

BÀI GIẢNG

XÁC SUẤT VÀ THỐNG KÊ ĐẠI HỌC

(Số đvhp: 2 – số tiết: 30)

Biên soạn: Đoàn Vương Nguyên

PHẦN I. LÝ THUYẾT XÁC SUẤT

(*Probability theory*)

Chương 1. Xác suất của Biến cố

Chương 2. Biến ngẫu nhiên

Chương 3. Phân phối Xác suất thông dụng – Vector ngẫu nhiên rời rạc

Chương 4. Định lý giới hạn trong Xác suất

PHẦN II. LÝ THUYẾT THỐNG KÊ

(*Statistical theory*)

Chương 5. Mẫu thống kê và Ước lượng tham số

Chương 6. Kiểm định Giả thuyết Thống kê

Tài liệu tham khảo

1. Nguyễn Phú Vinh – *Giáo trình Xác suất – Thống kê và Ứng dụng* – NXB Thống kê.
2. Đinh Ngọc Thanh – *Giáo trình Xác suất Thống kê* – ĐH Tôn Đức Thắng Tp.HCM.
3. Đặng Hùng Thắng – *Bài tập Xác suất; Thống kê* – NXB Giáo dục.
4. Lê Sĩ Đồng – *Xác suất – Thống kê và Ứng dụng* – NXB Giáo dục.
5. Đào Hữu Hồ – *Xác suất Thống kê* – NXB Khoa học & Kỹ thuật.
6. Đậu Thế Cấp – *Xác suất Thống kê – Lý thuyết và các bài tập* – NXB Giáo dục.
7. Phạm Xuân Kiêu – *Giáo trình Xác suất và Thống kê* – NXB Giáo dục.
8. Nguyễn Cao Văn – *Giáo trình Lý thuyết Xác suất & Thống kê* – NXB Ktế Quốc dân.
9. Nguyễn Đức Phương – *Xác suất & Thống kê* – Lưu hành nội bộ.
10. F.M.Dekking – *A modern introduction to Probability and Statistics* – Springer Publication (2005).

LÝ THUYẾT XÁC SUẤT

(*Probability theory*)

Chương 1. XÁC SUẤT CỦA BIẾN CỐ

Bài 1. Biến cố ngẫu nhiên

Bài 2. Xác suất của biến cố

Bài 3. Công thức tính xác suất

Bài 1. BIẾN CỐ NGẪU NHIÊN

1.1. Hiện tượng ngẫu nhiên

Người ta chia các hiện tượng xảy ra trong đời sống hàng ngày thành hai loại: *tất nhiên* và *ngẫu nhiên*.

- Những hiện tượng mà khi được thực hiện trong cùng một điều kiện sẽ cho ra kết quả như nhau được gọi là những hiện tượng *tất nhiên*. Chẳng hạn, đun nước ở điều kiện bình thường đến 100°C thì nước sẽ bốc hơi; một người nhảy ra khỏi máy bay đang bay thì người đó sẽ rơi xuống là *tất nhiên*.
- Những hiện tượng mà cho dù khi được thực hiện trong cùng 1 điều kiện vẫn có thể sẽ cho ra các kết quả khác nhau được gọi là những hiện tượng *ngẫu nhiên*. Chẳng hạn, gieo hạt lúa ở điều kiện bình thường

thì hạt lúa có thể nảy mầm cũng có thể không nảy mầm.

Hiện tượng ngẫu nhiên chính là đối tượng khảo sát của lý thuyết xác suất.

1.2. Phép thử và biến cố

- Để quan sát các hiện tượng ngẫu nhiên, người ta cho các hiện tượng này xuất hiện nhiều lần. Việc thực hiện một quan sát về một hiện tượng ngẫu nhiên nào đó, để xem hiện tượng này có xảy ra hay không được gọi là một *phép thử* (test).
- Khi thực hiện một phép thử, ta không thể dự đoán được kết quả xảy ra. Tuy nhiên, ta có thể liệt kê tất cả các kết quả có thể xảy ra.
 - Tập hợp tất cả các kết quả có thể xảy ra trong một phép thử được gọi là *không gian mẫu* của phép thử đó, ký hiệu là Ω .
 - Mỗi phần tử $\omega \in \Omega$ được gọi là một *biến cố sơ cấp*.
 - Mỗi tập $A \subset \Omega$ được gọi là một *biến cố*.

VD 1. Xét một sinh viên thi hết môn XSTK, thì hành động của sinh viên này là một phép thử.

- Tập hợp tất cả các điểm số:

$$\Omega = \{0; 0,5; 1; 1,5; \dots; 9,5; 10\}$$

mà sinh viên này có thể đạt là không gian mẫu.

- Các biến cố sơ cấp là các phần tử:

$$\omega_1 = 0 \in \Omega, \omega_2 = 0,5 \in \Omega, \dots, \omega_{21} = 10 \in \Omega.$$

- Các biến cố là các tập con của Ω :

$$A = \{4; 4,5; \dots; 10\}, B = \{0; 0,5; \dots; 3,5\}, \dots$$

- Các biến cố A, B có thể được phát biểu lại là:

- A : “sinh viên này thi đạt môn XSTK”;
- B : “sinh viên này thi hỏng môn XSTK”.

- Trong một phép thử, biến cố mà chắc chắn sẽ xảy ra được gọi là *biến cố chắc chắn*, ký hiệu là Ω . Biến cố không thể xảy ra được gọi là *biến cố rỗng*, ký hiệu là \emptyset .

VD 2. Từ nhóm có 6 nam và 4 nữ, ta chọn ngẫu nhiên ra 5 người.

- Biến cố “chọn được ít nhất 1 nam” là chắc chắn.
- Biến cố “chọn được 5 người nữ” là rỗng.

1.3. Quan hệ giữa các biến cố

1.3.1. Quan hệ tương đương

Trong 1 phép thử

- Biến cố A được gọi là *kéo theo* biến cố B nếu khi A xảy ra thì B xảy ra, ký hiệu là

$$A \subset B$$

- Hai biến cố A và B được gọi là *tương đương* với nhau nếu $A \subset B$ và $B \subset A$, ký hiệu là

$$A = B$$

VD 3. Cho trước 5 hộp trong đó 2 hộp có quà. Ông X mở lần lượt 3 hộp.

Gọi

A_i : “hộp được mở lần thứ i có quà” ($i = 1, 2, 3$);

B : “Ông X mở được hộp có quà”;

C : “Ông X mở được 2 hộp có quà”;

D : “Ông X mở được ít nhất 1 hộp có quà”.

Khi đó, ta có: $A_i \subset B$, $B \not\subset C$, $C \subset B$ và $B = D$.

1.3.2. Tổng và tích của hai biến cố

- Tổng của hai biến cố A và B là một biến cố, biến cố này xảy ra khi A xảy ra hay B xảy ra trong một phép thử (ít nhất một trong hai biến cố xảy ra), ký hiệu là

$$A \cup B \text{ hay } A + B$$

- Tích của hai biến cố A và B là một biến cố, biến cố này xảy ra khi cả A và B cùng xảy ra trong một phép thử, ký hiệu là

$$A \cap B \text{ hay } AB$$

VD 4. Một người thợ săn bắn 2 viên đạn vào một con thú và con thú sẽ chết nếu nó bị trúng cả 2 viên đạn. Gọi

A_i : “viên đạn thứ i trúng con thú” ($i = 1, 2$);

A : “con thú bị trúng đạn”;

B : “con thú bị chết”.

Khi đó, ta có: $A = A_1 \cup A_2$ và $B = A_1 \cap A_2$.

VD 5. Xét phép thử gieo hai hạt lúa. Gọi

N_i : “hạt lúa thứ i nảy mầm”;

K_i : “hạt lúa thứ i không nảy mầm” ($i = 1, 2$);

A : “có 1 hạt lúa nảy mầm”.

Khi đó, không gian mẫu của phép thử là $\Omega = \{K_1K_2; N_1K_2; K_1N_2; N_1N_2\}$.

Các biến cố tích sau đây là các biến cố sơ cấp:

$$\omega_1 = K_1K_2, \omega_2 = N_1K_2, \omega_3 = K_1N_2, \omega_4 = N_1N_2.$$

Biến cố A không phải là sơ cấp vì $A = N_1K_2 \cup K_1N_2$.

1.3.3. Biến cố đối lập

Trong 1 phép thử, biến cố \bar{A} được gọi là biến cố *đối lập* (hay biến cố bù) của biến cố A nếu và chỉ nếu khi A xảy ra thì \bar{A} không xảy ra và ngược lại, khi A không xảy ra thì \bar{A} xảy ra. Vậy ta có

$$\bar{A} = \Omega \setminus A$$

VD 6. Từ lô hàng chứa 12 chính phẩm và 6 phế phẩm, người ta chọn ngẫu nhiên ra 15 sản phẩm.

Gọi A_i : “chọn được i chính phẩm”, $i = 9; 10; 11; 12$.

Không gian mẫu là $\Omega = A_9 \cup A_{10} \cup A_{11} \cup A_{12}$.

Biến cố đối lập của A_{10} là $\bar{A}_{10} = \Omega \setminus A_{10} = A_9 \cup A_{11} \cup A_{12}$.

1.4. Hệ đầy đủ các biến cố

1.4.1. Hai biến cố xung khắc

Hai biến cố A và B được gọi là *xung khắc với nhau* trong 1 phép thử nếu A và B không cùng xảy ra.

VD 7. Hai sinh viên A và B cùng thi môn XSTK.

Gọi A : “sinh viên A thi đỗ”;

B : “chỉ có sinh viên B thi đỗ”;

C : “chỉ có 1 sinh viên thi đỗ”.

Khi đó, A và B là xung khắc; B và C không xung khắc.

Chú ý. A và B xung khắc nhưng không đối lập.

1.4.2. Hệ đầy đủ các biến cố

Trong một phép thử, họ gồm n biến cố $\{A_i\}$, $i = \overline{1, n}$ được gọi là hệ đầy đủ khi và chỉ khi có duy nhất biến cố A_{i_0} , $i_0 \in \{1; 2; \dots; n\}$ của họ xảy ra. Nghĩa là:

- 1) $A_i \cap A_j = \emptyset, \forall i \neq j$;
- 2) $A_1 \cup A_2 \cup \dots \cup A_n = \Omega$.

VD 8. 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, hệ $\{A; \bar{A}\}$ là đầy đủ với biến cố A tùy ý.

BÀI 2. XÁC SUẤT CỦA BIẾN CỐ

2.1. Khái niệm xác suất

Quan sát các biến cố đối với một phép thử, mặc dù không thể khẳng định một biến cố có xảy ra hay không nhưng người ta có thể *phỏng đoán khả năng xảy ra* của các biến cố này là ít hay nhiều.

Khả năng xảy ra khách quan của một biến cố được gọi là *xác suất* (probability) của biến cố đó.

Xác suất của biến cố A , ký hiệu là $P(A)$, có thể được định nghĩa bằng nhiều dạng sau:

- dạng cổ điển;
- dạng thống kê;
- dạng tiên đề Kolmogorov;
- dạng hình học.

2.2. Định nghĩa xác suất dạng cổ điển

Xét một phép thử với không gian mẫu $\Omega = \{\omega_1; \dots; \omega_n\}$ và biến cố $A \subset \Omega$ có k phần tử. Nếu n biến cố sơ cấp có cùng khả năng xảy ra (đồng khả năng) thì xác suất của biến cố A được định nghĩa

$$P(A) = \frac{\text{Số trường hợp } A \text{ xảy ra}}{\text{Số trường hợp có thể xảy ra}} = \frac{k}{n}$$

VD 1. Một công ty cần tuyển 2 nhân viên. Có 4 người nữ và 2 người nam nộp đơn ngẫu nhiên (khả năng trúng tuyển là như nhau). Tính xác suất để:

- 1) cả hai người trúng tuyển đều là nữ;
- 2) có ít nhất một người nữ trúng tuyển.

Giải. Gọi A : “cả hai người trúng tuyển đều là nữ”;
 B : “có ít nhất một người nữ trúng tuyển”.

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

VD 2. Từ 1 hộp chứa 86 sản phẩm tốt và 14 phế phẩm người ta chọn ngẫu nhiên ra 25 sản phẩm.

Tính xác suất chọn được:

1) cả 25 sản phẩm đều tốt;

2) đúng 20 sản phẩm tốt.

Giải. Gọi A : “chọn được 25 sản phẩm tốt”, B : “chọn được đúng 20 sản phẩm tốt”.

VD 3. Trong một vùng dân cư, tỉ lệ người mắc bệnh tim là 9%; mắc bệnh huyết áp là 12%; mắc cả bệnh tim và huyết áp là 7%. Chọn ngẫu nhiên 1 người trong vùng đó. Tính xác suất để người này không mắc bệnh tim và không mắc bệnh huyết áp?

Giải. Gọi A : “người được chọn không mắc cả hai bệnh trên”.

2.3. Định nghĩa xác suất dạng thống kê

Nếu khi thực hiện một phép thử nào đó n lần (đủ lớn), ta thấy có k lần biến cố A xuất hiện thì xác suất của biến cố A theo nghĩa thống kê là

$$P(A) \approx \frac{k}{n}$$

VD 4

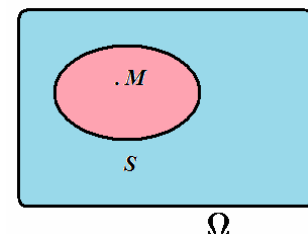
- Pearson đã gieo một đồng tiền cân đối, đồng chất 12.000 lần thấy có 6.019 lần xuất hiện mặt sấp (tần suất là 0,5016); gieo 24.000 lần thấy có 12.012 lần xuất hiện mặt sấp (tần suất là 0,5005).
- Laplace đã nghiên cứu tỉ lệ sinh trai – gái ở London, Petecbua và Berlin trong 10 năm và đưa ra tần suất sinh bé gái là 21/43.
- Cramer đã nghiên cứu tỉ lệ sinh trai – gái ở Thụy Điển trong năm 1935 và kết quả có 42.591 bé gái được sinh ra trong tổng số 88.273 trẻ sơ sinh, tần suất là 0,4825.

2.4. 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 .

Gọi A : “điểm M rơi vào miền $S \subset \Omega$ ”, ta có:

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



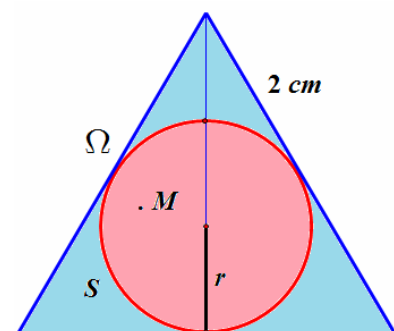
VD 5. Tìm xác suất của điểm M rơi vào hình tròn nội tiếp tam giác đều có cạnh 2 cm.

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

Diện tích của tam giác: $dt(\Omega) = \frac{2^2 \cdot \sqrt{3}}{4} = \sqrt{3} \text{ cm}^2$.

Bán kính của hình tròn: $r = \frac{1}{3} \cdot \frac{2\sqrt{3}}{2} = \frac{\sqrt{3}}{3} \text{ cm}$

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



VD 6. 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: $0 \leq x \leq 1, 0 \leq y \leq 1$.

Suy ra Ω là hình vuông có cạnh là 1 đơn vị.

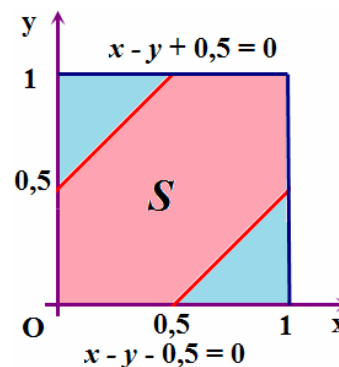
Từ điều kiện, ta có:

$$|x - y| \leq 0,5 \Leftrightarrow \begin{cases} x - y \leq 0,5 \\ x - y \geq -0,5 \end{cases} \Leftrightarrow \begin{cases} x - y - 0,5 \leq 0 \\ x - y + 0,5 \geq 0 \end{cases}$$

Suy ra, miền gặp nhau gặp nhau của hai người là S :

$$\{0 \leq x \leq 1, 0 \leq y \leq 1, x - y - 0,5 \leq 0, x - y + 0,5 \geq 0\}.$$

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



2.5. Tính chất của xác suất

- 1) $0 \leq P(A) \leq 1$, mọi biến cố A ;
- 2) $P(\emptyset) = 0$;
- 3) $P(\Omega) = 1$;
- 4) Nếu $A \subset B$ thì $P(A) \leq P(B)$.

BÀI 3. CÔNG THỨC TÍNH XÁC SUẤT

3.1. Công thức cộng xác suất

Xét một phép thử, ta có các công thức cộng xác suất sau

- Nếu A và B là hai biến cố tùy ý thì

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

- Nếu A và B là hai biến cố xung khắc thì

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

- Nếu họ $\{A_i\}$ ($i = 1, \dots, n$) xung khắc từng đôi thì

$$P(A_1 \cup A_2 \cup \dots \cup A_n) = P(A_1) + P(A_2) + \dots + P(A_n)$$

VD 1. Một nhóm có 30 nhà đầu tư các loại, trong đó có 13 nhà đầu tư vàng, 17 nhà đầu tư chứng khoán và 10 nhà đầu tư cả vàng lẫn chứng khoán. Một đối tác gặp ngẫu nhiên 1 nhà đầu tư trong nhóm. Tìm xác suất để người đó gặp được nhà đầu tư vàng hay chứng khoán?

.....

.....

.....

.....

.....

.....

.....

.....

▪ Đặc biệt

$$P(A) = 1 - P(\bar{A}); P(A) = P(A.B) + P(A.\bar{B})$$

VD 2. Một hộp phần có 10 viên trong đó có 3 viên màu đỏ. Lấy ngẫu nhiên từ hộp ra 3 viên phần. Tính xác suất để lấy được ít nhất 1 viên phần màu đỏ.

3.2. Xác suất có điều kiện

Xét phép thử: có 3 người A , B và C thi tuyển vào một công ty. Gọi

A : “ A thi đỗ”; B : “ B thi đỗ”; C : “ C thi đỗ”; H : “có 2 người thi đỗ”.

Khi đó, không gian mẫu Ω là $\{ABC, \bar{A}BC, A\bar{B}C, AB\bar{C}, \bar{A}\bar{B}C, \bar{A}B\bar{C}, A\bar{B}\bar{C}, \bar{A}\bar{B}\bar{C}\}$.

Ta có: $A = \{ABC, \bar{A}BC, AB\bar{C}, A\bar{B}\bar{C}\} \Rightarrow P(A) = \frac{4}{8}$; $H = \{\bar{A}BC, A\bar{B}C, AB\bar{C}\} \Rightarrow P(H) = \frac{3}{8}$.

Lúc này, biến cố “2 người thi đỗ trong đó có A ” là $AH = \{A\bar{B}C, AB\bar{C}\}$ và $P(AH) = \frac{2}{8}$.

Bây giờ, ta xét phép thử là: A , B , C thi tuyển vào một công ty và biết thêm thông tin có 2 người thi đỗ. Không gian mẫu trở thành H và A trở thành AH .

Gọi $A|H$: “ A thi đỗ biết rằng có 2 người thi đỗ” thì ta được $P(A|H) = \frac{2}{3} = \frac{P(AH)}{P(H)}$.

3.2.1. Định nghĩa xác suất có điều kiện

Trong một phép thử, xét hai biến cố bất kỳ A và B với $P(B) > 0$. Xác suất của biến cố A sau khi biến cố B đã xảy ra được gọi là xác suất của A với điều kiện B , ký hiệu và công thức tính là

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

VD 3. Từ 1 hộp chứa 3 bi đỏ và 7 bi xanh người ta bốc ngẫu nhiên ra 2 bi.

Gọi A : “bốc được bi đỏ”; B : “bốc được bi xanh”. Hãy tính $P(A|B)$, $P(B|A)$?

Nhận xét. Khi tính $P(A|B)$ 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) $0 \leq P(A|B) \leq 1, \forall A \subset \Omega$; 2) nếu $A \subset C$ thì $P(A|B) \leq P(C|B)$; 3) $P(A|B) = 1 - P(\bar{A}|B)$.

3.2.2. Công thức nhân xác suất**3.2.2.1. Sự độc lập của hai biến cố**

Trong một phép thử, hai biến cố A và B được gọi là *độc lập* nếu B có xảy ra hay không cũng không ảnh hưởng đến *khả năng xảy ra* A và ngược lại.

Chú ý. Nếu A và B độc lập với nhau thì các cặp biến cố :

\bar{A} và B , A và \bar{B} , \bar{A} và \bar{B} cũng độc lập với nhau.

3.2.2.2. Công thức nhân

Trong một phép thử, ta có:

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

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

- Nếu A và B là hai biến cố *không độc lập* (phụ thuộc) thì

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

- Nếu n biến cố $A_i (i = 1, \dots, n)$ *phụ thuộc* thì

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

VD 4. Một người có 5 bóng đèn trong đó có 2 bóng bị hỏng. Người đó thử ngẫu nhiên 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.

.....

VD 5. Một sinh viên học hệ niên chế được thi lại 1 lần nếu lần thi thứ nhất bị rớt (2 lần thi độc lập). Biết rằng xác suất để sinh viên này thi đỗ lần 1 và lần 2 tương ứng là 60%, 80%. Tính xác suất sinh viên này thi đỗ?

.....

VD 6. Có hai người A và B cùng đặt lệnh (độc lập) để mua cổ phiếu của một công ty với xác suất mua được tương ứng là 0,8 và 0,7. Biết rằng có người mua được, xác suất để người A mua được cổ phiếu này là:

A. $\frac{19}{47}$;

B. $\frac{12}{19}$;

C. $\frac{40}{47}$;

D. $\frac{10}{19}$.

.....

.....

.....

.....

.....

.....

.....

.....

VD 7*. Ông A bắn lần lượt 2 viên đạn vào 1 mục tiêu và mục tiêu sẽ bị phá hủy nếu bị trúng cả 2 viên đạn. Xác suất viên đạn thứ nhất trúng mục tiêu là 0,8. Nếu viên thứ nhất trúng mục tiêu thì xác suất viên thứ hai trúng là 0,7. Nếu viên thứ nhất không trúng thì xác suất viên thứ hai trúng mục tiêu là 0,3. Biết rằng ông A bắn trúng, tính xác suất để mục tiêu bị phá hủy ?

.....

.....

.....

.....

.....

VD 8*. Trong dịp tết, ông 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 ông A bán được ít nhất 1 cây mai, xác suất để ông A bán được cả hai cây mai là:

A. 0,6342;

B. 0,6848;

C. 0,4796;

D. 0,8791.

VD 9. Hai người A và B cùng chơi trò chơi như sau: Cả hai luân phiên lấy mỗi lần 1 viên bi từ một hộp đựng 2 bi trắng và 4 bi đen (bi được lấy ra không trả lại hộp). Người nào lấy được bi trắng trước thì thắng cuộc. Giả sử A lấy trước, tính xác suất A thắng cuộc ?

.....

.....

.....

.....

.....

.....

3.2.3. Công thức xác suất đầy đủ và Bayes

3.2.3.1. Công thức xác suất đầy đủ

Xét họ n biến cố $\{A_i\}$ ($i = 1, 2, \dots, n$) đầy đủ và B là một biến cố bất kỳ trong phép thử, ta có

$$P(B) = P(A_1)P(B|A_1) + \dots + P(A_n)P(B|A_n) = \sum_{i=1}^n P(A_i)P(B|A_i)$$

Chứng minh

$$\begin{aligned} P(B) &= P(B \cap \Omega) = P\left[B \cap (A_1 \cup A_2 \cup \dots \cup A_n)\right] = P(BA_1 \cup BA_2 \cup \dots \cup BA_n) \\ &= P(A_1B) + P(A_2B) + \dots + P(A_nB) = P(A_1)P(B|A_1) + \dots + P(A_n)P(B|A_n). \blacksquare \end{aligned}$$

VD 10. Một cửa hàng bán hai loại bóng đèn cùng kích cỡ gồm: 70 bóng màu trắng với tỉ lệ bóng hỏng là 1%, 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 ?

Giải. Gọi B : “khách chọn được bóng đèn tốt”,

A_1 : “khách chọn được bóng đèn màu trắng”,

A_2 : “khách chọn được bóng đèn màu vàng”.

Suy ra hệ $\{A_1, A_2\}$ là đầy đủ. Ta có:

$$P(B) = P(A_1)P(B | A_1) + P(A_2)P(B | A_2) = \frac{70}{70 + 30} \cdot 0,99 + \frac{30}{70 + 30} \cdot 0,98 = 0,987.$$

VD 11. Chuồng thỏ I có 3 con thỏ trắng và 4 con thỏ đen, chuồng II có 5 thỏ trắng và 3 thỏ đen. Quan sát thấy có 1 con thỏ chạy từ chuồng I sang chuồng II, sau đó có 1 con thỏ chạy ra từ chuồng II. Tính xác suất để con thỏ chạy ra từ chuồng II là thỏ trắng ?

VD 12*. Có một kho bia kém chất lượng chứa các thùng giống nhau (24 lon/thùng) gồm 2 loại: loại I để lần mỗi thùng 5 lon quá hạn sử dụng và loại II để lần mỗi thùng 3 lon quá hạn. Biết rằng số thùng bia loại I bằng 1,5 lần số thùng bia loại II. Chọn ngẫu nhiên 1 thùng trong kho và từ thùng đó lấy ra 10 lon. Tính xác suất chọn phải 2 lon bia quá hạn sử dụng ?

3.2.3.2. Công thức Bayes

Xét họ n biến cố $\{A_i\}$ ($i = 1, 2, \dots, n$) đầy đủ và B là biến cố bất kỳ trong phép thử. Khi đó, xác suất để biến cố A_i xảy ra sau khi B đã xảy ra là

$$P(A_i | B) = \frac{P(A_i)P(B | A_i)}{P(B)}$$

VD 13. Xét tiếp VD 10. Giả sử khách hàng chọn mua được bóng đèn tốt. Tính xác suất để người này mua được bóng đèn màu vàng ?

Giải. Đặt tên biến cố như VD 10, ta có:

$$P(A_2 | B) = \frac{P(A_2)P(B | A_2)}{P(B)} = \frac{0,3 \cdot 0,98}{0,987} = \frac{14}{47}.$$

VD 14*. Có 20 thùng hàng giống nhau gồm 3 loại: 8 thùng loại I, 7 thùng loại II và 5 thùng loại III. Mỗi thùng hàng có 10 sản phẩm và số sản phẩm tốt tương ứng cho mỗi loại lần lượt là 8, 7 và 5. Chọn ngẫu nhiên 1 thùng hàng và từ thùng đó lấy ra 3 sản phẩm.

- 1) Tính xác suất có 2 sản phẩm lấy ra là tốt.
- 2) Tính xác suất có 2 sản phẩm lấy ra là tốt và của thùng hàng loại II.
- 3) Giả sử có 2 sản phẩm lấy ra là tốt, tính xác suất 2 sản phẩm này là của thùng hàng loại II.

VD 15. Nhà máy X có 3 phân xưởng A, B, C tương ứng sản xuất ra 20%, 30% và 50% tổng sản phẩm của nhà máy. Giả sử tỉ lệ sản phẩm hỏng do các phân xưởng A, B, C tương ứng sản xuất ra là 1%, 2%, 3%. Chọn ngẫu nhiên 1 sản phẩm do nhà máy X sản xuất ra.

- 1) Tính xác suất (tỉ lệ) sản phẩm này là hỏng.
- 2) Tính xác suất sản phẩm này hỏng và do phân xưởng A sản xuất ra.
- 3) Biết rằng sản phẩm được chọn là hỏng, tính xác suất sản phẩm này là do phân xưởng A sản xuất ra.

VD 16. Tỉ lệ ô tô tải, ô tô con và xe máy đi qua đường X có trạm bơm dầu là 5 : 2 : 13. Xác suất để ô tô tải, ô tô 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 ô tô con ?

- A. $\frac{11}{57}$; B. $\frac{10}{57}$; C. $\frac{8}{57}$; D. $\frac{7}{57}$.

Chương 2. BIẾN NGẪU NHIÊN

Bài 1. Biến ngẫu nhiên và hàm mật độ

Bài 2. Hàm phân phối xác suất

Bài 3. Tham số đặc trưng của biến ngẫu nhiên

BÀI 1. BIẾN NGẪU NHIÊN VÀ HÀM MẬT ĐỘ

1.1. Khái niệm biến ngẫu nhiên

Xét một phép thử với không gian mẫu Ω . Giả sử, ứng với mỗi biến cố sơ cấp $\omega \in \Omega$, ta liên kết với một số thực $X(\omega) \in \mathbb{R}$, thì X được gọi là một *biến ngẫu nhiên* (hay đại lượng ngẫu nhiên).

Tổng quát, biến ngẫu nhiên (BNN) X của một phép thử với không gian mẫu Ω là một ánh xạ

$$\begin{aligned} X : \Omega &\rightarrow \mathbb{R} \\ \omega &\mapsto X(\omega) = x. \end{aligned}$$

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

- Nếu tập giá trị $\{X(\omega) \mid \omega \in \Omega\}$ của X là hữu hạn hay đếm được thì ta gọi X là BNN *rời rạc*.

Đặt $X(\omega_i) = x_i$ ($i = 1, 2, \dots, x_n, \dots$), ta ký hiệu $X = \{x_1, x_2, \dots, x_n, \dots\}$.

- Nếu tập giá trị $\{X(\omega) \mid \omega \in \Omega\}$ lấp đầy một khoảng trên trục số thì ta gọi X là BNN *liên tục*.

- Cho biến ngẫu nhiên X và hàm số $y = \varphi(x)$. Khi đó, biến ngẫu nhiên $Y = \varphi(X)$ được gọi là *hàm của biến ngẫu nhiên X* . Và Y cũng là một biến ngẫu nhiên.

VD 1. Một hộp chứa 3 lá thăm màu đỏ và 2 lá thăm màu đen. Một người bốc lần lượt 2 lá thăm từ hộp đó. Nếu bốc được lá thăm đỏ thì được thưởng 100 ngàn đồng; nếu bốc lá thăm đen thì bị phạt 70 ngàn đồng. Gọi A_i : “bốc được lá thăm đỏ lần thứ i ” ($i = 1, 2$), X là số lá thăm đỏ bốc được và Y là số tiền có được.

- Không gian mẫu là $\Omega = \{\bar{A}_1\bar{A}_2, \bar{A}_1A_2, A_1\bar{A}_2, A_1A_2\}$.
- X là biến ngẫu nhiên và $X = \{0; 1; 2\}$.
- $Y = 100X - 70(2 - X)$ (ngàn đồng) là hàm của X và $Y = \{-140; 30; 200\}$.

Chú ý

Trong thực nghiệm, các biến ngẫu nhiên thường là rời rạc. Khi biến ngẫu nhiên rời rạc X có các giá trị đủ nhiều trên 1 khoảng của \mathbb{R} , thì ta xem X là biến ngẫu nhiên liên tục. Thực chất là, các biến ngẫu nhiên liên tục được dùng làm xấp xỉ cho các biến ngẫu nhiên rời rạc khi tập giá trị của biến ngẫu nhiên rời rạc đủ lớn.

1.2. Hàm mật độ

1.2.1. Biến ngẫu nhiên rời rạc

Xét BNN $X = \{x_1, x_2, \dots, x_n, \dots\}$ ($x_1 < x_2 < \dots < x_n < \dots$) với xác suất tương ứng là

$$P(X = x_i) = p_i \quad (i = 1, 2, \dots).$$

Ta định nghĩa

- Bảng phân phối xác suất của X là

X	x_1	x_2	\dots	x_n	\dots
P	p_1	p_2	\dots	p_n	\dots

- Hàm mật độ của X (tham khảo) là

$$f(x) = \begin{cases} p_i & \text{khi } x = x_i \quad \forall i. \\ 0 & \text{khi } x \neq x_i \end{cases}$$

Chú ý

- $p_i \geq 0$ và $\sum p_i = 1$ ($i = 1, 2, \dots$)
- Nếu $x \notin \{x_1, x_2, \dots, x_n, \dots\}$ thì $P(X = x) = 0$
- $P(a < X \leq b) = \sum_{a < x_i \leq b} p_i$

VD 2. Cho BNN rời rạc X có bảng phân phối xác suất

X	-1	0	1	3	5
P	$3a$	a	0,1	$2a$	0,3

- 1) Tìm a và tính $P(-1 < X \leq 3)$.
- 2) Lập bảng phân phối xác suất của hàm $Y = X^2$.

.....

.....

.....

.....

VD 3. Một xạ thủ có 4 viên đạn, bắn lần lượt từng viên vào một mục tiêu một cách độc lập. Xác suất trúng mục tiêu ở mỗi lần bắn là 0,8. Biết rằng, nếu có 1 viên trúng mục tiêu hoặc hết đạn thì dừng. Gọi X là số viên đạn xạ thủ đã bắn, lập bảng phân phối xác suất của X ?

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

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

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

1.2.2. Biến ngẫu nhiên liên tục

Hàm số $f(x)$ không âm, xác định trên \mathbb{R} được gọi là *hàm mật độ* của biến ngẫu nhiên liên tục X nếu

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

Chú ý. Hàm số $f(x)$ là hàm mật độ của biến ngẫu nhiên liên tục X khi và chỉ khi

$$f(x) \geq 0, \forall x \in \mathbb{R} \text{ và } \int_{-\infty}^{+\infty} f(x)dx = 1.$$

Nhận xét

- Khi $f(x)$ liên tục trên lân cận của điểm a , ta có:

$$P(a - \varepsilon \leq X \leq a + \varepsilon) = \int_{a-\varepsilon}^{a+\varepsilon} f(x)dx \Rightarrow P(X = a) = \lim_{\varepsilon \rightarrow 0} \int_{a-\varepsilon}^{a+\varepsilon} f(x)dx = 0.$$

$$\text{Vậy } P(a \leq X < b) = P(a < X \leq b) = P(a < X < b) = \int_a^b f(x)dx.$$

- Ý nghĩa hình học, xác suất của biến ngẫu nhiên X nhận giá trị trong $[a; b]$ bằng diện tích hình thang cong giới hạn bởi $x = a$, $x = b$, $y = f(x)$ và Ox .

Chú ý

$$\bullet \int_{-\infty}^{+\infty} f(x)dx = F(x) \Big|_{-\infty}^{+\infty} = F(+\infty) - F(-\infty). \quad \bullet \int_{-\infty}^{+\infty} f(x)dx = F(x) \Big|_{-\infty}^{+\infty} = \lim_{x \rightarrow +\infty} F(x) - \lim_{x \rightarrow -\infty} F(x).$$

VD 5. 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 và tính $P(0,5 \leq X < 3)$?

.....

.....

.....

.....

.....

.....

.....

.....

VD 6. Cho BNN X có hàm mật độ $f(x) = \begin{cases} 0, & x < 2 \\ \frac{k}{x^2}, & x \geq 2. \end{cases}$ Tính $P(-3 < X < 5)$?

.....

.....

.....

.....

.....

.....

.....

.....

Bài 2. HÀM PHÂN PHỐI XÁC SUẤT**2.1. Định nghĩa**

Hàm phân phối xác suất (hay hàm phân phối tích lũy) của biến ngẫu nhiên X , ký hiệu $F(x)$, là xác suất để X nhận giá trị nhỏ hơn x với mọi $x \in \mathbb{R}$. Nghĩa là

$$F(x) = P(X < x), \forall x \in \mathbb{R}$$

Nhận xét 1

- Nếu biến ngẫu nhiên X là rời rạc có phân phối xác suất $P(X = x_i) = p_i$ thì

$$F(x) = \sum_{x_i < x} p_i$$

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

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

Nhận xét 2

- 1) Giả sử BNN rời rạc X nhận các giá trị trong $[x_1; x_n]$ ($x_1 < x_2 < \dots < x_n$) và có phân phối xác suất $P(X = x_i) = p_i$ ($i = 1, 2, \dots, n$). Ta có hàm phân phối xác suất của X là

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

Chứng minh

- Với $x \leq x_1$: $F(x) = P(X < x) = P(X < x_1) = P(\phi) = 0$.
- Với $x_1 < x \leq x_2$: $F(x) = P(X < x) = P(X < x_2) = P(X = x_1) = p_1$.
- Với $x_2 < x \leq x_3$: $F(x) = P(X < x) = P(X < x_3) = P(X = x_1) + P(X = x_2) = p_1 + p_2$.
-
- Với $x > x_n$: $F(x) = P(X \leq x) = P(X \leq x_n) = P(X = x_1) + P(X = x_2) + \dots + P(X = x_n) = p_1 + p_2 + \dots + p_n = 1$. ■

- **Quy ước.** Nếu BNN X liên tục thì miền xác định của $F(x)$ được lấy theo hàm mật độ $f(x)$.

$$2) \text{ Nếu BNN } X \text{ có hàm mật độ } f(x) = \begin{cases} \varphi(x), & x \in [a; b] \\ 0, & x \notin [a; b] \end{cases} \text{ thì } F(x) = \begin{cases} 0 & \text{khi } x < a \\ \int_a^x \varphi(t)dt & \text{khi } a \leq x \leq b \\ 1 & \text{khi } b < x. \end{cases}$$

$$3) \text{ Nếu BNN } X \text{ có hàm mật độ } f(x) = \begin{cases} 0, & x < a \\ \varphi(x), & x \geq a \end{cases} \text{ thì } F(x) = \begin{cases} 0 & \text{khi } x < a \\ \int_a^x \varphi(t)dt & \text{khi } x \geq a. \end{cases}$$

$$4) \text{ Nếu BNN } X \text{ có hàm mật độ } f(x) = \begin{cases} \varphi(x), & x \leq a \\ 0, & x > a \end{cases} \text{ thì } F(x) = \begin{cases} \int_{-\infty}^x \varphi(t)dt & \text{khi } x \leq a \\ 1 & \text{khi } x > a. \end{cases}$$

VD 1. Cho BNN X có bảng phân phối xác suất là

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

Lập hàm phân phối xác suất $F(x)$ của X .

.....

.....

.....

.....

.....

VD 2. BNN X có hàm mật độ $f(x) = \begin{cases} 0, & x \notin [0; 1] \\ 3x^2, & x \in [0; 1] \end{cases}$. Tìm hàm phân phối xác suất $F(x)$ của X .

.....

.....

.....

.....

.....

VD 3. BNN X có hàm mật độ $f(x) = \begin{cases} 0, & x < 100 \\ \frac{100}{x^2}, & x \geq 100 \end{cases}$. Tìm hàm phân phối xác suất $F(x)$ của X .

.....

.....

.....

.....

.....

2.2. Tính chất của hàm phân phối xác suất

- 1) Hàm $F(x)$ xác định với mọi $x \in \mathbb{R}$.
- 2) $0 \leq F(x) \leq 1, \forall x \in \mathbb{R}; F(-\infty) = 0; F(+\infty) = 1$.
- 3) $F(x)$ không giảm và liên tục trái tại mọi $x \in \mathbb{R}$. Đặc biệt, với X liên tục thì $F(x)$ liên tục $\forall x \in \mathbb{R}$.
- 4) $P(a \leq X < b) = F(b) - F(a), \forall a, b \in \mathbb{R}$.

Chú ý

- Nếu X là BNN rời rạc thì

$$p_i = F(x_{i+1}) - F(x_i), \forall i$$

- Nếu X là BNN liên tục thì

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

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

$$F'(x) = f(x)$$

VD 4. Tính xác suất $P(X \geq 400)$ trong VD 3.

.....

.....

.....

.....

.....

VD 5. X có hàm mật độ $f(x) = \begin{cases} \frac{3}{28}x^2, & x \in [-1; 3] \\ 0, & x \notin [-1; 3]. \end{cases}$ Hàm phân phối xác suất của X là:

$$\begin{aligned} \text{A. } F(x) &= \begin{cases} 0, & x < -1 \\ \frac{x^3}{28}, & -1 \leq x \leq 3 \\ 1, & 3 < x. \end{cases} & \text{B. } F(x) &= \begin{cases} 0, & x < -1 \\ \frac{x^3}{28}, & -1 \leq x < 3 \\ 1, & 3 \leq x. \end{cases} \\ \text{C. } F(x) &= \begin{cases} 0, & x < -1 \\ \frac{x^3}{28} - \frac{1}{28}, & -1 \leq x \leq 3 \\ 1, & 3 < x. \end{cases} & \text{D. } F(x) &= \begin{cases} 0, & x < -1 \\ \frac{x^3}{28} + \frac{1}{28}, & -1 \leq x \leq 3 \\ 1, & 3 < x. \end{cases} \end{aligned}$$

VD 6. BNN X có hàm $F(x) = \begin{cases} 0, & x \leq -2 \\ ax^3 + 2b, & x \in (-2; 3] \\ 1, & x > 3. \end{cases}$

1) Tìm các hằng số a và b ?

2) Tính $P(\sqrt{2} < Y \leq \sqrt{5})$ với $Y = \sqrt{X^2 + 1}$.

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

BÀI 3. THAM SỐ ĐẶC TRƯNG 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 3 loại đặc trưng số là

- Các đặc trưng số cho xu hướng trung tâm của BNN: *Trung vị, Mode, Kỳ vọng,...*
- 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.

3.1. Mode

Mode của biến ngẫu nhiên X , ký hiệu $\text{Mod } X$, là giá trị $x_0 \in X$ thỏa mãn:

- $\max_{x \in X} P(X = x) = P(X = x_0)$ nếu X là rời rạc, và
- $\max_{x \in \mathbb{R}} f(x) = f(x_0)$ nếu X liên tục có hàm mật độ $f(x)$.

Chú ý

- $\text{Mod } X$ còn được gọi là *giá trị tin chắc nhất* của X .
- Biến ngẫu nhiên X có thể có nhiều $\text{Mod } X$.

VD 1. 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

Ta có $\text{Mod } X = 2$.

VD 2. Tìm $\text{Mod } X$, biết X có bảng phân phối xác suất

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

VD 3. Tìm $\text{Mod } X$, biết X có hàm mật độ xác suất $f(x) = \begin{cases} \frac{3}{64}x^2(4-x), & x \in [0; 4] \\ 0, & x \notin [0; 4]. \end{cases}$

3.2. KỲ VỌNG

3.2.1. Định nghĩa

Kỳ vọng của biến ngẫu nhiên X , ký hiệu EX hay $M(X)$, là một số thực được xác định như sau

- Nếu X là rời rạc với xác suất $P(X = x_i) = p_i$ thì

$$EX = \sum_i x_i p_i$$

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

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

Đặc biệt

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

$$EX = x_1 p_1 + x_2 p_2 + \dots + x_n p_n$$

VD 4. Cho BNN X có bảng phân phối xác suất

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

Tính kỳ vọng của X ?

.....

.....

.....

VD 5. Một lô hàng có 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 6. Tìm kỳ vọng của BNN X có hàm mật độ $f(x) = \begin{cases} \frac{3}{4}(x^2 + 2x), & x \in [0; 1] \\ 0, & x \notin [0; 1]. \end{cases}$

.....

.....

.....

.....

VD 7. Tìm kỳ vọng của BNN X có hàm mật độ $f(x) = \begin{cases} e^{kx}, & x \leq 0 \\ 0, & x > 0. \end{cases}$

.....

.....

.....

.....

.....

Chú ý

- Nếu X là BNN liên tục trên $[a; b]$ thì $EX \in [a; b]$.
- Nếu $X = \{x_1, \dots, x_n\}$ thì $EX \in [\min\{x_1, \dots, x_n\}; \max\{x_1, \dots, x_n\}]$.

VD 8. Cho BNN X có bảng phân phối xác suất

X	1	2	4	5	7
P	a	0,2	b	0,2	0,1

Tìm giá trị của tham số a và b để $EX = 3,5$?

.....

.....

.....

VD 9. Cho biến ngẫu nhiên X có hàm mật độ $f(x) = \begin{cases} ax + bx^2, & x \in [0; 1] \\ 0, & x \notin [0; 1]. \end{cases}$

Cho biết $EX = 0,6$ hãy tính $P(X < 0,5)$?

3.2.2. Tính chất của Kỳ vọng

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

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

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

$$4) E(X.Y) = EX.EY \text{ nếu } X, Y \text{ độc lập.}$$

VD 10. Cho hai BNN X, Y độc lập có bảng ppxs:

X	-1	1	3
P	0,3	0,1	0,6

Y	-1	2
P	0,6	0,4

Tính $E(X^2.Y - 3XY + 5Y + 7)$.

3.2.3. Ý nghĩa của Kỳ vọng

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

VD 11. Thống kê cho biết tỉ lệ tai nạn xe máy ở thành phố H là 0,001. Công ty bảo hiểm A đề nghị bán loại bảo hiểm tai nạn xe máy cho ông B ở thành phố H trong 1 năm với số tiền chi trả là 10 (triệu đồng), phí bảo hiểm là 0,1 (triệu đồng). Hỏi trung bình công ty A lãi bao nhiêu khi bán bảo hiểm cho ông B ?

VD 12. Một cửa hàng điện máy lời 2,3 triệu đồng khi bán được 1 máy giặt, nhưng nếu máy giặt bị hỏng trước thời hạn bảo hành thì bị lỗ 4,5 triệu. Biết rằng cửa hàng lời trung bình 1,96 triệu đồng khi bán được 1 máy giặt. Tính tỉ lệ máy giặt phải bảo hành?

VD 13. Ông A tham gia một trò chơi đỏ, đen như sau:

Trong một hộp có 4 bi đỏ và 6 bi đen. Mỗi lần ông A lấy ra 1 bi: nếu là đỏ thì được thưởng 100 (ngàn đồng), nếu là đen thì bị mất 70 (ngàn đồng). Hỏi trung bình mỗi lần lấy bi ông A bị mất bao nhiêu tiền?

VD 14. 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. 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 nhận 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.

VD 15. Nhu cầu hàng ngày của một khu phố về 1 loại thực phẩm tươi sống có bảng phân phối xác suất

Nhu cầu (kg)	31	32	33	34
P	0,15	0,25	0,45	0,15

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

3.2.4. Kỳ vọng của hàm của biến ngẫu nhiên

Giả sử $Y = \varphi(X)$ là hàm của biến ngẫu nhiên X .

- Nếu X là biến ngẫu nhiên rời rạc thì

$$EY = \sum_i y_i \cdot p_i = \sum_i \varphi(x_i) \cdot p_i$$

- Nếu X là biến ngẫu nhiên liên tục thì

$$EY = \int_{-\infty}^{+\infty} y \cdot f(x) dx = \int_{-\infty}^{+\infty} \varphi(x) \cdot f(x) dx$$

Chú ý. Khi biến ngẫu nhiên X là rời rạc thì ta nên lập bảng phân phối xác suất của Y , rồi tính EY .

VD 16. Cho BNN X có bảng phân phối xác suất

X	-1	0	1	2
P	0,1	0,3	0,35	0,25

Tính EY với $Y = X^2 - 3$?

VD 17. Cho BNN X có hàm mật độ $f(x) = \begin{cases} \frac{2}{x^2}, & x \in [1; 2] \\ 0, & x \notin [1; 2]. \end{cases}$ Tính EY với $Y = X^5 - \frac{2}{X}$.

.....

.....

.....

3.3. PHƯƠNG SAI

3.3.1. Định nghĩa

Phương sai của BNN X , ký hiệu $VarX$ hay $D(X)$, là một số thực không âm được xác định bởi

$$VarX = E(X - EX)^2 = E(X^2) - (EX)^2$$

- Nếu BNN X là rời rạc và $P(X = x_i) = p_i$ thì

$$VarX = \sum_i x_i^2 \cdot p_i - \left(\sum_i x_i \cdot p_i \right)^2$$

- Nếu BNN X là liên tục và có hàm mật độ $f(x)$ thì

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

VD 18. Cho BNN X có bảng phân phối xác suất

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

Ta có:

$$VarX = (1^2 \cdot 0,2 + 2^2 \cdot 0,7 + 3^2 \cdot 0,1) - (1 \cdot 0,2 + 2 \cdot 0,7 + 3 \cdot 0,1)^2 = 0,29.$$

VD 19. Tính phương sai của X , biết hàm mật độ $f(x) = \begin{cases} \frac{3}{4}(x^2 + 2x), & x \in [0; 1] \\ 0, & x \notin [0; 1]. \end{cases}$

.....

.....

.....

.....

VD 20. BNN X có hàm mật độ $f(x) = \begin{cases} \frac{3}{4}(1 - x^2), & |x| \leq 1 \\ 0, & |x| > 1. \end{cases}$ Tính phương sai của Y , cho biết $Y = 2X^2$.

.....

.....

.....

.....

.....

3.3.2. Tính chất của Phương sai

- 1) $VarC = 0, C \in \mathbb{R}$;
- 2) $Var(CX) = C^2 \cdot VarX$;
- 3) $Var(X \pm Y) = VarX + VarY$ nếu X và Y độc lập.

3.3.3. Ý nghĩa của Phương sai

- $(X - EX)^2$ là bình phương sai biệt giữa giá trị của X so với trung bình của nó. Và phương sai là trung bình của sai biệt này, nên phương sai cho ta hình ảnh về sự phân tán của các số liệu: phương sai càng nhỏ thì số liệu càng tập trung xung quanh trung bình của chúng.
- 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 là

$$\sigma = \sqrt{VarX}$$

VD 21. Năng suất (sp/phút) của hai máy tương ứng là các BNN X và Y , có 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

Từ bảng phân phối xác suất, ta tính được:

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

Vì $EX < EY$, $VarX > VarY$ nên nếu phải chọn mua 1 trong 2 loại máy này thì ta chọn mua máy Y .

Chú ý. Trong trường hợp $\begin{cases} EX < EY \\ VarX < VarY \end{cases}$ hay $\begin{cases} EX > EY \\ VarX > VarY \end{cases}$ thì ta không thể so sánh được. Để giải

quyết vấn đề này, trong thực tế người ta dùng tỉ số tương đối $\frac{\sigma}{\mu} \cdot 100\%$ (μ là trung bình) để so sánh sự ổn định của các BNN X và Y . Tỉ số tương đối càng nhỏ thì độ ổn định càng cao.

VD 22. Điểm thi hết môn XSTK của lớp A và B tương ứng là các BNN X và Y . Người ta tính được:

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

Ta có: $\frac{\sigma_x}{EX} \cdot 100\% = 17,89\%$; $\frac{\sigma_y}{EY} \cdot 100\% = 15,06\%$.

Vậy lớp B học đều (ổn định) hơn lớp A .

3.4. Một số đặc trưng khác (tham khảo)

Xét BNN X có kỳ vọng, phương sai là μ và σ^2 .

3.4.1. Hệ số đối xứng của X

$$\gamma_1(X) = \frac{E(X - \mu)^3}{\sigma^3}.$$

Khi $\gamma_1(X) = 0$ thì phân phối của X là đối xứng; lệch phải khi $\gamma_1(X) > 0$ và lệch trái khi $\gamma_1(X) < 0$.

3.4.2. Hệ số nhọn của X

$$\gamma_2(X) = \frac{E(X - \mu)^4}{\sigma^4}.$$

Khi $\gamma_2(X)$ càng lớn thì phân phối của X càng nhọn.

Chương 3. PHÂN PHỐI XÁC SUẤT THÔNG DỤNG

Bài 1. Phân phối Siêu bội

Bài 2. Phân phối Nhị thức

Bài 3. Phân phối Poisson

Bài 4. Phân phối Chuẩn

Bài 5. Vector ngẫu nhiên rời rạc

BÀI 1. PHÂN PHỐI SIÊU BỘI

1.1. Định nghĩa

- Xét tập có N phần tử gồm N_A phần tử có tính chất A và $N - N_A$ phần tử có tính chất \bar{A} . Từ tập đó, ta chọn ra n phần tử. Gọi X là số phần tử có tính chất A lần trong n phần tử đã chọn thì X có phân phối Siêu bội, ký hiệu là $X \in H(N, N_A, n)$ hay $X \sim H(N, N_A, n)$.
- Xác suất trong n phần tử chọn ra có k phần tử A là

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

trong đó

$$0 \leq k \leq n \text{ và } n - (N - N_A) \leq k \leq N_A.$$

VD 1. Một hộp phấn gồm 10 viên, trong đó có 7 viên màu trắng. Lấy ngẫu nhiên 5 viên phấn từ hộp này. Gọi X là số viên phấn trắng lấy được. Lập bảng phân phối xác suất và tính kỳ vọng của X ?

.....

.....

.....

.....

.....

.....

.....

.....

1.2. Các số đặc trưng của $X \sim H(N, N_A, n)$

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

$$\text{trong đó: } p = \frac{N_A}{N}, \quad q = 1 - p.$$

VD 2. Một cửa hàng bán 100 bóng đèn, trong đó có 12 bóng hỏng. Một người chọn mua ngẫu nhiên 15 bóng đèn từ cửa hàng này. Hỏi trung bình người đó mua được bao nhiêu bóng đèn tốt ?

.....

.....

.....

VD 3. Tại một công trình có 100 người đang làm việc, trong đó có 70 kỹ sư. Chọn ngẫu nhiên 40 người từ công trình này. Gọi X là số kỹ sư chọn được.

1) Tính xác suất chọn được từ 27 đến 29 kỹ sư ?

2) Tính EX và $VarX$?

BÀI 2. PHÂN PHỐI NHỊ THỨC

2.1. Phân phối Bernoulli

2.1.1. Định nghĩa

- Phép thử Bernoulli là phép thử mà ta chỉ quan tâm đến 2 biến cố A và \bar{A} , với $P(A) = p$.
- Xét biến ngẫu nhiên:

$$X = \begin{cases} 1 & \text{khi } A \text{ xảy ra} \\ 0 & \text{khi } \bar{A} \text{ xảy ra,} \end{cases} \quad P(\bar{A}) = 1 - p = q.$$

Khi đó ta nói X có phân phối Bernoulli với tham số p , ký hiệu là $X \in B(p)$ hay $X \sim B(p)$.
Bảng phân phối xác suất của X là

X	0	1
P	q	p

2.1.2. Các số đặc trưng của $X \sim B(p)$

$$EX = p; \quad VarX = pq$$

VD 1. Một câu hỏi trắc nghiệm có 4 phương án trả lời, trong đó chỉ có 1 phương án đúng. Một sinh viên chọn ngẫu nhiên 1 phương án để trả lời câu hỏi đó.

Gọi A : “sinh viên này trả lời đúng”. Khi đó, việc trả lời câu hỏi của sinh viên này là một phép thử

Bernoulli và $p = P(A) = \frac{1}{4}$, $q = \frac{3}{4}$.

Gọi BNN $X = \begin{cases} 1 & \text{khi sinh viên này trả lời đúng} \\ 0 & \text{khi sinh viên này trả lời sai,} \end{cases}$ thì $X \in B\left(\frac{1}{4}\right)$ và $EX = \frac{1}{4}$, $VarX = \frac{1}{4} \cdot \frac{3}{4} = \frac{3}{16}$.

2.2. Phân phối Nhị thức

2.2.1. Định nghĩa

- Xét dãy n phép thử Bernoulli độc lập. Với phép thử thứ i ($i = 1, \dots, n$), ta xét biến ngẫu nhiên

$$X_i \in B(p). \text{ Nghĩa là, } X_i = \begin{cases} 1 & \text{khi lần thứ } i \text{ } A \text{ xảy ra} \\ 0 & \text{khi lần thứ } i \text{ } A \text{ không xảy ra.} \end{cases}$$

- Gọi X là số lần biến cố A xuất hiện trong n phép thử. Khi đó, $X = X_1 + \dots + X_n$ và ta nói X có phân phối Nhị thức, ký hiệu là $X \in B(n, p)$ hay $X \sim B(n, p)$.

- Xác suất trong n lần thử có k lần A xảy ra là

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

VD 2. Một đề thi XSTK gồm 20 câu hỏi trắc nghiệm như trong VD 1. Sinh viên B làm bài một cách ngẫu nhiên. Biết rằng, nếu trả lời đúng 1 câu thì sinh viên B được 0,5 điểm và nếu trả lời sai 1 câu thì bị trừ 0,125 điểm. Tính xác suất để sinh viên B đạt điểm 5 ?

2.2.2. Các số đặc trưng của $X \sim B(n, p)$

$EX = np; \text{ Var}X = npq$ $\text{Mod } X = x_0 \in \mathbb{N}: np - q \leq x_0 \leq np - q + 1$
--

VD 3. Ông B trồng 100 cây bạch đàn với xác suất cây chết là 0,02. Gọi X là số cây bạch đàn chết.

- 1) Tính xác suất có từ 3 đến 5 cây bạch đàn chết ?
- 2) Tính trung bình số cây bạch đàn chết và $VarX$?
- 3) Hỏi ông B cần phải trồng tối thiểu mấy cây bạch đàn để xác suất có ít nhất 1 cây chết lớn hơn 50% ?

VD 4. Một nhà vườn trồng 126 cây lan quý, xác suất nở hoa của mỗi cây trong 1 năm là 0,67.

- 1) Giá bán 1 cây lan quý nở hoa là 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 nhiều hơn 100 cây lan quý nở hoa thì nhà vườn phải trồng tối thiểu mấy cây lan quý ?

VD 5. Có 10 hộp phần màu giống nhau, mỗi hộp chứa 20 viên phần gồm hai loại: 3 hộp loại I, mỗi hộp có 12 viên phần đỏ; 7 hộp loại II, mỗi hộp có 8 viên phần đỏ. Chọn ngẫu nhiên 1 hộp và từ hộp đó lấy lần lượt ra 5 viên phần (lấy viên nào xong thì trả lại vào hộp). Tính xác suất chọn được 3 viên phần đỏ ?

VD 6*. Một lô hàng chứa 20 sản phẩm trong đó có 4 phế phẩm. Chọn liên tiếp 3 lần từ lô hàng (mỗi lần chọn có hoàn lại), mỗi lần chọn ra 4 sản phẩm. Tính xác suất để trong 3 lần chọn có ít nhất 1 lần chọn phải 2 phế phẩm ?

BÀI 3. PHÂN PHỐI POISSON

3.1. Bài toán dẫn đến phân phối Poisson

- Giả sử các vụ tai nạn giao thông ở vùng A xảy ra một 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 đó có nhiều nhất 1 vụ tai nạn xảy ra, và khả năng xảy ra tai nạn giao thông trong mỗi khoảng thời gian bằng $\frac{\lambda}{n}$. Khi đó, $X \in B\left(n, \frac{\lambda}{n}\right)$.

$$\begin{aligned} \bullet \text{ Ta có: } P(X = k) &= C_n^k \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} = \frac{n!}{k!(n-k)!} \cdot \frac{\lambda^k}{n^k} \cdot \frac{1}{(n-\lambda)^k} \cdot \left(1 - \frac{\lambda}{n}\right)^n \\ &= \frac{\lambda^k}{k!} \cdot \frac{n(n-1)\dots(n-k+1)}{(n-\lambda)^k} \cdot \left(1 - \frac{\lambda}{n}\right)^n. \end{aligned}$$

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

3.2. Định nghĩa phân phối Poisson

Biến ngẫu nhiên X được gọi là có phân phối Poisson với tham số $\lambda > 0$, ký hiệu là $X \in P(\lambda)$ hay $X \sim P(\lambda)$, nếu $X = \{0, 1, 2, \dots, n, \dots\}$ với xác suất

$$p_k = P(X = k) = \frac{e^{-\lambda} \cdot \lambda^k}{k!} \quad (k = 0, 1, \dots, n, \dots)$$

Trong đó, λ là trung bình số lần xuất hiện biến cố ta quan tâm trong một khoảng xác định (khoảng thời gian hoặc một khoảng đơn vị tính nào đó).

VD. Quan sát tại một sân bay thấy trung bình 16 phút có 2 máy bay hạ cánh. Suy ra trong 1 giờ trung bình có $\lambda = \frac{60 \cdot 2}{16} = 7,5$ máy bay hạ cánh.

VD. Trung bình cứ 100 sinh viên thi hết môn XSTK có 71 sinh viên thi đạt. Suy ra 120 sinh viên thi hết môn XSTK thì trung bình có 85,2 sinh viên thi đạt.

3.3. Các số đặc trưng của $X \sim P(\lambda)$

$$EX = VarX = \lambda$$

$$\text{Mod } X = x_0 \in \mathbb{N}: \lambda - 1 \leq x_0 \leq \lambda$$

VD 1. Quan sát tại siêu thị A thấy trung bình 5 phút có 18 khách đến mua hàng.

- 1) Tính xác suất để trong 7 phút có 25 khách đến siêu thị A ?
- 2) Tính xác suất để trong 2 phút có từ 3 đến 5 khách đến siêu thị A ?
- 3) Tính số khách chắc chắn nhất sẽ đến siêu thị A trong 1 giờ ?

VD 2. Quan sát thấy trung bình 2 phút có 6 ô tô đi qua trạm thu phí. Biết xác suất có ít nhất 1 ô tô đi qua trạm thu phí trong t phút bằng 0,9. Giá trị của t (phút) là:

- A. 0,9082 B. 0,8591 C. 0,8514 D. 0,7675.

VD 3*. Cứ mỗi lần đi câu cá thì ông A chọn ngẫu nhiên 1 trong 2 nơi để câu. Nếu đi câu ở địa điểm I thì trung bình cứ 10 lần móc mỗi, ông A câu được 2 con cá; câu ở địa điểm II thì trung bình cứ 12 lần móc mỗi, ông A câu được 3 con cá. Hôm nay ông A đi câu, ông đã móc mỗi 20 lần và câu được 5 con cá. Tính xác suất ông A câu được 5 con cá ở địa điểm II ?

VD 4*. Tại một xưởng dệt, trung bình dệt 10 m vải loại B thì bị lỗi 13 chỗ. Chọn lần lượt 5 xấp vải loại B của xưởng, mỗi xấp dài 6 m. Tính xác suất để 3 trong 5 xấp vải ấy, mỗi xấp vải có đúng 7 chỗ bị lỗi ?

VD 5*. Quan sát thấy trung bình 1 ngày (24 giờ) có 12 chuyến tàu vào cảng A. Chọn ngẫu nhiên 6 giờ trong 1 ngày. Tính xác suất để 2 trong 6 giờ ấy, mỗi giờ có đúng 1 tàu vào cảng A ?

BÀI 4. PHÂN PHỐI CHUẨN

4.1. Phân phối chuẩn

4.1.1. Định nghĩa

Biến ngẫu nhiên liên tục X được gọi là có phân phối chuẩn (*Normal distribution*) với hai tham số μ và σ^2 ($\sigma > 0$), ký hiệu là $X \in N(\mu; \sigma^2)$ hay $X \sim N(\mu; \sigma^2)$, nếu hàm mật độ xác suất của X có dạng

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

4.1.2. Các số đặc trưng của $X \sim N(\mu, \sigma^2)$

$$\text{Mod } X = EX = \mu; \quad \text{Var} X = \sigma^2$$

4.1.3. Xác suất của $X \sim N(\mu, \sigma^2)$

$$P(a \leq X \leq b) = \int_a^b f(x)dx = \frac{1}{\sigma\sqrt{2\pi}} \int_a^b e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

Nhận xét. Đổi biến $z = \frac{x-\mu}{\sigma}$, ta có: $\frac{1}{\sigma\sqrt{2\pi}} \int_a^b e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \int_{\frac{a-\mu}{\sigma}}^{\frac{b-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz$.

4.2. Phân phối chuẩn tắc

4.2.1. Định nghĩa

BNN Z có phân phối chuẩn với hai tham số $\mu = 0$ và $\sigma^2 = 1$ được gọi là có *phân phối chuẩn tắc*, ký hiệu là $Z \in N(0; 1)$ hay $Z \sim N(0; 1)$.

Hàm mật độ xác suất của Z là $f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$, $z \in \mathbb{R}$

(Giá trị của hàm $f(z)$ được cho trong bảng phụ lục A).

4.2.2. Xác suất của $Z \sim N(0; 1)$

• **Hàm Laplace.** Hàm số $\varphi(x) = \int_0^x f(z)dz$ được gọi là hàm Laplace.

(Giá trị của hàm $\varphi(x)$ được cho trong bảng phụ lục B).

• Tính chất của hàm Laplace

- Hàm $\varphi(x)$ đồng biến trên \mathbb{R} ;
- $\varphi(-x) = -\varphi(x)$ (hàm $\varphi(x)$ lẻ);
- $\varphi(-\infty) = -0,5$; $\varphi(+\infty) = 0,5$.

• Công thức tính xác suất

$$P(\alpha \leq Z \leq \beta) = \int_{\alpha}^{\beta} f(z)dz = \varphi(\beta) - \varphi(\alpha)$$

Chú ý

- $P(Z < \beta) = 0,5 + \varphi(\beta)$; $P(Z > \alpha) = 0,5 - \varphi(\alpha)$.
- Nếu $x \geq 4$ thì $\varphi(x) \approx 0,5$.
- Nếu $X \sim N(\mu; \sigma^2)$ thì $Z = \frac{X - \mu}{\sigma} \sim N(0; 1)$.

Vậy, công thức tính xác suất của phân phối chuẩn là

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

VD 1. Thời gian X (tháng) đạt chuẩn chiều cao của loại cây giống A tại một vườn ươm là biến ngẫu nhiên có phân phối $N(8; 3)$. Tỷ lệ (xác suất) đạt chuẩn chiều cao của loại cây giống A tại vườn ươm này trong khoảng từ 6 tháng đến 8,2 tháng là:

- A. 27,65% B. 31,15% C. 42,27% D. 45,78%.

VD 2. 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 tỷ 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 3. 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. 0,2144; C. 0,1313; D. 0,1060.

VD 4. Cho BNN X có phân phối chuẩn với $EX = 10$, $P(10 < X < 20) = 0,3$. Tính $P(0 < X \leq 15)$?

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

1) Tính xác suất khách phải chờ từ 3,5 phút đến 5 phút ?

2) Tính thời gian tối thiểu t nếu xác suất khách phải chờ vượt quá t là không quá 5% ?

VD 6*. Tuổi thọ của 1 loại máy lạnh A là BNN X (năm) có phân phối $N(10; 6,25)$. Khi bán 1 máy lạnh A thì lãi được 1,4 triệu đồng nhưng nếu máy lạnh phải bảo hành thì lỗ 3,8 triệu đồng. Vậy để có tiền lãi trung bình khi bán mỗi máy lạnh loại này là 1 triệu đồng thì cần phải quy định thời gian bảo hành là bao nhiêu ?

BÀI 5. PHÂN PHỐI XÁC SUẤT CỦA VECTOR NGẪU NHIÊN RỜI RẠC

5.1 Bảng phân phối xác suất đồng thời của (X, Y)

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

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

5.2. Phân phối xác suất thành phần (phân phối lề)

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	\dots	x_m
P	$p_{1\bullet}$	$p_{2\bullet}$	\dots	$p_{m\bullet}$

trong đó $p_{i\bullet} = p_{i1} + p_{i2} + \dots + p_{in}$ (tổng dòng i của bảng phân phối xác suất đồng thời).

Kỳ vọng của X là

$$EX = x_1 p_{1\bullet} + x_2 p_{2\bullet} + \dots + x_m p_{m\bullet}$$

• Bảng phân phối xác suất của Y

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

trong đó $p_{\bullet j} = p_{1j} + p_{2j} + \dots + p_{mj}$ (tổng cột j của bảng phân phối xác suất đồng thời).

Kỳ vọng của Y là

$$EY = y_1 p_{\bullet 1} + y_2 p_{\bullet 2} + \dots + y_n p_{\bullet n}$$

VD 1. Phân phối xác suất đồng thời của vector ngẫu nhiên (X, Y) cho bởi bảng:

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

- Tính $P(X = 6)$ và $P(X \geq 7, Y \geq 2)$.
- Lập bảng phân phối xs thành phần và tính EX , EY .

Giải

- $P(X = 6) = 0,1 + 0,05 + 0,15 = 0,3$.

$$P(X \geq 7, Y \geq 2) = P\{(7,2)\} + P\{(7,3)\} + P\{(8,2)\} + P\{(8,3)\} = 0,15 + 0,1 + 0,2 + 0,1 = 0,55.$$

2) Bảng phân phối của X là

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

$$EX = 6.0,3 + 7.0,3 + 8.0,4 = 7,1.$$

Bảng phân phối của Y là

Y	1	2	3
P	0,25	0,40	0,35

$$EY = 1.0,25 + 2.0,4 + 3.0,35 = 2,1.$$

5.3. Phân phối xác suất có điều kiện

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

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

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

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

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

Kỳ vọng của X với điều kiện $Y = y_j$ là:

$$EX = \frac{1}{p_{\bullet j}}(x_1 p_{1j} + x_2 p_{2j} + \dots + x_m p_{mj})$$

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

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

Kỳ vọng của Y với điều kiện $X = x_i$ là:

$$EY = \frac{1}{p_{i\bullet}}(y_1 p_{i1} + y_2 p_{i2} + \dots + y_n p_{in})$$

VD 2. Cho bảng phân phối xác suất đồng thời của (X, Y) :

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

1) Lập bảng phân phối xác suất của X với điều kiện $Y = 2$ và tính kỳ vọng của X .

2) Lập bảng phân phối xác suất của Y với điều kiện $X = 8$ và tính kỳ vọng của Y .

Giải. 1) Ta có:

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

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

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

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

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

$$EX = 6 \cdot \frac{1}{6} + 7 \cdot \frac{1}{2} + 8 \cdot \frac{1}{3} = \frac{43}{6}.$$

2) Bảng phân phối xác suất của Y với điều kiện $X = 8$:

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

$$EY = 1 \cdot 0,5 + 2 \cdot 0,25 + 3 \cdot 0,25 = 1,75.$$

VD 3. Cho vector ngẫu nhiên rời rạc (X, Y) có bảng phân phối xác suất 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 $P(X - Y = 1)$.

2) Tính xác suất $P(X > 0 | Y = 1)$.

3) Tính trung bình của X và Y .

4) Tính trung bình của Y khi $X = 1$.

Giải. 1) Ta có: $P(X - Y = 1) = P\{(1, 0)\} + P\{(2, 1)\} = \frac{4}{18} + \frac{1}{18} = \frac{5}{18}.$

2) $P(X > 0 | Y=1) = P(X=1 | Y=1) + P(X=2 | Y=1) = \frac{P\{(1, 1)\}}{P(Y=1)} + \frac{P\{(2, 1)\}}{P(Y=1)} = \frac{4}{7}.$

3) 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	$\frac{11}{18}$	$\frac{7}{18}$

$$\text{Vậy } EX = 0 \cdot \frac{4}{18} + 1 \cdot \frac{7}{18} + 2 \cdot \frac{7}{18} = \frac{21}{18} \text{ và } EY = \frac{7}{18}.$$

4) Bảng phân phối xác suất của Y khi $X = 1$ là:

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

$$\text{Vậy } EY = \frac{3}{7}.$$

VD 4. Chi phí quảng cáo X (triệu đồng) và doanh thu Y (triệu đồng) của một công ty có bảng phân phối xác suất đồng thời như sau:

$\begin{matrix} Y \\ X \end{matrix}$	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

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.

Chương 5. ĐỊNH LÝ GIỚI HẠN TRONG XÁC SUẤT

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

Bài 2. Các loại xấp xỉ phân phối xác suất

BÀI 1. MỘT SỐ LOẠI HỘI TỤ TRONG XÁC SUẤT VÀ CÁC ĐỊNH LÝ (tham khảo)

1.1. Hội tụ theo xác suất – Luật số lớn

1.1.1. Định nghĩa

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

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

Ký hiệu là $X_n \xrightarrow{P} X$ ($n \rightarrow \infty$).

- Dãy các biến ngẫu nhiên $\{X_i\}$ ($i = 1, \dots, n, \dots$) được gọi là tuân theo *luật số lớn* (dạng Tchébyshev) nếu

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

1.1.2. Định lý (Bất đẳng thức Tchébyshev)

Nếu biến ngẫu nhiên X có $EX = \mu$ và $VarX = \sigma^2$ thì

$$\forall \varepsilon > 0 : P\left(\left|X - \mu\right| \geq \varepsilon\right) \leq \frac{\sigma^2}{\varepsilon^2} \Leftrightarrow P\left(\left|X - \mu\right| < \varepsilon\right) \geq 1 - \frac{\sigma^2}{\varepsilon^2}.$$

Chứng minh

- Nếu X là biến ngẫu nhiên rời rạc, ta có:

$$\begin{aligned} \sigma^2 &= \sum_x (x - \mu)^2 f(x) = \sum_{|x - \mu| < \varepsilon} (x - \mu)^2 f(x) + \sum_{|x - \mu| \geq \varepsilon} (x - \mu)^2 f(x) \\ &\geq \sum_{|x - \mu| \geq \varepsilon} (x - \mu)^2 f(x) \geq \varepsilon^2 \sum_{|x - \mu| \geq \varepsilon} f(x) = \varepsilon^2 P\left(\left|X - \mu\right| \geq \varepsilon\right). \end{aligned}$$

- Nếu X là biến ngẫu nhiên liên tục, ta có:

$$\sigma^2 = \int_{-\infty}^{+\infty} (x - \mu)^2 f(x) dx = \int_{|x - \mu| < \varepsilon} (x - \mu)^2 f(x) dx + \int_{|x - \mu| \geq \varepsilon} (x - \mu)^2 f(x) dx$$

$$\geq \int_{|x-\mu| \geq \varepsilon} (x-\mu)^2 f(x) dx \geq \varepsilon^2 \int_{|x-\mu| \geq \varepsilon} f(x) dx = \varepsilon^2 P(|X-\mu| \geq \varepsilon).$$

$$\text{Vậy } \sigma^2 \geq \varepsilon^2 P(|X-\mu| \geq \varepsilon) \Leftrightarrow P(|X-\mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}. \blacksquare$$

Ý nghĩa của định lý

$\forall \varepsilon > 0$ cho trước, xác suất để X nhận giá trị trong khoảng $(\mu - \varepsilon; \mu + \varepsilon)$ ít nhất phải bằng $1 - \frac{\sigma^2}{\varepsilon^2}$.

1.1.3. Định lý luật số lớn Tchébyshev

• Định lý

Nếu dãy các BNN $\{X_i\}$ ($i = 1, \dots, n, \dots$) độc lập từng đôi có EX_i hữu hạn và $VarX_i \leq C$ (hằng số) thì

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

• Hệ quả

Nếu dãy các BNN $\{X_i\}$ ($i = 1, \dots, n, \dots$) 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 các BNN độc lập cùng phân phối và có phương sai hữu hạn.
- Để đo một 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

1.2.1. Định nghĩa

Dãy các biến ngẫu nhiên $\{X_i\}$ ($i = 1, \dots, n, \dots$) được gọi là *hội tụ yếu* hay *hội tụ theo phân phối* đến biến ngẫu nhiên X nếu $\lim_{n \rightarrow \infty} F_n(x) = F(x), \forall x \in C(F)$. 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 $X_n \xrightarrow{d} X$.

1.2.2. Định lý giới hạn trung tâm (định lý Liapounop)

Cho X_1, \dots, X_n là các BNN độc lập có cùng phân phối xs, với kỳ vọng μ và phương sai σ^2 hữu hạn. Nếu $S_n = X_1 + \dots + X_n$ thì $ES_n = n\mu$, $VarS_n = n\sigma^2$ và khi $n \rightarrow \infty$ thì $S_n \xrightarrow{d} X \sim N(n\mu; n\sigma^2)$.

• 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,...

BÀI 2. CÁC LOẠI XẤP XỈ PHÂN PHỐI XÁC SUẤT

2.1. Xấp xỉ phân phối Siêu bội bởi Nhị thức

Xét BNN X có phân phối Siêu bội $H(N; N_A; n)$. Nếu N khá lớn và n rất nhỏ so với N thì

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

Chú ý

Khi cỡ mẫu n khá nhỏ so với kích thước N (khoảng $5\%N$) của tổng thể thì việc lấy mẫu có hoàn lại hay không hoàn lại là như nhau.

VD 1. Trong kho, người ta đã để lẫn 500 sản phẩm loại B với 1500 sản phẩm loại A . Chọn ngẫu nhiên 40 sản phẩm từ kho này. Tính xác suất chọn được 30 sản phẩm loại A ?

.....

VD 2. 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 50 cây lan thì được 10 cây có hoa màu đỏ.

2) Có thể tính xác suất để khi chọn ngẫu nhiên 300 cây lan thì có 45 cây hoa màu đỏ được không ?

.....

2.2. Xấp xỉ phân phối Nhị thức bởi Poisson

Xét BNN X có phân phối Nhị thức $B(n; p)$. Nếu n đủ lớn và p gần bằng 0 (hoặc gần bằng 1) thì

$$X \sim P(\lambda), \quad \lambda = np$$

Chú ý. Xấp xỉ trên sẽ có hiệu quả khi $np < 5$ hay $nq < 5$.

VD 3. Một lô hàng thịt đông lạnh đóng gói nhập khẩu có chứa 3% bị nhiễm khuẩn. Tìm xác suất để khi chọn ngẫu nhiên 2.000 gói thịt từ lô hàng này có từ 40 đến 42 gói bị nhiễm khuẩn ?

.....

VD 4. Giải câu 2) trong VD 2.

.....

2.3. Xấp xỉ phân phối Nhị thức bởi phân phối Chuẩn

Xét biến ngẫu nhiên $X \in B(n; p)$. Nếu n đủ lớn, p không quá gần 0 và 1 thì

$$X \sim N(\mu; \sigma^2) \text{ với } \mu = np, \sigma^2 = npq.$$

Khi đó

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

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

$$P(k_1 \leq X \leq k_2) \approx \varphi\left(\frac{k_2 - \mu}{\sigma}\right) - \varphi\left(\frac{k_1 - \mu}{\sigma}\right)$$

(giá trị được cho trong bảng B với $\varphi(-x) = -\varphi(x)$).

Chú ý. Khi $k = \mu$, ta sử dụng công thức hiệu chỉnh

$$P(X = k) \approx P(k - 0,5 \leq X \leq k + 0,5)$$

VD 5. Trong một đợt thi tuyển công chức ở một thành phố có 1.000 người dự thi với tỉ lệ thi đạt là 80%.

Tính xác suất để:

1) có 172 người không đạt;

2) có khoảng 170 đến 180 người không đạt.

VD 6. Một kho chứa 10.000 sản phẩm trong đó có 2.000 sản phẩm không được kiểm tra chất lượng. Chọn ngẫu nhiên từ kho ra 400 sản phẩm. Tính xác suất để trong 400 sản phẩm đó:

1) có 80 sản phẩm không được kiểm tra;

2) có từ 70 đến 100 sản phẩm không được kiểm tra.

VD 7. Người ta đã phát ra 480 giấy mời dự hội nghị khách hàng. Biết rằng sức chứa của khán phòng là 400 khách và thường chỉ có 80% khách hàng đến dự. Tính xác suất để tất cả khách hàng đến dự đều có chỗ ngồi ?

VD 8. 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 9. 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 1.000 con từ 20.000 con cá da trơn đó. Tính xác suất khách hàng chọn được từ 182 đến 230 con cá tra ?

A. 0,8143;

B. 0,9133;

C. 0,9424;

D. 0,9765.

PHẦN II. LÝ THUYẾT THỐNG KÊ (Statistical theory)

Chương 5. MẪU THỐNG KÊ VÀ ƯỚC LƯỢNG THAM SỐ

Bài 1. Lý thuyết mẫu

Bài 2. Ước lượng khoảng

BÀI 1. LÝ THUYẾT MẪU

1.1. Tổng thể và Mẫu

- Tập hợp tất cả 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 (thường rất lớn).
- Từ tổng thể ta chọn ra n phần tử thì n phần tử đó được gọi là một *mẫu* có *kích thước* n (*cỡ mẫu*).
- Mẫu được chọn ngẫu nhiên một cách khách quan được gọi là *mẫu ngẫu nhiên*.
- Có hai cách lấy mẫu:
 - **Mẫu có hoàn lại:** phần tử vừa quan sát xong được trả lại cho tổng thể trước khi quan sát lần sau.
 - **Mẫu không hoàn lại:** phần tử vừa quan sát xong không được trả lại cho tổng thể.
 Khi mẫu có kích thước lớn thì ta không phân biệt mẫu có hoàn lại hay không hoàn lại.
- **Mẫu định tính** là mẫu mà ta chỉ quan tâm đến các phần tử của nó có tính chất A nào đó hay không.
- **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.
- Gọi X_1, X_2, \dots, X_n là những kết quả quan sát. Ta xem như đã quan sát n lần, mỗi lần ta được một biến ngẫu nhiên X_i ($i = 1, \dots, n$).

Do ta thường lấy mẫu trong tổng thể có rất nhiều phần tử nên X_1, X_2, \dots, X_n được xem là độc lập và có cùng phân phối xác suất.

1.2. Sắp xếp mẫu dựa vào số liệu thực nghiệm

1.2.1. Sắp xếp theo dạng bảng

VD 1. Kiểm tra ngẫu nhiên 50 sinh viên. Ta sắp xếp điểm số X thu được theo thứ tự tăng dần và số sinh viên n có điểm tương ứng vào bảng như sau

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

1.2.2. Sắp xếp theo dạng khoảng

VD 2. Đo chiều cao X (cm) của $n = 100$ thanh niên.

Vì chiều cao khác nhau nên để tiện việc sắp xếp, người ta chia chiều cao thành nhiều khoảng.

Các thanh niên có chiều cao trong cùng 1 khoảng được xem là cao như nhau. Khi đó, ta có bảng số liệu ở dạng khoảng như sau:

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

Khi cần tính toán, người ta chọn *số trung bình* của mỗi khoảng để đưa số liệu trên về dạng bảng:

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

Chú ý. Đối với trường hợp số liệu được cho dưới dạng liệt kê thì ta sắp xếp lại ở dạng bảng.

1.3. Các đặc trưng mẫu

Xét một mẫu ngẫu nhiên (X_1, X_2, \dots, X_n) , ta có các đặc trưng mẫu như sau

1.3.1. Trung bình mẫu

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

Để đơn giản, ta dùng ký hiệu $\bar{X} = \bar{X}_n$.

1.3.2. Phương sai mẫu

- Phương sai mẫu

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

- Phương sai mẫu hiệu chỉnh

$$S^2 = S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

- Trong tính toán cụ thể, ta sử dụng công thức

$$S^2 = \frac{n}{n-1} \left[\overline{(X^2)} - (\bar{X})^2 \right] = \frac{n}{n-1} \hat{S}^2$$

$$\text{trong đó } \overline{(X^2)} = \frac{1}{n} \sum_{i=1}^n X_i^2.$$

1.3.3. Tỷ lệ mẫu

Xét mẫu định tính với các biến ngẫu nhiên X_i ($i = 1, \dots, n$) có phân phối Bernoulli $B(1; p)$:

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

Nếu mẫu có m phần tử có tính chất A thì tỷ lệ mẫu là

$$F = F_n = \frac{X_1 + X_2 + \dots + X_n}{n} = \frac{m}{n}$$

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

Các đặc trưng mẫu \bar{X} , S^2 , F là các thống kê dùng để nghiên cứu các đặc trưng μ , σ^2 , p tương ứng của tổng thể. Từ luật số lớn ta có:

$$F \rightarrow p, \bar{X} \rightarrow \mu, S^2 \rightarrow \sigma^2 \text{ (theo xác suất).}$$

.....

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 $n = 5$:

12; 13; 11; 14; 11.

1.1. 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 fx500MS);

MODE → MODE → 1 (chọn SD đối với fx570MS).

– Nhập liên tục các số:

12 M+ 13 M+ 11 M+ 14 M+ 11 M+

• **Xuất kết quả:**

– SHIFT → 2 → 1 (\bar{x}) → = **12.2** (kết quả \bar{x} là *trung bình mẫu*).

– SHIFT → 2 → 2 (σn) → = **1.1662** (kết quả σn là *độ lệch chuẩn của mẫu \hat{s}*).

– SHIFT → 2 → 3 ($\sigma n - 1$) → = **1.3038** ($\sigma n - 1$ là *độ lệch chuẩn của mẫu có hiệu chỉnh s*).

1.2. 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 **4 (Stat)** → **2 (OFF-chế độ không tần số)**.

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

12= 13= 11= 14= 11= → AC

• **Xuất kết quả:**

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

– SHIFT → 1 → 5 (var) → 2 → = (\bar{x})

– SHIFT → 1 → 5 (var) → 3 → = ($\sigma n = \hat{s}$).

– SHIFT → 1 → 5 (var) → 4 → = ($\sigma n - 1 = s$).

2. Số liệu có tần số

VD 2. Cho mẫu có cỡ mẫu là $n = 9$ như sau:

X	12	11	15
n	3	2	4

2.1. 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 fx500MS);

MODE → MODE → 1 (chọn SD đối với fx570MS).

– Nhập các số:

12 → SHIFT → , → 3 → M+

11 → SHIFT → , → 2 → M+

15 → SHIFT → , → 4 → M+

• Xuất kết quả, ta làm như 1.1.

Đáp số: $\bar{x} = 13.1111$, $\hat{s} = 1.7285$, $s = 1.8333$.

2.1. 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 → 4 (Stat) → 1 (ON – chế độ có tần số)

– MODE → 3 (stat) → 1 (1-var)

– 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 1.2.

VD 3. Điều tra năng suất của 100 ha lúa trong vùng A, 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

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 của mẫu chưa hiệu chỉnh và độ lệch chuẩn 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

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

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

KHÁI NIỆM CHUNG VỀ ƯỚC LƯỢNG

• Ước lượng là phỏng đoán một giá trị chưa biết của tổng thể dựa vào quan sát trên mẫu lấy ra từ tổng thể đó. Thông thường, ta cần ước lượng về trung bình, tỉ lệ, phương sai, hệ số tương quan của tổng thể.

• Có hai hình thức ước lượng:

▪ **Ước lượng điểm:** kết quả cần ước lượng được cho bởi một trị số.

▪ **Ước lượng khoảng:** kết quả cần ước lượng được cho bởi một khoảng.

• Ước lượng điểm có ưu điểm là cho ta một giá trị cụ thể, có thể dùng để tính các kết quả khác, nhưng nhược điểm là không cho biết sai số của ước lượng. Ước lượng khoảng thì ngược lại.

BÀI 2. ƯỚC LƯỢNG KHOẢNG

Trong bài này, ta chỉ xét đến ước lượng trung bình, phương sai trong phân phối chuẩn $N(\mu; \sigma^2)$ và ước lượng tỉ lệ trong phân phối Bernoulli $B(1; p)$.

2.1. Định nghĩa

- Xét thống kê T ước lượng tham số θ về một đặc tính X nào đó của tổng thể, khoảng $(\theta_1; \theta_2)$ được gọi là *khoảng ước lượng* nếu với xác suất $1 - \alpha$ cho trước thì $P(\theta_1 < \theta < \theta_2) = 1 - \alpha$.
- Xác suất $1 - \alpha$ được gọi là *độ tin cậy* của ước lượng, $2\varepsilon = \theta_2 - \theta_1$ được gọi là *độ dài* của khoảng ước lượng và ε được gọi là *độ chính xác* của ước lượng.
- Bài toán đi tìm khoảng ước lượng $(\theta_1; \theta_2)$ cho θ được gọi là bài toán *ước lượng khoảng*.

2.2. Ước lượng khoảng cho trung bình tổng thể

Xét đặc tính X của tổng thể có trung bình μ chưa biết. Với độ tin cậy $1 - \alpha$ cho trước, ta đi tìm khoảng ước lượng $(\mu_1; \mu_2)$ cho μ thỏa $P(\mu_1 < \mu < \mu_2) = 1 - \alpha$.

Trong thực hành, ta có 4 trường hợp sau

- **Trường hợp 1.** Cỡ mẫu $n > 30$ và phương sai tổng thể σ^2 đã biết.

- Từ mẫu ta tính trung bình mẫu \bar{x} .
- Từ $1 - \alpha \Rightarrow \frac{1 - \alpha}{2} = \varphi(z_{\alpha/2}) \xrightarrow{\text{tra bảng B}} z_{\alpha/2}$.
- Khoảng ước lượng là $(\bar{x} - \varepsilon; \bar{x} + \varepsilon)$, trong đó

$$\varepsilon = z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

- **Trường hợp 2.** Cỡ mẫu $n > 30$ và phương sai tổng thể σ^2 chưa biết.

- Tính \bar{x} , $z_{\alpha/2}$ và s (độ lệch chuẩn mẫu đã hiệu chỉnh).
- Từ $1 - \alpha \Rightarrow \frac{1 - \alpha}{2} = \varphi(z_{\alpha/2}) \xrightarrow{\text{tra bảng B}} z_{\alpha/2}$.
- Khoảng ước lượng là $(\bar{x} - \varepsilon; \bar{x} + \varepsilon)$, trong đó

$$\varepsilon = z_{\alpha/2} \frac{s}{\sqrt{n}}$$

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

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

- **Trường hợp 3.** Cỡ mẫu $n \leq 30$, σ^2 đã biết và X có phân phối chuẩn.

Ta làm như trường hợp 1.

- **Trường hợp 4.** Cỡ mẫu $n \leq 30$, σ^2 chưa biết và X có phân phối chuẩn.

- Từ mẫu ta tính \bar{x} , s .
- Từ $1 - \alpha \Rightarrow \alpha \xrightarrow{\text{tra bảng } C} t_{\alpha/2}^{n-1}$ (nhớ giảm bậc thành $n - 1$ rồi mới tra bảng!)
- Khoảng ước lượng là $(\bar{x} - \varepsilon; \bar{x} + \varepsilon)$, trong đó

$$\varepsilon = t_{\alpha/2}^{n-1} \frac{s}{\sqrt{n}}$$

- **Sai số chuẩn (tham khảo)**

Nếu chúng ta chọn mẫu ngẫu nhiên N lần (N là số rất lớn), mỗi lần với n đối tượng thì chúng ta sẽ có N số trung bình. Độ lệch chuẩn của N số trung bình này được gọi là sai số chuẩn. Sai số chuẩn phản ánh độ dao động hay biến thiên của các số trung bình mẫu.

- **Ý nghĩa của ước lượng khoảng (tham khảo)**

Xét đặc tính X của tổng thể có trung bình μ chưa biết. Ta chọn mẫu ngẫu nhiên gồm n phần tử của tổng thể và tính được \bar{x} , s . Khi đó

- $1 - \alpha$ số phần tử của tổng thể có đặc tính X dao động trong khoảng từ $\bar{x} - z_{\alpha/2} \times s$ đến $\bar{x} + z_{\alpha/2} \times s$;
- số trung bình về đặc tính X của tất cả các phần tử của tổng thể dao động trong khoảng từ

$$\bar{x} - z_{\alpha/2} \frac{s}{\sqrt{n}} \text{ đến } \bar{x} + z_{\alpha/2} \frac{s}{\sqrt{n}} \text{ với xác suất là } 1 - \alpha.$$

- Độ tin cậy của một ước lượng khoảng là xác suất để khoảng ước lượng đó chứa tham số cần ước lượng.
- Độ lệch chuẩn phản ánh độ biến thiên của một số phần tử trong một tổng thể. Còn sai số chuẩn phản ánh độ dao động của các số trung bình chọn từ tổng thể.

VD 1. Lượng Vitamin có trong trái cây A là biến ngẫu nhiên X (mg) có độ lệch chuẩn 3,98 mg. Phân tích 250 trái cây A thì thu được lượng Vitamin trung bình là 20mg. Với độ tin cậy 95%, hãy ước lượng lượng Vitamin trung bình có trong mỗi trái cây A ?

.....

VD 2. Biết chiều cao của con người là biến ngẫu nhiên X (cm) có phân phối chuẩn $N(\mu; 100)$. Với độ tin cậy 95%, nếu muốn ước lượng chiều cao trung bình của dân số có độ chính xác không quá 1 cm thì phải cần đo ít nhất mấy người ?

.....

VD 3. Kiểm tra tuổi thọ (giờ) của 50 bóng đèn do nhà máy A sản xuất ra, người ta được bảng số liệu

Tuổi thọ	3.300	3.500	3.600	4.000
Số bóng đèn	10	20	12	8

- 1) Hãy ước lượng tuổi thọ trung bình của loại bóng đèn do nhà máy A sản xuất với độ tin cậy 97% ?
- 2) Dựa vào mẫu trên để ước lượng tuổi thọ trung bình của loại bóng đèn do nhà máy A sản xuất có độ chính xác 59,02 giờ thì đảm bảo độ tin cậy là bao nhiêu ?
- 3) Dựa vào mẫu trên, nếu muốn ước lượng tuổi thọ trung bình của loại bóng đèn do nhà máy A sản xuất có độ chính xác nhỏ hơn 40 giờ với độ tin cậy 98% thì cần phải kiểm tra tối thiểu bao nhiêu bóng đèn nữa ?

VD 4. Chiều cao của loại cây A là biến ngẫu nhiên có phân phối chuẩn. Người ta đo ngẫu nhiên 20 cây A thì thấy chiều cao trung bình 23,12 m và độ lệch chuẩn của mẫu chưa hiệu chỉnh là 1,25 m.

Tìm khoảng ước lượng chiều cao trung bình của loại cây A với độ tin cậy 95%?

VD 5*. Để nghiên cứu nhu cầu về loại hàng X ở phường A người ta tiến hành khảo sát 400 trong toàn bộ 2500 gia đình. Kết quả khảo sát là:

Nhu cầu (kg/tháng)	0,5	1,5	2,5	3,5	4,5	5,5	6,5	7,5
Số gia đình	10	35	86	132	78	31	18	10

- 1) Hãy ước lượng nhu cầu trung bình 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 ước lượng nhu cầu trung bình về loại hàng X của phường A với độ chính xác lớn hơn 3 tấn/năm và độ tin cậy 99% thì cần khảo sát tối đa bao nhiêu gia đình trong phường A ?

VD 6. 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 trung bình đường kính của trục máy với độ tin cậy 97% ?
- 2) Dựa vào mẫu trên để ước lượng trung bình đường kính của trục máy có độ chính xác 0,006cm thì đảm bảo độ tin cậy là bao nhiêu ?
- 3) Dựa vào mẫu trên, nếu muốn ước lượng trung bình đường kính của trục máy có độ chính xác lớn hơn 0,003cm với độ tin cậy 99% thì cần phải đo tối đa bao nhiêu trục máy nữa ?

VD 7*. Tiến hành khảo sát 420 trong tổng số 3.000 gia đình ở một phường 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 trung bình tổng khối lượng sản phẩm X do công ty A sản xuất được tiêu thụ ở phường này trong một tháng với độ tin cậy 95%?

- A. (5612,7kg; 6012,3kg); B. (5893,3kg; 6312,9kg);
C. (5307,3kg; 5763,9kg); D. (5210,4kg; 5643,5kg).

2.3. Ước lượng khoảng cho tỉ lệ tổng thể

- 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 ước lượng p là $(p_1; p_2)$ thỏa mãn $P(p_1 < p < p_2) = 1 - \alpha$.
- Từ cỡ mẫu n và số phần tử có tính chất A trong mẫu là m , ta tính được tỉ lệ mẫu $f = \frac{m}{n}$.
- Khoảng ước lượng cho p là $(f - \varepsilon; f + \varepsilon)$, trong đó

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

Nhận xét. Nếu $p \in (p_1; p_2)$ và $p = \frac{N_A}{N}$ thì

- 1) $N_A = p \cdot N \Rightarrow p_1 \cdot N < N_A < p_2 \cdot N$; 2) $N = \frac{N_A}{p} \Rightarrow \frac{N_A}{p_2} < N < \frac{N_A}{p_1}$.

VD 8. Một trại gà tây đang nuôi 250.000 con gà trống 22 tuần tuổi. Cân thử 160 con gà trống này thì thấy có 138 con đã đạt chuẩn (nặng hơn 12 kg). Với độ tin cậy 95%, hãy ước lượng số gà trống của trại đã đạt chuẩn ?

.....

.....

VD 9. Để ước lượng số cá có trong một hồ người ta bắt lên 10.000 con, đánh dấu rồi thả lại xuống hồ. Sau một thời gian, lại bắt lên 8.000 con cá thấy 564 con có đánh dấu. Với độ tin cậy 97%, hãy ước lượng tỉ lệ cá có đánh dấu và số cá có trong hồ ?

.....

.....

.....

.....

.....

.....

.....

.....

VD 10. Người ta chọn ngẫu nhiên 500 chiếc tivi trong một kho chứa TV thì thấy có 27 TV Sony.

- 1) Dựa vào mẫu trên, để ước lượng tỉ lệ TV Sony trong kho có độ chính xác là $\varepsilon = 1,77\%$ thì đảm bảo độ tin cậy của ước lượng là bao nhiêu?
 - 2) Dựa vào mẫu trên, nếu muốn có độ chính xác của ước lượng tỉ lệ TV Sony nhỏ hơn 0,01 với độ tin cậy 95% thì cần chọn thêm ít nhất bao nhiêu TV nữa?
-
-
-
-
-
-
-
-

VD 11. Lấy ngẫu nhiên 200 sản phẩm trong kho hàng A thì 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 $\varepsilon = 3,5\%$ thì đảm bảo độ tin cậy của ước lượng là bao nhiêu?
- 2) Nếu muốn độ chính xác ước lượng tỉ lệ phế phẩm nhỏ hơn 1% với độ tin cậy 93% thì cần kiểm tra thêm ít nhất bao nhiêu sản phẩm nữa?

VD 12. Thửa ruộng có năng suất lúa trên 5,5 (tấn/ha) là ruộng có năng suất cao. 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

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. 92%; B. 94%; C. 96%; D. 98%.
-

Chương 6. KIỂM ĐỊNH GIẢ THUYẾT THỐNG KÊ

(Tests of Statistical Hypothesis)

Bài 1. Khái niệm về kiểm định giả thuyết thống kê

Bài 2. Kiểm định so sánh đặc trưng với một số

Bài 3. Kiểm định so sánh hai đặc trưng

BÀI 1. KHÁI NIỆM VỀ KIỂM ĐỊNH GIẢ THUYẾT THỐNG KÊ

- Mô hình tổng quát của bài toán kiểm định là: ta nêu lên hai mệnh đề trái ngược nhau, một mệnh đề được gọi là *giả thuyết* H và mệnh đề còn lại được gọi là *ngịch thuyết* (hay *đối thuyết*) \bar{H} .
- Giải quyết một bài toán kiểm định là: bằng cách dựa vào quan sát mẫu, ta nêu lên một quy tắc hành động, ta chấp nhận giả thuyết H hay bác bỏ giả thuyết H .
- Khi ta chấp nhận giả thuyết H , nghĩa là ta tin rằng H đúng; khi bác bỏ H , nghĩa là ta tin H sai. Do chỉ dựa trên một mẫu quan sát ngẫu nhiên, nên ta không thể khẳng định chắc chắn điều gì cho tổng thể.
- Trong chương này, ta chỉ xét loại kiểm định tham số (so sánh đặc trưng với 1 số, so sánh hai đặc trưng của hai tổng thể).

1.1. Khái niệm chung

- Giả sử ta khảo sát tham số $\theta \in \Theta$ chưa biết về đặc tính X của tổng thể, với $\Theta = \Theta_0 \cup \Theta_1$ và $\Theta_0 \cap \Theta_1 = \emptyset$.
- Đặt giả thuyết $H_0 : \theta \in \Theta_0$ và đối thuyết $H_1 : \theta \in \Theta_1$. *Bài toán kiểm định giả thuyết* là đưa ra tiêu chuẩn khi nào thì chấp nhận H_0 và khi nào chấp nhận H_1 .

1.2. Các loại sai lầm trong kiểm định

Khi thực hiện kiểm định giả thuyết, ta dựa vào quan sát ngẫu nhiên một số trường hợp rồi suy rộng ra cho tổng thể. Sự suy rộng này có khi đúng, có khi sai. Thống kê học phân biệt 2 loại sai lầm sau

- Sai lầm loại I** là ta bác bỏ H_0 trong khi H_0 đúng và $P(\text{bác bỏ } H_0 \mid H_0 \text{ đúng}) = \alpha$.
- Sai lầm loại II** là ta bác bỏ H_1 trong khi H_1 đúng và $P(\text{bác bỏ } H_1 \mid H_1 \text{ đúng}) = \beta$.

▪ Mối liên hệ giữa hai loại sai lầm

- Khi thực hiện kiểm định, ta luôn muốn xác suất phạm phải sai lầm càng ít càng tốt. Tuy nhiên, nếu hạ thấp α thì β sẽ tăng lên và ngược lại. Trong thực tế, giữa hai loại sai lầm này, loại nào tác hại hơn thì ta nên tránh.
- Trong thống kê, người ta quy ước rằng sai lầm loại 1 tác hại hơn loại 2 nên cần tránh hơn. Do đó, ta chỉ xét các phép kiểm định có α không vượt quá một giá trị ấn định trước, thông thường là 1%; 3%; 5%;... Giá trị α còn được gọi là *mức ý nghĩa* của kiểm định.

1.3. Cơ sở lý thuyết của kiểm định

- Để giải quyết bài toán kiểm định, ta quan sát mẫu ngẫu nhiên X_1, \dots, X_n và đưa ra giả thuyết H .
- Từ mẫu trên, ta chọn thống kê $T = f(X_1, \dots, X_n; \theta_0)$ sao cho nếu khi H đúng thì phân phối xác suất của T hoàn toàn xác định.
- Với mức ý nghĩa α , ta tìm được khoảng ước lượng $[a; b]$ cho T ở độ tin cậy $1 - \alpha$.

Khi đó:

- nếu $t \in [a; b]$ thì ta chấp nhận giả thuyết H ;
- nếu $t \notin [a; b]$ thì ta bác bỏ giả thuyết H .
- Nếu đồ thị hàm mật độ T đối xứng qua trục xác suất thì ta chọn khoảng đối xứng $[-t_\alpha; t_\alpha]$, với:

$$P(T \leq -t_\alpha) = P(T \geq t_\alpha) = \frac{\alpha}{2}.$$

Vậy, khi xét nửa bên phải của trục xác suất thì ta được:

- nếu $t \leq t_\alpha$ thì ta chấp nhận giả thuyết H ;
- nếu $t > t_\alpha$ thì ta bác bỏ giả thuyết H .
- Nếu đồ thị hàm mật độ T không đối xứng thì ta chọn khoảng tin cậy $[0; C]$, với $P(T \geq C) = \alpha$.
 - Nếu $t \leq C$ thì ta chấp nhận giả thuyết H , và
 - nếu $t > C$ thì ta bác bỏ giả thuyết H .

BÀI 2. KIỂM ĐỊNH SO SÁNH ĐẶC TRƯNG CỦA TỔNG THỂ VỚI MỘT SỐ

2.1. Kiểm định so sánh trung bình với một số

Xét đặc tính X của tổng thể. Giả sử cần so sánh trung bình μ của tổng thể về đặc tính X với số μ_0 , ta đặt giả thuyết $H: \mu = \mu_0$. Ta có 4 trường hợp sau

- **Trường hợp 1.** Cỡ mẫu $n > 30$, σ^2 đã biết.

- Từ mức ý nghĩa $\alpha \Rightarrow \frac{1-\alpha}{2} = \varphi(z_{\alpha/2}) \xrightarrow{B} z_{\alpha/2}$.

- Tính giá trị thống kê $z = \frac{|\bar{x} - \mu_0|}{\frac{\sigma}{\sqrt{n}}}$.

- Nếu $z \leq z_{\alpha/2}$ thì ta chấp nhận H , nghĩa là $\mu = \mu_0$; nếu $z > z_{\alpha/2}$ thì ta bác bỏ H , $\mu \neq \mu_0$.

- **Trường hợp 2.** Cỡ mẫu $n > 30$, σ^2 chưa biết.

Ta làm như trường hợp 1 nhưng thay σ bởi s .

- **Trường hợp 3.** Cỡ mẫu $n \leq 30$, σ^2 đã biết và X có phân phối chuẩn.
Ta làm như trường hợp 1.

- **Trường hợp 4.** Cỡ mẫu $n \leq 30$, σ^2 chưa biết và X có phân phối chuẩn.

- Từ cỡ mẫu n và mức ý nghĩa $\alpha \xrightarrow{\text{tra bảng } C} t_{\alpha/2}^{n-1}$.

- Tính giá trị thống kê $t = \frac{|\bar{x} - \mu_0|}{\frac{s}{\sqrt{n}}}$.

- Nếu $t \leq t_{\alpha/2}^{n-1}$ thì ta chấp nhận giả thuyết H ; nếu $t > t_{\alpha/2}^{n-1}$ thì ta bác bỏ giả thuyết H .

Chú ý. Trong tất cả các trường hợp bác bỏ, ta so sánh \bar{x} và μ_0 :

- nếu $\bar{x} > \mu_0$ thì ta kết luận $\mu > \mu_0$;
- nếu $\bar{x} < \mu_0$ thì ta kết luận $\mu < \mu_0$.

VD 1. Sở Điện lực A báo cáo rằng: trung bình mỗi hộ hàng tháng phải trả 250 ngàn đồng tiền điện, với độ lệch chuẩn là 20 ngàn. Người ta khảo sát ngẫu nhiên 500 hộ thì tính được trung bình hàng tháng mỗi hộ trả 252 ngàn đồng tiền điện. Trong kiểm định giả thuyết H : “trung bình mỗi hộ phải trả hàng tháng là 250 ngàn đồng tiền điện” với mức ý nghĩa $\alpha = 1\%$, hãy cho biết giá trị thống kê và kết luận?

.....

.....

.....

.....

.....

.....

.....

VD 2. Nhà Giáo dục học B muốn nghiên cứu xem số giờ tự học trung bình hàng ngày của sinh viên có thay đổi không so với mức 1 giờ/ngày cách đây 10 năm. Ông B khảo sát ngẫu nhiên 120 sinh viên và tính được trung bình là 0,82 giờ/ngày với $s = 0,75$ giờ/ngày. Với mức ý nghĩa 3%, hãy cho biết kết luận của ông B ?

.....

.....

.....

.....

.....

.....

.....

VD 3. Trong một nhà máy gạo, trọng lượng đóng bao theo quy định của 1 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 trung bình mỗi bao gạo của nhà máy này là 50 kg” có giá trị thống kê z và kết luận là:

- $z = 1,7205$; chấp nhận H với mức ý nghĩa 6%.
- $z = 1,7205$; bác bỏ H , trọng lượng thực tế của bao gạo nhỏ hơn 50 kg với mức ý nghĩa 6%.
- $z = 1,9732$; chấp nhận H với mức ý nghĩa 4%.
- $z = 1,9732$; bác bỏ H , trọng lượng thực tế của bao gạo nhỏ hơn 50 kg với mức ý nghĩa 4%.

VD 4. Một công ty cho biết mức lương trung bình của 1 kỹ sư ở công ty là 5,7 triệu đồng/tháng với độ lệch chuẩn 0,5 triệu đồng/tháng. Kỹ sư A đã thăm dò 18 kỹ sư ở công ty này thì thấy lương trung bình là 5,45 triệu đồng/tháng. Kỹ sư A quyết định rằng: nếu mức lương trung bình thực sự bằng hay cao hơn mức lương công ty đưa ra thì nộp đơn xin làm. Với mức ý nghĩa 5%, cho biết kết luận của kỹ sư A ?

.....

.....

.....

.....

.....

.....

.....

VD 5. Người ta kiểm tra ngẫu nhiên 38 cửa hàng của công ty A và có bảng doanh thu trong 1 tháng là

X (triệu đồng/tháng)	200	220	240	260
Số cửa hàng	8	16	12	2

Kiểm định giả thuyết H : “doanh thu trung bình hàng tháng của mỗi cửa hàng công ty là 230 triệu đồng”, mức ý nghĩa tối đa để giả thuyết H được chấp nhận là:

- A. 3,4%; B. 4,2%; C. 5,6%; D. 7,8%.

VD 6. Điểm trung bình môn Toán của SV 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

Kiểm định giả thuyết H : “điểm trung bình môn Toán của sinh viên năm nay bằng năm trước”, mức ý nghĩa tối đa để H được chấp nhận là:

- A. 13,94%; B. 13,62%; C. 11,74%; D. 11,86%.

VD 7. Thời gian X (phút) giữa hai chuyến xe bus trong thành phố là biến ngẫu nhiên có phân phối chuẩn. Công ty xe bus nói rằng: trung bình cứ 5 phút lại có 1 chuyến xe bus. Người ta chọn ngẫu nhiên 8 thời điểm và ghi lại thời gian (phút) giữa hai chuyến xe bus là:

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

Với mức ý nghĩa 7%, hãy kiểm định lời nói trên ?

VD 8*. Chiều cao cây giống X 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 chiều cao trung bình của cây hơn 1 m thì đem ra trồng. Với mức ý nghĩa 5%, kiểm định giả thuyết H : “cây giống của vườn ươm cao trung bình 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 so sánh tỉ lệ với một số

Giả sử ta cần so sánh tỉ lệ p về tính chất A nào đó của tổng thể X với số p_0 , ta đặt giả thuyết $H: p = p_0$.

• Từ mức ý nghĩa $\alpha \Rightarrow \frac{1-\alpha}{2} = \varphi(z_{\alpha/2}) \xrightarrow{B} z_{\alpha/2}$.

• Tính tỉ lệ mẫu $f = \frac{m}{n}$ và giá trị thống kê $z = \frac{|f - p_0|}{\sqrt{\frac{p_0(1-p_0)}{n}}}$.

▪ Nếu $z \leq z_{\alpha/2}$ thì ta chấp nhận H , nghĩa là $p = p_0$.

▪ Nếu $z > z_{\alpha/2}$ thì ta bác bỏ H , nghĩa là $p \neq p_0$.

Khi đó: $f > p_0 \Rightarrow p > p_0$; $f < p_0 \Rightarrow p < p_0$.

VD 9. Một báo cáo cho biết có 58% người tiêu dùng Việt Nam quan tâm đến hàng Việt. Khảo sát ngẫu nhiên 1.000 người dân Việt Nam thấy có 612 người được hỏi là có quan tâm đến hàng Việt. Với mức ý nghĩa 5%, hãy kiểm định lại báo cáo trên ?

VD 10. Khảo sát ngẫu nhiên 400 sinh viên về mức độ nghiêm túc trong giờ học thì thấy 13 sinh viên thừa nhận có ngủ trong giờ học. Trong kiểm định giả thuyết $H: \text{"có 2\% sinh viên ngủ trong giờ học"}$, mức ý nghĩa tối đa là bao nhiêu để H được chấp nhận ?

VD 11*. Để kiểm tra một loại súng thể thao, người ta cho bắn 1.000 viên đạn vào một tấm bia thì có 670 viên trúng mục tiêu. Sau đó, người ta cải tiến kỹ thuật và kiểm tra lại thì thấy tỉ lệ trúng của súng lúc này

là 70%. Trong kiểm định giả thuyết H : “tỉ lệ bắn trúng của loại súng thể thao này trước cải tiến là 70%”, với mức ý nghĩa 3% có giá trị thống kê và kết luận là:

- A. $z = 2,0702$ và cải tiến kỹ thuật là tốt; B. $z = 2,0702$ và cải tiến kỹ thuật là chưa tốt;
C. $z = 2,0176$ và cải tiến kỹ thuật là tốt; D. $z = 2,0176$ và cải tiến kỹ thuật là chưa tốt.

VD 12. 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 để H được chấp nhận là:

- A. 7,86%; B. 6,48%; C. 5,24%; D. 4,32%.

BÀI TẬP TỔNG HỢP

Bài 1. Kiểm tra lượng kẹo X được bán ra hàng ngày tại một cửa hàng, có kết quả

X (kg)	25	30	35	40	45	50	55
Số ngày	9	23	27	30	25	20	5

- Dựa vào mẫu trên, nếu muốn ước lượng lượng kẹo trung bình được bán ra hàng ngày ở cửa hàng này có độ chính xác là 1,2309 kg thì đảm bảo độ tin cậy là bao nhiêu?
- Hãy ước lượng tỉ lệ những ngày cửa hàng bán được nhiều hơn 40 kg với độ tin cậy 90%?
- Bằng cách hạ giá bán, cửa hàng đã bán được lượng kẹo trung bình hàng ngày là 40,5 kg. Với mức ý nghĩa 5% hãy cho kết luận thực tế về việc hạ giá này?

Bài 2. Lượng chất đạm của gà ta thả vườn đạt chuẩn là cao hơn hay bằng 252g mỗi con. Khảo sát lượng chất đạm X (g) của loại gà này tại một nông trại, có kết quả

X (g)	248	250	252	254	256	258
Số con	5	17	31	45	19	3

- Ước lượng lượng chất đạm trung bình có trong mỗi con gà của nông trại trên với độ tin cậy 95%?
- Để có thể nói tỉ lệ gà đạt chuẩn về chất đạm tại nông trại trên là 87% thì mức ý nghĩa tối đa là bao nhiêu?
- Người ta thử nuôi nhốt loại gà này cho đến lúc gà có cùng trọng lượng với cách nuôi như trên thì thấy lượng đạm trung bình là 249g. Với mức ý nghĩa 5% hãy cho kết luận thực tế của cách nuôi nhốt?

BÀI 3. KIỂM ĐỊNH SO SÁNH HAI ĐẶC TRƯNG CỦA HAI TỔNG THỂ

3.1. So sánh hai trung bình của hai tổng thể

Xét hai đặc tính X, Y của hai tổng thể. Giả sử ta cần so sánh hai trung bình tương ứng là μ_x và μ_y , ta đặt giả thuyết $H: \mu_x = \mu_y$. Có 4 trường hợp và việc chấp nhận hay bác bỏ H ta đều làm như kiểm định so sánh trung bình với 1 số.

▪ **Trường hợp 1.** Cỡ mẫu $n_x > 30$, $n_y > 30$ và phương sai σ_x^2 , σ_y^2 đã biết.

$$\text{Ta tính thống kê } z = \frac{|\bar{x} - \bar{y}|}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}} \text{ và so sánh với } z_{\alpha/2}.$$

- **Trường hợp 2.** Cỡ mẫu $n_x, n_y > 30$ và σ_x^2, σ_y^2 chưa biết.

Ta thay σ_x^2, σ_y^2 trong trường hợp 1 bởi s_x^2, s_y^2 .

- **Trường hợp 3.** Cỡ mẫu $n_x, n_y \leq 30, \sigma_x^2, \sigma_y^2$ đã biết và X, Y có phân phối chuẩn.

Ta làm như trường hợp 1.

- **Trường hợp 4.** Cỡ mẫu $n_x, n_y \leq 30, \sigma_x^2, \sigma_y^2$ chưa biết và X, Y có phân phối chuẩn.

• Tính phương sai chung của hai 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 $t = \frac{|\bar{x} - \bar{y}|}{s \cdot \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}}$.

• Từ $\alpha \xrightarrow{\text{tra bảng } C} t_{\alpha/2}^{n_x + n_y - 2}$ và so sánh với t .

VD 1. Người ta tiến hành bón hai loại phân X, Y cho cây cà chua. Với 60 cây được bón phân X thì thu được trung bình 32,2 quả và độ lệch chuẩn của mẫu đã hiệu chỉnh là 8,5 quả; 72 cây được bón phân Y thu được trung bình 28,4 quả và độ lệch chuẩn của mẫu đã hiệu chỉnh là 9,3 quả. Với mức ý nghĩa 5%, hãy cho biết kết luận về hai loại phân bón trên ?

.....

.....

.....

.....

.....

.....

.....

VD 2. Để so sánh mức lương trung bình của nhân viên nữ X (USD/giờ) và nam Y (USD/giờ) ở một công ty đa quốc gia, người ta tiến hành khảo sát ngẫu nhiên 100 nữ và 75 nam thì có kết quả:

$$\bar{x} = 7,23, s_x = 1,64 \text{ và } \bar{y} = 8,06, s_y = 1,85.$$

Với mức ý nghĩa 3%, kiểm định giả thuyết H : “mức lương trung bình của nữ và nam ở công ty này là như nhau” có giá trị thống kê và kết luận là:

- A. $z = 4,0957$; mức lương TB của nữ, nam như nhau;
- B. $z = 4,0957$; mức lương TB của nữ thấp hơn nam;
- C. $z = 3,0819$; mức lương TB của nữ, nam như nhau;
- D. $z = 3,0819$; mức lương TB của nữ thấp hơn nam.

VD 3. Tuổi thọ (năm) của pin là biến ngẫu nhiên có phân phối chuẩn. Một công ty sản xuất thử nghiệm 10 chiếc pin loại X và 12 chiếc pin loại Y thì có kết quả:

$$\bar{x} = 4,8, s_x = 1,1 \text{ và } \bar{y} = 4,3, s_y = 0,3.$$

Với mức ý nghĩa 8%, ta có thể kết luận tuổi thọ trung bình của loại pin X cao hơn loại pin Y được không ?

.....

.....

VD 4. Tuổi thọ (tháng) của thiết bị là biến ngẫu nhiên có phân phối chuẩn. Người ta kiểm tra ngẫu nhiên tuổi thọ của 15 thiết bị loại A , có kết quả:

114; 78; 96; 137; 78; 103; 126; 86; 99; 114; 72; 104; 73; 86; 117.

Kiểm tra tuổi thọ 17 thiết bị loại B thấy có trung bình là 84 tháng và độ lệch chuẩn đã hiệu chỉnh là 19 tháng. Kiểm định giả thuyết H : “tuổi thọ trung bình của thiết bị loại A và B là như nhau với mức ý nghĩa 3%” có giá trị thống kê và kết luận là:

- A. $t = 2,1616$; tuổi thọ trung bình của 2 loại thiết bị như nhau;
- B. $t = 2,1616$; tuổi thọ trung bình của loại thiết bị A lớn hơn;
- C. $t = 2,4616$; tuổi thọ trung bình của 2 loại thiết bị như nhau;
- D. $t = 2,4616$; tuổi thọ trung bình của loại thiết bị A lớn hơn.

3.2. So sánh hai tỉ lệ của hai tổng thể

Xét hai tổng thể X, Y . Giả sử ta cần so sánh hai tỉ lệ tương ứng p_x và p_y về một tính chất A nào đó, ta đặt giả thuyết $H: p_x = p_y$. Việc chấp nhận hay bác bỏ H ta làm như kiểm định so sánh tỉ lệ với 1 số.

- Từ 2 mẫu ta tính $f_x = \frac{m_x}{n_x}$, $f_y = \frac{m_y}{n_y}$ và $p_0 = \frac{m_x + m_y}{n_x + n_y}$.

- Tính giá trị thống kê $z = \frac{|f_x - f_y|}{\sqrt{p_0 q_0 \left(\frac{1}{n_x} + \frac{1}{n_y} \right)}}$, rồi so sánh với $z_{\alpha/2}$.

• Kết luận:

- Nếu $z \leq z_{\alpha/2}$ thì ta chấp nhận $H \Rightarrow p_x = p_y$.
- Nếu $z > z_{\alpha/2}$ và $f_x < f_y$ thì ta bác bỏ $H \Rightarrow p_x < p_y$.
- Nếu $z > z_{\alpha/2}$ và $f_x > f_y$ thì ta bác bỏ $H \Rightarrow p_x > p_y$.

VD 5. 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 một tính chất A thì được $f_x = 0,27$ và $f_y = 0,3$. Với mức ý nghĩa 9%, hãy so sánh tỉ lệ p_x, p_y của hai tổng thể?

VD 6. 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 3%.

VD 7. Một công ty điện tử nghiên cứu thị trường về sở thích xem tivi của cư dân trong thành phố. Hỏi 400 người ở quận X thì 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. 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 để H được chấp nhận là:

A. 0,96%;

B. 2,84%;

C. 4,06%;

D. 6,14%.

VD 8. 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 5% có giá trị thống kê 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.

3.3. So sánh hai trung bình ở dạng vector (X, Y) (tham khảo)

- Đặt $d = Y - X$ và giả thuyết $H: \mu_d = 0$.

- Tính giá trị thống kê $t = \frac{|\bar{d}|}{s_d} \sqrt{n}$ (n là số cặp có trong mẫu).

- Tùy vào n và phương sai, ta xét các trường hợp giống như so sánh trung bình với 1 số.

VD 9. Giả sử người ta dùng thuốc A cho 10 người. Đo nhịp tim/phút trước và sau khi dùng thuốc của từng người, có bảng kết quả

Người	1	2	3	4	5	6	7	8	9	10
Trước: X	70	77	78	72	81	78	73	74	79	80
Sau: Y	76	75	78	77	85	81	76	74	85	80
d	6	-2	0	5	4	3	3	0	6	0

Với mức ý nghĩa 5%, thuốc A có làm thay đổi nhịp tim trước khi dùng so với sau khi dùng hay không?

Giải. Đặt $d = Y - X$ và giả thuyết $H: \mu_d = 0$. Do $n = 10$, phương sai chưa biết nên bài toán là TH4.

Từ bảng số liệu, ta tính được: $\bar{d} = 2,5$; $s_d = 2,8382$.

Mức ý nghĩa $\alpha = 0,05 \xrightarrow{c} t_{0,05}^9 = 2,262$. Thống kê $t = \frac{|\bar{d}|}{s_d} \sqrt{n} = \frac{2,5}{2,8382} \sqrt{10} = 2,7855$.

Vì $t > t_{0,05}^9$ nên ta bác bỏ H . Vậy thuốc A đã làm thay đổi nhịp tim.

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-----

MỘT SỐ BÀI TẬP TRẮC NGHIỆM XÁC SUẤT

I. XÁC SUẤT CỦA BIẾN CỐ

Câu 1. Có 3 sinh viên A , B và C cùng thi môn XSTK.

Gọi biến cố A_i : “có i sinh viên thi đỗ” ($i = 0, 1, 2, 3$); C : “sinh viên C thi đỗ”.

Biến cố A_1C là:

- A. Sinh viên C thi đỗ;
C. Có 1 sinh viên thi đỗ;
B. Chỉ có sinh viên C thi đỗ;
D. Sinh viên C thi không đỗ.

Câu 2. Có 3 sinh viên A , B và C cùng thi môn XSTK.

Gọi biến cố A_i : “có i sinh viên thi đỗ” ($i = 0, 1, 2, 3$); A : “sinh viên A thi đỗ”.

Biến cố $A_2\bar{A}$ là:

- A. Sinh viên A thi hỏng;
C. Có 2 sinh viên thi đỗ;
B. Chỉ có sinh viên A thi đỗ;
D. Chỉ có sinh viên A thi hỏng.

Câu 3. Có 3 sinh viên A , B và C cùng thi môn XSTK.

Gọi biến cố A_i : “có i sinh viên thi đỗ” ($i = 0, 1, 2, 3$); B : “sinh viên B thi đỗ”.

Biến cố $A_1\bar{B}$ là:

- A. Sinh viên B thi hỏng;
C. Sinh viên A hoặc C thi đỗ;
B. Chỉ có 1 sinh viên thi đỗ;
D. Chỉ có 1 sinh viên hoặc A hoặc C thi đỗ.

Câu 4. Có 3 sinh viên A , B và C cùng thi môn XSTK.

Gọi biến cố A_i : “có i sinh viên thi đỗ” ($i = 0, 1, 2, 3$); C : “sinh viên C thi đỗ”.

Biến cố $A_0\bar{C}$ là:

- A. Sinh viên C thi hỏng;
C. Có 2 sinh viên thi đỗ;
B. Chỉ có sinh viên C thi hỏng;
D. Cả 3 sinh viên thi hỏng.

Câu 5. Có 3 sinh viên A , B và C cùng thi môn XSTK.

Gọi biến cố A_i : “có i sinh viên thi đỗ” ($i = 0, 1, 2, 3$); B : “sinh viên B thi đỗ”.

Biến cố $\bar{A}_0\bar{B}$ là:

- A. Sinh viên B thi hỏng;
C. Sinh viên A hoặc C thi đỗ;
B. Có 2 sinh viên thi đỗ;
D. Sinh viên A và C thi đỗ.

Câu 6. Có 3 sinh viên A , B và C cùng thi môn XSTK.

Gọi biến cố A_i : “có i sinh viên thi đỗ” ($i = 0, 1, 2, 3$); B : “sinh viên B thi đỗ”.

Hãy chọn đáp án đúng ?

- A. $\bar{A}_0\bar{B} \subset \bar{A}_1\bar{B}$; B. $\bar{A}_1\bar{B} \subset \bar{A}_2$; C. $\bar{A}_0\bar{B} = A_1\bar{B}$; **D.** $\bar{A}_3B \subset \bar{A}_3$.

Câu 7. Có 3 sinh viên A_1 , A_2 , A_3 cùng thi môn XSTK.

Gọi biến cố A_i : “sinh viên A_i thi đỗ” ($i = 1, 2, 3$); H : “có sinh viên thi hỏng”.

Hãy chọn đáp án đúng ?

- A.** $A_1H = A_1\bar{A}_2\bar{A}_3 \cup A_1A_2\bar{A}_3 \cup A_1\bar{A}_2A_3$; B. $A_1H = \bar{A}_1\bar{A}_2\bar{A}_3 \cup A_1\bar{A}_2\bar{A}_3 \cup A_1A_2\bar{A}_3 \cup A_1\bar{A}_2A_3$;
C. $A_1H = A_1\bar{A}_2\bar{A}_3 \cup \bar{A}_1A_2\bar{A}_3 \cup \bar{A}_1\bar{A}_2A_3$; D. $A_1H = A_1A_2\bar{A}_3 \cup A_1\bar{A}_2A_3 \cup \bar{A}_1A_2A_3$.

Câu 8. Có 3 sinh viên A_1, A_2, A_3 cùng thi môn XSTK.

Gọi biến cố A_i : “sinh viên A_i thi đỗ” ($i = 1, 2, 3$); H : “2 sinh viên thi hỏng trong đó có A_1 ”.

Hãy chọn đáp án đúng ?

A. $A_1 \bar{A}_2 \bar{A}_3 \cup \bar{A}_1 A_2 \bar{A}_3 \cup \bar{A}_1 \bar{A}_2 A_3 \subset H$;

B. $H = \bar{A}_1 \bar{A}_2 \bar{A}_3 \cup \bar{A}_1 A_2 \bar{A}_3 \cup \bar{A}_1 \bar{A}_2 A_3$;

C. $H = A_1 \bar{A}_2 \bar{A}_3 \cup \bar{A}_1 A_2 \bar{A}_3 \cup \bar{A}_1 \bar{A}_2 A_3$;

☒ D. $H \subset A_1 \bar{A}_2 \bar{A}_3 \cup \bar{A}_1 A_2 \bar{A}_3 \cup \bar{A}_1 \bar{A}_2 A_3$.

Câu 9. Có 3 sinh viên A_1, A_2, A_3 cùng thi môn XSTK.

Gọi biến cố A_i : “sinh viên A_i thi đỗ” ($i = 1, 2, 3$); H : “có 1 sinh viên thi hỏng”.

Hãy chọn đáp án đúng ?

A. $P(A_1 \bar{A}_2 A_3 | H) \geq P(A_1 \bar{A}_2 | H)$;

☒ B. $P(A_1 \bar{A}_2 | H) = P(A_1 \bar{A}_2 A_3 | H)$;

C. $P(A_1 \bar{A}_2 | H) \geq P(A_1 \bar{A}_2 A_3 | H)$;

D. $A_1 H = A_1 A_2 \bar{A}_3 \cup A_1 \bar{A}_2 A_3 \cup \bar{A}_1 A_2 A_3$.

Câu 10. Có 3 sinh viên A_1, A_2, A_3 cùng thi môn XSTK.

Gọi biến cố A_i : “sinh viên A_i thi đỗ” ($i = 1, 2, 3$); H : “có 1 sinh viên thi hỏng”.

Hãy chọn đáp án đúng ?

A. $\bar{A}_1 = H$;

B. $\bar{A}_2 A_3 \subset H$;

☒ C. $\bar{A}_1 A_2 A_3 \subset H$;

D. $A_1 A_2 \bar{A}_3 = H$.

Câu 11. Một hộp đựng 10 quả cầu gồm: 2 quả màu đỏ, 3 quả vàng và 5 quả xanh. Chọn ngẫu nhiên từ hộp đó ra 4 quả cầu. Xác suất chọn được 1 quả màu đỏ, 1 quả vàng và 2 quả xanh là:

☒ A. 0,2857;

B. 0,1793;

C. 0,1097;

D. 0,0973.

Câu 12. Một hộp đựng 10 quả cầu gồm: 2 quả màu đỏ, 3 quả vàng và 5 quả xanh. Chọn ngẫu nhiên từ hộp đó ra 4 quả cầu. Xác suất chọn được 2 quả màu xanh là:

A. 0,2894;

☒ B. 0,4762;

C. 0,0952;

D. 0,0476.

Câu 13. Một hộp đựng 10 quả cầu gồm: 2 quả màu đỏ, 3 quả vàng và 5 quả xanh. Chọn ngẫu nhiên từ hộp đó ra 4 quả cầu thì thấy có 3 quả màu xanh. Xác suất chọn được 1 quả màu đỏ là:

☒ A. 40%;

B. 50%;

C. 60%;

D. 80%.

Câu 14. Một hộp đựng 10 quả cầu gồm: 2 quả màu đỏ, 3 quả vàng và 5 quả xanh. Chọn ngẫu nhiên từ hộp đó ra 4 quả cầu thì thấy có 2 quả màu xanh. Xác suất chọn được ít nhất 1 quả màu đỏ là:

A. 40%;

☒ B. 70%;

C. 26%;

D. 28%.

Câu 15. Một cầu thủ ném lần lượt 3 quả bóng vào rổ một cách độc lập với xác suất vào rổ tương ứng là 0,7; 0,8; 0,9. Biết rằng có 2 quả bóng vào rổ. Xác suất để quả bóng thứ nhất vào rổ là:

A. 0,5437;

B. 0,5473;

C. 0,4753;

☒ D. 0,4573.

Câu 16. Một cầu thủ ném lần lượt 3 quả bóng vào rổ một cách độc lập với xác suất vào rổ tương ứng là 0,7; 0,8; 0,9. Biết rằng quả bóng thứ nhất vào rổ. Xác suất để có 2 quả bóng vào rổ là:

A. 20%;

B. 24%;

☒ C. 26%;

D. 28%.

Câu 17. Một xạ thủ bắn lần lượt 2 viên đạn vào một con thú và con thú chỉ chết khi bị trúng 2 viên đạn. Xác suất viên đạn thứ nhất trúng con thú là 0,8. Nếu viên thứ nhất trúng con thú thì xác suất trúng của viên thứ hai là 0,7 và nếu trượt thì xác suất trúng của viên thứ hai là 0,1. Biết rằng con thú còn sống. Xác suất để viên thứ hai trúng con thú là:

A. 0,0714;

B. 0,0741;

☒ C. 0,0455;

D. 0,0271.

Câu 18. Một trung tâm Tai–Mũi–Họng có tỉ lệ bệnh nhân Tai, Mũi, Họng tương ứng là 25%, 40%, 35%; tỉ lệ bệnh nặng phải mổ tương ứng là 1%, 2%, 3%. Xác suất để chọn ngẫu nhiên được một bệnh nhân bị bệnh Mũi phải mổ từ trung tâm này là:

- ☒ A. 0,008; B. 0,021; C. 0,312; D. 0,381.

Câu 19. Một trung tâm Tai–Mũi–Họng có tỉ lệ bệnh nhân Tai, Mũi, Họng tương ứng là 25%, 40%, 35%; tỉ lệ bệnh nặng phải mổ tương ứng là 1%, 2%, 3%. Xác suất để chọn ngẫu nhiên được một bệnh nhân phải mổ từ trung tâm này là:

- A. 0,008; ☒ B. 0,021; C. 0,312; D. 0,381.

Câu 20. Một trung tâm Tai–Mũi–Họng có tỉ lệ bệnh nhân Tai, Mũi, Họng tương ứng là 25%, 40%, 35%; tỉ lệ bệnh nặng phải mổ tương ứng là 1%, 2%, 3%. Chọn ngẫu nhiên một bệnh nhân từ trung tâm này thì được người bị mổ. Xác suất để người được chọn bị bệnh Mũi là:

- A. 0,008; B. 0,021; C. 0,312; ☒ D. 0,381.

II. BIẾN NGẪU NHIÊN

Câu 1. Cho BNN rời rạc X có bảng phân phối xác suất:

X	-1	0	2	4	5
P	0,15	0,10	0,45	0,05	0,25

Giá trị của $P[(-1 < X \leq 2) \cup (X = 5)]$ là:

- A. 0,9; ☒ B. 0,8; C. 0,7; D. 0,6.

Câu 2. Cho BNN rời rạc X có bảng phân phối xác suất:

X	1	2	3	4
P	0,15	0,25	0,40	0,20

Giá trị kỳ vọng của X là:

- A. 2,6; B. 2,8; ☒ C. 2,65; D. 1,97.

Câu 3. Cho BNN rời rạc X có bảng phân phối xác suất:

X	1	2	3	4
P	0,15	0,25	0,40	0,20

Giá trị phương sai của X là:

- A. 5,3; B. 7,0225; C. 7,95; ☒ D. 0,9275.

Câu 4. Một kiện hàng có 6 sản phẩm tốt và 4 phế phẩm. Chọn ngẫu nhiên từ kiện hàng đó ra 2 sản phẩm.

Gọi X là số phế phẩm trong 2 sản phẩm chọn ra. Bảng phân phối xác suất của X là:

A)

X	0	1	2
P	$\frac{2}{15}$	$\frac{8}{15}$	$\frac{1}{3}$

C)

X	0	1	2
P	$\frac{1}{3}$	$\frac{7}{15}$	$\frac{3}{15}$

☒ B)

X	0	1	2
P	$\frac{1}{3}$	$\frac{8}{15}$	$\frac{2}{15}$

D)

X	0	1	2
P	$\frac{1}{3}$	$\frac{4}{15}$	$\frac{2}{5}$

Câu 5. Cho BNN rời rạc X có hàm phân phối xác suất:

$$F(x) = \begin{cases} 0 & \text{khi } x \leq 1 \\ 0,19 & \text{khi } 1 < x \leq 2 \\ 1 & \text{khi } 2 < x. \end{cases}$$

Bảng phân phối xác suất của X là:

A)

X	0	1	2
P	0	0,19	0,81

C)

X	1	2
P	0,29	0,71

B)

X	0	1	2
P	0,19	0,51	0,3

D)

X	1	2
P	0,19	0,81

Câu 6. Lô hàng I có 3 sản phẩm tốt và 2 phế phẩm, lô hàng II có 2 sản phẩm tốt và 2 phế phẩm. Chọn ngẫu nhiên từ lô hàng I ra 1 sản phẩm và bỏ vào lô hàng II, sau đó từ lô hàng II chọn ngẫu nhiên ra 2 sản phẩm. Gọi X là số sản phẩm tốt chọn được từ lô hàng II. Bảng phân phối xác suất của X là:

A)

X	0	1	2
P	$\frac{11}{50}$	$\frac{30}{50}$	$\frac{9}{50}$

C)

X	0	1	2
P	$\frac{9}{50}$	$\frac{30}{50}$	$\frac{11}{50}$

B)

X	0	1	2
P	$\frac{11}{50}$	$\frac{9}{50}$	$\frac{30}{50}$

D)

X	0	1	2
P	$\frac{9}{50}$	$\frac{11}{50}$	$\frac{30}{50}$

Câu 7. Kiện hàng I có 3 sản phẩm tốt và 2 phế phẩm, kiện hàng II có 2 sản phẩm tốt và 4 phế phẩm. Chọn ngẫu nhiên từ kiện hàng I ra 1 sản phẩm và từ kiện hàng II ra 1 sản phẩm. Gọi X là số phế phẩm chọn được. Hàm phân phối xác suất $F(x) = P(X < x)$ của X là:

$$A. F(x) = \begin{cases} 0, & x < 0 \\ \frac{1}{5}, & 0 \leq x < 1 \\ \frac{11}{15}, & 1 \leq x < 2 \\ 1, & 2 \leq x \end{cases}$$

$$B. F(x) = \begin{cases} 0, & x \leq 0 \\ \frac{1}{5}, & 0 < x \leq 1 \\ \frac{11}{15}, & 1 < x \leq 2 \\ 1, & 2 < x \end{cases}$$

$$C. F(x) = \begin{cases} 0, & x \leq 0 \\ \frac{1}{5}, & 0 < x \leq 1 \\ \frac{8}{15}, & 1 < x \leq 2 \\ 1, & 2 < x \end{cases}$$

$$D. F(x) = \begin{cases} 0, & x < 0 \\ \frac{1}{5}, & 0 \leq x < 1 \\ \frac{8}{15}, & 1 \leq x < 2 \\ 1, & 2 \leq x \end{cases}$$

Câu 8. Cho BNN liên tục X có hàm mật độ xác suất $f(x) = \begin{cases} \frac{2}{3}x, & x \in [-1; 2] \\ 0, & x \notin [-1; 2]. \end{cases}$

Hàm phân phối xác suất $F(x) = P(X < x)$ của X là:

$$A. F(x) = \begin{cases} 0 & khi & x < -1 \\ \frac{1}{3}(x^2 - 1) & khi & -1 \leq x \leq 2 \\ 1 & khi & 2 < x. \end{cases}$$

$$B. F(x) = \begin{cases} 0 & khi & x < -1 \\ \frac{1}{3}(x^2 + 1) & khi & -1 \leq x \leq 2 \\ 1 & khi & 2 < x. \end{cases}$$

$$C. F(x) = \begin{cases} 0 & khi & x < -1 \\ \frac{1}{3}x^2 & khi & -1 \leq x \leq 2 \\ 1 & khi & 2 < x. \end{cases}$$

$$D. F(x) = \begin{cases} 0 & khi & x \leq -1 \\ \frac{1}{3}x^2 & khi & -1 < x \leq 2 \\ 1 & khi & 2 < x. \end{cases}$$

Câu 9. Biến ngẫu nhiên X có hàm mật độ xác suất $f(x) = \begin{cases} \frac{3}{16}x^2, & x \in (-2; 2) \\ 0, & x \notin (-2; 2) \end{cases}$.

Giá trị của $P(\sqrt{2} < Y \leq \sqrt{5})$ với $Y = \sqrt{X^2 + 1}$ là:

- A. 0,3125; B. 0,4375; **C.** 0,875; D. 0,625.

Câu 10. Theo thống kê trung bình cứ 1.000 người dân ở độ tuổi 40 thì sau 1 năm có 996 người còn sống. Một công ty bảo hiểm nhân thọ bán bảo hiểm 1 năm cho những người ở độ tuổi này với giá 1,5 triệu đồng, nếu người mua bảo hiểm chết thì số tiền bồi thường là 300 triệu đồng. Giả sử công ty bán được 40.000 hợp đồng bảo hiểm loại này (mỗi hợp đồng ứng với 1 người mua bảo hiểm) trong 1 năm.

Hỏi trong 1 năm lợi nhuận trung bình thu được của công ty về loại bảo hiểm này là bao nhiêu ?

- A. 1,2 tỉ đồng; B. 1,5 tỉ đồng; **C.** 12 tỉ đồng; D. 15 tỉ đồng.

Câu 11. Theo thống kê trung bình cứ 1.000 người đi xe máy thì có 25 người bị tai nạn trong 1 năm. Một công ty bảo hiểm bán bảo hiểm loại này cho 20.000 người trong 1 năm với giá 98 ngàn đồng và mức chi trả khi bị tai nạn là 3 triệu đồng.

Hỏi trong 1 năm lợi nhuận trung bình thu được của công ty về loại bảo hiểm này là bao nhiêu ?

- A. 445 triệu đồng; B. 450 triệu đồng; C. 455 triệu đồng; **D.** 460 triệu đồng.

Câu 12. Một cửa hàng điện máy bán 1 chiếc máy lạnh A thì lời 850.000 đồng nhưng nếu chiếc máy lạnh đó phải bảo hành thì lỗ 1.000.000 đồng. Biết xác suất máy lạnh A phải bảo hành của cửa hàng là $p = 15\%$, tính mức lời trung bình khi bán 1 chiếc máy lạnh A ?

- A. 722.500 đồng; B. 675.500 đồng; C. 605.500 đồng; **D.** 572.500 đồng.

Câu 13. Một cửa hàng điện máy bán 1 chiếc tivi thì lời 500.000 đồng nhưng nếu chiếc tivi đó phải bảo hành thì lỗ 700.000 đồng. Tính xác suất tivi phải bảo hành của cửa hàng để mức lời trung bình khi bán 1 chiếc tivi là 356.000 đồng ?

- A. 10%; **B.** 12%; C. 15%; D. 23%.

Câu 14. Cho BNN liên tục X có hàm mật độ xác suất $f(x) = \begin{cases} a(3x - x^2), & 0 \leq x \leq 3 \\ 0, & x \notin [0; 3] \end{cases}$.

Giá trị trung bình của X là:

- A. $EX = 1,2$; B. $EX = 1,4$; **C.** $EX = 1,5$; D. $EX = 2,4$.

Câu 15. Cho BNN liên tục X có hàm mật độ xác suất $f(x) = \begin{cases} a(3x - x^2), & 0 \leq x \leq 3 \\ 0, & x \notin [0; 3] \end{cases}$.

Giá trị phương sai của X là:

- A. $VarX = 0,64$; B. $VarX = 1,5$; C. $VarX = 2,7$; **D.** $VarX = 0,45$.

Câu 16. Cho BNN liên tục X có hàm mật độ xác suất $f(x) = \begin{cases} a(3x - x^2), & 0 \leq x \leq 3 \\ 0, & x \notin [0; 3] \end{cases}$.

Giá trị trung bình của Y với $Y = 3X^2$ là:

- A.** $EY = 8,1$; B. $EY = 7,9$; C. $EY = 4,5$; D. $EY = 5,4$.

Câu 17. Cho BNN liên tục X có hàm mật độ xác suất $f(x) = \begin{cases} a(3x - x^2), & 0 \leq x \leq 3 \\ 0, & x \notin [0; 3] \end{cases}$.

Giá trị phương sai của Y với $Y = 3X^2$ là:

- A. $VarY = 38,0329$; **B.** $VarY = 38,5329$; C. $VarY = 38,9672$; D. $VarY = 39,0075$.

Câu 18. Cho BNN liên tục X có hàm mật độ xác suất $f(x) = \begin{cases} a(3x - x^2), & 0 \leq x \leq 3 \\ 0, & x \notin [0; 3] \end{cases}$.

Giá trị của $ModX$ là:

- ☒ A. $ModX = 1,5$; B. $ModX = 0$; C. $ModX = 1$; D. $ModX = 3$.

Câu 19. Cho BNN liên tục X có hàm mật độ xác suất $f(x) = \begin{cases} a(3x - x^2), & 0 \leq x \leq 3 \\ 0, & x \notin [0; 3] \end{cases}$.

Giá trị của xác suất $p = P(1 < X \leq 2)$ là:

- ☒ A. $p = 0,4815$; B. $p = 0,4915$; C. $p = 0,5015$; D. $p = 0,5115$.

Câu 20. BNN liên tục X có hàm phân phối xác suất $F(x) = \begin{cases} 0, & x \leq 1 \\ \frac{x-1}{2}, & 1 < x \leq 3 \\ 1, & 3 < x. \end{cases}$

Giá trị phương sai của X là:

- A. $VarX = \frac{1}{4}$; B. $VarX = \frac{1}{6}$; C. $VarX = \frac{1}{2}$; ☒ D. $VarX = \frac{1}{3}$.

III. PHÂN PHỐI XÁC SUẤT THÔNG DỤNG

Câu 1. Một thùng bia có 24 chai trong đó để lần 3 chai quá hạn sử dụng. Chọn ngẫu nhiên từ thùng đó ra 4 chai bia. Xác suất chọn phải ít nhất 1 chai bia quá hạn sử dụng là:

- A. 0,4123; B. 0,5868; ☒ C. 0,4368; D. 0,5632.

Câu 2. Chủ vườn lan đã để nhâm 10 chậu lan có hoa màu đỏ với 10 chậu lan có hoa màu tím (lan chưa nở hoa). Một khách hàng chọn ngẫu nhiên 7 chậu từ 20 chậu lan đó. Xác suất khách chọn được nhiều hơn 5 chậu lan có hoa màu đỏ là:

- A. 0,0586; B. 0,0486; C. 0,0386; ☒ D. 0,0286.

Câu 3. Chủ vườn lan đã để nhâm 20 chậu lan có hoa màu đỏ với 100 chậu lan có hoa màu tím (lan chưa nở hoa). Một khách hàng chọn ngẫu nhiên 15 chậu từ 120 chậu lan đó. Gọi X là số chậu lan có hoa màu tím khách chọn được. Giá trị của EX và $VarX$ là:

- A. $EX = 3, VarX = \frac{36}{17}$; B. $EX = \frac{25}{2}, VarX = \frac{135}{68}$;
☒ C. $EX = \frac{25}{2}, VarX = \frac{125}{68}$; D. $EX = \frac{5}{2}, VarX = \frac{125}{68}$.

Câu 4. Một hiệu sách bán 40 cuốn truyện A , trong đó có 12 cuốn in lậu. Một khách hàng chọn ngẫu nhiên 4 cuốn truyện A . Hỏi khả năng cao nhất khách chọn được bao nhiêu cuốn truyện A không phải in lậu?

- A. 1 cuốn; B. 2 cuốn; ☒ C. 3 cuốn; D. 4 cuốn.

Câu 5. Một hộp chứa 100 viên phấn trong đó có 10 viên màu đỏ. Hỏi nếu không nhìn vào hộp bốc tùy ý 1 lần bao nhiêu viên để xác suất có 4 viên màu đỏ là 0,0272?

- A. 10 viên; B. 12 viên; ☒ C. 14 viên; D. 16 viên.

Câu 6. Xác suất có bệnh của những người chờ khám bệnh tại 1 bệnh viện là 12%. Khám lần lượt 20 người này, xác suất có ít hơn 2 người bị bệnh là:

- ☒ A. 0,2891; B. 0,7109; C. 0,3891; D. 0,6109.

Câu 7. Xác suất có bệnh của những người chờ khám bệnh tại 1 bệnh viện là 72%. Khám lần lượt 61 người này, hỏi khả năng cao nhất có mấy người bị bệnh?

- A. 41 người; B. 42 người; C. 43 người; ☒ D. 44 người.

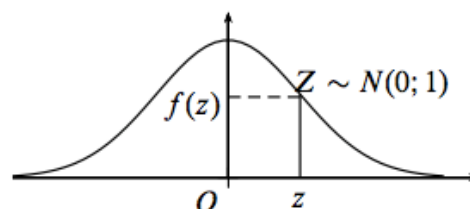
Câu 8. Một gia đình nuôi gà mái đẻ với xác suất đẻ trứng của mỗi con gà trong 1 ngày là 0,75. Để trung bình mỗi ngày có nhiều hơn 122 con gà mái đẻ trứng thì số gà tối thiểu gia đình đó phải nuôi là:

- A. 151 con; B. 162 con; ☒ C. 163 con; D. 175 con.

- Câu 9.** Trong một đợt xổ số người ta phát hành 100.000 vé trong đó có 10.000 vé trúng thưởng. Hỏi 1 người muốn trúng ít nhất 1 vé với xác suất lớn hơn 95% thì cần phải mua tối thiểu bao nhiêu vé ?
 A. 2 vé; B. 12 vé; C. 27 vé; **D. 29 vé.**
- Câu 10.** Một trạm điện thoại trung bình nhận được 900 cuộc gọi trong 1 giờ. Xác suất để trạm nhận được đúng 32 cuộc gọi trong 2 phút là:
A. 0,0659; B. 0,0481; C. 0,0963; D. 0,0624.
- Câu 11.** Tại bệnh viện A trung bình 3 giờ có 8 ca mổ. Hỏi số ca mổ chắc chắn nhất sẽ xảy ra tại bệnh viện A trong 10 giờ là bao nhiêu ?
 A. 25 ca; **B.** 26 ca; C. 27 ca; D. 28 ca.
- Câu 12.** Một bến xe khách trung bình có 70 xe xuất bến trong 1 giờ. Xác suất để trong 5 phút có từ 4 đến 6 xe xuất bến là:
 A. 0,2133; B. 0,2792; C. 0,3209; **D.** 0,4663.
- Câu 13.** Cho biến ngẫu nhiên $X \in N(4; 2,25)$. Giá trị của xác suất $P(X > 5,5)$ là:
A. 0,1587; B. 0,3413; C. 0,1916; D. 0,2707.
- Câu 14.** Thống kê điểm thi X (điểm) môn XSTK của sinh viên tại trường Đại học A cho thấy X là biến ngẫu nhiên với $X \in N(5,25; 1,25)$. Tỷ lệ sinh viên có điểm thi môn XSTK của trường A từ 4 đến 6 điểm là:
 A. 56,71%; B. 68,72%; C. 64,72%; **D.** 61,72%.
- Câu 15.** Thời gian X (tháng) từ lúc vay đến lúc trả tiền của 1 khách hàng tại ngân hàng A là biến ngẫu nhiên có phân phối $N(18; 16)$. Tính tỷ lệ khách hàng trả tiền cho ngân hàng A trong khoảng từ 12 đến 16 tháng ?
A. 24,17%; B. 9,63%; C. 25,17%; D. 10,63%.
- Câu 16.** Chiều cao của nam giới đã trưởng thành là biến ngẫu nhiên X (cm) có phân phối $N(165; 25)$. Tỷ lệ nam giới đã trưởng thành cao từ 1,65m đến 1,75m là:
 A. 1,6%; B. 42,75%; C. 45,96%; **D.** 47,73%.
- Câu 17.** Một lô hàng thịt đông lạnh đóng gói nhập khẩu với tỷ lệ bị nhiễm khuẩn là 1,6%. Kiểm tra lần lượt ngẫu nhiên 2000 gói thịt từ lô hàng này. Tính xác suất có đúng 36 gói thịt bị nhiễm khuẩn ?
 A. 0,1522; B. 0,2522; C. 0,0922; **D.** 0,0522.
- Câu 18.** Trong một kho lúa giống có tỷ lệ hạt lúa lai tạp là 2%. Tính xác suất sao cho khi chọn lần lượt 1000 hạt lúa giống trong kho thì có từ 17 đến 19 hạt lúa lai tạp ?
A. 0,2492; B. 0,3492; C. 0,0942; D. 0,0342.
- Câu 19.** Một khách sạn nhận đặt chỗ của 585 khách hàng cho 500 phòng vào ngày 2/9 vì theo kinh nghiệm của những năm trước cho thấy có 15% khách đặt chỗ nhưng không đến. Biết mỗi khách đặt 1 phòng, tính xác suất có từ 494 đến 499 khách đặt chỗ và đến nhận phòng vào ngày 2/9 ?
 A. 0,0273; B. 0,1273; **C.** 0,2273; D. 0,3273.
- Câu 20.** Tỷ lệ thanh niên đã tốt nghiệp THPT của quận A là 75%. Trong đợt tuyển quân đi nghĩa vụ quân sự năm nay, quận A đã gọi ngẫu nhiên 325 thanh niên. Tính xác suất để có từ 80 đến 84 thanh niên bị loại do chưa tốt nghiệp THPT ?
 A. 13,79%; **B.** 20,04%; C. 26,32%; D. 28,69%.

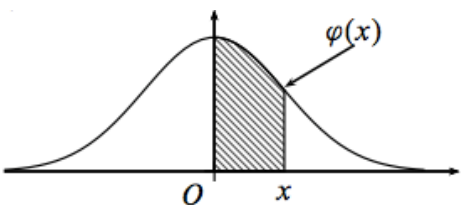
.....Hết.....

Bảng A. Giá trị $f(z)$

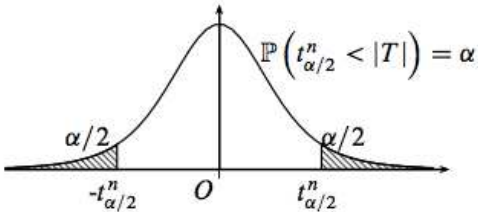


z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,3989	0,3989	0,3989	0,3988	0,3986	0,3984	0,3982	0,3980	0,3977	0,3973
0,1	0,3970	0,3965	0,3961	0,3956	0,3951	0,3945	0,3939	0,3932	0,3925	0,3918
0,2	0,3910	0,3902	0,3894	0,3885	0,3876	0,3867	0,3857	0,3847	0,3836	0,3825
0,3	0,3814	0,3802	0,3790	0,3778	0,3765	0,3752	0,3739	0,3725	0,3712	0,3697
0,4	0,3683	0,3668	0,3653	0,3637	0,3621	0,3605	0,3589	0,3572	0,3555	0,3538
0,5	0,3521	0,3503	0,3485	0,3467	0,3448	0,3429	0,3410	0,3391	0,3372	0,3355
0,6	0,3332	0,3312	0,3292	0,3271	0,3251	0,3230	0,3209	0,3187	0,3166	0,3144
0,7	0,3123	0,3101	0,3079	0,3056	0,3034	0,3011	0,2989	0,2966	0,2943	0,2920
0,8	0,2897	0,2874	0,2850	0,2827	0,2803	0,2780	0,2756	0,2732	0,2709	0,2685
0,9	0,2661	0,2637	0,2613	0,2589	0,2565	0,2541	0,2516	0,2492	0,2468	0,2444
1,0	0,2420	0,2396	0,2371	0,2347	0,2323	0,2299	0,2275	0,2251	0,2227	0,2203
1,1	0,2179	0,2155	0,2131	0,2107	0,2083	0,2059	0,2036	0,2012	0,1989	0,1965
1,2	0,1942	0,1919	0,1895	0,1872	0,1849	0,1826	0,1804	0,1781	0,1758	0,1736
1,3	0,1714	0,1691	0,1669	0,1647	0,1626	0,1604	0,1582	0,1561	0,1539	0,1518
1,4	0,1497	0,1476	0,1456	0,1435	0,1415	0,1394	0,1374	0,1354	0,1334	0,1315
1,5	0,1295	0,1276	0,1257	0,1238	0,1219	0,1200	0,1182	0,1163	0,1145	0,1127
1,6	0,1109	0,1092	0,1074	0,1057	0,1040	0,1023	0,1006	0,0989	0,0973	0,0957
1,7	0,0940	0,0925	0,0909	0,0893	0,0878	0,0863	0,0848	0,0833	0,0818	0,0804
1,8	0,0790	0,0775	0,0761	0,0748	0,0734	0,0721	0,0707	0,0694	0,0681	0,0669
1,9	0,0656	0,0644	0,0632	0,0620	0,0608	0,0596	0,0584	0,0573	0,0562	0,0551
2,0	0,0540	0,0529	0,0519	0,0508	0,0498	0,0488	0,0478	0,0468	0,0459	0,0449
2,1	0,0440	0,0431	0,0422	0,0413	0,0404	0,0396	0,0387	0,0379	0,0371	0,0363
2,2	0,0355	0,0347	0,0339	0,0332	0,0325	0,0317	0,0310	0,0303	0,0297	0,0290
2,3	0,0283	0,0277	0,0270	0,0264	0,0258	0,0252	0,0246	0,0241	0,0235	0,0229
2,4	0,0224	0,0219	0,0213	0,0208	0,0203	0,0198	0,0194	0,0189	0,0184	0,0180
2,5	0,0175	0,0171	0,0167	0,0163	0,0158	0,0154	0,0151	0,0147	0,0143	0,0139
2,6	0,0136	0,0132	0,0129	0,0126	0,0122	0,0119	0,0116	0,0113	0,0110	0,0107
2,7	0,0104	0,0101	0,0099	0,0096	0,0093	0,0091	0,0088	0,0086	0,0084	0,0081
2,8	0,0079	0,0077	0,0075	0,0073	0,0071	0,0069	0,0067	0,0065	0,0063	0,0061
2,9	0,0060	0,0058	0,0056	0,0055	0,0053	0,0051	0,0050	0,0048	0,0047	0,0046
3,0	0,0044	0,0043	0,0042	0,0040	0,0039	0,0038	0,0037	0,0036	0,0035	0,0034
3,1	0,0033	0,0032	0,0031	0,0030	0,0029	0,0028	0,0027	0,0026	0,0025	0,0025
3,2	0,0024	0,0023	0,0022	0,0022	0,0021	0,0020	0,0020	0,0019	0,0018	0,0018
3,3	0,0017	0,0017	0,0016	0,0016	0,0015	0,0015	0,0014	0,0014	0,0013	0,0013
3,4	0,0012	0,0012	0,0012	0,0011	0,0011	0,0010	0,0010	0,0010	0,0009	0,0009
3,5	0,0009	0,0008	0,0008	0,0008	0,0008	0,0007	0,0007	0,0007	0,0007	0,0006
3,6	0,0006	0,0006	0,0006	0,0005	0,0005	0,0005	0,0005	0,0005	0,0005	0,0004
3,7	0,0004	0,0004	0,0004	0,0004	0,0004	0,0004	0,0003	0,0003	0,0003	0,0003
3,8	0,0003	0,0003	0,0003	0,0003	0,0003	0,0002	0,0002	0,0002	0,0002	0,0002
3,9	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0001	0,0001

Bảng B. Giá trị

[illegible]

Bảng C. Giá trị phân phối Student n bậc tự do $\mathbb{P}(-t_{\alpha/2}^n < T < t_{\alpha/2}^n) = \alpha$



$\alpha / 2 \backslash n$	0,005	0,010	0,015	0,020	0,025	0,030	0,035	0,040	0,045	0,050	0,055	0,060	0,065	0,070	0,075
1	63,657	31,821	21,205	15,895	12,706	10,579	9,058	7,916	7,026	6,314	5,730	5,242	4,829	4,474	4,165
2	9,925	6,965	5,643	4,849	4,303	3,896	3,578	3,320	3,104	2,920	2,760	2,620	2,495	2,383	2,282
3	5,841	4,541	3,896	3,482	3,182	2,951	2,763	2,605	2,471	2,353	2,249	2,156	2,072	1,995	1,924
4	4,604	3,747	3,298	2,999	2,776	2,601	2,456	2,333	2,226	2,132	2,048	1,971	1,902	1,838	1,778
5	4,032	3,365	3,003	2,757	2,571	2,422	2,297	2,191	2,098	2,015	1,941	1,873	1,810	1,753	1,699
6	3,707	3,143	2,829	2,612	2,447	2,313	2,201	2,104	2,019	1,943	1,874	1,812	1,754	1,700	1,650
7	3,499	2,998	2,715	2,517	2,365	2,241	2,136	2,046	1,966	1,895	1,830	1,770	1,715	1,664	1,617
8	3,355	2,896	2,634	2,449	2,306	2,189	2,090	2,004	1,928	1,860	1,797	1,740	1,687	1,638	1,592
9	3,250	2,821	2,574	2,398	2,262	2,150	2,055	1,973	1,899	1,833	1,773	1,718	1,666	1,619	1,574
10	3,169	2,764	2,527	2,359	2,228	2,120	2,028	1,948	1,877	1,812	1,754	1,700	1,650	1,603	1,559
11	3,106	2,718	2,491	2,328	2,201	2,096	2,007	1,928	1,859	1,796	1,738	1,686	1,636	1,591	1,548
12	3,055	2,681	2,461	2,303	2,179	2,076	1,989	1,912	1,844	1,782	1,726	1,674	1,626	1,580	1,538
13	3,012	2,650	2,436	2,282	2,160	2,060	1,974	1,899	1,832	1,771	1,715	1,664	1,616	1,572	1,530
14	2,977	2,624	2,415	2,264	2,145	2,046	1,962	1,887	1,821	1,761	1,706	1,656	1,609	1,565	1,523
15	2,947	2,602	2,397	2,249	2,131	2,034	1,951	1,878	1,812	1,753	1,699	1,649	1,602	1,558	1,517
16	2,921	2,583	2,382	2,235	2,120	2,024	1,942	1,869	1,805	1,746	1,692	1,642	1,596	1,553	1,512
17	2,898	2,567	2,368	2,224	2,110	2,015	1,934	1,862	1,798	1,740	1,686	1,637	1,591	1,548	1,508
18	2,878	2,552	2,356	2,214	2,101	2,007	1,926	1,855	1,792	1,734	1,681	1,632	1,587	1,544	1,504
19	2,861	2,539	2,346	2,205	2,093	2,000	1,920	1,850	1,786	1,729	1,677	1,628	1,583	1,540	1,500
20	2,845	2,528	2,336	2,197	2,086	1,994	1,914	1,844	1,782	1,725	1,672	1,624	1,579	1,537	1,497
21	2,831	2,518	2,328	2,189	2,080	1,988	1,909	1,840	1,777	1,721	1,669	1,621	1,576	1,534	1,494
22	2,819	2,508	2,320	2,183	2,074	1,983	1,905	1,835	1,773	1,717	1,665	1,618	1,573	1,531	1,492
23	2,807	2,500	2,313	2,177	2,069	1,978	1,900	1,832	1,770	1,714	1,662	1,615	1,570	1,529	1,489
24	2,797	2,492	2,307	2,172	2,064	1,974	1,896	1,828	1,767	1,711	1,660	1,612	1,568	1,526	1,487
25	2,787	2,485	2,301	2,167	2,060	1,970	1,893	1,825	1,764	1,708	1,657	1,610	1,566	1,524	1,485
26	2,779	2,479	2,296	2,162	2,056	1,967	1,890	1,822	1,761	1,706	1,655	1,608	1,564	1,522	1,483
27	2,771	2,473	2,291	2,158	2,052	1,963	1,887	1,819	1,758	1,703	1,653	1,606	1,562	1,521	1,482
28	2,763	2,467	2,286	2,154	2,048	1,960	1,884	1,817	1,756	1,701	1,651	1,604	1,560	1,519	1,480

Bảng C. (tiếp theo)

$\alpha / 2$ n	0,005	0,010	0,015	0,020	0,025	0,030	0,035	0,040	0,045	0,050	0,055	0,060	0,065	0,070	0,075
29	2,756	2,462	2,282	2,150	2,045	1,957	1,881	1,814	1,754	1,699	1,649	1,602	1,558	1,517	1,479
30	2,750	2,457	2,278	2,147	2,042	1,955	1,879	1,812	1,752	1,697	1,647	1,600	1,557	1,516	1,477
40	2,704	2,423	2,250	2,123	2,021	1,936	1,862	1,796	1,737	1,684	1,635	1,589	1,546	1,506	1,468
50	2,678	2,403	2,234	2,109	2,009	1,924	1,852	1,787	1,729	1,676	1,627	1,582	1,539	1,500	1,462
60	2,660	2,390	2,223	2,099	2,000	1,917	1,845	1,781	1,723	1,671	1,622	1,577	1,535	1,496	1,458
70	2,648	2,381	2,215	2,093	1,994	1,912	1,840	1,776	1,719	1,667	1,619	1,574	1,532	1,493	1,456
80	2,639	2,374	2,209	2,088	1,990	1,908	1,836	1,773	1,716	1,664	1,616	1,572	1,530	1,491	1,453
90	2,632	2,368	2,205	2,084	1,987	1,905	1,834	1,771	1,714	1,662	1,614	1,570	1,528	1,489	1,452
100	2,626	2,364	2,201	2,081	1,984	1,902	1,832	1,769	1,712	1,660	1,613	1,568	1,527	1,488	1,451
1000	2,581	2,330	2,173	2,056	1,962	1,883	1,814	1,752	1,697	1,646	1,600	1,556	1,515	1,477	1,441