai và thứ ba. Gọi B là biến cố viên bi được màu trắng.

$$P(A_1) = P(A_2) = P(A_3) = \frac{1}{3}; \quad P(B|A_1) = \frac{2}{5}, \quad P(B|A_2) = \frac{4}{5}, \quad P(B|A_3) = \frac{3}{7}$$

$$\Rightarrow P(B) = \frac{1}{3} \left(\frac{2}{5} + \frac{4}{5} + \frac{3}{7} \right) \Rightarrow P(A_1|B) = \frac{P(A_1)P(B|A_1)}{P(B)} = \frac{14}{57}.$$

1.32 Gọi A là biến cố sản phẩm kiểm tra có kết luận đạt tiêu chuẩn chất lượng.

Gọi B_T là biến cố sản phẩm đạt tiêu chuẩn chất lượng.

Gọi B_H là biến cố sản phẩm không đạt tiêu chuẩn chất lượng.

$$P(B_T) = 0,85; \quad P(B_H) = 0,15. \quad \text{Hệ } \{B_T, B_H\} \text{ đầy đủ}$$

$$P(A|B_T) = 0,9; \quad P(\bar{A}|B_H) = 0,95 \Rightarrow P(\bar{A}|B_T) = 0,1; \quad P(A|B_H) = 0,05.$$

a) $P(A) = 0,7725$

b) $P(B_H|A) = 0,0097$

c) $P((A \cap B_T) \cup (\bar{A} \cap B_H)) = 0,9075$.

1.33 Nếu $P(A|B) = \frac{P(A \cap B)}{P(B)} > P(A)$ thì $P(A \cap B) > P(A)P(B)$.

$$\text{Do đó } P(B|A) = \frac{P(A \cap B)}{P(A)} > \frac{P(A)P(B)}{P(A)} = P(B).$$

HƯỚNG DẪN VÀ ĐÁP ÁN CHƯƠNG 2

2.1	2.2	2.3	2.4	2.5	2.6	2.7	2.8	2.9	2.10
Sai	Sai	Đúng	Sai	Đúng	Sai	Sai	Đúng	Đúng	Sai

2.11	2.12	2.13	2.14	2.15
Sai	Sai	Đúng	Đúng	Sai

2.16 $0 \leq F(x) \leq 1 \Rightarrow 0 \leq a \leq 1$ và $b > 0$

2.17 $E(X) = -0,3; D(X) = 15,21.$

2.18 $x_3 = 29,1; p_3 = 0,2.$

2.19 $E(X_1) = 3,1; E(X_2) = 3,4; D(X_1) = 1,09; D(X_2) = 1,44.$

$E(X_1 - X_2) = -0,3; D(X_1 - X_2) = 2,53.$

2.20 $E(\bar{X}) = 0,8; D(\bar{X}) = 0,12.$

2.21 a) $D(Z) = 61.$ b) $D(Z) = 41.$

2.22 $p_1 = 0,4; p_2 = 0,1; p_3 = 0,5.$

2.23 $E(X) = 0.$

2.24 a) $\forall i \int_0^4 x^2(4-x)dx = \frac{64}{3} \Rightarrow k = \frac{3}{64}. P\{X < 1\} = \int_0^1 \frac{3}{64} x^2(4-x)dx = \frac{13}{256}.$

b) $EX = \int_0^4 \frac{3}{64} x^3(4-x)dx = \frac{3}{64} \left(x^4 - \frac{x^5}{5} \right) \Big|_{x=0}^{x=4} = 3 \left(4 - \frac{16}{5} \right) = \frac{12}{5}$

$EX^2 = \int_0^4 \frac{3}{64} x^4(4-x)dx = \frac{3}{64} \left(\frac{4x^5}{5} - \frac{x^6}{6} \right) \Big|_{x=0}^{x=4} = 3 \times 64 \left(\frac{1}{5} - \frac{1}{6} \right) = \frac{32}{5}$

$\Rightarrow DX = EX^2 - (EX)^2 = \frac{32}{5} - \left(\frac{12}{5} \right)^2 = \frac{16}{25}.$

2.25 a) $k = 2; F_X(x) = \begin{cases} 1 - e^{-2x} & \text{nếu } x > 0 \\ 0 & \text{nếu } x \leq 0 \end{cases}$

b) $E(X) = 0,5; E(X^2) = 0,5; DX = 0,25.$

2.26 a) $k = \frac{6}{29}; P\{X > 2\} = \frac{15}{29}, P\{0,5 < X < 1,5\} = \frac{19}{116}$

b) $F_X(x) = \begin{cases} 0 & \text{nếu } x \leq 1 \\ (2x^3 - 2)/29 & \text{nếu } 1 \leq x \leq 2 \\ (3x^2 + 2)/29 & \text{nếu } 2 \leq x \leq 3 \\ 1 & \text{nếu } x \geq 3 \end{cases}$

2.27 Gọi X là tiền thắng của mỗi vé, X là biến ngẫu nhiên có bảng phân bố xác suất

X	0	5	25	100
P	0,9775	0,02	0,002	0,0005

$E(X) = 0,2$ \$. Vậy giá vé tối thiểu là 0,2\$.

2.28 a) Bảng phân bố xác suất của X

X	-2	-1	0	1	2
P	0,09	0,30	0,37	0,20	0,04

$$EX = (-2) \times 0,09 + (-1) \times 0,30 + 0 \times 0,37 + 1 \times 0,20 + 2 \times 0,04 = -0,2$$

$$EX^2 = (-2)^2 \times 0,09 + (-1)^2 \times 0,30 + 0^2 \times 0,37 + 1^2 \times 0,20 + 2^2 \times 0,04 = 1,02$$

$$DX = EX^2 - (EX)^2 = 1,02 - (-0,2)^2 = 0,98$$

b) $P\{Y=0\} = 0,37$

$$P\{Y=1\} = P\{X=1\} + P\{X=-1\} = 0,5$$

$$P\{Y=2\} = P\{X=2\} + P\{X=-2\} = 0,13$$

$$EY = 0 \times 0,37 + 1 \times 0,5 + 2 \times 0,13 = 0,76.$$

2.29 Gọi X là số ô tô bị hỏng trong thời gian làm việc

X	0	1	2
P	0,72	0,26	0,02

$$EX = 0 \times 0,72 + 1 \times 0,26 + 2 \times 0,02 = 0,3.$$

$$EX^2 = 0^2 \times 0,72 + 1^2 \times 0,26 + 2^2 \times 0,02 = 0,34.$$

$$DX = EX^2 - (EX)^2 = 0,34 - (0,3)^2 = 0,25.$$

$$2.30 \quad p_X(x) = \begin{cases} 1/36 & \text{nếu } x=2 \text{ hoặc } x=12 \\ 2/36 & \text{nếu } x=3 \text{ hoặc } x=11 \\ 3/36 & \text{nếu } x=4 \text{ hoặc } x=10 \\ 4/36 & \text{nếu } x=5 \text{ hoặc } x=9 \\ 5/36 & \text{nếu } x=6 \text{ hoặc } x=8 \\ 6/36 & \text{nếu } x \neq 2,3,\dots,12 \\ 0 & \text{nếu } x=7 \end{cases} \quad F_X(x) = \begin{cases} 0 & \text{nếu } x < 2 \\ 1/36 & \text{nếu } 2 \leq x < 3 \\ 3/36 & \text{nếu } 3 \leq x < 4 \\ 6/36 & \text{nếu } 4 \leq x < 5 \\ \dots & \\ 35/36 & \text{nếu } 11 \leq x < 12 \\ 1 & \text{nếu } x \geq 12 \end{cases}$$

2.31 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: $A = A_1 \cap A_2 \cap A_3 \Rightarrow \bar{A} = \bar{A}_1 \cup \bar{A}_2 \cup \bar{A}_3$.

$$P(\bar{A}_i) = \frac{2^5}{3^5} = \frac{32}{243}; \quad P(\bar{A}_i \cap \bar{A}_j) = \frac{1}{243}; \quad P(\bar{A}_1 \cap \bar{A}_2 \cap \bar{A}_3) = 0$$

$$P(\bar{A}) = P(\bar{A}_1) + P(\bar{A}_2) + P(\bar{A}_3) - P(\bar{A}_1 \cap \bar{A}_2) - P(\bar{A}_1 \cap \bar{A}_3) - P(\bar{A}_2 \cap \bar{A}_3) + P(\bar{A}_1 \cap \bar{A}_2 \cap \bar{A}_3) = \frac{31}{81}$$

$$\Rightarrow P(A) = \frac{50}{81}.$$

$$\text{b) } P\{X=0\} = P(\bar{A}_1) = \frac{2^5}{3^5} = \frac{32}{243}; P\{X=k\} = \frac{C_5^k 2^{5-k}}{3^5}$$

X	0	1	2	3	4	5
P	$\frac{32}{243}$	$\frac{80}{243}$	$\frac{80}{243}$	$\frac{40}{243}$	$\frac{10}{243}$	$\frac{1}{243}$

Y	0	1	2	3	4	5
P	$\frac{1}{243}$	$\frac{10}{243}$	$\frac{40}{243}$	$\frac{80}{243}$	$\frac{80}{243}$	$\frac{32}{243}$

$$\text{2.32 a) Điều kiện } \begin{cases} p_i > 0 \\ \sum_i p_i = 1 \end{cases} \Rightarrow \begin{cases} k > 0 \\ 10k^2 + 9k = 1 \end{cases} \Rightarrow \begin{cases} k > 0 \\ \begin{cases} k = -1 \\ k = 1/10 \end{cases} \end{cases} \Rightarrow k = 1/10.$$

$$\text{b) } P\{X \geq 5\} = \frac{1}{100} + \frac{2}{100} + \frac{7}{100} + \frac{1}{10} = \frac{2}{10}; P\{X < 3\} = \frac{3}{10}.$$

$$\text{c) } EX = \frac{1}{10} + \frac{4}{10} + \frac{6}{10} + \frac{12}{10} + \frac{5}{100} + \frac{12}{100} + 7\left(\frac{7}{100} + \frac{1}{10}\right) = 3,66.$$

$$\text{d) } EX^2 = \frac{1}{10} + \frac{8}{10} + \frac{18}{10} + \frac{48}{10} + \frac{25}{100} + \frac{72}{100} + 49\left(\frac{7}{100} + \frac{1}{10}\right) = 16,8.$$

$$DX = EX^2 - (EX)^2 = 16,8 - (3,66)^2 = 3,404.$$

$$\text{2.33 a) Gọi } X \text{ là "số phế phẩm gặp phải": } \begin{array}{c|cc} X & 0 & 1 \\ \hline P & 0,6 & 0,4 \end{array}$$

$$EX = \frac{2}{5}; DX = \frac{6}{25}.$$

$$\text{b) Gọi } Y \text{ là "số chính phẩm gặp phải" } \Rightarrow Y = 2 - X: \begin{array}{c|cc} Y & 1 & 2 \\ \hline P & 0,4 & 0,6 \end{array}$$

$$EY = E(2 - X) = 2 - \frac{2}{5} = \frac{8}{5}; DY = \frac{6}{25}.$$

2.34 Gọi X là "số nữ có trong nhóm được chọn"

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

$$EX = 0 \times \frac{5}{30} + 1 \times \frac{15}{30} + 2 \times \frac{9}{30} + 3 \times \frac{1}{30} = \frac{36}{30} = \frac{6}{5}.$$

2.35 Thẳng 2 trong 4 ván dễ hơn.

2.36 a) X tuân theo quy luật nhị thức $B(n; p)$ với $n = 5$ và $p = 0,8$.

b) $EX = 4$; $DX = 0,8$. c) $\text{Mod}X = 4$; $P\{X = 4\} = 0,4096$.

2.37 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.

2.38 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 $P = \sum_{k=0}^3 C_{10}^k \left(\frac{1}{5}\right)^k \left(\frac{4}{5}\right)^{10-k} = 0,879$.

2.39 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 $P\{X \leq 1\} = 0,031$

c) Xác suất thu được tín hiệu $P\{X \geq 1\} = 1 - P\{X = 0\} = 1 - 0,002 = 0,998$

2.40 Không đúng; $P = 0,41$.

2.41 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

$$P\{X \geq 1\} = 1 - P\{X = 0\} = 1 - e^{-1/3} = 0,2835.$$

b) Gọi Y là số cuộc gọi trong khoảng thời gian 3 phút thì Y 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

$$P\{Z \leq 1\}^3 = 0,406^3 = 0,0067.$$

2.42 a) $\sum_{k=1}^{\infty} p_X(k) = \sum_{k=1}^{\infty} q^{k-1} p = p \sum_{k=0}^{\infty} q^k = \frac{p}{1-q} = 1$

$$P\{X > k\} = \sum_{i=k+1}^{\infty} q^{i-1} p = \frac{q^k p}{1-q} = q^k. \text{ Vậy } P\{X \leq k\} = 1 - P\{X > k\} = 1 - q^k$$

Do đó $F_X(x) = P\{X \leq x\} = 1 - q^k$ nếu $k \leq x < k+1$.

b) Xét hàm số $g(x) = \sum_{n=0}^{\infty} ax^n = \frac{a}{1-x}$ với $|x| < 1$ thì

$$\sum_{n=1}^{\infty} anx^{n-1} = g'(x) = \frac{a}{(1-x)^2}, \quad \sum_{n=2}^{\infty} an(n-1)x^{n-2} = g''(x) = \frac{2a}{(1-x)^3}.$$

Áp dụng kết quả này ta được $E(X) = \frac{p}{(1-q)^2} = \frac{p}{p^2} = \frac{1}{p},$

$$E[X(X-1)] = \frac{2pq}{(1-q)^3} = \frac{2pq}{p^3} = \frac{2(1-p)}{p^2}$$

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

$$D(X) = \frac{2-p}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2} = \frac{q}{p^2}.$$

2.43 Tra bảng phụ lục II ta được

a) 0,3849 b) 0,2517 c) 0,6636 d) 0,1828 e) 0,8997

2.44 Tra bảng phụ lục II ta được

a) 1,16 b) 1,09 c) -1,35 d) -1,69

2.45 $P\{8 < X < 12\} = \Phi\left(\frac{12-10}{2}\right) - \Phi\left(\frac{8-10}{2}\right) = 0,6826.$

2.46 $P = 0,3.$

2.47 a) 95,44%; b) 4,56%.

2.48 $E(\bar{X}) = \mu; D(\bar{X}) = \frac{\sigma^2}{n} \Rightarrow P\{|\bar{X} - \mu| < \varepsilon\} = 2\Phi\left(\frac{\varepsilon\sqrt{n}}{\sigma}\right) - 1.$

2.49 $P\{X^2 \leq x\} = 0, \forall x \leq 0.$

$$\forall x > 0; P\{X^2 \leq x\} = P\{-\sqrt{x} \leq X \leq \sqrt{x}\} = \frac{1}{\sqrt{2\pi}} \int_{-\sqrt{x}}^{\sqrt{x}} e^{-\frac{t^2}{2}} dt = \frac{2}{\sqrt{2\pi}} \int_0^{\sqrt{x}} e^{-\frac{t^2}{2}} dt$$

$$\text{Đổi biến số } t = u^2 \Rightarrow dt = 2udu; \quad \frac{2}{\sqrt{2\pi}} \int_0^{\sqrt{x}} e^{-\frac{t^2}{2}} dt = \frac{1}{\sqrt{2\pi}} \int_0^x u^{-\frac{1}{2}} e^{-\frac{u}{2}} du.$$

$$\text{Vậy hàm mật độ } f_{X^2}(x) = \begin{cases} \frac{1}{\sqrt{2\pi}} x^{-\frac{1}{2}} e^{-\frac{x}{2}} & x > 0 \\ 0 & x < 0 \end{cases}.$$

2.50 a) Hàm phân bố $F_X(x) = \int_0^x \frac{t}{\sigma^2} e^{-t^2/(2\sigma^2)} dt = 1 - e^{-x^2/(2\sigma^2)}$, $x \geq 0$.

b) $EX = \int_0^\infty \frac{x^2}{\sigma^2} e^{-x^2/(2\sigma^2)} dx = \sigma \int_0^\infty y^2 e^{-y^2/2} dy = \frac{\sigma}{2}$.

$EX^2 = \int_0^\infty \frac{x^3}{\sigma^2} e^{-x^2/(2\sigma^2)} dx = \int_0^\infty 2\sigma^2 y e^{-y^2} dy = 2\sigma^2$. $DX = EX^2 - (EX)^2 = \frac{7\sigma^2}{4}$.

2.51 a)

A	B	A	B	X	P
D				-5	$\frac{3}{7} = \frac{15}{35}$
T	D			10	$\frac{4}{7} \cdot \frac{3}{6} = \frac{10}{35}$
T	T	D		-15	$\frac{4}{7} \cdot \frac{3}{6} \cdot \frac{3}{5} = \frac{6}{35}$
T	T	T	D	20	$\frac{4}{7} \cdot \frac{3}{6} \cdot \frac{2}{5} \cdot \frac{3}{4} = \frac{3}{35}$
T	T	T	T	-25	$\frac{4}{7} \cdot \frac{3}{6} \cdot \frac{2}{5} \cdot \frac{1}{4} = \frac{1}{35}$

X	-25	-15	-5	10	20
P	$\frac{1}{35}$	$\frac{6}{35}$	$\frac{15}{35}$	$\frac{10}{35}$	$\frac{3}{35}$

$EX = -\frac{6}{7}$.

b) $-150 \cdot \frac{6}{7} = -128,57 \text{USD}$.

HƯỚNG DẪN VÀ ĐÁP ÁN CHƯƠNG 3

3.1	3.2	3.3	3.4	3.5	3.6	3.7	3.8	3.9	3.10
Sai	Đúng	Đúng	Đúng	Sai	Sai	Sai	Đúng	Đúng	Sai
3.11	3.12	3.13	3.14	3.15	3.16	3.17	3.18	3.19	3.20
Sai	Đúng	Đúng	Đúng	Sai	Sai	Đúng	Đúng	Đúng	Đúng

3.21

$X \backslash Y$	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

$$P\{X = 1\} = 0,5, \quad P\{Y = 1\} = 0,45 \quad \text{và} \quad P\{X = 1, Y = 1\} = 0,15 \neq 0,5 \cdot 0,45.$$

3.22

X	x_1	x_2
P	0,56	0,44

Y	y_1	y_2	y_3
P	0,26	0,38	0,36

3.23 Bảng phân bố xác suất của biến ngẫu nhiên Z

Z	1	2	3	4	6
P	0,12	0,43	0,03	0,35	0,07

$$EX = 1,7; \quad EY = 1,7; \quad EZ = 2,89.$$

3.24 Bảng phân bố xác suất đồng thời của X và Y

$X \backslash Y$	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

$$P\{X > Y\} = 0,19.$$

3.25

$Y X = 26$	23	27
P	0,357	0,643

$X Y = 27$	26	30	41	50
P	0,1268	0,4225	0,1549	0,2958

3.26 $E[Y|X = 1] = 5$. $EX = 2,93$; $EY = 4,5$; $DX = 4,83$; $DY = 2,25$.

3.27 a. $\alpha = 15$; $EX = -0,2$; $EY = 0$.

b. $\text{cov}(X, Y) = 0 \Rightarrow \rho(X, Y) = 0$.

c. X, Y không độc lập vì $P\{X=1\} = \frac{2}{15}$, $P\{Y=1\} = \frac{5}{15}$ nhưng $P\{X=1, Y=1\} = 0$.

3.28 Hàm khối lượng xác suất của biến ngẫu nhiên rời rạc X là $p_X(k) = e^{-\lambda} \frac{\lambda^k}{k!}$.

$$P(B) = P\{X=0, 2, 4, \dots\} = \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^{2k}}{(2k)!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^{2k}}{(2k)!}$$

$$e^{\lambda} = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = \sum_{k=0}^{\infty} \frac{\lambda^{2k}}{(2k)!} + \sum_{k=0}^{\infty} \frac{\lambda^{2k+1}}{(2k+1)!}$$

$$e^{-\lambda} = \sum_{k=0}^{\infty} \frac{(-\lambda)^k}{k!} = \sum_{k=0}^{\infty} \frac{\lambda^{2k}}{(2k)!} - \sum_{k=0}^{\infty} \frac{\lambda^{2k+1}}{(2k+1)!}$$

$$\sum_{k=0}^{\infty} \frac{\lambda^{2k}}{(2k)!} = \frac{e^{\lambda} + e^{-\lambda}}{2} \Rightarrow P(B) = e^{-\lambda} \frac{e^{\lambda} + e^{-\lambda}}{2} = \frac{1 + e^{-2\lambda}}{2}$$

$$p_X(k|B) = \frac{P(\{X=k\} \cap B)}{P(B)}; \quad \{X=2k\} \cap B = \{X=2k\}, \quad \{X=2k+1\} \cap B = \emptyset.$$

$$p_X(k|B) = \begin{cases} \frac{P\{X=k\}}{P(B)} = \frac{2e^{-\lambda}\lambda^k}{(1+e^{-2\lambda})k!} & k \text{ chẵn} \\ \frac{P(\emptyset)}{P(B)} = 0 & k \text{ lẻ} \end{cases}$$

3.29 a) $k = \frac{1}{42}$; $P\{X=2, Y=1\} = \frac{5}{42}$; $P\{X \geq 1, Y \leq 2\} = \frac{24}{42} = \frac{4}{7}$

b) $p_{Y|X=2}(y|X=2) = \begin{cases} \frac{4+y}{22} & y=0,1,2,3 \\ 0 & y \neq 0,1,2,3 \end{cases}; \quad E[Y|X=2] = \frac{19}{11}.$

3.30 a) $k=3$;

b) $f_X(x) = \begin{cases} 3x^2 & \text{nếu } 0 < x < 1 \\ 0 & \text{nếu ngược lại} \end{cases}; \quad f_Y(y) = \begin{cases} \frac{3}{2}(1-y^2) & \text{nếu } 0 < y < 1 \\ 0 & \text{nếu ngược lại} \end{cases}.$

c) X và Y không độc lập vì $P\left\{X < \frac{1}{2}, Y > \frac{1}{2}\right\} = 0$ nhưng $P\left\{X < \frac{1}{2}\right\} \neq 0, P\left\{Y > \frac{1}{2}\right\} \neq 0$.

3.31 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

$$f(x|y) = \begin{cases} e^{-x} & \text{nếu } x > 0, \\ 0 & \text{nếu } x \leq 0. \end{cases}$$

3.32 a) $k = \frac{1}{\pi^2}$;

b) $F(x, y) = \left(\frac{1}{\pi} \arctg x + \frac{1}{2} \right) \left(\frac{1}{\pi} \arctg y + \frac{1}{2} \right)$;

c) $F_X(x) = \lim_{y \rightarrow \infty} F(x, y) = \frac{1}{\pi} \arctg x + \frac{1}{2}$; $F_Y(y) = \lim_{x \rightarrow \infty} F(x, y) = \frac{1}{\pi} \arctg y + \frac{1}{2}$;

Vì $F(x, y) = F_X(x)F_Y(y)$ nên ta kết luận X và Y độc lập.

d) $P\{1 < X < \sqrt{3}, 0 < Y < 1\} = P\{1 < X < \sqrt{3}\}P\{0 < Y < 1\} = \frac{1}{12} \cdot \frac{1}{4} = \frac{1}{48}$.

3.33 $F_X(x|a < X \leq b) = P\{X \leq x|a < X \leq b\} = \frac{P\{X \leq x; a < X \leq b\}}{P\{a < X \leq b\}}$

$$\{X \leq x; a < X \leq b\} = \begin{cases} \emptyset & x \leq a \\ \{a < X \leq x\} & a < x \leq b \\ \{a < X \leq b\} & x > b \end{cases} \Rightarrow F_X(x|a < X \leq b) = \begin{cases} 0 & x \leq a \\ \frac{F_X(x) - F_X(a)}{F_X(b) - F_X(a)} & a < x \leq b \\ 1 & x > b \end{cases}$$

3.34 Tương tự bài tập 3.27 ta có $F_X(x|X > 0) = \begin{cases} 0 & x \leq 0 \\ 2F_X(x) - 1 & x > 0 \end{cases}$

$$f_X(x|X > 0) = \begin{cases} 0 & x < 0 \\ \frac{2}{\sigma\sqrt{2\pi}} e^{-x^2/(2\sigma^2)} & x \geq 0 \end{cases} ; E[X|X > 0] = \frac{2}{\sigma\sqrt{2\pi}} \int_0^\infty x e^{-x^2/(2\sigma^2)} dx = \sigma\sqrt{\frac{2}{\pi}}$$

3.35 a) $k = \frac{1}{8}$; $f_X(x) = \begin{cases} \frac{1}{4}(x+1) & \text{nếu } 0 < x < 2 \\ 0 & \text{nếu ngược lại} \end{cases}$, $f_Y(y) = \begin{cases} \frac{1}{4}(y+1) & \text{nếu } 0 < y < 2 \\ 0 & \text{nếu ngược lại} \end{cases}$

X, Y không độc lập.

b) $f_{X|Y}(x|y) = \frac{1}{2} \frac{x+y}{y+1}$ $0 < x < 2, 0 < y < 2$; $f_{Y|X}(y|x) = \frac{1}{2} \frac{x+y}{x+1}$ $0 < x < 2, 0 < y < 2$.

c) $P\{0 < Y < 1/2 | X = 1\} = \int_0^{1/2} f_{Y|X}(y|x=1) dy = \frac{1}{2} \int_0^{1/2} \left(\frac{1+y}{2} \right) dy = \frac{5}{32}$.

3.36 a) $k = \frac{1}{96}$; $f_X(x) = \begin{cases} \frac{x}{8} & \text{nếu } 0 < x < 4 \\ 0 & \text{nếu ngược lại} \end{cases}$, $f_Y(y) = \begin{cases} \frac{y}{12} & \text{nếu } 1 < y < 5 \\ 0 & \text{nếu ngược lại} \end{cases}$.

$f_{XY}(x, y) = f_X(x)f_Y(y)$ do đó X, Y độc lập.

b) $P\{1 < X < 2, 2 < Y < 3\} = \int_1^2 \int_2^3 \frac{xy}{96} dx dy = \frac{5}{128}$; $P\{X \geq 3; Y \leq 2\} = \int_3^4 \int_1^2 \frac{xy}{96} dx dy = \frac{7}{128}$.

$$c) P\{X+Y < 3\} = \iint_{\{x+y < 3\} \cap R_{XY}} \frac{xy}{96} dx dy = \frac{1}{96} \int_0^2 dx \int_1^{3-x} xy dy = \frac{1}{48}.$$

$$3.37 \text{ a) } \frac{45}{512} \quad \text{b) } \frac{1}{14}.$$

$$3.38 \text{ E}(2X - 3Y) = 2E(X) - 3E(Y) = 10;$$

$$D(2X - 3Y) = 4D(X) + 9D(Y) - 12\sqrt{D(X)D(Y)}\rho_{X,Y} = 57,6.$$

$$3.39 \text{ a) Xét } F(t) = E[X - tY]^2 \geq 0, \forall t$$

$$F(t) = E[X]^2 - 2tE[XY] + t^2E[Y]^2 \Rightarrow \Delta' = (E[XY])^2 - (E[X^2])(E[Y^2]) \leq 0.$$

b) Áp dụng bất đẳng thức Cauchy-Schwarz ta có

$$(E[(X - EX)(Y - EY)])^2 \leq (E[X - EX]^2)(E[Y - EY]^2)$$

$$\Rightarrow (\text{cov}[X, Y])^2 \leq (DX)(DY) \Rightarrow \rho^2 \leq 1.$$

3.40 Gọi X là số máy hỏng trong ca. X có phân bố nhị thức $EX = 0,5$, $DX = 1$. Áp dụng bất đẳng thức Trêbusép ta có

$$P\{|X - 0,05| < 2\} \geq 1 - \frac{0,475}{2^2} = 0,88; \quad P\{|X - 0,05| \geq 2\} \leq \frac{0,475}{2^2} = 0,12.$$

3.41 Đặt $S = \sum_{n=1}^{12} X_n$; $ES = 12 \cdot 16 = 192$, $DS = 12$. Theo bất đẳng thức Trêbusép

$$P\{|S - 192| \leq \varepsilon\} \geq 1 - \frac{DS}{\varepsilon^2} \geq 0,99. \text{ Chọn } a = 157,36; \quad b = 226,64.$$

3.42 Đặt $S = \sum_{n=1}^{10000} X_n$; $ES = 0$, $DS = \frac{10000}{12}$. Theo bất đẳng thức Trêbusép

$$P\{|S| \geq 500\} \leq \frac{DS}{500^2} = \frac{1}{300}.$$

3.43 Áp dụng công thức (4.6) cho biến ngẫu nhiên \bar{X} ta được.

$$P\left\{\left|\bar{X} - \mu\right| \leq \frac{\sigma}{10}\right\} \geq 1 - \frac{\sigma^2/n}{\sigma^2/100} = 1 - \frac{100}{n}.$$

$$\text{Chọn } \frac{100}{n} \leq 0,05 \text{ hoặc } n \geq 100/0,05 = 200 \text{ thì } P\left\{\left|\bar{X} - \mu\right| \leq \frac{\sigma}{10}\right\} \geq 0,95.$$

3.44 Ta biết rằng S là biến ngẫu nhiên có phân bố nhị thức tham số $p = \frac{1}{6}$. $ES = \frac{n}{6}$ và

$$DS = \frac{5n}{36}. \text{ Theo bất đẳng thức Trêbusép}$$

$$P\{|S - ES| < \sqrt{n}\} \geq 1 - \frac{DS}{n} = 1 - \frac{5}{36} = \frac{31}{36} \Leftrightarrow P\left\{\frac{n}{6} - \sqrt{n} < S < \frac{n}{6} + \sqrt{n}\right\} \geq \frac{31}{36}.$$

3.45 Đặt $S = \sum_{n=1}^{12} X_n$. Ta cần tìm M nhỏ nhất để $P\left\{\sum_{n=1}^{12} X_n \leq M\right\} \geq 0,99$.

Ta có $ES = 192$, $DS = 12$. Theo bất đẳng thức Trêbusép

$$P\{|S - 192| \leq \varepsilon\} \geq 1 - \frac{DS}{\varepsilon^2} \geq 0,99 \Rightarrow \varepsilon = 34,64. \text{ Vậy } M = 192 + 34,64 = 226,64.$$

3.46 Áp dụng bất đẳng thức Trêbusép tính được xác suất $P \geq 0,9131$

3.47 Áp dụng bất đẳng thức Trêbusép cần kiểm tra 23.750 chi tiết.

3.48 Gọi X là số sản phẩm hỏng. Ta có $X \sim \mathbf{B}(250; 0,02)$. 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) $P\{X \leq 2\} = 0,1247$

3.49 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

$$P\{X < k, 1000 - X < k\} \geq 0,99 \Leftrightarrow P\{1000 - k < X < k\} \geq 0,99.$$

Ta xem X có phân bố chuẩn với $\mu = 500$, $\sigma = \sqrt{250}$. Vậy ta phải có

$$\Phi\left(\frac{k - 500}{\sqrt{250}}\right) - \Phi\left(\frac{500 - k}{\sqrt{250}}\right) \geq 0,99 \Leftrightarrow 2\Phi\left(\frac{k - 500}{\sqrt{250}}\right) \geq 1,99 \Leftrightarrow \Phi\left(\frac{k - 500}{\sqrt{250}}\right) \geq \Phi(2,58).$$

Từ đó $k \geq 500 + 2,58\sqrt{250} = 540,49$. Vậy $k = 541$.

3.50 a) Rõ ràng rằng số lỗi trong 10 ký số nhận được là biến ngẫu nhiên X có phân bố nhị thức với tham số $(n; p) = (10; 0,01)$

$$P\{X > 1\} = 1 - P\{X = 0\} - P\{X = 1\} = 1 - C_{10}^0(0,01)^0(0,99)^{10} - C_{10}^1(0,01)^1(0,99)^9 = 0,0042$$

b) Sử dụng công thức (4.16) với $\lambda = np = 10(0,01) = 0,1$

$$P\{X > 1\} = 1 - P\{X = 0\} - P\{X = 1\} = 1 - e^{-0,1} \frac{(0,1)^0}{0!} - e^{-0,1} \frac{(0,1)^1}{1!} = 0,0047.$$

3.51 Theo công thức (2.24) ta có thể xem $X = X_1 + \dots + X_n$, trong đó X_1, \dots, X_n là các biến ngẫu nhiên độc lập có cùng phân bố Poisson tham số $\frac{\lambda}{n}$. Kỳ vọng và phương sai $EX = DX = \lambda$.

Áp dụng định lý giới hạn trung tâm, công thức (4.11) ta được $P\left\{\frac{X - \lambda}{\sqrt{\lambda}} \leq y\right\} \approx \Phi(y)$

$$\text{Mặt khác } P\left\{\frac{X - \lambda}{\sqrt{\lambda}} \leq y\right\} = P\{X \leq \sqrt{\lambda}y + \lambda\} \Rightarrow P\{X \leq x\} \approx \Phi\left(\frac{x - \lambda}{\sqrt{\lambda}}\right).$$

HƯỚNG DẪN VÀ ĐÁP ÁN CHƯƠNG 4

4.1	4.2	4.3	4.4	4.5	4.6	4.7
Đúng	Sai	Đúng	Đúng	Đúng	Sai	Đúng

4.8 Mẫu ngẫu nhiên có kích thước 10: $W = (X_1, X_2, \dots, X_{10})$.

$$P\left\{\bar{X} = \frac{1}{2}\right\} = P\left\{\frac{1}{10} \sum_{i=1}^{10} X_i = \frac{1}{2}\right\} = P\left\{\sum_{i=1}^{10} X_i = 5\right\}.$$

Vì X có phân bố nhị thức nên

$$P\left\{\sum_{i=1}^{10} X_i = 5\right\} = P_{10}(5) = C_{10}^5 (0,5)^5 \cdot (0,5)^{10-5} = C_{10}^5 (0,5)^{10}.$$

4.9 X có phân bố chuẩn $N(\mu; \sigma^2)$ nên \bar{X} có phân bố chuẩn $N(\mu; \frac{\sigma^2}{n})$.

Vậy
$$P\left\{|\bar{X} - \mu| < \varepsilon\right\} = P\left\{\mu - \varepsilon < \bar{X} < \mu + \varepsilon\right\} = \Phi\left(\frac{\varepsilon\sqrt{n}}{\sigma}\right) - \Phi\left(\frac{-\varepsilon\sqrt{n}}{\sigma}\right) = 2\Phi\left(\frac{\varepsilon\sqrt{n}}{\sigma}\right) - 1.$$

Do đó
$$P\left\{|\bar{X} - 20| < 0,2\right\} = 2\Phi\left(\frac{0,2\sqrt{100}}{1}\right) - 1 = 2\Phi(2) - 1 = 0,9545.$$

4.10 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

$$F_{10}(x) = \begin{cases} 0 & x \leq 1 \\ 1/5 & 1 < x \leq 2 \\ 3/5 & 2 < x \leq 3 \\ 4/5 & 3 < x \leq 4 \\ 1 & x > 4 \end{cases}$$

$$\bar{x} = 2,4; s^2 = 1,15, s = 1,072.$$

4.11 Đặt $u_i = x_i - 26 \Rightarrow \bar{x} = \bar{u} + 26 = 26; s^2 = \frac{1}{99} \left(1080 - \frac{0^2}{100} \right) = 10,909.$

4.12 $\bar{x} = 19,672; s^2 = 0,1692; s = 0,413.$

4.13 b) $\bar{x} = 67,45; s^2 = 8,6136; s = 2,9349.$

4.14 a) 0,1056

b) 0,5714.

4.15 a) $E(\bar{X}) = \mu = 22,40$; $D(\bar{X}) = \frac{\sigma^2}{n} \Rightarrow \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{0,048}{\sqrt{6}} = 0,008$.

b) $\bar{X} \sim N(\mu; \frac{\sigma^2}{n})$;

$$P\{22,39 < \bar{X} < 22,41\} = \Phi\left(\frac{22,41 - 22,40}{0,008}\right) - \Phi\left(\frac{22,39 - 22,40}{0,008}\right) = 2\Phi(1,25) - 1 = 0,789$$

$$P\{\bar{X} > 22,42\} = 1 - \Phi\left(\frac{22,42 - 22,40}{0,008}\right) = 1 - \Phi(2,5) = 0,0054.$$

$$P\left\{\bar{X} < 22,38\right\} \cup \left\{\bar{X} > 22,41\right\} = \Phi(-2,5) + 1 - \Phi(1,25) = 2 - \Phi(2,5) - \Phi(1,25) = 0,066.$$

4.16 a) 237 (vì $300 \cdot 0,789 = 236,7$) b) 2 (vì $300 \cdot 0,0054 = 1,62$) c) 20 (vì $300 \cdot 0,066 = 19,8$).

4.17 $E(\bar{X}) = \mu = 800$; $D(\bar{X}) = \frac{\sigma^2}{n} \Rightarrow \sigma_{\bar{X}} = \frac{56}{\sqrt{8}} = 7$. $\bar{X} \sim N(\mu; \frac{\sigma^2}{n})$

a) $P\{790 < \bar{X} < 810\} = \Phi\left(\frac{810 - 800}{7}\right) - \Phi\left(\frac{790 - 800}{7}\right) = 2\Phi(1,43) - 1 = 0,8472$

b) $P\{\bar{X} < 788\} = \Phi\left(\frac{788 - 800}{7}\right) = \Phi(-1,71) = 1 - \Phi(1,71) = 0,0436$

c) $P\{\bar{X} > 820\} = 1 - \Phi\left(\frac{820 - 800}{7}\right) = 1 - \Phi(2,86) = 0,002$. d) $\varepsilon = 1,96 \cdot \frac{\sigma}{\sqrt{n}}$.

4.18 a) 0,0029

b) 0,0,9596

c) 0,1446.

4.19 $\bar{x} = 11,1417$; $s^2 = 0,5465$; $s = 0,7393$.

4.20 a) $\bar{x} = 64,164$; $s^2 = 33,3023$; $s = 5,77$

b) 72,1%; 93,3%; 99,76%.

HƯỚNG DẪN VÀ ĐÁP ÁN CHƯƠNG 5

5.1	5.2	5.3	5.4	5.5	5.6		
Đúng	Sai	Sai	Đúng	Đúng	Sai		
5.7	5.8	5.9	5.10	5.11	5.12	5.13	5.14
Sai	Sai	Đúng	Đúng	Sai	Sai	Sai	Sai

5.15 Mẫu ngẫu nhiên có kích thước 10: $W = (X_1, X_2, \dots, X_{10})$.

$$P\left\{\bar{X} = \frac{1}{2}\right\} = P\left\{\frac{1}{10} \sum_{i=1}^{10} X_i = \frac{1}{2}\right\} = P\left\{\sum_{i=1}^{10} X_i = 5\right\}.$$

Vì X có phân bố nhị thức nên

$$P\left\{\sum_{i=1}^{10} X_i = 5\right\} = P_{10}(5) = C_{10}^5 (0,5)^5 \cdot (0,5)^{10-5} = C_{10}^5 (0,5)^{10}.$$

5.16 X có phân bố chuẩn $N(\mu; \sigma^2)$ nên \bar{X} có phân bố chuẩn $N(\mu; \frac{\sigma^2}{n})$.

Vậy
$$P\left\{\left|\bar{X} - \mu\right| < \varepsilon\right\} = P\left\{\mu - \varepsilon < \bar{X} < \mu + \varepsilon\right\} = \Phi\left(\frac{\varepsilon\sqrt{n}}{\sigma}\right) - \Phi\left(\frac{-\varepsilon\sqrt{n}}{\sigma}\right) = 2\Phi\left(\frac{\varepsilon\sqrt{n}}{\sigma}\right) - 1.$$

Do đó
$$P\left\{\left|\bar{X} - 20\right| < 0,2\right\} = 2\Phi\left(\frac{0,2\sqrt{100}}{1}\right) - 1 = 2\Phi(2) - 1 = 0,9545.$$

5.17 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

$$F_{10}(x) = \begin{cases} 0 & x \leq 1 \\ 1/5 & 1 < x \leq 2 \\ 3/5 & 2 < x \leq 3 \\ 4/5 & 3 < x \leq 4 \\ 1 & x > 4 \end{cases}$$

$$\bar{x} = 2,4; s^2 = 1,15, s = 1,072.$$

5.18 Áp dụng bất đẳng thức Trêbusép ta có:

$$\begin{aligned} P\left\{\left|\widehat{\Theta}_n - \theta\right| \geq \varepsilon\right\} &\leq \frac{E\left[\left(\widehat{\Theta}_n - \theta\right)^2\right]}{\varepsilon^2} = \frac{1}{\varepsilon^2} E\left[\left\{\left(\widehat{\Theta}_n - E\widehat{\Theta}_n\right) + \left(E\widehat{\Theta}_n - \theta\right)\right\}^2\right] \\ &= \frac{1}{\varepsilon^2} E\left[\left(\widehat{\Theta}_n - E\widehat{\Theta}_n\right)^2 + \left(E\widehat{\Theta}_n - \theta\right)^2 + 2\left(\widehat{\Theta}_n - E\widehat{\Theta}_n\right)\left(E\widehat{\Theta}_n - \theta\right)\right] \\ &= \frac{1}{\varepsilon^2} \left[D\left(\widehat{\Theta}_n\right) + E\left(E\widehat{\Theta}_n - \theta\right)^2 + 2E\left(\widehat{\Theta}_n - E\widehat{\Theta}_n\right)\left(E\widehat{\Theta}_n - \theta\right)\right]. \end{aligned}$$

Vì vậy nếu $\lim_{n \rightarrow \infty} E\left(\widehat{\Theta}_n\right) = \theta$ và $\lim_{n \rightarrow \infty} D\left(\widehat{\Theta}_n\right) = 0$ thì $\lim_{n \rightarrow \infty} P\left\{\left|\widehat{\Theta}_n - \theta\right| \geq \varepsilon\right\} = 0$.

5.19 $f = \frac{1082}{2000}$; Điều kiện $\begin{cases} nf = 1082 > 10 \\ n(1-f) = 918 > 10 \end{cases}$

$$f - U_{\alpha/2} \frac{\sqrt{f(1-f)}}{\sqrt{n}} = \frac{1082}{2000} - 2,33 \frac{\sqrt{918 \times 1082}}{2000\sqrt{2000}} = 0,515$$

Vậy tối thiểu có 51,5% số phiếu bầu cho ứng cử viên A.

$$5.20 \quad \bar{x} = \frac{\sum x_i}{n} = \frac{34,15}{35} = 0,976.$$

$$s^2 = \frac{1}{n-1} \left[\sum x_i^2 - \frac{(\sum x_i)^2}{n} \right] = \frac{1}{34} \left[33,8943 - \frac{(34,15)^2}{35} \right] = 0,01687$$

$$\Rightarrow s = 0,1299; U_{\alpha/2} \frac{s}{\sqrt{n}} = 1,96 \times \frac{0,1299}{\sqrt{35}} = 0,043. \text{ Khoảng tin cậy 95\%: } [0,933; 1,019].$$

$$5.21 \quad \text{Tần suất mẫu } f = \frac{53}{400}, \text{ điều kiện } \begin{cases} nf = 53 > 10 \\ n(1-f) = 347 > 10 \end{cases}$$

Gọi p là xác suất bắt được con cá có đánh dấu, khoảng tin cậy 95% của p :

$$U_{\alpha/2} \frac{\sqrt{f(1-f)}}{\sqrt{n}} = 1,96 \frac{\sqrt{53 \times 347}}{400\sqrt{400}} = 0,0332$$

Khoảng ước lượng $[0,0993; 0,1657]$

Mặt khác $p = \frac{2000}{N}$, trong đó N là số cá trong hồ.

$$\text{Vậy } 0,0993 < \frac{2000}{N} < 0,1657 \Rightarrow \frac{2000}{0,1657} < N < \frac{2000}{0,0993}.$$

$$\Rightarrow 12070 < N < 20141$$

$$5.22 \quad \text{Đặt } u_i = \frac{x_i - 18,25}{5} \Rightarrow \bar{x} = 5 \frac{\sum r_i u_i}{n} + 18,25 = 5 \frac{-1,8}{40} + 18,25 = 18,025$$

$$s^2 = \frac{5^2}{n-1} \left[\sum r_i u_i^2 - \frac{(\sum r_i u_i)^2}{n} \right] = \frac{25}{39} \left[0,76 - \frac{(-1,8)^2}{40} \right] = 0,435$$

$$\Rightarrow s = 0,66; U_{\alpha/2} \frac{s}{\sqrt{n}} = 1,64 \times \frac{0,66}{\sqrt{40}} = 0,171$$

a) Khoảng tin cậy 90%: $[17,854; 18,196].$

b) Kích thước mẫu cần thiết $n \geq \frac{U_{\alpha/2}^2 s^2}{\varepsilon^2} = 116,99$ chọn $n = 117$

$$5.23 \quad \text{Đặt } u_i = x_i - 50 \Rightarrow \bar{x} = \frac{\sum r_i u_i}{n} + 50 = \frac{-8}{27} + 50 = 49,704$$

$$U_{\alpha/2} \frac{\sigma}{\sqrt{n}} = 1,96 \times \frac{1}{\sqrt{27}} = 0,377$$

a) Khoảng tin cậy 95%: $[49,327 ; 50,081]$.

b) Kích thước mẫu cần thiết $n \geq \frac{U_{\alpha/2}^2 \sigma^2}{\varepsilon^2} = 384,16$ chọn $n = 385$

5.24 Khoảng tin cậy 95% của phương sai được tính theo công thức (5.53).

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

Tra bảng χ^2 với 25 bậc tự do và với giả thiết $\hat{S}^2 = 0,5$ ta tìm được khoảng tin cậy:

$$\left(\frac{25 \cdot 0,5}{40,646} ; \frac{25 \cdot 0,5}{13,120} \right) = (0,3075 ; 0,9520).$$

5.25 Gọi μ là trọng lượng trung bình của một bao sản phẩm được đóng gói. Ta kiểm định giả thiết $H_0 : \mu = 100$; đối thiết $H_1 : \mu < 100$

Tiêu chuẩn kiểm định $T = \frac{(100 - \bar{X})\sqrt{n}}{S};$

Miền bác bỏ $W_\alpha = \{T > 2,086\}.$

Đặt $u_i = \frac{x_i - 99,25}{5} \Rightarrow \sum r_i u_i = 0,4 ; \sum r_i u_i^2 = 0,42$

$$\Rightarrow \bar{x} = 5 \times \frac{0,4}{29} + 99,25 = 99,319 ; \quad s^2 = 25 \times \frac{1}{28} \left[0,42 - \frac{0,4^2}{29} \right] = 0,37 \Rightarrow s = 0,608$$

$$T_{qs} = \frac{(100 - 99,319)\sqrt{29}}{0,608} = 6,032 \in W_\alpha.$$

Vậy bác bỏ H_0 chấp nhận H_1 , nghĩa là sản phẩm bị đóng thiếu.

5.26 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 $T = \frac{(\bar{X} - 14)\sqrt{n}}{S};$

Miền bác bỏ $W_\alpha = \{|T| > 1,96\}.$

Đặt $u_i = \frac{x_i - 15}{2} \Rightarrow \sum r_i u_i = 0 ; \sum r_i u_i^2 = 300$

$$\Rightarrow \bar{x} = 15 ; \quad s^2 = 4 \times \frac{1}{249} \left[300 - \frac{0}{300} \right] = 4,819 \Rightarrow s = 2,195$$

$$\Rightarrow T_{qs} = \frac{(115 - 14)\sqrt{300}}{2,195} = 7,89 \in W_\alpha.$$

Vậy bác bỏ H_0 chấp nhận H_1 , nghĩa là cần thay đổi định mức.

5.27 Gọi μ là mức hao phí xăng trung bình của ô tô chạy từ A đến B. Ta kiểm định giả thiết $H_0: \mu = 50$; đối thiết $H_1: \mu < 50$

$$\text{Tiêu chuẩn kiểm định} \quad T = \frac{(50 - \bar{X})\sqrt{n}}{S};$$

$$\text{Miền bác bỏ} \quad W_\alpha = \{T > 2,052\}.$$

$$\text{Theo mẫu ta có } \bar{x} = \frac{1387,5}{28} = 49,5536;$$

$$s^2 = \frac{1}{27} \left(6876375 - \frac{1387,5^2}{28} \right) = \frac{8,1696}{27} = 0,3026 \Rightarrow s = 0,55$$

$$T_{qs} = \frac{(50 - 49,53)\sqrt{30}}{0,55} = 4,2948 \in W_\alpha.$$

Vậy bác bỏ H_0 chấp nhận H_1 , nghĩa là mức hao phí xăng có giảm xuống.

5.28 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 $H_1: \mu > 1300$

$$\text{Tiêu chuẩn kiểm định} \quad T = \frac{(\bar{X} - 1300)\sqrt{n}}{S};$$

$$\text{Miền bác bỏ} \quad W_\alpha = \{T > 1,96\}.$$

$$\text{Từ mẫu cụ thể ta có} \quad T = \frac{(1378 - 1300)\sqrt{40}}{215} = 2,294 > 1,96$$

Vậy bác bỏ H_0 chấp nhận H_1 , nghĩa là hệ thống máy tính mới xử lý tốt hơn.

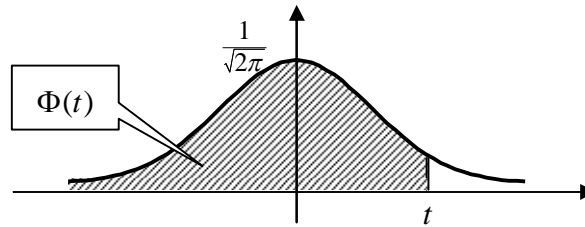
PHỤ LỤC 1: GIÁ TRỊ HÀM MẬT ĐỘ XÁC SUẤT PHÂN BỐ CHUẨN TẮC

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

	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	0065	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

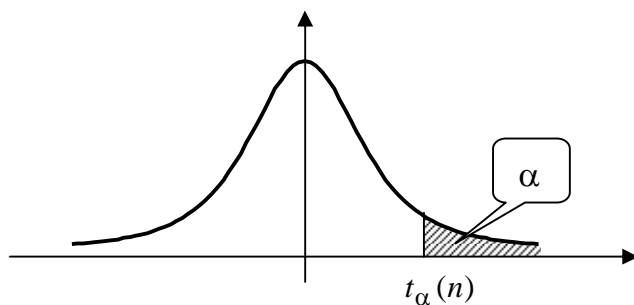
PHỤ LỤC 2: GIÁ TRỊ HÀM PHÂN BỐ CHUẨN TẮC

$$\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-\frac{x^2}{2}} dx$$



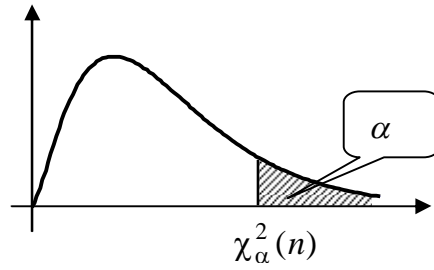
t	0	1	2	3	4	5	6	7	8	9
0,0	0,5000	5040	5080	5120	5160	5199	5239	5279	5319	5359
0,1	5398	5438	5478	5517	5557	5596	5636	5675	5714	5753
0,2	5793	5832	5871	5910	5948	5987	6026	6064	6103	6141
0,3	6179	6217	6255	6293	6331	6368	6406	6443	6480	6517
0,4	6554	6591	6628	6664	6700	6736	6772	6808	6844	6879
0,5	0,6915	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 3: GIÁ TRỊ TỚI HẠN CỦA PHÂN BỐ STUDENT



Bậc tự do	$\alpha = 0,05$	$\alpha = 0,025$	$\alpha = 0,01$	$\alpha = 0,005$	$\alpha = 0,002$
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	2,528	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 4: GIÁ TRỊ TỚI HẠN CỦA PHÂN BỐ “KHI BÌNH PHƯƠNG”



Bậc tự do	$\chi^2_{0,995}$	$\chi^2_{0,99}$	$\chi^2_{0,975}$	$\chi^2_{0,95}$	$\chi^2_{0,05}$	$\chi^2_{0,025}$	$\chi^2_{0,01}$	$\chi^2_{0,005}$
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	40,289	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	49,645
28	12,461	13,565	15,308	16,928	41,337	44,461	48,278	50,993
29	13,121	14,256	16,047	17,708	42,557	45,722	49,588	52,336
30	13,787	14,930	16,791	18,493	43,773	46,979	50,892	53,672

PHỤ LỤC 5: GIÁ TRỊ HÀM PHÂN BỐ POISSON

$$P\{X \leq k\} = \sum_{i=0}^k \frac{e^{-\lambda} \lambda^i}{i!}$$

$\lambda \backslash k$	0,1	0,2	0,3	0,4	0,5	0,6
0	0,904837	0,818731	0,740818	0,670320	0,606531	0,548812
1	0,995321	0,982477	0,963063	0,938448	0,909796	0,878099
2	0,999845	0,998853	0,996400	0,992074	0,985612	0,976885
3	0,999996	0,999943	0,999734	0,999224	0,998248	0,996642
4	1,000000	0,999998	0,999984	0,999939	0,999828	0,999606
5		1,000000	0,999999	0,999996	0,999986	0,999962
6				1,000000	0,999999	0,999997
7						1,000000

$\lambda \backslash k$	0,7	0,8	0,9	1,0	2,0	3,0
0	0,496585	0,449329	0,406570	0,367877	0,135335	0,049787
1	0,844195	0,808792	0,772483	0,735759	0,406006	0,199148
2	0,965858	0,952577	0,937144	0,919699	0,676677	0,423190
3	0,994246	0,990920	0,986542	0,981012	0,857124	0,647232
4	0,999214	0,998589	0,997657	0,996340	0,947348	0,815263
5	0,999909	0,999816	0,999658	0,999403	0,983437	0,916082
6	0,999990	0,999980	0,999958	0,999917	0,995467	0,966491
7	0,999999	0,999998	0,999997	0,999990	0,998904	0,988095
8			1,000000	0,999999	0,999753	0,996196
9				1,000000	0,999954	0,998897
10					0,999992	0,999707
11					0,999999	0,999928
12					1,000000	0,999983
13						0,999996
14						0,999999
15						1,000000

Phụ lục

$\lambda \backslash k$	4,0	5,0	6,0	7,0	8,0	9,0
0	0,018316	0,006738	0,002479	0,000912	0,000335	0,000123
1	0,091579	0,040428	0,017352	0,007295	0,003019	0,001234
2	0,238105	0,124652	0,061970	0,029636	0,013754	0,006232
3	0,433472	0,265026	0,151205	0,081765	0,042380	0,021228
4	0,785132	0,615960	0,445681	0,300708	0,191236	0,115690
5	0,889326	0,762183	0,606304	0,449711	0,313374	0,206780
6	0,948866	0,866628	0,743981	0,598711	0,452961	0,323896
7	0,978636	0,931806	0,847239	0,729091	0,592548	0,455652
8	0,991867	0,968172	0,916077	0,830496	0,716625	0,587408
9	0,997159	0,986305	0,957380	0,901479	0,815887	0,705988
10	0,999084	0,984547	0,979909	0,946650	0,888077	0,803008
11	0,999726	0,997981	0,991173	0,973000	0,936204	0,875773
12	0,999923	0,999202	0,996372	0,987188	0,965820	0,926149
13	0,999979	0,999774	0,998600	0,994282	0,982744	0,958533
14	0,999994	0,999931	0,999491	0,997593	0,991770	0,977964
15	0,999998	0,999980	0,999825	0,999041	0,996283	0,988894
16	0,999999	0,999994	0,999943	0,999637	0,998407	0,994680
17	0,999999	0,999998	0,999982	0,999869	0,999351	0,997573
18	0,999999	0,999999	0,999994	0,999955	0,999748	0,998943
19	1,000000	0,999999	0,999998	0,999985	0,999907	0,999560
20		1,000000	0,999999	0,999995	0,999967	0,999824
21			0,999999	0,999998	0,999989	0,999932
22			1,000000	0,999999	0,999997	0,999974
23				0,999999	0,999998	0,999990
24				1,000000	0,999999	0,999996
25					1,000000	0,999998
26						0,999999
27						1,000000

BẢNG CHỈ DẪN THUẬT NGỮ

Bản phân bố xác suất biên	73	Hàm khối lượng xác suất đồng thời	72
Bảng phân bố ghép lớp	125	Hàm mật độ xác suất đồng thời	76
Bảng phân bố tần số thực nghiệm	124	Hàm mật độ xác suất biên	76
Bảng phân bố tần suất thực nghiệm	124	Hàm đặc trưng	61
Bất đẳng thức Markov	111	Hàm phân bố xác suất đồng thời	70
Bất đẳng thức Trêbusép	112	Hàm của một biến ngẫu nhiên	84
Biểu đồ tần số hình gậy	126	Hàm của hai biến ngẫu nhiên	87
Biểu đồ đa giác tần suất	126	Hàm hợp lý	136
Biến cố sơ cấp	6	Hàm phân bố thực nghiệm của mẫu	124
Biến cố	6	Hàm mẫu của quá trình ngẫu nhiên	162
Biến cố chắc chắn	6	Hệ số nhọn	60
Biến cố không thể	7	Hiệp phương sai	80
Biến cố đối	7	Hệ đầy đủ biến cố	8
Biến cố xung khắc	7	Hệ số tương quan	8
Biến cố độc lập	8	Hoán vị	10
Biến ngẫu nhiên	31	Hội tụ theo xác suất	110
Biến ngẫu nhiên kỳ vọng có điều kiện	101	Hội tụ theo phân bố	111
Cá thể	122	Khoảng tin cậy	139
Chỉnh hợp	10	Không gian mẫu	6
Công thức xác suất đầy đủ	21	Không gian trạng thái	162
Công thức Bayes	21	Kỳ vọng	52, 80
Dấu hiệu nghiên cứu	122	Kỳ vọng có điều kiện	97
Định lý giới hạn trung tâm	115	Luật số lớn Trêbusép	113
Độ chính xác của ước lượng	140	Luật số lớn Bernoulli	114
Độ lệch chuẩn mẫu	130	Ma trận hiệp phương sai	81
Giá trị quan sát tiêu chuẩn kiểm định	149	Mẫu ngẫu nhiên	123
Giả thiết thống kê	148	Miền bác bỏ	149
Hàm phân bố xác suất	32	Mốt	59
Hàm khối lượng xác suất	35	Moment	60
Hàm mật độ xác suất	42	Mức ý nghĩa của kiểm định	149
Hệ số bất đối xứng	60	Phép thử	6
		Phép thử Bernoulli	24
		Phân bố Bernoulli	38

Phân bố nhị thức	38	Bảng phân bố ghép lớp	125
Phân bố Poisson	40	Bảng phân bố tần số thực nghiệm	124
Phân bố đều	44	Bảng phân bố tần suất thực nghiệm	124
Phân bố mũ	45	Bất đẳng thức Markov	111
Phân bố Erlang	46	Bất đẳng thức Trêbusép	112
Phân bố chuẩn	47	Biểu đồ tần số hình gậy	126
Phân bố “khi bình phương”	50	Biểu đồ đa giác tần suất	126
Phân bố Student	51	Biến cố sơ cấp	6
Phương sai	56	Biến cố	6
Phân vị	57	Biến cố chắc chắn	6
Phân bố có điều kiện	97	Biến cố không thể	7
Phân bố chuẩn nhiều chiều	102	Biến cố đối	7
Phương pháp ước lượng hợp lý cực đại	136	Biến cố xung khắc	7
Phương pháp ước lượng moment	138	Biến cố độc lập	8
Phương sai mẫu	129	Biến ngẫu nhiên	31
Quy tắc hai xích ma, ba xích ma	50	Biến ngẫu nhiên kỳ vọng có điều kiện	101
Sai lầm loại một sai lầm loại hai	150	Cá thể	122
Tần suất mẫu	130	Chỉnh hợp	10
Tính độc lập của biến ngẫu nhiên	79	Công thức xác suất đầy đủ	21
Thủ tục kiểm định giả thiết thống kê	151	Công thức Bayes	21
Tích biến cố	7	Dấu hiệu nghiên cứu	122
Tổ hợp	11	Định lý giới hạn trung tâm	115
Tổ chức đồ	127	Độ chính xác của ước lượng	140
Tổng thể	122	Độ lệch chuẩn mẫu	130
Tổng biến cố	7	Giá trị quan sát của tiêu chuẩn kiểm định	149
Trung bình mẫu	128	Giả thiết thống kê	148
Trung vị	58	Hàm phân bố xác suất	32
Ước lượng điểm	133	Hàm khối lượng xác suất	35
Ước lượng không chệch	134	Hàm mật độ xác suất	42
Ước lượng hiệu quả	134	Hệ số bất đối xứng	60
Ước lượng vững	135	Hàm khối lượng xác suất đồng thời	72
Véc tơ ngẫu nhiên	70	Hàm mật độ xác suất đồng thời	76
Xác suất có điều kiện	18	Hàm mật độ xác suất biên	76
Bản phân bố xác suất biên	73		

Hàm đặc trưng	61	Phân bố mũ	45
Hàm phân bố xác suất đồng thời	70	Phân bố Erlang	46
Hàm của một biến ngẫu nhiên	84	Phân bố chuẩn	47
Hàm của hai biến ngẫu nhiên	87	Phân bố “khi bình phương”	50
Hàm hợp lý	136	Phân bố Student	51
Hàm phân bố thực nghiệm của mẫu	124	Phương sai	56, 80
Hàm mẫu của quá trình ngẫu nhiên	162	Phân vị	57
Hệ số nhọn	60	Phân bố có điều kiện	97
Hiệp phương sai	80	Phân bố chuẩn nhiều chiều	102
Hệ đầy đủ biến cố	8	Phương pháp ước lượng hợp lý cực đại	136
Hệ số tương quan	8	Phương pháp ước lượng moment	138
Hoán vị	10	Phương sai mẫu	129
Hội tụ theo xác suất	110	Quy tắc hai xích ma, ba xích ma	50
Hội tụ theo phân bố	111	Sai lầm loại một sai lầm loại hai	150
Khoảng tin cậy	139	Tần suất mẫu	130
Không gian mẫu	6	Tính độc lập của biến ngẫu nhiên	79
Không gian trạng thái	162	Thủ tục kiểm định giả thiết thống kê	151
Kỳ vọng	52, 80	Tích biến cố	7
Kỳ vọng có điều kiện	97	Tổ hợp	11
Luật số lớn Trêbusép	113	Tổ chức đồ	127
Luật số lớn Bernoulli	114	Tổng thể	122
Ma trận hiệp phương sai	81	Tổng biến cố	7
Mẫu ngẫu nhiên	123	Trung bình mẫu	128
Miền bác bỏ	149	Trung vị	58
Một	59	Ước lượng điểm	133
Moment	60	Ước lượng không chệch	134
Mức ý nghĩa của kiểm định	149	Ước lượng hiệu quả	134
Phép thử	6	Ước lượng vững	135
Phép thử Bernoulli	24	Véc tơ ngẫu nhiên	70
Phân bố Bernoulli	38	Xác suất có điều kiện	18
Phân bố nhị thức	38		
Phân bố Poisson	40		
Phân bố đều	44		

TÀI LIỆU THAM KHẢO

- [1]. Lê Bá Long, 2009. Giáo trình Xác suất và thống kê. NXB Thông tin và Truyền thông.
- [2]. Tổng Đình Quỳ, *Hướng dẫn giải bài tập xác suất thống kê*, NXB Đại Học Quốc Gia Hà Nội, 2004.
- [3]. Đặng Hùng Thắng, 1997. *Mở đầu về lý thuyết xác suất và các ứng dụng*. NXB GD.
- [4]. Đặng Hùng Thắng, *Bài tập xác suất*, NXB Giáo dục – 1998.
- [5]. Đặng Hùng Thắng, *Thống kê và ứng dụng*, NXB Giáo dục, 1999.
- [6]. Nguyễn Duy Tiến, Vũ Việt Yên, 2000. *Lý thuyết xác suất*. NXB GD.
- [7]. Trần Mạnh Tuấn, *Xác suất và Thống kê, lý thuyết và thực hành tính toán*, NXB Đại Học Quốc Gia Hà Nội, 2004.
- [8]. Nguyễn Cao Văn và Trần Thái Ninh, *Bài giảng xác suất và thống kê toán*, NXB Thống kê, Hà Nội 1999.
- [9]. 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.
- [10]. B.V. Gnedenko, *The theory of probability*, Mir publishers, Moscow 1976.
- [11]. HWEI P. HSU, Ph.D., *Theory and problems Probability, Random variables, and Random processes*, Schaum's outline series. McGRAW-HILL, 1996.
- [12]. Murray R. Spiegel, John J. Schiller, R. Alu Srinivasan, *Probability and Statistics*, Schaum's outline series. McGRAW-HILL, 2009.