UBND TỈNH QUẢNG BÌNH

TRƯỜNG ĐẠI HỌC QUẢNG BÌNH

NHẬP MÔN LÝ THUYẾT XÁC SUẤT THỐNG KÊ

Biên soạn: *Th.s PHAN TRỌNG TIẾN*

Quảng Bình, tháng 4 năm 2009

Mục lục

Mục lục 1

Lời nói đầu 2

Chương 1 Các khái niệm cơ bản về xác suất 3

§1 Bổ sung về giải tích tổ hợp . . . 3

§2 Phép thử ngẫu nhiên . . . 7

§3 Xác suất . . . 8

§4 Cách tính xác suất . . . 8

§5 Quy tắc cộng và nhân xác suất . . . 11

§6 Hệ biến cố đầy đủ và xác suất toàn phần . . . 15

§7 Công thức Bayes . . . 16

Chương 2 Biến ngẫu nhiên 19

§1 Biến ngẫu nhiên rời rạc . . . 19

§2 Bảng phân phối và hàm phân phối . . . 20

§3 Các sỐ ĐẶc trƯng . . . 21

§4 Biến ngẫu nhiên rời rạc có vô số giá trị . . . 24

§5 Một số phân phối rời rạc thường gặp . . . 25

Chương 3 Mẫu quan sát và bài toán ước lượng 31

§1 Tổng thể và mẫu quan sát . . . 31

§2 Ước lượng tham số của tổng thể . . . 33

§3 Xác định kích thước mẫu . . . 36

Chương 4 Kiểm định giả thiết 41

§1 Giả thiết và đối thiết . . . 41

§2 Kiểm định giá trị trung bình μ của biến phân phối chuẩn $N(\mu, \sigma^2)$ . . . 42

§3 Kiểm định xác suất . . . 44

4 Xác suất . . . 46

5 Biến ngẫu nhiên . . . 49

6 Bài toán ước lượng, kiểm định . . . 50

Tài liệu tham khảo 55

LỜI NÓI ĐẦU

Lý thuyết Xác suất và thống kê toán học là một ngành toán học ra đời vào khoảng thế kỷ XVII. Đối tượng nghiên cứu của Xác suất - Thống kê là các hiện tượng ngẫu nhiên, các quy luật ngẫu nhiên mà chúng ta thường gặp trong thực tế. Khác với một số môn Toán học trừu tượng, lý thuyết Xác suất - Thống kê được xây dựng dựa trên các công cụ toán học hiện đại như Giải tích hàm, Lý thuyết độ đo, ... nhưng lại gắn liền với các bài toán thực tế cuộc sống, trong tự nhiên và xã hội.

Ngày nay, lý thuyết Xác suất - Thống kê Toán học đã được đưa vào giảng dạy ở hầu hết các ngành đào tạo trong các trường Đại học và Cao đẳng trên thế giới và trong nước. Nó đang là một trong những ngành khoa học phát triển cả về lý thuyết cũng như ứng dụng. Nó được ứng dụng rộng rãi trong hầu hết các lĩnh vực khoa học tự nhiên, khoa học xã hội, trong kinh tế, kỹ thuật, y học, ...

Bài giảng Xác suất - Thống kê này được biên soạn cho sinh viên Đại học không chuyên ngành Toán với thời lượng 30 tiết. Chính vì vậy, chúng tôi không đi sâu vào việc chứng minh những lý thuyết toán học phức tạp mà trình bày các kiến thức cơ bản như là công cụ và tập trung đưa ra các ví dụ minh họa.

Bài giảng gồm có 4 chương:

Chương 1: Phần đầu đề cập các khái niệm cơ bản trong giải tích tổ hợp. Phần sau trình bày khái niệm xác suất và các tính chất của xác suất.

Chương 2: Trình bày biến ngẫu nhiên, bảng phân phối, hàm phân phối và các số đặc trưng. Một số phân phối thường gặp cũng được giới thiệu trong chương này.

Chương 3 và Chương 4 trình bày bài toán ước lượng và kiểm định cho các tham số của biến ngẫu nhiên.

Tác giả rất mọng nhận được sự góp ý từ phía Thầy Cô và các bạn sinh viên để bài giảng được hoàn thiện hơn.

Tác giả

Chương 1

CÁC KHÁI NIỆM CƠ BẢN VỀ XÁC SUẤT

§1 BỔ SUNG VỀ GIẢI TÍCH TỔ HỢP

Phần này không nằm trong nội dung môn Xác suất thống kê mà thuộc về các kiến thức chung đã được học ở Phổ thông, tuy nhiên để hiểu được các phép tính xác suất, thống kê ở các chương sau thì cần phải học, hoặc phải ôn lại các khái niệm cơ bản như: chỉnh hợp, hoán vị, tổ hợp, chỉnh hợp lặp.

1.1 Quy tắc nhân

Giả sử một công việc nào đó được thực hiện qua n bước. Bước thứ i có x_i cách sau khi các bước $1, 2, ..., i-1$ đã làm, khi đó để thực hiện công việc đó có $x_1.x_2...x_n$ cách.

Ví dụ 1.1. Một bé có thể mang họ cha là Lê hay họ mẹ là Đỗ, chữ lót có thể là Văn, Đồng, Bích hoặc Đình tên có thể là Nhân, Nghĩa, Trí, Đức. Hỏi có bao nhiêu cách để đặt tên đầy đủ cho bé?

Giải. Xem việc đặt tên cho bé được thực hiện qua 3 bước. Bước 1 đặt họ: có 2 cách để đặt họ. Sau khi đặt họ thì thực hiện bước 2 đặt chữ lót: có 4 cách để đặt chữ lót. Đặt xong họ và chữ lót tiếp tục thực hiện bước 3 đặt tên: có 4 cách đặt tên. Tên đầy đủ của bé sẽ có được khi thực hiện xong cả ba bước trên. Số cách thực hiện là 2.4.4=32 cách.

1.2 Hoán vị

Định nghĩa 1.2. Cho tập A có n ($n \geq 1$) phần tử. Một cách sắp xếp có thứ tự n phần tử này được gọi là một *hoán vị* các phần tử của tập A.

Ký hiệu P_n là số các hoán vị của tập hợp có n phần tử. Ta có

Định lý 1.3. *Số các hoán vị của một tập hợp có n phần tử là*

$$P_n = n! = n(n-1)(n-2)...1.$$

Ví dụ 1.4. Có bao nhiêu số có 4 chữ số khác nhau được thiết lập từ các chữ số 1, 2, 3, 4?
Giải. Mỗi cách sắp xếp 4 chữ số 1, 2, 3, 4 theo một thứ tự nào đó ta được một số gồm 4 chữ số khác nhau. Nó chính là một hoán vị của 4 chữ số đó.

Vậy số các số khác nhau gồm 4 chữ số là 4!=24.

Ví dụ 1.5. Cửa hàng có 3 cái mũ màu xanh, đỏ, tím. Có 3 khách đến mua mũ mỗi người mua một chiếc. Hỏi cô bán hàng có mấy cách để bán mũ?
Giải. Mỗi cách bán ba chiếc mũ cho ba khách là một hoán vị của ba phần tử. Vậy số cách để cô bán hàng bán mũ là $P_3 = 3! = 6$.

Ví dụ 1.6. Có 6 cụ ông sắp hàng ngang để tập thể dục buổi sáng, sau buổi tập đầy phấn khích các cụ quyết định từ ngày hôm sau sẽ ra tập tiếp và mỗi ngày sẽ sắp hàng theo một trật tự khác những lần tập trước. Hỏi sau nhiều nhất bao nhiêu ngày các cụ mới quay lại cách xếp hàng đầu tiên?
Giải. Coi mỗi cách sắp hàng là một cách sắp xếp 6 cụ vào 6 chỗ, tức là một hoán vị của 6 cụ, có thể tìm được tất cả có $6! = 720$ cách xếp hàng. Như vậy phải 720 ngày sau, tức là gần 2 năm sau 6 cụ mới xếp hàng lại theo đúng cách sắp hàng đầu tiên.

1.3 Chỉnh hợp không lặp

Cho tập hợp $A = \{1, 2, 3\}$. Lập các bộ có thứ tự gồm hai phần tử trong ba phần tử đã cho:
Giải. Các bộ có thứ tự gồm hai phần tử trong ba phần tử của A là

$$\{1, 2\}, \{2, 1\}, \{2, 3\}, \{3, 2\}, \{1, 3\}, \{3, 1\}.$$

Mỗi bộ có thứ tự gồm hai phần tử trên được gọi là một chỉnh hợp không lặp chập 2 của 3 phần tử đã cho.

Định nghĩa 1.7. Cho tập A có n phần tử. Một *chỉnh hợp không lặp* chập k $(1 \leq k \leq n)$ của n phần tử đã cho là một bộ có thứ tự gồm k phần tử trong n phần tử.
Ký hiệu số các chỉnh hợp không lặp chập k của n phần tử là A_n^k.

Định lý 1.8. *Số các chỉnh hợp không lặp chập k của n phần tử bằng*

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

Ví dụ 1.9. Cho năm chữ số 1, 2, 3, 4, 5. Hỏi có bao nhiêu số khác nhau gồm 3 chữ số lấy từ 5 chữ số trên
Giải. Số các số khác nhau gồm 3 chữ số lấy từ 5 chữ số bằng số các chỉnh hợp không lặp chập của 5 phần tử, tức là: $A_5^3 = \dfrac{5!}{(5-3)!} = 60$.

Ví dụ 1.10. Có 8 đội bóng chuyền thi đấu để tranh ba huy chương vàng, bạc, đồng. Nếu 8 đội thực lực như nhau thì có thể có bao nhiêu dự báo về danh sách bộ ba được huy chương?
Giải. Vì thực lực như nhau nên có thể có 8 cách dự báo đội được huy chương vàng, sau đó còn 7 cách dự báo đội được huy chương bạc, cuối cùng có 6 cách dự báo đội được huy chương đồng, như vậy tất cả có $8.7.6 = 336$ chính là số chỉnh hợp không lặp chập 3 của 8 đội. Hai dự báo khác nhau nếu trong danh sách 3 đội được huy chương có ít nhất tên một đội khác nhau hoặc vẫn cùng tên 3 đội nhưng thứ tự khác nhau do đó có sự thay đổi tên đội tương ứng với loại huy chương.

Ví dụ 1.11. Một tổ có 10 người, chọn lần lượt 3 người đi làm việc, người thứ nhất là nhóm trưởng. Người thứ hai theo dõi các chỉ tiêu kinh tế. Người thứ ba theo dõi các chỉ tiêu kỹ thuật. Giả sử 10 người trong tổ có khả năng làm việc như nhau thì có bao nhiêu cách phân công việc trong nhóm
Giải. Có $A_{10}^3 = 720$ cách.

1.4 Chỉnh hợp lặp

Định nghĩa 1.12. Cho tập A có n ($n \geq 1$) phần tử. Ta gọi *chỉnh hợp lặp* chập k ($k \geq 1$) của n phần tử của A là một tập có thứ tự gồm k phần tử lấy từ n phần tử của A, mà phần tử của tập đó có thể có mặt nhiều nhất k lần. Ký hiệu số các chỉnh hợp lặp chập k của n phần tử của A là $\widetilde{A}_n^k$.

Định lý 1.13. *Số chỉnh hợp lặp chập k của n phần tử được tính theo công thức:*

$$\widetilde{A}_n^k = n^k \tag{1.1}$$

Từ nay về sau khi nói đến "chỉnh hợp" thì ta hiểu đó là chỉnh hợp không lặp. Còn "chỉnh hợp lặp" sẽ được hiểu là chỉnh hợp có lặp.

Ví dụ 1.14. Có bao nhiêu cách phân ngẫu nhiên 12 khách lên 3 toa tàu?
Giải. Số cách để 12 khách lên 3 toa tàu là số chỉnh hợp lặp chập 12 của 3 phần tử đã cho. Bởi vì mỗi hành khách có thể có 3 cách để lên tàu, nên có 3^{12} cách.

Ví dụ 1.15. Số máy điện thoại của một tỉnh gồm bảy chữ số. Mỗi chữ số được chọn trong mười số 0, 1, ... , 9 như vậy có thể tạo ra $10.10.10.10.10.10.10 = 10^7$ số máy điện thoại.

Ví dụ 1.16. Vé xổ số có bốn chữ số, như vậy có tất cả 10^4 vé xổ số có bốn chữ số.

có bao nhiêu cách trao 15 phần thưởng cho 5 người dự thi. Mỗi cách phân 15 sản phẩm cho 5 người là một chỉnh hợp chập 15 của 5.
Vậy số cách đề phân ngẫu nhiên 15 phần thưởng cho 5 người là: 5^{15}.

1.5 Tổ hợp

Cho tập hợp $A = \{1, 2, 3\}$. Lập các tập con gồm hai phần tử (không kể thứ tự) của tập M. Ta có

$$\{1, 2\}, \{1, 3\}, \{3, 2\}$$

tập hợp con khác nhau.

Mỗi tập con gồm 2 phần tử ở trên được gọi là 1 tổ hợp chập 2 của 3 phần tử.

Định nghĩa 1.17. Cho tập A gồm n ($n \in \mathbb{N}$) phần tử. Một *tổ hợp* chập k ($0 \leq k \leq n$) của n phần tử đã cho của A là một tập con của A gồm k phần tử không kể thứ tự. Ký hiệu số các tổ hợp chập k của n phần tử là C_n^k.

Định lý 1.18. *Số các tổ hợp chập k của n phần tử đã cho là*

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

Chú ý: Người ta chứng minh được các công thức sau

$$C_n^k = C_n^{n-k}$$
$$C_n^k = C_{n-1}^k + C_{n-1}^{k-1}.$$

Ví dụ 1.19. Có mấy cách cử 3 người trong một tổ gồm 12 người đi lao động?
Giải. Số cách phân công 3 người trong 12 người đi lao động bằng số các tổ hợp chập 3 của 12 phần tử. Vậy có $C_{12}^3 = 220$ cách.

Bài Tập phần giải tích tổ hợp

1.1. Có thể tạo ra bao nhiêu vectơ có gốc và đỉnh tại 2 trong 4 điểm đã cho?

1.2. Giải các phưng trình
a) $A_n^3 = 20n$; b) $A_n^2 - A_n^1 = 3$; c)$3A_n^2 + 42 = A_{2n}^2$. (n là ẩn).

1.3. Có bao nhiêu số gồm ba chữ số khác nhau lấy từ năm chữ số 0, 2, 4, 6, 8?

1.4. Một lớp có 50 học viên. Cần chọn ra lớp trưởng, lớp phó học tập và lớp phó đời sống. Nếu ai cũng có khả năng được chọn vào các chức vụ trên thì có bao nhiêu cách chọn?

1.5. Có thể lập được bao nhiêu số chẵn gồm năm chữ số khác nhau lấy từ sáu chữ số 0, 1, 2, 3, 4, 5?

1.6. Có 24 đội bóng tham gia thi đấu, hai đội phải đấu với nhau một lượt đi, một lượt về. Ban tổ chức phải tổ chức bao nhiêu trận đấu?

1.7. Có 3 cụ ông, 2 cụ bà và 5 em bé ngồi quanh một bàn tròn. Hỏi có bao nhiêu cách sắp xếp chỗ ngồi sao cho các cụ ông ngồi cạnh nhau, các cụ bà ngồi cạnh nhau và các em bé cũng ngồi cạnh nhau?

1.8. Có 3 cặp vợ chồng đi xem văn nghệ và ngồi vào 6 ghế trên một hàng ngang có bao nhiêu cách sắp xếp sao cho vợ chồng luôn ngồi cạnh nhau?

1.9. Có 3 quyển sách Toán, 4 quyển sách Lí, 2 quyển sách Hoá và 5 quyển sách Sinh.
a) Có bao nhiêu cách xếp 14 quyển sách lên giá sách?
b) Có bao nhiêu cách xếp sao cho sách cùng môn học ở cạnh nhau?

1.10. Có bao nhiêu số có năm chữ số lấy từ năm chữ số 1, 2, 3, 4, 5 trong đó hai chữ số 1 và 2 không đứng cạnh nhau?

1.11. Trong mặt phẳng có n điểm trong đó không có 3 điểm nào thẳng hàng. Hỏi có thể kẽ được bao nhiêu đường thẳng, mỗi đường thẳng đi qua 2 điểm trong số n điểm đã cho?

1.12. Cho đa giác lồi n $(n \geq 4)$ đỉnh $D_1, D_2, ..., D_n$. Có tất cả bao nhiêu đường chéo?

1.13. Có 12 điểm nằm trên một đường tròn .
a) Hỏi có bao nhiêu tam giác nội tiếp đường tròn có đỉnh nằm trong số các điểm đã cho?
b) Hỏi có bao nhiêu tứ giác nội tiếp đường tròn có đỉnh nằm trong số các điểm đã cho?

1.14. Một hộp đựng 6 bi trắng và 4 bi đen.
a) Có bao nhiêu cách lấy ra 5 bi.
b) Có bao nhiêu cách lấy ra 5 bi trong đó có 2 bi trắng.
c) Có bao nhiêu cách lấy ra 5 bi trong đó ít nhất có 2 bi trắng.
d) Có bao nhiêu cách lấy ra 5 bi trong đó có nhiều nhất 2 bi trắng.

1.15. Có 10 người gồm 7 nam và 3 nữ.
a) Có bao nhiêu cách chọn một uỷ ban gồm 3 người.
b) Có bao nhiêu cách chọn để trong uỷ ban nói trên có một nữ.
c) Có bao nhiêu cách chọn để trong uỷ ban nói trên có ít nhất một nữ.

§2 PHÉP THỬ NGẪU NHIÊN

Trong nghiên cứu tự nhiên và xã hội ta phải theo dõi các hiện tượng, phải cân, đong, đo, đếm, làm thí nghiệm ... những việc này, trong điều kiện cho phép, phải lặp lại nhiều lần. Ta gọi chung các công việc này là phép thử. Khi lặp lại phép thử ta thấy có phép thử luôn cho cùng một kết quả, ví dụ đun nước ở điều kiện cao độ và áp suất bình thường thì đến 100^0C nước sẽ sôi, trứng gà trong đàn gà không có trống khi ấp sẽ không nở, hạt giống nếu xử lý ở nhiệt độ quá cao hoặc nồng độ hoá chất quá cao sẽ không nẩy mầm, ... Ta gọi đó là các kết quả tất yếu.

Ngoài loại phép thử cho kết quả tất yếu ra còn có rất nhiều phép thử khi lặp lại sẽ cho các kết quả khác nhau. Số kết quả đó có thể hữu hạn, có thể vô hạn, có thể là các giá trị rời rạc hay liên tục, ví dụ sinh con có thể trai hay gái, ấp trứng có thể nở hoặc không, trồng 10 cây thì số cây sống có thể là 0, 1, ... , 10, làm thí nghiệm có thể thành công hoặc thất bại.

Một hành động mà kết quả của nó không thể dự báo trước được gọi là một *phép thử ngẫu nhiên*.

Ký hiệu phép thử ngẫu nhiên là T. Các kết quả của T không thể nói trước được một cách chắc chắn, nhưng ta có thể liệt kê ra tất cả các kết quả có thể có của T.

Tập tất cả các kết quả của T được gọi là *không gian mẫu* và thường ký hiệu nó bằng chữ Ω.

Khi thực hiện phép thử, kết quả của phép thử gọi là *biến cố sơ cấp (sự kiện sơ cấp)* ký hiệu là ω, như vậy $\omega \in \Omega$.

Mỗi tập con A của Ω được gọi là *một biến cố*. Mỗi kết quả $\omega \in A$ được gọi là một *kết quả thuận lợi cho* A.

Khi kết quả của T là một phần tử của A thì có nghĩa là A xảy ra.

Biến cố không thể là biến cố không bao giờ xảy ra. Nó tương ứng với tập con $\emptyset$ của Ω.

Biến cố chắc chắn là biến cố luôn xảy ra. Nó tương ứng với toàn bộ tập Ω.

Ví dụ 2.1. Gieo một con xúc xắc, biến cố sơ cấp là ra mặt 1, 2, 3, 4, 5, 6. biến cố ra mặt chẵn A bao gồm ba biến cố sơ cấp (2, 4, 6). Biến cố ra mặt lẻ B bao gồm ba biến cố sơ cấp (1, 3, 5).

Nếu gieo hai con xúc xắc thì các biến cố sơ cấp là 36 cặp số (1, 2), (1, 3), ..., (6, 6).
biến cố "Có mặt 6" bao gồm 11 biến cố sơ cấp: (1, 6), (2, 6), ... , (6, 1), ..., (6, 6).
biến cố "Tổng số điểm trên hai con xúc xắc là 10" gồm ba biến cố sơ cấp (4, 6), (5, 5), (6, 4). Biến cố "Điểm trên hai con xúc xắc bằng nhau" bao gồm 6 biến cố sơ cấp (1, 1), (2, 2), (3, 3), (4, 4), (5, 5), (6, 6).

Kiểm tra 3 sản phẩm. Biến cố "không có quá 3 sản phẩm tốt có trong 3 sản phẩm kiểm tra" là biến cố chắc chắn. Biến cố "có 4 phế phẩm có trong 3 sản phẩm kiểm tra" là biến cố không thể. Biến cố "có 2 sản phẩm tốt trong 3 sản phẩm kiểm tra" là biến cố ngẫu nhiên.

§3 XÁC SUẤT

Theo dõi nhiều lần một phép thử và các biến cố liên quan đến phép thử ta thấy có biến cố hay xuất hiện, hay xảy ra, có biến cố ít xuất hiện, ít xảy ra, biến cố tất yếu luôn xảy ra còn biến cố không thể không bao giờ xảy ra.

Ví dụ gieo một con xúc xắc, biến cố ra mặt chẵn và biến cố ra mặt lẻ có mức độ xuất hiện như nhau, biến cố "ra số chia được cho 3" ít xuất hiện hơn. Biến cố ra mặt 6 lại còn ít xuất hiện hơn nữa. Biến cố "ra một số ít hơn 7" là biến cố tất yếu. Còn biến cố "ra một số lớn hơn 6" là biến cố không thể.

Như vậy trong một phép thử mỗi biến cố có một mức độ (hay khả năng) xuất hiện mà chúng ta muốn đánh giá (thay nó) bằng một con số.

Giả sử A là biến cố của phép thử nào đó. Mặc dù khi tiến hành phép thử ta không thể nói trước biến cố A xảy ra hay không nhưng ta thừa nhận rằng: có một số đo khả năng xảy ra của biến cố A, ký hiệu $p(A)$. Khi đó $p(A) = 1$ nếu A là biến cố chắc chắn và $p(A) = 0$ nếu A là biến cố không thể.

Định nghĩa 3.1. *Xác suất* của một biến cố là một số đo lường khả năng xuất hiện của biến cố đó. Số đó luôn nằm giữa 0 và 1. Xác suất của một biến cố càng nhỏ (càng gần 0) thì biến cố đó càng ít khả năng xảy ra. Xác suất của một biến cố càng lớn (càng gần 1) thì biến cố có nhiều khả năng xảy ra.

Tính chất

Nếu A là biến cố ngẫu nhiên thì: $0 < p(A) < 1$

Nếu A là biến cố chắc chắn thì: $p(A) = 1$

Nếu A là biến cố không thể thì: $p(A) = 0$

Như vậy nếu A là biến cố bất kỳ thì $0 \leq p(A) \leq 1$.

§4 CÁCH TÍNH XÁC SUẤT

Có nhiều cách tính xác suất, có cách tính chặt chẽ theo hệ liên để giúp xây dựng xác suất thành một ngành toán học với lý thuyết và ứng dụng phong phú, có cách tính trực quan hơn và dựa vào các môn học khác như tính xác suất theo Cơ học, theo Hình học, theo Đại số ... trong tài liệu này chúng ta dùng hai cách tính xác suất: cách tính thống kê và cách tính đồng khả năng.

4.1 Cách tính thống kê

Xác định điều kiện đầu xong ta lặp lại phép thử nhiều lần, càng nhiều càng tốt giữ lại số lần thử n và số lần có biến cố A, gọi là tần số $n(A)$.

Tần suất của biến cố A, ký hiệu là $f(A)$ được tính theo công thức

$$f(A) = \frac{n(A)}{n}. \tag{4.1}$$

Tần suất không phải là xác suất nhưng nếu không có cách nào khác để tính xác suất lại lấy tần suất $f(A)$ làm xác suất $p(A)$. Nếu có điều kiện làm hàng loạt phép thử và tính tần suất của biến cố A trong các loạt đó người ta thấy tần suất khá ổn định (khác nhau rất ít) và thường dao động quanh một số xác định. Khi số phép thử tăng lên (và khá lớn) thì biên độ (sai khác giữa tần suất và số nói trên) có khuynh hướng nhỏ dần đi càng ngày càng ít xuất hiện các biên độ lớn. Số xác định nói trên được lấy làm xác suất.

Ví dụ 4.1. Để tính xác suất ra mặt sấp khi gieo một đồng tiền, ta có các kết quả sau, dao động quanh 0,5

Người thực hiện	Số lần gieo	Số lần ra mặt sấp	Tần suất
Buýt phông	4040	2048	0,5080
Piếc sơn	12000	6019	0,5016
Piếc sơn	24000	12012	0,5005

Ví dụ 4.2. Ở Trung Quốc, từ năm 228 trước Công nguyên, đã tìm thấy tần suất sinh con trai là $\frac{1}{2}$. Laplaxơ theo dõi các thành phố Luân Đôn, Pê-téc-bua và Bec-lin và công bố tần suất sinh con trai là $\frac{22}{43}$. Cramơ cho tần suất sinh con trai ở Thuỵ Điển là 0,508. Ở Việt Nam năm 1961 tần suất sinh con trai là 0,51.

Cách tính thống kê đơn giản, cho kết quả dễ hiểu, dễ giải thích, phù hợp với thực tế nhưng không chính xác và không thể dùng để tính xác suất trong những trường hợp phức tạp hoặc các trường hợp không thể trực tiếp lặp lại phép thử.

Trong thực tế, khi xem xét một số lượng lớn sản phẩm, chúng ta thường dùng phần trăm (%). Ví dụ số học sinh thi đỗ là 90%, số sản phẩm đạt tiêu chuẩn là 80%, số người bị bệnh trong một đợt dịch là 30% ... Theo cách tính thống kê có thể đổi % sang xác suất như sau: nếu số đạt tiêu chuẩn là P% thì khi chọn ngẫu nhiên một sản phẩm xác suất để sản phẩm đó đạt tiêu chuẩn là $\frac{P}{100}$. Cách tính thống kê đã được dùng từ xa xưa để tính xác suất sinh con trai, con gái. Xác suất xuất hiện các hiện tượng lạ trong tự nhiên như lũ lụt, sóng thần, nhật thực, nguyệt thực, ...

4.2 Cách tính cổ điển hay cách tính đồng khả năng

Tiến hành một phép thử và giả sử n kết quả (biến cố sơ cấp) của phép thử có khả năng xuất hiện như nhau, gọi đó là phép thử có n kết quả đồng khả năng. Khi đó người ta lấy xác suất của mỗi kết quả là $\frac{1}{n}$.

Từ chấp nhận này có thể tính được xác suất $p(A)$ của một biến cố bất kì A như sau:

Xác suất của biến cố A là tỷ số giữa $n(A)$ số kết quả thuận lợi cho A và số kết quả đồng khả năng n

$$p(A) = \frac{n(A)}{n} \tag{4.2}$$

Cách tính đồng khả năng đơn giản, cho kết quả dễ hiểu, dễ giải thích, phù hợp với thực tế, nhưng không chặt chẽ và không thể dùng để tính xác suất trong những trường hợp phức tạp, hoặc các trường hợp không thể chấp nhận giả thiết đồng khả năng. Trong nhiều ví dụ, ở các phần sau chúng ta tính xác suất theo cách đồng khả năng và trong các bài tập cũng có rất nhiều bài tính xác suất theo cách đồng khả năng.

Ví dụ 4.3. Gieo một đồng tiền có thể coi hai kết quả sấp (S), ngửa (N), là 2 biến cố sơ cấp đồng khả năng, mỗi biến cố có xác suất $\frac{1}{2}$. Nếu gieo một lúc hai đồng tiền thì có thể coi 4 kết quả sau là đồng khả năng: $(S, S), (S, N)$, $(N, S), (N, N)$, một biến cố sơ cấp có xác suất $\frac{1}{4}$. Nếu gọi A là biến cố "hai đồng tiền cùng mặt" thì xác suất $p(A) = \frac{2}{4} = \frac{1}{2}$ vì A gồm 2 biến cố sơ cấp (S, S) và (N, N).

Ví dụ 4.4. Vé xổ số có bốn chữ số, khi quay số trúng thưởng có 1 vé trúng giải nhất. Tính xác suất để mua 1 vé thì vé đó sẽ trúng giải nhất. Có tất cả $10^4 = 10000$ vé bốn chữ số, có thể coi đó là 10000 kết quả đồng khả năng khi quay số trúng thưởng. Như vậy mỗi vé có xác suất trúng giải nhất như nhau và bằng $\frac{1}{10000}$.

Tính xác suất để vé trúng giải nhất là vé có bốn chữ số khác nhau? Trong 10000 vé có $A_{10}^4 = 5040$ vé có bốn chữ số khác nhau (vé xổ số có thể bắt đầu bằng số 0) như vậy xác suất để vé trúng giải nhất có bốn chữ số khác nhau là $\frac{5040}{10000} = 0,504$.

Ví dụ 4.5. Gieo đồng thời hai con xúc xắc cân đối và đồng chất. Tìm xác suất để:
a) Tổng số chấm ở mặt trên hai con xúc xắc bằng 8.
b) Hiệu các số chấm ở mặt trên hai con xúc xắc có giá trị tuyệt đối bằng 2.
c) Số chấm ở mặt trên hai con xúc xắc bằng nhau.
Giải. Số kết quả đồng khả năng là $n = 6.6 = 36$.
a) Gọi A là biến cố "tổng số chấm ở mặt trên hai con xúc xắc bằng 8", khi đó

$$A = \{(2, 6), (6, 2), (4, 4), (5, 3), (3, 5)\}$$

xác suất phải tìm $p(A) = \frac{5}{36}$.
b) Gọi B là biến cố "hiệu các số chấm ở mặt trên hai con xúc xắc có giá trị tuyệt đối bằng 2", khi đó $B = \{(1, 3), (3, 1), (2, 4), (4, 2), (3, 5), (5, 3)\}$ và $p(B) = \frac{8}{36} = \frac{2}{9}$.
c) Gọi C là biến cố "số chấm ở mặt trên hai con xúc xắc bằng nhau", khi đó $C = \{(1, 1), (2, 2), (3, 3), (4, 4), (5, 5), (6, 6)\}$ và $p(C) = \frac{6}{36} = \frac{1}{6}$.

Ví dụ 4.6. Một công ty cần tuyển hai nhân viên. Có 6 người nộp đơn trong đó có 4 nam và 2 nữ. Giả thiết rằng khả năng trúng tuyển của 6 người là như nhau.
a) Tính xác suất để 2 người trúng tuyển đều là nam.
b) Tính xác suất để 2 người trúng tuyển đều là nữ.
c) Tính xác suất để có ít nhất 1 nữ trúng tuyển.
Giải. Số kết quả đồng khả năng $C_6^2 = 15$.
a) Chỉ có một trường hợp là 2 nam trúng tuyển nên xác suất cần tìm là $p = \frac{1}{15}$.
b) Số cách chọn 2 nữ trúng tuyển trong số 4 nữ là $C_4^2 = 6$. Xác suất cần tìm $p = \frac{6}{15} = \frac{2}{5}$.
c) Chỉ có 1 trường hợp 2 nam trúng tuyển nên trong 14 trường hợp còn lại ta đều có ít nhất 1 nữ trúng tuyển. Xác suất cần tìm $p = \frac{14}{15}$.

Ví dụ 4.7. Gieo đồng thời 3 con xúc xắc được chế tạo cân đối, đồng chất. Tính xác suất để tổng số nốt xuất hiện của 3 con là 9.

Giải. Mỗi kết quả của phép thử là một bộ ba (a, b, c), trong đó a, b, c là các số nguyên dương từ 1 đến 6. Vậy số kết quả đồng khả năng là$6^3 = 216$. Các bộ ba có tổng bằng 9 là:
(1,2,6) và 5 hoán vị của nó; (1,3,5) và 5 hoán vị của nó (1,4,4) và 2 hoán vị của nó; (2,2,5) và 2 hoán vị của nó (2,3,4) và 5 hoán vị của nó; (3,3,3). Suy ra số trường hợp thuận lợi là 6+6+3+6+3+1=25. Vậy $p(A) = \dfrac{25}{216}$.

§5 QUY TẮC CỘNG VÀ NHÂN XÁC SUẤT

Sau khi tính xác suất của các biến cố tương đối đơn giản, chúng ta xem xét các biến cố phức tạp hơn. Để làm được việc này, ta xét một số phép tính trên các biến cố.

Gọi A và B là hai biến cố xác định trên tập hợp các biến cố sơ cấp Ω. *Hội* của hai biến cố A và B ký hiệu $A \cap B$ là biến cố bao gồm các biến cố sơ cấp vừa của biến cố A, vừa của biến cố B. (Hội $A \cap B$ còn được gọi là biến cố "A và B" hoặc giao của A và B).

Như vậy hội của hai biến cố A, B là biến cố "cả A và B đều xảy ra".

Ví dụ 5.1. Gieo một xúc xắc, biến cố A "ra số chẵn" và biến cố B "ra một số chia được cho 3" có hội là biến cố sơ cấp "ra mặt 6", nói cách khác nếu kết quả vừa là số chẵn (có biến cố A) vừa là số chia được cho 3 (có biến cố B) thì hội $A \cap B$ là biến cố "ra mặt 6".

Ví dụ 5.2. Gọi A là biến cố người đại diện của tổ là người được xếp loại giỏi về học tập, B là biến cố người đại diện của tổ là người biết chơi bóng chuyền thì $A \cap B$ là biến cố người đại diện của tổ là người vừa học giỏi vừa biết chơi bóng chuyền.

Biến cố đối lập của biến cố A, ký hiệu $\overline{A}$, là biến cố bao gồm các biến cố sơ cấp trong Ω nhưng không thuộc A. Như vậy $\overline{A} = \Omega \backslash A$.

Từ định nghĩa trên ta thấy biến cố $\overline{A}$ là biến cố đối lập của biến cố A thì A cũng là biến cố đối lập của $\overline{A}$. Ta nói A và $\overline{A}$ là hai biến cố đối lập của nhau.

Ví dụ 5.3. Gieo một xúc xắc nếu gọi A là biến cố "ra mặt chẵn thì biến cố đối lập $\overline{A}$ của A là biến cố "ra mặt lẻ".

Ví dụ 5.4. Khi thi thì biến cố A "thi đỗ" có biến cố đối lập $\overline{A}$ là "thi trượt".

Hai biến cố A và B được gọi là *xung khắc* nếu hội của chúng là rỗng $A \cap B = \emptyset$.

Khi tiến hành phép thử hai biến cố xung khắc không có biến cố sơ cấp chung nào nên không thể xuất hiện đồng thời.

Ví dụ 5.5. Biến cố A "ra mặt chẵn" và biến cố C "ra mặt lẻ" là 2 biến cố xung khắc khi gieo một con xúc xắc.

Ví dụ 5.6. Biến cố A "ra mặt chẵn" và biến cố B "ra một số chia được cho 3" không xung khắc.

Ví dụ 5.7. Nếu trong hộp có 3 loại bi màu trắng, màu xanh, màu đỏ thì biến cố rút được bi xanh và biến cố rút được bi đỏ là 2 biến cố xung khắc nhưng không đối lập.

Ví dụ 5.8. Khi thi thì biến cố A "đạt điểm giỏi" và biến cố B "đạt điểm khá" là hai biến cố xung khắc, nhưng không đối lập, vì còn nhiều điểm khác. biến cố A và biến cố C "trên trung bình" không xung khắc.

Qua các ví dụ trên ta thấy đối lập là trường hợp riêng của xung khắc. Đối lập thì xung khắc nhưng xung khắc chưa chắc đã đối lập.

Hợp của hai biến cố A và B, ký hiệu $A \cup B$, là biến cố bao gồm tất cả các biến cố sơ cấp của biến cố A và biến cố B.

Khi tiến hành phép thử thì biến cố $A \cup B$ xuất hiện khi có ít nhất một trong hai biến cố A và B xuất hiện.

Nếu phân tích kỹ có thể thấy có ba trường hợp: A xuất hiện nhưng B không xuất hiện $A \cap \overline{B}$, B xuất hiện nhưng A không xuất hiện $\overline{A} \cap B$, cả A và B đều xuất hiện $A \cap B$.

Hợp $A \cup B$ còn được gọi là biến cố "A hoặc B".

Ví dụ 5.9. Khi gieo xúc xắc nếu gọi A là biến cố "ra mặt chẵn", B là biến cố "ra một số chia hết cho 3" thì biến cố $A \cup B$ gồm bốn biến cố sơ cấp (2, 3, 4, 6).

Ví dụ 5.10. Trong Ví dụ 5.7, khi rút bi trong hộp nếu gọi A là biến cố rút được bi trắng thì biến cố đối lập $\overline{A}$ là biến cố rút được bi xanh hoặc bi đỏ $\overline{A} = B \cup C$.

Quy tắc cộng đơn giản

Ta thừa nhận quy tắc cộng đơn giản sau đây:

$$p(A \cup B) = p(A) + p(B) \text{ nếu } A \text{ và } B \text{ xung khắc.} \tag{5.1}$$

Hệ quả 5.10.1. Gọi $\overline{A}$ là biến cố đối lập của biến cố A, ta có

$$p(\overline{A}) = 1 - p(A). \tag{5.2}$$

Thật vậy, từ $A \cup \overline{A} = \Omega$, A và $\overline{A}$ đối lập ta có

$$1 = p(\Omega) = p(A \cup \overline{A}) = p(A) + p(\overline{A}) \Rightarrow p(\overline{A}) = 1 - p(A).$$

Ví dụ 5.11. Trong hộp có 3 bi trắng, 4 bi xanh và 5 bi đỏ, gọi A là biến cố rút được bi trắng, B là biến cố rút được bi xanh, C là biến cố rút được bi đỏ. $\overline{A}$ là biến cố "bi rút ra không phải bi trắng", $B \cup C$ là biến cố "rút được bi xanh hoặc bi đỏ".

Ta có $p(A) = \dfrac{3}{12}$, $p(B) = \dfrac{4}{12}$, $p(C) = \dfrac{5}{12}$, $p(\overline{A}) = 1 - \dfrac{3}{12} = \dfrac{9}{12}$.

Vì $\overline{A} = B \cup C$ nên $p(B \cup C) = \dfrac{9}{12}$.

Cũng có thể tính theo quy tắc cộng đơn giản: $p(B \cup C) = p(B) + p(C) = \dfrac{4}{12} + \dfrac{5}{12} = \dfrac{9}{12}$.

Ví dụ 5.12. Trong kỳ thi quy định "điểm giỏi" là điểm trên 8 (không cho điểm lẻ). Một học sinh vào thi, A là biến cố "đạt điểm 10", B là biến cố "đạt điểm 9". Giả sử với em đó xác suất $p(A) = 0,3, p(B) = 0,4$. Gọi C là biến cố "đạt điểm giỏi", C là hợp của A và B

$$p(C) = p(A \cup B) = p(A) + p(B) = 0,3 + 0,4 = 0,7$$

Ví dụ 5.13. 70% sản phẩm của xí nghiệp thuộc loại I, 20% thuộc loại II. số còn lại thuộc loại III. Khi kiểm tra để xuất khẩu thì chỉ chấp nhận sản phẩm loại I hoặc II.

Gọi A là biến cố khi kiểm tra thì sản phẩm thuộc loại I, B là biến cố khi kiểm tra thì sản phẩm thuộc loại II, C là biến cố khi kiểm tra thì sản phẩm được chấp nhận cho xuất khẩu.

$$p(C) = p(A \cup B) = 0,7 + 0,2 = 0,9.$$

Xác suất có điều kiện

Xét một phép thử được tiến hành trong một điều kiện đầu nào đó và hai biến cố A, B.

Xác suất có điều kiện $p(B/A)$ là xác suất của B khi đã xảy ra biến cố A.
Có thể coi như B được tính khi phép thử được tiến hành trong điều kiện đầu mới gồm điều kiện đầu cũ cộng thêm sự xuất hiện (có mặt) biến cố A.

Ví dụ 5.14. Lấy lại Ví dụ 5.11 $p(A) = \dfrac{3}{12}; p(B) = \dfrac{4}{12}$. Nếu bây giờ biết bi lấy ra không phải bi trắng (biến cố $\overline{A}$) thì $p(B/\overline{A}) = \dfrac{4}{9}$. Trong ví dụ này $p(B/\overline{A}) \neq p(B)$.

Ví dụ 5.15. Gọi A là biến cố rút được con pích, B là biến cố rút được con át trong cỗ bài tu lơ khơ 52 quân với 4 loại: cơ, rô, nhép, pích. $p(A) = \dfrac{13}{52}$; $p(B) = \dfrac{4}{52} = \dfrac{1}{13}$.
Nếu biết con bài rút ra là con pích (biến cố A) thì xác suất rút được con át $p(B/A) = \dfrac{1}{13}$.
Trong ví dụ này $p(B) = p(B/A)$.

Một túi đựng 5 quả cầu, (trong đó có 2 quả màu trắng). Lấy ngẫu nhiên (không hoàn lại) lần lượt từ túi ra 2 quả cầu. Tính xác suất để lần thứ hai được quả cầu trắng biết rằng lần thứ nhất lấy được quả cầu trắng.
Giải. Gọi A là biến cố "lần thứ hai lấy được quả cầu trắng" và gọi B là biến cố "lần thứ hai nhất lấy được quả cầu trắng". Ta cần tìm $p(A/B)$.
Ta thấy lần thứ nhất đã lấy được quả cầu trắng (B đã xãy ra) nên trong túi còn 4 quả cầu, trong đó có 1 quả trắng. Vậy $p(A/B) = \frac{1}{4} = 0.25$

Quy tắc nhân xác suất

Nếu A và B là hai biến cố trong một phép thử ta thừa nhận quy tắc nhân sau:

$$p(A \cap B) = p(A).p(B/A) = p(B).p(A/B). \tag{5.3}$$

Ví dụ 5.16. Trong hộp có 10 phiếu, 2 phiếu ghi "trúng thưởng". Một người rút lần lượt 2 phiếu, tính xác suất để cả 2 phiếu đều trúng thưởng.
Giải. Gọi A là biến cố phiếu đầu trúng thưởng, B là biến cố phiếu thứ hai trúng thưởng, C là biến cố 2 phiếu đều trúng thưởng. Có thể tính như sau:

Khi rút phiếu đầu có 10 phiếu trong đó có 2 phiếu trúng $p(A) = \dfrac{2}{10}$. Khi đã xảy ra A thì còn lại 9 phiếu trong đó có 1 phiếu trúng do đó $p(B/A) = \dfrac{1}{9}$. Từ đó suy ra:

$$p(C) = p(A \cap B) = \frac{2}{10}.\frac{1}{9} = \frac{1}{45}.$$

Ví dụ 5.17. Sản phẩm trước khi xuất khẩu phải qua hai lần kiểm tra. Bình quân 80% sản phẩm làm ra qua được lần kiểm tra I, 90% sản phẩm đã qua lần kiểm tra I và thì sẽ qua được kiểm tra II. Tính xác suất để sản phẩm được xuất khẩu.
Giải. Gọi A là biến cố qua được kiểm tra I, B là biến cố qua được kiểm tra II, C là biến cố đạt tiêu chuẩn xuất khẩu. $p(A) = 0,8; p(B/A) = 0,9$. $p(C) = p(A).p(B/A) = 0,8.0,9 = 0,72$.

Biến cố độc lập

Nếu biến cố B có xác suất có điều kiện $p(B/A)$ bằng xác suất $p(B)$ thì B được gọi là biến cố *không phụ thuộc* biến cố A.

Có thể chứng minh ngay nếu B không phụ thuộc A thì A không phụ thuộc B, thực vậy theo quy tắc nhân tổng quát: $p(A \cap B) = p(A).p(B/A) = p(B).p(A/B)$. Nếu $p(B/A) = p(B)$ thì thay vào hệ thức trên suy ra $p(A/B) = p(A)$, tức là A không phụ thuộc B. Qua chứng minh này chúng ta thấy tính phụ thuộc là tương hỗ nên sẽ thay thuật ngữ "không phụ thuộc" bằng thuật ngữ "độc lập".

Hai biến cố A, B trong cùng một phép thử gọi là *độc lập* khi $p(A/B) = p(A)$ (hoặc $p(B/A) = p(B)$).

Nếu A và B độc lập thì có thể chứng minh A và $\overline{B}$ độc lập, $\overline{A}$ và B độc lập, $\overline{A}$ và $\overline{B}$ độc lập.

Trong thực tế nếu hai biến cố A và B trong cùng một phép thử không ảnh hưởng đến nhau thì thường thừa nhận tính độc lập.

Quy tắc nhân đơn giản.

Nếu A và B độc lập thì từ quy tắc nhân (5.3) suy ra quy tắc nhân đơn giản sau:

$$p(A \cap B) = p(A).p(B) \tag{5.4}$$

Ví dụ 5.18. Hai người đi bắn, xác suất để người thứ nhất bắn trúng đích là $p(A) = 0,7$, xác suất để người thứ hai bắn trúng đích là $p(B) = 0,8$.
Xác suất để cả hai người bắn trúng $p(A \cap B) = 0,7.0,6 = 0,56$.

Ví dụ 5.19. Sản phẩm mới làm ra phải gửi đi kiểm nghiệm ở hai phòng thí nghiệm độc lập. Nếu cả hai phòng chấp nhận thì sản phẩm được sản xuất đại trà. Xác suất để sản phẩm được phòng thí nghiệm A chấp nhận lại 0,8. Xác suất để được phòng thí nghiệm B chấp nhận là 0,9. Vậy xác suất để sản phẩm được đem ra sản xuất đại trà là 0,8 . 0,9 = 0,72.

Quy tắc cộng tổng quát

Nếu A và B là hai biến cố trong một phép thử thì có thể chứng minh quy tắc cộng tổng quát sau:

$$p(A \cup B) = p(A) + p(B) - p(A \cap B). \tag{5.5}$$

Nếu A và B xung khắc thì $p(A \cap B) = 0$ nên (5.5) trùng với quy tắc cộng đơn giản (5.1).

Ví dụ 5.20. Trong Ví dụ 5.18 nếu gọi C là biến cố đích bị bắn trúng thì $C = A \cup B$, $p(C) = p(A \cup B) = 0,7 + 0,8 - 0,56 = 0,94$. Có thể tính cách khác:

$$p(C) = p(A \cap \overline{B}) + p(\overline{A} \cap B) + p(A \cap B)$$

Ví dụ 5.21. Trong Ví dụ 5.19, gọi C là biến cố "có phòng thí nghiệm chấp nhận sản phẩm mới" $C = A \cup B$, $p(C) = 0,8 + 0,9 - 0,8.0,9 = 0,98$
Có thể lập luận như sau: C là biến cố đối lập của biến cố "cả hai phòng thí nghiệm đều không chấp nhận sản phẩm mới":
$p(C) = 1 - p(\overline{A} \cap \overline{B}) = 1 - 0,2.0,1 = 0,98$

§6 HỆ BIẾN CỐ ĐẦY ĐỦ VÀ XÁC SUẤT TOÀN PHẦN

Cho một hệ các biến cố $A_1, A_2, ..., A_n$ trong một phép thử. Nếu hệ thoả mãn hai điều kiện:
a) Từng đôi một xung khắc, tức là $A_i \cap A_j = \emptyset$ với $i \neq j$ $(i, j = \overline{1, n})$;
b) hợp của tất cả các biến cố là biến cố tất yếu, tức là $A_1 \cup A_2 \cup ... \cup A_n = \Omega$ thì hệ được gọi là hệ đầy đủ hay hệ toàn phần.

Có thể trình bày lại hai điều kiện trên dưới dạng: Có một và chỉ một trong các biến cố A_i xảy ra khi tiến hành phép thử. (Có một là điều kiện b), còn chỉ có một là điều kiện a)).

Khi có một hệ biến cố đầy đủ thì có thể tính xác suất của một biến cố bất kì B trong phép thử đó theo công thức:

$$p(B) = p(A_1)p(B/A_1) + p(A_2)p(B/A_2) + ... + p(A_n)p(B/A_n) \tag{6.1}$$

$$p(B) = \sum_{i=1}^{n} p(A_i)p(B/A_i).$$

(6.1) còn gọi là Công thức xác suất toàn phần.

Ví dụ 6.1. Cửa hàng nhận trứng của 3 cơ sở nuôi gà theo tỉ lệ: $25\%, 35\%$, và 40%. Nếu tỉ lệ trứng hỏng của 3 cơ sở là $5\%, 4\%$ và 2% thì xác suất để một quả trứng mua tại cửa hàng bị hỏng là bao nhiêu?
Giải. Khi mua một quả trứng của cửa hàng thì có một và chỉ một trong 3 biến cố xảy ra: biến cố A_1 "trứng của cơ sở I", biến cố A_2 "trứng của cơ sở II", biến cố A_3 "trứng của cơ sở III". Xác suất của ba biến cố trên lần lượt là: 0,25; 0,35; 0,40.

Gọi B là biến cố trứng mua ở cửa hàng bị hỏng. Xác suất trứng hỏng tại ba cơ sở lần lượt là $p(B/A_1) = 0,05; p(B/A_2) = 0,04; p(B/A_3) = 0,02$; $p(B) = 0,25.0,05 + 0,35.0,04 + 0,40.0,02 = 0,0345$.

Ví dụ 6.2. Có 2 hộp bên ngoài giống nhau, hộp thứ nhất chứa 1 sản phẩm hỏng và 9 sản phẩm tốt, hộp thứ hai chứa 2 sản phẩm hỏng và 8 sản phẩm tốt. Lấy ngẫu nhiên một hộp, sau đó lấy ngẫu nhiên một sản phẩm. Tính xác suất để được sản phẩm tốt.

Gọi A_1 là biến cố lấy được hộp thứ nhất, A_2 là biến cố lấy được hộp thứ hai, vì chọn ngẫu nhiên nên $p(A_1) = \frac{1}{2}; p(A_2) = \frac{1}{2}$. Gọi B là biến cố "sản phẩm tốt" ta có:

$$p(B) = \frac{1}{2}.\frac{9}{10} + \frac{1}{2}.\frac{8}{10} = \frac{17}{20} = 0,85.$$

§7 CÔNG THỨC BAYES

Cho một hệ biến cố đầy đủ $A_1, A_2, ..., A_n$. Xác suất của biến cố B tính theo công thức 6.1.

Viết lại công thức nhân tổng quát

$$p(A_i \cap B) = p(A_i).p(B/A_i) = p(B).p(A_i/B),\ (i = \overline{1,n})$$

suy ra:

$$p(A_i/B) = \frac{p(A_i).p(B/A_i)}{p(B)},\ (i = \overline{1,n}) \tag{7.1}$$

Công thức 7.1 có tên là công thức Bayes, công thức này cho phép tính $p(A_i/B)$ là gọi là xác suất hậu nghiệm, còn xác suất $p(A_i)$ được gọi là xác suất tiền nghiệm. Trong Ví dụ 6.1

$$p(A_1/B) = \frac{0,25.0,05}{0,0345} = 0,3623.$$

Có thể hiểu xác suất hậu nghiệm $p(A_i/B)$ như sau: Vào cửa hàng mua một quả trứng, xác suất mua phải quả trứng hỏng bằng 0,0345, nói cách khác số trứng hỏng của cửa hàng là $3,45\%$. Bây giờ nếu quả trứng ta mua đúng là quả trứng hỏng thì xác suất để quả đó là quả trứng nhập của cơ sở I bằng 0,3623.

Trong Ví dụ 6.2

$$p(A_2/B) = \frac{1/2.8/10}{17/20} = \frac{8}{17} \approx 0,47.$$

Có thể hiểu xác suất hậu nghiệm $p(A_2/B)$ là như sau: Lấy ngẫu nhiên một hộp, sau từ hộp nó lấy ngẫu nhiên ra một sản phẩm và được sản phẩm tốt, thế thì xác suất để hộp mà ta lấy ra là hộp thứ hai bằng 0,47.

Bài Tập

1.16. Có 10 vé đánh số từ 1 đến 10. Rút ngẫu nhiên 6 vé, tính xác suất để trong đó:
a) Có vé số 1.
b) Có vé số 1 và số 2.

1.17. Số điện thoại ở một vùng có 5 chữ số, quay ngẫu nhiên một số, tính xác suất để:
a) Được số có 5 chữ số khác nhau;
b) Số mà các chữ số đều lẻ.

1.18. Có 20 câu hỏi thi, mỗi học sinh chọn một đề gồm 3 câu. Học sinh chỉ học 12 câu, tính xác suất để ít nhất làm được một câu.

1.19. Trong bình có 2 bi trắng, 4 bi đen. Lấy lần lượt các bi ra khỏi bình. Tính xác suất để bi cuối cùng là bi đen.

1.20. Có 2 hộp, hộp thứ nhất đựng 3 bi trắng, 7 bi đỏ, 15 bi xanh, hộp thứ hai đựng 10 bi trắng, 6 bi đỏ, 9 bi xanh. Từ mỗi hộp ta lấy ngẫu nhiên ra một bi, tính xác suất để được hai bi cùng màu.

1.21. Trong một vùng tỉ lệ người mắc bệnh tim là 9% mắc bệnh huyết áp là 12%, mắc cả hai bệnh là 7%. Gặp ngẫu nhiên một người trong vùng, tính xác suất để người đó không mắc cả hai bệnh nói trên.

1.22. Một lớp học có 30 học sinh trong đó có 4 giỏi, 8 khá và 10 trung bình số còn lại loại yếu. Chọn ngẫu nhiên 3 học sinh, tính xác suất:
a) Cả 3 đều yếu;
b) Có ít nhất một học sinh giỏi;
c) Có đúng một học sinh khá.

1.23. Một lô hàng có 6 chính phẩm và 4 phế phẩm được chia thành hai phần bằng nhau. Tính xác suất để mỗi phần đều có số chính phẩm như nhau.

1.24. Hộp có 10 sản phẩm trong đó có 2 phế phẩm. Lấy ngẫu nhiên 2 sản phẩm, tính xác suất để cả hai đều là phế phẩm trong 3 trường hợp:
a) Lấy lần lượt hai sản phẩm;
b) Lấy một lúc hai sản phẩm;
c) Lấy có hoàn lại (tức là lấy sản phẩm thứ nhất ra xem sau đó hoàn lại, trộn đều rồi mới lấy sản phẩm thứ hai).

1.25. Phải gieo hai đồng tiền bao nhiều lần để biến cố "ít nhất một lần ra hai mặt sấp" có xác suất không nhỏ hơn 0,99.

1.26. Có 18 xạ thủ trong đó có 5 người bắn trúng đích với xác suất 0,8 (giỏi), 7 người bắn trúng với xác suất 0,7 (khá), 4 người bắn trúng với xác suất 0,6 (trung bình), 2 người bắn trúng với xác suất 0,5 (đạt). Chọn ngẫu nhiên một người vào bắn.
a) Tính xác suất để người đó bán trượt.
b) Nếu người đó bắn trượt thì nhiều khả năng người đó thuộc nhóm nào?

1.27. Hai máy cùng sản xuất 1 loại sản phẩm, tỉ lệ phế phẩm của máy I là 0,03, của máy II là 0,02. Kho chứa $\frac{2}{3}$ sản phẩm của máy I, $\frac{1}{3}$ sản phẩm của máy II. Lấy ngẫu nhiên một sản phẩm.
a) Tính xác suất để được sản phẩm tốt.
b) Nếu được sản phẩm tốt thì nhiều khả năng sản phẩm đó là của máy nào?

1.28. Tỉ lệ người nghiện thuốc lá ở một vùng là 30%. Biết tỉ lệ viêm họng trong số người nghiện thuốc lá là 60% còn trong người không nghiện là 40%. Gặp ngẫu nhiên một người.
a) Tính xác suất để đó là người viêm họng.
b) Nếu người đó viêm họng thì tính xác suất để đó là người nghiện thuốc lá.

1.29. Trong một bệnh viện tỉ lệ bệnh nhân của các tỉnh như sau: 25% của tỉnh $A, 35\%$ của tỉnh $B, 40\%$ của tỉnh C. Tỉ lệ kỹ sư trong số bệnh nhân của tỉnh A là 2%, của tỉnh B là 3%, của tỉnh C là $3,5\%$. Gặp ngẫu nhiên một bệnh nhân, tính xác suất để:
a) Bệnh nhân đó là một kỹ sư ;
b) Nhiều khả năng kỹ sư đó là người tỉnh nào?

1.30. Lô hàng xuất khẩu có 100 kiện hàng, trong đó 60 kiện của xí nghiệp I và 40 kiện của xí nghiệp II. Tỉ lệ phế phẩm của hai xí nghiệp là 30% và 10% Lấy ngẫu nhiên một kiện rồi lấy ra một sản phẩm.
a) Tính xác suất để được một phế phẩm.
b) Nếu sản phẩm lấy ra là một phế phẩm thì nhiều khả năng kiện hàng lấy ra là của xí nghiệp nào?

1.31. Có 2 hộp đựng cam, hộp I có 9 quả tốt, 1 quả hỏng, hộp II có 6 quả tốt, 2 quả hỏng. Lấy ngẫu nhiên một quả từ hộp I bỏ sang hộp II. sau đó lấy ngẫu nhiên ở hộp II ra hai quả. Tính xác suất để cả hai quả đều hỏng.

1.32. Một cỗ máy có 3 bộ phận, xác suất hỏng trong ngày lần lượt là 0,2; 0,4; 0,3. Trong ngày có hai bộ phận hỏng, tính xác suất để đó là bộ phận 1 và 2.

1.33. Một người có 3 chỗ ưa thích như nhau để câu cá. Xác suất để câu được cá của một lần thả câu ở những chỗ đó lần lượt là 0,6; 0,7; 0,8. Người đó chọn ngẫu nhiên một chỗ và thả câu 3 lần thì câu được một con cá. Tính xác suất để chỗ câu đó là chỗ 1.

1.34. Tỉ lệ thuốc hỏng ở lô A là 0,10, ở lô B là 0,08 và ở lô C là 0,15. Chọn ngẫu nhiên một lô sau đó lấy ra 3 lọ. Tính xác suất để ít nhất có một lọ hỏng.

1.35. Hộp A có 15 lọ thuốc tốt, 5 lọ hỏng. Hộp B có 17 lọ tốt, 3 lọ hỏng. Hộp C có 10 lọ tốt 10 lọ hỏng.

a) Lấy ở mỗi hộp một lọ, tính xác suất được 2 lọ tốt, 1 lọ hỏng.
b) Chọn ngẫu nhiên một hộp sau đó lấy từ đó ra 3 lọ, tính xác suất được 2 lọ tốt và 1 lọ hỏng.
c) Trộn chung 3 hộp rồi từ đó lấy ra 3 lọ, tính xác suất được 2 lọ tốt, 1 lọ hỏng.
d) Kiểm tra từng hộp cho đến khi phát hiện đủ 3 lọ hỏng. Tính xác suất để việc kiểm tra dừng lại ở lần kiểm tra thứ 5.

Chương 2

BIẾN NGẪU NHIÊN

§1 BIẾN NGẪU NHIÊN RỜI RẠC

Ví dụ 1.1. Gieo một đồng tiền. Gọi X là kết quả với quy ước nếu ra mặt ngửa thì ghi 0, ra mặt sấp thì ghi 1. Xác suất ra 0 là $\frac{1}{2}$. Xác suất ra 1 là $\frac{1}{2}$. Ghi lại kết quả dưới dạng bảng:

X	0	1
p	$\frac{1}{2}$	$\frac{1}{2}$

Cũng gieo đồng tiền nhưng quy ước nếu ngửa thì coi như thua và phải nộp 10 đ, sấp coi như thắng và nhận được 10 đ. Số tiền thu được Y sẽ là -10 đ hoặc 10 đ với xác suất bằng nhau và bằng $\frac{1}{2}$.

Y	-10	10
p	$\frac{1}{2}$	$\frac{1}{2}$

Ví dụ 1.2. Tung một con xúc xắc, gọi X là kết quả với quy ước nếu ra mặt 1 thì ghi số 1 ra một 2 thì ghi số 2, ... ra mặt 6 thì ghi số 6. Như vậy X có thể lấy các giá trị 1, 2, 3, 4, 5, 6 với xác suất bằng nhau và bằng $\frac{1}{6}$.

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

Nếu chỉ quan tâm đến số chẵn hay lẻ thì quy ước ghi kết quả Y như sau: nếu ra mặt lẻ thì ghi 0, nếu ra mặt chẵn thì ghi 1. Như vậy biến Y có thể lấy các giá trị 0 và 1 với xác suất bằng nhau và bằng $\frac{1}{2}$.

Y	0	1
p	$\frac{1}{2}$	$\frac{1}{2}$

Nếu quan tâm đến việc ra mặt 6 thì quy ước ghi kết quả Z như sau: 0 nếu ra mặt nhỏ hơn 6, 1 nếu ra mặt 6. Như vậy Z sẽ lấy giá trị 0 với xác suất $\frac{5}{6}$ và lấy giá trị 1 với xác suất $\frac{1}{6}$

Z	0	1
p	$\frac{5}{6}$	$\frac{1}{6}$

Ví dụ 1.3. Trồng 10 cây, xác suất sống của mỗi cây là 0,8. Coi việc trồng các cây là các phép thử lặp (y hệt và độc lập), số cây sống X có thể là 0, 1, 2, . . . , 10 với các xác suất khác nhau tính theo công thức (sẽ trình bày kĩ ở phần phân phối nhị thức):

$$p_k = p(X = k) = C_n^k 0,8^k.0,2^{10-k}, k = \overline{1,10}$$

X	0	1	2	...	k	...	10
p	p_0	p_1	p_2	...	p_k	...	p_{10}

Ví dụ 1.4. Trong hộp có 4 bi trắng, 2 bi đen, lấy ngẫu nhiên 2 bi. Gọi X là số bi trắng có trong 2 bi lấy ra, ta thấy X có thể là 0, 1, 2 với các xác suất tính lần lượt như sau (sẽ trình bày ở phần phân phối siêu bội):

$$p(X = 0) = \frac{C_4^0.C_2^2}{C_6^2} = \frac{1}{15};\ p(X = 1) = \frac{C_4^1.C_2^1}{C_6^2} = \frac{8}{15};\ p(X = 2) = \frac{C_4^2.C_2^0}{C_6^2} = \frac{6}{15}.$$

X	0	1	2
p	$\frac{1}{15}$	$\frac{8}{15}$	$\frac{6}{15}$

Qua các ví dụ trên ta thấy:
Cho một phép thử có tập hợp các biến cố sơ cấp Ω và một hàm X xác định trên các biến cố sơ cấp. Nếu biết được tất cả các giá trị x_i của X và các xác suất tương ứng $p_i = p(X = x_i)$, nhưng không biết khi tiến hành phép thử X sẽ lấy giá trị nào trong các x_i thì X được gọi là đại lượng ngẫu nhiên hay biến ngẫu nhiên liên kết với phép thử đã cho.

Một ĐLNN được gọi là rời rạc nếu tập giá trị mà nó có thể nhận là tập hữu hạn hoặc vô hạn đếm được. Đối với ĐLNN rời rạc ta có thể liệt kê được các giá trị của nó.
Một ĐLNN được gọi là liên tục nếu tập giá trị mà nó có thể nhận được có thể lấp kín cả một khoảng trên trục số. Đối với ĐLNN liên tục, không thể liệt kê được các giá trị của nó.
Ví dụ: Chiều cao của học sinh trong một lớp học, khối lượng của một loại hoa quả là những ĐLNN liên tục.

§2 BẢNG PHÂN PHỐI VÀ HÀM PHÂN PHỐI

Cho một biến ngẫu nhiên X có các giá trị có thể x_i và các xác suất tương ứng p_i. Ghi lại x_i và p_i vào một bảng, gọi là bảng (hay dãy) phân phối:

X	x_1	x_2	...	x_n
p	p_1	p_2	...	p_n

(2.1)

Các biến cố $(X = x_i)$ $i = 1, 2, ..., n$ là các biến cố xung khắc có tổng xác suất bằng 1, như vậy các biến cố nói trên là một hệ biến cố đầy đủ.

Ngoài cách theo dõi X bằng bảng phân phối, có thể theo dõi X bằng hàm phân phối $F(x)$.

Hàm phân phối $F(x)$ được định nghĩa như sau: Cho $x, F(x)$ là xác suất của biến cố $X < x$, tức là xác suất để biến ngẫu nhiên X lấy các giá trị nhỏ hơn x (hay còn gọi là bên trái x)

$$F(x) = p(X < x).$$

Nếu có dãy phân phối (2.1) thì có thể tìm được hàm phân phối $F(x)$: $x \leq x_1$ bên trái x_1 không có giá trị nào của X nên $F(x_1) = p(X < x_1) = 0$, $x_1 < x \leq x_2$ bên trái x_2 có giá trị x_1 nên $F(x_2) = p(X < x_2) = p(X = x_1) = p_1$, $x_2 < x \leq x_3$ bên trái x_3 có giá trị x_1 và x_2 nên $F(x_3) = p(X < x_3) = p(X = x_1 \cup x_2) = p_1 + p_2$.
$x > x_k$ tất cả các giá trị có thể của X đều ở bên trái x nên $F(x) = p(X > x_n) = 1$.

$$F(x) = \begin{cases} 0 & \text{nếu } x \leq x_1 \\ p_1 & \text{nếu } x_1 < x \leq x_2 \\ p_1 + p_2 & \text{nếu } x_2 < x \leq x_3 \\ p_1 + p_2 + p_3 & \text{nếu } x_3 < x \leq x_4 \\ \ldots & \\ 1 & \text{nếu } x_n < x. \end{cases}$$

Trong Ví dụ 1.2 ta có dãy phân phối:

Z	0	1
p	$\frac{5}{6}$	$\frac{1}{6}$

Hàm phân phối:

$$F(x) = \begin{cases} 0 & \text{nếu } x \leq 0 \\ \frac{5}{6} & \text{nếu } 0 < x \leq 1 \\ 1 & \text{nếu } 1 < x. \end{cases}$$

Trong Ví dụ 1.4

X	0	1	2
p	$\frac{1}{15}$	$\frac{8}{15}$	$\frac{6}{15}$

Hàm phân phối:

$$F(x) = \begin{cases} 0 & \text{nếu } x \leq 0 \\ \frac{1}{15} & \text{nếu } 0 < x \leq 1 \\ \frac{9}{15} & \text{nếu } 1 < x \leq 2 \\ 1 & \text{nếu } 2 < x. \end{cases}$$

§3 CÁC SỐ ĐẶC TRƯNG

Đối với các biến ngẫu nhiên nếu có bảng phân phối (hoặc hàm phân phối) thì coi như có sự hiểu biết đầy đủ về biến.

Trong một số vấn đề không cần phải biết đầy đủ như vậy mà chỉ cần biết một số số đặc trưng cho dãy phân phối về một khía cạnh nào đó.

Người ta chia các số đặc trưng thành 2 nhóm: nhóm đặc trưng cho vị trí và nhóm đặc trưng cho độ phân tán.

Nhóm đặc trưng cho vị trí gồm một số số như: kỳ vọng, trung vị, mod, tứ phân vị dưới, tứ phân vị trên ...

Nhóm đặc trưng cho độ phân tán (hay còn gọi là đặc trưng cho độ tập trung) gồm phương sai, độ lệch chuẩn, biên độ, hệ số biến động, ...

Ở đây chúng ta chỉ xem xét hai số đặc trưng là kỳ vọng và phương sai.

3.1 Kỳ vọng

Kỳ vọng, ký hiệu là $M(X)$ hay MX hay EX, được tính theo công thức:

$$MX = \sum_{i=1}^{N} x_i p_i. \tag{3.1}$$

Phương sai, ký hiệu là $D(X)$ hay DX, VX, Var$X$ được tính theo công thức:

$$DX = \sum_{i=1}^{N} (x_i - MX)^2 p_i. \tag{3.2}$$

Khai triển bình phương ta có cách tính thứ hai:

$$DX = \sum_{i=1}^{N} x_i^2 p_i - (MX)^2. \tag{3.3}$$

Trong Ví dụ 1.1
$MX = 0.\frac{1}{2} + 1.\frac{1}{2} = \frac{1}{2}$; $DX = (0-\frac{1}{2})^2.\frac{1}{2} + (1-\frac{1}{2})^2.\frac{1}{2} = \frac{1}{4}$ hay $DX = 0^2.\frac{1}{2} + 1^2.\frac{1}{2} - (\frac{1}{2})^2 = \frac{1}{4}$.
Trong Ví dụ 1.2
$MZ = 0.\frac{5}{6} + 1.\frac{1}{6} = \frac{1}{6}$; $DZ = (0-\frac{1}{6})^2.\frac{5}{6} + (1-\frac{1}{6})^2.\frac{1}{6} = \frac{5}{36}$ hay $DZ = 0^2.\frac{5}{6} + 1^2\frac{1}{6} - (\frac{1}{6})^2 = \frac{5}{36}$.
Trong Ví dụ 1.4
$MX = 0.\frac{1}{15} + 1.\frac{8}{15} + 2.\frac{6}{15} = \frac{20}{15} = \frac{4}{3}$; $DX = 0^2.\frac{1}{15} + 1^2.\frac{8}{15} + 2^2.\frac{6}{15} - (\frac{4}{3})^2 = \frac{16}{45}$.

3.2 Tính chất của kỳ vọng và phương sai

Có thể chứng minh kỳ vọng có 3 tính chất sau:
a) Nếu C là hằng số thì $MC = C$
b) Nếu a là hằng số thì $M(aX) = aMX$

c) Nếu X và Y là 2 biến ngẫu nhiên thì $M(X+Y) = MX + MY$
*a) Coi C là trường hợp đặc biệt của biến ngẫu nhiên lấy 1 giá trị C với xác suất 1, do đó $MC = C.1 = C$.
*b) Đại lượng aX có các giá trị ax_i với xác suất p_i do đó

$$M(aX) = \sum_{i=1}^{n} ax_i p_i = a\sum_{i=1}^{n} x_i p_i = aMX$$

*c) Thừa nhận tính chất này.

Từ 3 tính chất trên có thể chứng minh: nếu a và b là hai hằng số thì $M(aX+b) = aMX+b$
* Thật vậy $M(aX+b) = M(aX) + M(b) = aMX + b$.

Có thể chứng minh phương sai DX có ba tính chất sau:
a) $DC = 0$
b) $D(aX) = a^2 DX$
c) $D(X+Y)$ nói chung khác $DX + DY$, nhưng nếu X và Y là hai biến ngẫu nhiên độc lập theo nghĩa: các biến cố $(X = x_i), i = \overline{1,k}$ và $(Y = y_i), j = \overline{1,l}$ là các biến cố độc lập, nói cách khác hai biến X và Y liên kết với hai phép thử độc lập thì:

$$D(X+Y) = DX + DY.$$

Cách chứng minh tưng tự như đối với kỳ vọng (ở đây thừa nhận).
Từ b) và c) có thể suy ra $D(-Y) = DY$.
Từ 3 tính chất có thể suy ra:

$$D(aX+b) = a^2 DX.$$

Ví dụ 3.1. Tung hai đồng tiền, X là biến ngẫu nhiên liên kết với đồng tiền thứ nhất. Y là biến ngẫu nhiên liên kết với đồng tiền thứ hai, X và Y lấy giá trị 0 và 1 với xác suất $\frac{1}{2}$ tuỳ theo đồng tiền ra mặt ngửa hay sấp, còn Z là tổng $X+Y$, coi X và Y độc lập, ta có dãy phân phối của X, Y và Z

X	0	1
p	$\frac{1}{2}$	$\frac{1}{2}$

Y	0	1
p	$\frac{1}{2}$	$\frac{1}{2}$

Z	0	1	2
p	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

$$\begin{aligned}
&MZ = 0.\frac{1}{4} + 1.\frac{1}{2} + 2.\frac{1}{4} = 1; && DZ = 0^2.\frac{1}{4} + 1^2.\frac{1}{2} + 2^2.\frac{1}{4} - 1^2 = \frac{1}{2} \\
&MX + MY = \frac{1}{2} + \frac{1}{2} = 1; && DX + DY = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}; \\
&MZ = MX + MY = 1; && DZ = DX + DY = \frac{1}{2}.
\end{aligned}$$

Ví dụ 3.2. Tung hai con xúc xắc, X là số điểm trên con xúc xắc thứ nhất, Y là số điểm trên con xúc xắc thứ hai. Z là tổng số điểm trên hai xúc xắc: $Z = X + Y$

$$MX = MY = \frac{1+2+3+4+5+6}{6} = \frac{7}{2};$$

$$DX = DY = \frac{1^2+2^2+3^2+4^2+5^2+6^2}{6} - (\frac{7}{2})^2 = \frac{35}{12}$$

Z có phân phối

Z	2	3	4	5	6	7	8	9	10	11	12
p	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

$MZ = (2.1+3.2+4.3+5.4+6.5+7.6+8.5+9.4+10.3+11.2+12.1) : 36 = 7 = MX + MY$

$$DZ = \frac{2^2.1 + 3^2.2 + 4^2.3 + 5^2.4 + 6^2.5 + 7^2.6 + 8^2.5 + 9^2.4 + 10^2.3 + 11^2.2 + 12^2.1}{36} - 7^2$$
$$= \frac{35}{6} = DX + DY.$$

§4 BIẾN NGẪU NHIÊN RỜI RẠC CÓ VÔ SỐ GIÁ TRỊ

Trong §1 ta đưa ra khái niệm biến ngẫu nhiên rời rạc có một số hữu hạn giá trị $x_1, x_2, ..., x_n$.
Sau đây là hai ví dụ về biến ngẫu nhiên rời rạc có vô số giá trị.

Ví dụ 4.1. Một người đi bắn, xác suất trúng đích là 0,4. Người đó quyết tâm bắn cho đến khi bắn trúng mới về, giả thiết thêm là số đạn không bị hạn chế. Gọi X là số đạn đã dùng cho đến khi về, ta có bảng phân phối:

X	1	2	...	k	...
p	0,4	0,6.0,4	...	$0,6^{k-1}.0,4$	...

Ví dụ 4.2. Một lô hàng gồm rất nhiều sản phẩm, tỉ lệ phế phẩm là 20%. Người kiểm tra chọn lần lượt các sản phẩm ra cho đến khi phát hiện phế phẩm thì dừng. Gọi X là số sản phẩm đã kiểm tra cho đến khi kết thúc, ta có bảng phân phối:

X	1	2	...	k	...
p	0,2	0,8.0,2	...	$0,8^{k-1}$.0,2	...

Đối với biến rời rạc có vô số giá trị ta cũng có các số đặc trưng như đối với biến rời rạc có hữu hạn giá trị, tuy nhiên việc tính toán khó hơn.
Gọi p là xác suất thành công trong một phép thử, $q = 1 - p$ là xác suất thất bại. Làm các phép thử lần lượt cho đến khi thành công ta có dãy phân phối

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

Dùng cách tính tổng một chuỗi ta có $MX = \frac{1}{p}; DX = \frac{q}{p^2}$
Trong Ví dụ 4.1: $p = 0,6; q = 0.4$; $MX = \frac{1}{0,4}; DX = \frac{0,6}{0,16}$.
Trong Ví dụ 4.2: $p = 0,2; q = 0,8$; $MX = \frac{1}{0,2}; DX = \frac{0,8}{0,04}$.

Bài tập

2.1. Tỉ lệ học sinh lên lớp của một trường là 0,9. Gặp ngẫu nhiên hai em học sinh, gọi X là số em được lên lớp trong hai em đó. Viết bảng phân phối và hàm phân phối của X. Tính kỳ vọng MX và phương sai DX.

2.2. Trong số 10 hạt giống đem trồng có 7 hạt ra hoa vàng, 3 hạt ra hoa trắng. Lấy ngẫu nhiên 2 hạt. Gọi X là số hạt ra hoa vàng, tìm bảng phân phối và hàm phân phối của X. Tính MX và DX.

2.3. Tỉ lệ chính phẩm do một máy sản xuất ra là 90%. Kiểm tra 5 sản phẩm. gọi X là số phế phẩm trong 5 sản phẩm. Viết bảng phân phối và hàm phân phối của X. Tính MX và DX.

2.4. Một bác sĩ thú y chữa bệnh cho bò với xác suất chữa khỏi 0,8. Một nhóm 5 con bò bị bệnh được đem đến để bác sĩ chữa, gọi X là số con khỏi bệnh. Viết bảng phân phối và hàm phân phối của X. Tính MX và DX.

2.5. Một học sinh đi thi ngoại ngữ để lấy chứng chỉ, xác suất thi đỗ là 0,3, nếu không đỗ thì phải thi lại cho đến khi đỗ thì thôi. Gọi X là số lần thi. Viết bảng phân phối của X và kỳ vọng của X.

2.6. Một người trồng 2 cây cảnh, xác suất để cây thứ nhất ra hoa là 0,4, xác suất để cây thứ hai ra hoa là 0,6. Gọi X là số cây ra hoa, viết bảng phân phối và hàm phân phối của X, tính kỳ vọng MX và phương sai DX.

§5 MỘT SỐ PHÂN PHỐI RỜI RẠC THƯỜNG GẶP

Trong các mục trước chúng ta đã đề cập đến biến ngẫu nhiên rời rạc, bảng phân phối và hàm phân phối.

Trong các ngành, nghề khi khảo sát các hiện tượng tự nhiên hoặc xã hội sẽ gặp các biến ngẫu nhiên có các phân phối khác nhau nhưng thông thường hay gặp một số phân phối sau đây:

a) Phân phối Bec-nu-li (hay còn gọi là phân phối (0, 1), ký hiệu là $A(p)$).

Biến ngẫu nhiên X phân phối Bec-nu-li nếu bảng phân phối có dạng:

X	0	1
p	$q = 1 - p$	p

Phân phối này có kỳ vọng $MX = p$ và phương sai $DX = pq$.

Phân phối Bec-nu-li gắn liền với một phép thử có hai kết quả đối lập, một kết quả quy ước gọi là 1 hay thành công, có xác suất p, kết quả kia quy ước gọi là 0 hay thất bại, có xác suất $q = 1 - p$.

Ví dụ 5.1. Gieo xúc xắc, gọi X là số lần ra mặt chẵn. X lấy giá trị 1 (chẵn) với xác suất $p = \frac{1}{2}$, giá trị 0 (lẻ) với xác suất $q = \frac{1}{2}$; $MX = \frac{1}{2}$; $DX = \frac{1}{4}$.

Sinh con, gọi X là số con trai. X lấy giá trị 1 (trai) với xác suất $p = \frac{1}{2}$, giá trị 0 (gái) với xác suất $q = \frac{1}{2}$; $MX = \frac{1}{2}$; $DX = \frac{1}{2}$.

Ấp một quả trứng, gọi X là số trứng nở. X lấy giá trị 1 (nở) với xác suất $p = 0,8$, giá trị 0 (không nở) với xác suất $q = 0,2; MX = 0,8; DX = 0,16$.

Một học sinh đi thi, gọi X là kết quả thi. X lấy giá trị 1 (đỗ) với xác suất $p = 0,9$. Giá trị 0 (trượt) với xác suất $q = 0,1; MX = 0,9; DX = 0,09$.

Kiểm tra một sản phẩm, gọi X là số sản phẩm tốt, X lấy giá trị 1 (sản phẩm tốt) với xác suất $p = 0,8$, giá trị 0 (sản phẩm hỏng) với xác suất $q = 0,2$; $MX = 0,8; DX = 0,16$.

b) Phân phối nhị thức

Biến ngẫu nhiên X phân phối nhị thức nếu bảng phân phối có dạng:

X	0	1	...	k	...	n
p	p_0	p_1	...	p_k	...	p_n

$$p_k = p(X = k) = C_n^k p^k q^{n-k}. \tag{5.1}$$

Phân phối này có:

$$\text{Kỳ vọng } MX = np. \tag{5.2}$$

$$\text{Phương sai } DX = npq. \tag{5.3}$$

Giá trị có xác suất lớn nhất Mod$X$ là số nguyên thoả mãn bất đẳng thức kép

$$np - q \leq \text{Mod}X \leq np + p. \tag{5.4}$$

Phân phối nhị thức gắn liền với việc lặp lại n lần một phép thử có hai biến cố đối lập (thành công và thất bại) với X là số lần thành công. Lặp ở đây có nghĩa là dãy phép thử được tiến hành trong cùng điều kiện và độc lập với nhau. Phân phối nhị thức thường ký hiệu là $B(n, p)$.

Ví dụ 5.2. Gia đình có 2 con, xác suất sinh con trai là 0,5. Coi các lần sinh là các phép thử độc lập. Số con trai X phân phối $B(2, 0, 5)$ với $p = \frac{1}{2}; q = \frac{1}{2}; n = 2$.

X	0	1	2
p	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

$MX = 1; DX = \frac{1}{2}; \text{Mod}X = 1.$

Ví dụ 5.3. Gieo 4 hạt đậu, xác suất để một hạt cho cây ra hoa vàng là 0,75, ra hoa trắng là 0,25. Số cây đậu ra hoa vàng X phân phối nhị thức $B(4; 0, 75)$.

X	0	1	2	3	4
p	$0,25^2$	$4.0,75.0,25^3$	$6.0,75.0,25^2$	$4.0,75^3.0,25$	$0,75^4$

$MX = 4.0,75 = 3; DX = 4.0,75.0,25 = 0,75; \text{Mod}X = 3.$

c) Phân phối siêu bội

Cho $N, M (M < N)$ và một số $n \leq \min(M, N - M)$.

Biến ngẫu nhiên X phân phối siêu bội hay siêu hình học nếu bảng phân phối có dạng:

X	0	1	...	k	...	N
p	p_0	p_1	...	p_k	...	p_n

$$p_k = \frac{C_M^k C_{N-M}^{n-k}}{C_N^n}, \ k = \overline{0, n}. \tag{5.5}$$

Phân phối này có:

$$\text{Kỳ vọng} \qquad MX = n.\frac{M}{N} \tag{5.6}$$

$$\text{Phương sai} \qquad DX = n.\frac{M}{N}.\frac{N-M}{N}.\frac{N-n}{N-1} \tag{5.7}$$

Cho một hộp đựng N bi trong đó có M bi trắng, $N - M$ bi đen. Lấy ngẫu nhiên một lúc hoặc lấy lần lượt không hoàn lại một nhóm n bi. Số bi trắng X trong nhóm phân phối siêu bội.

Phân phối siêu bội thường ký hiệu là $M(N, n)$.

Nếu không có điều kiện $n \le \min(M, N-M)$ thì các giá trị có thể của biến X không phải từ 0 đến n mà ít hơn (bớt một số giá trị đầu hay bớt một số giá trị cuối), nhưng các xác suất vẫn tính theo (5.5) *và vẫn gọi là phân phối siêu bội. Kỳ vọng và phương sai vẫn tính theo* (5.6) *và* (5.7).

Gọi tỉ số bi trắng trong hộp là $p = \frac{M}{N}$.

Nếu lấy có hoàn lại n lần (tức là lấy một bi, xem xong hoàn trả vào hộp, trộn đều sau đó lấy ngẫu nhiên ra một bi khác) thì số bi trắng X phân phối nhị thức $B(n, p)$. Như vậy siêu bội và phân phối nhị thức có những nét giống nhau chỉ khác ở chỗ nếu lấy n bi không hoàn lại thì số bi trắng X phân phối siêu bội còn nếu có hoàn lại thì X phân phối nhị thức. Sự khác nhau trở nên không đáng kể nếu tổng số bi N và số bi trắng M là các số rất lớn.

Ví dụ 5.4. Chọn một uỷ ban gồm 3 người trong số 3 nữ và 5 nam. Gọi X là số nữ trong uỷ ban, X có phân phối siêu bội:

X	0	1	2	3
p	$\frac{10}{56}$	$\frac{30}{56}$	$\frac{15}{56}$	$\frac{1}{56}$

$MX = \frac{9}{8}; DX = \frac{225}{448}$.

Ví dụ 5.5. Hộp có 15 quả cam trong đó có 5 quả hỏng, lấy 2 quả. Gọi X là số cam hỏng trong 2 quả đó ta có:

X	0	1	2
p	$\frac{45}{105}$	$\frac{50}{105}$	$\frac{10}{105}$

$MX = \frac{2}{3}; DX = \frac{26}{63}$.

d)Phân phối Poát-xông

Biến ngẫu nhiên X phân phối Poát-xông nếu bảng phân phối có dạng:

X	0	1	2	...	k	...
p	p_0	p_1	p_2	...	p_k	...

$$p_k = \frac{e^{-\mu}}{k!}\mu^k \text{ (μ là một hằng số, } k = 0, 1, ...\infty). \tag{5.8}$$

Phân phối này có:

$$MX = DX = \mu. \tag{5.9}$$

Ví dụ 5.6. Chuyển 5000 quả trứng vào kho với xác suất vỡ của mỗi quả là 0,0004. Tính xác suất để khi vận chuyển có không quá một quả bị vỡ. Gọi X là số quả bị vỡ, ở đây có thể dùng phân phối nhị thức nhưng vì $n = 5000$ quá lớn, $p = 0,0004$ lại quá bé nên có thể coi X phân phối xấp xỉ phân phối poát - xông với $\mu = np = 2$ từ đó có thể tính:

Xác suất để có không quá một quả bị vỡ bằng xác suất để $X = 0(p_0)$ cộng xác suất để $X = 1(p_1)$.

$$p_0 = \frac{e^{-2}.2^0}{0!} = \frac{1}{e^2}; p_1 = \frac{e^{-2}.2^1}{1!} = \frac{2}{e^2}$$

$$P(0 \le X \le 1) = p_0 + p_1 = \frac{3}{e^2} = 0,406.$$

$MX = DX = \mu = 2$.

Ví dụ 5.7. Gieo 10000 hạt giống, xác suất để hạt lép là $p = 0,0005$. Tính xác suất để có đúng 6 hạt lép. Lấy $\mu = np = 5$, ta có $p_6 = \frac{e^2.5^6}{6!}$

e) Phân phối hình học

Biến cố A có xác suất xuất hiện trong một phép thử là p. Lần lượt thực hiện phép thử cho đến khi A xuất hiện. Số lần thực hiện phép thử cho đến khi A xuất hiện là biến cố X có phân phối hình học. Bảng phân phối có dạng:

X	1	2	...	k	...
p	p_1	p_2	...	p_k	...

$$p_k = P(X = k) = q^{k-1}p \text{ với } q = 1 - p,\ k = 1, 2, \dots \tag{5.10}$$

$$\text{Kỳ vọng} \qquad MX = \frac{1}{p} \tag{5.11}$$

$$\text{Phương sai} \qquad DX = \frac{q}{p^2} \tag{5.12}$$

Ví dụ 5.8. Lô hàng khá lớn có 20% phế phẩm. Kiểm tra lần lượt cho đến khi phát hiện phế phẩm.

Gọi X là số sản phẩm đã kiểm tra, X phân phối hình học với $p = 0,2$.
Kì vọng $MX = \frac{1}{0,2} = 5$; Phương sai $DX = \frac{0,8}{0,04} = 20$.

Ví dụ 5.9. Phát tín hiệu trên lạc với trạm bạn, xác suất nhận được là 0,4. Nếu trạm bạn báo đã nhận được tín hiệu thì dừng, nếu không thì phát tiếp. Gọi X là số tín hiệu đã phát cho đến khi dừng, X phân phối hình học với $p = 0,4$. Kì vọng $MX = \frac{1}{0,4} = 2,5$; Phương sai $DX = \frac{0,6}{0,16}$.

Bài Tập

2.7. Trồng 5 cây, xác suất để cây ra hoa là 0,2. Tính xác suất đề ít nhất có 3 cây ra hoa.

2.8. Ấp 12 quả trứng, xác suất trứng nở là $\frac{1}{3}$. Tính xác suất để:
a) Có 4 trứng nở.
b) Có từ 3 đến 6 trứng nở.

2.9. Việc sản xuất ra các sản phẩm được tiến hành độc lập. Phải sản xuất mỗi đợt bao nhiêu sản phẩm để trung bình có 10 sản phẩm đạt tiêu chuẩn, biết rằng xác suất để một sản phẩm làm ra đạt tiêu chuẩn là 0,8.

2.10. Mỗi chậu ươm 2 hạt giống hai loại, hạt loại 1 có xác suất nẩy mầm 0,8, hạt loại 2 có xác suất nẩy mầm 0,6. Có tất cả 15 chậu, gọi X là số chậu mà cả 2 hạt đều nẩy mầm. Tìm giá trị hay gặp nhất của X.

2.11. Trồng cây, xác suất sống mỗi cây là 0,4. Phải trồng bao nhiêu cây để nhiều khả năng nhất có 25 cây sống.

2.12. Khi tiêm phòng dịch thì cứ một lô gà 50 con thường thấy có 30 con không mắc bệnh. Tính xác suất để gà không mắc bệnh khi tiêm chủng.

2.13. Có 9 học sinh vào quán ăn, 4 trong số đó dưới 16 tuổi. Lúc đầu chủ quán mang ra 5 cốc bia và nói là dành cho các em lớn tuổi, 4 cốc nước hoa quả sẽ mang ra sau. Các học sinh lại phân phát 5 cốc một cách ngẫu nhiên.
Tính xác suất để có 2 em dưới 16 tuổi uống bia.

2.14. Một chuồng gà có 3 mái và 5 trống. Bắt ra 5 con, gọi X là số gà mái. Tìm bảng phân phối của X.

2.15. Khảo sát ở một vùng thấy cứ 1000 người thì có một người nghiện rượu. Tính xác suất để trong một khu vực dân cư 8000 người có ít hơn 7 người nghiện rượu.

2.16. Trung bình trên $10000m^2$ ruộng có 100 con chuột đồng. Tính xác suất để tại một mảnh ruộng rộng $1000m^2$ có nhiều hơn 15 con chuột.

2.17. Xác suất sinh con trai là 0,515. Tính xác suất để trong 200 trẻ mới sinh có 95 em gái.

2.18. Xác suất để khi trồng thì cây sống là 0,8. Trồng 100 cây, tính xác suất để:
a) Có từ 75 đến 90 cây sống.
b) Ít hơn 75 cây sống.
c) Nhiều hơn 90 cây sống.

Chương 3

MẪU QUAN SÁT VÀ BÀI TOÁN ƯỚC LƯỢNG

§1 TỔNG THỂ VÀ MẪU QUAN SÁT

1.1 Tổng thể

Tổng thể (còn được gọi là tập hợp chính), là tập hợp tất cả các phần tử do mục đích và phạm vi vấn đề cần nghiên cứu qui định.

Đối với tổng thể, ta sử dụng một số khái niệm và ký hiệu sau:

i) N: Số phần tử của tổng thể và được gọi là kích thước (cỡ) của tổng thể.

ii) $\mathcal{H}$: Dấu hiệu mà ta khảo sát (trong kinh tế được gọi là chỉ tiêu, trong vật lý gọi là đại lượng). Cần nhấn mạnh rằng, ta không nghiên cứu trực tiếp bản thân tổng thể mà chỉ nghiên cứu dấu hiệu $\mathcal{H}$ của nó.

iii) x_i, $i = \overline{1, k}$: là những giá trị của dấu hiệu $\mathcal{H}$ đo được trên các phần tử của tổng thể, x_i là thông tin mà ta cần đến, còn phần tử của tổng thể là vật mang thông tin.

iv) n_i, $i = \overline{1, k}$: tần số của x_i là số phần tử của tổng thể có chung giá trị x_i ấy.

v) p_i, $i = \overline{1, k}$: tần suất của x_i là tỷ số giữa tần số của x_i và kích thước của tổng thể $p_i = \dfrac{n_i}{n}$.

Biểu diễn sự tương ứng của các giá trị x_i và tần suất p_i được gọi là bảng cơ cấu của tổng thể theo dấu hiệu $\mathcal{H}$.

Bảng này có dạng:

Giá trị của $\mathcal{H}$	x_1	x_2	...	x_i	...	x_k
Tần suất p_i	p_1	p_2	...	p_i	...	p_k

Bảng này mô tả đầy đủ dấu hiệu $\mathcal{H}$, nhưng phải sử dụng nhiều số liệu. Vì vậy để phân tích dấu hiệu $\mathcal{H}$ người ta thường tóm tắt bảng trên bằng các số đặc trưng sau đây:

a) *Trung bình* của dấu hiệu $\mathcal{H}$ hay trung bình của tổng thể, ký hiệu là $\overline{x}$ và được xác định bởi:

$$\overline{x} = \frac{\sum_{i=1}^{k} n_i x_i}{n} = \sum_{i=1}^{k} p_i x_i.$$

b) *Phương sai mẫu* của dấu hiệu $\mathcal{H}$ ký hiệu s^2 và được xác định bởi

$$s^2 = \frac{\sum_{i=1}^{k} n_i x_i^2 - n\overline{x}^2}{(n-1)}.$$

c) Căn bậc hai của s^2 là độ lệch chuẩn s.

Ví dụ 1.1. Điều tra năng suất lúa trên diện tích 100 hecta trồng lúa của một vùng ta thu được bảng số liệu sau

Năng suất (tạ/ha)	25	30	35	40	45
Số ha có năng suất tương ứng	10	20	25	20	25

Tính giá trị trung bình mẫu, phương sai mẫu và độ lệch chuẩn của mẫu cụ thể trên.

Giải. Trung bình mẫu

$$\overline{x} = \frac{\sum_{i=1}^{k} n_i x_i}{n} = \frac{3650}{100} = 36,5.$$

Phương sai mẫu

$$s^2 = \frac{\sum_{i=1}^{k} n_i x_i^2 - n\overline{x}^2}{(n-1)} = \frac{1}{99}(137500 - 100.36,5^2) = 43,2.$$

Độ lệch chuẩn

$$s = \sqrt{s^2} = \sqrt{43,2} \approx 6,55.$$

1.2 Khái niệm mẫu

Khi nghiên cứu một đặc điểm, tính chất nào đó của tổng thể ta có thể tiến hành theo hai phương pháp sau:

a) *Phương pháp điều tra toàn bộ*: Mọi phần tử của tổng thể đều được khảo sát.

Ưu điểm: các kết luận rút ra phản ánh đúng bản chất của tổng thể.

Nhược điểm:

- Chi phí lớn về tiền của, thời gian, nhân lực, phương tiện ...
- Quá trình điều tra cũng chính là quá trình phá hủy các phần tử được điều tra.
- Có những trường hợp ta không xác định được toàn bộ N phần tử của tổng thể ...

Chính vì lý do trên nên phương pháp điều tra toàn bộ ít được thực hiện.

b) *Phương pháp điều tra mẫu:*

- Từ tổng thể ta lấy ra n phần tử (tập con của tổng thể) $n << N$ và đo lường giá trị của dấu hiệu $\mathcal{H}$ trên chúng.
- Từ đó rút ra các kết luận khoa học trên mẫu rồi suy rộng cho toàn bộ tổng thể.

Ưu điểm:

- Thu thập, xử lý và khai thác nhanh.
- Toàn diện.

Yêu cầu: Mẫu phải đại diện được cho tổng thể do đó khi lấy mẫu phải đảm bảo tính ngẫu nhiên của mẫu, không chọn mẫu theo một tiêu chuẩn chủ quan định trước.

1.3 Cách chọn mẫu

a) Mẫu có hoàn lại (có lặp):

Trong tổng thể gồm N phần tử ta chọn một phần tử khảo sát và ghi lại kết quả X_1. Trả lại phần tử đó vào tổng thể trước khi chọn phần tử tiếp theo để khảo sát ... , cứ lặp lại như thế đến lần thứ n ta nhận được một mẫu với số liệu về dấu hiệu đang khảo sát là $(X_1, X_2, ..., X_n)$. Mẫu này được gọi là mẫu ngẫu nhiên hoàn lại (có lặp).

b) Mẫu ngẫu nhiên không hoàn lại (không lặp):

Từ tổng thể gồm N phần tử, ta chọn ra một phần tử, khảo sát và ghi lại kết quả X_1. Bỏ phần tử đó sang một bên trước khi chọn phần tử tiếp theo để khảo sát tiếp, ... cứ lặp lại như thế cho đến lần thứ n ta được mẫu với số liệu về dấu hiệu đang khảo sát là $(X_1, X_2, ..., X_n)$. Mẫu này được gọi là mẫu ngẫu nhiên không hoàn lại (không lặp).

Chú ý: Hai mẫu nói trên được gọi là mẫu ngẫu nhiên đơn giản. Nhờ các định lý giới hạn trong lý thuyết xác suất người ta đã chứng minh được rằng khi số phần tử tổng thể vừa đủ lớn thì có thể coi hai mẫu có lặp và không lặp là như nhau.

Có thể kể thêm một số phương pháp sau:

c) Mẫu được chọn theo phương pháp cơ học:

d) Phương pháp điển hình:

e) Phương pháp phân dãy:

f) Sắp xếp các số liệu thực nghiệm theo nhóm:

§2 ƯỚC LƯỢNG THAM SỐ CỦA TỔNG THỂ

Như chúng ta biết, các số đặc trưng của dấu hiệu $\mathcal{H}$ như trung bình, phương sai ... được sử dụng rộng rãi trong phân tích kinh tế, xã hội và các lĩnh vực khác. Nhưng các số đặc trưng này thường chưa biết, vì vậy đặt ra vấn đề cần ước lượng chúng bằng phương pháp mẫu. Sau khi đã mô hình hoá dấu hiệu $\mathcal{H}$ bằng một ĐLNN và cơ cấu tổng thể bằng qui luật phân phối xác suất của X, ta có thể phát biểu vấn đề thực tế nêu trên dưới dạng toán học như sau: Cho ĐLNN X có thể đã biết hoặc chưa biết qui luật phân phối xác suất của X, nhưng chưa biết tham số θ nào đó của nó. Hãy ước lượng θ bằng phương pháp mẫu (dựa trên cở sở một mẫu thống kê nào đó). Bài toán này là một trong những bài toán cơ bản của thống kê toán. Vì θ là một hằng số nên có thể dùng một số nào đó để ước lượng θ, ước lượng như vậy được gọi là *ước lượng điểm* (nếu ta đưa chọn số dùng để ước lượng θ lên trục số thì nó tương ứng với một điểm). Ngoài ước lượng điểm người ta còn dùng phương pháp *ước lượng khoảng*, tức là chỉ ra một khoảng số $[\theta_1, \theta_2]$ nào đó có thể chứa được θ. Cận trên và cận dưới của khoảng được tính theo quy tắc cụ thể dựa trên các thống kê và dựa trên mức tin cậy P.

2.1 Ước lượng lỳ vọng μ của phân phối chuẩn khi biết phương sai σ^2

Các bước cần làm để ước lượng μ.

+ Chọn mẫu dung lượng n, tính trung bình cộng $\overline{x}$. Chọn mức tin cậy P. ($\alpha = 1 - P$ gọi là mức sai cho phép hay mức ý nghĩa)

+ Dùng bảng tính giá trị tới hạn $u(\frac{\alpha}{2})$, tức là giá trị u sao cho $\Phi(u) = 1 - \dfrac{\alpha}{2}$.

Chú ý: hàm Φ được xác định bởi $\Phi(t) = \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{-\infty}^{t} e^{-\frac{x^2}{2}} dx$

+ Ước lượng μ theo bất đẳng thức kép

$$\overline{x} - u(\frac{\alpha}{2})\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{x} + u(\frac{\alpha}{2})\frac{\sigma}{\sqrt{n}} \tag{2.1}$$

Ví dụ 2.1. Hãy tìm khoảng tin cậy cho chiều cao trung bình của sinh viên dựa trên một mẫu kích thước $n = 36$ với trung bình mẫu $\overline{x} = 160cm$, độ tin cậy $P = 95\%$. Giả sử độ lệch tiêu chuẩn $\sigma = 7cm$.
Giải. Ta có $\sigma = 7cm$, $n = 36$, $\alpha = 0,05$, $u(\frac{\alpha}{2}) = 1,96$. Vậy khoảng tin cậy $P = 95\%$ là

$$160 - 1,96\frac{7}{\sqrt{36}} \leq \mu \leq 160 + 1,96\frac{7}{\sqrt{36}}$$

$$157,71 \leq \mu \leq 162,29$$

Vậy với độ tin cậy $P = 95\%$, chiều cao trung bình μ nằm giữa 157,71 và 162,29.

Ví dụ 2.2. Cũng câu hỏi như trên nhưng cần tìm khoảng tin cậy có độ tin cậy là 99%.
Giải. Ta có $\sigma = 7cm$, $n = 36$, $\alpha = 0,01$, $u(\frac{\alpha}{2}) = 2,58$. Vậy khoảng tin cậy $P = 95\%$ là

$$160 - 2,58\frac{7}{\sqrt{36}} \leq \mu \leq 160 + 2,58\frac{7}{\sqrt{36}}$$

$$156,99 \leq \mu \leq 163,01.$$

Từ hai ví dụ trên, ta thấy. Trên cùng một kích thước mẫu, nếu độ tin cậy càng lớn thì độ dài khoảng tin cậy sẽ càng lớn.

2.2 Ước lượng kỳ vọng μ của phân phối chuẩn khi chưa biết phương sai σ^2, $n < 30$

Các bước để tìm ước lượng μ (với mức tin cậy $P = 1 - \alpha$)
+ Chọn mẫu dung lượng n, tính trung bình mẫu $\overline{x}$, tính phương sai mẫu s^2
+ Dùng bảng phân phối Student tính giá trị tới hạn $t\left(\frac{\alpha}{2}, n-1\right)$, tức là giá trị t ở cột $\frac{\alpha}{2}$ và dòng $n - 1$
Ước lượng μ theo bất đẳng thức kép

$$\overline{x} - t\left(\frac{\alpha}{2}, n-1\right)\frac{s}{\sqrt{n}} \leq \mu \leq \overline{x} + t\left(\frac{\alpha}{2}, n-1\right)\frac{s}{\sqrt{n}}. \tag{2.2}$$

Ví dụ 2.3. Để ước lượng năng suất một giống ngô, người ta theo dõi 25 mảnh ruộng. Sau khi thu hoạch được $\overline{x} = 10,6$; $s = 2,082$ (đơn vị tạ/ha). Giả thiết năng suất ngô phân phối chuẩn. Mức tin cậy $P = 0,95$.
Giải. Tra bảng ta được $t(0,025, 24) = 2,061$. Khoảng ước lượng

$$10,6 - 2,061\frac{2,082}{\sqrt{25}} \leq \mu \leq 10,6 + 2,061\frac{2,082}{\sqrt{25}}$$

$$9,74 \leq \mu \leq 11,46.$$

2.3 Ước lượng kỳ vọng μ của phân phối chuẩn khi chưa biết phương sai σ^2, $n \geq 30$

Trong nhiều bài toán thực tế, ta không biết phương sai của tập hợp chính, nếu kích thước mẫu $n \geq 30$ thì ta có thể xấp xỉ σ bởi s.

Khi đó khoảng ước lượng sẽ là

$$\overline{x} - u(\frac{\alpha}{2})\frac{s}{\sqrt{n}} \leq \mu \leq \overline{x} + u(\frac{\alpha}{2})\frac{s}{\sqrt{n}} \tag{2.3}$$

Ví dụ 2.4. Một trường đại học muốn nghiên cứu xem trung bình một sinh viên tiêu hết bao nhiêu tiền điện thoại trong một tháng, chọn ngẫu nhiên 59 sinh viên và thu được $\overline{x} = 41,05$, $s = 27,99$. Hãy xây dựng khoảng tin cậy 95% cho số tiền gọi điện thoại trung bình μ hàng tháng của một sinh viên.

Giải. Từ các số liệu trên ta có khoảng ước lượng là

$$41,05 - 1,96\frac{27,99}{\sqrt{59}} \leq \mu \leq 41,05 + 1,96\frac{27,99}{\sqrt{59}}$$

$$33,92 \leq \mu \leq 48,18.$$

2.4 Ước lượng xác suất p của phân phối nhị thức

Một tổng thể gồm 2 loại cá thể A và $\overline{A}$ với số lượng rất lớn, tỷ lệ loại A là p (chưa biết). Lấy ngẫu nhiên một cá thể, có thể coi xác suất được cá thể loại A là p.

Lấy ngẫu nhiên n cá thể, trong đó có m cá thể loại A.

Nếu n nhỏ thì có các bảng tính sẵn để ước lượng p căn cứ vào n và m.

Nếu n lớn (lý thuyết $n > 30$ nhưng thực tế chỉ nên dùng khi $n > 100$) thì coi m như biến ngẫu nhiên X phân phối nhị thức $B(n,p)$, sau đó dựa trên việc tính gần đúng phân phối nhị thức bằng phân phối chuẩn có kỳ vọng np và phương sai npq, ta tìm được quy tắc thực hành sau:

+ Lấy mẫu dung lượng n, đếm số cá thể loại A, gọi là tần số m, tính tần suất $f = \dfrac{m}{n}$.

Dùng bảng tính giá trị tới hạn $u(\frac{\alpha}{2})$ sau đó ước lượng p theo bất đẳng thức kép:

$$f - u(\frac{\alpha}{2})\sqrt{\frac{f(1-f)}{n}} \leq p \leq f + u(\frac{\alpha}{2})\sqrt{\frac{f(1-f)}{n}}. \tag{2.4}$$

Ví dụ 2.5. Trước ngày bầu cử tổng thống, một cuộc thăm dò dư luận được tiến hành. Người ta chọn ngẫu nhiên 100 người để hỏi ý kiến thì có 60 người nói rằng họ sẽ bỏ phiếu cho ông A. Tìm khoảng tin cậy cử tri bỏ phiếu cho ông A với độ tin cậy 90%.

Giải. Ta có $n = 100$, $m = 60$, $f = \dfrac{m}{n} = 0,6$, $\alpha = 0,1$, $u(\frac{\alpha}{2}) = 1,64$. Vậy khoảng tin cậy cho p là

$$0,6 - 1,64\sqrt{\frac{0,6.0,4}{100}} \leq p \leq 0,6 + 1,64\sqrt{\frac{0,6.0,4}{100}}$$

hay

$$0,52 \leq p \leq 0,68.$$

Như vậy ta kết luận: Với độ tin cậy 90% ông A sẽ thu được từ 52% đến 68% phiếu bầu. Như vậy ông A sẽ thu được ít nhất 52% phiếu và do đó sẽ trúng cử. Khẳng định này với độ tin cậy 90%.

Ví dụ 2.6. Trong một mẫu ngẫu nhiên gồm 200 người dùng xe máy, có 162 người dùng xe 100 phân khối trở lên. Tìm khoảng tin cậy với mức tin cậy 95% cho tỷ lệ người dùng xe trên 100 phân khối.

Giải. Ta có $n = 200$, $m = 162$, $f = 0,81$, $u(\frac{\alpha}{2}) = 1,96$. Khoảng tin cậy cho p là

$$0,81 - 1,96\sqrt{\frac{0,81.0,19}{200}} \leq p \leq 0,81 + 1,96\sqrt{\frac{0,81.0,19}{200}}$$

hay

$$0,755 \leq p \leq 0,865.$$

§3 XÁC ĐỊNH KÍCH THƯỚC MẪU

Với độ tin cậy P đã cho, ta thấy có mối quan hệ giữa kích thước mẫu n và độ dại khoảng tin cậy. Kích thước mẫu càng lớn thì độ dài khoảng tin cậy càng hẹp nghĩa là độ chính xác của ước lượng càng cao, sai số của ta càng nhỏ. Tuy nhiên kích thước mẫu càng càng lớn thì đòi hỏi ở nhà nghiên cứu càng nhiều thời gian, tiền của và công sức. Vậy bài toán đặt ra là: cần chọn kích thước mẫu tối thiểu là bao nhiêu để đạt được độ chính xác mong muốn.

3.1 Tính dung lượng mẫu khi ước lượng kỳ vọng μ của phân phối chuẩn

Theo công thức 2.1 ta có

$$|\overline{x} - \mu| \leq u(\frac{\alpha}{2})\frac{\sigma}{\sqrt{n}}.$$

Độ dài khoảng ước lượng là $2u(\frac{\alpha}{2})\frac{\sigma}{\sqrt{n}}$ và nửa độ dài của khoảng ước lượng là $u(\frac{\alpha}{2})\frac{\sigma}{\sqrt{n}}$. Nếu muốn ước lượng đạt độ chính xác ε thì phải lấy

$$u(\frac{\alpha}{2})\frac{\sigma}{\sqrt{n}} \leq \varepsilon \Leftrightarrow n \geq \left(\frac{u(\frac{\alpha}{2})\sigma}{\varepsilon}\right)^2.$$

Ví dụ 3.1. Đo chiều dài X (đơn vị cm). Giả thiết X phân phối chuẩn với $\sigma = 1,8cm$. Phải lấy mẫu có dung lượng bao nhiêu để ước lượng của kỳ vọng μ có độ chính xác $0,5cm$, $P = 0,99$.

Giải. Ta có $\sigma = 1,8cm$, $\varepsilon = 0,5$, $\alpha = 0,01$, $u(\frac{\alpha}{2}) = 2,575$. Phải lấy cỡ mẫu

$$n \geq \frac{(2,575.1,8)^2}{0,5^2} \approx 85,93$$

chọn kích thước mẫu $n = 86$.

3.2 Tính dung lượng mẫu khi ước lượng xác suất p của phân phối nhị thức

Theo công thức 2.4 nửa độ dài khoảng ước lượng là $u(\frac{\alpha}{2})\sqrt{\frac{f(1-f)}{n}}$. Nếu muốn ước lượng đạt độ chính xác ε thì phải lấy

$$u(\frac{\alpha}{2})\sqrt{\frac{f(1-f)}{n}} \leq \varepsilon \Leftrightarrow n \geq u^2(\frac{\alpha}{2})\frac{f(1-f)}{\varepsilon^2}. \tag{3.1}$$

Ta luôn có $\frac{1}{4} \geq f(1-f)$ do đó 3.1 đạt được nếu ta chọn

$$n \geq \frac{u^2(\frac{\alpha}{2})}{4\varepsilon^2}.$$

Ví dụ 3.2. Một xí nghiệp muốn ước lượng tỉ lệ p số người tán thành một kế hoạch sản xuất mới. Để tỷ lệ tìm ra có độ chính xác $\varepsilon = 0,06$ ở mức $P = 0,95$ phải thăm dò ít nhất bao nhiêu người?

Giải. $P = 0,95$; $\alpha = 0,05$; $u(\frac{\alpha}{2}) = 1,96$; $n \geq \frac{u^2(\frac{\alpha}{2})}{4\varepsilon^2} = 266,78$. Như vậy phải thăm dò ít nhất $n = 267$ người.

Bài Tập

3.1. Điều tra 200 mảnh ruộng, mỗi mảnh $4m^2$, ta được bảng sau:

x_i	1,02	1,08	1,14	1,20	1,26	1,32
n_i	10	15	35	75	55	10

a) Tính trung bình cộng và phương sai mẫu.
b) Ước lượng năng suất trung bình nếu coi năng suất phân phối chuẩn ($P = 0,95$).

3.2. Phỏng vấn 400 người ở một khu vực rất đông người thì có 240 người ủng hộ một dự luật. Ước lượng tỉ lệ p số người ủng hộ dự luật ở mức $P = 0,95$.

3.3. Thời gian đóng bột vào một bao phân phối chuẩn với $\sigma = 0,3$ phút.
a) Theo dõi 36 bao thấy thời gian trung bình để đóng một bao là 1,2 phút. Hãy ước lượng thời gian trung bình μ ở mức tin cậy $P = 0,95$.
b) Nếu muốn độ dài khoảng tin cậy giảm đi 2 lần thì phải chọn mẫu cỡ bao nhiêu?

3.4. Theo dõi lương của 50 công nhân trong một nhà máy ta có $\overline{x} = 79$ (đơn vị nghìn đồng), độ lệch chuẩn $s = 12,84$, số công nhân có lương cao hơn 90 là 14. Giả sử lương phân phối chuẩn. Với mức tin cậy $P = 0,95$.
a) Ước lượng kì vọng μ.
b) Ước lượng tỉ lệ công nhân có lương trên 90.

3.5. Cân 50 em học sinh lớp 4 được trọng lượng trung bình $\overline{x} = 32$ kg; độ lệch chuẩn $s = 2,5$ kg. Giả sử trọng lượng phân phối chuẩn, hãy ước lượng kỳ vọng ở mức $P = 0,99$.

3.6. Trọng lượng cam phân phối chuẩn với độ lệch chuẩn $\sigma = 32,4g$. Cân thử 9 quả được trọng lượng trung bình $\overline{x} = 187,9g$.
a) Ước lượng kỳ vọng ở mức $P = 0,80$.
b) Nếu muốn khoảng tin cậy bằng $[165,8 - 210]$ thì phải chọn mức tin cậy P bằng bao nhiêu?

3.7. Để xác định trọng lượng trung bình của 15 bao bột mì được đóng bằng máy tự động, người ta chọn ngẫu nhiên 15 bao và tính được $\overline{x} = 39,8kg$ và $s^2 = 0,144$. Tìm khoảng tin cậy 99% của trọng lượng trung bình μ của các bao bột mì?

3.8. Để khảo sát chiều cao trung bình μ của thanh niên trong một vùng A nào đó, một mẫu gồm 16 thanh niên được chọn, chiều cao của các thanh niên này đo được như sau:

Chiều cao (cm)	tần số	Chiều cao (cm)	tần số
163	1	170	3
164	4	172	2
166	3	174	3

Tìm khoảng tin cậy 95% cho μ.

3.9. Để khảo sát chiều cao trung bình μ của thanh niên trong một vùng A nào đó, một mẫu gồm 36 thanh niên được chọn, chiều cao của các thanh niên này đo được như sau:

Chiều cao (cm)	tần số	Chiều cao (cm)	tần số
163	7	170	4
164	2	172	6
166	3	174	5
168	5	175	4

Tìm khoảng tin cậy 95% cho μ.

3.10. Trong một cuộc khảo sát thị trường của công ty sản xuất thuốc lá, 150 người nghiện thuốc lá được chọn ngẫu nhiên, số điếu thuốc hút trung bình trong 1 tuần của nhóm người này là 135 với độ lệch tiêu chuẩn là 36. Tìm khoảng tin cậy 99% cho số điếu thuốc hút trung bình trong 1 tuần của tất cả những người nghiện thuốc lá?

3.11. Trong một cuộc thăm dò ý kiến 100 khách hàng, người ta thấy 55 người nói yêu thích mặt hàng A. Tìm khoảng tin cậy 95% cho tỷ lệ người tiêu dùng ưa thích mặt hàng A?

3.12. Một cơ quan Cảnh sát Giao thông kiểm tra hệ thống phanh của 40 chiếc xe tải trên đường Quốc lộ, họ phát hiện 14 chiếc có phanh chưa đảm bảo an toàn.

a) Hãy tìm khoảng tin cậy 98% cho tỷ lệ xe tải có phanh chưa an toàn?

b) Hãy tìm khoảng tin cậy 98% cho tỷ lệ xe tải có phanh tốt?

3.13. Trước ngày bầu cử, để biết tỷ lệ phần trăm các cử tri còn do dự chưa biết bỏ phiếu cho ai, người ta hỏi n cử tri được chọn một cách ngẫu nhiên. Hỏi số n tối thiểu là bao nhiêu để khoảng tin cậy 95% có độ dài không vượt quá 0,04?

3.14. Ta muốn xây dựng một khoảng tin cậy 95% cho trọng lượng trung bình của các gói đường đóng bằng máy tự động. Một mẫu điều tra sơ bộ cho ta $\overline{x} = 11,8kg$ với độ lệch tiêu chuẩn $s = 0,7kg$. Hãy ước lượng khoảng tin cậy cho kì vọng. Hỏi cần phải lấy kích thước mẫu tối thiểu bao nhiêu để đạt được sai số không vượt quá $0,2kg$.

3.15. Một khách sạn lớn tiến hành một nghiên cứu để xác định tỷ lệ phần trăm các khách ở trọ với thời gian nhiều hơn 1 ngày. Người chủ khách sạn muốn đạt được độ tin cậy 95% và sai số không vượt quá 0,05. Anh ta ước lượng sơ bộ tỷ lệ này khoảng 30%. Hỏi cần lấy mẫu với kích thước bao nhiêu?

3.16. Tương tự như bài 4.49 nhưng ở đây người chủ khách sạn không có trước một thông tin gì về tỷ lệ cần ước lượng?

3.17. Để điều tra số cá trong một hồ, người ta đánh bắt 1000 con cá đánh dấu rồi thả xuống hồ. Lần sau bắt lại 200 con thì được 40 con đánh dấu. Với độ tin cậy 95% hãy

a) Ước lượng tỷ lệ cá được đánh dấu trong hồ?

b) Ước lượng số cá trong hồ?

3.18. Ở một nhà máy dệt, kiểm tra ngẫu nhiên 100 tấm vải (dài $30m$) thấy số khuyết tật trung bình là 2,5 với phương sai $\sigma^2 = 4$.

a) Hãy ước lượng số khuyết tật trung bình của mỗi tấm vải với độ tin cậy 95%.

b) Nếu độ chính xác của ước lượng là 0,14 (khuyết tật) thì độ tin cậy của ước lượng là bao nhiêu?

c) Với độ chính xác của ước lượng bé hơn 0,14, độ tin cậy 95% thì phải kiểm tra thêm ít nhất bao nhiêu sản phẩm nữa?

Chương 4

KIỂM ĐỊNH GIẢ THIẾT

§1 GIẢ THIẾT VÀ ĐỐI THIẾT

Ở Chương 3 đã nghiên cứu ĐLNN, khi chưa biết tham số của nó và đã xây dựng các phương pháp ước lượng các tham số đó. Chương này tiếp tục nghiên cứu ĐLNN trong trường hợp thông tin không đầy đủ thể hiện ở nhiều mặt, cụ thể là:

Chưa biết chính xác các tham số θ hoặc qui luật phân phối xác suất của ĐLNN X, nhưng có cơ sở nào đó để nêu lên giả thiết, chẳng hạn $\theta = \theta_0$ (θ_0 là hằng số đã biết), hay: X tuân theo qui luật phân phối chuẩn.

Khi nghiên cứu hai hay nhiều ĐLNN, một trong những vấn đề cần quan tâm nhất là: các đại lượng này độc lập với nhau hay có sự phụ thuộc tương quan? Các tham số của chúng có bằng nhau hay không?

Những câu hỏi này thường chưa được trả lời khẳng định mà mới nêu lên như một giả thiết. Vậy có thể định nghĩa:

Giả thiết thống kê là những giả thiết nói về các tham số, dạng qui luật phân phối hoặc tính độc lập của các ĐLNN.

Việc tìm ra kết luận về tính thừa nhận được hay không thừa nhận được của một giả thiết gọi là kiểm định giả thiết thống kê.

Đây là một trong những bài toán cơ bản của thông kê toán. Trước hết ta đề cập đến các tham số ĐLNN.

Giả sử cần nghiên cứu tham số θ của ĐLNN X và có cơ sở nào đó để nêu giả thiết $\theta = \theta_0$.

Giả thiết này được ký hiệu $H_0 : \theta = \theta_0$ (được gọi là giả thiết cần kiểm định hay giả thiết cơ bản).

Mệnh đề đối lập với giả thiết H_0 được gọi là đối thiết của H_0 và ký hiệu là H_1. Dạng tổng quát của H_1 là: $\theta \neq \theta_0$.

Trong nhiều trường hợp, đối thiết có thể phát biểu cụ thể hơn như: $H_1 : \theta < \theta_0$ hay $H_1 : \theta > \theta_0$.

Như vậy giả thiết kiểm định và đối thiết thường được nêu lên thành từng cặp. Chẳng hạn:

$$H_0 : \theta = \theta_0;\ H_1 : \theta \neq \theta_0 \text{ (đối thiết hai phía)}$$
$$H_0 : \theta = \theta_0;\ H_1 : \theta > \theta_0 \text{ (đối thiết phải)}$$
$$H_0 : \theta = \theta_0;\ H_1 : \theta < \theta_0 \text{ (đối thiết trái).}$$

Nhiệm vụ của lý thuyết kiểm định giả thiết thống kê là: Bằng thực nghiệm (thông qua mẫu cụ thể) kiểm tra tính đúng (sai) của giả thiết H_0. Khi kết luận thì hoặc chấp nhận H_0 hoặc bác bỏ H_0 và trong trường hợp này, tuy không hoàn toàn tương đương, nhưng coi như chấp nhận H_1, ta có thể phạm hai loại sai lầm.

Sai lầm loại 1. Bác bỏ H_0 khi thực ra H_0 đúng.

Sai lầm loại 2. Chấp nhận H_0 khi thực ra H_0 sai.

Sai lầm loại 1 tương tự như sai lầm của quan tòa khi "kết án nhầm người vô tội", còn sai lầm loại 2 tương tự như khi "tha bỗng kẻ dó tội".

Một kiểm định thống kê lý tưởng là kiểm định làm cực tiểu cả hai loại sai lầm. Tuy nhiên chúng ta làm giảm sai lầm loại 1 sẽ tăng sai lầm loại 2 và ngược lại. Trong một xã hội văn minh, người ta có xu hướng thừa nhận rằng việc kết án nhầm người vô tội là một sai lầm nghiêm trọng hơn nhiều so với sai lầm tha bỗng kẻ có tội. Trong bài toán kiểm định giả thiết thống kê cũng vậy. Ta coi sai lầm loại 1 là nghiêm trọng hơn sai lầm loại 2. Thành thử người ta cố định trước xác suất mắc sai lầm loại 1. Xác suất của mắc sai lầm loại 1 còn gọi là *mức ý nghĩa*, được ký hiệu là α. Sau đó sẽ cực tiểu sai lầm loại 2. Kiểm định nào có xác suất sai lầm loại 2 nhỏ nhất được xem là tốt nhất. Các kiểm định thống kê trình bày trong chương này đều đã được chứng minh chặt chẽ về Toán học là các kiểm định tốt nhất.

Trong phạm vi tài liệu này chỉ đề cập đến đối thiết hai phía.

§2 KIỂM ĐỊNH GIÁ TRỊ TRUNG BÌNH μ CỦA BIẾN PHÂN PHỐI CHUẨN $N(\mu, \sigma^2)$

Bài toán kiểm định $H_0 : \mu = \mu_0$ với đối thiết $H_1 : \mu \neq \mu_0$ ở mức ý nghĩa α được chia thành 2 trường hợp sau:

2.1 Đã biết phương sai

Đây là trường hợp khi tiến hành điều tra lại một tổng thể, người ta lấy phương sai của lần điều tra trước làm σ^2, hoặc việc kiểm định được tiến hành thường xuyên tại một cơ sở công nghiệp mà qua một quá trình dài đã tìm được phương sai σ^2 (chủ yếu phụ thuộc vào độ chính xác của các thiết bị đo lường và tay nghề của nhân viên sử dụng thiết bị).

Ta tiến hành các bước sau:

- Lấy mẫu, tính $\overline{x}$

- Tính giá trị U thực nghiệm

$$U_{tn} = \frac{\overline{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{(\overline{x} - \mu_0)\sqrt{n}}{\sigma}.$$

- Tính giá trị tới hạn $u\left(\frac{\alpha}{2}\right)$ từ bảng phân phối chuẩn.

Nếu $|U_{tn}|$ (giá trị tuyệt đối của U_{tn}) bé hơn $u\left(\frac{\alpha}{2}\right)$ thì chấp nhận H_0 nếu ngược lại thì bác bỏ H_0 tức là chấp nhận H_1.

Ví dụ 2.1. Nuôi 100 con lợn theo một chế độ ăn riêng, sau 4 tháng tăng trọng trung bình là 30kg, giả thiết tăng trọng phân phối chuẩn $N(\mu, 25)$. Hãy kiểm định giả thiết $H_0 : \mu = 32$

đối thiết $H_1 : \mu \neq 32$ mức $\alpha = 0,05$.

$$U_{tn} = \frac{(30-32)\sqrt{100}}{5} = -4$$

$$|U_{tn}| = 4;\ u\left(\frac{\alpha}{2}\right) = 1,96.$$

Kết luận: Bác bỏ H_0 như vậy tăng trọng trung bình không phải là $32kg$.

Ví dụ 2.2. Khảo sát 64 gia đình tìm được chi tiêu trung bình của mỗi gia đình là 2,03 triệu đồng/tháng. Giả sử chi tiêu của một gia đình phân phối chuẩn $N(\mu, 0,09)$, hãy kiểm định giả thiết $H_0 : \mu = 2$ đối thiết $H_1 : \mu \neq 2$ ở mức $\alpha = 0,1$.

$$U_{tn} = \frac{(2,03-2)\sqrt{64}}{0,3} = 0,8$$

$$|U_{tn}| = 0,8;\ u\left(\frac{\alpha}{2}\right) = 1,645.$$

Kết luận: Chấp nhận H_0: mức chi tiêu trung bình của một gia đình là 2 triệu đồng 1 tháng.

2.2 Không biết phương sai, mẫu $n \geq 30$

Trong trường hợp này các bước kiểm định giống như mục 2.1 nhưng σ được thay bởi s.

Ví dụ 2.3. Một nhóm nghiên cứu công bố rằng trung bình một người vào siêu thị A tiêu hết 140 ngàn đồng. Chọn ngẫu nhiên 50 người mua hàng ta tính được số tiền trung bình họ tiêu là 154 ngàn với độ lệch tiêu chuẩn là 62 ngàn. Với mức ý nghĩa $0,02$ hãy kiểm định xem công bố của nhóm nghiên cứu có đúng hay không?

Giải. Ta cần kiểm định giả thiết $H_0 : \mu = 140$ với đối thiết $H_1 : \mu \neq 140$. Ta có

$$U_{tn} = \frac{(154-140)\sqrt{50}}{62} = 1,59.$$

$$|U_{tn}| = 1,59;\ u\left(\frac{\alpha}{2}\right) = 2,33.$$

Chưa có cơ sở để loại bỏ H_0, báo cáo của nhóm nghiên cứu là đúng.

2.3 Không biết phương sai, mẫu $n < 30$

Đây là trường hợp phổ biến khi kiểm định giá trị trung bình của phân phối chuẩn. Ta tiến hành các bước sau:

- Lấy mẫu, tính $\overline{x}$ và s^2.

- Tính giá trị T thực nghiệm

$$T_{tn} = \frac{(\overline{x}-\mu_0)\sqrt{n}}{s}.$$

- Tìm giá trị tới hạn $t\left(\frac{\alpha}{2}, n-1\right)$ là trong bảng phân phối Student.

Nếu $|T_{tn}|$ (giá trị tuyệt đối của T_{tn}) bé hơn $t\left(\frac{\alpha}{2}, n-1\right)$ thì chấp nhận H_0 nếu ngược lại thì bác bỏ H_0 tức là chấp nhận H_1.

Ví dụ 2.4. Trong điều kiện chăn nuôi bình thường lượng sữa trung bình của một con bò sữa là 19 kg/ngày. Trong một đợt hạn, người ta theo dõi 25 con bò và được lượng sữa trung bình 17,5 kg/ngày, độ lệch chuẩn $s = 2,5kg$. Giả thiết lượng sữa phân phối chuẩn, hãy kiểm định giả thiết $H_0 : \mu = 19$ với đối thiết $\mu \neq 19$ ở mức $\alpha = 0,05$.

$$T_{tn} = \frac{(17,5 - 19)\sqrt{25}}{2,5} = -3;\ |T_{tn}| = 3;\ t\left(\frac{\alpha}{2}, n-1\right) = 2,064$$

Kết luận: Bác bỏ H_0 như vậy trọng lượng sữa trung bình không còn là 19 kg/ngày nữa.

Ví dụ 2.5. Thóc được đóng trong bao $50kg$. Sau một thời gian, để kiểm tra, người ta cân thử 25 bao và được trọng lượng trung bình 49,4kg/bao, độ lệch chuẩn $s = 3,6kg$. Giả thiết trọng lượng bao thóc phân phối chuẩn, hãy kiểm định giả thiết $H_0 : \mu = 50$ với đối thiết $\mu \neq 50$ ở mức $\alpha = 0,05$.

$$T_{tn} = \frac{(49,4 - 50)\sqrt{25}}{3,6} = -0,833;\ |T_{tn}| = 0,833;\ t\left(\frac{\alpha}{2}, n-1\right) = 2,064$$

Kết luận: Chấp nhận H_0 như vậy vẫn coi trọng lượng trung bình của một bao thóc là $50kg$.

§3 KIỂM ĐỊNH XÁC SUẤT

Tổng thể có hai loại cá thể A và $\overline{A}$, loại A chiếm tỉ lệ p (chưa biết). Ta muốn kiểm định giả thiết $H_0 : p = p_0$ với đối thiết $H_1 : p \neq p_0$.

Từ mẫu dung lượng n, tính số cá thể loại A được tần số m và tần suất $f = \frac{m}{n}$. Tính

$$U_{tn} = \frac{f - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}.$$

Tìm giá trị tới hạn $u\left(\frac{\alpha}{2}\right)$

Kết luận: Nếu $|U_{tn}| \leq u\left(\frac{\alpha}{2}\right)$ thì chấp nhận H_0 ngược lại thì bác bỏ H_0.

Ví dụ 3.1. Gieo thử 100 hạt, có 82 hạt nảy mầm.

Kiểm định giả thiết tỉ lệ nảy mầm $p = 0,80$, đối thiết $p \neq 0,8$ với $\alpha = 0,05$.

Giải. $n = 100$; $m = 82$; $f = \frac{m}{n} = 0,82$;

$$U_{tn} = \frac{0,82 - 0,80}{\sqrt{\frac{0,8.0,2}{100}}} = 0,5;\ u\left(\frac{\alpha}{2}\right) = 1,96.$$

Kết luận: Chấp nhận H_0: tỷ lệ nảy mầm là 0,80.

Ví dụ 3.2. Một đảng chính trị trong một cuộc bầu cử tổng thống ở Mỹ tuyên bố rằng 45% cử tri sẽ bỏ phiếu cho ứng cử viên A của họ. Chọn ngẫu nhiên 200 cử tri để thăm dò ý kiến cho thấy 80 người trong số đó tuyên bố sẽ bỏ phiếu cho ông A. Với mức $\alpha = 5\%$, hãy kiểm định xem dự đoán của đảng trên có đúng không.

Giải. Giả thiết $H_0 : p = 0,45$, đối thiết $H_1 : p \neq 0,45$. Ta có $f = \frac{80}{200} = 0,40$, $u\left(\frac{\alpha}{2}\right) = 1,96$,

$$U_{tn} = \frac{0,40 - 0,45}{\sqrt{\frac{0,45.0,55}{200}}} = -1,43.$$

Kết luận: Dự đoán của đảng trên đúng.

Bài Tập

4.1. Điều tra thấy chi phí trung bình $\overline{x}$ của 25 sinh viên xa nhà là 475000đ/ tháng, độ lệch chuẩn 30000đ. Coi chi phí một tháng phân phối chuẩn, Hãy kiểm định giả thiết $H_0 : \mu =$ 500000đ. Đối thiết $H_1 : \mu \neq$ 500000đ, $\alpha = 0,05$.

4.2. Theo hợp đồng khi bán các bao gạo được đóng trong bao $50kg$. Kiểm tra 16 bao được $\overline{x} = 49kg$, $s = 3,6kg$. Hỏi hợp đồng có được bên bán thực hiện nghiêm chỉnh hay không, $\alpha = 0,05$?

4.3. Giám đốc của một xí nghiệp cho biết lương trung bình của một công nhân thuộc xí nghiệp hiện nay là 600 ngàn đồng/1tháng. Chọn ngẫu nhiên 36 công nhân thấy lương trung bình là 529 ngàn đồng/1tháng. Với độ lệch tiêu chuẩn $\sigma = 40$ ngàn đồng/1tháng và mức ý nghĩa $\alpha = 5\%$. Kiểm tra xem lời báo cáo của giám đốc có tin cậy không?

4.4. Một nhóm nghiên cứu tuyên bố rằng trung bình một ngày một người vào siêu thị X tiêu hết 140 ngàn đồng. Chọn ngẫu nhiên 25 người và tính được số tiền trung bình tiêu là 154 ngàn đồng với độ lệch chuẩn mẫu $s = 62$. Với mức ý nghĩa $\alpha = 0,02$ hãy kiểm định xem nhóm nghiên cứu có đúng không?

4.5. Một phương pháp ăn kiêng được quảng cáo rằng sẽ giảm trọng lượng ít nhất là $20kg$ trong vòng 6 tháng. Một mẫu gồm 28 người theo chế độ ăn kiêng này giảm trọng lượng trung bình là $16kg$ với độ lệch tiêu chuẩn $s = 9kg$. Với mức ý nghĩa $\alpha = 0,01$, hãy nhận định xem phương pháp ăn kiêng trên có nói đúng không.

4.6. Theo hợp đồng, khi bán mỗi bao gạo được đóng $50kg$. Kiểm tra 16 bao thu được kết quả sau

Trọng lượng	tần số	Trọng lượng	tần số
47	1	51	3
48	4	53	2
49	3	54	3

Với mức ý nghĩa $\alpha = 0,05$ hỏi hợp đồng có được bên bán thực hiện nghiêm chỉnh hay không?

4.7. Một công ty A sản xuất bánh kẹo tuyên bố rằng $2/3$ số trẻ em thích ăn bánh của công ty. Trong một mẫu gồm 100 trẻ em được hỏi, có 55 em tỏ ra thích ăn bánh của công ty A. Với mức ý nghĩa $\alpha = 0,05$ số liệu nói trên có chứng tỏ là tuyên bố của công ty A có đúng không?

4.8. Một trại chăn nuôi A xuất chuồng 200 con lợn, trọng lượng của chúng được tổng hợp như sau

Trọng lượng (kg)	45-55	55-65	65-75	75-85	85-95	95-100
Số lợn	10	30	45	80	30	5

a) Ước lượng trọng lượng trung bình của trại chăn nuôi A với độ tin cậy 95%.

b) Nếu chủ trại chăn nuôi A thông báo trọng lượng trung bình của trại là $80kg$ thì có chấp nhận được không với mức ý nghĩa 10%?

c) Lợn có trọng lượng $85kg$ trở lên được coi là lợn chóng lớn, hãy ước lượng tỷ lệ lợn chóng lớn của trại chăn nuôi trên với mức ý nghĩa 90%.

PHỤ LỤC

A) Một số bài tập làm thêm:

4 Xác suất

4.9. Trong 30 đề thi, trong đó có 10 đề khó, 20 đề trung bình. Tìm xác suất để:
a. Một học sinh bốc 1 đề, gặp đề trung bình.
b. Một học sinh bốc 2 đề, gặp ít nhất một đề trung bình.

4.10. Một lớp có 60 sinh viên, trong đó có 20 nam và 40 nữ. Chọn ngẫu nhiên một nhóm gồm 8 sinh viên. Tính xác suất để:
a. Có 4 nam trong số 8 sinh viên được chọn?
b. Có nhiều nhất 3 sinh viên nam trong 8 sinh viên được chọn?
c. Có ít nhất 1 sinh viên nam trong 8 sinh viên được chọn?
d. Không có sinh viên nam trong 8 sinh viên được chọn?.

4.11. Một chiếc hộp đựng 6 quả cầu trắng, 4 quả cầu đỏ và 2 quả cầu đen. Chọn ngẫu nhiên 6 quả cầu. Tính xác suất để thu được 3 quả trắng, 2 quả đỏ và 1 quả đen?

4.12. Một lớp có 25 sinh viên trong đó có 5 giỏi, 10 khá và 10 trung bình. Chọn ngẫu nhiên 3 người, tìm xác suất để:

a) Cả 3 đều là SV yếu?

b) Có ít nhất 1 SV giỏi?

4.13. Một đoàn tàu có 4 toa. Có 4 hành khách từ sân ga lên tàu, mỗi người độc lập với nhau chọn ngẫu nhiên 1 toa. Tính xác suất để

a) 1 toa có 3 người, 1 toa có 1 người và hai toa còn lại không có ai?

b) Toa nào cũng có khách?

c) Cũng hỏi như câu b) nhưng số toa tàu là 3?

4.14. Một công ty có 60 nhân viên, trong đó có 20 nam và 40 nữ. Tỷ lệ nhân viên nữ có thể nói tiếng Anh lưu loát là 15% và tỷ lệ này đối với nam là 20%
a. Gặp ngẫu nhiên một nhân viên của công ty. Tìm xác suất để gặp được nhân viên nói tiếng Anh lưu loát?
b.) Biết người gặp được nói tiếng Anh lưu loát, tính xs để người đó là nữ?
c. Gặp ngẫu nhiên hai nhân viên của công ty. Tìm xác suất để có ít nhất một người nói tiếng Anh lưu loát trong số 2 người gặp?

4.15. Một công ty cần tuyển hai nhân viên. Có 6 người nạp đơn trong đó có 4 nữ và 2 nam. Khả năng được tuyển của mỗi người là như nhau.

a) Tính xác suất để cả hai nữ được chọn biết rằng ít nhất một nữ đã được chọn?

b) Giả sử Hoa là một trong 4 nữ. Tính xác suất để Hoa được chọn?

c) Tính xác suất để Hoa được chọn nếu biết rằng ít nhất một nữ đã được chọn?

4.16. Xét một lô sản phẩm trong đó số sản phẩm do nhà máy I sản xuất chiếm 20%, nhà máy II sản xuất chiếm 30%, nhà máy III sản xuất chiếm 50%. Xác suất phế phẩm của các nhà máy I, II, III tương ứng là 0,001; 0,005; 0,006. Tìm xác suất để lấy ngẫu nhiên được đúng 1 phế phẩm?

4.17. Có 4 hộp như nhau đựng cùng một chi tiết máy. Trong đó hộp 1 có 3 chi tiết xấu, 5 chi tiết tốt do máy I sản xuất. Ba hộp còn lại mỗi hộp đựng 4 chi tiết xấu, 6 chi tiết tốt do máy II sản xuất. Lấy ngẫu nhiên 1 hộp rồi từ đó lấy ra 1 chi tiết máy

a) Tính xác suất để chi tiết máy lấy ra là tốt?

b) Biết chi tiết lấy ra là tốt. Tính xác suất để chi tiết đó của lô I?

4.18. Một hộp có 4 sản phẩm tốt và 2 sản phẩm xấu. Lấy ngẫu nhiên lần lượt từ hộp ra 2 sản phẩm. Biết sản phẩm lấy ra lần 2 là sản phẩm tốt. Tính xác suất để sản phẩm lấy lần thứ 1 cũng tốt?

4.19. Có 3 chuồng thỏ. Chuồng thứ nhất có 3 thỏ trắng và 2 thỏ đen, chuồng thứ hai có 4 thỏ trắng và 3 thỏ đen, chuồng thứ ba có 3 thỏ trắng và 3 thỏ đen. Chọn ngẫu nhiên một chuồng và từ chuồng đó bắt ngẫu nhiên một con thỏ.

a) Tìm xác suất để bắt được thỏ trắng?

b) Biết thỏ bắt ra là thỏ trắng. Tìm xác suất để thỏ đó thuộc chuồng thứ nhất?

4.20. Có 3 chuồng thỏ: chuồng thứ nhất có 3 con thỏ trắng và 2 con thỏ nâu, chuồng thứ hai có 4 con thỏ trắng và 3 con thỏ nâu, chuồng thứ ba có 3 con thỏ trắng và 3 con thỏ nâu. Chọn ngẫu nhiên một chuồng và từ đó bắt ngẫu nhiên một con thỏ.
a. Tìm xác suất để bắt được con thỏ trắng?
b. Biết con thỏ bắt được là thỏ trắng. Tìm xác suất để con thỏ đó thuộc chuồng thứ nhất.

4.21. Có 20 xạ thủ trong đó có 7 người bắn trúng đích với xác suất 0,8 (giỏi), 5 người bắn trúng với xác suất 0,7 (khá) số còn lại bắn trúng với xác suất 0,5 (trung bình). Chọn ngẫu nhiên một người vào bắn.

a) Tìm xác suất để người đó bắn trượt?

b) Biết rằng người đó bắn trượt, có nhiều khả năng người đó xếp loại nào?

4.22. Có hai hộp đựng cam. Hộp I đựng 10 quả tốt và 3 quả hỏng, hộp II đựng 7 quả tốt và 2 quả hỏng. Lấy ngẫu nhiên **một** quả từ hộp I bỏ sang hộp II sau đó từ hộp II lấy ngẫu nhiên ra hai quả.

a) Tính xác suất để cả hai quả đều hỏng?

b) Tính xác suất để cả hai quả đều tốt?

c) Tính xác suất để có một quả tốt và một quả hỏng?

4.23. Có hai hộp đựng cam. Hộp I đựng 10 quả tốt và 3 quả hỏng, hộp II đựng 7 quả tốt và 2 quả hỏng. Lấy ngẫu nhiên **hai** quả từ hộp I bỏ sang hộp II sau đó từ hộp II lấy ngẫu nhiên ra hai quả.

a) Tính xác suất để cả hai quả đều hỏng?

b) Tính xác suất để cả hai quả đều tốt?

c) Tính xác suất để có một quả tốt và một quả hỏng?

4.24. Một chiếc máy bay có thể xuất hiện ở vị trí A với xác suất 2/3 và ở vị trí B với xác suất 1/3. Có 3 phương án bố trí 4 khẩu pháo bắn máy bay như sau:

Phương án 1: 3 khẩu tại A, 1 khẩu tại B.

Phương án 2: 2 khẩu tại A, 2 khẩu tại B.

Phương án 3: 1 khẩu tại A, 3 khẩu tại B.

Biết rằng xác suất bắn trúng máy bay của mỗi khẩu pháo là 0,7 và các khẩu pháo hoạt động độc lập với nhau. Hãy chọn phương án tốt nhất

4.25. Có hai kiện hàng:
Kiện thứ nhất: có 5 sản phẩm loại A, 1 sản phẩm loại B.
Kiện thứ hai : có 2 sản phẩm loại A, 4 sản phẩm loại B.
Từ mỗi kiện hàng chọn ngẫu nhiên ra 1 sản phẩm đem giao cho khách hàng. Sau đó các sản phẩm còn lại được dồn vào kiện thứ ba (trống).

a. Chọn ngẫu nhiên một sản phẩm từ kiện hàng thứ ba. Tính xác suất để lấy được là sản phẩm loại B?. b. Nếu ta chọn ngẫu nhiên 2 sản phẩm từ kiện ba. Tính xác suất để có ít nhất một sản phẩm loại B từ 2 sản phẩm đã chọn.

4.26. Có hai tổ. Tổ I có 6 nam và 4 nữ, tổ 2 có 3 nam và 7 nữ. Cần lập một nhóm 3 người từ tổ này. Để khách quan từ mỗi tổ ta chọn ngẫu nhiên ra 1 người, sau đó trong số còn lại của hai tổ ta chọn ngẫu nhiên ra một người.

a) Tính xác suất để người thứ 3 được chọn là nam?

b) Tính xác suất để người thứ 3 được chọn thuộc tổ II?

4.27. Một hộp đựng 15 quả bóng bàn trong đó có 9 quả còn mới. Lần đầu người ta lấy ngẫu nhiên ba quả để thi đấu, sau đó lại trả vào hộp. Lần 2 lấy ngẫu nhiên 3 quả. Tính xác suất để ba quả lần sau đều mới.
HD. Gọi $B_i, i = \overline{0,3}$ là biến cố trong 3 quả lấy ra có i quả mới

4.28. Có hai hộp, hộp 1 đựng 8 bi trắng và 2 bi đen; hộp 2 đựng 9 bi trắng và 1 bi đen, các viên bi cùng kích thước. Lấy ngẫu nhiên hai bi từ hộp 1 bỏ sang hộp 2. Sau đó lấy ngẫu nhiên 3 bi từ hộp 2. Tính xác suất để trong 3 bi lấy ra sau có hai bi trắng.
HD. $B_i, i = \overline{0,2}$ trong hai bi bỏ sang hộp hai có i bi trắng

5 Biến ngẫu nhiên

4.29. Bán ba viên đạn độc lập vào một mục tiêu. Xác suất trúng của mỗi viên tương ứng là 0,6; 0,4; 0,5.
Gọi X là số viên đạn không trúng mục tiêu. Tìm phân phối xác suất của X.

Trong các bài tập sau, mỗi câu đều được hỏi bốn ý

a) Lập bảng phân phối xác suất của Biến ngẫu nhiên được gọi?

b) Xác định hàm phân phối?

c) Tính kỳ vọng ($M(X), M(X+3), M(X-6)$...? (Có thể thay bằng câu hỏi trung bình của biến ngẫu nhiên đó)

d) Tính phương sai?

4.30. Một chiếc hộp gồm 3 bộ phận hoạt động độc lập nhau, xác suất trong khoảng thời gian t các bộ phận bị hỏng tương ứng bằng 0,2; 0,3; 0,5. Gọi X là số bộ phận bị hỏng.

4.31. Một hộp có 7 viên bi trong đó có 4 bi trắng và 3 bi đỏ. Lấy ngẫu nhiên 3 viên. Gọi X là số bi trắng lấy ra.

4.32. Bắn 5 viên đạn độc lập với nhau vào một mục tiêu (trong cùng một điều kiện như nhau). Xác suất trúng đích của mỗi lần bắn bằng 0,2. Gọi X là số đạn trúng mục tiêu.
Hỏi thêm: Tính xác suất để có đúng 3 viên trúng đích.

4.33. Một xạ thủ đem 5 viên đạn đi bắn, xạ thủ bắn từng viên vào bia với xác suất trúng vòng 10 là 0,85. Nếu bắn được ba viên liên tiếp trúng vòng 10 thì thôi không bắn nữa. Gọi X là số viên đạn đã bắn.

4.34. Một xạ thủ dùng 5 viên đạn để thử súng. Anh ta bắn từng viên vào tâm với xác suất trúng tâm là 0,95. Nếu có 2 viên liên tiếp trúng tâm thì thôi không bắn nữa. Gọi Y là số viên đạn còn thừa.

4.35. Một cơ quan mua về 15 chiếc máy trong đó có 4 chiếc bị khuyết tật. Phòng A nhận về 6 chiếc một cách ngẫu nhiên. Gọi X là số chiếc máy khuyết tật mà phòng A nhận được.

4.36. Một hộp có 3 bi trắng và 4 bi đỏ. Lấy ngẫu nhiên từ hộp từng viên cho đến khi lấy được bi trắng. Gọi X là số bi trắng lấy được.

4.37. Có hai lô sản phẩm. Lô 1: Có 8 chính phẩm và 2 phế phẩm
Lô 2: Có 7 chính phẩm và 3 phế phẩm
Từ lô thứ nhất lấy ngẫu nhiên 2 sản phẩm bỏ sang lô thứ hai, sau đó từ lô thứ hai lấy ra 2 sản phẩm Gọi X là số chính phẩm được lấy ra.

4.38. Có 3 kiện hàng, mỗi kiện chứa 10 sản phẩm. Số sản phẩm loại B trong mỗi kiện tưng ứng là 1,2,3.
a. Lấy ngẫu nhiên từ mỗi kiện ra một sản phẩm. Gọi X là số sản phẩm loại A có trong 3 sản phẩm lấy ra?.
b. Chọn ngẫu nhiên một kiện rồi từ kiện đã chọn lấy ngẫu nhiên 3 sản phẩm (lấy đồng thời). Gọi X là số sản phẩm loại B có trong 3 sản phẩm lấy ra?

6 Bài toán ước lượng, kiểm định

4.39. Đo áp lực $X(kg/cm^2)$ của 18 thùng chứa được kết quả:

X_i	19,6	19,5	19,9	20,0	19,8	20,5	21,0	18,5	19,7
n_i	1	2	2	4	2	3	2	1	1

Biết rằng áp lực là đại lượng ngẫu nhiên tuân theo phân phối chuẩn. Với độ tin cậy $p = 0,99$, hãy tìm khoảng ước lượng của áp lực trung bình μ?

4.40. Để xác định trọng lượng trung bình của 15 bao bột mì được đóng bằng máy tự động, người ta chọn ngẫu nhiên 15 bao và tính được $\overline{x} = 39,8kg$ và $s^2 = 0,144$. Tìm khoảng tin cậy 99% của trọng lượng trung bình μ của các bao bột mì?

4.41. Để khảo sát chiều cao trung bình μ của thanh niên trong một vùng A nào đó, một mẫu gồm 16 thanh niên được chọn, chiều cao của các thanh niên này đo được như sau:

Chiều cao (cm)	tần số	Chiều cao (cm)	tần số
163	1	170	3
164	4	172	2
166	3	174	3

Tìm khoảng tin cậy 95% cho μ.

4.42. Để khảo sát chiều cao trung bình μ của thanh niên trong một vùng A nào đó, một mẫu gồm 36 thanh niên được chọn, chiều cao của các thanh niên này đo được như sau:

Chiều cao (cm)	tần số	Chiều cao (cm)	tần số
163	7	170	4
164	2	172	6
166	3	174	5
168	5	175	4

Tìm khoảng tin cậy 95% cho μ.

4.43. Trong một cuộc khảo sát thị trường của công ty sản xuất thuốc lá, 150 người nghiện thuốc lá được chọn ngẫu nhiên, số điếu thuốc hút trung bình trong 1 tuần của nhóm người này là với độ lệch tiêu chuẩn là 36. Tìm khoảng tin cậy 99% cho số điếu thuốc hút trung bình trong 1 tuần của tất cả những người nghiện thuốc lá?

4.44. Trong một cuộc thăm dò ý kiến 100 khánh hàng, người ta thấy 55 người nói yêu thích mặt hàng A. Tìm khoảng tin cậy 95% cho tỷ lệ người tiêu dùng ưa thích mặt hàng A?

4.45. Một cơ quan Cảnh sát Giao thông kiểm tra hệ thống phanh của 40 chiếc xe tải trên đường Quốc lộ, họ phát hiện 14 chiếc có phanh chưa đảm bảo an toàn.

a) Hãy tìm khoảng tin cậy 98% cho tỷ lệ xe tải có phanh chưa an toàn?

b) Hãy tìm khoảng tin cậy 98% cho tỷ lệ xe tải có phanh tốt?

4.46. Trước ngày bầu cử, để biết tỷ lệ phần trăm các cử tri còn do dự chưa biết bỏ phiếu cho ai, người ta hỏi n cử tri được chọn một cách ngẫu nhiên. Hỏi số n tối thiểu là bao nhiêu để khoảng tin cậy 95% có độ dài không vượt quá 0,04?

4.47. Ta muốn xây dựng một khoảng tin cậy 95% cho trọng lượng trung bình của các gói đường đóng bằng máy tự động. Một mẫu điều tra sơ bộ cho ta $\overline{x} = 11,8kg$ với độ lệch tiêu chuẩn $s = 0,7kg$. Hỏi cần phải lấy kích thước mẫu tối thiểu bao nhiêu để đạt được sai số không vượt quá $0,2kg$.

4.48. Người ta muốn tìm khoảng tin cậy 95% cho tỷ lệ những gia đình có máy giặt với độ chính xác 0,04. Hỏi kích thước mẫu cần lấy là bao nhiêu? Giả sử rằng mẫu điều tra sơ bộ cho ta tần suất $f = 0,72$.

4.49. Một khách sạn lớn tiến hành một nghiên cứu để xác định tỷ lệ phần trăm các khách ở trọ với thời gian nhiều hơn 1 ngày. Người chủ khách sạn muốn đạt được độ tin cậy 95% và sai số không vượt quá 0,05. Anh ta ước lượng so bộ tỷ lệ này khoảng 30%. Hỏi cần lấy mẫu với kích thước bao nhiêu?

4.50. Tương tự như bài 4.49 nhưng ở đây người chủ khách sạn không có trước một thông tin gì về tỷ lệ cần ước lượng?

4.51. Để điều tra số cá trong một hồ, người ta đánh bắt 1000 con cá đánh dấu rồi thả xuống hồ. Lần sau bắt lại 200 con thì được 40 con đánh dấu. Với độ tin cậy 95% hãy

1. Ước lượng tỷ lệ cá được đánh dấu trong hồ?

2. Ước lượng số cá trong hồ?

4.52. Trọng lượng bao gạo là biến ngẫu nhiên có phân phối chuẩn với trọng lượng trung bình là 50 kg. Sau một khong thời gian người ta nghi ngờ trọng lượng của bao gạo có thay đổi. Cân 25 bao gạo người ta thu được kết quả sau:

Khối lượng (kg)	48 - 48,5	48,5 - 49	49 - 49,5	49,5 - 50	50 - 50,5
số bao gạo	2	5	10	6	2

4.53. Một công ty sản xuất hạt giống tuyên bố rằng một loại giống mới của họ có năng suất trung bình là 21,5 tạ/ha. Gieo thử hạt giống mới này tại 16 vườn thí nghiệm và thu được kết quả

19,2; 18,7; 22,4; 20,3; 16,8; 25,1; 17,0; 15,8; 21,0; 18,6; 23,7; 24,1; 23,4; 19,8; 21,7; 18,9.

Dựa vào kết quả này hãy xác nhận xem quảng cáo của công ty có đúng không với mức ý nghĩa $\alpha = 0,05$. Biết rằng năng suất giống cây trồng là một biến ngẫu nhiên có phân phối chuẩn $N(\mu; \sigma^2)$.

4.54. Ở một nhà máy dệt, kiểm tra ngẫu nhiên 100 tấm vải (dài $30m$) thấy số khuyết tật trung bình là 2,5 với phương sai $\sigma^2 = 4$.

a) Hãy ước lượng số khuyết tật trung bình của mỗi tấm vải với độ tin cậy 95%.

b) Nếu độ chính xác của ước lượng là 0,14 (khuyết tật) thì độ tin cậy của ước lượng là bao nhiêu?

c) Với độ chính xác của ước lượng bé hơn 0,14, độ tin cậy 95% thì phải kiểm tra thêm ít nhất bao nhiêu sản phẩm nữa?

4.55. Công ty xe buýt nói rằng cứ trung bình 5 phút lại có một chuyến xư. Chọn ngẫu nhiên 8 thời điểm và ghi lại thời gian giữa hai chuyến xe buýt ta thu được số liệu sau:

5,3; 4,5; 4,8; 5,1; 4,3; 4,8; 4,9 4,7.

Với mức ý nghĩa 5%, nhận định xem công ty xe buýt nói có đúng không.

4.56. Một trại chăn nuôi A xuất chuồng 200 con lợn, trọng lượng hơi tính bằng (Kg), ta có bảng số liệu sau:

Trọng lượng (x_i)	45-55	55-65	65-75	75-85	85-95	95-105
Số lợn	10	30	45	80	30	5

a) Ước lượng trọng lượng lợn (hơi) trung bình của trại chăn nuôi A với độ tin cậy 95%.
b) Nếu chủ trại chăn nuôi A thông báo trọng lượng (hơi) trung bình của trại chăn nuôi là trên 80 kg thì có chấp nhận được không với mức ý nghĩa 10%
c) Lợn có trọng lượng từ 85 kg trở lên được gọi là lợn chóng lớn, hãy ước lượng tỷ lệ lợn chóng lớn của trại chăn nuôi trên với mức ý nghĩa 99%.

4.57. Tại một khu vườn trồng xoài, để điều tra trọng lượng của các trái xoài, một người đã cân thử một 100 trái xoài và kết quả được cho ở bảng sau:

Trọng lượng (g)	Số trái	Trọng lượng (g)	Số trái
450 - 500	2	600 - 650	10
500 - 550	30	650 - 700	25
550 - 600	20	700 - 750	13

a) Nếu người đó cho biết trọng lượng trung bình của các trái xoài là 610g với độ tin cậy 95% thì có thể chấp nhận được không?
b) Những trái xoài có trọng lượng từ 650g trở lên được xem là loại I. Người đó cho biết tỉ lệ loại I là 25% với độ tin cậy 99% thì có đúng hay không?
c) Những trái xoài không phải là loại I thì là loại II. Với độ tin cậy 95% có thể khẳng định trọng lượng trung bình của các trái xoài loại II là 580g được không.

B) Một số đề kiểm tra tham khảo

Đề 1:

Câu 1. Đề cương ôn tập môn Toán có 20 câu. Lớp A có 40 sinh viên được chia thành ba nhóm gồm: nhóm chuẩn bị tốt có 15 sinh viên, nhóm chuẩn bị khá có 20 sinh viên và nhóm chuẩn bị trung bình có 5 sinh viên. Sinh viên chuẩn bị tốt học 18 câu, sinh viên chuẩn bị khá học 15 câu, sinh viên chuẩn bị trung bình học 10 câu. Gọi ngẫu nhiên một sinh viên của lớp và hỏi một câu trong đề cương.
a) Tìm xác suất để sinh viên được gọi trả lời được câu hỏi.
b) Biết rằng sinh viên đó trả lời được câu hỏi, tìm xác suất để sinh viên đó thuộc nhóm chuẩn bị khá.

Câu 2. Có hai xạ thủ A và B mỗi người cầm 3 viên đạn đi thử súng. Họ bắn liên tiếp vào một bia, nếu còn đạn và bia chưa bị bắn thì họ vẫn còn bắn. Xác suất bắn trúng bia của hai người A và B lần lượt là 0,8 và 0,9. Biết rằng người A bắn trước. Gọi X là số đạn người A đã bắn.

a) Lập bảng phân phối xác suất của biến ngẫu nhiên X.
b) Xác định hàm phân phối của biến ngẫu nhiên X.
c) Tính kỳ vọng và phương sai của biến ngẫu nhiên X.

Câu 3. Để khảo sát chiều cao sinh viên của trường Đại học A một mẫu gồm 36 sinh viên được chọn, chiều cao của các sinh viên này đo được như sau

chiều cao (cm)	tần số	chiều cao (cm)	tần số
163	6	170	4
165	3	172	6
166	2	174	5
168	6	175	4

Biết rằng chiều cao sinh viên có phân phối chuẩn $N(\mu, 16)$.
a) Ở mức tin cậy $P = 95\%$ hãy ước lượng kỳ vọng μ.
b) Để đạt được độ chính xác $\varepsilon = 1$ ở mức tin cậy 99% thì phải điều tra ít nhất bao nhiêu sinh viên?
c) Kiểm định giả thiết $H_0 : \mu = 180cm$; đối thiết $H_1 : \mu \neq 180cm$ ở mức tin cậy 99%.
Cho biết: với $\Phi(t) = \int\limits_{-\infty}^{t} e^{-x^2/2}dx$ thì $\Phi(1,96) = 0,975$; $\Phi(2,58) = 0,995$.

Ghi chú: Sinh viên **không** được sử dụng tài liệu khi làm bài.

Đề 2

Câu 1. Có ba hộp đựng cam. Hộp I đựng 10 quả tốt và 4 quả hỏng, hộp II đựng 15 quả tốt và 3 quả hỏng, hộp III đựng 12 quả tốt và 3 quả hỏng. Chọn ngẫu nhiên một hộp rồi từ hộp đó lấy ra 2 quả cam.

a) Tìm xác suất để hai quả lấy ra có 1 quả tốt và 1 quả hỏng.

b) Biết rằng trong 2 quả lấy ra có đúng 1 quả hỏng, tìm xác suất để hai quả này thuộc hộp I.

Câu 2. Một xạ thủ cầm 5 viên đạn đi bắn, xác suất bắn trúng vòng mười của anh ta là 0,8. Nếu bắn được ba viên liên tiếp trúng vòng mười hoặc hết đạn thì thôi không bắn nữa. Gọi X là số đạn còn thừa.

a) Lập bảng phân phối xác suất của biến ngẫu nhiên X.

b) Xác định hàm phân phối của biến ngẫu nhiên X.

c) Tính kỳ vọng và phương sai của biến ngẫu nhiên X.

Câu 3. Để khảo sát chiều cao trung bình μ của thanh niên trong một vùng A nào đó, một mẫu gồm 16 thanh niên được chọn, chiều cao của các thanh niên này đo được như sau:

Chiều cao (cm)	tần số	Chiều cao (cm)	tần số
163	1	170	3
164	4	172	2
166	3	174	3

Biết rằng chiều cao của thanh niên tuân theo phân phối chuẩn.

a) Hãy ước lượng khoảng cho kỳ vọng μ với độ tin cậy 95%.

b) Một kết luận nói rằng chiều cao thanh niên của vùng A là $170cm$. Ta nghi ngờ kết luận trên và muốn kiểm định nó.

Hãy phát biểu giả thiết và đối thiết của bài toán kiểm định. Với mức tin cậy $P = 99\%$ có chấp nhận kết luận chiều cao trung bình thanh niên của vùng A là $170cm$ hay không?

Cho biết, kí hiệu $t(\frac{\alpha}{2}, n-1)$ là giá trị tới hạn trong phân phối Student mức $\frac{\alpha}{2}$; $n-1$ bậc tự do; $\Phi(t) = \int\limits_{-\infty}^{t} e^{-x^2/2}dx$; $t(0,025;15) = 2,131$; $t(0,005;15) = 2,947$. $\Phi(1,96) = 0,975$; $\Phi(2,58) = 0,995$.

Ghi chú: Sinh viên **không** được sử dụng tài liệu khi làm bài.

Tài liệu tham khảo

[1] Nguyễn Đình Hiền, *Giáo trình xác suất thống kê*, NXB Đại học Sư phạm, 2004.

[2] Phạm Văn Kiều, *Xác suất thống kê*, NXB Đại học Sư phạm, 2005.

[3] Đặng Hùng Thắng, *Thống kê và ứng dụng*, NXBGD, 1999.

[4] Đặng Hùng Thắng, *Bài tập thống kê*, NXBGD, 1999.