XÁC SUẤT & THỐNG KÊ ĐẠI HỌC

PHÂN PHỐI CHƯƠNG TRÌNH
Số tiết: 45

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
Chương 4. Vector ngẫu nhiên
Chương 5. Định lý giới hạn trong Xác suất

PHẦN II. LÝ THUYẾT THỐNG KÊ
(Statistical theory)

Chương 6. Mẫu thống kê và Ước lượng tham số
Chương 7. Kiểm định Giả thuyết Thống kê
Chương 8. Bài toán Tương quan và Hồi quy

Tài liệu tham khảo

1. Nguyễn Phú Vinh – *Giáo trình Xác suất – Thống kê và Ứng dụng* – NXB Thống kê.
2. 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 Kiều – *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. F.M. Dekking – *A modern introduction to Probability and Statistics* – Springer Publication (2005).

Vào năm **1651,** Blaise Pascal nhận được bức thư của nhà quý tộc Pháp, De Méré, nhờ ông giải quyết các rắc rối nảy sinh trong trò chơi đánh bạc. Pascal đã toán học hoá các trò chơi đánh bạc này, nâng lên thành những bài toán phức tạp hơn và trao đổi với nhà toán học Fermat. Những cuộc trao đổi đó đã nảy sinh ra Lý thuyết Xác suất – Lý thuyết toán học về các hiện tượng ngẫu nhiên.

Blaise Pascal

Pierre de Fermat

* James BERNOULLI là người phát minh ra **Luật Số Lớn**. Chính vì lý do đó, ngày nay Hội Xác Suất Thống Kê Thế Giới mang tên BERNOULLI

* Leibnitz có nhiều đóng góp quan trọng trong việc xây dựng Lý thuyết Xác suất

James BERNOULLI

Gottfried Wilhelm Leibniz

PHẦN I. LÝ THUYẾT XÁC SUẤT
(Probability theory)

Chương 1. XÁC SUẤT CỦA BIẾN CỐ

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

---

§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 100^0C 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 một điều kiện vẫn có thể sẽ cho ra các kết quả khác nhau được gọi là những hiện tượng ngẫu nhiên.

Chẳng hạn, gieo một 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 của 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*** (events).

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:

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

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

Các phần tử:

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

là các biến cố sơ cấp.

Các tập con của Ω:

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

là các biến cố.

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 $\varnothing$.

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

a) 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. Quan sát 4 con gà mái đẻ trứng trong 1 ngày. Gọi

A_i: "có i con gà mái đẻ trứng trong 1 ngày", $i = \overline{0,\ 4}$.

A: "có 3 hoặc 4 con gà mái đẻ trứng trong 1 ngày".

B: "có nhiều hơn 2 con gà mái đẻ trứng trong 1 ngày".

Khi đó, ta có: $A_3 \subset B$, $A_2 \not\subset B$, $B \subset A$ và $A = B$.

b) 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 $A \cup B$ 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 $A \cap B$ hay AB.

VD 4. Một người thợ săn bắn hai viên đạn vào một con thú và con thú sẽ chết nếu nó bị trúng cả hai 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 $A = N_1K_2 \cup K_1N_2$.

c) Biến cố đối lập

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

Vậy ta có: $\boxed{\overline{A} = \Omega \setminus A.}$

VD 6. Từ 1 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", $i = 9, 10, 11, 12$.
Ta có không gian mẫu là:
$$\Omega = A_9 \cup A_{10} \cup A_{11} \cup A_{12},$$
và $\overline{A}_{10} = \Omega \setminus A_{10} = A_9 \cup A_{11} \cup A_{12}$.

1.4. Hệ đầy đủ các biến cố

a) Hai biến cố xung khắc

Hai biến cố A và B được gọi là ***xung khắc với nhau*** trong một 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
Trong VD 7, A và B xung khắc nhưng không đối lập.

b) 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}, $i_0 \in \{1; 2; \ldots; n\}$ của họ xảy ra. Nghĩa là:

1) $A_i \cap A_j = \varnothing, \forall\ i \neq j$ và 2) $A_1 \cup A_2 \cup \ldots \cup A_n = \Omega$.

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 $\{A;\ \overline{A}\}$ là đầy đủ với A tùy ý.

……………………………………………………

§2. XÁC SUẤT CỦA BIẾN CỐ

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.

2.1. Định nghĩa xác suất dạng cổ điển

Xét một phép thử với không gian mẫu $\Omega = \{\omega_1; ...; \omega_n\}$ 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 là:

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

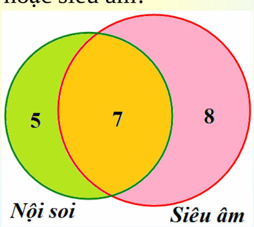
VD 1. Một công ty cần tuyển hai nhân viên. Có 4 người nữ và 2 người nam nộp đơn ngẫu nhiên (khả năng trúng tuyển của 6 người là như nhau). Tính xác suất để:

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

VD 2. Từ một hộp chứa 6 sản phẩm tốt và 4 phế phẩm người ta chọn ngẫu nhiên ra 5 sản phẩm.
Tính xác suất để có:
1) cả 5 sản phẩm đều tốt; 2) đúng 2 phế phẩm.

VD 3. Tại một bệnh viện có 50 người đang chờ kết quả khám bệnh. Trong đó có 12 người chờ kết quả nội soi, 15 người chờ kết quả siêu âm, 7 người chờ kết quả cả nội soi và siêu âm. Gọi tên ngẫu nhiên một người trong 50 người này, hãy tính xác suất gọi được người đang chờ kết quả nội soi hoặc siêu âm?

Biểu đồ Ven



2.2. Định nghĩa xác suất dạng thống kê

- Nếu khi thực hiện một phép thử nào đó n lần, thấy có k lần biến cố A xuất hiện thì tỉ số $\frac{k}{n}$ được gọi là ***tần suất*** của biến cố A.
- Khi n thay đổi, tần suất cũng thay đổi theo nhưng luôn dao động quanh một số cố định $p = \lim\limits_{n \to \infty} \frac{k}{n}$.
- Số p cố định này được gọi là ***xác suất*** của biến cố A theo nghĩa thống kê.

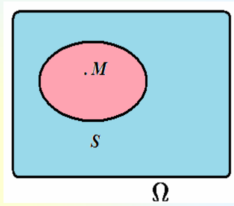
Trong thực tế, khi n đủ lớn thì $P(A) \approx \frac{k}{n}$.

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.

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

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



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

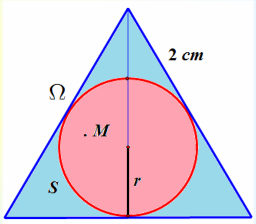
$$\boxed{P(A) = \frac{\text{độ đo } S}{\text{độ đo } \Omega}.}$$

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 là:

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



Bán kính của hình tròn là:

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

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

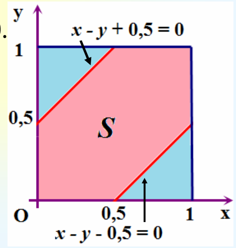
VD 6. Hai người bạn hẹn gặp nhau tại 1 địa điểm xác định trong khoảng từ 7h đến 8h. Mỗi người đến (và chắc chắn đến) điểm hẹn một cách độc lập, nếu không gặp người kia thì đợi 30 phút hoặc đến 8 giờ thì không đợi nữa. Tìm xác suất để hai người gặp nhau.

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:

$0 \le x \le 1,\ 0 \le y \le 1.$

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



Từ điều kiện, ta có:

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

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

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

$\{0 \le x \le 1, 0 \le y \le 1, x - y - 0,5 \le 0, x - y + 0,5 \ge 0\}.$

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

2.4. Tính chất của xác suất

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

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

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

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

………………………………………………………………

§3. CÔNG THỨC TÍNH XÁC SUẤT

3.1. Công thức cộng xác suất

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

- Nếu A và B là hai biến cố tùy ý:

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

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

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

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

$$\boxed{P\left(A_1 \cup A_2 \cup \ldots \cup A_n\right) = P(A_1) + P(A_2) + \ldots + P(A_n).}$$

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 một nhà đầu tư trong nhóm. Tìm xác suất để người đó gặp được nhà đầu tư vàng hoặc chứng khoán?

Đặc biệt

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

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

Chú ý

$$\boxed{\overline{A \cap B} = \overline{A} \cup \overline{B};\ \overline{A \cup B} = \overline{A} \cap \overline{B}.}$$

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?

3.2. XÁC SUẤT CÓ ĐIỀU KIỆN

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

A: "người A thi đỗ", B: "người B thi đỗ",
C: "người C thi đỗ", H: "có 2 người thi đỗ".

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

Ta có:

$$A = \{ABC, A\overline{B}C, AB\overline{C}, A\overline{B}\overline{C}\} \Rightarrow P(A) = \frac{4}{8};$$

$$H = \{\overline{A}BC, A\overline{B}C, AB\overline{C}\} \Rightarrow P(H) = \frac{3}{8}.$$

Lúc này, biến cố: "2 người thi đỗ trong đó có A" là:

$$AH = \{A\overline{B}C, AB\overline{C}\} \text{ và } P(AH) = \frac{2}{8}.$$

• 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: $P\left(A|H\right) = \frac{2}{3} = \frac{P(AH)}{P(H)}$.

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

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

VD 4. Một nhóm 10 sinh viên gồm 3 nam và 7 nữ trong đó có 2 nam 18 tuổi và 3 nữ 18 tuổi. Chọn ngẫu nhiên 1 sinh viên từ nhóm đó.

Gọi A: "sinh viên được chọn là nữ",
B: "sinh viên được chọn là 18 tuổi".

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

Nhận xét

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

Tính chất

1) $0 \le P\left(A|B\right) \le 1,\ \forall A \subset \Omega$;

2) nếu $A \subset C$ thì $P\left(A|B\right) \le P\left(C|B\right)$;

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

3.2.2. Công thức nhân xác suất

a) 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:

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

b) Công thức nhân

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

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

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

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

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

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

VD 5. 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 6. 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% và 80%. Tính xác suất sinh viên này thi đỗ?

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.

a) Công thức xác suất đầy đủ

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

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

VD 10. Một cửa hàng bán hai loại bóng đèn cùng kích cỡ gồm: 70 bóng màu trắng với tỉ lệ bóng hỏng là 1% và 30 bóng màu vàng với tỉ lệ hỏng 2%. Một khách hàng chọn mua ngẫu nhiên 1 bóng đèn từ cửa hàng này. Tính xác suất để người này mua được bóng đèn tốt ?

VD 11. Chuồng thỏ 1 có 3 con thỏ trắng và 4 con thỏ đen; chuồng 2 có 5 thỏ trắng và 3 thỏ đen. Quan sát thấy có 1 con thỏ chạy từ chuồng 1 sang chuồng 2, sau đó có 1 con thỏ chạy ra từ chuồng 2. Tính xác suất để con thỏ chạy ra từ chuồng 2 là thỏ trắng ?

b) Công thức Bayes

Xét họ n biến cố $\{A_i\}$ $(i = 1, 2, \ldots, n)$ đầy đủ và B là một 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:

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

VD 12. 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 ?

Phân biệt các bài toán áp dụng công thức Nhân – Đầy đủ – Bayes

Trong 1 bài toán, ta xét 3 biến cố A_1, A_2, B.

1) Nếu bài toán yêu cầu tìm xác suất của $A_1 \cap B$, $A_2 \cap B$ thì đây là bài toán công thức nhân. Xác suất là *xác suất tích của từng nhánh*.

2) Nếu bài toán yêu cầu tìm xác suất của B và $\{A_1, A_2\}$ đầy đủ thì đây là bài toán áp dụng công thức đầy đủ. Xác suất bằng *tổng 2 nhánh*.

Phân biệt các bài toán áp dụng công thức Nhân – Đầy đủ – Bayes

3) Nếu bài toán yêu cầu tìm xác suất của A_1, A_2 và cho biết B *đã xảy ra*, đồng thời hệ $\{A_1, A_2\}$ đầy đủ thì đây là bài toán áp dụng công thức Bayes. Xác suất là *tỉ số* giữa *nhánh cần tìm* với *tổng của hai nhánh*.

VD 13. 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% và 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 14. Tỉ lệ ôtô tải, ôtô con và xe máy đi qua đường X có trạm bơm dầu là 5 : 2 : 13. Xác suất để ôtô tải, ôtô con và xe máy đi qua đường này vào bơm dầu lần lượt là 0,1; 0,2 và 0,15. Biết rằng có 1 xe đi qua đường X vào bơm dầu, tính xác suất để đó là ôtô con ?

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

§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 1 số thực $X(\omega) \in \mathbb{R}$, thì X được gọi là một ***biến ngẫu nhiên***.

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

$$X : \Omega \to \mathbb{R}$$
$$\omega \mapsto X(\omega) = x.$$

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

VD 1. Người A mua một loại bảo hiểm tai nạn trong 1 năm với phí là 70 ngàn đồng. Nếu bị tai nạn thì công ty sẽ chi trả 3 triệu đồng. Gọi X là số tiền người A có được sau 1 năm mua bảo hiểm này. Khi đó, ta có

Phép thử là: "mua bảo hiểm tai nạn".
Biến cố là T: "người A bị tai nạn".
Không gian mẫu là $\Omega = \{T, \overline{T}\}$.

Vậy $X(T) = 2,93$ (triệu), $X(\overline{T}) = 0,07$ (triệu).

• Nếu $X(\Omega)$ là 1 tập hữu hạn $\{x_1, x_2, ..., x_n\}$ hay vô hạn đếm được thì X được gọi là ***biến ngẫu nhiên rời rạc***.
Để cho gọn, ta viết là $X = \{x_1, x_2, ..., x_n, ...\}$.

• Nếu $X(\Omega)$ là 1 khoảng của $\mathbb{R}$ (hay cả $\mathbb{R}$) thì X được gọi là ***biến ngẫu nhiên liên tục***.

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.

• 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.

1.2. Hàm mật độ

a) Biến ngẫu nhiên rời rạc

Cho BNN rời rạc $X : \Omega \to \mathbb{R}$, $X = \{x_1, x_2, ..., x_n, ...\}$.
Giả sử $x_1 < x_2 < ... < x_n < ...$ với xác suất tương ứng là $P(\{\omega : X(\omega) = x_i\}) \equiv P(X = x_i) = p_i$, $i = 1, 2, ...$

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 là**

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

Chú ý

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

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

X	– 1	0	1	3	5
P	3*a*	*a*	0,1	2*a*	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, hãy lập bảng phân phối xác suất của X?

b) Biến ngẫu nhiên liên tục

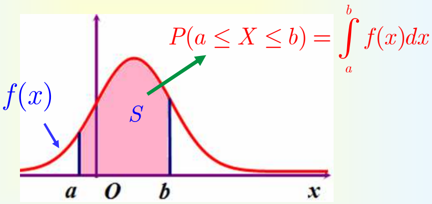
Hàm số $f: \mathbb{R} \to \mathbb{R}$ được gọi là ***hàm mật độ*** của biến ngẫu nhiên liên tục X nếu:

$$\boxed{P(a \leq X \leq b) = \int_a^b f(x)dx,\ \forall a, b \in \mathbb{R}.}$$

Nhận xét

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

- 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.



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 biến ngẫu nhiên X có hàm mật độ:

$$f(x) = \begin{cases} 0, & x < 2 \\ \dfrac{k}{x^2}, & x \geq 2. \end{cases}$$ Tính $P(-3 < X < 5)$?

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

Nhận xét 1

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

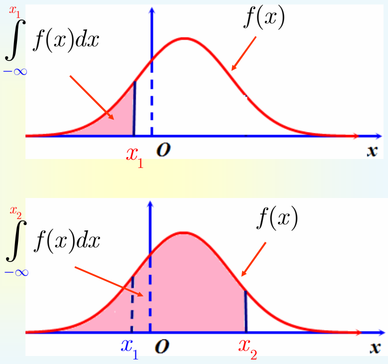
$$\boxed{F(x) = \sum_{x_i < x} p_i}.$$

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

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

$$F(x_1) = \int_{-\infty}^{x_1} f(x)dx$$

$$F(x_2) = \int_{-\infty}^{x_2} f(x)dx$$



Nhận xét 2

- Giả sử BNN rời rạc X nhận các giá trị trong $[x_1; x_n]$ và $x_1 < x_2 < ... < x_n$, $P(X = x_i) = p_i$ $(i = 1, 2, ..., n)$.

Ta có hàm phân phối của X là:

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

- Giả sử BNN liên tục X có hàm mật độ

$$f(x) = \begin{cases} \varphi(x), & x \in [a; b] \\ 0, & x \notin [a; b]. \end{cases}$$

Ta có hàm phân phối của X là:

$$F(x) = \begin{cases} 0 & \text{khi} \quad x \le a \\ \int_a^x \varphi(t)dt & \text{khi} \quad a < x \le b \\ 1 & \text{khi} \quad b < x. \end{cases}$$

- Giả sử BNN liên tục X có hàm mật độ

$$f(x) = \begin{cases} 0, & x < a \\ \varphi(x), & x \ge a. \end{cases}$$

Ta có hàm phân phối của X là:

$$F(x) = \begin{cases} 0 & \text{khi} \quad x \le a \\ \int_a^x \varphi(t)dt & \text{khi} \quad x > a. \end{cases}$$

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

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

Hãy lập hàm phân phối của X và vẽ đồ thị của $F(x)$?

VD 2. Cho BNN X có hàm mật độ là:

$$f(x) = \begin{cases} 0, & x \notin [0; 1] \\ 3x^2, & x \in [0; 1]. \end{cases}$$

Tìm hàm phân phối của X và vẽ đồ thị của $F(x)$?

VD 3. Cho BNN X có hàm mật độ là:

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

Tìm hàm phân phối $F(x)$ 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 phải tại mọi $x \in \mathbb{R}$.

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

Đặc biệt

- Nếu X là BNN rời rạc thì:

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

- Nếu X là BNN liên tục thì:

$$\boxed{\begin{aligned} P(a \leq X \leq b) &= P(a \leq X < b) = P(a < X \leq b) \\ &= P(a < X < b) = F(b) - F(a). \end{aligned}}$$

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

$$\boxed{F'(x) = f(x).}$$

§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. TRUNG VỊ và MODE

3.1.1. Trung vị (tham khảo)

Trung vị (*median*) của BNN X, ký hiệu $MedX$, là số thực m thỏa:

$$\boxed{P(X \leq m) = P(X \geq m).}$$

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

X	0	1	2	3
P	0,125	0,375	0,375	0,125

Ta có: $MedX = m$ thỏa $1 < m < 2$.

VD 2. Cho BNN X có hàm mật độ:

$$f(x) = \frac{2}{9}\begin{cases} 3x - x^2, & x \in [0;\, 3] \\ 0, & x \notin [0;\, 3]. \end{cases}$$

Ta có: $P(X \leq m) = P(X \geq m) \Rightarrow MedX = m \in [0;\, 3]$

$$\Rightarrow \frac{1}{2} = \frac{2}{9}\int_0^m (3x - x^2)dx \Rightarrow m = \frac{3}{2} \in [0;\, 3].$$

3.1.2. MODE

Mode của biến ngẫu nhiên X, ký hiệu $ModX$, là giá trị $x_0 \in X$ thỏa:

- $P(X = x_0)$ max nếu X là rời rạc, và
- $f(x_0)$ max nếu X liên tục có hàm mật độ $f(x)$.

Chú ý

- $ModX$ còn được gọi là ***giá trị tin chắc nhất*** của X.
- Biến ngẫu nhiên X có thể có nhiều $ModX$.

VD 3. 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: $ModX = 2$.

VD 4. Tìm $ModX$, biết X có bảng phân phối xác suất:

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

Giải. Ta có:

$$1 - 3p + 0,18 + 0,07 + 0,25 + p = 1 \Rightarrow p = 0,25.$$

Vậy $ModX = 1;\ 5;\ 8$.

VD 5. Tìm $ModX$, biết X có hàm mật độ xác suất:

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

Giải. Với $x \in [0;\ 4]$, ta có:

$$f'(x) = 0 \Leftrightarrow x = 0,\ x = \frac{8}{3}.$$

$$f(0) = f(4) = 0,\ f\left(\frac{8}{3}\right) = \frac{4}{9}$$

$$\Rightarrow \max_{\mathbb{R}} f(x) = f\left(\frac{8}{3}\right) = \frac{4}{9}.$$ Vậy $ModX = \dfrac{8}{3}$.

3.2. KỲ VỌNG

3.2.1. Định nghĩa

Kỳ vọng (*Expectation*) 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:

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

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

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

Đặc biệt

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

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

VD 6. 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?

Giải. Ta có:

$$EX = -1 \times 0,1 + 0 \times 0,2 + 2 \times 0,4 + 3 \times 0,3.$$

Vậy $EX = 1,6$.

VD 7. Một lô hàng gồm 10 sản phẩm tốt và 2 phế phẩm. Lấy ngẫu nhiên 4 sản phẩm từ lô hàng đó, gọi X là số sản phẩm tốt trong 4 sản phẩm lấy ra.

Tìm phân phối xác suất và tính kỳ vọng của X?

VD 8. Tìm kỳ vọng của BNN X có hàm mật độ:

$$f(x) = \begin{cases} \dfrac{3}{4}(x^2 + 2x), & x \in [0;\ 1] \\ 0, & x \notin [0;\ 1]. \end{cases}$$

VD 9. 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 10. 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. 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. Một 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. 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 nhận được bao nhiêu tiền?

3.2.3. 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:

$$\boxed{EY = \sum_i y_i.p_i = \sum_i \varphi(x_i).p_i}$$

- Nếu X là biến ngẫu nhiên liên tục thì:

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

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 15. 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 16. Cho BNN X có hàm mật độ xác suất:

$$f(x) = \begin{cases} \dfrac{2}{x^2}, & x \in [1;\ 2] \\ 0, & x \notin [1;\ 2]. \end{cases}$$

Tính EY với $Y = X^5 - \dfrac{2}{X}$?

3.3. PHƯƠNG SAI

3.3.1. Định nghĩa

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

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

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

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

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

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

VD 17. Cho BNN X có bảng phân phối xác suất:

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

Ta có: $VarX = (1^2.0,2 + 2^2.0,7 + 3^2.0,1)$

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

VD 18. 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 19. Cho BNN X có hàm mật độ xác suất:

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

Tính phương sai của Y, cho biết $Y = 2X^2$.

3.3.2. 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*** (*standard deviation*) là:

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

VD 20. Năng suất (sản phẩm/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 một trong hai loại máy này thì ta chọn mua máy Y.

§2. PHÂN PHỐI NHỊ THỨC

2.1. Phân phối Bernoulli

a) Định nghĩa

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

$$X = \begin{cases} 1 \textit{ khi } A \textit{ xuất hiện,} \\ 0 \textit{ khi } \overline{A} \textit{ xuất hiện,} \end{cases} \quad P(\overline{A}) = 1 - p = q.$$

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

b) Các số đặc trưng của *X* ~ *B*(*p*)

$$\boxed{EX = p;\ VarX = pq.}$$

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 $X = \begin{cases} 1 \; khi\ sinh\ viên\ này\ trả\ lời\ đúng, \\ 0 \; khi\ sinh\ viên\ này\ trả\ lời\ sai, \end{cases}$

thì $X \in B\left(\frac{1}{4}\right)$ và $EX = \frac{1}{4}$, $VarX = \frac{1}{4}.\frac{3}{4} = \frac{3}{16}$.

2.2. Phân phối Nhị thức

a) Định nghĩa

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

  Nghĩa là: $X_i = \begin{cases} 1 \; khi\ lần\ thứ\ i\ A\ xuất\ hiện, \\ 0 \; khi\ lần\ thứ\ i\ \overline{A}\ xuất\ hiện. \end{cases}$

- 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*** (*Binomial distribution*) với tham số n, p.

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

- Xác suất trong n lần thử có k lần A xuất hiện là:

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

<u>VD 2.</u> 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 ?

b) Các số đặc trưng của *X ~ B(n, p)*

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

<u>VD 3.</u> 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 10% ?

<u>VD 4.</u> 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 ?

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

§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, $X \in B\left(n, \frac{\lambda}{n}\right)$.

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

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

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

Suy ra:

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

3.2. Định nghĩa phân phối Poisson

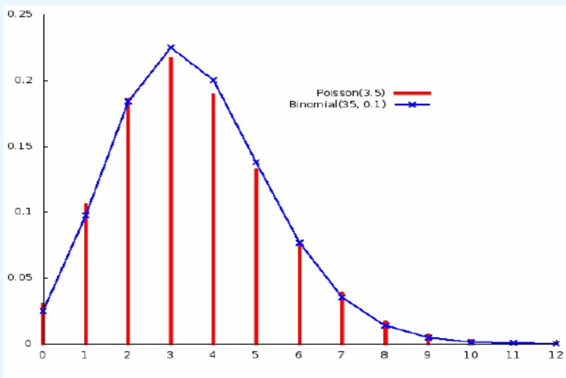
Biến ngẫu nhiên X được gọi là có phân phối Poisson tham số $\lambda > 0$, ký hiệu là $X \in P(\lambda)$ hay $X \sim P(\lambda)$, nếu X nhận các giá trị 0, 1, 2,…, n,… với xác suất:

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

Trong đó, λ là trung bình số lần xuất hiện biến cố nào đó mà ta quan tâm.

Nhận xét

- Phân phối Poisson không phải là phân phối xác suất chính xác. Tuy vậy, phân phối Poisson rất thuận tiện cho việc mô tả và tính toán.
- Phân phối Poisson thường gắn với yếu tố thời gian.



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

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

$$\boxed{EX = VarX = \lambda; \quad ModX = x_0 : \lambda - 1 \le x_0 \le \lambda.}$$

<u>VD 1.</u> 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 ?

<u>VD 2.</u> Quan sát thấy trung bình 1 phút có 3 ôtô đi qua trạm thu phí. Biết xác suất có ít nhất 1 ôtô đi qua trạm thu phí trong t phút bằng 0,9. Giá trị của t là:

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

<u>VD 3.</u> Quan sát thấy trung bình 1 ngày (24 giờ) có 12 chuyến tàu vào cảng A. Chọn ngẫu nhiên liên tiếp 6 giờ trong 1 ngày. Tính xác suất để 2 trong 6 giờ ấy, mỗi giờ có đúng 1 tàu vào cảng A.

§4. PHÂN PHỐI CHUẨN

4.1. Phân phối Chuẩn đơn giản

a) Định nghĩa

Biến ngẫu nhiên liên tục T được gọi là có phân phối Chuẩn đơn giản (hay *phân phối* ***Gauss***), ký hiệu là $T \in N(0;\ 1)$ hay $T \sim N(0;\ 1)$, nếu hàm mật độ xác suất của T có dạng:

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

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

b) Các số đặc trưng của $T \sim N(0;\ 1)$

$$\boxed{ModT = ET = 0;\ \ VarT = 1.}$$

c) Xác suất của $T \sim N(0;\ 1)$

• Hàm Laplace

Hàm $\varphi(x) = \int_0^x f(t)dt\ (t \ge 0)$ được gọi là hàm Laplace.

(Giá trị 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

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

Chú ý

- $P(T < b) = 0,5 + \varphi(b);\ P(T > a) = 0,5 - \varphi(a).$
- Nếu $x \ge 4$ thì $\varphi(x) \approx 0,5.$

4.2. Phân phối Chuẩn

a) Định nghĩa

Biến ngẫu nhiên liên tục X được gọi là có phân phối Chuẩn (*Normal distribution*) 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:

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

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

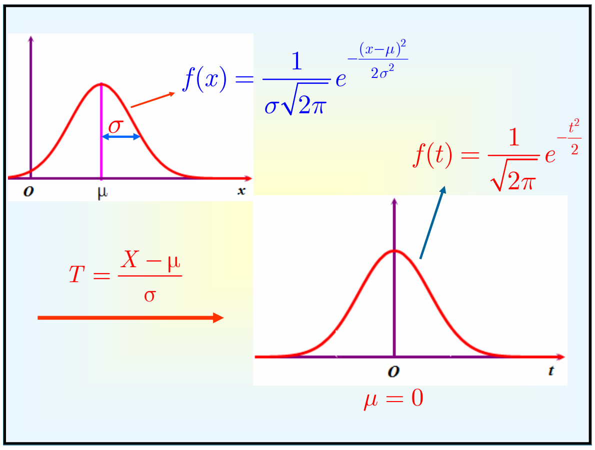
$$\boxed{ModX = EX = \mu;\ \ VarX = \sigma^2.}$$

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

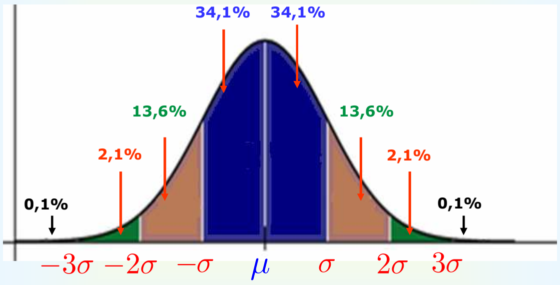
Nếu $X \in N(\mu;\ \sigma^2)$ thì $T = \dfrac{X - \mu}{\sigma} \in N(0;\ 1).$

Vậy, ta có công thức tính xác suất:

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



Phân bố xác suất



VD 1. 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 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 tỉ lệ học sinh thi đỗ là 25,14%. Độ lệch chuẩn là:

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

VD 3. Giả sử 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 4. Cho BNN X có phân phối chuẩn với $EX = 10$ và $P(10 < X < 20) = 0{,}3$. Tính $P(0 < X \le 15)$?

Phân phối Chi bình phương $\chi^2(n)$ (tham khảo)

Nếu $X_i \in N(0;\ 1)$ $(i = 1,\ldots,n)$ và các X_i độc lập thì

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

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

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

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

• Phân phối $\chi^2(n)$ do Karl Pearson đưa ra năm 1900.

Karl Pearson (1857 - 1936)

Phân phối Student $St(n)$ (tham khảo)

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

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

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

Trong đó, n được gọi là bậc tự do và giá trị của $St(n)$ được cho trong bảng C.

• Phân phối $St(n)$ do Willam.S.Gosset đưa ra năm 1908.

Willam. Sealy .Gosset 1876 - 1937

§1. Phân phối xác suất của vector ngẫu nhiên rời rạc

§2. Phân phối xác suất của vector ngẫu nhiên liên tục

Khái niệm vector ngẫu nhiên

• Một bộ có thứ tự n biến ngẫu nhiên $(X_1, \ldots, X_n)$ được gọi là một *vector ngẫu nhiên n chiều*.

• Vector ngẫu nhiên n chiều là liên tục hay rời rạc nếu các biến ngẫu nhiên thành phần là liên tục hay rời rạc. Chẳng hạn, một nhà máy sản xuất một loại sản phẩm, nếu xét đến kích thước của sản phẩm được đo bằng chiều dài X và chiều rộng Y thì ta có vector ngẫu nhiên hai chiều (X, Y). Còn nếu xét thêm cả chiều cao Z nữa thì ta có vector ngẫu nhiên ba chiều (X, Y, Z).

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

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

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

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

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

1.2. 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	$\ldots$	x_m
P	$p_{1\bullet}$	$p_{2\bullet}$	$\ldots$	$p_{m\bullet}$

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

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

Kỳ vọng của X là:

$$\boxed{EX = x_1 p_{1\bullet} + x_2 p_{2\bullet} + \cdots + x_m p_{m\bullet}.}$$

• Bảng phân phối xác suất của *Y*

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

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

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

Kỳ vọng của Y là:

$$\boxed{EY = y_1 p_{\bullet 1} + y_2 p_{\bullet 2} + \cdots + y_n p_{\bullet n}.}$$

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

X \ 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

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

2) Lập bảng phân phối xs thành phần và tính EX, EY.

Giải

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

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

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

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

1.3. Phân phối xác suất có điều kiện

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

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

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

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

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

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

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

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

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

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

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

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

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

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.

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 $P\left(X - Y = 1\right)$.
2) Tính xác suất $P(X > 0 \mid Y = 1)$.
3) Tính trung bình của X và Y.
4) Tính trung bình của Y khi $X = 1$.

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:

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

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.

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

2.1. Hàm mật độ đồng thời của (X, Y)

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

$$\iint_{\mathbb{R}^2} f(x,y)dxdy = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)dxdy = 1.$$

• Xác suất của vector (X,Y) trên tập $D \subset \mathbb{R}^2$ là:

$$P\{(X,Y) \in D\} = \iint_D f(x,y)dxdy.$$

2.2. Hàm mật độ thành phần

• Hàm mật độ của X là:

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

• Hàm mật độ của Y là:

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

Chú ý

Khi tìm hàm $f_X(x)$, ta lấy tích phân hàm $f(x,y)$ theo biến y và điều kiện x phải độc lập đối với y.

Tìm hàm $f_Y(y)$, ta làm tương tự.

Trung bình thành phần

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

2.3. Hàm mật độ có điều kiện

• Hàm mật độ có điều kiện của X khi biết $Y = y$ là:

$$f_X\left(x\middle|y\right) = \frac{f(x,y)}{f_Y(y)}.$$

• Hàm mật độ có điều kiện của Y khi biết $X = x$ là:

$$f_Y\left(y\middle|x\right) = \frac{f(x,y)}{f_X(x)}.$$

VD 1. Cho hàm $f(x,y) = \begin{cases} 10x^2y, \text{ khi } 0 \leq y \leq x \leq 1, \\ 0, \text{ nơi khác.} \end{cases}$

1) Chứng tỏ vector (X,Y) có hàm mật độ là $f(x,y)$.

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

3) Tìm hàm mật độ thành phần của X, Y.

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

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

VD 2. Cho hàm mật độ đồng thời của vector (X,Y) là:

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

1) Tính trung bình thành phần của X, Y.

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

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

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

Thời gian chơi thể thao trung bình là:

A. 0,3125 giờ; B. 0,5214 giờ;
C. 0,1432 giờ; D. 0,4132 giờ.

....

§1. Một số loại hội tụ trong xác suất và các định lý
§2. Các loại xấp xỉ phân phối xác suất

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

1.1. Hội tụ theo xác suất – Luật số lớn

a) Định nghĩa

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

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

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

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

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

b) Định lý (***Bất đẳng thức Tchébyshev***)

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

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

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

Chứng minh

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

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

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

$$\sigma^2=\int_{-\infty}^{+\infty}(x-\mu)^2 f(x)dx$$
$$=\int_{|x-\mu|<\varepsilon}(x-\mu)^2 f(x)dx+\int_{|x-\mu|\ge\varepsilon}(x-\mu)^2 f(x)dx$$
$$\ge\int_{|x-\mu|\ge\varepsilon}(x-\mu)^2 f(x)dx$$
$$\ge\varepsilon^2\int_{|x-\mu|\ge\varepsilon}f(x)dx=\varepsilon^2 P\left(\left|X-\mu\right|\ge\varepsilon\right).$$

Vậy $\sigma^2\ge\varepsilon^2 P\left(\left|X-\mu\right|\ge\varepsilon\right)\Leftrightarrow P\left(\left|X-\mu\right|\ge\varepsilon\right)\le\dfrac{\sigma^2}{\varepsilon^2}$ ■

Ý nghĩa của định lý

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

c) 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 $VarX_i\le C$ (hằng số) thì:

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

• **Hệ quả**

Nếu dãy các BNN $\{X_i\}$ $(i = 1,\ldots,n,\ldots)$ 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

a) Định nghĩa

Dãy các biến ngẫu nhiên $\{X_i\}$ $(i = 1,\ldots,n,\ldots)$ được gọi là ***hội tụ yếu*** hay ***hội tụ theo phân phối*** đến biến ngẫu nhiên X nếu $\lim\limits_{n\to\infty} F_n(x) = F(x), \forall x \in C(F)$.

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

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

Chú ý

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

b) Định lý giới hạn trung tâm (*định lý Liapounop*)

Cho dãy BNN $\{X_i\}$ $(i = 1,\ldots,n,\ldots)$ độc lập từng đôi.

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

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

thì $Y \in N(\mu,\ \sigma^2)$.

• **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,…

§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 p cố định, $N \to \infty$ và $\frac{N_A}{N} \to p = 1 - q$ thì:

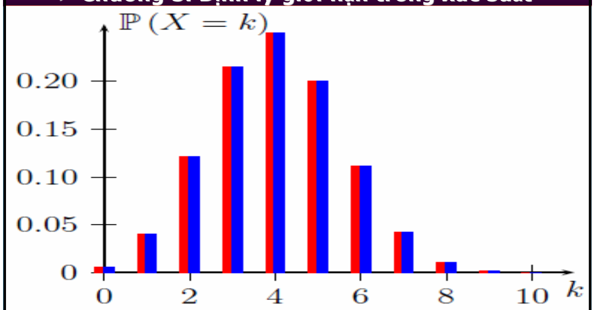
$$\frac{C_{N_A}^{k} C_{N-N_A}^{n-k}}{C_N^n} \xrightarrow{d} C_n^k p^k q^{n-k}.$$

• Ứng dụng, nếu N khá lớn và n rất nhỏ so với N thì:

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

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.



Đỏ: $X \in H(10.000;\ 4.000;\ 10)$,

Xanh: $X \in B(10;\ 0,4)$.

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

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

➢ Chương 5. Định lý giới hạn trong xác suất

2.2. Xấp xỉ phân phối Nhị thức bởi Poisson

Xét biến ngẫu nhiên X có phân phối Nhị thức $B(n;\ p)$.

• Khi $n \to \infty$, nếu $p \to 0$ và $np \to \lambda$ thì:

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

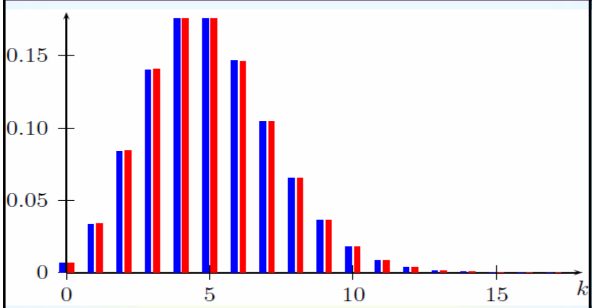
• Ứng dụng, đặt $\lambda = np$.

Nếu n đủ lớn và p gần bằng 0 (hoặc gần bằng 1) thì: $\boxed{X \sim P(\lambda).}$

Chú ý

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

➢ Chương 5. Định lý giới hạn trong xác suất



Xanh: $X \in B(1.000;\ 0,005)$,

Đỏ: $X \in P(5)$.

➢ Chương 5. Định lý giới hạn trong xác suất

VD 2. Một lô hàng thịt đông lạnh đóng gói nhập khẩu có chứa 0,4% bị nhiễm khuẩn. Tìm xác suất để khi chọn ngẫu nhiên 1.000 gói thịt từ lô hàng này có:

1) không quá 2 gói bị nhiễm khuẩn;
2) đúng 34 gói bị nhiễm khuẩn.

VD 3. Giải câu 3) trong VD 1.

➢ Chương 5. Định lý giới hạn trong xác suất

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

a) Định lý giới hạn địa phương Moivre – Laplace

Xét biến ngẫu nhiên X có phân phối Nhị thức $B(n;\ p)$.

Với $k = 0, 1, \ldots, n$ bất kỳ và $x = \dfrac{k - np}{\sqrt{npq}}$, ta có :

$$\lim_{n\to\infty} \frac{\sqrt{npq}.P_n(X = k)}{\dfrac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}} = 1.$$

➢ Chương 5. Định lý giới hạn trong xác suất

b) Định lý giới hạn tích phân Moivre – Laplace

Xét biến ngẫu nhiên X có phân phối Nhị thức $B(n;\ p)$.

Với mọi $a,\ b \in \mathbb{R}$ và $a < b$, ta có:

$$\lim_{n\to\infty} P(a \le X \le b) = \frac{1}{\sqrt{2\pi}} \int_{\frac{a-np}{\sqrt{npq}}}^{\frac{b-np}{\sqrt{npq}}} e^{-\frac{x^2}{2}}\, dx.$$

c) Ứng dụng xấp xỉ

Cho $X \in B(n;\ p)$. Nếu n khá lớn, $np \geq 5$ và $nq \geq 5$

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

Khi đó:

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

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

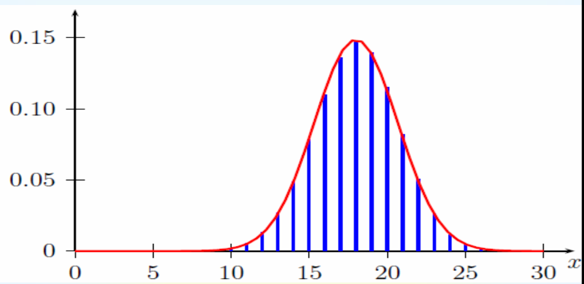
$$\boxed{P(k_1 \leq X \leq k_2) = \varphi\left(\frac{k_2-\mu}{\sigma}\right) - \varphi\left(\frac{k_1-\mu}{\sigma}\right).}$$

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

Chú ý

Khi $k = \mu$, ta sử dụng công thức hiệu chỉnh:

$$\boxed{P(X = k) \approx P(k - 0,5 \leq X \leq k + 0,5).}$$



Xanh: $X \in B(30;\ 0,6)$,

Đỏ: $X \in N(18;\ 7,2)$.

VD 4. 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 5. Trong 10.000 sản phẩm trên một dây chuyền sản xuất có 2.000 sản phẩm không được kiểm tra chất lượng. Tìm xác suất để trong 400 sản phẩm sản xuất ra:

1) có 80 sản phẩm không được kiểm tra;
2) có từ 70 đến 100 sản phẩm không được kiểm tra.

VD 6. 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.

Hướng dẫn:

$$X \in B(325;\ 0,9) \Rightarrow \mu = 292,5;\ \sigma^2 = 29,25$$

$$\Rightarrow X \sim N(292,5;\ 29,25).$$

1) $P(X = 300) = 0,0281$;

2) $P(0 \leq X \leq 300) = 0,9177$.

Tóm tắt xấp xỉ Chuẩn cho Nhị thức

$X \in B(n,\ p)$	$\begin{cases} np \geq 5 \\ nq \geq 5 \end{cases}$	$X \in N(\mu,\ \sigma^2)$
$EX = np$	$\mu = np$	$EX = \mu$
$VarX = npq$	$\sigma^2 = npq$	$VarX = \sigma^2$

....

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

Chương VI. MẪU THỐNG KÊ VÀ ƯỚC LƯỢNG THAM SỐ

§1. LÝ THUYẾT MẪU

1.1. Mẫu và tổng thể

- 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 $X_1, X_2, \ldots, X_n$ là những kết quả quan sát. Ta xem như đã quan sát n lần, mỗi lần ta được một biến ngẫu nhiên $X_i\ (i = 1, \ldots, n)$.

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

1.2. Sắp xếp mẫu dựa vào số liệu thực nghiệm

a) 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

b) 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.

VD 3. Theo dõi mức nguyên liệu hao phí để sản xuất ra một đơn vị sản phẩm ở một nhà máy, ta thu được các số liệu sau (đơn vị: gam):

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

Hãy sắp xếp số liệu trên dưới dạng bảng ?

1.3. Các đặc trưng mẫu

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

a) Trung bình mẫu

$$\boxed{\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i.}$$

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

b) Phương sai mẫu

• Phương sai mẫu:

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

• Phương sai mẫu hiệu chỉnh:

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

• Trong tính toán cụ thể, ta sử dụng công thức:

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

Với $\overline{X^2} = \frac{1}{n}\sum_{i=1}^{n} X_i^2$.

c) Tỉ lệ mẫu

Xét mẫu định tính với các biến X_i $(i = 1, ..., n)$ 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 tỉ lệ mẫu là:

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

d) Liên hệ giữa đặc trưng của mẫu và tổng thể

Các đặc trưng mẫu $\overline{X}$, 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:

$F \to p$, $\overline{X} \to \mu$, $S^2 \to \sigma^2$ (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.

a) Máy *fx* 500 – 570 MS

• ***Xóa bộ nhớ***: **SHIFT** $\to$ **MODE** $\to$ **3** $\to$ **=** $\to$ **=**

• ***Vào chế độ thống kê nhập dữ liệu***:

– **MODE** $\to$ **2** (chọn SD đối với *fx*500MS);

MODE $\to$ **MODE** $\to$ **1** (chọn SD đối với *fx*570MS).

– **Nhập các số:**

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

• ***Xuất kết quả:***
- **SHIFT** $\to$ **2** $\to$ **1** $\to$ **=**
  (kết quả $\overline{x}$ là ***trung bình mẫu***).
- **SHIFT** $\to$ **2** $\to$ **2** $\to$ **=**
  (kết quả $x\sigma n$ là ***độ lệch chuẩn của mẫu*** $\hat{s}$).
- **SHIFT** $\to$ **2** $\to$ **3** $\to$ **=**
  ($x\sigma n - 1$ là ***độ lệch chuẩn của mẫu có hiệu chỉnh*** s).

b) Máy *fx* 500 – 570 ES

• ***Xóa bộ nhớ*: SHIFT** $\to$ **9** $\to$ **3** $\to$ **=** $\to$ **=**

• ***Vào chế độ thống kê nhập dữ liệu:***
- **SHIFT** $\to$ **MODE** $\to$ dịch chuyển mũi tên tìm chọn mục **Stat** $\to$ **2** (chế độ không tần số).

- **MODE** $\to$ **3** (stat) $\to$ **1** (1-var) $\to$ (nhập các số):
  12= 13= 11= 14= 11= $\to$ **AC**

• ***Xuất kết quả:***
- **SHIFT** $\to$ **1** $\to$ **5** (var) $\to$ **1** $\to$ **=** (n: cỡ mẫu)
- **SHIFT** $\to$ **1** $\to$ **5** (var) $\to$ **2** $\to$ **=** ($\overline{x}$)
- **SHIFT** $\to$ **1** $\to$ **5** (var) $\to$ **3** $\to$ **=** ($x\sigma n = \hat{s}$).
- **SHIFT** $\to$ **1** $\to$ **5** (var) $\to$ **4** $\to$ **=** ($x\sigma n - 1 = s$).

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

a) Máy *fx* 500 – 570 MS

• ***Xóa bộ nhớ*: SHIFT** $\to$ **MODE** $\to$ **3** $\to$ **=** $\to$ **=**

• ***Vào chế độ thống kê nhập dữ liệu:***
- **MODE** $\to$ **2** (chọn SD đối với *fx*500MS);
  MODE $\to$ **MODE** $\to$ **1** (chọn SD đối với *fx*570MS).
- **Nhập các số:**
  12 $\to$ **SHIFT** $\to$ **,** $\to$ **3** $\to$ **M+**
  11 $\to$ **SHIFT** $\to$ **,** $\to$ **2** $\to$ **M+**
  15 $\to$ **SHIFT** $\to$ **,** $\to$ **4** $\to$ **M+**

• ***Xuất kết quả, ta làm như 1a).***

b) Máy *fx* 500 – 570 ES

• ***Xóa bộ nhớ*: SHIFT** $\to$ **9** $\to$ **3** $\to$ **=** $\to$ **=**

• ***Vào chế độ thống kê nhập dữ liệu:***
- **SHIFT** $\to$ **MODE (SETUP)** dịch chuyển mũi tên $\to$ **4** $\to$ **1**
- **MODE** $\to$ **3** (stat) $\to$ **1** (1-var)
- **Nhập các giá trị và tần số vào 2 cột trên màn hình:**

X	FREQ	
12	3	
11	2	
15	4	$\to$ **AC**

• ***Xuất kết quả, làm như 1b).***

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 mẫu chưa hiệu chỉnh và độ lệch chuẩn của mẫu có hiệu chỉnh.

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.

§3. 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 $B(1;\ p)$.

3.1. Định nghĩa

- Xét thống kê T ước lượng tham số θ, 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 cho θ được gọi là bài toán ***ước lượng khoảng***.

3.2. Ước lượng khoảng cho trung bình tổng thể μ

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

Trong thực hành, ta có 4 trường hợp sau.

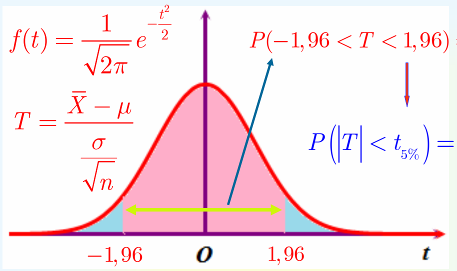
a) *Trường hợp 1.* Kích thước mẫu $n \ge 30$ và phương sai tổng thể σ^2 đã biết.

- Từ mẫu ta tính $\overline{x}$ (trung bình mẫu).

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

- Khoảng ước lượng là:

$$\boxed{\left(\overline{x}-\varepsilon;\ \overline{x}+\varepsilon\right),\ \varepsilon = t_\alpha . \frac{\sigma}{\sqrt{n}}.}$$

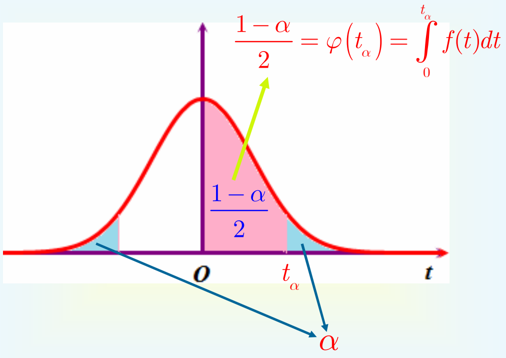


$-t_{5\%}$ $\qquad$ $t_{5\%}$

ε

$$-t_{5\%} < T < t_{5\%} \Rightarrow \overline{X} - t_{5\%}.\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + t_{5\%}.\frac{\sigma}{\sqrt{n}}$$

Tra bảng B



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

- Tính $\overline{x}$ và s (độ lệch chuẩn mẫu đã hiệu chỉnh).
- Từ $1-\alpha \Rightarrow \frac{1-\alpha}{2} = \varphi(t_\alpha) \xrightarrow{\text{tra bảng } B} t_\alpha$.
- Khoảng ước lượng là:

$$\boxed{\left(\overline{x}-\varepsilon;\ \overline{x}+\varepsilon\right),\ \varepsilon = t_\alpha\frac{s}{\sqrt{n}}.}$$

Chú ý

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

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

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

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

- Từ mẫu ta tính $\overline{x}$, s.
- Từ $1-\alpha \Rightarrow \alpha \xrightarrow{\text{tra bảng } C} t_\alpha^{n-1}$

  (nhớ giảm bậc thành $n-1$ rồi mới tra bảng!)
- Khoảng ước lượng là:

$$\boxed{\left(\overline{x}-\varepsilon;\ \overline{x}+\varepsilon\right),\ \varepsilon = t_\alpha^{n-1}.\frac{s}{\sqrt{n}}.}$$

CÁC BÀI TOÁN VỀ ƯỚC LƯỢNG KHOẢNG

Bài 1. Ước lượng khoảng (sử dụng công thức).

Bài 2. Tìm độ tin cậy (ta không xét TH4)

$$\varepsilon = t_\alpha\frac{s}{\sqrt{n}} \Rightarrow t_\alpha = \frac{\varepsilon\sqrt{n}}{s} \xrightarrow{B} \varphi(t_\alpha)$$

$$\varphi(t_\alpha) = \frac{1-\alpha}{2} \Rightarrow 1-\alpha = 2\varphi(t_\alpha)$$

Bài 3. Tìm cỡ mẫu (ta chỉ xét TH1 và TH2)

a) Nếu ε > ε' thì ta giải bất đẳng thức:

$$t_\alpha\frac{s}{\sqrt{N}} > \varepsilon' \Rightarrow N < \left(t_\alpha.\frac{s}{\varepsilon'}\right)^2 \Rightarrow N_{\max}.$$

b) Nếu ε < ε' thì ta giải bất đẳng thức:

$$t_\alpha\frac{s}{\sqrt{N}} < \varepsilon' \Rightarrow N > \left(t_\alpha.\frac{s}{\varepsilon'}\right)^2 \Rightarrow N_{\min}.$$

VD 1. Lượng Vitamin có trong một 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 20 mg. Với độ tin cậy 95%, hãy ước lượng lượng Vitamin trung bình có trong một trái cây A ?

VD 2. Biết chiều cao 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 sai số không quá 1 cm thì phải cần đo ít nhất mấy người ?

➢ Chương 6. Mẫu thống kê & Ước lượng tham số

VD 3. Kiểm tra tuổi thọ (tính bằng 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 ?

➢ Chương 6. Mẫu thống kê & Ước lượng tham số

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%?

➢ Chương 6. Mẫu thống kê & Ước lượng tham số

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

Nhu cầu (kg/tháng)	0,5	1,5	2,5	3,5
Số gia đình	10	35	86	132

Nhu cầu (kg/tháng)	4,5	5,5	6,5	7,5
Số gia đình	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 4,8 tấn/năm và độ tin cậy 99% thì cần khảo sát tối đa bao nhiêu gia đình trong phường A ?

➢ Chương 6. Mẫu thống kê & Ước lượng tham số

3.3. Ước lượng khoảng cho tỉ lệ tổng thể *p*

• Giả sử tỉ lệ p các phần tử có tính chất A của tổng thể chưa biết. Với độ tin cậy $1-\alpha$ cho trước, khoảng ước lượng p là $(p_1;\ p_2)$ thỏa $P(p_1 < p < p_2) = 1-\alpha$.

• Nếu biết tỉ lệ mẫu $f = f_n = \dfrac{m}{n}$ với n là cỡ mẫu, m là số phần tử ta quan tâm thì khoảng ước lượng cho p là:

$$\boxed{\left(f-\varepsilon;\ f+\varepsilon\right),\ \varepsilon = t_\alpha\sqrt{\frac{f(1-f)}{n}}.}$$

Trong đó t_α tìm được từ $\varphi(t_\alpha) = \dfrac{1-\alpha}{2}$ (tra bảng B).

➢ Chương 6. Mẫu thống kê & Ước lượng tham số

VD 8. Tỉnh X có 1.000.000 thanh niên. Người ta khảo sát ngẫu nhiên 20.000 thanh niên của tỉnh X về trình độ học vấn thì thấy có 12.575 thanh niên đã tốt nghiệp PTTH. Hãy ước lượng tỉ lệ thanh niên đã tốt nghiệp PTTH của tỉnh X với độ tin cậy 95%? Số thanh niên đã tốt nghiệp PTTH của tỉnh X trong khoảng nào?

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 $\varepsilon = 0,0177$ 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?

3.4. Ước lượng khoảng cho phương sai tổng thể σ²
(Tham khảo)

Giả sử tổng thể X có phân phối chuẩn với phương sai σ^2 chưa biết. Với độ tin cậy $1-\alpha$ cho trước, khoảng ước lượng cho σ^2 là $(\sigma_1^2;\ \sigma_2^2)$ thỏa:

$$P(\sigma_1^2 < \sigma^2 < \sigma_2^2) = 1 - \alpha.$$

Trong thực hành ta có hai trường hợp sau

a) *Trường hợp 1.* Trung bình tổng thể μ đã biết.

- Từ mẫu ta tính $n.\hat{s}^2$.

- Từ $1-\alpha \Rightarrow \dfrac{\alpha}{2}$, tra bảng D ta tìm được:

$$\chi_n^2\left(1-\frac{\alpha}{2}\right),\ \chi_n^2\left(\frac{\alpha}{2}\right).$$

- Khoảng ước lượng là $(\sigma_1^2;\ \sigma_2^2)$, trong đó:

$$\boxed{\sigma_1^2 = \frac{n.\hat{s}^2}{\chi_n^2\left(1-\frac{\alpha}{2}\right)},\ \sigma_2^2 = \frac{n.\hat{s}^2}{\chi_n^2\left(\frac{\alpha}{2}\right)}.}$$

b) *Trường hợp 2.* Trung bình tổng thể μ chưa biết.

- Từ mẫu ta tính $\overline{x},\ s^2$.
- Từ $1-\alpha \Rightarrow \dfrac{\alpha}{2}$, tra bảng D ta tìm được:

$$\chi_{n-1}^2\left(1-\frac{\alpha}{2}\right),\ \chi_{n-1}^2\left(\frac{\alpha}{2}\right).$$

- Khoảng ước lượng là $(\sigma_1^2;\ \sigma_2^2)$, trong đó:

$$\boxed{\sigma_1^2 = \frac{(n-1)s^2}{\chi_{n-1}^2\left(1-\frac{\alpha}{2}\right)},\ \sigma_2^2 = \frac{(n-1)s^2}{\chi_{n-1}^2\left(\frac{\alpha}{2}\right)}.}$$

VD 13. Khảo sát 16 sinh viên về điểm trung bình của học kỳ 2 thì tính được $s = 1,5$ điểm. Hãy ước lượng phương sai về điểm trung bình học kỳ 2 của sinh viên với độ tin cậy 97%, biết rằng điểm trung bình X của sinh viên là biến ngẫu nhiên có phân phối chuẩn.

VD 14. Mức hao phí nguyên liệu cho 1 đơn vị sản phẩm là biến ngẫu nhiên X (gram) có phân phối chuẩn. Quan sát 28 sản phẩm này người ta thu được bảng số liệu:

X (gram)	19,0	19,5	20,0	20,5
Số sản phẩm	5	6	14	3

Với độ tin cậy 95%, hãy ước lượng phương sai của mức hao phí nguyên liệu trên trong 2 trường hợp:
1) biết $EX = 20$ gram; b) chưa biết EX.

§1. Khái niệm về kiểm định giả thuyết thống kê
§2. Kiểm định so sánh đặc trưng với một số
§3. Kiểm định so sánh hai đặc trưng

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

1.1. Khái niệm chung

- 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 *nghịch thuyết* (hay đối thuyết) $\overline{H}$.
- 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 rằng 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).

1.2. 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:

a) Sai lầm loại I

- Sai lầm loại 1 là loại sai lầm mà ta phạm phải trong việc bác bỏ giả thuyết H khi H đúng.
- Xác suất của việc bác bỏ H khi H đúng là xác suất của sai lầm loại 1 và được ký hiệu là α.

b) Sai lầm loại II

- Sai lầm loại 2 là loại sai lầm mà ta phạm phải trong việc chấp nhận giả thuyết H khi H sai.
- Xác suất của việc chấp nhận giả thuyết H khi H sai là xác suất của sai lầm loại 2 và được ký hiệu là β.

c) 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 $X_1, \ldots, X_n$ và đưa ra giả thuyết H.
- Từ mẫu trên, ta chọn thống kê $T = f(X_1, \ldots, X_n; \theta_0)$ 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 tin cậy (hay 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 hàm mật độ của T đối xứng qua trục Oy thì ta chọn khoảng đối xứng $[-t_\alpha; t_\alpha]$, với:

$$P(T \le -t_\alpha) = P(T \ge t_\alpha) = \frac{\alpha}{2}.$$

Vậy, khi xét nửa bên phải của trục Oy thì ta được:
  - nếu $t \le t_\alpha$ thì ta chấp nhận giả thuyết H;
  - nếu $t > t_\alpha$ thì ta bác bỏ giả thuyết H.
- Nếu hàm mật độ của T không đối xứng qua trục Oy thì ta chọn khoảng tin cậy $[0; C]$, với $P(T \ge C) = \alpha$.
  - Nếu $t \le C$ thì ta chấp nhận giả thuyết H, và
  - nếu $t > C$ thì ta bác bỏ giả thuyết H.

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

2.1. Kiểm định so sánh trung bình với một số

Với số μ_0 cho trước, ta đặt giả thuyết $H: \mu = \mu_0$.

a) *Trường hợp* 1. Với $n \ge 30$, σ^2 đã biết.

- Từ mức ý nghĩa $\alpha \Rightarrow \dfrac{1-\alpha}{2} = \varphi(t_\alpha) \xrightarrow{B} t_\alpha$.
- Tính giá trị thống kê $t = \dfrac{|\bar{x} - \mu_0|}{\sigma}\sqrt{n}$.
- Nếu $t \le t_\alpha$ thì ta chấp nhận H, nghĩa là $\mu = \mu_0$;
nếu $t > t_\alpha$ thì ta bác bỏ H, nghĩa là $\mu \ne \mu_0$.

b) *Trường hợp* 2. Với $n \ge 30$, σ^2 chưa biết.
Ta làm như trường hợp 1 nhưng thay σ bằng s.

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

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

- Từ cỡ mẫu n và mức ý nghĩa $\alpha \xrightarrow{\text{tra bảng } C} t_{\alpha}^{n-1}$.
- Tính giá trị thống kê $t = \dfrac{|\overline{x} - \mu_0|}{s}\sqrt{n}$.
- Nếu $t \le t_{\alpha}^{n-1}$ thì ta chấp nhận giả thuyết H;
  $t > t_{\alpha}^{n-1}$ thì ta bác bỏ giả thuyết H.

Chú ý

Trong tất cả các trường hợp bác bỏ, ta so sánh $\overline{x}$ và μ_0:

- Nếu $\overline{x} > \mu_0$ thì ta kết luận $\mu > \mu_0$.
- Nếu $\overline{x} < \mu_0$ thì ta kết luận $\mu < \mu_0$.

VD 1. Sở Điện lực A báo cáo rằng: trung bình một 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 một hộ trả 252 ngàn đồng tiền điện.
Trong kiểm định giả thuyết H: "**trung bình một hộ phải trả hàng tháng là 250 ngàn đồng tiền điện**" với mức ý nghĩa $\alpha = 1\%$, hãy cho biết giá trị thống kê t 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 $\hat{s} = 0,75$ giờ/ngày.
Với mức ý nghĩa 3%, hãy cho biết kết luận của ông B?

VD 4. Một công ty cho biết mức lương trung bình của một kỹ sư ở công ty là 5,7 triệu đồng/tháng với độ lệch chuẩn 0,5 triệu đồng/tháng. Kỹ sư A dự định xin vào làm ở công ty này và đã thăm dò 18 kỹ sư thì thấy lương trung bình là 5,45 triệu đồng/tháng.
Kỹ sư A quyết định rằng: nếu mức lương trung bình bằng với mức công ty đưa ra thì nộp đơn xin làm.
Với mức ý nghĩa 2%, 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 một 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 7. Thời gian X (phút) giữa hai chuyến xe bus trong một 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 5%, hãy kiểm định lời nói trên ?

VD 8. Chiều cao cây giống X (m) trong một vườn ươm là biến ngẫu nhiên có phân phối chuẩn. Người ta đo ngẫu nhiên 25 cây giống này và có bảng số liệu:

X (m)	0,8	0,9	1,0	1,1	1,2	1,3
Số cây	1	2	9	7	4	2

Theo quy định của vườn ươm, khi nào cây cao hơn 1 m thì đem ra trồng. Với mức ý nghĩa 5%, kiểm định giả thuyết H: "**cây giống của vườn ươm cao 1 m**" có giá trị thống kê và kết luận là:

A. $t = 2,7984$, không nên đem cây ra trồng.
B. $t = 2,7984$, nên đem cây ra trồng.
C. $t = 1,9984$, không nên đem cây ra trồng.
D. $t = 1,9984$, nên đem cây ra trồng.

2.2. Kiểm định so sánh tỉ lệ với một số

- Với số p_0 cho trước, ta đặt giả thuyết $H: p = p_0$.
- Từ mức ý nghĩa $\alpha \Rightarrow \dfrac{1-\alpha}{2} = \varphi(t_\alpha) \xrightarrow{B} t_\alpha$.
- Từ mẫu cụ thể, ta tính tỉ lệ mẫu $f = \dfrac{m}{n}$ và

giá trị thống kê $t = \dfrac{|f - p_0|}{\sqrt{p_0 q_0}}\sqrt{n}$, $q_0 = 1 - p_0$.

- Nếu $t \le t_\alpha$ thì chấp nhận H, nghĩa là $p = p_0$.
- Nếu $t > t_\alpha$ thì bác bỏ H, nghĩa là $p \ne p_0$.

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

VD 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 536 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: "**có 2% sinh viên ngủ trong giờ học**", mức ý nghĩa tối đa là bao nhiêu để H được chấp nhận ?

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

3.1. So sánh hai trung bình của hai tổng thể *X, Y*

Ta 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 (cả 4 trường hợp ta đều đặt giả thuyết $H: \mu_x = \mu_y$).

a) *Trường hợp 1.* $n_x,\ n_y \ge 30$ và $\sigma_x^2,\ \sigma_y^2$ đã biết.

Ta tính thống kê $t = \dfrac{|\overline{x} - \overline{y}|}{\sqrt{\dfrac{\sigma_x^2}{n_x} + \dfrac{\sigma_y^2}{n_y}}}$ và so sánh với t_α.

b) *Trường hợp 2.* $n_x,\ n_y \ge 30$ và $\sigma_x^2,\ \sigma_y^2$ chưa biết.

Ta thay $\sigma_x^2,\ \sigma_y^2$ bằng $s_x^2,\ s_y^2$ trong trường hợp 1.

c) *Trường hợp 3.* $n_x,\ n_y < 30$ và $\sigma_x^2,\ \sigma_y^2$ đã biết đồng thời X, Y ***có phân phối chuẩn.***

Ta làm như trường hợp 1.

d) *Trường hợp* 4. $n_x,\ n_y < 30$ và $\sigma_x^2,\ \sigma_y^2$ chưa biết đồng thời X, Y ***có phân phối chuẩn***.

- Tính phương sai chung của 2 mẫu:

$$s^2 = \frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2}.$$

- Tính giá trị thống kê $t = \dfrac{\left|\overline{x} - \overline{y}\right|}{s.\sqrt{\dfrac{1}{n_x} + \dfrac{1}{n_y}}}$.
- Từ $\alpha \xrightarrow{\text{tra bảng } C} t_\alpha^{n_x + n_y - 2}$ và so sánh với t.

VD 1. Người ta 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 8,5 quả; 72 cây được bón phân Y thu được trung bình 28,4 quả và độ lệch chuẩn 9,3 quả. Với mức ý nghĩa 5%, hãy cho biết kết luận về hai loại phân bón trên ?

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:

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

Với mức ý nghĩa 1%, ta có thể kết luận tuổi thọ của loại pin X cao hơn loại pin Y được không ?

3.2. So sánh hai tỉ lệ của hai tổng thể *X, Y*

Ta thực hiện các bước sau:

- Đặt giả thuyết $H: p_x = p_y$.
- Từ 2 mẫu ta tính $f_x = \dfrac{m_x}{n_x},\ f_y = \dfrac{m_y}{n_y},\ p_0 = \dfrac{m_x + m_y}{n_x + n_y}$.
- Tính giá trị thống kê $t = \dfrac{\left|f_x - f_y\right|}{\sqrt{p_0 q_0 \left(\dfrac{1}{n_x} + \dfrac{1}{n_y}\right)}}$.

- Kết luận:
  - Nếu $t \le t_\alpha$ thì ta chấp nhận $H \Rightarrow p_x = p_y$.
  - Nếu $t > t_\alpha$ và $f_x < f_y$ thì ta bác bỏ $H \Rightarrow p_x < p_y$.
  - Nếu $t > t_\alpha$ 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 hai 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 1%.