Xác suất và thống kê

BỘ CÔNG THƯƠNG
TRƯỜNG ĐẠI HỌC CÔNG NGHIỆP TP. HCM

Nguyễn Đức Phương

Bài giảng
Xác suất & thống kê

MSSV: . . . .
Họ tên: . . . .

TP. HCM – Ngày 24 tháng 12 năm 2010

Mục lục

Chương 1

Biến cố, xác suất của biến cố

1.1 Phép thử, biến cố

- Phép thử là việc thực hiện một thí nghiệm hoặc quan sát một hiện tượng nào đó. Phép thử được gọi là ngẫu nhiên nếu ta không thể dự báo trước chính xác kết quả nào sẽ xảy ra.

- Mỗi kết quả của phép thử, ω được gọi là một biến cố sơ cấp.

Ví dụ 1.1. Thực hiện phép thử tung một đồng xu. Có hai kết quả có thể xảy ra khi tung đồng xu là xuất hiện mặt **sấp-S** hoặc mặt **ngửa-N**:

- Kết quả $\omega = S$ là một biến cố sơ cấp.
- Kết quả $\omega = N$ là một biến cố sơ cấp. □

- Tập hợp tất cả các kết quả, ω có thể xảy ra khi thực hiện phép thử gọi là không gian các biến cố sơ cấp, ký hiệu là Ω.

Ví dụ 1.2. Tung ngẫu nhiên một con xúc sắc. Quan sát số chấm trên mặt xuất hiện của xúc sắc, ta có 6 kết quả có thể xảy ra đó là:1, 2, 3, 4, 5, 6. Không gian các biến cố sơ cấp, $\Omega = \{1, 2, 3, 4, 5, 6\}$. Số phần tử của Ω, $|\Omega| = 6$. □

- Mỗi tập con của không gian các biến cố sơ cấp gọi là biến cố.

Ví dụ 1.3. Thực hiện phép thử tung một xúc sắc. Ta đã biết $\Omega = \{1, 2, 3, 4, 5, 6\}$

- Đặt $A = \{2, 4, 6\} \subset \Omega$, A gọi là biến cố "Số chấm trên mặt xuất hiện là số chẵn". Thay vì liệt kê các phần tử của A, ta đặt tên cho A

A: "Số chấm trên mặt xuất hiện là số chẵn"

- Ngược lại, nếu ta gọi biến cố:

  B: "Số chấm trên mặt xuất hiện lớn hơn 4"

  thì khi đó $B = \{5, 6\}$ □

- Xét biến cố A, khi thực hiện phép thử ta được kết quả ω.

- Nếu trong lần thử này kết quả $\omega \in A$ ta nói biến cố A xảy ra.
- Ngược lại nếu trong lần thử này kết quả $\omega \notin A$ ta nói biến cố A không xảy ra.

Ví dụ 1.4. Một sinh viên thi kết thúc môn xác suất thống kê.

Gọi các biến cố:

A: "Sinh viên này thi đạt" $A = \{4; \ldots; 10\}$

- Giả sử sinh viên này đi thi được kết quả $\omega = 6 \in A$ lúc này ta nói biến cố A xảy ra (Sinh viên này thi đạt).
- Ngược lại nếu sinh viên này thi được kết quả $\omega = 2 \notin A$ thì ta nói biến cố A không xảy ra (Sinh viên này thi không đạt). □

1.2 Quan hệ giữa các biến cố

a) Quan hệ kéo theo $(A \subset B)$: Nếu biến cố A xảy ra thì kéo theo biến cố B xảy ra.

Ví dụ 1.5. Theo dõi 3 bệnh nhân phỏng đang được điều trị. Gọi các biến cố:

Gọi các biến cố:

A_i: "Có i bệnh nhân tử vong", $i = 0, 1, 2, 3$
B : "Có nhiều hơn một bệnh nhân tử vong"

Ta có $A_2 \subset B$, $A_3 \subset B$, $A_1 \not\subset B$ □

b) Hai biến cố A và B được gọi là bằng nhau nếu $A \subset B$ và $B \subset A$, ký hiệu $A = B$.

c) Biến cố tổng $A + B$ $(A \cup B)$ xảy ra khi và chỉ khi A xảy ra hoặc B xảy ra trong một phép thử. (Ít nhất một trong hai biến cố xảy ra)

Ví dụ 1.6. Hai xạ thủ cùng bắn vào một mục tiêu, mỗi người bắn một phát. Gọi các biến cố:

Gọi các biến cố:

A: "Người thứ nhất bắn trung mục tiêu"
B: "Người thứ hai bắn trúng mục tiêu"

Biến cố $A + B$: "Có it nhất một người bắn trúng mục tiêu" □

d) Biến cố tích AB $(A \cap B)$ xảy ra khi và chỉ khi cả hai biến cố A và B cùng xảy ra trong một phép thử.

Ví dụ 1.7. Một sinh viên thi kết thúc 2 môn hoc. Gọi các biến cố:

Gọi các biến cố:

A: "Sinh viên thi đạt môn thứ nhất"
B: "Sinh viên thi đạt môn thứ hai"

Biến cố AB: "Sinh viên thi đạt cả hai môn" □

e) Hai biến cố A và B gọi là xung khắc nếu chúng không cùng xảy ra trong một phép thử $(AB = \emptyset)$.

f) Biến cố không thể: là biến cố không bao giờ xảy ra khi thực hiện phép thử, ký hiệu $\emptyset$.

g) Biến cố chắc chắn: là biến cố luôn xảy ra khi thực hiện phép thử, ký hiệu Ω.

h) Biến cố $\bar{A}$ được gọi là biến cố bù của biến cố A hay ngược lại khi và chỉ khi

$$\begin{cases} A \cap \bar{A} = \emptyset \\ A \cup \bar{A} = \Omega \end{cases}$$

1.3 Định nghĩa xác suất

Định nghĩa 1.1 (Định nghĩa cổ điển)**.** *Xét một phép thử đồng khả năng, có không gian các biến cố sơ cấp*

$$\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}, \ |\Omega| = n < \infty$$

$A \subset \Omega$ *là một biến cố. Xác suất xảy ra biến cố* A*, ký hiệu* $\mathbb{P}(A)$

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|} = \frac{\text{số trường hợp thuận lợi đối với } A}{\text{số trường hợp có thể}}$$

Ví dụ 1.8. Gieo một con xúc sắc cân đối. Tính xác suất số chấm trên mặt xuất hiện lớn hơn 4.

Giải. ______________________________

____

____

____

□

Ví dụ 1.9. Xếp ngẫu nhiên 5 sinh viên vào một ghế dài có 5 chỗ ngồi. Tính xác suất hai người định trước ngồi cạnh nhau.

Giải. ______________________________

____

____

____

□

Tính chất 1.2 (Tính chất của xác suất)**.** *Xác suất có các tính chất:*

i. $0 \leq \mathbb{P}(A) \leq 1$ *với mọi biến cố* A.

ii. $\mathbb{P}(\emptyset) = 0$, $\mathbb{P}(\Omega) = 1$.

iii. Nếu $A \subset B$ *thì* $\mathbb{P}(A) \leq \mathbb{P}(B)$.

iv. $\mathbb{P}(A) = 1 - \mathbb{P}\left(\bar{A}\right)$.

Ví dụ 1.10. Một lọ đựng 4 bi trắng và 6 bi đen. Từ lọ lấy ra ngẫu nhiên 3 bi, tính xác suất lấy được:

a) Hai bi trắng.

b) Ít nhất một bi trắng.

Giải. ____

$\square$

Chú ý: Trong câu b), chúng ta tính xác suất của biến cố bù sẽ đơn giản hơn. Ta có

$$\bar{B} : \text{“Lấy được không bi trắng”}$$

$$\mathbb{P}(B) = 1 - \mathbb{P}\left(\bar{B}\right) = 1 - \frac{C_4^0 C_6^3}{C_{10}^3}$$

1.4 Xác suất có điều kiện, sự độc lập

1.4.1 Xác suất có điều kiện

Định nghĩa 1.3 (Xác suất có điều kiện)**.** *$\mathbb{P}(A|B)$ là xác suất xảy ra biến cố A biết rằng biến cố B đã xảy ra ($\mathbb{P}(B) > 0$).*

Ví dụ 1.11. Một lọ có 4 viên bi trắng và 6 viên bi đen. Từ lọ này lấy lần lượt ra 2 viên bi, mỗi lần lấy một bi (lấy không hoàn lại). Tìm xác suất để lần lấy thứ hai được viên bi trắng biết lần lấy thứ nhất đã lấy được viên bi trắng.

Giải.

$$\begin{cases} 4 \text{ bi trắng} \\ 6 \text{ bi đen} \end{cases} \xrightarrow[\text{đã lấy 1 bi trắng}]{B \text{ xảy ra}} \begin{cases} 3 \text{ bi trắng} \\ 6 \text{ bi đen} \end{cases}$$

□

Ví dụ 1.12. Từ một bộ bài tây (4 chất, 52 lá), rút ngẫu nhiên ra 2 lá. Tính xác suất:

a) Rút được hai lá bài cơ.

b) Rút được 2 lá bài cơ biết rằng 2 lá bài này màu đỏ.

Giải.

□

Ví dụ 1.13. Một nhóm 100 người có:

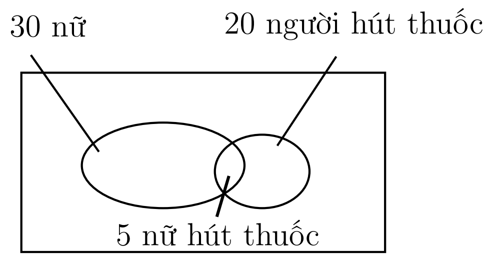
+ 20 người hút thuốc.

+ 30 nữ, trong đó có 5 người hút thuốc.

Chọn ngẫu nhiên một người trong nhóm 100 người này. Tính xác suất:

a. Người này hút thuốc biết rằng người này là nữ.

b. Người này là nữ biết rằng người này hút thuốc.



Giải. ______________________________

____

____

____

____

____

____

____

□

Công thức xác suất điều kiện

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(AB)}{\mathbb{P}(B)}, \quad \mathbb{P}(B) > 0$$

Tính chất 1.4. *Xác suất có điều kiện có các tính chất:*

i. $0 \leq \mathbb{P}(A|B) \leq 1$ *với mọi biến cố* A.

ii. *Nếu* $A \subset A'$ *thì* $\mathbb{P}(A|B) \leq \mathbb{P}(A'|B)$.

iii. $\mathbb{P}(A|B) = 1 - \mathbb{P}\left(\bar{A}|B\right)$.

Ví dụ 1.14. Một công ty cần tuyển 4 nhân viên. Có 10 người nộp đơn dự tuyển, trong đó có 4 nữ (khả năng trúng tuyển của các ứng cử viên là như nhau). Tính xác suất:

a) Cả 4 nữ trúng tuyển.

b) Có ít nhất một nữ trúng tuyển.

c) Cả 4 nữ trúng tuyển, biết rằng có ít nhất một nữ đã trúng tuyển.

Giải. ____

____

____

____

____

____

____

____

____

____

____

____

□

1.4.2 Sự độc lập của hai biến cố

A và B là hai biến cố độc lập nếu B có xảy ra hay không cũng không ảnh hưởng đến khả năng xảy ra A và ngược lại, nghĩa là:

$$\mathbb{P}(A|B) = \mathbb{P}(A) \text{ hoặc } \mathbb{P}(B|A) = \mathbb{P}(B)$$

Nhận xét: Nếu hai biến cố A và B độc lập thì các cặp biến cố A và $\bar{B}$; $\bar{A}$ và B; $\bar{A}$ và $\bar{B}$ độc lập.

Ví dụ 1.15. Tung một xúc sắc 2 lần. Gọi các biến cố:

Gọi các biến cố:

A: "Lần 1 xuất hiện mặt 6 chấm"
B: "Lần 2 xuất hiện mặt 6 chấm"

Hai biến cố A và B có độc lập?

Giải. ____

____

____

____

____

____

____

____

□

Ví dụ 1.16. Một lọ đựng 4 bi trắng và 6 bi đen, thực hiện hai lần lấy bi. Mỗi lần lấy 1 bi (lấy không hoàn lại). Đặt các biến cố:

Gọi các biến cố:

> $\boldsymbol{A}$: "Lần 1 lấy được bi đen"
> $\boldsymbol{B}$: "Lần 2 lấy được bi trắng"

Hai biến cố $\boldsymbol{A}$ và $\boldsymbol{B}$ có độc lập?

Giải. ____

____

____

____

____

____

____

____

□

1.5 Các công thức tính xác suất

1.5.1 Công thức cộng

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

Chú ý: Nếu A và B xung khắc ($AB = \emptyset$) thì

$$\mathbb{P}(A+B) = \mathbb{P}(A) + \mathbb{P}(B)$$

Ví dụ 1.17. Một lớp học có 20 học sinh trong đó có 10 học sinh giỏi toán, 8 học sinh giỏi văn và 6 học sinh giỏi cả toán và văn. Chọn ngẫu nhiên một học sinh, tính xác suất học sinh này giỏi ít nhất một môn.

Giải. ____________________

____

____

____

____

____

____

____

□

Công thức cộng 3 biến cố:

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

Chú ý: Nếu A, B, C xung khắc từng đôi một thì

$$\mathbb{P}(A+B+C) = \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C)$$

1.5.2 Công thức nhân

$$\mathbb{P}(AB) = \mathbb{P}(A)\,\mathbb{P}(B|A) = \mathbb{P}(B)\,\mathbb{P}(A|B)$$

Chú ý: Nếu A và B độc lập thì $\mathbb{P}(AB) = \mathbb{P}(A)\mathbb{P}(B)$

Mở rộng công thức nhân: Cho n biến cố $A_1, A_2, \ldots, A_n$

$$\mathbb{P}(A_1A_2\ldots A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2|A_1)\ldots\mathbb{P}(A_n|A_1A_2\ldots A_{n-1})$$

Chú ý: Nếu $A_i, i = 1, \ldots, n$ độc lập toàn bộ thì

$$\mathbb{P}(A_1\ldots A_n) = \mathbb{P}(A_1)\ldots\mathbb{P}(A_n)$$

Ví dụ 1.18. Một người có 4 con gà mái, 6 con gà trống nhốt trong một lồng. Hai người đến mua (người thứ nhất mua xong rồi đến lượt người thứ hai mua, mỗi người mua 2 con) và người bán bắt ngẫu nhiên từ lồng. Tính xác suất người thứ nhất mua được một gà trống và người thứ hai mua hai gà trống.

Giải.

□

Ví dụ 1.19. Trong một kỳ thi, mỗi sinh viên phải thi 2 môn. Một sinh viên A ước lượng rằng: xác suất đạt môn thứ nhất là 0,8. Nếu đạt môn thứ nhất thì xác suất đạt môn thứ hai là 0,6; nếu không đạt môn thứ nhất thì xác suất đạt môn thứ hai là 0,3. Tính xác suất sinh viên A:

a. Đạt môn thứ hai.

b. Đạt i môn, $i = 0, 1, 2$.

c. Đạt ít nhất một môn.

d. Đạt môn thứ hai biết rằng sinh viên này đạt một môn.

e. Đạt môn thứ hai biết rằng sinh viên này đạt ít nhất một môn.

Giải.

□

Ví dụ 1.20. Một người có 3 con gà mái, xác suất đẻ trứng trong ngày của con gà I, II, III lần lượt là 0,4; 0,7; 0,8. Tính xác suất:

a) Có i con gà đẻ trứng trong ngày, $i = 0, 1, 2, 3$.

b) Có ít nhất 1 con gà đẻ trứng trong ngày.

c) Có nhiều nhất 2 con gà đẻ trứng trong ngày.

d) Con gà thứ I đẻ trứng trong ngày biết rằng trong ngày đó có 1 con đẻ trứng.

e) Con gà thứ I đẻ trứng trong ngày biết rằng trong ngày đó có ít nhất 1 con đẻ trứng.

f) Con gà thứ I đẻ trứng trong ngày biết rằng trong ngày đó có nhiều nhất 2 con đẻ trứng.

Giải.

$\square$

1.5.3 Công thức xác suất đầy đủ

Định nghĩa 1.5 (Hệ đầy đủ). *n biến cố $A_1, A_2, \ldots, A_n$ được gọi là hệ đầy đủ nếu chúng xung khắc từng đôi một và luôn có ít nhất một biến cố xảy ra trong một phép thử. Nghĩa là*

$$\begin{cases} A_i \cap A_j = \emptyset, & \forall i \neq j \\ A_1 + A_2 + \cdots + A_n = \Omega \end{cases}$$

Ví dụ 1.21. Từ một lọ có 4 bi trắng và 6 bi đen lấy ra 2 bi.

Gọi các biến cố:

A_0: "Lấy được 0 bi đen"
A_1: "Lấy được 1 bi đen"
A_2: "Lấy được 2 bi đen"

Khi đó $A_0; A_1; A_2$ là hệ đầy đủ. □

Công thức xác suất đầy đủ: Cho $A_1; A_2; \ldots; A_n$ ($\mathbb{P}(A_i) > 0$) là hệ đầy đủ các biến cố và B là một biến cố bất kỳ. Xác suất xảy ra biến cố B

$$\mathbb{P}(B) = \mathbb{P}(A_1)\mathbb{P}(B|A_1) + \mathbb{P}(A_2)\mathbb{P}(B|A_2) + \cdots + \mathbb{P}(A_n)\mathbb{P}(B|A_n)$$

Ví dụ 1.22. Một đám đông có số đàn ông bằng nửa số đàn bà. Xác suất để đàn ông bị bệnh tim là 0,06 và đàn bà là 0,036. Chọn ngẫu nhiên 1 người từ đám đông, tính xác suất để người này bị bệnh tim.

Giải. ____

□

1.5.4 Công thức xác suất Bayes

Gải thiết giống công thức xác suất đầy đủ. Xác suất:

$$\mathbb{P}(A_i|B) = \frac{\mathbb{P}(A_i B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(A_i)\mathbb{P}(B|A_i)}{\mathbb{P}(B)}, \quad i = 1, 2, \ldots, n$$

Ví dụ 1.23. Một lớp có số học sinh nam bằng 3 lần số học sinh nữ. Tỷ lệ học sinh nữ giỏi toán là 30% và tỷ lệ học sinh nam giỏi toán là 40%. Chọn ngẫu nhiên một học sinh trong lớp này. Tính xác suất:

a. Học sinh này giỏi toán.

b. Học sinh này là nam biết rằng học sinh này giỏi toán.

Giải.

□

Ví dụ 1.24. Có hai chuồng gà: Chuồng I có 10 gà trống và 8 gà mái; Chuồng II có 12 trống và 10 mái. Có hai con gà chạy từ chuồng I sang chuồng II. Sau đó có hai con gà chạy ra từ chuồng II. Tính xác suất:

a. Hai con gà chạy từ chuồng I sang chuồng II là 2 con trống và hai con gà chạy ra từ chuồng II cũng là hai con trống.

b. Hai con gà chạy ra từ chuồng II là hai con trống.

c. Biết rằng hai con gà chạy ra từ chuồng II là hai con trống, tính xác suất hai con gà chạy từ chuồng I sang chuồng II là 2 con gà trống.

Giải.

□

1.6 Bài tập chương 1

Bài tập 1.1. Một nhóm khảo sát sở thích tiết lộ thông tin là trong năm qua:

- 45% người xem Tivi thích xem phim tình cảm Hàn quốc.
- 25% người xem Tivi thích xem phim hành động Mỹ.
- 10% thích xem cả hai thể loại trên.

Tính tỷ lệ nhóm người thích xem ít nhất một trong hai thể loại trên. **(60%)**

Giải. ____

____

____

____

Bài tập 1.2. Có ba lô hàng mỗi lô có 20 sản phẩm, số sản phẩm loại A có trong lô I, II, III lần lượt là: 12; 14; 16. Bên mua chọn ngẫu nhiên từ mỗi lô hàng 3 sản phẩm, nếu lô nào cả 3 sản phẩm đều loại A thì bên mua nhận mua lô hàng đó. Tính xác suất:

a. Lô thứ i được mua, $i = 1, 2, 3$. **(0,193; 0,3193; 0,4912)**

b. Có i lô được mua, $i = 0, 1, 2, 3$. **(0,2795; 0,4678; 0,2225; 0,0303)**

c. Có nhiều nhất hai lô được mua. **(0,9697)**

d. Có ít nhất một lô được mua.**(0,7205)**

e. Giả sử có ít nhất một lô được mua. Tính xác suất trong đó lô II được mua. **(0,4432)**

f. Giả sử có ít nhất một lô được mua. Tính xác suất trong đó lô I và II được mua.**(0,0855)**

g. Giả sử có một lô được mua. Tính xác suất lô II được mua. **(0,2803)**

Giải. ____

____

____

____

____

____

____

____

____

____

____

____

Bài tập 1.3. Một hộp bóng bàn có 15 bóng mới và 8 bóng cũ. Lần thứ I lấy ra 2 bóng để sử dụng sau đó cho vào lại hộp; lần thứ II lấy ra 3 bóng. Tính xác suất

- **a.** Lần thứ I lấy được i bóng cũ, $i = 0, 1, 2$. **(0,4150; 0,4743; 0,1107)**
- **b.** Lần I lấy 1 bóng cũ và lần II là 3 bóng mới. **(0,0975)**
- **c.** Lần thứ II lấy được 3 bóng mới. **(0,1929)**
- **d.** Biết lần thứ II lấy được 3 bóng mới, tính xác suất lần thứ I lấy được 1 bóng cũ. **(0,5054)**

Giải.

Bài tập 1.4. Có 3 bình đựng bi: bình I có 4 bi trắng và 6 bi đen; bình II có 7 bi trắng và 3 bi đen; bình III có 6 bi trắng và 8 bi đen. Từ bình I và bình II, mỗi bình lấy 1 bi và bỏ sang bình III. Tiếp theo, từ bình III lấy ra tiếp 3 bi. Tính xác suất:

a. Hai bi lấy ra từ bình I và II có i bi trắng, $i = 0, 1, 2$. **(0,25; 0,5167; 0,2333)**

b. Ba bi lấy ra từ bình III có hai bi trắng. **(0,3347)**

c. Giả sử ba bi lấy từ bình III có hai bi trắng, tính xác suất hai bi lấy từ bình I và II là hai bi đen. **(0,2001)**

Giải.

Bài tập 1.5. Một thùng kín đựng 2 loại thuốc: Số lượng lọ thuốc loại A bằng 2/3 thuốc số lượng lọ thuốc loại B. Tỉ lệ lọ thuốc A, B đã hết hạn sử dụng lần lượt là 10% và 8%. Từ thùng lấy ngẫu nhiên một lọ thuốc.

a. Tính xác suất lấy được lọ thuốc A hết hạn sử dụng. **(0,04)**

b. Tính xác suất lọ thuốc lấy ra từ thùng đã hết hạn sử dụng. **(0,0448)**

c. Giả sử lấy được lọ thuốc còn hạn sữ dụng, tính xác suất lọ này là lọ thuốc B. **(0,6231)**

Giải. ____

Bài tập 1.6. * Một người bắn 3 phát đạn vào một mục tiêu một cách độc lập. Xác suất trúng mục tiêu ở mỗi phát lần lượt là 0,55; 0,6; 0,7. Xác suất mục tiêu bị hạ khi bi trúng 1, 2, 3 phát đạn lần lượt là 0,2; 0,4; 0,8. Tính xác suất:

a. Có i phát trúng mục tiêu, $i = 0, 1, 2, 3$. **(0,054; 0,273; 0,442; 0,231)**

b. Có nhiều nhất 2 phát trúng mục tiêu. **(0,769)**

c. Tính xác suất mục tiêu bị hạ. **(0,4162)**

d. Giả sử có 2 phát trúng mục tiêu, tính xác suất phát thứ I trúng mục tiêu. **(0,5724)**

e. Giả sử mục tiêu bị hạ. Tính xác suất phat thứ nhất trúng mục tiêu. **(0,7189)**

*Sinh viên hệ cao đẳng không phải làm các câu c, e, f.

f. Biết rằng có nhiều nhất 2 phát trúng mục tiêu, tính xác suất mục tiêu bị hạ. **(0,3575)**

Giải.

Bài tập 1.7. Nhà máy có hai phân xưởng, sản lượng của phân xưởng I gấp 3 lần sản lượng của phân xưởng II. Tỉ lệ phế phẩm của phân xưởng I, II lần lượt là 7% và 12%. Chọn ngẫu nhiên một sản phẩm của nhà máy, tính:

a. Xác suất chọn được sản phẩm tốt do phân xưởng I sản xuất. **(0,6975)**

b. Xác suất chọn được phế phẩm. **(0,0825)**

c. Giả sử chọn được sản phẩm tốt, tính xác suất sản phẩm này do phân xưởng I sản xuất. **(0,7602)**

Giải.

Bài tập 1.8. Một người buôn bán bất động sản đang cố gắng bán một mảnh đất lớn. Ông ta tin rằng nếu nền kinh tế tiếp tục phát triển, khả năng mảnh đất được mua là 80%; ngược lại nếu nền kinh tế ngừng phát triển, ông ta chỉ có thể bán được mảnh đất đó với xác suất 40%. Theo dự báo của một chuyên gia kinh tế, xác suất nền kinh tế tiếp tục tăng trưởng là 65%. Tính xác suất để bán được mảnh đất. **(0,66)**

Giải.

Bài tập 1.9. † Có hai hộp đựng bi: hộp I có 5 bi trắng và 7 bi đen; hộp II có 6 bi trắng và 4 bi đen. Lấy 1 bi từ hộp I bỏ sang hộp II, rồi từ hộp II lấy ra 1 bi. Tính xác suất

a. Bi lấy từ hộp II là bi trắng. **(7/12)**

b. Giả sử bi lấy từ hộp II là bi trắng, tính xác suất bi lấy từ hộp I là bi trắng. **(5/11)**

c. Giả sử bi lấy ra từ hộp II là bi trắng, tính xác suất bi này của hộp I. $\left(\frac{5}{12}\frac{1}{11}/\frac{7}{12}\right)$

d. Giả sử bi lấy ra từ hộp II là bi trắng, tính xác suất bi này của hộp II. $\left(\frac{6}{11}/\frac{7}{12}\right)$

Giải.

†Sinh viên hệ cao đẳng không phải làm các câu c, d.

Chương 2

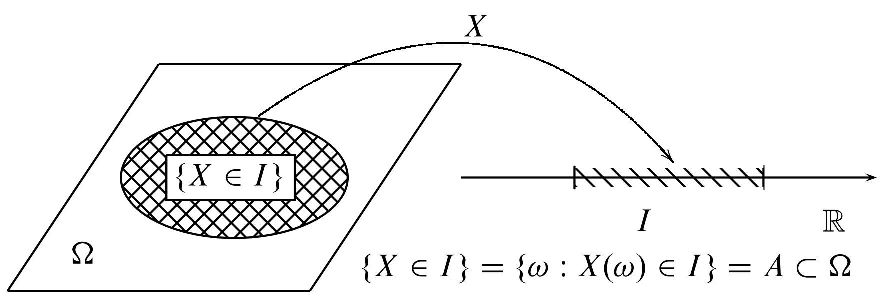
Biến ngẫu nhiên

2.1 Khái niệm biến ngẫu nhiên

- Xét một phép thử có không gian các biến cố sơ cấp Ω. Đặt

$$\begin{array}{rccl} X & : & \Omega & \longrightarrow & \mathbb{R} \\ & & \omega & \longmapsto & X(\omega) = x \end{array}$$

X được gọi là biến ngẫu nhiên, x gọi là giá trị của biến ngẫu nhiên X.



Hình 2.1: Biến ngẫu nhiên X

Ví dụ 2.1. Thực hiện phép thử gieo đồng thời 2 đồng xu cân đối, chúng ta có không gian các biến cố sơ cấp

$$\Omega = \{N_1N_2; N_1S_2; S_1N_2; S_1S_2\}$$

Đặt $X(\omega)$ là số đồng xu sấp khi kết quả phép thử là ω. Ta có:

$$X(N_1N_2) = 0; \quad X(N_1S_2) = 1; \quad X(S_1N_2) = 1; \quad X(S_1S_2) = 2$$

Khi đó ta gọi X là biến ngẫu nhiên số đồng xu sấp khi tung 2 đồng xu. $\square$

- Có hai loại biến ngẫu nhiên:

- Biến ngẫu nhiên rời rạc là biến ngẫu nhiên mà giá trị có thể của nó là một tập hữu hạn hoặc vô hạn đếm được.
- Biến ngẫu nhiên liên tục là biến ngẫu nhiên mà giá trị có thể của nó lấp đầy một khoảng trên trục số.

Ví dụ 2.2.

- Số chấm trên mặt xuất hiện khi tung một xúc sắc là biến ngẫu nhiên rời rạc (giá trị của X là tập hữu hạn).
- Số cuộc gọi đến tổng đài điện thoại trong 1 giờ là biến ngẫu nhiên rời rạc (giá trị của X là tập vô hạn đếm được).
- Thời gian hoàn thành 1 sản phẩm của một công nhân là biến ngẫu nhiên liên tục. □

2.2 Phân phối xác suất của biến ngẫu nhiên

2.2.1 X là biến ngẫu nhiên rời rạc

Để mô tả phân phối xác suất của biến ngẫu nhiên rời rạc người ta sử dụng bảng phân phối xác suất:

X	x_1	x_2	$\cdots$	x_n	$\cdots$
$\mathbb{P}$	$f(x_1)$	$f(x_2)$	$\cdots$	$f(x_n)$	$\cdots$

Trong đó:

- Dòng 1 liệt kê giá trị có thể của X.
- $f(x_i) = \mathbb{P}(X = x_i), i = 1, 2, \ldots$ gọi là xác suất X nhận giá trị x_i.
- Nếu $x_0 \notin \{x_1, \ldots, x_n, \ldots\}$ thì $f(x_0) = 0$.

Ví dụ 2.3. Thực hiện phép thử tung một xúc sắc. Gọi X là số chấm trên mặt xuất hiện của xúc sắc. X có bảng phân phối như sau:

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

Nhận xét:

- $f(x_1) + f(x_2) + \cdots + f(x_n) + \cdots = 1.$
- $\mathbb{P}(a < X < b) = \sum_{a < x_i < b} f(x_i).$ □

Ví dụ 2.4. Cho biến ngẫu nhiên rời rạc X có bảng phân phối xác suất cho như sau:

X	-1	1	3	5
$\mathbb{P}$	a	$2a$	$3a$	$4a$

a. Xác định a.
b. Xác định $\mathbb{P}(X = 2)$.
c. Xác định $\mathbb{P}(-1 < X < 4)$.

Giải. ____

____

____

____

□

Ví dụ 2.5. Một xạ thủ có 4 viên đạn, bắn lần lượt từng viên vào một mục tiêu một cách độc lập. Xác suất trúng mục tiêu ở mỗi lần bắn là 0,7. Nếu có một viên trúng mục tiêu hoặc hết đạn thì dừng. Gọi X là số viên đạn đã bắn, lập bảng phân phối xác suất của X.

Giải. ____

____

____

____

____

____

____

____

□

Ví dụ 2.6. Một xạ thủ có 6 viên đạn, bắn lần lượt từng viên vào một mục tiêu một cách độc lập. Xác suất trúng mục tiêu ở mỗi lần bắn là 0,7. Nếu có 3 viên trúng mục tiêu hoặc hết đạn thì dừng. Gọi X là số viên đạn đã bắn, lập bảng phân phối xác suất của X.

Giải.

□

Ví dụ 2.7. Một lọ có 3 bi trắng và 7 bi đen. Từ lọ này lấy ra ngẫu nhiên 4 bi. Gọi X là số bi đen lẫn trong 4 bi lấy ra, lập bảng phân phối xác suất của X.

Giải.

$\square$

2.2.2 X là biến ngẫu nhiên liên tục

Định nghĩa 2.1 (Hàm mật độ). *Hàm số $f(x) \geq 0, \forall x \in \mathbb{R}$ được gọi là hàm mật độ của biến ngẫu nhiên liên tục X nếu*

$$\mathbb{P}(X \in A) = \int_A f(x)dx, \forall A \subset \mathbb{R}$$

Chú ý. Với định nghĩa hàm mật độ ta có

i. Nếu X là biến ngẫu nhiên liên tục thì xác suất X thuộc một tập $A \subset \mathbb{R}$ được tính bằng tích phân của hàm mật độ $f(x)$ trên tập A.

ii. Mọi hàm mật độ phải thỏa hai điều kiện $f(x) \geq 0$ và $\int\limits_{-\infty}^{+\infty} f(x)\mathrm{d}x = 1$

Ví dụ 2.8. Cho hàm số

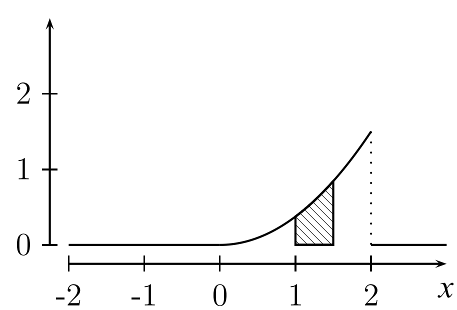
$$f(x) = \begin{cases} \dfrac{3}{8}x^2 & \text{khi } 0 \leq x \leq 2 \\ 0 & \text{nơi khác} \end{cases}$$

a. Chứng tỏ $f(x)$ là hàm mật độ của biến ngẫu nhiên X.

b. Tính xác suất $\mathbb{P}(1 \leq X \leq 3/2)$.

c. Tính xác suất $\mathbb{P}(1 \leq X \leq 3)$.

Giải.



□

2.2.3 Hàm phân phối xác suất

Định nghĩa 2.2 (Hàm phân phối xác suất)**.** *Hàm phân phối xác suất của biến ngẫu nhiên X, ký hiệu $F(x)$*

$$F(x) = \mathbb{P}(X < x)$$

Nhận xét:

- Nếu X là biến ngẫu nhiên rời rạc thì

$$F(x) = \mathbb{P}(X < x) = \sum_{x_i < x} f(x_i)$$

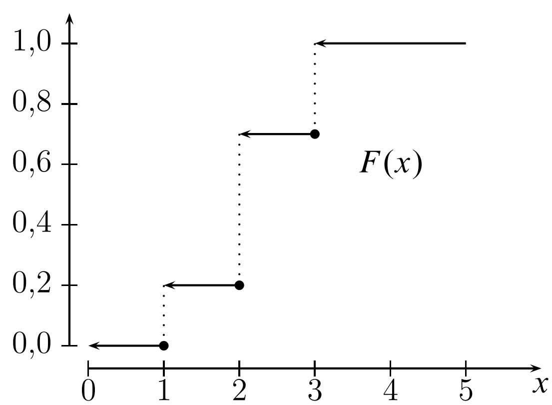
- Nếu X là biến ngẫu nhiên liên tục có hàm mật độ $f(x)$ thì

$$F(x) = \mathbb{P}(X < x) = \int_{-\infty}^{x} f(t)dt$$

Ví dụ 2.9. Cho biến ngẫu nhiên X có bảng phân phối như sau:

X	1	2	3
$\mathbb{P}$	0,2	0,5	0,3

a. Tìm hàm phân phối $F(x)$ của X.
b. Vẽ đồ thị của $F(x)$.



Giải. ____

□

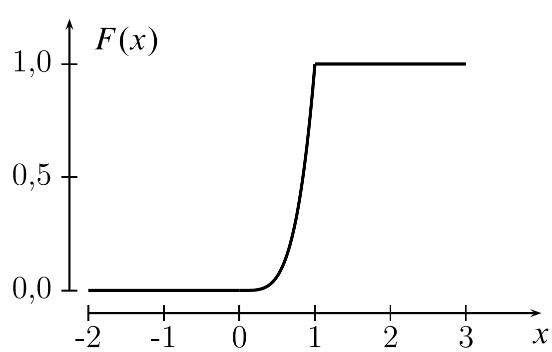
Ví dụ 2.10. Cho biến ngẫu nhiên liên tục X có hàm mật độ

$$f(x) = \begin{cases} kx^3 & \text{khi } 0 \le x \le 1 \\ 0 & \text{nơi khác} \end{cases}$$

a. Xác định k.

b. Tìm hàm phân phối xác suất $F(x)$.

c. Vẽ đồ thị hàm phân phối $F(x)$.



Giải.

$\square$

Tính chất 2.3. *Hàm phân phối xác suất $F(x)$ có các tính chất:*

i. $0 \le F(x) \le 1, \forall x \in \mathbb{R}; F(-\infty) = 0; F(+\infty) = 1$

ii. $F(x)$ *là hàm không giảm (nếu* $x_1 < x_2$ *thì* $F(x_1) \le F(x_2)$*).*

iii. $\mathbb{P}(a \le X < b) = F(b) - F(a)$.

iv. Nếu X *là biến ngẫu nhiên liên tục có hàm mật độ* $f(x)$ *thì:*

- $F'(x) = f(x)$
- $\mathbb{P}(X = x) = 0, \quad \forall x \in \mathbb{R}$ *và*

$$\begin{aligned}\mathbb{P}(b \le X < a) &= \mathbb{P}(a < X < b)\\ &= \mathbb{P}(a < X \le b)\\ &= \mathbb{P}(a \le X \le b) = F(b) - F(a)\end{aligned}$$

Ví dụ 2.11. Một phân xưởng có 2 máy hoạt động độc lập. Xác suất trong 1 ngày làm việc các máy đó hỏng tương ứng là 0,3 và 0,4. Gọi X là số máy hỏng trong 1 ngày làm việc.

a. Lập bảng phân phối xác suất của X.

b. Tìm hàm phân phối xác suất của X.

Giải.

□

2.3 Các đặc trưng số của biến ngẫu nhiên

2.3.1 Kỳ vọng - $\mathbb{E}X$

Định nghĩa 2.4 (Kỳ vọng)**.** *Kỳ vọng của biến ngẫu nhiên X, ký hiệu $\mathbb{E}X$:*

- *X là biến ngẫu nhiên rời rạc có bảng phân phối xác suất*

X	x_1	x_2	$\cdots$	x_n	$\cdots$
$\mathbb{P}$	$f(x_1)$	$f(x_2)$	$\cdots$	$f(x_n)$	$\cdots$

  Kỳ vọng $\mathbb{E}X = x_1 f(x_1) + \cdots + x_n f(x_n) + \cdots$

- *X là biến ngẫu nhiên liên tục có hàm mật độ $f(x)$*

$$\textit{Kỳ vọng } \mathbb{E}X = \int_{-\infty}^{+\infty} xf(x)dx$$

Ví dụ 2.12. Anh A nuôi 5 con lợn có cân nặng (kg) 55, 55, 60, 70, 70. Chọn ngẫu nhiên một con và mang cân, gọi X là cân nặng.

a. Lập bảng phân phối xác suất của X.
b. Tính kỳ vọng của X.
c. Lập bảng phân phối xác suất của X^2.
d. Tính kỳ vọng của X^2.

Giải. ____

□

Ý nghĩa của kỳ vọng: Kỳ vọng của X là *trung bình* các giá trị của X theo xác suất.

Tính chất 2.5. *Kỳ vọng có các tính chất:*

i. $\mathbb{E}c = c$, *c là hằng số.*

ii. $\mathbb{E}(cX) = c\mathbb{E}X$.

iii. $\mathbb{E}(X + Y) = \mathbb{E}X + \mathbb{E}Y$.

iv. $\mathbb{E}(XY) = \mathbb{E}X \cdot \mathbb{E}Y$ *khi* X *và* Y *độc lập.*

v. Cho $Y = h(X)$ *là hàm của biến ngẫu nhiên* X.

- *Khi* X *là biến ngẫu nhiên rời rạc*

$$\mathbb{E}Y = \mathbb{E}h(X) = h(x_1)f(x_1) + \cdots + h(x_n)f(x_n) + \cdots$$

- *Khi* X *là biến ngẫu nhiên liên tục có hàm mật độ* $f(x)$ *thì*

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

Ví dụ 2.13. Thời gian học rành nghề sửa ti vi của một người là một biến ngẫu nhiên - X (năm) có hàm mật độ.

$$f(x) = \begin{cases} \frac{9}{40}x^2 + \frac{1}{5} & \text{khi } x \in (0;2) \\ 0 & \text{khi } x \notin (0;2) \end{cases}$$

a. Tính thời gian trung bình một người học rành nghề sửa tivi.

b. Tính $\mathbb{E}(2X + 3)$.

c. Tính $\mathbb{E}(X^2)$.

Giải. ____

□

2.3.2 Phương sai - $\mathbb{V}arX$

Định nghĩa 2.6 (Phương sai)**.** *Phương sai của biến ngẫu nhiên X, ký hiệu $\mathbb{V}arX$*

$$\mathbb{V}arX = \mathbb{E}\left(\mathbb{E}X - X\right)^2 = \mathbb{E}X^2 - \left(\mathbb{E}X\right)^2$$

Ví dụ 2.14. Anh A nuôi 5 con lợn có cân nặng (kg) 55, 55, 60, 70, 70. Chọn ngẫu nhiên một con và mang cân, gọi X là cân nặng. Tính phương sai của X.

Giải.

□

Ý nghĩa phương sai: Phương sai là trung bình của bình phương sai khác giữa các giá trị của X so với trung bình của nó. Do đó phương sai dùng để đo độ phân tán các giá trị của X so với trung bình của nó. Nghĩa là phương sai lớn thì độ phân tán lớn và ngược lại.

Do đơn vị của phương sai bằng bình phương đơn vị của X. Để có cùng đơn vị, ta định nghĩa độ lệch chuẩn

$$\sigma = \sqrt{\mathbb{V}arX}$$

Ví dụ 2.15. Giả thiết giống ví dụ 2.13. Thời gian học rành nghề sửa ti vi của một người là một biến ngẫu nhiên - X (năm) có hàm mật độ.

$$f(x) = \begin{cases} \dfrac{9}{40}x^2 + \dfrac{1}{5} & \text{khi } x \in (0;2) \\ 0 & \text{khi } x \notin (0;2) \end{cases}$$

Tính phương sai của X.

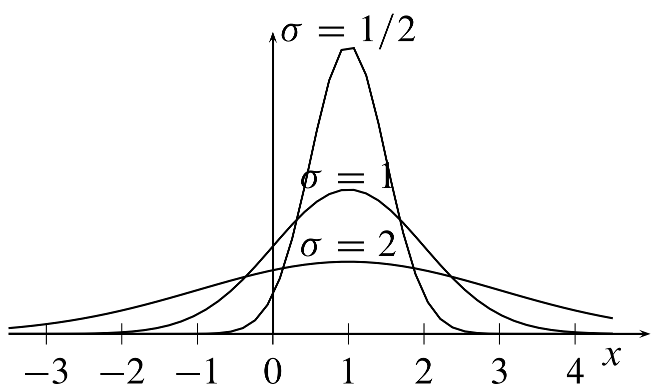
Giải. ____

____

____

____

□



Tính chất 2.7. *Phương sai có các tính chất:*

i. $\mathbb{V}ar(c) = 0, c$ *là hằng số.*

ii. $\mathbb{V}ar(cX) = c^2\mathbb{V}arX.$

iii. $\mathbb{V}ar(X + Y) = \mathbb{V}arX + \mathbb{V}arY$, *nếu* X *và* Y *độc lập.*

2.3.3 ModX

Định nghĩa 2.8. *Mod của biến ngẫu nhiên* S, *ký hiệu* $ModX$

- *X là biến ngẫu nhiên rời rạc*

$$\boldsymbol{ModX} = \{x_i | \mathbb{P}(X = x_i) \max\}$$

- *X là biến ngẫu nhiên liên tục có hàm mật độ $f(x)$*

$$\boldsymbol{ModX} = \{x_0 | f(x_0) \max\}$$

Ví dụ 2.16. Cho biến ngẫu nhiên X có bảng phân phối xác suất cho như sau:

X	1	2	3	4
$\mathbb{P}$	0,1	0,3	0,4	0,2

$\boldsymbol{ModX} = 3$ vì $\mathbb{P}(X = 3) \max$ □

Ví dụ 2.17. Cho biến ngẫu nhiên liên tục X có hàm mật độ

$$f(x) = \begin{cases} x - \dfrac{x^3}{4} & \text{khi } x \in [0;2] \\ 0 & \text{khi } x \notin [0;2] \end{cases}$$

Xác định $\boldsymbol{ModX}$.

Giải.

□

2.4 Bài tập chương 2

Bài tập 2.1. Cho biến ngẫu nhiên rời rạc X có bảng phân phối xác suất

X	a	$0,1$	$0,3$	$0,4$	2
$\mathbb{P}$	$0,3$	$0,2$	$0,2$	$0,2$	$0,1$

a. Giá trị của tham số a để $\mathbb{E}X = 0,3$. **(-0,2)**

b. Tìm hàm phân phối xác suất của X.

Giải.

Bài tập 2.2. Theo thống kê, một người Mỹ 25 tuổi sẽ sống thêm trên 1 năm có xác suất là 0,992 và người đó chết trong vòng 1 năm tới là 0,008. Một công ty bảo hiểm A đề nghị người đó bảo hiểm sinh mạng cho 1 năm với số tiền chi trả là 10000 USD, phí bảo hiểm là 100 USD. Hỏi trung bình công ty A lãi bao nhiêu khi bán bảo hiểm cho người đó? **(20USD)**

Giải.

Bài tập 2.3. Người thợ chép tranh mỗi tuần chép hai bức tranh độc lập A và B với xác suất hỏng tương ứng là 0,03 và 0,05. Biết rằng nếu thành công thì người thợ sẽ kiếm lời từ bức tranh A là 1,3 triệu đồng và B là 0,9 triệu đồng, nhưng nếu hỏng thì bị lỗ do bức tranh A là 0,8 triệu đồng và do B là 0,6 triệu đồng. Hỏi trung bình người thợ kiếm được bao nhiêu tiền chép tranh mỗi tuần? **(2,062)**

Giải.

Bài tập 2.4. Nhu cầu hằng ngày của 1 khu phố về 1 loại thực phẩm tươi sống có bảng phân phối xác suất

Nhu cầu (kg)	31	32	33	34
$\mathbb{P}$	$0,15$	$0,25$	$0,45$	$0,15$

Một cửa hàng trong khu phố nhập về mỗi ngày 34 kg loại thực phẩm này với giá 25.000 đồng/kg và bán ra với giá 40.000 đồng/kg. Nếu bị ế, cuối ngày cửa hàng phải bán hạ giá còn 15.000 đồng/kg mới bán hết hàng. Tính tiền lời trung bình của cửa hàng này về loại thực phẩm trên trong 1 ngày. **(460 ngàn đồng)**

Giải.

Bài tập 2.5. Tuổi thọ (X-tuổi) của người dân ở một địa phương là biến ngẫu nhiên có hàm phân phối cho như sau

$$F(x) = \begin{cases} 0 & \text{khi } x \leq 0 \\ 1 - e^{-\lambda x} & \text{khi } 0 < x \end{cases} \quad \text{với } \lambda = 0,013$$

Tính:

a. Tỷ lệ người dân thọ từ 60 đến 70 tuổi. **(0,1361)**
b. Xác định hàm mật độ của X.
c. Tính tuổi thọ trung bình và $\mathbb{V}arX$. $\left(1/\lambda; 1/\lambda^2\right)$

Giải.

Bài tập 2.6. Tuổi thọ (X-tháng) của một bộ phận của một dây chuyền sản xuất là biến ngẫu nhiên liên tục có hàm mật độ:

$$f(x) = \begin{cases} \dfrac{25}{2}(10+x)^{-2} & \text{khi } x \in (0;40) \\ 0 & \text{khi } x \notin (0;40) \end{cases}$$

a. Xác suất tuổi thọ của bộ phận này nhỏ hơn 6 tháng. **(0,4688)**

b. Tuổi thọ trung bình của dây chuyền này. **(10,118 tháng)**

c. Tìm hàm phân phối xác suất của X.

Giải.

Bài tập 2.7. Tuổi thọ của một loại côn trùng nào đó là một đại lượng ngẫu nhiên liên tục X (đơn vị tháng) có hàm mật độ

$$f(x) = \begin{cases} kx^2(4-x) & \text{khi } 0 \le x \le 4 \\ 0 & \text{nơi khác} \end{cases}$$

- **a.** Tìm hằng số k. **(3/64)**
- **b.** Tìm $F(x)$.
- **c.** Tìm $\mathbb{E}(X)$, $\mathbb{V}ar(X)$ và $Mod(X)$. **(12/5; 16/25; 8/3)**
- **d.** Tính xác suất để côn trùng chết trước một tháng tuổi. **(13/256)**

Giải.

Bài tập 2.8. X là biến ngẫu nhiên có hàm mật độ

$$f(x) = \begin{cases} kx^2 & 0 < x < 1 \\ 0 & \text{nơi khác} \end{cases}$$

a. Tìm k để hàm $f(x)$ là hàm mật độ khi đó tìm kỳ vọng và phương sai của X. **(3; 3/4; 3/80)**

b. Tính $\mathbb{P}(1/2 < X < 3/2), \mathbb{P}(X \leq 1/2)$. **(7/8)**

c. Biết $Y = X^3$, tìm $\mathbb{P}(1/64 < Y < 1/8)$. **(7/64)**

Giải.

Bài tập 2.9. Cho hàm số

$$f(x) = \begin{cases} kx(2-x) & \text{khi } 1 < x < 2 \\ 0 & \text{nơi khác} \end{cases}$$

a. Xác định giá trị của k để $f(x)$ là hàm mật độ của biến ngẫu nhiên X. Với k vừa tìm được tính kỳ vọng và phương sai của biến ngẫu nhiên X. **(3/2; 11/8; 19/320)**

b. Tìm hàm phân phối $F(x)$ của biến ngẫu nhiên X.

c. Tính xác suất $\mathbb{P}(Y > 2X)$ với $Y = X^3$. $\left(\boldsymbol{\frac{6}{5}\sqrt{2} - \frac{3}{5}}\right)$

Giải.

Chương 3

Một số phân phối xác suất thông dụng

3.1 Phân phối Bernoulli

Xét một phép thử, trong phép thử này ta chỉ qua tâm đến 2 biến cố A và $\bar{A}$, với $\mathbb{P}(A) = p$. Phép thử như thế này còn gọi là phép thử Bernoulli. Đặt biến ngẫu nhiên

$$X = \begin{cases} 1 & \text{Nếu } A \text{ xảy ra}; \mathbb{P}(X = 1) = p \\ 0 & \text{Nếu } A \text{ không xảy ra}; \mathbb{P}(X = 0) = 1 - p = q \end{cases}$$

Biến ngẫu nhiên X gọi là có phân phối nhị thức tham số p, ký hiệu $X \sim B(p)$. Ta có bảng phân phối xác suất của $X \sim B(p)$

X	0	1
$\mathbb{P}$	q	p

Tính chất 3.1. *Các đặc trưng của $X \sim B(p)$*

i. $\mathbb{E}X = p$.

ii. $\mathbb{V}arX = pq$.

Ví dụ 3.1. Trả lời ngẫu nhiên một câu hỏi trắc nghiệm có 4 đáp án, trong đó chỉ có một đáp án đúng. Gọi biến ngẫu nhiên:

$$X = \begin{cases} 1 & \text{Nếu trả lời đúng}; \mathbb{P}(X = 1) = 1/4 \\ 0 & \text{Nếu trả lời sai}; \mathbb{P}(X = 0) = 3/4 \end{cases}$$

$X \sim B(p); \mathbb{E}X = 1/4; \mathbb{V}arX = 3/16.$ $\square$

3.2 Phân phối Nhị thức

Xét dãy n phép thử Bernoulli độc lập và cùng phân phối,

$$X_i = \begin{cases} 1 & \text{Lần } i \; A \text{ xảy ra}; \mathbb{P}(X_i = 1) = p \\ 0 & \text{Lần } i \; A \text{ không xảy ra}; \mathbb{P}(X_i = 0) = 1 - p = q \end{cases}, i = \overline{1, n}$$

Đặt $X = X_1 + \cdots + X_n$: gọi là số lần A xảy ra trong n lần thực hiện phép thử. X được gọi là có phân phối Bernoulli tham số n, p; ký hiệu $X \sim B(n; p)$.

Ví dụ 3.2. Một xạ thủ bắn 3 phát đạn vào một mục tiêu một cách độc lập, xác suất trúng mục tiêu ở mỗi lần bắn là 0,7. Gọi các biến ngẫu nhiên:

$$X_i = \begin{cases} 1 & \text{Lần } i \text{ bắn trúng MT}; \mathbb{P}(X_i = 1) = 0,7 \\ 0 & \text{Lần } i \text{ bắn không trúng MT}; \end{cases}, i = 1, 2, 3$$

$X = X_1 + X_2 + X_3$, $X \sim B(3; 0,7)$. X là số phát trúng mục tiêu trong 3 phát, giá trị có thể của X là $0, 1, 2$. Xác suất có 2 phát trúng mục tiêu:

$$\begin{aligned} \mathbb{P}(X = 2) &= + \left\| \begin{array}{ll} 0,7.0,7.0,3 = (0,7)^2.0,3 & \text{Phát 1,2 trúng MT} \\ 0,7.0,3.0,7 = (0,7)^2.0,3 & \text{Phát 1,3 trúng MT} \\ 0,3.0,7.0,7 = (0,7)^2.0,3 & \text{Phát 2,3 trúng MT} \end{array} \right. \\ &= 3.(0,7)^2.0,3 = C_3^2 (0,7)^2 0,3 \end{aligned}$$

□

Công thức tính xác suất của $X \sim B(n; p)$

Xác suất trong n lầ thực hiện phép thử Bernoulli có k lần A xảy ra

$$\mathbb{P}(X = k) = C_n^k p^k q^{n-k}, \quad k = 0, 1, \ldots, n$$

Tính chất 3.2. *Các đặc trưng của $X \sim B(n; p)$*

i. $\mathbb{E}X = np$.

ii. $\mathbb{V}arX = npq$.

iii. $np - q \leq ModX \leq np - q + 1$.

Ví dụ 3.3. Một đề thi có 10 câu hỏi, mỗi câu có 4 đáp án trong đó chỉ có một đáp án đúng. Sinh viên A trả lời một cách ngẫu nhiên tất cả các câu. Gọi X là số câu trả lời đúng trong 10 câu:

a. Xác định phân phối xác suất của X.

b. Tính xác suất sinh viên A trả lời đúng từ 2 đến 3 câu.

c. Tính xác suất sinh viên A trả lời đúng ít nhất một câu.

d. Số câu trung bình sinh viên A trả lời đúng và $\mathbb{V}arX$.

e. Số câu sinh viên A có khả năng trả lời đúng lớn nhất.

f. Đề thi cần có ít nhất bao nhiêu câu để xác suất sinh viên A trả lời đúng ít nhất một câu $\geq 0,99$.

Giải.

□

3.3 Phân phối Siêu bội

Ví dụ 3.4. Từ một lọ có 3 bi trắng và 7 bi đen lấy ra 4 bi. Gọi X là số bi đen lẫn trong 4 bi lấy ra, lập bảng phân phối xác suất của X.

$$10 \text{ bi} \begin{cases} 3 & \text{bi trắng} \\ 7 & \text{bi đen} \end{cases} \xrightarrow[\text{có } k \text{ bi đen}]{\text{Lấy ra 4 bi}} \begin{cases} 4-k & \text{bi trắng} \\ k & \text{bi đen} \end{cases}$$

Mô hình siêu bội: Từ một tập có N phần tử gồm:

- N_A phần tử A.
- $N - N_A$ phần tử khác phần tử A.

Từ tập N lấy ra n phần tử. Gọi X là số phần tử A lẫn trong n phần tử lấy ra, X gọi là có phân phối siêu bội tham số N, N_A, n, ký hiệu $X \sim H(N, N_A, n)$

$$N \begin{cases} N_A & \text{Phần tử A} \\ N - N_A & \text{Phần tử } \bar{A} \end{cases} \xrightarrow[\text{được } kPTA]{\text{Lấy ra } n \text{ PT}} \begin{cases} k & \text{Phần tử } A \\ n-k & \text{Phần tử } \bar{A} \end{cases}$$

Công thức tính xác suất cho $X \sim H(N, N_A, n)$

Xác suất trong n phần tử lấy ra từ tập N có k phần tử A :

$$\mathbb{P}(X = k) = \frac{C_{N_A}^k C_{N-N_A}^{n-k}}{C_N^n}, \text{trong đó} \begin{cases} 0 \le k \le n \\ n - (N - N_A) \le k \le N_A \end{cases}$$

Tính chất 3.3. *Các đặc trưng của* $X \sim H(N, N_A, n)$

i. $\mathbb{E}X = np; \left(p = \dfrac{N_A}{N}\right).$

ii. $\mathbb{V}arX = npq\dfrac{N-n}{N-1}.$

Ví dụ 3.5. Có 20 chi tiết máy, trong đó có 15 chi tiết máy tốt. Từ 20 chi tiết này lấy ra ngẫu nhiên 4 chi tiết máy (lấy một lần), gọi X là số chi tiết tốt lẫn trong 4 chi tiết lấy ra.

a. Xác định phân phối xác suất của X.

b. Tính xác suất lấy được 3 chi tiết tốt.

c. Tính trung bình số chi tiết tốt lấy được và $\mathbb{V}arX$.

Giải.

□

3.4 Phân phối Poisson

Định nghĩa 3.4 (Phân phối Poisson)**.** *Biến ngẫu nhiên X được gọi là có phân phối Poisson tham số λ (ký hiệu $X \sim P(\lambda)$ nếu biến ngẫu nhiên X nhận giá trị $k = 0, 1, \ldots$ với*

$$\mathbb{P}(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

Tính chất 3.5. *Các đặc trưng của $X \sim P(\lambda)$*

i. $\mathbb{E}X = \lambda$.

ii. $\mathbb{V}arX = \lambda$.

iii. $\lambda - 1 \leq ModX \leq \lambda$.

Chú ý: Biến ngẫu nhiên X là số lần xuất hiện A tại những thời điểm ngẫu nhiên trong khoảng $(t_1; t_2)$ thỏa 2 điều sau:

- Số lần xuất hiện biến cố A trong khoảng $(t_1; t_2)$ không ảnh hưởng đến xác suất xuất hiện A trong khoảng thời gian kế tiếp.
- Số lần xuất hiện biến cố A trong 1 khoảng thời gian bất kỳ tỉ lệ với độ dài của khoảng đó.

Khi đó biến ngẫu nhiên X có phân phối Poisson.

Ví dụ 3.6. Tại một siêu thi, trung bình cứ 5 phút có 10 khách đến quầy tính tiền.

a. Tính xác suất để trong 1 phút có 3 khách đến quầy tính tiền.

b. Tính xác suất để trong 1 phút có từ 1 đến 3 khách đến quầy tính tiền.

c. Số khách có khả năng đến quầy tính tiền lớn nhất trong 1 giờ.

Giải.

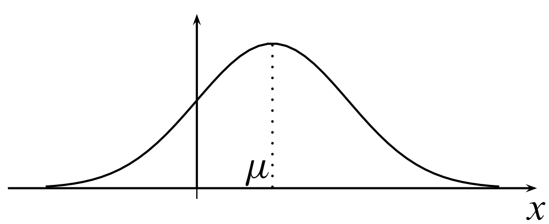
□

3.5 Phân phối Chuẩn

Định nghĩa 3.6 (Phân phối chuẩn)**.** *Biến ngẫu nhiên X được gọi là có phân phối chuẩn tham số μ và σ^2, ký hiệu $X \sim N\left(\mu;\sigma^2\right)$, nếu X có hàm mật độ:*

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

Đồ thị hàm mật độ của $X \sim N(\mu;\sigma^2)$



Nhận xét: Đồ thị hàm mật độ chuẩn có dạng hình "chuông" đối xứng qua $x = \mu$

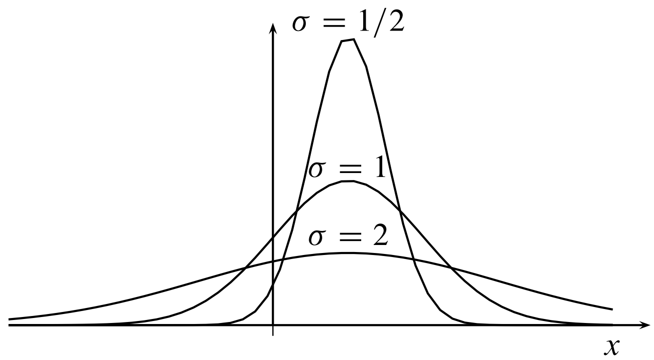
Tính chất 3.7. *Các đặc trưng của $X \sim N\left(\mu;\sigma^2\right)$*

i. $\mathbb{E}X = \mu$.

ii. $\mathbb{V}arX = \sigma^2$.

iii. $ModX = \mu$.

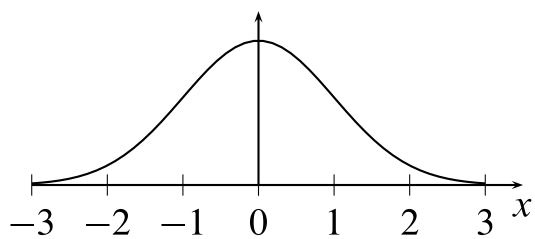
Các đồ thị hàm mật độ biến ngẫu nhiên chuẩn với trung bình là μ và $\sigma = 2, \sigma = 1, \sigma = 1/2$.



Định nghĩa 3.8 (Phân phối chuẩn: $\mu = 0; \sigma^2 = 1$)**.** *Hàm mật độ của biến ngẫu nhiên* $Z \sim N(0; 1)$ *có dạng*

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}, \quad z \in \mathbb{R}$$

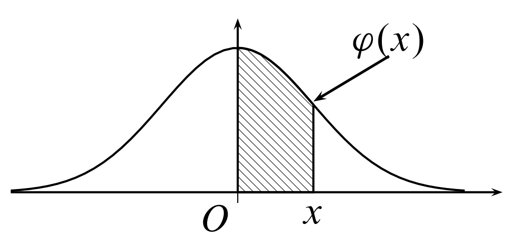
Hình sau là đồ thị hàm mật độ của $z \sim N(0; 1)$



Định nghĩa 3.9 (Hàm Laplace)**.** *Cho biến ngẫu nhiên* $Z \sim N(0; 1)$. *Đặt hàm*

$$\varphi(x) = \int_0^x \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz, \quad x \geq 0$$

gọi là hàm Laplace. (Giá trị của $\varphi(x)$ *được cho trong bảng A.2)*



Tính chất 3.10. *Hàm Laplace* $\varphi(x)$ *có các tính chất:*

i. $\varphi(-x) = -\varphi(x)$.

ii. $\varphi(+\infty) = 0,5; \varphi(-\infty) = -0,5$.

iii. Nếu $Z \sim N(0; 1)$ *thì* $\mathbb{P}(a < Z < b) = \varphi(b) - \varphi(a)$.

iv. Nếu $X \sim N\left(\mu;\sigma^2\right)$ *thì biến ngẫu nhiên* $Z = \dfrac{X-\mu}{\sigma} \sim N(0;1)$. *và*

$$\mathbb{P}(a < X < b) = \varphi\left(\frac{b-\mu}{\sigma}\right) - \varphi\left(\frac{a-\mu}{\sigma}\right)$$

Ví dụ 3.7. Cho biến ngẫu nhiên $X \sim N(0;1)$, tính các xác suất.

a. $\mathbb{P}(-1 < X < 2)$.

b. $\mathbb{P}(1,5 < X)$.

c. $\mathbb{P}(X < -1)$.

□

Ví dụ 3.8. Cho biến ngẫu nhiên $X \sim N\left(3;2^2\right)$. Tính các xác suất:

a. $\mathbb{P}(1 < X)$.

b. $\mathbb{P}(|X-1| < 2)$.

c. $\mathbb{P}(|X-1| < 1)$.

□

Ví dụ 3.9. Điểm Toeic của sinh viên sắp tốt nghiệp ở trường đại học có phân phối chuẩn với giá trị trung bình 560 và độ lệch chuẩn 78. Tính:

a. Tỷ lệ sinh viên có điểm nằm giữa 600 và 700.

b. Tỷ lệ sinh viên có điểm Toeic trên 500.

c. Giả sử nhà trường muốn xác định điểm Toeic tối thiểu để sinh viên có thể ra trường với tỉ lệ 80%. Tính điểm Toeic tối thiểu (lấy phần nguyên).

Giải.

□

Ví dụ 3.10. Tuổi thọ của máy cắt cỏ là biến ngẫu nhiên có phân phối chuẩn với trung bình là 82 tháng. Nhà sản xuất bảo hành sản phẩm khi bán ra là 33 tháng. Giả sử 2,5% sản phẩm bị trả lại (hỏng) trong thời gian bảo hành. Tính:

a. Độ lệch chuẩn của tuổi thọ sản phẩm này.

b. Xác suất một máy loại này có tuổi thọ trên 50 tháng.

c. Một cửa hàng bán 10 máy cắt cỏ loại này. Tính:

- i) Xác suất có 2 máy hỏng trong thời gian bảo hành.
- ii) Số máy trung bình hỏng trong thời gian bảo hành.

Giải.

□

3.6 Bài tập chương 3

Bài tập 3.1. Một nhà vườn trồng 121 cây mai với xác suất nở hoa của mỗi cây trong dịp tết năm nay là 0,75. Giá bán 1 cây mai nở hoa là 0,5 triệu đồng.

a. Tính số cây trung bình nở hoa trong dịp tết. **(90,75 cây)**

b. Giả sử nhà vườn bán hết những cây mai nở hoa, tính số tiền trong dịp tết năm nay nhà vườn thu được chắc chắn nhất. **(45,5 triệu đồng)**

Giải.

Bài tập 3.2. Chủ vườn lan đã để nhầm 20 chậu lan có hoa màu đỏ với 100 chậu lan có hoa màu tím (lan chưa nở hoa). Một khách hàng chọn ngẫu nhiên 15 chậu từ 120 chậu lan đó (chọn 1 lần).

a. Tính xác suất có từ 5 đến 6 chậu lan có hoa màu đỏ. **(0,0723)**

b. Gọi X là số chậu lan có hoa màu đỏ khách chọn được. Tính giá trị của $\mathbb{E}X$ và $\mathbb{V}arX$. **(5/2; 125/68)**

Giải.

Bài tập 3.3. Tại bệnh viện A trung bình 3 giờ có 8 ca mổ. Tính

a. Số ca mổ chắc chắn nhất sẽ xảy ra tại bệnh viện A trong 25 giờ. **(66 ca)**

b. Tính xác suất trong 5 giờ có từ 10 đến 12 ca mổ. **(0,2821)**

Giải.

Bài tập 3.4. Một lô hàng chứa 20 sản phẩm trong đó có 4 phế phẩm. Chọn liên tiếp 3 lần (có hoàn lại) từ lô hàng, mỗi lần chọn ra 4 sản phẩm. Tính xác suất để trong 3 lần chọn có:

a. Đúng 1 lần chọn được không quá 1 phế phẩm. **(0,066)**

b. Trung bình số lần chọn được không quá 1 phế phẩm. **(2,514)**

Giải.

Bài tập 3.5. Giá cà phê trên thị trường là biến ngẫu nhiên có phân phối chuẩn với trung bình là 26000 đồng/kg và độ lệch chuẩn 2000 đồng. k là giá trị tại đó cà phê có giá lớn hơn k với xác suất 90%. Tính giá trị k. **(23420 đồng)**

Giải.

Bài tập 3.6. Thời gian mang thai của sản phụ là biến ngẫu nhiên có phân phối chuẩn với trung bình 280 ngày. Cho biết tỷ lệ một sản phụ mang thai trên 290 ngày là 25,14%, tính độ lệch chuẩn của thời gian mang thai. **(15 ngày)**

Giải.

Bài tập 3.7. Chiều dài của loại linh kiện điện tử A tại cửa hàng B là biến ngẫu nhiên X (mm) có phân phối chuẩn $N(12; 2,5)$. Một công ty cần mua loại linh kiện này với chiều dài từ 11,98mm đến 13mm và họ chọn lần lượt 7 chiếc từ cửa hàng B. Tính xác suất để trong 7 chiếc được chọn có:

a. Từ 5 đến 6 chiếc sử dụng được. **(1,06%)**

b. Ít nhất một chiếc sử dụng được. **(0,8531)**

Giải.

Bài tập 3.8. Thời gian chơi thể thao trong một ngày của một thanh niên là biến ngẫu nhiên X (giờ/ngày) có hàm mật độ

$$f(x) = \begin{cases} A\sin\left(\dfrac{\pi}{3}x\right) & \text{khi } 0 < x < 1 \\ 0 & \text{nơi khác} \end{cases}$$

a. Tính hằng số A. $(2\pi/3)$

b. Tính thời gian chơi thể thao trung bình. **(0,6530 giờ/ngày)**

c. Tính xác suất một thanh niên có thời gian chơi thể thao chưa tới 30 phút/ngày. **(0,2679)**

d. Trung bình có bao nhiêu thanh niên chơi thể thao hơn 30 phút/ngày trong 100 thanh niên. **26,79 thanh niên**

e. Ta phải chọn ít nhất bao nhiêu thanh niên để gặp được ít nhất 1 người có thời gian chơi thể thao chưa tới 30 phút/ngày xảy ra với xác suất hơn 95%. **(10 thanh niên)**

Giải.

Bài tập 3.9. Tuổi thọ của người dân ở một địa phương là một biến ngẫu nhiên - X (tuổi) có hàm mật độ

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{khi } x > 0 \\ 0 & \text{nơi khác} \end{cases}, \quad \lambda = 0,013$$

a. Tính tuổi thọ trung bình của người dân ở địa phương. **(76,9231 tuổi)**

b. Tính tỉ lệ người dân thọ trên 60 tuổi. **(0,4584)**

c. Trung bình có bao nhiêu người thọ trên 60 tuổi tuổi trong 1000 dân. **(458,4)**

Giải.

Bài tập 3.10. Thời gian học rành nghề sửa ti vi của một người là một biến ngẫu nhiên - X (năm) có hàm mật độ.

$$f(x) = \begin{cases} Ax^2 + \dfrac{1}{5} & \text{khi } 0 < x < 2 \\ 0 & \text{nơi khác} \end{cases}$$

a. Xác định hằng số A. **(9/40)**

b. Thời gian học rành nghề trung bình của một người. **(1,3 năm)**

c. Tính xác suất một người học rành nghề dưới 6 tháng. **(0,1094)**

d. Chọn ngẫu nhiên 5 học viên, tính xác suất có 2 người học rành nghề dưới 6 tháng. **(0,0845)**

Giải.

Chương 4

Luật số lớn và các định lý giới hạn

4.1 Hội tụ theo xác suất và phân phối

Định nghĩa 4.1 (Hội tụ theo xác suất). *Cho dãy biến ngẫu nhiên* $\{X_n\}$ *và biến ngẫu nhiên* X. *Ta nói* $\{X_n\}$ *hội tụ theo xác suất đến* X, *ký hiệu* $X_n \xrightarrow{\mathbb{P}} X$, *nếu với mọi* $\varepsilon > 0$ *thì*

$$\lim_{n\to+\infty} \mathbb{P}\left(|X_n - X| < \varepsilon\right) = 1$$

Nếu $X_n \xrightarrow{\mathbb{P}} X$ thì với n lớn chúng ta có $X_n \approx X$ với xác suất gần 1. Thông thường, X_n hội tụ theo xác suất đến biến ngẫu nhiên X là hằng số ($X_n \xrightarrow{\mathbb{P}} \theta$, θ là hằng số) nghĩa là khi n lớn thì hầu như biến ngẫu nhiên X_n không có sự thay đổi.

Định nghĩa 4.2 (Hội tụ theo phân phối). *Định nghĩa hội tụ theo phân phối Cho dãy biến ngẫu nhiên* $\{X_n\}$ *và biến ngẫu nhiên* X. *Ta nói* $\{X_n\}$ *hội tụ theo phân phối đến* X, *ký hiệu* $X_n \xrightarrow{F} X$, *nếu*

$$\lim_{n\to+\infty} \mathbb{P}\left(X_n < x\right) = \mathbb{P}\left(X < x\right) = F(x)$$

tại mọi điểm liên tục của hàm phân phối $F(x)$

Nếu $X_n \xrightarrow{F} X$ thì với n đủ lớn chúng ta có thể xấp xỉ phân phối của X_n bởi phân phối của X. Vậy hội tụ theo phân phối rất tiện lợi cho việc xấp xỉ phân phối của biến ngẫu nhiên X_n.

Định nghĩa 4.3 (Hội tụ hầu chắc chắn). *Cho dãy biến ngẫu nhiên* $\{X_n\}$ *và biến ngẫu nhiên* X. *Ta nói* $\{X_n\}$ *hội tụ hầu chắc chắn đến* X, *ký hiệu* $X_n \xrightarrow{a.s.} X$, *nếu* $X_n \not\to X$ *với xác suất là không.*

4.2 Bất đẳng thức Markov, Chebyshev

4.2.1 Bất đẳng thức Markov

Nếu X là biến ngẫu nhiên nhận giá trị không âm thì với mọi hằng số dương ε ta có

$$\mathbb{P}(X \geq \varepsilon) \leq \frac{\mathbb{E}(X)}{\varepsilon}$$

Chứng minh. X là biến ngẫu nhiên có hàm mật độ $f(x)$ thì

$$\begin{aligned}\mathbb{E}(X) = \int_0^{+\infty} xf(x)dx &= \int_0^{\varepsilon} xf(x)dx + \int_{\varepsilon}^{+\infty} xf(x)dx \\ &\geq \int_{\varepsilon}^{+\infty} xf(x)dx \geq \int_{\varepsilon}^{+\infty} \varepsilon f(x)dx = \varepsilon\mathbb{P}(X \geq \varepsilon)\end{aligned}$$

Nhân hai vế của bất phương trình với $1/\varepsilon$ thì ta được kết quả. □

4.2.2 Bất đẳng thức Chebyshev

Nếu X là biến ngẫu nhiên có kỳ vọng là μ và phương sai σ^2 hữu hạn thì với mọi hằng số dương ε bé tùy ý ta có

$$\mathbb{P}(|X-\mu| \geq \varepsilon) \leq \frac{\mathbb{V}ar(X)}{\varepsilon^2}$$

hay tương đương

$$\mathbb{P}(|X-\mu| < \varepsilon) > \frac{\mathbb{V}ar(X)}{\varepsilon^2}$$

Chứng minh. Ta thấy $\left(X-\mu^2\right)$ là biến ngẫu nhiên không âm và $\varepsilon > 0$. Sử dụng bất đẳng thức *Markov* với $\varepsilon := \varepsilon^2$ ta được

$$\mathbb{P}\left[(X-\mu)^2 \geq \varepsilon^2\right] \leq \frac{\mathbb{E}(X-\mu)^2}{\varepsilon}$$

Vì $(X-\mu)^2 \geq \varepsilon^2$ khi và chỉ khi $|X-\mu| \geq \varepsilon$ nên

$$\mathbb{P}(|X-\mu| \geq \varepsilon) \leq \frac{\mathbb{V}ar(X)}{\varepsilon^2}$$

Bất đẳng thức *Markov* và *Chebyshev* cho ta phương tiện thấy được giới hạn xác suất khi biết kỳ vọng và phương sai của biến ngẫu nhiên chưa biết phân phối xác suất. □

Ví dụ 4.1. Giả sử số phế phẩm của một nhà máy làm ra trong một tuần là một biến ngẫu nhiên với kỳ vọng là $\mu = 50$.

a. Có thể nói gì về xác suất sản phẩm hư tuần này vượt quá 75.

b. Nếu phương sai của phế phẩm trong tuần này là $\sigma^2 = 25$ thì có thể nói gì về xác suất sản phẩm tuần này sẽ ở giữa 40 và 60.

Giải.

a. Theo bất đẳng thức *Markov* $\mathbb{P}(X > 75) \geq \dfrac{\mathbb{E}(X)}{75} = \dfrac{50}{75} = \dfrac{2}{3}$

b. Theo bất đẳng thức *Chebyshev* $\mathbb{P}(|X - 50| \geq 10) \leq \dfrac{\sigma^2}{10^2} = \dfrac{25}{100} = \dfrac{1}{4}$. Do đó

$$\mathbb{P}(40 < X < 60) = \mathbb{P}(|X - 50| < 10) > 1 - \frac{1}{4} = \frac{3}{4}$$

□

4.3 Luật số lớn

Định lý 4.4 (Luật số lớn)**.** *Gọi $X_1, \ldots, X_n$ là các biến ngẫu nhiên độc lập và cùng phân phối xác suất với kỳ vọng $\mu = \mathbb{E}(X)$ và phương sai $\sigma^2 = \mathbb{V}ar(X)$ hữu hạn. Đặt $S_n = X_1 + \cdots + X_n$. Khi đó với mọi $\varepsilon > 0$,*

$$\mathbb{P}\left(\left|\frac{S_n}{n} - \mu\right| \geq \varepsilon\right) \to 0$$

khi $n \to +\infty$.

Chứng minh. Bởi vì $X_1, \ldots, X_n$ là độc lập và cùng phân phối, ta có $\mathbb{V}ar\left(\dfrac{S_n}{n}\right) = \dfrac{\sigma^2}{n}$ và $\mathbb{E}\left(\dfrac{S_n}{n}\right) = \mu$. Áp dụng bất đẳng thức *Chebyshev*, với mọi $\varepsilon > 0$,

$$\mathbb{P}\left(\left|\frac{S_n}{n} - \mu\right| \geq \varepsilon\right) \leq \frac{\sigma^2}{n\varepsilon^2}$$

Cố định ε và khi $n \to +\infty$

$$\mathbb{P}\left(\left|\frac{S_n}{n} - \mu\right| \geq \varepsilon\right) \to 0$$

S_n/n là trung bình của các biến ngẫu nhiên X_i, $(i = 1, \ldots, n)$, do đó người ta thường gọi luật số lớn là luật "trung bình". □

4.4 Định lý giới hạn trung tâm

Định lý 4.5. *Cho $X_1, \ldots, X_n$ là các biến ngẫu nhiên độc lập và cùng phân phối với kỳ vọng μ và phương sai σ^2 hữu hạn. Ta đặt*

$$S_n = X_1 + \cdots + X_n$$

Khi $n \to \infty$ thì biến ngẫu nhiên

$$S_n \xrightarrow{F} X, \qquad \text{với } X \sim N\left(\mathbb{E}\left(S_n\right); \mathbb{V}ar\left(S_n\right)\right)$$

Nhận xét: Định lý trên cho ta kết quả là khi n lớn phân phối của biến ngẫu nhiên S_n được xấp xỉ bằng phân phối chuẩn $N\left(\mathbb{E}\left(S_n\right); \mathbb{V}ar\left(S_n\right)\right)$. Để đơn giản ta viết $S_n \dot{\sim} N\left(\mathbb{E}\left(S_n\right); \mathbb{V}ar\left(S_n\right)\right)$, dấu "$\dot{\sim}$" nghĩa là "xấp xỉ phân phối".

Ví dụ 4.2. Tung 1000 lần 1 xúc sắc, tính xác suất tổng số chấm trong 1000 lần tung lớn hơn 3600.

Giải.

$\square$

4.5 Liên hệ giữa các phân phối xác suất

4.5.1 Liên hệ giữa phân phối nhị thức và chuẩn

Cho $X_1, \ldots, X_n$ là các biến ngẫu nhiên độc lập và $X_i \sim B(p)$. Ta có

$$X = X_1 + \cdots + X_n \sim B(n; p)$$

Theo định lý giới hạn trung tâm $X \dot{\sim} N(np; \sqrt{npq}^2)$ khi $n \to \infty$. Khi đó:

- $\mathbb{P}(a \leq X \leq b) = \varphi\left(\frac{b - np}{\sqrt{npq}}\right) - \varphi\left(\frac{a - np}{\sqrt{npq}}\right)$
- $\mathbb{P}(X = k) = \frac{1}{\sqrt{npq}} f(z)$, ($f(x)$-A.1) trong đó $z = \frac{k - np}{\sqrt{npq}}$

Chú ý: Xấp xỉ trên chúng ta thường sử dụng khi p không quá gần 0 hoặc 1.

Ví dụ 4.3. Trong một kho lúa giống có tỉ lệ hạt lúa lai là 20%. Tính xác suất sao cho khi chọn lần lượt 1000 hạt lúa giống trong kho thì có:

a. Đúng 192 hạt lúa lai.

b. Có từ 185 đến 195 hạt lúa lai.

Giải.

□

4.5.2 Liên hệ giữa siêu bội và nhị thức

Cho $X \sim H(N, N_A, n)$. Nếu cố định n, $N \to \infty$ và $\dfrac{N_A}{N} \to p$ thì $X \dot{\sim} B(n,p)$, $p = \dfrac{N_A}{N}$

Nhận xét: Khi $X \sim H(N, N_A, n)$, nếu N khá lớn và $n << N, (n < 0,05N)$ thì

$$\mathbb{P}(X = k) = \frac{C_{N_A}^k C_{N-N_A}^{n-k}}{C_N^n} \approx C_n^k p^k q^{n-k}, \left(p = \frac{N_A}{N}\right)$$

Ví dụ 4.4. Một ao cá có 10.000 cá da trơn, trong đó có 1.000 con cá tra.

a. Tính xác suất để khi bắt ngẫu nhiên 20 con từ ao thì được 5 con cá tra.

b. Tính xác suất để khi bắt ngẫu nhiên 50 con từ ao thì được 10 con cá tra.

Giải.

□

4.5.3 Liên hệ giữa nhị thức và Poisson

Cho $X \sim B(n, p)$ và khi $n \to \infty$ thì $X \dot\sim P(\lambda)$ trong đó $\lambda = np$

Nhận xét: Khi $X \sim B(n, p)$ và khi n khá lớn thì

$$\mathbb{P}(X = k) = C_n^k p^k q^{n-k} \approx \frac{\lambda^k e^{-\lambda}}{k!}$$

Chú ý: Xấp xỉ trên chúng ta thường dùng khi n lớn và p gần 0 hoặc 1.

Ví dụ 4.5. Một lô hàng thịt đông lạnh đóng gói nhập khẩu có chứa 0,6% bị nhiễm khuẩn. Tìm xác suất để khi chọn ngẫu nhiên 1.000 gói thịt từ lô hàng này có:

a. Không quá 2 gói bị nhiễm khuẩn.

b. Đúng 40 gói bị nhiễm khuẩn.

Giải.

□

Chương 5

Véctơ ngẫu nhiên

5.1 Khái niệm véctơ ngẫu nhiên

- Một bộ có thứ tự n biến ngẫu nhiên $(X_1, \ldots, X_n)$ gọi là một véctơ ngẫu nhiên n chiều.
- Véctơ ngẫu nhiên n chiều là liên tục hay rời rạc nếu, các biến ngẫu nhiên thành phần là liên tục hay rời rạc.

Ví dụ 5.1. Năng xuất lúa ở một thửa ruộng ở địa phương A là biến ngẫu nhiên X, nếu xét đến lượng phân Y thì ta có véctơ ngẫu nhiên hai chiều (X, Y), còn nếu xét thêm lượng nước Z thì ta có véctơ ngẫu nhiên 3 chiều (X, Y, Z). $\square$

Trong giới hạn của chương trình ta chỉ xét véctơ ngẫu nhiên hai chiều, ký hiệu (X, Y).

5.2 Phân phối xác suất của (X, Y)

5.2.1 (X, Y) là véctơ ngẫu nhiên rời rạc

a) Phân phối xác suất đồng thời: Véctơ ngẫu nhiên rời rạc (X, Y) được biểu diễn bằng bảng phân phối xác suất đồng thời:

X \ Y	y_1	y_2	$\cdots$	y_j	$\cdots$	y_n	Tổng dòng
x_1	$f(x_1, y_1)$	$f(x_1, y_2)$	$\cdots$	$f(x_1, y_j)$	$\cdots$	$f(x_1, y_n)$	$f(x_1, \bullet)$
x_2	$f(x_2, y_1)$	$f(x_2, y_2)$	$\cdots$	$f(x_2, y_j)$	$\cdots$	$f(x_2, y_n)$	$f(x_2, \bullet)$
$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\cdots$	$\vdots$	$\vdots$
x_i	$f(x_i, y_1)$	$f(x_i, y_2)$	$\cdots$	$f(x_i, y_j)$	$\cdots$	$f(x_i, y_n)$	$f(x_i, \bullet)$
$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\cdots$	$\vdots$	$\vdots$
x_m	$f(x_m, y_1)$	$f(x_m, y_2)$	$\cdots$	$f(x_m, y_j)$	$\cdots$	$f(x_m, y_n)$	$f(x_m, \bullet)$
Tổng cột	$f(\bullet, y_1)$	$f(\bullet, y_2)$	$\cdots$	$f(\bullet, y_j)$	$\cdots$	$f(\bullet, y_n)$	1

Trong đó:

- $f(x_i, y_j) = \mathbb{P}\left(X = x_i; Y = y_j\right)$
- $\sum\limits_{i=1}^{m} \sum\limits_{j=1}^{n} f(x_i; y_j) = 1$

Ví dụ 5.2. Biến ngẫu nhiên rời rạc X nhận các giá trị 6, 7 và 8. Biến ngẫu nhiên Y nhận các giá trị 1, 2, 3. Phân phối đồng thời của véctơ ngẫu nhiên (X, Y) cho bởi bảng

X \ Y	1	2	3
6	0,1	0,15	0,05
7	0,1	0,2	0,1
8	0,05	0,2	0,05

Tính:

a. $\mathbb{P}(X = 6; Y = 2)\,; \mathbb{P}(X = 4; Y = 6)\,.$

b. $\mathbb{P}(X \geq 7; Y \geq 2)\,.$

Giải. ____

□

b) Phân phối xác suất thành phần (lề)

- Bảng phân phối xác suất của X

X	x_1	x_2	$\cdots$	x_m
$\mathbb{P}(X = x)$	$f(x_1, \bullet)$	$f(x_3, \bullet)$	$\cdots$	$f(x_m, \bullet)$

Trong đó $f(x_i, \bullet)$ là tổng dòng i.

- Bảng phân phối xác suất của Y

Y	y_1	y_2	$\cdots$	y_n
$\mathbb{P}(Y = y)$	$f(\bullet, y_1)$	$f(\bullet, y_2)$	$\cdots$	$f(\bullet, y_n)$

Trong đó $f(\bullet, y_j)$ là tổng cột j.

Ví dụ 5.3. Cho bảng phân phối xác suất đồng thời như sau:

X \ Y	1	2	3
6	0,1	0,15	0,05
7	0,1	0,2	0,1
8	0,05	0,2	0,05

a. Lập bảng phân phối xác suất của X.

b. Tính $\mathbb{P}(X > 6)$.

c. Lập bảng phân phối xác suất của Y.

d. Tính $\mathbb{P}(Y < 3)$.

Giải.

$\square$

c) Phân phối xác suất có điều kiện

- Bảng phân phối xác suất của X với điều kiện $Y = y_j$

X	x_1	x_2	$\cdots$	x_m
$\mathbb{P}(X = x \mid Y = y_j)$	$\dfrac{f(x_1, y_j)}{f(\bullet, y_j)}$	$\dfrac{f(x_2, y_j)}{f(\bullet . y_j)}$	$\cdots$	$\dfrac{f(x_m, y_j)}{f(\bullet, y_j)}$

- Bảng phân phối xác suất của Y với điều kiện $X = x_i$

Y	y_1	y_2	$\cdots$	y_n
$\mathbb{P}(Y = y \mid X = x_i)$	$\dfrac{f(x_i, y_1)}{f(x_i, \bullet)}$	$\dfrac{f(x_i, y_2)}{f(x_i, \bullet)}$	$\cdots$	$\dfrac{f(x_i, y_n)}{f(x_i, \bullet)}$

Ví dụ 5.4. Cho bảng phân phối xác suất đồng thời như sau:

X \ Y	1	2	3
6	0,1	0,15	0,05
7	0,1	0,2	0,1
8	0,05	0,2	0,05

a. Lập bảng phân phối xác suất của X biết $Y = 2$.

b. Tính xác suất $\mathbb{P}(X > 6|Y = 2)$.
c. Lập bảng phân phối xác suất của Y biết $X = 6$.
d. Tính xác suất $\mathbb{P}(Y > 1|X = 6)$.

Giải.

$\square$

5.2.2 (X, Y) là véctơ ngẫu nhiên liên tục

a) Hàm mật độ đồng thời

Định nghĩa 5.1 (Hàm mật độ đồng thời)**.** *Hàm số* $f(x, y) \geq 0, \forall (x, y) \in \mathbb{R}^2$ *được gọi là hàm mật độ đồng thời của* (X, Y) *nếu*

$$\mathbb{P}((X, Y) \in A) = \iint\limits_A f(x, y) dxdy, \quad A \subset \mathbb{R}^2$$

Chú ý. Với định nghĩa hàm mật độ đồng thời của véctơ ngẫu nhiên ta có

i. Nếu (X, Y) là véctơ ngẫu nhiên liên tục thì xác suất (X, Y) thuộc một tập $A \subset \mathbb{R}^2$ được tính bằng tích phân của hàm mật độ $f(x, y)$ trên tập A.

ii. Mọi hàm mật độ đồng thời của véctơ ngẫu nhiên (X,Y) phải thỏa hai điều kiện $f(x, y) \geq 0$ và

$$\mathbb{P}\left((X, Y) \in R^2\right) = \iint\limits_{R^2} f(x, y)dxdy = 1$$

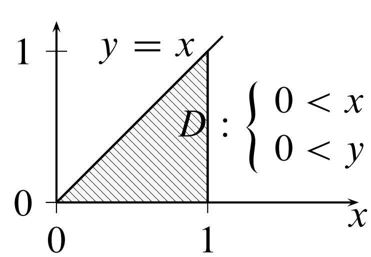
Ví dụ 5.5. Cho hàm số

$$f(x, y) = \begin{cases} 10x^2y & \text{khi } 0 < y < x < 1 \\ 0 & \text{nơi khác} \end{cases}$$

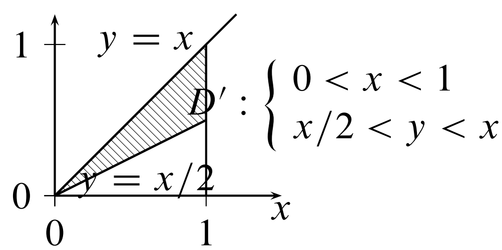
a. Chứng tỏa $f(x, y)$ là hàm mật độ (X, Y).

b. Tính $\mathbb{P}(2Y > X)$.

Giải.



$$D : \begin{cases} 0 < x < 1 \\ 0 < y < x \end{cases} \quad \text{hoặc} \quad \begin{cases} 0 < y < 1 \\ y < x < 1 \end{cases}$$



□

b) Hàm mật độ thành phần (lề)

- Hàm mật độ của X.

$$f_X(x) = \int_{-\infty}^{+\infty} f(x, y)dy$$

- Hàm mật độ của Y.

$$f_Y(y) = \int_{-\infty}^{+\infty} f(x, y)dx$$

Ví dụ 5.6. Cho véctơ ngẫu nhiên (X, Y) có hàm mật độ

$$f(x, y) = \begin{cases} 10x^2y & \text{khi } 0 < y < x < 1 \\ 0 & \text{nơi khác} \end{cases}$$

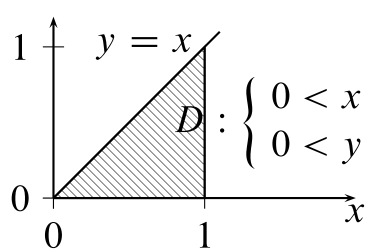
a. Tìm hàm mật độ của X.

b. Tìm hàm mật độ của Y.

c. Tính $\mathbb{P}(X > 1/2)$ và $\mathbb{E}X$.

d. Tính $\mathbb{P}(Y < 1/2)$ và $\mathbb{E}X$.

Giải.



$$D : \begin{cases} 0 < x < 1 \\ 0 < y < x \end{cases} \quad \text{hoặc} \begin{cases} 0 < y < 1 \\ y < x < 1 \end{cases}$$

□

c) Hàm mật độ có điều kiện

- Hàm mật độ của X với điều kiện $Y = y$

$$f_X(x|Y = y) = \frac{f(x, y)}{f_Y(y)}$$

- Hàm mật độ của Y với điều kiện $X = x$

$$f_Y(y|X = x) = \frac{f(x, y)}{f_X(x)}$$

Ví dụ 5.7. Cho véctơ ngẫu nhiên (X, Y) có hàm mật độ

$$f(x, y) = \begin{cases} 10x^2y & \text{khi } 0 < y < x < 1 \\ 0 & \text{nơi khác} \end{cases}$$

a. Tìm hàm mật độ của X với điều kiện $Y = 1/2$.

b. Tìm hàm mật độ của Y với điều kiện $X = 1/3$.

c. Tính $\mathbb{P}(X > 2/3|Y = 1/2)$ và $\mathbb{E}(X|Y = 1/2)$.
d. Tính $\mathbb{P}(Y < 1/4|X = 1/3)$ và $\mathbb{E}(Y|X = 1/3)$.

$\square$

5.3 Bài tập chương 5

Bài tập 5.1. Chi phí quảng cáo (X: triệu đồng) và doanh thu (Y: triệu đồng) của một cửa hàng có bảng phân phối đồng thời cho như sau:

X \ Y	500 (400-600)	700 (600-800)	900 (800-1000)
30	0,10	0,05	0
50	0,15	0,20	0,05
80	0,05	0,05	0,35

a. Lập bảng phân phối xác suất chi phí chi cho quảng cáo.

b. Cho doanh thu là 500 triệu, lập bảng phân phối xác suất chi phí quảng cáo.

c. Lập bảng phân phối xác suất doanh thu của cửa hàng.

d. Cho biết chi phí quảng cáo là 30 triệu, lập bảng phân phối xác suất của doanh thu.

e. Tính chi phí chi cho quảng cáo trung bình.

f. Cho doanh thu là 500 triệu, tính chi phí quảng cáo trung bình.

g. Tính doanh thu trung bình của cửa hàng.

h. Cho chi phí quảng cáo là 30 triệu, tính doanh thu trung bình.

Giải.

Bài tập 5.2. Năng suất lúa X(tấn/ha) và lượng phân Urê Y(x 100 kg) có hàm mật độ đồng thời

$$f(x,y) = \begin{cases} \dfrac{1}{40}y^2 + \dfrac{xy}{20} & \text{khi } 0 \le 3y \le x \le 6 \\ 0 & \text{nơi khác} \end{cases}$$

a. Tìm hàm mật độ xác suất của năng suất lúa.

b. Tìm hàm mật độ xác suất của lượng phân Urê.

c. Tính năng suất lúa trung bình.

d. Tính lượng phân bón trung bình.

e. Tìm hàm mật độ xác suất của năng suất khi lượng phân bón 1 (x 100kg).

f. Tìm hàm mật độ xác suất của lượng phân bón khi năng suất 3 (tấn/ha).

g. Cho biết lượng phân bón 1(x100kg), tính xác suất năng suất lúa dưới 4(tấn/ha).

h. Cho biết lượng phân bón 1(x100 kg), tính năng suất lúa trung bình.

i. Cho biết năng suất lúa 3(tấn/ha), tính lượng phân bón trung bình.

Giải.

Chương 6

Lý thuyết mẫu

6.1 Tổng thể, mẫu

Ta cần nghiên cứu đặc tính X (cân nặng, chiều cao ...) của tập lớn gồm N phần tử (N phần tử này được gọi là tổng thể). Thông thường ta không quan sát hết tất cả các phần tử của tập hợp này bởi vì các lý do:

- Làm hư hại tất cả các phần tử (kiểm tra đồ hộp, bắn thử đạn)
- Thời gian và kinh phí không cho phép – Số phần tử quá lớn (Nghiên cứu một đặc điểm nào của trẻ ta không thể đợi nghiên cứu toàn bộ trẻ em trên thế giới rồi mới đưa ra kết luận).

Do đó người ta lấy từ tổng thể này ra n phần tử (n phần tử này được gọi là mẫu) và quan sát đặc tính X để tính các đặc trưng trên mẫu sau đó sử dụng công cụ toán học để đưa ra kết luận cho tổng thể mà ta không có điều kiện khảo sát tất cả các phần tử.

Muốn mẫu lấy ra đại diện tốt cho tổng thể thì mẫu phải thỏa mãn hai điều kiện chính:

- Mẫu phải chọn ngẫu nhiên từ tổng thể.
- Các phân phối của mẫu phải được chọn độc lập nhau.

Khi quan sát phần tử thứ i, ta gọi X_i là biến ngẫu nhiên giá trị quan sát đặc tính X trên phần tử thứ i. Trong trường hợp cụ thể, giả sử X_i có giá trị x_n thì bộ n giá trị cụ thể $(x_1, \ldots, x_n)$ được gọi là mẫu cụ thể, cỡ mẫu cụ thể là n. Bộ n biến ngẫu nhiên độc lập $(X_1, \ldots, X_n)$ gọi là mẫu ngẫu nhiên.

Ví dụ 6.1. Khảo sát điểm môn xác suất thống kê của sinh viên lớp A có 100 sinh viên, tiến hành lấy mẫu có cỡ mẫu là 5. Gọi $X_i, i = 1, \ldots, 5$ là điểm của sinh viên thứ i trong 5 sinh viên được khảo sát. Nếu $X_1 = 3, X_2 = 7, X_3 = 8, X_4 = 5, X_5 = 7$ thì ta có mẫu cụ thể $(3, 7, 8, 5, 7)$. □

Tính chất 6.1 (Mẫu ngẫu nhiên). *Cho ngẫu nhiên* $(X_1, \ldots, X_n)$, *trong đó* X_i *giá trị quan sát đặc tính* X *trên phần tử thứ* i. *Khi đó:*

i. Các X_i *có cùng phân phối như* X.

ii. Các X_i *độc lập nhau.*

6.2 Mô tả dữ liệu

6.2.1 Phân loại mẫu ngẫu nhiên

Mẫu ngẫu nhiên còn được phân làm 2 loại:

- Mẫu chỉ quan tâm các phần tử của nó có tính chất A hay không gọi là *mẫu định tính*. Giả sử tỷ lệ phần tử A trên tổng thể là p, ta đặt

$$X_i = \begin{cases} 1 & \text{Nếu phần tử thứ } i \text{ loại } A \\ 0 & \text{Nếu phần tử thứ } i \text{ khác loại } A \end{cases}, \quad i = 1, \ldots, n$$

  Khi đó các X_i độc lập và cùng phân phối xác suất với X, $X_i \sim B(p)$.

- Mẫu mà ta quan tâm đến các yếu tố về lượng như là chiều cao, cân nặng, mức hao phí nhiên liệu của một loại động cơ,... gọi là *mẫu định lượng*.

6.2.2 Sắp xếp số liệu

Giả sử mẫu cụ thể $(x_1, \ldots, x_n)$ có k giá trị khác nhau $x_1, \ldots, x_k, (k \leq n)$ và x_i có tần số n_i (với $n_1 + \cdots + n_k = n$). khi đó, số liệu được sắp xếp theo thứ tự tăng dần của x_i như sau:

X	x_1	x_2	$\cdots$	x_k
n_i	n_1	n_2	$\cdots$	n_k

Bảng này gọi là *bảng tần số dạng điểm.*

Ví dụ 6.2. Khảo sát tuổi (X) trẻ bắt đầu đến trường ở một địa phương, lấy mẫu cỡ 10 ta có mẫu cụ thể như sau:

$$4, 5, 6, 7, 6, 6, 5, 5, 6, 6$$

Có bảng tần số dạng điểm:

X	4	5	6	7
n_i	1	3	5	1

□

Giả sử mẫu cụ thể $(x_1, \ldots, x_n)$ có nhiều giá trị khác nhau (quan sát từ biến ngẫu nhiên liên tục) thường người ta phân dữ liệu theo khoảng:

X	$a_0 - a_1$	$a_1 - a_2$	$\cdots$	$a_{k-1} - a_k$
n_i	n_1	n_2	$\cdots$	n_k

Bảng này gọi là *bảng tần số dạng khoảng.* Trong đó n_k là số quan sát có giá trị thuộc khoảng $(a_{k-1}; a_k]$. Khi tính toán ta đưa về *bảng tần số dạng điểm* bằng cách lấy giá trị chính giữa của mỗi khoảng $x_k = \dfrac{x_{k-1} + x_k}{x}$.

Ví dụ 6.3. Khảo sát thời gian (tuần) mang thai của thai phụ không hút thuốc. Tiến hành lấy mẫu, người ta có số liệu cho như bảng sau:

Thời gian	34 − 36	36 − 38	38 − 40	40 − 42	42 − 44
Số thai phụ	7	10	59	41	4

□

Bảng tần số dạng điểm có dạng:

Thời gian	35	37	39	41	43
Số thai phụ	7	10	59	41	4

6.3 Các đặc trưng của mẫu

Giả sử ta cần nghiên cứu đặc tính X. Ký hiệu các tham số $\mu = \mathbb{E}X$ và $\sigma^2 = \mathbb{V}arX$. Trong thống kê các tham số này là các *tham số lý thuyết.*

6.3.1 Trung bình mẫu

Xét mẫu ngẫu nhiên $(X_1, \dots, X_n)$ lấy từ X.

Định nghĩa 6.2 (Trung bình mẫu)**.** *Biến ngẫu nhiên*

$$\bar{X} = \frac{1}{n}(X_1 + \cdots + X_n)$$

được gọi là trung bình mẫu.

Từ các tính chất của mẫu ngẫu nhiên, ta có:

Tính chất 6.3. *Trung bình mẫu có tính chất:*

i. $\mathbb{E}\bar{X} = \frac{1}{n}(\mathbb{E}X_1 + \cdots + \mathbb{E}X_n) = \frac{n\mu}{n} = \mu.$

ii. $\mathbb{V}ar\bar{X} = \frac{1}{n^2}(\mathbb{V}arX_1 + \cdots + \mathbb{V}arX_n) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$

Cho mẫu cụ thể $(x_1, \dots, x_n)$, trung bình mẫu $\bar{x} = \frac{1}{n}(x_1 + \cdots + x_n)$ và trung bình của bình phương $\overline{x^2} = \frac{1}{n}(x_1^2 + \cdots + x_n^2)$

Chú ý. Khi số liệu cho dưới dạng bảng tần số thì $\bar{x} = \frac{1}{n}(x_1 n_1 + \cdots x_k n_k)$ và trung bình của bình phương là $\overline{x^2} = \frac{1}{n}(x_1^2 n_1 + \cdots x_k^2 n_k)$

6.3.2 Phương sai mẫu

Xét mẫu ngẫu nhiên $(X_1, \dots, X_n)$ lấy từ X.

Định nghĩa 6.4 (Phương sai mẫu)**.** *Biến ngẫu nhiên*

$$\hat{S}^2 = \frac{1}{n}\left((X_1 - \bar{X})^2 + \cdots + (X_n - \bar{X})^2\right)$$

được gọi là phương sai mẫu.

Tính chất 6.5. *Phương sai mẫu có các tính chất*

i. $\hat{S}^2 = \mathbb{E}X^2 - (\mathbb{E}X)^2$

ii. $\mathbb{E}\hat{S}^2 = \frac{n-1}{n}\sigma^2.$

Cho mẫu cụ thể $(x_1, \dots, x_n)$, phương sai mẫu $\hat{s}^2 = \overline{x^2} - \bar{x}^2$.

6.3.3 Phương sai mẫu có hiệu chỉnh

Xét mẫu ngẫu nhiên $(X_1, \dots, X_n)$ lấy từ X.

Định nghĩa 6.6 (Phương sai mẫu có hiệu chỉnh)**.** *Biến ngẫu nhiên*

$$S^2 = \frac{1}{n-1}\left((X_1 - \bar{X})^2 + \cdots + (X_n - \bar{X})^2\right)$$

được gọi là phương sai mẫu có hiệu chỉnh.

Tính chất 6.7. *Phương sai mẫu có các tính chất*

i. $S^2 = \dfrac{n}{n-1}\hat{S}^2$
ii. $\mathbb{E}S^2 = \sigma^2$.

Cho mẫu cụ thể $(x_1, \dots, x_n)$, phương sai mẫu có hiệu chỉnh $s^2 = \dfrac{n}{n-1}\hat{s}^2$.

Ta thấy phương sai mẫu và phương sai mẫu có đơn vị đo bằng bình phương đơn vị đo của đặc tính X. Để chuyển về cùng đơn vị ta có khái niệm:

- Độ lệch chuẩn của mẫu, $\hat{s} = \sqrt{\hat{s}^2}$
- Độ lệch chuẩn của mẫu có hiệu chỉnh, $s = \sqrt{s^2}$

Ví dụ 6.4. Khảo sát chiều cao (cm) của nữ sinh trong một trường đại học ta có số liệu như sau

$$153; \quad 160; \quad 145; \quad 162; \quad 165; \quad 158$$

Tính: $\bar{x}, \hat{s}^2, s^2, \hat{s}, s$

Giải. Trung bình mẫu

$$\bar{x} = \frac{1}{6}(153 + 160 + 145 + 162 + 165 + 158) = 157,1666$$

Trung bình của bình phương

$$\overline{x^2} = \frac{1}{6}(153^2 + 160^2 + 145^2 + 162^2 + 165^2 + 158^2) = 24744,5$$

Phương sai mẫu

$$\hat{s}^2 = \overline{x^2} - \bar{x}^2 = 24744,5 - 157,1666^2 = 43,1598$$

Phương sai mẫu có hiệu chỉnh $s^2 = \dfrac{n}{n-1}\hat{s}^2 = \dfrac{6}{5}43,1598 = 51,7907$

Độ lệch chuẩn của mẫu $\hat{s} = \sqrt{\hat{s}^2} = \sqrt{43,1598}$

Độ lệch chuẩn của mẫu có hiệu chỉnh $s = \sqrt{s^2} = \sqrt{51,7907}$

Chú ý. Hướng dẫn sử dụng máy tính cầm tay tính các đặc trưng mẫu

a. Máy FX500MS (tương tự cho máy FX570MS)

- Bước 1: Ấn phím **Mod** đến khi màn hình xuất hiện chữ **SD** và chọn số tương ứng với mục **SD**
- Bước 2: Nhập số liệu
  153; M+; 160; M+; 145; M+; 162; M+; 165; M+; 158; M+
- Bước 3: Sau khi đã nhập hết các số liệu tiếp theo bạn nhấn phím **on**
- Bước 4: Xuất kết quả nhấn **Shift -> 2**
  - Tính $\bar{x}(\bar{x})$: **1; =**
  - Tính $\hat{s}(x\sigma n)$: **2; =**
  - Tính $s(x\sigma n-1)$: **3; =**

b. Máy FX500ES (tương tự cho FX570ES)

- Bước 1: Shift; Mode; ↓; chọn (Stat); chọn (Off) (Số liệu nhập vào không có tần số)
- Bước 2: Mod; chọn (Stat); chọn (1-Var)
- Bước 3: Nhập số liệu
  153; =; 160; =; 145; =; 162; =; 165; =; 158; =
- Sau khi đã nhập hết các số liệu tiếp theo bạn nhấn phím **on**
- Xuất kết quả **Shift; 1; chọn (Var)**
  - Tính $n(n)$: **1; =**
  - Tính $\bar{x}(\bar{x})$: **2; =**
  - Tính $\hat{s}(x\sigma n)$: **3; =**
  - Tính $s(x\sigma n-1)$: **4; =**

Ví dụ 6.5. Điểm môn xác suất thống kê của một số sinh viên khoa A cho như sau

Điểm	5	6	7	8	9	10
Số SV	2	4	12	15	6	2

a. Tính $\bar{x}$.

$$\bar{x} = \frac{1}{41}(5 \cdot 2 + 6 \cdot 4 + 7 \cdot 12 + 8 \cdot 15 + 9 \cdot 6 + 10 \cdot 2) = 7,6097$$

b. Tính $\hat{s}^2$.

$$\overline{x^2} = \frac{1}{41}(5^2 \cdot 2 + 6^2 \cdot 4 + 7^2 \cdot 12 + 8^2 \cdot 15 + 9^2 \cdot 6 + 10^2 \cdot 2) = 59,2195$$

suy ra $\hat{s}^2 = \overline{x^2} - \bar{x}^2 = 59,2195$ -$7,6097^2 = 1,3119$.

□

Chú ý. Hướng dẫn sử dụng máy tính cầm ta tính các đặc trưng mẫu (mẫu có tần số)

a. Máy FX500MS (tương tự cho máy FX570MS)

- Bước 1: Ấn phím **Mod** đến khi màn hình xuất hiện chữ **SD** và chọn số tương ứng với mục **SD**
- Bước 2: Nhập số liệu
  5; Shift;, ; 2; M+;
  6; Shift;, ; 4; M+;
  7; Shift;, ; 12; M+;
  8; Shift;, ; 15; M+;
  9; Shift;, ; 6; M+;
  10; Shift;, ; 2; M+
- Bước 4: Sau khi đã nhập hết các số liệu tiếp theo bạn nhấn phím **on**
- Bước 3: Xuất kết quả nhấn **Shift; 2**
  * Tính $\bar{x}(\bar{x})$: **1; =**
  * Tính $\hat{s}(x\sigma n)$: **2; =**
  * Tính $s(x\sigma n-1)$: **3; =**

b. Máy FX500ES (tương tự cho FX570ES)

– Bước 1: Shift; Mode; ↓; chọn (Stat); chọn (On) (Số liệu nhập vào có tần số)
– Bước 2: Mod; chọn (Stat); chọn (1-Var)
– Bước 3: Nhập số liệu

Cột x: 5 ; =; 6; =; 7; =; 8; =; 9; =; 10; =

Cột Freq: 2; =; 4; =; 12; =; 15; =; 6; =; 2; =

– Sau khi đã nhập hết các số liệu tiếp theo bạn nhấn phím **on**
– Xuất kết quả **Shift; 1; chọn (Var)**
  * Tính $n(n)$: **1; =**
  * Tính $\bar{x}(\bar{x})$: **2; =**
  * Tính $\hat{s}(x\sigma n)$: **3; =**
  * Tính $s(x\sigma n-1)$: **4; =**

Ví dụ 6.6. Năng suất lúa trong 1 vùng là đại lượng ngẫu nhiên có phân phối chuẩn. Gặt ngẫu nhiên 115 ha của vùng này, người ta thu được bảng số liệu:

Năng suất (tạ / ha)	40-42	42 – 44	44 – 46	46 – 48	48 – 50	50 – 52
Diện tích (ha)	7	13	25	35	30	5

Tính $\bar{x}$; $\hat{s}^2$. □

6.4 Phân phối xác suất của trung bình mẫu

a. Trường hợp $X \sim N(\mu;\sigma^3)$

Gọi $(X_1,\ldots,X_n)$ là mẫu ngẫu nhiên lấy từ X, khi đó $X_i \sim N(\mu;\sigma^2)$ và

$$\bar{X} \sim N\left(\mu;\frac{\sigma^2}{n}\right) \tag{6.1}$$

Trong trường hợp chưa biết σ^2 ta có

$$\frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}} \sim T^{n-1} \tag{6.2}$$

b. Trường hợp cỡ mẫu lớn*

$$\bar{X} \dot{\sim} N\left(\mu; \frac{\sigma^2}{n}\right) \tag{6.3}$$

Trong trường hợp chưa biết σ^2 ta có

$$\bar{X} \dot{\sim} N\left(\mu; \frac{S^2}{n}\right) \tag{6.4}$$

Chú ý. Khi mẫu $(X_1, \ldots, X_n)$ là mẫu định tính, tỷ lệ phần tử A trên tổng thể là p.

$$X_i = \begin{cases} 1 & \text{Nếu phần tử thứ } i \text{ loại } A, \mathbb{P}(X_i = 1) = p \\ 0 & \text{Nếu phần tử thứ } i \text{ khác loại } A, \mathbb{P}(X_i = 0) = q \end{cases}, \quad i = 1, \ldots, n$$

Các biến ngẫu nhiên X_i độc lập và $X_i \sim B(p)$, theo 4.5.2 ta có

$$X = X_1 + \ldots + X_n \dot{\sim} N\left(np; \sqrt{npq}^2\right) \text{ hay } \frac{X/n - p}{\sqrt{\frac{npq}{n}}} \dot{\sim} N(0; 1) \tag{6.5}$$

Trong đó X/n gọi là tỷ lệ phần tử A của mẫu, thường được ký hiệu F.

6.5 Đại lượng thống kê

Giả sử có mẫu ngẫu nhiên $(X_1, \ldots, X_n)$ từ biến ngẫu nhiên X.

Định nghĩa 6.8. *Hàm số $\theta(X_1, \ldots, X_n)$ phụ thuộc vào mẫu được gọi là đại lượng thống kê. (Người ta còn gọi ngắn gọn là thống kê).*

Ví dụ 6.7. Trung bình mẫu, phương sai mẫu, tỷ lệ mẫu là các thống kê. □

*Trong thống kê, cỡ mẫu gọi là lớn khi $n \geq 30$.

Chương 7

Ước lượng tham số

7.1 Khái niệm chung

Giả sử biến ngẫu nhiên X có tham số θ chưa biết, dựa vào mẫu ngẫu nhiên $(X_1, \ldots, X_n)$ ta đưa ra thống kê $\hat{\theta} = \theta(X_1, \ldots, X_n)$ để ước lượng giá trị của θ. Có hai phương pháp:

- Ước lượng điểm: Chỉ ra giá trị θ_0 để ước lượng cho θ.
- Ước lượng khoảng: Chỉ ra một khoảng $(\theta_1; \theta_2)$ chứa θ sao cho $\mathbb{P}(\theta_1 < \theta < \theta_2) = 1 - \alpha$ cho trước, trong đó $1 - \alpha$ gọi là độ tin cậy của ước lượng.

7.2 Ước lượng điểm

Định nghĩa 7.1 (Ước lượng không chệch)**.** *Thống kê $\hat{\theta}$ được gọi là ước lượng không chệch cho tham số θ nếu $\mathbb{E}(\hat{\theta}) = \theta$.*

Ví dụ 7.1. Giả sử biến ngẫu nhiên X có giá trị trung bình là μ. Từ X ta lập mẫu ngẫu nhiên $(X_1, \ldots, X_n)$. Khi đó $\bar{X}$ là ước lượng không chệch* cho μ □

Ta nhận thấy thống kê $\hat{\theta} = \dfrac{1}{2}(X_1 + X_n)$ cũng là một ước lượng không chệch cho θ. Vì vậy có thể nói có nhiều ước lượng không chệch cho θ. Vấn đề cần một tiêu chuẩn để chọn một thống kê $\hat{\theta}$ trong lớp các ước lượng không chệch cho θ.

*Theo tính chất 6.3

Định nghĩa 7.2 (Ước lượng hiệu quả). *Ước lượng không chệch $\hat{\theta}$ được gọi là ước lượng có hiệu quả của tham số θ nếu $\mathbb{V}ar\left(\hat{\theta}\right)$ nhỏ nhất trong các ước lượng không chệch của θ.*

Chú ý. Người ta chứng minh được rằng nếu $\hat{\theta}$ là ước lượng hiệu quả của θ thì phương sai của nó là

$$\mathbb{V}ar\left(\hat{\theta}\right) = \frac{1}{n\mathbb{E}\left(\frac{\partial \ln f(x,0)}{\partial \theta}\right)}$$

Trong đó $f(x,\theta)$ là hàm mật độ xác suất của biến ngẫu nhiên gốc.

Các thống kê $\bar{X}, S^2, F$ là ước lượng hiệu quả cho tham số μ, σ^2, p. Ta có quy tắc thực hành ước lượng điểm như sau:

Tham số lý thuyết	Đặc trưng mẫu	Ước lượng
$\mathbb{E}X = \mu$	$\bar{x}$	$\mu \approx \bar{x}$
$\mathbb{V}arX = \sigma^2$	s^2	$\sigma^2 \approx s^2$
p (tỷ lệ phần tử A)	f=tỷ lệ phần tử A trên mẫu	$p \approx f$

7.3 Ước lượng khoảng

7.3.1 Mô tả phương pháp.

Gọi θ là tham số của X chưa biết. Với mẫu cụ thể $(x_1, \ldots, x_n)$ ta tìm khoảng $(\theta_1; \theta_2)$ chứa θ sao cho $\mathbb{P}(\theta_1 < \theta < \theta_2) = 1 - \alpha$ cho trước.

- Khoảng $(\theta_1; \theta_2)$ gọi là khoảng tin cậy.
- $|\theta_1 - \theta_2|$ gọi là độ dài khoảng tin cậy.
- $1 - \alpha$ gọi là độ tin cậy.

7.3.2 Ước lượng khoảng cho trung bình

Gọi μ là trung bình của X chưa biết ta tìm khoảng $(\mu_1; \mu_2)$ chứa μ sao cho $\mathbb{P}(\mu_1 < \mu < \mu_2) = 1 - \alpha$. Khoảng tin cậy $(\mu_1; \mu_2) = (\bar{x} - \varepsilon; \bar{x} + \varepsilon)$, với ε gọi là độ chính xác của ước lượng. Trong đó ε tính như sau[†]

[†]Công thức tính độ chính xác được giải thích ở phụ lục B.1.1

Cỡ mẫu / $\mathbb{V}arX$	$n \geq 30$	$n < 30, X \sim N(\mu;\sigma^2)$
Biết σ^2	$\varepsilon = \dfrac{\sigma}{\sqrt{n}} t_{\frac{1-\alpha}{2}}$ ($t_{\frac{1-\alpha}{2}}$ tra bảng A.2)	$\varepsilon = \dfrac{\sigma}{\sqrt{n}} t_{\frac{1-\alpha}{2}}$ ($t_{\frac{1-\alpha}{2}}$ tra bảng A.2)
Không biết σ^2	$\varepsilon = \dfrac{s}{\sqrt{n}} t_{\frac{1-\alpha}{2}}$ ($t_{\frac{1-\alpha}{2}}$ tra bảng A.2)	$\varepsilon = \dfrac{s}{\sqrt{n}} t_{\alpha}^{n-1}$ (t_{α}^{n-1} tra bảng A.3).

Ví dụ 7.2. Khảo sát về thời gian tự học X (giờ/tuần) trong tuần của một số sinh viên hệ chính quy ở trường đại học A trong thời gian gần đây, người ta thu được bảng số liệu

X	5	6	7	8	9	10
Số SV	10	35	45	36	10	8

Ước lượng thời gian tự học trung bình của một sinh viên với độ tin cậy 95% cho hai trường hợp:

a. Biết $\sigma = 2$

b. Chưa biết σ

Giải. Từ mẫu ta tính được $n = 144; \bar{x} = 7,1736; s = 1,2366$.

Gọi μ là thời gian tự học trung bình của sinh viên. Khoảng ước lượng cho μ với độ tin cậy 95% có dạng

$$(\mu_1;\mu_2) = (\bar{x} - \varepsilon; \bar{x} + \varepsilon)$$

Tiếp theo ta tính ε cho từng trường hợp:

a. Biết $\sigma = 2$

$$\varepsilon = \frac{\sigma}{\sqrt{n}} t_{\frac{1-\alpha}{2}} = \frac{2}{\sqrt{144}} 1,96 = 0,3267$$

Vậy khoảng ước lượng

$$(\mu_1;\mu_2) = (7,1736 - 0,3267; 7,1736 + 0,3267) = (6,8469; 7,5003)$$

Chú ý. Cho trước độ tin cậy là $1-\alpha = 0,95$ cho nên ta có $\frac{1-\alpha}{2} = 0,475$. Tra bảng A.2 ta có $t_{0,475} = 1,96$.

b. Không biết σ

$$\varepsilon = \frac{s}{\sqrt{n}} t_{\frac{1-\alpha}{2}} = \frac{1,2366}{\sqrt{144}} 1,96 = 0,202$$

Vậy khoảng ước lượng $(\mu_1; \mu_2) = (7,1736 - 0,202; 7,1736 + 0,202) = (6,9716; 7,3756)$

Chú ý. Với $t_{0,475} = 1,96$ được tính như câu a. □

Ví dụ 7.3. Khảo sát cân nặng (kg) của gà khi xuất chuồng, người ta cân một số con và kết quả cho như sau:

2,1; 1,8; 2,0; 2,3; 1,7; 1,5; 2,0; 2,2; 1,8

Giả sử cân nặng của gà là biến ngẫu nhiên có phân phối chuẩn. Với độ tin cậy 95% ước lượng cân nặng trung bình của gà khi xuất chuồng:

a. Biết $\sigma = 0,3$.

b. Không biết σ.

Giải. Từ mẫu ta tính được $n = 9; \bar{x} = 1,9333; s = 0,2549$.

Gọi μ là cân nặng trung bình của gà khi xuất chuồng.

a. Cho biết $\sigma = 0,3$

$$\varepsilon = \frac{\sigma}{\sqrt{n}} t_{\frac{1-\alpha}{2}} = \frac{0,3}{\sqrt{9}} 1,96 = 0,196$$

Vậy khoảng ước lượng

$$(\mu_1; \mu_2) = (1,9333 - 0,196; 1,9333 + 0,196) = (1,7373; 2,1293)$$

b. Không biết σ

$$\varepsilon = \frac{s}{\sqrt{n}} t_{\alpha}^{n-1} = \frac{0,2549}{\sqrt{9}} 2,306 = 0,1959$$

Vậy khoảng ước lượng $(\mu_1; \mu_2) = (1,9333 - 0,1959; 1,9333 + 0,1959) = (1,7374; 2,1292)$

Chú ý. Cho trước độ tin cậy là $1-\alpha = 0,95$ cho nên ta có $\alpha = 0,05$. Tra bảng A.3 ta có $t^8_{0,05} = 2,306$. □

Chú ý. Các chỉ tiêu ước lượng trung bình. Ta nhận thấy trong ước lượng trung bình có 3 chỉ tiêu chính $\varepsilon, 1-\alpha, n$. Nếu biết hai chỉ tiêu thì sẽ xác định được chỉ tiêu thứ 3.

a. Xác định cỡ mẫu n nhỏ nhất sao cho độ chính xác không lớn hơn ε và độ tin cậy là $1-\alpha$ (ở đây ta luôn giả sử cỡ mẫu lớn). Ta có

$$n \geq \left(\frac{\sigma}{\varepsilon} t_{\frac{1-\alpha}{2}}\right)^2 \left(\text{ hoặc } n \geq \left(\frac{s}{\varepsilon} t_{\frac{1-\alpha}{2}}\right)^2 \right)$$

n nhỏ nhất thỏa điều kiện trên là

$$n = \left| \left(\frac{\sigma}{\varepsilon} t_{\frac{1-\alpha}{2}}\right)^2 \right| + 1 \left(\text{ hoặc } n = \left| \left(\frac{s}{\varepsilon} t_{\frac{1-\alpha}{2}}\right)^2 \right| + 1 \right)$$

b. Xác định độ tin cậy của ước lượng khi biết độ chính xác của ước lượng. Trước hết xác định giá trị $t_{\frac{1-\alpha}{2}} = \dfrac{\varepsilon\sqrt{n}}{s}$. Và từ đây dễ dàng tính được $1-\alpha$.

Ví dụ 7.4. Cân thử 121 sản phẩm (đơn vị tính bằng kg) ta tính được $s^2 = 5,76$.

a. Xác định độ chính xác nếu muốn ước lượng trọng lượng trung bình với độ tin cậy 95%.

b. Xác định cỡ mẫu nhỏ nhất để lượng trọng lượng trung bình với độ tin cậy 95% và độ chính xác nhỏ hơn 0,4.

c. Xác định độ tin cậy nếu muốn ước lượng trung bình với độ chính xác là $\varepsilon = 0,5$.

Giải.

a. Xác định độ chính xác:

$$\varepsilon = \frac{s}{\sqrt{n}} t_{\frac{1-\alpha}{2}} = \frac{2,4}{\sqrt{121}} 1,96 = 0,4276$$

b. Xác định cỡ mẫu n.

$$n = \left|\left(\frac{s}{\varepsilon} t_{\frac{1-\alpha}{2}}\right)^2\right| + 1 = \left|\left(\frac{2,4}{0,4} 1.96\right)^2\right| + 1 = 139$$

c. Xác định độ tin cậy, trước hết ta tính

$$t_{\frac{1-\alpha}{2}} = \frac{\varepsilon\sqrt{n}}{s} = \frac{0,5\sqrt{121}}{2,4} = 2,29$$

Tra bảng A.2 ta tính được $\frac{1-\alpha}{2} = 0,489$. Từ đó suy ra $1 - \alpha = 0,978$ □

7.3.3 Ước lượng khoảng cho tỷ lệ

Gọi p là tỷ lệ phần tử A chưa biết ta tìm khoảng $(p_1; p_2)$ chứa p sao cho $\mathbb{P}(p_1 < p < p_2) = 1 - \alpha$. Khoảng tin cậy

$$(p_1; p_2) = (f - \varepsilon; f + \varepsilon)$$

trong đó

- f là tỷ lệ phần tử A tính trên mẫu.
- ε gọi là độ chính xác của ước lượng được tính như sau[‡]

$$\varepsilon = \sqrt{\frac{f(1-f)}{n}} t_{\frac{1-\alpha}{2}}$$

Ví dụ 7.5. Khảo sát tỷ lệ phế phẩm do một nhà máy sản xuất ra, người ta quan sát 800 sản phẩm thấy có 8 phế phẩm. Với độ tin cậy 95% hãy ước lượng tỷ lệ phế phẩn của nhà máy.

Giải. Gọi

f là tỷ lệ phế phẩm trên mẫu. $\left(f = \dfrac{8}{800} = 0,01\right)$.

p là tỷ lệ phế phẩm của nhà máy.

[‡]Công thức tính độ chính xác được giải thích ở phụ lục B.1.2

Độ chính xác của ước lượng tỷ lệ

$$\varepsilon = \sqrt{\frac{f(1-f)}{n}} t_{\frac{1-\alpha}{2}} = \sqrt{\frac{0,01(1-0,01)}{800}} 1,96 = 0,0069$$

Vậy khoảng ước lượng cho p với độ tin cậy 95% là

$$(p_1; p_2) = (0,01 - 0,0069; 0,01 + 0,0069) = (0,0031; 0,0169)$$

□

Chú ý. Xác định các chỉ tiêu ước lượng

a Xác định cỡ mẫu n nhỏ nhất sao cho độ chính xác không lớn hơn ε và độ tin cậy là $1-\alpha$ Ta có $n \geq \frac{f(1-f)}{\varepsilon^2}\left(t_{\frac{1-\alpha}{2}}\right)^2$. n nhỏ nhất thỏa điều kiện trên là

$$n = \left| \frac{f(1-f)}{\varepsilon^2}\left(t_{\frac{1-\alpha}{2}}\right)^2 \right| + 1$$

b Xác định độ tin cậy của ước lượng khi biết độ chính xác của ước lượng. Trước hết xác định giá trị

$$t_{\frac{1-\alpha}{2}} = \varepsilon\sqrt{\frac{n}{f(1-f)}}.$$

Và từ đây dễ dàng tính được $1-\alpha$ bằng bảng A.2.

Ví dụ 7.6. Quan sát 800 sản phẩm do một xí nghiệp sản xuất ra thấy có 128 mẫu loại A.

a. Xác định độ chính xác nếu muốn ước lượng tỷ lệ sản phẩm loại A với độ tin cậy 95%.

b. Xác định cỡ mẫu nhỏ nhất để ước lượng tỷ lệ sản phẩm loại A với độ chính xác nhỏ hơn 0,023 và độ tin cậy 95%.

c. Xác định độ tin cậy nếu muốn ước lượng tỷ lệ sản phẩm A với độ chính xác là 0,022.

Giải. Gọi:

f là tỷ lệ sản phẩm loại A tính trên mẫu $\left(f = \frac{128}{800} = 0,16\right)$.

p là tỷ lệ sản phẩm loại A do xí nghiệp sản xuất ra.

a. Độ chính xác của ước lượng

$$\varepsilon = \sqrt{\frac{f(1-f)}{n}} t_{\frac{1-\alpha}{2}} = \sqrt{\frac{0,16(1-0,16)}{800}} 1,96 = 0,0254$$

b. Xác định n

$$n = \left| \frac{f(1-f)}{\varepsilon^2} \left(t_{\frac{1-\alpha}{2}} \right)^2 \right| + 1 = \left| \frac{0,16(1-0,16)}{0,023^2} 1,96^2 \right| + 1 = 977$$

c. Xác định độ tin cậy $1-\alpha$

$$t_{\frac{1-\alpha}{2}} = \varepsilon \sqrt{\frac{n}{f(1-f)}} = 0,022 \sqrt{\frac{800}{0,016(1-0,016)}} = 1,69$$

Tra bảng A.2 ta tính được $\frac{1-\alpha}{2} = 0,4545$. Từ đó suy ra $1-\alpha = 0,909$ □

7.4 Bài tập chương 7

Bài tập 7.1. Kiểm tra ngẫu nhiên 25 bóng đèn của một hãng điện tử, thấy tuổi thọ trung bình là 5000 giờ, độ lệch chuẩn của mẫu có hiệu chỉnh là 200 giờ. Giả sử tuổi thọ của bóng đèn có phân phối chuẩn. Tính khoảng ước lượng tuổi thọ trung bình của loại bóng đèn trên với độ tin cậy 95%. **(4917,44 giờ; 5082,56 giờ)**

Bài tập 7.2. Kiểm tra ngẫu nhiên 25 bóng đèn của một hãng điện tử, thấy độ lệch chuẩn của mẫu có hiệu chỉnh là 200 giờ. Giả sử tuổi thọ của bóng đèn có phân phối chuẩn. Sử dụng mẫu trên để ước lượng tuổi thọ trung bình của loại bóng đèn trên với độ chính xác là 73,12 giờ thì đảm bảo độ tin cậy bao nhiêu? **92%** ______

Bài tập 7.3. Thăm dò 25 người đang sử dụng điện thoại di động về số tiền phải trả trong 1 tháng, thấy số tiền trung bình một người phải trả là 200 ngàn đồng, độ lệch chuẩn của mẫu có hiệu chỉnh là 50 ngàn đồng. Giả sử số tiền phải trả trong một tháng có phân phối chuẩn. Với độ tin cậy là 95% tính khoảng ước lượng số tiền trung bình một người sử dụng điện thoại di động phải trả. **(179,36 ngàn đồng; 220,64 ngàn đồng)** ______

Bài tập 7.4. Thăm dò 25 người đang sử dụng điện thoại di động về số tiền phải trả trong 1 tháng, thấy độ lệch chuẩn của mẫu có hiệu chỉnh là 50 ngàn đồng. Giả sử số tiền phải trả trong một tháng có phân phối chuẩn. Với độ chính xác là 19,74 ngàn đồng thì độ tin cậy bao nhiêu? **94%** ______

Bài tập 7.5. Biết chiều dài của một loại sản phẩm là biến ngẫu nhiên có phân phối chuẩn. Đo ngẫu nhiên 10 sản phẩm loại này thì được chiều dài trung bình là 10,02m và độ lệch chuẩn của mẫu chưa hiệu chỉnh là 0,04m. Tính khoảng ước lượng chiều dài trung bình của loại sản phẩm này với độ tin cậy 95%. **(9,9898m; 10,0502m)**

Chương 8

Kiểm định giả thiết

8.1 Bài toán kiểm định giả thiết

8.1.1 Giả thiết không, đối thiết

Trong chương này chúng ta sẽ đề cập đến bài toán thống kê liên quan đến tham số θ, với giá trị của nó không biết thuộc không gian tham Θ. Tuy nhiên chúng ta sẽ giả sử Θ có thể được phân chia thành hai tập tách biệt Θ_0 và Θ_1 và nhiệm vụ của người làm thống kê phải quyết định xem θ thuộc Θ_0 hay Θ_1.

Chúng ta đặt H_0 để ký hiệu giả thiết $\theta \in \Theta_0$, và H_1 ký hiệu giả thiết $\theta \in \Theta_1$. Bởi vì Θ_0 và Θ_1 tách biệt và $\Theta_0 \cap \Theta_1 = \Theta$, chính xác chỉ có giả thiết H_0 hoặc H_1 là đúng. Chúng ta phải quyết định chấp nhận H_0 để bác bỏ H_1 hoặc ngược lại. Bài toán thuộc dạng này được gọi là kiểm định giả thiết.

Đến đây, chúng ta thấy vai trò của giả thiết H_0 và H_1 cơ bản giống nhau. Trong hầu hết các bài toán kiểm định, hai giả thiết này hơi khác. Để phân biệt giữa hai giả thiết này ta gọi H_0 gọi là *giả thiết không* và H_1 gọi là *đối thiết*. Chúng ta sẽ dùng các thuật ngữ này trong phần còn lại của chương.

8.1.2 Miền tới hạn

Ta xét bài toán với giả thiết có dạng như sau:

$$\begin{cases} \text{Giả thiết không } H_0: & \theta \in \Theta_0 \\ \text{Đối thiết } H_1: & \theta \in \Theta_1 \end{cases}$$

Giả sử trước khi chúng ta quyết định giả thiết nào sẽ được chấp nhận, chúng ta có mẫu ngẫu nhiên $X_1, \ldots, X_n$ được trích từ phân phối của đặc tính X với tham số θ chưa biết. Chúng ta ký hiệu Ω là không gian mẫu, Ω chứa tất cả các kết quả có thể xảy ra khi lấy mẫu ngẫu nhiên.

Trong quá trình kiểm định, chúng ta sẽ chia Ω thành hai tập con. Một tập chứa tất cả các giá trị của X sao cho ta chấp nhận H_0, và tập còn lại chứa tất cả các giá trị của X sao cho ta bác bỏ H_0 và chấp nhận H_1. Tập các giá trị của X để H_0 bị bác bỏ gọi là *miền tới hạn*, ký hiệu C .

Với mỗi giá trị $\theta \in \Theta$ ta đặt hàm *lực lượng* $\pi(\theta)$ là xác suất dẫn đến bác bỏ H_0, ngược lại $1 - \pi(\theta)$ là xác suất dẫn đến chấp nhận H_0. Nếu ký hiệu C là miền tới hạn của kiểm định, hàm $\pi(\theta)$ được xác định bởi quan hệ

$$\pi(\theta) = \mathbb{P}\left(X \in C|\theta\right), \quad \forall \theta \in \Theta$$

Bởi vì $\pi(\theta)$ là xác suất ứng với mỗi θ thì H_0 bị bác bỏ, trong trường hợp lý tưởng hàm $\pi(\theta) = 0$ với mọi $\theta \in \Theta_0$ và $\pi(\theta) = 1$ với mọi $\theta \in \Theta_1$. Nếu hàm $\pi(\theta)$ có các giá trị này thì bất chấp giá trị thực tế θ nào ta luôn có kết luận đúng với xác suất 1.

8.1.3 Hai loại sai lầm

Khi chọn một trong hai quyết định trên sẽ nẩy sinh ra hai sai lầm:

- Sai lầm loại I: Bác bỏ H_0 khi H_0 đúng, xác suất sai lầm loại I là

$$\mathbb{P}\left(C|H_0\right) = \mathbb{P}\left((X_1, \ldots, X_n) \in C|H_0\right)$$

- Sai lầm loại II: Chấp nhận H_0 khi H_0 sai, xác suất sai lầm loại II là

$$\mathbb{P}\left(\bar{C}|H_1\right) = \mathbb{P}\left((X_1, \ldots, X_n) \notin C|H_1\right)$$

Ví dụ 8.1. Cần nghiên cứu tác dụng phụ của một loại thuốc mới vừa được nghiên cứu ta đặt giả thiết và đối thiết như sau

$$\begin{cases} \text{Giả thiết } H_0: & \text{Thuốc có tác dụng phụ} \\ \text{Đối thiết } H_1: & \text{Thuốc không có tác dụng phụ} \end{cases}$$

Thực tế / Kết luận	Thuốc có tác dụng phụ	Thuốc không có tác dụng phụ
Chấp nhận H_0	Kết luận đúng	Sai lầm loại II
Bác bỏ H_0	Sai lầm loại I	Kết luận đúng

Việc đặt giả thiết như trên khi sai lầm loại I xảy ra là tai hại hơn sai lầm loại II (thuốc có tác dụng phụ mà kết luận thuốc không có tác dụng phụ). □

Lẽ tự nhiên là ta chọn miền C sao cho cực tiểu cả hai xác suất phạm sai lầm. Song không thể cực tiểu đồng thời cả hai sai lầm khi cỡ mẫu cố định, bởi vì hai xác suất trên hiên hệ nhau bởi:

$$\mathbb{P}(C|H_0) + \mathbb{P}\left(\bar{C}|H_0\right) = 1; \mathbb{P}(C|H_1) + \mathbb{P}\left(\bar{C}|H_1\right) = 1.$$

Do đó C cực tiểu $\mathbb{P}(C|H_0)$ chưa chắc đã cực tiểu $\mathbb{P}\left(\bar{C}|H_1\right)$

8.1.4 Phương pháp chọn miền tới hạn

Ta cố định một loại xác suất sai lầm và tìm miền C sao cho xác suất phạm sai lầm kia đạt giá trị nhỏ nhất. Thông thường ta cố định xác suất sai lầm loại I: $\mathbb{P}(C|H_0) \leq \alpha$, ta sẽ chọn miền C sao cho $\mathbb{P}\left(\bar{C}|H_1\right)$ đạt cực tiểu hay $\mathbb{P}(C|H_1)$ cực đại, nghĩa là tim C sao cho:

$$\begin{cases} \mathbb{P}(C|H_0) \leq \alpha \\ \mathbb{P}(C|H_1) \text{ đạt cực đại} \end{cases} \text{hay} \begin{cases} \pi(\theta) \leq \alpha \text{ với } \theta \in \Theta_0 \\ \pi(\theta) \text{ đạt cực đại với } \theta \in \Theta_1 \end{cases} \tag{8.1}$$

Ta gọi α là *mức ý nghĩa* của kiểm định, khi cố định α và có hàm lực lượng $\pi(\theta), \forall\theta \in \Theta_1$ lớn nhất thì qui tắc này gọi là qui tắc mạnh nhất.

8.2 Kiểm định giả thiết về trung bình

Giả sử μ (chưa biết) là trung bình của biến ngẫu nhiên X, cần kiểm định*

$$\begin{cases} \text{Giả thiết } H_0: & \mu = \mu_0 \\ \text{Đối thiết } H_1: & \mu \neq \mu_0 \end{cases}$$

*Xem giải thích phụ lục B.2.1

Cỡ mẫu / $\mathbb{V}arX$	$n \geq 30$	$n < 30, X \sim N(\mu;\sigma^2)$
Biết σ^2	$t = \dfrac{\lvert\bar{x}-\mu_0\rvert}{\sigma}\sqrt{n}$ $t_{\frac{1-\alpha}{2}}$ (Bảng A.2)	$t = \dfrac{\lvert\bar{x}-\mu_0\rvert}{\sigma}\sqrt{n}$ $t_{\frac{1-\alpha}{2}}$ (Bảng A.2)
Không biết σ^2	$t = \dfrac{\lvert\bar{x}-\mu_0\rvert}{s}\sqrt{n}$ ($t_{\frac{1-\alpha}{2}}$ (Bảng A.2)	$t = \dfrac{\lvert\bar{x}-\mu_0\rvert}{s}\sqrt{n}$ t_{α}^{n-1} (Bảng A.3)

Kết luận

- Chấp nhận giả thiết H_0 khi $t \leq t_{\frac{1-\alpha}{2}}$ hoặc $\left(t \leq t_{\alpha}^{n-1}\right)$
- Bác bỏ giả thiết H_0 khi $t > t_{\frac{1-\alpha}{2}}$ hoặc $\left(t > t_{\alpha}^{n-1}\right)$

Ví dụ 8.2. Cân thử 15 con gà tây ở 1 trại chăn nuôi khi xuất chuồng ta tính được $\bar{x} = 3,62kg$. Cho biết $\sigma^2 = 0,01$.

a. Giám đốc trại tuyên bố trọng lương trung bình của gà tây là $3,5kg$ thì có tin được không với mức ý nghĩa $\alpha =$1%.

b. Giả sử người ta dùng thức ăn mới và khi xuất chuồng trọng lượng trung bình của gà tây là 3,9 kg. Cho kết luận về loại thức ăn này với mức ý nghĩa $\alpha = 1\%$.

Giải.

a. Gọi μ cân nặng trung bình của gà khi xuất chuồng. Cần kiểm định:

$$\begin{cases} \text{Giả thiết } H_0: & \mu = 3,5kg \\ \text{Đối thiết } H_1: & \mu \neq 3,5kg \end{cases}$$

$$t = \frac{|\bar{x}-\mu_0|}{\sigma}\sqrt{n} = \frac{|3,62-3,5|}{0,1}\sqrt{15} = 4,6 \text{ và } t_{\frac{1-\alpha}{2}} = 2,58$$

$\left(t > t_{\frac{1-\alpha}{2}}\right)$ nên bác bỏ giả thiết. Vậy giám đốc báo cáo sai.

b. Gọi μ cân nặng trung bình của gà tây khi xuất chuồng (trước khi sử dụng thức ăn mới)

$$\begin{cases} \text{Giả thiết } H_0: & \mu = 3,9kg \\ \text{Đối thiết } H_1: & \mu \neq 3,9kg \end{cases}$$

$$t = \frac{|\overline{x}_n - \mu_0|}{\sigma}\sqrt{n} = \frac{|3,62 - 3,9|}{0,1}\sqrt{15} = 10,84$$

$\left(t > t_{\frac{1-\alpha}{2}}\right)$ nên bác bỏ giả thiết. Vậy thức ăn mới có tác dụng tốt. □

8.3 Kiểm định giả thiết về tỷ lệ

Giả sử p(chưa biết) là tỷ lệ phần tử loại A, cần kiểm định†

$$\begin{cases} \text{Giả thiết } H_0: & p = p_0 \\ \text{Đối thiết } H_1: & p \neq p_0 \end{cases}$$

Qui tắc thực hành như sau: Tính giá trị

$$t = \frac{|f - p_0|}{\sqrt{p_0(1-p_0)}}\sqrt{n} \text{ và } t_{\frac{1-\alpha}{2}} (\text{Bảng A.2})$$

Trong đó f là tỷ lệ phần tử A trên mẫu

Kết luận:

- Chấp nhận giả thiết H_0 khi $t \leq t_{\frac{1-\alpha}{2}}$.
- Bác bỏ giả thiết H_0 khi $t > t_{\frac{1-\alpha}{2}}$.

Ví dụ 8.3. Để kiểm tra một loại súng thể thao, người ta cho bắn 1000 viên đạn vào bia thấy có 540 viên trúng mục tiêu. Sau đó, bằng cải tiến kỹ thuật người ta tính được tỷ lệ trúng mục tiêu là 70%. Hãy cho kết luận về cải tiến với mức ý nghĩa 1%.

Giải. Gọi

- p là tỷ lệ bắn trúng trước cải tiến.
- f là tỷ lệ bắn trúng trên mẫu (trước cải tiến).

†Xem giải thích ở phụ lục B.2.2

Cần kiểm định giả thiết

$$\begin{cases} \text{Giả thiết } H_0: & p = 0,7 \\ \text{Đối thiết } H_1: & p \neq 0,7 \end{cases}$$

Tiến hành kiểm tra giả thiết

$$t = \frac{|f - p_0|}{\sqrt{p_0(1-p_0)}}\sqrt{n} = \frac{|0,54 - 0,7|}{\sqrt{0,7.0,3}}\sqrt{1000} = 11,04$$

$1-\alpha = 0,99$ tra bảng A.2 ta được $t_{\frac{1-\alpha}{2}} = 2,58$. Kết luận cải tiến có tác dụng tốt. □

Ví dụ 8.4. Kiểm tra 800 sinh viên thấy có 128 sinh viên giỏi. Trường báo cáo tổng kết là có 40% sinh viên giỏi thì có thể chấp nhận được không với mức ý nghĩa 5%.

Giải. Gọi

- p tỷ lệ sinh viên giỏi thực tế (chưa biết)
- f tỷ lệ sinh viên giỏi tính trên mẫu $f = \dfrac{128}{800} = 0,16$

$$\begin{cases} \text{Giả thiết } H_0: & p = 40\% \\ \text{Đối thiết } H_1: & p \neq 40\% \end{cases}$$

Tiến hành kiểm tra giả thiết

$$t = \frac{|f - p_0|}{\sqrt{p_0(1-p_0)}}\sqrt{n} = \frac{|0,16 - 0,4|}{\sqrt{0,4.0,6}}\sqrt{800} = 13,871$$

$1-\alpha = 0,95$ tra bảng A.2 ta được $t_{\frac{1-\alpha}{2}} = 1,96$. Kết luận báo cáo là sai sự thật, tỷ lệ sinh viên giỏi trong thực tế thấp hơn nhiều. □

8.4 So sánh hai giá trị trung bình

Giả sử X_1 và X_2 là hai biến ngẫu nhiên độc lập có giá trị trung bình là μ_1 và μ_2 . Cần kiểm định

$$\begin{cases} \text{Giả thiết } H_0: & \mu_1 = \mu_2 \\ \text{Đối thiết } H_1: & \mu_1 \neq \mu_2 \end{cases}$$

Ký hiệu các đặc trưng của mẫu 1, 2 lấy từ tổng thể 1, tổng thể 2.

Mẫu	Cỡ mẫu	Trung bình mẫu	Độ lệch chuẩn có hiệu chỉnh
I	n_1	$\bar{x}_1$	s_1
II	n_2	$\bar{x}_2$	s_2

Cỡ mẫu / $\mathbb{V}arX_1; \mathbb{V}arX_2$	$n_1; n_2 \geq 30$	$n_1 < 30; X_1 \sim N(\mu_1; \sigma_1^2)$ $n_2 < 30; X_2 \sim N(\mu_2; \sigma_2^2)$
Biết $\sigma_1^2; \sigma_2^2$	$t = \dfrac{\lvert\bar{x}_1 - \bar{x}_2\rvert}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$ $t_{\frac{1-\alpha}{2}}$ (Bảng A.2)	$t = \dfrac{\lvert\bar{x}_1 - \bar{x}_2\rvert}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$ $t_{\frac{1-\alpha}{2}}$ (Bảng A.2)
Không biết $\sigma_1^2; \sigma_2^2$	$t = \dfrac{\lvert\bar{x}_1 - \bar{x}_2\rvert}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$ $t_{\frac{1-\alpha}{2}}$ (Bảng A.2)	$t = \dfrac{\lvert\bar{x}_1 - \bar{x}_2\rvert}{\sqrt{\dfrac{s^2}{n_1} + \dfrac{s^2}{n_2}}}$ $t_{\alpha}^{n_1+n_2-2}$ (Bảng A.3)

Trong đó $s^2 = \dfrac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}$ gọi là phương sai gộp.

Kết luận:

- Chấp nhận giả thiết H_0 khi $t \leq t_{\frac{1-\alpha}{2}}$ hoặc $\left(t \leq t_{\alpha}^{n_1+n_2-2}\right)$
- Bác bỏ giả thiết H_0 khi $t > t_{\frac{1-\alpha}{2}}$ hoặc $\left(t > t_{\alpha}^{n_1+n_2-2}\right)$

Ví dụ 8.5. Cân thử 100 trái cây ở nông trường I ta tính được $\bar{x}_1 = 101,2$; $s_1^2 = 571,7$ và 361 trái cây ở nông trường II tính được $\bar{x}_2 = 66,39$; $s_2^2 = 29,72$. So sánh trọng lượng trung bình của trái cây ở hai nông trường với mức ý nghĩa 1%.

Giải. Gọi μ_1, μ_2 cân nặng trung bình của trái cây ở nông trường I và II.

Cần kiểm định

$$\begin{cases} \text{Giả thiết } H_0: & \mu_1 = \mu_2 \\ \text{Đối thiết } H_1: & \mu_1 \neq \mu_2 \end{cases}$$

Mẫu	Cỡ mẫu	Trung bình mẫu	Độ lệch chuẩn có hiệu chỉnh
I	$n_1 = 100$	$\bar{x}_1 = 101,2$	$s_1^2 = 571,7$
II	$n_2 = 361$	$\bar{x}_2 = 66,39$	$s_2^2 = 29,72$

Tính giá trị

$$t = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}} = \frac{|101,2 - 66,39|}{\sqrt{\dfrac{571,7}{100} + \dfrac{29,72}{361}}} = 14,4549$$

Tra bảng A.2 ta được $t_{\frac{1-\alpha}{2}} = t_0,495 = 2,58$. Vậy $t > t_{\frac{1-\alpha}{2}}$ cho nên bác bỏ giả thiết H_0 hay cân nặng trung bình của trái cây ở hai địa phương không bằng nhau. □

Ví dụ 8.6. Đo đường kính 20 trục máy do máy I sản xuất và 22 trục máy do máy II sản xuất ta tính được $\bar{x}_1 = 251,7$; $s_1^2 = 52,853$ và $\bar{x}_2 = 249,8$; $s_2^2 = 56,2$. Có thể xem đường kính trung bình của các trục máy ở 2 máy như nhau với mức ý nghĩa 1% không?

Giải.

8.5 So sánh hai tỷ lệ

Gọi $p_1; p_2$ tỷ lệ phần tử A trên tổng thể 1 và 2 chưa biết. Ta cần kiểm định

$$\begin{cases} \text{Giả thiết } H_0: & p_1 = p_2 \\ \text{Đối thiết } H_1: & p_1 \neq p_2 \end{cases}$$

Tính: $f = \dfrac{n_1 f_1 + n_2 f_2}{n_1 + n_2}$ (Tỷ lệ phần tử A chung của 2 mẫu), trong đó $f_1; f_2$ tỷ lệ phần tử A trên mẫu 1, 2.

$$t = \frac{|f_1 - f_2|}{\sqrt{f(1-f)\left(\dfrac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Kết luận:

- Chấp nhận giả thiết H_0 khi $t \leq t_{\frac{1-\alpha}{2}}$.
- Bác bỏ giả thiết H_0 khi $t > t_{\frac{1-\alpha}{2}}$.

Ví dụ 8.7. Từ hai đám đông tiến hành 2 mẫu với $n_1 = 100, n_2 = 120$ tính được tỷ lệ phần tử loại A trên mẫu 1, 2 lần lượt $f_1 = 0,2$ và $f_2 = 0,3$. Với mức ý nghĩa $\alpha = 1\%$ cho kết luận tỷ lệ phần tử A của 2 đám đông có như nhau không.

Giải. Tính $f = \dfrac{20 + 36}{100 + 120} = 0,255$.

Gọi p_1, p_2 (chưa biết) tỷ lệ phần tử A trên tổng thể 1, 2. Cần kiểm định giả thiết

$$\begin{cases} \text{Giả thiết } H_0: & p_1 = p_2 \\ \text{Đối thiết } H_1: & p_1 \neq p_2 \end{cases}$$

$$t = \frac{|0,2 - 0,3|}{\sqrt{0,255.0,745\left(\dfrac{1}{100} + \dfrac{1}{120}\right)}} = 1,695$$

Với $\alpha = 1\%$ tra bảng A.2 tính được $t_{\frac{1-\alpha}{2}} = 2,58$. Kết luận chấp nhận giả thiết H_0 hay tỷ lệ phần tử A trên 2 mẫu như nhau. □

Ví dụ 8.8. Kiểm tra 120 sinh viên trường A thấy có 80 sinh viên giỏi, 150 sinh viên trường B có 90 sinh viên giỏi. Hỏi tỷ lệ sinh viên giỏi của 2 trường như nhau không? Biết mức ý nghĩa là 5%.

Giải.

Ví dụ 8.9. Kiểm tra 230 sản phẩm của ca ngày thấy có 4 sản phẩm hỏng. Còn kiểm tra 160 sản phẩm của ca đêm thấy có 3 sản phẩm hỏng. Kết luận tỷ lệ sản phẩm hỏng phụ thuộc vào ca có đúng không với mức ý nghĩa 1%.

Giải.

8.6 Bài tập chương 8

Bài tập 8.1. Biết chiều dài của một loại sản phẩm là biến ngẫu nhiên có phân phối chuẩn. Đo ngẫu nhiên 10 sản phẩm loại này thì được chiều dài trung bình là 10,02m và độ lệch chuẩn của mẫu chưa hiệu chỉnh là 0,04m. Kiểm định giả thuyết H: "chiều dài trung bình của loại sản phẩm này là 10,0543m" có giá trị kiểm định t là bao nhiêu và cho kết luận với mức ý nghĩa 3%. **t = 2,5703; chiều dài trung bình của loại sản phẩm này là 10,0543m với mức ý nghĩa 3%** ______________________________

Bài tập 8.2. Khảo sát về thời gian tự học (giờ/tuần) của sinh viên hệ chính quy ở trường đại học A trong học kỳ này. Tiến hành lấy mẫu, người ta thu được bảng số liệu:

Thời gian	3 – 5	5 – 7	7 – 9	9 – 11	11 – 13
Số sinh viên	5	14	16	8	6

a. Tìm khoảng ước lượng thời gian tự học trung bình trong tuần của sinh viên trường A với độ tin cậy 95%. **(7,1817giờ/tuần; 8,4917giờ/tuần)**

b. Để ước lượng thời gian tự học trung bình trong tuần với độ tin cậy 95% và độ chính xác nhỏ hơn $\varepsilon = 0,6$(giờ/tuần) thì cỡ mẫu nhỏ nhất là bao nhiêu? **59**

c. Sử dụng mẫu ban đầu để ước lượng thời gian tự học trung bình trong tuần với độ chính xác $\varepsilon = 0,6$(giờ/tuần) thì đảm bảo độ tin cậy là bao nhiêu? **92,82%**

d. Những sinh viên có thời gian tự học từ 9(giờ/tuần) trở lên gọi là sinh viên "chăm học". Với độ tin cậy 95% khoảng ước lượng tỷ lệ sinh viên chăm học là bao nhiêu? **(15,92%; 41,22%)**

e. Những sinh viên có thời gian tự học từ 9(giờ/tuần) trở lên gọi là sinh viên "chăm học". Để ước lượng tỷ lệ sinh viên "chăm học" với độ tin cậy

95% và độ chính xác nhỏ hơn $\varepsilon = 0,12$ thì cỡ mẫu nhỏ nhất là bao nhiêu? **55**

f. Những sinh viên có thời gian tự học từ 9(giờ/tuần) trở lên gọi là sinh viên "chăm học". Sử dụng mẫu trên để ước lượng tỷ lệ sinh viên "chăm học" với độ chính xác $\varepsilon = 0,12$ thì đảm bảo độ tin cậy là bao nhiêu? **93,71%**

g. Tính giá trị thống kê t để kiểm định giả thuyết H: "thời gian tự học trung bình của sinh viên trường A là 8,4(giờ/tuần)" và cho kết luận với mức ý nghĩa 5%. **t = 1,6855; thời gian tự học trung bình của sinh viên trường A là 8,4(giờ/tuần) với mức ý nghĩa 5%**

h. Trong kiểm định giả thuyết H: "thời gian tự học trung bình của sinh viên trường A là 8,4(giờ/tuần)", mức ý nghĩa tối đa để giả thuyết H được chấp nhận là bao nhiêu? **9,1%**

i. Những sinh viên có thời gian tự học từ 9(giờ/tuần) trở lên gọi là sinh viên "chăm học". Tính giá trị thống kê t để kiểm định giả thuyết H: "tỷ lệ sinh viên chăm học ở trường A là 18%" và cho kết luận với mức ý nghĩa 5%. **t = 1,9261; tỷ lệ sinh viên chăm học ở trường A là 18% với mức ý nghĩa 5%**

j. Những sinh viên có thời gian tự học từ 9(giờ/tuần) trở lên gọi là sinh viên "chăm học". Trong kiểm định giả thuyết H: "tỷ lệ sinh viên chăm học ở trường A là 18%", mức ý nghĩa tối đa để giả thuyết H được chấp nhận là bao nhiêu? **5,36%**

k. Trường B khảo sát 64 sinh viên về thời gian tự học. Người ta tính được độ lệch chuẩn của mẫu có hiệu chỉnh là 2(giờ/tuần) và trung bình mẫu là 8,5(giờ/tuần). Tính giá trị thống kê t để kiểm định giả thuyết H: "thời gian tự học trung bình trong tuần của sinh viên hai trường là như nhau" và cho kết luận với mức ý nghĩa 5%. **t = 1,5893; thời gian tự học trung bình trong tuần của sinh viên hai trường là như nhau mức ý nghĩa 5%** ________________

l. Trường B khảo sát 64 sinh viên về thời gian tự học. Người ta tính được độ lệch chuẩn của mẫu có hiệu chỉnh là 2(giờ/tuần) và trung bình mẫu là 8,5(giờ/tuần). Trong kiểm định giả thuyết H: "thời gian tự học trung bình trong tuần của sinh viên hai trường là như nhau", mức ý nghĩa tối đa để giả thuyết H được chấp nhận là bao nhiêu? **11,18%** ________________

m. Những sinh viên có thời gian tự học từ 9(giờ/tuần) trở lên gọi là sinh viên "chăm học". Trường B khảo sát 64 sinh viên về thời gian tự học thấy có 28 sinh viên "chăm học". Tính giá trị thống kê t để kiểm định giả thuyết H: "tỷ lệ sinh viên "chăm học" của hai trường là như nhau" và cho kết luận với mức ý nghĩa 5%. **t = 1,6546; tỷ lệ sinh viên chăm học của hai trường là như nhau với mức ý nghĩa 5%** ________________

n. Những sinh viên có thời gian tự học từ 9(giờ/tuần) trở lên gọi là sinh viên "chăm học". Trường B khảo sát 64 sinh viên về thời gian tự học thấy có 28 sinh viên "chăm học". Trong kiểm định giả thuyết H: "tỷ lệ sinh viên "chăm học" của hai trường là như nhau", mức ý nghĩa tối đa để giả thuyết H được chấp nhận là bao nhiêu? **9,7%**

Bài tập 8.3. Khảo sát thời gian (tuần) mang thai của thai phụ không hút thuốc. Tiến hành lấy mẫu, người ta có số liệu cho như bảng sau:

Thời gian	34 – 36	36 – 38	38 – 40	40 – 42	42 – 44
Số thai phụ	7	10	59	41	4

Khoảng ước lượng thời gian mang thai trung bình của thai phụ với độ tin cậy 95% là:

A. (39,1049 tuần; 39,7215 tuần). B. (38,1049 tuần; 38,7215 tuần).

C. (37,1049 tuần; 37,7215 tuần). D. (40,1049 tuần; 40,7215 tuần).

Bài tập 8.4. Khảo sát thời gian (tuần) mang thai của thai phụ không hút thuốc. Tiến hành lấy mẫu, người ta có số liệu cho như bảng sau:

Thời gian	$34-36$	$36-38$	$38-40$	$40-42$	$42-44$
Số thai phụ	7	10	59	41	4

Để ước lượng thời gian mang thai trung bình của thai phụ với độ tin cậy 95% và độ chính xác nhỏ hơn $\varepsilon = 0,25$(tuần) thì cỡ mẫu nhỏ nhất là:

A. 175. B. 185. C. 195. D. 165.

Bài tập 8.5. Khảo sát thời gian (tuần) mang thai của thai phụ không hút thuốc. Tiến hành lấy mẫu, người ta có số liệu cho như bảng sau:

Thời gian	$34-36$	$36-38$	$38-40$	$40-42$	$42-44$
Số thai phụ	7	10	59	41	4

Sử dụng mẫu trên để ước lượng thời gian mang thai trung bình của thai phụ với độ chính xác $\varepsilon = 0,25$(tuần) thì đảm bảo độ tin cậy:

A. 86,82%. B. 87,82%. C. 88,82%. D. 89,82%.

Bài tập 8.6. Khảo sát thời gian (tuần) mang thai của thai phụ không hút thuốc. Tiến hành lấy mẫu, người ta có số liệu cho như bảng sau:

Thời gian	34 – 36	36 – 38	38 – 40	40 – 42	42 – 44
Số thai phụ	7	10	59	41	4

Những thai phụ có thời gian mang thai dưới 36 tuần là thai phụ sinh non. Với độ tin cậy 95% khoảng ước lượng tỷ lệ thai phụ sinh non:

A. (2,63%; 10,95%). B. (3,63%; 11,95%).

C. (4,63%; 12,95%). D. (1,63%; 9,95%).

Bài tập 8.7. Khảo sát thời gian (tuần) mang thai của thai phụ không hút thuốc. Tiến hành lấy mẫu, người ta có số liệu cho như bảng sau:

Thời gian	34 – 36	36 – 38	38 – 40	40 – 42	42 – 44
Số thai phụ	7	10	59	41	4

Những thai phụ có thời gian mang thai dưới 36 tuần là thai phụ sinh non. Để ước lượng tỷ lệ thai phụ sinh non với độ tin cậy 95% và độ chính xác nhỏ hơn $\varepsilon = 0,04$ thì cỡ mẫu nhỏ nhất là:

A. 121. B. 141. C. 151. D. 131.

Bài tập 8.8. Khảo sát thời gian (tuần) mang thai của thai phụ không hút thuốc. Tiến hành lấy mẫu, người ta có số liệu cho như bảng sau:

Thời gian	34 – 36	36 – 38	38 – 40	40 – 42	42 – 44
Số thai phụ	7	10	59	41	4

Những thai phụ có thời gian mang thai dưới 36 tuần là thai phụ sinh non. Sử dụng mẫu trên để ước lượng tỷ lệ thai phụ sinh non với độ chính xác $\varepsilon = 0,04$ thì đảm bảo độ tin cậy:

A. 91,99%. B. 95,99%. C. 93,99%. D. 97,99%.

Bài tập 8.9. Khảo sát thời gian (tuần) mang thai của thai phụ không hút thuốc. Tiến hành lấy mẫu, người ta có số liệu cho như bảng sau:

Thời gian	34 – 36	36 – 38	38 – 40	40 – 42	42 – 44
Số thai phụ	7	10	59	41	4

Giá trị thống kê t để kiểm định giả thuyết H: "thời gian mang thai trung bình của thai phụ là 39,7 tuần" là:

A. t = 1,8231; thời gian mang thai trung bình của thai phụ là 39,7 tuần với mức ý nghĩa 7%.

B. t = 1,8231; thời gian mang thai trung bình của thai phụ là 39,7 tuần với mức ý nghĩa 5%.

C. t = 2,8231; thời gian mang thai trung bình của thai phụ lớn hơn 39,7 tuần với mức ý nghĩa 5%.

D. t = 2,8231; thời gian mang thai trung bình của thai phụ nhỏ hơn 39,7 tuần với mức ý nghĩa 3%.

Bài tập 8.10. Khảo sát thời gian (tuần) mang thai của thai phụ không hút thuốc. Tiến hành lấy mẫu, người ta có số liệu cho như bảng sau:

Thời gian	34 – 36	36 – 38	38 – 40	40 – 42	42 – 44
Số thai phụ	7	10	59	41	4

Trong kiểm định giả thuyết H: "thời gian mang thai trung bình của thai phụ là 39,7 tuần", mức ý nghĩa tối đa để giả thuyết H được chấp nhận là:

A. 6,72%. B. 7,72%. C. 8,72%. D. 9,72%.

Bài tập 8.11. Khảo sát thời gian (tuần) mang thai của thai phụ không hút thuốc. Tiến hành lấy mẫu, người ta có số liệu cho như bảng sau:

Thời gian	34 – 36	36 – 38	38 – 40	40 – 42	42 – 44
Số thai phụ	7	10	59	41	4

Những thai phụ có thời gian mang thai dưới 36 tuần là thai phụ sinh non. Giá trị thống kê t để kiểm định giả thuyết H: "tỷ lệ thai phụ sinh non là 12%" là:

A. t = 2,1037; tỷ lệ thai phụ sinh non thấp hơn 12% với mức ý nghĩa 5%.

B. t = 2,1037; tỷ lệ thai phụ sinh non lớn hơn 12% với mức ý nghĩa 5%.

C. t = 1,1037; tỷ lệ thai phụ sinh non cao hơn 12% với mức ý nghĩa 5%.

D. t = 1,1037; tỷ lệ thai phụ sinh non là 12% với mức ý nghĩa 5%.

Bài tập 8.12. Khảo sát thời gian (tuần) mang thai của thai phụ không hút thuốc. Tiến hành lấy mẫu, người ta có số liệu cho như bảng sau:

Thời gian	34 – 36	36 – 38	38 – 40	40 – 42	42 – 44
Số thai phụ	7	10	59	41	4

Những thai phụ có thời gian mang thai dưới 36 tuần là thai phụ sinh non. Trong kiểm định giả thuyết H: "tỷ lệ thai phụ sinh non là 12%", mức ý nghĩa tối đa để giả thuyết H được chấp nhận là:

A. 3,48%. B. 4,48%. C. 5,48%. D. 6,48%.

Bài tập 8.13. Khảo sát thời gian (tuần) mang thai của thai phụ không hút thuốc. Tiến hành lấy mẫu, người ta có số liệu cho như bảng sau:

Thời gian	34 – 36	36 – 38	38 – 40	40 – 42	42 – 44
Số thai phụ	7	10	59	41	4

Khảo sát thời gian mang thai của 100 thai phụ có hút thuốc và tính được thời gian mang thai trung bình là 38,5 tuần và độ lệch chuẩn của mẫu có hiệu chỉnh 3,5 tuần. Giá trị thống kê t để kiểm định giả thuyết H: "Thời gian mang thai của thai phụ hút thuốc và không hút thuốc là như nhau" là:

A. t = 1,3798; Thời gian mang thai của thai phụ hút thuốc và không hút thuốc là như nhau với mức ý nghĩa 5%.

B. t = 1,3798; Thời gian mang thai của thai phụ hút thuốc nhỏ hơn với mức ý nghĩa 5%.

C. t = 2,3798; Thời gian mang thai của thai phụ hút thuốc lớn hơn với mức ý nghĩa 5%.

D. t = 2,3798; Thời gian mang thai của thai phụ hút thuốc nhỏ hơn với mức ý nghĩa 5%.

Bài tập 8.14. Khảo sát thời gian (tuần) mang thai của thai phụ không hút thuốc. Tiến hành lấy mẫu, người ta có số liệu cho như bảng sau:

Thời gian	34 – 36	36 – 38	38 – 40	40 – 42	42 – 44
Số thai phụ	7	10	59	41	4

Khảo sát thời gian mang thai của 100 thai phụ có hút thuốc và tính được thời gian mang thai trung bình là 38,5 tuần và độ lệch chuẩn của mẫu có hiệu chỉnh 3,5 tuần. Trong kiểm định giả thuyết H: "Thời gian mang thai của thai phụ hút thuốc và không hút thuốc là như nhau", mức ý nghĩa tối đa để giả thuyết H được chấp nhận là

A. 2,74%. B. 3,74%. C. 1,74%. D. 4,74%.

Bài tập 8.15. Khảo sát thời gian (tuần) mang thai của thai phụ không hút thuốc. Tiến hành lấy mẫu, người ta có số liệu cho như bảng sau:

Thời gian	34 – 36	36 – 38	38 – 40	40 – 42	42 – 44
Số thai phụ	7	10	59	41	4

Những thai phụ có thời gian mang thai dưới 36 tuần là thai phụ sinh non. Khảo sát thời gian mang thai của 100 thai phụ có hút thuốc và tính được thời gian mang thai thấy có 16 thai phụ sinh non. Giá trị thống kê t để kiểm định giả thuyết H: "tỷ lệ sinh non của thai phụ có hút thuốc và không hút thuốc là như nhau" là:

A. t = 2,4753; tỷ lệ sinh non của thai phụ không hút thuốc lớn hơn với mức ý nghĩa 5%.

B. t = 2,4753; tỷ lệ sinh non của thai phụ có hút thuốc lớn hơn với mức ý nghĩa 5%.

C. t = 1,4753; tỷ lệ sinh non của thai phụ không hút thuốc lớn hơn với mức ý nghĩa 5%.

D. t = 1,4753; tỷ lệ sinh non của thai phụ có hút thuốc lớn hơn với mức ý nghĩa 5%.

Bài tập 8.16. Khảo sát thời gian (tuần) mang thai của thai phụ không hút thuốc. Tiến hành lấy mẫu, người ta có số liệu cho như bảng sau:

Thời gian	34 – 36	36 – 38	38 – 40	40 – 42	42 – 44
Số thai phụ	7	10	59	41	4

Những thai phụ có thời gian mang thai dưới 36 tuần là thai phụ sinh non. Khảo sát thời gian mang thai của 100 thai phụ có hút thuốc và tính được thời gian mang thai thấy có 16 thai phụ sinh non. Trong kiểm định giả thuyết H: "tỷ lệ sinh non của thai phụ có hút thuốc và không hút thuốc là như nhau", mức ý nghĩa tối đa để giả thuyết H được chấp nhận là:

A. 1,32%. B. 2,32%. C. 3,32%. D. 4,32%.

Đáp án câu hỏi trắc nghiệm

8.3 A	**8.5** C	**8.7** D	**8.9** B	**8.11** A	**8.13** D	**8.15** B
8.4 B	**8.6** D	**8.8** C	**8.10** A	**8.12** A	**8.14** C	**8.16** A

Chương 9

Tương quan, hồi qui

9.1 Mở đầu

9.1.1 Số liệu trong phân tích tương quan, hồi qui

Quan trắc n đối tượng và ở mỗi đối tượng chúng ta "đo" 2 đại lượng X, Y. Số liệu cụ thể của n đối tượng cụ thể như sau:

$$(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$$

Ví dụ 9.1. Khảo sát chiều cao $Y(cm)$ của 10 đứa trẻ tuổi X(tháng tuổi). Mỗi đứa trẻ ta ghi nhận một cặp $(X; Y)$ và các giá trị như sau:

$$\begin{array}{ccccc} (18; 76,0) & (19; 77,0) & (19; 76,3) & (20; 77,3) & (21; 77,7) \\ (22; 78,8) & (22; 78,2) & (23; 79,0) & (24; 80,2) & (25; 80,6) \end{array}$$

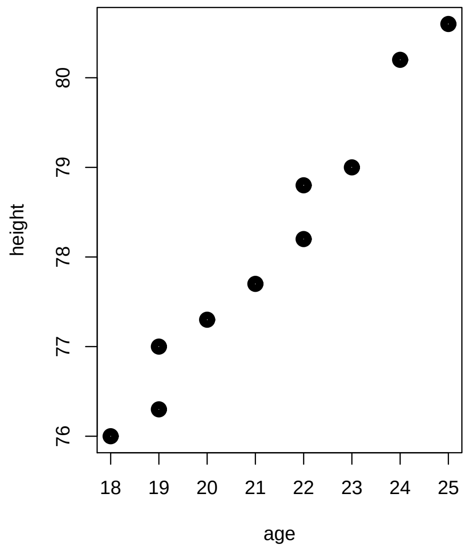
Thông thường các giá trị trên còn được xếp thành bảng như sau

X	18	19	19	20	21	22	22	23	24	25
Y	76,0	77,0	76,3	77,3	77,7	78,8	78,2	79,0	80,2	80,6

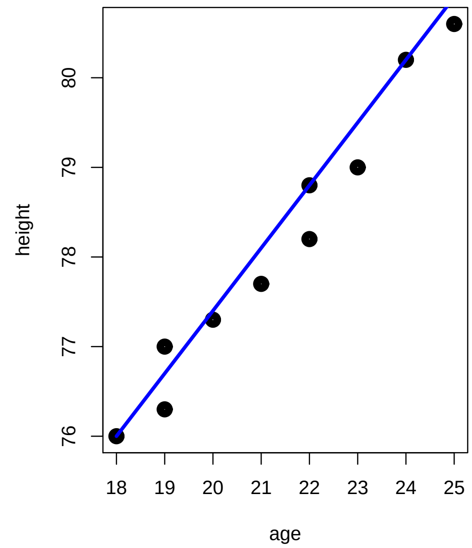
□

9.1.2 Biểu đồ tán xạ

Khi quan sát một đối tượng ta có cặp giá trị $(x_i; y_i)$. Để có được hình ảnh về sự phân tán của các cặp giá trị $(x_i; y_i)$ ta có thể biểu diễn các cặp giá trị này trên hệ trục $0xy$. Để minh họa, với số liệu ..... ta có biểu đồ tán xạ như sau



Hình a



Hình b

Ta nhận thấy hai đứa trẻ bất kỳ mặc dù cùng tuổi nhưng có chiều cao khác nhau (ngẫu nhiên) tuy nhiên xu hướng ở đây là chiều cao tăng theo độ tuổi (tất nhiên) hay chiều cao Y thay đổi một cách có hệ thống theo độ tuổi X.

Biểu đồ trên đây gợi ý cho thấy mối liên hệ giữa độ tuổi (X) và chiều cao (Y) là một đường thẳng (tuyến tính - như hình b). Để "đo lường" mối liên hệ này, chúng ta có thể sử dụng hệ số tương quan

9.2 Hệ số tương quan

Định nghĩa 9.1. *Giả sử ta có mẫu n quan trắc $(x_1, y_1), \ldots, (x_n, y_n)$. Hệ số tương quan* Pearson *được ước tính bằng công thức như sau*

$$r_{xy} = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\hat{s}_x \hat{s}_y}$$

Trong đó $\overline{xy} = \frac{1}{n} \sum_{i=1}^{n} x_i y_i$

Ý nghĩa hệ số tương quan

- r_{xy} đo mức độ quan hệ tuyến tính giữa x; y và $-1 \le r_{xy} \le 1$.

- $r_{xy} = 0$ hai biến số không có quan hệ tuyến tính, $r_{xy} = \pm 1$ thì hai biến số có quan hệ tuyến tính tuyệt đối (các cặp $(x_i; y_i)$ thuộc một đường thẳng).
- $r_{xy} < 0$ quan hệ giữa x, y là nghịch biến (có nghĩa là khi x tăng thì y giảm)
- $r_{xy} > 0$ quan hệ giữa x, y là đồng biến (có nghĩa là khi x tăng cao thì y tăng)

Ví dụ 9.2. Nghiên cứu đo lường độ cholesterol (Y) trong máu của 10 đối tượng nam của người độ tuổi (X). Kết quả đo lường như sau:

X	20	52	30	57	28	43	57	63	40	49
Y	1,9	4	2,6	4,5	2,9	3,8	4,1	4,6	3,2	4

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{451}{10} = 45,1; \quad \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{35,6}{10} = 3,56$$

$$\widehat{s}_x = 11,785; \quad \widehat{s}_y = 0,8333$$

$$\overline{xy} = \frac{1}{n}\sum_{i=1}^{n} x_i y_i = \frac{1695,4}{10} = 169,54$$

$$r_{xy} = \frac{\overline{xy} - \overline{x}.\overline{y}}{\widehat{s}_x.\widehat{s}_y} = \frac{169,54 - 33,9 \cdot 3,56}{11,785 \cdot 0.8333} = 0,914$$

□

9.3 Tìm đường thẳng hồi qui

Để tiện việc theo dõi và mô tả mô hình, gọi độ tuổi cho cá nhân i là x_i và cholesterol là y_i ở đây $i = 1, 2 \ldots 10$. Mô hình hồi tuyến tính phát biểu rằng:

$$y_i = a + bx_i + \varepsilon_i$$

Nói cách khác, phương trình trên giả định rằng độ cholesterol của một cá nhân bằng một hằng số a cộng với một hệ số b liên quan đến độ tuổi, và một sai số ε_i. Trong phương trình trên, a là chặn (intercept, tức giá trị lúc x_i=0), và b là độ dốc (slope hay gradient).

Các thông số a, b phải được ước tính từ dữ liệu. Phương pháp để ước tính các thông số này là phương pháp bình phương nhỏ nhất (least squares method). Như tên gọi, phương pháp bình phương nhỏ nhất tìm giá trị a, b sao cho tổng bình phương sai số

$$\sum_{i=1}^{n} [y_i - (a + bx_i)]^2$$

là nhỏ nhất. Sau vài thao tác toán, có thể chứng minh dễ dàng rằng, ước lượng cho a, bđáp ứng điều kiện đó là

$$b = \frac{\overline{xy} - \bar{x}.\bar{y}}{\widehat{s}_x^{\,2}}; \quad a = \bar{y} - b\bar{x}$$

Cuối cùng ta được đường hồi qui $y = a + bx$

Chú ý: $\dfrac{y - \overline{y}}{\widehat{s}_y} = r_{xy}\dfrac{x - \overline{x}}{\widehat{s}_x}$

Ví dụ 9.3. xác định phương trình hồi qui mẫu giữa tuổi và cholesterol. Từ

$$\frac{y - \overline{y}}{\widehat{s}_y} = r_{xy}\frac{x - \overline{x}}{\widehat{s}_x}$$

thay các giá trị $\bar{y}, \bar{x}, \widehat{s}_x, \widehat{s}_y, r_{xy}$ được tính ở ví dụ trên vào ta có kết quả

$$y = 0,9311 + 0,05988x$$

□

9.4 Sử dụng máy tính cầm tay

Ví dụ 9.4. Bài toán cho dạng cặp (x_i, y_i) như sau:

X	20	52	30	57	28	43	57	63	40	49
Y	1,9	4	2,6	4,5	2,9	3,8	4,1	4,6	3,2	4

Tìm hệ số tương quan r_{xy}, đường hồi qui mẫu $y = a + bx$.

a. Máy FX500MS (máy FX570MS tương tự)

- Bước 1: Nhấn phím Mod đến lúc màn hình xuất hiện **REG**; chọn (REG); Chọn (Lin)
- Bước 2: Nhập liệu 20; ,; 1.9; M+ $\cdots$
- Bước 3: Xuất kết quả Shift; chọn (S-Var); chọn (mũi tên phải 2 lần); 1(A =a); 2(B=b); 3(r=r_{xy})

b. Máy FX500ES(tương tự FX570ES)

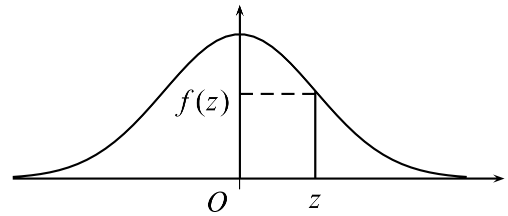
- Bước 1: SHIFT; MODE; $\downarrow$; chọn (Stat); chọn (Off)
- Bước 2: MODE; chọn (stat); chọn (A+Bx); (nhập các giá trị của X, Y vào 2 cột)
  * Nhập giá trị của X 20= 52= $\cdots$
  * Nhập giá trị của Y 1.9= 4= $\cdots$
- Bước 3: Xuất kết quả SHIFT; chọn phím (Stat); chọn (Reg); 1(A =a); 2(B=b); 3(r=r_{xy}).

Kết quả $r_{xy} = 0,9729$; $y = 0,9311 + 0,0599x$.

Phụ lục A

Các bảng giá trị xác suất

A.1 Giá trị hàm mật độ chuẩn đơn giản $f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$

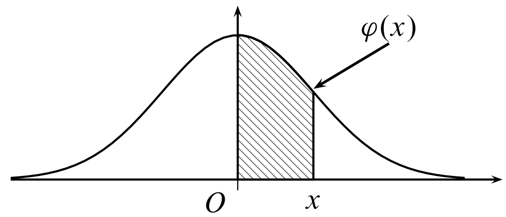


z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,3989	0,3989	0,3989	0,3988	0,3986	0,3984	0,3982	0,3980	0,3977	0,3970
0,1	0,3970	0,3965	0,3961	0,3956	0,3951	0,3945	0,3939	0,3932	0,3925	0,3911
0,2	0,3910	0,3902	0,3894	0,3885	0,3876	0,3867	0,3857	0,3847	0,3836	0,3815
0,3	0,3814	0,3802	0,3790	0,3778	0,3765	0,3752	0,3739	0,3725	0,3712	0,3684
0,4	0,3683	0,3668	0,3653	0,3637	0,3621	0,3605	0,3589	0,3572	0,3555	0,3522
0,5	0,3521	0,3503	0,3485	0,3467	0,3448	0,3429	0,3410	0,3391	0,3372	0,3334
0,6	0,3332	0,3312	0,3292	0,3271	0,3251	0,3230	0,3209	0,3187	0,3166	0,3125
0,7	0,3123	0,3101	0,3079	0,3056	0,3034	0,3011	0,2989	0,2966	0,2943	0,2899
0,8	0,2897	0,2874	0,2850	0,2827	0,2803	0,2780	0,2756	0,2732	0,2709	0,2663
0,9	0,2661	0,2637	0,2613	0,2589	0,2565	0,2541	0,2516	0,2492	0,2468	0,2422
1,0	0,2420	0,2396	0,2371	0,2347	0,2323	0,2299	0,2275	0,2251	0,2227	0,2181
1,1	0,2179	0,2155	0,2131	0,2107	0,2083	0,2059	0,2036	0,2012	0,1989	0,1944
1,2	0,1942	0,1919	0,1895	0,1872	0,1849	0,1826	0,1804	0,1781	0,1758	0,1716
1,3	0,1714	0,1691	0,1669	0,1647	0,1626	0,1604	0,1582	0,1561	0,1539	0,1499
1,4	0,1497	0,1476	0,1456	0,1435	0,1415	0,1394	0,1374	0,1354	0,1334	0,1297
1,5	0,1295	0,1276	0,1257	0,1238	0,1219	0,1200	0,1182	0,1163	0,1145	0,1111
1,6	0,1109	0,1092	0,1074	0,1057	0,1040	0,1023	0,1006	0,0989	0,0973	0,0942
1,7	0,0940	0,0925	0,0909	0,0893	0,0878	0,0863	0,0848	0,0833	0,0818	0,0791
1,8	0,0790	0,0775	0,0761	0,0748	0,0734	0,0721	0,0707	0,0694	0,0681	0,0657
1,9	0,0656	0,0644	0,0632	0,0620	0,0608	0,0596	0,0584	0,0573	0,0562	0,0541
2,0	0,0540	0,0529	0,0519	0,0508	0,0498	0,0488	0,0478	0,0468	0,0459	0,0441
2,1	0,0440	0,0431	0,0422	0,0413	0,0404	0,0396	0,0387	0,0379	0,0371	0,0356
2,2	0,0355	0,0347	0,0339	0,0332	0,0325	0,0317	0,0310	0,0303	0,0297	0,0284
2,3	0,0283	0,0277	0,0270	0,0264	0,0258	0,0252	0,0246	0,0241	0,0235	0,0224
2,4	0,0224	0,0219	0,0213	0,0208	0,0203	0,0198	0,0194	0,0189	0,0184	0,0176
2,5	0,0175	0,0171	0,0167	0,0163	0,0158	0,0154	0,0151	0,0147	0,0143	0,0136
2,6	0,0136	0,0132	0,0129	0,0126	0,0122	0,0119	0,0116	0,0113	0,0110	0,0104
2,7	0,0104	0,0101	0,0099	0,0096	0,0093	0,0091	0,0088	0,0086	0,0084	0,0079
2,8	0,0079	0,0077	0,0075	0,0073	0,0071	0,0069	0,0067	0,0065	0,0063	0,0060
2,9	0,0060	0,0058	0,0056	0,0055	0,0053	0,0051	0,0050	0,0048	0,0047	0,0044
3,0	0,0044	0,0043	0,0042	0,0040	0,0039	0,0038	0,0037	0,0036	0,0035	0,0033

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
3,1	0,0033	0,0032	0,0031	0,0030	0,0029	0,0028	0,0027	0,0026	0,0025	0,0024
3,2	0,0024	0,0023	0,0022	0,0022	0,0021	0,0020	0,0020	0,0019	0,0018	0,0017
3,3	0,0017	0,0017	0,0016	0,0016	0,0015	0,0015	0,0014	0,0014	0,0013	0,0012
3,4	0,0012	0,0012	0,0012	0,0011	0,0011	0,0010	0,0010	0,0010	0,0009	0,0009
3,5	0,0009	0,0008	0,0008	0,0008	0,0008	0,0007	0,0007	0,0007	0,0007	0,0006
3,6	0,0006	0,0006	0,0006	0,0005	0,0005	0,0005	0,0005	0,0005	0,0005	0,0004
3,7	0,0004	0,0004	0,0004	0,0004	0,0004	0,0004	0,0003	0,0003	0,0003	0,0003
3,8	0,0003	0,0003	0,0003	0,0003	0,0003	0,0002	0,0002	0,0002	0,0002	0,0002
3,9	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0001	0,0001

Bảng A.1: Giá trị hàm mật độ chuẩn hóa

A.2 Giá trị hàm $\varphi(x) = \int\limits_0^x \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right) dz$

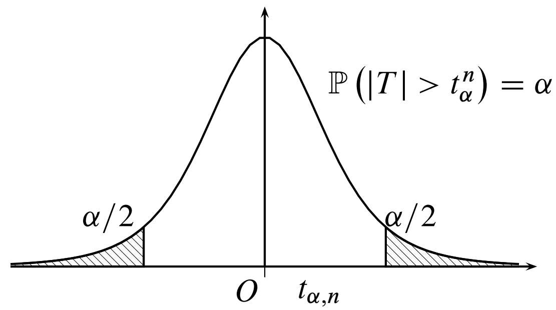


x	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0753
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	0,2257	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2517	0,2549
0,7	0,2580	0,2611	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2939	0,2967	0,2995	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,4	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319
1,5	0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418	0,4429	0,4441
1,6	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
1,7	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633
1,8	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4706
1,9	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	**0,475**	0,4756	0,4761	0,4767
2,0	0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4812	0,4817
2,1	0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	0,4850	0,4854	0,4857
2,2	0,4861	0,4864	0,4868	0,4871	0,4875	0,4878	0,4881	0,4884	0,4887	0,4890
2,3	0,4893	0,4896	0,4898	0,4901	0,4904	0,4906	0,4909	0,4911	0,4913	0,4916
2,4	0,4918	0,4920	0,4922	0,4925	0,4927	0,4929	0,4931	0,4932	0,4934	0,4936
2,5	0,4938	0,4940	0,4941	0,4943	0,4945	0,4946	0,4948	0,4949	0,4951	0,4952
2,6	0,4953	0,4955	0,4956	0,4957	0,4959	0,4960	0,4961	0,4962	0,4963	0,4964
2,7	0,4965	0,4966	0,4967	0,4968	0,4969	0,4970	0,4971	0,4972	0,4973	0,4974
2,8	0,4974	0,4975	0,4976	0,4977	0,4977	0,4978	0,4979	0,4979	0,4980	0,4981
2,9	0,4981	0,4982	0,4982	0,4983	0,4984	0,4984	0,4985	0,4985	0,4986	0,4986
3,0	0,4987	0,4987	0,4987	0,4988	0,4988	0,4989	0,4989	0,4989	0,4990	0,4990

x	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
3,1	0,4990	0,4991	0,4991	0,4991	0,4992	0,4992	0,4992	0,4992	0,4993	0,4993
3,2	0,4993	0,4993	0,4994	0,4994	0,4994	0,4994	0,4994	0,4995	0,4995	0,4995
3,3	0,4995	0,4995	0,4995	0,4996	0,4996	0,4996	0,4996	0,4996	0,4996	0,4997
3,4	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4998
3,5	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998
3,6	0,4998	0,4998	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,7	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,8	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,9	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000

Bảng A.2: Giá trị hàm φ của phân phối chuẩn đơn giản

A.3 Giá trị phân vị của luật Student ($T \sim T_n$)



n \ α	0,14	0,13	0,12	0,11	0,10	0,09	0,08	0,07	0,06	0,05	0,04	0,03	0,02	0,01
1	4,474	4,829	5,242	5,730	6,314	7,026	7,916	9,058	10,579	12,706	15,895	21,205	31,821	63,657
2	2,383	2,495	2,620	2,760	2,920	3,104	3,320	3,578	3,896	4,303	4,849	5,643	6,965	9,925
3	1,995	2,072	2,156	2,249	2,353	2,471	2,605	2,763	2,951	3,182	3,482	3,896	4,541	5,841
4	1,838	1,902	1,971	2,048	2,132	2,226	2,333	2,456	2,601	2,776	2,999	3,298	3,747	4,604
5	1,753	1,810	1,873	1,941	2,015	2,098	2,191	2,297	2,422	2,571	2,757	3,003	3,365	4,032
6	1,700	1,754	1,812	1,874	1,943	2,019	2,104	2,201	2,313	2,447	2,612	2,829	3,143	3,707
7	1,664	1,715	1,770	1,830	1,895	1,966	2,046	2,136	2,241	2,365	2,517	2,715	2,998	3,499
8	1,638	1,687	1,740	1,797	1,860	1,928	2,004	2,090	2,189	2,306	2,449	2,634	2,896	3,355
9	1,619	1,666	1,718	1,773	1,833	1,899	1,973	2,055	2,150	2,262	2,398	2,574	2,821	3,250
10	1,603	1,650	1,700	1,754	1,812	1,877	1,948	2,028	2,120	2,228	2,359	2,527	2,764	3,169
11	1,591	1,636	1,686	1,738	1,796	1,859	1,928	2,007	2,096	2,201	2,328	2,491	2,718	3,106
12	1,580	1,626	1,674	1,726	1,782	1,844	1,912	1,989	2,076	2,179	2,303	2,461	2,681	3,055
13	1,572	1,616	1,664	1,715	1,771	1,832	1,899	1,974	2,060	2,160	2,282	2,436	2,650	3,012
14	1,565	1,609	1,656	1,706	1,761	1,821	1,887	1,962	2,046	2,145	2,264	2,415	2,624	2,977
15	1,558	1,602	1,649	1,699	1,753	1,812	1,878	1,951	2,034	2,131	2,249	2,397	2,602	2,947

Bảng A.3: Giá trị phân vị của luật Student (tiếp theo)

n \ α	0,14	0,13	0,12	0,11	0,10	0,09	0,08	0,07	0,06	0,05	0,04	0,03	0,02	0,01
16	1,553	1,596	1,642	1,692	1,746	1,805	1,869	1,942	2,024	2,120	2,235	2,382	2,583	2,921
17	1,548	1,591	1,637	1,686	1,740	1,798	1,862	1,934	2,015	2,110	2,224	2,368	2,567	2,898
18	1,544	1,587	1,632	1,681	1,734	1,792	1,855	1,926	2,007	2,101	2,214	2,356	2,552	2,878
19	1,540	1,583	1,628	1,677	1,729	1,786	1,850	1,920	2,000	2,093	2,205	2,346	2,539	2,861
20	1,537	1,579	1,624	1,672	1,725	1,782	1,844	1,914	1,994	2,086	2,197	2,336	2,528	2,845
21	1,534	1,576	1,621	1,669	1,721	1,777	1,840	1,909	1,988	2,080	2,189	2,328	2,518	2,831
22	1,531	1,573	1,618	1,665	1,717	1,773	1,835	1,905	1,983	2,074	2,183	2,320	2,508	2,819
23	1,529	1,570	1,615	1,662	1,714	1,770	1,832	1,900	1,978	2,069	2,177	2,313	2,500	2,807
24	1,526	1,568	1,612	1,660	1,711	1,767	1,828	1,896	1,974	2,064	2,172	2,307	2,492	2,797
25	1,524	1,566	1,610	1,657	1,708	1,764	1,825	1,893	1,970	2,060	2,167	2,301	2,485	2,787
26	1,522	1,564	1,608	1,655	1,706	1,761	1,822	1,890	1,967	2,056	2,162	2,296	2,479	2,779
27	1,521	1,562	1,606	1,653	1,703	1,758	1,819	1,887	1,963	2,052	2,158	2,291	2,473	2,771
28	1,519	1,560	1,604	1,651	1,701	1,756	1,817	1,884	1,960	2,048	2,154	2,286	2,467	2,763
29	1,517	1,558	1,602	1,649	1,699	1,754	1,814	1,881	1,957	2,045	2,150	2,282	2,462	2,756
30	1,516	1,557	1,600	1,647	1,697	1,752	1,812	1,879	1,955	2,042	2,147	2,278	2,457	2,750
40	1,506	1,546	1,589	1,635	1,684	1,737	1,796	1,862	1,936	2,021	2,123	2,250	2,423	2,704
60	1,496	1,535	1,577	1,622	1,671	1,723	1,781	1,845	1,917	2,000	2,099	2,223	2,390	2,660
80	1,491	1,530	1,572	1,616	1,664	1,716	1,773	1,836	1,908	1,990	2,088	2,209	2,374	2,639
100	1,488	1,527	1,568	1,613	1,660	1,712	1,769	1,832	1,902	1,984	2,081	2,201	2,364	2,626
1000	1,477	1,515	1,556	1,600	1,646	1,697	1,752	1,814	1,883	1,962	2,056	2,173	2,330	2,581

Bảng A.3: Giá trị phân vị của luật Student

Phụ lục B

Giải thích lý thuyết

B.1 Ước lượng khoảng

B.1.1 Ước lượng khoảng cho trung bình

Trường hợp $X \sim X(\mu; \sigma^2)$, biết σ

Từ 6.1 trang 99 ta có

$$\bar{X} \sim N\left(\mu; \frac{\sigma^2}{n}\right) \text{ suy ra } T = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0; 1)$$

Gọi $t_{\frac{1-\alpha}{2}}$ là giá trị của T sao cho

$$\mathbb{P}\left(t_{\frac{1-\alpha}{2}} < T < t_{\frac{1-\alpha}{2}}\right) = 1 - \alpha$$

Thay T vào ta được

$$\mathbb{P}\left(\bar{X} - \frac{\sigma}{\sqrt{n}} t_{\frac{1-\alpha}{2}} < \mu < \bar{X} + \frac{\sigma}{\sqrt{n}} t_{\frac{1-\alpha}{2}}\right) = 1 - \alpha$$

Vậy ta có $\mu_1 = \bar{X} - \frac{\sigma}{\sqrt{n}} t_{\frac{1-\alpha}{2}}$ và $\mu_2 = \bar{X} + \frac{\sigma}{\sqrt{n}} t_{\frac{1-\alpha}{2}}$

Các trường hợp còn lại giải tương tự.

B.1.2 Ước lượng khoảng cho tỷ lệ

Từ 6.5 trang 100 ta có

$$X = X_1 + \ldots + X_n \dot{\sim} N\left(np; \sqrt{np(1-p)}^2\right) \text{ hay } \frac{X - np}{\sqrt{np(1-p)}} \dot{\sim} N(0;1) \tag{B.1}$$

Bởi vì $F = X/n$ là ước lượng điểm cho p cho nên $\sqrt{n(X/n)(1 - X/n)}$ sẽ xấp xỉ cho $\sqrt{np(1-p)}$, cho nên B.1 trở thành

$$T = \frac{X - np}{\sqrt{n(X/n)(1 - X/n)}} \dot{\sim} N(0;1)$$

Gọi $t_{\frac{1-\alpha}{2}}$ là giá trị của T sao cho

$$\mathbb{P}\left(t_{\frac{1-\alpha}{2}} < T < t_{\frac{1-\alpha}{2}}\right) = 1 - \alpha$$

Thay T vào ta được

$$\mathbb{P}\left(X/n - \sqrt{\frac{X/n(1-X/n)}{n}} < t_{\frac{1-\alpha}{2}} < p < X/n + \sqrt{\frac{X/n(1-X/n)}{n}} < t_{\frac{1-\alpha}{2}}\right) = 1-$$

Chú ý. Khi có mẫu cụ thể ta thay $F = X/n$ bằng giá trị f, là tỷ lệ phần tử A trên mẫu.

B.2 Kiểm định giả thiết

B.2.1 So sánh trung bình với một số

Gọi μ là trung bình của X, cần kiểm định giả thiết:

$$\begin{cases} \text{Giả thiết không } H_0: & \mu = \mu_0 \\ \text{Đối thiết } H_1: & \mu = \mu_1 \end{cases}$$

Bởi vì $\bar{X}$ là ước lượng điểm cho μ, do đó ta sẽ chấp nhận giả thiết nếu $\bar{X}$ và μ_0 không quá khác nhau. Do đó miền bác bỏ sẽ có dạng

$$C = \left\{(X_1, \ldots, X_n) : |\bar{X} - \mu_0| > c\right\} \tag{B.2}$$

với c là một giá trị nào đó.

Nếu cho trước mức ý nghĩa α, chúng ta sẽ xác định giá trị tới hạn c trong (B.2) sao cho sai lầm loại I bằng với α. Do đó, c phải thoải

$$\mathbb{P}\left(|\bar{X}-\mu_0|>c|H_0\right)=\alpha \text{ hay } \mathbb{P}\left(|\bar{X}-\mu_0|>c|\mu=\mu_0\right)=\alpha \tag{B.3}$$

Ở đây chỉ xét trường hợp à $X\sim N(\mu;\sigma^2)$ và đã biết σ. Khi $\mu=\mu_0$ thì theo (6.1) trang 99 ta có

$$T=\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}=\frac{\bar{X}-\mu_0}{\frac{\sigma}{\sqrt{n}}}\sim N(0;1)$$

Bây giờ (B.3) trở thành

$$\mathbb{P}\left(|T|>\frac{c\sqrt{n}}{\sigma}\right)=\alpha$$

Ta biết rằng $T\sim N(0;1)$ thì $\mathbb{P}\left(|T|>t_{\frac{1-\alpha}{2}}\right)=\alpha$. Cho nên ta chọn $\frac{c\sqrt{n}}{\sigma}=t_{\frac{1-\alpha}{2}}$. Vậy ta bác bỏ H_0 khi

$$T=\frac{|\bar{X}-\mu_0|}{\frac{\sigma}{\sqrt{n}}}>t_{\frac{1-\alpha}{2}}$$

B.2.2 So sánh tỷ lệ với một số

Giống như B.2.1, ở đây ta xem thống kê

$$X=X_1+\ldots+X_n\dot{\sim}N\left(np;\sqrt{np(1-p)}^2\right) \text{ hay } T=\frac{X-np}{\sqrt{np(1-p)}}\dot{\sim}N(0;1)$$

Tài liệu tham khảo

[1] Nguyễn Phú Vinh. *Xác Suất - Thống Kê Và Ứng Dụng*

[2] Đinh Văn Gắng. (1999). *Lý thuyết xác suất và thống kê toán*. NXB Giáo dục.

[3] Tô Anh Dũng. (2007). *Lý thuyết xác suất và thống kê toán*. NXB ĐHQG TP.HCM.

[4] Nguyễn Bác Văn. (1999). *Xác suất và xử lý số liệu thống kê*. NXB Giáo dục.

[5] Đặng Hấn. (1986). *Xác suất thống kê*. NXB Thống kê.

[6] Sheldon M. Ross. (1987). *Introduction to probability and statistics for engineers and scientists*. A John Wiley & Sons Publication.

[7] F.M. Dekking. (2005). *A modern introduction to Probability and Statistics*. Springer Publication.

[8] T.T. Song. (2004). *Fundamentals of probability and statistics for engineers*. A John Wiley & Sons Publication.

[9] Ronald N. Forthofer. (2007). *Biostatistics: Aguide to design, analysis, and discovery*. Academic Press.

[10] Y. Suhov. (2005). *Volume I: Basic probability and statistics*. Cambridge University Press.

[11] Michaelr. Chernick. (2003). *Introductory biostatistics for the health sciences*. A John Wiley & Sons Publication.

[12] E.L. Lehmann. (2005). *Testing statistical hypotheses: Third Edition*. Springer Publication.