

California Polytechnic State University, San Luis Obispo

JAY L. DEVORE

Tài liệu môn học

**Probability and Statistics for Engineering and
Sciences**

XÁC SUẤT THỐNG KÊ ỨNG DỤNG

Biên soạn:

Chương 2,3,4 - Nguyễn Ngọc Tứ

Bộ môn Toán - ĐH SPKT, Tp. Hồ Chí Minh - Năm 2017

Mục lục

1	PHÉP TÍNH XÁC SUẤT	3
1.1	Không gian mẫu và biến cố	3
1.2	Các tiên đề và tính chất của xác suất	6
1.3	Giải tích tổ hợp	10
1.4	Xác suất có điều kiện	14
1.5	Sự độc lập	19
2	BIẾN NGẪU NHIÊN RỜI RẠC VÀ PHÂN PHỐI XÁC SUẤT	25
2.1	Biến ngẫu nhiên	25
2.2	Phân phối xác suất của biến ngẫu nhiên rời rạc	27
2.3	Kỳ vọng và phương sai	32
2.4	Phân phối nhị thức	35
2.5	Phân phối siêu bội và phân phối nhị thức âm	42
2.6	Phân phối Poisson	45
3	BIẾN NGẪU NHIÊN LIÊN TỤC VÀ PHÂN PHỐI XÁC SUẤT	52
3.1	Hàm mật độ xác suất	52
3.2	Hàm phân phối tích lũy và các số đặc trưng	55
3.3	Phân phối chuẩn	61
3.4	Phân phối mũ và Gamma	67
3.5	Một số phân phối liên tục khác	71
4	PHÂN PHỐI XÁC SUẤT ĐỒNG THỜI VÀ MẪU NGẪU NHIÊN	81
4.1	Phân phối đồng thời của các biến ngẫu nhiên	81
4.1.1	Hai biến ngẫu nhiên rời rạc	82
4.1.2	Hai biến ngẫu nhiên liên tục	83

4.1.3	Các biến ngẫu nhiên độc lập	85
4.1.4	Nhiều hơn 2 biến	86
4.1.5	Phân phối điều kiện	87
4.2	Giá trị Kỳ vọng, Phương sai và Hiệp phương sai	87
4.2.1	Hiệp phương sai	89
4.2.2	Tương quan	90
4.3	Các phân phối thống kê	91
4.3.1	Mẫu ngẫu nhiên	92
4.3.2	Tìm phân phối mẫu	93
4.3.3	Những thí nghiệm mô phỏng	93
4.4	Phân phối của trung bình mẫu	93
4.4.1	Trường hợp phân phối chuẩn tổng thể	94
4.4.2	Định lý giới hạn trung tâm	95
4.4.3	Những ứng dụng khác của định lý giới hạn trung tâm	97
4.5	Phân phối của tổ hợp tuyến tính	98
4.5.1	Hiệu của hai biến ngẫu nhiên	99
4.5.2	Trường hợp của biến ngẫu nhiên liên tục	100

Chương 1

PHÉP TÍNH XÁC SUẤT

1.1 Không gian mẫu và biến cố

Không gian mẫu của một phép thử

Một **phép thử** là một hành động hay một quá trình mà cho kết quả một cách ngẫu nhiên.

Định nghĩa 1.1.1. Không gian mẫu của một phép thử là tập hợp tất cả các kết quả có thể xảy ra của phép thử, thông thường được kí hiệu là $\Omega, \mathcal{S}$.

Ví dụ 1.1 Một phép thử đơn giản nhất là chỉ cho kết quả có hai kết quả. Chẳng hạn, tung một đồng xu thì chỉ có hai khả năng xảy ra là xuất hiện mặt sấp H hoặc mặt ngửa T . Khi đó, ta có không gian mẫu của phép thử tung đồng xu là $\Omega = \{H, T\}$. Một trường hợp khác trong thực tế là giới tính của một trẻ sơ sinh mà có không gian mẫu là $\Omega = \{M, F\}$.

Ví dụ 1.2 Nếu ta tung đồng xu ba lần và ghi lại các kết quả xảy ra của mỗi lần thì một kết quả của cả ba lần tung là dãy H hoặc T . Vì thế,

$$\Omega = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}.$$

Ví dụ 1.3 Gieo hai con xúc xắc (6 mặt) đồng thời thì ta có 36 khả năng xảy ra với không gian mẫu là

$$\Omega = \{(i, j), \text{ với } i, j = 1, \dots, 6\}.$$

Định nghĩa 1.1.2. Một **biến cố** là một tập hợp của một số kết quả trong không gian mẫu Ω . Một biến cố **đơn** nếu nó chỉ chứa một kết quả và một biến cố **kép** nếu nó chứa nhiều hơn một kết quả.

Ví dụ 1.4 Xét một phép thử về việc chuyển hướng của ba phương tiện giao thông tại một ngã ba, nghĩa là xe sẽ rẽ trái (L) hoặc phải (R) tại ngã ba. Không gian mẫu sẽ gồm 8 trường hợp

$$\Omega = \{LLL, LLR, LRL, LRR, RLL, RLR, RRL, RRR\}.$$

Do đó, ta có được 8 biến cố đơn và một số biến cố kép, chẳng hạn

$$\begin{aligned} A &= \{RLL, LRL, LLR\} = \{\text{biến cố chỉ một xe rẽ phải}\} \\ B &= \{LLL, RLL, LRL, LLR\} = \{\text{biến cố nhiều nhất một xe rẽ phải}\} \\ C &= \{LLL, RRR\} = \{\text{biến cố cả 3 xe rẽ cùng hướng}\} \end{aligned}$$

Ví dụ 1.5 Tiếp tục ví dụ 1.1, ta có 36 biến cố đơn $E_1 = (1, 1), \dots, E_{36} = (6, 6)$. Ví dụ một số biến cố kép

$$\begin{aligned} A &= \{(1, 1), (2, 2), (3, 3), (4, 4), (5, 5), (6, 6)\} \\ &= \{\text{biến cố 2 xúc xuất hiện mặt giống nhau}\} \\ B &= \{(1, 3), (2, 2), (3, 1)\} = \{\text{biến cố tổng 2 mặt bằng 4}\}. \end{aligned}$$

Mối quan hệ của lý thuyết tập hợp

Định nghĩa 1.1.3. i) Phần bù của biến cố A , kí hiệu A' , là tập hợp tất cả các kết quả trong không gian mẫu Ω mà không chứa trong biến cố A .

ii) Hợp của hai biến cố A và B , kí hiệu $A \cup B$, đọc là " A hoặc B ", là biến cố chứa tất cả các biến cố của A hoặc của B , hoặc đồng thời của cả 2 biến cố, (nghĩa là chỉ cần ít nhất một biến cố xảy ra.)

iii) Giao của hai biến cố A và B , kí hiệu $A \cap B$, đọc là " A và B ", là biến cố chứa tất cả các kết quả của cả 2 biến cố A và B .

iv) Biến cố rỗng, kí hiệu $\emptyset$, là biến cố không chứa kết quả nào của phép thử. Khi $A \cap B = \emptyset$ thì ta gọi A và B là hai biến cố rời nhau.

Ví dụ 1.6 Tiếp tục ví dụ 1.1, ta có các biến cố sau $A = \{1, 2, 3, 4\}$, $B = \{3, 4, 5, 6\}$, $C = \{1, 3, 5\}$, $D = \{6\}$. Khi đó,

$$A' = \{5, 6\}, A \cup B = \{1, 2, 3, 4, 5, 6\}, A \cup C = \{1, 2, 3, 4, 5\}, A \cap B = \{3, 4\}$$

$$A \cap C = \{1, 3\}, (A \cap C)' = \{1, 3\}, A \cap D = \emptyset, C \cap D = \emptyset.$$

Bài tập: Không gian mẫu và biến cố

1.1.1 Bốn trường đại học 1, 2, 3, 4 đang tham gia một giải đấu bóng rổ. Trong vòng đầu tiên 1 sẽ đấu với 2 và 3 sẽ đấu với 4. Sau đó, hai đội chiến thắng sẽ đấu để giành chức vô địch và hai đội thua cũng sẽ đấu với nhau. Một kết quả được biểu thị bằng 1324 (1 chơi với 2, 3 chơi với 4 trong trận đầu tiên sau đó 1 chơi với 3, 2 chơi với 4).

- Xác định không gian mẫu.
- Gọi A là biến cố 1 thắng. Hãy xác định A .
- Gọi B là biến cố 2 có tham gia để giành chức vô địch. Hãy xác định B .
- Xác định $A + B$; $A \cdot B$ và A' là gì?

1.1.2 Giả sử rằng các phương tiện đi theo lối ra một xa lộ đặc biệt có thể rẽ phải (R), rẽ trái (L) hoặc đi thẳng (S). Quan sát hướng của 3 xe.

- Liệt kê tất cả các kết quả có trong biến cố A cả 3 xe đều đi cùng một hướng.
- Liệt kê tất cả các kết quả có trong biến cố B 3 xe có hướng đi khác nhau.
- Liệt kê tất cả các kết quả có trong biến cố C hai trong ba xe có hướng rẽ phải.
- Liệt kê tất cả các kết quả có trong biến cố D hai trong ba xe có cùng hướng đi.
- Liệt kê kết quả có trong D' ; $C + D$ và CD .

1.1.3 Một hệ thống được kết nối với nhau bởi ba thành phần như trong sơ đồ dưới. Bởi vì các hệ thống con 2-3 được mắc song song với nhau nên hệ thống con sẽ hoạt động nếu ít nhất một trong hai thành phần 2 hoặc 3 hoạt động. Để toàn bộ hệ thống hoạt động thì thành phần 1 phải hoạt động và có thêm hoạt động của thành phần 2 hoặc 3. Thử nghiệm hệ thống bằng cách xác định sự hoạt động của mỗi thành phần là S (thành công) cho một thành phần hoạt động và F (thất bại) cho một thành phần không hoạt động.

- Liệt kê tất cả các kết quả có trong A là biến cố hai trong ba thành phần hoạt động.
- Liệt kê tất cả các kết quả có trong B là biến cố ít nhất hai thành phần hoạt động.
- Liệt kê tất cả các kết quả có trong C là biến cố hệ thống hoạt động.
- Liệt kê kết quả có trong C' ; $A + C$; AC ; $B + C$; BC .

1.1.4 Một gia đình bao gồm ba người A, B, và C đến một trạm y tế có 3 phòng 1, 2 và 3. Trong một tuần nhất định, mỗi thành viên trong gia đình thăm phòng khám

một lần và được phân ngẫu nhiên đến phòng bất kì. Thí nghiệm ngẫu nhiên bao gồm ghi số phòng cho mỗi thành viên. Một kết quả là (1, 2, 1) cho A đến phòng 1, B đến phòng 2, và C đến phòng 1.

- a. Xác định không gian mẫu.
- b. Liệt kê tất cả các kết quả có trong trường hợp cả ba thành viên đến cùng một phòng.
- d. Liệt kê tất cả các kết quả có trong trường hợp tất cả thành viên đi đến các phòng khác nhau.
- e. Liệt kê tất cả các kết quả có trong trường hợp không có ai đi đến phòng 2.

1.2 Các tiên đề và tính chất của xác suất

Cho một phép thử và một không gian mẫu tương ứng Ω , xác suất của một biến cố A là độ đo chính xác khả năng xảy ra của biến cố A đó.

Tiên đề 1.2.1. 1. **Tính không âm:** $P(A) \geq 0$, với biến cố A bất kì.

2. **Tính cộng tính:** Nếu A và B là hai biến cố rời nhau thì

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

Hơn thế nữa, nếu $A_1, A_2, \dots$ là tập hợp vô hạn các biến cố rời nhau thì

$$P(A_1 \cup A_2 \cup A_3 \dots) = \sum_{j=1}^{\infty} P(A_j).$$

3. **Tính chuẩn hóa:** $P(\Omega) = 1$.

Mệnh đề 1.2.2. 1. $P(\emptyset) = 0$.

2. Cho biến cố A bất kì, $P(A) + P(A') = 1$.

3. $P(A) \leq 1$.

4. Cho ba biến cố A, B và C bất kì

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

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

Ví dụ 1.7 Trong một thị trấn, 60% hộ gia đình sử dụng dịch vụ Internet của công ty **A**, 80% hộ gia đình sử dụng dịch vụ truyền hình cáp cũng của công ty **A**, và 50% hộ gia đình sử dụng cả hai dịch vụ của công ty này. Chọn ngẫu nhiên một hộ gia đình, tính xác suất để hộ gia đình này sử dụng ít nhất một dịch vụ của công ty **A** và xác suất để hộ gia đình này chỉ sử dụng duy nhất một loại dịch vụ?

Đặt $A = \{\text{sử dụng Internet}\}$ và $B = \{\text{sử dụng truyền hình cáp}\}$, ta có

$$P(A) = 0.6, P(B) = 0.8 \text{ và } P(A \cap B) = 0.5.$$

Khi đó,

$$\begin{aligned} &P(\text{biến cố sử dụng ít nhất một loại dịch vụ}) \\ &= P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.6 + 0.8 - 0.5 = 0.9. \\ &P(\text{sử dụng chỉ một dịch vụ}) \\ &= P(A \cup B) - P(A \cap B) = 0.9 - 0.5 = 0.4. \end{aligned}$$

Định nghĩa 1.2.3. Nếu một không gian mẫu có n kết quả có thể xảy ra như nhau (nghĩa là mỗi biến cố đơn có xác suất như nhau) thì xác suất của một biến cố A bất kì xảy ra được cho bởi

$$P(A) = \frac{\text{Số lượng kết quả chứa trong } A}{n} = \frac{n(A)}{n}.$$

Ví dụ 1.8 Nam có 6 sách tiểu thuyết và 6 sách khoa học giả tưởng chưa đọc. Ba quyển sách đầu tiên của mỗi loại là sách bìa cứng và 3 sách còn lại là bìa mềm. Nam chọn ngẫu nhiên một sách tiểu thuyết và một sách khoa học giả tưởng mang đi du lịch. Tính xác suất Nam chọn được cả hai sách bìa mềm?

Giả sử ta đánh số $1, \dots, 6$ chọn sách tiểu thuyết và tương tự cho sách khoa học giả tưởng. Khi đó, ta có 36 khả năng chọn được hai quyển sách với xác suất như nhau. Và số cách chọn được hai sách bìa mềm là 9 cách như sau $\{(4, 4), (4, 5), (4, 6), (5, 4), (5, 5), (5, 6), (6, 4), (6, 5), (6, 6)\}$. Vậy xác suất của biến cố A chọn hai sách bìa mềm là

$$P(A) = \frac{n(A)}{n} = \frac{9}{36}.$$

Bài tập: Các tiên đề và tính chất của xác suất

1.2.1 Một công ty quỹ tương hỗ cung cấp cho khách hàng một loạt các quỹ: quỹ thị trường tiền tệ, ba quỹ trái phiếu khác nhau (ngắn, trung, dài hạn), hai quỹ chứng

khoán (vừa và cao) và một quỹ cân bằng. Khách sở hữu cổ phần chỉ trong một quỹ, tỷ lệ phần trăm của các khác hàng trong các quỹ như sau:

Thị trường tiền tệ :	20%	Trái phiếu dài hạn :	18%
Trái phiếu ngắn hạn :	15%	Trái phiếu trung hạn :	25%
Chứng khoán vừa :	10%	Chứng khoán cao :	5%
Cân bằng :	7%		

- a. Tính xác suất một cá nhân sở hữu cổ phần trong quỹ cân bằng.
- b. Tính xác suất một cá nhân sở hữu cổ phần trong quỹ trái phiếu.
- c. Tính xác suất một cá nhân không sở hữu cổ phần trong quỹ chứng khoán.

1.2.2 Một công ty tư vấn máy tính hiện nay đã đề xuất ba dự án. Cho A_i là dự án thứ i được chọn với $i = 1, 2, 3$ và giả sử rằng $P(A_1) = 0,22$; $P(A_2) = 0,25$; $P(A_3) = 0,28$; $P(A_1.A_2) = 0,11$; $P(A_1.A_3) = 0,05$; $P(A_2.A_3) = 0,07$; $P(A_1.A_2.A_3) = 0,01$.

Tính xác suất của các biến cố sau:

- a. $A_1 + A_2$
- b. $A'_1.A'_2$
- c. $A_1 + A_2 + A_3$
- d. $A'_1.A'_2.A'_3$
- e. $A'_1.A'_2.A_3$
- f. $A'_1.A'_2 + A_3$

1.2.3 Quan sát 2 loại máy sấy quần áo (sử dụng khí gas hoặc điện) được mua bởi năm khách hàng khác nhau tại một cửa hàng.

a. Nếu xác suất một máy sấy điện được mua là 0,428; tính xác suất ít nhất hai máy sấy điện được mua?

b. Nếu $P(\text{cả năm người mua máy sấy loại sử dụng khí gas}) = 0,116$ và $P(\text{cả năm người mua máy sấy loại điện}) = 0,005$. Tính xác suất ít nhất một máy của mỗi loại được mua?

1.2.4 Cho A là biến cố: yêu cầu hỗ trợ từ tư vấn phần mềm thống kê với bộ SPSS, B là biến cố: yêu cầu hỗ trợ từ SAS. Giả sử $P(A) = 0,3$; $P(B) = 0,5$.

- a. Tại sao $P(A) + P(B) \neq 1$
- b. Tính $P(A')$
- c. Tính $P(A + B)$
- d. Tính $P(A'B')$

1.2.5 Kiểm tra các mối hàn trên mạch in bằng mắt người có thể cho kết quả rất chủ quan. Một phần của vấn đề bắt nguồn từ một số loại lỗi (ví dụ như khả năng hiển thị, lỗ rỗng). Do đó, ngay cả những thanh tra viên có trình độ cao cũng có thể không đồng ý về việc bố trí của một khớp cụ thể. Trong một đợt kiểm tra 10.000 khớp, thanh tra A đã phát hiện được 724 khiếm khuyết, thanh tra viên B tìm thấy 751 khớp có lỗi và có 1159 khớp bị đánh giá là có khuyết điểm bởi ít nhất một trong

hai thanh tra. Giả sử rằng trong 10.000 khớp được chọn ngẫu nhiên.

a. Tính xác suất công ty chọn ra được một khớp bị đánh giá khiếm khuyết bởi cả hai thanh tra.

b. Tính xác suất một khớp bị đánh giá là bị lỗi bởi thanh tra viên B nhưng không bởi thanh tra A.

1.2.6 Một công ty bảo hiểm cung cấp bốn mức khấu trừ khác nhau cho hợp đồng bảo hiểm nhà là không có, thấp, trung bình và cao; ba cấp độ bảo hiểm ô tô là thấp, trung bình và cao. Bảng dưới là tỷ lệ cho các loại bảo hiểm khác nhau của cả hai loại bảo hiểm. Ví dụ như tỷ lệ cá nhân có khấu trừ mức thấp đối với bảo hiểm ô tô và thấp bảo hiểm nhà ở là 0,06.

Bảo hiểm nhà ở				
Bảo hiểm ô tô	Không	Thấp	Trung bình	Cao
Thấp	0,04	0,06	0,05	0,03
Trung bình	0,07	0,10	0,20	0,10
Cao	0,02	0,03	0,15	0,15

Lấy ngẫu nhiên một cá nhân có cả hai loại hình bảo hiểm ô tô và nhà ở.

a. Tính xác suất cá nhân này có một khoản khấu trừ bảo hiểm ô tô mức trung bình và khoản khấu trừ bảo hiểm nhà ở mức cao.

b. Tính xác suất cá nhân được lấy ra có khoản khấu trừ bảo hiểm ô tô mức thấp. Tính xác suất cá nhân được lấy ra có khoản khấu trừ bảo hiểm nhà ở mức thấp.

c. Tính xác suất cá nhân này có khoản khấu trừ bảo hiểm ô tô và bảo hiểm nhà ở cùng mức.

d. Dựa vào câu trả lời câu c, tính xác suất cá nhân được lấy ra có hai khoản khấu trừ bảo hiểm là khác mức.

e. Xác suất cá nhân được lấy ra có ít nhất một khấu trừ bảo hiểm mức thấp là bao nhiêu?

g. Dựa vào câu trả lời của câu e, xác suất cá nhân được lấy ra có mức khấu trừ bảo hiểm không thấp là bao nhiêu?

1.2.7 Ba sự lựa chọn phổ biến cho một loại xe mới là: GPS tích hợp (A), cửa sổ trời (B) và hộp số tự động (C). Nếu 40% người mua yêu cầu A, 55% yêu cầu B, 70% yêu cầu C, 63% yêu cầu A hoặc B, 77% yêu cầu A hoặc C, 80% yêu cầu B hoặc C và 85% yêu cầu A hoặc B hoặc C. Tính xác suất của những biến cố sau [Gọi ý: “A hoặc B” là biểu thị cho ít nhất một trong hai lựa chọn A, B được yêu cầu].

- a. Tính xác suất người mua tiếp theo sẽ yêu cầu ít nhất một trong ba lựa chọn.
- b. Tính xác suất người mua tiếp theo sẽ không chọn ba lựa chọn này.
- c. Tính xác suất người mua tiếp theo sẽ chỉ yêu cầu hộp số tự động và không yêu cầu một trong hai lựa chọn khác.
- d. Tính xác suất người mua tiếp theo sẽ yêu cầu một trong ba lựa chọn.

1.2.8 Một bộ phận học thuật có năm thành viên là: Anderson, Box, Cox, Cramer và Fisher phải lựa chọn hai thành viên của mình để phục vụ trong một ủy ban xét duyệt nhân sự. Bởi vì công việc sẽ tốn nhiều thời gian, nên người ta quyết định rằng người đại diện sẽ được lựa chọn bằng cách đặt tên trên những mảnh giấy giống hệt nhau và chọn ngẫu nhiên hai mảnh giấy

- a. Tính xác suất chọn được Anderson và Box.
- b. Tính xác suất ít nhất một trong hai thành viên có tên bắt đầu bằng C được chọn.
- c. Nếu năm thành viên trên đã từng giảng dạy 3, 6, 7, 10 và 14 năm tại trường đại học thì xác suất hai đại diện được chọn có tổng số ít nhất 15 năm kinh nghiệm giảng dạy ở đó là bao nhiêu?

1.3 Giải tích tổ hợp

Mệnh đề 1.3.1. Quy tắc nhân: Giả sử ta có một tập hợp có thứ tự gồm có k phần tử và có n_1 cách chọn cho phần tử đầu tiên; với mỗi cách chọn phần tử đầu tiên, ta lại có n_2 cách chọn phần tử thứ hai; ...; với mỗi cách chọn phần tử thứ $k-1$, ta lại có n_k cách chọn phần tử thứ k . Do đó có $n_1 n_2 \dots n_k$ cách chọn k phần tử.

Ví dụ 1.9 Một người chủ nhà yêu cầu cung cấp dịch vụ điện và nước. Có 12 đơn vị dịch vụ về nước và 9 đơn vị về điện. Hỏi có bao nhiêu cách để người chủ nhà chọn dịch vụ điện và nước?

Có 12 cách chọn dịch vụ về nước và với mỗi cách chọn dịch vụ về nước thì người chủ nhà lại có 9 cách chọn dịch vụ về điện. Vì thế người chủ nhà có $12 \times 9 = 108$ cách.

Định nghĩa 1.3.2. Giả sử ta xét cách chọn k phần tử từ một tập hợp có n phần tử khác nhau. Nếu thứ tự chọn lựa có kể đến thì cách chọn lựa này gọi là một **hoán vị**; ngược lại nếu không kể đến thứ tự chọn lựa thì ta gọi là một **tổ hợp**, kí hiệu là $C_{k,n}$ hoặc $\binom{n}{k}$.

Giả sử ta có n phần tử khác nhau, cho k là một số dương và $k \leq n$. Ta đếm số cách khác nhau từ việc chọn k phần tử từ tập hợp n phần tử và sắp xếp thành một dãy, nghĩa là số lượng dãy của k phần tử phân biệt. Khi đó, ta có n cách chọn phần tử đầu tiên. Với cách chọn phần tử đầu tiên, ta chỉ còn $n - 1$ cách chọn phần tử thứ hai, v.v. ... Khi ta chọn phần tử thứ k thì ta đã chọn được $k - 1$ phần tử, vì thế ta sẽ có $n - (k - 1)$ cách chọn phần tử cuối cùng. Do đó, theo quy tắc nhân, số lượng các dãy có thể xảy ra mà ta gọi là k -**hoán vị** hay còn gọi là **chỉnh hợp** chập k của n , là

$$\begin{aligned} n(n-1) \dots (n-k+1) &= \frac{n(n-1) \dots (n-k+1)(n-k) \dots 2 \cdot 1}{(n-k) \dots 2 \cdot 1} \\ &= \frac{n!}{(n-k)!} := P_{k,n} \end{aligned}$$

Đặc biệt, khi $k = n$, số lượng dãy có thể mà ta gọi là **hoán vị** là

$$n \cdot (n-1) \cdot (n-2) \dots 2 \cdot 1 = n!$$

Để đếm số lượng tổ hợp, ta chú ý rằng việc chọn một k -hoán vị chính là việc chọn lần đầu tiên của một tổ hợp của k phần tử và sau đó sắp xếp lại chúng có thứ tự. Vì thế, ta có $k!$ cách chọn có thứ tự của k phần tử đã chọn trước và ta chỉ ra rằng số lượng của k -hoán vị bằng với số lượng của tổ hợp nhân với $k!$. Do đó, số lượng của tổ hợp là

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

Ví dụ 1.10 Có mười trợ giảng để chấm thi cho môn giải tích tại một trường đại học. Bài kiểm tra đầu tiên có bốn câu hỏi và giáo sư muốn chọn trợ giảng khác nhau chấm điểm cho từng câu hỏi (mỗi trợ giảng chỉ chấm một câu hỏi). Hỏi có bao nhiêu cách chọn trợ giảng để chấm thi.

Số cách chọn là số lượng của 4-hoán vị của tập hợp 10 phần tử:

$$P_{k,n} = \frac{10!}{(10-4)!} = 5040.$$

Ví dụ 1.11 Một danh sách bài hát trong iPod thì chứa 100 bài hát và có 10 bài hát của nhóm Beatle. Giả sử danh sách nhạc này phát ngẫu nhiên. Tính xác suất để bài hát của nhóm Beatle xuất hiện ở lần iPod phát ra bài hát thứ năm?

Để bài hát của nhóm Beatle xuất hiện ở lần thứ năm thì bốn lần đầu không xuất hiện bài hát của nhóm Beatle. Do đó, số cách để xuất hiện bài hát của nhóm

Beatle ở lần thứ năm là $90 \cdot 89 \cdot 88 \cdot 87 \cdot 10$. Và số cách xuất hiện năm bài hát bất kì là $100 \cdot 99 \cdot 98 \cdot 97 \cdot 96$. Vì thế, ta có xác suất xuất hiện bài hát của nhóm Beatle ở lần thứ 5 là

$$\frac{90 \cdot 89 \cdot 88 \cdot 87 \cdot 10}{100 \cdot 99 \cdot 98 \cdot 97 \cdot 96} =$$

Bài tập: Giải tích tổ hợp

1.3.1 Một cửa hàng âm thanh cung cấp mức giá đặc biệt cho một bộ hoàn chỉnh các thiết bị âm thanh gồm 4 thiết bị (receiver, compact disc player, speakers, turntable). Một người mua sẽ được lựa chọn các nhà sản xuất cho mỗi bộ phận.

Receiver: Kenwood, Onkyo, Pioneer, Sony, Sherwood

Compact disc player: Onkyo, Pioneer, Sony, Technics

Speakers: Boston, Infinity, Polk

Turntable: Onkyo, Sony, Teac, Technics

Bảng trên cho thấy các hãng của từng bộ phận để cho khách hàng có thể lựa chọn để kết nối với nhau (mỗi bộ phận chỉ được chọn 1 hãng). Hãy trả lời các câu hỏi dưới đây:

- a. Có bao nhiêu cách tạo thành một bộ âm thanh hoàn chỉnh?
- b. Có bao nhiêu cách tạo thành một bộ âm thanh hoàn chỉnh trong đó thiết bị receiver và compact disc player là của hãng Sony?
- c. Có bao nhiêu cách tạo thành một bộ âm thanh hoàn chỉnh mà các thành phần của nó không phải của hãng Sony?
- d. Có bao nhiêu cách tạo thành một bộ âm thanh hoàn chỉnh trong đó có ít nhất một thiết bị của hãng Sony?
- e. Nếu lựa chọn ngẫu nhiên các thiết bị để tạo thành một bộ âm thanh hoàn chỉnh thì xác suất mà bộ âm thanh có ít nhất một thiết bị của hãng Sony là bao nhiêu? Xác suất để có đúng một thiết bị của hãng Sony là bao nhiêu?

1.3.2 Lỗi bàn phím có thể là lỗi do mạch điện hoặc lỗi cơ học. Một cơ sở sửa chữa bàn phím hiện đang có 25 bàn phím lỗi, trong đó có 6 bàn phím lỗi do mạch điện và 19 bàn phím lỗi do cơ học.

- a. Có bao nhiêu cách chọn ngẫu nhiên 5 bàn phím?
- b. Có bao nhiêu cách chọn ngẫu nhiên 5 bàn phím trong đó có đúng hai bàn phím bị lỗi do mạch điện?
- c. Tính xác suất trong 5 bàn phím được chọn ngẫu nhiên có ít nhất 4 bàn phím bị lỗi do cơ học.

1.3.3 Một cơ sở sản xuất sử dụng 20 công nhân làm việc ca ngày, 15 công nhân làm việc ca chiều và 10 công nhân làm việc ca đêm. Một người kiểm soát chất lượng chọn 6 nhân viên để kiểm tra. Giả sử việc lựa chọn được thực hiện sao cho bất kỳ nhóm 6 công nhân nào cũng có cơ hội được lựa chọn như nhau.

a. Có bao nhiêu cách lựa chọn được 6 công nhân là của ca ngày. Tính xác suất lựa được 6 công nhân là của ca ngày?

b. Xác suất để 6 công nhân được lựa chọn đều đến từ chung một ca làm việc là bao nhiêu?

c. Xác suất để 6 công nhân được lựa chọn đến từ ít nhất hai ca làm việc khác nhau là bao nhiêu?

d. Xác suất để ít nhất một trong 3 ca làm việc sẽ không có công nhân đại diện là bao nhiêu?

1.3.4 Một nhà thử nghiệm đang nghiên cứu ảnh hưởng của nhiệt độ, áp suất và chất xúc tác lên hiệu suất của một phản ứng hóa học nhất định. Ba nhiệt độ khác nhau, bốn áp suất khác nhau và năm chất xúc tác khác nhau đang được xem xét.

a. Nếu sử dụng bất kỳ nhiệt độ, áp suất và chất xúc tác vào thử nghiệm thì có thể phải chạy thử nghiệm bao nhiêu lần?

b. Có bao nhiêu thử nghiệm nếu sử dụng nhiệt độ thấp nhất và áp suất thấp nhất?

c. Giả sử có năm lần thử nghiệm khác nhau sẽ được thực hiện vào ngày thứ nhất. Nếu năm lựa chọn được chọn ngẫu nhiên trong số tất cả các khả năng có thể xảy ra sao cho bất kỳ nhóm năm nào cũng có xác suất được chọn như nhau, xác suất chất xúc tác khác nhau được sử dụng cho mỗi lần chạy là bao nhiêu?

1.3.5 5 điện thoại di động, 5 điện thoại không dây và 5 điện thoại có dây vừa được chuyển đến một trung tâm bảo trì. Giả sử rằng các điện thoại này được bảo trì theo thứ tự được phân bố ngẫu nhiên là 1, 2, 3, ..., 15.

a. Tính xác suất tất cả điện thoại không dây được bảo trì trong 10 điện thoại được bảo trì đầu tiên.

b. Tính xác suất sau khi bảo trì được 10 điện thoại thì chỉ còn hai trong số ba loại điện thoại còn lại cần được bảo trì.

c. Tính xác suất 6 điện thoại được bảo trì đầu tiên chỉ thuộc hai loại.

1.3.6 Số PIN cá nhân của thẻ ATM gồm 4 chữ số được tạo từ các số 0, 1, 2, ..., 9.

a. Có bao nhiêu số PIN khác nhau nếu không có hạn chế về sự lựa chọn các chữ số?

b. Theo một đại diện của chi nhánh tại địa phương của Chase Bank, có những hạn chế thực tế về việc lựa chọn các chữ số. Các lựa chọn sau đây bị cấm:

- i. Tất cả bốn chữ số giống hệt nhau.
- ii. Các chữ số tăng dần hoặc giảm dần, chẳng hạn như 6543.
- iii. Bất kỳ trình tự nào bắt đầu từ 19 (năm sinh là quá dễ đoán).

Tính xác một trong những mã PIN được chọn ngẫu nhiên là mã PIN hợp lệ.

c. Có người đã đánh cắp một thẻ ATM và biết chữ số đầu tiên và cuối của mã PIN là 8 và 1. Anh ta có ba lần thử trước khi thẻ được giữ lại bởi máy ATM. Vì vậy, anh ta chọn ngẫu nhiên các chữ số thứ hai và thứ ba cho lần thử đầu tiên, sau đó chọn ngẫu nhiên một cặp các chữ số khác nhau cho lần thử thứ hai và chọn ngẫu nhiên một cặp số khác cho lần thử thứ ba (người này biết về những hạn chế được mô tả trong phần trên vì vậy chỉ lựa chọn trong các khả năng hợp lệ). Tính xác suất người này có thể truy cập vào tài khoản.

d. Tính lại xác suất trong (c) nếu các chữ số đầu tiên và cuối cùng tương ứng là 1 và 1.

1.4 Xác suất có điều kiện

Định nghĩa 1.4.1. Cho trước hai biến cố A và B sao cho $P(B) > 0$. Khi đó, xác suất của biến cố A với điều kiện biến cố B xảy ra trước được xác định bởi

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad (1.1)$$

Ví dụ 1.12 Người ta khảo sát những người đi mua điện thoại ở một cửa hàng điện tử thì nhận thấy rằng có 60% mua thêm thẻ nhớ, 40% mua thêm sạc dự phòng và 30% mua cả hai. Chọn một người đi mua hàng bất kỳ, và gọi

$A = \{\text{mua thêm bộ nhớ}\}$

$B = \{\text{mua thêm sạc}\}$

$C = \{\text{mua thêm cả hai}\}$

Khi đó, $P(A) = 0.6, P(B) = 0.4, P(C) = 0.3$ Giả sử, người này đã mua sạc dự phòng thì xác suất người này cũng mua thẻ nhớ là

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0.3}{0.4} = 0.75$$

Nghĩa là, trong tất cả những người mua sạc dự phòng có 75% trong số đó mua thêm thẻ nhớ. Tương tự,

$$P(\text{sạc}|\text{thẻ nhớ}) = P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{0.3}{0.6} = 0.5$$

Lưu ý rằng: $P(A|B) \neq P(A); P(B|A) \neq P(B)$.

Công thức nhân xác suất: Từ công thức (1.1), ta có

$$P(A \cap B) = P(A|B) \cdot P(B) = P(B|A) \cdot P(A)$$

Mở rộng cho n biến cố $A_1, A_2, \dots, A_n$, ta có

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

Ví dụ 1.13 Một công ty sản xuất máy nghe nhạc cho DVD gồm 3 loại khác nhau: 50% loại I, 30% loại II, 20% loại III. Công ty bảo hành một năm cho các sản phẩm của mình. Qua khảo sát công ty thu thập được số liệu cần bảo hành trong một năm cho 3 loại sản phẩm là: 25% cho loại I, 20% cho loại II và 10% cho loại III.

1. Tính xác suất người mua sản phẩm loại I phải bảo hành trong năm đầu tiên?
2. Người mua hàng chọn ngẫu nhiên một sản phẩm từ 3 loại sản phẩm, tính xác suất sản phẩm này phải bảo hành?
3. Một người đến bảo hành sản phẩm, tính xác suất sản phẩm này thuộc loại nào có khả năng nhất?

Gọi $A_i = \{\text{sản phẩm loại } i, \text{ với } i = 1, 2, 3\}$. Gọi $B = \{\text{sản phẩm phải bảo hành}\}$. Khi đó, ta có:

$$P(A_1) = 0.5, P(A_2) = 0.3, P(A_3) = 0.2, P(B|A_1) = 0.25, P(B|A_2) = 0.2, P(B|A_3) = 0.1.$$

Do đó, người mua sản phẩm loại I phải bảo hành trong năm đầu tiên là

$$P(A_1 B) = P(B|A_1) P(A_1) = 0.25 \cdot 0.5 = 0.125.$$

và một người mua bất kì một trong ba sản phẩm phải bảo hành là

$$P(B) = P(B|A_1) + P(B|A_2) + P(B|A_3) = 0.125 + 0.06 + 0.02 = 0.205$$

Khi đó,

$$P(A_1|B) = \frac{P(A_1B)}{P(B)} = \frac{0.125}{0.205} = 0.61$$

$$P(A_2|B) = \frac{P(A_2B)}{P(B)} = \frac{0.06}{0.205} = 0.29$$

$$P(A_3|B) = \frac{P(A_3B)}{P(B)} = \frac{0.02}{0.205} = 0.1$$

Công thức toàn phần

Cho $A_1, A_2, \dots, A_k$ là họ các biến cố đầy đủ và đôi một rời nhau. Khi đó, với biến cố B bất kì, ta có

$$\begin{aligned} P(B) &= P(B|A_1)P(A_1) + \dots + P(B|A_k)P(A_k) \\ &= \sum_{i=1}^k P(B|A_i)P(A_i) \end{aligned} \quad (1.2)$$

Định lý 1.4.2. Công thức Bayes

Cho $A_1, A_2, \dots, A_k$ là họ các biến cố đầy đủ và đôi một rời nhau với xác suất tiên nghiệm $P(A_i), i = 1, \dots, k$. Khi đó, với biến cố B bất kì với điều kiện $P(B) > 0$, ta có xác suất hậu nghiệm của biến cố A_j với điều kiện biến cố B đã xảy ra là

$$P(A_j|B) = \frac{P(A_jB)}{P(B)} = \frac{P(B|A_j)P(A_j)}{\sum_{i=1}^k P(B|A_i)P(A_i)}, \quad j = 1, \dots, k. \quad (1.3)$$

Ví dụ 1.14 *Tỉ lệ của một bệnh hiếm.* Thống kê về một loại bệnh hiếm gặp cho thấy trong 1000 người lớn thì chỉ có 1 người mắc bệnh. Để kiểm tra điều này người ta tiến hành xét nghiệm cho kết quả chính xác 99% nếu một người bị nhiễm bệnh, và sai số 2% đối với một người không nhiễm bệnh. Chọn ngẫu nhiên một người và xét nghiệm cho kết quả dương tính, tính xác suất người này có bệnh thật sự?

Gọi A_1 = người có bệnh, A_2 = người không bệnh và B = xét nghiệm dương tính. Khi đó, ta có $P(A_1) = 0.001, P(A_2) = 0.999, P(B|A_1) = 0.99$ và $P(B|A_2) = 0.02$. Vì thế, xác suất cần tìm là

$$P(A_1|B) = \frac{P(A_1B)}{P(B)} = \frac{P(B|A_1)P(A_1)}{P(B|A_1)P(A_1) + P(B|A_2)P(A_2)} = 0.047$$

Bài tập: Xác suất có điều kiện

1.4.1 Ở một trạm xăng nào đó có 40% người sử dụng Gas thông thường (A_1), 35% sử dụng Gas + (A_2) và 25% sử dụng Gas premium (A_3). Trong số những khách hàng sử dụng Gas thông thường thì chỉ có 30% người đổ đầy bình chứa của họ (B), tương tự cho người sử dụng Gas + và Gas premium là 60% và 50%.

a. Tính xác suất khách hàng tiếp theo sẽ yêu cầu đổ Gas + đầy vào bình chứa của họ.

b. Tính xác suất khách hàng tiếp theo sẽ yêu cầu đổ đầy vào bình chứa của họ?

c. Nếu khách hàng tiếp theo đổ đầy vào bình chứa của họ, xác suất họ yêu cầu đổ Gas thông thường, Gas +, Gas premium lần lượt là bao nhiêu?

1.4.2 Giả thuyết rằng 70% máy bay hạng nhẹ biến mất trong khi đang bay ở một quốc gia nào đó sau đó sẽ được phát hiện. Trong số máy bay được phát hiện có 60% máy bay có định vị khẩn cấp, trong số máy bay không được phát hiện thì có 90% máy bay không có định vị khẩn cấp. Giả sử một chiếc máy bay hạng nhẹ đã biến mất.

a. Nếu nó có định vị khẩn cấp, xác suất nó không được phát hiện là bao nhiêu?

b. Nếu nó không có định vị khẩn cấp, xác suất nó được phát hiện là bao nhiêu?

1.4.3 Các chi tiết của một loại sản phẩm nào đó được vận chuyển đến nhà cung cấp bằng cách vận chuyển 10 chi tiết trong 1 lô. Giả sử rằng 50% tất cả các lô đó không chứa các chi tiết bị lỗi, 30% chứa một chi tiết bị lỗi và 20% chứa hai chi tiết bị lỗi. Hai chi tiết từ một lô được chọn ngẫu nhiên và kiểm tra. Tính xác suất lô được chọn có 0, 1 và 2 chi tiết bị lỗi dựa vào điều kiện sau đây:

a. Cả 2 chi tiết được kiểm tra là không bị lỗi.

b. Một trong hai chi tiết được kiểm tra là bị lỗi.

1.4.4 Giả sử thống kê các du khách đang trong kỳ nghỉ thấy có 40% kiểm tra email công việc, 30% sử dụng điện thoại di động để kết nối với công việc, 25% mang theo laptop, 23% kiểm tra email công việc và sử dụng điện thoại để kết nối với công việc, 51% không sử dụng điện thoại, email và laptop. Ngoài ra trong 100 người có 80 người mang laptop thì họ cũng kiểm tra email công việc và trong 100 người có 70 người sử dụng điện thoại để kết nối với công việc thì họ cũng mang theo laptop.

a. Một khách du lịch được chọn ngẫu nhiên và thấy rằng người này có kiểm tra email công việc, tính xác suất người này sử dụng điện thoại để kết nối với công việc.

b. Tính xác suất một khách du lịch mang theo laptop cũng sử dụng điện thoại để kết nối với công việc.

c. Nếu một khách du lịch được chọn ngẫu nhiên mà người này có kiểm tra email công việc và mang theo laptop, tính xác suất người đó sử dụng điện thoại để kết nối với công việc.

1.4.5 Trong vài năm qua đã có rất nhiều tranh cãi về các loại giám sát phù hợp để ngăn ngừa khủng bố. Giả sử một hệ thống giám sát có 99% khả năng xác định chính xác một tên khủng bố và 99,9% khả năng xác định chính xác một người không phải là khủng bố trong tương lai. Giả sử rằng trong 300 triệu dân có tới 1000 kẻ khủng bố trong tương lai. Chọn ngẫu nhiên một người trong số 300 triệu người, người này bị hệ thống xác định là một kẻ khủng bố trong tương lai. Khả năng người được chọn đó thực sự là tên khủng bố trong tương lai là bao nhiêu? Liệu giá trị của xác suất này có khiến chúng ta tin tưởng vào hệ thống giám sát hay không? Giải thích.

1.4.6 Một người bạn sống ở Los Angeles thường xuyên có các chuyến đi đến Washington, D.C để tham quan. Tỷ lệ % cô ấy sử dụng các hãng hàng không 1, 2, 3 lần lượt là 50%; 30% và 20%. Đối với hãng hàng không số 1, các chuyến bay trễ đến D.C là 30% và trễ đến L.A là 10%. Đối với hãng hàng không số 2, tỷ lệ này là 25% và 20%, trong khi đối với hãng hàng không số 3 tỷ lệ này là 40% và 25%. Nếu chúng ta biết rằng trong một chuyến đi đặc biệt, cô ấy đã đến muộn ở một trong hai điểm đến, tính xác suất lần bay sau người phụ nữ này sử dụng các hãng hàng không số 1, số 2 và số 3? Giả sử rằng cơ hội đến muộn ở LA là không bị ảnh hưởng bởi những gì xảy ra trên chuyến bay tới DC .

1.4.7 Trong bài tập 2.4.1 ta bổ sung thêm cách sử dụng thẻ tín dụng của khách hàng như sau:

70% tổng số khách hàng sử dụng Gas thông thường và đồ đầy bình chứa sử dụng thẻ tín dụng.

50% tổng số khách hàng sử dụng Gas thông thường và không đồ đầy bình chứa sử dụng thẻ tín dụng.

60% tổng số khách hàng sử dụng Gas + và đồ đầy bình chứa sử dụng thẻ tín dụng.

50% tổng số khách hàng sử dụng Gas + và đồ không đầy bình chứa sử dụng thẻ tín dụng.

50% tổng số khách hàng sử dụng Gas premium và đồ đầy bình chứa sử dụng thẻ tín dụng.

40% tổng số khách hàng sử dụng Gas premium và không đồ đầy bình chứa sử dụng thẻ tín dụng.

Tính xác suất cho khách hàng tiếp theo đến

- a. Sử dụng Gas + và đổ đầy bình chứa và có sử dụng thẻ tín dụng.
- b. Sử dụng Gas premium và không đổ đầy bình chứa và có sử dụng thẻ tín dụng.
- c. Sử dụng Gas premium và có sử dụng thẻ tín dụng.
- d. Đổ đầy bình và sử dụng thẻ tín dụng.
- e. Sử dụng thẻ tín dụng.
- g. Nếu khách hàng tiếp theo sử dụng thẻ tín dụng, xác suất họ sử dụng Gas premium là?

1.5 Sự độc lập

Định nghĩa 1.5.1. Hai biến cố A và B gọi là độc lập nếu $P(A|B) = P(A)$; ngược lại gọi là phụ thuộc.

Mệnh đề 1.5.2. A và B độc lập nếu và chỉ nếu

$$P(AB) = P(A) \cdot P(B).$$

Ví dụ 1.15 Một công ty sản xuất máy giặt và máy sấy có tỉ lệ bảo hành sản phẩm trong 1 năm lần lượt là 30% và 10%. Một người mua cả hai sản phẩm này, tính xác suất cả hai máy đều phải bảo hành?

Gọi A là biến cố bảo hành máy giặt và B là biến cố bảo hành máy sấy. Ta có: $P(A) = 0.3, P(B) = 0.1$. Khi đó, xác suất để bảo hành 2 máy là

$$P(AB) = P(A) \cdot P(B) = 0.03.$$

Định nghĩa 1.5.3. Các biến cố $A_1, \dots, A_n$ gọi là độc lập nếu

$$P(\cap_{i \in S} A_i) = \prod_{i \in S} P(A_i), \quad \text{với } S \text{ là tập con của } \{1, 2, \dots, n\}.$$

Sự độc lập

1.5.1 Một công ty về dầu có hai dự án đang hoạt động, một ở Châu Á và một ở Châu Âu. Gọi A là biến cố dự án Châu Á thành công, B là biến cố dự án Châu Âu thành công. Giả sử A, B độc lập với $P(A) = 0,4; P(B) = 0,7$.

- a. Nếu dự án Châu Á không thành công, thì xác suất dự án Châu Âu không thành công là bao nhiêu?
- b. Tính xác suất có ít nhất một trong hai dự án sẽ thành công.

c. Biết có ít nhất một dự án thành công, tính xác suất chỉ có dự án Châu Á thành công.

1.5.2 Một hệ thống bao gồm 4 thành phần kết nối với nhau theo 2 nhánh. Nhánh ở trên gồm các thành phần 1 và 2 được kết nối song song với nhau, do đó để nhánh trên hoạt động thì ít nhất một trong hai hệ thống 1, 2 phải hoạt động. Nhánh dưới gồm các thành phần 3, 4 nối tiếp với nhau, nhánh dưới sẽ hoạt động nếu cả 2 thành phần 3, 4 cùng hoạt động. Hệ thống sẽ hoạt động nếu một trong hai nhánh hoạt động. Nếu các thành phần hoạt động độc lập và tính xác suất nhánh trên và nhánh dưới cùng hoạt động. Tính xác suất hệ thống hoạt động.

1.5.3 Khi kiểm tra các xe tại trạm kiểm tra khí thải thì thấy có 70% xe có lượng khí thải vượt quá quy định. Giả sử các xe kế tiếp vượt qua kiểm tra hoặc không vượt qua kiểm tra là độc lập với nhau. Tính các xác suất sau:

- a. 3 xe tiếp theo vượt qua kiểm tra.
- b. Ít nhất một trong 3 xe không vượt qua kiểm tra.
- c. Có đúng 1 xe trong 3 xe vượt qua kiểm tra.
- d. Tối đa 1 xe trong 3 xe vượt qua kiểm tra.
- e. Biết rằng có ít nhất 1 trong 3 xe vượt qua kiểm tra, tính xác suất cả 3 xe đều vượt qua.

1.5.4 a. Một công ty gỗ nhận vận chuyển 10.000 tấm gỗ. Giả sử 20% những tấm này còn quá tươi để sử dụng trong công trình thi công quan trọng. 2 tấm được chọn một cách ngẫu nhiên, lần lượt từng cái một. Gọi A là biến cố tấm đầu tiên chọn còn tươi và B là biến cố tấm được chọn tiếp theo cũng tươi. Tính $P(A), P(B), P(AB)$ với A, B độc lập với nhau.

b. Với A, B độc lập với nhau và $P(A) = P(B) = 0,2$ thì $P(AB) = ?$ Khác biệt giữa câu trả lời này và câu trả lời ở câu (a) là bao nhiêu ? Với mục đích là tính toán $P(AB)$ liệu ta có thể giả sử rằng A, B ở câu (a) là độc lập để có thể có được xác suất chính xác không.

c. Cho một lô bao gồm 10 tấm trong đó có 2 tấm gỗ còn tươi. Vậy thì việc giả sử rằng A, B độc lập bây giờ có mang lại câu trả lời xấp xỉ với câu trả lời $P(AB)$ không? Điểm khác biệt chính so với câu (a) là gì? Khi nào bạn nghĩ rằng một giả định về tính độc lập sẽ có giá trị khi nhận được một câu trả lời gần đúng với $P(AB)$?

1.5.5 Lựa chọn ngẫu nhiên một người để thử nghiệm ba loại xe khác nhau 1, 2, 3 với biến cố A_i là thích xe số $i, i = 1, 2, 3$. Giả sử:

$$P(A_1) = 0,55; \quad P(A_2) = 0,65; \quad P(A_3) = 0,7;$$

$$P(A_1 + A_2) = 0,8; \quad P(A_2.A_3) = 0,4; \quad P(A_1 + A_2 + A_3) = 0,88$$

- a. Tính xác suất người được chọn thích xe số 1 và 2.
- b. Tính $P(A_2|A_3)$.
- c. A_2 và A_3 có độc lập với nhau hay không? Trả lời theo 2 cách khác nhau.
- d. Nếu người được chọn không thích xe số 1, xác suất người đó thích một trong hai chiếc còn lại là bao nhiêu?

Bài tập tổng hợp

TH 2.1 Một công ty sản xuất nhỏ có làm việc ca đêm. Có 20 thợ máy làm việc cho công ty.

- a. Giả sử một đêm làm việc có 3 thợ máy tạo thành một nhóm, hỏi có bao nhiêu cách lập nhóm làm ca đêm?
- b. Nếu các thợ máy được xếp hạng 1, 2, . . . , 20 theo năng lực từng người, hỏi có bao nhiêu nhóm mà các nhóm đó không có thợ máy tốt nhất?
- c. Có bao nhiêu nhóm sẽ có ít nhất 1 trong số 10 thợ máy tốt nhất?
- d. Nếu một trong những nhóm được lựa chọn ngẫu nhiên để làm việc vào một ca đêm đặc biệt, tính xác suất thợ máy tốt nhất sẽ không làm việc tối hôm đó?

TH 2.2 Một vệ tinh dự kiến phóng từ Cape Canaveral ở Florida và một cái khác ở căn cứ quân sự Vandenberg ở California. Gọi A là biến cố vệ tinh ở Vandenberg phóng thành công và B là biến cố vệ tinh ở Cape Canaveral phóng thành công. Giả sử, $P(A) > P(B)$; $P(A + B) = 0,626$ và $P(AB) = 0,144$. Tính $P(A)$ và $P(B)$.

TH 2.3 Máy phát đang gửi một tin nhắn bằng cách sử dụng một mã nhị phân là một chuỗi các số 0 và 1. Mỗi bit truyền (0 hoặc 1) phải đi qua ba role để tiếp cận người nhận. Ở mỗi role, xác suất mà một bit gửi đi sẽ khác với bit đã nhận được (một sự đảo ngược) là 0,2. Giả sử rằng các role hoạt động độc lập với nhau.

Máy phát $\rightarrow$ Role 1 $\rightarrow$ Role 2 $\rightarrow$ Role 3 $\rightarrow$ Người nhận

- a. Nếu một bit 1 được gửi từ máy phát, tính xác suất bit 1 được gửi qua được cả ba role.
- b. Nếu một bit 1 được gửi từ máy phát, tính xác suất bit 1 này được nhận bởi người nhận? [Gợi ý: tám kết quả thực nghiệm có thể được hiển thị trên một sơ đồ

cây với ba thể hệ của các chi nhánh, một thể hệ cho mỗi lần tiếp.]

c. Giả sử 70% số bit được gửi từ máy phát là những bit 1. Nếu một người nhận được bit 1, tính xác suất một bit 1 đã được gửi đi.

TH 2.4 Người A có năm người bạn thân (B, C, D, E, và F). A đã nghe một tin đồn từ bên ngoài và đã mời năm người bạn đến một bữa tiệc để thông báo tin đồn. Để bắt đầu, A chọn ngẫu nhiên một trong số năm và nói tin đồn với cá nhân đã chọn. Cá nhân đó sau đó chọn ngẫu nhiên một trong bốn người còn lại và lặp lại tin đồn. Tiếp tục, một cá nhân mới được lựa chọn từ những người chưa từng nghe tin đồn của cá nhân vừa nghe nó, cho đến khi mọi người được biết tin đồn.

- Tính xác suất tin đồn được lặp lại theo thứ tự B, C, D, E, và F.
- Tính xác suất F là người thứ ba được nghe tin đồn.
- Tính xác suất F là người cuối cùng nghe thấy tin đồn.
- Nếu ở mỗi giai đoạn người hiện đang "có" tin đồn không biết ai đã nghe nó và chọn ngẫu nhiên người nhận tiếp theo từ tất cả năm cá nhân, tính xác suất F vẫn chưa nghe thấy gì tin đồn sau khi nó đã được nói mười lần trong bữa tiệc?

TH 2.5 Một kỹ sư hóa học quan tâm đến việc xác định một tạp chất trong một loại sản phẩm. Biết độ chính xác của xét nghiệm kiểm tra như sau: xác suất phát hiện tạp chất nếu nó có trong sản phẩm là 0,8 và xác suất không phát hiện thấy tạp chất nếu nó không có trong sản phẩm là 0,9. Tỷ lệ sản phẩm này có tạp chất là 60%. Xét nghiệm tạp chất một sản phẩm được chọn ngẫu nhiên. Biết trong ba xét nghiệm riêng biệt chỉ có hai lần xét nghiệm phát hiện có tạp chất, tính xác suất tạp chất có mặt trong sản phẩm này.

TH 2.6 Một loại ốc vít được sử dụng trong sản xuất máy bay được tạo răng để khóa lại và tránh nổi lỏng do rung động trong quá trình sử dụng. Giả sử 95% số ốc vít vượt qua được quá trình kiểm tra ban đầu. Trong số 5% số ốc vít không qua kiểm tra ban đầu có 20% bị lỗi nghiêm trọng nên phải bỏ đi, số còn lại được gửi đến sửa lại trong đó có 40% không sửa được 60% còn lại được điều chỉnh và thông qua kiểm tra.

- Tính xác suất một ốc vít được lựa chọn ngẫu nhiên vượt qua kiểm tra ban đầu hoặc sau khi được điều chỉnh và vượt qua kiểm tra.
- Biết một ốc ít đã được thông qua quá trình kiểm tra, xác suất nó đã vượt

qua kiểm tra ban đầu và không cần điều chỉnh lại là bao nhiêu?

TH 2.7 Một phần trăm dân số một vùng có người mắc một loại bệnh cụ thể. Một loại xét nghiệm cho ra kết quả dương tính đối với 90% người có bệnh và 5% đối với những người không mắc bệnh. Giả sử các xét nghiệm được áp dụng độc lập với hai mẫu máu khác nhau từ cùng một cá thể được lựa chọn ngẫu nhiên.

- a. Tính xác suất mà cả hai kết quả xét nghiệm đều có chung một kết luận.
- b. Nếu cả hai xét nghiệm dương tính, tính xác suất cá nhân được lựa chọn là người có mắc bệnh.

TH 2.8 Một hệ thống bao gồm hai thành phần. Xác suất thành phần thứ hai hoạt động tốt là 0,9; xác suất ít nhất một trong số hai thành phần này hoạt động tốt là 0,96 và xác suất cả hai thành phần hoạt động tốt là 0,75. Biết thành phần đầu tiên hoạt động tốt, tính xác suất thành phần thứ hai cũng hoạt động tốt.

TH 2.9 Một công ty nhất định sẽ gửi 40% bưu kiện qua đêm qua dịch vụ chuyển phát nhanh E1. Trong những bưu kiện này, 2% đến sau thời gian giao hàng được đảm bảo (đến trễ). Nếu một hồ sơ lưu trữ qua đêm được chọn ngẫu nhiên từ tệp của công ty, tính xác suất lô hàng đó đã đi qua E1 và đến trễ. Giả sử rằng 50% bưu kiện được gửi qua dịch vụ chuyển phát nhanh E2 và 10% còn lại được gửi qua E3. Trong số gửi qua E2, chỉ có 1% đến trễ, trong khi 5% lô hàng do E3 xử lý đến muộn.

- a. Chọn ngẫu nhiên một lô hàng, tính xác suất nó đến muộn?
- b. Nếu bưu kiện được chọn ngẫu nhiên đã đến đúng giờ, tính xác suất nó không được gửi qua E1?

TH 2.10 Một công ty sử dụng ba dây chuyền lắp ráp khác nhau là A1, A2, và A3 để sản xuất một loại sản phẩm cụ thể. Những sản phẩm sản xuất bởi dây chuyền A1, A2, A3 cần phải khắc phục khuyết điểm chiếm tỷ lệ lần lượt là 5%, 8% và 10%. Giả sử rằng 50% sản phẩm được sản xuất theo dây chuyền A1, 30% được sản xuất theo dây chuyền A2, và 20% sản xuất theo dây chuyền A3. Chọn ngẫu nhiên một sản phẩm thấy sản phẩm này cần phải điều chỉnh lại, tính xác suất sản phẩm này được sản xuất bởi dây chuyền A1? Dây chuyền A2? Dây chuyền A3?

TH 2.11 Giả thuyết rằng ta không quan tâm đến khả năng sinh nhật vào ngày

29 tháng 2. Giả sử một người được lựa chọn ngẫu nhiên có khả năng đã được sinh ra vào bất kỳ một ngày trong 365 ngày khác nhau.

a. Chọn ngẫu nhiên mười người, tính xác suất tất cả đều có những ngày sinh khác nhau. Tính xác suất có ít nhất hai người trong 10 người này có cùng ngày sinh.

b. Với k (người) thay thế 10 (người) ở câu (a). Tìm giá trị k nhỏ nhất trong đó có ít nhất 50-50 cơ hội mà hai hoặc nhiều người sẽ có cùng ngày sinh nhật.

c. Nếu 10 người được lựa chọn ngẫu nhiên, tính xác suất có ít nhất hai người có cùng ngày sinh.

TH 2.12 Một phương pháp được sử dụng để phân biệt giữa đá granit (G) và đá bazan (B) là kiểm tra tia hồng ngoại một phần của quang phổ năng lượng mặt trời phản xạ từ bề mặt đá. Cho R_1, R_2 và R_3 là cường độ phổ đo được ở ba bước sóng khác nhau; thông thường, đối với đá granite $R_1 < R_2 < R_3$, trong khi $R_3 < R_1 < R_2$ đối với bazan. Khi thực hiện phép đo từ xa (sử dụng máy bay), lệnh của Ris có thể nảy sinh khi đá này là bazan hoặc đá granit. Ta có các thông tin thu được như sau:

	Granite	Bazan
$R_1 < R_2 < R_3$	60%	10%
$R_1 < R_3 < R_2$	25%	20%
$R_3 < R_1 < R_2$	15%	70%

Giả sử chọn ngẫu nhiên một tảng đá trong một vùng với $P(\text{granite}) = 0,25$; $P(\text{bazan}) = 0,75$.

a. Chỉ ra rằng $P(\text{granite} | R_1 < R_2 < R_3) > P(\text{bazan} | R_1 < R_2 < R_3)$. Nếu các phép đo cho ra $R_1 < R_2 < R_3$, bạn sẽ phân loại đá là loại granit hay bazan?

b. Tương tự câu hỏi phần (a.) với các trường hợp: $R_1 < R_3 < R_2$ và $R_3 < R_1 < R_2$.

c. Dựa vào các phân loại trong câu (a.) và (b.) tính khả năng phân loại là sai. [Gợi ý: G có thể bị phân loại thành B và ngược lại.]

Chương 2

BIẾN NGẪU NHIÊN RỜI RẠC VÀ PHÂN PHỐI XÁC SUẤT

2.1 Biến ngẫu nhiên

Định nghĩa 2.1.1. Xét một phép thử bất kì với không gian mẫu Ω cho trước, một **biến ngẫu nhiên** là một quy luật kết hợp giữa một số và một kết quả của không gian mẫu. Theo ngôn ngữ toán học, một biến ngẫu nhiên là một hàm số có miền xác định là không gian mẫu và miền giá trị là tập số thực. Thông thường, ta hay kí hiệu biến ngẫu nhiên bằng chữ in hoa như X, Y, Z .

Ví dụ 2.1 Giả sử tung một đồng xu cân đối, đồng chất. Ta có không gian mẫu là $\Omega = \{H, T\}$. Khi đó, ta định nghĩa một biến ngẫu nhiên X bởi

$$X(H) = 1, X(T) = 0.$$

Định nghĩa 2.1.2. Bất kì một biến ngẫu nhiên nào chỉ nhận các giá trị 0 và 1 thì đều được gọi là **biến ngẫu nhiên Bernoulli**.

Ví dụ 2.2 Xét thí nghiệm kiểm tra 9 pin chỉ dừng lại khi một pin được chấp nhận. Ta có không gian mẫu là $\Omega = \{S, FS, FFS, \dots\}$. Khi đó, ta định nghĩa biến ngẫu nhiên X là số lần thí nghiệm cho đến khi một pin được chấp nhận.

$$X(S) = 1, X(FS) = 2, X(FFS) = 3, \dots, X(FFFFFFFFFS) = 9.$$

Định nghĩa 2.1.3. Một biến ngẫu nhiên **rời rạc** nếu tập giá trị của nó là tập hữu hạn hay *vô hạn đếm được*.

Một biến ngẫu nhiên gọi là **liên tục** nếu thỏa mãn hai điều kiện sau:

1. Tập giá trị của nó là tập hợp chứa tất cả các giá trị trên một khoảng hay là hợp rời nhau của các khoảng giá trị.
2. Xác suất của một điểm bất kì luôn là 0, ($P(X = c) = 0, \forall c \in \mathbb{R}$).

Bài tập: Biến ngẫu nhiên

2.1.1 Tại một thời điểm xác định, mỗi xe đi vào một giao lộ được quan sát xem liệu nó rẽ trái (L), phải (R) hay đi thẳng về phía trước (A). Thử nghiệm kết thúc ngay sau khi quan sát được xe rẽ trái. Gọi X là số xe quan sát. Các giá trị có thể có của X là bao nhiêu? Liệt kê năm giá trị của X và các kết quả tương ứng với các giá trị này.

2.1.2 Với mỗi biến ngẫu nhiên được định nghĩa ở đây, mô tả tập hợp các giá trị có thể của biến và nêu rõ biến có rời rạc hay không?

- a. X là số trứng vỡ trong một thùng trứng được lựa chọn ngẫu nhiên.
- b. Y là số sinh viên trong danh sách một lớp học xác định vắng mặt trong buổi học đầu tiên.
- c. Z là số lần lắc gậy đánh gôn trước khi đánh quả bóng gôn của một người chơi.
- d. U là chiều dài của một con rắn chuồng được chọn ngẫu nhiên.
- e. W là số tiền bản quyền kiếm được từ việc bán một ấn bản đầu tiên gồm 10000 cuốn sách giáo khoa.
- g. O là lực căng dây (psi) của một cây vợt tennis được chọn ngẫu nhiên.

2.1.3 Mỗi khi một thành phần được kiểm tra, kết quả kiểm tra là thành công (S) hoặc thất bại (F). Giả sử thành phần được kiểm tra liên tục cho đến khi được kết luận thành công xảy ra trên ba thử nghiệm liên tiếp. Gọi Y là số lượng các thử nghiệm cần thiết để đạt được điều này. Liệt kê tất cả các kết quả tương ứng với năm giá trị nhỏ nhất có thể của Y và nêu rõ giá trị Y tương ứng với mỗi giá trị đó.

2.1.4 Xác định số bơm đang hoạt động ở trạm 6 bơm và 4 bơm. Xác định các giá trị có thể có của các biến ngẫu nhiên sau.

- a. T là tổng số máy bơm đang hoạt động.
- b. X là sự chênh lệch giữa số máy bơm đang hoạt động ở trạm 1 và trạm 2.
- c. U là số bơm đang hoạt động lớn nhất ở mỗi trạm.
- d. Z là trạm bơm có đúng 2 máy bơm đang hoạt động.

2.2 Phân phối xác suất của biến ngẫu nhiên rời rạc

Định nghĩa 2.2.1. Phân phối xác suất hay hàm khối lượng xác suất của một biến ngẫu nhiên rời rạc X nhận giá trị x bất kì được xác định bởi

$$p(x) = P(X = x) = P(\forall s \in \Omega : X(s) = x).$$

Lưu ý: $p(x) \geq 0$ và $\sum_{\text{tất cả giá trị của } x} p(x) = 1$.

Ví dụ 2.3 Sáu kiện hàng được vận chuyển đến người tiêu dùng. Biết rằng trong các kiện hàng này có một số sản phẩm bị lỗi được cho trong bảng sau

Kiện hàng	1	2	3	4	5	6
Số sản phẩm lỗi	0	2	0	1	2	0

Gọi X là số sản phẩm lỗi khi kiểm tra một kiện hàng. X nhận ba giá trị là 0, 1, và 2. Khi đó, ta có

$$\begin{aligned} p(0) &= P(X = 0) = P(\text{kiện 1 hoặc kiện 3 hoặc kiện 6}) = \frac{3}{6} \\ p(1) &= P(X = 1) = P(\text{kiện 4}) = \frac{1}{6} \\ p(2) &= P(X = 2) = P(\text{kiện 2 hoặc kiện 5}) = \frac{2}{6} \end{aligned}$$

Ví dụ 2.4 Một nhóm gồm 5 người tham gia hiến máu nhân đạo là a, b, c, d, e trong đó chỉ 2 mẫu máu a và b thuộc nhóm máu O+. Giả sử người ta lấy lần lượt từng mẫu máu cho đến khi được mẫu nhóm O+ thì dừng lại. Gọi Y là số lần chọn mẫu

mẫu để chọn được nhóm O+. Khi đó, ta có hàm khối lượng xác suất của Y là

$$p(1) = P(Y = 1) = P(\text{chọn được mẫu a hoặc b}) = \frac{2}{5} = 0.4$$

$$p(2) = P(Y = 2) = P(\text{chọn được mẫu c, d hoặc e đầu, sau đó được a hoặc b}) \\ = \frac{3}{5} \cdot \frac{2}{4} = 0.3$$

$$p(3) = P(Y = 3) = P(\text{chọn được mẫu c, d hoặc e 2 lần đầu, sau đó được a hoặc b}) \\ = \frac{3}{5} \cdot \frac{2}{4} \cdot \frac{2}{3} = 0.2$$

$$p(4) = P(Y = 4) = P(\text{chọn được mẫu c, d hoặc e trong ba lần đầu, sau đó được a hoặc b}) \\ = \frac{3}{5} \cdot \frac{2}{4} \cdot \frac{1}{3} = 0.1$$

$$p(y) = 0 \text{ nếu } y \neq 1, 2, 3, 4.$$

Bảng phân phối xác suất của Y là

y	1	2	3	4
p(y)	0.4	0.3	0.2	0.1

Định nghĩa 2.2.2. Hàm phân phối tích lũy (cdf) $F(x)$ của một biến ngẫu nhiên rời rạc X với hàm khối lượng xác suất $p(x)$ được xác định như sau:

$$F(x) = P(X \leq x) = \sum_{y \leq x} p(y), \forall x \in \mathbb{R}. \quad (2.1)$$

Ví dụ 2.5 Một cửa hàng bán thẻ nhớ có các loại sau: 1 GB, 2 GB, 4 GB, 8 GB hoặc 16 GB. Gọi biến ngẫu nhiên Y là số lượng thẻ nhớ được bán được cho trong bảng phân phối xác suất sau:

y	1	2	4	8	16
p(y)	0.05	0.1	0.35	0.4	0.1

Khi đó, $F(y)$ được xác định như sau

$$F(1) = P(Y \leq 1) = P(Y = 1) = p(1) = 0.05$$

$$F(2) = P(Y \leq 2) = P(Y = 1, 2) = p(1) + p(2) = 0.15$$

$$F(3) = P(Y \leq 3) = P(Y = 1, 2, 4) = p(1) + p(2) + p(4) = 0.5$$

$$F(8) = P(Y \leq 8) = P(Y = 1, 2, 4, 8) = p(1) + p(2) + p(4) + p(8) = 0.9$$

$$F(16) = P(Y \leq 16) = 1.$$

Cho giá trị y bất kì, ta xác định giá trị của $F(y)$ bằng với giá trị của F tại giá trị gần nhất với biến ngẫu nhiên Y về phía bên trái của y . Chẳng hạn,

$$\begin{aligned} F(2.7) &= P(Y \leq 2.7) = P(Y \leq 2) = F(2) = 0.15 \\ F(7.999) &= P(Y \leq 7.999) = P(Y \leq 4) = F(4) = 0.5 \end{aligned}$$

Nếu y nhỏ hơn 1 thì $F(y) = 0$, và nếu y lớn hơn 16 thì $F(y) = 1$. Khi đó, cdf là

$$F(y) = \begin{cases} 0 & y < 1 \\ 0.05 & 1 \leq y < 2 \\ 0.15 & 2 \leq y < 4 \\ 0.5 & 4 \leq y < 8 \\ 0.9 & 8 \leq y < 16 \\ 1 & 16 \leq y \end{cases}$$

Mệnh đề 2.2.3. Cho hai số thực a và b bất kì với $a \leq b$

$$P(a \leq X \leq b) = F(b) - F(a-)$$

trong đó, $a-$ là giá trị lớn nhất có thể của X mà bé hẳn hơn a . Đặc biệt, nếu các giá trị của X là số nguyên và a và b cũng là số nguyên thì

$$P(a \leq X \leq b) = F(b) - F(a - 1)$$

Nếu $a = b$ thì $P(X = a) = F(a) - F(a - 1)$.

Ví dụ 2.6 Gọi X là số ngày nghỉ ốm trong năm của một nhân viên trong một công ty. Giả sử số ngày phép tối đa của một người là 14. Khi đó, X có thể nhận các giá trị $0, 1, \dots, 14$. Giả sử $F(0) = 0.58, F(1) = 0.72, F(2) = 0.76, F(3) = 0.81, F(4) = 0.88$, và $F(5) = 0.94$,

$$P(2 \leq X \leq 5) = F(5) - F(1) = 0.22$$

và

$$P(X = 3) = F(3) - F(2) = 0.05.$$

Phân phối xác suất của biến ngẫu nhiên rời rạc

2.2.1 Một cơ sở dịch vụ ô tô chuyên về điều chỉnh động cơ thấy rằng 45% của tất cả các điều chỉnh được thực hiện trên xe ô tô bốn xi-lanh, 40% trên xe sáu xi-lanh,

và 15% trên xe tám xi-lanh. Gọi X là số xi lanh trên xe tiếp theo được điều chỉnh.

a. Tìm hàm xác suất của X ?

b. Tính xác suất chiếc xe tiếp theo điều chỉnh có ít nhất sáu xi-lanh? Hơn sáu xi-lanh?

2.2.2 Các hãng hàng không đôi khi cho đặt vé nhiều hơn số chỗ trên một chuyến bay. Giả sử một chiếc máy bay với 50 chỗ ngồi, cho phép 55 hành khách đặt vé. Gọi Y là biến ngẫu nhiên đặc trưng cho số lượng hành khách đã mua vé thực sự xuất hiện trên chuyến bay. Hàm xác suất của Y xác định bởi bảng phân phối xác suất

y	45	46	47	48	49	50	51	52	53	54	55
$p(y)$	0,05	0,10	0,12	0,14	0,25	0,17	0,06	0,05	0,03	0,02	0,01

a. Tính xác suất chuyến bay sẽ đáp ứng được cho tất cả hành khách đã mua vé có thực hiện chuyến bay.

b. Xác suất không phải tất cả hành khách đã mua vé xuất hiện có thể được cung cấp chỗ trên chuyến bay là bao nhiêu?

c. Nếu bạn là người đầu tiên trong danh sách chờ (có nghĩa là bạn sẽ là người đầu tiên lên máy bay nếu có bất kỳ ghế nào sau khi tất cả hành khách đã mua vé đã có chỗ), xác suất mà bạn sẽ có thể lên chuyến bay là bao nhiêu? Xác suất này là bao nhiêu nếu bạn là người thứ ba trong danh sách chờ?

2.2.3 Một đường bưu điện đặt hàng của một doanh nghiệp máy tính có sáu đường dây điện thoại. Gọi X là số đường dây được sử dụng tại một thời điểm xác định. Giả sử hàm xác suất của X được xác định trong bảng phân phối xác suất sau

x	0	1	2	3	4	5	6
$p(x)$	0,10	0,15	0,20	0,25	0,20	0,06	0,04

Tính xác suất của các biến cố sau.

a. Nhiều nhất ba đường dây đang được sử dụng.

b. Ít hơn ba đường dây đang được sử dụng.

c. Ít nhất ba đường dây đang được sử dụng.

d. Số đường dây đang sử dụng ít nhất là hai và nhiều nhất là năm dòng.

e. Có nhiều nhất 5 đường dây đang được sử dụng biết rằng có ít nhất 2 đường dây đã được sử dụng.

g. Có ít nhất 2 đường dây sử dụng biết có không quá 5 đường dây được sử dụng.

2.2.4 Một văn phòng kế hoạch quận yêu cầu một nhà thầu nộp một, hai, ba, bốn, hoặc năm bộ hồ sơ (tùy thuộc vào bản chất của dự án) khi xin giấy phép xây dựng.

Gọi Y là số mẫu đơn được yêu cầu của người nộp đơn tiếp theo. Biết rằng xác suất số bộ hồ sơ được yêu cầu y là tỷ lệ thuận với y , nghĩa là $p(y) = ky$.

- a. Xác định giá trị của k .
- b. Tính xác suất nhà thầu này bị yêu cầu tối đa ba hồ sơ.
- c. Tính xác suất số hồ sơ bị yêu cầu là từ hai đến bốn bộ.

2.2.5 Một số vùng của California đặc biệt dễ bị động đất. Giả sử rằng trong một khu vực đô thị, 25% chủ nhà mua bảo hiểm về thiệt hại do động đất. Chọn ngẫu nhiên bốn chủ nhà trong khu vực; gọi X là số người trong 4 người này có bảo hiểm động đất.

- a. Tìm phân bố xác suất của X . [Gợi ý: Cho S biểu thị một chủ nhà có bảo hiểm và F không có. Một kết quả có thể có là SFSS, với xác suất $(0,25) (0,75) (0,25) (0,25)$ giá trị của tương ứng với trường này là 3. Ngoài ra còn có 15 kết quả khác.]
- b. X nhận giá trị nào với khả năng lớn nhất?
- c. Tính xác suất có ít nhất hai trong bốn người được chọn có bảo hiểm động đất.

2.2.6 Điện áp của pin mới có thể được chấp nhận (A) hoặc không được chấp nhận (U). Một đèn pin xác định cần có hai pin và các pin được lựa chọn độc lập và kiểm tra cho đến khi hai pin được chấp nhận. Giả sử rằng 90% của tất cả các pin có điện áp chấp nhận được. Gọi Y là số lượng pin phải được kiểm tra.

- a. Tính $p(2)$, tức là $P(Y=2)$.
- b. Tính $p(3)$. [Gợi ý: Có hai kết quả khác nhau dẫn đến $Y=2$.]
- c. Tính $p(5)$ và tìm một công thức chung cho $p(y)$.

2.2.7 Thư viện đăng ký hai tạp chí tin tức hàng tuần khác nhau, mỗi tạp chí được cho là sẽ đến vào thứ Tư. Trong thực tế, mỗi tạp chí có thể đến vào thứ Tư, thứ Năm, thứ Sáu hoặc thứ Bảy. Giả sử hai tạp chí này đến độc lập với nhau và cho xác suất mỗi loại tạp chí đến vào thứ Tư, Năm, Sáu và Bảy lần lượt là 0,3; 0,4; 0,2 và 0,1. Gọi Y là số ngày kể từ ngày thứ Tư mà cả hai tạp chí đến thư viện (giá trị Y có thể là 0, 1, 2 hoặc 3). Tìm hàm xác suất của Y . [Gợi ý: Có 16 kết quả có thể xảy ra; $Y(\text{thứ Tư, thứ Tư})=0$, $Y(\text{thứ Sáu, thứ Năm})=2$, v.v].

2.2.8 Ba cặp vợ chồng và hai cá nhân riêng lẻ được mời đến một hội thảo đầu tư và đã đồng ý tham dự. Giả sử xác suất bất kỳ cặp vợ chồng hoặc cá nhân cụ thể đến trễ là 0,4 (một cặp vợ chồng sẽ đi cùng nhau trong cùng một phương tiện, vì vậy cả hai vợ chồng sẽ cùng đến đúng giờ hoặc cả hai sẽ cùng đến trễ). Giả sử rằng các cặp vợ chồng và cá nhân khác nhau đến đúng giờ hoặc muộn độc lập với nhau.

Gọi X là số người đến trễ buổi hội thảo.

a. Xác định hàm xác suất của X .

b. Tìm hàm phân phối xác suất (cdf) của X và tính $P(2 \leq X \leq 6)$.

2.2.9 Một tổ chức tiêu dùng đánh giá xe ô tô mới thường báo cáo số lượng khuyết tật trong mỗi chiếc xe được kiểm tra. Gọi X là số lỗi chính trong một chiếc xe được chọn ngẫu nhiên của một loại nhất định. Hàm phân phối xác suất (cdf) của X là như sau:

$$F(x) = \begin{cases} 0 & x < 0 \\ 0,06 & 0 \leq x < 1 \\ 0,19 & 1 \leq x < 2 \\ 0,39 & 2 \leq x < 3 \\ 0,67 & 3 \leq x < 4 \\ 0,92 & 4 \leq x < 5 \\ 0,97 & 5 \leq x < 6 \\ 1 & 6 \leq x \end{cases}$$

Tính các xác suất sau:

a. $P(X = 2)$ b. $P(X > 3)$ c. $P(2 \leq X \leq 5)$ d. $P(2 < X < 5)$

2.2.10 Alvie Singer sống ở vị trí 0 trong sơ đồ đi kèm và có bốn người bạn sống ở A, B, C và D. Một ngày Alvie quyết định đến thăm bạn, vì vậy anh ta ném một đồng xu công bằng hai lần để quyết định xem ai trong số bốn người đến thăm. Khi ở nhà của một người bạn, anh ta sẽ trở về nhà hoặc đến một trong hai ngôi nhà liền kề (chẳng hạn như 0, A, hoặc C khi ở B) với mỗi khả năng có xác suất là $\frac{1}{4}$. Bằng cách này, Alvie tiếp tục thăm bạn bè cho đến khi anh trở về nhà. a. Gọi X là số lần Alvie tới thăm 1 người bạn. Xác định hàm xác suất của X .

b. Gọi Y là số đoạn thẳng mà Alvie đi qua bắt đầu từ nhà cho đến khi quay về nhà. Tìm hàm xác suất của Y .

c. Giả sử rằng những người bạn nữ sống ở A và C và bạn nam ở B và D. Gọi Z là số lần đến thăm những người bạn nữ, tìm hàm xác suất của Z .

2.3 Kỳ vọng và phương sai

Định nghĩa 2.3.1. Giả sử X là biến ngẫu nhiên rời rạc có tập giá trị D và hàm khối lượng xác suất $p(x)$. **Giá trị trung bình** hay còn gọi là **kỳ vọng** của biến

ngẫu nhiên X , kí hiệu $E(X)$ hoặc μ_X hoặc đơn giản là μ được xác định bởi

$$E(X) = \mu_X = \sum_{x \in D} x \cdot p(x)$$

Ví dụ 2.7 Từ ví dụ 3.5, ta tìm được kì vọng của biến ngẫu nhiên Y là

$$\mu_Y = 1 \cdot 0.05 + 2 \cdot 0.1 + 4 \cdot 0.35 + 8 \cdot 0.4 + 16 \cdot 0.1 = 6.45$$

Mệnh đề 2.3.2. Giả sử X là biến ngẫu nhiên rời rạc có tập giá trị D và hàm khối lượng xác suất $p(x)$. Khi đó, kỳ vọng của hàm $h(X)$, kí hiệu $E[h(X)]$ được cho bởi

$$E[h(X)] = \sum_{x \in D} h(x) \cdot p(x)$$

Ví dụ 2.8 Một cửa hàng vi tính nhập 3 laptop về với giá 500\$ mỗi máy. Họ sẽ bán mỗi máy này cho khách hàng giá 1000\$. Nhưng nếu không bán được thì nhà sản xuất sẽ mua lại mỗi máy với giá 200\$. Gọi X là số máy tính cửa hàng đã bán và giả sử rằng $p(0) = 0.1, p(1) = 0.2, p(2) = 0.3, p(3) = 0.4$. Gọi $h(X)$ là lợi nhuận thu được của cửa hàng. Khi đó, ta có

$$h(X) = \text{thu nhập} - \text{chi phí} = 1000X + (3 - X) \cdot 200 - 1500 = 800X - 900.$$

Vì thế, ta có kì vọng lợi nhuận thu được là

$$E[h(X)] = h(0) \cdot p(0) + h(1) \cdot p(1) + h(2) \cdot p(2) + h(3) \cdot p(3) = 700.$$

Nhân xét. Giả sử $h(X)$ là một hàm tuyến tính nghĩa là $h(X) = aX + b$. Khi đó, ta có

$$E[h(X)] = E(aX + b) = a \cdot E(X) + b$$

Định nghĩa 2.3.3. Cho X là biến ngẫu nhiên có hàm khối lượng xác suất $p(x)$. Khi đó, **phương sai** của X , kí hiệu $V(X)$ hoặc σ_X^2 , hoặc đơn giản là σ^2 được xác định bởi

$$V(X) = \sum_D (x - EX) \cdot p(x) = E[(X - EX)^2] = E(X^2) - (EX)^2$$

Độ lệch chuẩn của X có công thức là

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

Ví dụ 2.9 Một thư viện giới hạn số lần kiểm tra video cho một người là 6 lần trong cùng một thời điểm. Giả sử một người muốn kiểm tra video và gọi X là số lần kiểm tra. Ta có bảng phân phối xác suất của X

x	1	2	3	4	5	6
$p(x)$	0.3	0.25	0.15	0.05	0.1	0.15

Ta tìm được kì vọng là $EX = 2.85$. Khi đó, ta tính được phương sai của X là

$$\begin{aligned} V(X) &= \sigma^2 = \sum_{x=1}^6 (x - 2.85)^2 \cdot p(x) \\ &= (1 - 2.85)^2 \cdot 0.3 + \dots + (6 - 2.85)^2 \cdot 0.15 = 3.2275 \end{aligned}$$

Độ lệch chuẩn của X là $\sigma = \sqrt{3.2275} = 1.8$

Một cách tính phương sai khác:

$$E(X^2) = \sum_{x=1}^6 x^2 \cdot p(x) = 1^2 \cdot 0.3 + \dots + 6^2 \cdot 0.15 = 11.35.$$

Khi đó, $\sigma^2 = 11.35 - 2.85^2 = 3.2275$.

Mệnh đề 2.3.4.

$$V(aX + b) = a^2 \cdot \sigma_X^2 \quad \text{và} \quad \sigma_{aX+b} = |a| \cdot \sigma_X.$$

Kỳ vọng và phương sai

2.3.1 Chọn ngẫu nhiên một cá nhân có bảo hiểm ô tô từ một công ty nào đó. Gọi Y là số lần vi phạm cho mỗi cá nhân được nêu ra trong 3 năm qua

y	0	1	2	3
$p(y)$	0,60	0,25	0,10	0,05

a. Tính kì vọng, phương sai của Y .

b. Giả sử một cá nhân có vi phạm Y chịu một khoản phụ phí là $100Y^2$ đô. Tính số tiền phụ phí trung bình của mỗi cá nhân.

2.3.2 Một đại lý bán ba mẫu khác nhau của tủ đá thẳng đứng có thể tích không gian lưu trữ lần lượt là 13,5; 15,9 và 19,1 feet khối. Gọi X là thể tích không gian lưu trữ của tủ đông mà khách hàng mua tiếp theo. Giả sử X có hàm xác suất xác định bởi bảng phân phối xác suất.

x	13,5	15,9	19,1
p(x)	0,2	0,5	0,3

- a. Tính kì vọng, phương sai của X.
- b. Giả sử giá của một tủ đông có X feet khối là $25X - 8,5$; tính giá tiền trung bình của người khách tiếp theo mua tủ đông.
- c. Tính phương sai cho giá tiền $25X - 8,5$ của người khách tiếp theo.
- d. Giả sử rằng mặc dù công suất định mức của tủ đông là X, công suất thực tế là $h(X) = X - 0,01X^2$. Dung lượng thực tế trung bình của tủ đông được mua bởi khách hàng tiếp theo là bao nhiêu?

2.3.3 Giả sử rằng số lượng cây của một loại thực vật cụ thể được tìm thấy trong một khu vực lấy mẫu hình chữ nhật (được gọi là một quadrat bởi các nhà sinh thái học) trong một khu vực địa lý nhất định là một biến ngẫu nhiên X với hàm xác suất

$$p(x) = \begin{cases} c/x^3 & x = 1, 2, 3, \dots \\ 0 & \text{trường hợp khác} \end{cases}$$

Xác định c và giá trị EX.

2.3.4 Giả sử có n ứng viên cho một công việc được đánh thứ tự là 1, 2, 3, ..., n. Gọi X là số thứ tự của n ứng viên được chọn ngẫu nhiên. X có hàm xác suất

$$p(x) = \begin{cases} 1/n & x = 1, 2, 3, \dots, n \\ 0 & \text{trường hợp khác} \end{cases}$$

X được gọi có phân phối đều rời rạc. Tính kì vọng và phương sai của X. [Gợi ý: tổng của n số tự nhiên đầu tiên là $n(n+1)/2$, tổng của các bình phương của n số tự nhiên đầu tiên là $n(n+1)(2n+1)/6$.]

2.3.5 Gọi X là số chấm xuất hiện khi tung một con xúc sắc đồng chất. Nếu trước khi tung bạn được nhận $\frac{1}{3,5}$ đô la hoặc $h(X) = \frac{1}{X}$ đô la, bạn sẽ chọn nhận số tiền cố định hay nhận số tiền do may rủi?

2.4 Phân phối nhị thức

Định nghĩa 2.4.1. Một phép thử gọi là **phép thử nhị thức** nếu thỏa mãn 2 điều kiện sau đây:

1. Phép thử này gồm n phép thử nhỏ hơn và chúng độc lập với nhau.

2. Mỗi phép thử nhỏ hơn này chỉ có hai khả năng xảy ra, ta kí hiệu là S và F , và xác suất xảy ra mỗi biến cố S là như nhau và ta kí hiệu là p .

Ví dụ 2.10 Tung một đồng xu n lần liên tiếp nhau một cách độc lập. Ta kí hiệu H là biến cố xuất hiện mặt ngửa và T là biến cố xuất hiện mặt sấp. Khi đó, ta thấy phép thử này thỏa mãn 2 điều kiện trên nên nó là một phép thử nhị thức.

Định nghĩa 2.4.2. Biến ngẫu nhiên nhị thức X kết hợp với một phép thử nhị thức chứa n phép thử được xác định bởi

$$X = \text{số lượng của } S \text{ trong } n \text{ phép thử.}$$

Chẳng hạn, cho $n = 3$. Khi đó, ta có 8 kết quả có thể

$$SSS \quad SSF \quad SFS \quad SFF \quad FSS \quad FSF \quad FFS \quad FFF$$

Từ định nghĩa của X , ta có $X(SSF) = 2, X(SFF) = 1, \dots$. Giá trị của X trong n phép thử có thể nhận là $x = 0, 1, 2, \dots, n$. Chúng ta thường kí hiệu $X \sim \text{Bin}(n, p)$ để chỉ biến ngẫu nhiên X là biến ngẫu nhiên nhị thức thực hiện n phép thử với xác suất thành công p .

Kí hiệu.

Hàm khối lượng xác suất của một biến ngẫu nhiên nhị thức X phụ thuộc vào hai tham số n và p nên ta kí hiệu hàm khối lượng xác suất của X là $b(x; n, p)$.

Chẳng hạn, với $n = 4$ và ta tính xác suất của $SSFS$

$$\begin{aligned} P(SSFS) &= P(S) \cdot P(S) \cdot P(F) \cdot P(S) \quad (\text{các phép thử độc lập}) \\ &= p \cdot p \cdot (1 - p) \cdot p \quad (P(S) = p) \\ &= p^3 \cdot (1 - p) \end{aligned}$$

Định lý 2.4.3.

$$b(x; n, p) = \begin{cases} \binom{n}{x} p^x (1 - p)^{n-x} & x = 0, 1, 2, \dots, n \\ 0 & \text{nếu khác} \end{cases}$$

Ví dụ 2.11 Sáu người uống coca chọn ngẫu nhiên hai loại coca S và F . Gọi $p = P(\text{chọn coca } S) = 0.5$. Khi đó, ta có

$$X = \text{số người chọn coca } S, X \sim \text{Bin}(6, 0.5).$$

Chẳng hạn, ta tính

$$P(X = 3) = b(3; 6, 0.5) = \binom{6}{3} 0.5^3 0.5^3 = 20 \cdot 0.5^6 = 0.313$$

Xác suất ít nhất ba người chọn coca S

$$P(X \geq 3) = \sum_{x=3}^6 b(x; 6, 0.5) = \sum_{x=3}^6 \binom{6}{x} 0.5^x 0.5^{6-x} = 0.656$$

và xác suất nhiều nhất một người chọn S

$$P(X \leq 1) = \sum_{x=0}^1 b(x; 6, 0.5) = 0.109$$

Kí hiệu.

Cho $X \sim \text{Bin}(n, p)$, hàm phân phối tích lũy của X là

$$B(x; n, p) = P(X \leq x) = \sum_{y=0}^x b(y; n, p) \quad x = 0, 1, \dots, n$$

Ví dụ 2.12 Một nhà xuất bản thống kê rằng 20% số lượng bản in ấn bị mắc lỗi. Giả sử gọi X là số sách bị lỗi trong 15 sách chọn ngẫu nhiên. Khi đó, X có phân phối nhị thức với $n = 15$ và $p = 0.2$

1. Xác suất nhiều nhất 8 sách bị lỗi

$$P(X \leq 8) = \sum_{y=0}^8 b(y; 15, 0.2) = B(8; 15, 0.2) = .999$$

2. Xác suất có chính xác 8 sách bị lỗi

$$P(X = 8) = P(X \leq 8) - P(X \leq 7) = B(8; 15, 0.2) - B(7; 15, 0.2) = 0.003$$

3. Xác suất ít nhất 8 sách bị lỗi

$$P(X \geq 8) = 1 - P(X \leq 7) = 1 - B(7; 15, 0.2) = 1 - 0.996 = 0.004$$

4. Xác suất có từ 4 đến 7 sách bị lỗi

$$P(4 \leq X \leq 7) = P(X \leq 7) - P(X \leq 3) = 0.996 - 0.648 = 0.348$$

Mệnh đề 2.4.4. Kỳ vọng và phương sai của X .

Cho $X \sim \text{Bin}(n, p)$. Khi đó $EX = n \cdot p$, $V(X) = np(1 - p) = npq$ (với $q = 1 - p$).

Phân phối nhị thức

2.4.1 Khi các bo mạch được sử dụng trong quá trình sản xuất máy nghe đĩa compact được kiểm tra, tỷ lệ phần trăm bo mạch bị lỗi là 5%. Gọi X là số bo mạch bị lỗi trong một mẫu các bo mạch được chọn ngẫu nhiên có kích thước 25, vì vậy $X \sim \text{Bin}(25; 0,05)$. Tính các xác suất:

- a. $P(X \leq 2)$
- b. $P(X \geq 5)$
- c. $P(1 \leq X \leq 4)$
- d. Tính xác suất không có bo mạch nào trong 25 bo mạch này bị lỗi.
- e. Tính kỳ vọng và phương sai của X .

2.4.2 Một công ty sản xuất ly pha lê theo kinh nghiệm biết rằng 10% sản phẩm của họ có lỗi bị xếp xuống loại 2.

- a. Tính xác suất trong 6 ly pha lê được chọn ngẫu nhiên chỉ có 1 ly loại 2.
- b. Tính xác suất trong 6 ly pha lê được chọn ngẫu nhiên có ít nhất 2 ly loại 2.
- c. Nếu kiểm tra từng ly cho đến khi được 4 ly loại 2 thì xác suất phải kiểm tra nhiều nhất 5 ly là bao nhiêu?

2.4.3 Một số điện thoại cụ thể được sử dụng để nhận cả cuộc gọi thoại và tin nhắn fax. Giả sử rằng 25% các cuộc gọi đến bao gồm các tin nhắn fax. Xét một mẫu của 25 cuộc gọi đến. Tính xác suất

- a. Tối đa 6 cuộc gọi đến là tin nhắn fax?
- b. Chính xác 6 cuộc gọi đến là tin nhắn fax?
- c. Ít nhất 6 cuộc gọi đến là tin nhắn fax?
- d. Hơn 6 cuộc gọi đến là tin nhắn fax?
- e. Tính số cuộc gọi thông báo fax trung bình, độ lệch chuẩn trong số 25 cuộc gọi đến.
- g. Tính xác suất trong số 25 cuộc gọi số lượng có liên quan đến việc truyền fax lớn hơn số cuộc gọi thông báo fax trung bình 2 độ lệch chuẩn.

2.4.4 Giả sử rằng 30% tất cả sinh viên phải mua một văn bản cho một khóa học cụ thể muốn có một bản sao mới, Trong khi 70% khác muốn mua bản đã được sử dụng. Chọn ngẫu nhiên 25 người mua.

- a. Giá trị trung bình và độ lệch chuẩn của số người muốn có một bản sao mới của cuốn sách là bao nhiêu?

b. Tính xác suất số người muốn bản sao mới nhiều hơn hai độ lệch chuẩn so với giá trị trung bình.

c. Hiệu sách có 15 bản mới và 15 bản đã sử dụng. Nếu 25 người đến từng người một để mua văn bản này, xác suất mà tất cả 25 sẽ nhận được loại sách họ muốn từ số sách hiện tại là bao nhiêu? [Gợi ý: Gọi X là số người muốn có một bản sao mới. Giá trị nào của X thì tất cả 25 người sẽ nhận được những gì họ muốn?]

d. Giả sử các bản sao mới có giá 100 đô la và các bản sao đã sử dụng có giá 70 đô la. Giả sử hiệu sách hiện có 50 bản sao mới và 50 bản sao đã sử dụng. Giá trị kỳ vọng của tổng doanh thu từ việc bán 25 bản tiếp theo được mua là bao nhiêu?

2.4.5 Một loại vợt tennis đặc biệt có phiên bản cỡ trung và phiên bản quá khổ. Sáu mươi phần trăm của tất cả các khách hàng tại một cửa hàng nào đó muốn có phiên bản quá khổ.

a. Trong số mười khách hàng được lựa chọn ngẫu nhiên muốn loại vợt này, tính xác suất ít nhất sáu người muốn có phiên bản quá khổ.

b. Trong số mười khách hàng được chọn ngẫu nhiên, tính xác suất số lượng người muốn phiên bản quá khổ là nhỏ hơn 1 độ lệch chuẩn so với giá trị trung bình.

c. Cửa hàng hiện có bảy vợt cho mỗi phiên bản. Tính xác suất tất cả mười khách hàng tiếp theo muốn vợt này có thể nhận được phiên bản mà họ muốn từ số vợt hiện có.

2.4.6 Hai mươi phần trăm của tất cả các điện thoại của một loại nhất định được gửi cho dịch vụ bảo hành. Trong số này, 60% có thể được sửa chữa, trong khi 40% khác phải được thay thế bằng chiếc điện thoại mới. Nếu một công ty mua mười chiếc điện thoại loại này, thì xác suất mà chính xác hai chiếc sẽ bị thay thế theo bảo hành là bao nhiêu?

2.4.7 Giả sử rằng 90% các pin từ nhà cung cấp A có điện áp chấp nhận được. Một loại đèn pin nhất định yêu cầu hai loại pin D và đèn pin sẽ chỉ hoạt động nếu cả hai pin của nó có điện áp chấp nhận được. Trong số mười đèn pin được chọn ngẫu nhiên, xác suất mà ít nhất chín đèn sẽ hoạt động là bao nhiêu?

2.4.8 Mỗi lô hàng, lấy từ một số lượng lớn các sản phẩm của một nhà phân phối, được chấp nhận khi tỷ lệ các sản phẩm bị lỗi nhiều nhất là 0,10. Nhà phân phối quyết định chọn ngẫu nhiên 10 sản phẩm từ lô hàng và chỉ chấp nhận lô hàng đó nếu số sản phẩm bị lỗi trong mẫu tối đa là 2.

a. Xác suất mà lô hàng sẽ được chấp nhận khi tỷ lệ sản phẩm bị lỗi trong thực tế là 0,01 là bao nhiêu? 0,05? 0,10? 0,20? 0,25?

b. Tính xác suất nhà phân phối mua được ít nhất 8 lô hàng sau khi kiểm tra 10 lô hàng trong trường hợp tỷ lệ sản phẩm bị lỗi trong thực tế là 0,01.

2.4.9 Khách hàng tại trạm xăng thanh toán bằng thẻ tín dụng (A), thẻ ghi nợ (B) hoặc tiền mặt (C). Giả sử rằng khách hàng liên tiếp thực hiện các lựa chọn độc lập, với $P(A)=0,5$; $P(B)=0,2$ và $P(C)=0,3$.

a. Trong số 100 khách hàng tiếp theo, tính giá trị trung bình và phương sai của số khách thanh toán bằng thẻ ghi nợ.

b. Trong số 100 khách hàng tiếp theo, tính giá trị trung bình và phương sai của số khách không thanh toán bằng tiền mặt.

2.4.10 Xe limousine đưa rước ra sân bay có thể chứa tối đa 4 hành khách trên bất kỳ chuyến đi nào. Công ty sẽ chấp nhận tối đa sáu đặt chỗ cho một chuyến đi và hành khách phải đặt trước. Từ các dữ liệu trước đó, 20% tất cả những người đặt chỗ nhưng không thực hiện chuyến đi. Trả lời các câu hỏi sau đây với giả sử các khách đặt xe là độc lập.

a. Nếu có sáu khách đặt chỗ, xác suất có ít nhất một cá nhân có đặt chỗ nhưng không thể được cung cấp chỗ trong chuyến đi là bao nhiêu?

b. Nếu sáu khách đặt chỗ, số lượng khách trung bình sẽ sẵn đi khi xe limousine khởi hành là bao nhiêu?

c. Giả sử phân bố xác suất số khách đặt trước được đưa ra trong bảng kèm theo.

Số khách đặt chỗ	3	4	5	6
Xác suất	0,1	0,2	0,3	0,4

Gọi X là số khách đi trên một chuyến xe được chọn ngẫu nhiên. Tìm hàm xác suất của X .

Phân phối nhị thức âm và siêu bội

2.4.11 Một cửa hàng điện tử đã nhận được một lô 20 radio có kết nối cho iPod hoặc iPhone. Mười hai trong số này có hai khe cắm (để chúng có thể kết nối cả hai thiết bị) và tám cái còn lại có một khe cắm duy nhất. Giả sử rằng sáu trong số 20 radio được chọn ngẫu nhiên được đặt trên một kệ và những cái còn lại được đặt trong kho. Gọi X là số radio trong số các radio đặt trên kệ trưng bày có hai khe kết nối.

a. Xác định phân phối của X (tên và giá trị của tất cả các tham số)?

b. Tính $P(X = 2)$, $P(X \leq 2)$ và $P(X \geq 2)$.

c. Tính giá trị trung bình và độ lệch chuẩn của X .

2.4.12 Có 12 tủ lạnh thuộc một loại nhất định đã được trả lại cho một nhà phân phối vì tiếng ồn, hỏng khớp nối, rung khi tủ lạnh đang chạy. Giả sử rằng 7 trong số

các tủ lạnh này có máy nén bị lỗi và 5 máy còn lại có vấn đề ít nghiêm trọng hơn. Tủ lạnh được kiểm tra theo thứ tự ngẫu nhiên, gọi X là số tủ lạnh trong số 6 tủ kiểm tra đầu tiên có máy nén bị lỗi. Tính

- a. $P(X = 5); P(X \leq 4)$.
- b. Tính xác suất X có giá trị lớn hơn giá trị trung bình của nó 1 độ lệch chuẩn.
- c. Hãy xem xét một lô hàng lớn của 400 tủ lạnh, trong đó 40 máy nén có lỗi. Nếu X là số trong số 15 tủ lạnh được lựa chọn ngẫu nhiên có máy nén bị lỗi, tính $P(X \leq 8)$.

2.4.13 Một giảng viên đã dạy hai lớp học phần thống kê kỹ thuật trong học kỳ trước, lớp đầu tiên với 20 học sinh và lớp thứ hai với 30 học sinh. Trước khi chấm điểm giáo viên sắp thứ tự các bài kiểm tra cuối kỳ của sinh viên một cách ngẫu nhiên. Xét 15 bài kiểm tra được xếp hạng đầu tiên.

- a. Tính xác suất có đúng 10 bài kiểm tra trong số này của sinh viên lớp học phần thứ hai.
- b. Tính xác suất có ít nhất 10 trong số các bài này là từ lớp học phần thứ hai?
- c. Tính xác suất mà ít nhất 10 trong số này là từ cùng một lớp học phần.
- d. Tính giá trị trung bình và độ lệch chuẩn của số bài trong số 15 bài này là từ lớp học phần thứ hai.
- e. Tính giá trị trung bình và độ lệch chuẩn của số bài không nằm trong số 15 bài tiên này là từ lớp học phần thứ hai.

2.4.14 Một nhà địa chất đã thu thập 10 mẫu đá bazan và 10 mẫu đá granit. Nhà địa chất hướng dẫn một trợ lý phòng thí nghiệm chọn ngẫu nhiên 15 mẫu để phân tích.

- a. Xác định hàm xác suất của số lượng mẫu đá granite được chọn để phân tích.
- b. Tính xác suất tất cả các mẫu của một trong hai loại đá được chọn để phân tích.
- c. Tính xác suất số lượng mẫu đá granite được chọn để phân tích là lệch không quá 1 độ lệch chuẩn so với giá trị trung bình của nó?

2.4.15 Một giám đốc nhân sự phỏng vấn 11 kỹ sư cao cấp cho bốn việc làm đã lên kế hoạch sáu cuộc phỏng vấn cho ngày đầu tiên và năm cho ngày thứ hai của cuộc phỏng vấn. Giả sử các ứng cử viên được phỏng vấn theo thứ tự ngẫu nhiên.

- a. Tính xác suất x ($x = 0, 1, 2, 3, 4$) ứng cử viên trong số bốn ứng cử viên hàng đầu được phỏng vấn vào ngày đầu tiên.
- b. Tính kì vọng số ứng cử viên trong số bốn ứng cử viên hàng đầu sẽ được phỏng

vẫn vào ngày đầu tiên.

2.4.16 Giả sử rằng $p = P(\text{sinh con trai}) = 0,5$. Một cặp vợ chồng mong muốn có hai con gái. Họ sẽ sinh con cho đến khi có được hai người con gái.

- Xác suất mà gia đình có x người con trai là bao nhiêu?
- Xác suất mà gia đình có bốn người con là bao nhiêu?
- Xác suất mà gia đình có tối đa là bốn người con là bao nhiêu?
- Tính kì vọng số con trai của gia đình này. Tính kì vọng số con của gia đình này.

2.4.17 Một gia đình quyết định sinh con cho đến khi có được 3 người con cùng giới tính thì dừng. Giả sử $P(\text{sinh con trai}) = P(\text{sinh con gái}) = 0,5$. Gọi X là số con của gia đình này. Tính EX .

2.4.18 Ba anh em và vợ của họ quyết định có con cho đến khi mỗi gia đình có hai con gái. Gọi X là tổng số con gái của ba gia đình này. Xác định hàm xác suất và kì vọng của X và so sánh EX với số con gái trung bình mỗi gia đình.

2.5 Phân phối siêu bội và phân phối nhị thức âm

PHÂN PHỐI SIÊU BỘI

Một biến ngẫu nhiên có phân phối siêu bội nếu thỏa mãn các điều kiện sau:

- Tập hợp mẫu chứa N phần tử.
- Mỗi phần tử có hai khả năng xảy ra là xảy ra (S) hoặc không xảy ra (F) và có M lần xảy ra trong tập hợp mẫu.
- Một mẫu có n phần tử được chọn một cách không lặp lại sao cho mỗi lần chọn tập n phần tử này đều có khả năng như nhau.

Xét biến ngẫu nhiên X là số lần xảy ra (S) trong mẫu có kích thước n . Phân phối xác suất của X phụ thuộc vào tham số n , M , và N . Vì thế, ta có $P(X = x) = h(x; n, M, N)$.

Ví dụ 2.13 Một trung tâm bảo hành điện máy nhận được 20 đơn hàng trong đó có 8 yêu cầu sửa chữa máy in và 12 yêu cầu thay hộp mực. Sau khi sửa xong, trung tâm chọn ngẫu nhiên 5 đơn hàng và kiểm tra sự hài lòng của khách hàng về dịch vụ của mình. Tính xác suất có x ($x = 0, 1, 2, 3, 4, 5$) đơn hàng yêu cầu thay hộp mực.

Ta có, tập hợp mẫu là $N = 20$, mẫu kiểm tra là $n = 5$ và số lần xảy ra với S (thay mực) là $M = 12$ và $N - M = 8$. Khi đó, chẳng hạn cho $x = 2$, ta có

$$P(X = x) = h(x; 5, 12, 20) = \frac{\text{số kết quả xảy ra khi } X=2}{\text{số kết quả có thể xảy ra}} = \frac{\binom{12}{2} \binom{8}{3}}{\binom{20}{5}}$$

Mệnh đề 2.5.1. Nếu X là số lần xảy ra (S) trong mẫu ngẫu nhiên có kích thước n được rút từ một tập hợp cho trước có M (S) và $(N - M)$ (F). Khi đó, hàm phân phối xác suất của X gọi là **phân phối siêu bội** được cho bởi

$$P(X = x) = h(x; n, M, N) = \frac{\binom{M}{x} \binom{N - M}{n - x}}{\binom{N}{n}} \quad (2.2)$$

với x là số nguyên thỏa mãn $\max(0, n - N + M) \leq x \leq \min(n, M)$.

Mệnh đề 2.5.2. Kỳ vọng và phương sai của biến ngẫu nhiên X có phân phối siêu bội

$$E(X) = n \cdot \frac{M}{N} \quad V(X) = \frac{N - n}{N - 1} \cdot n \cdot \frac{M}{N} \cdot \left(1 - \frac{M}{N}\right)$$

Ví dụ 2.14 Năm con hổ được bắt trong một vùng được đeo định vị và sau đó thả lại tự nhiên. Sau đó, người ta bắt ngẫu nhiên 10 con hổ trong vùng đó kiểm tra xem có bao nhiêu con được định vị. Gọi X là số hổ được gắn định vị trong 10 con hổ. Giả sử rằng trong vùng này chỉ còn 25 con hổ. Tính

a) $X = 2$ b) $X \leq 2$ c) $EX, V(X)$?

Ta có: $n = 10, M = 5, N = 25$. Do đó,

$$P(X = x) = h(x; 10, 5, 25) = \frac{\binom{5}{x} \binom{20}{10 - x}}{\binom{25}{10}}, \quad x = 0, 1, 2, 3, 4, 5$$

$$\text{a) } P(X = 2) = h(2; 10, 5, 25) = \frac{\binom{5}{2} \binom{20}{8}}{\binom{25}{10}} = 0.385$$

$$b) P(X \leq 2) = P(X = 0, 1, 2) = \sum_{x=0}^2 h(x; 10, 5, 25) = 0.699$$

$$c) EX = 10 \cdot \frac{5}{25} = 2, \quad V(X) = \frac{15}{24} \cdot 10 \cdot \frac{5}{25} \left(1 - \frac{5}{25}\right) = 1$$

PHÂN PHỐI NHỊ THỨC ÂM

Một biến ngẫu nhiên có phân phối nhị thức âm nếu thỏa mãn các điều kiện sau:

1. Thí nghiệm là dãy các phép thử độc lập.
2. Mỗi phép thử chỉ chứa 2 khả năng là thành công (S) hoặc thất bại (F).
3. Khả năng xảy ra S trong mọi phép thử là như nhau và bằng p .
4. Thí nghiệm tiến hành các phép thử sao cho đạt được r lần S, với r là số dương cụ thể.

Giá trị của biến ngẫu nhiên X là số lần F xảy ra trước khi đạt tới lần thứ r thành công (S). X gọi là **biến ngẫu nhiên nhị thức âm** vì số lần thành công là cố định trong khi số lần thực hiện phép thử là ngẫu nhiên.

Mệnh đề 2.5.3. Hàm khối lượng xác suất của biến ngẫu nhiên nhị thức âm X với tham số $r =$ số lần thành công (S) và $p = P(S)$ là

$$nb(x; r, p) = \binom{x+r-1}{r-1} p^r (1-p)^x \quad x = 0, 1, 2, \dots$$

Ví dụ 2.15 Một bác sĩ nhi khoa muốn tìm 5 cặp vợ chồng tham gia chương trình ăn uống mới cho đứa con đầu tiên của những cặp vợ chồng này. Gọi $p = P(\text{chọn ngẫu nhiên được cặp vợ chồng đồng ý tham gia})$. Cho $p = 0.2$, hãy tìm xác suất để 15 cặp vợ chồng được hỏi tham gia chương trình biết rằng sẽ có 5 cặp đồng ý? Nghĩa là với $S = \{\text{đồng ý tham gia}\}$, tính xác suất có 10 cặp từ chối trước khi có được cặp thứ 5 đồng ý tham gia.

Ta có: $r = 5, p = 0.2, x = 10$. Khi đó

$$nb(10; 5, 0.2) = \binom{14}{4} 0.2^5 0.8^{10} = 0.034$$

Xác suất để nhiều nhất có 10 cặp từ chối là

$$P(X \leq 10) = \sum_{x=0}^{10} nb(x; 5, 0.2) = 0.2^5 \sum_{x=0}^{10} \binom{x+4}{4} 0.8^x = 0.164$$

Mệnh đề 2.5.4. Cho X là biến ngẫu nhiên nhị thức âm với hàm khối lượng xác suất $nb(x; r, p)$. Khi đó, ta có

$$EX = \frac{r(1-p)}{p} \quad V(X) = \frac{r(1-p)}{p^2}$$

2.6 Phân phối Poisson

Định nghĩa 2.6.1. Một biến ngẫu nhiên rời rạc X gọi là có **phân phối Poisson** với tham số $\mu (\mu > 0)$ nếu hàm khối lượng xác suất của X có dạng

$$p(x; \mu) = \frac{e^{-\mu} \cdot \mu^x}{x!} \quad x = 0, 1, 2, 3 \dots$$

Ví dụ 2.16 Gọi X là số lượng chuột bị bắt bởi một cái bẫy trong một khoảng thời gian nhất định. Giả sử X có phân phối Poisson với $\mu = 4.5$. Khi đó, xác suất để chính xác 5 con chuột dính bẫy là

$$P(X = 5) = \frac{e^{-4.5} \cdot 4.5^5}{5!} = 0.1708$$

Xác suất mà nhiều nhất 5 con chuột dính bẫy là

$$P(X \leq 5) = \sum_{x=0}^5 \frac{e^{-4.5} \cdot 4.5^x}{x!} = e^{-4.5} \left[1 + 4.5 + \frac{4.5^2}{2!} + \dots + \frac{4.5^5}{5!} \right] = 0.7029$$

Phân phối Poisson như là một giới hạn

Mệnh đề 2.6.2. Giả sử một biến ngẫu nhiên X có hàm khối lượng xác suất nhị thức là $b(x; n, p)$. Ta cho $n \rightarrow \infty$ và $p \rightarrow 0$ sao cho $np \rightarrow \mu$ với $\mu > 0$. Khi đó, $b(x; n, p) \rightarrow p(x, \mu)$.

Dựa vào mệnh đề này, ta chú ý rằng bất kì phép thử nhị thức có n đủ lớn và p đủ nhỏ thì $b(x; n, p) \simeq p(x, \mu)$ với $\mu = np$. Ở đây, ta có thể áp dụng quy luật này khi $n > 50$ và $np < 5$.

Ví dụ 2.17 Một nhà xuất bản chuẩn bị phát hành một quyển sách tiểu thuyết với yêu cầu sai sót lỗi đánh máy trong mỗi trang sách có ít nhất một lỗi với xác suất là 0.005 và sai sót mỗi trang là độc lập nhau. Giả sử quyển sách này có 400 trang. Tính xác suất có chỉ có chính xác một trang có lỗi? nhiều nhất ba trang có lỗi?

Giả sử S là một trang chứa ít nhất một lỗi và F là một trang lỗi. Khi đó, X là số lượng trang chứa ít nhất một lỗi và X cũng là biến ngẫu nhiên nhị thức với $n = 400, p = 0.005$, vì thế $np = 2$. Khi đó,

$$P(X = 1) = b(1; 400, 0.005) \simeq p(1; 2) = \frac{e^{-2}2^1}{1!} = 0.270671$$

Giá trị X nếu tính theo phân phối nhị thức là $b(1; 400, 0.005) = 0.270669$, vì thế xấp xỉ này rất tốt. Tương tự,

$$\begin{aligned} P(X \leq 3) &\simeq \sum_{x=0}^3 p(x, 2) = \sum_{x=0}^3 \frac{e^{-2}2^x}{x!} \\ &= 0.135335 + 0.270671 + 0.270671 + 0.180447 = 0.8571 \end{aligned}$$

và gần bằng với giá trị nhị thức là $P(X \leq 3) = 0.8576$.

Mệnh đề 2.6.3. Nếu X có phân phối Poisson với tham số μ thì $EX = V(X) = \mu$.

Phân phối Poisson

2.6.1 Gọi X là số lỗi trên bề mặt của một lò hơi được lựa chọn ngẫu nhiên thuộc một loại xác định, X có phân phối Poisson với tham số $\mu = 5$. Tính các xác suất sau:

$$P(X \leq 8); \quad P(X = 8); \quad P(9 \leq X); \quad P(5 \leq X \leq 8); \quad P(5 < X < 8)$$

2.6.2 Giả sử rằng số lượng tài xế di chuyển giữa một điểm xuất phát cụ thể và đích đến trong một khoảng thời gian được chỉ định có phân phối Poisson với tham số $\mu = 20$. Tính xác suất số lượng tài xế:

- Nhiều nhất là 10.
- Vượt quá 20.
- Thuộc khoảng từ 10 đến 20.
- Trong khoảng 2 độ lệch chuẩn so với giá trị trung bình.

2.6.3 Một bài báo trên tờ Los Angeles Times (ngày 3 tháng 12 năm 1993) báo cáo rằng 1 trong 200 người mang gen khiếm khuyết gây ung thư ruột kết di truyền. Gọi X là số người mang gen này trong một mẫu gồm 1000 người, Xác định phân phối gần đúng của X. Sử dụng phân phối này để tính gần đúng xác suất

- Trong mẫu có từ 5 đến 8 người mang gen này.
- Trong mẫu có ít nhất 8 người mang gen này.

2.6.4 Giả sử rằng chỉ có 10% các máy tính thuộc loại nào đó có CPU gặp lỗi trong thời gian bảo hành. Hãy xem xét một mẫu gồm 10.000 máy tính.

a. Giá trị kỳ vọng và độ lệch chuẩn của số lượng máy tính trong mẫu có khiếm khuyết là bao nhiêu?

b. Xác suất (gần đúng) hơn 10 máy tính được lấy mẫu có lỗi là bao nhiêu?

c. Xác suất (gần đúng) không có máy tính được lấy mẫu là có lỗi là bao nhiêu?

2.6.5 Giả sử số lượng máy bay nhỏ đến một sân bay nhất định tuân theo phân phối Poisson với tốc độ $\alpha = 8$ mỗi giờ. Khi đó số lượng máy bay nhỏ đến trong một khoảng thời gian t là một biến ngẫu nhiên có phân phối Poisson với tham số $\mu = 8t$.

a. Tính xác suất có chính xác 6 chiếc máy bay nhỏ đến trong khoảng thời gian 1 giờ; Ít nhất 6? Ít nhất 10?

b. Tính giá trị kỳ vọng và độ lệch chuẩn của số lượng máy bay nhỏ đến trong khoảng thời gian 90 phút.

c. Tính xác suất có ít nhất 20 máy bay nhỏ đến trong khoảng thời gian 2,5 giờ. Tính xác suất có tối đa 40 máy bay nhỏ bay đến trong thời gian này?

2.6.6 Số người đến để điều trị tại phòng cấp cứu có thể được lập mô hình bằng phân phối Poisson với tham số tốc độ tới điều trị là 5 lần mỗi giờ.

a. Tính xác suất có chính xác bốn lượt đến điều trị một giờ cụ thể.

b. Tính xác suất có ít nhất bốn người đến điều trị trong một giờ cụ thể.

c. Bạn kỳ vọng có bao nhiêu người đến điều trị trong thời gian 45 phút?

2.6.7 Khi thử nghiệm một loại bảng mạch điện cụ thể, xác suất đoản bất kỳ sẽ bị hỏng là 0,01. Giả sử một bảng mạch có 200 đoản.

a. Dự kiến có bao nhiêu đoản thất bại và độ lệch chuẩn của số đoản sẽ thất bại là bao nhiêu?

b. Xác suất (gần đúng) mà ít nhất 10 đoản sẽ thất bại trên một bảng mạch điện được chọn ngẫu nhiên là bao nhiêu?

c. Nếu năm bảng mạch được chuyển đến một khách hàng cụ thể, thì khả năng ít nhất bốn bảng sẽ hoạt động đúng cách là bao nhiêu? (Một bảng hoạt động đúng chỉ khi tất cả các đoản của nó hoạt động.)

2.6.8 Giả sử rằng số cây được phân bố trong một khu rừng tuân theo phân phối Poisson với tham số α , số cây trung bình trên một mẫu Anh, bằng 80.

a. Tính xác suất trong một lô đất mẫu Anh nhất định, sẽ có tối đa 16 cây.

b. Nếu rừng bao phủ có diện tích 85.000 mẫu Anh, số cây dự kiến trong rừng là bao nhiêu?

c. Giả sử bạn chọn một điểm trong rừng và xây dựng một vòng tròn bán kính 0,1 dặm. Gọi X là số cây trong vùng tròn đó. Tìm hàm xác suất của X . [Biết 1 dặm vuông = 640 mẫu.]

2.6.9 Giả sử số xe ô tô đến trạm kiểm tra thiết bị phương tiện tuân theo phân phối Poisson với tốc độ $\alpha = 10$ mỗi giờ. Giả sử rằng một chiếc xe đến sẽ không có vi phạm thiết bị với xác suất 0,5.

a. Tính xác suất có đúng mười xe đến trong giờ và tất cả mười xe không có vi phạm.

b. Đối với bất kỳ $y \geq 10$ cố định, xác suất có y xe đến trong một giờ trong đó có 10 xe không vi phạm là bao nhiêu?

c. Tính xác suất mà mười chiếc xe "không vi phạm" đến trong giờ tiếp theo. [Gợi ý: Tổng các xác suất trong phần (b) từ $y=10$ đến ∞ ?.]

Bài tập Tổng hợp

TH 3.1 Một hộp gồm bảy thẻ được đánh dấu 1, 2, ..., 7. Chọn ngẫu nhiên 3 trong số các thẻ này. Biến ngẫu nhiên W là tổng các số trên 3 tấm thẻ lấy ra. Tìm hàm xác suất của W từ đó tính trung bình và phương sai của W . [Gợi ý: Xem các kết quả là không có thứ tự, do đó (1, 3, 7) và (3, 1, 7) không phải là kết quả khác nhau.]

TH 3.2 Trong số các khách hàng mua dụng cụ mở cửa nhà để xe tự động, 75% mua mô hình điều khiển bằng dây xích. Gọi X là số khách hàng mua mô hình điều khiển bằng dây xích này trong số 15 người mua tiếp theo.

- Tìm hàm xác suất của X .
- $P(X > 10)$; $P(6 \leq X \leq 10)$
- Tính kì vọng và phương sai của X .

d. Nếu cửa hàng hiện có trong kho 10 mô hình điều khiển bằng dây xích và 8 mô hình điều khiển bằng trực, xác suất mà các yêu cầu của 15 khách hàng này có thể được đáp ứng từ số mô hình hiện tại là bao nhiêu?

TH 3.3 Một người gần đây đã lên kế hoạch cho một chuyến đi cắm trại. Anh ta có hai đèn pin, một cái cần một pin 6-V và một cái khác sử dụng hai pin cỡ D. Trước đây anh đã đóng gói hai pin 6-V và bốn pin kích thước D trong chiếc xe cắm trại của mình. Giả sử xác suất mà bất kỳ pin cụ thể nào hoạt động là p và pin hoạt động hoặc không hoạt động là độc lập với nhau. Người này chỉ muốn lấy một đèn pin. Đối với những giá trị nào của p , anh này nên dùng đèn pin 6-V.

TH 3.4 Hệ thống k-out-of-n là hệ thống sẽ hoạt động khi và chỉ khi ít nhất k của n thành phần riêng lẻ trong hệ thống hoạt động. Nếu các thành phần riêng lẻ hoạt động độc lập với nhau, mỗi thành phần có xác suất hoạt động là 0,9; xác suất hệ thống 3-out-of-5 hoạt động là bao nhiêu?

TH 3.5 Một nhà sản xuất chip mạch tích hợp muốn kiểm soát chất lượng sản phẩm của mình bằng cách từ chối bất kỳ lô nào trong đó tỷ lệ chip lỗi quá cao. Do đó trong mỗi lô (10.000 chip), 25 chip sẽ được chọn và kiểm tra. Nếu ít nhất 5 trong số 25 bị lỗi, toàn bộ lô sẽ bị từ chối.

- Tính xác suất một lô sẽ bị từ chối nếu 5% số lượng chip trong lô thực tế bị

lỗi?

b. Trả lời câu hỏi được đặt ra trong (a) nếu tỷ lệ phần trăm của các chip bị lỗi trong lô là 10%.

c. Trả lời câu hỏi được đặt ra trong (a) nếu tỷ lệ phần trăm của các chip bị lỗi trong lô là 20%.

d. Điều gì xảy ra với xác suất trong (a) - (c) nếu lô hàng sẽ bị từ chối khi số chip bị lỗi tăng từ 5 lên 6?

TH 3.6 Trong số những người đi qua máy dò kim loại sân bay có 0,5% kích hoạt nó. Gọi X là số người kích hoạt máy dò trong số 5000 người được chọn ngẫu nhiên đi qua máy dò.

a. Xác định hàm xác suất (xấp xỉ) của X .

b. Tính $P(X = 5)$; $P(5 \leq X)$.

TH 3.7 Một công ty tư vấn giáo dục đang cố gắng quyết định xem học sinh trung học chưa bao giờ sử dụng máy tính cầm tay có thể giải quyết một loại vấn đề nào đó dễ dàng hơn bằng máy tính sử dụng logic Ba Lan ngược hay không sử dụng logic này. Chọn ngẫu nhiên một mẫu gồm 25 học sinh chưa từng sử dụng máy tính cầm tay và các học sinh này thực hành trên cả hai máy tính. Sau đó, mỗi học sinh được yêu cầu giải quyết một vấn đề trên máy tính Ba Lan ngược lại và một vấn đề tương tự trên máy kia. Đặt $p = P(S)$ là xác suất một học sinh xử lý vấn đề nhanh hơn bằng cách sử dụng logic Ba Lan ngược hơn là không sử dụng logic này và gọi X là số người sử dụng logic ngược nhanh hơn không sử dụng logic này trong nhóm các học sinh chọn ra.

a. Nếu $p = 0,5$ thì $P(7 \leq X \leq 18)$ là bao nhiêu?

b. Nếu $p = 0,8$ thì $P(7 \leq X \leq 18)$ là bao nhiêu?

TH 3.8 Xét một căn bệnh có thể được xác định bằng cách thực hiện một xét nghiệm máu. Gọi p là xác suất mà một cá nhân được chọn ngẫu nhiên có bệnh. Giả sử n cá nhân được lựa chọn độc lập để thử nghiệm. Một cách để tiến hành là thực hiện các xét nghiệm riêng biệt trên mỗi mẫu máu. Một cách khác có khả năng kinh tế hơn là xét nghiệm nhóm, đã được giới thiệu trong Thế chiến II để xác định những người đàn ông mắc bệnh giang mai trong số những người được chỉ định quân đội. Đầu tiên, lấy một phần của từng mẫu máu, kết hợp các mẫu thử này và thực hiện

chỉ một xét nghiệm. Nếu không có ai mắc bệnh, kết quả sẽ là âm tính và chỉ cần một xét nghiệm là đủ kết luận cho cả nhóm. Nếu có ít nhất một cá nhân bị bệnh, xét nghiệm trên mẫu kết hợp sẽ mang lại kết quả dương tính, khi đó thực hiện các thử nghiệm riêng lẻ với từng mẫu thành phần. Nếu $p=0,1$ và $n=3$, số xét nghiệm trung bình khi sử dụng quy trình này là bao nhiêu? Số lượng dự kiến khi $n=5$ là?

TH 3.9 Đặt p_1 là xác suất mà bất kỳ mã ký hiệu cụ thể nào được truyền sai thông qua một hệ thống truyền thông. Giả sử rằng trên các biểu tượng khác nhau, lỗi xảy ra độc lập với nhau. Giả sử rằng p_2 là xác suất một biểu tượng sai lầm sẽ được sửa khi nhận được. Gọi X là số lượng các ký hiệu chính xác trong một khối tin bao gồm n ký hiệu (sau khi quá trình hiệu chỉnh đã kết thúc). Xác định phân bố xác suất của X .

TH 3.10 Một phiên tòa có một bồi thẩm đoàn với tám thành viên của bồi thẩm đoàn đã ủng hộ một phán quyết có tội và bốn người khác đã tha bổng. Nếu các bồi thẩm rời khỏi phòng bồi thẩm theo thứ tự ngẫu nhiên và mỗi người trong số bốn người đầu tiên rời khỏi phòng được một phóng viên phỏng vấn. Gọi X là số bồi thẩm đoàn ủng hộ phán quyết có tội trong số 4 người được hỏi. Tìm hàm xác suất của X và tính EX .

TH 3.11 Dịch vụ đặt hàng trước sử dụng năm nhân viên tiếp nhận thông tin, những người nhận thông tin là độc lập với nhau, việc nhận thông tin của mỗi người theo phân phối Poisson với tỷ lệ $\alpha = 2$ mỗi phút.

a. Tính xác suất trong khoảng thời gian 1 phút nhất định, người đầu tiên sẽ không nhận được yêu cầu nào?

b. Tính xác suất trong khoảng thời gian 1 phút nhất định, có đúng bốn trong số năm nhân viên không nhận được yêu cầu nào?

c. Viết một biểu thức tính xác suất trong khoảng thời gian 1 phút nhất định, tất cả các nhân viên đều nhận được số yêu cầu như nhau.

Chương 3

BIẾN NGẪU NHIÊN LIÊN TỤC VÀ PHÂN PHỐI XÁC SUẤT

3.1 Hàm mật độ xác suất

Nhắc lại: Một biến ngẫu nhiên X gọi là liên tục nếu thỏa mãn:

- (1) các giá trị có thể có của X nằm trong một khoảng số thực hoặc là hợp của các khoảng rời nhau;
- (2) $P(X = c) = 0$ với bất kì số c là một giá trị của X .

Ví dụ 3.1

- (1) Đo độ sâu của một hồ nước tại một vị trí chọn ngẫu nhiên. Khi đó, gọi $X =$ là độ sâu tại một vị trí thì X là biến ngẫu nhiên liên tục. Khoảng giá trị của X nằm trong vị trí thấp nhất và sâu nhất của hồ nước.
- (2) Một hợp chất hóa học được chọn ngẫu nhiên và ta đo mức pH X của nó. Khi đó, X là một biến ngẫu nhiên liên tục vì bất kì mức pH nào cũng nằm trong khoảng 0 và 14.

Định nghĩa 3.1.1. Cho X là một biến ngẫu nhiên liên tục. Khi đó, hàm mật độ xác suất của X là hàm số $f(x)$ sao cho với bất kì hai giá trị a, b nào và $a \leq b$ thì

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

Hơn thế nữa, $f(x)$ phải thỏa mãn hai điều kiện sau

1. $f(x) \geq 0$ với mọi x .

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

Ví dụ 3.2 Cho X là biến ngẫu nhiên liên tục có hàm mật độ xác suất là

$$f(x) = \begin{cases} \frac{1}{360} & 0 \leq x < 360, \\ 0 & \text{nếu khác.} \end{cases}$$

Rõ ràng, $f(x) \geq 0$ và $\int_{-\infty}^{\infty} f(x)dx = \int_0^{360} f(x)dx = 1$. Chẳng hạn, ta tính

$$P(90 \leq X \leq 180) = \int_{90}^{180} \frac{1}{360} dx = 0.25$$

Định nghĩa 3.1.2. Một biến ngẫu nhiên liên tục X gọi là một **phân phối đều** trên một khoảng $[A, B]$ nếu hàm mật độ của X là

$$f(x; A, B) = \begin{cases} \frac{1}{B - A} & A \leq x \leq B \\ 0 & \text{nếu khác.} \end{cases}$$

Bài tập: Hàm mật độ xác suất

3.1.1 Cho X là biến ngẫu nhiên có hàm mật độ

$$f(x) = \begin{cases} 0,075x + 0,2 & 3 \leq x \leq 5 \\ 0 & \text{nếu khác} \end{cases}$$

a. Vẽ đồ thị và kiểm tra tổng diện tích miền phẳng nằm dưới đồ thị hàm mật độ và trên trục hoành có bằng 1 không.

b. Tính $P(X \leq 4)$ và so sánh với $P(X < 4)$.

c. Tính $P(3,5 \leq X \leq 4,5)$ và $P(4,5 < X)$.

3.1.2 Cho X là biến ngẫu nhiên có hàm mật độ

$$f(x) = \begin{cases} 0,09375(4 - x^2) & -2 \leq x \leq 2 \\ 0 & \text{nếu khác} \end{cases}$$

Tính các xác suất sau

a. $P(X < 0)$ **b.** $P(-1 < X < 1)$. **d.** $P(X < -0,5 \text{ hoặc } X > 0,5)$.

3.1.3 Một giáo sư đại học không bao giờ kết thúc bài giảng đúng giờ và luôn kết thúc bài trong vòng 2 phút sau giờ học. Gọi X là số giờ trôi qua sau hết giờ đến khi giáo sư kết thúc bài giảng, giả sử hàm mật độ của X có dạng

$$f(x) = \begin{cases} kx^2 & 0 \leq x \leq 2 \\ 0 & \text{nếu khác} \end{cases}$$

- a. Tìm k .
- b. Tính xác suất giáo sư kết thúc bài sau hết giờ trong vòng 1 phút.
- c. Tính xác suất giáo sư kết thúc bài sau hết giờ trong khoảng thời gian từ 60 đến 90 giây.
- d. Tính xác suất giáo sư kết thúc bài sau hết giờ trong khoảng thời gian ít nhất là 90 giây.

3.1.4 theo dõi thực tế trọng lượng của một loại sản phẩm được quy định là có trọng lượng là 3 gam, có thể coi trọng lượng này là một biến ngẫu nhiên X liên tục có hàm mật độ

$$f(x) = \begin{cases} k[1 - (x - 3)^2] & 2 \leq x \leq 4 \\ 0 & \text{nếu khác} \end{cases}$$

- a. Tìm k .
- b. Tính xác suất trọng lượng loại sản phẩm này thực tế lớn hơn trọng lượng quy định.
- c. Tính xác suất chênh lệch giữa trọng lượng trong thực tế so với trọng lượng quy định của loại sản phẩm này nhỏ hơn 0,25 gam.
- d. Tính xác suất chênh lệch giữa trọng lượng trong thực tế so với trọng lượng quy định của loại sản phẩm này nhỏ hơn 0,5 gam.

3.1.5 Thời gian X (phút) cho một thợ lý phòng thí nghiệm chuẩn bị thiết bị cho một thử nghiệm nhất định được cho là có phân phối đều với $A=25$ và $B=35$.

- a. Xác định hàm mật độ xác suất của X .
- b. Tính xác suất thời gian chuẩn bị ít hơn 33 phút.
- c. Tính xác suất thời gian chuẩn bị ít hơn thời gian chuẩn bị trung bình 2 phút.
- d. Biết thợ lý đã chuẩn bị được 22 phút tính xác suất thời gian chuẩn bị không quá 30 phút.

3.1.6 Sinh viên H để đến trường cần bắt hai chuyến xe buýt. Giả sử thời gian chờ mỗi chuyến có phân phối đều với $A=0$ và $B=5$. Khi đó tổng thời gian chờ là biến

ngẫu nhiên Y có hàm mật độ

$$f(y) = \begin{cases} \frac{1}{25}y & 0 \leq y < 5 \\ \frac{2}{5} - \frac{1}{25}y & 5 \leq y \leq 10 \\ 0 & \text{nếu khác} \end{cases}$$

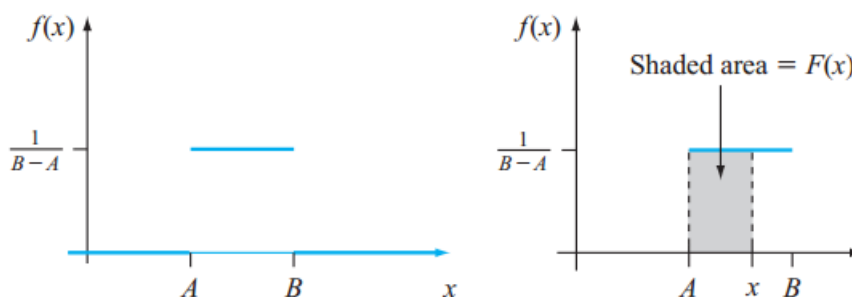
- a. Tính xác suất tổng thời gian chờ xe của H nhiều nhất là 3 phút.
- b. Tính xác suất tổng thời gian chờ xe của H trong khoảng 3 đến 8 phút.
- c. Biết sinh viên H đã chờ được 2 phút tính xác suất tổng thời gian chờ của H ít nhất là 5 phút.
- d. Biết sinh viên H đã chờ được 3 phút tính xác suất tổng thời gian chờ của H nhiều nhất là 8 phút.

3.2 Hàm phân phối tích lũy và các số đặc trưng

Định nghĩa 3.2.1. Hàm phân phối tích lũy $F(x)$ của biến ngẫu nhiên liên tục X được xác định bởi: với mọi số thực x thì

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(y)dy.$$

Ví dụ 3.3 Cho X là biến ngẫu nhiên có phân phối đều trên khoảng $[A, B]$. Hàm mật độ của X chỉ ra trong hình phía dưới.



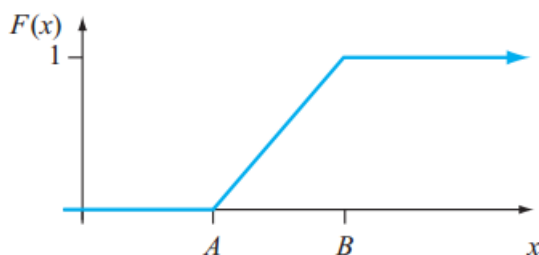
Nếu $x < A$, $F(x) = 0$ bởi vì không có diện tích trong đồ thị của hàm mật độ phía bên trái $x < A$. Nếu $x \geq B$, $F(x) = 1$ bởi vì toàn bộ diện tích tích lũy phía bên trái của $x > B$. Cuối cùng, nếu $A \leq x \leq B$,

$$F(x) = \int_{-\infty}^x f(y)dy = \int_A^x \frac{1}{B-A}dy = \frac{x-A}{B-A}$$

Khi đó, ta có hàm phân phối tích lũy là

$$F(x) = \begin{cases} 0 & x < A \\ \frac{x - A}{B - A} & A \leq x < B \\ 1 & x \geq B \end{cases}$$

Đồ thị của hàm $F(x)$ như sau.



Mệnh đề 3.2.2. Cho X là một biến ngẫu nhiên liên tục có hàm mật độ $f(x)$ và hàm phân phối $F(x)$. Khi đó, với mọi số thực a , ta có

$$P(X > a) = 1 - F(a),$$

và cho hai số a và b với $a < b$, ta có

$$P(a \leq X \leq b) = F(b) - F(a).$$

Ví dụ 3.4 Cho X là biến ngẫu nhiên liên tục có hàm mật độ là

$$f(x) = \begin{cases} \frac{1}{8} + \frac{3}{8}x & 0 \leq x \leq 2 \\ 0 & \text{nếu khác} \end{cases}$$

Xét giá trị x bất kì trong đoạn $[0, 2]$, ta có

$$F(x) = \int_{-\infty}^x f(y) dy = \int_0^x \left(\frac{1}{8} + \frac{3}{8}y \right) dy = \frac{x}{8} + \frac{3}{16}x^2$$

Khi đó,

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{x}{8} + \frac{3}{16}x^2 & 0 \leq x \leq 2 \\ 1 & x > 2 \end{cases}$$

Chẳng hạn, ta tính

$$\begin{aligned} P(1 \leq X \leq 1.5) &= F(1.5) - F(1) \\ &= \left[\frac{1}{8} 1.5 + \frac{3}{16} 1.5^2 \right] - \left[\frac{1}{8} 1 + \frac{3}{16} 1^2 \right] \\ &= \frac{19}{64} = 0.297 \end{aligned}$$

và

$$\begin{aligned} P(X > 1) &= 1 - P(X \leq 1) = 1 - F(1) = 1 - \left[\frac{1}{8} 1 + \frac{3}{16} 1^2 \right] \\ &= \frac{11}{16} = 0.688 \end{aligned}$$

Mệnh đề 3.2.3. Nếu X là biến ngẫu nhiên liên tục với hàm mật độ $f(x)$ và hàm phân phối $F(x)$ thì với mọi x sao cho đạo hàm $F'(x)$ tồn tại và $F'(x) = f(x)$.

Phân vị của phân phối liên tục

Định nghĩa 3.2.4. Cho p là một số thuộc khoảng $[0, 1]$. **Phân vị thứ (100p)** của một hàm phân phối của một biến ngẫu nhiên liên tục X , kí hiệu $\eta(p)$, được xác định bởi

$$p = F(\eta(p)) = \int_{-\infty}^{\eta(p)} f(y) dy \quad (3.1)$$

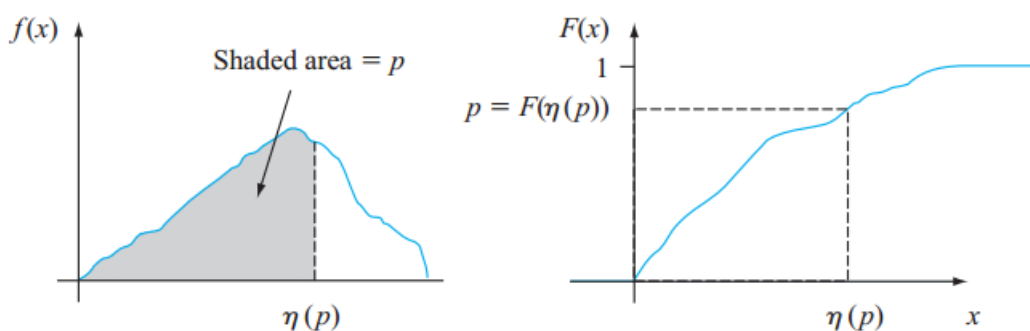


Figure 4.10 The (100p)th percentile of a continuous distribution

Ví dụ 3.5 Cho X là biến ngẫu nhiên liên tục có hàm mật độ

$$f(x) = \begin{cases} \frac{3}{2}(1 - x^2) & 0 \leq x \leq 1, \\ 0 & \text{nếu khác.} \end{cases}$$

Khi đó, cho x thuộc khoảng $[0, 1]$ ta có,

$$F(x) = \int_0^x \frac{3}{2}(1 - y^2)dy = \frac{3}{2} \left(x - \frac{x^3}{3} \right)$$

Phân vị thứ $(100p)$ của phân phối này thỏa mãn biểu thức sau

$$p = F(\eta(p)) = \frac{3}{2} \left(\eta(p) - \frac{\eta(p)^3}{3} \right)$$

nghĩa là,

$$(\eta(p))^3 - 3\eta(p) + 2p = 0$$

Chẳng hạn, phân vị thứ 50, tức là $p = 0.5$ và ta có phương trình $\eta^3 - 3\eta + 1 = 0$; với nghiệm tương ứng là $\eta = \eta(0.5) = 0.347$.

Định nghĩa 3.2.5. Trung vị của một phân phối liên tục, kí hiệu $\tilde{\mu}$, là phân vị thứ 50, sao cho $\tilde{\mu}$ thỏa mãn $0.5 = F(\tilde{\mu})$. Nghĩa là, trung vị là giá trị chia phân phối thành 2 phần bằng nhau.

Kì vọng

Định nghĩa 3.2.6. Kì vọng hay giá trị trung bình của một biến ngẫu nhiên liên tục X với hàm mật độ $f(x)$ là

$$\mu_X = E(X) = \int_{-\infty}^{\infty} x \cdot f(x)dx$$

Ví dụ 3.6 Cho X là biến ngẫu nhiên liên tục có hàm mật độ

$$f(x) = \begin{cases} \frac{3}{2}(1 - x^2) & 0 \leq x \leq 1, \\ 0 & \text{nếu khác.} \end{cases}$$

Khi đó,

$$E(X) = \int_{-\infty}^{\infty} x \cdot f(x)dx = \int_0^1 x \cdot \frac{3}{2}(1 - x^2)dx = \frac{3}{8}$$

Mệnh đề 3.2.7. Cho X là biến ngẫu nhiên liên tục và $h(X)$ là một hàm bất kì của X . Khi đó,

$$E[h(X)] = \mu_{h(X)} = \int_{-\infty}^{\infty} h(x) \cdot f(x)dx$$

Định nghĩa 3.2.8. Phương sai của một biến ngẫu nhiên liên tục X với hàm khối lượng $f(x)$ và kì vọng μ là

$$\sigma_X^2 = Var(X) = \int_{-\infty}^{\infty} (x - \mu)^2 \cdot f(x) dx = E[(X - \mu)^2]$$

Độ lệch chuẩn của X là $\sigma_X = \sqrt{Var(X)}$.

Ví dụ 3.7 Cho X là biến ngẫu nhiên liên tục có hàm mật độ

$$f(x) = \begin{cases} \frac{3}{2}(1 - x^2) & 0 \leq x \leq 1, \\ 0 & \text{nếu khác.} \end{cases}$$

Khi đó, ta có

$$E(X^2) = \int_{-\infty}^{\infty} x^2 \cdot f(x) dx = \int_0^1 x^2 \cdot \frac{3}{2}(1 - x^2) dx = \frac{1}{5}$$

và

$$Var(X) = \frac{1}{5} - \left(\frac{3}{8}\right)^2 = \frac{19}{320} \quad \text{và} \quad \sigma_X = 0.244$$

Bài tập: Hàm phân phối tích lũy và các số đặc trưng

3.2.1 Cho X là biến ngẫu nhiên có hàm phân phối xác suất

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{x^2}{4} & 0 \leq x \leq 2 \\ 1 & 2 \leq x \end{cases}$$

Tính:

- a. $P(X \leq 1)$
- b. $P(0,5 \leq X \leq 1)$
- c. $P(X > 1,5)$
- d. Tìm trung vị $\tilde{\mu}$ (tức là giải $0,5 = F(\tilde{\mu})$)
- e. Tính $F'(x)$ để lấy được hàm mật độ $f(x)$.
- g. Tính EX , $V(X)$, σ_X .

3.2.2 Biến ngẫu nhiên X có phân phối đều trên khoảng $[7,5; 20]$

- a. Xác định trung bình và phương sai của X .
- b. Xác định hàm phân phối xác suất của X

- c. Tính $P(X > EX)$
- d. Tính $P(10 < X < 15)$
- e. Tính $E(X^2)$; $E(X^3)$.

3.2.3 Cho X là biến ngẫu nhiên có hàm mật độ

$$f(x) = \begin{cases} 90x^8(1-x) & 0 \leq x \leq 1 \\ 0 & \text{nếu khác} \end{cases}$$

- a. Vẽ đồ thị $f(x)$. Sau đó tính $F(x)$ và vẽ nó.
- b. Tính $P(X \leq 0,5)$.
- c. Tính $P(0,25 < X \leq 0,5)$.
- d. Xác định phân vị thứ 75 của phân phối của X .
- e. Tính $E(X)$, σ_X .
- g. Tính xác suất X lớn hơn giá trị trung bình 1 độ lệch chuẩn.

3.2.4 Biến ngẫu nhiên X liên tục có hàm mật độ xác suất có dạng

$$f(x) = \begin{cases} kx^2 & 0 \leq x \leq 2 \\ 0 & \text{nếu khác} \end{cases}$$

- a. Vẽ đồ thị $f(x)$. Sau đó tính $F(x)$ và vẽ nó.
- b. Tính $P(X \leq 0,5)$.
- c. Tính $P(0,25 < X \leq 0,5)$.
- d. Xác định phân vị thứ 75 của phân phối của X .
- e. Tính $E(X)$, σ_X .
- g. Tính xác suất X lớn hơn giá trị trung bình 1 độ lệch chuẩn.

3.2.5 Cho X là biến ngẫu nhiên có hàm phân phối xác suất

$$F(x) = \begin{cases} 0 & x \leq 0 \\ \frac{x}{4} \left[1 + \ln \left(\frac{4}{x} \right) \right] & 0 < x \leq 4 \\ 1 & x > 4 \end{cases}$$

- a. Tính $P(X \leq 1)$.
- b. Tính $P(1 \leq X \leq 3)$.
- c. Tìm hàm mật độ của X .

3.2.6 Một nhà sinh thái học muốn đánh dấu một khu vực lấy mẫu hình tròn có bán kính 10m. Tuy nhiên, bán kính thực tế là một biến ngẫu nhiên R với hàm mật độ xác suất

$$f(r) = \begin{cases} \frac{3}{4}[1 - (10 - r)^2] & 9 \leq r \leq 11 \\ 0 & \text{trường hợp khác} \end{cases}$$

Tính diện tích trung bình của khu vực lấy mẫu.

3.2.7 Nhu cầu khí propane hàng tuần (tính bằng 1000 galông) của một cơ sở cụ thể là một biến ngẫu nhiên X với hàm mật độ xác suất là

$$f(x) = \begin{cases} 2(1 - \frac{1}{x^2}) & 1 \leq x \leq 2 \\ 0 & \text{trường hợp khác} \end{cases}$$

- a. Xác định phân phối xác suất của X .
- b. Xác định biểu thức cho giá trị phân vị mức thứ $(100p)\%$. Giá trị trung vị $\tilde{\mu}$ là bao nhiêu?
- c. Tính $E(X)$ và $V(X)$.
- d. Nếu 1,5 nghìn gallon đang tồn kho vào đầu tuần và không có nguồn cung mới nào trong tuần, thì sẽ có bao nhiêu trong số 1,5 nghìn gallon dự kiến sẽ còn lại vào cuối tuần?

3.3 Phân phối chuẩn

Định nghĩa 3.3.1. Một biến ngẫu nhiên liên tục X gọi là có phân phối chuẩn với hai tham số μ và σ , trong đó $-\infty < \mu < \infty$ và $0 < \sigma$, nếu hàm mật độ của X là

$$f(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/(2\sigma^2)}, \quad -\infty < x < \infty$$

Định nghĩa 3.3.2. Phân phối chuẩn với hai tham số $\mu = 0, \sigma = 1$ gọi là **phân phối chuẩn tắc**. Một biến ngẫu nhiên Z có phân phối chuẩn tắc với hàm mật độ xác suất là

$$f(z; 0, 1) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \quad -\infty < z < \infty$$

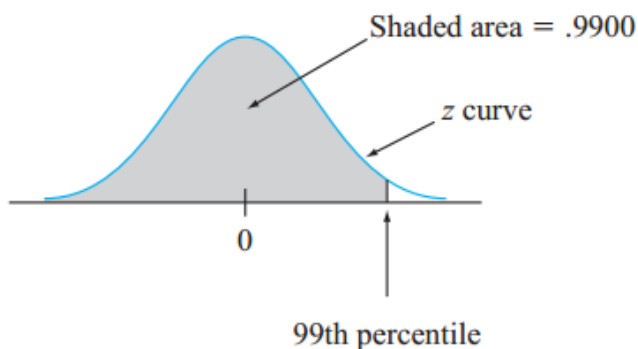
Hàm phân phối của Z là $P(Z \leq z) = \int_{-\infty}^z f(y; 0, 1)dy = \Phi(z)$.

Ví dụ 3.8 Cho Z là biến ngẫu nhiên có phân phối chuẩn tắc. Tính các xác suất sau: a) $P(Z \leq 1.25)$, b) $P(Z > 1.25)$ c) $P(Z \leq -1.25)$ d) $P(-0.38 \leq Z \leq 1.25)$

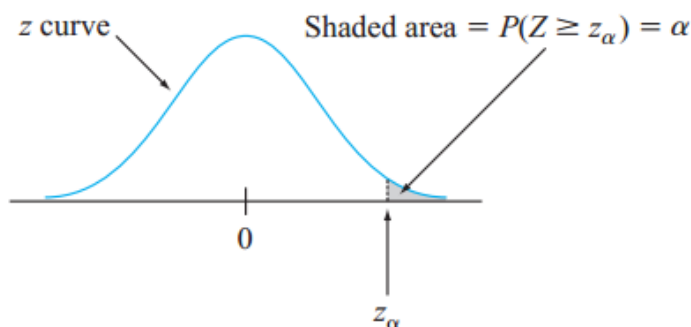
- a. $P(Z \leq 1.25) = \Phi(1.25)$, tra bảng ta được $\Phi(1.25) = 0.8944$. Do đó, $P(Z \leq 1.25) = 0.8944$
- b. $P(Z > 1.25) = 1 - P(Z \leq 1.25) = 0.1056$
- c. $P(Z \leq -1.25) = \Phi(-1.25) = 0.1056$
- d. $P(-0.38 \leq Z \leq 1.25) = P(Z \leq 1.25) - P(Z \leq -0.38) = \Phi(1.25) - \Phi(-0.38) = 0.8944 - 0.3520 = 0.5424$.

Phân vị của phân phối chuẩn

Ví dụ 3.9 Phân vị thứ 99 của phân phối chuẩn là giá trị nằm trên trục hoành sao cho diện tích dưới đường cong z nằm bên trái của giá trị này là 0.99.



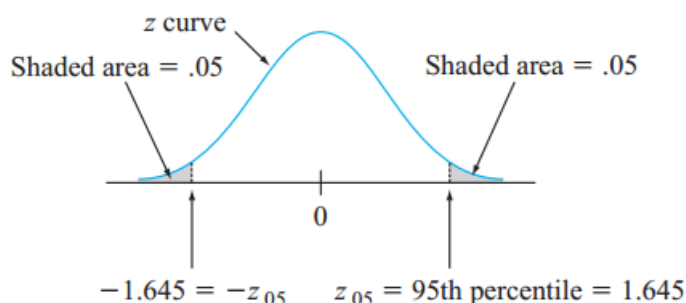
Ký hiệu. z_α biểu thị giá trị trên trục z mà α là diện tích vùng dưới đường cong z (đường cong của phân phối chuẩn) nằm bên phải z_α như hình dưới đây. Khi đó, z_α chính là phân vị thứ $100(1 - \alpha)$ của phân phối chuẩn.



Ta có bảng phân vị của phân phối chuẩn

Percentile	90	95	97.5	99	99.5	99.9	99.95
α (tail area)	.1	.05	.025	.01	.005	.001	.0005
$z_\alpha = 100(1 - \alpha)$ th percentile	1.28	1.645	1.96	2.33	2.58	3.08	3.27

Ví dụ 3.10 $z_{0.05}$ là phân vị thứ $100(1 - 0.05) = 95$ của phân phối chuẩn vì thế $z_{0.05} = 1.645$. Do tính đối xứng của phân phối chuẩn, diện tích dưới đường cong của phân phối chuẩn nằm bên trái của $-z_{0.05}$ cũng bằng với diện tích dưới đường cong của phân phối chuẩn nằm bên phải của $z_{0.05}$ (xem hình bên dưới).



Mệnh đề 3.3.3. Cho X là biến ngẫu nhiên có phân phối chuẩn với trung bình μ và độ lệch chuẩn σ . Khi đó,

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

có phân phối chuẩn tắc. Do đó,

$$P(a \leq X \leq b) = P\left(\frac{a - \mu}{\sigma} \leq Z \leq \frac{b - \mu}{\sigma}\right) = \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right)$$

$$P(X \leq a) = \Phi\left(\frac{a - \mu}{\sigma}\right) \quad P(X \geq b) = 1 - \Phi\left(\frac{b - \mu}{\sigma}\right)$$

Ví dụ 3.11 Cho X là biến ngẫu nhiên có phân phối chuẩn với giá trị trung bình 1.25 và độ lệch chuẩn 0.46. Tính xác suất: a) $P(1 \leq X \leq 1.75)$ b) $P(X > 2)$

a) Ta có

$$1 \leq X \leq 1.75$$

nếu và chỉ nếu

$$\frac{1 - 1.25}{0.46} \leq \frac{X - 1.25}{0.46} \leq \frac{1.75 - 1.25}{0.46}$$

Khi đó,

$$\begin{aligned} P(1 \leq X \leq 1.75) &= P\left(\frac{1 - 1.25}{0.46} \leq Z \leq \frac{1.75 - 1.25}{0.46}\right) \\ &= P(-0.54 \leq Z \leq 1.09) = \Phi(1.09) - \Phi(-0.54) \\ &= 0.8621 - 0.2946 = 0.5675 \end{aligned}$$

$$\text{b) } P(X > 2) = P\left(Z > \frac{2 - 1.25}{0.46}\right) = P(Z > 1.63) = 1 - \Phi(1.63) = 0.0516$$

Xấp xỉ phân phối nhị thức

Mệnh đề 3.3.4. Cho X là biến ngẫu nhiên có phân phối nhị thức với n phép thử với xác suất thành công p . Khi đó nếu đồ thị của phân phối nhị thức không quá lệch (không quá mất đối xứng) thì ta có thể xấp xỉ X như một phân phối chuẩn với $\mu = np$ và $\sigma = \sqrt{np(1-p)}$. Đặc biệt, nếu x là một giá trị của biến ngẫu nhiên X thì

$$P(X \leq x) = B(x; n, p) \simeq \Phi\left(\frac{x + 0.5 - np}{\sqrt{np(1-p)}}\right)$$

Trong các bài tập cụ thể, tính xấp xỉ này được áp dụng nếu $np \geq 10$ và $n(1-p) \geq 10$ thỏa mãn. Bởi vì đây là điều kiện cần để phân phối nhị thức có tính chất đối xứng.

Ví dụ 3.12 Giả sử một trường đại học có 25% sinh viên được nhận hỗ trợ tài chính. Gọi X là số lượng sinh viên được nhận hỗ trợ trong 50 sinh viên sao cho $p = 0.25$. Khi đó, ta có $\mu = 12.5$ và $\sigma = 3.06$. Vì $np = 50 \cdot 0.25 = 12.5 \geq 10$ và $n(1-p) = 37.5 \geq 10$ nên có thể xấp xỉ phân phối nhị thức. Xác suất mà nhiều nhất 10 sinh viên nhận hỗ trợ

$$P(X \leq 10) = B(10; 50, 0.25) \simeq \Phi\left(\frac{10 + 0.5 - 12.5}{3.06}\right) = \Phi(-0.65) = 0.2578$$

Xác suất có từ 5 đến 15 sinh viên được nhận hỗ trợ tài chính

$$\begin{aligned} P(5 \leq X \leq 15) &= B(15; 50, 0.25) - B(4; 50, 0.25) \\ &\simeq \Phi\left(\frac{15.5 - 12.5}{3.06}\right) - \Phi\left(\frac{4.5 - 12.5}{3.06}\right) = 0.8320 \end{aligned}$$

Bài tập: Phân phối chuẩn

3.3.1 Cho X là biến ngẫu nhiên có phân phối chuẩn tắc. Tính

a. $P(0 \leq X \leq 2, 17)$

d. $P(Z \leq 2, 7)$

- b. $P(-2, 5 \leq X \leq 0)$
c. $P(-2, 5 \leq X \leq 2, 5)$

3.3.2 Trong mỗi trường hợp sau, xác định giá trị của hằng số c thỏa mãn

- a.** $\Phi(c) = 0,9838$
b. $P(0 \leq Z \leq c) = 0,291$
c. $P(c < Z) = 0,121$
d. $P(-c \leq Z \leq c) = 0,668$
e. $P(c \leq |Z|) = 0,016$.

3.3.3 Xác định các phân vị **a)** thứ 91 **b)** thứ 9 **c)** thứ 75 **d)** thứ 25 **e)** thứ 6 của phân phối chuẩn.

3.3.4 Xác định giá trị z_α khi

- a.** $\alpha = 0,0055$ **b.** $\alpha = 0,663$ **c.** $\alpha = 0,09$

3.3.5 Biến ngẫu nhiên X có phân phối chuẩn với trung bình bằng 15 và độ lệch chuẩn bằng 1,25. Tính

- a.** $P(X \leq 15)$
- b.** $P(X \geq 10)$
- c.** $P(14 \leq X \leq 18)$
- d.** $P(|Z - 15| \leq 3)$

3.3.6 Một xe máy có vận tốc tuân theo phân phối chuẩn với giá trị trung bình là 46,8 km/h và độ lệch chuẩn 1,75 km/h.

- Tính xác suất vận tốc tối đa nhiều nhất là 50km/h.
- Tính xác suất vận tốc tối đa ít nhất là 50km/h.
- Tính xác suất vận tốc tối đa khác giá trị trung bình nhiều nhất là 1,5 lần độ chuẩn.

3.3.7 Đường kính (inch) của một số cây có phân phối chuẩn với $\mu = 8,8$ và $\sigma = 2,8$.

- Tính xác suất chọn được cây có đường kính ít nhất 10 inch? vượt qua 10 inch?
- Tính xác suất chọn được cây có đường kính không vượt quá 20 inch biết cây này có đường kính ít nhất là 10 inch?
- Tính xác suất chọn được cây có đường kính ít nhất là 5 inch và không quá 10 inch?
- Tìm giá trị c sao cho khoảng $(8, 8 - c; \quad 8, 8 + c)$ bao gồm 98% các giá trị đường kính?

e. Nếu bốn cây được chọn ngẫu nhiên, tính xác suất có ít nhất một cây có đường kính vượt qua 10 inch?

3.3.8 Giả sử nồng độ chloride (mmol/L) trong máu có phân phối chuẩn với trung bình 104 và độ lệch chuẩn 5.

a. Tính xác suất nồng độ chloride trong máu bằng 105? ít hơn 105? nhiều nhất 105?

b. Tính xác suất nồng độ chloride trong máu chênh lệch so với giá trị trung bình 1 độ lệch chuẩn? Xác suất này có phụ thuộc vào μ và σ không?

3.3.9 Có hai máy cắt các nút bần được sử dụng cho các chai rượu. Máy thứ nhất cắt các nút bần với đường kính là biến ngẫu nhiên có phân phối chuẩn với trung bình là 3cm và độ lệch chuẩn 0,1cm. Máy thứ hai cắt các nút bần với đường kính là biến ngẫu nhiên có phân phối chuẩn với trung bình là 3,04cm và độ lệch chuẩn 0,02cm. Nút bần cắt ra được chấp nhận nếu có đường kính từ 2,9cm đến 3,1cm. Vậy máy nào sử dụng tốt hơn?

3.3.10 a. Xác định phân vị thứ 91, thứ 6 của phân phối chuẩn có $\mu = 30$ và $\sigma = 5$.

b. Độ rộng của một đường cắt bằng axit là biến ngẫu nhiên có phân phối chuẩn với $\mu = 3\mu m$ và $\sigma = 0,14$. Giá trị nào của độ rộng của đường cắt bằng axit chia các giá trị độ rộng thành 2 phần 10% và 90%.

3.3.11 Thiết bị tự động mở dù của một dù nhảy quân sự với quy định là dù mở ở độ cao 200 m so với mặt đất. Giả sử độ cao mở dù thực sự có phân phối chuẩn với giá trị trung bình 200 m và độ lệch chuẩn 30 m. Thiết bị sẽ bị thiệt hại nếu dù mở ra ở độ cao dưới 100 m. Tính xác suất có ít nhất một trong năm chiếc dù rơi độc lập bị thiệt hại?

3.3.12 Một máy sản xuất vòng bi ban đầu thiết lập để đường kính trung bình của vòng bi là 0,5 inch. Một vòng bi được chấp nhận nếu đường kính của nó lệch khoảng 0,004 inch so với giá trị mục tiêu này. Tuy nhiên, giả sử thiết lập đã thay đổi trong quá trình sản xuất nên vòng bi sản xuất ra có đường kính có phân bố chuẩn với giá trị trung bình là 0,499 inch và độ lệch chuẩn 0,002 inch. Tính tỷ lệ phần trăm của vòng bi sản xuất ra sẽ không được chấp nhận được?

3.3.13 Quan sát các em bé sinh ra trong khoảng thời gian "chuẩn" từ 37-43 tuần tuổi thai. Dữ liệu được giả định rằng đối với những đứa trẻ sinh ra ở Hoa Kỳ, trọng lượng khi sinh có phân bố chuẩn với mức trung bình là 3432g và độ lệch tiêu chuẩn 482g.

a. Tính xác suất một đứa trẻ được chọn ngẫu nhiên khi sinh ra có trọng lượng vượt quá 4000g? giữa 3000g đến 4000g?

b. Tính xác suất một đứa trẻ được chọn ngẫu nhiên khi sinh ra có trọng lượng

dưới 2000g hoặc lớn hơn 5000g.

c. Xác định giá trị trọng lượng đặc trưng của 0,1% các trọng lượng các em bé mới sinh.

3.3.14 Cho X là biến ngẫu nhiên có phân phối nhị thức với các tham số $n = 2500$ và p . Tính mỗi xác suất sau bằng phương pháp xấp xỉ chuẩn trong các trường hợp $p = 0,5; 0,6; 0,8$.

a. $P(1500 \leq X \leq 2000)$.

b. $P(X \leq 1500)$.

c. $P(2000 \leq X)$.

3.3.15 Giả sử chỉ có 75% số người lái xe có đeo dây an toàn. Chọn ngẫu nhiên một mẫu ngẫu nhiên gồm 500 tài xế. Tính xác suất

a. Số tài xế đeo dây an toàn từ 360 đến 400 xế.

b. Có ít hơn 400 người đeo dây an toàn.

3.4 Phân phối mũ và Gamma

Phân phối mũ

Định nghĩa 3.4.1. Một biến ngẫu nhiên X gọi là có phân phối mũ với tham số $\lambda (\lambda > 0)$ nếu hàm mật độ xác suất của X là

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0, \\ 0 & \text{nếu khác.} \end{cases}$$

Hàm phân phối của X là

$$F(x; \lambda) = \begin{cases} 1 - e^{-\lambda x} & x \geq 0, \\ 0 & x < 0 \end{cases}$$

Giá trị trung bình và phương sai của X lần lượt là $\mu = \frac{1}{\lambda}; \sigma^2 = \frac{1}{\lambda^2}$

Ví dụ 3.13 Cho X là biến ngẫu nhiên có phân phối mũ với giá trị trung bình là 6.

Tính xác suất: a) $P(X \leq 10)$ b) $P(5 \leq X \leq 10)$

Ta có: $6 = EX = \frac{1}{\lambda}$, do đó $\lambda = 0.1667$.

a) $P(X \leq 10) = F(10; 0.1667) = 1 - e^{-(0.1667)(10)} = 1 - 0.189 = 0.811$

b) $P(5 \leq X \leq 10) = F(10; 0.1667) - F(5; 0.1667) = (1 - e^{-1.667}) - (1 - e^{-8.335}) = 0.246$

Phân phối Gamma

Định nghĩa 3.4.2. Cho $\alpha > 0$, khi đó **hàm gamma** $\Gamma(\alpha)$ được xác định bởi

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx$$

Một số tính chất quan trọng của hàm gamma:

1. Cho bất kì $\alpha > 1$, $\Gamma = (\alpha - 1) \cdot \Gamma(\alpha - 1)$.
2. Cho n là một số nguyên dương bất kì, ta có $\Gamma(n) = (n - 1)!$
3. $\Gamma(\frac{1}{2}) = \sqrt{\pi}$

Đặt

$$f(x; \alpha) = \begin{cases} \frac{x^{\alpha-1} e^{-x}}{\Gamma(\alpha)} & x \geq 0, \\ 0 & \text{nếu khác.} \end{cases} \quad (3.2)$$

Khi đó, ta thấy $f(x; \alpha) \geq 0$ và $\int_0^{\infty} f(x; \alpha) dx = \Gamma(\alpha)/\Gamma(\alpha) = 1$, vì thế $f(x; \alpha)$ thỏa mãn hai tính chất cơ bản của hàm mật độ xác suất.

Định nghĩa 3.4.3. Một biến ngẫu nhiên liên tục X gọi là có **phân phối gamma** nếu hàm mật độ của X là

$$f(x; \alpha, \beta) = \begin{cases} \frac{1}{\beta^{\alpha} \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} & x \geq 0, \\ 0 & \text{nếu khác.} \end{cases} \quad (3.3)$$

trong đó α và β là hai tham số dương.

Hàm gamma chuẩn tắc khi $\beta = 1$ và hàm mật độ của X lúc đó là (3.2)

Kì vọng và phương sai của biến ngẫu nhiên X có hàm phân phối gamma $f(x; \alpha, \beta)$ là

$$E(X) = \mu = \alpha\beta \quad Var(X) = \sigma^2 = \alpha\beta^2$$

Hàm phân phối của biến ngẫu nhiên X có phân phối gamma chuẩn tắc là

$$F(x; \alpha) = \int_0^x \frac{y^{\alpha-1} e^{-y}}{\Gamma(\alpha)} dy \quad x > 0$$

còn được gọi là **hàm gamma không đầy đủ**.

Ví dụ 3.14 Cho X là biến ngẫu nhiên có phân phối gamma chuẩn tắc với $\alpha = 2$. Ta có

$$P(a \leq X \leq b) = F(b) - F(a)$$

khi X là biến ngẫu nhiên liên tục. Do đó,

$$P(3 \leq X \leq 5) = F(5; 2) - F(3; 2) = 0.960 - 0.801 = 0.159$$

và

$$P(X > 4) = 1 - P(X \leq 4) = 1 - F(4; 2) = 1 - 0.908 = 0.092$$

Mệnh đề 3.4.4. Cho X là biến ngẫu nhiên có hàm phân phối gamma với tham số α và β . Khi đó, với mọi $x > 0$, hàm phân phối của X là

$$P(X \leq x) = F(x; \alpha, \beta) = F\left(\frac{x}{\beta}; \alpha\right)$$

với $F(\bullet; \alpha)$ là hàm gamma không đầy đủ.

Ví dụ 3.15 Giả sử vòng đời X của một con chuột được tính theo tuần sống trong môi trường có bức xạ 240rads có hàm phân phối gamma với $\alpha = 8$ và $\beta = 15$. Khi đó, thời gian sống sót trung bình là $EX = 8 \cdot 15 = 120$ tuần, trong khi $Var(X) = 8 \cdot 15^2 = 1800$ và $\sigma_X = \sqrt{1800} = 42.43$ tuần. Xác suất mà con chuột sống từ 60 đến 129 tuần là

$$\begin{aligned} P(60 \leq X \leq 120) &= P(X \leq 120) - P(X \leq 60) \\ &= F(120/15; 8) - F(60/15; 8) \\ &= F(8; 8) - F(4; 8) = 0.547 - 0.051 = 0.496 \end{aligned}$$

Xác suất con chuột sống ít nhất 30 tuần

$$\begin{aligned} P(X \geq 30) &= 1 - P(X < 30) = 1 - P(X \leq 30) \\ &= 1 - F(30/15; 8) = 0.999 \end{aligned}$$

Phân phối Chi-bình phương

Định nghĩa 3.4.5. Cho ν là một số thực dương. Khi đó, một biến ngẫu nhiên X được gọi là **phân phối Chi-bình phương** với tham số ν nếu hàm mật độ xác suất

của X là hàm mật độ gamma có tham số $\alpha = \nu/2$ và $\beta = 2$. Hàm mật độ của một biến ngẫu nhiên có phân phối Chi-bình phương là

$$f(x; \nu) = \begin{cases} \frac{1}{2^{\nu/2}\Gamma(\nu/2)} x^{\nu/2-1} e^{-x/2} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Tham số ν gọi là **bậc tự do** của X . Ta dùng kí hiệu χ^2 thường dùng thay thế "Chi-bình phương".

Bài tập: Phân phối mũ và Gamma

3.4.1 Đặt X là thời gian giữa hai lượt đến liên tiếp tại cửa sổ của một ngân hàng địa phương, là nơi khách hàng được phục vụ mà không cần rời khỏi xe. Nếu X có phân phối mũ với tham số $\lambda = 1$ (giống hệt với phân phối gamma chuẩn với $\alpha = 1$), hãy tính

- Thời gian kỳ vọng, độ lệch chuẩn của thời gian giữa hai lần đến liên tiếp.
- Tính $P(X \leq 4)$; $P(2 \leq X \leq 5)$.

3.4.2 Biến ngẫu nhiên X biểu thị khoảng cách (m) di chuyển từ vị trí sinh của một con vật đến vị trí lãnh thổ đầu tiên mà nó gặp phải. Giả sử rằng đối với chuột kangaroo, X có phân phối mũ với tham số $\lambda = 0,01386$.

- Xác suất khoảng cách tối đa là 100 m là bao nhiêu? Tối đa 200 m? Từ 100 đến 200 m?
- Tính xác suất khoảng cách vượt quá khoảng cách trung bình nhiều hơn 2 độ lệch chuẩn.
- Xác định giá trị của trung vị của X .

3.4.3 Một người tiêu dùng đang lựa chọn giữa hai cách gọi đường dài. Cách đầu tiên tính phí là 10 cen trong mỗi phút. Giá cước cách thứ hai tính phí 99 cen cho cuộc gọi tối đa 20 phút và sau đó 10 cen cho mỗi phút vượt quá 20 phút (giả sử cuộc gọi kéo dài số phút không đến 20 phút tính tương ứng với toàn bộ phí của 20 phút là 99 cen). Giả sử thời lượng cuộc gọi của người tiêu dùng có phân phối mũ với tham số λ .

- Hãy giải thích cách lựa chọn kế hoạch gọi điện phải phụ thuộc vào thời lượng cuộc gọi trung bình.
- Kế hoạch nào tốt hơn nếu thời lượng cuộc gọi trung bình là 10 phút? 15 phút? [Gợi ý: Cho $h_1(x)$ biểu thị chi phí cho gói đầu tiên khi thời lượng cuộc gọi là x phút và cho $h_2(x)$ là chi phí cho gói thứ hai. Đưa ra các biểu thức cho hai hàm chi phí này và sau đó xác định chi phí trung bình cho mỗi gói.]

3.4.4 Tính a. $\Gamma(6)$ b. $\Gamma(5/2)$ c. $F(4; 5)$ d. $F(5; 4)$ e. $F(0; 4)$

3.4.5 Cho biến ngẫu nhiên X có phân phối gamma chuẩn với $\alpha = 7$, tính:

- a. $P(X \leq 5)$ b. $P(X < 5)$ c. $P(X > 8)$ d. $P(3 \leq X \leq 8)$
 e. $P(3 < X < 8)$ g. $P(X < 4 \text{ hoặc } X > 6)$

3.4.6 Giả sử thời gian sử dụng thiết bị đầu cuối kết nối với một cơ sở máy tính địa phương của một sinh viên được chọn ngẫu nhiên có phân phối gamma với trung bình là 20 phút và phương sai 80 phút².

- a. Xác định giá trị của α và β .
 b. Tính xác suất mà học sinh sử dụng thiết bị đầu cuối tối đa là 24 phút.
 c. Tính xác suất mà một sinh viên sử dụng thiết bị đầu cuối từ 20 đến 40 phút.

3.4.7 Tuổi thọ X (tính theo tuần) của một loại bóng bán dẫn nhất định có phân phối gamma với trung bình là 24 tuần và độ lệch chuẩn 12 tuần.

- a. Tính xác suất tuổi thọ một bóng bán dẫn sẽ kéo dài từ 12 đến 24 tuần.
 b. Tính xác suất tuổi thọ một bóng bán dẫn sẽ kéo dài tối đa 24 tuần. Trung vị của tuổi thọ một bóng bán dẫn có dưới 24 tuần hay không?
 c. Phân vị thứ 99 của tuổi thọ một bóng bán dẫn là bao nhiêu?
 d. Giả sử thời gian thử nghiệm quan sát tuổi thọ của loại bóng bán dẫn này thực sự sẽ bị chấm dứt sau t tuần. Giá trị của t là bao nhiêu để chỉ có 0,5% của tất cả các bóng bán dẫn vẫn sẽ hoạt động khi chấm dứt thử nghiệm?

3.5 Một số phân phối liên tục khác

Phân phối Weibull

Định nghĩa 3.5.1. Một biến ngẫu nhiên X gọi là có **phân phối Weibull** với tham số α and β ($\alpha > 0, \beta > 0$) nếu hàm mật độ của X là

$$f(x; \alpha, \beta) = \begin{cases} \frac{\alpha}{\beta^\alpha} x^{\alpha-1} e^{-(x/\beta)^\alpha} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Kì vọng và phương sai của X

$$\mu = \beta \Gamma\left(1 + \frac{1}{\alpha}\right) \quad \sigma^2 = \beta^2 \left[\Gamma\left(1 + \frac{2}{\alpha}\right) - \left(\Gamma\left(1 + \frac{1}{\alpha}\right) \right)^2 \right]$$

Hàm phân phối của X

$$F(x; \alpha, \beta) = \begin{cases} 0 & x < 0 \\ 1 - e^{-(x/\beta)^\alpha} & x \geq 0 \end{cases}$$

Ví dụ 3.16 Cho X là biến ngẫu nhiên có phân phối Weibull với $\alpha = 2, \beta = 10$. Tính xác suất:

$$P(X \leq 10) = F(10; 2, 10) = 1 - e^{-(10/10)^2} = 0.632$$

Phân phối lô-ga chuẩn

Định nghĩa 3.5.2. Một biến ngẫu nhiên không âm X gọi là có **phân phối lô-ga chuẩn** nếu biến ngẫu nhiên $Y = \ln(X)$ có phân phối chuẩn. Khi đó, hàm mật độ của X là

$$f(x; \mu, \sigma) = \begin{cases} \frac{1}{\sqrt{2\pi}\sigma x} e^{-[\ln(x)-\mu]^2/(2\sigma^2)} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Kì vọng và phương sai của X là

$$E(X) = e^{\mu+\sigma^2/2} \quad Var(X) = e^{2\mu+\sigma^2} \cdot (e^{\sigma^2} - 1)$$

Vì $\ln(X)$ có phân phối chuẩn nên hàm phân phối của X có thể biểu diễn qua hàm phân phối $\Phi(z)$ của biến ngẫu nhiên chuẩn tắc Z .

$$\begin{aligned} F(x; \mu, \sigma) &= P(X \leq x) = P[\ln(X) \leq \ln(x)] \\ &= P\left(Z \leq \frac{\ln(x) - \mu}{\sigma}\right) = \Phi\left(\frac{\ln(x) - \mu}{\sigma}\right) \quad x \geq 0 \end{aligned}$$

Ví dụ 3.17 Cho X là biến ngẫu nhiên có phân phối lô-ga chuẩn với $\mu = 0.353, \sigma = 0.754$. Khi đó, ta tính được kì vọng và phương sai của X

$$EX = e^{0.353+0.754^2/2} = 1.891$$

$$Var(X) = e^{2 \cdot 0.353+0.754^2} \cdot (e^{0.754^2} - 1) = 2.7387$$

Khi đó,

$$\begin{aligned} P(1 \leq X \leq 2) &= P(\ln(1) \leq \ln(X) \leq \ln(2)) = P(0 \leq \ln(X) \leq 0.693) \\ &= P\left(\frac{0 - 0.353}{0.754} \leq Z \leq \frac{0.693 - 0.353}{0.754}\right) = \Phi(0.47) - \Phi(-0.45) = 0.354 \end{aligned}$$

Phân phối Beta

Định nghĩa 3.5.3. Một biến ngẫu nhiên X có **phân phối beta** với tham số $\alpha >, \beta > 0, A, B$ nếu hàm mật độ của X có dạng

$$f(x; \alpha, \beta, A, B) = \begin{cases} \frac{1}{B-A} \cdot \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha) \cdot \Gamma(\beta)} \left(\frac{x-A}{B-A}\right)^{\alpha-1} \left(\frac{B-x}{B-A}\right)^{\beta-1} & A \leq x \leq B \\ 0 & \text{nếu khác} \end{cases}$$

Trường hợp, $A = 0, B = 1$, ta gọi X có **phân phối beta chuẩn**.

Kì vọng và phương sai của X

$$\mu = A + (B-A) \cdot \frac{\alpha}{\alpha+\beta} \quad \sigma^2 = \frac{(B-A)^2 \alpha \beta}{(\alpha+\beta)^2 (\alpha+\beta+1)}$$

Ví dụ 3.18 Cho X là biến ngẫu nhiên có phân phối beta với các tham số sau $A = 2, B = 5, \alpha = 2, \beta = 3$. Khi đó,

$$\begin{aligned} P(X \leq 3) &= \int_2^3 \frac{1}{3} \cdot \frac{4!}{1!2!} \cdot \frac{x-2}{3} \left(\frac{5-x}{3}\right) dx \\ &= \frac{4}{27} \int_2^3 (x-2)(5-x)^2 dx = 0.407 \end{aligned}$$

BÀI TẬP

1. Cho X là biến ngẫu nhiên có hàm mật độ

$$f(x) = \begin{cases} 0.075x + 0.2 & 3 \leq x \leq 5 \\ 0 & \text{nếu khác} \end{cases}$$

- Vẽ đồ thị và kiểm tra tổng diện tích dưới hàm mật độ có bằng 1 không.
- Tính $P(X \leq 4)$ và so sánh với $P(X < 4)$.
- Tính $P(3.5 \leq X \leq 4.5)$ và $P(4.5 < X)$.

3. Cho X là biến ngẫu nhiên có hàm mật độ

$$f(x) = \begin{cases} 0.09375(4-x^2) & -2 \leq x \leq 2 \\ 0 & \text{nếu khác} \end{cases}$$

- Vẽ đồ thị $f(x)$.
- Tính $P(X < 0)$.
- Tính $P(-1 < X < 1)$.
- Tính $P(X < -0.5 \text{ hoặc } X > 0.5)$.

11. Cho X là biến ngẫu nhiên có hàm phân phối

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{x^2}{4} & 0 \leq x \leq 2 \\ 1 & 2 \leq x \end{cases}$$

Tính:

- a. $P(X \leq 1)$
- b. $P(0.5 \leq X \leq 1)$
- c. $P(X > 1.5)$
- d. Tìm trung vị $\tilde{\mu}$ (tức là giải $0.5 = F(\tilde{\mu})$)
- e. Tìm $F'(x)$ để lấy được hàm mật độ $f(x)$.
- f. $EX, V(X), \sigma_X$.

15. Cho X là biến ngẫu nhiên có hàm mật độ

$$f(x) = \begin{cases} 90x^8(1-x) & 0 \leq x \leq 1 \\ 0 & \text{nếu khác} \end{cases}$$

- a. Vẽ đồ thị $f(x)$. Sau đó tính $F(x)$ và vẽ nó.
 - b. $P(X \leq 0.5)$ là gì?
 - c. Từ câu (a), tính $P(0.25 < X \leq 0.5)$
 - d. Phân vị thứ 75 của phân phối là bao nhiêu?
 - e. Tính $E(X), \sigma_X$.
 - f. Tính xác suất X lớn hơn 1 độ lệch chuẩn từ giá trị trung bình?
19. Cho X là biến ngẫu nhiên có hàm phân phối

$$F(x) = \begin{cases} 0 & x \leq 0 \\ \frac{x}{4} \left[1 + \ln \left(\frac{4}{x} \right) \right] & 0 < x \leq 4 \\ 1 & x > 4 \end{cases}$$

- a. $P(X \leq 1)$?
- b. $P(1 \leq X \leq 3)$
- c. Tìm hàm mật độ của X .

29. Trong mỗi trường hợp sau, xác định giá trị của hằng số c
- a. $\Phi(c) = 0.9838$ b. $P(0 \leq Z \leq c) = 0.291$
 - c. $P(c \leq Z) = 0.121$ d. $P(-c \leq Z \leq c) = 0.668$
 - e. $P(c \leq |Z|) = 0.016$.
31. Xác định giá trị z_α khi
- a. $\alpha = 0.0055$ b. $\alpha = 0.663$ c. $\alpha = 0.09$
33. Một xe máy có vận tốc tuân theo phân phối chuẩn với giá trị trung bình là 46.8 km/h và độ lệch chuẩn 1.75 km/h.
- a. Tính xác suất vận tốc tối đa nhiều nhất là 50km/h.
 - b. Tính xác suất vận tốc tối đa ít nhất là 50km/h.
 - c. Tính xác suất vận tốc tối đa khác giá trị trung bình nhiều nhất là 1.5 độ lệch chuẩn.
35. Đường kính (inch) của một số cây có phân phối chuẩn với $\mu = 8.8$ và $\sigma = 2.8$.
- a. Tính xác suất chọn được cây có đường kính ít nhất 10 inch? vượt qua 10 inch?
 - b. Tính xác suất chọn được cây có đường kính vượt qua 20 inch?
 - c. Tính xác suất chọn được cây có đường kính nằm trong khoảng từ 5 inch đến 10 inch?
 - d. Tìm giá trị c sao cho khoảng $(8.8 - c, 8.8 + c)$ bao gồm 98% của tất cả các giá trị đường kính?
 - e. Nếu bốn cây được chọn ngẫu nhiên, tính xác suất có ít nhất một cây có đường kính vượt qua 10 inch?
37. Giả sử nồng độ máu chloride (mmol/L) có phân phối chuẩn với trung bình 104 và độ lệch chuẩn 5.
- a. Tính xác suất nồng độ máu chloride bằng 105? ít hơn 105? nhiều nhất 105?
 - b. Tính xác suất nồng độ máu chloride khác với giá trị trung bình 1 độ lệch chuẩn? Xác suất này có phụ thuộc vào μ và σ không?
 - c. Làm thế nào để đặc trưng nhất 0.1% nồng độ chloride?
41. Thiết bị tự động mở dù của một dù nhảy quân sự là 200 m trên mặt đất. Giả sử độ cao mở dù thực sự có phân phối chuẩn với giá trị trung bình 200 m và

độ lệch chuẩn 30 m. Thiết bị sẽ bị thiệt hại nếu dù mở ra ở độ cao dưới 100 m. Tính xác suất có ít nhất một trong năm chiếc dù rơi độc lập bị thiệt hại?

45. Một máy sản xuất vòng bi ban đầu thiết lập để đường kính trung bình của vòng bi là 0.500 inch. Một vòng bi được chấp nhận nếu đường kính của nó lệch khoảng 0,004 inch so với giá trị mục tiêu này. Tuy nhiên, giả sử thiết lập đã thay đổi trong quá trình sản xuất, để vòng bi có đường kính phân bố chuẩn với giá trị trung bình là 0.499 inch và độ lệch chuẩn 0.002 inch. Tính tỷ lệ phần trăm của vòng bi sản xuất ra sẽ không được chấp nhận được?

49. Xem xét các em bé sinh ra trong phạm vi "chuẩn" từ 37-43 tuần tuổi thai. Dữ liệu mở rộng hỗ trợ giả định rằng đối với những đứa trẻ sinh ra ở Hoa Kỳ, trọng lượng khi sinh thường phân bố với mức trung bình là 3432 g và độ lệch tiêu chuẩn 482 g.

a. Tính xác suất sinh một đứa trẻ được chọn ngẫu nhiên vượt quá 4000 g? giữa 3000 đến 4000 g?

b. Tính xác suất sinh một đứa trẻ được chọn ngẫu nhiên dưới 2000 g hoặc lớn hơn 5000 g.

c. Tính xác suất là trọng lượng sinh của một em bé được chọn vượt quá 7 lb?

d. Làm thế nào đặc trưng nhất 0,1% của trọng lượng sinh em bé?

53. Cho X là biến ngẫu nhiên có phân phối nhị thức với các tham số $n = 25$ và p . Tính mỗi xác suất sau bằng phương pháp xấp xỉ chuẩn trong các trường hợp $p = 0.5, 0.6, 0.8$.

a. $P(15 \leq X \leq 20)$.

b. $P(X \leq 15)$.

c. $P(20 \leq X)$.

55. Giả sử chỉ có 75% số người lái xe có đeo dây an toàn. Một mẫu ngẫu nhiên của 500 tài xế được chọn. Tính xác suất

a. Giữa 360 và 400 các tài xế đeo dây an toàn?

b. Ít hơn 400 người đeo dây an toàn?

Bài tập: Một số phân phối liên tục khác

3.5.1 Tuổi thọ X (đơn vị: trăm giờ) của một loại ống chân không nhất định có phân

bố Weibull với các tham số $\alpha = 2$ và $\beta = 3$. Tính

- a. $E(X); V(X)$ b. $P(X \leq 6)$ c. $P(1,5 \leq X \leq 6)$

3.5.2 Các tác giả của bài báo “A Probabilistic Insulation Life Model for Combined Thermal-Electrical Stresses” (IEEE Trans. on Elect. Insulation, 1985: 519–522) cho rằng “phân bố Weibull được sử dụng rộng rãi trong các vấn đề thống kê liên quan đến lão hóa chất rắn vật liệu cách điện.” Họ đề xuất việc sử dụng phân phối như một mô hình cho tuổi thọ (tính theo giờ) của các mẫu cách điện rắn chịu điện áp AC. Các giá trị của các thông số phụ thuộc vào điện áp và nhiệt độ; giả sử $\alpha = 2,5$ và $\beta = 200$ (các giá trị được đề xuất bởi dữ liệu trong bài viết).

- a. Tính xác suất tuổi thọ của mẫu vật nhiều nhất là 250? Ít hơn 250? Hơn 300?
b. Tính xác suất tuổi thọ của mẫu vật là từ 100 đến 250?
c. Xác định mốc thời gian mà 50% các mẫu có tuổi thọ vượt quá mốc thời gian này.

3.5.3 Thời gian hoàn thành một công việc nào đó là một biến ngẫu nhiên X có phân bố loga với $\mu = 4,5$ và $\sigma = 0,8$.

- a. Giá trị trung bình và độ lệch chuẩn của thời gian này là bao nhiêu?
b. Tính xác suất X có giá trị tối đa là 100.
c. Tính xác suất X có giá trị tối thiểu là 200. Lớn hơn 200.

3.5.4 Một giải thích dựa trên cơ chế hỏng hóc vật liệu nào đó với giả thiết rằng độ bền dẻo X của vật liệu có phân bố loga chuẩn. Giả sử các tham số là $\mu = 5$ và $\sigma = 0,1$.

- a. Tính $E(X); V(X)$ và $m\tilde{u}_X$.
b. Tính $P(X > 125); P(110 \leq X \leq 125)$.
c. Nếu quan sát ngẫu nhiên 10 mẫu vật thì số mẫu trung bình có độ bền dẻo ít nhất là 125 là bao nhiêu?

d. Vật liệu không đạt yêu cầu nếu có độ bền dẻo ít hơn giá trị m . Giá trị bền dẻo m tối thiểu là bao nhiêu để có không quá 5% vật liệu không đạt yêu cầu.

3.5.5 Xác định α và β để hàm mật độ của phân phối beta chuẩn là đối xứng.

3.5.6 Giả sử tỷ lệ X của diện tích bề mặt trong một ô tiêu chuẩn xác định được chọn ngẫu nhiên được bao phủ bởi cây có phân bố beta chuẩn với $\alpha = 5$ và $\beta = 2$.

- a. Tính $E(X)$ và $V(X)$.
b. Tính $P(X \leq 0,2); P(0,2 \leq X \leq 0,4)$.
c. Tỷ lệ dự kiến của khu vực lấy mẫu không được bao phủ bởi cây là bao nhiêu?

Bài tập Tổng hợp

TH 4.1 Cho X là thời gian cần thiết một đầu đọc/ghi định vị bản ghi mong muốn trên bộ nhớ đĩa máy tính khi đầu đã được định vị vị trí chính xác. Nếu các đĩa xoay một lần mỗi 25 mini giây, giả định X được phân bố đều trên khoảng $[0, 25]$.

- Tính $P(10 \leq X \leq 20)$; $P(X \geq 10)$.
- Xác định hàm phân phối xác suất của X .
- Tính $E(X)$ và σ_X .

TH 4.2 Tuổi thọ của một loại thiết bị là biến ngẫu nhiên X (đơn vị: năm) có hàm mật độ xác suất

$$f(x) = \begin{cases} \frac{k}{(x+4)^3} & x > 0 \\ 0 & \text{trường hợp khác} \end{cases}$$

- Xác định k .
- Tính xác suất tuổi thọ của thiết bị này từ 2 đến 5 năm.
- Tính xác suất tuổi thọ của thiết bị này vượt quá tuổi thọ trung bình của thiết bị.
- Tính xác suất tuổi thọ của thiết bị không vượt quá 1,5 tuổi thọ trung bình biết thiết bị đã sử dụng qua tuổi thọ trung bình.
- Nếu thiết bị đã qua sử dụng x (năm) nếu thanh lý được trả lại $h(x) = 100/(4+x)$ đôla. Hỏi giá trị thanh lý trung bình của loại thiết bị này là bao nhiêu?

TH 4.3 Điện áp của một điốt thuộc một loại nhất định được chọn ngẫu nhiên có phân phối với chuẩn với giá trị trung bình là 40 V và độ lệch chuẩn 1,5 V.

- Tính xác suất điện áp của một điốt từ 39 đến 42V.
- Xác định giá trị mà chỉ có 15% các điốt có điện áp vượt quá giá trị đó?
- Nếu bốn điốt được lựa chọn độc lập, xác suất mà ít nhất một có một điện áp vượt quá 42V là bao nhiêu?

TH 4.4 Kiểm tra các bo mạch được sử dụng trong quá trình sản xuất máy nghe đĩa compact thấy rằng có 5% có lỗi. Giả sử một lô có 250 bảng mạch với điều kiện hoạt động là độc lập.

a. Tính xác suất có ít nhất 10% số bảng trong lô bị lỗi. b. Tính xác suất có chính xác 10 bảng mạch trong lô bị lỗi.

TH 4.5 Kích thước hạt Al cho một hợp kim nhôm-Indium bao gồm 2% indi có phân phối chuẩn với giá trị trung bình 96 và độ lệch chuẩn 14.

- a. Tính xác suất kích thước hạt vượt quá 100.
- b. Tính xác suất kích thước hạt là từ 50 đến 80.
- c. Xác định khoảng (a, b) mà 90% của tất cả các kích thước hạt (5% dưới a và 5% ở trên b) thuộc vào.

TH 4.6 Thời gian phản ứng (tính bằng giây) đối với một kích thích nhất định là một biến ngẫu nhiên liên tục có hàm mật độ xác suất

$$f(x) = \begin{cases} \frac{k}{x^2} & 1 \leq x \leq 3 \\ 0 & \text{trường hợp khác} \end{cases}$$

- a. Xác định k từ đó xác định hàm phân phối xác suất của biến ngẫu nhiên này.
- b. Tính xác suất thời gian phản ứng nhiều nhất là 2,5 giây? từ 1,5 đến 2,5 giây.
- c. Tính thời gian phản ứng trung bình và độ lệch chuẩn.
- d. Nếu một cá nhân mất hơn 1,5 giây để phản ứng, một ánh sáng sẽ sáng và tiếp tục cho đến khi thêm một giây nữa hoặc cho đến khi người đó phản ứng lại (tùy điều kiện nào xảy ra trước). Xác định lượng thời gian trung bình mà ánh sáng vẫn sáng. [Gợi ý: Đặt $h(X)$ = thời gian ánh sáng bật như một hàm của thời gian phản ứng X.]

TH 4.7 Đặt X biểu thị nhiệt độ mà tại đó một phản ứng hóa học nhất định xảy ra. Giả sử rằng X có hàm mật độ xác suất

$$f(x) = \begin{cases} k(4 - x^2) & -1 \leq x \leq 2 \\ 0 & \text{trường hợp khác} \end{cases}$$

- a. Xác định k từ đó xác định hàm phân phối xác suất của X.
- b. 0 có phải là nhiệt độ trung vị mà tại đó phản ứng xảy ra không? Nếu không, nhiệt độ trung vị nhỏ hơn hoặc lớn hơn 0?
- c. Giả sử phản ứng này được thực hiện độc lập một lần trong 10 phòng thí

thí nghiệm khác nhau. Đặt Y là số phòng thí nghiệm trong số mười phòng thí nghiệm có nhiệt độ vượt quá 1. Xác định phân phối của Y . (Đưa ra tên và giá trị các tham số.)

TH 4.8 Giả sử mô hình phân phối chiều dài đường dẫn tự do của photon trong một số trường hợp nhất định là phân phối theo hàm mũ với tham số $\lambda = 0,93$.

- a. Xác định độ dài đường dẫn kỳ vọng và độ lệch chuẩn của độ dài đường dẫn.
- b. Tính xác suất chiều dài đường dẫn vượt quá 3,0. Tính xác suất mà độ dài đường dẫn là giữa 1,0 và 3,0.
- c. Xác định giá trị đường dẫn mà chỉ có 10% các độ dài đường dẫn bị vượt quá.

TH 4.9 Độ bền khi kéo (MPa) của vật liệu tổng hợp trong điều kiện quy định có thể được mô hình hóa bằng phân phối Weibull với $\alpha = 9$ và $\beta = 180$.

- a. Tính xác suất mẫu được chọn ngẫu nhiên có độ bền khi kéo vượt quá 175? từ 150 đến 175?
- b. Nếu hai mẫu được lựa chọn ngẫu nhiên và độc lập, thì xác suất có ít nhất một mẫu có cường độ từ 150 đến 175 là bao nhiêu?
- c. Xác định giá trị độ bền khi kéo phân chia 10% các mẫu vật yếu nhất khỏi 90% còn lại?

TH 4.10 Cho biến ngẫu nhiên X có hàm mật độ xác suất $f(x)$ và cần xác định hàm xác suất của $Y = h(X)$. Giả sử $h(x)$ là một hàm có hàm ngược $x = k(y)$. Khi đó hàm mật độ xác suất của Y là

$$g(y) = f[k(y)] \cdot |k'(y)|$$

- a. X có phân phối đều trên $[0; 1]$, xác định hàm mật độ của $Y = -\ln(X)$.
- a. X có phân phối chuẩn tắc, xác định hàm mật độ của $Y = \sigma X + \mu$.
- a. X có phân phối gamma với hai tham số α và β , xác định hàm mật độ xác suất của biến ngẫu nhiên $Y = cX$.

Chương 4

PHÂN PHỐI XÁC SUẤT ĐỒNG THỜI VÀ MẪU NGẪU NHIÊN

Giới thiệu

Trong chương 3 và 4, chúng ta đã phát triển các mô hình xác suất cho một biến ngẫu nhiên đơn biến. Nhiều vấn đề về xác suất thống kê liên quan đến các biến ngẫu nhiên đồng thời. Trong chương này, đầu tiên chúng ta thảo luận về các mô hình xác suất đồng thời của các biến ngẫu nhiên, đặc biệt nhấn mạnh vào trường hợp trong đó các biến số độc lập với nhau. Sau đó chúng ta nghiên cứu các giá trị kỳ vọng của các hàm số của các biến ngẫu nhiên, gồm hiệp phương sai và tương quan để đo mức độ liên kết giữa hai biến.

Ba phần cuối của chương xem xét các hàm của n biến ngẫu nhiên $X_1, X_2, \dots, X_n$, và tập trung đặc biệt vào trung bình của chúng $(X_1 + \dots + X_n)/n$. Một hàm như vậy không những là một biến ngẫu nhiên mà còn là một thống kê. Các phương pháp từ xác suất sẽ được sử dụng để lấy thông tin cho phân phối của một thống kê. Ta được kết quả đầu tiên là Định lý giới hạn trung tâm (Central Limit Theorem), cơ sở cho nhiều quy trình suy luận liên quan đến kích cỡ mẫu lớn.

4.1 Phân phối đồng thời của các biến ngẫu nhiên

Có rất nhiều tình huống thử nghiệm trong đó có nhiều hơn một biến ngẫu nhiên được người quan sát quan tâm. Trước tiên chúng ta xem xét phân phối xác suất đồng thời cho hai biến ngẫu nhiên rời rạc, sau đó cho hai biến liên tục và cuối

cùng cho nhiều hơn hai biến.

4.1.1 Hai biến ngẫu nhiên rời rạc

Hàm xác suất khối của một biến ngẫu nhiên rời rạc đơn X xác định xác suất khối được đặt trên mỗi giá trị có thể của X . Hàm xác suất khối đồng thời của hai biến rời rạc X và Y mô tả xác suất khối được đặt trên mỗi cặp giá trị (x, y) .

Định nghĩa 4.1.1. Cho X và Y là hai biến rời rạc được định nghĩa trên không gian mẫu Ω của một phép thử. Hàm xác suất khối $p(x, y)$ cho mỗi cặp số (x, y) được định nghĩa bởi

$$p(x, y) = P(X = x \text{ and } Y = y)$$

thỏa $p(x, y) \geq 0$ và $\sum_x \sum_y p(x, y) = 1$.

Ví dụ 4.1 Cho biến ngẫu nhiên rời rạc X là chiều cao sinh viên (mét) với tập giá trị 1,5 ; 1,6 ; 1,7, biến ngẫu nhiên Y là cân nặng của sinh viên (kg) với tập giá trị là 50; 60; 65 và bảng phân phối xác suất đồng thời của X và Y như sau:

		Y		
$p_{X,Y}$		50	60	65
X	1,5	0,1	0,2	0,1
	1,6	0,2	0,05	0,1
	1,7	0,05	0,1	0,1

Khi đó tính các xác suất

$$p(1, 5; 60) = P(X = 1, 5; Y = 60) = 0, 2$$

$$P(Y \geq 60) = p(1, 5; 60) + p(1, 6; 60) + p(1, 7; 60) + p(1, 5; 65) + p(1, 6; 65) + p(1, 7; 65) = 0, 2 + 0, 05 + 0, 1 + 0, 1 + 0, 1 + 0, 1 = 0, 65$$

$$P(X = 1, 6) = 0, 2 + 0, 05 + 0, 1 = 0, 35$$

Định nghĩa 4.1.2. Hàm xác suất khối biên duyên của X , được kí hiệu là p_X , được định nghĩa $p_X(x) = \sum_{y:p(x,y)>0} p(x, y)$ với mỗi giá trị có thể của x .

Tương tự, hàm xác suất khối biên duyên của Y là $p_Y(y) = \sum_{x:p(x,y)>0} p(x, y)$ với mỗi giá trị có thể của y .

Việc sử dụng từ "biên duyên" ở đây là kết quả của thực tế rằng nếu hàm xác suất khối đồng thời được biểu diễn trong một bảng chữ nhật, thì các tổng xác suất các hàng (cột) chính là xác suất khối biên duyên của X và các tổng xác suất của các cột (hàng) chính là xác suất khối biên duyên của Y .

Ví dụ 4.2 (sử dụng dữ liệu của ví dụ 5.1)

		Y		
		50	60	65
X	1,5	0,1	0,2	0,1
	1,6	0,2	0,05	0,1
	1,7	0,05	0,1	0,1

$p_X(1, 5) = p(1, 5; 50) + p(1, 5; 60) + p(1, 5; 65) = 0, 1 + 0, 2 + 0, 1 = 0, 4$
 $p_X(1, 6) = p(1, 6; 50) + p(1, 6; 60) + p(1, 6; 65) = 0, 2 + 0, 05 + 0, 1 = 0, 35$
 $p_X(1, 7) = p(1, 7; 50) + p(1, 7; 60) + p(1, 7; 65) = 0, 05 + 0, 1 + 0, 1 = 0, 25$
 Vậy hàm xác suất khối biên duyên của X là:

$$p_X(x) = \begin{cases} 0, 4 & \text{nếu } X = 1, 5 \\ 0, 35 & \text{nếu } X = 1, 6 \\ 0, 25 & \text{nếu } X = 1, 7 \end{cases}$$

Tương tự hãy tính hàm xác suất khối biên duyên cho Y ?

4.1.2 Hai biến ngẫu nhiên liên tục

Xác suất mà giá trị quan sát thấy của một biến ngẫu nhiên liên tục X nằm trong một tập hợp một chiều A (ví dụ một khoảng) có được bằng cách lấy tích phân hàm mật độ $f(x)$ trên tập A . Tương tự như vậy, xác suất mà cặp (X, Y) của các biến ngẫu nhiên liên tục rơi vào tập hai chiều A (ví dụ hình chữ nhật) có được bằng cách lấy tích phân một hàm gọi tên là *hàm mật độ đồng thời*.

Định nghĩa 4.1.3. Cho X và Y là các biến ngẫu nhiên liên tục. Một **hàm mật độ đồng thời** $f(x, y)$ cho hai biến này là một hàm thỏa $f(x, y) \geq 0$ và $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) dx dy = 1$. Khi đó với tập 2 chiều A ta có:

$$P[(X, Y) \in A] = \int \int_A f(x, y) dx dy$$

Đặc biệt, nếu A là hình chữ nhật 2 chiều $(x, y) : a \leq x \leq b, c \leq y \leq d$ thì

$$P[(X, Y) \in A] = P(a \leq x \leq b, c \leq y \leq d) = \int_a^b \int_c^d f(x, y) dx dy$$

Hàm mật độ biên duyên của mỗi biến có thể có được bằng cách tương tự như ta tìm cho trường hợp biến 2 chiều rời rạc. Hàm mật độ biên duyên của X tại giá trị x có được khi ta cố định x trong cặp (x, y) và lấy tích phân hàm mật độ đồng thời theo y . Còn nếu lấy tích phân hàm mật độ đồng thời theo x thì được hàm mật độ biên duyên theo Y .

Định nghĩa 4.1.4. Hàm mật độ xác suất biên duyên của X và Y , được kí hiệu $f_X(x)$ và $f_Y(y)$, được cho bởi công thức:

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f(x, y) dy \text{ với } -\infty < x < \infty \\ f_Y(y) &= \int_{-\infty}^{\infty} f(x, y) dx \text{ với } -\infty < y < \infty \end{aligned}$$

Ví dụ 4.3 Một ngân hàng mở hai loại dịch vụ A và B, ngẫu nhiên chọn 1 ngày, gọi X là tỉ lệ thời gian mà dịch vụ A được sử dụng, và Y là tỉ lệ thời gian dịch vụ B được dùng. Thì tập hợp (X, Y) là một hình chữ nhật $D = \{(x, y) : 0 \leq x \leq 1, 0 \leq y \leq 1\}$. Giả sử hàm mật độ đồng thời của (X, Y) cho bởi

$$f(x, y) = \begin{cases} \frac{6}{5}(x + y^2) & \text{với } 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0 & \text{nơi khác} \end{cases}$$

Chú ý rằng điều kiện của hàm mật độ đồng thời là $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy = 1$ và $f(x, y) \geq 0$ đã thỏa.

a/ Xác suất không dịch vụ nào bận hơn $1/4$ thời gian là:

$$P(0 \leq X \leq \frac{1}{4}, 0 \leq Y \leq \frac{1}{4}) = \int_0^{1/4} \int_0^{1/4} \frac{6}{5}(x + y^2) dx dy = \frac{6}{5} \int_0^{1/4} \int_0^{1/4} x dx dy + \frac{6}{5} \int_0^{1/4} \int_0^{1/4} y^2 dx dy = 0,0109$$

$$\text{b/ Tính } f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy = \int_0^1 \frac{6}{5}(x + y^2) dy = \frac{6}{5}x + \frac{2}{5}$$

Hàm mật độ biên duyên của X (chính là phân phối xác suất của thời gian bận của dịch vụ A mà không quan tâm đến dịch vụ B) là:

$$f_X(x) = \begin{cases} \frac{6}{5}x + \frac{2}{5} & \text{với } 0 \leq x \leq 1 \\ 0 & \text{nơi khác} \end{cases}$$

c/ Tìm hàm mật độ biên duyên của Y ? Tính $P(\frac{1}{4} \leq Y \leq \frac{3}{4})$?

Ví dụ 4.4 Một công ty hạt có thị trường hộp các loại hạt pha trộn chứa hạnh nhân, hạt điều và đậu phộng. Giả sử trọng lượng tịnh của mỗi hộp có thể là 1 kg, nhưng trọng lượng của từng loại hạt là ngẫu nhiên. Bởi vì ba trọng số cộng lại bằng 1, một mô hình xác suất đồng thời cho hai loại hạt sẽ cung cấp thông tin cần thiết về trọng lượng cho loại hạt thứ ba. Đặt X là trọng lượng của hạnh nhân trong một chiếc hộp đã chọn và Y là trọng lượng hạt điều. Thì vùng để hàm mật độ dương là $D = (x, y) : 0 \leq x \leq 1, 0 \leq y \leq 1, x + y \leq 1$. Cho hàm mật độ đồng thời có dạng

$$f(x, y) = \begin{cases} kxy & \text{với } 0 \leq x \leq 1, 0 \leq y \leq 1, x + y \leq 1 \\ 0 & \text{nơi khác} \end{cases}$$

a/ Tính k ? (đs: 24)

b/ Tính xác suất mà hai loại hạt bất kì chiếm 50 phần trăm ? (Gợi ý: $A = \{(x, y) : 0 \leq x \leq 1, 0 \leq y \leq 1, x + y \leq 0,5\}$) (đs: 0,0625)

4.1.3 Các biến ngẫu nhiên độc lập

Trong Chương 2, chúng ta đã chỉ ra một cách xác định tính độc lập của hai biến cố là thông qua điều kiện $P(A \cap B) = P(A).P(B)$. Đây là một định nghĩa tương tự cho sự độc lập của hai biến ngẫu nhiên.

Định nghĩa 4.1.5. Hai biến ngẫu nhiên X và Y là độc lập nếu mỗi cặp giá trị x và y thỏa

$$p(x, y) = p_X(x).p_Y(y) \text{ khi } X \text{ và } Y \text{ rời rạc}$$

hoặc

$$f(x, y) = f_X(x).f_Y(y) \text{ khi } X \text{ và } Y \text{ liên tục.}$$

Nếu hai công thức trên không thỏa cho tất cả (x, y) thì X và Y được gọi là phụ thuộc.

Ví dụ 4.5 Trong ví dụ 5.1 cho biến rời rạc, tồn tại $(p(1, 6; 50) = 0,2) \neq (p(1, 6) \times p(50) = 0.35 \times 0.35)$ nên X và Y không độc lập.

Ví dụ 4.6 Trong ví dụ 5.3 ; tồn tại $f_X(\frac{3}{4}) = f_Y(\frac{3}{4}) = \frac{9}{16}; f(\frac{3}{4}, \frac{3}{4}) = 0 \neq \frac{9}{16} \cdot \frac{9}{16}$ nên hai biến trên không độc lập.

Sự độc lập của hai biến ngẫu nhiên khá quan trọng khi mô tả thí nghiệm nghiên cứu khi mà X và Y không ảnh hưởng lẫn nhau. Ngoài ra ta còn có kết quả sau:
 $P(a \leq X \leq b, c \leq Y \leq d) = P(a \leq X \leq b).P(c \leq Y \leq d)$

4.1.4 Nhiều hơn 2 biến

Để mô hình hoá các hàm của hơn hai biến ngẫu nhiên, chúng ta mở rộng khái niệm về phân phối đồng thời của hai biến.

Định nghĩa 4.1.6. Nếu $X_1, X_2, \dots, X_n$ là tất cả biến ngẫu nhiên rời rạc, hàm xác suất đồng thời của các biến ngẫu nhiên là :

$$p(x_1, x_2, \dots, x_n) = P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)$$

Nếu các biến liên tục, hàm mật độ đồng thời của $X_1, X_2, \dots, X_n$ là hàm $p(x_1, x_2, \dots, x_n)$ chẳng hạn với bất kì n khoảng $[a_1, b_1], \dots, [a_n, b_n]$,

$$P(a_1 \leq X_1 \leq b_1, \dots, a_n \leq X_n \leq b_n) = \int_{a_1}^{b_1} \dots \int_{a_n}^{b_n} f(x_1, \dots, x_n) dx_n \dots dx_1$$

Trong một phép thử nhị thức, mỗi thí nghiệm chỉ cho một kết quả trong 2 kết quả có thể xảy ra. Bây giờ xét một phép thử gồm n thí nghiệm độc lập và giống nhau, mà trong đó mỗi phép thử chỉ cho một kết quả trong r kết quả có thể xảy ra. Cho $p_i = P$ (kết quả thứ i của bất kì phép thử cụ thể), và định nghĩa biến ngẫu nhiên X_i = số lần thí nghiệm mà có kết quả i ($i = 1, \dots, r$). Một thí nghiệm như vậy được gọi là **phép thử đa thức - multinomial experiment** và hàm xác suất khối của $X_1, X_2, \dots, X_n$ được gọi là **phân phối đa thức - multinomial distribution**. Bằng cách sử dụng lý luận tính tương tự với một biến trong trường hợp phân phối nhị thức, hàm xác suất đồng thời của $X_1, X_2, \dots, X_n$ là

$$p(x_1, \dots, x_r) = \begin{cases} \frac{n!}{(x_1!)(x_2!)\dots(x_r!)} \cdot p_1^{x_1} \dots p_r^{x_r}, & x_i = 0, 1, 2, \dots; x_1 + \dots + x_r = n \\ 0, & \text{nơi khác} \end{cases}$$

Trường hợp $r = 2$ sẽ cho phân phối nhị thức, với X_1 = số lần thành công và $X_2 = n - X_1$ = số lần thất bại.

Định nghĩa 4.1.7. Biến ngẫu nhiên $X_1, X_2, \dots, X_n$ được gọi là **độc lập** nếu với mỗi tập con $X_{i_1}, X_{i_2}, \dots, X_{i_k}$ của các biến (mỗi cặp, mỗi bộ ba, ...) thì hàm xác suất đồng thời hay hàm mật độ đồng thời bằng với tích của các hàm xác suất biên duyên hay các hàm mật độ biên duyên..

Vì vậy, nếu các biến độc lập với $n = 4$, thì hàm xác suất đồng thời hay hàm mật độ đồng thời của 2 biến là tích của hai hàm biên duyên, tương tự cho trường hợp 3 và 4 biến. Quan trọng nhất là khi ta đã có n biến độc lập thì hàm xác suất đồng thời hay hàm mật độ đồng thời là tích của n hàm biên duyên.

4.1.5 Phân phối điều kiện

Định nghĩa 4.1.8. Cho X và Y là 2 biến ngẫu nhiên liên tục với hàm mật độ biên duyên $f(x, y)$ và hàm mật độ biên duyên của X là $f_X(x)$. Thì với bất kì X có giá trị x mà $f_X(x) > 0$, hàm mật độ điều kiện của Y cho bởi $X=x$ là

$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)} \quad -\infty < y < \infty$$

Nếu X và Y là biến rời rạc, thay hàm mật độ bằng hàm xác suất khối trong định nghĩa này thì được hàm xác suất khối điều kiện của Y khi $X=x$.

Ví dụ 4.7 Dựa vào dữ liệu của ví dụ 5.3.

a/ Hàm mật độ điều kiện của Y cho bởi $X = 0.8$ là

$$f_{Y|X}(y|0,8) = \frac{f(0,8;y)}{f_X(0,8)} = \frac{1,2(0,8+y^2)}{1,2(0,8)+0,4} = \frac{1}{34}(24 + 30y^2) \text{ với } 0 < y < 1$$

b/ Xác suất để dịch vụ A bận trong nửa thời gian đầu khi biết $X = 0.8$

$$P(Y \leq 0,5|X = 0,8) = \int_{-\infty}^{0,5} f_{Y|X}(y|0,8)dy = \int_0^{0,5} \frac{1}{34}(24 + 30y^2)dy = 0,39$$

c/ Kỳ vọng có điều kiện $X = 0,8$ là:

$$E(Y|X = 0,8) = \int_{-\infty}^{\infty} y \cdot f_{Y|X}(y|0,8)dy = \frac{1}{34} \int_0^1 y(24 + 30y^2)dy = 0.574$$

4.2 Giá trị Kỳ vọng, Phương sai và Hiệp phương sai

Bất kỳ hàm $h(X)$ của một biến ngẫu nhiên đơn thì chính nó cũng là một biến ngẫu nhiên. Tuy nhiên, để tính toán $E[h(X)]$ thì không cần phải có phân phối xác suất của $h(X)$; thay vào đó, $E[h(X)]$ được tính như một trung bình trọng số (a weighted average) của các giá trị $h(X)$, trong đó hàm trọng số (the weight function) là hàm xác suất khối $p(x)$ hoặc hàm mật độ $f(x)$ của X . Kết quả tương tự cho một hàm phân phối đồng thời 2 biến $h(x, y)$.

Mệnh đề 4.2.1. Cho X và Y là biến ngẫu nhiên đồng thời với hàm xác suất khối $p(x, y)$ hoặc hàm mật độ $f(x, y)$ tương ứng với trường hợp biến rời rạc hoặc biến liên tục. Thì giá trị kỳ vọng của hàm $h(X, Y)$ được kí hiệu là $E[h(X, Y)]$ hoặc $\mu_{h(X,Y)}$ được tính như sau:

$$E[h(X, Y)] = \begin{cases} \sum_x \sum_y h(x, y) \cdot p(x, y), & \text{nếu } X \text{ và } Y \text{ rời rạc} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x, y) \cdot f(x, y) dx dy, & \text{nếu } X \text{ và } Y \text{ liên tục} \end{cases}$$

Ví dụ 4.8 Năm người bạn đã mua vé cho một buổi hòa nhạc. Nếu vé là chỗ ngồi từ 1-5 trong một hàng nào đó và vé được phân phối ngẫu nhiên cho năm người, thì kì vọng cho số ghế khi chia cho 2 người là bao nhiêu? Đặt X và Y kí hiệu cho số ghế của người đầu tiên và người thứ 2. Cặp (X, Y) có thể có các giá trị $(1, 2), (1, 3), \dots, (5, 4)$, và hàm xác suất khối đồng thời của (X, Y) là

$$p(x, y) = \begin{cases} \frac{1}{20} & x = 1, \dots, 5; y = 1, \dots, 5; x \neq y \\ 0 & \text{nơi khác} \end{cases}$$

Lúc đó số chỗ ngồi chia cho 2 người là hàm $h(X, Y) = |X - Y| - 1$, ta có bảng tương ứng sau:

		x				
$h_{X,Y}$		1	2	3	4	5
y	1	—	0	1	2	3
	2	0	—	0	1	2
	3	1	0	—	0	1
	4	2	1	0	—	0
	5	3	2	1	0	—

Vì vậy $E[h(X, Y)] = \sum_{(x, y)} h(x, y) \cdot p(x, y) = \sum_{x=1}^5 \sum_{y=1, x \neq y}^5 (|x - y| - 1) \cdot \frac{1}{20} = 1$

Ví dụ 4.9 Hàm mật độ đồng thời của X (lượng hạnh nhân) và Y (lượng hạt điều) trong 1kg hạt là:

$$f(x, y) = \begin{cases} 24xy & \text{với } 0 \leq x \leq 1, 0 \leq y \leq 1, x + y \leq 1 \\ 0 & \text{nơi khác} \end{cases}$$

Nếu 1kg hạt hạnh nhân có giá 1\$; 1 kg hạt điều có giá 1,5\$; 1 kg hạt đậu phộng có giá 0.5\$; thì tổng giá trị của 1 hộp là:

$$h(X, Y) = 1.X + 1,5.Y + 0,5.(1 - X - Y) = 0,5 + 0,5.X + Y$$

Kì vọng cho tổng giá trị là:

$$E[h(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x, y) \cdot f(x, y) dx dy = \int_0^1 \int_0^{1-x} (0,5 + 0,5x + y) \cdot 24xy dy dx = 1,1\$$$

4.2.1 Hiệp phương sai

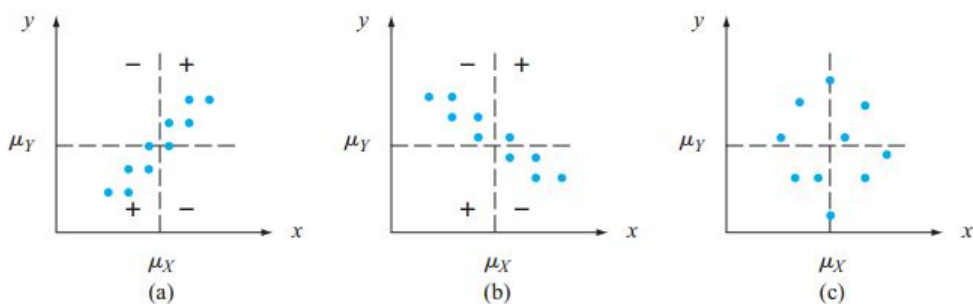
Định nghĩa 4.2.2. Hiệp phương sai giữa hai biến ngẫu nhiên X và Y là

$$Cov(X, Y) = E[(X - \mu_X)(Y - \mu_Y)] = \begin{cases} \sum_x \sum_y (x - \mu_X)(y - \mu_Y)p(x, y) & , \quad X, Y \text{ rời rạc} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_X)(y - \mu_Y)f(x, y)dxdy, & X, Y \text{ liên tục} \end{cases}$$

Đó là, vì $X - \mu_X$ và $Y - \mu_Y$ là độ lệch của hai biến từ các giá trị trung bình tương ứng, hiệp phương sai là kì vọng của tích hai độ lệch. Chú ý rằng $Cov(X, X) = E[(X - \mu_X)^2] = V(X)$.

Giải thích định nghĩa: giả sử X và Y có một mối liên hệ dương với nhau, nghĩa là các giá trị lớn của X có khuynh xảy ra với các giá trị lớn của Y và các giá trị nhỏ của X với giá trị nhỏ của Y . Thì hầu như xác suất khối hay hàm mật độ sẽ được liên quan đến $x - \mu_X$ và $y - \mu_Y$, khi cả hai giá trị X, Y đều âm hay cả hai đều dương. Do đó với liên kết dương mạnh, $Cov(X, Y)$ sẽ hầu như dương (quite positive). Đối với một liên kết âm mạnh dấu của $X - \mu_X$ và $Y - \mu_Y$ sẽ có xu hướng ngược nhau, lúc đó thì $Cov(X, Y)$ hầu như âm. Nếu X, Y không có liên kết mạnh, tích âm và dương có xu hướng triệt tiêu lẫn nhau thì $Cov(X, Y) = 0$.

Hình dưới chỉ ra những khả năng khác nhau. Hiệp phương sai phụ thuộc vào các tập kết quả và xác suất. Hiệp phương sai có thể được thay đổi mà không cần điều chỉnh những tập kết quả, và có thể làm thay đổi nhiều đến giá trị của $Cov(X, Y)$.



$p(x, y) = 1/10$ for each of ten pairs corresponding to indicated points:
(a) positive covariance; (b) negative covariance; (c) covariance near zero

Ví dụ 4.10 Cho hàm xác suất khối và hàm phân phối biên duyên của hai biến X, Y như sau:

		y					y				
		$p_{X,Y}$	0	100	200			$p_Y(y)$	0	100	200
x	100	0,2	0,1	0,2	$p_X(x)$	100	250	0,25	0,25	0,5	
	250	0,05	0,15	0,3							

Áp dụng công thức tính $\text{Cov}(X, Y)$? (Đs: 1875)

Mệnh đề 4.2.3. $\text{Cov}(X, Y) = E(X, Y) - \mu_X \cdot \mu_Y$

Ví dụ 4.11

$$f(x, y) = \begin{cases} 24xy & \text{với } 0 \leq x \leq 1, 0 \leq y \leq 1, x + y \leq 1 \\ 0 & \text{nơi khác} \end{cases}$$

$$f_X(x) = \begin{cases} 12x(1 - x^2) & \text{với } 0 \leq x \leq 1 \\ 0 & \text{nơi khác} \end{cases}$$

với $f_Y(y)$ cũng có được tương tự bằng cách thay x bởi y trong $f_X(x)$.

Hãy tính μ_X , μ_Y , $E(X, Y)$ từ đó suy ra $\text{Cov}(X, Y)$? (đs: -2/75)

4.2.2 Tương quan

Định nghĩa 4.2.4. Hệ số tương quan của X và Y , được kí hiệu là $\text{Corr}(X, Y)$, $\rho_{X,Y}$ hay chỉ là ρ và được định nghĩa như sau

$$\rho_{X,Y} = \frac{\text{Cov}(X,Y)}{\sigma_X \cdot \sigma_Y}$$

Ví dụ 4.12 Từ dữ kiện của ví dụ 5.10. Hãy tính hệ số tương quan của X và Y ? (Đs: 0,301)

Mệnh đề 4.2.5.

1. Nếu a và c cùng âm hoặc cùng dương thì

$$\text{Corr}(aX + b, cY + d) = \text{Corr}(X, Y)$$

2. Với bất kì hai biến ngẫu nhiên X và Y thì $-1 \leq \text{Corr}(X, Y) \leq 1$

Mệnh đề 4.2.6.

1. Nếu X và Y độc lập, thì $\rho = 0$ nhưng $\rho = 0$ thì không suy ra được là có sự độc lập.
2. $\rho = 1$ hay -1 khi và chỉ khi $Y = aX + b$ với a, b là số nào đó và $a \neq 0$.

Mệnh đề này chỉ ra rằng ρ là một thước đo về bậc của mối quan hệ **tuyến tính** giữa X và Y , và chỉ khi hai biến này quan hệ tuyến tính thì ρ có thể âm hoặc dương tùy ý. Giá trị tuyệt đối của $\rho < 1$ chỉ ra rằng mối quan hệ không tuyến tính hoàn toàn, nhưng vẫn có mối quan hệ phi tuyến tính trong đó. Cũng như vậy khi $\rho = 0$ thì không chỉ ra rằng X và Y độc lập, mà chỉ là X và Y không có quan hệ tuyến tính. Khi $\rho = 0$, X và Y được gọi là **không tương quan**. Hai biến có thể không tương quan nhưng chưa chắc phụ thuộc vì có quan hệ phi tuyến tính mạnh ở đây, vì vậy cần cẩn thận đưa ra kết luận khi gặp $\rho = 0$.

4.3 Các phân phối thống kê

Các quan sát trong một mẫu đơn được kí hiệu trong chương 1 là $x_1, x_2, \dots, x_n$. Xem như chọn hai mẫu khác nhau có kích thước n từ cùng một phân phối tổng thể. x_i trong mẫu thứ hai sẽ hầu như luôn khác nhau một chút với các giá trị của mẫu đầu tiên. Ví dụ, một mẫu đầu tiên của $n = 3$ xe ô tô loại đặc biệt có thể cho hiệu quả nhiên liệu $x_1 = 30,7, x_2 = 29,4, x_3 = 31,1$, trong khi một mẫu thứ hai có thể cho $x_1 = 28,8, x_2 = 30,0$ và $x_3 = 32,5$. Trước khi chúng ta thu được dữ liệu, có sự không chắc chắn về giá trị của mỗi x_i . Vì điều này không chắc chắn, trước khi thu thập dữ liệu, chúng ta xem mỗi quan sát như là một biến ngẫu nhiên và kí hiệu mẫu bằng $X_1, X_2, \dots, X_n$ (chữ in hoa cho biến ngẫu nhiên).

Sự thay đổi này trong các giá trị được quan sát giúp suy ra các giá trị của bất kỳ hàm nào của các quan sát mẫu - chẳng hạn như trung bình mẫu, độ lệch tiêu chuẩn mẫu. Trước tiên là để có $x_1, \dots, x_n$, thì có một tất định cho giá trị $\bar{x}$, giá trị $s \dots$

Định nghĩa 4.3.1. Một **thống kê** là bất kỳ lượng nào mà giá trị có thể được tính từ dữ liệu mẫu. Trước khi thu thập dữ liệu, có một tất định về giá trị của bất kỳ thống kê đặc biệt sẽ cho kết quả. Do đó, thống kê là một biến ngẫu nhiên và sẽ được kí hiệu bởi một kí tự viết hoa; một kí tự viết thường được sử dụng để đại diện cho giá trị đã tính toán hoặc đã quan sát của thống kê.

Vì vậy trung bình mẫu trong thống kê được kí hiệu là $\bar{X}$, và giá trị tính được của thống kê này là $\bar{x}$. Tương tự S đại diện cho độ lệch chuẩn mẫu và giá trị được tính kí hiệu là s .

Bất kỳ thống kê nào (bản thân nó là một biến ngẫu nhiên) đều có một phân phối xác suất. Đáng chú ý là phân phối xác suất của trung bình mẫu $\bar{X}$.

Phân phối xác suất của một thống kê đôi khi gọi là phân phối mẫu để nhấn mạnh rằng nó mô tả cách mà thống kê khác nhau về giá trị trên tất cả các mẫu được lựa chọn.

4.3.1 Mẫu ngẫu nhiên

Phân bố xác suất của bất kỳ thống kê cụ thể phụ thuộc không chỉ vào phân phối tổng thể (chuẩn, đều,...) và kích thước mẫu n mà còn phụ thuộc phương pháp lấy mẫu. Xem xét lựa chọn một mẫu kích thước $n = 2$ từ tổng thể chỉ gồm ba giá trị 1, 5 và 10, và giả sử rằng thống kê được quan tâm là phương sai mẫu. Nếu lấy mẫu được thực hiện "có hoàn lại," thì $S^2 = 0$ khi $X_1 = X_2$. Tuy nhiên, S^2 không thể bằng 0 nếu mẫu "không hoàn lại". Vì vậy, $P(S^2 = 0) = 0$ với một phương pháp lấy mẫu, và xác suất này là dương với phương pháp lấy mẫu khác. Định nghĩa tiếp theo của chúng ta mô tả một phương pháp lấy mẫu thường gặp trong thực tế.

Định nghĩa 4.3.2. Các biến ngẫu nhiên $X_1, X_2, \dots, X_n$ được gọi là mẫu ngẫu nhiên (đơn) cỡ mẫu n nếu:

1. Các X_i là biến ngẫu nhiên độc lập.
2. Các X_i có cùng phân phối mẫu.

Nếu lấy mẫu có hoàn lại từ tổng thể vô hạn, thì điều kiện 1 và 2 được thỏa mãn. Những điều kiện này cũng gần như thỏa mãn nếu lấy mẫu không hoàn lại, tuy nhiên mẫu cỡ n thì nhỏ hơn nhiều so với tổng thể cỡ N . Trong thực tế, nếu $n/N < 0.05$ (nhiều nhất 5% tổng thể được lấy mẫu), chúng ta có thể tiến hành khi X_i là biến ngẫu nhiên. Hiệu quả của phương pháp lấy mẫu này là phân phối xác suất của bất kỳ thống kê nào có thể dễ dàng thu được hơn so với bất kỳ phương pháp lấy mẫu nào khác.

Có hai phương pháp chung để thu thập thông tin về phân phối mẫu của một thống kê. Một phương pháp liên quan đến tính toán dựa trên các quy tắc xác suất, và thứ hai là thực hiện một thí nghiệm mô phỏng.

4.3.2 Tìm phân phối mẫu

Các quy tắc xác suất có thể được sử dụng để có được phân phối của một thống kê miễn là nó là một hàm "khá đơn giản" của các biến ngẫu nhiên X_i và hoặc là có vài giá trị X khác nhau trong tổng thể hoặc phân phối xác suất có dạng "tốt".

4.3.3 Những thí nghiệm mô phỏng

Phương pháp thứ hai để thu thập thông tin về phân phối mẫu của thống kê là thực hiện một thí nghiệm mô phỏng. Phương pháp này thường được sử dụng khi lấy đạo hàm thông qua các quy tắc xác suất là quá khó hoặc phức tạp để thực hiện. Một thí nghiệm như vậy luôn luôn được thực hiện với sự trợ giúp của máy tính. Các đặc điểm sau của thí nghiệm yêu cầu:

1. Thống kê của lãi suất $(\bar{X}, S, \text{ một giá trị trung bình đã được cắt xén (trimmed),...})$
2. Phân phối của tổng thể (chuẩn với $\mu = 100, \sigma = 15$; phân phối đều với giới hạn dưới $A = 5$ và giới hạn trên $B = 10$, ...)
3. Kích cỡ mẫu n (ví dụ, $n = 10$ hoặc $n = 50$)
4. Số lần lặp lại k (số mẫu cần lấy)

Sau đó sử dụng phần mềm thích hợp để lấy k mẫu ngẫu nhiên khác nhau, mỗi kích thước n , từ phân phối tổng thể đã được thiết kế. Đối với mỗi mẫu, tính toán giá trị của thống kê và xây dựng một biểu đồ tần số của các giá trị k . Biểu đồ này cung cấp một xấp xỉ phân phối mẫu của thống kê. Giá trị càng lớn hơn của k , thì xấp xỉ càng tốt hơn (phân phối mẫu thật sự sẽ xuất hiện khi $k \rightarrow \infty$). Trong thực hành, $k = 500$ hoặc 1000 thường là đủ nếu thống kê "khá đơn giản".

4.4 Phân phối của trung bình mẫu

Tầm quan trọng của trung bình mẫu $\bar{X}$ là để rút ra kết luận cho trung bình tổng thể μ . Một số phương pháp suy luận dựa trên tính chất của phân phối mẫu $\bar{X}$. Các tính chất tính toán và thí nghiệm mô phỏng đã nêu trong các phần trước khi ta đề cập đến mối quan hệ $E(\bar{X})$ và μ hay $V(\bar{X}), \sigma^2$, và n

Mệnh đề 4.4.1. Cho $X_1, X_2, \dots, X_n$ là một mẫu ngẫu nhiên từ một phân phối với trung bình μ và độ lệch chuẩn σ . Thì

1. $E(\bar{X}) = \mu_{\bar{X}} = \mu$
2. $V(\bar{X}) = \sigma_{\bar{X}}^2 = \sigma^2/n$ và $\sigma_{\bar{X}} = \sigma/\sqrt{n}$

Thêm nữa, với $T_o = X_1 + \dots + X_n$ (tổng của mẫu), $E(T_o) = n\mu, V(T_o) = n\sigma^2, \sigma_{T_o} = \sqrt{n}\sigma$

Các chứng minh cho những kết quả trên sẽ được trình bày ở phần tiếp theo. Theo kết quả 1, phân phối mẫu của $\bar{X}$ (tức là xác suất) tập trung chính xác ở trung bình của tổng thể mà mẫu được chọn từ đó. Kết quả 2 cho thấy rằng phân phối $\bar{X}$ trở nên tập trung hơn về μ khi kích thước mẫu n tăng lên. Sự khác biệt quan trọng là phân phối của T_o trở nên lan rộng ra khi cỡ mẫu n tăng lên.

Ví dụ 4.13 Trong một thí nghiệm kiểm tra sức chịu mài mòn trên một mẫu titan, kì vọng số chu kỳ để phát ra âm thanh đầu tiên (dùng chỉ ra vết nứt đầu tiên) là $\mu = 28.000$ và độ lệch tiêu chuẩn của số chu kỳ là $\sigma = 5000$. Đặt $X_1, X_2, \dots, X_{25}$ là một mẫu ngẫu nhiên có kích thước 25, trong đó mỗi X_i là số chu kỳ trên một mẫu titan khác nhau ngẫu nhiên được chọn. Thì giá trị kì vọng của số trung bình mẫu của chu kỳ cho đến khi phát âm thanh đầu tiên là $E(\bar{X}) = \mu = 28.000$ và tổng kì vọng cho số chu kỳ cho 25 mẫu titan là $E(T_o) = n\mu = 25(28000) = 700000$. Độ lệch chuẩn của $\bar{X}$ (sai số tiêu chuẩn của trung bình) và của T_o là:

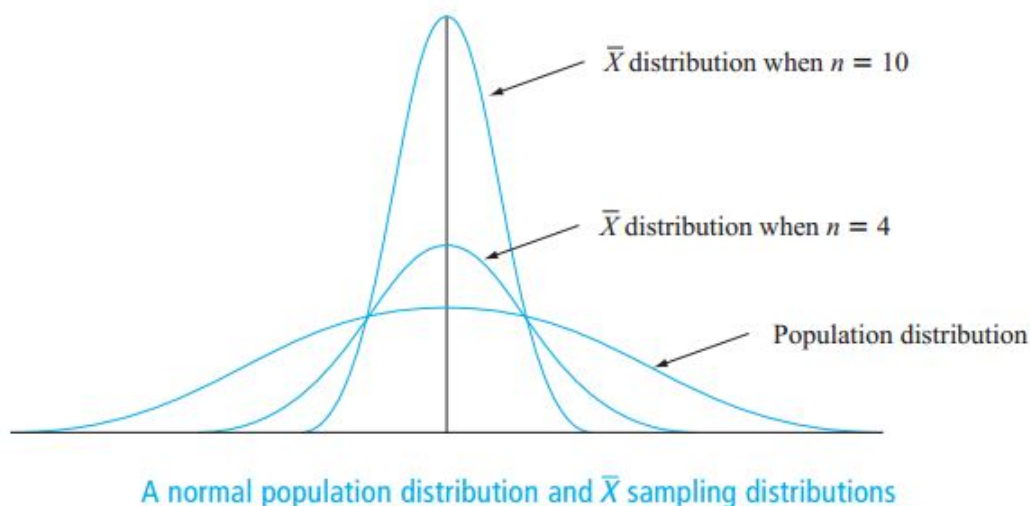
$$\begin{aligned}\sigma_{\bar{X}} &= \sigma/\sqrt{n} = \frac{5000}{\sqrt{25}} = 1000 \\ \sigma_{T_o} &= \sqrt{n}\sigma = \sqrt{25} \cdot 5000 = 25000\end{aligned}$$

Nếu cỡ mẫu tăng lên $n = 100$, $E(\bar{X})$ thì không đổi, nhưng $\sigma_{\bar{X}} = 500$ (bằng một nửa giá trị trước đó) (kích cỡ mẫu phải được tăng gấp bốn lần để giảm một nửa độ lệch tiêu chuẩn của $\bar{X}$).

4.4.1 Trường hợp phân phối chuẩn tổng thể

Mệnh đề 4.4.2. Cho $X_1, X_2, \dots, X_n$ là một mẫu ngẫu nhiên từ phân phối chuẩn với trung bình μ và độ lệch chuẩn σ . Thì với mọi n , $\bar{X}$ được phân phối chuẩn (với trung bình μ và độ lệch chuẩn $\sigma/\sqrt{n}$), như là T_o (với trung bình $n\mu$ và độ lệch chuẩn $\sqrt{n}\sigma$).

Chúng ta có dữ kiện để biết phân phối $\bar{X}$ và T_o khi phân phối tổng thể là phân phối chuẩn. Đặc biệt, xác suất như $P(a \leq \bar{X} \leq b)$ và $P(c \leq T_o \leq d)$ có thể tính được bằng cách chuẩn hóa. Hình dưới giải thích cho mệnh đề



Ví dụ 4.14 Thời gian cho một con chuột của một phân loài nhất định tìm đường ra một mê cung là một phân phối chuẩn với $\mu = 1,5$ phút và $\sigma = 0,35$ phút. Giả sử năm con chuột được chọn để quan sát. Đặt $X_1, \dots, X_5$ cho thời gian của chúng trong mê cung.

a/ Giả sử X_i là một mẫu ngẫu nhiên từ phân bố chuẩn này, tìm xác suất mà tổng thời gian $T_0 = X_1 + \dots + X_5$ nằm trong khoảng từ 6 đến 8 phút?

Theo mệnh đề, T_0 có phân phối chuẩn với $\mu_{T_0} = n\mu = 5 \cdot (1,5) = 7,5$ và phương sai $\sigma_{T_0}^2 = n\sigma^2 = 5(0,1225) = 0,6125$, vì vậy $\sigma_{T_0} = 0,783$.

$$P(6 \leq T_0 \leq 8) = P\left(\frac{6-7,5}{0,783} \leq Z \leq \frac{8-7,5}{0,783}\right) = P(-1,92 \leq Z \leq 0,64) = \Phi(0,64) - \Phi(-1,92) = 0,7115$$

b/ Xác định xác suất mà thời gian trung bình mẫu $\bar{X}$ nhiều nhất là 2 phút, với $\mu_{\bar{X}} = \mu = 1,5$ và $\sigma_{\bar{X}} = \sigma/\sqrt{n} = 0,35/\sqrt{5} = 0,1565$. Thì

$$P(\bar{X} \leq 2) = P\left(Z \leq \frac{2-1,5}{0,1565}\right) = P(Z \leq 3,19) = \Phi(3,19) = 0,9993$$

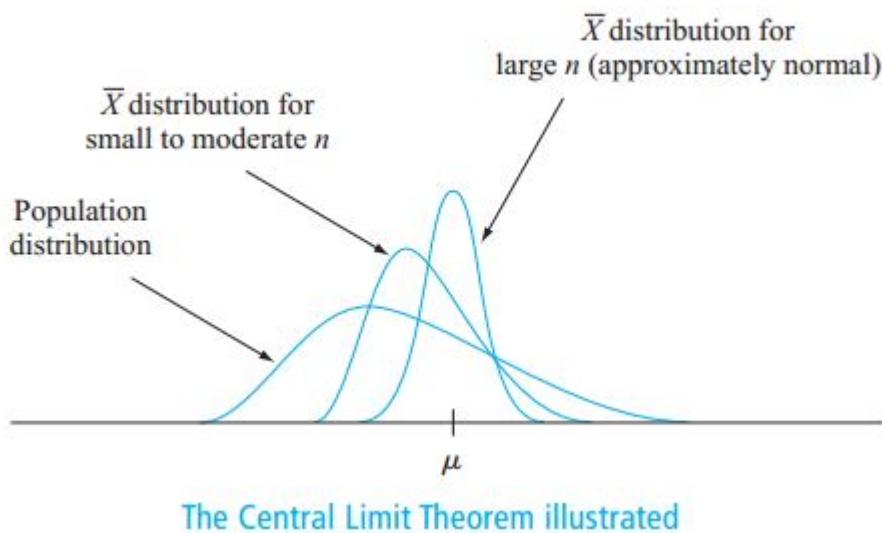
4.4.2 Định lý giới hạn trung tâm

Khi các X_i có phân phối chuẩn, thì $\bar{X}$ cũng có phân phối chuẩn cho mỗi mẫu cỡ n . Các kết quả từ ví dụ trên chỉ ra rằng khi phân phối tổng thể không chuẩn, lấy trung bình sẽ cho một phân phối tạo ra hình chuông hơn là mẫu được lấy. Nếu n lớn, một đường cong chuẩn thích hợp sẽ xấp xỉ gần đúng phân phối thật của $\bar{X}$.

Phát biểu cho ý này được trình bày trong định lý xác suất quan trọng sau đây.

Định lý 4.4.3. Cho $X_1, X_2, \dots, X_n$ là một mẫu ngẫu nhiên từ phân phối với trung bình μ và phương sai σ^2 . Thì nếu n đủ lớn, thì $\bar{X}$ có xấp xỉ gần đúng một phân phối chuẩn với $\mu_{\bar{X}} = \mu$ và $\sigma_{\bar{X}}^2 = \sigma^2/n$ và T_o cũng có xấp xỉ gần đúng một phân phối chuẩn với $\mu_{T_o} = n\mu, \sigma_{T_o}^2 = n\sigma^2$. Giá trị n càng lớn thì xấp xỉ càng tốt.

Hình dưới minh họa Định lý Giới hạn Trung tâm. Theo ĐLGHTT, khi n lớn và ta muốn tính xác suất như $P(a \leq \bar{X} \leq b)$, ta chỉ cần "giả vờ" là $\bar{X}$ là chuẩn và chuẩn hóa nó, sử dụng bảng phân phối chuẩn. Kết quả sẽ được xấp xỉ đúng.



Ví dụ 4.15 Lượng tạp chất cụ thể trong một lô sản phẩm hoá học nhất định là biến ngẫu nhiên có giá trị trung bình 4,0 g và độ lệch chuẩn 1,5 g. Nếu 50 lô được chuẩn bị độc lập, tính xác suất để trung bình mẫu tạp chất $\bar{X}$ nằm giữa 3,5 và 3,8g? Ta có $n = 50$ đủ lớn để áp dụng ĐLGHTT. $\bar{X}$ xấp xỉ về phân phối chuẩn với trung bình $\mu_{\bar{X}} = 4$ và $\sigma_{\bar{X}} = 1,5/\sqrt{50} = 0,2121$ vì vậy

$$P(3,5 \leq \bar{X} \leq 3,8) \approx P\left(\frac{3,5-4}{0,2121} \leq Z \leq \frac{3,8-4}{0,2121}\right) = \Phi(-0,94) - \Phi(-2,36) = 0,1645$$

Ví dụ 4.16 Một tổ chức tiêu dùng nhất định thường báo cáo số lượng khiếm khuyết lớn của mỗi chiếc ô tô mới mà nó thử nghiệm. Giả sử số lỗi này là một biến ngẫu nhiên với giá trị trung bình là 3.2 và độ lệch chuẩn 2.4. Trong số 100 xe được lựa chọn ngẫu nhiên, tính xác suất trung bình mẫu của những lỗi này nhiều hơn 4? (đs: 0,0004)

Định lý giới hạn trung tâm cung cấp cái nhìn sâu sắc về lý do tại sao nhiều biến ngẫu nhiên có phân phối xác suất gần với phân phối chuẩn. Ví dụ, sai số đo lường trong một thí nghiệm khoa học có thể được coi là tổng của một số điểm nhiễu loạn đáng kể và sai số của độ đo nhỏ.

Một khó khăn thực tế trong việc áp dụng DLGHTT là khi nào n là đủ lớn. Vấn đề là độ chính xác của xấp xỉ cho một n cụ thể phụ thuộc vào hình dạng phân phối cơ bản ban đầu được lấy mẫu. Nếu phân phối cơ bản gần với một đường cong chuẩn, thì xấp xỉ tốt ngay cả đối với một n nhỏ, trong khi nếu nó xa phân phối chuẩn, thì cần n lớn.

Nguyên tắc chung: Nếu $n > 30$ định lý giới hạn trung tâm được sử dụng.

Có những phân phối tổng thể mà thậm chí n từ 40 hoặc 50 thì không đủ, nhưng những phân phối này hiếm gặp trong thực hành. Mặt khác, nguyên tắc chung thì bảo toàn đối với nhiều phân phối tổng thể, khi đó giá trị n ít hơn 30 thì đã đủ. Ví dụ, trong trường hợp phân phối đều của tổng thể, DLGHTT cho xấp xỉ tốt với $n \geq 12$.

4.4.3 Những ứng dụng khác của định lý giới hạn trung tâm

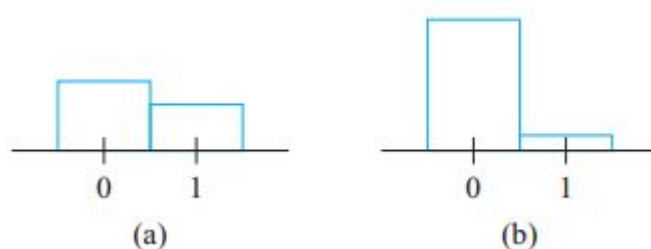
DLGHTT có thể được sử dụng để làm rõ cho xấp xỉ chuẩn đối với phân phối nhị thức được thảo luận trong Chương 4. Nhớ lại rằng một biến nhị thức X là số lần thành công trong một thí nghiệm nhị thức bao gồm n thử nghiệm thành công / thất bại độc lập với $p = P(S)$ cho bất kỳ thử nghiệm cụ thể nào. Xác định một biến ngẫu nhiên mới X_1 bởi

$$X_1 = \begin{cases} 1 & , \text{ kết quả đầu thành công} \\ 0 & , \text{ kết quả đầu thất bại} \end{cases}$$

và định nghĩa $X_1, X_2, X_3, \dots, X_n$ tương tự cho $n - 1$ thử nghiệm khác. Mỗi X_i cho biết một phép thử tương ứng có thành công hay không.

Bởi vì các thử nghiệm là độc lập và $P(S)$ là hằng số qua các thử nghiệm, các biến ngẫu nhiên X_i thì độc lập cùng phân phối (mẫu ngẫu nhiên được lấy từ phân phối Bernoulli). DLGTTT chỉ ra rằng nếu n là đủ lớn, cả tổng và trung bình của các X_i đều xấp xỉ phân phối chuẩn. Khi các X_i được cộng lại, tổng S được thêm 1 nếu biến đó xuất hiện và F được thêm 0, vì vậy $X_1 + X_2 + \dots + X_n = X$. Trung bình mẫu của các X_i là X/n chính là tỷ lệ thành công của mẫu. Nghĩa là, cả X và X/n đều

xấp xỉ chuẩn khi n lớn. Kích cỡ mẫu cần thiết cho phép xấp xỉ này phụ thuộc vào giá trị của p : Khi p gần 0,5, sự phân bố của mỗi X_i thì đối xứng (xem hình dưới), trong khi phân phối là khá lệch khi p gần 0 hoặc 1. Sử dụng xấp xỉ chỉ khi $np \geq 10$ và $n(1 - p) \geq 10$ đảm bảo rằng n đủ lớn để vượt qua bất kỳ sự sai lệch nào trong phân phối Bernoulli bên dưới:



Nhắc lại: X có phân phối log chuẩn nếu $\ln(X)$ có phân phối chuẩn.

Mệnh đề 4.4.4. Cho $X_1, X_2, \dots, X_n$ là một mẫu ngẫu nhiên từ một phân phối mà chỉ có giá trị dương thỏa $[P(X_i > 0) = 1]$. Thì với n đủ lớn, tích $Y = X_1 X_2 \dots X_n$ sẽ xấp xỉ phân phối log chuẩn

4.5 Phân phối của tổ hợp tuyến tính

Trung bình mẫu $\bar{X}$ và tổng của mẫu T_o là những trường hợp đặc biệt của một loại biến ngẫu nhiên xuất hiện rất thường xuyên trong các ứng dụng thống kê.

Định nghĩa 4.5.1. Cho một tập n các biến ngẫu nhiên $X_1, \dots, X_n$ và n hằng số $a_1, \dots, a_n$ của biến ngẫu nhiên.

$$Y = a_1 X_1 + \dots + a_n X_n = \sum_{i=1}^n a_i X_i$$

được gọi là **tổ hợp tuyến tính** của biến ngẫu nhiên X_i

Mệnh đề 4.5.2. Cho $X_1, X_2, \dots, X_n$ có giá trị trung bình tương ứng $\mu_1, \dots, \mu_n$ và phương sai tương ứng $\sigma_1^2, \dots, \sigma_n^2$.

1. X_i có độc lập hay không đều có:

$$E(a_1 X_1 + a_2 X_2 + \dots + a_n X_n) = a_1 E(X_1) + a_2 E(X_2) + \dots + a_n E(X_n) = a_1 \mu_1 + \dots + a_n \mu_n$$

2. Nếu $X_1, \dots, X_n$ độc lập

$$V(a_1X_1 + a_2X_2 + \dots + a_nX_n) = a_1^2V(X_1) + a_2^2V(X_2) + \dots + a_n^2V(X_n) = a_1^2\sigma_1^2 + \dots + a_n^2\sigma_n^2$$

và

$$\sigma_{a_1X_1+a_2X_2+\dots+a_nX_n} = \sqrt{a_1^2\sigma_1^2 + \dots + a_n^2\sigma_n^2}$$

3. Với bất kì $X_1, \dots, X_n$

$$V(a_1X_1 + a_2X_2 + \dots + a_nX_n) = \sum_{i=1}^n \sum_{j=1}^n a_i a_j \text{Cov}(X_i, X_j)$$

Ví dụ 4.17 Một trạm xăng bán ba loại xăng: regular, extra và super. Giá tương ứng các loại xăng này là 3 \$; 3,2\$ và 3,4\$ cho 1 gallon (4,4 lít). Gọi X_1, X_2, X_3 kí hiệu cho lượng gallon bán được trong 1 ngày nào đó. Giả sử mẫu X_i độc lập với $\mu_1 = 1000$, $\mu_2 = 500$, $\mu_3 = 300$, $\sigma_1 = 100$, $\sigma_2 = 80$, $\sigma_3 = 50$. Doanh thu kí hiệu là $Y = 3.X_1 + 3,2X_2 + 3,4X_3$. Hãy tính $E(Y), V(Y), \sigma(Y)$? (đs: 5620 ; 184,436 , 42,46)

4.5.1 Hiệu của hai biến ngẫu nhiên

Một trường hợp đặc biệt quan trọng của tổ hợp tuyến tính khi cho $n = 2, a_1 = 1, a_2 = -1$:

$$Y = a_1X_1 + a_2X_2 = X_1 - X_2$$

Từ đó ta có hệ quả sau:

Hệ quả 4.5.3. $E(X_1 - X_2) = E(X_1) - E(X_2)$ với hai biến ngẫu nhiên bất kì X_1 và X_2 .

$V(X_1 - X_2) = V(X_1) + V(X_2)$ nếu X_1 và X_2 độc lập.

Ví dụ 4.18 Một nhà sản xuất ô tô trang bị một mô hình với một động cơ có 6 xy-lanh và động cơ 4 xy-lanh. Đặt X_1 và X_2 là lượng dầu cần tương ứng cho động cơ 6 xy-lanh và 4 xy-lanh. Với $\mu_1 = 22, \mu_2 = 26, \sigma_1 = 1, 2; \sigma_2 = 1, 5$ hãy tính $E(X_1 - X_2), V(X_1 - X_2), \sigma_{X_1-X_2}$? (đs: -4 ; 3,69 ; 1,92)

4.5.2 Trường hợp của biến ngẫu nhiên liên tục

Khi X_i là mẫu ngẫu nhiên lấy từ phân phối chuẩn, $\bar{X}$ và T_o thì cũng có phân phối chuẩn. Đây là kết quả tổng quát hơn liên quan đến tổ hợp tuyến tính.

Mệnh đề 4.5.4. *Nếu $X_1, X_2, \dots, X_n$ độc lập, có phân phối chuẩn (có thể là trung bình khác nhau, phương sai khác nhau), thì bất kỳ tổ hợp tuyến tính của X_i cũng có phân phối chuẩn. Đặc biệt, hiệu của $X_1 - X_2$ cũng được phân phối chuẩn.*

Ví dụ 4.19 Theo ví dụ trên, cho tổng thu nhập khi bán 3 loại xăng ở 1 trạm là $Y = 3X_1 + 3,2X_2 + 3,4X_3$ và ta tính $\mu_Y = 5620$, $\sigma_Y = 429,46$. Nếu X_i có phân phối chuẩn thì xác suất để doanh thu hơn 4500 là:

$$P(Y > 45000) = P(Z > \frac{4500-5620}{429,26}) = P(Z > -2,61) = 1 - \Phi(-2,61) = 0,9955$$

CHÚ Ý: Trường hợp $n = 2$ thì ta có kết quả sau

$$V(a_1X_1 + a_2X_2) = a_1^2V(X_1) + a_2^2V(X_2) + 2a_1a_2Cov(X_1, X_2)$$