DƯƠNG NGỌC HẢO

XÁC SUẤT THỐNG KÊ

MỤC LỤC

Chương 1
GIẢI TÍCH TỔ HỢP

Chương này chủ yếu ôn lại kiến thức về giải tích tổ hợp mà sinh viên đã được học ở lớp 12. Sinh viên phải vận dụng được những kiến thức này trong việc xác định số chỉnh hợp, tổ hợp trong một hoàn cảnh cụ thể nào đó khi được yêu cầu.

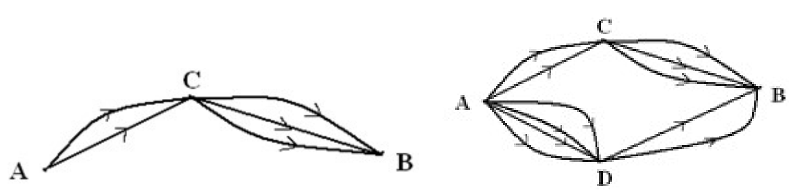
1.1. Qui tắc đếm

Giả sử một công việc hoàn thành theo k giai đoạn. Ở giai đoạn thứ i có n_i cách hoàn thành, $i = \overline{1,k}$. Khi đó, số cách hoàn thành công việc là:

$$\begin{aligned} n &= \prod_{i=1}^{k} n_i \\ &= n_1.n_2 \ldots n_k. \end{aligned} \tag{1.1}$$

Nếu có sự phân chia trường hợp và các trường hợp này *rời nhau*, nghĩa là chúng không có kết quả chung, thì ta sẽ *cộng* số cách hoàn thành ở mỗi trường hợp để được số cách hoàn thành công việc.

Ví dụ 1.1: Có mấy cách đi từ A đến B theo sơ đồ trong hình 1.1.



Sơ đồ a ***Sơ đồ b***

Hình 1.1.

Ví dụ 1.2: Có bao nhiêu cách xếp 3 người lên 4 toa tàu?

Ví dụ 1.3: Có bao nhiêu tập con của tập hợp $A = \{a, b, c\}$?

1.2. Chỉnh hợp

a. Định nghĩa: Một chỉnh hợp chập k của n phần tử (khác nhau) là một nhóm gồm k phần tử khác nhau được lấy từ n phần tử đó và được xếp theo một thứ tự nhất định.

b. Công thức tính: Bằng qui tắc đếm, dễ thấy công thức tính số chỉnh hợp chập k của n phần tử là:

$$A_n^k = \frac{n!}{(n-k)!}, \ k \le n \qquad (1.2)$$

Ví dụ 1.4: Có mấy cách chọn ngẫu nhiên 2 người, một người lau bảng, một người quét lớp cho một buổi trực nhật từ một tổ có 5 người?

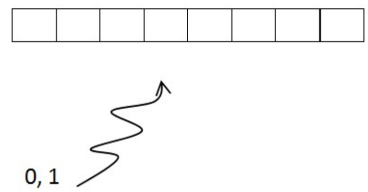
Chỉnh hợp lặp:

Khi k phần tử lấy ra có sự trùng lại nhưng vẫn có thứ tự thì ta có chỉnh hợp lặp chập k của n phần tử. Lúc này, số chỉnh hợp chập k của n phần tử được cho bởi công thức:

$$\overline{A}_n^k = n^k \text{ , } k \text{ là số tự nhiên tùy ý.} \qquad (1.4)$$

Ghi chú: Thường trường hợp này rất dễ nhầm lẫn vai trò của n và k nên nếu thấy lúng túng thì ta sẽ xác định lại số chỉnh hợp lặp bằng qui tắc đếm.

Ví dụ 1.5: Một bit (đơn vị thông tin cơ bản nhất trong tin học) có hai trạng thái 0 và 1. Một byte (gồm 8 bit) có thể biểu diễn bao nhiêu trạng thái? (Tương tự cho KB, MB, GB).



Hình 1.2.

1.3. Hoán vị

a. Định nghĩa: Một hoán vị của n phần tử khác nhau là một cách xếp thứ tự n phần tử đó.

b. Công thức tính: Số hoán vị của n phần tử là:

$$P_n = n! \qquad (1.5)$$

Số hoán vị của n phần tử chính là số chỉnh hợp chập n của n phần tử.

Ví dụ 1.6: Có mấy cách treo 4 bức tranh lên 4 vị trí cố định trên tường?

c. Hoán vị lặp: Nếu nhóm n phần tử chỉ gồm k loại phần tử khác nhau, loại phần tử thứ i có m_i phần tử trong n phần tử, $\sum_{i=1}^{k} m_i = n$, và các phần tử cùng loại không phân biệt nhau. Khi đó, ta có hoán vị lặp và công thức tính số hoán vị lặp sẽ là:

$$P_n = \frac{n!}{m_1!....m_k!} \qquad (1.6)$$

Lý do chia cho các $m_i!$ là do chúng ta có nhóm không đổi khi hoán vị các phần tử cùng loại.

Ví dụ 1.7: Số các số tự nhiên 7 chữ số trong đó có ba chữ số 1, hai chữ số 2 và hai chữ số 5 là số hoán vị của các phần tử: 1, 1, 1, 2, 2, 5, 5. Tức là có:

1.4. Tổ hợp

a. Định nghĩa: Một tổ hợp chập k của n phần tử khác nhau là một nhóm gồm k phần tử khác nhau được lấy từ n phần tử đó, thứ tự không quan trọng.

Ví dụ 1.8: Trong một giải đấu bóng đá gồm 4 đội thi đấu vòng tròn tính điểm, mỗi trận đấu sẽ được xác định bởi hai đội khác nhau mà không phân biệt thứ tự của đội bóng tham gia trận đấu đó. Do đó, mỗi trận đấu là một tổ hợp chập 2 của 4 đội bóng tham gia giải đấu.

Cũng cần để ý rằng, nếu đây là giải bóng đá vô địch quốc gia thì thể lệ thi đấu thường là hai lượt trận, "sân nhà" và "sân khách", nên mỗi trận đấu sẽ được xác định bởi hai đội bóng và thứ tự của đội trong trận đấu đó. Chẳng hạn, trận đấu đội A gặp đội B diễn ra trên sân nhà của đội A, trong khi trận đội B gặp đội A diễn ra trên sân nhà của đội B. Lúc này, mỗi trận đấu trong giải vô định quốc gia là một chỉnh hợp chập 2 của số đội tham gia.

b. Công thức tính: Số tổ hợp chập k của n phần tử, $k \le n$, là:

$$C_n^k = \frac{n!}{k!(n-k)!}\ ,\ k \le n. \qquad (1.7)$$

Ghi chú: Số tổ hợp chập k của n phần tử còn được kí hiệu là $\binom{n}{k}$ và còn được đọc là "số tổ hợp n chập k phần tử".

Ví dụ 1.9: Có bao nhiêu cách lập một tổ gồm 3 người từ 10 người đã cho?

Ví dụ 1.10: Có bao nhiêu cách lập một tổ gồm 3 người trong đó có 1 nữ từ 10 người, trong đó có 4 nữ?

Tổ hợp lặp:

Mỗi cách chọn ra k vật từ n loại vật khác nhau, trong đó mỗi loại vật có thể được chọn lại nhiều lần, được gọi là tổ hợp lặp chập k của n. Số các tổ lặp chập k của n là C_{n+k-1}^k.

Ví dụ: Một người đi mua 5 tờ giấy màu thủ công và chỉ muốn mua trong số các màu xanh, đỏ và tím. Hỏi có mấy cách chọn mua khác nhau?

Theo công thức tổ hợp lặp, số cách chọn mua khác nhau là:

$$C_{3+5-1}^{5} = C_{7}^{5} = 21$$

1.5. Nhị thức Newton

Ta đã biết hằng đẳng thức đáng nhớ:

$$(x+a)^2 = x^2 + 2xa + a^2$$

$$(x+a)^3 = x^3 + 3x^2a + 3xa^2 + a^3$$

Newton đã chứng minh được công thức tổng quát như sau:

$$(x+a)^n = \sum_{k=0}^{n} C_n^k x^{n-k} a^k = \sum_{k=0}^{n} C_n^k x^n a^{n-k}. \qquad (1.8)$$

Ta có thể vận dụng công thức trên trong việc đếm như trong ví dụ sau:

Ví dụ 1.11: Cho tập hợp A có n phần tử. Hãy đưa ra cách đếm tất cả các tập con của tập A.

Giải: Lần lượt đếm, tất cả các tập con của tập A có 0,1,..., n phần tử. Số tập có k phần tử là số tổ hợp chập k của n, vì thứ tự của các phần tử trong tập hợp không quan trọng. Theo qui tắc đếm và theo nhị thức Newton, số tập con của tập A sẽ là:

$$C_n^0 + C_n^1 + \cdots + C_n^{n-1} + C_n^n = (1+1)^n = 2^n$$

Ví dụ 1.12: Một bàn dài có hai dãy ghế đối diện nhau, mỗi dãy có 6 ghế. Người ta muốn xếp chỗ ngồi cho 6 học sinh trường A và 6 học sinh trường B vào bàn nói trên. Hỏi có bao nhiêu cách xếp sao cho bất cứ hai học sinh nào ngồi đối diện nhau thì khác trường nhau?

Giải:

Ghế 1						
Ghế 2						

Hình 1.3

Ta có các trường hợp sau:

+ Trường hợp 1: Dãy ghế 1 gồm toàn học sinh trường A: Có 6!.6! cách xếp.

+ Trường hợp 2: Dãy ghế 1 gồm 5 học sinh trường A và 1 học sinh trường B (hình 1.4).

- Giai đoạn 1: Chọn ra 1 vị trí để xếp 1 học sinh trường B vào ghế 1: Có $C_6^1 = 6$ cách.

Ghế 1		B				
Ghế 2	B		B	B	B	B

Hình 1.4

- Giai đoạn 2: Xếp 6 học sinh trường B vào các vị trí chỗ ngồi đã chọn gồm 1 chỗ trên ghế 1 và 5 chỗ trên ghế 2: $6!$.
- Giai đoạn 3: Xếp 6 học sinh trường A vào các chỗ còn trống trên cả hai ghế: $6!$.

Theo qui tắc nhân, trường hợp này sẽ có $C_6^1 6!6!$ cách xếp.

+ Trường hợp i, $i = \overline{3,6}$: Tương tự trường hợp 2, dãy ghế 1 gồm $6 - i + 1$ học sinh trường A và $i - 1$ học sinh trường B sẽ có $C_6^{i-1} 6!6!$ cách xếp.

+ Trường hợp 7: Dãy ghế 1 gồm toàn học sinh trường B: Có 6!.6! cách xếp.

Vậy theo qui tắc đếm và theo nhị thức Newton, ta có số cách xếp thỏa yêu cầu đề bài là:

$$6!6!\left(1 + C_6^1 + C_6^2 + \cdots + C_6^5 + 1\right) = \left(6!\right)^2 (2)^6 = 33{,}177{,}600.$$

BÀI TẬP

1.1. Tìm n nếu biết:

a) $C_n^2 = 45$. (b) $\dfrac{A_n^4}{C_{n-1}^3} = 60$.

1.2. Chứng minh rằng: $C_{n+1}^r = C_n^{r-1}$.

1.3. Từ 5 chữ số 1, 2, 3, 4, 5 có thể lập được bao nhiêu số tự nhiên (3 chữ số):

a) Có 3 chữ số?

b) Có 3 chữ số khác nhau?

c) Có 3 chữ số giống nhau?

d) Có 3 chữ số mà tận cùng là chữ số 4?

e) Có 3 chữ số khác nhau mà tận cùng là chữ số 4?

f) Có 3 chữ số không tận cùng là chữ số 4?

1.4. Tung 3 con súc sắc (hột xí ngầu) phân biệt. Dựa vào số chấm xuất hiện trên mỗi con súc sắc, hỏi có bao nhiêu trường hợp:

a) Có thể xảy ra?

b) Xuất hiện 3 mặt khác nhau?

c) Xuất hiện 3 mặt giống nhau?

d) Xuất hiện mặt 5 chấm ở con súc sắc thứ nhất?

e) Xuất hiện ba mặt: 4, 5, 6 chấm?

f) Xuất hiện đúng một mặt 5 chấm?

Hình 1.5: Hình ảnh con súc sắc, hay hột xí ngầu.

1.5. Có 5 người lên 7 toa tàu một cách ngẫu nhiên. Có bao nhiêu trường hợp:

a) Có thể xảy ra?

b) 5 người cùng lên toa thứ 3?

c) 5 người cùng lên một toa?

d) 5 người lên 5 toa đầu và mỗi người một toa?

1.6. Xếp ngẫu nhiên 5 người vào 1 chiếc ghế dài có 5 chỗ. Có bao nhiêu cách xếp:

a) Năm người vào ghế trên?

b) Sao cho A và B ngồi hai đầu ghế?

c) Sao cho A ngồi cạnh B?

d) Sao cho A ngồi bên phải B?

e) Sao cho A và B không ngồi cạnh nhau?

f) Sao cho A ngồi chính giữa B và C?

1.7. Một lô sản phẩm gồm 6 sản phẩm loại A và 7 sản phẩm loại B. Từ lô sản phẩm trên, người ta lấy ngẫu nhiên 5 sản phẩm cùng một lúc để kiểm tra. Hỏi có bao nhiêu cách lấy sao cho trong đó có:

a) Đúng 2 sản phẩm loại A?

b) Đúng 1 sản phẩm loại B?

c) Ít nhất 1 sản phẩm loại A?

1.8. Một hộp chứa 3 quả cầu trắng, 2 đen và 5 đỏ. Lấy ngẫu nhiên từ hộp ra 6 quả cầu cùng một lúc. Hỏi có bao nhiêu cách lấy:

a) Sao cho trong đó có 2 cầu trắng, 2 cầu đen và 2 cầu đỏ?

b) Sao cho có đúng hai cầu trắng?

1.9. Trên mặt phẳng có 20 điểm không có 3 điểm nào thẳng hàng. Qua mỗi cặp điểm phân biệt ta vẽ được một đường thẳng. Hỏi có bao nhiêu đường thẳng như vậy?

1.10. Lập công thức tính số đường chéo của một n giác lồi ($n \geq 4$).

1.11. Một lớp có 45 học sinh trong đó có 20 nam. Hỏi có bao nhiêu cách lập nhóm gồm:

a) 3 học sinh.

b) 3 học sinh gồm 2 nam 1 nữ.

1.12. Có 5 bức tranh để trên bàn. Hỏi có mấy cách:

a) Lấy ra 3 bức để treo lên tường?

b) Lấy ra 3 bức và treo lên 3 vị trí định sẵn trên tường?

1.13. Từ các chữ số 0, 1, 2, 3, 4, 5 có thể lập được bao nhiêu số tự nhiên gồm 4 chữ số đôi một khác nhau.

1.14. Từ các chữ số 0, 1, 2, 3, 4, 5 có thể lập được bao nhiêu số tự nhiên chẵn gồm 4 chữ số đôi một khác nhau.

1.15. Từ các chữ số 0, 1, 2, 3, 4, 5 có thể lập được bao nhiêu số tự nhiên gồm 4 chữ số sao cho chữ số 1 xuất hiện 2 lần, còn 2 chữ số còn lại khác nhau và khác 1.

1.16. Từ các chữ số 0,1, 2, 3, 4, 5 có thể lập được bao nhiêu số tự nhiên chẵn gồm 4 chữ số đôi một khác nhau mà chữ số đầu tiên là chữ số lẻ.

1.17. Có 3 đường thẳng song song nằm ngang cắt 5 đường thẳng song song thẳng đứng. Hỏi có bao nhiêu hình chữ nhật được tạo thành (có giải thích)? Mở rộng cho m và n đường thẳng.

1.18. Một lô sản phẩm gồm có 6 sản phẩm loại A và 4 sản phẩm loại B. Người ta lấy ngẫu nhiên 4 sản phẩm. Có bao nhiêu cách lấy được ít nhất một sản phẩm loại B.

Chương 2
PHÉP THỬ VÀ BIẾN CỐ

Chương này trình bày các khái niệm cơ bản nhất của lý thuyết xác suất, đó là khái niệm phép thử và biến cố. Sau khi học xong chương này, sinh viên phải nắm được hai khái niệm trên cùng với các quan hệ giữa các biến cố để từ đó có cơ sở tìm hiểu lý thuyết xác suất.

2.1. Khái niệm về phép thử và biến cố

- **Phép thử** được hiểu là một nhóm các hành động, hoặc thí nghiệm do ta tiến hành hoặc dự định tiến hành nhằm nghiên cứu một vấn đề nào đó. Phép thử có thể đơn giản chỉ là gieo một con súc sắc, nhưng đôi khi rất phức tạp như việc lai tạo các giống mới trong nông nghiệp.

- **Biến cố**, hay **sự kiện**, được hiểu là một sự vật, hiện tượng trong đời sống hằng ngày. Có thể hiểu một biến cố là một trong các kết cục có thể có của một phép thử tương ứng.

Ví dụ 2.1: Phép thử là tung một đồng xu, khi đó thông thường ta sẽ có 2 kết cục hay 2 biến cố (2 sự kiện) là biến cố "được mặt sấp", và biến cố "được mặt ngửa" của đồng xu.

- **Biến cố ngẫu nhiên**, được ký hiệu bằng các chữ in: $A, B, \ldots$, là biến cố ta không biết trước có xảy ra hay không khi thực hiện phép thử tương ứng.

Biến cố *luôn xảy ra* khi thực hiện phép thử tương ứng gọi là **biến cố chắc chắn**, ký hiệu là Ω, còn biến cố *luôn không xảy ra* gọi là **biến cố không thể**, kí hiệu là $\varnothing$.

Ví dụ 2.2: Gieo một con súc sắc.

Gọi: A là biến cố xuất hiện mặt có số chấm nhỏ hơn 7.

B là biến cố xuất hiện mặt 7 chấm.

C là biến cố xuất hiện mặt 1 chấm.

Khi đó A là biến cố chắc chắn.

B là biến cố không thể.

C là biến cố ngẫu nhiên.

2.2. Quan hệ giữa các biến cố

 Định nghĩa 2.1: Ta nói biến cố A kéo theo biến cố B, kí hiệu là $A \subset B$ hay $A \Rightarrow B$, nếu A xảy ra làm cho B xảy ra. Ta còn nói *A là biến cố thuận lợi cho biến cố B*.

Ví dụ 2.3: Gieo một con súc sắc. Khi đó nếu gọi A_i là biến cố xuất hiện mặt $i, i = \overline{1,6}$ và B là biến cố xuất hiện mặt chẵn thì B là biến cố có các biến cố thuận lợi là A_2, A_4, A_6.

 Định nghĩa 2.2: Ta nói biến cố A bằng biến cố B**,** kí hiệu là $A = B$**,** nếu như A xảy ra làm cho B xảy ra và ngược lại. Ta còn nói là: *biến cố A biểu diễn theo biến cố B*.

 Định nghĩa 2.3: Biến cố "không xảy ra biến cố A", kí hiệu là $\overline{A}$, được gọi là biến cố đối lập với biến cố A.

Ví dụ 2.4: A là biến cố "lớp tôi có ít nhất 2 nữ".

Khi đó $\overline{A}$ là biến cố "lớp tôi có nhiều nhất 1 nữ".

2.3. CÁC PHÉP TOÁN VỀ BIẾN CỐ

 Định nghĩa 2.4: Tổng hai biến cố A và B là biến cố C, ký hiệu: $C = A + B$ hay $C = A \cup B$**,** xảy ra khi và chỉ khi có ít nhất một trong hai biến cố A và B xảy ra. Tương tự cho tổng n biến cố.

 Định nghĩa 2.5: Tích hai biến cố A và B là biến cố C (kí hiệu là $C = A.B$ hay $C = A \cap B$) xảy ra khi và chỉ khi A và B đồng thời xảy ra. Tương tự cho tích n biến cố. Nếu $A\ B = \varnothing$ thì ta nói A và B là hai **biến cố xung khắc**.

Ví dụ 2.5: Có hai hộp bi được đánh số 1, 2 chứa cả bi đỏ và bi trắng. Lấy ngẫu nhiên mỗi hộp 1 bi và gọi A_i là biến cố lấy được bi đỏ từ hộp thứ i, $i = 1,2$. Khi đó:

Biến cố lấy được bi đỏ là: $A_1 + A_2$.

Biến cố lấy được hai bi đỏ là: A_1A_2.

Biến cố lấy được (đúng) một bi đỏ là: $\overline{A}_1A_2 + A_1\overline{A}_2$. Biến cố này xung khắc với biến cố A_1A_2

 Tính chất:

- $A+\overline{A}=\Omega$
- $A\overline{A}=\varnothing$
- $\overline{AB}=\overline{A}+\overline{B}$
- $\overline{A+B}=\overline{A}.\overline{B}$

2.4. Thêm một số khái niệm

Định nghĩa 2.6: Hệ n biến cố $\left\{A_i,\ i=\overline{1,n}\right\}$ được gọi là *hệ đầy đủ* khi các biến cố xung khắc từng đôi và tổng của chúng là một biến cố chắc chắn.

Ví dụ 2.6: Các biến cố $A_i, i=\overline{1,6}$ trong **Error! Reference source not found.** lập thành một hệ đầy đủ.

Hệ biến cố đồng khả năng: Khi thực hiện phép thử, nếu trong các biến cố không có biến cố nào ưu tiên hay xảy ra hơn các biến cố khác thì ta có các biến cố đồng khả năng.

 Định nghĩa 2.7: *Biến cố sơ cấp* là biến cố mà không có biến cố nào kéo theo nó và tập các biến cố này lập thành một hệ đầy đủ. Biến cố sơ cấp còn gọi là *biến cố cơ bản.*

 Định nghĩa 2.8: Tập tất cả các biến cố sơ cấp lập thành *không gian các biến cố sơ cấp*.

Chẳng hạn trong Ví dụ 2.3, $A_i, i=\overline{1,6}$ là các biến cố sơ cấp và $\left\{A_1, A_2, ..., A_6\right\}$ là không gian các biến cố sơ cấp của phép thử gieo súc sắc.

Ví dụ 2.7: Một hộp có 10 quả bóng bàn được đánh số thứ tự trong đó có 3 quả màu trắng, 7 quả màu cam. Lấy ngẫu nhiên 1 quả bóng bàn từ hộp.

Nếu quan tâm đến việc lấy được quả bóng số mấy thì ta có 10 biến cố đồng khả năng. Các biến cố này là các biến cố sơ cấp và lập thành một hệ đầy đủ.

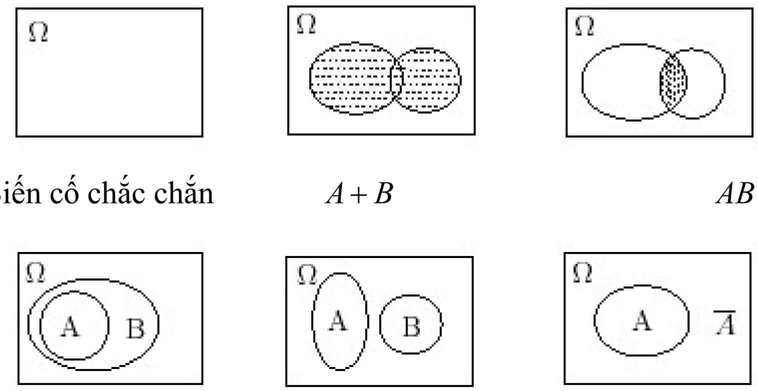
Nếu quan tâm đến việc lấy được quả bóng màu gì thì ta có 2 biến cố không đồng khả năng. Các biến cố này cũng lập thành một hệ đầy đủ.

2.5. Hình ảnh liên hệ giữa biến cố và tập hợp

Bảng sau tóm tắt hình ảnh tương ứng giữa quan hệ tập hợp và quan hệ biến cố.

Ký hiệu	Tập hợp	Xác suất
e_i, ω_i	Điểm	Biến cố sơ cấp
A	Tập hợp	Biến cố
Ω	Không gian	Biến cố chắc chắn, không gian biến cố sơ cấp
ϕ	Tập rỗng	Biến cố không thể
$A \subset B,\ A \Rightarrow B$	A là tập con của B	A xảy ra kéo theo B xảy ra
$A = B$	A bằng B	A biểu diễn theo B
$A \cap B,\ A.B$	tập các phần tử thuộc cả A và B	Biến cố xảy ra khi cả A và B xảy ra
$A \cup B,\ A + B$	tập các phần tử thuộc A hoặc B	Biến cố xảy ra khi A hoặc B xảy ra
$\bigcup_i A_i, \sum_i A_i$	Hợp nhiều tập hợp	Biến cố xảy ra khi có ít nhất một biến cố xảy ra.
$\bigcap_i A_i, \prod_i A_i$	Giao nhiều tập hợp	Biến cố xảy ra khi có ít nhất một biến cố xảy ra.
$A \cap B = \phi$	A rời B	Biến cố A và B xung khắc.
$\overline{A}$	Bù của tập A	Biến cố đối lập của b/cố A
$A_i.A_j = \varnothing, \forall i \neq j$ và $\sum_i A_i = \Omega$	Không gian chia thành các tập rời nhau	Hệ đầy đủ các biến cố.

Biểu diễn các biến cố bằng biểu đồ Venn:



Biến cố chắc chắn $A+B$ AB

$A \Rightarrow B$ A , B xung khắc Biến cố đối lập

Hình 2.1.

Ví dụ 2.8: Có 3 xạ thủ cùng bắn một con thú, mỗi người bắn một viên đạn.

Gọi A là biến cố xạ thủ 1 bắn trúng con thú.

B là biến cố xạ thủ 2 bắn trúng con thú.

C là biến cố xạ thủ 3 bắn trúng con thú.

D là biến cố con thú bị trúng đạn.

E là biến cố con thú không bị trúng đạn.

F là biến cố con thú bị trúng đúng 2 viên đạn

G là biến cố con thú bị trúng đúng 1 viên đạn

H là biến cố con thú bị trúng ít nhất 1 viên đạn.

Hãy mô tả các biến cố $\overline{A}, \overline{B}, \overline{C}$ và viết các biến cố D, E, F, G theo A, B, C và $\overline{A}, \overline{B}, \overline{C}$

Giải:

$\overline{A}$ là biến cố xạ thủ thứ nhất bắn trượt con thú.

$\overline{B}$ là biến cố xạ thủ thứ hai bắn trượt con thú.

$\overline{C}$ là biến cố xạ thủ thứ ba bắn trượt con thú.

$D = A + B + C$

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

$F = AB\overline{C} + A\overline{B}C + \overline{A}BC$

$G = A\overline{B}\overline{C} + \overline{A}B\overline{C} + \overline{A}\overline{B}C$

$H = D = \overline{E} = \overline{\overline{A}\overline{B}\overline{C}}$

$G \subset D$ (hay $G \Rightarrow D$)

BÀI TẬP

Trong các bài tập sau, sinh viên lưu ý cách phát biểu biến cố, cách đặt (cách gọi) các biến cố và cách viết một biến cố theo một nhóm các biến cố (tức là biểu diễn một biến cố qua một nhóm các biến cố)

2.1. Kiểm tra một lô sản phẩm. Gọi A là biến cố "có ít nhất một sản phẩm bị hư", B là biến cố "có từ hai sản phẩm bị hư trở lên". Hãy mô tả các biến cố $\overline{A}$ và $\overline{B}$

2.2. Trong số các sinh viên cùng lớp, ta chọn ngẫu nhiên 1 sinh viên. Gọi A là biến cố "sinh viên được chọn là nam", B là biến cố "sinh viên được chọn thuộc khu vực I", C là biến cố "sinh viên được chọn ở nội trú"

a) Hãy mô tả biến cố: $A \cap B \cap \overline{C}, \overline{A} \cap \overline{B} \cap C$

b) Với điều kiện nào ta có: $\overline{A} = B$

c) Với điều kiện nào ta có: $A \cap B \cap C = A$

d) Với điều kiện nào ta có: $\overline{C} \subset B$

2.3. Cho A, B, C là các biến cố ngẫu nhiên. Hãy viết các biểu thức chỉ các biến cố sau:

a) Chỉ có A xảy ra.

b) A và B xảy ra nhưng C không xảy ra.

c) Cả 3 biến cố đều xảy ra.

d) Cả 3 biến cố đều không xảy ra.

e) Có ít nhất một biến cố xảy ra.

f) Có ít nhất một biến cố không xảy ra.

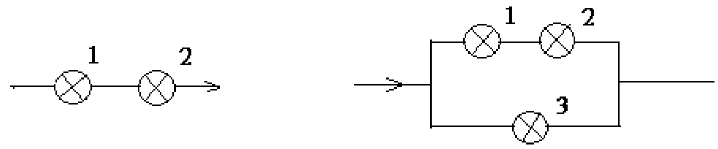
g) Có ít nhất 2 biến cố xảy ra.

h) Có nhiều nhất 1 biến cố xảy ra.

i) Có không ít hơn 2 biến cố xảy ra.

j) Có không nhiều hơn 2 biến cố xảy ra.

2.4. Một mạch điện như hình vẽ, kí hiệu A_i là biến cố bóng đèn thứ i bị hỏng, $i = 1,2,3$. Hãy viết các biến cố sau theo A_i và $\overline{A}_i$, $i = 1,2,3$



Hình 1 Hình 2

Hình 2.2.

a) Mạch có dòng điện chạy qua.

b) Mạch mất điện.

H2

2.5. Ba người đi săn cùng bắn mỗi người một phát đạn vào một con mồi. Gọi A_i là biến cố "người thứ i bắn trúng con mồi", $i=1,2,3$; A là biến cố "con mồi bị trúng đạn", B là biến cố "con mồi chỉ bị trúng một viên đạn", và C là biến cố "con mồi không bị trúng đạn". Hãy biểu diễn các biến cố A, B, C qua các biến cố A_i và $\overline{A_i}$, $i=1,2,3$.

2.6. Giả sử một môn học A ở học kỳ I có 2 lần thi. Xét một thí sinh, kí hiệu A_i là biến cố thí sinh này thi qua ở lần thi thứ i, $i=1,2$. Hãy biểu diễn các biến cố:

a) Thí sinh thi không đạt môn A ở học kỳ I.

b) Thí sinh này thi đạt môn A ở học kỳ I.

2.7. Một xạ thủ bắn 3 phát đạn vào bia. Ký hiệu A_i là biến cố bắn viên thứ i trúng bia, $i=1,2,3$. Hãy biểu diễn các biến cố:

a) Có một viên đạn trúng bia.

b) Không có viên đạn nào trúng bia.

c) Có ít nhất một viên đạn trúng bia.

2.8. Cho A và B là 2 biến cố đã biết, hãy tìm biến cố X từ đẳng thức sau:

$$\overline{X\cup A}\cup\overline{X\cup\overline{A}}=B$$

Chương 3
XÁC SUẤT VÀ CÔNG THỨC TÍNH

Chương này trình bày khái niệm xác suất, một số định nghĩa xác suất và các công thức tính xác suất. Học xong chương này, sinh viên phải nắm vững khái niệm xác suất và vận dụng thành thạo các qui tắc tính xác suất.

3.1. Khái niệm xác suất

Để so sánh hoặc đánh giá một hoặc nhiều biến cố về khả năng xuất hiện trong một phép thử tương ứng, người ta gán cho mỗi biến cố một con số không âm sao cho với hai biến cố bất kỳ, biến cố nào có khả năng xuất hiện nhiều hơn thì gán cho số lớn hơn, các biến cố có cùng khả năng xuất hiện thì gán cho cùng một con số.

Số gán cho biến cố A, ký hiệu là $P(A)$, được gọi là xác suất của biến cố A.

Sau đây ta tìm hiểu cách định nghĩa xác suất của một biến cố.

3.2. Các định nghĩa xác suất

3.2.1 Định nghĩa xác suất cổ điển

$$P(A) = \frac{m_A}{n} \tag{3.1}$$

trong đó n là số các biến cố sơ cấp đồng khả năng có thể xảy ra khi thực hiện phép thử, còn m_A là số các biến cố sơ cấp thuận lợi cho biến cố A.

Ví dụ3f 3. Gieo một con súc sắc. Số các trường hợp đồng khả năng có thể xảy ra là $n = 6$. Do đó, xác suất xuất hiện một mặt nào đó là $\frac{1}{6}$.

Ví dụ3f 3. Lấy ngẫu nhiên một bi trong hộp có 4 bi đỏ và 7 bi đen, khi đó xác suất lấy được một bi cụ thể nào đó là $\frac{1}{11}$, còn xác suất để lấy được bi đỏ là $\frac{4}{11}$.

Ví dụ3f 3. Một hộp có 10 bi, trong đó có 3 bi đỏ và 7 bi đen.

a) Lấy ngẫu nhiên 1 bi, tính xác suất lấy được bi đỏ.

b) Lấy ngẫu nhiên 2 bi, tính xác suất lấy được 2 bi đỏ.

c) Lấy ngẫu nhiên 2 bi, tính xác suất lấy được 1 bi đỏ.

Giải:

a) Lấy ngẫu nhiên 1 bi trong 10 bi nên có $n = 10$ trường hợp có thể xảy ra, trong đó có 3 trường hợp thuận lợi cho việc lấy được bi đỏ nên xác suất lấy được bi đỏ là 3/10.

b) Lấy ngẫu nhiên 2 bi trong 10 bi nên có $n = C_{10}^2 = 45$ trường hợp có thể xảy ra.

Gọi A là biến cố lấy được 2 bi đỏ. Ta có số thuận lợi cho A là $m_A = C_3^2$ =3. Do đó

$$P(A) = \frac{m_A}{n} = \frac{3}{45} = \frac{1}{15}.$$

c) Lấy ngẫu nhiên 2 bi trong 10 bi nên có $n = C_{10}^2 = 45$ trường hợp có thể xảy ra. Gọi B là biến cố lấy được 1 bi đỏ. Theo qui tắc đếm ta có số thuận lợi cho B là $m_B = 3 \times 7 = 21$. Do đó:

$$P(B) = \frac{m_B}{n} = \frac{21}{45} = \frac{7}{15}.$$

Ghi chú: Định nghĩa xác suất theo lối cổ điển có một số hạn chế như sau:

- Nó chỉ xét trong trường hợp phép thử có hữu hạn các biến cố sơ cấp.

- Không phải lúc nào ta cũng có hệ biến cố đồng khả năng.

Định nghĩa xác suất sau sẽ khắc phục các hạn chế nêu trên.

3.2.2 Định nghĩa xác suất theo thống kê

Xét biến cố A là kết cục của một phép thử ngẫu nhiên. Giả sử ta lặp lại phép thử này n lần thì thấy biến cố A xuất hiện m lần (m được gọi là tần số xuất hiện của biến cố A). khi n khá lớn thì người ta nhận thấy một điều là tỉ số $\frac{m}{n}$ sẽ xấp xỉ một số p cố định (tỉ số $\frac{m}{n}$ được gọi là tần suất của biến cố A). Khi đó ta nói p là xác suất của biến cố A, nghĩa là:

$$P\left(A\right) = p = \lim_{n \to +\infty} \frac{m}{n} \qquad (3.2)$$

Trong thực tế, người ta coi như $p \approx \frac{m}{n}$ khi n đủ lớn.

Chẳng hạn, khi quan sát ngẫu nhiên 100 người có hút thuốc lá, thấy có 91 người bị viêm phổi. Khi đó có thể nói rằng nếu bạn hút thuốc lá thì xác suất bạn bị viêm phổi *sẽ khoảng* 91% (hay 0.91).

Ví dụ3f 3. Để nghiên cứu khả năng xuất hiện mặt sấp khi tung một đồng tiền, người ta tiến hành tung đồng tiền đó nhiều lần và thu được kết quả sau:

Người làm thí nghiệm	Số lần tung (n)	Số lần được mặt sấp (m)	Tần suất ($\frac{m}{n}$)
Buyffon	4,040	2,048	0.5069
Pearson	12,000	6,019	0.5016
Pearson	24,000	12,012	0.5005

Bảng 3.1

Bảng trên cho thấy, khi số lần tung càng lớn thì tần suất xuất hiện mặt sấp $\frac{m}{n}$ càng gần với $\frac{1}{2}$

3.2.3 Định nghĩa xác suất theo hình học

Giả sử tập hợp (vô hạn) các kết cục đồng khả năng của một phép thử có thể biểu thị bởi miền hình học D, còn các kết cục thuận lợi cho biến cố A cho bởi miền D_0.

Khi đó ta định nghĩa xác suất của biến cố A là:

$$P(A) = \frac{S(D_0)}{S(D)} \qquad (3.3)$$

Với S là các phép đo tương ứng trong từng không gian. S có thể là độ dài của đường, là diện tích của miền, là thể tích của vật thể,...

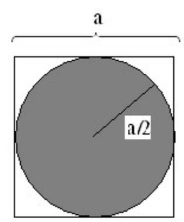
Ví dụ3f 3. Lấy ngẫu nhiên một điểm M trong hình vuông cạnh a. Tính xác suất lấy M thuộc hình tròn nội tiếp hình vuông trên.

Giải:

Gọi A là biến cố điểm M thuộc hình tròn nội tiếp hình vuông.

Ta có:

$$P(A) = \frac{S_{ht}}{S_{hv}} = \frac{\pi.\left(\frac{a}{2}\right)^2}{a^2} = \frac{\pi}{4}$$



Hình 3.1

Ví dụ: Tìm xác suất của điểm M rơi vào hình tròn nội tiếp tam giác đều có cạnh 2 cm

3.3. Tính chất cơ bản[1]

(i) $0 \le P(A) \le 1$ (3.4)

(ii) $P(\varnothing) = 0$ (3.5)

(iii) $P(\Omega) = 1$ (3.6)

(iv) Nếu $A_i,\ i = \overline{1,n}$ là các biến cố xung khắc từng đôi thì

[1] *Các tính chất này chính là các tiên đề trong định nghĩa xác suất của nhà toán học người Nga Kolmogorov.*

$$P\left(\sum_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P\left(A_i\right) \quad (3.7)$$

Tính chất (iv) còn được gọi là tính chất cộng tính của xác suất.

Nhận xét: Do A và $\overline{A}$ thỏa tính chất (iv) và thoả $A \cup \overline{A} = \Omega$ nên từ tính chất (iii) ta có công thức xác suất của biến cố đối lập:

$$P\left(\overline{A}\right) = 1 - P\left(A\right) \quad (3.8)$$

Ví dụ3f 3. Lấy ngẫu nhiên 3 bi trong một hộp có 12 bi đỏ và 8 bi đen. Tính xác suất lấy được ít nhất một bi đỏ.

Giải: Đặt A là biến cố lấy được ít nhất một bi đỏ.

Khi đó thấy ngay $\overline{A}$ là biến cố "không lấy được bi đỏ nào", tức là lấy được 3 bi đen. Ta có:

$$P\left(\overline{A}\right) = \frac{C_8^3}{C_{20}^3} = \frac{14}{285}$$

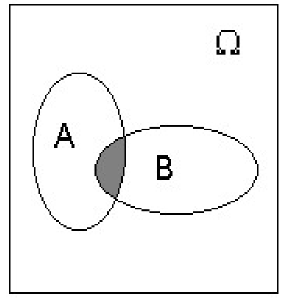
Nên công thức (3.8) ta có

$$P\left(A\right) = 1 - P\left(\overline{A}\right) = \frac{271}{285}.$$

Ví dụ 2.30. Có 20 máy tính trong một cửa hàng, trong số đó, 15 chiếc hoàn toàn mới và 5 chiếc được tân trang lại. Sáu máy tính được mua cho một phòng thí nghiệm sinh viên. Từ cái nhìn đầu tiên, chúng không thể phân biệt được, vì vậy sáu máy tính được chọn ngẫu nhiên. Tính xác suất trong số các máy tính được chọn, hai máy tính được tân trang lại.

3.4. Xác suất có điều kiện

Ví dụ mở đầu: Lấy ngẫu nhiên một điểm M trong hình Ω (hình 3.2). Gọi A là biến cố điểm M thuộc miền A, B là biến cố điểm M thuộc miền B.



Hình 3.2

Giả sử biến cố B đã xảy ra. Khi đó, ta thấy rằng các trường hợp có thể xảy ra được biểu diễn bởi hình B và miền thuận lợi cho biến cố A là $A \cap B$ nên theo định nghĩa xác suất hình học, xác suất để A xảy ra khi biết B xảy ra, ký hiệu là $P\left(A/B\right)$, phải là:

$$P\left(A/B\right) = \frac{S(A \cap B)}{S(B)}$$

$P\left(A/B\right)$ được gọi là xác suất có điều kiện (trong trường hợp này, nó là xác suất của biến cố A với điều kiện biến cố B đã xảy ra).

Tất nhiên, $P\left(A/B\right)$ thường khác với $P\left(A\right) = \dfrac{S(A)}{S(\Omega)}$

Nếu đem chia tử và mẫu cho S(Ω) thì ta có:

$$P\left(A/B\right) = \frac{\dfrac{S(A \cap B)}{S(\Omega)}}{\dfrac{S(B)}{S(\Omega)}} = \frac{P(AB)}{P(B)} \text{ , với } P\left(B\right) > 0$$

Từ đây ta có công thức định nghĩa xác suất có điều kiện.

Định nghĩa xác suất có điều kiện

Giả sử $P\left(B\right) > 0$, khi đó

$$P\left(A/B\right) = \frac{P(AB)}{P(B)} \qquad (3.9)$$

Ví dụ3f 3. Gieo một con súc sắc. Gọi A là biến cố xuất hiện mặt 1 chấm, B là biến cố xuất hiện mặt lẻ. Khi đó:

$$P\left(A/B\right) = \frac{P(AB)}{P(B)} = \frac{P(A)}{P(B)} \text{ (Do } A \text{ kéo theo } B\text{)}$$

$$= \frac{1/6}{1/2} = \frac{1}{3}$$

Nếu tính trực tiếp bằng công thức xác suất cổ điển thì ta thấy: Do biết B đã xảy ra nên ta chỉ có 3 biến cố sơ cấp có thể xảy ra là 1, 3, và 5 chấm, còn số thuận lợi cho A là 1 nên

$$P(A/B) = \frac{1}{3}$$

Như vậy, trong nhiều trường hợp, chúng ta vẫn dùng công thức xác suất cổ điển để tính xác suất có điều kiện.

Ví dụ3f 3. Có 40 phiếu đề thi xác suất, mỗi phiếu đề thi chỉ có một câu hỏi được phân bố theo bảng sau:

Số lượng	Câu khó	Câu dễ
Câu lý thuyết	6	8
Câu bài tập	12	14

Bảng 3.2

Rút ngẫu nhiên 1 phiếu đề thi. Ký hiệu các biến cố:

A : Rút được câu khó

B : Rút được câu lý thuyết.

Khi đó thấy ngay:

- Xác suất rút được câu khó là:
- Xác suất rút được câu lý thuyết là:
- Xác suất rút được câu lý thuyết khó là:
- Xác suất rút được câu khó khi biết đó là câu lý thuyết là

Ví dụ3f 3. Một người có 1 hộp bi gồm 3 bi đỏ và 4 bi đen. Giả sử bị rơi mất 1 bi màu đỏ, hãy tính xác suất để khi lấy ngẫu nhiên ra 2 bi thì người đó có được 2 bi đỏ.

Ví dụ 2.32 (Độ tin cậy của một xét nghiệm). Xét nghiệm cho một loại virus (bao gồm cả một cuộc tấn công virus trên mạng máy tính) thường có độ chính xác nhất định. Nó đáng tin cậy 95% cho bệnh nhân nhiễm bệnh và đáng tin cậy 99% cho những người khỏe mạnh. Nghĩa là, nếu một bệnh nhân có virus (sự kiện V), xét nghiệm có virus (sự kiện S) sẽ có xác suất là P(S | V) = 0.95 và nếu bệnh nhân không có virut, thì P (not S/not V) = 0.99.

Xem xét một bệnh nhân có kết quả xét nghiệm dương tính (nghĩa là, xét nghiệm cho thấy bệnh nhân có virus). Biết rằng đôi khi xét nghiệm là sai, một cách tự nhiên, bệnh nhân rất muốn biết xác suất rằng họ thực sự có virus, đây là xác suất có điều kiện. Viết công thức của xác suất này.

Các công thức tính xác suất

3.5.1 Công thức nhân:

Từ công thức xác suất có điều kiện, ta có

Công thức nhân 2 biến cố :

$$P(AB) = P(A).P(B/A) = P(B).P(A/B) \qquad (3.10)$$

Công thức nhân 3 biến cố :

$$P(ABC) = P(A).P(B/A).P(C/AB) \qquad (3.11)$$

Tổng quát, ta công thức nhân n biến cố như sau:

$$P(A_1A_2A_3...A_n) = P(A_1)P(A_2/A_1)P(A_3/A_1A_2)...P(A_n/A_1A_2...A_{n-1}) \qquad (3.12)$$

Công thức nhân từ 3 biến cố trở lên có thể chứng minh được bằng phương pháp qui nạp.

Ví dụ3f 3. Một lô hàng có 20 sản phẩm, trong đó có 3 phế phẩm. Lấy ngẫu nhiên lần lượt 2 lần, mỗi lần 1 sản phẩm và không hoàn lại, tính xác suất lấy được 2 phế phẩm.

Ví dụ: Giả sử lấy ngẫu nhiên 4 tờ giấy (không hoàn lại) từ một chiếc hộp chứa 25 tờ giấy, trong đó có 15 tờ giấy có chữ X được viết trên mỗi tờ và 10 tờ giấy có chữ Y được viết trên mỗi tờ giấy. Khi đó xác suất quan sát được chuỗi XYXX là

P (XYXX) =P(X).P(Y/X).P(X/XY).P(X/XYX)=

$\frac{15}{25}.\frac{10}{24}.\frac{14}{23}.\frac{13}{22} \approx 0.09$

3.5.2 Sự độc lập của các biến cố

a. Định nghĩa biến cố độc lập

Ta nói:

- A và B là hai biến cố *độc lập* nếu

$$P(AB) = P(A).P(B) \tag{3.13}$$

- Các biến cố $A_1, A_2, ..., A_n$, $n > 2$, độc lập từng đôi nếu

$$P(A_iA_j) = P(A_i).P(A_j),\ \forall i \neq j \tag{3.14}$$

- Các biến cố $A_1, A_2, ..., A_n$, $n > 2$ độc lập với nhau (độc lập trong toàn thể) nếu:

$$P\left(\prod_{i\in S} A_i\right) = \prod_{i\in S} P(A_i), \text{ với mọi tập con } S \subset \{1, 2, ..., n\} \tag{3.15}$$

Thấy ngay từ (3.15) có thể suy ra (3.14), còn ngược lại thì không.

Ví dụ3f 3. Xét phép thử có không gian các biến cố sơ cấp là

$$\Omega = \{1, 2, 3, 4\}$$

Đặt $A_1 = \{1, 2\}$, $A_2 = \{1, 3\}$, $A_3 = \{1, 4\}$

Ta thấy ngay $P(A_1) = P(A_2) = P(A_3) = \frac{1}{2}$

$$P(A_1A_2) = \frac{1}{4} = P(A_1)P(A_2)$$

$$P(A_1A_3) = \frac{1}{4} = P(A_1)P(A_3)$$

$$P(A_2A_3) = \frac{1}{4} = P(A_2)P(A_3)$$

Nghĩa là A_1, A_2, A_3 độc lập từng đôi, thế nhưng:

$$P(A_1A_2A_3) = \frac{1}{4} \neq \frac{1}{8} = P(A_1)P(A_2)P(A_3)$$

Nghĩa là A_1, A_2, A_3 không độc lập trong toàn thể.

Trong tài liệu này, khái niệm về tính độc lập của nhiều biến cố hiểu theo nghĩa "độc lập trong toàn thể", nghĩa là ta có:

$$A_1, A_2, ..., A_n \text{ độc lập} \Leftrightarrow P\left(\prod_{i\in S} A_i\right) = \prod_{i\in S} P(A_i), \forall S \subset \{1, 2, ..., n\}$$

Nói cách khác: các biến cố $A_1, A_2, ..., A_n$ độc lập nếu như với mọi tích có thể có của các biến cố, ta có xác suất của tích bằng tích các xác suất.

b. Tính chất: Nếu hai biến cố A và B độc lập thì

$P(A/B) = P(A), \quad P(B/A) = P(B).$

Các cặp biến cố ($A, \overline{B}$); ($\overline{A}, B$); ($\overline{A}, \overline{B}$) cũng độc lập.

Từ tính chất trên, ta thấy rõ hơn ý nghĩa của sự độc lập giữa hai biến cố.

c. Ý nghĩa

Nếu hai biến cố độc lập thì biến cố này có xảy ra hay không, không ảnh hưởng đến khả năng xảy ra của biến cố kia và ngược lại. Như vậy, đây chính là khái niệm "độc lập" mà ta thường hiểu.

*** Lưu ý**: việc kiểm tra tính độc lập của các biến cố thường dựa vào thực tế và trực giác như trong ví dụ dưới đây.

Ví dụ3f 3. Khi không nói rõ thì các biến cố kết quả của các phép thử sau là các biến cố độc lập.

- Việc hoạt động của n máy tính là các biến cố độc lập. Rõ ràng, nếu không có điều kiện đặc biệt nào khác thì ta thấy ngay các kết quả này "không liên quan gì đến nhau", nghĩa là chúng độc lập nhau. Giải thích tương tự cho các ví dụ dưới.
- Tung một đồng xu n lần, kết quả của mỗi lần tung là các biến cố độc lập.
- Việc lập trình bị lỗi của các chương trình con là độc lập,…

Ví dụ3f 3. Có hai hộp bi, hộp 1 có 10 bi trong đó có 3 bi đỏ; hộp 2 có 15 bi trong đó có 4 bi đỏ. Lấy ngẫu nhiên mỗi hộp 1 bi. Tính xác suất được 2 bi đỏ.

Ví dụ 2.18 (Độ tin cậy của lưu trữ dữ liệu). Giả sử một máy tính có xác suất hỏng ổ cứng là 1%. Để tăng độ tin cậy lưu trữ dữ liệu, người ta tạo hai bản sao lưu (backup), mỗi bản có xác suất bị hỏng là 2% và cả ba bản lưu dữ liệu này đều độc lập với nhau. Thông tin được lưu trữ chỉ bị mất nếu không may khi cả ba thiết bị gặp sự cố cùng một lúc. Tính xác suất dữ liệu được an toàn?

Ví dụ này cho thấy, khi các thành phần của hệ thống được kết nối song song, thì ít nhất một thành phần hoạt động để toàn bộ hệ thống hoạt động. Độ tin cậy của một hệ thống như vậy khá cao và việc sao lưu dữ liệu luôn có thể được coi là thiết bị được kết nối song song. Ngược lại, hệ thống được kết nối nối tiếp là hệ thống dễ bị "tổn thương" hơn.

3.5.3 Công thức cộng

Công thức cộng 2 biến cố

$$P(A+B)=P(A)+P(B)-P(AB) \tag{3.16}$$

Công thức cộng 3 biến cố

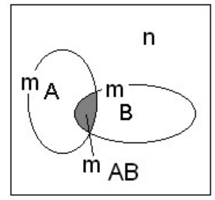
$$\begin{aligned}P(A+B+C)=&P(A)+P(B)+P(C)\\&-P(AB)-P(AC)-P(BC)+P(ABC)\end{aligned} \tag{3.17}$$

Công thức cộng tổng quát

$$\begin{aligned}P\left(\sum_{i=1}^{n}A_i\right)=&\sum_{i=1}^{n}P(A_i)-\sum_{i<j}P(A_iA_j)+\sum_{i<j<k}P(A_iA_jA_k)\\&-\cdots+(-1)^{n+1}P(A_1A_2\ldots A_n)\end{aligned} \tag{3.18}$$

Chứng minh: Để dễ hình dung, ta chỉ chứng minh công thức tổng của hai biến cố. Các công thức còn lại có thể làm tương tự.

Giả sử phép thử có n biến cố đồng khả năng có thể xảy ra, trong đó có m_A biến cố thuận lợi cho A, có m_B biến cố thuận lợi cho B, và có m_{AB} biến cố thuận lợi cho biến cố AB. Khi đó số biến cố thuận lợi cho biến cố $A+B$ là $m_A + m_B - m_{AB}$



Hình 3.3

Do đó:

$$P(A+B) = \frac{m_A + m_B - m_{AB}}{n} = \frac{m_A}{n} + \frac{m_B}{n} - \frac{m_{AB}}{n}$$
$$= P(A) + P(B) - P(AB)$$

Hệ quả:

Nếu A_i, $i = \overline{1,n}$ là các biến cố xung khắc từng đôi thì

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

Đây chính là tính chất cộng tính của xác suất (3.7).

Ví dụ3f 3. Một dây chuyền sản xuất gồm 2 công đoạn hoạt động độc lập. Xác suất để mỗi công đoạn ngừng hoạt động trong khoảng thời gian t lần lượt là 0.01; 0.02. Biết rằng dây chuyền sẽ ngừng hoạt động nếu có ít nhất 1 công đoạn ngừng hoạt động, hãy tính xác suất dây chuyền ngừng hoạt động trong khoảng thời gian t.

Giải:

Gọi A là biến cố dây chuyền ngừng hoạt động trong khoảng thời gian t.

Gọi A_i là biến cố công đoạn i ngừng hoạt động trong khoảng thời gian t, $i = 1,2$.

Theo đề bài thì A_1, A_2 độc lập và $P\left(A_1\right) = 0.01$; $P\left(A_2\right) = 0.02$.

Ta có

$$P\left(A\right) = P\left(A_1 + A_2\right)$$
$$= P\left(A_1\right) + P\left(A_2\right) - P\left(A_1 A_2\right)$$

$$\overset{\text{đl}}{=} 0.01 + 0.02 - (0.01)(0.02)$$

$$= 0.0298 .$$

Một cách làm khác là:

$$P(\overline{A}) = P(\overline{A}_1\overline{A}_2)$$

$$\overset{\text{đl}}{=} P(\overline{A}_1)P(\overline{A}_2)$$

$$= (0.99)(0.98) = 0.9702.$$

Do đó, theo công thức xác suất của biến cố đối lập ta có

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

$$= 1 - 0.9702 = 0.0298.$$

Ví dụ3f 3. Bỏ ngẫu nhiên 4 bức thư (khác nhau) vào 4 phong bì có điền sẵn 4 địa chỉ. Tính xác suất có ít nhất một lá đến đúng người nhận.

Giải:

Gọi A là biến cố có ít nhất một lá thư đến đúng người nhận, A_i là biến cố lá thư thứ i đến đúng người nhận, $i = \overline{1,4}$.

Ta có $P(A) = P(A_1 + A_2 + A_3 + A_4)$

$$= P(A_1) + P(A_2) + P(A_3) + P(A_4)$$
$$- P(A_1A_2) - P(A_1A_3) - P(A_1A_4) - P(A_2A_3) - P(A_2A_4) - P(A_3A_4)$$
$$+ P(A_1A_2A_3) + P(A_1A_2A_4) + P(A_1A_3A_4) + P(A_2A_3A_4)$$
$$- P(A_1A_2A_3A_4)$$
$$= 4.\frac{1}{4} - 6.\frac{1}{12} + 4.\frac{1}{24} - \frac{1}{24} = 1 - \frac{1}{2} + \frac{1}{6} - \frac{1}{24} = \frac{15}{24}$$

Ví dụ3f 3. Có hai hộp bi, hộp 1 có 10 bi trong đó có 3 bi đỏ; hộp 2 có 15 bi trong đó có 4 bi đỏ. Lấy ngẫu nhiên mỗi hộp 1 bi. Tính xác suất được 1 bi đỏ.

Giải:

Gọi A là biến cố lấy được 1 bi đỏ.

Gọi A_i là biến cố lấy được bi đỏ ở hộp i, $i = 1,2$.

Khi đó A_1 và A_2 độc lập và $P\left(A_1\right)=3/10$; $P\left(A_2\right)=4/15$.

Ta có: $A=\overline{A}_1A_2+A_1\overline{A}_2$

$$\Rightarrow P(A)\overset{xk}{=}P(\overline{A}_1A_2)+P(A_1\overline{A}_2)$$

$$\overset{\text{đl}}{=}\frac{7}{10}.\frac{4}{15}+\frac{3}{10}.\frac{11}{15}=\frac{61}{150}$$

Ví dụ3f 3. Nam và Phong cùng ăn cơm trưa tại căn tin trường. Cuối bữa ăn, họ thay phiên nhau tung một đồng xu (cân đối) để quyết định xem ai sẽ là người trả tiền bữa ăn theo qui tắc: Nếu ai tung được mặt sấp trước thì người đó phải trả tiền. Giả sử Nam là người tung đồng xu trước. Hãy tính xác suất Phong phải trả tiền.

Giải:

Đặt các biến cố:

A=" Phong phải trả tiền",

N_i= " Nam trả tiền ở lần tung thứ $i, i=2k+1$",

P_i = " Phong trả tiền ở lần tung thứ i, $i=2k$ ".

Thấy ngay các biến cố N_i và các biến cố P_j độc lập và:

$$P(N_i)=P\left(P_j\right)=\frac{1}{2},\ \forall i,j$$

Ta có:

$$P(A)=P\left(\overline{N}_1P_2+\overline{N}_1\overline{P}_2\overline{N}_3P_4+\overline{N}_1\overline{P}_2\overline{N}_3\overline{P}_4\overline{N}_5P_6+\cdots\right)$$

$$\overset{\text{xk,đl}}{=}\frac{1}{2}.\frac{1}{2}+\frac{1}{2}.\frac{1}{2}.\frac{1}{2}.\frac{1}{2}+\frac{1}{2}.\frac{1}{2}.\frac{1}{2}.\frac{1}{2}.\frac{1}{2}.\frac{1}{2}+\cdots$$

$$=\frac{1}{4}+\frac{1}{16}+\frac{1}{64}+\cdots$$

Ta biết rằng tổng các số hạng của dãy cấp số nhân lùi vô hạn với số hạng đầu tiên a_0 và công bội $q\in\left(-1,1\right)$ cho bởi công thức

$$S=a_0\left(1+q+q^2+q^3+\cdots\right)=a_0\frac{1}{1-q}$$

Do đó ta tính được $P(A) = \frac{1}{4} . \frac{1}{1 - \frac{1}{4}} = \frac{1}{3}$.

Nhận xét: theo như ví dụ trên thì nếu người nào tung đồng xu trước thì khả năng không phải trả tiền cho bữa trưa của người đó sẽ cao hơn người tung sau.

Ví dụ 2.14. Trong quá trình xây dựng hệ thống mạng, mất mạng xảy ra vào thứ Hai với xác suất 0,7 và vào thứ Ba với xác suất 0,5. Khi đó sự kiến mất điện vào thứ Hai và thứ Ba có xung khắc nhau không?

(nếu chúng xung khắc thì xác suất mất điện vào 2 ngày này sẽ là 0,7+0,5=1,2>1 là điều vô lý)

VD: Cho các biến cố A,B,C với các xác suất P(A)=0.5, P(B)=0.6, P(C)=0.7, P(A/B)=0.2, P(B+C)=0.85. Hỏi

a. Các biến cố A, B có độc lập nhau không?

b. Các biến cố B,C có độc lập nhau không?

BT: Hai người *A* và *B* cùng chơi trò chơi như sau: Cả hai luân phiên lấy mỗi lần 1 viên bi từ một hộp đựng 2 bi trắng và 4 bi đen (bi được lấy ra không trả lại hộp). Người nào lấy được bi trắng trước thì thắng cuộc. Giả sử *A* lấy trước, tính xác suất *A* thắng cuộc?

3.5.4 Công thức Bernoulli

Phép thử độc lập: Giả sử có các phép thử $\{T_i,\ i=1,2,...,n\}$. Nếu các biến cố của một phép thử độc lập với các biến cố của phép thử còn lại thì ta nói n phép thử $\{T_i,\ i=1,2,...,n\}$ là độc lập.

Ví dụ3f 3. Tung 1 đồng xu n lần ta có n phép thử độc lập.

Công thức Bernoulli: Giả sử có n phép thử độc lập $\{T_i,\ i=1,2,...,n\}$, một biến cố A xảy ra ở mỗi phép thử với xác suất là p không đổi (0< p <1). Khi đó xác suất để biến cố A xuất hiện đúng k lần trong n phép thử độc lập trên được tính theo công thức:

$$P_n(k,p) = C_n^k p^k (1-p)^{n-k} \qquad (3.19)$$

Hay

$$P_n(k,p) = C_n^k p^k q^{n-k},\ q = 1-p \qquad (3.20)$$

Công thức trên gọi là **công thức Bernoulli** (đọc là "Béc-nu-li"), hay còn gọi là **công thức xác suất nhị thức**, còn dãy phép thử trên gọi là **dãy phép thử Bernoulli**.

Trong công thức trên, p còn được gọi là "xác suất thành công", còn $q = 1-p$ được hiểu là "xác suất thất bại".

Chứng minh:

Xác suất để biến cố A xảy ra k lần trong một dãy n phép thử độc lập bằng $p^k(1-p)^{n-k}$.

Vì có C_n^k dãy như vậy nên xác suất để biến cố A xảy ra k lần trong dãy n phép thử độc lập bằng $\mathrm{C}_\mathrm{n}^\mathrm{k}\ C_n^k p^k (1-p)^{n-k}$

Nếu phải tính xác suất để biến cố A xuất hiện từ k đến m lần trong dãy n phép thử như trên thì người ta thường ký hiệu và tính xác suất này như sau:

$$P_n(k...m,p) = \sum_{i=k}^{m} P_n(i;p) = \sum_{i=k}^{m} C_n^i p^i q^{n-i} \text{ , với } q = 1-p \qquad (3.21)$$

Ví dụ3f 3. Một bác sĩ chữa khỏi bệnh D cho một người với xác suất là 95%. Giả sử có 10 người bị bệnh D đến bác sĩ này chữa bệnh, tính xác suất:

a) Có 8 người khỏi bệnh.

b) Có nhiều nhất 9 người khỏi bệnh.

Giải:

a) Gọi A là biến cố có 8 người khỏi bệnh. Áp dụng công thức Béc-nu-li với $n=10$, $p=95\%$, k=8, ta có

$$P\left(A\right)=C_{10}^{8}(95\%)^{8}(5\%)^{2}.$$

b) Gọi B là biến cố có nhiều nhất 9 người khỏi bệnh. Khi đó $\overline{B}$ là biến cố tất cả 10 người đều khỏi bệnh. Theo công thức Bernoulli

$$P\left(\overline{B}\right)=\left(95\%\right)^{10}\approx 0.59$$

Từ đó,

$$P\left(B\right)=1-P\left(\overline{B}\right)\approx 1-0.59=0.41$$

Ví dụ3f 3. (Bài toán tìm n) Xét việc gieo một con súc sắc. Hãy tính xem cần gieo bao nhiêu lần thì xác suất được ít nhất 1 lần xuất hiện mặt sáu chấm không nhỏ hơn 0.9.

Giải:

Đặt $X=$ số lần xuất hiện mặt sáu chấm.

$n=$ số lần gieo súc sắc.

Theo đề bài thì

$$P\left(X\geq 1\right)\geq 0.9$$

$$\Leftrightarrow 1-P\left(X=0\right)\geq 0.9$$

Hay

$$P\left(X=0\right)\leq 0.1$$

Theo công thức Bernoulli:

$$P\left(X=0\right)=C_{n}^{0}.\left(\frac{1}{6}\right)^{0}\left(\frac{5}{6}\right)^{n}=\left(\frac{5}{6}\right)^{n}$$

Nghĩa là ta có

$$\left(\frac{5}{6}\right)^n \leq 0.1$$

Từ đó suy ra

$$n \geq \log_{5/6} 0.1 = \frac{\lg 0.1}{\lg \frac{5}{6}} = 12.63$$

Vậy cần gieo súc sắc ít nhất 13 lần.

Ví dụ3f 3. (Bài toán xác suất có điều kiện) Giả sử một công ty du lịch ở TP. HCM thường nhận các hợp đồng qua fax từ một trong hai nơi: Trong TP. HCM hoặc ngoài TP. HCM. Xác suất công ty nhận một fax trong TP. HCM là 0.75. Một ngày kia, công ty nhận được 10 fax biết có ít nhất 6 fax trong TP.HCM. Hãy tính xác suất chỉ có một fax đến từ ngoài TP.HCM.

Giải: Đặt $X =$ Số fax đến từ ngoài Tp. HCM.

Ta sẽ áp dụng công thức Bernoulli với $n = 10$, $p = 0.25$.

Do biết có ít nhất 6 fax trong Tp. HCM nên đã biết $X \leq 4$, vậy xác suất cần tính là:

$$P\left(X = 1 / X \leq 4\right) = \frac{P\left(X = 1 \cap X \leq 4\right)}{P\left(X \leq 4\right)} = \frac{P\left(X = 1\right)}{P\left(X \leq 4\right)} = \frac{0.1877}{0.9219} = 0.2036$$

Trong đó

$$P\left(X \leq 4\right) = P_n\left(0...4; 0.25\right) = \sum_{i=0}^{4} P_n(i; 0.25)$$

BT: Có 4 sinh viên SV1, SV2, SV3, SV4 cùng làm 1 câu trắc nghiệm có 4 chọn lựa với xác suất làm đúng của mỗi sinh viên lần lượt là 0,9; 0,6; 0,6; 0,7. Biết sinh viên SV1 chọn đáp án A, các sinh viên SV2 và SV3 chọn đáp án B, SV4 chọn đáp án C. Hãy tính xác suất

a). chọn lựa A là đáp án đúng,

b). chọn lựa B là đáp án đúng,

c). chọn lựa C là đáp án đúng.

3.5.5 Công thức xác suất đầy đủ

Giả sử ta có hệ đầy đủ $\{A_i, i = 1, 2, ..., n\}$, tức là

$$\bigcup_{i=1}^{n} A_i = \Omega \text{ và } A_i.A_j = \varnothing, \forall i \neq j$$

Và biết được $P(A_i),\ P(B / A_i),\ i = \overline{1, n}$. Khi đó

$$P(B) = \sum_{i=1}^{n} P(A_i)P(B / A_i) \qquad (3.22)$$

Công thức trên gọi là **công thức xác suất đầy đủ**.

Công thức Bayès:

$$P(A_k / B) = \frac{P(A_k)P(B / A_k)}{\sum_{i=1}^{n} P(A_i)P(B / A_i)} \qquad (3.23)$$

Công thức Bayès còn gọi là công thức xác suất giả thiết.

Chứng minh:

Do $\{A_i, i = 1, 2, ..., n\}$ lập thành một hệ điều đầy đủ nên ta có:

$$B = B \cap \left(\sum_{i=1}^{n} A_i \right) = \sum_{i=1}^{n} A_i B$$

Và $A_i B,\ i = \overline{1, n}$ xung khắc từng đôi nên ta được công thức xác suất đầy đủ:

$$P(B) = P\left(\sum_{i=1}^{n} A_i B \right) \overset{xk}{=} \sum_{i=1}^{n} P(A_i B) = \sum_{i=1}^{n} P(A_i).P(B / A_i)$$

Theo công thức xác suất có điều kiện ta có

$$P(A_k / B) = \frac{P(A_k B)}{P(B)}$$

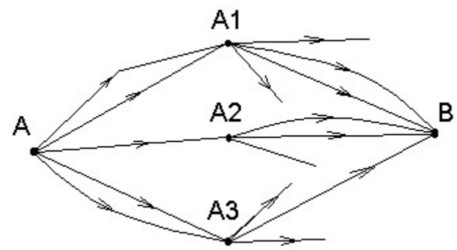
Áp dụng công thức xác suất đầy đủ ở trên và công thức nhân, ta được công thức Bayès:

$$P\left(A_k / B\right) = \frac{P(A_k)P(B / A_k)}{\sum_{i=1}^{n} P(A_i)P(B / A_i)}$$

Ví dụ3f 3. Giả sử một người xuất phát từ A bằng cách chọn ngẫu nhiên các con đường trước mặt theo sơ đồ dưới.

a) Tính xác suất người này đến được B.

b) Giả sử người này đã đến B. Tính xác suất người này đến từ A_1



Hình 3.4

Giải:

a) Nếu gọi B là biến cố người này đến được B, gọi A_i là biến cố người này đến được A_i, $i = 1,2,3$ thì $\{A_1, A_2, A_3\}$ là hệ biến số đầy đủ. Dựa vào sơ đồ hình 3.4 ta có:

$$P\left(A_1\right) = P\left(A_3\right) = \frac{2}{5};\ P\left(A_2\right) = \frac{1}{5};$$

$$P\left(B / A_1\right) = \frac{2}{4};\ \ P\left(B / A_2\right) = \frac{2}{3};\ P\left(B / A_3\right) = \frac{1}{3}$$

Theo công thức xác suất đầy đủ:

$$P\left(B\right) = \sum_{k=1}^{3} P(A_k)P(B / A_k) = \frac{2}{5}.\frac{2}{4} + \frac{1}{5}.\frac{2}{3} + \frac{2}{5}.\frac{1}{3} = \frac{7}{15}$$

b) Theo công thức Bayès, xác suất người này đến được B từ A_1 là

$$P\left(A_1 / B\right) = \frac{P(A_1).P(B / A_1)}{P(B)} = \frac{2/5.2/4}{7/15} = \frac{3}{7}.$$

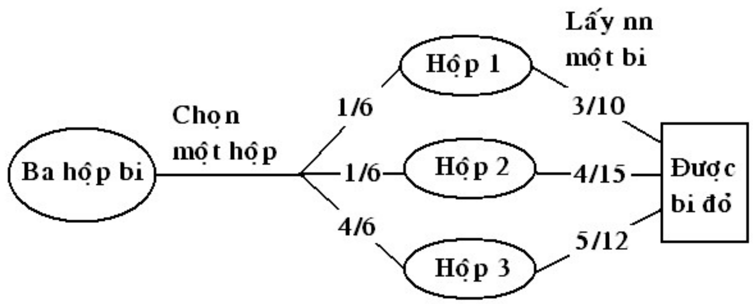
Ví dụ3f 3. Có ba hộp bi. Hộp 1 có 10 bi trong đó có 3 bi đỏ; hộp 2 có 15 bi trong đó có 4 bi đỏ; hộp 3 có 12 bi trong đó có 5 bi đỏ. Gieo một con súc sắc, nếu xuất hiện mặt 1 chấm thì chọn hộp 1, nếu xuất hiện mặt 2 chấm thì chọn hộp 2, nếu xuất hiện các mặt còn lại thì chọn hộp 3. Từ hộp được chọn, lấy ngẫu nhiên một bi.

a) Tính xác suất được bi đỏ.

b) Giả sử lấy được bi đỏ. Tính xác suất để bi đỏ này thuộc hộp 2.

Giải:

a) Ta tóm tắt bài toán bằng sơ đồ sau, trong đó số trên các nhánh chính là xác suất xảy ra các biến cố tiếp theo.



Hình 3.5

Gọi A là biến cố lấy được bi đỏ.

Gọi A_i là biến cố chọn được hộp i, $i = 1,2,3$. Khi đó $\{A_1, A_2, A_3\}$ là hệ đầy đủ, nên theo công thức xác suất đầy đủ thì:

$$P(A) = \sum_{k=1}^{3} P(A_k)P(A/A_k) = \frac{1}{6}.\frac{3}{10} + \frac{1}{6}.\frac{4}{15} + \frac{4}{6}.\frac{5}{12} = \frac{67}{180}$$

b) Xác suất để bi đỏ lấy ra thuộc hộp 2 khi biết đã lấy được bi đỏ là:

$$P(A_2/A) = \frac{P(A_2A)}{P(A)} = \frac{P(A_2)P(A/A_2)}{P(A)} = \frac{\frac{1}{6}\frac{4}{15}}{\frac{67}{180}} = \frac{8}{67}$$

Ví dụ3f 3. Một "test" kiểm tra sự hiện diện của virus A cho kết quả dương tính nếu bệnh nhân thực sự nhiễm virus A. Tuy nhiên, test này cũng có sai sót, đôi khi cho kết quả dương tính đối với người không bị nhiễm virus, tỷ lệ sai sót là 1/20,000. Giả sử cứ 10,000 người thì có 1 người nhiễm virus A. Tìm tỷ lệ người có kết quả dương tính thực sự nhiễm virus A.

Giải:

Gọi A là biến cố người bị bệnh nhiễm virus A, T là biến cố test cho kết quả dương tính.

Khi đó tỷ lệ người có kết quả dương tính thực sự nhiễm virus chính là xác suất $P(A/T)$

Ta có: $P(A) = 0.0001$; $P(T/A) = 1$; $P(T/\overline{A}) = \frac{1}{20,000}$

Theo công thức Bayes:

$$P(A/T) = \frac{P(A).P(T/A)}{P(A).P(T/A) + P(\overline{A}).P(T/\overline{A})}$$

$$= \frac{0.0001 \times 1}{0.0001 \times 1 + 0.9999 \times \dfrac{1}{20,000}} = \frac{20,000}{29,999}$$

BT: Chuồng thỏ I có *3 con thỏ trắng và 4 con thỏ đen*, chuồng II có *5 thỏ trắng và 3 thỏ đen*. Quan sát thấy có 1 con thỏ chạy từ chuồng I sang chuồng II, sau đó có 1 con thỏ chạy ra từ chuồng II. Tính xác suất để con thỏ chạy ra từ chuồng II là thỏ trắng ?

BÀI TẬP

3.1. Tính xác suất khi xếp ngẫu nhiên một bộ sách gồm 5 tập lên giá sách thì nó được xếp đúng thứ tự (từ nhỏ đến lớn, hoặc ngược lại)?

3.2. Gieo đồng thời hai con súc sắc cân đối. Tính xác suất tổng số nốt xuất hiện là 7, biết có ít nhất một con súc sắc xuất hiện mặt năm chấm.

3.3. Gieo đồng thời hai con súc sắc cân đối. Tìm xác suất sao cho:

a) Tổng số chấm xuất hiện là 8.

b) Hiệu số chấm xuất hiện có giá trị tuyệt đối bằng 3.

3.4. Một cái hộp chứa 100 tấm thẻ đánh số từ 1 đến 100. Rút ngẫu nhiên từ hộp ra một tấm thẻ. Tính xác suất số của tấm thẻ rút ra đó không chứa chữ số 5.

3.5. Từ 5 chữ số 1,2,3,4,5 ta lập ngẫu nhiên một số có 3 chữ số. Tính xác suất được một số:

a) Có 3 chữ số khác nhau.

b) Có 3 chữ số giống nhau.

c) Tận cùng bằng chữ số 4.

3.6. Tung 3 con súc sắc phân biệt. Dựa vào số chấm xuất hiện trên mỗi mặt, hãy tính xác suất xuất hiện

a) 3 mặt khác nhau.

b) Mặt năm chấm ở con súc sắc thứ 2.

c) Đúng một mặt năm chấm.

d) 3 mặt chẵn.

3.7. Có 5 người lên 5 toa tàu một cách ngẫu nhiên. Tính xác suất

a) 5 người lên 5 toa khác nhau.

b) A và b cùng lên một toa.

c) Chỉ có hai trong 5 người lên toa đầu tiên.

3.8. Xếp ngẫu nhiên 5 người vào một bàn dài có 5 chỗ. Tính xác suất:

a) Hai người A và B ngồi cạnh nhau.

b) Người A ngồi chính giữa hai người B và C.

3.9. Giả sử có n người ngồi ngẫu nhiên thành một hàng ngang. Tìm xác suất 2 người cho trước luôn luôn ngồi cách nhau r người (giả sử n và r thoả điều kiện để thực hiện được).

3.10. Có 9 tấm thẻ được đánh số từ $0,1,...,8$. Lấy ngẫu nhiên 4 tấm và xếp thành một hàng. Tìm xác suất được một số chẵn.

3.11. Một hộp có 3 bi đỏ và 7 bi đen cùng kích thước. Lấy ngẫu nhiên 2 bi. Tính xác suất được:

a) 2 bi đỏ.

b) 1 bi đỏ và 1 bi đen.

c) Ít nhất 1 bi đen.

3.12. Một lô hàng gồm 22 sản phẩm trong đó có 5 phế phẩm. Lấy ngẫu nhiên từ lô hàng ra 5 sản phẩm. Tìm xác suất trong 5 sản phẩm lấy ra có đúng 2 sản phẩm tốt.

3.13. Một lô hàng có n sản phẩm trong đó có m phế phẩm. Tìm xác suất để khi lấy ngẫu nhiên ra k sản phẩm thì có đúng h phế phẩm?

3.14. Một nhóm xin việc gồm 15 cử nhân mới ra trường, 10 người được chọn ngẫu nhiên. Gọi p là xác suất mà 4 trong số 5 người xin việc có kết quả tốt nghiệp cao nhất được chọn. Hãy tính p.

3.15. Một thủ kho có một chùm chìa khoá gồm 9 chiếc bề ngoài giống hệt nhau, trong đó chỉ có hai chiếc mở được cửa kho. Anh ta thử ngẫu nhiên từng chìa (chìa nào không đúng thì bỏ ra). Tính xác suất anh ta mở được cửa ở lần thử thứ ba.

3.16. Lớp A có 30 sinh viên gồm 20 nam & 10 nữ.

Lớp B có 30 sinh viên gồm 5 nam & 25 nữ.

Gọi mỗi lớp một sinh viên. Dựa vào số nam (hoặc nữ) gọi ra, hãy tính xác suất các trường hợp có thể xảy ra.

3.17. Bắn 3 viên đạn một cách độc lập vào một bia. Xác suất trúng bia của mỗi viên lần lượt là 0.6; 0.9; 0.7. Tìm xác suất:

a) Có đúng một viên trúng đích.

b) Có ít nhất một viên trúng đích.

3.18. Có 3 hộp phấn. Hộp I có 15 viên tốt và 5 viên xấu; Hộp II có 10 viên tốt và 4 viên xấu; Hộp III có 20 viên tốt và 10 viên xấu. Lấy ngẫu nhiên mỗi hộp 1 viên phấn. Tìm xác suất được ít nhất một viên phấn tốt.

3.19. Trong một hộp có 10 phiếu, trong đó có 1 phiếu trúng thưởng. Có 10 người lần lượt rút thăm. Hỏi rút trước hay rút sau có lợi? Tại sao?

3.20. Tung một con súc sắc 3 lần. Tìm xác suất mặt 6 chấm xuất hiện 2 lần.

3.21. Một nữ công nhân quản lý 5 máy dệt. Xác suất một máy dệt trong khoảng thời gian t cần đến sự chăm sóc của nữ công nhân bằng 1/3. Tính xác suất trong khoảng thời gian t:

a) Có 2 máy cần đến sự chăm sóc.

b) Số máy cần đến sự chăm sóc không bé hơn 2 và không lớn hơn 3.

3.22. Giả sử một người phải làm 10 câu hỏi của một bài kiểm tra trắc nghiệm và các lần chọn hoàn toàn độc lập nhau. Giả sử người này không học bài và chọn ngẫu nhiên các chọn lựa, tìm xác suất người đó chọn đúng 5 câu trả lời, biết mỗi câu hỏi có 4 chọn lựa trong đó có một chọn lựa đúng nhất.

3.23. Một cầu thủ nổi tiếng về đá phạt đền 11 mét. Xác suất thực hiện thành công quả sút 11 mét của cầu thủ này là 90%. Có người nói rằng nếu cầu thủ này thực hiện 10 cú "sút" thì chắc chắn có 9 quả vào gôn. Khẳng định đó đúng không? Tại sao? Hãy tính xác suất để khi cầu thủ đó sút 10 quả thì có 9 quả vào gôn?

3.24. Có hai lô hàng: lô 1 và lô 2 có số sản phẩm lần lượt là 10 và 8 sản phẩm, trong đó mỗi lô có một phế phẩm. Lấy ra 1 sản phẩm từ lô thứ nhất bỏ vào lô thứ hai sau đó từ lô này lấy ra 1 sản phẩm. Tính xác suất sản phẩm lấy ra sau cùng là phế phẩm.

3.25. Một trạm tín hiệu chỉ phát hai loại tín hiệu A và B với xác suất tương ứng là 0,8 và 0,2. Do có nhiễu trên đường truyền nên 1/6 tín hiệu A bị méo và thu được như tín hiệu B, còn 1/8 tín hiệu B bị méo thành tín hiệu A.

a) Tìm xác suất thu được tín hiệu A.

b) Giả sử thu được tín hiệu A, tìm xác suất thu được đúng tín hiệu lúc phát.

3.26. Một hộp đựng 10 bi trong đó có 6 bi trắng, một hộp khác chứa 20 bi trong đó có 4 bi trắng. Từ mỗi hộp rút ngẫu nhiên một bi. Sau đó, trong 2 bi thu được lấy ngẫu nhiên 1 bi. Tìm xác suất bi lấy ra sau cùng là bi trắng.

3.27. Một xí nghiệp với 2 phân xưởng với các tỷ lệ phế phẩm tương ứng là 1% và 2%. Biết rằng phân xưởng I sản xuất 40%, còn phân xưởng II sản xuất 60% sản phẩm.

a) Chọn ngẫu nhiên một sản phẩm từ kho của xí nghiệp. Tìm xác suất được một phế phẩm. Bạn có nhận xét gì về xác suất này?

b) Giả sử lấy được một phế phẩm, tìm xác suất để nó do phân xưởng I sản xuất ra.

c) Giả sử lấy được một phế phẩm, theo bạn thì phế phẩm này có thể do phân xưởng nào sản xuất.

3.28. Có 3 xạ thủ cùng bắn vào một con thú, mỗi người bắn một viên đạn, với xác suất bắn trúng lần lượt là 0.6; 0.7; 0.8. Biết rằng nếu bị trúng một phát đạn thì xác suất để con thú bị tiêu diệt là 0.5; bị trúng 2 phát đạn thì xác suất để con thú bị tiêu diệt là 0.8, còn nếu bị trúng 3 phát đạn thì chắc chắn con thú bị tiêu diệt.

a) Tính xác suất con thú bị tiêu diệt.

b) Hãy tính xác suất con thú bị tiêu diệt do bị trúng 2 phát đạn.

3.29. Điền các giá trị phù hợp vào ô trống

	$P(A)$	$P(B)$	$P(A\cup B)$	$P(A\cap B)$	$P(A/B)$	$P(B/A)$
a)	$\frac{3}{4}$		$\frac{9}{10}$	$\frac{1}{5}$		
b)				$\frac{2}{3}$	$\frac{8}{9}$	$\frac{4}{5}$
c)	$\frac{1}{5}$	$\frac{1}{2}$				$\frac{1}{20}$
d)	$\frac{5}{17}$	$\frac{3}{17}$		$\frac{1}{17}$		

Chương 4
BIẾN SỐ NGẪU NHIÊN

Chương 4 cung cấp cho sinh viên các khái niệm và các tính chất về biến số ngẫu nhiên cùng với các đặc trưng và một số luật phân phối thông dụng của nó.

Sau khi học chương này, sinh viên phải lập được luật phân phối xác suất của biến ngẫu nhiên rời rạc, tính được các đặc trưng và giải thích được ý nghĩa của chúng, vận dụng được các luật phân phối xác suất: Nhị thức, Poisson, Chuẩn để tính xác suất.

4.1. KHÁI NIỆM

Một biến số được gọi là **biến số ngẫu nhiên** (BSNN) nếu biến số đó nhận một trong các giá trị có thể có của nó tuỳ thuộc vào kết quả của một phép thử tương ứng.

Ta còn nói tắt BSNN là *biến ngẫu nhiên* (BNN). BSNN còn được gọi là **đại lượng ngẫu nhiên** (ĐLNN).

Ví dụ3f 4. Gọi X là số chấm xuất hiện khi gieo một con súc sắc. Rõ ràng là X có thể nhận các giá trị: 1, 2,.., 6, nhưng X nhận giá trị nào thì còn tùy thuộc vào việc gieo súc sắc. Vậy X là một BSNN.

Nhận xét: Trong ví dụ trên, kí hiệu ($X=k$) chỉ biến cố ngẫu nhiên "số chấm xuất hiện khi gieo một con súc sắc là k" với xác suất là

$$P\left[X=k\right]=\frac{1}{6},\ k=1,2,\ldots,6$$

Biến số ngẫu nhiên X gọi là **biến ngẫu nhiên rời rạc** nếu tập giá trị của X *đếm được*; còn nếu tập giá trị lấp đầy một khoảng nào đó trên trục số thì ta có **biến ngẫu nhiên liên tục**.

Ví dụ3f 4.

- Thấy ngay biến ngẫu nhiên X trong ví dụ 4.1 là biến rời rạc.

• Nếu biến số ngẫu nhiên X là số sinh viên đi học tại một thời điểm nào đó trong ngày thì ta có X là BNN rời rạc.

• Nếu biến số ngẫu nhiên X là nhiệt độ tại một thời điểm trong ngày thì ta có BSNN liên tục, vì nếu nhìn mức thuỷ ngân trong nhiệt kế, để tăng nhiệt độ từ a^0c đến b^0c thì nhiệt độ phải thay đổi qua tất cả các giá trị nằm giữa a^0 và b^0

Vấn đề đặt ra là X nhận một giá trị nào đó, khi X là biến rời rạc, hoặc nhận giá trị trong một khoảng nào đó, khi X là biến liên tục, với xác suất bằng bao nhiêu, nghĩa là ta cần xác định **luật phân phối xác suất**, hay phân bố xác suất, của X. Luật phân phối xác suất còn được gọi tắt là **luật phân phối** của X **.**

Ghi chú: Từ đây về sau, ta sẽ kí hiệu biến ngẫu nhiên bằng các chữ in ($X, Y, ...$), còn giá trị mà biến ngẫu nhiên có thể nhận được kí hiệu bằng các chữ thường ($x, y, ...$) khi không chỉ rõ giá trị.

4.2. LUẬT PHÂN PHỐI XÁC SUẤT CỦA BIẾN NGẪU NHIÊN

4.2.1. Bảng phân phối xác suất: (chỉ dùng cho biến rời rạc) là bảng có dạng:

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

Bảng 4.1.

Với x_i, $i = \overline{1,n}$ là các giá trị có thể của X, được xếp tăng dần:

$$x_1 < x_2 < ... < x_n,$$

p_i: là xác suất X nhận giá trị x_i, $i = \overline{1,n}$, $p_i = P(X = x_i)$, thoả điều kiện:

$$\sum_{i=1}^{n} p_i = 1.$$

Ghi chú:

- Trong bảng trên, những giá trị nào của X không được liệt kê tức là X nhận giá trị đó với xác suất bằng 0.

- Tập các biến cố $\left[X = x_i\right],\ i = \overline{1,n}$ lập thành một hệ đầy đủ.

Ví dụ3f 4. Luật phân phối xác suất của X là số chấm xuất hiện khi gieo một con súc sắc là:

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

Ví dụ3f 4. Một cơ quan có 3 xe ôtô: 1 xe 4 chỗ, 1 xe 50 chỗ và 1 xe tải. Xác suất để trong một ngày làm việc các xe được sử dụng lần lượt là 0.8; 0.4; 0.9. Hãy lập luật phân phối xác suất cho số loại xe được sử dụng trong một ngày của cơ quan. Ghi chú: Loại xe này phân biệt với loại xe kia bởi số người tối đa có thể chở trên xe.

Giải:

Gọi X là số loại xe được sử dụng trong một ngày của cơ quan. Ta thấy X có thể nhận các giá trị: 0, 1, 2, 3.

Gọi $A_1,\ A_2,\ A_3$ lần lượt là các sự kiện xe 4 chỗ, xe 50 chỗ, xe tải được sử dụng trong một ngày. Khi đó các biến cố $A_1,\ A_2,\ A_3$ độc lập và $P\left(A_1\right) = 0.8;\ P\left(A_2\right) = 0.4;\ P\left(A_3\right) = 0.9$

Xác suất X nhận các giá trị 0, 1, 2, 3 được tính như sau:

$$P\left[X = 0\right] = P(\overline{A}_1\overline{A}_2\overline{A}_3) \overset{\text{đl}}{=} P(\overline{A}_1)P(\overline{A}_2)P(\overline{A}_3)$$

$$= (0.2)(0.6)(0.1) = 0.012$$

$$P\left[X = 1\right] = P(A_1\overline{A}_2\overline{A}_3 + \overline{A}_1A_2\overline{A}_3 + \overline{A}_1\overline{A}_2A_3)$$

$$\overset{xk}{=} P(A_1\overline{A}_2\overline{A}_3) + P(\overline{A}_1A_2\overline{A}_3) + P(\overline{A}_1\overline{A}_2A_3)$$

$$\overset{\text{đl}}{=} P(A_1)P(\overline{A}_2)P(\overline{A}_3) + P(\overline{A}_1)P(A_2)P(\overline{A}_3) + P(\overline{A}_1)P(\overline{A}_2)P(A_3)$$

$$= (0.8)(0.6)(0.1)+(0.2)(0.4)(0.1)+(0.2)(0.6)(0.9)$$

$$= 0.164$$

$$P\left[X = 2\right] = P(A_1A_2\overline{A}_3 + A_1\overline{A}_2A_3 + \overline{A}_1A_2A_3)$$

$$\overset{xk,\text{đl}}{=} (0.8)(0.4)(0.1)+(0.8)(0.6)(0.9)+(0.2)(0.4)(0.9)$$

$$= 0.536$$

$$P[X=3] = P(A_1A_2A_3)$$

$$\overset{\text{đl}}{=} (0.8)(0.4)(0.9)$$

$$=0.288.$$

Vậy luật phân phối xác suất của X là

X	0	1	2	3
P	0.012	0.164	0.536	0.288

Thỏa

$$\sum_{i=1}^{4} p_i = 0.012 + 0.164 + 0.536 + 0.288 = 1$$

4.2.2. Hàm xác suất

Nếu X có phân phối xác suất rời rạc như bảng 4.1 thì hàm xác suất của biến ngẫu nhiên rời rạc X, kí hiệu là $p(x)$, được xác định như sau:

$$p(x) = \begin{cases} p_1 & , x = x_1 \\ p_2 & , x = x_2 \\ \dots & \\ p_n & , x = x_n \\ 0 & , \text{nơi khác} \end{cases} \qquad (4.1)$$

Ví dụ3f 4. Cho X có luật phân phối xác suất như trong ví dụ 4.3. Khi đó, hàm xác suất của X là:

$$p(x) = \begin{cases} 0.012 & , x = 0 \\ 0.164 & , x = 1 \\ 0.536 & , x = 2 \\ 0.288 & , x = 3 \\ 0 & , \text{nơi khác} \end{cases}$$

4.2.3. Hàm phân phối

a. Định nghĩa**:** Hàm phân phối xác suất của biến ngẫu nhiên X kí hiệu là $F\left(x\right)$ và được định nghĩa như sau:

$$F\left(x\right)=P\left[X<x\right],\ \forall x\in\mathbb{R}. \tag{4.2}$$

$F\left(x\right)$ còn được gọi là *hàm xác suất tích lũy* của X.

Ví dụ3f 4. Cho X có luật phân phối xác suất như sau:

X	0	1	2	3
P_x	0.012	0.164	0.536	0.288

Khi đó, hàm phân phối của X là:

$$F(x)=P[X<x]=\begin{cases} 0 & ,\ x\le 0 \\ 0.012 & ,\ 0<x\le 1 \\ 0.176 & ,1<x\le 2 \\ 0.712 & ,\ 2<x\le 3 \\ 1 & ,\ 3<x \end{cases}$$

***b. Tính chất cơ bản**:*

(i) $F\left(x\right)\in\left[0,1\right]$, $F\left(-\infty\right)=0$, $F\left(+\infty\right)=1$.

(ii) $F\left(x\right)$ không giảm trên $\mathbb{R}$.

(iii) $F\left(x\right)$ liên tục trái tại mọi $x\in\mathbb{R}$. Trong trường hợp X là biến ngẫu nhiên liên tục thì $F\left(x\right)$ liên tục trên $\mathbb{R}$.

(iv) $P\left(a\le X<b\right)=F\left(b\right)-F\left(a\right)$. (4.3)

Ví dụ3f 4. Cho X là biến ngẫu nhiên liên tục có hàm phân phối

$$F(x)=\begin{cases} 0 & ,x\le -1 \\ a\left(b+x-\dfrac{x^3}{3}\right) & ,\ -1<x\le 1. \\ 1 & ,x>1 \end{cases}$$

Hãy tìm các hằng số a và b.

Giải:

Do X là biến ngẫu nhiên liên tục nên $F(x)$ liên tục trên $\mathbb{R}$

Tại $x\notin\{-1,1\}$, $F(x)$ là hàm sơ cấp nên $F(x)$ liên tục

$\Rightarrow F(x)$ liên tục trên $\mathbb{R}$ $\Leftrightarrow F(x)$ liên tục tại $x=\pm 1$

$$\Leftrightarrow \begin{cases} \lim\limits_{x\to -1-} F(x)=\lim\limits_{x\to -1+} F(x)=F(-1) \\ \lim\limits_{x\to 1-} F(x)=\lim\limits_{x\to 1+} F(x)=F(1) \end{cases} \Leftrightarrow \begin{cases} a\left(b-\dfrac{2}{3}\right)=0 \\ a\left(b+\dfrac{2}{3}\right)=1 \end{cases} \Leftrightarrow \begin{cases} b=\dfrac{2}{3} \\ a=\dfrac{3}{4} \end{cases}$$

Vậy, ta có:

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

4.2.4. Hàm mật độ

a. Định nghĩa:

Cho X là biến ngẫu nhiên liên tục có hàm phân phối $F(x)$ khả vi, khi đó hàm mật độ xác suất của X, kí hiệu là $f(x)$ (hay còn kí hiệu là $p(x)$), được xác định như sau

$$f(x)=F'(x),\ x\in\mathbb{R} \text{ (Có thể trừ ra một số điểm)} \qquad (4.4)$$

b. Một số tính chất cơ bản:

(i) $f(x)\ge 0$, $\forall x\in\mathbb{R}$

(ii) $F(x) = \int\limits_{-\infty}^{x} f(t)dt$ (4.5)

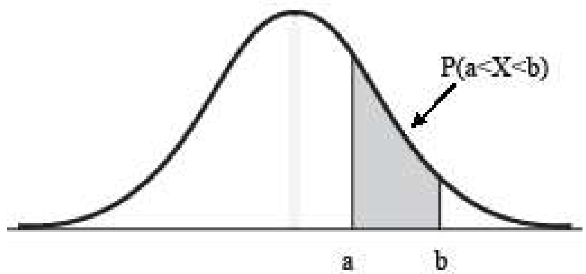
(iii) $\int\limits_{-\infty}^{+\infty} f(x)dx = 1$ (4.6)

(iv) $P(a \le X \le b) = P(a \le X < b)$

$= P(a < X \le b)$

$= P(a < X < b)$

$= \int\limits_{a}^{b} f(x)dx$ (4.7)



Hình 4.1

(v) $P(X = a) = 0$ nếu X là biến liên tục

Nhận xét: từ tính chất (v), ta thấy rằng một biến cố có xác suất bằng 0 chưa hẳn là biến cố không thể.

c. Ý nghĩa: Hàm mật độ của biến X, mà cụ thể là đồ thị của nó, cho hình ảnh của sự tập trung xác suất của X trên từng khoảng giá trị (xem hình 4.1).

Ví dụ3f 4. Cho X là biến ngẫu nhiên liên tục có hàm mật độ xác suất là:

$$f(x) = \begin{cases} 0 & , x \le 0 \\ x & , 0 < x \le 1 \\ 2 - x, & 1 < x \le 2 \\ 0 & , 2 < x \end{cases}$$

a) Viết hàm phân phối xác suất của X.

b) Tính $P\left[X < \frac{1}{2}\right]$.

Giải:

a) $$F(x) = P[X < x] = \begin{cases} 0 & , x \le 0 \\ \int_0^x t\,dt & , 0 < x \le 1 \\ \int_0^1 t\,dt + \int_1^x \left(2 - t\right) dt, & 1 < x \le 2 \\ 1 & , 2 < x \end{cases}$$

$$= \begin{cases} 0 & , x \le 0 \\ \frac{1}{2}x^2 & , 0 < x \le 1 \\ -\frac{1}{2}x^2 + 2x - 1, & 1 < x \le 2 \\ 1 & , 2 < x \end{cases}$$

b) $P\left[X < \frac{1}{2}\right] = \int_{-\infty}^{1/2} f(x)dx = F\left(\frac{1}{2}\right) = \frac{1}{2}\left(\frac{1}{2}\right)^2 = \frac{1}{8}$

4.3. CÁC ĐẶC TRƯNG SỐ CỦA BSNN

4.3.1. Kỳ vọng toán

a. Định nghĩa*:* Kỳ vọng toán của BSNN X được kí hiệu là $E\left(X\right)$ và được xác định như sau:

$$E\left(X\right) = \begin{cases} \sum x_i p_i \quad \text{nếu } X \text{ rời rạc} \\ \int_{-\infty}^{+\infty} x.f(x)dx \text{ nếu } X \text{ liên tục} \end{cases} \qquad (4.8)$$

nếu vế phải hội tụ tuyệt đối. Ngược lại, ta nói BSNN X không có kỳ vọng $E(X)$.

Ví dụ3f 4. Một hộp có 3 bi nặng 10g, 5 bi nặng 50g và 2 bi nặng 20g. Chọn ngẫu nhiên 1 bi và gọi X là trọng lượng bi đó. Tính $E(X)$

Giải:

Dễ thấy tập giá trị của X là: $\{10,\ 20,\ 50\}$.

Theo công thức xác suất cổ điển ta có:

$$P[X=10]=\frac{3}{10}\ ;\ P[X=20]=\frac{2}{10}\ ;\ P[X=50]=\frac{5}{10}$$

Nên:

$$\begin{aligned} E(X) &= 10.\frac{3}{10}+20.\frac{2}{10}+50.\frac{5}{10} \\ &= \frac{1}{10}(3\times 10+2\times 20+5\times 50)=32\,(g) \end{aligned}$$

Nhận xét: Đây chính là trọng lượng trung bình của các bi trong hộp.

***b. Ý nghĩa**:* $E(X)$ là giá trị trung bình (về mặt xác suất) của biến số ngẫu nhiên X khi thực hiện phép thử tương ứng.

Ví dụ3f 4. Có một trò chơi như sau: Một hộp có 5 bi, trong đó có 4 bi đỏ và 1 bi đen. Hai người A và B lần lượt lấy ngẫu nhiên một viên bi không hoàn lại cho đến khi người nào lấy trúng bi đen thì ván chơi kết thúc và người lấy trúng bi đen phải trả cho người kia một số tiền bằng với số bi lấy ra nhân với 1000 đồng.

Giả sử người A lấy bi trước.

a) Hãy tính số tiền trung bình người A thu về sau một ván chơi. Nhận xét về tính công bằng của trò chơi này?

b) Nếu muốn người A được lợi trong trò chơi này thì theo bạn ta cần đđiều chỉnh giả thiết nào trong đề bài và điều chỉnh như thế nào?

Giải:

Nhận xét: nếu gọi X là số tiền người A thu về sau một ván thì đề bài yêu cầu ta tính $E(X)$

Vì $E(X)$ là số tiền mà người A thu về sau một ván chơi nên nếu $E(X)<0$ thì người A bị thiệt, nếu $E(X)=0$ thì người A huề vốn, còn nếu $E(X)>0$ thì người này được lợi từ trò chơi này.

a) Gọi A_i là sự kiện lấy bi lần thứ i trúng bi đen, $i=1,2,3,4,5$.

Các trường hợp có thể xảy ra đối với người A trong một ván là:

A_1: Nghĩa là người A lấy bi lần đầu tiên và trúng ngay bi đen. Người A phải trả cho người B 1000 đồng.

$\overline{A}_1A_2$: Nghĩa là người A lấy bi lần đầu tiên và không lấy phải bi đen. Đến lượt người B lấy bi thì trúng bi đen. Người B phải trả cho người A số tiền là 2(bi)$\times$1000 (đđồng)=2000 đđồng.

$\overline{A}_1\overline{A}_2A_3$: Giải thích, người A phải trả cho người B 3000 đđồng.

$\overline{A}_1\overline{A}_2\overline{A}_3A_4$: Người B phải trả cho người A 4000 đđồng.

$\overline{A}_1\overline{A}_2\overline{A}_3\overline{A}_4A_5$: Người A phải trả cho người B 5000 đđồng.

Ứng với các trường hợp trên, X sẽ nhận các giá trị: –1000, +2000, –3000, +4000, –5000 với các xác suất tương ứng là:

$$P[X=-1000]=P(A_1)=\frac{1}{5}.$$

$$P[X=2000]=P(\overline{A}_1A_2)=P(\overline{A}_1).P(A_2/\overline{A}_1)=\frac{4}{5}.\frac{1}{4}=\frac{1}{5}$$

$$P[X=-3000]=P(\overline{A}_1\overline{A}_2A_3)=P(\overline{A}_1).P(\overline{A}_2/\overline{A}_1).P(A_3/\overline{A}_1\overline{A}_2)$$

$$=\frac{4}{5}.\frac{3}{4}.\frac{1}{3}=\frac{1}{5}$$

$$P[X=4000]=P(\overline{A}_1\overline{A}_2\overline{A}_3A_4)$$

$$=P(\overline{A}_1).P(\overline{A}_2/\overline{A}_1).P(\overline{A}_3/\overline{A}_1\overline{A}_2).P(A_4/\overline{A}_1\overline{A}_2\overline{A}_3)$$

$$=\frac{4}{5}.\frac{3}{4}.\frac{2}{3}.\frac{1}{2}=\frac{1}{5}$$

$$P[X=-5000]=P\left(\overline{A}_1\overline{A}_2\overline{A}_3\overline{A}_4A_5\right)=P\left(\overline{A}_1\overline{A}_2\overline{A}_3\overline{A}_4\right)$$

$$= P(\overline{A}_1).P(\overline{A}_2 / \overline{A}_1).P(\overline{A}_3 / \overline{A}_1\overline{A}_2).P(\overline{A}_4 / \overline{A}_1\overline{A}_2\overline{A}_3)$$
$$= \frac{4}{5}.\frac{3}{4}.\frac{2}{3}.\frac{1}{2} = \frac{1}{5}$$

Vậy luật phân phối xác suất của X là

X	-5000	-3000	-1000	2000	4000
P	1/5	1/5	1/5	1/5	1/5

Số tiền trung bình người A thu về sau 1 ván là:

$$E(X) = \frac{1}{5}\left[(-5000) + (-3000) + (-1000) + 2000 + 4000\right]$$

$$= -\frac{3000}{5} \text{ (đồng)}$$

Do $E(X) < 0$ nên ta kết luận trò chơi không công bằng; người A bị thiệt khi tham gia.

b) Muốn người A được lợi trong trò chơi này thì ta chỉ cần thay đổi điều kiện sao cho số tiền trung bình người A thu về sau một ván chơi dương, chẳng hạn như thay đổi thứ tự rút bi của người A và người B.

c. Tính chất

(i) $E(c) = c$, với c là hằng số.

(ii) $E(c.X) = c.E(X)$

(iii) $E(X+Y) = E(X) + E(Y)$

(iv) $E(X.Y) = E(X).E(Y)$ nếu như X và Y độc lập.

(X và Y độc lập khi X và Y nhận các giá trị độc lập nhau)

Ngoài ra ta còn có kết quả sau: Nếu $Y = g(X)$ thì

* $E(Y) = \sum_i g(x_i)P[X = x_i]$ nếu X là BSNN rời rạc

* $E(Y) = \int_{-\infty}^{+\infty} g(x)f(x)dx$ nếu X là BSNN liên tục có hàm mật độ là $f(x)$.

4.3.2. Phương sai

***a. Định nghĩa**:* Phương sai của BSNN ngẫu nhiên X được kí hiệu là $D(X)$ và được định nghĩa như sau:

$$D(X) = E\left[X - E(X)\right]^2 \tag{4.9}$$

Thấy ngay phương sai của BNN X có đơn vị đo là đơn vị của X bình phương. Chẳng hạn, nếu X có đơn vị đo là m thì $D(X)$ có đơn vị đo là m^2.

Để có một đại lượng có cùng đơn vị đo với X, người ta đưa vào khái niệm **Độ lệch chuẩn** $\sigma(X)$ được xác định như sau:

$$\sigma(X) = \sqrt{D(X)} \tag{4.10}$$

Trong thực hành, ta thường dùng công thức sau để tính phương sai $D(X)$:

$$D(X) = E\left(X^2\right) - E^2(X) \tag{4.11}$$

với

$$E(X^2) = \begin{cases} \sum\limits_i x_i^2 p_i & \text{, nếu } X \text{ rời rạc} \\ \int\limits_{-\infty}^{+\infty} x^2 f(x)dx, & \text{nếu } X \text{ liên tục} \end{cases} \tag{4.12}$$

Thật vậy,

$$\begin{aligned} D(X) &= E\left(X^2\right) - E^2(X) \\ &= E\left[X^2 - 2X.E(X) + E^2(X)\right] \\ &= E\left(X^2\right) - 2.E(X).E(X) + E^2(X) \end{aligned}$$

(do các tính chất của kỳ vọng)

$$= E\left(X^2\right) - E^2(X)$$

Ví dụ3f 4. Cho X là BNN có luật phân phối:

X	10	20	50
P	3/10	2/10	5/10

Tính $D(X)$.

Giải: Ta có

$$E(X) = 10.\frac{3}{10} + 20.\frac{2}{10} + 50.\frac{5}{10} = 32.$$

$$E(X^2) = 10^2.\frac{3}{10} + 20^2.\frac{2}{10} + 50^2.\frac{5}{10} = 1360.$$

$$D(X) = E(X^2) - E^2(X) = 1360 - 32^2 = 336.$$

Ví dụ3f 4. Cho biến ngẫu nhiên X có hàm mật độ xác suất:

$$f(x) = \begin{cases} \frac{3}{4}\left(1-x^2\right) & \text{, nếu } |x| \le 1 \\ 0 & \text{, nếu } |x| > 1 \end{cases}$$

Tính $E(X)$, $D(X)$.

Giải:

$$E(X) = \int_{-\infty}^{+\infty} xf(x)dx = \int_{-1}^{1} x.\frac{3}{4}\left(1-x^2\right)dx = 0$$

(Tích phân trên miền đối xứng của hàm lẻ)

$$E(X^2) = \int_{-\infty}^{+\infty} x^2 f(x)dx$$

$$= \int_{-1}^{1} x^2\frac{3}{4}\left(1-x^2\right)dx = 2.\frac{3}{4}\left(\frac{x^3}{3} - \frac{x^5}{5}\right)\Bigg|_0^1 = \frac{1}{5}$$

Vậy $D(X) = E(X^2) - E^2(X) = \frac{1}{5}$

Nhận xét: Từ công thức định nghĩa phương sai ta thấy: Nếu X nhận các giá trị càng xa giá trị kỳ vọng thì $\left[X - E(X)\right]^2$ càng lớn và như vậy $D(X)$ càng lớn, từ đó rút ra ý nghĩa của phương sai như sau:

***b. Ý nghĩa**:* $D(X)$ là thông số đo độ phân tán các giá trị mà X có thể nhận xung quanh giá trị kỳ vọng.

Thấy ngay $\sigma(X)$ cũng là thông số đo độ phân tán nhưng có cùng đơn vị đo với X.

c. Tính chất

(i) $D(X) \geq 0$.

(ii) $D(c) = 0$. (c là hằng số)

(iii) $D(cX) = c^2 D(X)$.

(iv) $D(X+Y) = D(X) + D(Y)$ khi X và Y độc lập.

Các tính chất trên có được qua công thức định nghĩa và tính chất của kỳ vọng.

4.3.3. Mod(X)

Mod(X) là giá trị của biến ngẫu nhiên X ứng với xác suất lớn nhất đối với biến ngẫu nhiên rời rạc hoặc ứng với giá trị làm cho hàm mật độ cực đại đối với biến ngẫu nhiên liên tục. Mod(X) còn gọi là **giá trị tin chắc nhất**, hay **giá trị có khả năng nhất**, của X .

Chú ý: Một biến ngẫu nhiên có thể có một hoặc nhiều mode.

Ví dụ3f 4. Nếu X là điểm thi môn xác suất thống kê của các sinh viên trong trường thì Mod(X) là điểm mà có nhiều sinh viên nhận được nhất.

4.3.4. Med(X)

Med(X) (Median) là giá trị của biến ngẫu nhiên X chia phân phối thành hai phần có xác suất bằng nhau và bằng $\frac{1}{2}$. Med(X) còn gọi là *điểm trung vị*.

Ta có: $P[X < Med(X)] = \frac{1}{2}$

Hoặc $P[X \leq Med(X)] = \frac{1}{2}$

Trong ứng dụng, Med(X) là đặc trưng vị trí rất tốt, nhất là khi trong số liệu có nhiều sai sót.

Ví dụ3f 4.

Cho biến ngẫu nhiên X có luật phân phối như sau

a)

X	1	2	3	4	5	6

P	1/6	1/6	1/6	1/6	1/6	1/6

Mod(X) là giá trị nào cũng được trong các giá trị 1, 2,.., 6.

Med(X) =3, hoặc 4.

b).

X	0	1	2	3
P	0,012	0,164	0,536	0,288

Mod(X)=2.

Không có Med(X).

Ví dụ3f 4. Cho biến ngẫu nhiên liên tục X có hàm mật độ xác suất là:

$$f(x) = \begin{cases} A.x.(3-x) & , \text{ nếu } x \in [0,3] \\ 0 & , \text{ nếu } x \notin [0,3] \end{cases}$$

a) Hãy tìm hằng số A.

b) Với hằng số A tìm được: Tính $P[X>2]$, $E(X)$, $D(X)$, Mod(X), Med(X) (nếu có).

c) Viết hàm phân phối xác suất của X.

Giải:

a) Ta thấy $f(x) \geq 0$ khi $A \geq 0$, còn

$$\int_{-\infty}^{+\infty} f(x)dx = \int_{-\infty}^{0} f(x)dx + \int_{0}^{3} f(x)dx + \int_{3}^{+\infty} f(x)dx$$

$$= \int_{0}^{3} Ax(3-x)dx = \frac{9}{2}A$$

$$\Rightarrow \int_{-\infty}^{+\infty} f(x)dx = 1 \Leftrightarrow A = \frac{2}{9}$$

b) Tính $P[X>2]$: $P[X>2] = \int_{2}^{3} \frac{2}{9}x(3-x)dx = \frac{7}{27}$

Tính $E(X)$: $E(X) = \int_0^3 \frac{2}{9}x^2(3-x)dx = \frac{3}{2}$

Tính $D(X)$: Ta có $E(X^2) = \int_0^3 \frac{2}{9}x^3(3-x)dx = \frac{27}{10}$

Nên $D(X) = E(X^2) - E^2(X) = \frac{27}{10} - \left(\frac{3}{2}\right)^2 = \frac{9}{20}$

Tìm Mod(X): Lấy đạo hàm của $f(x)$ trên khoảng (0,3)

$\Rightarrow f'(x) = \frac{2}{3} - \frac{4}{9}x.$

Lập bảng xét dấu suy ra $f(x)$ đạt cực đại tại $x = \frac{3}{2}$.

$\Rightarrow$ Mod(X)=$\frac{3}{2}$.

Tìm Med(X):

Đặt a= Med(X). Điều kiện: $a \in [0,3]$ vì nếu không a sẽ chia phân phối thành hai phần trong đó có một phần có xác suất bằng 1 (và phần còn lại có xác suất bằng 0).

Xét phương trình

$$\int_{-\infty}^{a} f(x)dx = \frac{1}{2},\ a \in [0;3]$$

$$\Leftrightarrow \int_0^a \frac{2}{9}x(3-x)dx = \frac{1}{2}$$

$$\Leftrightarrow -\frac{2}{27}a^3 + \frac{1}{3}a^2 = \frac{1}{2}$$

$$\Leftrightarrow a = \frac{3}{2},\ a = \frac{3}{2} + \frac{3}{2}\sqrt{3},\ a = \frac{3}{2} - \frac{3}{2}\sqrt{3}$$

So với điều kiện, ta chỉ nhận a=Med(X) = $\frac{3}{2}$

c) Hàm phân phối của X là:

$$F(x) = \begin{cases} 0 & \text{, nếu } x < 0 \\ -\dfrac{2}{27}x^3 + \dfrac{1}{3}x^2 & \text{, nếu } x \in [0,3] \\ 1 & \text{, nếu } x > 3 \end{cases}$$

4.4. MỘT SỐ LUẬT PHÂN PHỐI XÁC SUẤT THÔNG DỤNG

4.4.1. Luật phân phối của BSNN rời rạc

4.4.1.1. Phân phối nhị thức

***a. Định nghĩa*: BNN X gọi là có phân phối nhị thức với hai tham số là n và p, kí hiệu là $X \sim B(n,p)$, nếu như :

$$P\left[X = i\right] = C_n^i p^i (1-p)^{n-i} \text{ , } i = 0,1,2,...,n. \qquad (4.13)$$

Ví dụ3f 4. Một xạ thủ bắn 3 viên đạn vào một mục tiêu với xác suất bắn trúng mục tiêu của mỗi lần bắn là 0.8. Nếu gọi X là số đạn trúng mục tiêu thì $X \sim B(3,\ 0.8)$ vì xác suất của biến cố $\left[X = i\right]$ được tính theo công thức Bernoulli

b. Tính chất cơ bản*:*

Nếu $X \sim B(n,p)$ thì

$$E(X) = np \qquad (4.14)$$

$$D(X) = npq \qquad (4.15)$$

Trong ví dụ 4.16, số đạn trung bình bắn trúng mục tiêu là:

$$E(X) = 3 \times 0.8 = 2.4$$

4.4.1.2. Phân phối Poisson

Phân phối Poisson thường được biết như phân phối của các biến cố hiếm, tức là biến cố ít khi xảy ra, như: số tai nạn giao thông tại một giao lộ; số trận động đất, số người chết do bị ngựa đá,… trong một năm. Đây là các biến rời rạc nhưng lại được xét trong một khoảng thời gian (hoặc không gian) liên tục, nhưng hữu hạn.

***a. Định nghĩa**:* BNN X được gọi là có phân phối Poisson với tham số là λ, kí hiệu là $X \sim P(\lambda)$, nếu như

$$P[X=k]=\frac{e^{-\lambda}\lambda^k}{k!},\ k=0,1,2,...,n,... \tag{4.16}$$

b. Tính chất

Nếu $X \sim P(\lambda)$ thì

$$E(X)=D(X)=\lambda \tag{4.17}$$

Nếu $X_i,\ i=\overline{1,n}$ là các biến ngẫu nhiên độc lập và $X_i \sim P(\lambda_i),\ i=\overline{1,n}$ thì

$$\sum_{i=1}^{n} X_i \sim P\left(\sum_{i=1}^{n}\lambda_i\right) \tag{4.18}$$

Ví dụ3f 4. Số tai nạn trung bình mỗi năm xảy ra tại một giao lộ là 5. Tính xác suất có đúng 3 tai nạn tại giao lộ đó trong năm nay.

Giải:

Đặt X là số tai nạn xảy ra tại giao lộ trong một năm. Khi đó có thể coi như X có luật phân phối Poisson. Vì số tai nạn trung bình mỗi năm tại giao lộ là 5 nên $X \sim P(5)$. Do đó xác suất có đúng 3 tai nạn tại giao lộ đó trong năm nay sẽ là

$$P(X=3)=\frac{e^{-5}\times 5^3}{3!}=\frac{0.0067\times 125}{6}=0.1404$$

Ví dụ3f 4. Theo dõi trong một khoảng thời gian dài người ta nhận thấy rằng số tai nạn trong một ngày tuân theo luật Poisson có trung bình là 2 đối với những ngày trong tuần và là 3 đối với những ngày cuối tuần (là 2 ngày thứ bảy và chủ nhật).

Quan sát ngẫu nhiên một ngày.

a) Tính xác suất có đúng 3 tai nạn xảy ra trong ngày đó.

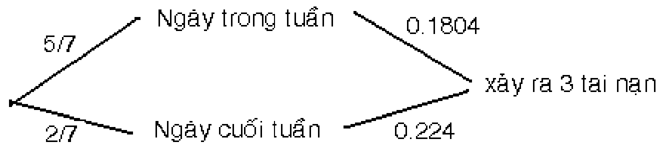
b) Nếu trong ngày đó không có tai nạn, tính xác suất ngày đó là ngày trong tuần.

Giải:

(Sinh viên tự tìm hiểu những xác suất có trong sơ đồ)

Đặt X = Số tai nạn xảy ra trong một ngày. Khi đó ta có X tuân theo luật phân phối Poisson: $X \sim P(\lambda)$. Tham số của phân phối Poisson trong bài toán này tuỳ thuộc vào ngày mà chúng ta quan sát: $\lambda = 2$ nếu là ngày trong tuần; $\lambda = 3$ nếu là ngày cuối tuần.

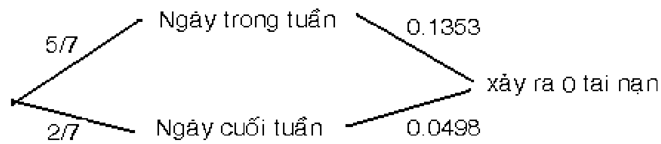
a) Ta có sơ đồ dưới.



Theo công thức xác suất đầy đủ:

$$P(X=3) = \frac{5}{7} \times 0.1804 + \frac{2}{7} \times 0.224 = 0.1929$$

b)



Dựa vào sơ đồ trên, xác suất để ngày đó không có tai nạn là:

$$P(X=0) = \frac{5}{7} \times 0.1353 + \frac{2}{7} \times 0.0498 = 0.1109\,.$$

Do đó, xác suất để ngày không có tai nạn là ngày trong tuần là:

$$P\left(\text{Ngày trong tuần} \middle/ X=0\right) = \frac{P(\text{Ngày trong tuần})\,P\left(X=0 \middle/ \text{Ngày trong tuần}\right)}{P(X=0)}$$

$$= \frac{\frac{5}{7} \times 0.1353}{0.1109} = \frac{0.0966}{0.1109} = 0.8711$$

Sinh viên thử nghĩ về bài toán tìm λ trong phân phối Poisson. Chẳng hạn: Cho biết tỉ lệ các mẫu vật liệu may có ít nhất 1 lỗi là 90%. Tìm số lỗi trung bình trên một vật liệu may biết số lỗi đó tuân theo luật Poisson.

c. *Công thức xấp xỉ*

Khi n khá lớn và p khá bé thì phân phối nhị thức $B(n,p)$ sẽ xấp xỉ phân phối Poisson $P(\lambda)$ với $\lambda = np$, nghĩa là ta có công thức xấp xỉ:

$$C_n^k p^k q^{n-k} \approx \frac{\lambda^k}{k!}e^{-\lambda}. \tag{4.19}$$

Chứng minh:

Ta có

$$P(X=k) = \frac{n!}{k!(n-k)!}p^k q^{n-k}$$

$$= \frac{n!}{k!(n-k)!}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k}$$

$$= \frac{n(n-1)...(n-k)}{n^k}.\frac{\lambda^k}{k!}.\frac{\left(1-\frac{\lambda}{n}\right)^n}{\left(1-\frac{\lambda}{n}\right)^k}$$

Do n khá lớn, p khá bé nên:

$$\left(1-\frac{\lambda}{n}\right)^n \approx e^{-\lambda}, \quad \frac{n(n-1)...(n-k)}{n^k} \approx 1, \quad \left(1-\frac{\lambda}{n}\right)^k \approx 1$$

Do đó: $P[X=k] \approx \dfrac{\lambda^k}{k!}e^{-\lambda}$.

Ví dụ3f 4. Ở một trường học, người ta nhận thấy rằng xác suất để 1 học sinh khi đi học bị bệnh và phải nằm điều trị tại phòng y tế của trường là 0.0004. Biết rằng trong một buổi học, trung bình có 7000 học sinh, hãy:

a) Tính xác suất để trong một buổi học có 3 học sinh phải nằm điều trị tại phòng y tế.

b) Theo bạn thì phòng y tế cần trang bị bao nhiêu giường điều trị.

Giải:

Nếu gọi X là số học sinh cần nằm điều trị trong một buổi học thì $X \sim B(7000;\ 0.0004)$

Do $n = 7000$ khá lớn và $p = 0.0004$ khá nhỏ nên coi như X tuân theo luật phân phối Poisson:

$$X \sim P\left(\lambda = np = 7,000 \times 0.0004 = 2.8\right)$$

a) Xác suất để trong một buổi học có 3 học sinh phải nằm điều trị tại phòng y tế là:

$$P\left[X = 3\right] = \frac{2.8^3}{3!}e^{-2.8} \approx 0.2225.$$

b) Số giường điều trị cần trang bị chính là số học sinh trung bình cần nằm điều trị. Vậy số giường trung bình phòng y tế cần trang bị là $\lambda = np = 2.8 \approx 3$.

4.4.2. Luật phân phối của BSNN liên tục

4.4.2.1. Phân phối chuẩn

a. *Định nghĩa*

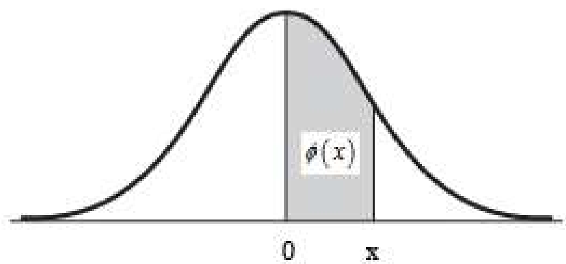
BNN X có phân phối chuẩn, hay phân phối bình thường (Normal Distribution), với hai tham số là μ và σ^2, kí hiệu là $X \sim N\left(\mu, \sigma^2\right)$, nếu X có hàm mật độ dạng:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \ x \in \mathbb{R}. \tag{4.20}$$

Nếu $\mu = 0$ và $\sigma^2 = 1$ thì $X \sim N\left(0,1\right)$ được gọi là phân phối chuẩn đơn giản hay phân phối Gauss. Xác suất $P\left[0 \le X \le x\right]$ (hình 4.2) trong trường hợp này được tính sẵn trong bảng hàm Laplace ở cuối tài liệu.

$$P\left[0 \le X \le x\right] = \phi(x) = \frac{1}{\sqrt{2\pi}}\int_0^x e^{-\frac{t^2}{2}}dt \tag{4.21}$$

Để ý hàm $\phi(x)$ là hàm lẻ, tức là $\phi(-x) = -\phi(x)$.



Hình 4.2

b. *Tính chất:*

Cho $X \sim N\left(\mu,\sigma^2\right)$, khi đó

(i) $E\left(X\right)=\mu$; $D\left(X\right)=\sigma^2$ (4.22)

(ii) $Z=\dfrac{X-\mu}{\sigma}\sim N\left(0,1\right)$ (4.23)

(iii) $P\left[a\le X\le b\right]=\phi\left(\dfrac{b-\mu}{\sigma}\right)-\phi\left(\dfrac{a-\mu}{\sigma}\right)$ (4.24)

Hệ quả: Qui tắc kσ (qui tắc k -xích ma)

Giả sử $X\sim N\left(\mu,\sigma^2\right)$. Khi đó

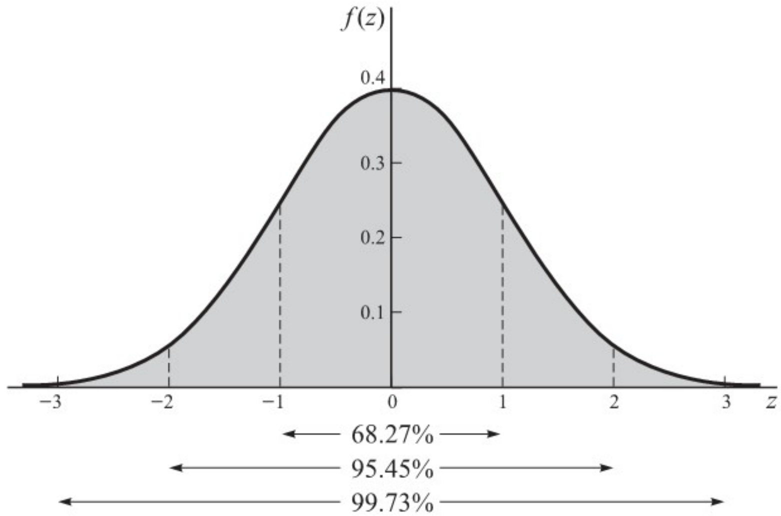
$$P\left[\left|X-\mu\right|<k\sigma\right]=P\left[-k<\frac{X-\mu}{\sigma}<k\right]$$

$$=\phi(k)-\phi(-k)=2\phi(k) \quad (4.25)$$

Khi $k=3$, ta có qui tắc 3-xích ma :

$$2\phi(3)=0.9973$$

Qui tắc này hay được dùng trong tính toán kỹ thuật vì xác suất, hay độ tin cậy, cao.



Hình 4.3: Qui tắc $k\sigma$, Z~N(0,1).

Cách tra bảng hàm Laplace: Để được đáp số ở ví dụ sau, ta phải tra bảng hàm Laplace như sau:

ϕ(0.13) tra ở hàng 0.1; cột 3.

ϕ(0.05) tra ở hàng 0.0; cột 5.

Ví dụ3f 4. Độ dài của chi tiết máy do cùng một máy tiện ra tuân theo luật phân phối chuẩn với trung bình là 50 cm, độ lệch chuẩn là 5cm. Tìm tỉ lệ chi tiết máy do máy trên tiện ra có độ dài từ:

a) 47.5 cm đến 50.1 cm

b) 50.25 cm đến 50.65cm

Giải:

Gọi X là độ dài của chi tiết máy. Theo đề bài X ~N(50; 25)

Đặt

$$Z = \frac{X - 50}{5}.$$

Khi đó

$$Z \sim N(0,1).$$

a) X =47.5 $\Rightarrow$ Z =-0.5 và X =50.1 $\Rightarrow$ Z =0.02

Tra bảng tích phân Laplace, suy ra:

$$P[47.5 \le X \le 50.1] = P[-0.5 \le Z \le 0.02]$$

$$= \phi(0.02) - \phi(-0.5)$$
$$= \phi(0.02) + \phi(0.5)$$

$$\approx 0.00798 + 0.19147 = 0.19945 .$$

b) Tương tự câu a) ta sẽ có

$$P[50.25 \le X \le 50.67] = \phi\left(\frac{50.67-50}{5}\right) - \phi\left(\frac{50.25-50}{5}\right)$$

$$= \phi\ (0.13) - \phi\ (0.05)$$

$$= 0.05172 - 0.01994$$

$$= 0.03178.$$

Sinh viên nghĩ thêm về bài toán: giả sử quan sát ngẫu nhiên 10 chi tiết, tính xác suất có đúng 3 chi tiết có độ dài từ 47.5 đến 50.1cm.

BT: Tuổi thọ X (năm) của một linh kiện máy tính là biến ngẫu nhiên tuân theo phân phối chuẩn với trung bình là 4.2 năm và phương sai là 2. Khi bán một linh kiện thì lãi được 100.000 đồng, nhưng nếu phải bảo hành thì lỗ 300.000 đồng. Cần quy định thời gian bảo hành là bao nhiêu để tiền lãi trung bình khi bán một linh kiện là 40.000 đồng.

1.5.4. Xấp xỉ phân phối nhị thức sang phân phối chuẩn

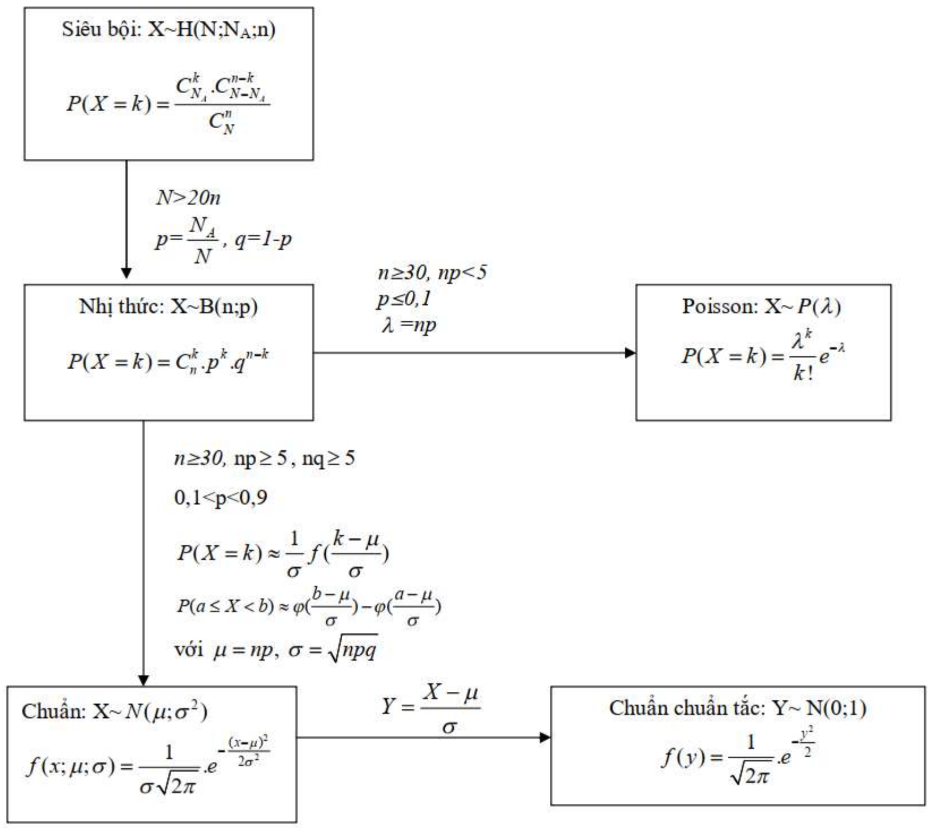
Biến nhị thức đại diện cho trường hợp đặc biệt của $S_n = X_1 + \cdots + X_n$, trong đó tất cả X_i độc lập với xác suất thành công p. Đối với các giá trị vừa phải của p (thường lấy giái trị $0.05 \leqslant p \leqslant 0.95$) và với n lớn, công thức (1.40) trở thành:

$$B(n;p) \approx N(\mu, \sigma^2), \text{ với } \mu = np, \sigma^2 = npq \tag{1.41}$$

Ví dụ 1.24. *Một loại virus máy tính mới tấn công một thư mục gồm 300 tệp. Mỗi tệp bị hỏng với xác suất 0.2 độc lập với các tệp khác. Tính xác suất ít hơn 50 tập tin bị hỏng?*

Đặt X = Số tệp bị hỏng trong thư mục. Khi đó $X \sim B(n,p)$ với $n = 300$, $p = 0.2$. Áp dụng xấp xỉ (1.8) ta có $X \sim N(np = 60, sigma^2 = 48 \approx (6.928)^2)$
Do đó,

$$P(X < 50) = P(X < 49.5) =$$



BT: Một vườn lan có 10.000 cây sắp nở hoa, trong đó có 1.000 cây hoa màu đỏ.

1) Tính xác suất để khi chọn ngẫu nhiên 20 cây lan thì được 5 cây có hoa màu đỏ.

2) Tính xác suất để khi chọn ngẫu nhiên 50 cây lan thì được 10 cây có hoa màu đỏ.

3) Có thể tính xác suất để khi chọn ngẫu nhiên 200 cây lan thì có 50 cây hoa màu đỏ được không ?

BT: Một lô hàng thịt đông lạnh đóng gói nhập khẩu có chứa 0,4% bị nhiễm khuẩn. Tìm xác suất để khi chọn ngẫu nhiên 1.000 gói thịt từ lô hàng này có:

1) không quá 2 gói bị nhiễm khuẩn;

2) đúng 34 gói bị nhiễm khuẩn.

Để ý việc hiệu chỉnh liên tục khi xấp xỉ BNN rời rạc sang BNN liên tục

4.4.2.2. Phân phối "khi bình phương"

Xét n BSNN độc lập $X_i \sim N(0,1)$, $i = \overline{1,n}$. Khi đó

$$X = \sum_{i=1}^{n} X_i^2$$

sẽ có phân phối khi bình phương ở n bậc tự do.

Kí hiệu: $X \sim \chi^2(n)$.

Nếu $X \sim \chi^2(n)$ thì

$$E(X) = n, \tag{4.26}$$

$$D(X) = 2n. \tag{4.27}$$

Bảng xác suất $P\left[X < \chi^2(n,\alpha)\right] = \alpha$ được tính sẵn phía cuối tài liệu. Khi n khá lớn thì phân phối $\chi^2(n)$ sẽ tiến về phân phối chuẩn.

4.4.2.3. Phân phối Student

Giả sử $X \sim N(0,1)$ và $Y \sim \chi^2(n)$. Khi đó biến ngẫu nhiên

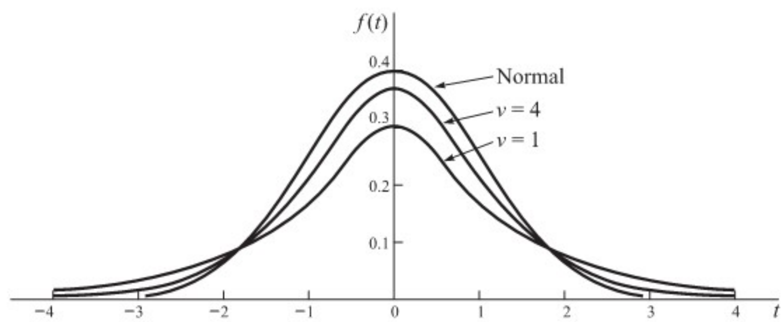
$T = \dfrac{X}{\sqrt{Y}}\sqrt{n}$ sẽ có phân phối Student ở n bậc tự do và kí hiệu là $T \sim t(n)$

Nếu $T \sim t(n)$ thì

$$E(T) = 0 \text{ và } D(T) = \frac{n}{n-2} \tag{4.28}$$

Phân phối Student có cùng dạng và tính đối xứng như phân phối chuẩn nhưng nó phản ánh sự thay đổi của phân phối sâu sắc hơn.

Trong thực hành, khi $n > 30$, phân phối $t(n)$ có đường cong mật độ rất gần với đường cong mật độ của phân phối chuẩn đơn giản $N(0,1)$ (xem hình) nên ta thường dùng phân phối $N(0,1)$ cho trường hợp này.



Hình 4.4: Đồ thị phân phối Student và phân phối chuẩn, v là số bậc tự do của phân phối Student.

Bảng xác suất $P(|X| > t(n,\alpha)) = \alpha$ được tính sẵn phía cuối tài liệu.

4.5. GIỚI THIỆU VỀ ĐẠI LƯỢNG NGẪU NHIÊN HAI CHIỀU RỜI RẠC

4.5.1. Khái niệm

Biến ngẫu nhiên ta xét ở phần trước gọi là biến ngẫu nhiên một chiều. Trong thực tế, ta còn gặp các biến ngẫu nhiên mà giá trị có thể có của nó được xác định bởi bộ 2, 3, hoặc n số ($n>3$).

Ví dụ3f 4. Khi nghiên cứu điểm rơi của một vật trên mặt phẳng, điểm rơi được xác định bởi hai thành phần tọa độ (X,Y).

Khi nghiên cứu về thể lực của một người, người ta quan tâm đến chiều cao và cân nặng.

Đại lượng ngẫu nhiên hai chiều là đại lượng ngẫu nhiên mà các giá trị có thể có của nó được xác định bằng hai số, ký hiệu là (X,Y).

Hai biến số ngẫu nhiên X và Y được gọi là các thành phần của đại lượng ngẫu nhiên hai chiều.

Đại lượng ngẫu nhiên 2 chiều gọi là rời rạc (liên tục) nếu các thành phần của nó là các đại lượng ngẫu nhiên rời rạc (liên tục).

4.5.2. Bảng phân phối xác suất

Bảng phân phối xác suất của hai đại lượng ngẫu nhiên X và Y rời rạc có dạng:

$X \backslash Y$	y_1	y_2		y_j		y_m
x_1	p_{11}	p_{12}		p_{1j}		p_{1m}
x_2	p_{21}	p_{22}		p_{2j}		p_{2m}
x_i	p_{i1}	p_{i2}		p_{ij}		p_{im}
x_n	p_{n1}	p_{n2}		p_{nj}		p_{nm}

Bảng 4.2.

Trong đó:

$x_i, i = \overline{1,n}$ là các giá trị có thể có của thành phần X,

$y_j, j = \overline{1,m}$ là các giá trị có thể có của thành phần Y,

$p_{ij} = P\left[\left(X,Y\right) = \left(x_i, y_j\right)\right]$, $i = \overline{1,n};\ j = \overline{1,m}$ thỏa $\sum\limits_{i,j} p_{ij} = 1$.

Nhận xét, khi biết phân phối xác suất của biến ngẫu nhiên 2 chiều thì ta cũng có thể suy ra được bảng phân phối xác suất của mỗi thành phần.

VD:

1. *(3,0 điểm)* Cho phân phối xác suất đồng thời của hai biến ngẫu nhiên rời rạc X, Y như hình bên.
a. Tìm phân phối xác suất của X.
b. Tính P(X+Y>3).
c. Tìm phân phối xác suất của X khi Y=2.

		x		
	P(x,y)	0	1	2
y	1	0.05	0.15	0.07
	2	0.1	0.15	0.1
	3		0.18	0.12
	4	0.03		0.05

4.5.3. Kỳ vọng và phương sai của các thành phần trong biến ngẫu nhiên rời rạc hai chiều

Giả sử đại lượng ngẫu nhiên hai chiều $\left(X,Y\right)$ có luật phân phối xác suất cho trong bảng hai chiều như trên, khi đó:

Kỳ vọng và phương sai của X:

$$E(X) = \sum_{i=1}^{n}\sum_{j=1}^{m} x_i p_{ij}\ ;\ D(X) = \sum_{i=1}^{n}\sum_{j=1}^{m} x_i^2 p_{ij} - E^2(X) \qquad (4.29)$$

Kỳ vọng và phương sai của Y:

$$E(Y) = \sum_{j=1}^{m}\sum_{i=1}^{n} y_j p_{ij}\ ;\ D(X) = \sum_{j=1}^{m}\sum_{i=1}^{n} y_j^2 p_{ij} - E^2(Y) \qquad (4.30)$$

4.5.4. Covarian và hệ số tương quan tuyến tính

Covarian, hay hiệp phương sai, của X và Y, ký hiệu là $Cov(X,Y)$, được định nghĩa như sau:

$$Cov(X,Y) = E\left[\left(X - E(X)\right)\left(Y - E(Y)\right)\right] \tag{4.31}$$

Nếu (X,Y) rời rạc thì

$$Cov(X,Y) = \sum_{i=1}^{n}\sum_{j=1}^{m}\left(x_i - E(X)\right)\left(y_j - E(Y)\right)P_{ij}. \tag{4.32}$$

Hệ số tương quan tuyến tính ký hiệu là ρ và được định nghĩa như sau:

$$\rho = \frac{Cov(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}} \tag{4.33}$$

Một tính chất quan trọng của hệ số này là: Nếu X và Y độc lập thì ρ=0.

Trong chương 8 ta sẽ nghiên cứu kỹ tính chất và ý nghĩa của hệ số này.

Đọc thêm:

LUẬT SỐ LỚN
và ĐỊNH LÝ GIỚI HẠN TRUNG TÂM

Luật số lớn là những định lý nói lên tính ổn định của trung bình cộng của một số lớn các BSNN có kỳ vọng bằng 0. Biến số có kỳ vọng khác 0 có thể đưa về biến số có kỳ vọng bằng 0 bằng cách trừ đi kỳ vọng của nó.

Định lý giới hạn trung tâm là các định lý nói về sự hội tụ (dưới một số điều kiện nào đó) của hàm phân phối xác suất của tổng các biến độc lập X_i, $i = \overline{1,n}$: $\sum_{i=1}^{n} X_i$ (đã được biến đổi).

1. Bất đẳng thức Trê-bư-sép

Nếu biến ngẫu nhiên có phương sai hữu hạn thì với mọi $\varepsilon > 0$ cho trước

$$P\left[\left|X - E(X)\right| \geq \varepsilon\right] \leq \frac{D(X)}{\varepsilon^2} \tag{4.34}$$

2. Luật số lớn dạng Trê-bư-sép

a. Định nghĩa: Ta nói dãy các biến ngẫu nhiên $\{X_n, n = 1, 2.., +\infty\}$ hội tụ theo xác suất tới a nếu với n đủ lớn ta luôn có:

$$P\left[\left|X_n - a\right| < \varepsilon\right] > 1 - \delta \tag{4.35}$$

Hay

$$\lim_{n \to +\infty} P\left[\left|X_n - a\right| \geq \varepsilon\right] = 0 \tag{4.36}$$

Với $\varepsilon > 0$ bé tùy ý và δ nói chung phụ thuộc n và ε. Khi đó ta ký hiệu $X_n \xrightarrow[n \to +\infty]{XS} a$

Luật số lớn dạng Trê-bư-sép:

Nếu dãy các biến ngẫu nhiên $\{X_n, n = 1,.., +\infty\}$ có phương sai bị chặn thì khi n đủ lớn, trung bình cộng của các giá trị của dãy sẽ hội tụ theo xác suất về trung bình cộng của các kỳ vọng tương ứng, tức là:

$$P\left[\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} E(X_i)\right| < \varepsilon\right] > 1 - \delta \tag{4.37}$$

Hay

$$\lim_{n \to +\infty} P\left[\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} E(X_i)\right| \geq \varepsilon\right] = 0 \tag{4.38}$$

Từ đây ta có ngay các hệ quả sau:

Hệ quả 1: Giả sử $E\left(X_n\right) = a,\ n = 1, 2, ...$ thì

$$\lim_{n\to+\infty} P\left[\left|\frac{1}{n}\sum_{i=1}^{n} X_i - a\right| \geq \varepsilon\right] = 0 \ , \ \forall\varepsilon \geq 0 \qquad (4.39)$$

Hệ quả 2: Trong thí nghiệm xây dựng công thức Bernoulli, nếu gọi X là số lần thành công (số lần xuất hiện biến cố A) khi thực hiện lặp lại phép thử n lần với xác suất thành công ở mỗi phép thử là p thì:

$$\lim_{n\to+\infty} P\left[\left|\frac{X}{n} - p\right| \geq \varepsilon\right] = 0 \ , \ \forall\varepsilon \geq 0 \qquad (4.40)$$

Hệ quả này chứng minh tính đúng đắn của định nghĩa xác suất theo thống kê.

3. Định lý (*Định lý giới hạn trung tâm với các thành phần cùng phân phối*)

Giả sử các biến ngẫu nhiên X_i, $i = 1, 2, ...$ độc lập và có cùng phân phối với kỳ vọng $E(X) = a$ và phương sai $D(X) = \sigma^2$. Khi đó hàm phân phối xác suất của biến ngẫu nhiên

$$\frac{\frac{1}{n}\sum_{i=1}^{n} X_i - a}{\frac{\sigma}{\sqrt{n}}}$$

sẽ hội tụ đến hàm phân phối chuẩn $\phi(x)$ tại mọi x, khi $n \to +\infty$.

Sự hội tụ như trên nói lên vai trò đặc biệt quan trọng của phân phối chuẩn trong các hiện tượng thực tế.

BÀI TẬP

4.1. Một tổ có 6 nam và 4 nữ. Chọn ngẫu nhiên ra 3 người. Tìm luật phân phối xác suất của số nữ trong nhóm được chọn.

4.2. Một xạ thủ bắn 3 viên đạn một cách độc lập vào một mục tiêu. Xác suất để mỗi viên đạn trúng mục tiêu là 0,7. Gọi X là số viên đạn trúng mục tiêu, hãy lập luật phân phối của X.

4.3. Một xạ thủ bắn 3 viên đạn một cách độc lập vào một mục tiêu cho đến khi trúng (hoặc hết đạn) thì dừng bắn. Tìm phân phối xác suất của số đạn đã bắn, biết xác suất bắn trúng mục tiêu của mỗi lần bắn là 0.8.

4.4. Có 3 xạ thủ cùng bắn vào một mục tiêu, mỗi người bắn 1 viên đạn, trong cùng một số điều kiện nhất định. Xác suất để mỗi xạ thủ bắn trúng mục tiêu lần lượt là 0.6; 0.7; 0.9. Gọi X là số viên đạn trúng mục tiêu, hãy lập luật phân phối của X. Tính $E(X)$, $D(X)$ và Mod(X).

4.5. Một hộp gồm có 6 lọ thuốc, trong đó có 2 lọ quá đát (date). Một người kiểm tra lần lượt từng lọ cho đến khi phát hiện 2 lọ hỏng thì dừng.

a) Tính xác suất để người ấy dừng ở lần kiểm tra thứ 3.

b) Sau hai lần kiểm tra, hãy lập luật phân phối xác suất cho số lọ quá đát còn lại trong hộp.

4.6. Hai xạ thủ A và B tập bắn. Mỗi người bắn hai phát. Xác suất trúng đích của A trong mỗi lần bắn là 0.4; của B là 0.5.

Gọi X là số phát trúng của A trừ đi số phát trúng của B. Hãy tìm phân phối xác suất của X.

4.7. Trong một chiếc hộp có 4 tấm thẻ đánh số từ 1 đến 4. Chọn ngẫu nhiên hai tấm thẻ rồi cộng hai số ghi trên thẻ lại với nhau. Gọi X là kết quả. Tìm phân phối xác suất cho X.

4.8. Cho biến ngẫu nhiên X có hàm mật độ dạng

$$f(x) = \begin{cases} \frac{1}{2}\sin x & \text{, nếu } x \in [0,\pi] \\ 0 & \text{, nếu } x \notin [0,\pi] \end{cases}$$

Hãy viết hàm phân phối xác suất của X

4.9. Cho hàm mật độ của biến ngẫu nhiên X có dạng:

a) $$f(x) = \begin{cases} Ax & \text{, nếu } x \in [0,1] \\ 0 & \text{, nếu } x \notin [0,1] \end{cases}$$

b) $$f(x) = \begin{cases} A.\sin x & \text{, nếu } x \in [0,\pi] \\ 0 & \text{, nếu } x \notin [0,\pi] \end{cases}$$

c) $$f(x) = \begin{cases} ACos\pi\, x & \text{, nếu } x \in \left[0,\frac{1}{2}\right] \\ 0 & \text{, nếu } x \notin \left[0,\frac{1}{2}\right] \end{cases}$$

d) $$f(x) = \begin{cases} A\frac{1}{x^4} & \text{, nếu } x \geq 1 \\ 0 & \text{, nếu } x < 1 \end{cases}$$

Hãy xác định hằng số A. Viết hàm phân phối của X. Tính $E(X)$, $D(X)$ (nếu có)

4.10. Tuổi thọ của một loài côn trùng A là một biến ngẫu nhiên X (đơn vị là *tháng*) với hàm mật độ như sau:

$$f(x) = \begin{cases} kx^2(4-x) \text{ , nếu } 0 \le x \le 4 \\ 0 \text{ , nếu } x \notin [0;4]. \end{cases}$$

(i) Tìm hằng số k và vẽ đồ thị của $f(x)$.

(ii) Quan sát ngẫu nhiên một con côn trùng thuộc loài A, tìm xác suất để côn trùng chết trước khi nó được 1 tháng tuổi.

(iii) Quan sát ngẫu nhiên một con côn trùng thuộc loài A sống qua 1 tháng tuổi, tính xác suất nó chết trước khi được 2 tháng tuổi.

4.11. Trọng lượng của một con gà 6 tháng tuổi là một biến ngẫu nhiên X (đơn vị tính là Kg) có hàm mật độ:

$$f(x) = \begin{cases} k\left(x^2-1\right), \text{ với } 1 \le x \le 3 \\ 0 \text{ , với } x > 3, x < 1 \end{cases}, \ k \text{ là hằng số.}$$

a) Tìm k

b) Với k tìm được, tìm:

- Trọng lượng trung bình của gà 6 tháng tuổi.
- Tỉ lệ gà chậm lớn, biết gà 6 tháng tuổi chậm lớn là gà có trọng lượng nhỏ hơn 2 kg.
- Hàm phân phối của X.

4.12. Biến ngẫu nhiên liên tục X được gọi là có phân phối đều trên đoạn $[a,b]$ nếu hàm mật độ của X có dạng

$$p(x) = \begin{cases} c \text{ , nếu } x \in [a;b] \\ 0 \text{ , nếu } x \notin [a;b] \end{cases}$$

a) Xác định hằng số c.

b) Tìm hàm phân phối $F(x)$ của X và vẽ đồ thị của hàm phân phối này.

c) Tính $P[\alpha<X<\beta]$ và $P[\alpha'<X<\beta']$ với $\alpha, \beta, \alpha', \beta'$ là các giá trị thuộc [a;b] và $\beta-\alpha=\beta'-\alpha'$

d) Tính các đặc trưng $E(X)$, $D(X)$, $\sigma(X)$ và nêu ý nghĩa của các số đặc trưng này.

Tính Mod(X) và Med(X).

e) Tìm hàm mật độ $q(y)$ của BSNN $Y = 2X - 3$.

4.13. Cho biến ngẫu nhiên liên tục X có hàm phân phối:

$$F(x) = \begin{cases} 0 & \text{, nếu } x < -\dfrac{\pi}{2} \\ a + b\sin x & \text{, nếu } -\dfrac{\pi}{2} \le x \le \dfrac{\pi}{2} \\ 1 & \text{, nếu } x > \dfrac{\pi}{2} \end{cases} \quad \text{(a, b là hằng số)}$$

a) Tìm a và b.

b) Với a,b tìm được, tính hàm mật độ $p(x)$ của X, Mod(X), Med(X), P[X>π/4]

4.14. Có hai hộp bi. Hộp I chứa 10 bi gồm 3 bi đỏ và 7 bi đen. Hộp II chứa 5 bi gồm 2 bi đỏ và 3 bi đen. Lấy ngẫu nhiên 1 bi từ hộp I bỏ vào hộp II, rồi từ hộp II lấy ngẫu nhiên 1 bi.

a) Tính xác suất để bi lấy từ hộp II là bi đỏ.

b) Lập luật phân phối xác suất cho số bi đỏ có trong hộp II sau khi bỏ vào 1 bi lấy từ hộp I.

4.15. Có 2 hộp bi, hộp thứ i có 5i +2 bi trong đó có 2i bi đỏ (i=1,2). Lấy ngẫu nhiên mỗi hộp 1 bi

a) Lập luật phân phối xác suất cho số bi đỏ có trong 2 bi lấy ra.

b) Lấy tiếp 1 bi trong hộp 1. Tính xác suất để được bi đỏ.

4.16. Trò chơi "bầu cua" công bằng hay thiên vị? Tại sao?

Luật chơi: *Giả sử đặt a (đồng) vào ô B. Gieo ngẫu nhiên 3 cục xí ngầu. Nếu xuất hiện i mặt B thì sẽ được thưởng i lần a đồng (i=1,2,3), ngược lại, nếu không xuất hiện mặt B nào thì mất số tiền đặt vào.*

4.17. Trong một cuộc xổ số người ta phát hành 100,000 vé trong đó có 10,000 vé trúng giải. Cần phải mua ít nhất bao nhiêu vé để với xác suất không nhỏ hơn 0.95 ta sẽ trúng ít nhất 1 vé?

4.18. Tại nhà máy A, trung bình một tháng có hai tai nạn lao động. Coi số tai nạn xảy ra trong một tháng là biến ngẫu nhiên tuân theo luật phân phối Poisson với λ=2. Tính xác suất để

a) Trong khoảng thời gian ba tháng xảy ra nhiều nhất 3 tai nạn,

b) Trong ba tháng liên tiếp, mỗi tháng xảy ra nhiều nhất 1 tai nạn.

Một trạm cho thuê xe taxi có 3 chiếc xe. Hàng ngày trạm phải nộp thuế 8 USD cho 1 chiếc xe bất kể xe đó có được thuê hay không. Mỗi chiếc được cho thuê với giá 20 USD. Giả sử số xe được yêu cầu cho thuê của trạm trong một ngày là đại lượng ngẫu nhiên X có phân phối Poisson với tham số $\lambda = 2.8$.

a) Tính số tiền trung bình trạm thu được trong một ngày.

b) Giải bài toán trên trong trường hợp trạm có 4 chiếc xe.

c) Theo bạn, trạm nên có 3 hay 4 chiếc xe ?

HD: Lập luật ppxs cho số xe được thuê trong ngày, qua đó suy ra luật ppxs của số tiền trạm thu về trong một ngày (xem như khi cho thuê từ 3 xe trở lên thì thu lời 20x3-8x3=36USD)

4.19. Một viên đạn có tầm xa trung bình là 300m. Giả sử tầm xa đó là một biến ngẫu nhiên tuân theo luật phân phối chuẩn với σ =10. Hãy tìm tỉ lệ đạn bay quá tầm trung bình từ 15 đến 25 mét.

4.20. Một giống chuột có trọng lượng X (g) tuân theo luật phân phối chuẩn với trung bình 250 g và phương sai 100g^2. Bắt ngẫu nhiên được một con chuột thuộc giống trên, tính xác suất để nó có trọng lượng từ 225.7g đến 253.5g.

BT. Kho linh kiện máy tính có 1000 sản phẩm, trong đó có 70 sản phẩm bị lỗi. Người ta giao ngẫu nhiên 100 linh kiện cho khách hàng.
a). Tính xác suất khách hàng nhận được toàn sản phẩm tốt.
b). Kỳ vọng và phương sai của số sản phẩm lỗi mà khách hàng nhận là bao nhiêu?

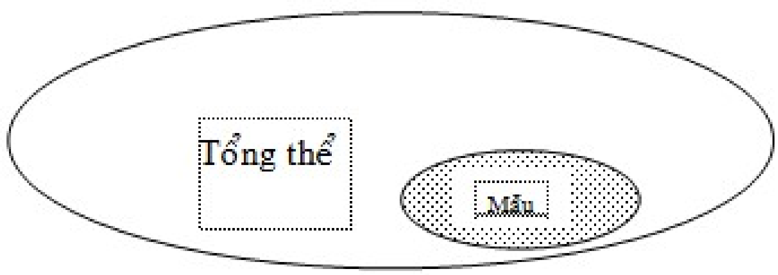
Chương 5
LÝ THUYẾT MẪU

5.1. KHÁI NIỆM

Ví dụ mở đầu: giả sử chúng ta muốn tính chiều cao trung bình của người Việt Nam ở độ tuổi nào đó. Cách tốt nhất để tính ra chiều cao trung bình này là đo chiều cao của tất cả công dân Việt Nam ở độ tuổi đó (đám đông), tuy nhiên cách này không thể thực hiện được vì một số lý do như :

- Về kinh tế: tốn kém, vì phải chi phí rất nhiều cho việc đi lấy số liệu.
- Việc xác định tất cả công dân ở độ tuổi đó khá khó khăn, vì nếu thực hiện thì phải không được bỏ sót một người nào.
- Thời gian: dài, …

Ngoài ra, đối với các đặc tính khác, quan sát còn có tính hủy hoại, như việc kiểm tra xem que diêm còn tốt không, khi có kết luận thì cũng không còn dùng được nữa.



Thống kê học đề nghị một phương pháp là: chọn ra n người (n người này được gọi là một mẫu cỡ n), tính toán và kết luận cho chiều cao trung bình của công dân Việt Nam ở độ tuổi đó. Tất nhiên, sự suy rộng này có thể đúng, cũng có thể sai. Để hạn chế sai lầm khi suy rộng, mẫu phải chọn khách quan (tức xác suất để mỗi phần từ được chọn vào mẫu phải bằng nhau)

Thông thường ta có hai cách chọn mẫu: hoàn lại và không hoàn lại. Tuy nhiên, khi tập cần nghiên cứu có rất nhiều phần tử thì có thể coi như hai phương pháp này như nhau.

- **Đám đông** (hay **tổng thể,** hay **tập tổng quát**) là tập hợp tất cả các phần tử cùng loại (hiểu theo nghĩa: các phần tử có đặc tính hay định lượng chung) mà ta quan tâm nghiên cứu đến một hoặc vài tính chất nào đó.

Trong ví dụ trên, tất cả công dân quốc tịch Việt Nam ở độ tuổi đang xét là "đám đông". Một nhóm gồm n phần tử được chọn từ đám đông sao cho phản ánh trung thực đặc điểm của "đám đông" mà ta đang nghiên cứu gọi là mẫu và kí hiệu là $\left(x_1, x_2, ..., x_n\right)$, trong đó n gọi là cỡ mẫu.

- Định nghĩa mẫu ngẫu nhiên: một **mẫu ngẫu nhiên** cỡ n cho dấu hiệu X của đám đông là một nhóm n biến ngẫu nhiên $X_1, X_2, ..., X_n$ độc lập và có cùng phân phối với X.

Vậy $x_1, x_2, ..., x_n$ là một thể hiện của $X_1, X_2, ..., X_n$. Ta gọi $\left(x_1, x_2, ..., x_n\right)$ là mẫu cụ thể.

- Đối với đám đông, ta thường quan tâm hai mặt: lượng và chất

+ **Về lượng**, ta thường đánh giá về trung bình μ và phương sai σ^2. Chẳng hạn như điểm trung bình, thời gian trung bình đi đến trường, … và độ ổn định (phương sai).

+ **Về chất**, thường ta quan tâm đến tỉ lệ (hay xác suất) p các phần tử có tính chất A nào đó. Ví dụ như tỉ lệ học sinh giỏi, tỉ lệ phế phẩm, …

- **Thông kê** là một đại lượng mà giá trị chỉ phụ thuộc vào các quan sát (tức là phụ thuộc vào mẫu), không phụ thuộc các tham số chưa biết.

Ví dụ3f 5. Các biểu thức dùng để tính các điều sau là các thống kê:

- Điểm trung bình học tập.
- Tỉ lệ học sinh giỏi, …

Để có thể cô đọng và nhanh chóng nắm bắt được những thông tin quan trọng chứa đựng trong mẫu, ta đưa ra một vài chỉ số gọi là các số đặc trưng của mẫu.

5.2. CÁC ĐẶC TRƯNG CỦA MẪU NGẪU NHIÊN

Giả sử có mẫu ngẫu nhiên cỡ n: $\left(X_1, X_2, ..., X_n\right)$

5.2.1. Trung bình mẫu

$$\overline{X} = \frac{X_1 + X_2 + \ldots + X_n}{n} \tag{5.1}$$

5.2.2. Phương sai mẫu

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 \tag{5.2}$$

Bằng các biến đổi đơn giản ta có công thức khác cho phương sai mẫu như sau:

$$S^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2 \tag{5.3}$$

Phương sai mẫu hiệu chỉnh:

$$S^{*2} = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 \tag{5.4}$$

Sử dụng công thức (5.2) ta có:

$$S^{*2} = \frac{n}{n-1} S^2 \tag{5.5}$$

Khi có mẫu cụ thể thì các giá trị đặc trưng sẽ được ký hiệu như sau:

Trung bình mẫu :

$$\overline{x} = \frac{x_1 + x_2 + \cdots x_n}{n} \tag{5.6}$$

Phương sai mẫu:

$$\sigma_n^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \overline{x}^2 \tag{5.7}$$

Phương sai mẫu hiệu chỉnh:

$$\sigma_{n-1}^2 = \frac{n}{n-1}\sigma_n^2 \tag{5.8}$$

Mục đích chính của việc sử dụng các ký hiệu này là để chúng ta có các ký hiệu trùng với các ký hiệu trung bình, phương sai và phương sai hiệu chỉnh trong máy tính bỏ túi (calculator). Ngoài ký hiệu trên, ta vẫn có thể dùng $\overline{x}, s^2, s^{*2}$ (chữ thường) để chỉ các giá trị trung bình, phương sai, phương sai hiệu chỉnh được tính từ mẫu cụ thể.

5.2.3. Tỉ lệ các phần tử có tính chất A của mẫu

$$F_n = \frac{X_1 + X_2 + \cdots + X_n}{n} \tag{5.9}$$

trong đó

$$X_i = \begin{cases} 1 & \text{nếu quan sát } i \text{ có tính chất A} \\ 0 & \text{nếu quan sát } i \text{ không có tính chất A} \end{cases}$$

Trong thực hành, với mẫu cụ thể ta thường tính như sau:

$$f_n = \frac{\text{Số phần tử có tính chất A}}{\text{cỡ mẫu}} \tag{5.10}$$

* **Số liệu lặp**: Thực tế, ta thường gặp các số liệu lặp lại (chẳng hạn, có 10 người (n_i) có chiều cao là 1.64m(x_i)). Giả sử giá trị x_i lặp lại n_i lần, $i = \overline{1,k}$. Khi đó mẫu thường được trình bày dưới dạng bảng phân phối tần số như sau:

x_i	x_1	x_2	$\cdots$	x_k
n_i	n_1	n_2	$\cdots$	n_k

Bảng 5.1: Bảng phân phối tần số

Còn các công thức tương ứng sẽ là:

$$n = n_1 + n_2 + \cdots + n_k \tag{5.11}$$

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{k} n_i x_i \tag{5.12}$$

$$\sigma_n^2 = \frac{1}{n}\sum_{i=1}^{k} n_i x_i^2 - \overline{x}^2 \tag{5.13}$$

Ví dụ3f 5. Giả sử đếm số lần xảy ra sự cố kỹ thuật trên một máy của 10 máy may trong một buổi làm việc ta thu được số liệu dạng liệt kê như sau: 3, 0, 0, 3, 2, 4, 3, 4, 4, 0. Khi đó ta có số liệu lặp và có thể trình bày gọn lại như sau:

Số lỗi	0	1	2	3	4	>4
Số máy	3	0	1	3	3	0

Ví dụ3f 5. Cân ngẫu nhiên 45 con heo 3 tháng tuổi trong một trại chăn nuôi ta được kết quả sau:

x_i (kg)	35	37	39	41	43	45	47
n_i (con)	2	6	10	11	8	5	3

a) Hãy tính $\overline{x}$, σ_n^2 , σ_{n-1}^2.

b) Giả sử heo có trọng lượng ≥38kg là heo đạt tiêu chuẩn. Hãy tính tỉ lệ heo đạt tiêu chuẩn của mẫu trên.

Nhận xét: Trong mẫu trên, với cặp số liệu đầu tiên ta hiểu là: Có 2 con heo 3 tháng tuổi nặng 35kg; các cặp số liệu còn lại được hiểu tương tự.

Giải:

a) Lập bảng tính như bảng 5.2.

Từ kết quả tính toán ở bảng tính 5.2 ta tính được:

Trung bình mẫu:

$$\overline{x} = \frac{1843}{45} \approx 40.96.$$

Phương sai mẫu:

$$\sigma_n^2 = \frac{75909}{45} - (40.96)^2 \approx 9.51$$

Phương sai mẫu hiệu chỉnh:

$$\sigma_{n-1}^2 = \frac{45}{44} 9.51 \approx 9.73.$$

x_i	n_i	$n_i.x_i$	$n_i.x_i^2$
35	2	70	2450
37	6	222	8214
39	10	390	15210
41	11	451	18491
43	8	344	14792
45	5	225	10125
47	3	141	6627
Tổng	45	1843	75909

	n	$\sum_{i=1}^{k} n_i x_i$	$\sum_{i=1}^{k} n_i x_i^2$

Bảng 5.2

b) Tỉ lệ heo đạt tiêu chuẩn của mẫu: Ta tính trực tiếp từ bảng số liệu của đề cho theo công thức (5.10). Từ bảng số liệu đề cho ta thấy ngay 10 con heo nặng 39 kg, 11 con nặng 41, 8 con nặng 43, 5 con nặng 45, và 3 con nặng 47 kg là những con heo đạt tiêu chuẩn nên tỉ lệ heo đạt tiêu chuẩn của mẫu là:

$$f_n = \frac{10+11+8+5+3}{45} = \frac{37}{45} \approx 82.22\% .$$

♣ **Hướng dẫn sử dụng máy tính bỏ túi hiệu Casio Fx-500A** tính trung bình và phương sai mẫu.

Bước 1: Chuyển sang chế độ thống kê (chế độ **SD**):

$\boxed{\text{on}}$ $\boxed{\text{mode}}$ $\boxed{\cdot}$

Bước 2: Kiểm tra bộ nhớ xem đã xóa nhớ chưa bằng cách lấy giá trị n (shift 6)

Nếu $n = 0$: đã xóa nhớ, sang bước 3. Ngược lại, chuyển qua lại giữa các chế độ để xoá nhớ.

Bước 3: Nhập lần lượt các cặp số liệu (x_i, n_i) theo thứ tự x_i rồi n_i như sau:

$\boxed{x_i}$ $\boxed{\times}$ $\boxed{n_i}$ $\boxed{\text{M+}}$

(màn hình chỉ hiện ra giá trị x_i chứ không hiện ra tích $x_i n_i$)

Lặp lại như vậy cho đến hết bộ số liệu.

Bước 4: Lấy kết quả: Đọc các ký hiệu trên máy và lấy kết quả.

Ví dụ: Muốn lấy n, ta gõ: $\boxed{\text{shift}}$ $\boxed{6}$

Ghi chú: Các máy tính có ký hiệu tương tự đều có thể sử dụng với thao tác phù hợp.

Ví dụ3f 5. Nghiên cứu trọng lượng của một giống vịt mới ta có kết quả:

Cân nặng (kg)	1.25	1.5	1.75	2	2.25	2.5	2.75	3
Số con	2	6	24	35	39	24	14	6

Tính các đặc trưng mẫu của bộ số liệu trên

Giải:

Từ bộ số liệu, bằng máy tính ta tính được:

$n = 150$; $\sum_i n_i x_i = 327.75$ $\sum_i n_i x_i^2 = 737.4375$

Suy ra $\overline{x} = \dfrac{327.75}{150} = 2.185$

$$\sigma_n^2 = \frac{737.4375}{150} - (2.185)^2 \approx 4.92 - 4.77 = \underline{0.15}$$

$$\sigma_{n-1}^2 = \frac{n}{n-1}\sigma^2 = \frac{150}{149}.(0.15) \approx \underline{0.151}$$

Nhận xét: Những giá trị gạch dưới trong ví dụ trên có sai số không đáng kể so với kết qủa được tính trực tiếp bằng máy tính.

* **Số liệu cho theo khoảng**: Trong thực tế, đôi khi ta còn gặp số liệu cho bởi các lớp (a_i, b_i) (tức là với những giá trị quan sát gần nhau, người ta thường gộp vào một lớp để tiện trình bày), khi đó ta tính toán các đặc trưng mẫu bằng các giá trị ở chính giữa lớp đó, nghĩa là:

$$x_i = \frac{a_i + b_i}{2} \tag{5.14}$$

Ví dụ3f 5. Đo chiều cao của một số nam học sinh lớp 10 tại một trường phổ thông trung học ta có kết quả:

Chiều cao(m)	1.50-1.55	1.55-1.60	1.60-1.65	1.65-1.70
Số học sinh	4	27	23	12

Khi đó ta tính các đặc trưng mẫu thông qua bộ số liệu sau:

Chiều cao(m)	1.525	1.575	1.625	1.675
Số học sinh	4	27	23	12

♣ Tính đặc trưng trung bình và phương sai bằng cách biến đổi.

Ta thực hiện phép đổi biến để đưa số liệu quan sát về gần 0. Giả sử số liệu cách đều nhau một khoảng là h. Kí hiệu a là giá trị có số lần lặp cao nhất.

Đặt

$$y_i = \frac{x_i - a}{h} \tag{5.15}$$

Ta tính được $\overline{y}$ và $y\sigma_n^2$ (phương sai mẫu theo biến y)

Khi đó mối liên hệ giữa $\overline{x}$, $x\sigma_n^2$, $\overline{y}$ và $y\sigma_n^2$ cho bởi:

$$\overline{y} = \frac{1}{h}\left(\overline{x} - a\right); \tag{5.16}$$

$$y\sigma_n^2 = \frac{1}{h^2} x\sigma_n^2 \tag{5.17}$$

Ví dụ3f 5. Tính các đặc trưng trung bình và phương sai mẫu của bộ số liệu sau:

Chiều cao(m)	1.525	1.575	1.625	1.675
Số học sinh	4	27	23	12

Giải:

Ta thấy ngay $h = 0.05$. Chọn $a = 1.575$.

Lập bảng tính ta có:

x	n_i	y_i	$n_i y_i$	$n_i y_i^2$
1.525	4	-1	-4	4
1.575	27	0	0	0
1.625	23	1	23	23
1.675	12	2	24	48
	66	2	43	75

Dễ dàng tính được

$$\overline{y} = \frac{43}{66} \approx 0.652$$

$$y\sigma_n^2 = \frac{75}{66} - (0.652)^2 \approx 0.712$$

Từ đó:

$$\overline{x} \approx 0.05 \times 0.652 + 1.575 = 1.6076$$

$$x\sigma_n^2 \approx 0.05^2 \times 0.712 \approx 0.00178$$

5.3. LUẬT PHÂN PHỐI XÁC SUẤT CỦA CÁC ĐẶC TRƯNG MẪU

Từ luật số lớn và định lý giới hạn trung tâm, người ta chứng minh được các kết quả sau:

5.3.1. Phân phối của tỉ lệ mẫu

Khi n khá lớn, ta có $F_n \sim N\left(p;\dfrac{pq}{n}\right)$

5.3.2. Phân phối của trung bình mẫu

Giả sử dấu hiệu X có luật phân phối là $N\left(\mu,\sigma^2\right)$ và $\left(X_1, X_2, ..., X_n\right)$ là một mẫu ngẫu nhiên cỡ n.

Trường hợp 1: $n \geq 30$, phương sai σ^2 đã biết, khi đó :

$$\bar{X} \sim N\left(\mu,\frac{\sigma^2}{n}\right)$$

Trường hợp 2: $n \geq 30$, phương sai σ^2 chưa biết, khi đó

$$\bar{X} \sim N\left(\mu,\frac{S^{*2}}{n}\right)$$

Trường hợp 3: $n < 30$, phương sai σ^2 đã biết, khi đó

$\bar{X} \sim N\left(\mu,\dfrac{\sigma^2}{n}\right)$ (như trường hợp 1).

Trường hợp 4: $n < 30$, phương sai σ^2 chưa biết, khi đó

$\dfrac{\bar{X}-\mu}{S^*}.\sqrt{n} \sim t\left(n-1\right)$: phân phối Student ở $n-1$ bậc tự do.

Lưu ý: khi $n \geq 30$ thì có thể bỏ qua giả thiết X tuân theo luật phân phối chuẩn.

5.3.3. Phân phối của phương sai mẫu

Giả sử dấu hiệu X có luật phân phối là $X \sim N\left(\mu,\sigma^2\right)$ và $\left(X_1, X_2, ..., X_n\right)$ là một mẫu ngẫu nhiên cỡ n.

$$\frac{n-1}{\sigma^2}S^{*2} \sim \chi^2(n-1)$$

BÀI TẬP

5.1. Đo lượng huyết tương của 7 người mạnh khoẻ ta có:

2.86 3.37 2.76 2.62 3.49 3.05 3.12

Hãy xác định các đặc trưng trung bình và phương sai mẫu trên.

5.2. Theo dõi số xe gắn máy bán ra trong một tuần ở 45 đại lý, người ta thu được kết quả như sau:

Số xe bán trong 1 tuần	1	2	3	4	5	6
Số đại lý	15	12	9	5	3	1

Tính trung bình và phương sai của mẫu trên.

5.3. Theo dõi 336 trường hợp tàu cập cảng, người ta quan tâm đến khoảng thời gian X (giờ) giữa hai lần tàu vào cảng và thu được bảng số liệu sau:

Thời gian X (giờ)	Số trường hợp
4 _ 12	143
12 _ 20	75
20 _ 28	53
28 _ 36	27
36 _ 44	14
44 _ 52	9
52 _ 60	5
60 _ 68	4
68 _ 76	3
76 _ 80	3

Hãy tính trung bình và phương sai của mẫu trên.

Chương 6
LÝ THUYẾT ƯỚC LƯỢNG

Ước lượng là một bài toán thường gặp trong thực tế. Chương này sẽ cung cấp cho sinh viên các kiến thức cơ bản về lý thuyết ước lượng để từ đó chủ động hơn khi đặt bài toán ước lượng khi làm việc cũng như trong học tập.

Yêu cầu sinh viên phải nắm được các bài toán ước lượng cơ bản về tỉ lệ, trung bình, phương sai cùng với các bài toán liên quan như: xác định cỡ mẫu, độ tin cậy, v.v…

6.1. KHÁI NIỆM

- Ước lượng là phỏng đoán giá trị chưa biết dựa vào quan sát.
- Có hai hình thức ước lượng đó là
  + *Ước lượng điểm*: Giá trị ước lượng cho bởi một số cụ thể. Chẳng hạn, ta phỏng đoán một người A nào đó cao khoảng 1.65m.
  + *Ước lượng khoảng*: Giá trị ước lượng cho bởi một khoảng (a,b). Chẳng hạn, ta phỏng đoán một người A nào đó cao khoảng từ 1.65 m đến 1.7 m.

6.2. ƯỚC LƯỢNG ĐIỂM

$\hat{a}$ được gọi là ước lượng điểm của tham số a chưa biết nếu ta coi như

$$a \approx \hat{a}$$

6.2.1. Ước lượng không chệch

Thống kê T gọi là ước lượng không chệch của tham số θ nếu như $E(T) = \theta$.

Ví dụ3f 6. Cho dấu hiệu $X \sim N(\mu, \sigma^2)$, giả sử $(X_1, X_2, ..., X_n)$ là một mẫu ngẫu nhiên cho dấu hiệu X. Ta sẽ chứng minh rằng 2 thống kê T_1 và T_2 đều là các ước lượng không chệch cho μ với:

$$T_1 = \frac{X_1 + X_2}{2} \quad ; \quad T_2 = \frac{X_1 + X_2 + X_3}{3}$$

Thật vậy, do định nghĩa mẫu ta có: $X_1, X_2, ..., X_n$ độc lập và có phân phối là $N\left(\mu, \sigma^2\right)$. Từ đó có $E(X_1) = \ldots = E(X_n) = \mu$ và $D\left(X_1\right) = \ldots = D\left(X_n\right) = \sigma^2$.

Nên $$E\left(T_1\right) = E\left(\frac{x_1 + x_2}{2}\right)$$

$$= \frac{1}{2}\left(E\left(X_1\right) + E\left(x_2\right)\right) = \frac{1}{2}(\mu + \mu) = \mu$$

Tương tự : $E\left(T_2\right) = \mu$.

Cả T_1 và T_2 đều là ước lượng không chệch cho μ, muốn biết dùng thống kê nào tốt hơn, ta xét tiếp đến phương sai của chúng: một cách "hợp lý", *thống kê nào có phương sai nhỏ hơn là thống kê tốt hơn*.

$$D\left(T_1\right) = D\left(\frac{X_1 + X_2}{2}\right) = \frac{1}{4}\left(DX_1 + DX_2\right) \overset{\text{đl}}{=} \frac{1}{4}(\sigma^2 + \sigma^2) = \frac{\sigma^2}{2}$$

Tương tự: $D\left(T_2\right) = \dfrac{\sigma^2}{3}$

Thấy ngay là: $D\left(T_2\right) \leq D\left(T_1\right)$

Vậy nếu dùng thống kê T_2 thì tốt hơn T_1. Điều này cũng có nghĩa là: *Nếu quan sát càng nhiều thì ước lượng càng tốt.*

Người ta chứng minh được các kết quả sau:

- Trung bình mẫu $\overline{X}$ là ước lượng không chệch cho trung bình μ của đám đông.

- Phương sai mẫu hiệu chỉnh S^{*2} là ước lượng không chệch cho phương sai σ^2 của đám đông (Điều này lý giải từ "hiệu chỉnh" trong phương sai mẫu).

- Tỉ lệ mẫu F_n là ước lượng không chệch cho tỉ lệ p (các phần tử có tính chất A) của đám đông.

6.2.2. Ước lượng hiệu quả

Ước lượng hiệu quả của một tham số θ là một ước lượng không chệch cho θ và có phương sai bé nhất trong các ước lượng không chệch.

6.2.3. Ước lượng vững

Ta nói thống kê T là một ước lượng vững của tham số θ nếu như T "hầu chắc chắn" hội tụ về θ, theo nghĩa: xác suất để T lệch khỏi θ bằng 0. Người ta gọi sự hội tụ này là *hội tụ theo nghĩa xác suất*.

Người ta chứng minh được các tính chất cho trong bảng sau:

Bảng tóm tắt các ước lượng điểm thường dùng:

Tham số cần ƯL	ƯL điểm	Công thức	Tính chất
Trung bình μ	$\hat{\mu}$	$\hat{\mu} \approx \overline{X}$	Không chệch, vững, hiệu quả
Phương sai σ^2	$\hat{\sigma}^2$	$\hat{\sigma}^2 \approx S^{*2}$	Không chệch, vững
Tỉ lệ p	$\hat{p}$	$\hat{p} \approx F_n$	Không chệch, hiệu quả, vững

Bảng 6.1

Trong thực hành, người ta thường dùng:

- $\overline{x}$ để dự đoán μ
- σ^2_{n-1} để dự đoán σ^2
- f_n để dự đoán p.

Ví dụ3f 6.

Với bộ số liệu trong ví dụ 5.1, nếu yêu cầu dự đoán trọng lượng trung bình của heo 3 tháng tuổi cùng với phương sai của nó và tỉ lệ heo đạt tiêu chuẩn thì ta trả lời như sau:

Dự đoán trọng lượng trung bình của heo 3 tháng tuổi là 40.96 kg với phương sai dự đoán là 9.73 (kg^2).

Dự đoán tỉ lệ heo đạt tiêu chuẩn là 82.22%.

Ví dụ3f 6.

Nghiên cứu trọng lượng của một giống vịt mới ta có kết quả:

Cân nặng (kg)	1.25	1.5	1.75	2	2.25	2.5	2.75	3
Số con	2	6	24	35	39	24	14	6

a. Hãy ước lượng trọng lượng trung bình của giống vịt mới và phương sai của trọng lượng đó.

b. Giả sử vịt có trọng lượng dưới 1.6 kg là vịt loại II. Hãy ước lượng tỉ lệ vịt loại II.

Giải:

a. Ta tính được $\overline{x} = 2.185$ (kg) và $\sigma^2_{n-1} = 0.151$ (kg^2) nên dự đoán trọng lượng trung bình của giống vịt là 2.185 với phương sai là 0.151.

b. Rõ ràng tỉ lệ vịt loại II của mẫu là

$$f_n \frac{2+6}{150} \approx 0.0533 = 5.33\%$$

Ta dự đoán tỉ lệ vịt loại II là 5.33%.

Ước lượng điểm có ưu điểm là cho một giá trị có thể dùng so sánh, tính toán được nhưng nhược điểm là không thể biết được độ tin cậy cũng như xác suất để ước lượng đó chính xác tới mức nào.

6.3. ƯỚC LƯỢNG KHOẢNG

6.3.1. Định nghĩa

Tham số θ có khoảng ước lượng là (θ_1, θ_2) ở độ tin cậy $\gamma = 1 - \alpha$ nếu như

$$P(\theta_1 < \theta < \theta_2) = \gamma \quad (6.1)$$

trong đó:

$\theta_1 - \theta_2$ được gọi là độ dài của khoảng ước lượng ở độ tin cậy γ (hay khoảng tin cậy γ) cho θ.

$\frac{\theta_2 - \theta_1}{2}$ thường được coi như là **sai số**, hay **độ chính xác**, của khoảng ước lượng.

Thông thường ta xét độ tin cậy γ khá lớn: 90%, 95%, 99%,…

6.3.2. Khoảng tin cậy của các tham số thường gặp

6.3.2.1. Khoảng ước lượng cho tỉ lệ p của đám đông

Giả sử đám đông có tỉ lệ p các phần tử có tính chất A nào đó chưa biết, khi đó khoảng tin cậy γ của p là

$$\left(f_n-\varepsilon, f_n+\varepsilon\right) \tag{6.2}$$

Trong đó

$$\varepsilon = t_{\gamma/2}\sqrt{\frac{\hat{p}\left(1-\hat{p}\right)}{n}} \tag{6.3}$$

Với $t_{\gamma/2}$ tìm trong bảng phân phối $N\left(0,1\right)$ sao cho $\phi\left(t_{\gamma/2}\right)=\frac{\gamma}{2}$, n là cỡ mẫu, còn $\hat{p}=f_n$ là tỉ lệ các phần tử có tính chất A của mẫu.

Thật vậy, khi n khá lớn ta có $F_n \sim N\left(p;\frac{pq}{n}\right)$. Do đó với γ cho trước, ta tìm được $t_{\gamma/2}$ trong bảng hàm Laplace sao cho

$$P\left(-t_{\gamma/2}<\frac{F_n-p}{\sqrt{\frac{pq}{n}}}<t_{\gamma/2}\right)=\gamma$$

$$\Leftrightarrow P\left(F_n-t_{\gamma/2}\sqrt{\frac{pq}{n}}<p<F_n+t_{\gamma/2}\sqrt{\frac{pq}{n}}\right)=\gamma$$

Để đơn giản, thay p ở hai đầu mút bởi ước lượng điểm của nó là: $\hat{p}=F_n$. Từ đó ta có các công thức (6.2) và (6.3).

Ghi chú: Công thức trên đòi hỏi n phải khá lớn. Vì thế, trong tài liệu ta luôn giả thiết rằng n đủ lớn để sử dụng được các công thức trên. Sinh viên tham thêm trường hợp cỡ mẫu n nhỏ trong tài liệu [10].

Ví dụ3f 6. Phỏng vấn 400 người ở một khu vực thấy có 240 người ủng hộ dự luật A. Với độ tin cậy 95%, hãy ước lượng tỉ lệ người ủng hộ dự luật A trong khu vực.

Giải:

Theo đề, tỉ lệ người ủng hộ dự luật A của mẫu là

$$f_n = 240/400 = 0.6$$

Ở độ tin cậy $\gamma = 95\%$ suy ra $t_{\gamma/2} = 1.96$

và độ chính xác:

$$\varepsilon = 1.96\,\varepsilon = 1.96\sqrt{\frac{0.6(1-0.6)}{400}} \approx 0.048$$

Từ đó thay vào công thức ước lượng (6.2) ta được:

$$\left(f_n - \varepsilon; f_n + \varepsilon\right) = \left(0.5522; 0.6478\right)$$

Vậy dự đoán tỉ lệ người ủng hộ dự luật A là từ 55.22% đến 64.78% ở độ tin cậy 95%.

♣ Xác định cỡ mẫu trong bài toán ước lượng tỉ lệ

Bài toán: Muốn ước lượng tỉ lệ p các phần tử có tích chất A với sai số không quá ε_0 ở độ tin cậy γ cho trước thì phải quan sát ít nhất bao nhiêu phần tử.

Giải: Vì ε là sai số của ước lượng tỉ lệ nên ta phải có

$$\varepsilon \le \varepsilon_0$$

Thay (6.3) vào bất đẳng thức trên cho

$$t_{\gamma/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} \le \varepsilon_0$$

Bình phương hai vế của bất đẳng thức trên và thực hiện biến đổi đơn giản ta được

$$n \ge \left(t_{\gamma/2}\right)^2 \frac{\hat{p}\hat{q}}{\varepsilon_0^2} \qquad (6.4)$$

Trong trường hợp chưa có mẫu thăm dò, tức là chưa có $\widehat{p} = f_n$ *thì ta xây dựng công thức cỡ mẫu như sau: Do hàm số* $x(1-x) \le \frac{1}{4}$, $\forall x$ *(sinh viên tự chứng minh) nên công thức cỡ mẫu trong trường hợp này là:*

$$n \ge \left(t_{\gamma/2}\right)^2 \frac{1}{4\varepsilon_0^2} \qquad (6.5)$$

Nhận xét:

 Số phần tử quan sát phải là số nguyên, do đó, Nếu dùng các công thức trên, ta được các giá trị không nguyên thì phải đem phần nguyên của nó cộng với 1.

 Do $n \ge \frac{\left(t_{\gamma/2}\right)^2}{4\varepsilon_0^2} \ge \left(t_{\gamma/2}\right)^2 \frac{\hat{p}(1-\hat{p})}{\varepsilon_0^2}$ nên ta chỉ dùng công thức (6.5) khi không có mẫu thăm dò, tức là khi không thể dùng được công thức (6.4).

 Trên ý tưởng này, sinh viên tự đặt bài toán tìm độ tin cậy khi biết cỡ mẫu và sai số của ước lượng.

Ví dụ3f 6. Kiểm tra ngẫu nhiên 130 sản phẩm của một kho hàng thấy có 23 phế phẩm. Hãy ước lượng tỉ lệ phế phẩm của kho hàng trên với độ tin cậy 95%. Nếu muốn độ chính xác đạt 3% với độ tin cậy trên thì phải quan sát thêm ít nhất bao nhiêu sản phẩm nữa?

Giải: Theo đề, ta có $n = 130$ và tỉ lệ phế phẩm của mẫu là:

$$f_n = \widehat{p} = 23/130 \approx 0.177$$

Với độ tin cậy $\gamma = 95\%$, tra bảng hàm Laplace ta được $t_{\gamma/2} = 1.96$. Thay vào công thức tính độ chính xác, ta được:

$$\varepsilon = t_{\gamma/2}\sqrt{\frac{f_n\left(1-f_n\right)}{n}}$$
$$= 1.96\sqrt{\frac{0.177\left(1-0.177\right)}{130}} \approx 0.067$$

Vậy khoảng tin cậy 95% cho tỉ lệ phế phẩm trong kho hàng là:

$$\left(0.177-0.067; 0.177+0.067\right) = \left(0.11; 0.244\right)$$

Hay: $(11\%; 24.4\%)$

Áp dụng công thức xác định cỡ mẫu cho trường hợp có quan sát (có mẫu thăm dò), ta có:

$$n \geq \left(t_{\gamma/2}\right)^2 \frac{\hat{p}\left(1-\hat{p}\right)}{\varepsilon_0^2}$$

$$= \left(1.96\right)^2 \frac{0.177\left(1-0.177\right)}{\left(0.03\right)^2} \approx 621.79$$

Vậy cần quan sát thêm ít nhất: $622 - 130 = 492$ sản phẩm nữa.

Ví dụ3f 6. Để có thể ước lượng được số chim cánh cụt ở một vùng A, người ta làm như sau: Bắt 1000 con chim cánh cụt, đeo vòng vào chân và thả trở lại vùng đó. Sau một thời gian, người ta bắt ngẫu nhiên lại 200 con và thấy có 78 con có đeo vòng chân. Dự đoán tỉ lệ chim có đeo có đeo vòng ở vùng A rồi suy ra số chim cánh cụt của vùng đó với độ tin cậy 90%.

Giải: Rõ ràng, nếu ước lượng được tỉ lệ chim có đeo vòng (tức là xác suất bắt được chim có đeo vòng) thì ta hoàn toàn có thể suy ra số chim của vùng đó, vì

$$P(\text{Bắt được 1 con chim có đeo vòng}) = \frac{\text{Số chim có đeo vòng}}{\text{Tổng số chim của vùng}}$$

Theo đề, ta có $n = 200$, $f_n = \dfrac{78}{200} = 0.39$

Với $\gamma = 90\% \Rightarrow t_{\gamma/2} = 1.65$

$$\Rightarrow \varepsilon = 1.65 \frac{\sqrt{\dfrac{78}{200}\left(1-\dfrac{78}{200}\right)}}{\sqrt{200}} \approx 0.057$$

Suy ra khoảng tin cậy 90% cho tỉ lệ chim có đeo vòng là $(0.333; 0.447)$

Suy ra khoảng tin cậy 90% cho số chim của vùng A là:

$$\left(\frac{1000}{0.447}; \frac{1000}{0.333}\right) = (2238 \text{ con}; 3004 \text{ con})$$

6.3.2.2. Khoảng tin cậy cho trung bình μ của đám đông

Giả thiết dấu hiệu $X \sim N\left(\mu, \sigma^2\right)$ có trung bình μ chưa biết. Khoảng tin cậy γ cho trung bình đám đông μ là:

$$\left(\overline{x} - \varepsilon; \overline{x} + \varepsilon\right) \tag{6.6}$$

Trong đó ε được xác định như sau:

Trường hợp 1: $n \geq 30$, phương sai σ^2 đã biết, khi đó :

$$\varepsilon = t_{\gamma/2} \frac{\sigma}{\sqrt{n}} \tag{6.7}$$

với $t_{\gamma/2}$ tìm trong bảng hàm Laplace sao cho $\phi\left(t_{\gamma/2}\right) = \frac{\gamma}{2}$

Trường hợp 2: $n \geq 30$, phương sai σ^2 chưa biết,

$$\varepsilon = t_{\gamma/2} \frac{\sigma_{n-1}}{\sqrt{n}} \tag{6.8}$$

với $t_{\gamma/2}$ tìm trong bảng hàm Laplace sao cho $\phi\left(t_{\gamma/2}\right) = \frac{\gamma}{2}$

Trường hợp 3: $n < 30$, phương sai σ^2 đã biết: ta làm như trường hợp 1.

Trường hợp 4: $n < 30$, phương sai σ^2 chưa biết,

$$\varepsilon = t_{n-1,\alpha} \frac{\sigma_{n-1}}{\sqrt{n}} \tag{6.9}$$

với $t_{n-1,\alpha}$ tìm trong bảng phân phối Student $t\left(n-1\right)$, mức α hai phía, nghĩa là với $X \sim t\left(n-1\right)$ thì $P\left[\left|X\right| > t_{(n-1,\alpha)}\right] = \alpha$

Ví dụ3f 6. Một phân xưởng muốn ước lượng thời gian trung bình để sản xuất 1 ram giấy. Giả sử lượng thời gian đó tuân theo luật phân phối chuẩn với $\sigma = 0.3$ phút. Trên một tập mẫu gồm 36 ram, thời gian trung bình tính được là 1.2 phút/ram. Tính khoảng tin cậy 95% cho thời gian sản xuất trung bình trên.

Giải: Theo đề bài ta có:

$n = 36 > 30$, $\overline{x} = 1.2$; $\sigma = 0.3$ và $\gamma = 95\%$ (trường hợp 1)

Theo bảng hàm Laplace :

$$\phi\left(t_{\gamma/2}\right) = \gamma / 2 = 0.475 \Rightarrow t_{\gamma/2} = 1.96$$

$$\varepsilon = 1.96\frac{0.3}{\sqrt{36}} = 0.098$$

$$\Rightarrow \left(\overline{x} - \varepsilon; \overline{x} + \varepsilon\right) = \left(1.2 - 0.098; 1.2 + 0.098\right)$$

$$= \left(1.102; 1.298\right)$$

Vậy khoảng tin cậy 95% cho thời gian sản xuất trung bình là (1.102; 1.298) (phút/ram)

Ví dụ3f 6. Kết quả thống kê về trọng lượng của số bột dùng làm bánh của 14 ngày đối với một lò bánh, ta được kết quả: $\overline{x} = 17.3$ kg; $\sigma_{n-1} = 4.5$ kg. Hãy ước lượng trọng lượng bột trung bình ở độ tin cậy 99%.

Giải: Ta có $n = 14 < 30$, $\overline{x} = 17.3$; $\sigma_{n-1} = 4.5$ (trường hợp 4)

Tra bảng phân phối Student suy ra $t_{n-1,\alpha} = t_{13,0.01} = 3.012$.

$$\varepsilon = 3.012\frac{4.5}{\sqrt{14}}$$

$$\Rightarrow \left(\overline{x} - \varepsilon; \overline{x} + \varepsilon\right) = \left(136.77; 209.23\right)$$

Vậy khoảng tin cậy 99% cho trọng lượng bột trung bình là (136.77; 209.23) (kg).

Ví dụ3f 6. Quan sát mức xăng hao phí của 25 xe máy thuộc cùng một loại xe, chạy trên cùng một quãng đường, người ta thu được kết quả:

Mức xăng (l)	1.9-2.1	2.1-2.3	2.3-2.5	2.5-2.6
Số xe	5	9	8	3

Hãy tìm khoảng tin cậy 99% cho mức xăng hao phí trung bình của loại xe trên.

Giải: Ta tính toán trên bộ số liệu:

Mức xăng (l)	2.0	2.2	2.4	2.55
Số xe	5	9	8	3

Thu được kết quả: $n = 25$; $\sum x = 56.65$; $\sum x^2 = 129.1475$ nên

$$\overline{x} = \frac{56.65}{25} = 2.266;$$

$$\sigma_{n-1} = \sqrt{\frac{25}{24}\left(\frac{129.1475}{25} - 2.266^2\right)} \approx 0.18$$

Vì $n < 30$, phương sai chưa biết (trường hợp 4) nên tra bảng phân phối Student ta có:

$$t_{n-1;\alpha} = t_{24;0.01} = 2.797$$

$$\Rightarrow \varepsilon = 2.797\frac{0.18}{\sqrt{25}} \approx 0.1$$

Vậy khoảng tin cậy 99% cho mức xăng hao phí trung bình là:

$$\left(\overline{x} - \varepsilon; \overline{x} + \varepsilon\right) = \text{(2.166 lít ; 2.366 lít)}$$

♣ **Công thức xác định cỡ mẫu**

Bài toán: Muốn sai số của ước lượng trung bình không quá ε_0 ở độ tin cậy γ cho trước thì cần quan sát ít nhất bao nhiêu phần tử.

Tương tự trong ước lượng tỉ lệ, ta sẽ có:

$$n \ge \frac{\left(t_{\gamma/2}\right)^2 \sigma^2}{\varepsilon_0^2} \tag{6.10}$$

Hoặc nếu σ^2 chưa biết thì

$$n \ge \frac{\left(t_{\gamma/2}\right)^2 \sigma_{n-1}^2}{\varepsilon_0^2} \tag{6.11}$$

Ví dụ3f 6. Kiểm tra tuổi thọ tính bằng giờ của 50 bóng đèn do một nhà máy sản xuất ra, người ta có kết quả như sau:

Tuổi thọ (x_i)	3300	3500	3600	4000
Số bóng đèn (n_i)	10	20	12	8

Giả thiết rằng tuổi thọ của bóng đèn có phân phối chuẩn.

a) Hãy ước lượng tuổi thọ trung bình của loại bóng đèn do nhà máy trên sản xuất với độ tin cậy 95%.

b) Nếu yêu cầu ước lượng phải đạt độ chính xác 50 giờ và độ tin cậy 95% thì cần phải kiểm tra ít nhất bao nhiêu bóng đèn?

Giải :

Tính toán trên mẫu đã cho, ta có kết quả sau:

$$n = 50\ ;\ \overline{x} = 3564\ ,\ \sigma_{n-1} \approx 217.37\ .$$

a) Vì phương sai chưa biết và $n > 30$ nên tính được $\varepsilon = 60.25$ nên khoảng tin cậy 95% cho tuổi thọ trung bình của loại bóng đèn trên là (3503.75; 3624.25).

b) Áp dụng công thức xác định mẫu, ta có:

$$n \geq \frac{\left(t_{\gamma/2}\right)^2 \sigma_{n-1}}{{\varepsilon_0}^2} = \left(1.96\frac{217.37}{50}\right)^2 \approx 73 \text{ (bóng)}$$

Vậy phải quan sát ít nhất 73 bóng đèn.

Ví dụ3f 6. Cân ngẫu nhiên 45 con heo 3 tháng tuổi trong một trại chăn nuôi ta được kết quả sau:

X_i	35	37	39	41	43	45	47
n_i	2	6	10	11	8	5	3

Giả sử trọng lượng X(kg) tuân theo luật phân phối chuẩn.

a) Hãy ước lượng trọng lượng trung bình của các con heo 3 tháng tuổi trong trại trên.

b) Hãy ước lượng trọng lượng trung bình của các con heo 3 tháng tuổi trong trại trên với độ tin cậy 95%.

c) Giả sử heo có trọng lượng ≥ 38kg là heo đạt tiêu chuẩn. Hãy tìm ước lượng không chệch và khoảng tin cậy 90% cho tỉ lệ heo đạt tiêu chuẩn.

Giải: Từ bộ số liệu ta tính được

$$n = 45\ ;\ \sum x_i = 1843\ ;\ \sum x_i^2 = 75909\ .$$

Vậy trung bình mẫu là $\overline{x} = \dfrac{1843}{45} \approx 40.96$.

Phương sai mẫu là $\sigma_n^2 = \dfrac{75909}{45} - (40.96)^2 \approx 9.51$.

Phương sai mẫu hiệu chỉnh $\sigma_{n-1}^2 = \dfrac{45}{44}9.51 \approx 9.73$.

a) Dự đoán trọng lượng trung bình của heo 3 tháng tuổi là 40.96 kg

b) Đây là bài toán ước lượng khoảng cho trung bình đám đông.

Vì phương sai chưa biết và $n = 45 > 30$ (trường hợp 2) nên tra bảng hàm Laplace ta được $t_{\gamma/2} = t_{0.475} = 1.96$.

Suy ra $\varepsilon = 1.96.\dfrac{\sqrt{9.73}}{\sqrt{45}} \approx 0.91$

$\Rightarrow \left(\overline{x} - \varepsilon; \overline{x} + \varepsilon\right) =$ (40,5 kg; 41,87 kg).

Vậy khoảng tin cậy 95% cho trọng lượng trung bình của heo là (40,5 kg; 41,87 kg).

c) Tỉ lệ heo đạt tiêu chuẩn của mẫu là

$$f_n = \frac{10+11+8+5+3}{45} = \frac{37}{45} \approx 0.8222$$

Vậy ước lượng không chệch cho tỉ lệ heo đạt tiêu chuẩn là 82.22%

* Công thức khoảng tin cậy 90% cho tỉ lệ là $\left(f_n - \varepsilon; f_n + \varepsilon\right)$.

Tra bảng hàm Laplace ta được $t_{0,45} = 1.65$. Vì vậy

$$\varepsilon = 1.65\sqrt{\frac{0.8222\left(1-0.8222\right)}{45}} \approx 0.094.$$

Vậy khoảng tin cậy 90% cho tỉ lệ heo đạt tiêu chuẩn là (72.82%; 91.62%).

6.3.2.3. Khoảng tin cậy γ cho phương sai đám đông σ^2

Với giả thiết đặc điểm X có phân phối chuẩn với phương sai chưa biết, khoảng tin cậy γ cho phương sai σ^2 của đám đông là:

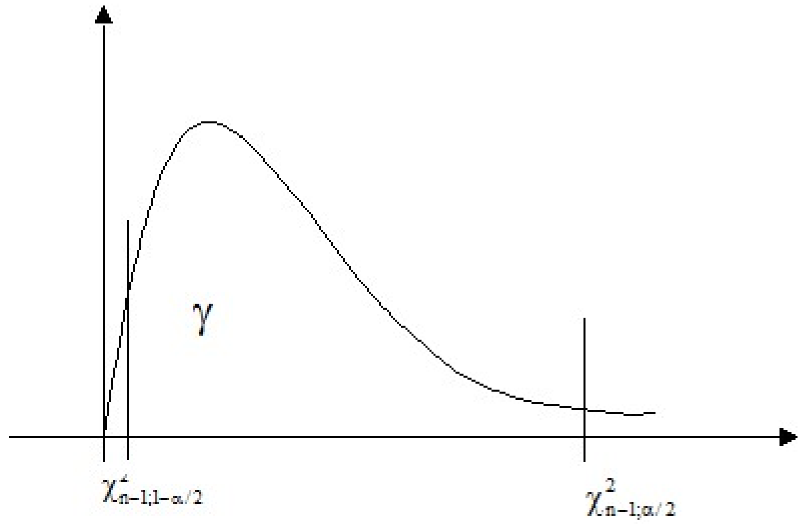
$$\left(\frac{n-1}{\chi^2_{n-1;\alpha/2}}\sigma^2_{n-1}; \frac{n-1}{\chi^2_{n-1;1-\alpha/2}}\sigma^2_{n-1}\right) \tag{6.12}$$

Với $\chi^2_{n-1;\alpha/2}$ và $\chi^2_{n-1;1-\alpha/2}$ tra trong bảng phân phối χ^2.

Ví dụ3f 6. Độ dày của bản kim loại tuân theo luật phân phối chuẩn. Đo 10 bản kim loại người ta tính được phương sai hiệu chỉnh của mẫu là 0.1367. Hãy xác định khoảng tin cậy 95% cho phương sai của độ dày đó.

Giải:

Với độ tin cậy 95%, tra bảng phân phối "khi bình phương" ta được:



Hình 6.1

$$\chi^2_{n-1,\alpha/2} = \chi^2_{9;0.025} = 19.023$$

$$\chi^2_{n-1,1-\alpha/2} = \chi^2_{9;0.975} = 2.7$$

Vậy khoảng tin cậy 95% cho phương sai là:

$$\left(\frac{9}{19.023}0.1367; \frac{9}{2.7}0.1367\right) \approx (0.064; 0.456).$$

BÀI TẬP

5.4. Đo độ dài của một loại trục xe, ta có kết quả:

Nhóm	18.4-18.6	18.6-18.8	18.8-19	19-19.2	19.2-19.4	19.4-19.6	19.6-19.8
n_i	1	6	22	41	19	7	4

Hãy ước lượng độ dài trung bình và phương sai của trục xe.

5.5. Đo sức bền chịu lực của một loại ống công nghiệp người ta thu được bộ số liệu sau: 4500 6500 5200 4800 4900 5125 6200 5375. Từ kinh nghiệm nghề nghiệp người ta cũng biết rằng sức bền đó có phân phối chuẩn với độ lệch chuẩn $\sigma = 300$. Hãy xây dựng khoảng tin cậy 95% cho sức bền trung bình của loại ống trên.

5.6. Trước bầu cử Tổng thống người ta phỏng vấn ngẫu nhiên 1800 cử tri thì thấy 1180 người ủng hộ một ứng cử viên A. Với độ tin cậy 95%, hỏi cử tri đó thu được tối thiểu bao nhiêu phần trăm số phiếu bầu ?

5.7. a) Muốn ước lượng tỉ lệ bị bệnh sốt rét ở một vùng A với sai số không quá 3% ở độ tin cậy $\gamma = 95\%$ thì phải quan sát ít nhất bao nhiêu người ?

b) Giả sử quan sát 200 người thấy có 24 người bị bệnh sốt rét. Hãy ước lượng tỉ lệ bị bệnh sốt rét ở vùng A ở độ tin cậy $\gamma = 97\%$. Nếu muốn sai số ước lượng không quá 3% ở độ tin cậy $\gamma = 95\%$ thì phải quan sát ít nhất bao nhiêu người?

5.8. Biết tỉ lệ nẩy mầm của một loại hạt giống là 0.9. Với độ tin cậy 0.95 nếu ta muốn độ dài khoảng tin cậy của tỉ lệ nẩy mầm không vượt quá 0.02 thì cần phải gieo ít nhất bao nhiêu hạt ?

5.9. Để ước lượng xác suất mắc bệnh A với độ tin cậy 95% và sai số không vượt quá 2% thì cần phải khám ít nhất bao nhiêu người, biết rằng tỷ lệ mắc bệnh A thực nghiệm đã cho bằng 0.8.

5.10. Muốn biết trong một hồ nước có bao nhiêu cá, người ta bắt lên 1000 con, đánh dấu xong lại thả xuống hồ. Sau một thời gian, người ta bắt lên 200 con và thấy có 30 con cá có đánh dấu của lần bắt trước. Dựa vào kết quả đó, hãy ước lượng số cá trong hồ với độ tin cậy 95%.

5.11. Để có thể dự đoán được số lượng cò thường nghỉ tại vườn nhà mình, ông chủ vườn bắt 89 con, đeo khoen cho chúng rồi thả đi. Sau một thời gian, ông bắt ngẫu nhiên được 120 con và thấy có 7 con có đeo vòng khoen. Hãy dự đoán số cò giúp ông chủ vườn ở độ tin cậy 99%. Tìm hiểu lý do xem tại sao khoảng dự đoán cho số cò lại lớn như vậy.

5.12. Sản lượng ngày của một phân xưởng là biến ngẫu nhiên tuân theo luật chuẩn. Kết quả thống kê của 9 ngày cho ta

27 26 21 28 25 30 26 23 26

Hãy xác định các khoảng tin cậy 90% cho sản lượng trung bình và cho phương sai tương ứng ?

5.13. Trên tập mẫu gồm 100 số liệu người ta tính được $\overline{x} = 0.1$; $\sigma_{n-1} = 0.014$. Xác định khoảng tin cậy 99% cho giá trị trung bình thật.

5.14. Cân thử 100 quả trứng ta có bộ số liệu sau:

Trọng lượng(g)	32	33	34	35	36	37	38	39	40
Số quả	2	3	15	28	30	8	5	5	4

a) Hãy ước lượng trọng lượng trung bình các quả trứng ở độ tin cậy 90%.

b) Trứng có khối lượng dưới 34g được coi là trứng loại II. Tìm ước lượng không chệch cho tỉ lệ trứng loại II và khoảng tin cậy 95% của tỉ lệ đó.

Chương 7
KIỂM ĐỊNH GIẢ THIẾT

Kiểm định cùng với ước lượng là hai trong số các bài toán cơ bản của thống kế toán học. Học xong chương này sinh viên phải giải được các bài toán kiểm định về: tỉ lệ, trung bình của một tính chất nào đó, và tính độc lập giữa 2 thuộc tính.

7.1. KHÁI NIỆM

Kiểm định giả thiết là một bài toán quan trọng trong đời sống cũng như trong thống kê toán. Ta thường gặp một cặp giả thiết đối nghịch nhau; bằng quan sát và khả năng của mình, ta phải xác định xem giả thiết nào là đúng.

- Giả thiết thống kê là các giả thiết về μ, σ^2, p, … của đám đông đang xét.
- Nội dung bài toán kiểm định: Cho hai giả thiết H_0 và H_1 (thường đối nghịch nhau) Dựa vào các số liệu thu được, ta phải ra quyết định xem giả thiết H_0 đúng hay sai. Giả thiết H_1 đối nghịch với giả thiết H_0 gọi là đối thiết của H_0. Việc đưa ra quyết định chấp nhận hay bác bỏ một giả thiết thống kê gọi là làm kiểm định (hay kiểm định thống kê)

Tất nhiên, khi chấp nhận H_0 thì cũng có nghĩa là bác bỏ H_1 và ngược lại.

Ví dụ3f 7. Khi ta cảm thấy mệt mỏi, ta nghi rằng "mình bị bệnh" – đây là giả thiết H_0 (giả thiết H_1 là "Mình không bị bệnh") – còn việc đi khám bệnh để xác định xem nghi ngờ này có đúng không chính là việc kiểm định giả thiết.

- Khi làm kiểm định, ta có thể mắc các sai lầm sau đây:
  - Sai lầm loại I: Bác bỏ một giả thiết đúng (bác H_0 khi H_0 đúng).
  - Sai lầm loại II: Chấp nhận một giả thiết sai (Nhận H_0 khi H_0 sai).

Kết luận / Thực tế	Chấp nhận H_0	Bác bỏ H_0
H_0 đúng	Kết luận đúng	Sai lầm I
H_0 sai	Sai lầm II	Kết luận đúng

Bảng 7.1: Sai lầm loại I và loại II

Ví dụ3f 7. Giả sử với một người đến khám bệnh, bác sĩ đặt giả thiết H_0: "Bệnh nhân này bị bệnh" (H_1: "Bệnh nhân này không bị bệnh") . Khi đó, sai lầm loại I là kết luận của bác sĩ rằng: Anh không bị bệnh trong khi thực sự là anh ta bị bệnh. Còn sai lầm loại II là kết luận: Anh bị bệnh trong khi thực sự là anh ta không bị bệnh.

Trong ví dụ trên, ta thấy rằng nếu H_0: "Tôi bị bệnh SARS" thì sai lầm I là rất tai hại vì khi đó, do không phòng bị nên bệnh sẽ lây nhiễm ra cộng đồng. (SARS- Bệnh viêm đường hô hấp cấp- là loại bệnh có khả năng lây nhiễm nhanh và tỉ lệ tử vong cao).

Tất nhiên, khi kiểm định một giả thiết, ta cần giảm thiểu tối đa xác suất phạm cả hai sai lầm- đây là điều mà trong thực tế ta không thể làm được vì khi xác suất phạm sai lầm I giảm thì xác suất phạm sai lầm II sẽ tăng và ngược lại.

Ví dụ3f 7. Giả sử ta có giả thiết H_o: "Tôi bị bệnh SARS". Vì tính chất đặc biệt nghiêm trọng của bệnh SARS nên ta có xu hướng sợ phạm phải sai lầm I (tức muốn xác suất phạm sai lầm I là bé nhất), từ đó sẽ xuất hiện xu hướng dễ dãi kết luận rằng trường hợp này bị bệnh SARS (Chẳng hạn, chỉ cần có thân nhiệt cao bất thường là cách ly ngay. Tất nhiên khi đó ta rất dễ cách ly một người không bị bệnh), nghĩa là dễ phạm phải sai lầm II, tức xác suất phạm sai lầm II tăng lên.

Nguyên tắc kiểm định:

Nguyên tắc chung của kiểm định giả thiết thống kê là dựa trên *nguyên lý xác suất nhỏ:* Một sự kiện có xác suất khá bé thì có thể coi như nó không xảy ra khi thực hiện phép thử tương ứng.

Ví dụ3f 7. Chẳng hạn như cho bạn biết ảnh của một người trong thành phố đông dân như Tp. HCM thì gần như chắc chắn bạn sẽ không thể gặp được người đó.

Qui ước: Sai lầm I là tai hại hơn và cần tránh trước. Vì vậy, với xác suất α nhỏ cho trước, ta cần ra quyết định sao cho:

$$P(\text{Phạm sai lầm I}) \le \alpha \tag{7.1}$$

α trong công thức (7.1) được gọi là *mức ý nghĩa của kiểm định*.

Để kiểm định giả thiết H_0, ta giả sử H_0 đúng và xây dựng một thống kê T_{qs} (qs=quan sát) dùng để đo sự khác biệt giữa lý thuyết và thực tế. Một cách hợp lý, nếu T_{qs} càng lớn thì ta càng nghiêng về phía bác giả thiết H_0. T_{qs} lớn như thế nào là đủ sẽ dựa vào mức α, còn khi T_{qs} nhỏ thì ta chỉ kết luận rằng chưa đủ điều kiện để bác giả thiết H_0, đây sẽ là trường hợp mà ta cần quan sát thêm (trường hợp nghi ngờ).

Ví dụ3f 7. Trong giả thiết về bệnh SARS ở trên, để hạn chế sai lầm loại 1, bác sĩ khám bệnh sẽ luôn nghi rằng H_0 đúng, rồi từ đó căn cứ vào những biểu hiện và những xét nghiệm (tức là mẫu), đem so sánh với những biểu hiện của bệnh (tức là xây dựng T_{qs}). Nếu thấy có khá nhiều biểu hiện không trùng (tức T_{qs} lớn) thì bác sĩ sẽ nghiên về phía kết luận rằng bệnh nhân này bị không bệnh SARS (tức là bác giả thiết H_0).

7.2. MỘT SỐ BÀI TOÁN KIỂM ĐỊNH THƯỜNG GẶP

Trong những bài toán so sánh sau đây ta chỉ xét những bài toán **kiểm định 2 phía**, rồi sau đó sử dụng ước lượng điểm để kết luận tiếp trong trường hợp bác giả thiết H_0 , nếu thấy cần thiết.

7.2.1. So sánh tỉ lệ *p* với một số

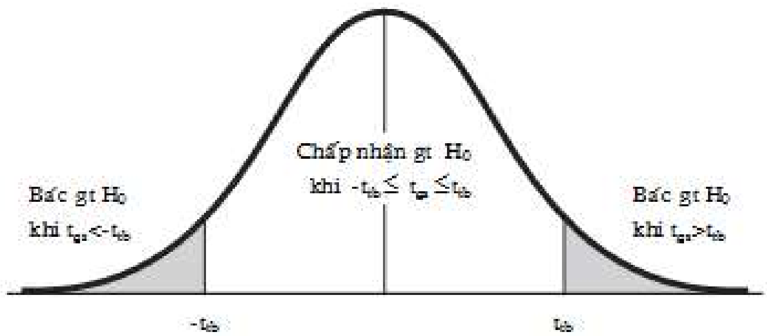
Đám đông có tỉ lệ p các phần tử có tính chất A chưa biết, dựa vào quan sát ta cần so sánh p với số p_0 đã cho ở mức ý nghĩa $\alpha = 1-\gamma$.

Do $F_n \sim N\left(p, \frac{pq}{n}\right)$ khi n khá lớn nên ta có cách giải sau:

Cách giải:

- Đặt giả thiết H_0: $p = p_0$, đối thiết H_1: $p \neq p_0$.
- Ở mức ý nghĩa α , ta tìm được số $t_{\text{tra bảng}} = t_{(1-\alpha)/2}$ trong bảng hàm Laplace sao cho $\phi(t_{(1-\alpha)/2}) = (1-\alpha)/2$.
- Tính tiêu chuẩn kiểm định:

$$t_{qs} = \frac{|f_n - p_0|}{\sqrt{\frac{p_0(1-p_0)}{n}}} \tag{7.2}$$



Hình 7.1. (t_{tb}= $t_{trbảng}$)

 Kiểm định:

Nếu $t_{qs} > t_{\text{tra bảng}}$ thì ta bác giả thiết H_0. Khi đó, để kết luận tiếp ta dựa vào ước lượng điểm như sau:

Nếu $f_n > p_0$ thì kết luận $p > p_0$.

Nếu $f_n < p_0$ thì kết luận $p < p_0$.

Nếu $t_{qs} \le t_{\text{tra bảng}}$ thì ta chưa đủ cơ sở để bác giả thiết H_0 nên ta tạm chấp nhận H_0 trong trường hợp này.

Ví dụ3f 7. Một tòa báo thanh niên thông báo có 25% học sinh phổ thông trung học là đọc giả thường xuyên. Một mẫu ngẫu nhiên gồm 200 học sinh được chọn cho thấy có 45 em đọc báo thường xuyên. Kiểm định tính chính xác của thông báo trên ở mức ý nghĩa 0.05.

Giải:

Thấy ngay rằng: Việc kết luận về tính chính xác của thông báo trên tương ứng với việc chấp nhận hay bác bỏ giả thiết H_0 sau:

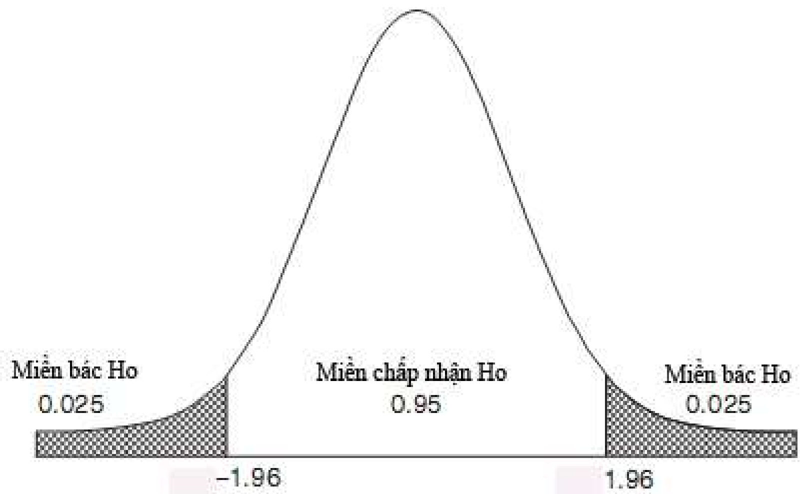
Đặt giả thiết H_0: $p = 25\%$, H_1: $p \neq 25\%$ trong đó p là tỉ lệ học sinh phổ thông là đọc giả thường xuyên của báo (p chưa biết).

Ở mức ý nghĩa $\alpha = 0.05$ ta tìm được $t_{\text{tra bảng}} = t_{\frac{1-\alpha}{2}} = 1.96$. (1)

Với tỉ lệ mẫu là $f_n = 45/200$.

Tiêu chuẩn kiểm định:

$$t_{qs} = \frac{|f_n - p_0|}{\sqrt{\dfrac{p_0(1-p_0)}{n}}} = \frac{\left|\dfrac{45}{200} - 0.25\right|}{\sqrt{\dfrac{0.25(1-0.25)}{200}}} \approx 0.806 \quad (2)$$



Hình 7.2.

Từ (1) và (2) suy ra $t_{qs} < t_{\text{tra bảng}}$ nên ta chưa đủ cơ sở để bác bỏ giả thiết H_0, nghĩa là thông báo trên có thể đúng, ở mức ý nghĩa 0.05.

7.2.2. So sánh 2 tỉ lệ.

Giả sử 2 đám đông 1 và 2 có tỉ lệ các phần tử có tính chất A nào đó là p_1, p_2 chưa biết. Dựa vào quan sát ta phải so sánh p_1 và p_2 với mức α cho trước.

Người ta chứng minh được rằng nếu đặt $D = p_1 - p_2$ thì khi n_1, n_2 khá lớn:

$$D \sim N\left(p_1 - p_2, \frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}\right) \quad (7.3)$$

Từ đó thấy rằng để so sánh hai tỉ lệ, ta chỉ cần kiểm định giả thiết “ $D = 0$ ” hay “ $p_1 = p_2$ ”.

Đây chính là bài toán ta xét ở trên. Vậy ta có cách giải cho bài toán này như sau:

Cách giải:

- Đặt giả thiết H0: $p_1 = p_2$, H1: $p_1 \neq p_2$
- Tra bảng hàm Laplace để được $t_{\text{tra bảng}} = t_{(1-\alpha)/2}$.
- Tính tiêu chuẩn kiểm định:

$$t_{qs} = \frac{\left| f_{n_1} - f_{n_2} \right|}{\sqrt{\overline{f}\left(1-\overline{f}\right)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \qquad (7.4)$$

Với n_1, f_{n_1}, và n_2, f_{n_2} lần lượt là cỡ mẫu và tỉ lệ mẫu của đám đông 1 và 2. Còn

$$\overline{f} = \frac{n_1.f_{n_1} + n_2.f_{n_2}}{n_1 + n_2} = \frac{m_1 + m_2}{n_1 + n_2} \qquad (7.5)$$

là tỉ lệ các phần tử có tính chất A chung của cả hai mẫu (m_1 ; m_2 là lần lượt là số phần tử có tính chất A của đám đông 1 và đám đông 2).

- Kiểm định:

  Nếu $t_{qs} > t_{\text{tra bảng}}$ thì ta bác giả thiết H0. Khi đó:

  Nếu $f_{n_1} > f_{n_2}$ thì kết luận $p_1 > p_2$.

  Nếu $f_{n_1} < f_{n_2}$ thì kết luận $p_1 < p_2$.

  Nếu $t_{qs} \leq t_{\text{tra bảng}}$ thì ta tạm chấp nhận giả thiết H0.

Ví dụ3f 7. Nghiên cứu về dự định đăng ký thi đại học về khối ngành kỹ thuật của học sinh phổ thông ở thành thị và nông thôn với câu hỏi: Em có dự định thi vào đại học khối ngành kỹ thuật hay không, ta có kết quả:

Nơi điều tra	Số HS trả lời **Có**	Số HS trả lời **Không**
Thành thị	240	300
Nông thôn	185	200

a) Hãy so sánh tỉ lệ học sinh dự định thi đại học về khối ngành kỹ thuật ở thành thị và nông thôn với mức ý nghĩa 0.05.

b) Có kết luận cho rằng tỉ lệ học sinh phổ thông có dự định thi đại học khối ngành kỹ thuật là 50%. Điều đó đúng không? (cho kết luận với mức ý nghĩa 0.05).

Giải:

Ở thành thị:

$$n_1 = 240 + 300 = 540\,;\; m_1 = 240 \Rightarrow f_{n_1} = 240/540 = 0.4444$$

Ở nông thôn:

$$n_2 = 185 + 200 = 385\,;\; m_2 = 185 \Rightarrow f_{n_2} = 185/385 = 0.4805$$

a) Ta có $\overline{f} = \dfrac{240+185}{540+385} = \dfrac{425}{925} \approx 0.4595$

Gọi p_1 và p_2 lần lượt là tỉ lệ học sinh dự định thi vào khối ngành kỹ thuật ở thành thị và nông thôn. Vậy ta cần kiểm định giả thiết H_0: $p_1 = p_2$

Với $\gamma = 1 - 0.05 \Rightarrow t_{\text{tra bảng}}=1.96$
(1)

Tiêu chuẩn kiểm định:

$$t_{qs} = \frac{|0.4444 - 0.4805|}{\sqrt{0.4595(1-0.4595)\left(\frac{1}{540}+\frac{1}{385}\right)}} \approx 1.1 \qquad (2)$$

Từ (1) và (2) $\Rightarrow t_{qs} < t_{\text{trabảng}}$. Vậy không đủ cơ sở để bác giả thiết H_0, nghĩa là tạm xem như tỉ lệ học sinh dự định thi vào khối ngành kỹ thuật ở thành thị và nông thôn là như nhau.

b) Theo đề bài thì ta cần kiểm định giả thiết H_0: $p = 50\%$ trong đó p là tỉ lệ học sinh phổ thông có dự định thi đại học khối ngành kỹ thuật.

Tương tự, ta có $t_{\text{trabảng}}=1.96$
(3)

Tiêu chuẩn kiểm định: $t_{qs} = \dfrac{|0.4595 - 0.5|}{\sqrt{\dfrac{0.5\times 0.5}{925}}} = 2.46 \qquad (4)$

Từ (3) và (4) $\Rightarrow t_{qs} > t_{\text{trabảng}}$. Vậy ta bác giả thiết H_0. Nghĩa là, ở mức ý nghĩa 0.05, kết luận trên sai.

7.2.3. So sánh trung bình μ với một số

Giả sử biến gốc $X \sim N\left(\mu,\sigma^2\right)$. Ta cần so sánh trung bình μ với μ_0 cho trước ở mức ý nghĩa α khi đã có mẫu quan sát. Từ phân phối của $\overline{X}$, ta có cách giải cho bài toán này như sau:

Cách giải:

Đặt giả thiết H_0: $\mu = \mu_0$, H_1: $\mu \neq \mu_0$

Ta chia trường hợp tương tự như các trường hợp đã xét trong ước lượng khoảng cho trung bình. Ở mức ý nghĩa α, ta xác định $t_{\text{tra bảng}}$ và tính tiêu chuẩn kiểm định t_{qs} như sau:

Trường hợp 1: $n \geq 30$, σ^2 đã biết

- Tra $t_{\text{tra bảng}} = t_{(1-\alpha)/2}$ trong bảng hàm Laplace sao cho

$$\phi\left(t_{(1-\alpha)/2}\right) = (1-\alpha)/2$$

- $$t_{qs} = \frac{\left|\overline{X} - \mu_0\right|}{\sigma}\sqrt{n} \qquad (7.6)$$

Trường hợp 2: $n \geq 30$, σ^2 chưa biết

- $t_{\text{tra bảng}} = t_{(1-\alpha)/2}$ (bảng hàm Laplace)

- $$t_{qs} = \frac{\left|\overline{X} - \mu_0\right|}{\sigma_{n-1}}\sqrt{n} \qquad (7.7)$$

Trường hợp 3: $n < 30$, σ^2 đã biết: ta làm như trường hợp 1.

Trường hợp 4: $n < 30$, σ^2 chưa biết

- Tra $t_{\text{tra bảng}} = t_{n-1,\alpha}$ trong bảng phân phối Student ở $n-1$ bậc tự do, mức α hai phía.

- $$t_{qs} = \frac{\left|\overline{X} - \mu_0\right|}{\sigma_{n-1}}\sqrt{n} \qquad (7.8)$$

Kiểm định:

Nếu $t_{qs} > t_{\text{tra bảng}}$ thì ta bác bỏ giả thiết H_0. Khi đó:

Nếu $\overline{X} > \mu_0$ thì kết luận $\mu > \mu_0$.

Nếu $\overline{X} < \mu_0$ thì kết luận $\mu < \mu_0$.

Nếu $t_{qs} \le t_{\text{tra bảng}}$ thì ta tạm chấp nhận giả thiết H_0.

*** Liên hệ giữa bài toán kiểm định và bài toán ước lượng khoảng:**

Trong bài toán kiểm định đang xét, nếu như giá trị trung bình mẫu $\overline{X}$ tính được không thể bác giải thiết H_0 ở mức ý nghĩa α thì nghĩa là ta có

$$-t_{trabang} \le \frac{\left|\overline{X} - \mu_0\right|}{\sigma_{n-1}} \sqrt{n} \le t_{trabang},$$

điều này tương đương với

$$\overline{X} - t_{trabang} \frac{\sigma}{\sqrt{n}} \le \mu_0 \le \overline{X} + t_{trabang} \frac{\sigma}{\sqrt{n}}.$$

Nghĩa là nếu μ_0 nằm trong khoảng ước lượng $1-\alpha$ của trung bình μ của đám đông thì ta chưa đủ cơ sở để bác giả thiết H_0.

Ví dụ3f 7. Quan sát thông số X (tuân theo luật phân phối chuẩn) của 36 phần tử, ta tính được $\overline{x} \approx 2576$; $\sigma_{n-1}^2 \approx 0.348$. Hãy so sánh thông số trung bình với 2.4 ở mức ý nghĩa 0.05.

Giải:

Gọi μ là thông số trung bình. Ta cần kiểm định giả thiết

$$H_0:\ \mu = 2.4,\quad H_1:\ \mu \ne 2.4$$

Vì $n > 30$, phương sai chưa biết nên tra bảng hàm Laplace, ta được

$$t_{trabang} = t_{0.475} = 196 \ (1)$$

Tiêu chuẩn kiểm định: $t_{qs} = \dfrac{\left|2.576 - 2.4\right|}{\sqrt{0.348}} \sqrt{36} \approx 1.79 \ (2)$

(1)&(2) $\to$ $t_{qs} < t_{trabang} \to$ Chưa đủ cơ sở để bác giả thiết H_0. Vậy tạm coi như thông số trung bình bằng 2.4.

Ví dụ3f 7. Quan sát mức xăng hao phí của 25 xe máy thuộc cùng một loại xe, chạy trên cùng một quãng đường, người ta thu được kết quả:

Mức xăng (l)	2.0	2.2	2.4	2.55
Số xe	5	9	8	3

Với mức ý nghĩa $\alpha = 0.05$ hãy so sánh mức xăng hao phí trung bình của loại xe máy trên với 2.35 lít. Giả thiết rằng mức xăng hao phí tuân theo luật chuẩn.

Giải: Tính toán trên bộ số liệu ta được:

$n = 25$; $\sum n_i x_i = 56.65$; $\sum n_i x_i^2 = 129.1475$ nên

$$\overline{x} = \frac{\sum n_i x_i}{n} = \frac{56.65}{25} = 2.266$$

$$\sigma_{n-1} = \sqrt{\frac{n}{n-1}\sigma_n^2} = \sqrt{\frac{n}{n-1}\left[\frac{\sum n_i x_i^2}{n} - \left(\overline{x}\right)^2\right]}$$

$$= \sqrt{\frac{25}{24}\left(\frac{129.1475}{25} - 2.266^2\right)} \approx 0.18$$

Gọi μ là mức xăng hao phí trung bình của loại xe máy.

Ta cần kiểm định giả thiết H_0: $\mu = 2.35$, H_1: $\mu \neq 2.35$

Vì $n < 30$, phương sai chưa biết (trường hợp 4) nên tra bảng phân phối Student ta có: $t_{n-1;\alpha} = t_{24;0.05} = 2.064$ (1)

Tiêu chuẩn kiểm định:

$$t_{qs} = \frac{\left|2.266 - 2.35\right|}{0.18}\sqrt{25} \approx 2.3 \ (2)$$

Từ (1) &(2) $\Rightarrow t_{qs} > t_{trabang}$: Bác giả thiết H_0.

Mà:

$$\overline{x} < 2.35 .$$

Nên với mức ý nghĩa 0.05, ta kết luận mức xăng hao phí trung bình của loại xe máy đang xét nhỏ hơn 2.35 lít.

7.2.4. So sánh hai trung bình

Hai đám đông có trung bình μ_x và μ_y đối với dấu hiệu nào đó có phân phối chuẩn và có cùng phương sai. Lấy hai mẫu ứng với hai đám đông, ta có:

$$X_1, X_2, \ldots, X_{n_1} \sim N\left(\mu_x, x\sigma^2\right)$$

$$Y_1, Y_2, \ldots, Y_{n_2} \sim N\left(\mu_y, y\sigma^2\right)$$

Và tính được $\overline{x}$, $x\sigma^2_{n_1-1}, \overline{y}$, $y\sigma^2_{n_2-1}$.

Ta cần kiểm định giả thiết H0: $\mu_x = \mu_y$ (H1: $\mu_x \neq \mu_y$) với mức ý nghĩa α. Người ta chứng minh được rằng khi $x\sigma^2, y\sigma^2$ đã biết thì biến ngẫu nhiên

$$\overline{Z} = \overline{X} - \overline{Y} \sim N\left(\mu_x - \mu_y, \sqrt{\frac{x\sigma^2}{n_1} + \frac{y\sigma^2}{n_2}}\right) \tag{7.9}$$

Còn nếu $x\sigma^2, y\sigma^2$ chưa biết thì có thể giả thiết thêm rằng chúng bằng nhau và bằng σ^2. Khi đó, nếu n_1, n_2 lớn thì biến ngẫu nhiên

$$\overline{Z} = \overline{X} - \overline{Y} \sim N\left(\mu_x - \mu_y, \sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right) \tag{7.10}$$

Với ước lượng của σ^2 là

$$\hat{\sigma}^2 = \frac{\left(n_1 - 1\right).x\sigma^2_{n_1-1} + \left(n_2 - 1\right).y\sigma^2_{n_2-1}}{n_1 + n_2 - 2} \tag{7.11}$$

Khi n_1, n_2 nhỏ thì ta $\overline{Z}$ tuân theo luật phân phối Student ở $n_1 + n_2 - 2$ bậc tự do, trung bình và phương sai vẫn như trên.

Vậy để kiểm định giả thiết H0: $\mu_x = \mu_y$, ta chuyển về bài toán so sánh trung bình của Z với số 0, tức là kiểm định giả thiết

H0: $\mu_x - \mu_y = 0$ (H1: $\mu_x - \mu_y \neq 0$)

Cách làm như sau:

Bước 1: Tìm $t_{trabang}$ và t_{qs} theo các trường hợp sau:

Trường hợp 1: $n_1, n_2 \geq 30$; $x\sigma^2$ và $y\sigma^2$ đã biết

- $t_{trabang} = t_{(1-\alpha)/2}$, bảng hàm Laplace

 $t_{qs} = \dfrac{|\overline{x} - \overline{y}|}{\sqrt{\dfrac{x\sigma^2}{n_1} + \dfrac{y\sigma^2}{n_2}}}$ (7.12)

Trường hợp 2: $n_1, n_2 \geq 30$; $x\sigma^2$ và $y\sigma^2$ chưa biết

 $t_{trabang} = t_{(1-\alpha)/2}$ bảng hàm Laplace

 $t_{qs} = \dfrac{|\overline{x} - \overline{y}|}{\sqrt{\dfrac{x\sigma_{n1-1}^2}{n_1} + \dfrac{y\sigma_{n2-1}^2}{n_2}}}$ (7.13)

Trường hợp 3: $n_1, n_2 \leq 30$; $x\sigma^2$ và $y\sigma^2$ đã biết: làm như trường hợp 1

Trường hợp 4: $n_1, n_2 \leq 30$; $x\sigma^2$ và $y\sigma^2$ chưa biết.

 $t_{trabang} = t_{n1+n2-2;\alpha}$ tra bảng phân phối Student ở $n_1 + n_2 - 2$ bậc tự do.

 $t_{qs} = \dfrac{|\overline{x} - \overline{y}|}{\sqrt{\dfrac{(n_1-1).x\sigma_{n_1-1}^2 + (n_2-1).y\sigma_{n_2-1}^2}{n_1+n_2-2}\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$ (7.14)

Bước 2: Nếu $t_{qs} > t_{trabang}$ thì ta bác bỏ giả thiết H0. Khi đó:

Nếu $\overline{x} > \overline{y}$ thì kết luận $\mu_1 > \mu_2$.

Nếu $\overline{x} < \overline{y}$ thì kết luận $\mu_1 < \mu_2$.

Nếu $t_{qs} \leq t_{trabang}$ thì ta tạm chấp nhận giả thiết H0.

Ví dụ3f 7. Nghiên cứu trọng lượng trẻ sơ sinh của 2 nhóm với mẹ không nghiện thuốc lá và nghiện thuốc lá, ta tính được:

$$n_1 = 15;\ \overline{x} = 3.5533;\ x\sigma_{n1-1} = 0.3707$$

và: $$n_2 = 14;\ \overline{y} = 3.2029;\ y\sigma_{n2-1} = 0.4927$$

Giả sử rằng trọng lượng trẻ của từng nhóm đều có phân phối chuẩn. Với mức $\alpha = 0.05$, có thể coi rằng trẻ sơ sinh của nhóm mẹ nghiện thuốc nhẹ cân hơn của nhóm không nghiện thuốc không?

Giải:

Gọi μ_1 và μ_2 lần lượt là trọng lượng trung bình của nhóm trẻ thuộc nhóm mẹ nghiện thuốc và nhóm mẹ không nghiện thuốc lá. Ta cần kiểm định giả thiết H0: $\mu_1 = \mu_2$, H1: $\mu_1 \neq \mu_2$.

Vì $n_1, n_2 < 30$ và phương sai chưa biết nên trường hợp ta xét rơi vào trường hợp 4.

$$t_{trabang} = t_{27;0.05} = 2.052 \quad (1)$$

$$t_{qs} = \frac{|3.5933 - 3.2029|}{\sqrt{\frac{14\times 0.3707^2 + 13\times 0.4927^2}{15+14-2}\left(\frac{1}{15}+\frac{1}{14}\right)}} \approx 2.42 \quad (2)$$

Từ (1) và (2) suy ra $t_{qs} > t_{trabang}$: Bác giả thiết H0. Mà $\overline{x} > \overline{y}$ nên ở mức ý nghĩa 0.05, ta kết luận: trọng lượng trung bình trẻ sơ sinh của nhóm mẹ không nghiện thuốc nặng hơn.

7.2.5. Kiểm định về tính độc lập của 2 biến ngẫu nhiên.

Một mẫu của hai đại lượng ngẫu nhiên X và Y (tức là trên một phần tử ta chú ý đến 2 dấu hiệu X và Y) là một bảng số 2 chiều:

Y \ X	x_1	x_2	$\cdots$	x_k	n_y
y_1	n_{11}	n_{12}	$\cdots$	n_{1k}	n_1
y_2	n_{21}	n_{22}	$\cdots$	n_{2k}	n_2
..	..	..		..	
y_ℓ	$n_{\ell 1}$	$n_{\ell 2}$	$\cdots$	$n_{\ell k}$	n_ℓ
m_x	m_1	m_2		m_k	n

Bảng 7.2.

Ta cần kiểm định giả thiết:

H0: “X và Y là hai đại lượng độc lập”

Với đối thiết là H1: “X và Y là hai đại lượng phụ thuộc lẫn nhau” ở mức ý nghĩa α.

Cách giải:

Từ định nghĩa xác suất theo thống kê ta có:

$$P\left[X = x_i\right] = \frac{m_i}{n},\ i = \overline{1,k},$$

$$p[Y = y_j] = \frac{n_j}{n},\ j = \overline{1,\ell},$$

$$P\left([X = x_i][Y = y_j]\right) = \frac{n_{ji}}{n},\ i = \overline{1,k},\ \ j = \overline{1,\ell}.$$

Vậy nếu X và Y độc lập thì

$$\frac{n_{ji}}{n} = \frac{m_j}{n}.\frac{n_j}{n}.$$

Từ đó ta có:

- Tiêu chuẩn kiểm định:

$$\chi^2_{qs} = \sum_{i=1}^{\ell}\ \sum_{j=1}^{k}\frac{\left(n_{ij} - v_{ij}\right)^2}{v_{ij}} \tag{7.15}$$

với $v_{ij} = \dfrac{n_i m_j}{n}$

- Tra bảng phân phối Khi bình phương ở $(k-1)(\ell-1)$ bậc tự do, ta được

$$\chi^2_{trabang} = \chi^2_{(k-1)(\ell-1),\alpha} \tag{7.16}$$

- Nếu $\chi^2_{qs} \geq \chi^2_{(k-1)(\ell-1),\alpha}$ thì ta bác giả thiết H_0, ngược lại ta chấp nhận giả thiết H_0.

Ví dụ3f 7. [2] Quan sát tính tốt xấu của 100 sản phẩm được sản xuất ở 3 ca I, II, III của một nhà máy, ta có kết quả sau:

Ca / Chất lượng	I	II	III	n_i
Tốt	25 (27.2)	32 (32.3)	28 (25.5)	85
Xấu	7 (4.8)	6 (5.7)	2 (4.5)	15
m_j	32	38	30	$n = 100$

Với mức $\alpha = 0.05$ có thể xem chất lượng sản phẩm không phụ thuộc vào ca sản xuất được không?

Giải:

Đặt giả thiết H_0: "Chất lượng sản phẩm không phụ thuộc vào ca sản xuất". Rõ ràng $(1-1)(k-1) = (2-1)(3-1) = 2$ và $\alpha = 5\% = 0.05$ nên tra bảng phân phối khi bình phương ta có

$$\chi^2_{2,\alpha} = \chi^2_{2;\,0.05} = 5.991 \quad (1)$$

Tính giá trị kiểm định, ta có :

$$v_{11} = \frac{n_1 m_1}{n} = \frac{85 \times 32}{100} = 27.2$$

Tương tự, ta tính được v_{ij} (là các giá trị trong ngoặc, được liệt kê trong bảng). Từ đó tính được:

$$\chi^2_{qs} = \frac{(25-27.2)^2}{27.2} + \frac{(32-32.3)^2}{32.3} + \frac{(28-25.5)^2}{25.5}$$
$$+ \frac{(7-4.8)^2}{4.8} + \frac{(6-5.7)^2}{5.7} + \frac{(2-4.5)^2}{4.5} \approx 2.84 \quad (2)$$

Từ (1) và (2) suy ra $\chi^2_{qs} < \chi^2_{2;\,0.05}$. Vậy ta chấp nhận H_0: có thể coi như chất lượng sản phẩm không phụ thuộc vào ca sản xuất với mức ý nghĩa 0.05.

7.2.6. Kiểm định về phân phối (đọc thêm)

Bài toán:

Dựa vào mẫu của biến ngẫu nhiên X, ta cần kiểm định giả thiết H_0: "X có phân phối F_0" (H_1: "X không có phân phối F_0")

Ở mức ý nghĩa α (F_0 là luật phân phối đã biết).

Phần này ta xét F_0 là một phân phối liên tục. Số liệu quan sát thường được sắp xếp vào các khoảng liên tiếp (thường là bằng nhau) và có dạng như sau:

Khoảng quan sát	$(x_0;x_1)$	$(x_1;x_2)$	…	$(x_{k-1};x_k)$
n_i	n_1	n_2	…	n_k

Trong đó $n_i > 4$ (có thể trừ n_1 và n_k), $3 < k < 16$ và x_0 và x_k có thể nhận giá trị vô cùng.

Qui tắc Pearson:

Bước 1: tra bảng phân phối khi bình phương:

$$t_{trabang} = \chi^2_{(k-r-1,\alpha)} \tag{7.17}$$

r là số các tham số chưa biết của phân phối F_0.

Bước 2: tính tiêu chuẩn kiểm định:

$$t_{qs} = \sum_{i=1}^{k} \frac{\left(n_i - np_i\right)^2}{np_i} \tag{7.18}$$

trong đó $p_i = P\left(x_{i-1} < X < x_i\right)$.

Bước 3: nếu $t_{qs} \le t_{trabang}$ thì ta chấp nhận H0. Ngược lại ta bác H0.

Ví dụ3f 7. [4] Đo độ nhạy một kênh truyền hình của 50 máy thu hình ta có kết quả cho trong bảng dưới. Hãy kiểm định giả thiết rằng độ nhạy của kênh truyền hình tuân theo phân phối chuẩn ở mức $\alpha = 0.05$.

Khoảng (μV)	Số máy (n_i)	Khoảng (μV)	Số máy (n_i)
75-125	1	375-425	6
125-175	2	425-475	5
175-225	4	475-525	2
225-275	9	525-575	2
275-325	8	575-625	1
325-375	8	625-675	2

Giải:

Gọi X là độ nhạy của kênh truyền hình, đặt giả thiết H0: $X \sim N\left(\mu, \sigma^2\right)$

Dùng ước lượng điểm để ước lượng μ, σ^2

Dễ dàng tính được: $n = 50$; $\overline{X} = 346$; $\sigma^2_{n-1} \approx 124.4^2$

Vậy H0: $X \sim N\left(346; 124.4^2\right)$

Do có nhiều $n_i < 5$ nên ta xếp lại mẫu:

x_i	<225	225-275	275-325	325-375	375-425	425-475	>475

n_i	7	9	8	8	6	5	7

Vậy ta có $k=7$, $r=2$ (Do phân phối N có 2 tham số)

Từ $\alpha=0.05$, tra bảng phân phối khi bình phương ta được:

$$t_{trabang}=\chi^2_{(7-2-1;0,05)}=7.815 \ (1)$$

Khi H đúng ta có:

$$p_i=P\left(x_{i-1}<X<x_i\right)=\phi\left(\frac{x_i-346}{124.4}\right)-\phi\left(\frac{x_{i-1}-346}{124.4}\right)$$

nên tra bảng hàm Laplace ta có (gần đúng):

$p_1=0.166$ $p_2=0.1183$ $p_3=0.1482$ $p_4=0.1585$ $p_5=0.1447$ $p_6=0.1151$ $p_7=0.1492$

Từ đó tính được:

$$t_{qs}=\sum_{i=1}^{7}\frac{\left(n_i-np_i\right)^2}{np_i}\approx 2.22 \ (2)$$

Từ (1) và (2) $\Rightarrow t_{qs}<t_{trabang}$

Vậy giả thiết X có phân phối chuẩn là chấp nhận được.

BÀI TẬP

7.1. Phương pháp sản xuất cũ có tỉ lệ phế phẩm là 7%. Sau khi áp dụng kĩ thuật mới, người ta chọn ngẫu nhiên ra 200 sản phẩm để kiểm tra thì thấy có 9 phế phẩm. Với mức ý nghĩa 0.05 có thể kết luận rằng việc áp dụng kĩ thuật mới có hiệu quả hơn hay không?

7.2. Tỉ lệ người mắc bệnh A ở một địa phương là 5%. Trong một lần kiểm tra sức khỏe ngẫu nhiên 300 người thấy có 24 người mắc bệnh A. Với mức ý nghĩa 0.05 có thể cho rằng tỉ lệ người bị bệnh A có xu hướng tăng lên hay không?

7.3. Một hiệu làm đầu cho rằng 90% khách hàng của họ hài lòng với chất lượng phục vụ. Nghi ngờ chủ hiệu nói quá lên, một nhà điều tra xã hội học phỏng vấn 150 khách hàng của hiệu thì thấy 132 người nói là hài lòng. Với mức $\alpha = 0.05$, có thể kết luận gì về nghi ngờ trên?

7.4. Một hãng điều tra dư luận cho biết có 68% cử tri sẽ bỏ phiếu cho ứng cử viên A. Chọn ngẫu nhiên 36 cử tri thì thấy có 26 người bỏ phiếu cho ứng cử viên A. Với $\alpha = 0.05$ bạn có kết luận gì về kết quả điều tra của hãng trên?

7.5. Trong một nước, tỉ lệ tử vong của bệnh A là 1.5%. Theo dõi 1000 trường hợp mắc bệnh A trong một ngành sản xuất, người ta thấy có 20 trường hợp tử vong. Với mức ý nghĩa $\alpha = 0.05$, hãy kiểm định xem đặc điểm nghề nghiệp của ngành sản xuất này có ảnh hưởng đến tỉ lệ tử vong của bệnh A hay không?

7.6. Một máy sản xuất tự động có tỉ lệ sản phẩm loại A lúc đầu là 48%. Người ta cải tiến máy và sau một thời gian áp dụng, người ta kiểm tra 40 lần, mỗi lần 10 sản phẩm và ghi được kết quả sau:

Số sản phẩm loại A trong một lần kiểm tra	**Số lần kiểm tra**
0	0
1	2
2	0
3	4
4	6
5	8
6	10
7	4
8	5
9	1
10	0

a) Ước lượng tỉ lệ sản phẩm loại A sau khi cải tiến máy với độ tin cậy 99%.

b) Hãy cho kết luận về hiệu quả của việc cải tiến máy với mức $\alpha = 0.05$

7.7. Theo dõi trọng lượng của một số trẻ sơ sinh tại một số nhà hộ sinh ở thành phố và ở nông thôn, người ta thấy: trong số 150 trẻ sơ sinh ở thành phố có 100 cháu nặng hơn 3000 gam, trong số 200 trẻ sơ sinh ở nông thôn có 98 cháu nặng hơn 3000 gam. Từ kết quả đó, hãy so sánh tỉ lệ trẻ sơ sinh có trọng lượng trên 3000 gam ở thành phố và ở nông thôn, với mức ý nghĩa 0.05.

7.8. Một máy cần đóng gói các sản phẩm có khối lượng 1 Kg. Nghi ngờ máy hoạt động không bình thường, người ta chọn ra một mẫu gồm 100 sản phẩm thì thấy như sau:

Khối lượng	0.95	0.97	0.99	1.01	1.03	1.05
Số gói	9	31	40	15	3	2

Với mức ý nghĩa 0.05, hãy kết luận về nghi ngờ trên?

7.9. Quan sát số gạo bán ra trong một ngày của một cửa hàng bán gạo sau một thời gian, người ta ghi được số liệu sau:

Số gạo (Tạ)	12	13	15	16	17	18	19
Số ngày	3	2	7	7	3	2	1

a) Tìm ước lượng không chệch cho số gạo trung bình bán được trong một ngày?

b) Sau khi tính toán, ông chủ cửa hàng nói rằng nếu trung bình một ngày bán không được 15 tạ gạo thì chẳng thà đóng cửa còn hơn. Dựa vào số liệu trên, anh (chị) hãy kết luận giúp ông chủ cửa hàng xem có nên tiếp tục bán hay không ở mức ý nghĩa $\alpha = 0.05$?

c) Giả sử những ngày bán được từ 13 tạ đến 17 tạ là những ngày "bình thường". Hãy ước lượng tỉ lệ ngày bình thường của cửa hàng ở độ tin cậy 99%. (Giả thiết rằng số gạo bán ra trong ngày cò phân phối chuẩn)

7.10. Một xí nghiệp đúc một số rất lớn các sản phẩm bằng gang với số khuyết tật trung bình ở mỗi sản phẩm là 3. Người ta cải tiến cách sản xuất và kiểm tra 36 sản phẩm, kết quả như sau:

7 sản phẩm không có khuyết tật nào
4 sản phẩm có 1 khuyết tật
4 sản phẩm có 2 khuyết tật
6 sản phẩm có 3 khuyết tật
8 sản phẩm có 4 khuyết tật
6 sản phẩm có 5 khuyết tật
1 sản phẩm có 6 khuyết tật

Giả thiết rằng số khuyết tật của các sản phẩm có phân phối chuẩn.

a) Hãy ước lượng số khuyết tật trung bình ở mỗi sản phẩm sau cải tiến, với độ tin cậy 95% (độ tin cậy 99%).

b) Hãy cho kết luận về hiệu quả của việc cải tiến sản xuất ở mức ý nghĩa 0.05.

7.11. Để nghiên cứu tác dụng của việc bón phân đạm theo công thức A đối với sản lượng của bắp, người ta làm thí nghiệm trên 5 cặp mảnh đất cạnh nhau, mỗi cặp gồm có 1 mảnh đối chứng (không bón phân đạm) và một mảnh có bón phân đạm theo công thức trên, các sản lượng thu được như sau (tính theo tạ/ha).

Mảnh đối chứng	55	53	30	37	49
Mảnh bón phân	60	58	29	39	47

Hãy cho kết luận về hiệu quả của việc bón phân đạm theo công thức A, với mức ý nghĩa 0.05.

7.12. Đo chỉ số mỡ sữa X (tính theo %) của 130 con bò lai F_1-Hà-An, ta được bảng số liệu sau đây:

a) Hãy ước lượng chỉ số mỡ sữa trung bình của giống bò lai trên ở độ tin cậy 99%.

b) Biết rằng chỉ số mỡ sữa trung bình của giống bò Hà thuần chủng là 4.95. Hãy cho kết luận về hiệu quả của việc lai tạo trên với mức ý nghĩa 0.01.

c) Muốn ước lượng chỉ số mỡ sữa trung bình của bò lai F_1 Hà-An ở độ chính xác 0.15% và độ tin cậy 99% thì phải lấy mẫu tối thiểu bằng bao nhiêu?

(Giả thiết rằng X có phân phối chuẩn)

Chỉ số mỡ sữa	Số bò lai
3.0-3.6	2
3.6-4.2	8
4.2-4.8	35
4.8-5.4	43
5.4-6.0	22
6.0-6.6	15
6.6-7.2	5

7.13. Trong một liên hiệp xí nghiệp, người ta chọn ngẫu nhiên 1000 công nhân và theo dõi số ngày nghỉ của họ trong 1 năm. Kết quả thu được như sau:

Với mức ý nghĩa 0.01, hãy kiểm định giả thiết cho rằng sự nghỉ việc không phụ thuộc vào giới tính.

Giới tính / Số ngày nghỉ	Nữ	Nam
0-5	300	500
5-20	80	70
>20	20	30

7.14. Điều tra tình hình bệnh tật trong một đợt dịch đối với 1000 người, ta có kết quả như sau:

Tình hình bệnh tật / Tình hình tiêm chủng	Mắc Bệnh	Không Mắc bệnh	n_i
Có tiêm chủng	8	192	200
Không tiêm chủng	92	708	800
m_j	100	900	1000

Hãy xét xem việc tiêm chủng và sự mắc bệnh có liên quan với nhau hay không? (Cho mức ý nghĩa $\alpha = 0.05$).

Chương 8
HỒI QUI TUYẾN TÍNH

8.1. TỔNG QUAN VỀ HỒI QUI

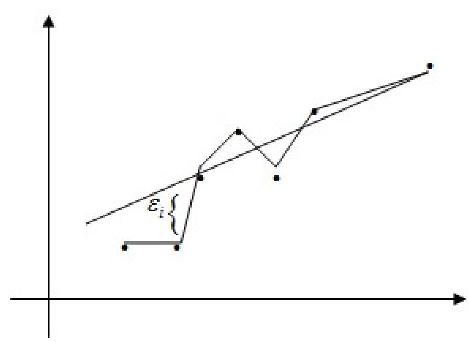
Trong thực tế ta thường gặp hai đại lượng ngẫu nhiên *X* và *Y* có mối liên hệ với nhau, trong đó thường sẽ có một đại lượng (*X*) dễ khảo sát được (hay kiểm soát được), còn đại lượng còn lại (*Y*) khó khảo sát hơn. Khi đó, ta muốn tìm mối liên hệ nào đó giữa *X* và *Y* để từ đó dự đoán được đại lượng khó khảo sát.

Giả sử quan sát hai dấu hiệu *X*, *Y* trong một bối cảnh nào đó (thường liên quan trong một phần tử, chẳng hạn như chiều cao và trọng lượng của cùng một người) ta có kết quả:

X	X_1	X_2	...	X_n
Y	Y_1	Y_2	...	Y_n

Bảng 8.1.

□ Để có một khái niệm sơ bộ về mối quan hệ giữa hai đại lượng ngẫu nhiên *X* và *Y*, người ta thường biểu diễn mỗi quan sát $(x_i; y_i)$ bởi một điểm trên mặt phẳng toạ độ. Các điểm này hợp lại thành *đám mây điểm* trên mặt phẳng. Nhìn vào đám mây điểm này, ta “đọc” được mối quan hệ giữa *X* và *Y*, nhưng quan trọng không kém là phải dự đoán được *Y* khi biết *X* (hoặc ngược lại). Để làm được điều này, người ta thường vẽ một đường cong để mô tả mối quan hệ giữa hai đại lượng và dùng nó để dự đoán.



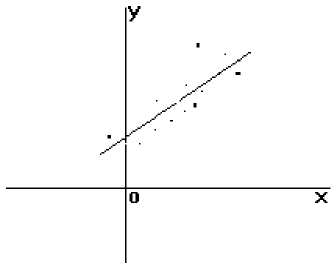
Hình 8.1.

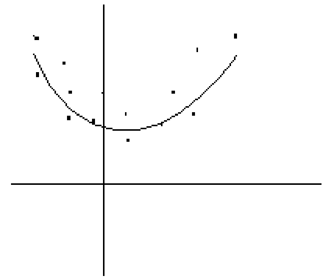
Trong hình trên, các đường thẳng gấp khúc nối tất cả các điểm quan sát gọi là **đường hồi qui thực nghiệm**. Đây chính là đường mà chúng ta thường gặp trong thực tế, chẳng hạn, khi người ta quan tâm đến sự biến động của giá một thùng dầu thô theo thời gian.

- Phương pháp chọn một đường cong (tức là chọn một phương trình) cho bộ dữ liệu gọi là phương pháp hồi qui, còn phương trình gọi là phương trình hồi qui (hay hàm hồi qui).
- Các phương trình hồi qui được xây dựng dựa trên ít nhất hai mục đích: dự đoán các quan sát mới và đánh giá mức độ tương quan giữa các biến tác động và biến đáp ứng.
- Phương pháp hồi qui được ứng dụng rộng rãi trong nhiều lĩnh vực: xã hội, khoa học kỹ thuật, thương mại, ...

Cách chọn phương trình: Nếu đám mây điểm có xu hướng tụ tập xung quanh một đường thẳng nào đó thì ta chọn hàm hồi qui là $Y = aX + b$ trường hợp này ta gọi là hồi qui tuyến tính của *Y* theo *X*; Nếu đám mây điểm có xu hướng tụ tập xung quanh một parabol thì ta chọn hàm hồi qui là $Y = aX^2 + bX + c$, v.v…

Nói chung, tuỳ theo "hình dáng" của đám mây điểm mà ta chọn cho nó phương trình thích hợp. Các trường hợp kể trên thuộc họ hồi qui có tham số. Để xác định được hàm hồi qui $y = f(x)$, người ta thường dùng phương pháp *bình phương bé nhất*. Ý tưởng của phương pháp này như sau:





Hồi qui tuyến tính *Hồi qui bậc hai*

Hình 8.2.

- Xác định trước dạng của hàm $f(x)$, chẳng hạn $f(x) = ax^2 + bx + c$, hàm $f(x)$ này sẽ xác định được khi tìm được các hệ số a, b, c.
- Đặt tổng bình phương

$$s = \sum \varepsilon_i^2 = \sum_i \left[y_i - f(x_i) \right]^2 \qquad (8.1)$$

Trong đó $\varepsilon_i = y_i - f(x_i)$ là độ lệch giữa giá trị quan sát thực tế và giá trị dự đoán bởi hàm hồi qui.

- Coi s là hàm theo các biến là các tham số cần xác định và tìm điểm cực tiểu của s. Từ đây ta sẽ tìm được các tham số cần xác định.

Sau đây ta xét trường hợp đơn giản nhất: Phương trình hồi qui là đường thẳng.

8.2. HỒI QUI TUYẾN TÍNH

Hồi qui tuyến tính là trường hợp phương trình hồi qui có dạng:

$$y = ax + b$$

Phương pháp *bình phương bé nhất*: ta phải xác định các hệ số a, b sao cho cực tiểu tổng bình phương

$$S = \sum \varepsilon_i^2 \,,\ \varepsilon_i = Y_i - \left(aX_i + b\right) \tag{8.2}$$

Coi S như là một hàm theo hai biến a, b ta sẽ tìm điểm cực tiểu của hàm S để được các ước lượng của a và b bằng cách giải hệ phương trình:

$$\begin{cases} \dfrac{\partial s}{\partial a} = 0 \\ \dfrac{\partial s}{\partial b} = 0 \end{cases}$$

$$\Leftrightarrow \begin{cases} \left(\sum X_i^2\right)a + \left(\sum X_i\right)b = \sum X_i Y_i \\ \left(\sum X_i\right)a + nb = \sum Y_i \end{cases} \tag{8.3}$$

Từ đó suy ra các ước lượng của a và b là:

$$A = \frac{n\sum X_i Y_i - \sum X_i \sum Y_i}{n\sum X_i^2 - \left(\sum X_i\right)^2}\,; \tag{8.4}$$

$$B = \overline{Y} - A\overline{X} \tag{8.5}$$

Trong công thức tính A ở trên, nếu đem tử và mẫu chia cho n^2 thì ta sẽ có công thức tính A như sau:

$$A = \frac{\dfrac{\sum X_i Y_i}{n} - \overline{X}\,\overline{Y}}{S_x^2} \tag{8.6}$$

Với S_x^2 là phương sai mẫu của X.

Khi có mẫu cụ thể, ta có công thức tính A và B như sau:

$$A = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - \left(\sum x_i\right)^2} \tag{8.7}$$

Hay

$$A = \frac{\dfrac{\sum x_i y_i}{n} - \overline{x}\,\overline{y}}{x\sigma_n^2} \tag{8.8}$$

$$B = \overline{y} - A\overline{x} \tag{8.9}$$

Trong đó $x\sigma_n^2$ là giá trị cụ thể của S_x^2.

Để xác định các hệ số của phương trình hồi qui tuyến tính, có thể tìm qua một trong hai công thức trên.

Ví dụ3f 8. Cho bộ số liệu:

X	0	1	2	3	4	5	6	7
Y	0	1	1	3	4	6	9	10

Lập phương trình hồi qui của y theo x.

Giải: (Ta sẽ tìm phương trình hồi qui theo công thức (8.3))

Tính được: $n = 8$;

$\sum x_i = 28; \sum y_i = 34; \sum x_i^2 = 140; \sum x_i y_i = 182.$

Tìm hệ số a,b bằng cách giải hệ: $\begin{cases} 140a + 28b = 182 \\ 28a + 8b = 34 \end{cases} \Leftrightarrow \begin{cases} a = 1.5 \\ b = -1 \end{cases}$

Vậy phương trình hồi qui cần tìm là: $y = 1.5x - 1$.

Lưu ý: cũng giống như trong lý thuyết mẫu, đôi khi ta cũng gặp các cặp số liệu trùng lại. khi các cặp quan sát có sự trùng lại thì thông thường số liệu sẽ được trình bày dưới dạng bảng hai chiều như sau:

Y \ X	x_1	x_2	…	x_k
y_1	n_{11}	n_{12}	…	n_{1k}
y_2	n_{21}	n_{22}	…	n_{2k}
..	..	..		..

y_ℓ	$n_{\ell 1}$	$n_{\ell 2}$	$\dots$	$n_{\ell k}$

Bảng 8.2

Trong đó n_{ij} là số lần lặp lại cặp giá trị $(x_j; y_i)$.

Với số liệu như bảng bên, công thức xác định hệ số A sẽ là

$$A = \frac{\dfrac{\sum_{i,j} n_{ij} x_j y_i}{n} - \overline{x}\,\overline{y}}{x\sigma_n^2} \tag{8.10}$$

Dễ dàng nhận thấy rằng: $\overline{x}$; $x\sigma_n^2$ được tính từ bảng số liệu:

X	x_1	x_2	$\dots$	x_k
n_x	$\sum_{i=1}^{\ell} n_{i1}$	$\sum_{i=1}^{\ell} n_{i2}$	$\dots$	$\sum_{i=1}^{\ell} n_{ik}$

Bảng 8.3.

Còn $\overline{y}$ được tính từ bảng số liệu:

Y	y_1	y_2	$\dots$	y_ℓ
n_y	$\sum_{j=1}^{k} n_{1j}$	$\sum_{j=1}^{k} n_{2j}$	$\dots$	$\sum_{j=1}^{k} n_{\ell j}$

Bảng 8.4.

8.3. HỆ SỐ TƯƠNG QUAN

Trong chương 4, ta đã biết khái niệm hệ số tương quan, giờ sẽ tìm hiểu tính chất, ý nghĩa và ứng dụng của nó.

Để đánh giá *mức độ phụ thuộc tuyến tính* của hai biến ngẫu nhiên X và Y, người ta dựa vào hệ số tương quan, ký hiệu là ρ, được xác định như sau:

$$\rho = \frac{E(X-\mu_X)(Y-\mu_Y)}{\sigma_x\,\sigma_y} \tag{8.11}$$

Tính chất cơ bản:

(i) $p \in [-1;1]$

(ii) Nếu $p = 0$ thì hai biến số trên không có tương quan tuyến tính.

(iii) Nếu $|p|$ càng gần 1 thì sự phụ thuộc tuyến tính giữa hai biến càng mạnh.

Ước lượng hệ số tương quan

Người ta sử dụng hệ số tương quan mẫu r_{xy} (còn gọi là hệ số tương quan Pearson) để ước lượng p :

$$r_{xy} = \frac{n\sum XY - \left(\sum X\right)\left(\sum Y\right)}{\sqrt{n\sum X^2 - \left(\sum X\right)^2}\sqrt{n\sum Y^2 - \left(\sum Y\right)^2}} \qquad (8.12)$$

Thay công thức (5.3) vào (8.12) được

$$r_{xy} = \frac{n\sum XY - \left(\sum X\right)\left(\sum Y\right)}{n^2.S_x.S_y} \qquad (8.13)$$

Chia tử và mẫu của (8.13) cho n^2 ta được

$$r_{xy} = \frac{\frac{\sum XY}{n} - \overline{X}.\overline{Y}}{S_x.S_y} \qquad (8.14)$$

Trong các công thức trên, $\sum XY$ dùng để chỉ tổng của các tích XY đối với từng cặp quan sát (X,Y) cho cả hai trường hợp: Số liệu liệt kê theo hàng và số liệu được cho trong bảng hai chiều.

Khi có mẫu cụ thể, các công thức (8.12), (8.13) và (8.14) trở thành:

$$r_{xy} = \frac{n\sum xy - \left(\sum x\right)\left(\sum y\right)}{\sqrt{n\sum x^2 - \left(\sum x\right)^2}\sqrt{n\sum y^2 - \left(\sum y\right)^2}} \qquad (8.15)$$

$$r_{xy} = \frac{n\sum xy - \left(\sum x\right)\left(\sum y\right)}{n^2.x\sigma_n.y\sigma_n} \qquad (8.16)$$

$$r_{xy} = \frac{\frac{\sum xy}{n} - \overline{x}.\overline{y}}{x\sigma_n.y\sigma_n} \qquad (8.17)$$

Trong đó $x\sigma_n$, $x\sigma_n$ lần lượt là độ lệch chuẩn mẫu của X và Y.

Từ (8.8) và (8.17), có thể rút ra mối liên hệ giữa hệ số tương quan và hệ số của hàm hồi qui tuyến tính là

$$A = r_{xy}\frac{y\sigma_n}{x\sigma_n} \qquad (8.18)$$

Đánh giá mối liên hệ tuyến tính:

Trong thực hành, người ta phân loại $\left|r_{xy}\right|$ đối với mối liên hệ tuyến tính như sau:

* 0.8 $\to$ 1 : Sự liên hệ cao, rất đáng tin cậy.
* 0.6 $\to$ 0.79 : Sự liên hệ từ vừa phải đến rõ rệt.
* 0.4 $\to$ 0.59 : Sự liên hệ ở mức trung bình.
* 0.2 $\to$ 0.39 : Sự liên hệ ở mức thấp.
* 0 $\to$ 0.19 : Sự liên hệ không đang kể, có thể do ngẫu nhiên.

Hơn nữa, nếu $r_{xy} > 0$ thì ta có tương quan thuận giữa hai đại lượng X và Y, tức là X và Y đồng biến; còn nếu $r_{xy} < 0$ thì ta có tương quan thuận giữa X và Y.

Ví dụ3f 8. Cho X là điểm toán thi vào đại học và *Y* là điểm thi cuối năm thứ I của 10 sinh viên ban toán (thang điểm 20), ta có kết quả điều tra:

X	2	8	5	10	14	12	18	6	8	10
Y	4	8	6	10	12	10	15	5	9	12

a) Lập phương trình hồi qui của Y theo X.

b) Dự đoán điểm thi cuối năm của sinh viên có điểm thi toán vào đại học là 15.

Giải:

a) Ta dễ dàng tính được $n = 10$

$$\sum xy = 982\,;\ \sum x = 93\,;\ \sum x^2 = 1057\,;\ \sum y = 91\,;$$

Thay vào công thức (8.7) ta tính được

$$A = 0.706 \text{ và } B = 2.534$$

Vậy phương trình hồi qui cần tìm là:

$$y = 0.706x + 2.534$$

b) Sinh viên có điểm thi vào ĐH là 15 nên $\rightarrow\ x = 15$, thay vào phương trình trên suy ra

$$y = 0.706 \times 15 + 2.534 \approx 13\,.$$

Vậy dự đoán điểm thi cuối năm thứ nhất của sinh viên trên là 13.

Ví dụ3f 8. Để nghiên cứu sự ảnh hưởng của phân bón (*X*) đến năng suất lúa (*Y*), người ta tiến hành thí nghiệm trên 100 mảnh ruộng và thu được kết quả như sau:

Lượng phân bón X (kg) / Năng suất lúa Y(kg)	10	12	14	16
100	22			
150	8	10	3	
200	14	15	12	
250				16

a) Ước lượng năng suất lúa trung bình với độ tin cậy 95%.

b) Tìm phương trình hồi qui tuyến tính giữa năng suất lúa với lượng phân bón. Đánh giá mức độ chặt chẽ của mối phụ thuộc tương quan tuyến tính này.

Giải:

Y(kg) \ X (kg)	**10**	**12**	**14**	**16**	**n_y**
100	22				**22**
150	8	10	3		**21**
200	14	15	12		**41**
250				16	**16**
n_x	**44**	**25**	**15**	**16**	**n = 100**

Từ bảng trên ta tính được:

$n = 100$; $\sum x_i = 1206$; $\sum x_i^2 = 15036$; $\overline{x} = 12.06$; $x\sigma_n = 2.217$;

$\sum y = 17550$; $\sum y^2 = 3332500$; $\overline{y} = 175.5$;

$y\sigma_n = 50.247$; $y\sigma_{n-1} = 50.5$

$\sum xy = 219900$

a) Thấy ngay, đây là bài toán xây dựng khoảng tin cậy 95% cho năng suất lúa trung bình.

- Do $n > 30$, σ_y^2 chưa biết nên ơ độ tin cậy 95%, tra bảng hàm Laplace, ta được $t_{\gamma/2} = t_{0.475} = 1.96$

- Sai số của khoảng tin cậy:

$$\varepsilon = t_{0.475}\frac{x\sigma_{n-1}}{\sqrt{n}} = 1.96\frac{50.5}{10} \approx 9.9$$

Vậy khoảng tin cậy 95% cho năng suất lúa trung bình là:

$$\left(\overline{y} - \varepsilon;\ \overline{y} + \varepsilon\right) = \text{(165.60; 185.4) (kg)}$$

b) Ta có:

$$A = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2}$$

$$= \frac{100 \times 219900 - 1206 \times 17550}{100 \times 15036 - 1206^2}$$
$$\approx 16.77$$
$$B = \overline{y} - A\overline{x}$$
$$= 175.5 - 16.77 \times 12.06$$
$$\approx -26.75$$

Vậy phương trình hồi qui tuyến tính của Y theo X là:

$$y = 16.77x - 26.75$$

Tính hệ số tương quan Pearson theo công thức (8.17), ta có

$$r_{xy} = \frac{\frac{219900}{100} - 12.06 \times 175.5}{2.217 \times 50.247} \approx 0.74$$

Vậy mối phụ thuộc tương quan tuyến tính này khá chặt chẽ và đây là tương quan thuận.

BÀI TẬP

1. Quan sát 40 lần cặp biến (X,Y) ta có bộ số liệu:

y_i \ x_i	0-0.2	0.2-0.4	0.4-0.6	0.6-0.8	0.8-1	1-1.2
10-15	4	2				
15-20		2		6		
20-25			2			
25-30				4		
30-35				4	6	
35-40					6	4

a) Xác định đường hồi qui tuyến tính của y theo x.

b) Tính hệ số tương quan tuyến tính giữa x và y

2. Số liệu thống kê nhằm nghiên cứu quan hệ giữa tổng sản phẩm nông nghiệp *Y* với tổng giá trị tài sản cố định *X* của 10 nông trại (tính trên 100 ha) như sau:

x_i	y_i	x_i	y_i
11.3	13.2	22.0	23.9
12.9	15.6	22.2	22.4
13.6	17.2	23.7	23.0
16.8	18.8	26.6	24.4
18.8	20.2	27.5	24.6

a) Hãy xác định phương trình hồi qui tuyến tính của *Y* theo *X*.

b) Dự đoán tổng sản phẩm nông nghiệp khi tổng tài sản cố định là 25.

3. Độ ẩm không khí X ảnh hưởng đến sự bay hơi nước Y trong sơn khi phun ra, sự hiểu biết về sự ảnh hưởng này sẽ giúp ta tiết kiệm được

lượng sơn bằng cách chỉnh súng phun sơn một cách thích hợp. Tiến hành 25 quan sát người ta thu được số liệu sau:

Thứ tự Quan sát	**Độ ẩm (%)**	**Độ bay hơi(%)**	**Thứ tự Quan sát**	**Độ ẩm (%)**	**Độ bay hơi(%)**
1	35.3	11	14	39.1	9.6
2	29.7	11.1	15	46.8	10.9
3	30.8	12.5	16	48.5	9.6
4	58.8	8.4	17.	59.3	10.1
5	61.4	9.3	18	70.0	8.1
6	71.3	8.7	19	70.0	6.8
7	74.4	6.4	20	74.4	8.9
8	76.7	8.5	21	72.1	7.7
9	70.7	7.8	22	58.1	8.5
10	57.5	9.1	23	44.6	8.9
11	46.4	8.2	24	33.4	10.4
12	28.9	12.2	25	28.6	11.1
13	28.1	11.9			

Lập hàm hồi qui tuyến tính của Y theo X.

ÔN TẬP

1. Cho $E = \{1,2,3,4,5\}$. Có thể lập được từ E bao nhiêu số tự nhiên

a) Có 3 chữ số?

b) Có 3 chữ số khác nhau?

c) Có 3 chữ số giống nhau?

d) Có 3 chữ số, tận cùng bằng chừ số 4?

e) Có 3 chữ số khác nhau, tận cùng băng chữ số 4?

f) Có 3 chữ số không tận cùng bằng chữ số 4?

g) Có 3 chữ số khác nhau, không tận cùng bằng chữ số 4?

2. Tung 3 súc sắc.

a) Có mấy trường hợp có thể xảy ra?

b) Có mấy trường hợp xuất hiện ba mặt có số chấm đôi một khác nhau?

c) Có mấy trường hợp xuất hiện ba mặt có số chấm như nhau?

d) Có mấy trường hợp xuất hiện mặt 5 ở súc sắc thứ nhất?

e) Có mấy trường hợp xuất hiện ba mặt 4, 5, 6?

f) Có mấy trường hợp xuất hiện đúng một mặt 5?

g) Có mấy trường hợp không xuất hiện mặt 5?

h) Có mấy trường hợp tổng số chấm xuất hiện trên mặt 3 súc sắc đó bằng 10?

i) Tổng số chấm bằng 8 biết rằng có ít nhất một con ra mặt 1 chấm?

j) Tổng số chấm bằng 8?

Hướng dẫn:

h) Xét các trường hợp (1;3;6), (1;4;5), (2;2;6), (2;3;5), (2;4;4), (3;4;3).

i) (1;1;6), (1;2;5), (1;3;4).

3. Xếp ngẫu nhiên 5 người, trong đó có hai người mà ta sẽ gọi là A và B lên 7 toa tàu có đánh số. Hỏi có bao nhiêu cách xếp

a) Có thể xảy ra?

b) Cả 5 người lên toa thứ 3?

c) Cả 5 người lên cùng 1 toa?

d) 5 người lên 5 toa đầu, mỗi người 1 toa khác nhau?

e) 5 người lên 5 toa khác nhau?

f) A, b cùng lên toa đầu?

g) A, b cùng lên 1 toa?

h) Toa đầu chỉ có a và b?

i) Có đúng 5 toa không có người lên?

j) Có đúng 3 toa không có người lên?

k) Chỉ có 3 toa đầu đều có người lên?

l) Chỉ có 3 toa có người lên?

4. Xếp ngẫu nhiên 7 cuốn sách trong đó có 3 sách toán vào 5 ngăn tủ có đánh số, mỗi ngăn tủ đều có thể chứa nhiều sách. Hỏi :

a) Có mấy cách xếp?

b) Có mấy cách xếp mà 3 sách toán ở cùng ngăn thứ 3?

c) Có mấy cách xếp mà 3 sách toán luôn cùng 1 ngăn?

d) Có mấy cách xếp mà 3 sách toán chỉ được xếp ở 3 ngăn đầu tiên?

e) Có mấy cách xếp 3 ngăn đầu đều có sách toán?

f) Có mấy cách xếp mà 3 sách toán ở 3 ngăn khác nhau?

g) Có mấy cách xếp mà 3 ngăn đầu mỗi ngăn chỉ có 1 sách toán (ngoài ra không có sách nào khác)?

h) Có mấy cách xếp mà ngăn nào cũng có sách xếp vào?

5. Xếp 5 người, trong đó có người A, B, và C vào một ghế dài 5 chỗ có đánh số. Hỏi :

a) Có mấy cách xếp?

b) Có mấy cách xếp mà A và B luôn cạnh nhau?

c) Có mấy cách xếp mà A ngồi bên phải cạnh B?

d) Có mấy cách xếp mà A ngồi bên phải B?

e) Có mấy cách xếp mà A và B không ngồi cạnh nhau?

f) Có mấy cách xếp mà A ngồi đầu ghế, B ngồi cuối ghế?

g) Có mấy cách xếp mà A và B ngồi 2 đầu ghế?

h) Có mấy cách xếp mà A luôn ngồi giữa, cạnh B và C?

i) Có mấy cách xếp mà A và B luôn cách nhau 1 người?

6. Một lô chứa 6 sản phẩm loại A, 4 sản phẩm loại B, lấy ngẫu nhiên cùng lúc ra 5 sản phẩm từ lô trên để kiểm tra.

a) Có mấy cách lấy?

b) Có mấy cách lấy trong đó có đúng 2 sản phẩm loại A?

c) Có mấy cách lấy có đúng 2 sản phẩm loại B?

7. Hộp chứa 3 quả cầu trắng, 2 quả cầu đen, 3 quả cầu xanh. Lấy cùng lúc ra 6 quả cầu.

a) Có mấy cách lấy?

b) Có mấy cách lấy mỗi loại đều có 2 quả?

c) Có mấy cách lấy có 2 quả cầu trắng, 4 quả cầu xanh?

d) Có mấy cách lấy có đúng 2 quả cầu trắng?

8. Hộp chứa 100 tấm thẻ đánh số từ 1 đến 100. Rút ngẫu nhiên 1 tấm thẻ. Tính xác suất được tấm thẻ có số đánh trên nó không chứa chữ số 5. Rút tiếp 1 tấm thẻ. Tính xác suất để tấm thẻ rút lần 2 cũng không chứa chữ số 5, biết rằng cả hai lần rút đều không hoàn lại.

9. Hộp chứa 7 quả cầu trắng, 3 quả cầu đen. Rút ngẫu nhiên cùng một lúc ra 6 quả cầu, tính xác suất để trong 6 quả cầu rút ra :

a) Có đúng 4 quả cầu trắng.

b) Có từ 5 đến 6 quả cầu trắng.

c) Tính xác suất để 4 quả cầu còn lại có cả trắng lẫn đen.

10. Hộp chứa 8 sản phẩm gồm 5 sản phẩm tốt và 3 sản phẩm xấu. Lấy ngẫu nhiên cùng một lúc ra 5 sản phẩm từ hộp.

a) Tính xác suất trong 5 sản phẩm lấy ra có đúng 4 sản phẩm tốt.

b) Lấy tiếp 1 sản phẩm (từ 3 sản phẩm còn lại). Tính xác suất được sản phẩm tốt.

11. Từ 5 chữ số 1, 2, 3, 4, 5 lập ngẫu nhiên một số tự nhiên có 3 chữ số. Tính xác suất được một số

a) Có 3 chữ số khác nhau.

b) Có 3 chữ số giống nhau.

c) Có 3 chữ số mà chữ số hàng đơn vị là 4.

d) Là bội của 4.

12. Tung 3 súc sắc. Tính xác suất

a) Xuất hiện ba mặt có số chấm đôi một khác nhau.

b) Xuất hiện mặt 5 ở súc sắc thứ 2.

c) Xuất hiện đúng một mặt 5 chấm.

d) Xuất hiện ba mặt chẵn.

e) Tổng số chấm ba mặt bằng 8.

f) Có xuất hiện mặt 1 chấm, tổng số chấm ba mặt bằng 8.

13. Xếp ngẫu nhiên 5 người lên một đoàn tàu 7 toa có đánh số. Tính xác suất:

a) 5 người lên 5 toa khác nhau.

b) Hai người là p và q trong 5 người trên lên cùng một toa.

c) Toa đầu chỉ có 2 người trong số họ lên.

d) 5 toa không có người lên.

e) 4 toa không có người lên.

14. Xếp ngẫu nhiên 5 người, trong đó có 3 người là P, Q và R vào một bàn tròn 5 chỗ có đánh số. Tính xác suất:

a) P, Q ngồi cạnh nhau,

b) P ngồi giữa và sát Q và R.

15. Khi gọi điện thoại, một khách hàng quên mất 2 chữ số cuối cùng của số điện thoại cần gọi, chỉ nhớ rằng hai số đó khác nhau và lấy ngẫu nhiên 2 chữ số thay vào. Tính xác suất anh ta gọi đúng số cần tìm qua lần gọi đầu tiên.

16. Hộp chứa 3 quả cầu trắng, 2 quả cầu đen và 5 quả cầu đỏ. Lấy ngẫu nhiên cùng lúc ra 6 quả cầu, tính xác suất 6 quả cầu lấy ra:

a) Có 2 quả cầu trắng, 2 quả cầu đen, 2 quả cầu đỏ.

b) Có 2 quả cầu trắng, 4 quả cầu đỏ.

c) Có đúng 2 quả cầu trắng.

d) Không có quả cầu trắng nào.

17. Hộp chứa 5 quả cầu trắng, 2 quả cầu xanh và 3 quả cầu đen. Rút ngẫu nhiên không hoàn lại 3 lần, mỗi lần 1 quả cầu từ hộp. Tính xác suất được:

a) Lần lượt cầu trắng, xanh, đen,

b) Ba quả cầu khác màu,

c) Ba quả cầu khác màu trong đó quả cầu thứ 3 màu đen.

18. Hộp chứa 7 quả cầu trắng, 3 quả cầu đen. Chia ngẫu nhiên 10 quả cầu ra 3 nhóm: 5 quả, 2 quả và 3 quả. Tính xác suất trong mỗi nhóm đều có 1 cầu đen.

19. Dưới tác dụng của phóng xạ, các nhiễm sắc thể của 1 tế bào có thể gãy thành 2 mảnh trong đó chỉ có 1 mảnh chứa tâm động. Sau đó các mảnh đôi một gắn lại với nhau một cách ngẫu nhiên, và tế bào sống sót nếu các cặp gắn lại chỉ chứa 1 tâm động.

Tính xác suất tế bào sống sót, biết rằng khi bị chiếu tia phóng xạ tế bào có n nhiễm sắc thể bị gãy đôi.

20. Hộp chứa 5 quả cầu trắng, 2 quả cầu xanh, 3 quả cầu đen. Từ hộp rút ngẫu nhiên không hoàn lại mỗi lần 1 cầu cho tới khi được cầu trắng thì dừng. Tính xác suất:

a) Quả cầu trắng được rút ra ở lần rút thứ 3.

b) Có 2 cầu xanh và 1 cầu đen được rút ra.

c) Không quả cầu đen nào được rút ra.

21. Hộp chứa 40 sản phẩm tốt, 3 sản phẩm hư. Rút lần lượt mỗi lần 1 sản phẩm không hoàn lại cho tới khi được 3 sản phẩm hư thì dừng. Gọi X là số sản phẩm tốt được rút ra.

Tìm $P(X = 17)$.

22. Cho một hộp chứa 7 quả cầu trắng, 3 quả cầu đen. Rút ngẫu nhiên 3 lần không hoàn lại, mỗi lần 1 quả cầu từ hộp. Tính xác suất có:

a) Quả cầu thứ nhất có màu trắng,

b) Quả cầu thứ hai có màu trắng,

c) 2 quả cầu trắng, 1 quả cầu đen.

23. Ba người bắn (độc lập) mỗi người 1 viên đạn vào 1 con mồi. Xác suất trúng con mồi của các viên đạn do người thứ 1, 2, 3 bắn lần lượt là 0.6; 0.7; 0.8. Tính xác suất con mồi

a) Bị trúng 3 viên đạn

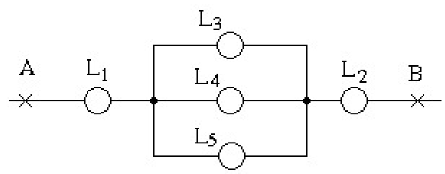
b) Không bị trúng đạn

c) Chỉ bị trúng 1 viên đạn

d) Bị trúng đạn

24. Một mạch điện tử AB gồm 5 linh kiện hoạt động độc lập với nhau và xác suất bị hư của các linh kiện L_1, L_2, L_3, L_4, L_5 trong thời gian T bằng 0.1; 0.2; 0.4; 0.7; 0.5.

Tính xác suất để mạch ngưng hoạt động trong khoảng thời gian T.



25. Lô sản phẩm gồm hai loại, kiểu dáng như nhau, trong đó số sản phẩm loại 1 gấp đôi số sản phẩm loại 2. Tỉ lệ hư của loại 1 là 3%, của loại 2 là 2%. Lấy từ lô ra 1 sản phẩm. Tính xác suất để được sản phẩm tốt.

26. Cửa hàng bán một loại sản phẩm, trong đó có 40% sản phẩm do phân xưởng I sản xuất, 60% do phân xưởng II sản xuất. Tỉ lệ sản phẩm tốt do các phân xưởng I, II sản xuất lần lượt là 80%, 90%. Mua 1 sản phẩm từ cửa hàng.

a) Tính xác suất sản phẩm mua được là sản phẩm tốt.

b) Sản phẩm mua được sau kiểm tra là sản phẩm xấu. Hỏi sản phẩm này có khả năng nhiều nhất do phân xưởng nào sản xuất ra.

27. Trong số 18 xạ thủ có 5 người bắn trúng bia với xác suất 0.8; 7 người bắn trúng bia với xác suất 0.7; 4 người bắn trúng bia với xác suất 0.6; 2 người bắn trúng bia với xác suất 0.5. Chọn ngẫu nhiên một xạ thủ, cho anh ta bắn 1 phát, kết quả không trúng bia.

a) Hỏi xạ thủ này có nhiều khả năng nhất thuộc về nhóm nào?

b) Cho xạ thủ này bắn thêm một phát nữa. Tính xác suất để anh ta bắn trúng (HD: Phải tính xác suất anh ta thuộc nhóm người nào khi biết anh ta bắn trật một lần rồi).

28. Có 2 hộp, hộp 1 chứa 8 quả cầu trắng và 2 quả cầu đen. Hộp 2 chứa 9 quả cầu trắng và 1 quả cầu đen, từ hộp 2 lấy ngẫu nhiên 2 quả cầu và bỏ vào hộp 1, sau đó từ hộp 1 lấy ngẫu nhiên cùng lúc ra 3 quả cầu.

a) Tính xác suất lấy được 2 quả cầu trắng và 1 quả cầu đen.

b) 3 quả cầu lấy ra đều có màu trắng. Tính xác suất để 3 quả cầu này đều của hộp 1.

29. Có 3 hộp, hộp I chứa 2 quả cầu trắng và 1 quả cầu đen. Hộp II chứa 3 quả cầu trắng và 1 quả cầu đen. Hộp III chứa 2 quả cầu trắng và 2 quả cầu đen.

a) Lấy ngẫu nhiên 1 hộp rồi từ đó lấy ra 1 quả cầu. Tính xác suất để được quả cầu trắng.

b) Lấy ngẫu nhiên 1 hộp, từ đó lấy ra 2 quả cầu.

Tính xác suất lấy được

i) 2 quả cầu trắng

ii) ít nhất 1 quả cầu trắng

c) Lấy ngẫu nhiên mỗi hộp 1 quả cầu. Trong 3 quả cầu lấy ra này, lấy ngẫu nhiên một quả thì được cầu trắng. Tính xác suất để quả cầu này thuộc hộp 1.

30. Một hộp đựng 10 sản phẩm: 6 sản phẩm do máy I sản xuất ra, và 4 sản phẩm do máy II sản xuất ra. Biết rằng hộp này bị mất một sản phẩm. Từ hộp lấy ngẫu nhiên cùng lúc ra 2 sản phẩm. Đặt $X = i$; $i = 0, 1, 2$; là số sản phẩm do máy II sản xuất có trong 2 sản phẩm lấy ra. Tính các xác suất $P\left(X = i\right)$. Tính xác suất để trong 2 sản phẩm lấy ra có ít nhất một sản phẩm do máy I sản xuất.

31. Trong 1 kho rượu, số lượng rượu loại A và loại B bằng nhau. Lấy ngẫu nhiên ra 1 chai và đưa cho 5 người sành rượu thử. Giả sử mỗi người đều có xác suất đoán đúng là 75%.

Có 4 người kết luận chai rượu loại A, 1 người kết luận chai rượu loại B. Hỏi khi đó xác suất để chai rượu loại A bằng mấy?

32. Một nhà máy sản xuất bóng đèn có tỉ lệ bóng đạt đại tiêu chuẩn 80%. Qua KCS (KCS là khâu *kiểm tra chất lượng sản phẩm*), một bóng đạt tiêu chuẩn được chấp nhận với xác suất 90%, một bóng không đạt tiêu chuẩn bị loại bỏ với xác suất 95%. Tính tỉ lệ bóng đạt tiêu chuẩn sau KCS (sau KCS, các sản phẩm được coi là đạt tiêu chuẩn).

33. Tại một phòng khám chuyên khoa, tỉ lệ người đến khám có bệnh là 0.8. Người ta áp dụng phương pháp mới để chẩn bệnh thì thấy: Nếu khẳng định có bệnh thì đúng 9 trên 10 trường hợp, còn nếu khẳng định không bệnh thì đúng 7 trên 10 trường hợp. Một người đến khám bệnh, tính xác suất để người này được

a) Chẩn đoán có bệnh.

b) Chẩn đoán đúng.

c) Chẩn đoán có bệnh là đúng.

34. Thống kê ở một thành phố, có bảng kết quả sau:

Số con trong một gia đình	0	1	2	3	4	5
Tỉ lệ % số gia đình có n con	15	20	30	20	10	5

Xác suất sinh ra trẻ là trai hay là gái đều bằng 0.5 và không phụ thuộc vào các trẻ khác.

a) Chọn ngẫu nhiên một gia đình. Tìm xác suất để gia đình đó có đúng 2 con gái (ngoài ra có thể có con trai).

b) Chọn ngẫu nhiên một đứa con. Tính xác suất để đứa con này thuộc gia đình có 2 con gái như trong câu a.

35. Thống kê ở một thành phố, có bảng kết quả sau:

Số con trong một gia đình	0	1	2	3	4
Tỉ lệ % số gia đình có n con	10	20	40	20	10

Xác suất sinh ra trẻ là trai hay là gái đều bằng 0.5 và không phụ thuộc vào các trẻ khác.

a) Chọn ngẫu nhiên một gia đình. Tính xác suất để gia đình đó có có ít nhất 1 con trai.

b) Chọn ngẫu nhiên một gia đình thì được gia đình có ít nhất 1 con trai. Tính xác suất để gia đình đó có ít nhất 1 con gái.

36. (Bài toán của Samuel Pépys) Biến cố nào trong các biến cố sau đây có xác suất lớn hơn:

a) Tung 6 súc sắc, có ít nhất 1 mặt 6 xuất hiện (A).

b) Tung 12 súc sắc, có ít nhất 2 mặt 6 xuất hiện (B).

c) Tung 18 súc sắc, có ít nhất 3 mặt 6 xuất hiện (C).

37. Hộp đựng 3 đồng xu, trong đó có 2 đồng xu công bằng và một đồng xu có 2 mặt hình. Chọn 1 đồng để tung, nếu được mặt hình (H) thì tung tiếp lần 2, nếu được mặt số (S) thì chọn đồng xu khác để tung tiếp.

a) Tính xác suất để hai lần tung đều ra mặt hình (H).

b) Giả sử đồng xu được tung 2 lần. Tính xác suất để đó là đồng xu có hai mặt hình (H).

38. Cặp sinh đôi được gọi là thật nếu do cùng một trứng sinh ra và trong trường hợp đó bao giờ cũng cùng giới tính. Nếu cặp sinh đôi do các trứng khác sinh ra thì xác suất để cùng giới tính bằng 0.5. Có một cặp sinh đôi cùng giới tính. Hỏi xác suất chúng là cặp sinh đôi thật bằng bao nhiêu, biết rằng xác suất để cặp sinh đôi cùng trứng bằng p.

39. Trên một toa tàu điện có n khách. Đến ga tiếp theo mỗi người có thể xuống với xác suất p.

Có 1 khách mới lên với xác suất $1-p_0$ và không có ai lên thêm với xác suất p_0. Tìm xác suất để sau lần dừng đó tàu vẫn có n khách.

40. Một lớp học có a học sinh giỏi, b học sinh khá, c học sinh yếu. Khi làm bài kiểm tra học sinh giỏi bao giờ cũng được điểm giỏi; học sinh khá có thể làm bài đạt loại giỏi hoặc khá với xác suất như nhau; một học sinh yếu có thể đạt khá, trung bình hoặc yếu với xác suất như nhau.

Sau kiểm tra, chọn ngẫu nhiên một học sinh thì được học sinh đạt điểm giỏi. Tính xác suất để đó là học sinh yếu.

41. (Bài toán Banach) Một người có 2 bao diêm, mỗi bao đều có n que. Khi cần, người đó lấy ra 1 bao và rút 1 que để dùng. Tính xác suất khi người đó phát hiện một bao rỗng thì bao kia còn lại đúng k que.

42. (Bài toán truyền máu) Biết rằng một người có máu AB có thể nhận máu của bất kì nhóm máu nào. Nếu người đó có nhóm máu còn lại (A, hoặc B, hoặc O) thì chỉ có thể nhận máu của người có cùng nhóm máu của mình hoặc người có nhóm máu O.

a) Chọn ngẫu nhiên 1 người cần tiếp máu và 1 người cho máu. Tính xác suất để sự truyền máu thực hiện được.

b) Chọn ngẫu nhiên 1 người cần tiếp máu và 2 người cho máu. Tính xác suất để sự truyền máu thực hiện được.

Cho biết tỉ lệ người có nhóm máu O, A, B, AB tương ứng là 0.337; 0.375; 0.209; 0.079.

43. Một máy tự động sản xuất ra 1 loại trục máy, có thể cho cả phế phẩm. Kiểm tra ngẫu nhiên 200 trục do máy này sản xuất ra, thấy có 25 phế phẩm. Ước lượng tỉ lệ phế phẩm do máy này sản xuất ra với độ tin cậy $\gamma = 95\%$.

44. Điều tra ngẫu nhiên 250 sinh viên, thấy có 150 sinh viên có đăng ký học thêm ngoài giờ. Hãy ước lượng tỉ lệ sinh viên đăng ký học thêm ngoài giờ với độ tin cậy 98%.

45. Điều tra ngẫu nhiên 4000 gia đình, thấy có 3200 gia đình có xe gắn máy. Ước lượng tỉ lệ các gia đình có xe gắn máy với độ tin cậy $\gamma = 95\%$.

46. Người ta đo chiều sâu của biển bằng 1 dụng cụ có sai số phép đo tuân theo phân phối chuẩn $N(0; 40^2)$. Hỏi phải đo bao nhiêu lần để kết quả đạt được có sai số không vượt quá 15(m) với độ tin cậy $\gamma = 90\%$.

47. Đo đường kính của một loại chi tiết máy do 1 máy tiện sản xuất ra (độ dài tính bằng mm) có bảng thống kê sau:

Độ dài (mm)	247	249	251	253	257
Số chi tiết	25	55	70	80	20

Ước lượng độ dài đường kính của loại trục máy này với độ tin cậy $\gamma = 95\%$. Muốn ước lượng đường kính trục máy với sai số không vượt quá $\varepsilon_0 = 4$ mm và với độ tin cậy 95%, có cần lấy thêm mẫu nữa không ? Nếu có, bằng bao nhiêu?

48. Thời gian sản xuất 1 loại chi tiết máy là 1 biến ngẫu nhiên tuân theo phân phối chuẩn $N\left(a, \sigma^2\right)$. Người ta quan sát 100 lần, thấy thời gian trung bình sản xuất ra 1 chi tiết máy là 5.5 giây với độ lệch quân phương (hiệu chỉnh) là 1.7 giây. Tìm khoảng ước lượng kì vọng a với độ tin cậy 90%.

49. Để xác định giá trị trung bình của một loại hàng bán trên thị trường, người ta điều tra ngẫu nhiên tại 100 cửa hàng với bảng số liệu sau:

Giá (nghìn đồng)	83	85	87	89	91	93	95	98	99	101
Số cửa hàng	6	7	12	15	30	10	8	6	4	2

a) Với độ tin cậy $\gamma = 95\%$, tìm khoảng tin cậy cho giá trung bình của loại hàng hóa tại thời điểm đang xét. Biết giá loại hàng tuân theo luật phân phối chuẩn.

b) Giả sử bán kính khoảng tin cậy trên là $\varepsilon = 0,5$ thì mẫu cần thiết là bao nhiêu ?

50. Kiểm tra chất lượng của cùng loại của 2 xí nghiệp có kết quả

Xí nghiệp	Số s/phẩm k/tra	Số phế phẩm
I	1800	60
II	1200	42

Với mức ý nghĩa $\alpha = 0.05$, có thể coi tỉ lệ phế phẩm ở 2 nhà máy này như nhau được không?

51. X (kg) là 1 chỉ tiêu của 1 loại sản phẩm thuộc xí nghiệp A. Điều tra một số sản phẩm của xí nghiệp này có kết quả sau:

x_I	50-55	55-60	60-65	65-70	70-75	75-80
n_I	5	10	25	30	18	12

a) Ước lượng trung bình chỉ tiêu X với độ tin cậy 92%.

b) Muốn ước lượng trung bình chỉ tiêu X với độ tin cậy với độ tin cậy 95% và độ chính xác không quá 1 kg có cần bổ sung mẫu không ? Nếu có, bằng bao nhiêu ?

c) Có tài liệu nói trung bình chỉ tiêu X bằng 70(kg) . Cho nhận xét ở mức ý nghĩa $\alpha = 0.05$

d) Các sản phẩm có chỉ tiêu $X \leq 70$ (kg) gọi là sản phẩm loại 2. Ở xí nghiệp B, điều tra 244 sản phẩm, thấy có 50 sản phẩm loại 2. Cho rằng tỉ lệ sản phẩm loại 2 ở hai xí nghiệp này như nhau được không ? Kết luận ở mức $\alpha = 0.05$

52. Trong một nhà máy may trang phục, người ta đo thời gian hoàn thành một công việc của 2 nhóm công nhân thực hiện các thao tác theo các trật tự khác nhau, được kết quả (tính bằng giây).

Nhóm A	220	235	214	197	206	214
Nhóm B	247	223	215	219	207	236

Coi số liệu tuân theo phân phối chuẩn (cùng 1 đồng hồ đo). Với mức ý nghĩa $\alpha = 0.05$, kiểm định xem thời gian trung bình để hoàn thành công việc của 2 nhóm có khác nhau nhiều không?

53. Theo dõi sự phát triển của 1 loại cây trồng có bảng sau:

Chiều cao x_i (cm)	275	325	375	425	475	525	575
Số cây n_i	5	20	25	30	30	23	14

a) Cho rằng chiều cao trung bình của cây là 450 cm được không? (kiểm định mức α = 0.05)

b) Tìm khoảng ước lượng cho chiều cao trung bình của cây với độ tin cậy $\gamma = 95\%$

54. Để khảo sát xem màu của mắt và màu của tóc có phải là các đặc tính độc lập không, người ta tiến hành quan sát 3200 người và thu được kết quả sau:

Màu tóc / Màu mắt	Vàng	Nâu	Đen và hung
Xanh da trời	872	380	112
Xanh lá cây, nâu	500	815	521

Hãy cho kết luận ở mức 0.05.

55. Tìm hàm hồi qui tuyến tính của Y theo X và tính hệ số tương quan đối với mẫu như sau:

X	2	3	7	6	5
Y	6	6	15	13	10

56. Tìm hàm hồi qui tuyến tính của Y theo X và tính hệ số tương quan đối với mẫu như sau:

$Y \backslash X$	5	10	15	20	25	30	35	40
100	2	1						
120	3	4	3					
140			5	10	8			
160						6	1	1
180							4	1

ĐÁP SỐ HOẶC HƯỚNG DẪN GIẢI MỘT SỐ BÀI TẬP

CHƯƠNG 1

1.3. a)125; b)60; c)5; d)25; e)12; f)100

1.4. a) 216; b) 120; c)6; d) 36; e) 6; f) 75

1.5. a) 16807 b) 1 c) 7 d) 120

1.6. a)120; b)12 c)48; d)60; e) 72; f) 16.

1.7. a)525 b)105 c)1266

1.8. a)30 b)105

1.9. 190

1.10. $(n^2 - 3n)/2$

1.11. a) 14190 b)4750

1.12. a)10 b)60

1.13. 300

1.14. 156

1.15. 108

1.16. 108

1.17. 30

1.18. 195

CHƯƠNG 2

2.1. $\overline{A}$: không có sản phẩm nào bị hư.

$\overline{B}$: Có nhiều nhất một sản phẩm bị hư.

2.2. $A \cap B \cap \overline{C}$: chọn được nam SV thuộc khu vực I nhưng không ở nội trú.

$\overline{A} \cap \overline{B} \cap C$:chọn được nữ SV nội trú không thuộc khu vực I

$\overline{A} = B$ xảy ra khi lớp này không có nam SV thuộc khu vực I, và tất cả nữ thuộc khu vực I.

$A \cap B \cap C = A$ xảy ra khi toàn bộ nam SV của lớp đều thuộc khu vực I và ở nội trú

$\overline{C} \subset B$ xảy ra khi mọi SV không ở nội trú của lớp đều thuộc khu vực I.

2.3. a) $A\overline{B}\overline{C}$; b) $AB\overline{C}$; c) ABC ; d) $\overline{A}\overline{B}\overline{C}$; e) Đối lập với biến cố ở câu d: $\overline{\overline{A}\overline{B}\overline{C}}$; f) $\overline{ABC}$;

g) $\overline{A}BC + A\overline{B}C + AB\overline{C} + ABC$; h) Đối lập với biến cố ở câu g; i) trùng với biến cố ở câu g; j) trùng với biến cố ở câu f, và đối lập với biến cố câu c.

2.4. Hình 1:a) $\overline{A}_1\overline{A}_2$; b) $A_1 + A_2$;

Hình 2: a) $\overline{A}_1\overline{A}_2 + \overline{A}_3$; b) $\left(A_1 + A_2\right)A_3$

2.5. $A = A_1 + A_2 + A_3$; $B = A_1\overline{A}_2\overline{A}_3 + \overline{A}_1A_2\overline{A}_3 + \overline{A}_1\overline{A}_2A_3$; $C = \overline{A}_1\overline{A}_2\overline{A}_3$

2.6. a) $\overline{A}_1\overline{A}_2$. b) $A_1 + \overline{A}_1A_2$

2.7. a) $A_1\overline{A}_2\overline{A}_3 + \overline{A}_1A_2\overline{A}_3 + \overline{A}_1\overline{A}_2A_3$;

b) $\overline{A}_1\overline{A}_2\overline{A}_3$; c) $\overline{\overline{A}_1\overline{A}_2\overline{A}_3}$

2.8. $X = \overline{B}$

CHƯƠNG 3

3.1. 1/60

3.2. 2/11

3.3. a) 5/36 b) 1/6

3.4. 81/100

3.5. a) 12/25 ; b) 1/25 ; c) 1/5

3.6. a) 5/9 ; b) 1/6 ; c) 25/72 ; d) 1/8

3.7. a) 24/625 ; b) 1/5 ; c) 128/625

3.8. a) 2/5 ; b) 2/15

3.9. $\dfrac{2(n-r-1)}{n(n-1)}$

3.10. 5/9

3.11. a) 1/15; b) 7/15 ; c) 14/15

3.12. 680/13167

3.13. $\dfrac{C_m^h C_{n-m}^{k-h}}{C_n^k}$

3.14. 50/143

3.15. 1/6

3.16. 1/9; 11/18; 5/18

3.17. a) 0.154 ; b) 0.988

3.18. 41/42

3.19. Rút ở lần nào cũng vậy

3.20. 5/72

3.21. a) 80/243 ; b) 40/81

3.22. ~0,0584

3.23. ~0,387

3.24. 11/90

3.25. a) 83/120 ; b) 80/83

3.26. 2/5

3.27. a)1.6% b) 0.25

3.28. a) 0.7916 ; b) 0.456

3.29.

	$P(A)$	$P(B)$	$P(A\cup B)$	$P(A\cap B)$	$P(A/B)$	$P(B/A)$
a)	3/4	7/20	9/10	1/5	4/7	4/15
b)	5/6	3/4	11/12	2/3	8/9	4/5

Hướng dẫn

3.1. Số cách xếp bộ sách lên giá sách là $n=5!=120$, trong đó có 2 cách xếp đúng thứ tự đó là: xếp theo thứ tự tăng dần và giảm dần theo tập. Vậy xác suất cần tính là 1/5!

3.2. Do biết trước có ít nhất 1 con xúc xắc xuất hiện mặt 5 nên số các trường hợp có thể xảy ra là 11, đó là:

(5,1),(5,2),(5,3),(5,4),(5,6),(1,5),(2,5),(3,5),(4,5),(6,5)

Rõ ràng trong các trường hợp này có 2 trường hợp có tổng số nốt xuất hiện là 7 nên xác suất để tổng số nốt xuất hiện bằng 7 là 2/11.

3.3. Lưu ý: gieo k con xúc xắc có 6^k trường hợp có thể xảy ra.

3.7. 5 người lên 5 toa tàu có $n=5^5$ cách.

a) Gọi A là biến cố 5 người lên 5 toa khác nhau. Số thuận lợi cho A là 5!.

$\Rightarrow P(A)=5!/5^5$

b) Gọi B là biến cố A và B cùng lên một toa. Để A và B cùng lên một toa, ta phải có:

- A,B cùng lên một toa: $C_5^1=5$ cách
- 3 người còn lại lên tàu: 5^3 cách nên số thuận lợi cho B là: $5.5^3 \Rightarrow P(B)=5.5^3/5^5$

c) Gọi C là biến cố chỉ có hai trong 5 người lên toa đầu tiên.

Muốn chỉ có hai trong 5 người lên toa đầu tiên, ta phải:

- Chọn ra 2 trong 5 người và xếp lên toa đầu tiên: C_5^2 cách.

- 3 người còn lại lên các toa khác: 5^3 cách.

Nên $P\left(C\right)=C_5^2.5^3/5^5$.

3.14. 10 người được chọn trong 15 người nên có $n=C_{15}^{10}$ cách. Theo đề 15 người xin việc này chia thành 2 nhóm: nhóm có kết quả tốt nghiệp cao nhất (5 người) và nhóm còn lại (10 người) nên $p=\dfrac{C_5^4.C_{10}^6}{C_{15}^{10}}$

3.15. Gọi A_i là biến cố người thủ kho mở được cửa ở lần thử thứ *i*, $i=1,2,3$

Khi đó xác suất để anh ta mở được cửa ở lần thử thứ 3 là:

$$P\left(A_3\right)=P\left(\overline{A}_1.\overline{A}_2.A_3\right)=P(\overline{A}_1).P\left(\overline{A}_2/\overline{A}_1\right).P(A_3/\overline{A}_1\overline{A}_2)=\frac{7}{9}.\frac{6}{8}.\frac{2}{7}=\frac{1}{6}$$

3.16. Gọi A là biến cố gọi được nữ sinh viên lớp A.

Gọi B là biến cố gọi được nữ sinh viên lớp B

Khi đó A, B độc lập và theo đề: $P(A)=\dfrac{10}{30}=\dfrac{1}{3}$; $P(B)=\dfrac{25}{30}=\dfrac{5}{6}$.

Ta có 3 trường hợp: Gọi được 0,1, 2 sinh viên nữ.

Gọi A_i là biến cố gọi được i nữ sinh viên, $i=0,1,2$

Vậy, xác suất các trường hợp là:

$$P(A_0)=P(\overline{A}.\overline{B})\overset{\text{đl}}{=}P(\overline{A}).P(\overline{B})=\frac{2}{3}.\frac{1}{6}=\frac{2}{18}=\frac{1}{9}.$$

$$P(A_1)=P\left(A.\overline{B}+\overline{A}.B\right)\overset{\text{xk,đl}}{=}P(A).P(\overline{B})+P(\overline{A}).P(B)=\frac{1}{3}.\frac{1}{6}+\frac{2}{3}.\frac{5}{6}=\frac{11}{18}$$

$$P(A_2)=P(A.B)\overset{\text{đl}}{=}P(A).P(B)=\frac{1}{3}.\frac{5}{6}=\frac{5}{18}$$

3.18. Gọi A là biến cố được ít nhất 1 viên phấn tốt.

Gọi A_i là biến cố lấy được viên phấn tốt từ hộp thứ i,i=1,2,3

Khi đó A_1,A_2,A_3 độc lập và $P(A_1)=\dfrac{15}{20}=\dfrac{3}{4}$; $P(A_2)=\dfrac{10}{14}=\dfrac{5}{7}$;

$$P(A_3)=\frac{20}{30}=\frac{2}{3}$$

Ta có: $P\left(\overline{A}\right)=P\left(\overline{A}_1.\overline{A}_2\overline{A}_3\right)\overset{\text{ñl}}{=}P\left(\overline{A}_1\right).P\left(\overline{A}_2\right).P\left(\overline{A}_3\right)=\dfrac{1}{4}.\dfrac{2}{7}.\dfrac{1}{3}=\dfrac{1}{42}$

Nên $P\left(A\right)=1-P\left(\overline{A}\right)=1-\dfrac{1}{42}=\dfrac{41}{42}$

3.25. a) Gọi A là biến cố thu được tín hiệu A

Gọi M là biến cố trạm phát tín hiệu A, Gọi N là biến cố trạm phát tín hiệu B. Khi đó *M, N* là hệ đầy đủ với $P\left(M\right)=0.8$; $P\left(N\right)=0.2$

Do $\dfrac{1}{6}$ tín hiệu A bị méo thành tín hiệu B nên $P\left(A/M\right)=\dfrac{5}{6}$ (xác suất thu A khi phát A) và do $\dfrac{1}{8}$ tín hiệu B bị méo thành tín hiệu A nên $P(A/N)=\dfrac{1}{8}$ (xác suất thu A khi phát B)

Theo công thức xác suất đầy đủ, ta có:

$$P\left(A\right)=P\left(M\right).P\left(A/M\right)+P\left(N\right).P\left(A/N\right)=0.8\times\frac{5}{6}+0.2\times\frac{1}{8}=\frac{83}{120}$$

b) Do thu được tín hiệu A nên xác suất thu được đúng tín hiệu lúc phát là:

$$P(M/A)=\frac{P(M).P(A/M)}{P(A)}=\frac{0.8\times\frac{5}{6}}{\frac{83}{120}}=\frac{80}{83}$$

3.26. Từ mỗi hộp rút ngẫu nhiên 1 bi nên nếu gọi A_i là biến cố lấy được i bi trắng từ hai hộp (i=0,1,2) thì theo công thức xác suất cổ điển, ta có:

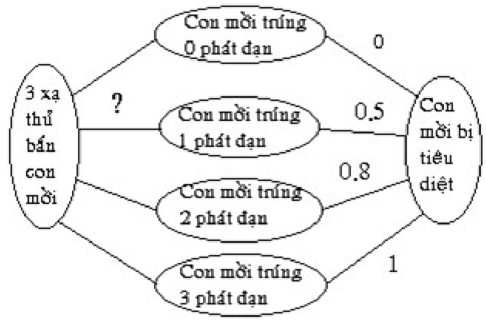
$$P(A_0)=\frac{4\times16}{10\times20}\ ;\ P(A_1)=\frac{6\times16+4\times4}{10\times20}\ ;\ P(A_2)=\frac{6\times4}{10\times20}$$

Gọi A là biến cố bi lấy ra sau cùng là bi trắng.

Rõ ràng: $P(A/A_0)=0;\ P(A/A_1)=\dfrac{1}{2};\ P(A/A_2)=1.$

Áp dụng công thức xác suất đầy đủ, ta được $P\left(A\right)=\dfrac{2}{5}$

3.28. Có thể thấy sơ đồ tóm tắt bài toán như hình dưới:



Nếu gọi A_i là biến cố con mồi bị trúng i phát đạn và gọi A là biến cố con mồi bị tiêu diệt thì ta đã biết: $P(A/A_i), i = 0,\ldots,3$

Chỉ còn phải tính $P(A_i)$ là xong. Để làm điều này, gọi B_j là biến cố xạ thủ j bắt trúng con mồi (j =1,2,3) và làm như bài tập 3.17.

Ta tính được:

$P(A_0) = 0.024$; $P(A_1) = 0.188$; $P(A_2) = 0.452$; $P(A_3) = 0.336$

Nên theo công thức xác suất đầy đủ, ta có $P(A) = 0.7916$

CHƯƠNG 4

4.1.

X	0	1	2	3
Px	1/6	1/2	3/10	1/30

4.2. Dễ thấy $X \sim B(3;0.7)$ nên luật phân phối xác suất của X là

X	0	1	2	3
Px	0.027	0.189	0.441	0.343

4.3. Gọi X là số đạn đã bắn. Rõ ràng X nhận các giá trị: 1,2,3.

Để tính $P[X = i]$, cần gọi A_i là sự kiện xạ thủ bắn lần thứ i trúng mục tiêu, i =1,2,3.

Khi đó A_1,A_2,A_3 độc lập; $P(A_i) = 0.8$ và $P[X = 1] = 0.8$

$$P[X = 2] = P(\overline{A}_1 A_2) \overset{\text{dl}}{=} P(\overline{A}_1).P(A_2) = 0.2 \times 0.8 = 0.16$$

Vì chỉ cần bắn trượt 2 viên đạn đầu thì ta sẽ bắn đến viên thứ 3 và sẽ dừng bắn bất kể viên thứ 3 bắn như thế nào do đây là viên đạn cuối cùng nên ta có

$$P\left[X=3\right]=P(\overline{A}_1\overline{A}_2)\overset{\text{đl}}{=}P(\overline{A}_1).P(\overline{A}_2)=0.2\times 0.2=0.04$$

Vậy luật phân phối xác suất của X là

X	1	2	3
Px	0.8	0.16	0.04

4.4.

X	0	1	2	3
Px	0.012	0.154	0.456	0.378

$E\left(X\right)=2.2$; $D\left(X\right)=0.54$

4.5. a) 2/15; b)

X	0	1	2
Px	1/15	8/15	6/15

4.6. Gọi A_i là sự kiện xạ thủ A bắn trúng đích *i* phát đạn (i = 0,1,2)

Gọi B_j là sự kiện xạ thủ B bắn trúng đích *j* phát đạn (j = 0,1,2)

Khi đó ta có các A_i, các B_j xung khắc từng đôi và A_i, B_j độc lập.

Theo công thức Bernoulli,

$$P(A_i)=C_2^i\left(0.4\right)^i\left(0.6\right)^{2-i} \text{ và } P(B_j)=C_2^j\left(0.5\right)^j\left(0.5\right)^{2-j}$$

Do X là số phát trúng của A trừ đi số phát trúng của B nên X nhận các giá trị: -2, -1, 0, 1, 2

$$\text{P[X=-2]=P(A}_0\text{.B}_2\text{)}\overset{\text{đl}}{=}P(A_0)P(B_2)=0.6^2\times 0.5^2=0.09$$

$$P[X=-1]=P(A_0B_1+A_1B_2)\overset{xk,\tilde{n}l}{=}P(A_0)P(B_1)+P(A_1)P(B_2)$$

$$=0.6^2\times 2\times 0.5^2+2\times 0.4\times 0.6\times 0.5^2=0.3$$

….

$\Rightarrow$ Phn phối xác suất của X:

X	-2	-1	0	1	2
Px	0.09	0.3	0.37	0.2	0.04

4.7.

X	3	4	5	6	7
Px	1/6	1/6	2/6	1/6	1/6

4.8. a) $A = 2;\ F(x) = \begin{cases} 0 & \text{nếu } x < 0 \\ x^2 & \text{nếu } x \in [0;1] \\ 1 & \text{nếu } x > 1 \end{cases}$;

$E(X) = 1;\ D(X) = 1/3$

b) $A = 1/2;\ F(x) = \begin{cases} 0 & \text{nếu } x < 0 \\ (1 - \cos x)/2 & \text{nếu } x \in [0;\pi] \\ 1 & \text{nếu } x > \pi \end{cases}$;

$E(X) = \pi/2;\ D(X) = \dfrac{\pi^2}{4} - 2$

c) $A = \pi;\ F(x) = \begin{cases} 0 & \text{nếu } x < 0 \\ \sin \pi x & \text{nếu } x \in \left[0, \dfrac{1}{2}\right] \\ 1 & \text{nếu } x > \dfrac{1}{2} \end{cases}$;

$E(X) = (\pi - 2)/(2\pi);\ D(X) = \dfrac{4\pi - 12}{4\pi^2}$

d) $A = 3;\ F(x) = \begin{cases} 0 & \text{nếu } x < 1 \\ 1 - \dfrac{1}{x^3} & \text{nếu } x \geq 1 \end{cases}$; $E(X) = 3/2;\ D(X) = 1.5$

4.9. $k = 3/64$, $p = 13/256$

4.10. $k = 3/20$, TLTB = 2,4(kg), $p = 20\%$

4.11. $c = 1/(b-a)$, $E(X) = Med(X) = (a+b)/2, D(X) = (a-b)^2/12$

4.12. $a = b = 1/2$

4.13. a) $p = 23/60$

4.14. a.

X	0	1	2

Px	$\frac{10}{21}$	$\frac{9}{21}$	$\frac{2}{21}$

b. $p = 2/7$

4.15. Trò chơi có lợi cho người tổ chức (người cầm cái).

HD: Tính kỳ vọng số tiền thu về sau 1 lần tham gia.

4.16. Mua ít nhất 29 vé.

4.17. Có thể coi như số tai nạn xảy ra trong một tháng là một BSNN Y tuân theo luật phân phối Poisson với tham số bằng 2.

a) Gọi X là số tai nạn xảy ra trong khoảng thời gian 3 tháng, theo đề bài ta có X~P(6)

$\Rightarrow P[X \le 3] = P[X = 0] + P[X = 1] + P[X = 2] + P[X = 3] \approx 0.151$

b) $P[Y \le 1] \approx 0.406$, suy ra đáp số: 0.067

4.18. a) Luật ppxs của số tiền thu về của trạm trong một ngày là:

X	-24	-4	16	36
Px	0.0608	0.1703	0.2384	0.5305

Suy ra *E(X)* = 20.8 (USD)

b) Tương tự số tiền trung bình trạm thu về trong một ngày khi có 4 chiếc xe là 18.9 (USD) c). Từ a) và b) ta kết luận trạm chỉ nên có 3 chiếc xe.

4.19. ≈ 0.061

4.20. ≈ 0.63.

CHƯƠNG 5

5.1. $\bar{x} = 3.039; \sigma_n^2 \approx 0.86; \sigma_{n-1}^2 \approx 0.1$

5.2. $\bar{x} = 2.38;\ \sigma_n^2 = 1.78$

5.3. $\bar{x} = 19.22;\ \sigma_n^2 = 212.532$

CHƯƠNG 6

6.1. 19.116; 0.055499; 0.235582

6.2. $\overline{x} = 5325$; (5117.11, 5532.89)

6.3. 63.36%

6.4. a) 1068 người b) (7%; 17%) c) 451 người

6.5. 3458 hạt

6.6. 1537 người.

6.7. $p \in (0.1005; 0.1999)$; số cá $\in$ (5013; 9950))

6.8. Số cò từ 784 đến 28401 con

6.9. $\mu \in$ (24.33718; 27.218371); $\sigma \in$ (3.58; 20.33)

6.10. (0.9964;0.1036)

6.11. a) (35.47; 36.01); b) 0.05; (0.0073; 0.0927)

CHƯƠNG 7

7.1. t_{qs} = 1.39; không đủ cơ sở để cho rằng việc áp dụng kỹ thuật mới hiệu quả hơn.

7.2. t_{qs} = 2.38>1.96: tỉ lệ này có xu hướng tăng lên.

7.3. t_{qs} = 0.8165: không đủ cơ sở để bác kết luận của chủ hiệu.

7.4. t_{qs} =0.543079: không đủ cơ sở để bác kết luận của hãng điều tra.

7.5. t_{qs} =1.3: chưa đủ cơ sở để kết luận đặc điểm nghề nghiệp của ngành sản xuất này có ảnh hưởng đến tỉ lệ tử vong của bệnh A.

7.6. a) (47.3% ; 60.18%) b) Có thể kết luận việc cải tiến làm tăng tỉ lệ sản phẩm loại A.

7.7. t_{qs} = 3.3: tỉ lệ trẻ sơ sinh có trọng lượng trên 3000 gam ở thành phố cao hơn ở nông thôn.

7.8. $\overline{x}$ =0.9656, $\sigma_{n-1} \approx 0.02$; $t_{qs} \approx 6.9$: Nghi ngờ trên là có cơ sở

7.9. a. 15.4 tạ; b. t_{qs} =–1.069; c.(53.96; 98.04)

7.10. a). $\overline{x} \approx 2.722$; $\sigma_{n-1} \approx 1.86$, KUL:(2.11;3.33)

b). t_{qs} = 0.8956; không đủ cơ sở để kết luận rằng việc cải tiến sản xuất có tác động đến sản xuất

7.11. t_{qs} = 0.24: Việc bón phân đạm không mang lại hiệu quả.

7.12. a) KUL: (4.97; 5.32); b) t_{qs} =2.913:Việc lai tạo có mang lại hiệu quả trong việc làm tăng chỉ số mỡ; c) $n \geq 175$

7.13. $t_{qs} \approx 13.19 > 9.21$: Có cơ sở để cho rằng sự nghỉ việc phụ thuộc vào giới tính.

7.14. $t_{qs} = 10$: Việc tiêm chủng và sự mắc bệnh có liên quan nhau.

CHƯƠNG 8

8.1. a) $y = 27.81$ x-7.56 ; b) 0.86

8.2. a) $y = 0.673$x+7.184; b) 2400.9

8.3. $y = -0.08$x+13.64

ÔN TẬP

1. a)125; b)60; c)5; d)25; e)12; f)100; g)48

2. a) 216; b)120; c)6; d)36; e) 6; f) 75; g) 125; h) 27; i)15; j) 21.

3. a) 7^5; b)1; c)7; d)120; e)2520; f)343; g)2401; h)216; i)630; j)8400; k)150; l)5250

4. a)78125; b)625; c)3125; d)16875; e)3750; f)37500; g)96; h)16800

5. a)120; b)48; c)24; d)60; e)72; f)6; g)12; h)12; i)36

6. a)252; b)60; c)120

7. a)120; b)30; c)15; d)105

8. 0.81; 0.81

9. a) 0.5; b) $\frac{1}{3}$; c) $\frac{5}{6}$

10. a) $\frac{15}{56}$; b) 0.625

11. a) 0.48; b) 0.04; c) 0.20; d) 0.20

12. a) $\frac{5}{9}$; b) $\frac{1}{6}$; c) $\frac{25}{72}$; (d) 0.125; e) $\frac{7}{72}$; f) $\frac{5}{72}$

13. a)≈ 0.15; b) $\frac{1}{7}$; c)≈ 0.1285; d) ≈ 0.0375; e) ≈ 0.3124

14. a) 0.5; b) $\frac{1}{6}$

15. $\frac{1}{90}$

16. a) $\frac{1}{7}$; b) $\frac{1}{14}$; c) 0.5; d) $\frac{1}{30}$

17. a) $\frac{1}{24}$; b) 0.25; c) $\frac{1}{12}$

18. 0.25

19. $\frac{2^n(n!)^2}{(2n)!}$

20. a) $\frac{5}{36}$; b) $\frac{1}{56}$; c) $\frac{5}{8}$

21. ≈ 0.01386

22. a) 0.7 b) 0.7 c) 0.525

23. a) 0.366 b) 0.024 c) 0.188 d) 0.976

24. 0.3808

25. ≈ 0.9733

26. a) 0.86

b) Có nhiều khả năng nhất do phân xưởng I sản xuất ra

27. a) có khả năng nhất thuộc về nhóm 7 người.

b) $\frac{373}{570} \approx 0.6544$

28. a) $\frac{117}{275}$ b) $\frac{70}{141}$

29. a) $\frac{23}{36}$ b) i/ $\frac{1}{3}$ ii/ $\frac{17}{18}$ c) ≈ 0.348

30. $P(X=0)=\frac{1}{3}$; $P(X=1)=\frac{8}{15}$; $P(X=2)=\frac{2}{15}$; $\frac{13}{15}$

31. $\frac{27}{28}$

32. $\frac{72}{73}$

33. a) 0.78 b) 0.86 c) ≈ 0.923

34. a) $\frac{13}{64}$ b) $\frac{65}{272}$

35. a) $\frac{107}{160}$ b) $\frac{70}{107}$

36. $P(A) \approx 0.665 > P(B) \approx 0.62 > P(C) \approx 0.5973$

37. a) 0.5 b) 0.5

38. $\frac{2p}{p+1}$

39. $(1-p)^{n-1}\left[p_0(1-p)+np(1-p_0)\right]$

40. $\frac{c}{3(a+b/2+c/3)}$

41. $C_{2n-k}^{n} \cdot 2^{k-2n}$

42. a) 0.5737 b) 0.7777

43. (0.079, 0.171)

44. (0.528, 0.672)

45. (0.788, 0.812)

46. 20 lần

47. (250.962mm, 251.598mm); Không cần lấy thêm mẫu.

48. a $\in$ (5.22 giây, 5.779 giây),

49. a) (89.94 nghìn đồng , 91.62 nghìn đồng), b) $n \geq 280$

50. Có thể xem tỉ lệ phế phẩm của hai nhà máy là như nhau.

51. a) (65.44 kg, 67.76 kg) ; b) Quan sát thêm 68 sp nữa.

c) Tài liệu cung cấp thông tin không đúng thực tế.

d) Không thể coi tỉ lệ sản phẩm loại 2 ở hai xí nghiệp này như nhau.

52. $t_{qs} \approx 1.3$: Thời gian trung bình giữa hai nhóm khác nhau không nhiều.

53. a) Tạm chấp nhận được. b) (424.75 cm, 451.1 cm)

54. $t_{qs} \approx 463.9$ Màu mắt và tóc có liên quan nhau (có thể do di truyền)

55. $y = 2.07x + 0.28$; $r_{xy} = 0.99$

BẢNG HÀM LAPLACE

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-t^2/2} dt$$

X	0	1	2	3	4	5	6	7	8	9
0.0	0.00000	00399	00798	01197	01596	01994	02393	02791	03189	03586
0.1	03983	04380	04776	05172	05567	05962	06356	06750	07143	07535
0.2	07926	08317	08707	09096	09484	09871	10257	10642	11026	11410
0.3	11791	12172	12552	12930	13308	13683	14058	14431	14803	15174
0.4	15543	15910	16276	16641	17003	17365	17724	18083	18439	18794
0.5	19147	19498	19847	20195	20540	20884	21226	21566	21905	22241
0.6	22575	22907	23237	23566	23892	24216	24538	24857	25175	25491
0.7	25804	26115	26424	26731	27035	27337	27637	27935	28231	28524
0.8	28815	29103	29389	29673	29955	30234	30511	30785	31057	31327
0.9	31594	31859	32122	32382	32639	32895	33147	33398	33646	33891
1.0	34135	34375	34614	34850	35083	35314	35543	35769	35993	36214
1.1	36434	36650	36864	37076	37286	37493	37698	37900	38100	38298
1.2	38493	38686	38877	39065	39251	39435	39617	39796	39973	40148
1.3	40320	40490	40658	40824	40988	41149	41309	41466	41621	41774
1.4	41924	42073	42220	42364	42507	42647	42786	42922	43056	43189
1.5	43319	43448	43575	43699	43822	43943	44062	44179	44295	44408
1.6	44520	44630	44738	44845	44950	45053	45154	45254	45352	45449
1.7	45544	45637	45728	45819	45907	45994	46080	46164	46246	46327
1.8	46407	46485	46562	46638	46712	46784	46856	46926	46995	47062
1.9	47128	47193	47257	47320	47381	47441	47500	47558	47615	47670
2.0	47725	47778	47831	47882	47933	47982	48030	48077	48124	48169
2.1	48214	48257	48300	48341	48382	48422	48461	48500	48537	48574
2.2	48610	48645	48679	48713	48745	48778	48809	48840	48870	48899
2.3	48928	48956	48983	49010	49036	49061	49086	49111	49134	49158
2.4	49180	49202	49224	49245	49266	49286	49305	49324	49343	49361
2.5	49379	49396	49413	49430	49446	49461	49477	49492	49506	49520
2.6	49534	49547	49560	49573	49585	49598	49609	49621	49632	49643
2.7	49653	49664	49674	49683	49693	49702	49711	49720	49728	49736
2.8	49744	49752	49760	49767	49774	49781	49788	49795	49801	49807
2.9	49813	49819	49825	49831	49836	49841	49846	49851	49856	49861
3.0	0.49865		**3.1**	48903	**3.2**	49931	**3.3**	49952	**3.4**	49966
3.5	49977		**3.6**	49984	**3.7**	49989	**3.8**	49993	**3.9**	49995
4.0	499968									

BẢNG PHÂN PHỐI STUDENT: $X \sim t(k)$

$$P\left(\left|X\right| > t_{k,\alpha}\right) = \alpha$$

α / k	0.20	0.10	0.05	0.02	0.01	0.001
1.	3.078	6.314	12.706	31.820	63.657	636.60
2.	1.886	2.920	4.303	6.965	9.925	31.600
3.	1.638	2.353	3.182	4.541	5.841	12.922
4.	1.533	2.132	2.776	3.747	4.604	8.610
5.	1.476	2.015	2.571	3.365	4.032	6.869
6.	1.440	1.943	2.447	3.143	3.707	5.959
7.	1.415	1.895	2.365	2.998	3.499	5.408
8.	1.397	1.860	2.306	2.896	3.355	5.041
9.	1.383	1.833	2.262	2.821	3.250	4.781
10.	1.372	1.812	2.228	2.764	3.169	4.587
11.	1.363	1.796	2.201	2.718	3.106	4.437
12.	1.356	1.782	2.179	2.681	3.055	4.318
13.	1.350	1.771	2.160	2.650	3.012	4.221
14.	1.345	1.761	2.145	2.624	2.977	4.140
15.	1.341	1.753	2.131	2.602	2.947	4.073
16.	1.337	1.746	2.120	2.583	2.921	4.015
17.	1.333	1.740	2.110	2.567	2.898	3.965
18.	1.330	1.734	2.101	2.552	2.878	3.922
19.	1.328	1.729	2.093	2.539	2.861	3.883
20.	1.325	1.725	2.086	2.528	2.845	3.850
21.	1.323	1.721	2.080	2.518	2.831	3.819
22.	1.321	1.717	2.074	2.508	2.819	3.792
23.	1.319	1.714	2.069	2.500	2.807	3.767
24.	1.318	1.711	2.064	2.492	2.797	3.745
25.	1.316	1.708	2.060	2.485	2.787	3.725
26.	1.315	1.706	2.056	2.479	2.779	3.707
27.	1.314	1.703	2.052	2.473	2.771	3.690
28.	1.313	1.701	2.048	2.467	2.763	3.674
29.	1.311	1.699	2.045	2.462	2.756	3.659
30.	1.282	1.645	1.960	2.326	2.750	3.291

BẢNG PHÂN PHỐI KHI BÌNH PHƯƠNG $X \sim \chi^2(n)$

$$P\left[X > \chi^2(n,\alpha)\right] = \alpha$$

α \ n	0.995	0.99	0.975	0.95	0.05	0.025	0.01	0.005
1.	0.000039	0.000157	0.00098	0.0039	3.841	5.024	6.635	7.879
2.	0.0100	0.0201	0.0506	0.103	5.991	7.378	9.210	10.597
3.	0.0717	0.115	0.216	0.352	7.815	9.348	11.345	12.838
4.	0.207	0.297	0.484	0.711	9.488	11.143	13.277	14.860
5.	0.412	0.554	0.831	1.145	11.071	12.833	15.086	16.749
6.	0.676	0.872	1.237	1.635	12.592	14.449	16.812	8.548
7.	0.989	1.239	1.169	2.167	14.067	16.013	18.475	20.278
8.	1.344	1.646	2.180	2.733	15.507	17.535	20.090	21.955
9.	1.735	2.088	2.700	3.325	16.919	19.023	21.666	23.590
10.	2.156	2.558	3.247	3.940	18.307	20.483	23.209	25.188
11.	2.603	3.053	3.816	4.575	19.675	21.920	24.725	26.758
12.	3.074	3.571	4.404	5.226	21.026	23.337	26.217	28.299
13.	3.565	4.107	5.009	5.892	22.362	24.736	27.688	29.820
14.	4.075	4.660	5.629	6.571	23.685	26.119	29.142	31.320
15.	4.601	5.229	6.262	7.261	24.996	27.489	30.578	32.801
16.	5.142	5.812	6.908	7.962	26.296	28.845	32.000	34.268
17.	5.697	6.408	7.564	8.672	27.587	30.191	33.409	35.717
18.	6.265	7.015	8.231	9.390	28.896	31.526	34.805	37.156
19.	6.844	7.633	8.907	10.117	30.144	32.853	36.191	38.581
20.	7.434	8.260	9.591	10.851	31.410	34.170	37.566	39.997
21.	8.034	8.897	10.283	11.591	32.671	35.479	38.932	41.400
22.	8.643	9.542	10.982	12.338	33.926	36.781	40.289	42.796
23.	9.260	10.196	11.689	13.091	35.172	38.075	41.638	44.184
24.	9.886	10.856	12.401	13.848	36.415	39.364	42.980	45.559
25.	10.520	11.524	13.120	14.611	37.652	40.646	44.314	46.930
26.	11.160	12. 198	13.844	15.379	38.885	41.924	45.643	48.290
27.	11.808	12. 878	14.573	16.151	40.113	43.195	46.963	49.647
28.	12.461	13.565	15.308	16.928	41.337	44.461	48.278	50.994
29.	13.121	14. 256	16.047	17.708	42.557	45.722	49.588	52.338
30.	13.787	14.953	16.791	18.493	43.773	46.979	50.892	53.673

MỘT SỐ THUẬT NGỮ TIẾNG ANH

□ Biến cố ngẫu nhiên	□ *Random event*
□ Biến ngẫu nhiên liên tục	□ *Continuous random variable*
□ Biến ngẫu nhiên rời rạc	□ *Discrete random variable*
□ Biến số ngẫu nhiên	□ *Random variable*
□ Chỉnh hợp	□ *Arrangement*
□ Đám đông	□ *Population*
□ Độ lệch tiêu chuẩn	□ *Standard deviation*
□ Hàm mật độ	□ *Density function.*
□ Hàm phân phối	□ *Distribution function.*
□ Hệ số tương quan tuyến tính	□ *Linear correlation coefficient*
□ Hồi qui phi tham số	□ *Nonparametric regression*
□ Hồi qui tuyến tính	□ *Linear regression.*
□ Hồi qui bậc 2	□ *Quadratic regression.*
□ Hoán vị	□ *Permutation*
□ Kết cục	□ *Outcome*
□ Không gian mẫu	□ *Sample space*
□ Khoảng tin cậy	□ *Confidence interval*
□ Kiểm định giả thiết	□ *Test hypothesis*
□ Kỳ vọng	□ *Expectation*
□ Mẫu	□ *Sample*
□ Phân phối nhị thức	□ *Binomial distribution*

Phân phối chuẩn	*Normal distribution*
Phân phối Poát-Xông	*Poission distribution*
Phân phối Student	*Student distribution*
Phép thử	*Trial, experiment*
Phương pháp bình phương bé nhất	*Least squares method*
Phương sai	*Variance*
Phương trình hồi qui	*Regression equation*
Sự quan sát	*Observation*
Sự kiện, biến cố	*Event*
Tỉ lệ	*Rate, proportion*
Thống kê	*Statistics*
Tổ hợp	*Combination*
Tổng bình phương	*Sum of squares*
Tương quan tuyến tính	*Linear correlation*
Trung bình	*Mean*
Ước lượng	*Estimation*
Ước lượng không chệch	*Unbias estimation*
Xác suất	*Probability*