XÁC SUẤT & THỐNG KÊ ĐẠI HỌC

PHÂN PHỐI CHƯƠNG TRÌNH
Số tiết: 45

PHẦN I. LÝ THUYẾT XÁC SUẤT
(*Probability theory*)

Chương 1. Các khái niệm cơ bản của xác suất
Chương 2. Biến ngẫu nhiên
Chương 3. Vector ngẫu nhiên
Chương 4. Định lý giới hạn trong xác suất

PHẦN II. LÝ THUYẾT THỐNG KÊ
(*Statistical theory*)

Chương 5. Lý thuyết mẫu
Chương 6. Ước lượng khoảng
Chương 7. Kiểm định Giả thuyết Thống kê
Chương 8. Bài toán Tương quan và Hồi quy

Tài liệu tham khảo

1. Nguyễn Phú Vinh – *Giáo trình Xác suất – Thống kê và Ứng dụng* – NXB Thống kê.
2. Nguyễn Thanh Sơn – Lê Khánh Luận – *Lý thuyết Xác suất và Thống kê toán* – NXBTKê.
3. Đậu Thế Cấp – *Xác suất – Thống kê – Lý thuyết và các bài tập* – NXB Giáo dục.
4. Lê Sĩ Đồng – *Xác suất – Thống kê và Ứng dụng* – NXB Giáo dục.
5. Đặng Hấn – *Xác suất và Thống kê* – NXB Giáo dục.
6. Phạm Xuân Kiều – *Giáo trình Xác suất và Thống kê* – NXB Giáo dục.
7. Nguyễn Cao Văn – *Giáo trình Lý thuyết Xác suất & Thống kê* – NXB Ktế Quốc dân.
8. Đào Hữu Hồ – *Xác suất Thống kê* – NXB Khoa học & Kỹ thuật.

➢ Bổ túc về Đại số Tổ hợp

1. Tính chất của các phép toán ∩, ∪

a) Tính giao hoán:
$$A \cap B = B \cap A,\ A \cup B = B \cup A.$$

b) Tính kết hợp:
$$(A \cap B) \cap C = A \cap (B \cap C),$$
$$(A \cup B) \cup C = A \cup (B \cup C).$$

c) Tính phân phối:
$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C),$$
$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C).$$

d) Tính đối ngẫu (De–Morgan):
$$\overline{A \cap B} = \overline{A} \cup \overline{B},\ \overline{A \cup B} = \overline{A} \cap \overline{B}.$$

➢ Bổ túc về Đại số Tổ hợp

2. Quy tắc nhân

- Giả sử một công việc nào đó được chia thành k giai đoạn. Có n_1 cách thực hiện giai đoạn thứ 1,..., có n_k cách thực hiện giai đoạn thứ k. Khi đó ta có:
  $n = n_1 \ldots n_k$ cách thực hiện toàn bộ công việc.
- Giả sử có k công việc $A_1, \ldots, A_k$ khác nhau. Có n_1 cách thực hiện A_1,..., có n_k cách thực hiện A_k. Khi đó ta có:
  $n = n_1 \ldots n_k$ cách thực hiện toàn bộ k công việc đó.

3. Quy tắc cộng

- Giả sử một công việc có thể thực hiện được k cách (trường hợp) loại trừ lẫn nhau: cách thứ nhất cho n_1 kết quả,…, cách thứ k cho n_k kết quả. Khi đó việc thực hiện công việc trên cho $n = n_1 + \ldots + n_k$ kết quả.

➢ Bổ túc về Đại số Tổ hợp

4. Phân biệt cách chọn *k* phần tử từ tập có *n* phần tử

Có 4 cách chọn ra k phần tử từ tập có n phần tử, n phần tử này luôn được coi là khác nhau mặc dù bản chất của chúng có thể giống nhau. Đó là:

- ➢ Chọn 1 lần ra k phần tử và *không* để ý đến thứ tự của chúng (Tổ hợp).
- ➢ Chọn 1 lần ra k phần tử và để ý đến thứ tự của chúng (Chỉnh hợp).
- ➢ Chọn k lần, mỗi lần 1 phần tử và *không hoàn lại* (số cách chọn như Chỉnh hợp).
- ➢ Chọn k lần, mỗi lần 1 phần tử và *có hoàn lại* (Chỉnh hợp lặp).

➢ Bổ túc về Đại số Tổ hợp

a) Tổ hợp

• Tổ hợp chập k của n phần tử $(0 \le k \le n)$ là một nhóm (bộ) *không phân biệt thứ tự* gồm k phần tử khác nhau được chọn từ n phần tử đã cho.
Số tổ hợp chập k của n phần tử được ký hiệu và tính theo công thức: $\boxed{C_n^k = \dfrac{n!}{k!(n-k)!}}$. Quy ước: 0! = 1.

Tính chất: $C_n^k = C_n^{n-k}$; $C_n^k = C_{n-1}^{k-1} + C_{n-1}^k$.

b) Chỉnh hợp

• Chỉnh hợp chập k của n phần tử $(0 \le k \le n)$ là một nhóm (bộ) *có thứ tự* gồm k phần tử khác nhau được chọn từ n phần tử đã cho.

➢ Bổ túc về Đại số Tổ hợp

Số chỉnh hợp chập k của n phần tử được ký hiệu và tính theo công thức:

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

c) Chỉnh hợp lặp

• Chỉnh hợp lặp k của n phần tử là một nhóm (bộ) *có thứ tự* gồm phần k tử *không nhất thiết* khác nhau được chọn từ n phần tử đã cho.
Số các chỉnh hợp lặp k của n phần tử là n^k.

Nhận xét:

Tổ hợp Chỉnh hợp Chỉnh hợp lặp

$$C_n^k < A_n^k = n(n-1)...(n-k+1) < n^k$$

PHẦN I. LÝ THUYẾT XÁC SUẤT

Chương 1. Các khái niệm cơ bản của xác suất

§1. Biến cố ngẫu nhiên
§2. Xác suất của biến cố
§3. Công thức tính xác suất

………………………

§1. BIẾN CỐ NGẪU NHIÊN

1.1. Phép thử và biến cố

• Phép thử là việc thực hiện 1 thí nghiệm hay quan sát một hiện tượng nào đó để xem có xảy ra hay không.
Phép thử mà ta không khẳng định được một cách chắc chắn kết quả trước khi thực hiện phép thử được gọi là phép thử ngẫu nhiên.

• Hiện tượng có xảy ra hay không trong phép thử được gọi là biến cố ngẫu nhiên.

➢ Chương 1. Các khái niệm cơ bản của xác suất

• Biến cố ngẫu nhiên thường được ký hiệu *A*, *B*, *C*…

VD 1

• Tung đồng tiền lên là một phép thử, biến cố là "*mặt sấp xuất hiện*" hay "*mặt ngửa xuất hiện*".

• Chọn ngẫu nhiên một số sản phẩm từ một lô hàng để kiểm tra là phép thử, biến cố là "*chọn được sản phẩm tốt*" hay "*chọn được phế phẩm*".

• Gieo một số hạt lúa là phép thử, biến cố là "*hạt lúa nảy mầm*" hay "*hạt lúa không nảy mầm*".

1.2. Phân loại biến cố

a) Biến cố sơ cấp và không gian các biến cố sơ cấp

• Trong một phép thử, các biến cố không thể phân nhỏ thành nhiều biến cố được gọi là *biến cố sơ cấp* (VD 6).
Ký hiệu các biến cố sơ cấp bởi các chữ ω_i.

vi du 1. Tung mot dong xu va mot con xuc sac (6 mat).Tim khong gian mau

➢ Chương 1. Các khái niệm cơ bản của xác suất

• Trong một phép thử, tập hợp tất cả các biến cố sơ cấp được gọi là *không gian các biến cố sơ cấp*. Ký hiệu không gian biến cố sơ cấp là $\Omega = \{\omega_i,\ i = 1, 2, ...\}$.

vidu

b) Biến cố chắc chắn và biến cố không thể

• Trong một phép thử, biến cố nhất định xảy ra (chắc chắn xảy ra) là biến cố *chắc chắn*, ký hiệu là Ω.

• Biến cố *không thể* (rỗng) là biến cố không thể xảy ra khi thực hiện phép thử, ký hiệu $\varnothing$.

VD 2. Từ một nhóm có 6 nam và 4 nữ chọn ra 5 người. Khi đó, biến cố "*chọn được 5 người nữ*" là *không thể*, biến cố "*chọn được ít nhất 1 nam*" là *chắc chắn*.

➢ Chương 1. Các khái niệm cơ bản của xác suất

1.3. Quan hệ giữa các biến cố

a) Quan hệ kéo theo

• Biến cố A được gọi là *kéo theo* biến cố B, ký hiệu $A \subset B$, khi và chỉ khi A xảy ra thì suy ra B xảy ra.

VD 3. Theo dõi 4 con gà mái đẻ trứng trong 1 ngày. Gọi:
A_i: "có i con gà mái đẻ trứng trong 1 ngày", $i = 0, 4$.
B: "có nhiều hơn 2 con gà mái đẻ trứng trong 1 ngày".
Ta có: $A_3 \subset B$, $A_4 \subset B$, $A_0 \not\subset B$, $A_1 \not\subset B$, $A_2 \not\subset B$.

b) Quan hệ tương đương

• Hai biến cố A và B được gọi là *tương đương* với nhau, ký hiệu $A = B$, khi và chỉ khi $A \subset B$ và $B \subset A$.

c) Tổng của hai biến cố

- Tổng của hai biến cố A và B là một biến cố được ký hiệu $A \cup B$ hay $A + B$, biến cố tổng xảy ra khi ít nhất một trong hai biến cố A và B xảy ra.

VD 4. Người thợ săn bắn hai viên đạn vào một con thú.
Gọi A_1: "viên đạn thứ nhất trúng con thú"
A_2: "viên đạn thứ hai trúng con thú"
A: "con thú bị bị trúng đạn" thì $A = A_1 \cup A_2$.

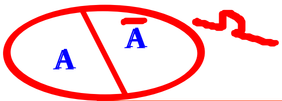
d) Tích của hai biến cố

- Tích của hai biến cố A và B là một biến cố được ký hiệu $A \cap B$ hay AB, biến cố tích xảy ra khi và chỉ khi biến cố A xảy ra và biến cố B xảy ra.

VD 5. Một người dự thi lấy bằng lái xe máy.
Gọi A: "người đó thi đạt vòng thi lý thuyết"
B: "người đó thi đạt vòng thi thực hành" và
C: "người đó lấy được bằng lái xe máy" thì $C = A \cap B$.

VD 6. Xét phép thử gieo 2 hạt lúa.
- Gọi A_i là biến cố "hạt thứ i nảy mầm" (i = 1, 2),
K_i là biến cố "hạt thứ i không nảy mầm" (i = 1, 2).
Khi đó, các biến cố tích sau đây là các biến cố sơ cấp:
$K_1 \cap K_2,\ A_1 \cap K_2,\ K_1 \cap A_2,\ A_1 \cap A_2$
và $\Omega = \{K_1K_2;\ A_1K_2;\ K_1A_2;\ A_1A_2\}$.
- Gọi B là biến cố "có 1 hạt nảy mầm" thì biến cố B không phải là biến cố sơ cấp vì $B = A_1K_2 \cup K_1A_2$.

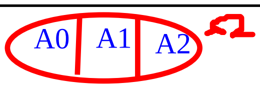
vi du. 1 dong xu va 1 con xuc sac.
A la bc xh mat chan, B la bc xh mat sap
.Tim A, B. Phat bieu va tim A+B, A.B



e) Biến cố đối lập

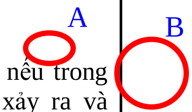
- Hiệu của hai biến cố A và B là một biến cố được ký hiệu $A \setminus B$, biến cố hiệu xảy ra khi và chỉ khi biến cố A xảy ra nhưng biến cố B không xảy ra.
- Đối lập của biến cố A là một biến cố được ký hiệu $\overline{A}$, khi $\overline{A}$ xảy ra thì A không xảy ra. Ta có $\overline{A} = \Omega \setminus A$.

VD 7. Một người bắn lần lượt 2 viên đạn vào 1 tấm bia.
Gọi A_i: "có i viên đạn trúng bia" (i = 0, 1, 2)
B: "có không quá 1 viên đạn trúng bia".
Khi đó: $\overline{B} = A_2,\ \overline{A_0} = A_1 \cup A_2$ và $\overline{A_1} = A_0 \cup A_2$.



1.4. Hệ đầy đủ các biến cố

a) Hai biến cố xung khắc



- Hai biến cố A và B được gọi là *xung khắc* nếu trong một phép thử, khi A xảy ra thì B không xảy ra và ngược lại khi B xảy ra thì A không xảy ra. AB= Ø

VD 8. Một hộp 10 viên phấn có 3 màu đỏ, vàng và xanh.
Chọn ngẫu nhiên 1 viên phấn từ hộp đó.
Gọi A: "chọn được viên phấn màu đỏ"
và B: "chọn được viên phấn màu xanh"
thì A và B là xung khắc.

Nhận xét $A.\bar{A} = \emptyset$

Hai biến cố đối lập là xung khắc, ngược lại không đúng.

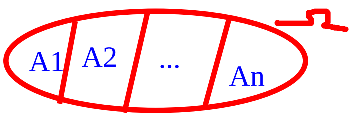
b) Hệ đầy đủ các biến cố

- Họ các biến cố $\{A_i\}$ (i = 1,…, n) được gọi là *hệ đầy đủ* các biến cố nếu thỏa mãn cả 2 điều sau:
1) Họ xung khắc, nghĩa là $A_i \cap A_j = \varnothing,\ \forall\ i \neq j$. doi mot xk
2) Có ít nhất 1 biến cố của họ xảy ra trong phép thử, nghĩa là $A_1 \cup A_2 \cup \ldots \cup A_n = \Omega$.

VD 9. Trộn lẫn 4 bao lúa vào nhau rồi bốc ra 1 hạt.
Gọi A_i: "hạt lúa bốc được là của bao thứ i", $i = \overline{1,\ 4}$.
Khi đó, hệ $\{A_1;\ A_2;\ A_3;\ A_4\}$ là đầy đủ.

Chú ý

Trong 1 phép thử, $\{A;\overline{A}\}$ là đầy đủ với biến cố A tùy ý.



§2. XÁC SUẤT CỦA BIẾN CỐ

2.1. Định nghĩa xác suất dạng cổ điển

a) Số trường hợp đồng khả năng

- Hai hay nhiều biến cố trong một phép thử có khả năng xảy ra như nhau được gọi là *đồng khả năng*.

VD 1. Trong dữ liệu máy tính của trường, ngân hàng đề có 100 đề thi. Cho máy chọn ngẫu nhiên 1 đề thì khả năng được chọn của mỗi đề thi là như nhau.

b) Định nghĩa

- Trong một phép thử có tất cả n biến cố sơ cấp đồng khả năng, trong đó có m khả năng thuận lợi cho biến cố A xuất hiện thì xác suất (*probability*) của A là:

$$\boxed{P(A) = \frac{m}{n} = \frac{\text{Số trường hợp thuận lợi cho } A \text{ xảy ra}}{\text{Số trường hợp có thể xảy ra}}.}$$

Vi du. Tung 2 con xuc xac.

➢ Chương 1. Các khái niệm cơ bản của xác suất

Nhận xét

$$0 \le P(A) \le 1,\ \forall A; \quad P(\varnothing) = 0; \quad P(\Omega) = 1.$$

VD 2. Một số điện thoại cố định tại thành phố H gồm 8 chữ số. Giả sử một người gọi một cách ngẫu nhiên đến một điện thoại cố định trong thành phố H có hai chữ số đầu là 83. Tính xác suất người đó gọi được số điện thoại:

1) Chữ số thứ ba là 7 và 5 chữ số còn lại đối xứng.
2) Chữ số thứ ba là 6, 5 chữ số còn lại khác nhau và chữ số cuối cùng là lẻ.

VD 3. Một hộp có 10 sản phẩm trong đó có 4 phế phẩm. Chọn ngẫu nhiên (1 lần) từ hộp đó ra 5 sản phẩm. Tính xác suất để có:

1) Cả 5 sản phẩm đều tốt; 2) Đúng 2 phế phẩm.

➢ Chương 1. Các khái niệm cơ bản của xác suất

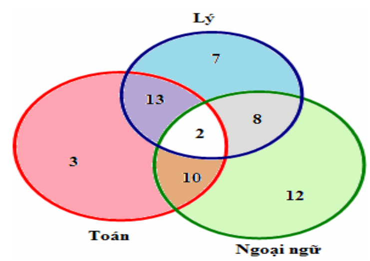
VD 4. Một bàn tròn trong một đám cưới có 10 chỗ ngồi. Giả sử mọi người ngồi vào chỗ một cách ngẫu nhiên (lấy sân khấu làm chuẩn). Tính xác suất để 1 cặp vợ chồng xác định trước ngồi cạnh nhau.

VD 5. Một lớp có 60 học sinh trong đó có 28 em giỏi Toán, 30 em giỏi Lý, 32 em giỏi Ngoại ngữ, 15 em vừa giỏi Toán vừa giỏi Lý, 10 em vừa giỏi Lý vừa giỏi Ngoại ngữ, 12 em vừa giỏi Toán vừa giỏi Ngoại ngữ, 2 em giỏi cả 3 môn. Chọn ngẫu nhiên một em học sinh của lớp.

Tính xác suất để:

1) Chọn được em giỏi ít nhất 1 môn.
2) Chọn được em chỉ giỏi môn Toán.
3) Chọn được em giỏi đúng 2 môn.

➢ Chương 1. Các khái niệm cơ bản của xác suất



Ưu điểm và hạn chế của định nghĩa dạng cổ điển

- **Ưu điểm:** Tính được chính xác giá trị của xác suất mà không cần thực hiện phép thử.
- **Hạn chế:** Trong thực tế có nhiều phép thử vô hạn các biến cố và biến cố không đồng khả năng.

➢ Chương 1. Các khái niệm cơ bản của xác suất

2.2. Định nghĩa xác suất dạng thống kê

- Thực hiện một phép thử nào đó n lần thấy có m lần biến cố A xuất hiện thì tỉ số $\frac{m}{n}$ được gọi là tần suất của biến cố A. Khi n thay đổi, tần suất cũng thay đổi nhưng luôn dao động quanh 1 số cố định $p = \lim\limits_{n \to +\infty} \frac{m}{n}$. Số p cố định này được gọi là xác suất của biến cố A theo nghĩa thống kê. Trong thực tế, khi n đủ lớn thì $P(A) \approx \frac{m}{n}$.

VD 6

- Pearson đã gieo một đồng tiền cân đối, đồng chất 12000 lần thấy có 6019 lần xuất hiện mặt sấp (tần suất 0,5016); gieo 24000 lần thấy có 12012 lần sấp (tần suất 0,5005).

➢ Chương 1. Các khái niệm cơ bản của xác suất

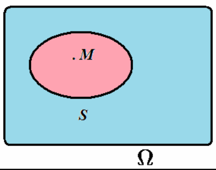
- Cramer đã nghiên cứu tỉ lệ sinh trai – gái ở Thụy Điển trong năm 1935 và kết quả có 42591 bé gái được sinh ra trong tổng số 88273 trẻ sơ sinh, tần suất là 0,4825.

Nhận xét

Định nghĩa xác suất theo dạng thống kê chỉ cho giá trị xấp xỉ và mức độ chính xác tùy thuộc vào số lần thực hiện phép thử.

2.3. Định nghĩa xác suất dạng hình học (tham khảo)

- Cho miền Ω. Gọi *độ đo* của Ω là *độ dài, diện tích, thể tích* (ứng với Ω là đường cong, miền phẳng, khối). Xét điểm M rơi ngẫu nhiên vào miền Ω.



➢ Chương 1. Các khái niệm cơ bản của xác suất

Gọi A là biến cố: "điểm M thuộc miền $S \subset \Omega$", ta có:

$$\boxed{P(A) = \frac{\text{độ đo } S}{\text{độ đo } \Omega}}.$$

VD 7. Tìm xác suất của điểm M rơi vào hình tròn nội tiếp tam giác đều có cạnh 2 *cm*.

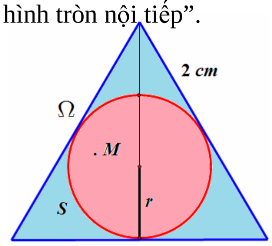
Giải. Gọi A: "điểm M rơi vào hình tròn nội tiếp".

Diện tích của tam giác là:

$$dt(\Omega) = \frac{2^2\sqrt{3}}{4} = \sqrt{3}\ cm^2.$$

Bán kính của hình tròn là:

$$r = \frac{1}{3}.\frac{2\sqrt{3}}{2} = \frac{\sqrt{3}}{3}\ cm$$



$$\Rightarrow dt(S) = \pi\left(\frac{\sqrt{3}}{3}\right)^2 = \frac{\pi}{3} \Rightarrow P(A) = \frac{\pi}{3\sqrt{3}} = 0,6046.$$

VD 8. Hai người bạn hẹn gặp nhau tại 1 địa điểm xác định trong khoảng từ 7h đến 8h. Mỗi người đến (và chắc chắn đến) điểm hẹn một cách độc lập, nếu không gặp người kia thì đợi 30 phút hoặc đến 8 giờ thì không đợi nữa. Tìm xác suất để hai người gặp nhau.

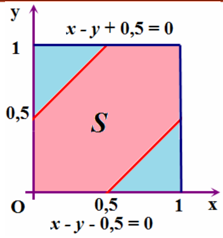
Giải. Chọn mốc thời gian 7h là 0.

Gọi *x, y* (giờ) là thời gian tương ứng của mỗi người đi đến điểm hẹn, ta có $0 \le x,\ y \le 1$ và:

$$|x-y| \le 0,5 \Leftrightarrow \begin{cases} x-y-0,5 \le 0 \\ x-y+0,5 \ge 0 \end{cases}.$$

Suy ra Ω là hình vuông và S là miền gặp nhau. Vậy:

$$P = \frac{dt(S)}{dt(\Omega)} = \frac{3}{4} = 75\%.$$



2.4. Ý nghĩa của xác suất

• Xác suất là số đo mức độ tin chắc, thường xuyên xảy ra của 1 biến cố trong phép thử.

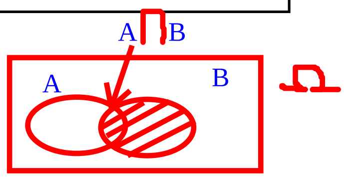
2.5. Tính chất của xác suất

1) Nếu A là biến cố tùy ý thì $0 \le P(A) \le 1$.

2) $P(\varnothing) = 0$.

3) $P(\Omega) = 1$.

4) Nếu $A \subset B$ thì $P(A) \le P(B)$.



§3. CÔNG THỨC TÍNH XÁC SUẤT

3.1. Công thức cộng xác suất

• Nếu *A* và *B* là hai biến cố tùy ý thì:

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

• Nếu *A* và *B* xung khắc thì:

$$\boxed{P(A \cup B) = P(A) + P(B).}$$

• Nếu họ $\{A_i\}$ (i = 1, 2,…, n) xung khắc từng đôi thì:

$$P(A_1 \cup A_2 \cup \ldots \cup A_n) = P(A_1) + P(A_2) + \ldots + P(A_n).$$

Đặc biệt

$$\boxed{P(\overline{A}) = 1 - P(A); \quad P(A) = P(AB) + P(A\overline{B}).}$$

VD 1. Một hộp phấn có 10 viên trong đó có 3 viên màu đỏ. Lấy ngẫu nhiên từ hộp ra 3 viên phấn.
Tính xác suất để lấy được ít nhất 1 viên phấn màu đỏ.

VD 2. Có 33 người dự thi lấy bằng lái xe 4 chỗ ngồi qua 2 vòng thi: vòng 1 thi lý thuyết và vòng 2 thi thực hành. Biết rằng có 17 người thi đỗ vòng 1, 14 người thi đỗ vòng 2 và 11 người trượt cả 2 vòng thi. Chọn ngẫu nhiên một người trong danh sách dự thi. Tìm xác suất để người đó chỉ thi đỗ 1 vòng thi.

Giải. Gọi *A*: "người đó chỉ thi đỗ 1 vòng thi",
A_i: "người đó thi đỗ vòng thứ i", $i = 1;\ 2$.

Ta có: $P(A) = P(A_1 \cup A_2) - P(A_1A_2)$
$= P(A_1) + P(A_2) - 2P(A_1A_2)$ (*).

Mặt khác: $P(A_1A_2) = 1 - P\left(A \cup \overline{A_1}.\overline{A_2}\right)$

$$= 1 - \left[P(A) + P\left(\overline{A_1}.\overline{A_2}\right) - P\left(A \cap \overline{A_1}.\overline{A_2}\right)\right]$$

$$\Rightarrow P(A_1A_2) = 1 - P(A) - \frac{11}{33} = \frac{2}{3} - P(A).$$

Thay $P(A_1A_2)$ vào (*) ta được:

$$P(A) = \frac{17}{33} + \frac{14}{33} - 2.\left[\frac{2}{3} - P(A)\right] \Rightarrow P(A) = \frac{13}{33}.$$

3.2. XÁC SUẤT CÓ ĐIỀU KIỆN

3.2.1. Định nghĩa

• Trong một phép thử, xét 2 biến cố bất kỳ *A* và *B* với $P(B) > 0$.

Xác suất có điều kiện của *A* với điều kiện ***B đã xảy ra*** được ký hiệu và định nghĩa:

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

VD 3. Một nhóm 10 sinh viên gồm 3 nam và 7 nữ trong đó có 2 nam 18 tuổi và 3 nữ 18 tuổi. Chọn ngẫu nhiên 1 sinh viên từ nhóm đó.
Gọi A: "sinh viên được chọn là nữ",
B: "sinh viên được chọn là 18 tuổi".

➢ Chương 1. Các khái niệm cơ bản của xác suất

Hãy tính $P(A)$, $P(B)$, $P(A\cap B)$, $P\left(A\middle|B\right)$, $P\left(B\middle|A\right)$?

Nhận xét

1) $P(AB)=P(A).P\left(B\middle|A\right)=P(B).P\left(A\middle|B\right)$.

2) Khi tính $P\left(A\middle|B\right)$ với điều kiện B đã xảy ra, nghĩa là ta đã hạn chế không gian mẫu Ω xuống còn B và hạn chế A xuống còn $A\cap B$.

Tính chất

1) $0\le P\left(A\middle|B\right)\le 1$; $P\left(A\middle|B\right)=0$ nếu *A*, *B* xung khắc

2) $P\left(B\middle|B\right)=1$; $P\left(\Omega\middle|B\right)=1$; $P\left(A\middle|B\right)=1$ nếu $B\subset A$

3) $P\left(\overline{A}\middle|B\right)=1-P\left(A\middle|B\right)$

4) Nếu A_1 và A_2 xung khắc thì:

$$P\left[\left(A_1\cup A_2\right)\middle|B\right]=P\left(A_1\middle|B\right)+P\left(A_2\middle|B\right).$$

➢ Chương 1. Các khái niệm cơ bản của xác suất

3.2.2. Công thức nhân xác suất

a) Sự độc lập của hai biến cố

A và *B* được gọi 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. tuc là: P(A)=P(A|B), P(B)=P(B|A)

Ví dụ, xét hai máy hoạt động trong hai dây chuyền khác nhau thì nếu có một máy hỏng cũng không ảnh hưởng đến hoạt động của máy còn lại.

…

Chú ý

Nếu *A*, *B* độc lập với nhau thì

$\overline{A}$, B độc lập; A, $\overline{B}$ độc lập và $\overline{A}$, $\overline{B}$ độc lập.

➢ Chương 1. Các khái niệm cơ bản của xác suất

b) Công thức nhân

• Nếu A và B là hai biến cố ***không độc lập*** thì:

$$\boxed{P(A\cap B)=P(B)P\left(A\middle|B\right)=P(A)P\left(B\middle|A\right).}$$

Nếu A và B là hai biến cố ***độc lập*** thì:

$$\boxed{P(A\cap B)=P(A).P(B).}$$

• Nếu n biến cố A_i, $i=1,\ldots,n$ không độc lập thì:

$$P\left(A_1A_2...A_n\right)=P\left(A_1\right)P\left(A_2\middle|A_1\right)...P\left(A_n\middle|A_1...A_{n-1}\right).$$

VD 4. Một người có 4 bóng đèn trong đó có 2 bóng bị hỏng. Người đó thử lần lượt từng bóng đèn (không hoàn lại) cho đến khi chọn được 1 bóng tốt. Tính xác suất để người đó thử đến lần thứ 2.

Ai la chon dc bong tot lan thu i, i=1,3.

➢ Chương 1. Các khái niệm cơ bản của xác suất

VD 5. Một người nhốt chung 5 con gà mái và 4 con gà trống trong 1 chiếc lồng đem đi bán. Người bán bắt ngẫu nhiên ra 1 con gà và bán nó, tiếp đến người bán cũng bắt ngẫu nhiên ra 1 con khác.

Tính xác suất người bán bắt được con gà thứ hai là gà trống nếu:

1) Con gà thứ nhất đã bán là gà mái. P($\overline{A2}$|A1)=0.5

2) Người bán không nhớ đã bán con gà trống hay mái. P(A2)=0.4

VD 6. Một cầu thủ bóng rổ có 4 quả bóng đang ném từng quả vào rổ. Nếu bóng vào rổ hoặc hết bóng thì cầu thủ ngừng ném. Biết các lần ném là độc lập và xác suất vào rổ của quả bóng thứ 1, 2, 3, 4 lần lượt là 90%, 80%, 85%, 70%. Tính xác suất cầu thủ ném được bóng vào rổ.

ví du: cho 3 vé sô trong đó có 2 vé trúng. chon lan luot 2 vé. tính xác xuât moi vé.

➢ Chương 1. Các khái niệm cơ bản của xác suất

VD 7. Một người nông dân tiến hành phun thuốc trừ sâu hại lúa 3 lần liên tiếp trong 1 tuần. Xác suất sâu chết sau lần phun thứ nhất là 0,5. Nếu sâu sống sót thì khả năng sâu chết sau lần phun thứ hai là 0,7; tương tự, sau lần phun thứ ba là 0,9. Tính xác suất sâu bị chết sau 3 lần phun thuốc.

VD 8. Trong dịp tết, một người *A* đem bán 1 cây mai lớn và 1 cây mai nhỏ. Xác suất bán được cây mai lớn là 0,9. Nếu bán được cây mai lớn thì xác suất bán được cây mai nhỏ là 0,7. Nếu cây mai lớn không bán được thì xác suất bán được cây mai nhỏ là 0,2. Biết rằng người *A* bán được ít nhất 1 cây mai, xác suất để người *A* bán được cả hai cây mai là:

A. 0,63; B. 0,6848; C. 0,4796; D. 0,87.

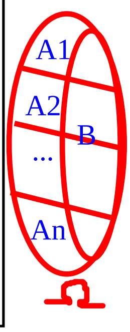
➢ Chương 1. Các khái niệm cơ bản của xác suất

3.2.3. Công thức xác suất đầy đủ và Bayes.

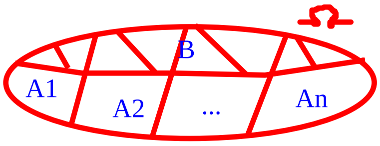
a) Công thức xác suất đầy đủ

• Cho họ các biến cố $\{A_i\}$, $i=\overline{1;\,n}$ đầy đủ và *B* là biến cố bất kỳ trong phép thử, ta có:

$$\boxed{\begin{aligned}P(B)&=\sum_{i=1}^{n}P(A_i)\left(B\middle|A_i\right)\\&=P(A_1)P\left(B\middle|A_1\right)+\ldots+P(A_n)P\left(B\middle|A_n\right).\end{aligned}}$$



VD 9. Một cửa hàng bán hai loại bóng đèn cùng kích cỡ gồm: 70 bóng màu trắng với tỉ lệ bóng hỏng là 1% và 30 bóng màu vàng với tỉ lệ hỏng 2%. Một khách hàng chọn mua ngẫu nhiên 1 bóng đèn từ cửa hàng này. Tính xác suất để người này mua được bóng đèn tốt ?



➤ Chương 1. Các khái niệm cơ bản của xác suất

b) Công thức Bayes

• Cho họ các biến cố $\{A_i\}$, $i = \overline{1; n}$ đầy đủ và *B* là biến cố bất kỳ trong phép thử.
Xác suất để A_i xuất hiện sau khi đã xuất hiện *B* là:

$$\boxed{P\left(A_i|B\right) = \frac{P(A_i)P\left(B|A_i\right)}{\sum_{i=1}^{n} P(A_i)P\left(B|A_i\right)} = \frac{P(A_i)P\left(B|A_i\right)}{P(B)}.}$$

VD 10. (Xét tiếp VD 9) Giả sử khách hàng chọn mua được bóng đèn tốt. Tính xác suất để người này mua được bóng đèn màu vàng ?

➤ Chương 1. Các khái niệm cơ bản của xác suất

VD 11. Tỉ lệ ôtô tải, ôtô con và xe máy đi qua đường X có trạm bơm dầu là 5 : 2 : 13. Xác suất để ôtô tải, ôtô con và xe máy đi qua đường này vào bơm dầu lần lượt là 0,1; 0,2 và 0,15. Biết rằng có 1 xe đi qua đường X vào bơm dầu, tính xác suất để đó là ôtô con ?

VD 12. Một thống kê cho thấy tỉ lệ 1 cặp trẻ sinh đôi khác trứng có cùng giới tính là 0,495; cặp trẻ sinh đôi cùng trứng thì luôn có cùng giới tính. Giả sử tỉ lệ cặp trẻ sinh đôi cùng trứng là p (tính trên tổng số các cặp trẻ sinh đôi). Nếu biết 1 cặp trẻ sinh đôi có ***cùng giới tính*** thì xác suất chúng được sinh đôi cùng trứng là 50/149, giá trị p là:
A. $p = 0,05$; B. $p = 0,1$; C. $p = 0,2$; D. $p = 0,23$.

Cong thuc Bernoulli

➤ Chương 2. Biến ngẫu nhiên

§1. Biến ngẫu nhiên và hàm phân phối
§2. Các đặc trưng số của biến ngẫu nhiên
§3. Một số luật phân phối xác suất thông dụng

………………………

§1. BIẾN NGẪU NHIÊN VÀ HÀM PHÂN PHỐI

1.1. Biến ngẫu nhiên

1.1.1. Khái niệm và phân loại biến ngẫu nhiên

a) Khái niệm

• Một biến cố được gọi là *ngẫu nhiên* nếu trong kết quả của phép thử nó sẽ nhận *một và chỉ một* trong các giá trị có thể có của nó tùy thuộc vào sự tác động của các nhân tố ngẫu nhiên.
• Các biến ngẫu nhiên được ký hiệu: *X*, *Y*, *Z*, … các giá trị tương ứng của chúng là *x*, *y*, *z*,…

➤ Chương 2. Biến ngẫu nhiên

VD 1. Khi tiến hành gieo *n* hạt đậu ta chưa thể biết có bao nhiêu hạt sẽ nảy mầm, số hạt nảy mầm có thể có là 0, 1, …, *n*.
Kết thúc phép thử gieo hạt thì ta sẽ biết chắc chắn có bao nhiêu hạt nảy mầm. Gọi *X* là số hạt nảy mầm thì là *X* biến ngẫu nhiên và *X* = {0, 1, 2, …, *n*}.

VD. Tung dong xu 10 lan. Goi X la so lan xh mat sap

b) Phân loại biến ngẫu nhiên X là bnn va X={0,1,2...10}

• Biến ngẫu nhiên (BNN) được gọi là *rời rạc* nếu các giá trị có thể có của nó lập nên 1 tập hợp hữu hạn hoặc đếm được.
• Biến ngẫu nhiên được gọi là *liên tục* nếu các giá trị có thể có của nó lấp đầy 1 khoảng trên trục số.

➤ Chương 2. Biến ngẫu nhiên

VD 2
• Biến *X* trong VD 1 là BNN rời rạc (tập hữu hạn).
• Gọi *Y* là số người đi qua 1 ngã tư trên đường phố thì *Y* là BNN rời rạc (tập đếm được).
• Bắn 1 viên đạn vào bia, gọi *X* (*cm*) là "khoảng cách từ điểm chạm của viên đạn đến tâm của bia" thì *X* là BNN liên tục.
• Gọi *Y* là "sai số khi đo 1 đại lượng vật lý" thì *Y* là BNN liên tục.

➤ Chương 2. Biến ngẫu nhiên

1.1.2. BNN rời rạc, bảng phân phối xác suất

• Cho biến ngẫu nhiên rời rạc *X*, $X = \{x_1, x_2, ..., x_n, ...\}$ với xác suất tương ứng là $P(X = x_i) = p_i$, $i = 1, 2, ...$
Ta có *phân phối xác suất* của *X* ở dạng bảng:

X	x_1	x_2	$\ldots$	x_n	$\ldots$
$P(X = x_i)$	p_1	p_2	$\ldots$	p_n	$\ldots$

Chú ý
1) $p_i \geq 0$; $\sum p_i = 1$, $i = 1, 2, ...$
2) Trong trường hợp các giá trị x_i, p_i có tính quy luật, thay cho việc lập bảng ta có thể mô tả bởi đẳng thức:
$P(X = x_i) = p_i$, $i = 1, 2, ...$

VD 3. Xác suất để 1 người thi đạt mỗi khi thi lấy bằng lái xe là 0,3. Người đó thi cho đến khi đạt mới thôi.
Gọi X là số lần người đó dự thi (mỗi lần thi là độc lập).
1) Lập bảng phân phối xác suất của X.
2) Tính xác suất để người đó phải thi không ít hơn 3 lần.

VD 4. Một hộp có 3 viên phấn trắng và 2 viên phấn đỏ. Một người lấy phấn ngẫu nhiên lần lượt (mỗi lần 1 viên và không trả lại) từ hộp đó ra cho đến khi lấy được 2 viên phấn đỏ. Gọi X là số lần người đó lấy phấn.
Hãy lập bảng phân phối xác suất của X ?

VD 5. Cho hai BNN X, Y độc lập với bảng ppxs như sau:

X	0	1	2
$P(X = x_i)$	0,3	0,4	0,3

Y	-1	1
$P(Y = y_j)$	0,4	0,6

Hãy lập bảng phân phối xác suất của X^2, $X+Y$, XY.

Giải

• Ta có $P(X^2 = x_i^2) = P(X = x_i)$, suy ra:

X^2	0	1	4
$P(X^2 = x_i^2)$	0,3	0,4	0,3

• Ta có $(X+Y=-1) = (X=0) \cap (Y=-1)$
$\Rightarrow P(X+Y=-1) = P(X=0).P(Y=-1) = 0,12;$
$P(X+Y=0) = P(X=1).P(Y=-1) = 0,16;$
$P(X+Y=1) = P(X=0).P(Y=1)$
$+ P(X=2).P(Y=-1) = 0,30;$

$P(X+Y=2) = P(X=1).P(Y=1) = 0,24;$
$P(X+Y=3) = P(X=2).P(Y=1) = 0,18.$

$X+Y$	-1	0	1	2	3
$P(X+Y=k)$	0,12	0,16	0,30	0,24	0,18

• Ta có $P(XY=-2) = P(X=2).P(Y=-1) = 0,12;$
$P(XY=-1) = P(X=1).P(Y=-1) = 0,16;$
$P(XY=0) = P(X=0).P(Y=-1)$
$+ P(X=0).P(Y=1) = 0,30;$
$P(XY=1) = P(X=1).P(Y=1) = 0,24;$
$P(XY=2) = P(X=2).P(Y=1) = 0,18.$

XY	-2	-1	0	1	2
$P(XY=k)$	0,12	0,16	0,30	0,24	0,18

VD 6. Cho bảng ppxs đồng thời của hai BNN X và Y:

X \ Y	-1	0	1
1	0,10	0,15	0,05
2	0,30	0,20	0,20

Hãy lập bảng phân phối xác suất của BNN Z nếu:
1) $Z = 2X - Y + 5$; 2) $Z = X^2 - Y^2$.

Giải. 1) $(X;Y) = (1;-1) \Rightarrow Z = 8,\ p = 0,1;$
$(X;Y) = (1;\ 0) \Rightarrow Z = 7,\ p = 0,15;$
$(X;Y) = (1;\ 1) \Rightarrow Z = 6,\ p = 0,05;$
$(X;Y) = (2;-1) \Rightarrow Z = 10,\ p = 0,3;$

$(X;Y) = (2;\ 0) \Rightarrow Z = 9,\ p = 0,2;$
$(X;Y) = (2;\ 1) \Rightarrow Z = 8,\ p = 0,2.$
Sắp xếp các giá trị của Z và xác suất tương ứng, ta có:

Z	6	7	8	9	10
$P(Z=k)$	0,05	0,15	0,30	0,20	0,30

2) $(X;Y) = (1;-1) \Rightarrow Z = 0,\ p = 0,1;$
$(X;Y) = (1;\ 0) \Rightarrow Z = 1,\ p = 0,15;$
$(X;Y) = (1;\ 1) \Rightarrow Z = 0,\ p = 0,05;$
$(X;Y) = (2;-1) \Rightarrow Z = 3,\ p = 0,3;$
$(X;Y) = (2;\ 0) \Rightarrow Z = 4,\ p = 0,2;$
$(X;Y) = (2;\ 1) \Rightarrow Z = 3,\ p = 0,2.$
Sắp xếp các giá trị của Z và xác suất tương ứng, ta có:

Z	0	1	3	4
$P(Z=k)$	0,15	0,15	0,50	0,20

1.1.3. Biến ngẫu nhiên liên tục, hàm mật độ

• Cho BNN liên tục X. Hàm $f(x)$, $x \in \mathbb{R}$ được gọi là *hàm mật độ xác suất* của X nếu thỏa hai điều kiện:

➢ $f(x) \geq 0,\ \forall x \in \mathbb{R}$;

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

• Khi đó, xác suất $P(a < X < b) = \int\limits_a^b f(x)dx.$

Chú ý

1) Đôi khi người ta dùng ký hiệu $f_X(x)$ để chỉ hàm mật độ xác suất (gọi tắt là hàm mật độ) của X.

➢ Chương 2. Biến ngẫu nhiên

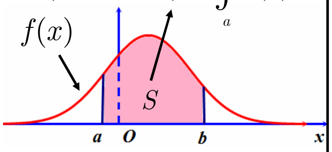
2) Do $P(X=a)=\int_a^a f(x)dx = 0$ nên ta suy ra:

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

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

3) Về mặt hình học, xác suất của BNN X nhận giá trị trong $(a;\ b)$ bằng diện tích hình thang cong giới hạn bởi $x=a,\ x=b,\ y=f(x)$ và trục Ox.

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



➢ Chương 2. Biến ngẫu nhiên

4) Nếu $f(x)$ thỏa $f(x) \ge 0,\ \forall x \in \mathbb{R}$ và $\int_{-\infty}^{+\infty} f(x)dx = 1$

thì $f(x)$ là hàm mật độ của BNN X nào đó.

VD 7. Chứng tỏ $f(x) = \begin{cases} 4x^3, & x \in (0;\ 1) \\ 0, & x \notin (0;\ 1) \end{cases}$ là hàm mật độ của biến ngẫu nhiên X.

VD 8. Cho biến ngẫu nhiên X có hàm mật độ:

$$f(x) = \begin{cases} 0, & x < 1 \\ \dfrac{k}{x^2}, & x \ge 1. \end{cases}$$

Tìm k và tính $P(-3 < X \le 2)$.

➢ Chương 2. Biến ngẫu nhiên

1.2. HÀM PHÂN PHỐI XÁC SUẤT

1.2.1. Định nghĩa

- Hàm phân phối xác suất (gọi tắt là hàm phân phối) của biến ngẫu nhiên X, ký hiệu $F(x)$ hoặc $F_X(x)$, là xác suất để X nhận *giá trị nhỏ hơn* x (với mọi $x \in \mathbb{R}$).

  Nghĩa là: $\boxed{F(x) = P(X < x),\ \forall x \in \mathbb{R}}$.

➢ Nếu biến ngẫu nhiên X rời rạc với xác suất $P(X = x_i) = p_i$ thì $F(x) = \sum_{x_i < x} p_i$.

➢ Nếu biến ngẫu nhiên X liên tục với hàm mật độ $f(x)$ thì $F(x) = \int_{-\infty}^{x} f(t)dt$.

➢ Chương 2. Biến ngẫu nhiên

Nhận xét

1) Giả sử BNN X chỉ nhận các giá trị trong $[x_1;\ x_n]$ và $x_1 < x_2 < x_3 < ... < x_n$, $P(X = x_i) = p_i\ \left(i = \overline{1,\ n}\right)$.

Ta có hàm phân phối của X:

$$F(x) = \begin{cases} 0 & \text{nếu } x \le x_1 \\ p_1 & \text{nếu } x_1 < x \le x_2 \\ p_1 + p_2 & \text{nếu } x_2 < x \le x_3 \\ \dots\dots\dots & \\ p_1 + p_2 + ... + p_{n-1} & \text{nếu } x_{n-1} < x \le x_n \\ 1 & \text{nếu } x > x_n. \end{cases}$$

➢ Chương 2. Biến ngẫu nhiên

2) Mối liên hệ của $F(x)$ với xác suất và hàm mật độ xác suất:

➢ $p_i = F(x_{i+1}) - F(x_i)$.

➢ Nếu X là biến ngẫu nhiên liên tục thì hàm $F(x)$ liên tục tại mọi $x \in \mathbb{R}$ và $F'(x) = f(x)$.

1.2.2. Tính chất cơ bản của hàm phân phối

1) Hàm $F(x)$ xác định với mọi $x \in \mathbb{R}$.

2) $0 \le F(x) \le 1, \forall x \in \mathbb{R};\ F(-\infty) = 0; F(+\infty) = 1$.

3) $F(x)$ không giảm: $F(x_1) \le F(x_2)$ nếu $x_1 < x_2$.

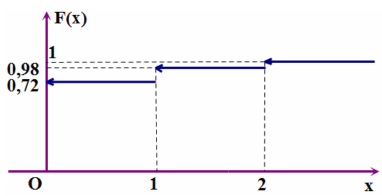
4) $P(a \le X < b) = F(b) - F(a)$.

➢ Chương 2. Biến ngẫu nhiên

VD 9. Một phân xưởng có 2 máy hoạt động độc lập. Xác suất trong 1 ngày làm việc các máy đó hỏng tương ứng là 0,1 và 0,2. Gọi X là số máy hỏng trong 1 ngày làm việc. Lập hàm phân phối xác suất của X và vẽ đồ thị.

BPPXS

Đồ thị:



➢ Chương 2. Biến ngẫu nhiên

VD 10. Tuổi thọ X (giờ) của 1 thiết bị có hàm mật độ là:

$$f(x) = \begin{cases} 0, & x < 100 \\ \dfrac{100}{x^2}, & x \geq 100. \end{cases}$$

1) Tìm hàm phân phối xác suất của X.
2) Thiết bị được gọi là loại A nếu tuổi thọ của nó kéo dài ít nhất là 400 giờ. Tính tỉ lệ thiết bị loại A.

VD 11. BNN X có $f(x) = \begin{cases} a\cos x, -\dfrac{\pi}{2} \leq x \leq \dfrac{\pi}{2} \\ 0, x < -\dfrac{\pi}{2} \,\&\, x > \dfrac{\pi}{2}. \end{cases}$

Tìm a và hàm phân phối xác suất $F(x)$.

➢ Chương 2. Biến ngẫu nhiên

VD 12. Thời gian chờ phục vụ của khách hàng là BNN X (phút) liên tục có hàm phân phối xác suất:

$$F(x) = \begin{cases} 0, & x \leq -2 \\ ax^3 + 8a, & x \in (-2;\ 3]. \\ 1, & x > 3. \end{cases}$$

1) Tìm hàm mật độ xác suất $f(x)$ của X.
2) Tính $P\left(\sqrt{2} < Y \leq \sqrt{5}\right)$ với $Y = \sqrt{X^2 + 1}$.

➢ Chương 2. Biến ngẫu nhiên

§2. CÁC ĐẶC TRƯNG SỐ CỦA BIẾN NGẪU NHIÊN

- Những thông tin cô đọng phản ánh từng phần về biến ngẫu nhiên giúp ta so sánh giữa các đại lượng với nhau được gọi là *các đặc trưng số*.
- Có ba loại đặc trưng số:
  - ➢ Các đặc trưng số cho xu hướng trung tâm của BNN: *Kỳ vọng toán, Trung vị, Mode,…*
  - ➢ Các đặc trưng số cho độ phân tán của BNN: *Phương sai, Độ lệch chuẩn,…*
  - ➢ Các đặc trưng số cho dạng phân phối xác suất.

2.1. KỲ VỌNG TOÁN (giá trị trung bình)

2.1.1. Định nghĩa

- Kỳ vọng toán (gọi tắt là kỳ vọng – *Expectation*) của biến ngẫu nhiên X, ký hiệu EX hay $M(X)$, là một con số được xác định như sau:

➢ Chương 2. Biến ngẫu nhiên

➢ Nếu X rời rạc với xác suất $P(X = x_i) = p_i$ thì:

$$\boxed{EX = \sum_i x_i p_i.}$$

➢ Nếu X liên tục có hàm mật độ $f(x)$ thì:

$$\boxed{EX = \int_{-\infty}^{+\infty} x.f(x)dx.}$$

Đặc biệt

- Nếu biến ngẫu nhiên rời rạc $X = \{x_1;\ x_2;\ldots;\ x_n\}$ với xác suất tương ứng là $p_1,\ p_2,\ldots,\ p_n$ thì:

$$\boxed{EX = x_1p_1 + x_2p_2 + \ldots + x_np_n.}$$

➢ Chương 2. Biến ngẫu nhiên

VD 1. Một lô hàng gồm 10 sản phẩm tốt và 2 phế phẩm. Lấy ngẫu nhiên 4 sản phẩm từ lô hàng đó, gọi X là số sản phẩm tốt trong 4 sản phẩm lấy ra.
Tìm phân phối xác suất và tính kỳ vọng của X.

VD 2. Tìm kỳ vọng của biến ngẫu nhiên X có hàm mật độ xác suất $f(x) = \begin{cases} \dfrac{3}{4}(x^2 + 2x), & x \in (0;\ 1) \\ 0, & x \notin (0;\ 1). \end{cases}$

Chú ý

1) Nếu X là BNN liên tục trên $[a;\ b]$ thì $EX \in [a;\ b]$.
2) Nếu $X = \{x_1,\ldots,x_n\}$ thì:
$$EX \in [\min\{x_1,\ldots,x_n\};\ \max\{x_1,\ldots,x_n\}].$$

➢ Chương 2. Biến ngẫu nhiên

VD 3. Cho BNN X có bảng phân phối xác suất:

X	0	0,1	0,3	0,4	0,7
P	a	0,2	b	0,2	0,1

Giá trị của tham số a và b để EX = 0,2 là:
A. a = 0,1 và b = 0,1; B. a = 0,4 và b = 0,1;
C. a = 0,2 và b = 0,3; D. a = 0,3 và b = 0,2.

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

$$f(x) = \begin{cases} ax + bx^2, & x \in (0;\ 1) \\ 0, & x \notin (0;\ 1). \end{cases}$$

Cho biết EX = 0,6. Hãy tính $P\left(X < \dfrac{1}{2}\right)$.

2.1.2. Ý nghĩa của Kỳ vọng

- Kỳ vọng của biến ngẫu nhiên X là ***giá trị trung bình*** (theo xác suất) mà X nhận được, nó phản ánh giá trị trung tâm của phân phối xác suất của X.
- Trong thực tế sản xuất hay kinh doanh nếu cần chọn phương án cho ***năng suất*** hay ***lợi nhuận*** cao, người ta chọn phương án sao cho ***năng suất kỳ vọng*** hay ***lợi nhuận kỳ vọng*** cao.

VD 5. 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?

VD 6. Một dự án xây dựng được viện *C* thiết kế cho cả 2 bên *A* và *B* xét duyệt một cách độc lập. Xác suất (khả năng) để *A* và *B* chấp nhận dự án này khi xét duyệt thiết kế là 70% và 80%. Nếu chấp nhận dự án thì bên *A* phải trả cho *C* là 400 triệu đồng, còn ngược lại thì phải trả 100 triệu đồng. Nếu chấp nhận dự án thì bên *B* phải trả cho *C* là 1 tỉ đồng, còn ngược lại thì phải trả 300 triệu đồng. Biết chi phí cho thiết kế của *C* là 1 tỉ đồng và 10% thuế doanh thu. Hỏi trung bình viện *C* có lãi bao nhiêu khi nhận thiết kế trên?

Giải. Gọi X (triệu đồng) là tiền lãi (đã trừ thuế) của *C*.

- Khi cả *A* và *B* đều chấp nhận dự án thì:
  $X = (400+1000).0,9 - 1000 = 260$
  và xác suất $p = 0,7.0,8 = 0,56$.

- Khi *A* chấp nhận và *B* không chấp nhận dự án thì:
  $X = (400+300).0,9 - 1000 = -370$
  và xác suất $p = 0,7.0,2 = 0,14$.
- Khi *A* không chấp nhận và *B* chấp nhận dự án thì:
  $X = (100+1000).0,9 - 1000 = -10$
  và xác suất $p = 0,3.0,8 = 0,24$.
- Khi cả *A* và *B* đều không chấp nhận dự án thì:
  $X = (100+300).0,9 - 1000 = -640$
  và xác suất $p = 0,3.0,2 = 0,06$.

Tiền lãi trung bình (đã trừ thuế) của viện *C* là:

$$EX = 260.0,56 - 370.0,14 - 10.0,24 - 640.0,06$$
$$= 53 \text{ (triệu đồng)}.$$

VD 7. 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?

A. 2,185 triệu đồng; B. 2,148 triệu đồng.
C. 2,116 triệu đồng; D. 2,062 triệu đồng.

2.1.3. Tính chất của Kỳ vọng

1) $EC = C,\ C \in \mathbb{R}$.

2) $E(CX) = C.EX,\ C \in \mathbb{R}$.

3) $E(X \pm Y) = EX \pm EY$.

4) $E(X.Y) = EX.EY$ nếu X, Y độc lập.

5) Khi $Y = \varphi(X)$ thì:

$$EY = \begin{cases} \sum_i \varphi(x_i).p_i, & \text{nếu } X \text{ rời rạc} \\ \int_{-\infty}^{+\infty} \varphi(x).f(x)dx, & \text{nếu } X \text{ liên tục.} \end{cases}$$

VD 7. Tính EY với $Y = \varphi(X) = X^2 - 3$, biết X có bảng phân phối xác suất:

X	–1	0	1	2
P	0,1	0,3	0,35	0,25

VD 8. Cho BNN X có hàm $f(x) = \begin{cases} \dfrac{2}{x^2}, & x \in [1;\ 2] \\ 0, & x \notin [1;\ 2] \end{cases}$.

1) Tính EX; 2) Tính EY với $Y = X^5 - \dfrac{2}{X}$.

➢ Chương 2. Biến ngẫu nhiên

2.3. Trung vị và Mode

2.3.1. Trung vị (tham khảo)

- Trung vị (*median*) của biến ngẫu nhiên *X*, ký hiệu med*X*, là số thực *m* thỏa:

$$P(X < m) \leq \frac{1}{2} \text{ và } P(X \geq m) \geq \frac{1}{2}.$$

➢ Với *X* rời rạc thì med*X* = x_i nếu:

$$F(x_i) \leq \frac{1}{2} \leq F(x_{i+1}).$$

➢ Với *X* liên tục thì med*X* = *m* nếu:

$$F(m) = \int_{-\infty}^{m} f(x)dx = 0,5.$$

➢ Chương 2. Biến ngẫu nhiên

Nhận xét

- Nếu có 2 trung vị $m_1 \neq m_2$ thì $\forall m \in [m_1; m_2]$ cũng là trung vị.
- Trung vị là điểm phân đôi xác suất thành 2 phần tương đối bằng nhau.

VD 9. Cho BNN *X* có bảng phân phối xác suất:

X	1	2	3	**4**	5
P	0,10	0,20	0,15	**0,12**	0,43

Khi đó ta có med*X* = 4.

VD 10. Cho BNN*X* có bảng phân phối xác suất:

X	–1	0	1	2
P	0,51	0,19	0,16	0,14

Ta có *medX* = –1 nhưng quá lệch!

➢ Chương 2. Biến ngẫu nhiên

VD 11. Cho BNN X có hàm mật độ $f(x) = \begin{cases} \frac{4}{x^5}, & x \geq 1 \\ 0, & x < 1. \end{cases}$

Tìm med*X*.

Giải. $\int_{-\infty}^{m} f(x)dx = 0,5 \Rightarrow 4\int_{1}^{m} \frac{dx}{x^5} = 0,5 \Rightarrow m = \sqrt[4]{2}.$

2.3.2. MODE

Định nghĩa. Mode của biến ngẫu nhiên X, ký hiệu $\mathrm{mod}\,X$, là giá trị của X thỏa:

➢ Nếu BNN X rời rạc thì:

$$\mathrm{mod}\,X = \left\{x_0 \middle| P(X = x_0) \ \max\right\}.$$

➢ Nếu BNN *X* liên tục có hàm mật độ $f(x)$ thì:

$$\mathrm{mod}\,X = \left\{x_0 \middle| f(x_0) \ \max\right\}.$$

➢ Chương 2. Biến ngẫu nhiên

Chú ý

- Mode còn được gọi là số ***có khả năng nhất*** (xác suất cao nhất).
- Nếu phân phối xác suất của BNN *X* đối xứng (nghĩa là $p_i = p_{n-i}$ hoặc đồ thị hàm mật độ $f(x)$ đối xứng) và có 1 mode thì cả 3 số đặc trưng: Kỳ vọng, Median và Mode trùng nhau.
- Nếu phân phối xác suất của BNN *X* đối xứng hoặc gần đối xứng thì dùng Kỳ vọng để định vị là tốt nhất. Ngược lại thì dùng Median và Mode để định vị.

➢ Chương 2. Biến ngẫu nhiên

VD 12. Cho BNN *X* có bảng phân phối xác suất:

X	0	1	**2**	4	5	8
P	0,10	0,20	**0,30**	0,05	0,25	0,10

Khi đó ta có mod *X* = 2.

VD 13. Cho BNN *X* có bảng phân phối xác suất:

X	1	2	4	5	8
P	$1-3p$	0,18	0,07	0,25	*p*

A. mod$X$ = 5; B. Mod$X$ = 5; 8;
C. mod$X$ = 1; 8; D. mod$X$ = 1; 5; 8.

VD 14. Tuổi thọ (*X*: tháng) của một loài côn trùng là biến ngẫu nhiên có hàm mật độ:

$$f(x) = \begin{cases} \frac{3}{64}x^2(4-x), & x \in [0;\ 4] \\ 0, & x \notin [0;\ 4]. \end{cases}$$ Tìm mod*X*.

➢ Chương 2. Biến ngẫu nhiên

2.4. PHƯƠNG SAI

2.4.1. Định nghĩa

- Phương sai (*Variance* hoặc *Dispersion*) của biến ngẫu nhiên X, ký hiệu $VarX$ hoặc $D(X)$, là một số thực không âm được xác định bởi:

$$\boxed{VarX = E\left(X - EX\right)^2 = E(X^2) - \left(EX\right)^2.}$$

➢ Nếu $P(X = x_i) = p_i$ thì:

$$\boxed{VarX = \sum_i x_i^2.p_i - \left(\sum_i x_i.p_i\right)^2.}$$

➢ Nếu X có hàm mật độ $f(x)$ thì:

$$\boxed{VarX = \int_{-\infty}^{+\infty} x^2.f(x)dx - \left(\int_{-\infty}^{+\infty} x.f(x)dx\right)^2.}$$

➢ Chương 2. Biến ngẫu nhiên

VD 15. Tính *Var X*, biết:

X	1	2	3
P	0,2	0,7	0,1

Giải. $VarX = (1^2.0,2 + 2^2.0,7 + 3^2.0,1)$
$-(1.0,2 + 2.0,7 + 3.0,1)^2 = 0,29.$

VD 16. Tính *Var X*, biết: $f(x) = \begin{cases} \dfrac{3}{4}(x^2 + 2x), & x \in (0;\ 1) \\ 0, & x \notin (0;\ 1). \end{cases}$

VD 17. Cho biến ngẫu nhiên *X* có hàm mật độ xác suất:

$$f(x) = \begin{cases} \dfrac{3}{4}(1 - x^2), & |x| \le 1 \\ 0, & |x| > 1. \end{cases}$$ Tìm $VarY$, $Y = 2X^2$.

➢ Chương 2. Biến ngẫu nhiên

2.4.2. Ý nghĩa của Phương sai

- Do $X - EX$ là độ lệch giữa giá trị của *X* so với trung bình của nó nên phương sai là trung bình của bình phương độ lệch đó.
  Phương sai dùng để đo mức độ phân tán của *X* quanh kỳ vọng. Nghĩa là: phương sai nhỏ thì độ phân tán nhỏ nên độ tập trung lớn và ngược lại.
- Trong kỹ thuật, phương sai đặc trưng cho độ sai số của thiết bị. Trong kinh doanh, phương sai đặc trưng cho độ rủi ro đầu tư.
- Do đơn vị đo của $VarX$ bằng bình phương đơn vị đo của X nên để so sánh được với các đặc trưng khác, người ta đưa vào khái niệm ***độ lệch tiêu chuẩn*** (*standard deviation*) là: $\boxed{\sigma = \sqrt{VarX}.}$

➢ Chương 2. Biến ngẫu nhiên

VD 18. Năng suất của hai máy tương ứng là các BNN *X* và *Y* (đơn vị: sản phẩm/phút), bảng phân phối xác suất:

X	1	2	3	4
P	0,3	0,1	0,5	0,1

Y	2	3	4	5
P	0,1	0,4	0,4	0,1

Nếu phải chọn mua 1 trong 2 loại máy này thì ta nên chọn máy nào?

Giải. $EX = 2,4;\ VarX = 1,04;\ EY = 3,5;\ VarY = 0,65.$

Do $EX < EY$, $VarX > VarY$ nên ta chọn máy Y.

Chú ý. Trong trường hợp $EX < EY$ và $VarX < VarY$ thì ta không thể so sánh được. Để giải quyết vấn đề này, trong thực tế người ta dùng ***tỉ số tương đối*** $\dfrac{\sigma}{\mu}.100\%$ ($\mu = EX$) để so sánh sự ổn định của các BNN.
Tỉ số tương đối càng nhỏ thì độ ổn định càng cao.

➢ Chương 2. Biến ngẫu nhiên

VD 19. Điểm thi hết môn XSTK của lớp *X* và *Y* tương ứng là các BNN *X* và *Y*. Từ bảng kết quả điểm thi người ta tính được:

$EX = 6,25;\ VarX = 1,25;\ EY = 5,75;\ VarY = 0,75.$

Ta có: $\dfrac{\sigma_x}{\mu_x}.100\% = 17,89\%;\ \dfrac{\sigma_y}{\mu_y}.100\% = 15,06\%.$

Vậy lớp *Y* học đều (ổn định) hơn lớp *X*.

2.4.3. Tính chất của Phương sai

1) $VarC = 0,\ C \in \mathbb{R}$

2) $Var(CX) = C^2.VarX$

3) Nếu *X* và *Y* độc lập thì:
$Var(X \pm Y) = VarX + VarY.$

➢ Chương 2. Biến ngẫu nhiên

§3. MỘT SỐ LUẬT PHÂN PHỐI XÁC SUẤT THÔNG DỤNG

3.1. Phân phối xác suất của biến ngẫu nhiên rời rạc

3.1.1. Phân phối Siêu bội (*Hypergeometric distribution*)

a) Định nghĩa

- Xét tập có *N* phần tử, trong đó có N_A phần tử có tính chất *A*. Từ tập đó lấy ra **1 lần** *n* phần tử.
- Gọi *X* là số phần tử có tính chất *A* lẫn trong *n* phần tử được lấy ra thì *X* có phân phối ***Siêu bội*** với xác suất:

$$\boxed{p_k = P(X = k) = \frac{C_{N_A}^k C_{N-N_A}^{n-k}}{C_N^n}.}$$

$n-k$

$N-N_A$ k

Ký hiệu: $X \in H(N, N_A, n)$ hay $X \sim H(N, N_A, n)$.

➢ Chương 2. Biến ngẫu nhiên

Trong đó: $\max\{0;\ n - (N - N_A)\} \le k \le \min\{n;\ N_A\}.$

VD 1. Trong một cửa hàng bán 10 bóng đèn có 3 bóng hỏng. Một người chọn mua ngẫu nhiên 5 bóng đèn từ cửa hàng này. Gọi *X* là số bóng đèn tốt người đó mua được. Lập bảng phân phối xác suất của *X*.

b) Các số đặc trưng

$$\boxed{EX = np;\quad VarX = npq\frac{N-n}{N-1}.}$$

Trong đó $p = \dfrac{N_A}{N}$, $q = 1 - p$.

VD 2. Một rổ mận có 20 trái trong đó có 6 trái bị hư. Chọn ngẫu nhiên từ rổ đó ra 4 trái. Gọi X là số trái mận hư chọn phải.
1) Lập bảng phân phối xác suất của X.
2) Tính EX, $VarX$ bằng hai cách: dùng bảng phân phối và công thức.

2) ***Cách 1.*** $EX = \sum_{k=0}^{4} x_k.p_k = \sum_{k=0}^{4} k.\dfrac{C_6^k C_{14}^{4-k}}{C_{20}^4} = \dfrac{6}{5}$.

$$VarX = \sum_{k=0}^{4} x_k^2 p_k - (EX)^2 = \frac{204}{95} - \left(\frac{6}{5}\right)^2 = \frac{336}{475}.$$

3.1.2. Phân phối Nhị thức (*Binomial distribution*)

a) Công thức Bernoulli

- Dãy phép thử Bernoulli là dãy có *n* phép thử thỏa 3 điều kiện:
  1) Các phép thử của dãy độc lập với nhau.
  2) Trong mỗi phép thử ta chỉ quan tâm đến 1 biến cố *A*, nghĩa là chỉ có *A* và $\overline{A}$ xuất hiện.
  3) Xác suất xuất hiện *A* trong mọi phép thử của dãy luôn là hằng số *p*:
  $P(A) = p,\ P\left(\overline{A}\right) = 1 - p = q,\ (0 < p < 1).$
- Cho dãy *n* phép thử Bernoulli, xác suất xuất hiện *k* lần biến cố *A* là: $\boxed{p_k = C_n^k p^k q^{n-k},\ \ p = P(A).}$

b) Định nghĩa phân phối Nhị thức

- Phân phối Nhị thức là phân phối của biến ngẫu nhiên rời rạc $X = \{0;\ 1;\ 2;\ldots;\ n\}$ với xác suất tương ứng là:

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

Ký hiệu: $X \in B(n,\ p)$ hay $X \sim B(n,\ p)$.

c) Các số đặc trưng

$$\boxed{\begin{array}{c} EX = np;\ \ VarX = npq; \\ ModX = x_0:\ np - q \le x_0 \le np - q + 1. \end{array}}$$

VD 3. Một bà mẹ sinh 2 con (mỗi lần sinh 1 con) với xác suất sinh con trai là 0,51. Gọi *X* là số con trai trong 2 lần sinh. Lập bảng phân phối xác suất của *X*, EX, VarX.

$X = \{0, 1, 2\}$ $X \sim B(n, p)$ $n = 2$, $p = 0{,}51$

$P(X = k) = C_n^k p^k q^{n-k}$

VD 4. Một máy sản xuất lần lượt từng sản phẩm với xác suất có phế phẩm là 0,01.
1) Cho máy sản xuất ra 10 sản phẩm, tính xác suất có 2 phế phẩm.
2) Máy cần sản xuất ít nhất bao nhiêu sản phẩm để xác suất có ít nhất 1 phế phẩm lớn hơn 3%.

VD 5. Cho *X* có hàm mật độ $f(x) = \begin{cases} 4x^3, & x \in (0;\ 1) \\ 0, & x \notin (0;\ 1) \end{cases}.$

Tính xác suất để trong 3 phép thử độc lập có 2 lần X nhận giá trị trong khoảng $(0,25;\ 0,5)$.

VD 6. Một nhà vườn trồng 26 cây lan quý, với xác suất nở hoa của mỗi cây trong 1 năm là 0,67.
1) Giá 1 cây lan nở hoa là 1,2 triệu đồng. Giả sử nhà vườn bán hết những cây lan nở hoa thì mỗi năm nhà vườn thu được chắc chắn nhất là bao nhiêu tiền?
2) Nếu muốn trung bình mỗi năm có 37 cây lan quý nở hoa thì nhà vườn phải trồng tối thiểu mấy cây?

VD 7. Một nhà tuyển dụng kiểm tra kiến thức lần lượt *n* ứng viên, với xác suất được chọn của mỗi ứng viên 0,56. Biết xác suất để nhà tuyển dụng chọn đúng 8 ứng viên là 0,0843 thì số người phải kiểm tra là bao nhiêu?

A. 9 người; B. 10 người;
C. 12 người; D. 13 người.

VD 8*. 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 có đúng 1 lần chọn có không quá 1 phế phẩm.

3.1.3. Phân phối Poisson

a) Bài toán dẫn đến phân phối Poisson

- Giả sử các vụ tai nạn giao thông ở vùng A xảy ra 1 cách ngẫu nhiên, độc lập với nhau và trung bình 1 ngày có λ vụ tai nạn. Gọi X là số vụ tai nạn giao thông xảy ra trong 1 ngày ở vùng A.
- Chia 24 giờ trong ngày thành n khoảng thời gian sao cho ta có thể coi rằng trong mỗi khoảng thời gian có nhiều nhất 1 vụ tai nạn xảy ra, và khả năng xảy ra

➢ Chương 2. Biến ngẫu nhiên

tai nạn giao thông trong mỗi khoảng thời gian bằng $\frac{\lambda}{n}$. Khi đó, $X \in B\left(n, \frac{\lambda}{n}\right)$.

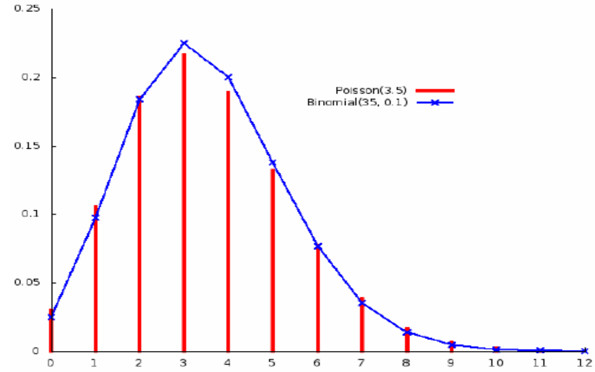
• Ta có: $P(X=k) = C_n^k \left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^{n-k}$

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

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

Suy ra: $P(X=k) \xrightarrow{n\to\infty} \frac{\lambda^k}{k!}.e^{-\lambda}$.

➢ Chương 2. Biến ngẫu nhiên



Phân phối P(3,5) và phân phối B(35; 0,1)

➢ Chương 2. Biến ngẫu nhiên

b) Định nghĩa

• Biến ngẫu nhiên X được gọi là có phân phối Poisson với tham số $\lambda > 0$ (λ là trung bình số lần xuất hiện biến cố nào đó mà ta quan tâm) nếu X nhận các giá trị 0, 1, 2,…, n,… với xác suất tương ứng:

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

Nhận xét

• Từ bài toán nêu ra ở trên, ta thấy phân phối Poisson không phải là phân phối xác suất chính xác vì số người là hữu hạn. Tuy vậy, phân phối Poisson là phân phối gần đúng rất thuận tiện cho việc mô tả và tính toán.

➢ Chương 2. Biến ngẫu nhiên

• Chẳng hạn, số xe qua 1 trạm hoặc số cuộc điện thoại tại 1 trạm công cộng trong 1 khoảng thời gian nào đó có phân phối Poisson.

c) Các số đặc trưng

$$\boxed{EX = VarX = \lambda;\ \text{mod}X = x_0: \lambda - 1 \le x_0 \le \lambda.}$$

VD 9. Quan sát tại siêu thị A trung bình 7 phút có 18 khách đến mua hàng.

1) Tính xác suất trong 2 phút có 3 khách đến siêu thị A ?

2) Tính số khách chắc chắn nhất sẽ đến siêu thị A trong 1 giờ ?

➢ Chương 2. Biến ngẫu nhiên

VD 10. Gia đình A có 5 ôtô cho thuê. Giả sử số ôtô được thuê trong 1 ngày của gia đình A là biến ngẫu nhiên X có phân phối Poisson $P(3)$. $X \sim P(3)$

1) Tính xác suất trong 1 ngày gia đình A có 4 ôtô được thuê ? $P(X=4) = \frac{e^{-3}3^4}{4!}$

2) Lập bảng phân phối xác suất của X ? $X = \{0, 1, \ldots, 5\}$

VD 11. Quan sát thấy trung bình 1 phút có 3 ôtô đi qua trạm thu phí. Biết xác suất có ít nhất 1 ôtô đi qua trạm thu phí trong t phút bằng 0,9. Giá trị của t là:

A. 0,9082 phút; B. 0,8591 phút;
C. 0,8514 phút; D. 0,7675 phút.

➢ Chương 2. Biến ngẫu nhiên

VD 12. Một trạm điện thoại trung bình nhận được 300 cuộc gọi trong 1 giờ.

1) Tính xác suất để trạm nhận được đúng 2 cuộc gọi trong 1 phút; đúng 5 cuộc gọi trong 3 phút.

2) Tính xác suất để 2 trong 3 phút liên tiếp, mỗi phút trạm nhận được nhiều nhất 1 cuộc gọi.

VD 13*. Quan sát thấy trung bình 1 ngày (24 giờ) có 12 chuyến tàu vào cảng A. Chọn ngẫu nhiên liên tiếp 6 giờ trong 1 ngày. Tính xác suất để 4 trong 6 giờ ấy, mỗi giờ có đúng 1 tàu vào cảng A.

> Chương 2. Biến ngẫu nhiên

3.2. Phân phối xác suất của biến ngẫu nhiên liên tục

3.2.1. Phân phối Chuẩn (*Normal distribution*)

a) Định nghĩa

• BNN X được gọi là có phân phối chuẩn với tham số μ và σ^2 $(\sigma > 0)$, ký hiệu $X \in N\left(\mu;\ \sigma^2\right)$, nếu hàm mật độ xác suất của X có dạng:

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

Các số đặc trưng

$$\boxed{\operatorname{mod} X = MedX = EX = \mu;\ VarX = \sigma^2.}$$

> Chương 2. Biến ngẫu nhiên

b) Phân phối Chuẩn đơn giản

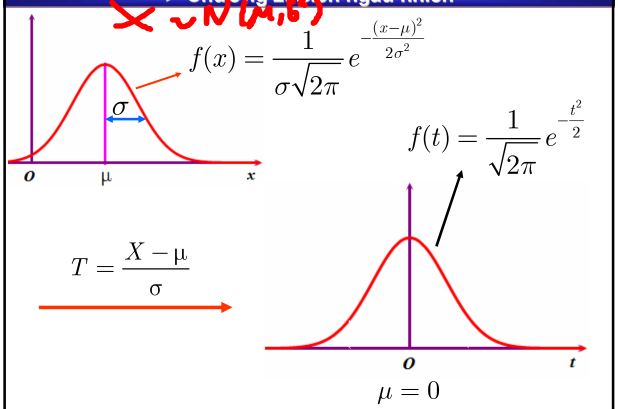
• Cho $X \in N\left(\mu;\ \sigma^2\right)$, đặt BNN $T = \dfrac{X-\mu}{\sigma}$ thì T có phân phối chuẩn đơn giản $\boxed{T \in N\left(0;\ 1\right).}$

• ***Hàm mật độ xác suất của biến ngẫu nhiên T***

$$\boxed{f(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}.}$$

(giá trị hàm Gauss $f(t)$ được cho trong bảng phụ lục A).

> Chương 2. Biến ngẫu nhiên



> Chương 2. Biến ngẫu nhiên

• ***Công thức tính xác suất của phân phối N(0; 1)***

$$\boxed{P(a \le T \le b) = \int_a^b f(t)dt = \varphi(b) - \varphi(a).}$$

Trong đó, hàm $\varphi(x) = \int_0^x f(t)dt\ (x \ge 0)$ được gọi là hàm Laplace (giá trị được cho trong bảng phụ lục B).

Chú ý

$$\boxed{P(T < x) = 0,5 + \varphi(x);\ P(T > x) = 0,5 - \varphi(x).}$$

> Chương 2. Biến ngẫu nhiên

Tính chất của hàm Laplace **(dùng để tra bảng)**

1) $\varphi(x)$ không giảm và $\varphi(-x) = -\varphi(x)$ (hàm số lẻ).
2) Nếu $x > 5$ thì $\varphi(x) \approx 0,5$.

c) Xác suất của phân phối Chuẩn tổng quát

• Cho $X \in N\left(\mu,\ \sigma^2\right)$. Để tính $P(a < X < b)$ ta đặt:

$$\alpha = \frac{a-\mu}{\sigma},\ \beta = \frac{b-\mu}{\sigma}$$

$$\Rightarrow P(a < X < b) = P(\alpha < T < \beta) = \varphi(\beta) - \varphi(\alpha).$$

• Sau đó, tra bảng phụ lục B ta được kết quả.

> Chương 2. Biến ngẫu nhiên

VD 14. Thời gian (X: phút) của một khách chờ được phục vụ tại một cửa hàng là BNN, $X \in N\left(4,5;\ 1,21\right)$.

1) Tính xác suất khách phải chờ để được phục vụ từ 3,5 phút đến 5 phút; không quá 6 phút.
2) Tính thời gian tối thiểu t nếu xác suất khách phải chờ vượt quá t là không quá 5%.

VD 15. Tốc độ chuyển dữ liệu từ máy chủ của ký túc xá đến máy tính của sinh viên vào buổi sáng chủ nhật có phân phối chuẩn với trung bình 60Kbits/s và độ lệch chuẩn 4Kbits/s. Xác suất để tốc độ chuyển dữ liệu lớn hơn 63Kbits/s là:

A. 0,2266; B. 0,2143; C. 0,1312; D. 0,1056.

VD 16. Trong một kỳ thi đầu vào ở trường chuyên *A* quy định điểm đỗ là tổng số điểm các môn thi không được thấp hơn 15 điểm. Giả sử tổng điểm các môn thi của học sinh là biến ngẫu nhiên có phân phối chuẩn với trung bình 12 điểm. Biết rằng tỉ lệ học sinh thi đỗ là 25,14%. Độ lệch chuẩn là:

A. 4 điểm; B. 4,5 điểm; C. 5 điểm; D. 5,5 điểm.

VD 17. Tuổi thọ (*X*: năm) của 1 loại bóng đèn *A* là biến ngẫu nhiên, $X \in N(4,2;\ 2,25)$. Khi bán 1 bóng đèn *A* thì lãi được 100 ngàn đồng nhưng nếu bóng đèn phải bảo hành thì lỗ 300 ngàn đồng. Vậy để có tiền lãi trung bình khi bán mỗi bóng đèn loại này là 30 ngàn đồng thì cần phải quy định thời gian bảo hành là bao nhiêu?

VD 18. Cho BNN *X* có phân phối chuẩn với *EX* = 10 và $P\left(10 < X < 20\right) = 0,3$. Tính $P\left(0 < X \le 15\right)$.

VD 19 (tham khảo). Một công ty cần mua 1 loại thiết bị có độ dày từ 0,118cm đến 0,122cm. Có 2 cửa hàng cùng bán loại thiết bị này với độ dày là các biến ngẫu nhiên có phân phối chuẩn $N(\mu, \sigma^2)$.

Giá bán của cửa hàng I là 300 USD/hộp/1000 cái và cửa hàng II là 260 USD/hộp/1000 cái.

Chỉ số độ dày trung bình μ (cm) và độ lệch chuẩn σ (cm) được cho trong bảng:

Cửa hàng	μ (cm)	σ (cm)
I	0,1200	0,0010
II	0,1200	0,0015

Hỏi công ty nên mua loại thiết bị này ở cửa hàng nào?

Giải. Gọi *X*, *Y* (cm) là độ dày thiết bị của cửa hàng I, II.

Ta có: $X \in N\left(0,12;\ 0,001^2\right)$ và $Y \in N\left(0,12;\ 0,0015^2\right)$.

Tỉ lệ sử dụng 1 thiết bị của mỗi cửa hàng là:

$P(0,118 \le X \le 0,122) = \varphi(2) - \varphi(-2) = 0,9544$,

$P(0,118 \le Y \le 0,122) = 2\varphi(1,33) = 0,8164$.

Giá trị thực của 1 thiết bị ở cửa hàng I là:

$$\frac{300}{0,9544.1000} = 0,3143 \text{ USD/cái.}$$

Giá trị thực của 1 thiết bị ở cửa hàng II là:

$$\frac{260}{0,8164.1000} = 0,3185 \text{ USD/cái.}$$

Vậy công ty nên mua thiết bị của cửa hàng I.

3.2.2. Phân phối $\chi^2(n)$ (tham khảo)

- Cho $X_i \in N(0;\ 1),\ i = \overline{1,\ n}$ và các X_i độc lập thì

$$X = \sum_{i=1}^{n} X_i^2 \in \chi^2(n) \text{ với hàm mật độ xác suất:}$$

$$f(x) = \begin{cases} 0, & x \le 0 \\ \dfrac{1}{2^{\frac{n}{2}}.\Gamma\left(\frac{n}{2}\right)}.e^{-\frac{x}{2}}x^{\frac{n}{2}-1}, & x > 0. \end{cases}$$

Trong đó: $\Gamma(n) = \int\limits_0^{+\infty} e^{-x}x^{n-1}dx,\ \Gamma(n+1) = n\Gamma(n)$,

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi},\ \Gamma(1) = 1.$$

3.2.3. Phân phối Student *T*(*n*) (với *n* bậc tự do)

- Cho $T \in N(0;\ 1)$ và $Y \in \chi^2(n)$ độc lập thì

$$X = \frac{T}{\sqrt{\frac{Y}{n}}} \in T(n) \text{ với hàm mật độ xác suất:}$$

$$f(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}.\Gamma\left(\frac{n}{2}\right)}\left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}},\ x \in \mathbb{R}.$$

Giá trị của *t*(*n*) được cho trong bảng *C*.

- Phân phối $T(n)$ do Willam.S.Gosset đưa ra năm 1908.

……………………………

§1. KHÁI NIỆM VECTOR NGẪU NHIÊN

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

VD. Một nhà máy sản xuất một loại sản phẩm, nếu xét đến kích thước của sản phẩm được đo bằng chiều dài X và chiều rộng Y thì ta có vector ngẫu nhiên hai chiều (X, Y), còn nếu xét thêm cả chiều cao Z nữa thì ta có vector ngẫu nhiên ba chiều (X, Y, Z).

- Trong khuôn khổ của chương trình ta chỉ xét vector ngẫu nhiên hai chiều, thường được ký hiệu là (X, Y).

§2. PHÂN PHỐI XÁC SUẤT CỦA VECTOR NGẪU NHIÊN RỜI RẠC

2.1 Bảng phân phối xác suất đồng thời

X \ Y	y_1	y_2	...	y_j	...	y_n	Tổng dòng
x_1	p_{11}	p_{12}	...	p_{1j}	...	p_{1n}	$p_{1\bullet}$
x_2	p_{21}	p_{22}	...	p_{2j}	...	p_{2n}	$p_{2\bullet}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_i	p_{i1}	p_{i2}	...	p_{ij}	...	p_{in}	$p_{i\bullet}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_m	p_{m1}	p_{m2}	...	p_{mj}	...	p_{mn}	$p_{m\bullet}$
Tổng cột	$p_{\bullet 1}$	$p_{\bullet 2}$	...	$p_{\bullet j}$	...	$p_{\bullet n}$	1

Trong đó $P\left(X = x_i;\, Y = y_j\right) = p_{ij}$ và $\sum_{i=1}^{m}\sum_{j=1}^{n} p_{ij} = 1$.

VD 1. BNN rời rạc X nhận các giá trị 6, 7 và 8. BNN Y nhận các giá trị 1, 2 và 3. Phân phối đồng thời của vector ngẫu nhiên (X,Y) cho bởi bảng:

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

Tính $P\left(X \ge 7,\, Y \ge 2\right)$ và $P\left(X = 6\right)$.

Giải

$P\left(X \ge 7,\, Y \ge 2\right) = 0,15 + 0,1 + 0,1 + 0,1 = 0,45$.

$P\left(X = 6\right) = 0,1 + 0,05 + 0,15 = 0,3$.

2.2. Bảng phân phối xác suất thành phần (lề) (biên)

Từ bảng phân phối xác suất đồng thời của (X,Y) ta có:

• **Bảng phân phối xác suất của *X***

X	x_1	x_2	...	x_m
P	$p_{1\bullet}$	$p_{2\bullet}$	...	$p_{m\bullet}$

Trong đó $p_{i\bullet} = p_{i1} + p_{i2} + \cdots + p_{in}$

(tổng dòng i của bảng phân phối xác suất đồng thời).

• **Bảng phân phối xác suất của *Y***

Y	y_1	y_2	...	y_n
P	$p_{\bullet 1}$	$p_{\bullet 2}$	...	$p_{\bullet n}$

Trong đó $p_{\bullet j} = p_{1j} + p_{2j} + \cdots + p_{mj}$

(tổng cột j của bảng phân phối xác suất đồng thời).

VD 2. Xác định phân phối thành phần của biến ngẫu nhiên X, Y trong VD 1.

Giải

• Bảng phân phối của X

X	6	7	8
P	0,3	0,3	0,4

• Bảng phân phối của Y

Y	1	2	3
P	0,35	0,30	0,35

2.3. Phân phối xác suất có điều kiện

• *Bảng phân phối xác suất của X với điều kiện $Y = y_j$:*

Từ công thức xác suất có điều kiện ta có xác suất

$$P\left(X=x_i \middle| Y=y_j\right) = \frac{P(X=x_i;\, Y=y_j)}{P(Y = y_j)} = \frac{p_{ij}}{p_{\bullet j}},\ i = \overline{1,m}.$$

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

X	x_1	x_2	...	x_m
$P\left(X=x_i \middle| Y=y_j\right)$	$\frac{p_{1j}}{p_{\bullet j}}$	$\frac{p_{2j}}{p_{\bullet j}}$	...	$\frac{p_{mj}}{p_{\bullet j}}$

• *Bảng phân phối xác suất của Y với điều kiện $X = x_i$:*

Y	y_1	y_2	...	y_n
$P\left(Y=y_j \middle| X=x_i\right)$	$\frac{p_{i1}}{p_{i\bullet}}$	$\frac{p_{i2}}{p_{i\bullet}}$	...	$\frac{p_{in}}{p_{i\bullet}}$

$$P\left(Y=y_j \middle| X=x_i\right) = \frac{P(X=x_i;\, Y=y_j)}{P(X = x_i)} = \frac{p_{ij}}{p_{i\bullet}},\ j = \overline{1,n}.$$

Chú ý

• Hai biến ngẫu nhiên X và Y được gọi là độc lập khi và chỉ khi $\forall x_i,\, y_j$ ta có:

$$P(X = x_i; Y = y_j) = P(X = x_i)P(Y = y_j).$$

VD 3. Xét bảng phân phối đồng thời của (X,Y)

X \ Y	1	2	3
6	0,10	0,05	0,15
7	0,05	0,15	0,10
8	0,20	0,10	0,10

Ta có:

$$P\left(X = 6 \,|\, Y = 2\right) = \frac{0,05}{0,05 + 0,15 + 0,1} = \frac{1}{6}.$$

➢ Chương 3. Vector ngẫu nhiên

$$P\left(X = 7 \mid Y = 2\right) = \frac{0,15}{0,05 + 0,15 + 0,1} = \frac{1}{2}.$$

$$P\left(X = 8 \mid Y = 2\right) = \frac{0,1}{0,05 + 0,15 + 0,1} = \frac{1}{3}.$$

Bảng phân phối xác suất của X với điều kiện $Y = 2$ là:

X	6	7	8
$P\left(X=x_i \mid Y=2\right)$	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{1}{3}$

Tương tự, bảng phân phối xác suất của Y với điều kiện $X = 8$ là:

Y	1	2	3
$P\left(Y=y_j \mid X=8\right)$	$0,50$	$0,25$	$0,25$

➢ Chương 3. Vector ngẫu nhiên

VD 4. Cho bảng phân phối xác suất đồng thời của vector ngẫu nhiên rời rạc (X, Y):

X \ Y	0	1	2
1	0,20	0,30	0,10
2	0,15	0,15	0,10

1) Lập bảng phân phối xác suất thành phần của X, Y.
2) Tính xác suất $P(X + Y = 2)$.
3) Lập bảng ppxs của Y với điều kiện $X = 2$.

Giải. 1) Bảng phân phối của X:

X	1	2
P	0,60	0,40

Bảng phân phối của Y:

Y	0	1	2
P	0,35	0,45	0,20

➢ Chương 3. Vector ngẫu nhiên

2) $P(X + Y = 2) = P(1;\, 1) + P(2;\, 0) = 0,45.$

3) Bảng phân phối xác suất của Y với điều kiện $X = 2$ là:

Y	0	1	2
$P(Y=y_j \mid X=2)$	$0,375$	$0,375$	$0,250$

VD 5. Cho vector ngẫu nhiên rời rạc (X, Y) có bảng phân phối đồng thời như sau:

(X,Y)	(0; 0)	(0; 1)	(1; 0)	(1; 1)	(2; 0)	(2; 1)
p_{ij}	$\frac{1}{18}$	$\frac{3}{18}$	$\frac{4}{18}$	$\frac{3}{18}$	$\frac{6}{18}$	$\frac{1}{18}$

1) Tính xác suất $P\left(X - Y = 1\right)$.
2) Lập bảng phân phối xác suất thành phần của X, Y.
3) Tính xác suất $P(X > 0 \mid Y = 1)$.
4) Lập bảng ppxs của Y với điều kiện $X = 1$.

➢ Chương 3. Vector ngẫu nhiên

Giải

1) $P\left(X - Y = 1\right) = P(1;\, 0) + P(2;\, 1) = \frac{4}{18} + \frac{1}{18} = \frac{5}{18}.$

2) Bảng phân phối thành phần của X và Y là:

X	0	1	2
P	$\frac{4}{18}$	$\frac{7}{18}$	$\frac{7}{18}$

Y	0	1
P	$\frac{11}{18}$	$\frac{7}{18}$

3) $P(X > 0 \mid Y = 1) = P(X = 1 \mid Y = 1)$

$+P(X = 2 \mid Y = 1) = \frac{4}{7}.$

4) Bảng phân phối xác suất của Y với điều kiện $X = 1$ là:

Y	0	1
$P(Y=y_j \mid X=1)$	$\frac{4}{7}$	$\frac{3}{7}$

➢ Chương 3. Vector ngẫu nhiên

VD 6. Bảng phân phối đồng thời của số lỗi vẽ màu X và số lỗi đúc Y của một loại sản phẩm nhựa ở một công ty cho bởi:

X \ Y	0	1	2
0	$0,48$	$0,10$	$0,06$
1	$0,06$	$0,05$	$0,05$
2	$0,02$	$0,04$	$0,01$
3	$0,02$	$0,01$	$0,10$

1) Nếu ta biết trên sản phẩm có 2 lỗi vẽ màu thì xác suất để không có lỗi đúc là bao nhiêu.
2) Nếu tổng số lỗi không vượt quá 2 và số lỗi đúc không vượt quá 1 thì hàng có thể bán ra thị trường. Tìm tỉ lệ các sản phẩm bán ra thị trường.

➢ Chương 3. Vector ngẫu nhiên

Giải. 1) $P(Y = 0 \mid X = 2) = \frac{P(X = 2;\, Y = 0)}{P(X = 2)} = \frac{0,02}{0,07} = \frac{2}{7}.$

2) Tỉ lệ các sản phẩm bán ra thị trường là:

$P(X + Y \le 2;\, Y \le 1) = P(X = 0;\, Y = 0)$

$+ P(X = 1;\, Y = 0) + P(X = 2;\, Y = 0)$

$+ P(X = 0; Y = 1) + P(X = 1; Y = 1) = 0,75.$

VD 7. Chi phí quảng cáo (X: triệu đồng) và doanh thu (Y: triệu đồng) của một cửa hàng có bảng phân phối đồng thời bên dưới. Nếu doanh thu là 700 triệu đồng thì chi phí quảng cáo trung bình là:

A. 60,5 triệu đồng; B. 48,3333 triệu đồng;
C. 51,6667 triệu đồng; D. 76,25 triệu đồng.

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	$0,05$	$0,05$	$0,35$

Giải. Bảng ppxs của X với điều kiện Y = 700 là:

X	30	50	80
$P\left(X=x_i \mid Y=700\right)$	$\frac{1}{6}$	$\frac{2}{3}$	$\frac{1}{6}$

$$\Rightarrow EX = 30.\frac{1}{6} + 50.\frac{2}{3} + 80.\frac{1}{6} = \frac{155}{3} \Rightarrow C.$$

§3. PHÂN PHỐI XÁC SUẤT CỦA VECTOR NGẪU NHIÊN LIÊN TỤC

3.1. Phân phối xác suất đồng thời

a) Định nghĩa

• Hàm hai biến $f(x,y) \ge 0$ xác định trên $\mathbb{R}^2$ được gọi là *hàm mật độ* của vector ngẫu nhiên (X,Y) nếu:

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)dxdy = 1.$$

• Với mọi tập $D \subset \mathbb{R}^2$ thì xác suất:

$$P\left[(X,Y) \in D\right] = \iint_D f(x,y)dxdy.$$

3.2. Phân phối xác suất thành phần

• Hàm mật độ của X:

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

• Hàm mật độ của Y:

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

Chú ý

• Hai biến ngẫu nhiên X và Y được gọi là độc lập khi và chỉ khi $f(x,y) = f_X(x)f_Y(y)$.

3.3. Phân phối xác suất có điều kiện

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

$$f_X(x \mid y) = \frac{f(x,y)}{f_Y(y)}.$$

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

$$f_Y(y \mid x) = \frac{f(x,y)}{f_X(x)}.$$

VD 1. Cho vector ngẫu nhiên (X,Y) có hàm mật độ đồng thời: $f(x,y) = \begin{cases} 10x^2y, \text{ khi } 0 < y < x < 1, \\ 0, \text{ nơi khác.} \end{cases}$

1) Tính xác suất $P\left(Y > \frac{1}{2}X\right)$.

2) Tìm hàm mật độ của X, Y.

3) Tìm hàm mật độ có điều kiện $f_X(x \mid y)$, $f_Y(y \mid x)$.

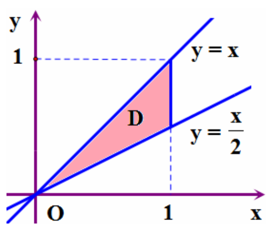
4) Tính xác xuất $P\left(Y < \frac{1}{8} \middle| X = \frac{1}{4}\right)$.

Giải

1) Đặt $D = \left\{(x,y) : y > \frac{x}{2}\right\}$.

Do hàm $f(x,y) = 10x^2y$ trên $0 < y < x < 1$ nên:

$$D = \left\{0 < x < 1,\ \frac{x}{2} < y < x\right\}$$

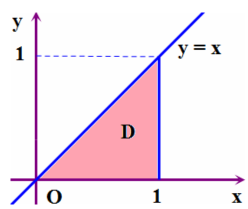


$$\Rightarrow P\left(Y > \frac{1}{2}X\right) = \iint_D f(x,y)dxdy$$

$$= \int_0^1 5x^2dx \int_{\frac{x}{2}}^{x} 2ydy = \frac{3}{4}.$$

2) • Hàm mật độ của X. Ta có:

$D = \left\{0 < x < 1,\ 0 < y < x\right\}$



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

$$= \int_0^x 10x^2ydy = 5x^4 \text{ với } 0 < x < 1.$$

$$\Rightarrow f_X(x) = \begin{cases} 5x^4, \text{ khi } 0 < x < 1, \\ 0, \text{ nơi khác.} \end{cases}$$

➢ Chương 3. Vector ngẫu nhiên

• Hàm mật độ của Y.

Ta có: $D = \{y < x < 1,\ 0 < y < 1\}$

$$\Rightarrow f_Y(y) = \int_{-\infty}^{+\infty} f(x,y)dx = \int_y^1 10x^2 y dx$$

$$= \frac{10}{3}y(1-y^3) \text{ với } 0 < y < 1.$$

$$\Rightarrow f_Y(y) = \begin{cases} \frac{10}{3}y(1-y^3), \text{ khi } 0 < y < 1, \\ 0, \text{ nơi khác.} \end{cases}$$

➢ Chương 3. Vector ngẫu nhiên

3)• Hàm mật độ $f_X(x \mid y) = \dfrac{f(x,y)}{f_Y(y)} = \dfrac{3x^2}{1-y^3}$

$$\Rightarrow f_X(x \mid y) = \begin{cases} \frac{3x^2}{1-y^3}, \text{ khi } 0 < y < x < 1, \\ 0, \text{ nơi khác.} \end{cases}$$

• Hàm mật độ $f_Y(y \mid x) = \dfrac{f(x,y)}{f_X(x)} = \dfrac{2y}{x^2}$

$$\Rightarrow f_Y(y \mid x) = \begin{cases} \frac{2y}{x^2}, \text{ khi } 0 < y < x < 1, \\ 0, \text{ nơi khác.} \end{cases}$$

➢ Chương 3. Vector ngẫu nhiên

4) Tính xác xuất $P\left(Y < \frac{1}{8} \middle| X = \frac{1}{4}\right)$. Theo câu 3) ta có:

$$f_Y\left(y \middle| x = \frac{1}{4}\right) = \begin{cases} 32y, \text{ khi } 0 < y < \frac{1}{4}, \\ 0, \text{ nơi khác.} \end{cases}$$

$$\Rightarrow P\left(Y < \frac{1}{8} \middle| X = \frac{1}{4}\right) = \int_{-\infty}^{\frac{1}{8}} f_Y\left(y \middle| x = \frac{1}{4}\right) dy$$

$$= \int_0^{\frac{1}{8}} 32y dy = \frac{1}{4}.$$

➢ Chương 3. Vector ngẫu nhiên

VD 2. Cho hàm mật độ đồng thời của X và Y

$$f(x,y) = \begin{cases} 6x, \text{ khi } 0 < x < 1;\ 0 < y < 1-x, \\ 0, \text{ nơi khác.} \end{cases}$$

1) Tìm hàm mật độ $f_X(x)$, $f_Y(y)$ của X, Y.

2) Tìm hàm mật độ có điều kiện $f_X\left(x \middle| y\right)$, $f_Y\left(y \middle| x\right)$.

3) Tính xác suất $P\left(X > 0,3 \middle| Y = 0,5\right)$.

Giải

Đặt $D : \begin{cases} 0 < x < 1, \\ 0 < y < 1-x. \end{cases} \equiv \begin{cases} 0 < y < 1, \\ 0 < x < 1-y. \end{cases}$

➢ Chương 3. Vector ngẫu nhiên

1) • $f_X(x) = \displaystyle\int_{-\infty}^{+\infty} f(x,y)dy = \int_0^{1-x} 6x dy$

$$= 6x(1-x),\ 0 < x < 1$$

$$\Rightarrow f_X(x) = \begin{cases} 6x(1-x), \text{ khi } 0 < x < 1, \\ 0, \text{ nơi khác.} \end{cases}$$

• $f_Y(y) = \displaystyle\int_{-\infty}^{+\infty} f(x,y)dx = \int_0^{1-y} 6x dx$

$$= 3(1-y)^2,\ 0 < y < 1$$

$$\Rightarrow f_Y(y) = \begin{cases} 3(1-y)^2, \text{ khi } 0 < y < 1, \\ 0, \text{ nơi khác.} \end{cases}$$

➢ Chương 3. Vector ngẫu nhiên

2) • $f_X\left(x \middle| y\right) = \dfrac{f(x,y)}{f_Y(y)} = \dfrac{2x}{(1-y)^2}$, $(x,y) \in D$

$$\Rightarrow f_X\left(x \middle| y\right) = \begin{cases} \frac{2x}{(1-y)^2}, \text{ khi } 0 < x < 1;\ 0 < y < 1-x, \\ 0, \text{ nơi khác.} \end{cases}$$

• $f_Y\left(y \middle| x\right) = \dfrac{f(x,y)}{f_X(x)} = \dfrac{1}{1-x}$, $(x,y) \in D$

$$\Rightarrow f_Y\left(y \middle| x\right) = \begin{cases} \frac{1}{1-x}, \text{ khi } 0 < x < 1;\ 0 < y < 1-x, \\ 0, \text{ nơi khác.} \end{cases}$$

➢ Chương 3. Vector ngẫu nhiên

3) Theo câu 2), $f_X\left(x \middle| y = 0,5\right) = \begin{cases} 8x, \text{ khi } 0 < x < 0,5 \\ 0, \text{ nơi khác.} \end{cases}$

$$\Rightarrow P\left(X > 0,3 \middle| Y = 0,5\right) = \int_{0,3}^{+\infty} f_X\left(x \middle| y = 0,5\right) dx$$

$$= \int_{0,3}^{0,5} 8x dx = 0,64.$$

VD 3. Tuổi thọ (X – năm) và thời gian chơi thể thao (Y – giờ) có hàm mật độ đồng thời được cho như sau:

$$f(x,y) = \begin{cases} \frac{15}{4} x(1-y^2), \text{ khi } 0 \le y < x \le 1, \\ 0, \text{ nơi khác.} \end{cases}$$

➢ Chương 3. Vector ngẫu nhiên

Thời gian chơi thể thao trung bình là:

A. 0,3125 giờ; B. 0,5214 giờ;

C. 0,1432 giờ; D. 0,4132 giờ.

Giải

$D = \left\{y < x \le 1,\ 0 \le y < 1\right\}$, ta có hàm mật độ của Y:

$$f_Y(y) = \frac{15}{4}\int_y^1 x(1-y^2)dx = \frac{15}{8}(1-y^2)^2,\ 0 \le y < 1$$

$$\Rightarrow E\left[f_Y(y)\right] = \int_{-\infty}^{+\infty} y.f_Y(y)dy$$

$$= \frac{15}{8}\int_0^1 (1-y^2)^2.ydy = 0,3125 \Rightarrow A.$$

➢ Chương 4. Định lý giới hạn trong xác suất

§1. Một số loại hội tụ trong xác suất và các định lý

§2. Các loại xấp xỉ phân phối xác suất

…………………………

§1. MỘT SỐ LOẠI HỘI TỤ TRONG XÁC SUẤT VÀ CÁC ĐỊNH LÝ

1.1. Hội tụ theo xác suất – Luật số lớn

a) Định nghĩa

• Dãy các biến ngẫu nhiên $\{X_i\}$ ($i = 1, 2,\ldots, n$) được gọi là *hội tụ theo xác suất* đến biến ngẫu nhiên X nếu:

$$\forall \omega \in \Omega, \forall \varepsilon > 0: \lim_{n\to\infty} P\left(\left|X_n(\omega) - X(\omega)\right| \ge \varepsilon\right) = 0.$$

Ký hiệu: $X_n \xrightarrow{P} X\ (n \to \infty)$.

➢ Chương 4. Định lý giới hạn trong xác suất

• Họ các biến ngẫu nhiên $\{X_i\}$ ($i = 1, 2,\ldots, n$) được gọi là tuân theo *luật số lớn* (dạng Tchébyshev) nếu:

$$\forall \varepsilon > 0: \lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^n X_i - \frac{1}{n}\sum_{i=1}^n EX_i\right| < \varepsilon\right) = 1$$

$$\Leftrightarrow \frac{1}{n}\sum_{i=1}^n \left(X_i - EX_i\right) \xrightarrow{P} 0.$$

b) Bất đẳng thức Tchébyshev

• Nếu biến ngẫu nhiên X có EX và $VarX$ hữu hạn thì:

$$\forall \varepsilon > 0: P\left(\left|X - EX\right| \ge \varepsilon\right) \le \frac{VarX}{\varepsilon^2}$$

hay

$$P\left(\left|X - EX\right| < \varepsilon\right) \ge 1 - \frac{VarX}{\varepsilon^2}.$$

➢ Chương 4. Định lý giới hạn trong xác suất

VD (tham khảo). Thu nhập trung bình hàng năm của dân cư 1 vùng là 700USD với độ lệch chuẩn 120USD. Hãy xác định một khoảng thu nhập hàng năm xung quanh giá trị trung bình của ít nhất 95% dân cư vùng đó.

Giải. Gọi X(USD) là thu nhập hàng năm của dân cư vùng đó. Ta có:

$$P\left(\left|X - EX\right| < \varepsilon\right) \ge 1 - \frac{VarX}{\varepsilon^2}$$

$$\Leftrightarrow P\left(\left|X - 700\right| < \varepsilon\right) \ge 1 - \frac{120^2}{\varepsilon^2} = 0,95$$

$$\Rightarrow \varepsilon = 536,656USD.$$

Vậy ít nhất 95% dân cư vùng đó có thu nhập hàng năm trong khoảng

$(EX - \varepsilon;\ EX + \varepsilon)$ = (163,344USD; 1236,656USD).

➢ Chương 4. Định lý giới hạn trong xác suất

c) Định lý luật số lớn Tchébyshev

Định lý

• Nếu họ các BNN $\{X_i\}$ ($i = 1, 2,\ldots, n$) độc lập từng đôi có EX_i hữu hạn và $VarX_i$ bị chặn trên bởi hằng số C thì:

$$\forall \varepsilon > 0: \lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^n X_i - \frac{1}{n}\sum_{i=1}^n EX_i\right| \ge \varepsilon\right) = 0.$$

Hệ quả

• Nếu họ các BNN $\{X_i\}$ ($i = 1, 2,\ldots, n$) độc lập từng đôi có $EX_i = \mu$ và $VarX_i = \sigma^2$ thì:

$$\frac{1}{n}\sum_{i=1}^n X_i \xrightarrow{P} \mu.$$

Ý nghĩa của định lý

- Thể hiện tính ổn định của trung bình số học các BNN độc lập cùng phân phối và có phương sai hữu hạn.
- Để đo 1 đại lượng vật lý nào đó ta đo *n* lần và lấy trung bình các kết quả làm giá trị thực của đại lượng cần đo.
- Áp dụng trong thống kê là dựa vào một mẫu khá nhỏ để kết luận tổng thể.

1.2. Hội tụ yếu – Định lý giới hạn trung tâm

a) Định nghĩa

- Dãy các biến ngẫu nhiên $\{X_i\}$ (i = 1, 2,…, *n*) được gọi là *hội tụ yếu* hay *hội tụ theo phân phối* đến BNN *X* nếu:

$$\lim_{n\to\infty} F_n(x) = F(x), \forall x \in C(F).$$

Trong đó, $C(F)$ là tập các điểm liên tục của $F(x)$.

Ký hiệu: $X_n \xrightarrow{d} X$ hay $F_n \xrightarrow{d} F$.

Chú ý. Nếu $X_n \xrightarrow{P} X$ thì $X_n \xrightarrow{d} X$.

b) Định lý Liapounop (giới hạn trung tâm)

Định lý

- Cho họ các BNN $\{X_i\}$ (i = 1, 2,…, *n*) độc lập từng đôi.

Đặt $Y = \sum_{i=1}^{n} X_i,\ \mu = \sum_{i=1}^{n} EX_i,\ \sigma^2 = \sum_{i=1}^{n} VarX_i.$

Nếu EX_i, $VarX_i$ hữu hạn và $\lim_{n\to\infty} \sum_{i=1}^{n} \dfrac{E\left|X_i - EX_i\right|^3}{\sigma^3} = 0$

thì $Y \in N\left(\mu,\ \sigma^2\right)$.

Ý nghĩa của định lý

- Sử dụng định lý giới hạn trung tâm Liapounop để tính xấp xỉ (gần đúng) xác suất.
- Xác định các phân phối xấp xỉ để giải quyết các vấn đề của lý thuyết ước lượng, kiểm định,…

§2. CÁC LOẠI XẤP XỈ PHÂN PHỐI XÁC SUẤT

2.1. Liên hệ giữa phân phối Siêu bội và Nhị thức

- Nếu *n* cố định, *N* tăng vô hạn và $\dfrac{N_A}{N} \to p\ (0 \neq p \neq 1)$

thì $\dfrac{C_{N_A}^k C_{N-N_A}^{n-k}}{C_N^n} \xrightarrow{d} C_n^k p^k q^{n-k}$.

Ứng dụng xấp xỉ phân phối Siêu bội bằng Nhị thức

- Cho $X \in H(N; N_A; n)$. Nếu *N* khá lớn và *n* rất nhỏ so với *N* (*n* < 5%.*N*) thì:

$$\boxed{X \sim B(n;p),\ p = \frac{N_A}{N}.}$$

VD 1. Một vườn lan có 10.000 cây sắp nở hoa, trong đó có 1.000 cây hoa màu đỏ.

1) Tính xác suất để khi chọn ngẫu nhiên 20 cây lan thì được 5 cây có hoa màu đỏ.
2) Tính xác suất để khi chọn ngẫu nhiên 50 cây lan thì được 10 cây có hoa màu đỏ.
3) Có thể tính xác suất để khi chọn ngẫu nhiên 200 cây lan thì có 50 cây hoa màu đỏ được không ?

Nhận xét

- Nếu dùng công thức của phân phối Siêu bội để giải câu

1) thì: $P(X=5) = \dfrac{C_{1.000}^5 C_{9.000}^{15}}{C_{10.000}^{20}} = 0,0318.$

2.2. Liên hệ giữa phân phối Nhị thức và Poisson

- Nếu $n \to +\infty,\ p \to 0,\ np \to \lambda$ thì:

$$C_n^k p^k q^{n-k} \xrightarrow{d} \frac{e^{-\lambda}.\lambda^k}{k!}.$$

Ứng dụng xấp xỉ phân phối Nhị thức bằng Poisson

- Cho *X* có phân phối nhị thức *B*(*n*, *p*), $\lambda = np$.

Khi đó:

➢ Nếu *n* lớn và *p* khá bé (gần bằng 0) thì:
$X \sim P(\lambda)$.

➢ Nếu *n* lớn và *p* cũng khá lớn ($p \approx 1$) thì:
$X \sim P(\lambda)$.

VD 2. 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:

1) Không quá 2 gói bị nhiễm khuẩn.
2) Đúng 40 gói bị nhiễm khuẩn.

Nhận xét

- Nếu dùng công thức của phân phối Nhị thức để giải câu 1) thì:

$$P(X \le 2) = 0,994^{1000} + 1000.0,006.(0,994)^{999} + C_{1000}^2(0,006)^2(0,994)^{998} = 0,0614.$$

VD 3. Giải câu 3) trong VD 1.

➢ Chương 4. Định lý giới hạn trong xác suất

Tóm tắt các loại xấp xỉ rời rạc

$X \in H(N,\ N_A,\ n) \xrightarrow[(n < 5\%N)]{p = \frac{N_A}{N}} X \in B(n,\ p)$

$P(X=k) = C_n^k p^k q^{n-k}$

n lớn

$\left(\frac{n}{N} < 0{,}05\right)$

$\lambda = n.\frac{N_A}{N}$

Sai số rất lớn

$\begin{bmatrix} p \approx 0 \\ p \approx 1 \end{bmatrix}$ $\lambda = np \approx npq$

$X \in P(\lambda)$

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

➢ Chương 4. Định lý giới hạn trong xác suất

2.3. Định lý giới hạn Moivre – Laplace

Định lý 1 (giới hạn địa phương)

• Gọi p_k là xác suất xuất hiện k lần biến cố A trong n phép thử Bernoulli với $P(A) = p$ (p không quá gần 0 và không quá gần 1) thì $\lim\limits_{n\to\infty} \frac{\sqrt{npq}.P_n(k)}{f(x_k)} = 1$.

Trong đó, $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, $x_k = \frac{k - np}{\sqrt{npq}}$ hữu hạn.

Định lý 2 (giới hạn Moivre – Laplace)

• Cho $X \in B(n,\ p)$ và $S_n = \frac{X - np}{\sqrt{npq}}$ thì:

$$S_n \xrightarrow{F} N(0,\ 1).$$

➢ Chương 4. Định lý giới hạn trong xác suất

Ứng dụng xấp xỉ Nhị thức bằng phân phối Chuẩn

• Cho $X \in B(n,\ p)$. Nếu n khá lớn, p không quá gần 0 và 1 thì $X \sim N(\mu;\ \sigma^2)$ với $\mu = np,\ \sigma^2 = npq$. Khi đó:

➢ $P(X = k) = \frac{1}{\sigma}.f\left(\frac{k-\mu}{\sigma}\right)$

(giá trị được cho trong bảng A với $f(-x) = f(x)$).

➢ $P(k_1 \le X \le k_2) = \varphi\left(\frac{k_2-\mu}{\sigma}\right) - \varphi\left(\frac{k_1-\mu}{\sigma}\right)$.

VD 4. Trong một đợt thi tuyển công chức ở thành phố A có 1000 người dự thi với tỉ lệ thi đạt là 80%. $= p$

Tính xác suất để:

1) có 172 người không đạt; $n = 1000$

$X \sim B(n, p)$

➢ Chương 4. Định lý giới hạn trong xác suất

2) có khoảng 170 đến 180 người không đạt.

VD 5. Một khách sạn nhận đặt chỗ của 325 khách hàng cho 300 phòng vào ngày 1/1 vì theo kinh nghiệm của những năm trước cho thấy có 10% khách đặt chỗ nhưng không đến. Biết mỗi khách đặt 1 phòng, tính xác suất:

1) Có 300 khách đến vào ngày 1/1 và nhận phòng.

2) Tất cả khách đến vào ngày 1/1 đều nhận được phòng.

VD 6. Một cửa hàng bán cá giống có 20.000 con cá loại da trơn trong đó để lẫn 4.000 con cá tra. Một khách hàng chọn ngẫu nhiên (1 lần) 1.000 con từ 20.000 con cá da trơn đó. Tính xác suất khách chọn được từ 182 đến 230 con cá tra ?

A. 0,8143; B. 0,9133; C. 0,9424; D. 0,9765.

➢ Chương 4. Định lý giới hạn trong xác suất

Tóm tắt xấp xỉ Chuẩn cho Nhị thức

$np > 5,\ nq > 5$

$(0 \not\approx p \not\approx 1)$

$X \in B(n,\ p)$		$X \in N(\mu,\ \sigma^2)$
$EX = np$	$\xleftrightarrow{\mu = np}$	$EX = \mu$
$VarX = npq$	$\xleftrightarrow{\sigma^2 = npq}$	$VarX = \sigma^2$

$\Rightarrow P(X = k) = \frac{1}{\sigma}f\left(\frac{k-\mu}{\sigma}\right)$, $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$

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

....

giam sai so

PHẦN II. LÝ THUYẾT THỐNG KÊ

(*Statistical theory*)

Chương 5. LÝ THUYẾT MẪU

§1. Khái niệm về phương pháp xác định mẫu
§2. Các đặc trưng của mẫu
§3. Phân phối xác suất của các đặc trưng mẫu
§4. Thực hành tính các đặc trưng mẫu cụ thể

....

§1. KHÁI NIỆM VỀ PHƯƠNG PHÁP XÁC ĐỊNH MẪU

1.1. Mẫu và tổng thể

• Tập hợp có các phần tử là các đối tượng mà ta nghiên cứu được gọi là *tổng thể*. Số phần tử của tổng thể được gọi là *kích thước* của tổng thể.

➢ Chương 5. Lý thuyết mẫu

• Từ tổng thể ta chọn ra n phần tử thì n phần tử đó được gọi là một *mẫu có kích thước* (*cỡ mẫu*) n.

• Mẫu được chọn ngẫu nhiên một cách khách quan được gọi là *mẫu ngẫu nhiên.*

• Khi mẫu có kích thước lớn thì ta không phân biệt mẫu có hoàn lại hay không hoàn lại.

VD 1. Khi nghiên cứu về số cá trong một hồ thì số cá trong hồ là kích thước của tổng thể. Từ hồ đó bắt lên 10 con cá thì được 1 *mẫu không hoàn lại* kích thước là 10. Nếu từ hồ đó bắt lên 1 con cá rồi thả xuống, sau đó tiếp tục bắt con khác, tiến hành 10 lần như thế ta được mẫu *có hoàn lại* kích thước 10.

➢ Chương 5. Lý thuyết mẫu

1.2. Phương pháp xác định mẫu

• *Mẫu định tính* là mẫu mà ta chỉ quan tâm đến các phần tử của nó có tính chất A nào đó hay không.

VD 2. Điều tra 100 hộ dân của một thành phố về thu nhập trong 1 năm. Nếu hộ có thu nhập dưới 10 triệu đồng/năm là hộ nghèo thì trong 100 hộ được điều tra ta quan tâm đến hộ nghèo (tính chất A). Mẫu điều tra này là mẫu định tính.

• *Mẫu định lượng* là mẫu mà ta quan tâm đến các yếu tố về lượng (như chiều dài, cân nặng,…) của các phần tử có trong mẫu.

VD 3. Cân 100 trái dưa gang được chọn ngẫu nhiên từ 1 cánh đồng ta được một mẫu định lượng.

➢ Chương 5. Lý thuyết mẫu

• Mẫu có kích thước n là tập hợp của n biến ngẫu nhiên độc lập $X_1, X_2,\ldots, X_n$ được lập từ biến ngẫu nhiên X và có cùng luật phân phối với X được gọi là *mẫu tổng quát*.

• Tiến hành quan sát (cân, đo,…) từng biến X_i và nhận được các giá trị cụ thể $X_i = x_i$, khi đó ta được *mẫu cụ thể* $x_1, x_2,\ldots, x_n$.

VD 4. Chiều cao của cây bạch đàn là biến ngẫu nhiên có phân phối chuẩn. Đo ngẫu nhiên 5 cây $X_1, X_2,\ldots, X_5$ ta được X_1=3,5m; X_2=3,2m; X_3=2,5m; X_4=4,1m; X_5=3m. Khi đó, $\{X_1, X_2,\ldots, X_5\}$ là mẫu tổng quát có phân phối chuẩn và {3,5m; 3,2m; 2,5m; 4,1m; 3m} là mẫu cụ thể.

Nhận xét

• Xác suất nghiên cứu về tổng thể để hiểu về mẫu còn thống kê thì ngược lại.

➢ Chương 5. Lý thuyết mẫu

➢ **Xét về lượng**

• Trung bình tổng thể là $\mu = EX$.

• Phương sai tổng thể $\sigma^2 = VarX$ là biểu thị cho mức độ biến động của biến X.

➢ **Xét về chất**

• Tổng thể được chia thành 2 loại phần tử: loại có tính chất A nào đó mà ta quan tâm và loại không có tính chất A.

• Gọi $X = 0$ nếu phần tử không có tính chất A và $X = 1$ nếu phần tử có tính chất A, p là tỉ lệ các phần tử có tính chất A thì:

$$X \in B(p),\ p = \frac{\text{Số phần tử có tính chất } A}{\text{Số phần tử của tổng thể}}.$$

➢ Chương 5. Lý thuyết mẫu

1.3. Sắp xếp số liệu thực nghiệm

1.3.1. Sắp xếp theo các giá trị khác nhau

• Giả sử mẫu $(X_1, X_2,\ldots, X_n)$ có k quan sát khác nhau là $X_1, X_2,\ldots, X_k$ $(k \le n)$ và X_i có tần số n_i (số lần lặp lại) với $n_1 + n_2 + \ldots + n_k = n$. Khi đó, số liệu được sắp xếp theo thứ tự tăng dần của X_i.

VD 5. Kiểm tra ngẫu nhiên 50 sinh viên, ta có kết quả:

X (điểm)	2	4	5	6	7	8	9	10
n (số SV)	4	6	20	10	5	2	2	1

1.3.2. Sắp xếp dưới dạng khoảng

• Giả sử mẫu $(X_1, X_2,\ldots, X_n)$ có nhiều quan sát khác nhau, khoảng cách giữa các quan sát không đồng đều hoặc các X_i khác nhau rất ít thì ta sắp xếp chúng dưới dạng khoảng.

➢ Chương 5. Lý thuyết mẫu

• Xét khoảng $\left(x_{\min},\ x_{\max}\right)$ chứa toàn bộ quan sát X_i.

Ta chia $\left(x_{\min},\ x_{\max}\right)$ thành các khoảng bằng nhau (còn gọi là lớp) theo nguyên tắc:

số khoảng tối ưu là $1 + 3,322\lg n$ và độ dài khoảng là

$$h = \frac{x_{\max} - x_{\min}}{1 + 3,322\lg n}.$$

VD 6. Đo chiều cao (X: cm) của $n = 100$ thanh niên, ta có bảng số liệu ở dạng khoảng:

X (cm)	148-152	152-156	156-160	160-164	164-168
n	5	20	35	25	15

Khi cần tính toán, ta sử dụng công thức $x_i = \dfrac{a_{i-1} + a_i}{2}$ để đưa số liệu trên về dạng bảng:

➢ Chương 5. Lý thuyết mẫu

X (cm)	150	154	158	162	166
n	5	20	35	25	15

Chú ý

• Đối với trường hợp số liệu được cho bởi cách liệt kê thì ta sắp xếp lại ở dạng bảng.

VD 7. Theo dõi mức nguyên liệu hao phí để sản xuất ra một đơn vị sản phẩm ở một nhà máy, ta thu được các số liệu sau (đơn vị: gam):

20; 22; 21; 20; 22; 22; 20; 19; 20; 22; 21;
19; 19; 20; 18; 19; 20; 20; 18; 19; 20; 20;
21; 20; 18; 19; 19; 21; 22; 21; 21; 20; 19.

Hãy sắp xếp số liệu trên dưới dạng bảng ?

X (gam)	18	19	20	21	22
n	3	8	11	6	5

➢ Chương 5. Lý thuyết mẫu

§2. CÁC ĐẶC TRƯNG MẪU (tham khảo)

2.1. Các đặc trưng mẫu

Giả sử tổng thể có trung bình $EX = \mu$, phương sai $VarX = \sigma^2$ và tỉ lệ các phần tử có tính chất *A* là *p*.

2.1.1. Tỉ lệ mẫu F_n

• Cho mẫu định tính kích thước *n*, ta gọi:

$$F_n = \frac{1}{n}\sum_{i=1}^{n} X_i,\ X_i = \begin{cases}0\\1\end{cases} \text{ là tỉ lệ mẫu tổng quát.}$$

• Cho mẫu định tính kích thước *n*, trong đó có *m* phần tử có tính chất *A*. Khi đó ta gọi:

$$f = f_n = \frac{m}{n} \text{ là tỉ lệ mẫu cụ thể.}$$

➢ Chương 5. Lý thuyết mẫu

Tính chất

1) Kỳ vọng của tỉ lệ mẫu bằng tỉ lệ tổng thể:

$$M(F_n) = M\left(\frac{X_1 + \ldots + X_n}{n}\right) = p.$$

2) Phương sai của tỉ lệ mẫu:

$$VarF_n = Var\left(\frac{X_1 + \ldots + X_n}{n}\right) = \frac{pq}{n}$$

(các X_i có phân phối Bernoulli).

2.1.2. Trung bình mẫu

• Trung bình mẫu: $\overline{X} = \overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

• Trung bình mẫu cụ thể: $\overline{x} = \overline{x_n} = \frac{1}{n}\sum_{i=1}^{n} x_i$.

➢ Chương 5. Lý thuyết mẫu

Tính chất

$$E\left(\overline{X}_n\right) = \mu = EX,\ Var\left(\overline{X}_n\right) = \frac{\sigma^2}{n} = \frac{VarX}{n}.$$

Chú ý

• Tỉ lệ mẫu $F_n = \frac{X_1 + \ldots + X_n}{n}$ và trung bình mẫu $\overline{X}_n = \frac{X_1 + \ldots + X_n}{n}$ khác nhau ở chỗ là trong F_n, các biến X_n chỉ có phân phối Bernoulli $B(p)$:

$$X_i = \begin{cases}0, & \text{nếu phần tử không có tính chất } A\\ 1, & \text{nếu phần tử có tính chất } A\end{cases}.$$

➢ Chương 5. Lý thuyết mẫu

2.1.3. Phương sai mẫu

• Phương sai mẫu: $\hat{S}^2 = \hat{S}_n^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X_n}\right)^2$.

Mẫu cụ thể: $\hat{s}^2 = \hat{s}_n^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}_n\right)^2$.

• Phương sai mẫu hiệu chỉnh:

$$S^2 = S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X_n}\right)^2.$$

Mẫu cụ thể: $s^2 = s_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \overline{x}_n\right)^2$.

Tính chất. $E\left(\hat{S}^2\right) = \frac{n-1}{n}\sigma^2,\ E\left(S^2\right) = \sigma^2$.

➢ Chương 5. Lý thuyết mẫu

• Trong tính toán ta sử dụng công thức:

$$s_n^2 = \frac{n}{n-1}\left[\overline{x_n^2} - \left(\overline{x}_n\right)^2\right],\ \overline{x_n^2} = \frac{1}{n}\sum_{i=1}^{n} x_i^2.$$

2.2. Liên hệ giữa đặc trưng của mẫu và tổng thể

• Các đặc trưng mẫu F_n, $\overline{X}_n$, S_n^2 là các thống kê dùng để nghiên cứu các đặc trưng p, μ, σ^2 tương ứng của tổng thể. Từ luật số lớn ta có:

$$F_n \to p,\ \overline{X}_n \to \mu,\ S_n^2 \to \sigma^2 \text{ (theo xác suất).}$$

• Trong thực hành, khi cỡ mẫu *n* khá lớn thì các đặc trưng mẫu xấp xỉ các đặc trưng tương ứng của tổng thể:

$$\overline{x} \approx \mu,\ f \approx p,\ \hat{s}^2 \approx \sigma^2,\ s^2 \approx \sigma^2.$$

§3. PHÂN PHỐI XÁC SUẤT CỦA CÁC ĐẶC TRƯNG MẪU (tham khảo)

3.1. Phân phối xác suất của tỉ lệ mẫu *F*

- $X \in B(p)$ và n khá lớn ($n \geq 100$) thì:

$$f = \frac{m}{n} \in N\left(p,\ \frac{pq}{n}\right) \Rightarrow T = \frac{f - p}{\sqrt{f(1-f)}}\sqrt{n} \in N(0,\ 1).$$

- $X_1 \in B(p_1),\ X_2 \in B(p_2)$ và $n_1,\ n_2$ khá lớn thì:

$$T = \frac{f_1 - f_2 - (p_1 - p_2)}{\sqrt{p_0(1-p_0)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \in N(0,\ 1).$$

Trong đó: $f_1 = \frac{m_1}{n_1},\ f_2 = \frac{m_2}{n_2},\ p_0 = \frac{m_1 + m_2}{n_1 + n_2}.$

3.2. Phân phối xác suất của trung bình mẫu

3.2.1. Trường hợp tổng thể *X* có phân phối chuẩn

- Do $E\overline{X} = \mu,\ Var\overline{X} = \frac{\sigma^2}{n}$ nên:

$$\overline{X} \in N\left(\mu,\ \frac{\sigma^2}{n}\right)\ hay\ \frac{\overline{X} - \mu}{\sigma}\sqrt{n} \in N\left(0,\ 1\right).$$

- Với mẫu cụ thể kích thước *n* đủ lớn, thì $\sigma^2 \approx S^2$ và:

$$\overline{X} \in N\left(\mu,\ \frac{S^2}{n}\right)\ hay\ \frac{\overline{X} - \mu}{S}\sqrt{n} \in N\left(0,\ 1\right).$$

- Khi $n < 30$ và σ^2 chưa biết thì $\frac{\overline{X} - \mu}{S}\sqrt{n} \in T(n-1)$ có phân phối Student với $n - 1$ bậc tự do.

3.2.2. Trường hợp *X* không có phân phối chuẩn

- Từ định lý giới hạn trung tâm, ta suy ra:

$$\frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \to T \in N\left(0,\ 1\right),\ \frac{\overline{X} - \mu}{\frac{S}{\sqrt{n}}} \to T \in N\left(0,\ 1\right).$$

- Với $n \geq 30$, ta có các phân phối xấp xỉ chuẩn như sau:

1) Nếu σ^2 đã biết thì:

$$\frac{\overline{X} - \mu}{\sigma}\sqrt{n} \sim N\left(0,\ 1\right),\ \overline{X} \sim N\left(\mu,\ \frac{\sigma^2}{n}\right).$$

2) Nếu σ^2 chưa biết thì:

$$\frac{\overline{X} - \mu}{S}\sqrt{n} \sim N\left(0,\ 1\right),\ \overline{X} \sim N\left(\mu,\ \frac{S^2}{n}\right).$$

3.3. Phân phối xác suất của phương sai mẫu

- Giả sử tổng thể $X \in N\left(\mu,\ \sigma^2\right)$, khi đó:

$$\frac{n}{\sigma^2}\hat{S}^2 = \frac{n-1}{\sigma^2}S^2 = \frac{1}{\sigma^2}\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2$$

sẽ có phân phối $\chi^2(n-1)$.

§4. THỰC HÀNH TÍNH CÁC ĐẶC TRƯNG CỦA MẪU CỤ THỂ

4.1. Tính tỉ lệ mẫu *f*

- Nếu trong mẫu có *m* phần tử có tính chất *A* mà ta quan tâm thì tỉ lệ mẫu là $f = \frac{m}{n}$.

4.2. Tính trung bình mẫu $\overline{x}$

- Nếu mẫu có *n* giá trị x_i thì trung bình mẫu là:

$$\overline{x} = \frac{x_1 + x_2 + ... + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

- Nếu x_i lặp lại n_i ($i = 1,\ldots, k \leq n$) lần thì trung bình mẫu là: $\overline{x} = \frac{1}{n}\sum_{i=1}^{k} x_i n_i.$

VD. Xét 10 kết quả quan sát:

102; 102; 202; 202; 202; 302; 302; 302; 302; 402.

Ta có: $\overline{x} = \frac{1}{10}(102.2 + 202.3 + 302.4 + 402.1).$

4.3. Tính phương sai mẫu $\hat{s}^2$

- Tính $\overline{x} = \frac{1}{n}(x_1 + x_2 + ... + x_n) = \frac{1}{n}\sum_{i=1}^{n} x_i.$

và $\overline{x^2} = \frac{1}{n}\left(x_1^2 + x_2^2 + ... + x_n^2\right) = \frac{1}{n}\sum_{i=1}^{n} x_i^2.$

- Phương sai mẫu là:

$$\hat{s}^2 = \overline{x^2} - \left(\overline{x}\right)^2.$$

- Phương sai mẫu có hiệu chỉnh là:

$$s^2 = \frac{n}{n-1}\hat{s}^2.$$

➢ Tính đặc trưng mẫu bằng máy tính bỏ túi

SỬ DỤNG MÁY TÍNH BỎ TÚI ĐỂ TÍNH CÁC ĐẶC TRƯNG CỦA MẪU

1. Số liệu đơn (không có tần số)

VD 1. Cho mẫu có cỡ mẫu là 5:

w = (12; 13; 11; 14; 11).

a) Máy *fx* 500 – 570 MS

• *Xóa bộ nhớ*: **SHIFT → MODE → 3 → = → =**

• *Vào chế độ thống kê nhập dữ liệu:*

– **MODE → 2** (chọn SD đối với *fx*500MS);
MODE → MODE → 1 (chọn SD đối với *fx*570MS).

– **Nhập các số:**

12 M+ 13 M+…. 11 M+

➢ Tính đặc trưng mẫu bằng máy tính bỏ túi

• *Xuất kết quả:*

– **SHIFT → 2 → 1 → =**
(xuất kết quả $\bar{x}$: trung bình mẫu).

– **SHIFT → 2 → 2 → =**
(xuất kết quả $\hat{s} = x\sigma n$: độ lệch chuẩn của mẫu).

– **SHIFT → 2 → 3 → =** (xuất kết quả $s = x\sigma n - 1$:
độ lệch chuẩn của mẫu có hiệu chỉnh).

b) Máy *fx* 500 – 570 ES

• *Xóa bộ nhớ*: **SHIFT → 9 → 3 → = → =**

• *Vào chế độ thống kê nhập dữ liệu:*

– **SHIFT → MODE →** dịch chuyển mũi tên tìm chọn mục **Stat → 2** (chế độ không tần số).

– **MODE → 3** (stat) **→ 1** (1-var) **→** (nhập các số):

12 = 13 =…. 11 = → AC

➢ Tính đặc trưng mẫu bằng máy tính bỏ túi

• *Xuất kết quả:*

– **SHIFT → 1 → 5** (var) **→ 1 → =** (n: cỡ mẫu)

– **SHIFT → 1 → 5** (var) **→ 2 → =** ($\bar{x}$:trung bình mẫu)

– **SHIFT → 1 → 5** (var) **→ 3 → =** ($x\sigma n$: độ lệch chuẩn của mẫu).

– **SHIFT → 1 → 5** (var) **→ 4 → =** ($x\sigma n - 1$: độ lệch chuẩn của mẫu có hiệu chỉnh).

2. Số liệu có tần số

VD 2. Cho mẫu như sau:

x_i	12	11	15
n_i	3	2	4

a) Máy *fx* 500 – 570 MS

• *Xóa bộ nhớ*: **SHIFT → MODE → 3 → = → =**

➢ Tính đặc trưng mẫu bằng máy tính bỏ túi

• *Vào chế độ thống kê nhập dữ liệu:*

– **MODE → 2** (chọn SD đối với *fx*500MS);
MODE → MODE → 1 (chọn SD đối với *fx*570MS).

– **Nhập các số:**

12 → SHIFT → , → 3 → M+
11 → SHIFT → , → 2 → M+
15 → SHIFT → , → 4 → M+

• *Xuất kết quả, làm như 1a).*

b) Máy *fx* 500 – 570 ES

• *Xóa nhớ vào chế độ thống kê nhập dữ liệu có tần số:*

– **SHIFT → MODE (SETUP)** dịch chuyển mũi tên **→ 4 → 1**

– **MODE → 3** (stat) **→ 1** (1-var)

➢ Tính đặc trưng mẫu bằng máy tính bỏ túi

– **Nhập các giá trị và tần số vào 2 cột trên màn hình:**

X	FREQ
12	3
11	2
15	4

→ AC

• *Xuất kết quả, làm như 1b).*

VD 3. Điều tra năng suất của 100 ha lúa trong vùng, ta có bảng số liệu sau:

Năng suất (tấn/ha)	3 - 3,5	3,5 - 4	4 - 4,5	4,5 - 5	5 - 5,5	5,5 - 6	6 - 6,5	6,5 - 7
Diện tích(ha)	7	12	18	27	20	8	5	3

➢ Tính đặc trưng mẫu bằng máy tính bỏ túi

Những thửa ruộng có năng suất ít hơn 4,4 tấn/ha là có năng suất thấp. Dùng máy tính bỏ túi để tính:

1) Tỉ lệ diện tích lúa có năng suất thấp.
2) Năng suất lúa trung bình, phương sai mẫu chưa hiệu chỉnh và độ lệch chuẩn của mẫu có hiệu chỉnh.

Giải. Bảng số liệu được viết lại:

Năng suất (tấn/ha)	3,25	3,75	4,25	4,75	5,25	5,75	6,25	6,75
Diện tích(ha)	7	12	18	27	20	8	5	3

1) $f = \dfrac{m}{n} = \dfrac{7+12+18}{100} = 37\%$.

2) $\bar{x} = 4,75$; $\hat{s}^2 = 0,685$; $s = 0,8318$.

➢ Chương 6. Ước lượng khoảng

§1. Ước lượng điểm
§2. Ước lượng khoảng
………………………..

§1. ƯỚC LƯỢNG ĐIỂM (tham khảo)

1.1. Thống kê

• Một hàm của mẫu tổng quát $T = T(X_1, X_2, \ldots, X_n)$ được gọi là 1 thống kê.

• Các vấn đề của thống kê toán được giải quyết chủ yếu nhờ vào việc xây dựng các hàm thống kê chỉ phụ thuộc vào mẫu tổng quát, không phụ thuộc các tham số.

1.2. Ước lượng điểm

• Ước lượng điểm của tham số θ (tỉ lệ, trung bình, phương sai,…) là thống kê $\hat{\theta} = \hat{\theta}(X_1, \ldots, X_n)$ chỉ phụ thuộc vào n quan sát $X_1, \ldots, X_n$, không phụ thuộc vào θ.

➢ Chương 6. Ước lượng khoảng

VD 1.

• Trung bình mẫu $\overline{X} = \dfrac{X_1 + X_2 + \ldots + X_n}{n}$ là ước lượng điểm của trung bình tổng thể μ.

• Tỉ lệ mẫu $F = \dfrac{X_1 + X_2 + \ldots + X_n}{n}$ là ước lượng điểm của tỉ lệ tổng thể p.

1.3. Ước lượng không chệch

• Thống kê $\hat{\theta}(X_1, \ldots, X_n)$ là ước lượng không chệch của θ nếu $E\left[\hat{\theta}(X_1, \ldots, X_n)\right] = \theta$.

VD 2.

• $E(\overline{X}) = \mu$ (trung bình mẫu là ước lượng không chệch của trung bình tổng thể μ).

➢ Chương 6. Ước lượng khoảng

• $EF = p$ (tỉ lệ mẫu là ước lượng không chệch của tỉ lệ tổng thể).

• $E(S^2) = \sigma^2$ (phương sai mẫu là ước lượng không chệch của phương sai tổng thể σ^2).

VD 3. Người ta cân 100 sản phẩm của 1 xí nghiệp A và có bảng số liệu:

X (gr)	498	502	506	510
n	40	20	20	20

Khi đó:

$$\overline{x} = \frac{498.40+502.20+506.20+510.20}{100} = 502,8(gr).$$

Dự đoán (ước lượng): Trọng lượng trung bình của các sản phẩm trong xí nghiệp là $\mu \approx 502,8(gr)$.

➢ Chương 6. Ước lượng khoảng

VD 4 (tham khảo). Từ mẫu tổng quát $W = (X_1, X_2)$ ta xét hai ước lượng của trung bình tổng thể μ sau:

$$\overline{X} = \frac{1}{2}X_1 + \frac{1}{2}X_2 \text{ và } \overline{X'} = \frac{1}{3}X_1 + \frac{2}{3}X_2.$$

1) Chứng tỏ $\overline{X}$ và $\overline{X'}$ là ước lượng không chệch của μ.
2) Ước lượng nào hiệu quả hơn?

Giải. 1) $E(\overline{X}) = E\left(\frac{1}{2}X_1 + \frac{1}{2}X_2\right)$

$$= \frac{1}{2}E(X_1) + \frac{1}{2}E(X_2) = \frac{1}{2}\mu + \frac{1}{2}\mu = \mu.$$

$$E(\overline{X'}) = E\left(\frac{1}{3}X_1 + \frac{2}{3}X_2\right) = \frac{1}{3}E(X_1) + \frac{2}{3}E(X_2)$$

$$= \frac{1}{3}\mu + \frac{2}{3}\mu = \mu \Rightarrow \text{(đpcm)}.$$

➢ Chương 6. Ước lượng khoảng

2) $Var(\overline{X}) = Var\left(\frac{1}{2}X_1 + \frac{1}{2}X_2\right)$

$$= \frac{1}{4}Var(X_1) + \frac{1}{4}Var(X_2) = \frac{\sigma^2}{4} + \frac{\sigma^2}{4} = \frac{\sigma^2}{2}.$$

$$Var(\overline{X'}) = Var\left(\frac{1}{3}X_1 + \frac{2}{3}X_2\right)$$

$$= \frac{1}{9}Var(X_1) + \frac{4}{9}Var(X_2) = \frac{\sigma^2}{9} + \frac{4\sigma^2}{9} = \frac{5\sigma^2}{9}$$

$$\Rightarrow Var(\overline{X}) < Var(\overline{X'}).$$

Vậy ước lượng $\overline{X}$ hiệu quả hơn.

➢ Chương 6. Ước lượng khoảng

§2. ƯỚC LƯỢNG KHOẢNG

2.1. Định nghĩa

• Khoảng $(\hat{\theta}_1;\ \hat{\theta}_2)$ của thống kê $\hat{\theta}$ được gọi là *khoảng tin cậy* của tham số θ nếu với xác suất $1-\alpha$ cho trước thì

$$P(\hat{\theta}_1 < \theta < \hat{\theta}_2) = 1 - \alpha.$$

• Xác suất $1-\alpha$ là độ tin cậy của ước lượng, $2\varepsilon = \hat{\theta}_2 - \hat{\theta}_1$ là độ dài của khoảng ước lượng và ε là độ chính xác của ước lượng. Khi đó: $\theta \in (\hat{\theta}_1;\ \hat{\theta}_2)$.

• Bài toán tìm khoảng tin cậy của θ là bài toán *ước lượng khoảng*.

➤ Chương 6. Ước lượng khoảng

2.2. Ước lượng khoảng cho trung bình tổng thể μ

• Giả sử tổng thể có trung bình μ chưa biết. Với độ tin cậy $1-\alpha$ cho trước, khoảng tin cậy cho μ là $(\mu_1;\ \mu_2)$ thỏa: $P(\mu_1 < \mu < \mu_2) = 1-\alpha$.

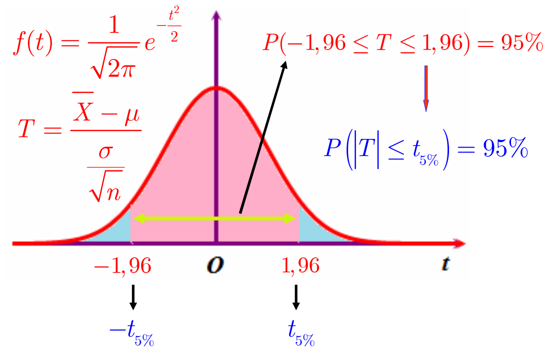
a) *Trường hợp 1.* Kích thước mẫu $n \geq 30$ và phương sai tổng thể σ^2 đã biết.

• Tính $\overline{x}$ (trung bình mẫu).

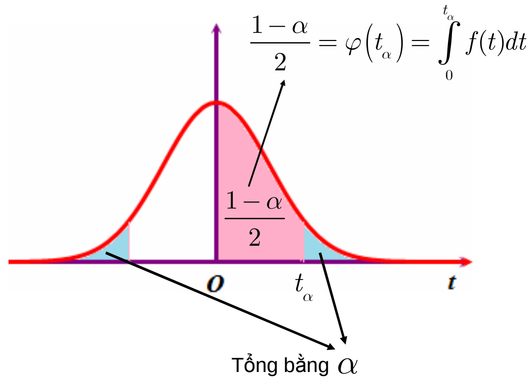
Từ $1-\alpha \Rightarrow \dfrac{1-\alpha}{2} = \varphi(t_\alpha) \xrightarrow{\text{tra bảng B}} t_\alpha$.

• Suy ra $\mu \in (\overline{x}-\varepsilon;\ \overline{x}+\varepsilon)$ với $\varepsilon = t_\alpha.\dfrac{\sigma}{\sqrt{n}}$.

➤ Chương 6. Ước lượng khoảng



➤ Chương 6. Ước lượng khoảng



➤ Chương 6. Ước lượng khoảng

VD 1. Điểm trung bình môn XSTK của sinh viên trường Đại học A là biến ngẫu nhiên có độ lệch chuẩn 0,26 điểm. Khảo sát ngẫu nhiên 100 sinh viên trường này thấy điểm trung bình môn XSTK là 5,12 điểm. Hãy ước lượng khoảng điểm trung bình môn XSTK của sinh viên trường A với độ tin cậy 98%?

b) *Trường hợp 2.* Kích thước mẫu $n \geq 30$ và phương sai tổng thể σ^2 chưa biết.

• Tính $\overline{x}$ và s (độ lệch chuẩn mẫu đã hiệu chỉnh).

• Từ $1-\alpha \Rightarrow \dfrac{1-\alpha}{2} = \varphi(t_\alpha) \xrightarrow{\text{tra bảng B}} t_\alpha$

$\Rightarrow \mu \in (\overline{x}-\varepsilon;\ \overline{x}+\varepsilon)$ với $\varepsilon = t_\alpha \dfrac{s}{\sqrt{n}}$.

➤ Chương 6. Ước lượng khoảng

Chú ý. Mối liên hệ giữa độ lệch chuẩn mẫu đã hiệu chỉnh s và chưa hiệu chỉnh $\hat{s}$ là:

$$\boxed{s^2 = \frac{n}{n-1}\hat{s}^2 \Rightarrow s = \sqrt{\frac{n}{n-1}\hat{s}^2}.}$$

VD 2. Đo đường kính của 100 trục máy do 1 nhà máy sản xuất thì được bảng số liệu:

Đường kính (cm)	9,75	9,80	9,85	9,90
Số trục máy	5	37	42	16

1) Hãy ước lượng khoảng trung bình đường kính của trục máy với độ tin cậy 97%?

2) Dựa vào mẫu trên để ước lượng khoảng trung bình đường kính của trục máy có độ chính xác 0,006cm thì đảm bảo độ tin cậy là bao nhiêu?

➤ Chương 6. Ước lượng khoảng

3) Dựa vào mẫu trên, nếu ước lượng khoảng trung bình đường kính của trục máy có độ chính xác lớn hơn 0,003cm với độ tin cậy 95% thì cần phải đo tối đa bao nhiêu trục máy?

c) *Trường hợp 3.* Kích thước mẫu $n < 30$, σ^2 đã biết và ***X có phân phối chuẩn*** thì ta làm như trường hợp 1.

d) *Trường hợp 4.* Kích thước mẫu $n < 30$, σ^2 chưa biết và ***X có phân phối chuẩn***.

• Tính $\overline{x}$, s.

• Từ $1-\alpha \Rightarrow \alpha \xrightarrow{\text{tra bảng C}} t_\alpha^{n-1}$

(nhớ giảm bậc thành $n-1$ rồi mới tra bảng!)

$\Rightarrow \mu \in (\overline{x}-\varepsilon;\ \overline{x}+\varepsilon)$ với $\varepsilon = t_\alpha^{n-1}.\dfrac{s}{\sqrt{n}}$.

VD 3. Giả sử chiều dài của 1 loại sản phẩm là biến ngẫu nhiên có phân phối chuẩn. Đo ngẫu nhiên 10 sản phẩm này thì được chiều dài trung bình 10,02*m* và độ lệch chuẩn của mẫu chưa hiệu chỉnh là 0,04*m*.
Tìm khoảng ước lượng trung bình chiều dài của loại sản phẩm này với độ tin cậy 95%?

VD 4. Năng suất lúa trong vùng *A* là biến ngẫu nhiên.
Gặt ngẫu nhiên 115 ha lúa của vùng này ta có số liệu:

Năng suất (tạ/ha)	40 – 42	42 – 44	44 – 46
Diện tích (ha)	7	13	25

Năng suất (tạ/ha)	46 – 48	48 – 50	50 – 52
Diện tích (ha)	35	30	5

1) Hãy tìm khoảng ước lượng trung bình cho năng suất lúa ở vùng *A* với độ tin cậy 95%?
2) Những thửa ruộng có năng suất lúa không vượt quá 44 tạ/ha ở vùng *A* là năng suất thấp (giả sử có phân phối chuẩn). Hãy ước lượng khoảng trung bình cho năng suất lúa của những thửa ruộng có năng suất thấp với độ tin cậy 99%?

Giải. 1) Số liệu được viết lại dưới dạng bảng:

Năng suất (tạ/ha)	41	43	45	47	49	51
Diện tích (ha)	7	13	25	35	30	5

VD 5. Để nghiên cứu nhu cầu về loại hàng *X* ở phường *A* người ta tiến hành khảo sát 400 trong toàn bộ 4000 gia đình. Kết quả khảo sát là:

Nhu cầu (kg/tháng)	0 – 1 (0,5)	1 – 2 (1,5)	2 – 3 (2,5)	3 – 4 (3,5)
Số gia đình	10	35	86	132

Nhu cầu (kg/tháng)	4 – 5 (4,5)	5 – 6 (5,5)	6 – 7 (6,5)	7 – 8 (7,5)
Số gia đình	78	31	18	10

1) Hãy ước lượng khoảng cho trung bình nhu cầu về loại hàng *X* của toàn bộ gia đình ở phường *A* trong 1 năm với độ tin cậy 95%?
2) Với mẫu khảo sát trên, nếu muốn có ước lượng khoảng trung bình nhu cầu về loại hàng *X* của phường *A* với độ chính xác nhỏ hơn 4,8 tấn/năm và độ tin cậy 99% thì cần khảo sát tối thiểu bao nhiêu gia đình trong phường *A*?

VD 6. Tiến hành khảo sát 500 trong tổng số 600.000 gia đình ở một thành phố thì thấy có 400 gia đình dùng loại sản phẩm *X* do công ty *A* sản xuất với bảng số liệu:

Số lượng (kg/tháng)	0,75	1,25	1,75	2,25	2,75	3,25
Số gia đình	40	70	110	90	60	30

Hãy ước lượng khoảng cho trung bình tổng khối lượng sản phẩm *X* do công ty *A* sản xuất được tiêu thụ ở thành phố này trong một tháng với độ tin cậy 95%?

A. (877,68 tấn; 982,32 tấn).
B. (1121,58 tấn; 1203,42 tấn).
C. (898,24 tấn; 993,21 tấn).
D. (1125,9 tấn; 1199,1 tấn).

2.3. Ước lượng khoảng cho tỉ lệ tổng thể *p*

- Giả sử tỉ lệ p các phần tử có tính chất *A* của tổng thể chưa biết. Với độ tin cậy $1-\alpha$ cho trước, khoảng tin cậy cho p là $(p_1;\ p_2)$ thỏa: $P(p_1 < p < p_2) = 1-\alpha$.
- Nếu biết tỉ lệ mẫu $f = f_n = \dfrac{m}{n}$ với *n* là cỡ mẫu, *m* là số phần tử ta quan tâm thì khoảng tin cậy cho p là:

$$\boxed{(f-\varepsilon;\ f+\varepsilon),\ \varepsilon = t_\alpha\sqrt{\frac{f(1-f)}{n}}.}$$

Trong đó t_α tìm được từ $\varphi(t_\alpha) = \dfrac{1-\alpha}{2}$ (tra bảng *B*).

VD 7. Một trường Đại học có 50.000 sinh viên. Điểm danh ngẫu nhiên 7000 sinh viên thấy có 765 sinh viên nghỉ học. Hãy ước lượng khoảng cho tỉ lệ sinh viên nghỉ học của trường với độ tin cậy 95%? Số sinh viên nghỉ học của trường trong khoảng nào?

VD 8. Để ước lượng số cá có trong một hồ người ta bắt lên 3000 con, đánh dấu rồi thả lại xuống hồ. Sau một thời gian, lại bắt lên 400 con cá thấy 60 con có đánh dấu. Với độ tin cậy 97%, hãy ước lượng khoảng cho tỉ lệ cá có đánh dấu và số cá có trong hồ?

VD 9. Lấy ngẫu nhiên 200 sản phẩm trong kho hàng *A* thấy có 21 phế phẩm.
1) Dựa vào mẫu trên, để ước lượng tỉ lệ phế phẩm trong kho *A* có độ chính xác là $\varepsilon = 0,035$ thì đảm bảo độ tin cậy của ước lượng là bao nhiêu?

➢ Chương 6. Ước lượng khoảng

2) Dựa vào mẫu trên, nếu muốn có độ chính xác của ước lượng tỉ lệ phế phẩm nhỏ hơn 0,01 với độ tin cậy 93% thì cần *kiểm tra thêm* ít nhất bao nhiêu sản phẩm nữa?

VD 10. Khảo sát năng suất (X: tấn/ha) của 100 ha lúa ở huyện A, ta có bảng số liệu:

X	3,25	3,75	4,25	4,75	5,25	5,75	6,25	6,75
S (ha)	7	12	18	27	20	8	5	3

Những thửa ruộng có năng suất lúa trên 5,5 tấn/ha là những thửa ruộng có năng suất cao. Sử dụng bảng khảo sát trên, để ước lượng tỉ lệ diện tích lúa có năng suất cao ở huyện A có độ chính xác là $\varepsilon = 8,54\%$ thì đảm bảo độ tin cậy là bao nhiêu?

A. 95%; B. 96%; C. 97%; D. 98%.

➢ Chương 7. Kiểm định Giả thuyết Thống kê

§1. Khái niệm về kiểm định giả thuyết thống kê
§2. Kiểm định giả thuyết về đặc trưng của tổng thể
§3. Kiểm định so sánh hai đặc trưng

……………………………

§1. KHÁI NIỆM VỀ KIỂM ĐỊNH GIẢ THUYẾT THỐNG KÊ

• Thông thường đối với tham số θ chưa biết của tổng thể ta có thể đưa ra nhiều giả thuyết về θ.
Vấn đề đặt ra là làm thế nào kiểm định được giả thuyết nào thích hợp với các số liệu của mẫu quan sát được.

1.1. Giả thuyết thống kê

• Giả thuyết thống kê (*Statistical Hypothesis*) là một giả sử hay một phát biểu có thể đúng, có thể sai liên quan đến tham số của một hay nhiều tổng thể.

➢ Chương 7. Kiểm định Giả thuyết Thống kê

1.2. Giả thuyết không (giả thuyết đơn) và giả thuyết ngược lại (đối thuyết)

• Giả thuyết không (*Null Hypothesis*) là sự giả sử mà ta muốn kiểm định, thường được ký hiệu là H_0.

• Giả thuyết ngược lại (*Alternative Hypothesis*) là việc bác bỏ *giả thuyết không* sẽ dẫn đến việc chấp nhận *giả thuyết ngược lại*. Giả thuyết ngược lại thường được ký hiệu là H_1.

Ta có các trường hợp sau:

Kiểm định giả thuyết $H_0: \theta = \theta_0$ với $H_1: \theta < \theta_0$.

Kiểm định giả thuyết $H_0: \theta = \theta_0$ với $H_1: \theta > \theta_0$.

Kiểm định giả thuyết $H_0: \theta = \theta_0$ với $H_1: \theta \neq \theta_0$.

➢ Chương 7. Kiểm định Giả thuyết Thống kê

1.3. Các loại sai lầm trong kiểm định

Khi kiểm định giả thuyết thống kê, ta có thể phạm phải 2 loại sai lầm sau

a) Sai lầm loại I (*type I error*)

• Là loại sai lầm mà ta phạm phải trong việc bác bỏ giả thuyết H_0 khi H_0 đúng. Xác suất của việc bác bỏ H_0 khi H_0 đúng là xác suất của sai lầm loại I và được ký hiệu là α. Số α còn được gọi là *mức ý nghĩa* (*level of significance*). Thông thường α = 0,05; 0,01; 0,001 …

b) Sai lầm loại II (*type II error*)

• Là loại sai lầm mà ta phạm phải trong việc chấp nhận giả thuyết H_0 khi H_0 sai. Xác suất của việc chấp nhận giả thuyết H_0 khi H_0 sai là xác suất của sai lầm loại II và được ký hiệu là β.

➢ Chương 7. Kiểm định Giả thuyết Thống kê

1.4. Miền bác bỏ và miền chấp nhận

• Tất cả các giá trị có thể có của các đại lượng thống kê trong kiểm định có thể chia làm 2 miền: *miền bác bỏ* và *miền chấp nhận*.

➢ Miền bác bỏ là miền chứa các giá trị làm cho giả thuyết H_0 bị bác bỏ.

➢ Miền chấp nhận là miền chứa các giá trị giúp cho giả thuyết H_0 không bị bác bỏ (được chấp nhận).

• Giá trị chia đôi hai miền được gọi là *giá trị giới hạn* (*critical value*).

1.5. Kiểm định một đầu và kiểm định 2 đầu

a) Kiểm định một đầu

• Khi đối thuyết H_1 có tính chất 1 phía thì việc kiểm định được gọi là *kiểm định 1 đầu*.

➢ Chương 7. Kiểm định Giả thuyết Thống kê

Có hai loại kiểm định 1 đầu:

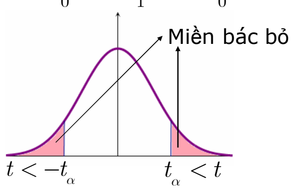
Kiểm định giả thuyết $H_0: \theta = \theta_0$ với $H_1: \theta < \theta_0$.

Kiểm định giả thuyết $H_0: \theta = \theta_0$ với $H_1: \theta > \theta_0$.

b) Kiểm định hai đầu

• Khi đối thuyết H_1 có tính chất 2 phía thì việc kiểm định được gọi là *kiểm định 2 đầu*:
Kiểm định giả thuyết $H_0: \theta = \theta_0$ với $H_1: \theta \neq \theta_0$.

• Từ đây về sau ta chỉ xét loại kiểm định hai đầu và để cho gọn ta chỉ đặt 1 giả thuyết là H.



§2. KIỂM ĐỊNH GIẢ THUYẾT VỀ ĐẶC TRƯNG CỦA TỔNG THỂ

2.1. Kiểm định giả thuyết trung bình tổng thể μ

Với trung bình μ_0 cho trước, tương tự bài toán ước lượng khoảng cho trung bình tổng thể, ta có 4 trường hợp sau (4 trường hợp đều đặt giả thuyết *H*: $\mu = \mu_0$).

a) *Trường hợp* 1. Với $n \geq 30$, σ^2 đã biết.

• Từ mức ý nghĩa $\alpha \Rightarrow \dfrac{1-\alpha}{2} = \varphi(t_\alpha) \xrightarrow{B} t_\alpha$.

• Tính giá trị thống kê $t = \dfrac{|\overline{x} - \mu_0|}{\dfrac{\sigma}{\sqrt{n}}}$.

• Nếu $t \leq t_\alpha$ ta chấp nhận *H*; nếu $t > t_\alpha$ ta bác bỏ *H*.

b) *Trường hợp* 2. Với $n \geq 30$, σ^2 chưa biết.

Ta làm như trường hợp 1 nhưng thay σ bằng s.

c) *Trường hợp* 3. Với $n < 30$, σ^2 đã biết và ***X* có phân phối chuẩn** (ta làm như trường hợp 1).

d) *Trường hợp* 4. Với $n < 30$, σ^2 chưa biết và ***X* có phân phối chuẩn.**

• Từ cỡ mẫu *n* và mức ý nghĩa $\alpha \xrightarrow{\text{tra bảng } C} t_\alpha^{n-1}$.

• Tính giá trị thống kê $t = \dfrac{|\overline{x} - \mu_0|}{\dfrac{s}{\sqrt{n}}}$.

• Nếu $t \leq t_\alpha^{n-1}$ ta chấp nhận *H*; $t > t_\alpha^{n-1}$ ta bác bỏ *H*.

Chú ý

• Trong tất cả các trường hợp bác bỏ, ta so sánh $\overline{x}$ và μ_0:

➢ Nếu $\overline{x} > \mu_0$ thì kết luận $\mu > \mu_0$.

➢ Nếu $\overline{x} < \mu_0$ thì kết luận $\mu < \mu_0$.

VD 1. Trong nhà máy bánh kẹo *A*, một máy tự động sản xuất ra các thanh chocolate với trọng lượng quy định là 250gram và độ lệch chuẩn là 5gram. Trong một ngày, bộ phận kiểm tra kỹ thuật chọn một mẫu ngẫu nhiên gồm 32 thanh chocolate và tính được trọng lượng trung bình của chúng là 248gram. Trong kiểm định giả thuyết *H*: "trọng lượng các thanh chocolate do máy tự động sản xuất ra đúng quy định" với mức ý nghĩa $\alpha = 0,05$. Hãy cho biết giá trị thống kê *t* và kết luận?

VD 2. Trọng lượng của loại sản phẩm *A* theo quy định là 6 kg. Kiểm tra ngẫu nhiên 121 sản phẩm *A* tính được trọng lượng trung bình là 5,795 kg và phương sai mẫu chưa hiệu chỉnh là 5,712 $(kg)^2$.
Với mức ý nghĩa 5%, hãy kiểm định giả thuyết *H*: "trọng lượng của sản phẩm *A* là 6 kg"?

VD 3. Trong một nhà máy gạo, trọng lượng đóng bao theo quy định của một bao gạo là 50 kg và độ lệch chuẩn là 0,3 kg. Cân thử 296 bao gạo của nhà máy này thì thấy trọng lượng trung bình là 49,97 kg. Kiểm định giả thuyết *H*: "trọng lượng mỗi bao gạo của nhà máy này là 50 kg" có giá trị thống kê *t* và kết luận là:

A. $t = 1,7205$; chấp nhận *H* với mức ý nghĩa 6%.

B. $t = 1,7205$; bác bỏ *H*, trọng lượng thực tế của bao gạo nhỏ hơn 50 kg với mức ý nghĩa 6%.

C. $t = 1,9732$; chấp nhận *H* với mức ý nghĩa 4%.

D. $t = 1,9732$; bác bỏ *H*, trọng lượng thực tế của bao gạo nhỏ hơn 50 kg với mức ý nghĩa 4%.

VD 4. Trọng lượng một loại gà ở trại chăn nuôi *A* khi xuất chuồng là 3,62 kg/con. Biết trọng lượng gà là biến ngẫu nhiên có phân phối chuẩn $N(\mu;\, 0,01)$. Sau một thời gian người ta cho gà ăn thức ăn mới và cân thử 15 con khi xuất chuồng thấy trọng lượng trung bình của gà là 3,69 kg/con. Với mức ý nghĩa 2%, hãy cho kết luận về loại thức ăn này?

VD 5. Điểm trung bình môn Toán của sinh viên năm trước là 5,72. Năm nay theo dõi 100 SV được số liệu:

Điểm	3	4	5	6	7	8	9
Số sinh viên	3	5	27	43	12	6	4

Trong kiểm định giả thuyết *H*: "điểm trung bình môn Toán của sinh viên năm nay bằng năm trước", mức ý nghĩa tối đa là bao nhiêu để *H* được chấp nhận?

A. $\alpha = 13,98\%$.

B. $\alpha = 13,62\%$.

C. $\alpha = 12,46\%$.

D. $\alpha = 11,84\%$.

VD 6. Chiều cao cây giống (X: m) trong một vườn ươm là biến ngẫu nhiên có phân phối chuẩn. Người ta đo ngẫu nhiên 25 cây giống này và có bảng số liệu:

X (m)	0,8	0,9	1,0	1,1	1,2	1,3
Số cây	1	2	9	7	4	2

Theo quy định của vườn ươm, khi nào cây cao hơn 1 m thì đem ra trồng. Với mức ý nghĩa 5%, kiểm định giả thuyết H: "cây giống của vườn ươm cao 1 m" có giá trị thống kê và kết luận là:

A. $t = 2,7984$, không nên đem cây ra trồng.
B. $t = 2,7984$, nên đem cây ra trồng.
C. $t = 1,9984$, không nên đem cây ra trồng.
D. $t = 1,9984$, nên đem cây ra trồng.

2.2. Kiểm định giả thuyết tỉ lệ tổng thể p

- Với tỉ lệ p_0 cho trước, ta đặt giả thuyết $H: p = p_0$.
- Từ mức ý nghĩa $\alpha \Rightarrow \dfrac{1-\alpha}{2} = \varphi(t_\alpha) \xrightarrow{B} t_\alpha$.
- Từ mẫu cụ thể, ta tính tỉ lệ mẫu $f = \dfrac{m}{n}$ và

giá trị thống kê $t = \dfrac{|f - p_0|}{\sqrt{\dfrac{p_0 q_0}{n}}}$.

➢ Nếu $t \le t_\alpha$ thì chấp nhận H, nghĩa là $p = p_0$.
➢ Nếu $t > t_\alpha$ thì bác bỏ H, nghĩa là $p \ne p_0$.
Khi đó: $f > p_0 \Rightarrow p > p_0$; $f < p_0 \Rightarrow p < p_0$.

VD 7. Kiểm tra ngẫu nhiên 800 sinh viên của trường A thấy có 128 sinh viên giỏi. Với mức ý nghĩa 5%, hãy kiểm định giả thuyết H: "tỉ lệ sinh viên giỏi của trường A là 20%"?

VD 8. Để kiểm tra một loại súng thể thao, người ta cho bắn 1000 viên đạn vào 1 tấm bia thấy có 670 viên trúng mục tiêu. Sau đó, bằng cải tiến kỹ thuật người ta nâng được tỉ lệ trúng của súng này lên 70%. Hãy cho kết luận về việc cải tiến trên với mức ý nghĩa 1%?

VD 9. Công ty A tuyên bố rằng có 40% người tiêu dùng ưa thích sản phẩm của mình. Một cuộc điều tra 400 người tiêu dùng thấy có 179 người ưa thích sản phẩm của công ty A. Trong kiểm định giả thuyết H: "có 40% người tiêu dùng thích sản phẩm của công ty A", mức ý nghĩa tối đa là bao nhiêu để H được chấp nhận?

§3. KIỂM ĐỊNH SO SÁNH HAI ĐẶC TRƯNG CỦA HAI TỔNG THỂ

3.1. So sánh hai trung bình μ_x và μ_y của X và Y

➢ **Tóm tắt 4 trường hợp**

- Tất cả 4 trường hợp đều đặt giả thuyết $H: \mu_x = \mu_y$.
- Việc chấp nhận hay bác bỏ giả thuyết H đều làm như bài toán kiểm định trung bình.

a) *Trường hợp 1.* $n_x,\ n_y \ge 30$ và $\sigma_x^2,\ \sigma_y^2$ đã biết.

Ta tính thống kê $t = \dfrac{|\overline{x} - \overline{y}|}{\sqrt{\dfrac{\sigma_x^2}{n_x} + \dfrac{\sigma_y^2}{n_y}}}$ và so sánh với t_α.

b) *Trường hợp 2.* $n_x,\ n_y \ge 30$ và $\sigma_x^2,\ \sigma_y^2$ chưa biết.

Ta thay $\sigma_x^2,\ \sigma_y^2$ bằng $s_x^2,\ s_y^2$ trong trường hợp 1.

c) *Trường hợp 3.* $n_x,\ n_y < 30$ và $\sigma_x^2,\ \sigma_y^2$ đã biết
đồng thời X, Y ***có phân phối chuẩn***.
Ta làm như trường hợp 1.

d) *Trường hợp 4.* $n_x,\ n_y < 30$ và $\sigma_x^2,\ \sigma_y^2$ chưa biết
đồng thời X, Y ***có phân phối chuẩn***.

- Tính phương sai mẫu chung của 2 mẫu:

$$s^2 = \frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2}.$$

- Tính giá trị thống kê $t = \dfrac{|\overline{x} - \overline{y}|}{s.\sqrt{\dfrac{1}{n_x} + \dfrac{1}{n_y}}}$.
- Từ $\alpha \xrightarrow{\text{tra bảng } C} t_\alpha^{n_x + n_y - 2}$ và so sánh với t.

VD 1. Người ta cân 100 trái cây A ở nông trường X và tính được $\overline{x} = 102gram$, $s_x^2 = 30$; cân 150 trái cây A ở nông trường Y và tính được $\overline{y} = 100gram$, $s_y^2 = 31$. Trong kiểm định giả thuyết H: "trọng lượng của trái cây ở 2 nông trường là như nhau", mức ý nghĩa tối đa là bao nhiêu để giả thuyết H được chấp nhận?

➢ Chương 7. Kiểm định Giả thuyết Thống kê

VD 2. Người ta đo ngẫu nhiên đường kính của 15 trục máy do máy X sản xuất và 17 trục máy do máy Y sản xuất (giả sử có phân phối chuẩn) tính được kết quả là:

$\overline{x} = 251,7mm$; $s_x^2 = 25$ và $\overline{y} = 249,8mm$; $s_y^2 = 23$.

Với mức ý nghĩa 1%, kiểm định giả thuyết *H*: "đường kính các trục máy do 2 máy sản xuất là như nhau" có giá trị thống kê và kết luận là:

A. $t = 2,0963$, chấp nhận *H*.
B. $t = 2,0963$, đường kính trục máy *X* lớn hơn.
C. $t = 1,0963$, chấp nhận *H*.
D. $t = 1,0963$, đường kính trục máy *X* lớn hơn.

➢ Chương 7. Kiểm định Giả thuyết Thống kê

VD 3. Trọng lượng trung bình của 23 trái dưa hấu do xã X trồng là 6,72kg/trái và $s_x = 0,32$kg. Trọng lượng trung bình của 19 trái dưa hấu do xã Y trồng là 6,46kg/trái và $s_y = 0,41$kg (giả sử trọng lượng dưa hấu có phân phối chuẩn). Với mức ý nghĩa 5%, có kết luận trọng lượng trái dưa hấu do xã X trồng nặng hơn dưa hấu do xã Y trồng được không?

3.2. So sánh hai tỉ lệ p_x, p_y của hai tổng thể *X, Y*

Ta thực hiện các bước sau:

- Đặt giả thuyết $H: p_x = p_y$.
- Từ 2 mẫu ta tính $f_x = \dfrac{m_x}{n_x}$, $f_y = \dfrac{m_y}{n_y}$, $p_0 = \dfrac{m_x + m_y}{n_x + n_y}$.

➢ Chương 7. Kiểm định Giả thuyết Thống kê

- Tính giá trị thống kê $t = \dfrac{|f_x - f_y|}{\sqrt{p_0 q_0 \left(\dfrac{1}{n_x} + \dfrac{1}{n_y}\right)}}$.
- **Kết luận**
  - ➢ Nếu $t \le t_\alpha$ thì chấp nhận $H \Rightarrow p_x = p_y$.
  - ➢ Nếu $t > t_\alpha$ và $f_x < f_y$ thì bác bỏ $H \Rightarrow p_x < p_y$;
  - ➢ Nếu $t > t_\alpha$ và $f_x > f_y$ thì bác bỏ $H \Rightarrow p_x > p_y$.

VD 4. Từ hai tổng thể *X* và *Y* người ta tiến hành kiểm tra 2 mẫu có kích thước $n_x = 1000$, $n_y = 1200$ về 1 tính chất *A* thì được $f_x = 0,27$ và $f_y = 0,3$. Với mức ý nghĩa 9%, hãy so sánh hai tỉ lệ p_x, p_y của hai tổng thể *X* và *Y*?

➢ Chương 7. Kiểm định Giả thuyết Thống kê

VD 5. Kiểm tra 120 sản phẩm ở kho I thấy có 6 phế phẩm; 200 sản phẩm ở kho II thấy có 24 phế phẩm. Hỏi chất lượng hàng ở hai kho có khác nhau không với:
1) Mức ý nghĩa 5%? 2) Mức ý nghĩa 1%?

VD 6. Một công ty điện tử tiến hành điều tra thị trường về sở thích xem tivi của cư dân trong 1 thành phố. Điều tra ngẫu nhiên 400 người ở quận *X* thì thấy có 270 người xem tivi ít nhất 1 giờ trong 1 ngày; 600 người ở quận *Y* có 450 người xem tivi ít nhất 1 giờ trong 1 ngày. Trong kiểm định giả thuyết *H*: "tỉ lệ cư dân xem tivi ít nhất 1 giờ trong 1 ngày ở quận *X* và *Y* như nhau", mức ý nghĩa tối đa là bao nhiêu để giả thuyết *H* được chấp nhận là:
A. 0,96%; B. 2,84%; C. 4,06%; D. 6,14%.

➢ Chương 7. Kiểm định Giả thuyết Thống kê

VD 7. Trước bầu cử, người ta thăm dò 1000 cử tri thì thấy có 400 người nói rằng sẽ bỏ phiếu cho ông *A*. Một tuần sau (vẫn chưa bầu cử), người ta tổ chức 1 cuộc thăm dò khác và thấy có 680 trong số 1500 cử tri được hỏi sẽ bỏ phiếu cho ông *A*. Kiểm định giả thuyết *H*: "tỉ lệ cử tri ủng hộ ông *A* ở hai lần là như nhau", với mức ý nghĩa 1% có giá trị thống kê *t* và kết luận là:

A. *t* = 2,6356; cử tri ngày càng ủng hộ ông *A*.
B. *t* = 2,6356; cử tri ủng hộ ông *A* không thay đổi.
C. *t* = 2,1349; cử tri ngày càng ủng hộ ông *A*.
D. *t* = 2,1349; cử tri ủng hộ ông *A* không thay đổi.

➢ Chương 8. Bài toán tương quan & Hồi quy

1. HỆ SỐ TƯƠNG QUAN MẪU

1.1. Định nghĩa

- *Hệ số tương quan mẫu* r là số *đo mức độ phụ thuộc tuyến tính* giữa hai mẫu ngẫu nhiên **cùng cỡ** X và Y.
- Giả sử ta có mẫu ngẫu nhiên cỡ n về vector ngẫu nhiên (X, Y) là (x_i, y_i); $i = 1; 2; ...; n$. Khi đó, hệ số tương quan mẫu r được tính theo công thức:

$$\boxed{r = \frac{\overline{xy} - \overline{x}.\overline{y}}{\hat{s}_x.\hat{s}_y};\ \overline{xy} = \frac{1}{n}\sum_{i=1}^{n} x_i y_i.}$$

➢ Chương 8. Bài toán tương quan & Hồi quy

1.2. Tính chất

1) $-1 \le r \le 1$.
2) Nếu $r = 0$ thì $X,\ Y$ không có quan hệ tuyến tính;
   Nếu $r = \pm 1$ thì $X,\ Y$ có quan hệ tuyến tính tuyệt đối.
3) Nếu $r < 0$ thì quan hệ giữa $X,\ Y$ là giảm biến.
4) Nếu $r > 0$ thì quan hệ giữa $X,\ Y$ là đồng biến.

VD 1. Kết quả đo lường độ cholesterol (*Y*) có trong máu của 10 đối tượng nam ở độ tuổi (*X*) như sau:

X	20	52	30	57	28	43	57	63	40	49
Y	1,9	4,0	2,6	4,5	2,9	3,8	4,1	4,6	3,2	4,0

Tính hệ số tương quan mẫu giữa X và Y.

➢ Chương 8. Bài toán tương quan & Hồi quy

Giải. Từ số liệu ở bảng trên, ta tính được:

$$\overline{xy} = \frac{20 \times 1,9 + ... + 49 \times 4,0}{10} = 167,26;$$

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 43,9;\ \hat{s}_x = 13,5385;$$

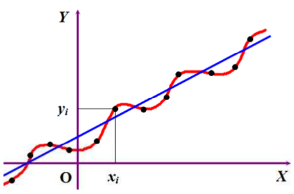
$$\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = 3,56;\ \hat{s}_y = 0,8333.$$

$$\text{Vậy } r = \frac{\overline{xy} - \overline{x}.\overline{y}}{\hat{s}_x.\hat{s}_y} = 0,9729.$$

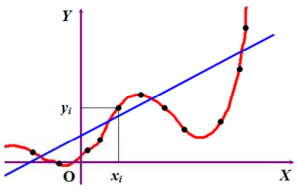
➢ Chương 8. Bài toán tương quan & Hồi quy

2. Đường hồi quy trung bình tuyến tính thực nghiệm

• Từ mẫu thực nghiệm về vector ngẫu nhiên $(X,\ Y)$, ta biểu diễn các cặp điểm $(x_i; y_i)$ lên mpOxy. Khi đó, đường cong nối các điểm là đường cong phụ thuộc của Y theo X mà ta cần tìm (xem hình *a*), *b*)).



Hình *a*



Hình *b*

➢ Chương 8. Bài toán tương quan & Hồi quy

• Đường thẳng là *đường hồi quy* thực nghiệm xấp xỉ tốt nhất các điểm mẫu đã cho, cũng là xấp xỉ đường cong cần tìm. Trong hình *a*) ta thấy xấp xỉ tốt (phụ thuộc tuyến tính chặt), hình *b*) xấp xỉ không tốt.

2.1. Phương pháp bình phương bé nhất

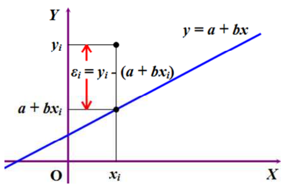
• Khi có sự phụ thuộc tuyến tính tương đối chặt giữa hai biến ngẫu nhiên X và Y ta cần tìm biểu thức $a + bX$ xấp xỉ Y tốt nhất theo nghĩa cực tiểu sai số bình phương trung bình $E(Y - a - bX)^2$, phương pháp này được gọi là *bình phương bé nhất*.

• Với mỗi cặp điểm (x_i, y_i) thì sai số xấp xỉ là:

$$\varepsilon_i = y_i - (a + bx_i) \text{ (xem hình } c\text{)).}$$

➢ Chương 8. Bài toán tương quan & Hồi quy

Ta đi tìm các ước lượng *a*, *b* sao cho $\sum_{i=1}^{n} \varepsilon_i^2$ đạt cực tiểu.



Hình *c*

Đặt $Q = \sum_{i=1}^{n} \varepsilon_i^2$

$$= \sum_{i=1}^{n} \left[y_i - (a + bx_i)\right]^2 \text{, ta có:}$$

$$\begin{cases} Q'_a = 0 \\ Q'_b = 0 \end{cases} \Leftrightarrow \begin{cases} na + b\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i & (1) \\ a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i & (2) \end{cases}$$

➢ Chương 8. Bài toán tương quan & Hồi quy

$$(1) \Leftrightarrow a = \frac{1}{n}\sum_{i=1}^{n} y_i - b.\frac{1}{n}\sum_{i=1}^{n} x_i = \overline{y} - b.\overline{x}.$$

Thay *a* vào (2), ta được:

$$\left(\overline{y} - b.\overline{x}\right)\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i$$

$$\Leftrightarrow b\left(\frac{1}{n}\sum_{i=1}^{n} x_i^2 - \overline{x}.\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \left(\frac{1}{n}\sum_{i=1}^{n} x_i y_i - \overline{y}.\frac{1}{n}\sum_{i=1}^{n} x_i\right)$$

$$\Leftrightarrow b\left(\overline{x^2} - \overline{x}^2\right) = \overline{xy} - \overline{x}.\overline{y} \Leftrightarrow b = \frac{\overline{xy} - \overline{x}.\overline{y}}{\hat{s}_x^2}.$$

➢ Chương 8. Bài toán tương quan & Hồi quy

- Vậy $b = \dfrac{\overline{xy} - \overline{x}.\overline{y}}{\hat{s}_x^2}$, $a = \overline{y} - b.\overline{x}$.

➢ Đường hồi quy tuyến tính của Y theo X là:

$$\boxed{y = a + bx.}$$

- Tương tự: $b = \dfrac{\overline{xy} - \overline{x}.\overline{y}}{\hat{s}_y^2}$, $a = \overline{x} - b.\overline{y}$.

➢ Đường hồi quy tuyến tính của X theo Y là:

$$\boxed{x = a + by.}$$

➢ Chương 8. Bài toán tương quan & Hồi quy

VD 2. Đo chiều cao (X: m) và khối lượng (Y: kg) của 5 học sinh nam, ta có kết quả:

X	1,45	1,60	1,50	1,65	1,55
Y	50	55	45	60	55

1) Tìm hệ số tương quan r.
2) Lập phương trình hồi quy tuyến tính của Y theo X.
3) Dự đoán nếu một học sinh cao 1,62m thì nặng khoảng bao nhiêu kg?

Giải. 1) $\overline{x} = 1,55$; $\hat{s}_x = 0,0707$; $\overline{y} = 53$; $\hat{s}_y = 5,099$;

$$\overline{xy} = 82,45 \Rightarrow r = \frac{82,45 - 1,55 \times 53}{0,0707 \times 5,099} = 0,8322.$$

➢ Chương 8. Bài toán tương quan & Hồi quy

2) $b = \dfrac{\overline{xy} - \overline{x}.\overline{y}}{\hat{s}_x^2} = \dfrac{82,45 - 1,55 \times 53}{(0,0707)^2} = 60,0181;$

$$a = \overline{y} - b\overline{x} = 53 - 60,0181 \times 1,55 = -40,0281.$$

Vậy $y = -40,0281 + 60,0181x$.

3) Học sinh cao 1,62m thì nặng khoảng:

$$y = -40,0281 + 60,0181 \times 1,62 = 57,2012 \text{ kg}.$$

VD 3. Số vốn đầu tư (X: triệu đồng) và lợi nhuận thu được (Y: triệu đồng) trong một đơn vị thời gian của 100 quan sát là:

X \ Y	0,3	0,7	1,0
1	20	10	
2		30	10
3		10	20

➢ Chương 8. Bài toán tương quan & Hồi quy

1) Lập phương trình hồi tuyến tính của X theo Y.
2) Dự đoán nếu muốn lợi nhuận thu được là 0,5 triệu đồng thì cần đầu tư bao nhiêu?

Giải. 1) Ta có $\overline{x} = 2$; $\hat{s}_x = 0,7746$; $\overline{y} = 0,71$;

$$\hat{s}_y = 0,2427;\ \overline{xy} = 1,56.$$

$$\Rightarrow b = \frac{\overline{xy} - \overline{x}.\overline{y}}{\hat{s}_y^2} = \frac{1,56 - 0,71 \times 2}{(0,2427)^2} = 2,3768;$$

$$a = \overline{x} - b\overline{y} = 2 - 2,3768 \times 0,71 = 0,3125.$$

Vậy $x = 0,3125 + 2,3768y$.

2) Nếu muốn lợi nhuận thu được là 0,5 triệu thì cần đầu tư khoảng:

$$x = 0,3125 + 2,3768 \times 0,5 = 1,5009 \text{ triệu đồng}.$$

➢ Chương 8. Bài toán tương quan & Hồi quy

VD 4. Số thùng bia (Y: thùng) được bán ra phụ thuộc vào giá bán (X: triệu đồng/ thùng). Điều tra 100 đại lý về 1 loại bia trong một đơn vị thời gian có bảng số liệu:

X \ Y	100	110	120
0,150	5	15	30
0,160	10	25	
0,165	15		

1) Tính hệ số tương quan r.
2) Lập phương trình hồi tuyến tính của X theo Y.
3) Dự đoán nếu muốn bán được 115 thùng bia thì giá bán mỗi thùng cỡ bao nhiêu?

➢ Chương 8. Bài toán tương quan & Hồi quy

Giải. 1) $\overline{x} = 0,1558$; $\hat{s}_x = 0,006$; $\overline{y} = 110$; $\hat{s}_y = 7,746$;

$$\overline{xy} = 17,1 \Rightarrow r = \frac{17,1 - 0,1558 \times 110}{0,006 \times 7,746} = -0,8176.$$

2) $b = \dfrac{\overline{xy} - \overline{x}.\overline{y}}{\hat{s}_y^2} = \dfrac{17,1 - 0,1558 \times 110}{(7,746)^2} = -0,0006;$

$$a = \overline{x} - b\overline{y} = 0,1558 + 0,0006 \times 110 = 0,2218.$$

Vậy $x = 0,2218 - 0,0006y$.

3) Nếu muốn bán được 115 thùng bia thì giá bán mỗi thùng khoảng:

$$x = 0,2218 - 0,0006 \times 115 = 0,1528 \text{ triệu đồng}.$$

➢ Sử dụng máy tính bỏ túi tìm đường hồi quy

3. Sử dụng máy tính bỏ túi tìm đường hồi qui

3.1. Số liệu không có tần số

a) Máy tính f_x500MS, f_x570MS

VD 5. 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	4,0	2,6	4,5	2,9	3,8	4,1	4,6	3,2	4,0

Tìm hệ số r, đường hồi quy Y theo X: $y = a + bx$.

Nhập số liệu:

MODE → **3 (REG)** → **1 (LIN)**

X, Y → **M+**

20, 1.9 → **M+**

… …

49, 4.0 → **M+**

➢ Sử dụng máy tính bỏ túi tìm đường hồi quy

Xuất kết quả:

SHIFT → 2 → (dịch chuyển mũi tên phải 2 lần)

→ **1** (***A*** chính là ***a*** trong phương trình)

→ **2** (***B*** chính là ***b*** trong phương trình)

→ **3** (***r*** chính là r).

Đáp số: $r = 0,9729$; $y = 0,9311 + 0,0599x$.

b) Máy tính f_x500ES, f_x570ES

Xét lại VD 5 ở trên.

Nhập số liệu:

SHIFT → **MODE** → dịch chuyển mũi tên tìm chọn mục **Stat** → **2** (chế độ không tần số)

MODE → **3** (stat) → **2** (*A*+*Bx*) → (nhập các giá trị của *X*, *Y* vào 2 cột)

➢ Sử dụng máy tính bỏ túi tìm đường hồi quy

X	Y
20	1.9
…	…
49	4.0

Xuất kết quả:

SHIFT →**1** → **7** → **1**(*A* chính là ***a*** trong phương trình)

→**2** (*B* chính là ***b*** trong phương trình)

→ **3** (r chính là r).

3.2. Số liệu có tần số

a) Máy tính f_x500MS, f_x570MS

VD 6. Xét bài toán cho ở dạng bảng (hình bên). Tìm hệ số r, đường hồi quy thực nghiệm Y theo X: $y = a + bx$.

Y \ X	21	23	25
3	2		
4	5	3	
5		11	8

➢ Sử dụng máy tính bỏ túi tìm đường hồi quy

Nhập số liệu:

MODE → **3 (REG)** → **1 (LIN)**

X, Y; n → **M+**

21, 3; 2 → **M+**

… …

25, 5; 8 → **M+**

Xuất kết quả:

SHIFT → 2 → (dịch chuyển mũi tên phải 2 lần)

→ **1** (***A*** chính là ***a*** trong phương trình)

→ **2** (***B*** chính là ***b*** trong phương trình)

→ **3** (***r*** chính là r).

Đáp số: $r = 0,7326$; $y = -2,6694 + 0,3145x$.

➢ Sử dụng máy tính bỏ túi tìm đường hồi quy

b) Máy tính f_x500ES, f_x570ES

Xét lại VD 6 ở trên

Nhập số liệu:

SHIFT → **MODE** → dịch chuyển mũi tên tìm chọn Mục **Stat** → **1** (chế độ có tần số)

MODE → **3** (stat) → **2** (*A*+*Bx*) → (nhập các giá trị của *X*, *Y*, tần số vào 3 cột)

X	Y	FREQ
21	3	2
...	…	…
25	5	8

➢ Sử dụng máy tính bỏ túi tìm đường hồi quy

Xuất kết quả:

SHIFT →**1** → **7** → **1** (kết quả là ***A***).

SHIFT →**1** → **7** → **2** (kết quả là ***B***).

SHIFT →**1** → **7** → **3** (kết quả là ***r***).

Chú ý

Sai số khi dùng máy tính bỏ túi là không tránh khỏi. Do đó, sinh viên nên chọn đáp án gần với kết quả của mình nhất khi làm bài trắc nghiệm.

………………..Hết……………….