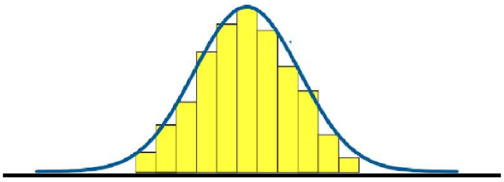
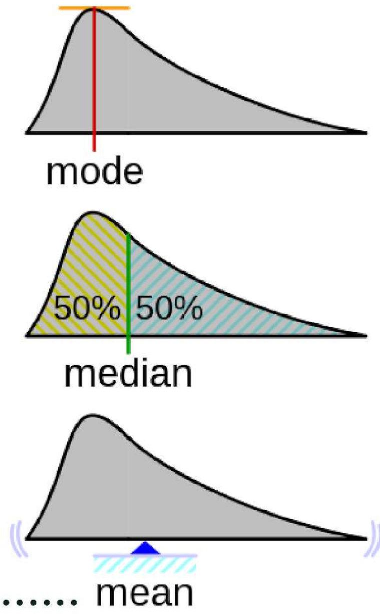


TÀI LIỆU HỌC TẬP

XÁC SUẤT THỐNG KÊ



Lớp..

Họ và tên: ....

MSSV: ....

Chương 1: Các khái niệm cơ bản về xác suất

Nguyễn Minh Trí

Trường Đại học Công nghệ Thông tin

Ngày 26 tháng 2 năm 2023

1.1 Phép thử và biến cố

Định nghĩa 1.1 (Quy tắc nhân) Cho một công việc được thực hiện bởi n giai đoạn. Giai đoạn thứ i có k_i cách thực hiện $(i = 1, 2, \dots, n)$. Số cách thực hiện công việc trên là $k_1 k_2 \dots k_n$.

Ví dụ 1.2 An định mua một laptop và một headphone. Cửa hàng UITLaptop có 5 laptop và 6 headphone phù hợp với yêu cầu của An. Hỏi An có bao nhiêu cách chọn laptop và headphone?

Giải. Số cách chọn laptop và headphone là $5.6 = 30$ (cách).

....

....

Định nghĩa 1.3 (Quy tắc cộng) Một công việc có thể chia thành k trường hợp để thực hiện

- Trường hợp 1 có n_1 cách thực hiện xong công việc
- Trường hợp 2 có n_2 cách thực hiện xong công việc
- $\ldots$
- Trường hợp k có n_k cách thực hiện xong công việc

và không có một cách thực hiện nào ở trường hợp này lại trùng với một cách thực hiện ở trường hợp khác. Khi đó, ta có $n_1 + n_2 + \ldots + n_k$ cách thực hiện xong công việc trên.

Ví dụ 1.4 Có bao nhiêu số tự nhiên chẵn có 3 chữ số khác nhau được tạo thành từ các chữ cố 0, 1, 2, 3, 4?

Giải. Trường hợp 1: Chữ số hàng trăm là chẵn.

- Chữ số hàng trăm: có 2 cách chọn
- Chữ số hàng đơn vị: có 2 cách chọn
- Chữ số hàng chục: có 3 cách chọn

Trường hợp 2: Chữ số hàng trăm là lẻ.

- Chữ số hàng trăm: có 2 cách chọn
- Chữ số hàng đơn vị: có 3 cách chọn
- Chữ số hàng chục: có 3 cách chọn

Số các số thỏa yêu cầu đề bài là $2.2.3 + 2.3.3 = 30$.

Định nghĩa 1.5 (Tổ hợp) Một tổ hợp chập k của n phần tử là một nhóm không phân biệt thứ tự gồm k $(k \le n)$ phần tử khác nhau lấy từ n phần tử đã cho.

Số tổ hợp chập k của n phần tử

$$\boxed{C_n^k = \frac{n!}{k!(n-k)!}.}$$

Ví dụ 1.6 Mỗi đề thi gồm 3 câu hỏi lấy trong 25 câu hỏi cho trước. Hỏi có thể lập được bao nhiêu đề thi có nội dung khác nhau?

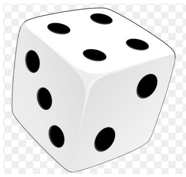
Giải. Mỗi đề thi là một cách chọn 3 câu hỏi từ 25 câu hỏi (không có thứ tự, không lặp lại). Số đề thi là $C_{25}^3 = \dfrac{25!}{3!(25-3)!} = 2300.$

1.2 Phép thử và biến cố

Định nghĩa 1.7

1. Một **phép thử**: Việc thực hiện một nhóm các điều kiện nhất định để quan sát một hiện tượng nào đó. (experiment).
2. **Không gian mẫu** (sample space): Tập hợp tất cả các kết quả (outcome) của một phép thử, ký hiệu Ω.
3. **Biến cố** (event): Một tập con của không gian mẫu. Ký hiệu $A, B, C, \ldots$
4. Biến cố **sơ cấp**: là một kết quả (outcome) của phép thử.
5. Biến cố **phức hợp**: biến cố có thể phân tích thành các biến cố sơ cấp.
6. Biến cố **ngẫu nhiên** là biến cố có thể xảy ra hoặc không thể xảy ra.

Ví dụ 1.8 Phép thử: gieo một con xúc xắc



- Biến cố A: mặt có số chấm chẵn $A = \{2, 4, 6\}$
- Biến cố B: mặt có số chấm là số nguyên tố

$$B = \{2, 3, 5\}$$

- Xuất hiện mặt 1 chấm là biến cố sơ cấp. Xuất hiện mặt có số chấm lẻ là một biến cố phức hợp vì nó có thể phân tích thành các biến cố: Xuất hiện mặt 1 chấm, xuất hiện mặt 3 chấm, xuất hiện mặt 5 chấm.
- Đặt A_i là biến cố "Xuất hiện mặt i chấm" với $i = 1, 2, \ldots, 6$. Không gian mẫu là $\Omega = \{A_1, A_2, A_3, A_4, A_5, A_6\}$.

Ví dụ 1.9 Chạy một chương trình và xem liệu nó có biên dịch được không.

- Phép thử: chạy một chương trình
- Không gian mẫu: $\Omega = \{$biên dich, không biên dịch$\}$
- Biến cố: E "biên dịch"

Ví dụ 1.10 Chạy lại một chương trình máy tính, đo "thời gian CPU" (tính bằng phút), tức là khoảng thời gian mà chương trình cần để được xử lý trong bộ xử lý trung tâm.

- Phép thử: Chạy một chương trình và đo thời gian CPU của nó.
- Không gian mẫu: $\Omega = \{t \in \mathbb{R} \mid t > 0\}$
- Biến cố $E = \{t \in \mathbb{R} \mid t > 5\}$ (thời gian CPU lớn hơn 5 phút)

Định nghĩa 1.11 Cho A, B là các biến cố ngẫu nhiên.

1. Biến cố A được gọi là **thuận lợi** cho biến cố B, ký hiệu $A \subset B$, nếu A xảy ra thì B xảy ra.
2. **Biến cố tích** của A và B là biến cố khi A và B cùng xảy ra, kí hiệu $A \cap B$ hoặc AB.
3. Nếu $A \cap B = \emptyset$ thì ta nói rằng A và B là hai **biến cố xung khắc.**
4. Biến cố **tổng** của A và B là biến cố khi có ít nhất một trong hai biến cố A, B xảy ra, kí hiệu $A \cup B$. Nếu A, B là hai biến cố xung khắc thì biến cố tổng của A và B được kí hiệu là $A + B$.
5. Biến cố **đối** (**đối kháng**) của biến cố A là biến cố xảy ra khi và chỉ khi A không xảy ra, kí hiệu $\overline{A}$.
6. Hiệu của hai biến cố A và B là biến cố khi A xảy ra nhưng B không xảy ra, kí hiệu $A \setminus B$.

Ví dụ 1.12 Phép thử: gieo một con xúc xắc.

- Biến cố A_i "Xuất hiện mặt i chấm," với $i = 1, 2, \ldots, 6$.
- Nếu $i \neq j$ thì A_i và A_j là hai biến cố xung khắc.
- Đặt L là biến cố "Xuất hiện mặt có số chấm lẻ". Các biến cố A_1, A_3, A_5 là thuận lợi đối với L và $L = A_1 + A_3 + A_5$.
- Đặt C là biến cố "Xuất hiện mặt có số chấm chẵn". Khi đó L và C là hai biến cố đối kháng.
- Biến cố đối kháng của biến cố "Xuất hiện mặt có số chấm nhỏ hơn 3" là biến cố "Xuất hiện mặt có số chấm lớn hơn hoặc bằng 3".

Định nghĩa 1.13 **Hệ đầy đủ**: Tập các biến cố $\{A_1, A_2, \ldots, A_n\}$ thỏa mãn hai điều kiện sau:

- Nếu $i \neq j$ thì $A_i \cap A_j = \emptyset$
- $A_1 + A_2 + \cdots + A_n = \Omega$.

Ví dụ 1.14 Một hộp có 3 viên bi màu đỏ, xanh và trắng. Chọn ngẫu nhiên 1 viên. Gọi A: "Chọn được viên màu đỏ", B: "Chọn được viên màu xanh" và C: "Chọn được viên màu trắng". Khi đó $\{A, B, C\}$ là một hệ đầy đủ.

Cho A, B và C là các biến cố của một phép thử. Khi đó

1. $A \cap B = B \cap A; A \cup B = B \cup A$
2. $A \cap (B \cap C) = (A \cap B) \cap C; (A \cup B) \cup C = A \cup (B \cup C)$
3. $A \cap \emptyset = \emptyset; A \cup \emptyset = A; A \cup \Omega = \Omega; A \cap \Omega = A$

1.3 Định nghĩa xác suất

- Xác suất (probability) của một biến cố A là một số nằm giữa 0 và 1, đo khả năng xuất hiện của biến cố A. Ký hiệu là $P(A)$.
- Các biến cố gọi là **đồng khả năng** nếu khi thực hiện phép thử thì chúng có cùng khả năng xảy ra.

Định nghĩa 1.15 Cho một phép thử có n biến cố sơ cấp đồng khả năng xuất hiện và đôi một xung khắc, trong đó có m biến cố sơ cấp thuận lợi đối với biến cố A. Khi đó xác suất của A là

$$P(A) = \frac{m}{n}.$$

Ví dụ 1.16 Xét phép thử là tung một con xúc xắc và quan sát số chấm xuất hiện. Tính xác suất xuất hiện mặt có số chấm lớn hơn 3.

Giải.

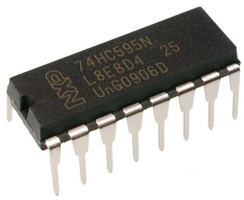
- Biến cố A_i : Xuất hiện mặt i chấm, với $i = 1, 2, 3, \dots, 6$.
- Không gian mẫu $\Omega = \{A_1, A_2, A_3, A_4, A_5, A_6\}$.
- Số biến cố sơ cấp: 6
- Biến cố B: Xuất hiện mặt có số chấm lớn hơn 3.
- Số biến cố sơ cấp thuận lợi đối với B là 3 (A_4, A_5, A_6)
- Xác suất của B là

$$P(B) = \frac{3}{6} = \frac{1}{2}.$$

Ví dụ 1.17 Một hộp chứa 14 chip IC tốt và 6 chip IC lỗi. Chọn ngẫu nhiên 5 chip từ hộp. Tính xác suất:

a. chọn được tất cả đều là chip tốt.

b. chọn được 3 chip tốt và 2 chip lỗi.



. . . .

. . . .

. . . .

. . . .

. . . .

. . . .

. . . .

. . . .

. . . .

. . . .

. . . .

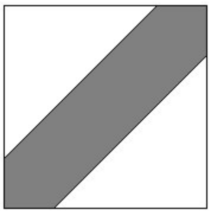
. . . .

. . . .

. . . .

Định nghĩa 1.18 (Định nghĩa theo quan điểm hình học) Giả sử một điểm được rơi ngẫu nhiên vào một miền D và A là một mền con của D. Khi đó xác suất để điểm rơi ngẫu nhiên vào miền A được xác định bởi công thức:

$$P = \frac{\text{"độ đo" miền } A}{\text{"độ đo" miền } D}$$



Từ "độ đo" được hiểu như là **độ dài, diện tích, thể tích** tùy theo từng bài toán cụ thể.

Ví dụ 1.19 Cho một đoạn dây có độ dài 5cm. An dùng kéo cắt ngẫu nhiên tại một vị trí trên đoạn dây. Tính xác suất để An cắt được một đoạn dây có độ dài nhỏ hơn hoặc bằng 1cm.

Giải.

- Biến cố A : Cắt được một đoạn dây có độ dài nhỏ hơn hoặc bằng 1cm.
- Để biến cố A xảy ra, vị trí cắt phải nằm cách một trong hai đầu mút một khoảng cách nhỏ hơn hoặc bằng 1cm.
- Do đó xác suất của biến cố A là

$$P(A) = \frac{2}{5} = 0,4.$$

Định nghĩa 1.20 (Định nghĩa theo quan điểm thống kê) Thực hiện n phép thử độc lập với các điều kiện giống nhau có k lần xuất hiện biến cố A. Tỉ số

$$f(A) = \frac{k}{n}$$

được gọi là **tần suất** xuất hiện biến cố A trong n phép thử.

Định lý 1.21 (Luật số lớn Bernoulli) Khi số phép thử n càng lớn thì tần suất xuất hiện biến cố A tiến về một giá trị xác định. Ta định nghĩa giá trị đó là xác suất của biến cố A.

Ví dụ 1.22 Để nghiên cứu khả năng xuất hiện mặt sấp khi tung một đồng xu, người ta tung một đồng xu nhiều lần (đồng xu không cần cân đối đồng chất nhưng các lần tung phải giống nhau):

Người thí nghiệm	Số lần tung (n)	Số mặt sấp (k)	Tần suất $\frac{k}{n}$
Buffon	4040	2048	0,5069
Pearson	12000	6019	0,5016
Pearson	24000	12012	0,5005

Ta thấy khi số phép thử tăng lên thì tần suất xuất hiện mặt sấp dao động quanh giá trị 0,5. Điều này cho phép ta hy vọng rằng khi số phép thử tăng lên vô hạn thì tần suất xuất hiện mặt sấp hội tụ về 0,5.

Nhận xét.

- Định nghĩa xác suất theo quan điểm thống kê của xác suất khắc phục được một nhược điểm của định nghĩa cổ điển là không dùng đến khái niệm đồng khả năng. Nó hoàn toàn dựa trên các quan sát thực tế để làm cơ sở kết luận về xác suất xảy ra của một biến cố
- Chỉ áp dụng được đối với các hiện tượng ngẫu nhiên mà tần suất của nó có tính ổn định. Hơn nữa, để xác định một cách tương đối chính xác giá trị của xác suất, ta phải tiến hành trên thực tế một số lượng đủ lớn các phép thử.

Định nghĩa 1.23 (Định nghĩa xác suất theo hệ tiên đề) Cho một phép thử có không gian mẫu Ω. Xác suất là một ánh xạ $P : \Omega \to [0, 1]$ thỏa mãn các điều kiện:

1. Với mỗi biến cố A, ta có $0 \leq P(A) \leq 1$
2. $P(\Omega) = 1$
3. Với bất kì họ đếm được các biến cố đôi một xung khắc $A_1, A_2, \ldots$ ta có
$$P(A_1 \cup A_2 \cup \ldots) = \sum_{i \geq 1} P(A_i)$$

Năm 1933, nhà toán học Andrey Nikolaevich Kolmogorov (1903 – 1987) đã đưa ra định nghĩa xác suất theo phương pháp tiên đề. Ông là một nhà toán học Liên Xô đã có nhiều đóng góp lớn trong lý thuyết xác suất và tôpô.
Nguồn hình:
https://www.kolmogorov.com/Kolmogorov.html

Tính chất. Xét không gian mẫu Ω và các biến cố A, B, C.

1. $0 \leq P(A) \leq 1$
2. $P(\emptyset) = 0$
3. $P(\Omega) = 1$
4. $P(\overline{A}) = 1 - P(A)$
5. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
6. Nếu A, B xung khắc thì $P(A \cup B) = P(A) + P(B)$
7. $P(A \cup B \cup C) =$
$P(A)+P(B)+P(C)-P(A\cap B)-P(B\cap C)-P(A\cap C)+P(A\cap B\cap C)$

Ví dụ 1.24 Một hệ thống máy tính sử dụng mật khẩu gồm 5 ký tự, mỗi ký tự là một trong 26 chữ cái (a-z) hoặc 10 chữ số (0-9). Ký tự đầu tiên phải là một chữ cái. Giả sử một hacker chọn một mật khẩu ngẫu nhiên. Xác định xác suất sao cho mật khẩu

a. bắt đầu bằng một nguyên âm (a, e, o, i ,u)
b. kết thúc là một số lẻ
c. bắt đầu bằng nguyên âm và kết thúc bằng một số lẻ.

Giải. a. Biến cố A : Mật khẩu bắt đầu bằng một nguyên âm

$$P(A) = \frac{5}{26}$$

b. Biến cố B: Mật khẩu kết thúc bằng một số lẻ

$$P(B) = \frac{5}{36}$$

c. Cần tính $P(A \cup B)$. Theo công thức

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

Biến cố $A \cap B$: Mật khẩu bắt đầu bằng một nguyên âm và kết thúc bằng một số lẻ

$$P(A \cap B) = \frac{5.36^3.5}{26.36^4} = \frac{25}{26.36}$$

Suy ra

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{5}{26} + \frac{5}{36} - \frac{25}{26.36} = 0,3045$$

Ví dụ 1.25 Một lớp học có 20 sinh viên trong đó có 10 sinh viên biết lập trình C++, 8 sinh viên biết lập trình Java và 6 sinh biết lập trình C++ và Java. Chọn ngẫu nhiên một sinh viên.
a. Tính xác suất để sinh này biết ít nhất một ngôn ngữ lập trình.
b. Tính xác suất để sinh viên này không biết ngôn ngữ lập trình nào.
Giải. ....

....

....

....

....

....

....

....

....

....

....

1.4 Xác suất có điều kiện. Công thức nhân xác suất

Bài toán: Tính xác suất biến cố A xảy ra khi biến cố B đã xảy ra.

Định nghĩa 1.26 (Xác suất có điều kiện) Giả sử trong một phép thử, ta có $P(B) > 0$. Xác suất của A với điều kiện B, ký hiệu là $P(A|B)$,

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

Ví dụ 1.27 Tại UIT, hàng tuần có 200 chương trình được viết bằng Python hoặc JavaScript, với các số liệu biên dịch sau

	Biên dịch được khi chạy lần đầu	Không biên dịch được khi chạy lần đầu	
Python	72	48	120
JavaScript	64	16	80
Tổng	136	64	200

Chọn ngẫu nhiên một chương trình, tính xác suất chương trình đó biên dịch được trong lần chạy đầu tiên biết rằng nó được viết bằng Python?

Giải.

- Biến cố T : chương trình viết bằng Python
- Biến cố J : chương trình viết bằng JavaScript
- Biến cố C : chương trình biên dịch được trong lần đầu chạy
- Tìm $P(C|T)$.
- Ta có
$$P(C|T) = \frac{P(C \cap T)}{P(T)}$$
- Biến cố $C \cap T$: chương trình được viết bằng Python và biên dịch được trong lần chạy đầu tiên.
- Từ bảng số liệu
$$P(C \cap T) = \frac{72}{200} \text{ và } P(T) = \frac{120}{200}$$
- Do đó
$$P(C|T) = \frac{72/200}{120/200} = \frac{3}{5}$$

Ví dụ 1.28 Một lớp học có 56 sinh viên, trong đó có 26 nam và 30 nữ. Kết quả thi môn Xác suất thống kê có 12 sinh viên đạt điểm giỏi (7 nam và 5 nữ). Chọn ngẫu nhiên một sinh viên trong lớp.

a. Tính xác suất để chọn được sinh viên đạt điểm giỏi.
b. Tính lại xác suất để chọn được sinh viên đạt điểm giỏi biết rằng sinh viên đó là nữ.

Giải. ....

....

....

....

....

....

....

....

....

....

....

....

....

....

....

....

....

....

....

....

....

....

....

....

Định nghĩa 1.29 (Biến cố độc lập) Hai biến cố A và B được gọi là **độc lập** nếu

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

Định lý 1.30
Cho A, B là các biến cố của một phép thử. Các điều sau đây là tương đương:

1. A và B độc lập (tức là $P(A \cap B) = P(A)P(B)$);
2. $P(B|A) = P(B)$;
3. $P(A|B) = P(A)$.

Nhận xét. Cho A, B là hai biến cố độc lập. Khi đó

1. $\overline{A}$ và B là hai biến cố độc lập.
2. A và $\overline{B}$ là hai biến cố độc lập.
3. $\overline{A}$ và $\overline{B}$ là hai biến cố độc lập.

Ví dụ 1.31 Trong một thành phố nọ có 51% nam và 49% nữ, và tỷ lệ sử dụng Facebook của họ được thể hiện trong bảng xác suất dưới đây:

	Nam	Nữ	Tổng cộng
Sử dụng Facebook	0,04	0,002	0,042
Không sử dụng Facebook	0,47	0,488	0,958
Tổng	0,51	0,49	1,00

a. Gặp ngẫu nhiên một người nam. Tính xác suất người nam này sử dụng Facebook.
b. Biến cố gặp người nam và biến cố gặp người sử dụng Facebook có độc lập không?

Giải. a. Biến cố A : Gặp người nam.
Biến cố B : Gặp người sử dụng Facebook. Ta cần tính $P(B|A)$.

. . . .

. . . .

. . . .

. . . .

Định lý 1.32 Cho $A_1, A_2, \dots, A_n$ là các biến cố bất kì.

1. Công thức nhân xác suất: $P(A_1 \cap A_2) = P(A_1)P(A_2|A_1)$
2. $P(A_1 \cap A_2 \cap \dots \cap A_n) =$
   $P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2)\dots P(A_n|A_1 \cap \dots \cap A_{n-1})$
3. Nếu $A_1, A_2, \dots, A_n$ là các biến cố đôi một độc lập nhau thì

$$P(A_1 \cap A_2 \cap \dots \cap A_n) = P(A_1)P(A_2)\dots P(A_n).$$

Ví dụ 1.33 Một đội 15 sinh viên trong đó có 5 sinh viên biết thiết kế web. Chia đội này thành 5 nhóm, mỗi nhóm 3 người. Tính xác suất nhóm nào cũng có một sinh viên biết thiết kế web.
Giải. Biến cố A : Nhóm nào cũng có một sinh viên biết thiết kế web;
Biến cố A_i : Nhóm i có một sinh viên học biết thiết kế web, $i = 1, \dots, 5$.
Khi đó $A = A_1A_2A_3A_4A_5$. Sử dụng công thức nhân

$P(A) =$ ....

....

1.5 Xác suất toàn phần. Công thức Bayes

Cho hệ đầy đủ các biến cố $\{A_1, A_2, \dots, A_n\}$ của một phép thử và B là một biến cố bất kì.
Công thức xác suất đầy đủ

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

Từ công thức xác suất có điều kiện, ta có

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

Định lý 1.34 (Định lý Bayes) Cho hệ đầy đủ các biến cố $\{A_1, A_2, \dots, A_n\}$ và B là một biến cố bất kì với $P(B) \neq 0$. Khi đó

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

Ví dụ 1.35 Trong một đại học nọ, người ta thấy rằng tỷ lệ sinh viên có chơi game là 30% và số sinh viên bị cận thị trong số những sinh viên chơi game là 60%, tỷ lệ sinh viên bị cận thị trong số sinh viên không chơi game là 40%.

a. Chọn ngẫu nhiên 1 sinh viên. Tính xác suất sinh viên đó bị cận thị.
b. Chọn ngẫu nhiên 1 sinh viên, biết rằng người đó bị cận thị. Tính xác suất sinh viên đó có chơi game.

Giải. Biến cố A : Chọn sinh viên bị cận thị.
Biến cố B_1 : Chọn sinh viên có chơi game
Biến cố B_2 : Chọn sinh viên không chơi game
Hệ biến cố $\{B_1, B_2\}$ là đầy đủ.

. . . .

. . . .

. . . .

. . . .

. . . .

. . . .

. . . .

. . . .

. . . .

Ví dụ 1.36 Hai nhà máy cùng sản xuất một loại sản phẩm. Tỉ lệ sản phẩm tốt của nhà máy thứ nhất là 0,9; của nhà máy thứ hai là 0,85. Từ một kho chứa 1/3 số sản phẩm của nhà máy thứ nhất và 2/3 số sản phẩm còn lại của nhà máy thứ hai, người ta lấy ra 1 sản phẩm để kiểm tra.

a. Tính xác suất để lấy được sản phẩm không tốt.
b. Giả sử sản phẩm lấy là sản phẩm tốt. Tính xác suất sản phẩm đó do nhà máy thứ 2 sản xuất.

Giải. Đặt A_i là "Sản phẩm do nhà máy i sản suất" với $i = 1, 2$. Khi đó A_1, A_2 là một hệ đầy đủ.
Đặt A là "Sản phẩm không tốt".

. . . .

. . . .

. . . .

. . . .

. . . .

. . . .

. . . .

. . . .

. . . .

Ví dụ 1.37 Tại một trường đại học nọ, có 50% các văn bản viết bằng Word, 30% được viết bằng Latex và 20% viết bằng HTML. Người ta thấy rằng

- có 40% văn bản Word có nhiều hơn 10 trang.
- có 20% văn bản Latex có nhiều hơn 10 trang.
- có 30% văn bản HTML có nhiều hơn 10 trang.

Chọn ngẫu nhiên một văn bản, tính xác suất văn bản đó có nhiều hơn 10 trang.

Giải. . . . .

. . . .

. . . .

. . . .

. . . .

. . . .

. . . .

. . . .

. . . .

BÀI TẬP

Bài 1.1 Hai xạ thủ mỗi người bắn một viên đạn vào cùng một bia. Xác suất trúng đích của người thứ nhất là 0,9 và của người thứ hai là 0,7. Tính các xác suất của biến cố:
a. Có đúng một phát trúng.
b. Cả hai phát đều trúng.
c. Có ít nhất một phát trúng.

Bài 1.2 Cho đoạn thẳng AB có độ dài 10 cm. Lấy một điểm C bất kỳ trên đoạn thẳng đó. Tính xác suất chênh lệch độ dài giữa hai đoạn thẳng AC và CB không vượt quá 4cm. (0,4)

Bài 1.3 Có 4 cục pin trong đó có một pin bị lỗi. Lấy ngẫu nhiên 2 cục pin để sử dụng. Tìm xác suất để pin thứ hai được chọn không bị lỗi, biết rằng pin thứ nhất không bị lỗi. (2/3)

Bài 1.4 Một công ty có 70% nhân viên biết C++, 60% nhân viên biết Python và 50% nhân viên biết cả hai ngôn ngữ này. Tỉ lệ nhân viên biết ít nhất một trong hai ngôn ngữ là bao nhiêu? Giả sử một nhân viên biết C++, hỏi xác suất nhân viên đó cũng biết Python là bao nhiêu?

Bài 1.5 Kiểm tra 20 bài tập xác suất thống kê của Bình, giáo viên thấy có 4 bài sai. Giáo viên chọn ngẫu nhiên 2 bài để chấm lấy điểm. Tìm xác suất
a. Cả 2 bài đều đúng. (0,623)
b. Có ít nhất một bài sai. (0,368)
c. Cả hai bài đều đúng biết rằng có ít nhất một bài đúng. (0,652)
Bài 1.6 Sinh viên tại một trường đại học có hai cơ hội để vượt qua một kỳ thi tốt nghiệp. Xác suất một sinh viên thi đậu lần thứ nhất là 0,8. Đối với những người trượt lần thứ nhất, xác suất đậu lần thứ hai là 0,6. Chọn ngẫu nhiên 1 sinh viên.
a. Tìm xác suất sinh viên đó thi đậu. (0,92)
b. Nếu sinh viên đó vượt qua bài thi thì xác suất sinh viên đó thi đậu trong lần thi đầu tiên là bao nhiêu? (0,87)
Bài 1.7 Huy quyết tâm thi ôn xác suất thống kê đạt điểm giỏi. Huy biết rằng nếu anh ấy ôn tập sớm hơn 2 tuần trước kì thi thì xác suất đạt điểm giỏi là 0,8. Nếu không ôn bài sớm hơn 2 tuần thì xác suất đạt điểm giỏi là 0,3. Tuy nhiên, vì bận học nhiều môn nên xác suất Huy ôn bài ít hơn 2 tuần là 40%. Tính xác suất để Huy đạt được điểm giỏi. (0,5)
Bài 1.8 Dũng có 3 tài khoản email. Dũng thấy rằng có 70% email được gửi đến tài khoản thứ 1, 20% email được gửi đến tài khoản thứ 2 và 10% email được gửi đến tài khoản thứ 3. Hơn nữa, các email rác gửi đến tài khoản 1, 2 và 3 lần lượt là 1%, 2% và 5%. Tính xác suất để chọn ngẫu nhiên được một email rác. (0,016)
Bài 1.9 Có một nhóm 3 sinh viên, mỗi người có một laptop giống hệt nhau để trên bàn. Mỗi sinh viên lấy ngẫu nhiên một chiếc laptop. Tính xác suất để
a. Sinh viên thứ nhất và sinh viên thứ hai lấy đúng laptop của mình.
b. Có ít nhất một sinh viên lấy đúng laptop của mình.

Bài 1.10 Cửa hàng cafe UIT-Cof bán 2 loại cafe: Thường và đặc biệt với bảng thông tin như sau

	Gói nhỏ	Gói trung	Gói lớn
Thường	14%	20%	26%
Đặc biệt	20%	10%	10%

Chọn ngẫu nhiên một gói cafe.
a. Tính xác suất để chọn được gói nhỏ. (0,34)
b. Tính xác suất để chọn được gói đặc biệt. (0,4)
c. Nếu chọn được một gói nhỏ thì xác suất chọn loại đặc biệt là bao nhiêu? (0,588)
d. Nếu chọn được một gói loại đặc biệt thì xác suất chọn gói nhỏ là bao nhiêu? (0,5)
Bài 1.11 Một hộp chứa 6 bi đỏ và 4 bi xanh, và hộp thứ hai chứa 7 bi đỏ và 3 bi xanh. Một bi được chọn ngẫu nhiên từ hộp thứ nhất và bỏ vào hộp thứ hai. Sau đó, một viên bi được chọn ngẫu nhiên từ hộp thứ hai và bỏ vào hộp thứ nhất.
a. Xác suất để một bi đỏ được chọn từ hộp thứ nhất và bi đỏ được chọn từ hộp thứ hai là bao nhiêu? (0,436)
b. Khi kết thúc quá trình lựa chọn, xác suất để số bi đỏ và bi xanh ở hộp thứ nhất bằng với số bi lúc đầu là bao nhiêu? (0,581)
Bài 1.12 Các đề tài nghiên cứu của một nhóm sinh viên được đánh giá bởi một hội đồng chuyên gia để quyết định xem chúng có xứng đáng được tài trợ hay không. Khi đề tài này được đệ trình lên một hội đồng chuyên gia độc lập thứ hai, có 30% số đề tài không được duyệt tài trợ. Giả sử xác suất mà một đề tài được hội đồng đầu tiên đánh giá là xứng đáng được cấp kinh phí là 20%. Tính xác suất của những biến cố sau
a. Một đề tài được chấp thuận bởi cả hai hội đồng. (0,14)
b. Một đề tài bị cả hai hội đồng từ chối. (0,56)
c. Một đề tài được phê duyệt bởi một hội đồng. (0,3)

Bài 1.13 Có hai lô hàng: lô thứ nhất có 7 sản phẩm loại A và 3 sản phẩm loại B; lô thứ hai có 8 sản phẩm loại A và 4 sản phẩm loại B.
a. Lấy ngẫu nhiên một lô hàng, rồi từ lô hàng đó lấy ra 2 sản phẩm. Tính xác suất trong 2 sản phẩm lấy ra có loại A.
b. Nếu 2 sản phẩm lấy ra có loại A thì xác suất để các sản phẩm loại A đó thuộc lô thứ 2 là bao nhiêu?
Bài 1.14 Một cửa hàng bán 3 loại laptop trong đó có 50% hiệu A, 30% hiệu B và 20% hiệu C. Các laptop đều được bảo hành một năm. Cửa hàng thấy rằng, có 25% laptop hiệu A, 20% laptop hiệu B và 10% laptop hiệu C bị hư trong thời gian bảo hành. Chọn ngẫu nhiên một laptop.
a. Tính xác suất chọn được một laptop hiệu A bị hư trong thời gian bảo hành. (0,125)
b. Tính xác suất chọn được một laptop bị hư trong thời gian bảo hành. (0,205)
c. Nếu có một laptop bị hư trong thời gian bảo hành thì xác suất laptop đó hiệu A là bao nhiêu? (0,61)
Bài 1.15 Lỗi máy in có liên quan đến ba loại sự cố: phần cứng, phần mềm và kết nối điện. Một nhà sản xuất máy in đã tính được các xác suất sau từ cơ sở dữ liệu về kết quả kiểm tra. Xác suất xảy ra sự cố liên quan đến phần cứng, phần mềm hoặc kết nối điện lần lượt là 0,1, 0,6 và 0,3. Xác suất lỗi máy in khi có sự cố phần cứng là 0,9, do sự cố phần mềm là 0,2 và do sự cố về điện là 0,5. Nếu một khách hàng thấy máy in của mình có lỗi thì nguyên nhân có khả năng nhất của sự cố là gì?

Index

Chương 2. Biến ngẫu nhiên

Nguyễn Minh Trí

Trường Đại học Công nghệ Thông tin

Ngày 26 tháng 2 năm 2023

2.1 Biến ngẫu nhiên

Định nghĩa 2.1 (Biến ngẫu nhiên) Cho một phép thử có không gian mẫu là Ω. Biến ngẫu nhiên X (random variable) là một phép gán một kết quả (outcome) của Ω với một số.

Ví dụ 2.2 Xét một phép thử gieo 3 đồng xu và tính số mặt sấp. Đặt X là số mặt sấp. Ta thấy X nhận các giá trị 0,1,2,3.

Ví dụ 2.3 Xét một phép thử bắn một viên đạn vào một tấm bia. Kết quả của phép thử là "A: Bắn trúng tấm bia" và "B: Bắn không trúng tấm bia". Ta đặt $X = 1$ nếu A xảy ra và $X = 0$ nếu B xảy ra. Khi đó X là biến ngẫu nhiên nhận các giá trị 0, 1.

Ví dụ 2.4 Gieo một con xúc xắc cho đến khi xuất hiện mặt 1 chấm. Đặt Y là số lần gieo xúc xắc. Khi đó $Y = 1, 2, 3, \ldots$

Ví dụ 2.5 Đặt T là thời gian (tính bằng giờ) tính từ bây giờ đến khi xuất hiện lần mất điện tiếp theo tại TPHCM. Khi đó T là một biến ngẫu nhiên và có giá trị là các số thực dương thuộc $[0, +\infty)$.

Định nghĩa 2.6

1. Biến ngẫu nhiên được gọi là **rời rạc** nếu tập giá trị của nó là một tập hữu hạn hoặc *vô hạn đếm được*[1] các phần tử.
2. Biến ngẫu nhiên được gọi là **liên tục** nếu tập giá trị của nó lấp kín một khoảng trên trục số.

Ví dụ 2.7

1. Đặt X là số chấm xuất hiện khi gieo một con xúc xắc. Khi đó X nhận các giá trị là 1,2,3,4,5,6. Do đó, X là một biến ngẫu nhiên rời rạc. (hữu hạn)
2. Gieo một con xúc xắc cho đến khi xuất hiện mặt 1 chấm. Đặt Y là số lần gieo xúc xắc. Khi đó Y là một biến ngẫu nhiên rời rạc và nhận các giá trị $1, 2, 3, \dots$ (vô hạn đếm được)
3. Đặt Y là nhiệt độ trong một ngày ở TPHCM. Khi đó Y là một số ngẫu nhiên từ 0^oC đến 50^oC. Do đó Y là một biến ngẫu nhiên liên tục.

[1] Một tập hợp là vô hạn đếm được nếu ta có thể gán mỗi phần tử của nó với một số tự nhiên

2.2 Biến ngẫu nhiên rời rạc

Định nghĩa 2.8 Cho X là một biến ngẫu nhiên rời rạc có miền giá trị là $\{x_1, x_2, \dots\}$.

1. Tập hợp các xác suất $P(X = x)$ của X được gọi là **phân phối** (distribution) của X

X	x_1	x_2	$\dots$
$P(X = x_i)$	p_1	p_2	$\dots$

2. **Hàm khối xác suất** (probability mass function), viết tắt *pmf*,
$$p(x) = P(X = x)$$
3. **Hàm phân phối xác suất tích luỹ** (cumulative distribution function), viết tắt *cdf*
$$F(x) = P(X \leq x).$$

Cho X là một biến ngẫu nhiên rời rạc và $x_1 < x_2 < \dots$ thì
$$F(x_n) = \sum_{i=1}^{n} P(X = x_i).$$

Ví dụ 2.9 Xét phép thử là gieo một con xúc xắc. Đặt X là số chấm xuất hiện và X nhận các giá trị nguyên từ 1 đến 6.

- Phân phối xác suất của X là

X	1	2	3	4	5	6
$P(X = x_i)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

- Hàm phân phối xác suất tích lũy của X

$$F(x) = \begin{cases} 0, & x < 1 \\ \frac{1}{6}, & 1 \leq x < 2 \\ \frac{2}{6}, & 2 \leq x < 3 \\ \frac{3}{6}, & 3 \leq x < 4 \\ \frac{4}{6}, & 4 \leq x < 5 \\ \frac{5}{6}, & 5 \leq x < 6 \\ 1, & 6 \leq x \end{cases}$$

Ví dụ 2.10 Một xạ thủ chỉ có 3 viên đạn. Anh ta được yêu cầu bắn từng phát cho đến khi trúng mục tiêu thì dừng bắn, biết rằng xác suất trúng của mỗi lần bắn là 0,6. Tìm phân phối xác suất của số đạn cần bắn.

Giải. Đặt X là số đạn cần bắn. Khi đó X là một biến ngẫu nhiên rời rạc và nhận các giá trị 1, 2, 3.

- Nếu $X = 1$ thì người đó bắn viên thứ nhất trúng và do đó $P(X = 1) = 0,6$.
- Nếu $X = 2$ thì viên thứ nhất bắn không trúng và viên thứ hai bắn trúng. Do đó $P(X = 2) = 0,4.0,6 = 0,24$.
- Nếu $X = 3$ thì 2 viên đạn đầu không trúng và viên thứ 3 có thể trúng hay không trúng. Do đó $P(X = 3) = 0,4.0,4 = 0,16$.

Phân phối xác suất của X là

X	1	2	3
$P(X = x_i)$	$0,6$	$0,24$	$0,16$

Ví dụ 2.11 Gieo một con xúc xắc và gọi X là số lần gieo đến khi gặp mặt 6 chấm. Lập bảng phân phối xác suất của X.

Giải. ....

....

....

....

....

....

....

Ví dụ 2.12 Cho X là một biến ngẫu nhiên rời rạc có bảng phân phối xác suất như sau

X	0,2	0,4	0,5	0,8	1
$P(X = x)$	0,1	0,2	0,2	0,3	0,2

a. Tính $P(X \leq 0,5)$
b. Tính $P(0,25 < X < 0,75)$
c. Tính $P(X = 0,2 | X < 0,6)$

Giải. ....

....

....

....

....

....

....

....

....

....

....

....

Ví dụ 2.13 Có hai lô hàng: lô 1 có 10 sản phẩm trong đó có 3 sản phẩm loại A; lô 2 có 12 sản phẩm trong đó có 4 sản phẩm loại A.

a. Lấy ngẫu nhiên mỗi lô một sản phẩm. Gọi X là số sản phẩm loại A lấy được. Lập bảng phân phối xác suất của X.

b. Lấy ngẫu nhiên một lô hàng, từ đó lấy ngẫu nhiên 3 sản phẩm. Đặt Y là số sản phẩm loại A lấy được. Lập bảng phân phối xác suất của Y.

Giải. a. Tập giá trị của X là $\{0, 1, 2\}$.

- $P(X = 0)$ là xác suất lấy 2 sản phẩm không có sản phẩm loại A

$$P(X = 0) = \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots$$

- $P(X = 2)$ là xác suất lấy 2 sản phẩm có 2 sản phẩm loại A

$$P(X = 2) = \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots$$

- $P(X = 1)$ là xác suất lấy 2 sản phẩm có 1 sản phẩm loại A

$$P(X = 1) = \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots$$

....

....

....

....

b. Dùng công thức xác suất toàn phần.
Gọi A_i là biến cố "Lấy lô i" với $i = 1, 2$. Ta có

$$P(A_1) = P(A_2) = \dots\dots\dots$$

....

....

....

....

....

....

....

....

. . . .

. . . .

. . . .

. . . .

. . . .

. . . .

. . . .

. . . .

. . . .

. . . .

. . . .

. . . .

. . . .

. . . .

. . . .

2.3 Biến ngẫu nhiên liên tục

Ví dụ 2.14 Chọn X là một số ngẫu nhiên trong đoạn $[0, 10]$. Tính xác suất

a. $P(3 \leq X < 6)$

b. $P(3 < X < 6)$

c. $P(X = 5,4321)$

Giải. . . . .

. . . .

. . . .

. . . .

. . . .

. . . .

. . . .

. . . .

. . . .

- Nếu X là một biến ngẫu nhiên liên tục thì $P(X = a) = ...$ với mọi a.
- Do đó hàm khối xác suất $p(x) = 0$ với mọi x.
- Hàm phân phối xác suất tích lũy (cdf)

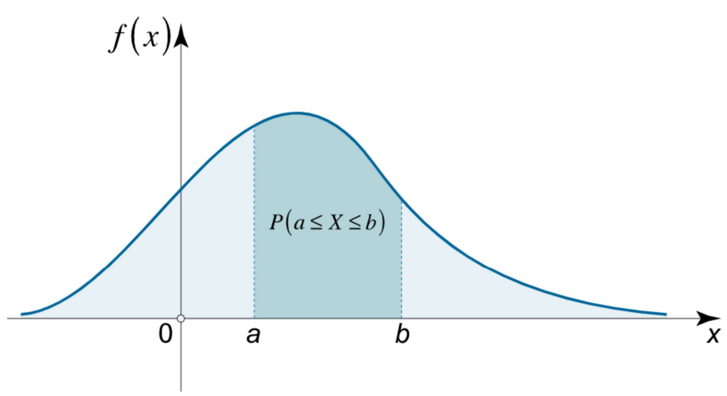
$$F(x) = P(X \leq x) = P(X < x)$$

Định nghĩa 2.15 Hàm mật độ xác suất (probability density function) của một biến ngẫu nhiên liên tục X là một hàm số $f(x)$ thỏa mãn

1. $f(x) \geq 0$,
2. $\displaystyle\int_{-\infty}^{+\infty} f(x)dx = 1$,
3. $\displaystyle P(a \leq X \leq b) = \int_a^b f(x)dx$

Nhận xét.

- Đồ thị của hàm mật độ không nằm dưới trục hoành.
- Diện tích giới hạn bởi đồ thị của $f(x)$ và trục hoành bằng 1.
- Xác suất của của biến ngẫu nhiên X thuộc đoạn $[a, b]$ bằng diện tích hình giới hạn bởi của đồ thị $f(x)$, trục hoành và hai đường thẳng $x = a; x = b$.



Định lý 2.16 Cho X là một biến ngẫu nhiên liên tục có hàm mật độ xác suất $f(x)$ và hàm phân phối xác suất tích lũy $F(x)$. Khi đó

1. $\displaystyle P(X = a) = \int_a^a f(x)dx = 0$
2. $P(a \leq X \leq b) = P(a < X \leq b) = P(a \leq X < b) = P(a < X < b)$.
3. $\displaystyle F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt.$
4. $\displaystyle P(X > a) = 1 - P(X \leq a) = \int_a^{+\infty} f(x)dx.$
5. $f(x) = F'(x)$

Ví dụ 2.17 Thời gian sử dụng của một thiết bị điện tử (tính bằng năm) là một biến ngẫu nhiên liên tục có hàm mật độ xác suất

$$f(x) = \begin{cases} \frac{k}{x^3}, & x \geq 1 \\ 0, & x < 1 \end{cases}$$

Tìm k, hàm phân phối xác suất và xác suất để thiết bị đó sử dụng được hơn 5 năm.

Giải.

- Tìm k. Ta có

$$1 = \int_{-\infty}^{+\infty} f(x)dx = \int_{1}^{+\infty} f(x)dx = \text{..}$$

  Do đó $k =$ ....

- Hàm phân phối xác suất tích lũy

$$F(x) = \int_{-\infty}^{x} f(t)dt = \int_{1}^{x} \frac{k}{t}dt = \text{..}$$

. . . .

. . . .

. . . .

. . . .

- Xác suất để thiết bị đó sử dụng được nhiều hơn 5 năm là

$$P(X > 5) = \int_{5}^{+\infty} f(x)dx = \text{...}$$

  Ta có thể sử dụng

$$P(X > 5) = 1 - F(5) = \text{.......................................}$$

Ví dụ 2.18 Thời gian phục vụ mỗi khách hàng tại một cửa hàng là một biến ngẫu nhiên liên tục X (phút) có hàm mật độ xác suất như sau

$$f(x) = \begin{cases} k.e^{-4x}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

a. Tìm k.
b. Tìm hàm phân phối xác suất của X.
c. Tính xác suất thời gian phục vụ một khách hàng trong khoảng thời gian từ 1 đến 2 phút.

Giải. ....

....

....

....

....

....

....

....

....

....

....

Ví dụ 2.19 Cho một biến ngẫu nhiên X là thời gian sử dụng của một loại pin (tính bằng năm). Giả sử hàm mật độ của X được cho bởi

$$f(x) = \begin{cases} \frac{1}{2}e^{-x/2} & x > 0 \\ 0, & x \leq 0. \end{cases}$$

a. Tính xác suất loại pin này sử dụng được ít nhất 2 năm hoặc nhiều hơn 4 năm.
b. Giả sử có một viên pin loại này đã được sử dụng 2 năm. Tính xác suất pin này được sử dụng ít nhất 3 năm.

Giải.

Ví dụ 2.20 Đặt X là thời gian (tháng) sử dụng của một loại thiết bị điện tử. Giả sử biến ngẫu nhiên X có hàm phân phối xác suất

$$F(x) = 1 - e^{-3x} \text{ với } x > 0.$$

a. Chọn ngẫu nhiên một thiết bị. Tính xác suất thiết bị đó sử dụng được ít nhất 12 tháng.
b. Tìm hàm mật độ của X.

Giải.

2.4 Các tham số của biến ngẫu nhiên

Định nghĩa 2.21 **Kỳ vọng** (expectation) của biến ngẫu nhiên X, ký hiệu $E(X)$, được xác định như sau:

- Nếu X là biến ngẫu nhiên rời rạc thì

$$E(X) = \sum_{i \geq 1} x_i P(X = x_i)$$

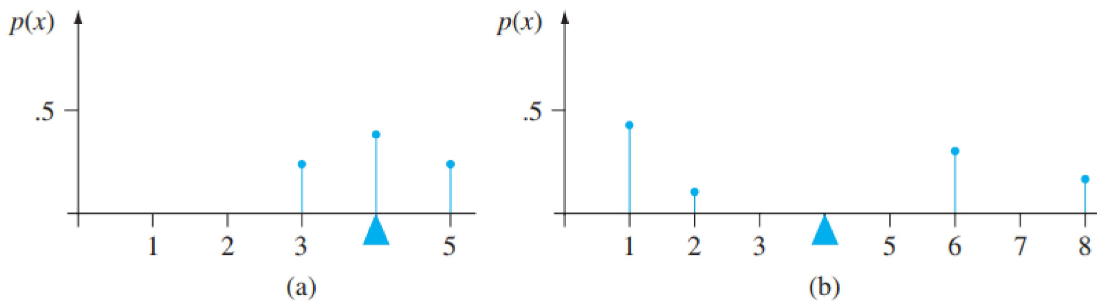
- Nếu X là biến ngẫu nhiên liên tục có hàm mật độ xác suất $f(x)$ thì

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

Kỳ vọng còn được gọi là giá trị trung bình (Mean), giá mong đợi (Expected Value).
Quan sát 2 hình dưới đây biểu diễn phân phối xác suất của 2 biến ngẫu nhiên. Giả sử tại vị trí tam giác là kỳ vọng của biến ngẫu nhiên.

X	3	4	5
$P(X = x_i)$	0,3	0,4	0,3

Y	1	2	6	8
$P(Y = y_i)$	0,4	0,1	0,3	0,2



Cả hai phân phối xác xuất có cùng kỳ vọng $E(X)$ nhưng phân phối của Hình (b) có độ phân tán (tức là độ biến thiên) lớn hơn so với của Hình (a). Chúng ta sẽ sử dụng **phương sai** (variance) của X để đánh giá mức độ phân tán trong (phân phối của) X.

Định nghĩa 2.22

- **Phương sai** (variance) của biến ngẫu nhiên X, kí hiệu $\text{Var}(X)$ hoặc σ^2,
  1. Nếu X là biến ngẫu nhiên rời rạc có miền giá trị $\{x_1, x_2, \dots\}$ thì

  $$\text{Var}(X) = \sigma^2 = \sum_{i=1}^{\infty} (x_i - E(X))^2 P(X = x_i) = E((X - E(X))^2)$$

  2. Nếu X là biến ngẫu nhiên liên tục có hàm mật độ xác suất $f(x)$ thì

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

- **Độ lệch chuẩn** (standard deviation) của X

  $$\text{Std}(X) = \sqrt{\text{Var}(X)} = \sigma.$$

- **Mode** của biến ngẫu nhiên X, ký hiệu Mod(X), là một giá trị của biến ngẫu nhiên X mà tại đó biến ngẫu nhiên X nhận xác suất lớn nhất (nếu X là rời rạc) hoặc tại đó hàm mật độ đạt giá trị lớn nhất (nếu X là biến liên tục).

Ví dụ 2.23 Đặt X là lượng cát (tính bằng tấn) do công ty CAT bán ra hàng tuần. Giả sử biến ngẫu nhiên X có hàm mật độ xác suất như sau

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

Giải.

- Kỳ vọng của X

  ....

  ....

  ....

- Phương sai của X

  ....

  ....

  ....

Ví dụ 2.24 Người ta thấy rằng xác suất một chương trình do lập trình viên Z biên dịch được là 25%. Mỗi ngày, lập trình viên Z viết 5 chương trình. Đặt X là số chương trình biên dịch được. Tính trung bình số chương trình biên dịch được của lập trình viên Z trong một ngày.

Giải. X nhận các giá trị . . . .

Xác suất

$$P(X = x) = \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

Bảng phân phối xác suất của X

. . . .

. . . .

Kỳ vọng

. . . .

. . . .

Định lý 2.25

1. Nếu X là một biến ngẫu nhiên rời rạc và $g(x)$ là một hàm số thì

$$E(g(X)) = \sum_x g(x)P(X = x)$$

2. Nếu X là một biến ngẫu nhiên liên tục có hàm mật độ xác suất $f(x)$ và $g(x)$ là một hàm số thì

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

3. $E(aX + b) = aE(X) + b$ với $a, b \in \mathbb{R}$
4. $\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$ với $a, b \in \mathbb{R}$
5. $\mathrm{Var}(X) = E(X^2) - E(X)^2$

Ví dụ 2.26 Một trung tâm bảo hành laptop thấy rằng số laptop hiệu Potpal cần sửa trong năm tối đa là 3 máy. Cho X là số latop Potpal cần sửa trong một năm. Dựa trên các số liệu đã có, bảng phân phối xác suất của X như sau

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

a. Trong một năm, trung bình có bao nhiêu laptop Potpal cần sửa?
b. Tìm phương sai, độ lệch chuẩn và mode.
c. Một cửa hàng Z có hợp đồng với trung tâm bảo hành đối với việc sửa laptop Potpal như sau: Phí dịch vụ hàng năm là 200$ và 50$ cho mỗi laptop được sửa. Tìm trung bình, phương sai và độ lệch chuẩn của số tiền mà cửa hàng Z trả cho trung tâm bảo hành mỗi năm.

Giải. ....

....

....

....

....

....

....

....

....

....

....

....

....

....

....

....

....

Ví dụ 2.27 Một cửa hàng bán một loại hóa chất thấy rằng lượng hóa chất bán ra hàng tuần (đơn vị tính là trăm lít) là X có hàm mật độ xác suất là

$$f(x) = \begin{cases} \frac{x}{4}, & 0 \le x \le 2 \\ \frac{1}{2}, & 2 < x \le 3 \\ 0, & \text{các trường hợp khác} \end{cases}$$

Giả sử mỗi tuần, chi phí bảo dưỡng cửa hàng là 50$. Cửa hàng mua mỗi lít hóa chất là 1,75$ và bán ra 2,7$.

a. Tính lượng hóa chất trung bình bán ra hàng tuần.

b. Tính lợi nhuận trung bình hàng tuần của cửa hàng.

c. Tính mode của X.

Giải. ....

....

....

....

....

....

....

....

....

....

....

....

....

....

....

....

....

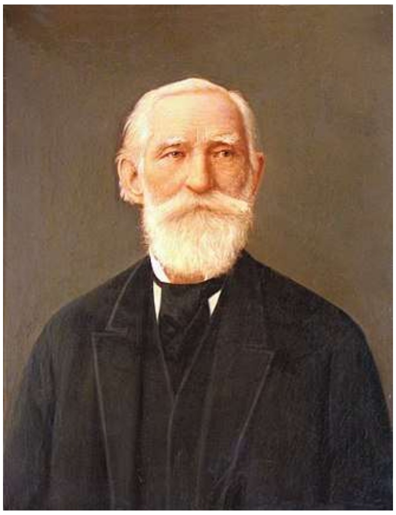
....

Định lý 2.28 (Định lý Chebyshev) Cho X là một biến ngẫu nhiên có kỳ vọng μ và phương sai σ^2. Khi đó, với mọi $\varepsilon > 0$, ta có

$$P(|X - \mu| > \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}.$$

Định lý Chebyshev cho thấy

$$P(\mu - \varepsilon \leq X \leq \mu + \varepsilon) \geq 1 - \frac{\sigma^2}{\varepsilon^2}.$$



Pafnuty Lvovich Chebyshev (1821 - 1894) là nhà toán học người Nga.

Ví dụ 2.29 Khảo sát tình hình sản xuất trong năm 2021, giám đốc một nhà máy thấy rằng trung bình một ngày làm được 120 sản phẩm với độ lệch chuẩn là 10.

a. Tỉ lệ những ngày nhà máy sản xuất được từ 100 đến 140 sản phẩm có nhiều hơn 70% không?
b. Tìm khoảng số lượng sản phẩm ngắn nhất để chứa ít nhất 90% mức sản xuất hàng ngày.

Giải. ....

....

....

....

....

....

....

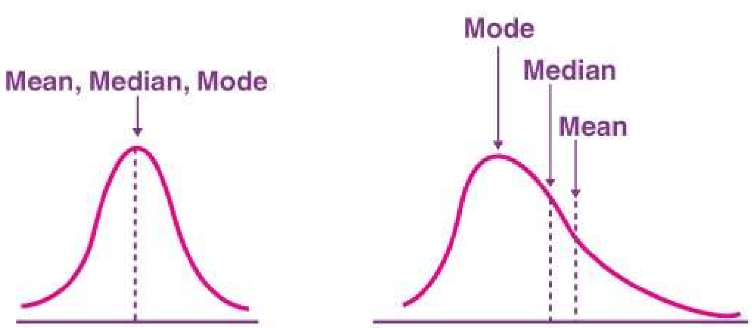
....

....

....

Định nghĩa 2.30 **Trung vị** (median) là giá trị của biến ngẫu nhiên X mà nó chia phân phối của biến ngẫu nhiên thành hai phần bằng nhau, kí hiệu Med(X).

1. Nếu X là biến ngẫu nhiên rời rạc thì giá trị x_k sẽ là trung vị nếu
$$F(x_{k-1}) < 0,5 \leq F(x_{k+1}).$$
2. Nếu X là biến ngẫu nhiên liên tục thì trung vị của X là giá trị m thỏa mãn
$$\int_{-\infty}^{m} f(x)dx = 0,5.$$
3. Nếu X là một biến ngẫu nhiên liên tục có hàm mật độ xác suất $f(x)$ là đối xứng qua điểm μ thì cả trung vị và kỳ vọng của biến ngẫu nhiên đều bằng μ.



2.5 Các phân phối xác suất thường gặp

2.5.1 Phân phối Bernoulli

Định nghĩa 2.31

1. Một biến ngẫu nhiên chỉ nhận hai giá trị: 1 (nếu thành công) và 0 (nếu thất bại) được gọi là **biến Bernoulli**
2. Một **phép thử Bernoulli** là phép thử chỉ có hai biến cố: A (thành công) và $\overline{A}$ (thất bại) trong đó $P(A) = p$.
3. Một phân phối xác suất của biến Bernoulli được gọi là **phân phối Bernoulli**

Ví dụ 2.32

- Bạn trả lời một câu hỏi dạng Đúng - Sai. Khi đó bạn trả lời đúng $(X = 1)$ hoặc sai $(X = 0)$. Xác suất thành công là 0,5.
- Trả lời một câu hỏi trắc nghiệm 4 lựa chọn, có duy nhất một lựa chọn đúng là một phép thử Bernoulli với xác suất thành công là 0,25.

- $p =$ xác suất thành công
- X : **biến Bernoulli** với tham số p, kí hiệu $X \sim B(p)$,
- Phân phối xác suất

$$p(x) = \begin{cases} p, & x = 1 \\ 1 - p, & x = 0 \\ 0, & \text{trường hợp khác} \end{cases}$$

  where $0 < p < 1$.
- $E(X) = p$
- $\mathrm{Var}(X) = p(1 - p)$

- Daniel Bernoulli (1700 - 1782)
- A famous Swiss mathematician and physicist.
- Nguồn ảnh: https://en.wikipedia.org

Ví dụ 2.33 Một câu hỏi trắc nghiệm có 4 phương án trả lời trong đó có một phương án đúng. Một sinh viên chọn ngẫu nhiên một phương án để trả lời câu hỏi đó. Đặt A là biến cố "sinh viên chọn phương án đúng". Khi đó việc trả lời của sinh viên là một phép thử Bernoulli và $p = P(A) = \frac{1}{4}$ và $P(\overline{A}) = \frac{3}{4}$.

Gọi X là biến ngẫu nhiên xác định bởi

$$X = \begin{cases} 1 \text{ nếu sinh viên trả lời đúng} \\ 0 \text{ nếu sinh viên trả lời sai.} \end{cases}$$

Khi đó $X \sim B(\frac{1}{4})$ và $E(X) = \frac{1}{4}, \mathrm{Var}(X) = \frac{3}{4}.\frac{1}{4} = \frac{3}{16}$.

2.5.2 Phân phối nhị thức

Định nghĩa 2.34 Phân phối nhị thức (hay phân phối Bernoulli) là phân phối của một biến ngẫu nhiên mà nó là số lần thành công trong một dãy các phép thử Bernoulli độc lập.

- $n =$ số phép thử
- $p =$ xác suât thành công
- Kí hiệu $X \sim B(n, p)$,
- Hàm khối xác suất

$$p(k) = C_n^k (1-p)^{n-x} p^k, k = 0, 1, \ldots, n,$$

- $E(X) = np$
- $\mathrm{Var}(X) = np(1-p)$

Ví dụ 2.35 Gieo một con xúc xắc n lần. Đặt Y là số lần xuất hiện mặt 1 chấm. Khi đó Y có phân phối nhị thức với tham số n và $p = \frac{1}{6}$.

Ví dụ 2.36 Theo một cuộc khảo sát tại một quốc gia nọ có 41% các hộ gia đình có sử dụng thiết bị kết nối không dây. Chọn ngẫu nhiên 20 hộ gia đình. Tính xác suất

a. có đúng 5 hộ có sử dụng thiết bị không dây.

b. có ít hơn 3 hộ có sử dụng thiết bị không dây.

c. có ít nhất 3 hộ có sử dụng thiết bị không dây.

d. số hộ có sử dụng thiết bị không dây từ 5 đến 7.

Giải.

Trong dãy phép thử Bernoulli, số x_0 mà tại đó xác suất đạt giá trị lớn nhất gọi là **số có khả năng nhất** (hay số lần xuất hiện chắc chắn nhất).

- Nếu $(n+1)p \in \mathbb{Z}$ thì có hai số có khả năng nhất $x_0 = (n+1)p - 1$ và $x_0 = (n+1)p$.
- Nếu $(n+1)p \notin \mathbb{Z}$ thì $x_0 = [(n+1)p - 1] + 1$, trong đó $[(n+1)p - 1]$ là phần nguyên của $(n+1)p - 1$.

Ví dụ 2.37 Tỷ lệ mắc Covid-19 ở một thành phố nọ là 10%. Trong đợt khám bệnh cho vùng đó, người ta đã khám 100 người. Hỏi có nhiều nhất là bao nhiêu người mắc Covid-19? Tính xác suất tương ứng.

Giải. Đây là một dãy phép thử Bernoulli với $n = 100; p = 0,1$ và cần tìm k sao cho $P_n(k)$ là lớn nhất. Ta có $(n+1)p = (100+1).0,1 \notin \mathbb{Z}$. Do đó, số có khả năng lớn nhất là

$x_0 =$ ....

....

và xác suất tương ứng là

....

....

2.5.3 Phân phối hình học

Definition 2.38 **Phân phối hình học** (geometric distribution) là phân phối của biến ngẫu nhiên mà nó là số phép thử Bernoulli cần thực hiện đến khi xuất hiện lần thành công đầu tiên.

- $p =$ xác suất thành công
- Kí hiệu $X \sim Geo(p)$,
- Hàm khối xác suất

$$p(k) = \begin{cases} (1-p)^{k-1}p, & k = 1, 2 \dots \\ 0, & \text{trường hợp khác} \end{cases}$$

- $E(X) = \dfrac{1}{p}$
- $\mathrm{Var}(X) = \dfrac{1-p}{p^2}$

Ví dụ 2.39 Gieo một con xúc xắc đến khi xuất hiện mặt 1 chấm.

a. Tính xác suất lần gieo thứ 4 xuất hiện mặt 1 chấm.
b. Tính xác suất xuất hiện mặt 1 chấm khi gieo ít hơn 10 lần
c. Trung bình số lần gieo đến khi xuất hiện mặt 1 chấm là bao nhiêu?
d. Tìm phương sai của phép thử này.

Giải. ....

....

....

....

....

....

....

....

....

....

2.5.4 Phân phối siêu bội

Định nghĩa 2.40 Xét một tập có N phần tử trong đó có M phần tử có tính chất $\mathcal{P}$ và $N-M$ phần tử không có tính chất $\mathcal{P}$. Chọn k phần tử khác nhau, không phân biệt thứ tự từ N phần tử. Gọi X là số phần tử có tính chất $\mathcal{P}$ trong k phần tử đã chọn. Khi đó X là một biến ngẫu nhiên và ta nói X có **phân phối siêu bội** (hypergeometric distribution) với ba tham số N, M, k.

- Ký hiệu $X \sim H(N, M, k)$
- Xác suất

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

- $E(X) = k\dfrac{M}{N}$
- $\mathrm{Var}(X) = \dfrac{N-k}{N-1} k \dfrac{M}{N} \cdot \left(1 - \dfrac{M}{N}\right)$

Ví dụ 2.41 Trong 100 laptop trưng bày tại cửa hàng POTPAL có 70 laptop hiệu PH. Chọn ngẫu nhiên 40 laptop trong số 100 laptop. Gọi X là số laptop hiệu PH trong số 40 laptop được chọn.

a. Tính xác suất để chọn được từ 27 đến 29 laptop PH.
b. Tính trung bình số laptop PH được chọn và $\mathrm{Var}(X)$.

Giải. ....

....

....

....

....

....

....

....

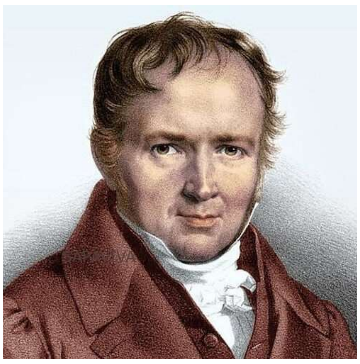
....

....

....

2.5.5 Phân phối Poisson

Cho X là một biến ngẫu nhiên rời rạc ký hiệu số lần xuất hiện của một sự kiện nào đó trong một khoảng thời gian (hoặc không gian). Khi đó ta nói X có phân phối Poisson



Siméon Denis Poisson (1781 - 1840) là nhà toán học, nhà vật lý người Pháp.
Nguồn ảnh: Internet.

Ví dụ 2.42

- Gọi X là số lỗi chính tả trên một trang in. (không gian là trang in.)
- Gọi X là số ô tô đi qua ngã tư Thủ Đức trong một phút. (thời gian là một phút.)
- Gọi X là số email tôi nhận được trong 1 ngày.

Định nghĩa 2.43 Một biến ngẫu nhiên rời rạc X được gọi là có **phân phối Poisson** với tham số $\lambda(\lambda > 0)$ nếu nó có phân phối xác suất là

$$P(X = x) = \frac{e^{-\lambda}.\lambda^x}{x!}, x = 0, 1, 2, \ldots$$

trong đó λ là trung bình số lần xuất hiện biến cố và x là số lần xuất hiện biến cố mà ta quan tâm.

- λ : trung bình số lần xuất hiện biến cố
- Ký hiệu $X \sim P(\lambda)$
- $E(X) = \lambda$
- $\mathrm{Var}(x) = \lambda$
- Giá trị của hàm phân phối xác suất tích lũy $F(X) = P(X \leq x)$ của biến ngẫu nhiên có phân phối Poisson có thể tìm trong bảng A3.

Ví dụ 2.44 Người ta thấy rằng trong một quyển sách nọ có trung bình 3 lỗi ở mỗi trang. Chọn ngẫu nhiên một trang, tính xác suất

a. trang đó có đúng 3 lỗi.

b. trang đó có nhiều hơn 1 lỗi.

Giải. Trung bình số lỗi của 1 trang

$$\lambda = \ldots\ldots\ldots\ldots\ldots$$

. . . .

. . . .

. . . .

. . . .

. . . .

. . . .

. . . .

. . . .

. . . .

. . . .

. . . .

Ví dụ 2.45 Giả sử rằng số lần truy cập trung bình vào một trang web là 2 lần trong 1 giây. Đặt X là số lần truy cập trong 1 giây. Giả sử X có phân phối Poisson.

a. Tính xác suất có 5 lần truy cập trong 1 giây.
b. Tính xác suất có nhiều nhất 5 lần truy cập trong 1 giây.
c. Tính xác suất có nhiều hơn 5 lần truy cập trong 2 giây.

Giải. a. Đặt X là số lần truy cập trong 1 giây. Khi đó X có phân phối Poisson với tham số $\lambda = 2$. Suy ra

$$P(X = 5) = \frac{e^{-2}.2^5}{5!} = 0,0361.$$

b. Dùng bảng A3, ta có

$$P(X \leq 5) = 0,983$$

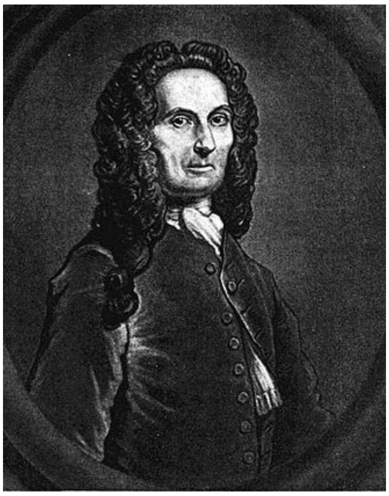
c. Trung bình số lần truy cập trong 2 giây là 4. Đặt Y là số lần truy cập vào trang web trong 2 giây. Khi đó $Y \sim P(4)$ và do đó

$$P(Y > 5) = 1 - P(Y \leq 5) = 1 - 0,785 = 0,215$$

(Dùng bảng A3, ta có $P(Y \leq 5) = 0,785$)

2.5.6 Phân phối chuẩn

Phân phối chuẩn (Normal distribution) là phân phối quan trọng nhất trong xác suất và thống kê. Phân phối chuẩn được giới thiệu bởi một người Anh gốc Pháp tên là Abraham de Moivre (1733). Sau đó, Gauss, một nhà toán học người Đức, đã dùng phân phối chuẩn để nghiên cứu các dữ liệu về thiên văn học (1809) và do vậy nó cũng được gọi là phân phối Gauss.

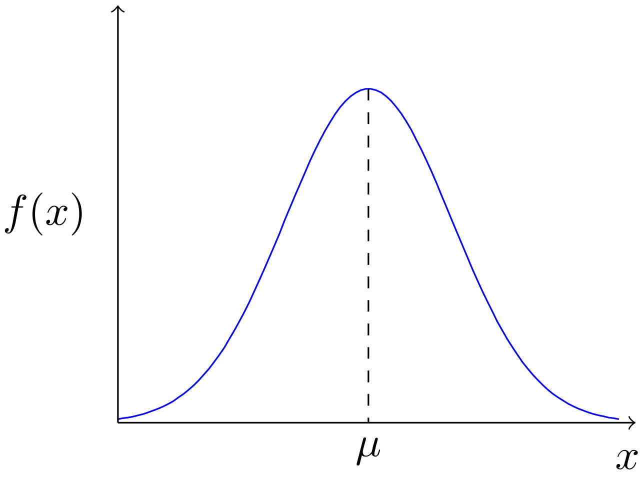



Johann Carl Friedrich Gauss (1777-1855) và Abraham de Moivre (1667-1754).
Nguồn ảnh: Wikipeadia.org

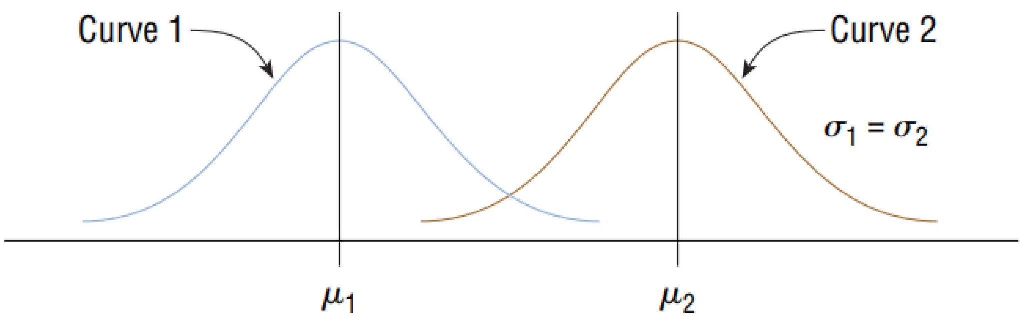
Định nghĩa 2.46 Một biến ngẫu nhiên liên tục X được gọi là có **phân phối chuẩn** (normal distribution) nếu hàm mật độ xác suất của nó có dạng

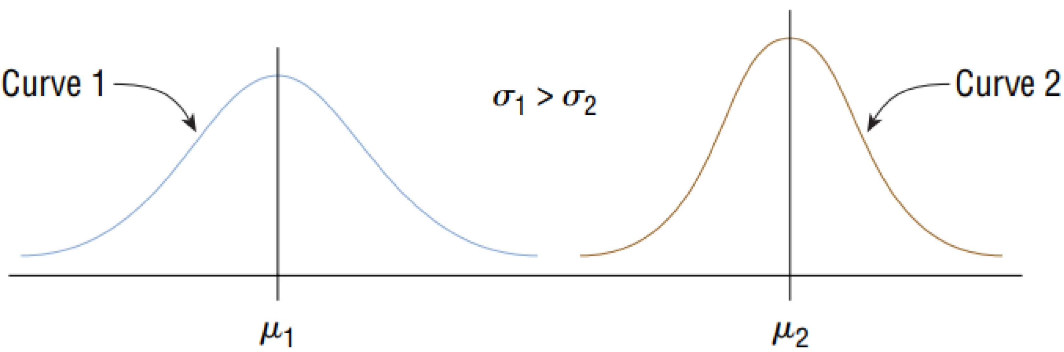
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- Ký hiệu $X \sim N(\mu; \sigma^2)$.
- $E(X) = \mu$
- $\mathrm{Var}(X) = \sigma^2$
- Độ lệch chuẩn $\mathrm{Std}(X) = \sigma$



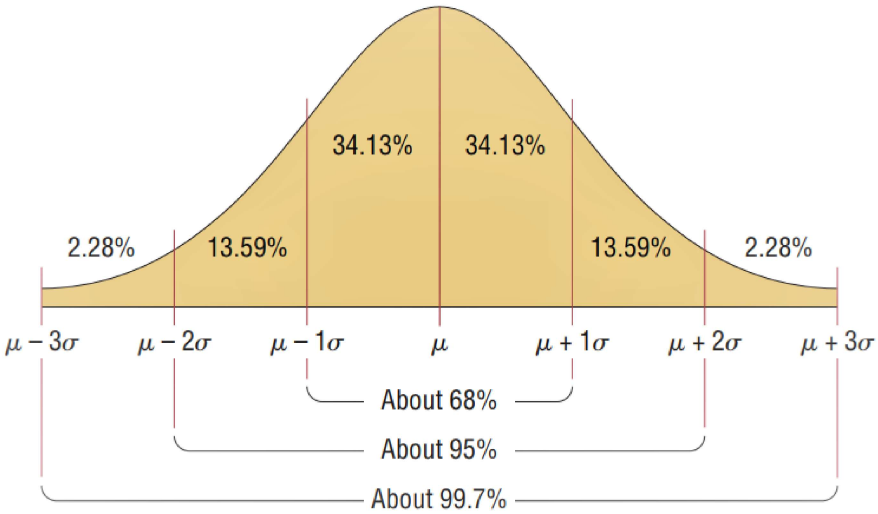
Hình dạng và vị trí của đường cong phân phối chuẩn (đồ thị của hàm mật độ xác suất) phụ thuộc vào hai tham số μ và σ.





Một số tính chất của phân phối chuẩn

1. Kỳ vọng, trung vị bằng nhau và nằm ở trung điểm của phân phối.
2. Đồ thị của hàm mật độ xác suất có hình chuông, đối xứng qua đường thẳng đứng đi qua kỳ vọng.
3. Đường cong liên tục và không chạm vào trục x.
4. Diện tích dưới phần của đường cong phân phối chuẩn trong đoạn $[\mu-\sigma;\mu+\sigma]$ là khoảng 0,68; trong đoạn $[\mu-2\sigma;\mu+2\sigma]$ là 0,95; và trong khoảng $[\mu-3\sigma;\mu+3\sigma]$ là khoảng 0,997.



Định nghĩa 2.47 Phân phối chuẩn với $\mu = 0$ và $\sigma = 1$ được gọi là **phân phối chuẩn chuẩn tắc** (standard normal distribution).
Các phân phối chuẩn có thể được chuyển thành phân phối chuẩn chuẩn tắc bằng cách đổi biến

$$Z = \frac{X-\mu}{\sigma}.$$

Biến ngẫu nhiên Z có hàm mật độ xác suất

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

và hàm phân phối xác suất

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{t^2}{2}} dt$$

Bài toán. Giả sử $X \sim N(\mu; \sigma^2)$. Tính $P(X \leq a)$.

1. Đổi biến $Z = \dfrac{X-\mu}{\sigma} \Rightarrow Z \sim N(0;1)$.
2. $P(X \leq a) = P(Z \leq \dfrac{a-\mu}{\sigma}) = \Phi(\dfrac{a-\mu}{\sigma})$
3. Tìm $\Phi(z) = P(Z \leq z)$ từ Bảng A4
4. Xác định giá trị của z như sau: cột đầu tiên bên trái là hai chữ số đầu tiên của z, hàng ngang đầu tiên là chữ số thứ ba của z.
5. Giao của cột và hàng này là giá trị $\Phi(z) = P(Z \leq z)$.

Ví dụ 2.48 Cho biến ngẫu nhiên X có phân phối chuẩn chuẩn tắc $X \sim N(0;1)$.

a. Tính $P(X \leq 1,35)$.

b. Tính $P(-1,37 \leq X \leq 1,68)$.

Giải. a. Tra bảng A4, ta thấy

Table A4, continued. Standard Normal distribution

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545

$$P(X \leq 1,35) = \Phi(1,35) = 0,9115.$$

b. Ta thấy

$$P(-1,37 \leq X \leq 1,68) = \Phi(1,68) - \Phi(-1,37) = 0,9535 - 0,0853 = 0,8682$$

Ví dụ 2.49 Giả sử thu nhập của các gia đình trong một tháng của một quốc gia có phân phối chuẩn với trung bình là 900$ và độ lệch chuẩn là 200$. Gặp ngẫu nhiên một gia đình.

a. Tính xác suất gia đình này có thu nhập từ 600$ đến 1200$.

b. Chính phủ sẽ trợ cấp cho 0,3% gia đình có mức thu nhập thấp nhất. Hỏi các gia đình này có thu nhập nhiều nhất là bao nhiêu đôla?

Giải. a. Đặt X là thu nhập của một gia đình. Theo giả thiết $X \sim N(900; 200^2)$ với $\mu = 900$ và $\sigma = 200$. Đặt $Z = \dfrac{X - 900}{200}$, khi đó

$$\begin{aligned} P(600 \le X \le 1200) &= P\left(\frac{600 - \mu}{\sigma} \le \frac{X - \mu}{\sigma} \le \frac{1200 - \mu}{\sigma}\right) \\ &= P(-1,5 \le Z \le 1,5) \\ &= \Phi(1,5) - \Phi(-1,5) \\ &= 0,9332 - 0,0668 = 0,8664. \end{aligned}$$

Xác suất gia đình này có thu nhập từ 600$ đến 1200$ là 86,64%.

b. Gọi t là thu nhập (tính bằng $)) của các gia đình thỏa mãn

$$P(X \le t) = 0,3\% = 0,003.$$

$$P(X \le t) = P(Z \le \frac{t - \mu}{\sigma}) = \Phi(\frac{t - \mu}{\sigma}) = 0,003$$

Tra bảng A4,

z	-0.09	-0.08	-0.07	-0.06	-0.05	-0.04	-0.03	-0.02	-0.01	-0.00
-(3.9+)	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
-3.8	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001
-3.7	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001
-3.6	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0002	.0002
-3.5	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002
-3.4	.0002	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003
-3.3	.0003	.0004	.0004	.0004	.0004	.0004	.0004	.0005	.0005	.0005
-3.2	.0005	.0005	.0005	.0006	.0006	.0006	.0006	.0006	.0007	.0007
-3.1	.0007	.0007	.0008	.0008	.0008	.0008	.0009	.0009	.0009	.0010
-3.0	.0010	.0010	.0011	.0011	.0011	.0012	.0012	.0013	.0013	.0013
-2.9	.0014	.0014	.0015	.0015	.0016	.0016	.0017	.0018	.0018	.0019
-2.8	.0019	.0020	.0021	.0021	.0022	.0023	.0023	.0024	.0025	.0026
-2.7	.0026	.0027	.0028	.0029	$\boxed{.0030}$	.0031	.0032	.0033	.0034	.0035
-2.6	.0036	.0037	.0038	.0039	.0040	.0041	.0043	.0044	.0045	.0047
-2.5	.0048	.0049	.0051	.0052	.0054	.0055	.0057	.0059	.0060	.0062

ta thấy $Z = -2,75$.

b. Gọi t là thu nhập (tính bằng $)) của các gia đình thỏa mãn

$$P(X \le t) = 0,3\% = 0,003.$$

Như vậy

$$P(X \le t) = P(Z \le \frac{t-\mu}{\sigma}) = \Phi(\frac{t-\mu}{\sigma}) = 0,003$$

Ta thấy $z = -2,75$. Suy ra

$$\frac{t-900}{200} = -2,75$$

hay $t = 900 - 2,75.200 = 350$. Như vậy, các gia đình này có thu nhập không quá 350$.

Ví dụ 2.50 Một thành phố lắp đặt đèn điện để chiếu sáng đường. Loại đèn này có tuổi thọ trung bình là 1000 giờ với độ lệch chuẩn là 200 giờ. Giả sử tuổi thọ của loại đèn này có phân phối chuẩn.

a. Tính xác suất đèn này bị hỏng trong 700 giờ.

b. Tính xác suất đèn sử dụng được từ 900 giờ đến 1300 giờ.

c. Sau bao lâu thì còn 10% đèn sử dụng được?

Giải. ....

....

....

....

....

....

....

....

....

....

....

....

....

....

....

....

....

....

....

....

....

....

....

2.5.7 Phân phối χ^2

- Nếu X là một biến ngẫu nhiên có phân phối chuẩn thì biến ngẫu nhiên $Y = X^2$ được gọi là có phân phối χ^2 (Chi-Squared Distribution) với bậc tự do là 1.
- Nếu các biến ngẫu nhiên $X_i \sim N(0;1)$ $(1 \leq i \leq n)$ độc lập thì biến ngẫu nhiên
$$Y = X_1^2 + X_2^2 + \cdots + X_n^2$$
được gọi là có phân phối **chi-bình phương** với bậc tự do là n.

Định nghĩa 2.51 Cho $\alpha > 0$, hàm gamma $\Gamma(\alpha)$ được xác định bởi

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} x^{-x} dx$$

1. Với $\alpha > 1$, ta có $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$.
2. Với n là một số nguyên dương, ta có $\Gamma(n) = (n-1)!$
3. $\Gamma(\frac{1}{2}) = \sqrt{\pi}$.

Định nghĩa 2.52 Cho ν là một số nguyên dương. Một biến ngẫu nhiên X được gọi là có **phân phối** χ^2 (chi bình phương) với tham số ν, ký hiệu $X \sim \chi^2_\nu$, nếu hàm mật độ xác suất của nó có dạng

$$f(x) = \begin{cases} \dfrac{1}{2^{\nu/2}\Gamma(\nu/2)} x^{(\nu/2)-1} e^{-x/2}, & x \geq 0 \\ 0, & x < 0. \end{cases}$$

Tham số ν được gọi là **bậc tự do** của X.

2.5.7 Phân phối Student

Định nghĩa 2.53 Một biến ngẫu nhiên liên tục X được gọi là có **phân phối Student** (Student distribution) với bậc tự do $\nu - 1$, ký hiệu $X \sim St(\nu)$, nếu hàm mật độ xác suất của nó có dạng

$$f(x) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\Gamma(\nu/2)} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

Ta có $X \sim \dfrac{N(0;1)}{\sqrt{\chi^2_\nu/\nu}}$.

Phân phối Student sẽ được dùng trong việc ước lượng khoảng của trung bình trong chương 5.

William Sealey Gosset (1876-1937) học toán và hóa học tại Đại học Oxford ở Anh. Năm 1899, ông chuyển đến Dublin, Ireland và làm nhân viên thống kê cho nhà máy bia Guinness. Trong quá trình nghiên cứu chất lượng lúa mạch và hoa bia, Gosset đề xuất việc sử dụng t-phân phối. Năm 1908, ông đã công bố ý tưởng của mình trong một bài báo khoa học bằng bút danh "Student" vì Guinness cấm nhân viên của mình xuất bản kết quả nghiên cứu của riêng họ. Do đó, t-phân phối thường được gọi là phân phối Student.

William Sealey Gosset.
Nguồn ảnh: wikipedia.org

2.5.8 Phân phối Fisher

Định nghĩa 2.54 Cho hai biến ngẫu nhiên độc lập có phân phối χ^2 với các bậc tự do lần lượt là ν_1 và ν_2. Khi đó tỉ số

$$\frac{\chi^2_{\nu_1}/\nu_1}{\chi^2_{\nu_2}/\nu_2}$$

được gọi là F-**phân phối** hay **phân phối Fisher** (F-distribution), ký hiệu là F_{ν_1,ν_2}.

Ronald Aylmer Fisher (1890 - 1962) là một nhà một nhà toán học, thống kê, nhà di truyền học và học thuật người Anh.

Ronald Aylmer Fisher năm 1913.
Nguồn ảnh: wikipedia.org

2.6 Hàm đặc trưng

Định nghĩa 2.55 **Hàm đặc trưng** của biến ngẫu nhiên X là một hàm số $\phi(t) : \mathbb{R} \to \mathbb{C}$ xác định bởi

- Nếu X là biến ngẫu nhiên rời rạc thì

$$\phi(t) = \sum_{j\geq 1} e^{itx_j} P(X = x_j)$$

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

$$\phi(t) = \int_{-\infty}^{+\infty} e^{itx} f(x)\mathrm{d}x,$$

trong đó $e^{itx} = \cos tx + i \sin tx$ và $i^2 = -1$.

Phân phối của một biến ngẫu nhiên có thể được xác định hoàn toàn thông qua hàm đặc trưng của nó. Hàm đặc trưng cũng cho ta một tiêu chuẩn để xác định xem khi nào biến ngẫu nhiên có hàm mật độ.

Biến ngẫu nhiên có phân phối	Tham số	Hàm đặc trưng
Bernoulli	p	$\phi(t) = 1 - p + pe^{it}$
Nhị thức	n, p	$\phi(t) = (1 - p + pe^{it})^n$
Siêu bội	p	$\phi(t) = \dfrac{p}{1 - e^{it}(1-p)}$
Poisson	λ	$\phi(t) = e^{\lambda(e^{it}-1)}$
Chuẩn	μ, σ^2	$\phi(t) = e^{it\mu - \frac{1}{2}\sigma^2 t^2}$
Chi-bình phương	ν	$\phi(t) = (1 - 2it)^{-\frac{\nu}{2}}$

BÀI TẬP

Bài 2.1 Chọn ngẫu nhiên 3 viên bi từ 1 túi có 6 bi trắng và 4 bi đen. Gọi X là số bi đen trong 3 bi vừa chọn. Tìm bảng phân phối xác suất của X.

Bài 2.2 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. Nếu bắn trúng mục tiêu thì dừng lại. Đặt X là số viên đạn xạ thủ đã bắn. Lập bảng phân phối xác suất và hàm phân phối xác suất của X.

Bài 2.3 Một túi có 10 bi đỏ và 6 bi đen. Chọn ngẫu nhiên ra 3 bi.

a. Gọi X là số bi đỏ. Tìm bảng phân phối xác suất của X.

b. Giả sử chọn được mỗi bi đỏ được 5$ và mỗi bi đen được 8$. Gọi Y là số tiền tổng cộng tương ứng với 3 bi được chọn ra. Tìm bảng phân phối xác suất của Y.

Bài 2.4 Trong một lô hàng có 2 loại sản phẩm: Loại thứ nhất có 10 hộp, mỗi hộp có 7 sản phẩm hiệu A và 3 sản phẩm hiệu B; loại thứ hai có 15 hộp, mỗi hộp có 8 sản phẩm hiệu A và 4 sản phẩm hiệu B. Lấy ngẫu nhiên một hộp, rồi từ hộp đó lấy ra 2 sản phẩm. Gọi X là số sản phẩm hiệu A. Tìm bảng phân phối và hàm phân phối xác suất của X.

Bài 2.5 Một hộp có 4 bi đen, 6 bi trắng và 4 bi đỏ. Lấy ngẫu nhiên 3 bi từ hộp. Giả sử nếu lấy bi đỏ được 1 điểm, lấy bi trắng được 0 điểm và lấy bi đen bị trừ 1 điểm. Tìm bảng phân phối xác suất của số điểm có được.

Bài 2.6 Cho X số lần uống cafe trong một ngày của các nhân viên công ty TUC. Bảng phân phối xác suất của X như sau

X	0	1	2	3	4	5
$P(X = x_i)$	0,28	0,37	0,17	0,12	0,05	0,01

a. Chọn ngẫu nhiên một nhân viên. Tính xác suất để nhân viên này uống hơn 2 ly cafe trong một ngày.
b. Tính trung bình (μ) và độ lệnh chuẩn (σ) của biến ngẫu nhiên X.
c. Tìm xác suất $X \in [\mu - 2\sigma, \mu + 2\sigma]$.
Bài 2.7 Trong một cuộc khảo sát, người ta thấy chiều cao trung bình của sinh viên trong một lớp học học là 170 cm. Chọn ngẫu nhiên 4 sinh viên của lớp. Đặt X là số sinh viên có thấp hơn 170cm.
a. Lập bảng phân phối xác suất của X.
b. Tìm hàm phân phối của X.
c. Tính xác suất gặp ít nhất 2 sinh viên thấp hơn 170cm.
d. Tính xác suất gặp từ 1 đến 3 sinh viên thấp hơn 170cm.
Bài 2.8 Cho biến ngẫu nhiên X có hàm mật độ xác suất

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

Tìm C và tính xác suất $P(0 < X < \dfrac{3}{2})$.
Bài 2.9 Thời gian phục vụ mỗi khách hàng tại một cửa hàng là một biến ngẫu nhiên liên tục X (phút) có hàm mật độ xác suất như sau

$$f(x) = \begin{cases} ke^{-4x}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

a. Tìm k và hàm phân phối xác suất của X.
b. Tính xác suất thời gian phục vụ một khách hàng trong khoảng thời gian từ 1 đến 2 phút.

Bài 2.10 Chi phí sửa chữa hàng tuần (X, đơn vị là trăm đô la) cho một loại máy bay có hàm mật độ xác suất cho bởi

$$f(x) = \begin{cases} cx(1-x), & 0 \leq x \leq 1 \\ 0, & x \notin [0;1] \end{cases}$$

a. Tìm c để hàm này trở thành hàm mật độ xác suất hợp lệ.
b Tìm hàm phân phối xác suất của X.
c. Xác suất chi phí sửa chữa sẽ vượt quá 75$ trong một tuần là bao nhiêu?
d. Xác suất để chi phí sửa chữa sẽ vượt quá 75$ trong một tuần với điều kiện chúng sẽ vượt quá 50$.
Bài 2.11 Một biến ngẫu nhiên X có hàm mật độ như sau

$$f(x) = \begin{cases} k.(30-x), & x \in (0;30) \\ 0, & x \notin (0;30) \end{cases}.$$

a. Tìm k.
b. Tìm hàm phân phối của X.
c. Tính kỳ vọng và phương sai của X.
Bài 2.12 Cho biến ngẫu nhiên X có hàm mật độ xác suất

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

và $E(X) = 0,6$.
a. Tìm hàm phân phối xác suất của X.
b. Tính $P(-1 < X < \dfrac{1}{2})$ và $\mathrm{Var}(X)$.

Bài 2.13 Cho biến ngẫu nhiên X có hàm phân phối xác suất

$$F(X) = \begin{cases} 0, & x < 0 \\ ax^2(9-2x), & 0 \le x \le 3 \\ 1, & x > 3 \end{cases}$$

a. Tìm a để $F(x)$ liên tục trên $\mathbb{R}$.
b. Tính $P(-1 < X < 1)$.
c. Tính $E(X), \mathrm{Var}(X)$.
Bài 2.14 Giả sử tỉ lệ sinh viên tốt nghiệp của một lớp học là 80%. Chọn ngẫu nhiên 20 sinh viên của lớp. Tính các xác suất sau
a. Có 15 sinh viên tốt nghiệp.
b. Có 8 sinh viên không tốt nghiệp.
c. Số sinh viên không tốt nghiệp cao nhất là bao nhiêu?
Bài 2.15 Tín hiệu thông tin được phát đi 3 lần độc lập nhau. Xác suất thu được tín hiệu ở mỗi lần là 0,4.
a. Tìm xác suất để nguồn thu nhận được thông tin đúng 2 lần.
b. Tìm xác suất để nguồn thu nhận được thông tin đó.
c. Nếu muốn xác suất thu được tin không nhỏ hơn $0,9$ thì phải phát đi ít nhất bao nhiêu lần?
Bài 2.16 Một phân xưởng có 12 máy dệt. Xác suất để mỗi máy dệt trong khoảng thời gian T gặp sự cố bằng 1/3. Tính xác suất
a. Trong khoảng thời gian T có 4 máy gặp sự cố.
b. Trong khoảng thời gian T số máy gặp sự cố không bé hơn 3, không lớn hơn 6.

Bài 2.17 Một gói phần mềm gồm 12 chương trình, trong đó có 5 chương trình phải được nâng cấp. Chọn ngẫu nhiên 4 chương trình để kiểm tra.
a. Xác suất ít nhất 2 trong số chúng phải được nâng cấp là bao nhiêu?
b. Số lượng chương trình trung bình cần nâng cấp trong 4 chương trình đã chọn là bao nhiêu?
Bài 2.18 Trung bình mỗi ngày có 10 khách hàng mới đăng kí sử dụng dich vụ internet của công ty ITN.
a. Tính xác suất có nhiều hơn 8 khách hàng mới đăng ký mới trong một ngày.
b. Xác suất có hơn 16 người đăng ký mới trong vòng 2 ngày là bao nhiêu?
Bài 2.19 Một công ty cung cấp RAM máy tính cho 10 cửa hàng. Xác suất một cửa hàng gọi điện thoại để đặt hàng trong một ngày là 0,2 (cả 10 cửa hàng đều giống nhau). Tính xác suất số cuộc gọi đặt hàng vào một ngày đã cho trong các trương hợp
a. Có nhiều nhất 3 cuộc gọi đặt hàng.
b. Có ít nhất 3 cuộc gọi đặt hàng.
c. Có đúng 3 cuộc gọi đặt hàng.
Bài 2.20 Số vụ tai nạn giao thông trung bình trên một đoạn đường cao tốc nhất định là 2 vụ mỗi tuần. Giả sử rằng số vụ tai nạn có phân phối Poisson.
a. Tìm xác suất không xảy ra tai nạn trên đoạn đường cao tốc này trong thời gian 1 tuần.
b. Tìm xác suất xảy ra nhiều nhất ba vụ tai nạn trên đoạn đường cao tốc này trong thời gian 2 tuần.
Bài 2.21 Một công ty đã bán 5000 ổ cứng SDD trong một năm. Xác suất một ổ cứng SDD của công ty bị hỏng trong một năm nhất định là 0,001. Hãy tìm xác suất có 4 ổ cứng bị hư trong một năm.
Bài 2.22 Một cuộc khảo sát cho thấy phụ nữ chi tiêu trung bình 146,21 đô la cho kỳ nghỉ Giáng sinh. Giả sử độ lệch chuẩn là 29,44 đô la. Tìm tỉ lệ phần trăm phụ nữ chi tiêu ít hơn 160 đô la. Giả sử số tiền chi tiêu trong lễ Giáng sinh có phân phối chuẩn.

Bài 2.23 Mỗi tháng, một hộ gia đình ở Việt Nam sử dụng nước trung bình 28 mét khối. Giả sử độ lệch chuẩn là 2 mét khối. Chọn ngẫu nhiên một hộ gia đình. Tìm xác suất
a. Hộ đó sử dụng từ 27 đến 31 mét khối nước mỗi tháng.
b. Hộ đó sử dụng hơn 30,2 mét khối nước mỗi tháng.
Giả sử lượng nước sử dụng mỗi tháng có phân phối chuẩn.
Bài 2.24 Để đủ điều kiện vào học Đại học Công nghệ Thông tin, thí sinh phải đạt điểm trong 5% cao nhất trong bài thi Đánh giá năng lực. Bài thi có điểm trung bình là 646,1 điểm và độ lệch chuẩn là 118,8 điểm. Tìm điểm thấp nhất có thể để đủ điều kiện được vào học trường Đại học Công nghệ Thông tin. Giả sử điểm bài thi Đánh giá năng lực có phân phối chuẩn.

Index

Chương 3. Luật số lớn và định lý giới hạn trung tâm

Nguyễn Minh Trí

Trường Đại học Công nghệ Thông tin

Ngày 26 tháng 2 năm 2023

3.1 Các dạng hội tụ của dãy các biến ngẫu nhiên

Giả sử chúng ta muốn tìm hiểu một biến ngẫu nhiên X, nhưng ta không thể quan sát X một cách trực tiếp. Thay vào đó, ta có thể thực hiện một số phép đo và đưa ra ước tính của X: thu được $X_1, X_2, X_3, \dots$. Ta hy vọng khi n tăng lên, X_n ngày càng gần X. Nói cách khác, ta hy vọng rằng X_n "hội tụ" về X.

Câu hỏi. Ta muốn biết liệu một dãy các biến ngẫu nhiên $X_1, X_2, X_3, \dots$ có "hội tụ" về một biến ngẫu nhiên X hay không. Nghĩa là, chúng ta muốn xem liệu X_n có càng ngày càng gần X khi n tăng lên.

Ví dụ 3.1 Xét phép thử ngẫu nhiên: Gieo một đồng xu. Không gian mẫu $\Omega = \{S, N\}$ (S là xuất hiện mặt sấp, N là xuất hiện mặt ngửa). Ta định nghĩa dãy các biến ngẫu nhiên $X_1, X_2, X_3, \dots$ như sau

$$X_n(s) = \begin{cases} \dfrac{1}{n+1}, & \text{nếu } s = S \\ 1, & \text{nếu } s = N \end{cases}$$

Dãy các biến ngẫu nhiên $X_1, X_2, \dots$ (xác định trên cùng một không gian mẫu) có thể "hội tụ" về một biến ngẫu nhiên X. Các kiểu hội tụ như sau:

1. Hội tụ theo phân phối (convergence in distribution)
2. Hội tụ theo xác suất (convergence in probability)
3. Hội tụ theo trung bình (convergence in mean)
4. Hội tụ gần như chắc chắn (hầu khắp nơi) (convergence almost surely)

Một dãy các biến ngẫu nhiên có thể hội tụ theo nghĩa này nhưng không hội tụ theo nghĩa khác. Một số trong số các kiểu hội tụ này "mạnh hơn" so với những kiểu khác và một số "yếu hơn". Tức là, ta nói hội tụ Loại A mạnh hơn hội tụ Loại B nếu có hội tụ Loại A thì có hội tụ loại Loại B.

Định nghĩa 3.2 Dãy các biến ngẫu nhiên $X_1, X_2, \ldots$ được gọi là **hội tụ theo xác suất** đến biến ngẫu nhiên X, kí hiệu $X_n \xrightarrow{P} X$, nếu với mọi $\varepsilon > 0$, ta có

$$\lim_{n\to\infty} P(|X_n - X| \geq \varepsilon) = 0$$

Ví dụ 3.3 Cho một dãy các biến ngẫu nhiên $X_1, X_2, \ldots$ trong đó X_n có hàm mật độ xác suất như sau

$$f_{X_n}(x) = \begin{cases} ne^{-nx}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

Khi đó $X_n \xrightarrow{P} 0$ (tức là $X_1, X_2, \ldots$ hội tụ theo xác suất đến biến ngẫu nhiên $X = 0$).

Giải. Với mọi $\varepsilon > 0$, ta có

$$\begin{aligned}
\lim_{n\to\infty} P(|X_n - 0| \geq \varepsilon) &= \lim_{n\to\infty} P(X_n \geq \varepsilon) \quad \text{vì } X_n \geq 0 \\
&= \lim_{n\to\infty} \left(\int_{\varepsilon}^{\infty} f_{X_n}(t)dt \right) \\
&= \lim_{n\to\infty} \left(\int_{\varepsilon}^{\infty} ne^{-nt}dt \right) \\
&= \lim_{n\to\infty} \left(-e^{-nt}\Big|_{\varepsilon}^{\infty} \right) \\
&= \lim_{n\to\infty} \left(e^{-n\varepsilon} \right) \\
&= 0
\end{aligned}$$

Định lý 3.4 Cho $\{X_n\}_{n\geq 1}$ hội tụ theo xác suất đến μ_1 và $\{Y_n\}_{n\geq 1}$ hội tụ theo xác suất đến μ_2. Khi đó

1. $\{X_n + Y_n\}_{n\geq 1}$ hội tụ theo xác suất đến $\mu_1 + \mu_2$.
2. $\{X_n Y_n\}_{n\geq 1}$ hội tụ theo xác suất đến $\mu_1\mu_2$.
3. $\{X_n / Y_n\}_{n\geq 1}$ hội tụ theo xác suất đến μ_1/μ_2 nếu $\mu_2 \neq 0$.
4. $\{\sqrt{X_n}\}_{n\geq 1}$ hội tụ theo xác suất đến $\sqrt{\mu_1}$ nếu $P(X_n \geq 0) = 1$ với mọi n.

Định nghĩa 3.5 Dãy các biến ngẫu nhiên $X_1, X_2, \ldots$ được gọi là **hội tụ theo phân phối** đến biến ngẫu nhiên X, kí hiệu $X_n \xrightarrow{d} X$, nếu

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$$

với mọi x sao cho $F_X(x)$ liên tục tại x.

Ví dụ 3.6 Cho một dãy các biến ngẫu nhiên $X_1, X_2, \ldots$ với hàm phân phối xác suất của X_n là

$$F_{X_n}(x) = \begin{cases} 1 - \left(1 - \dfrac{1}{n}\right)^{nx}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

Khi đó $X_n \xrightarrow{d} 1 - e^{-x}$ (tức là $X_1, X_2, \ldots$ **hội tụ theo phân phối** đến biến ngẫu nhiên X có hàm phân phối xác suất $F_X(x) = 1 - e^{-x}$ với $x > 0$).

Giải. Với mọi $x \leq 0$, ta có

$$F_{X_n}(x) = F_X(x) = 0$$

với mọi $n = 1, 2, \ldots$.
Xét $x \geq 0$, ta có

$$\begin{aligned}\lim_{n\to\infty} F_{X_n}(x) &= \lim_{n\to\infty}\left(1 - \left(1 - \frac{1}{n}\right)^{nx}\right)\\ &= 1 - \lim_{n\to\infty}\left(1 - \frac{1}{n}\right)^{nx}\\ &= 1 - e^{-x} = F_X(x)\end{aligned}$$

Như vậy $X_n \xrightarrow{d} X$.

Định lý 3.7 Cho dãy các biến ngẫu nhiên $X_1, X_2, X_3, \ldots$ và biến ngẫu nhiên X. Giả sử rằng X và X_n (với mọi n) có tập giá trị là các số nguyên không âm. Khi đó $X_n \xrightarrow{d} X$ nếu và chỉ nếu

$$\lim_{n\to\infty} P(X_n = k) = P(X = k)$$

với $k = 1, 2, \ldots$.

Ví dụ 3.8 Cho dãy các biến ngẫu nhiên $X_1, X_2, X_3, \ldots$ có phân phối nhị thức

$$X_n \sim B(n; \frac{\lambda}{n}), \quad \text{với } n \in \mathbb{N}, n > \lambda$$

trong đó $\lambda > 0$ là một hằng số. Khi đó $X_n \xrightarrow{d} X$ với $X \sim P(\lambda)$ (X có phân phối Poisson với tham số λ).

Giải. Theo Định lý 2.7, ta chứng minh $\lim\limits_{n\to\infty} P(X_n = k) = P(X = k)$

$$\begin{aligned}\lim_{n\to\infty} P(X_n = k) &= \lim_{n\to\infty} C_n^k \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \lambda^k \lim_{n\to\infty} \frac{n!}{k!(n-k)!} \frac{1}{n^k} \left(1 - \frac{\lambda}{n}\right)^{n-k}\end{aligned}$$

$$= \frac{\lambda^k}{k!} \lim_{n\to\infty} \frac{n(n-1)(n-2)\ldots(n-k+1)}{n^k} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k}$$

Chú ý rằng, khi k, λ là các hằng số, ta có

$$\lim_{n\to\infty} \frac{n(n-1)(n-2)\ldots(n-k+1)}{n^k} = 1$$

$$\lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^n = e^{-\lambda}; \ \lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^{-\lambda} = 1$$

Do đó

$$\lim_{n\to\infty} P(X_n = k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

Định nghĩa 3.9 Một dãy các biến ngẫu nhiên $X_1, X_2, \ldots$ **hội tụ hầu chắc chắn** về một biến ngẫu nhiên X, ký hiệu $X_n \xrightarrow{a.s} X$, nếu

$$P(\{s \in \Omega \mid \lim_{n\to\infty} X_n(s) = X(s)\}) = 1.$$

3.2 Luật số lớn và Định lý giới hạn trung tâm

Định lý 3.10 (Luật số lớn yếu-weak law of large numbers) Cho $X_1, X_2, \ldots, X_n$ là các biến ngẫu nhiên độc lập và phân phối giống nhau với $E(X_i) = \mu < \infty$ với mọi i. Đặt $\overline{X}_n = \dfrac{X_1 + X_2 + \ldots + X_n}{n}$. Khi đó $\overline{X}_n$ hội tụ theo xác suất đến μ.

Định lý 3.11 (Luật số lớn mạnh-strong law of large numbers) Cho $X_1, X_2, \ldots, X_n$ là các biến ngẫu nhiên độc lập và phân phối giống nhau với $E(X_i) = \mu < \infty$ với mọi i. Đặt $\overline{X}_n = \dfrac{X_1 + X_2 + \ldots + X_n}{n}$. Khi đó $\overline{X}_n$ hội tụ theo hầu chắc chắn đến μ.

Định lý giới hạn trung tâm (Central limit theorem) nói rằng, trong những điều kiện nhất định, **tổng** hoặc **trung bình** của một số lượng lớn các biến ngẫu nhiên "xấp xỉ" phân phối chuẩn.

Định lý 3.12 (Định lý giới hạn trung tâm) Cho $X_1, X_2, \ldots, X_n$ là các biến ngẫu nhiên độc lập và phân phối giống nhau với kỳ vọng $E(X_i) = \mu$ và phương sai $V(X_i) = \sigma^2 < \infty$ với mọi i. Đặt $S_n = X_1 + X_2 + \ldots + X_n$ và $Z_n = \dfrac{S_n - n\mu}{\sigma\sqrt{n}}$. Khi đó $Z_n \xrightarrow{d} Z$ trong đó $Z \sim N(0; 1)$.

Định lý 3.13 Cho $X_1, X_2, \ldots, X_n$ là các biến ngẫu nhiên độc lập và phân phối giống nhau với kỳ vọng $E(X_i) = \mu$ và phương sai $V(X_i) = \sigma^2 < \infty$ với mọi i. Đặt $\overline{X}_n = \dfrac{X_1 + X_2 + \ldots + X_n}{n}$ và $Y_n = \dfrac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}$. Khi đó $Y_n \xrightarrow{d} Z$ trong đó $Z \sim N(0; 1)$.

Nhận xét 3.2.4.

1. Nếu các biến ngẫu nhiên X_i độc lập và **đều có phân phối chuẩn** thì tổng S_n và S_n/n cũng có phân phối chuẩn.
2. Cho X_i là các biến ngẫu nhiên độc lập và **có phân phối giống nhau** với cùng trung bình μ và phương sai σ^2. Khi đó, với n đủ lớn (thông thường $n \geq 30$)

$$X_1 + X_2 + \ldots + X_n \approx N(n\mu; n\sigma^2)$$

và

$$\frac{X_1 + X_2 + \ldots + X_n}{n} \approx N(\mu; \sigma^2/n).$$

Ví dụ 3.14 Theo một cuộc khảo sát, thời gian xem tivi trung bình của các em bé từ 2 đến 5 tuổi là 25 giờ mỗi tuần. Giả sử thời gian xem tivi có phân phối chuẩn với độ lệch chuẩn là 3 giờ. Chọn ngẫu nhiên 20 em bé từ 2 đến 5 tuổi. Tính xác suất thời gian xem tivi trung bình của 20 bé lớn hơn 26,3 giờ.

Giải. Đặt X_i là thời gian xem tivi của bé thứ i. Thời gian xem trung bình của 20 bé là $Y = \dfrac{X_1 + \ldots + X_{20}}{20}$. Theo đề bài $\mu = 25$ và $\sigma = 3$. Theo Định lý giới hạn trung tâm, $Y \sim N(\mu; \sigma^2/n)$. Do đó

$$\begin{aligned}
P(Y > 26,3) &= P\left(\frac{Y-\mu}{\sigma/\sqrt{n}} > \frac{26,3-25}{3/\sqrt{20}}\right) \\
&= P(Z > 1,94) \\
&= 1 - \Phi(1,94) = 0,0262.
\end{aligned}$$

Ví dụ 3.15 Một đĩa cứng có dung lượng trống là 330 megabyte. Cho 300 hình ảnh độc lập, kích thước trung bình mỗi ảnh là 1 megabyte với độ lệch chuẩn là 0,5 megabyte. Xác suất ổ cứng này lưu được 300 hình là bao nhiêu?

Giải. Đặt X_i là dung lượng của hình ảnh thứ i và $S = \sum_{i=1}^{300} X_i$. Ta có $n = 300, E(X_i) = \mu = 1$ và $V(X_i) = \sigma^2 = 0,5^2$. Áp dụng Đinh lý giới hạn trung tâm, $S \approx N(n\mu; n\sigma^2)$ ta có

$$\begin{aligned}
P(S \le 330) &\approx P\left(\frac{S-n\mu}{\sqrt{n}\sigma} \le \frac{330-n\mu}{\sqrt{n}\sigma}\right) \\
&= P\left(\frac{S-300}{\sqrt{300}.0,5} \le \frac{330-300}{\sqrt{300}.0,5}\right) \\
&= P\left(\frac{S-300}{\sqrt{300}.0,5} \le 3,46\right) \\
&= \Phi(3,46) = 0,9997.
\end{aligned}$$

Ví dụ 3.16 Giả sử thời gian chơi thể thao (tính bằng giờ) của các sinh viên ở một trường đại học có phân phối chuẩn với trung bình là 2 giờ và độ lệch chuẩn là 0,5 giờ. Chọn ngẫu nhiên 50 sinh viên. Xác suất thời gian chơi thể thao trung bình của 50 sinh viên này từ 1,8 giờ đến 2,3 giờ là bao nhiêu?

Giải. ....

....

....

....

....

....

....

....

....

....

Ví dụ 3.17 Giả sử thời gian tự học ở nhà của các sinh viên trường UIT có phân phối giống nhau với trung bình 3 giờ và độ lệch chuẩn 1,15 giờ. Chọn ngẫu nhiên 75 sinh viên. Tính xác suất 75 sinh viên này có tổng thời gian tự học ở nhà ít hơn 200 giờ.

Giải. ....

....

....

....

....

....

....

....

....

....

3.3 Các công thức xấp xỉ

3.3.1 Xấp xỉ phân phối siêu bội bằng phân phối nhị thức

- Cho biến ngẫu nhiên X có phân phối siêu bội $X \sim H(N, M, n)$.
- Nếu $N \to \infty$ thì $H(N, M, n) \xrightarrow{P} B(n, \frac{M}{N})$.
- **Khi** $N \geq 20n$, ta có thể dùng phân phối nhị thức để xấp xỉ cho phân phối siêu bội.

Ví dụ 3.18 Một vườn lan có 10 000 cây lan sắp nở hoa trong đó có 1000 cây hoa màu đỏ. Chọn ngẫu nhiên 20 cây lan. Tính xác suất có 5 cây lan màu đỏ trong 20 cây.

Giải. Gọi X là số cây lan đỏ trong 20 cây. Khi đó $N = 10000, M = 1000, n = 20$ và $X \sim H(10000; 1000; 20)$. Vì $N = 10000 > 20n$ nên ta có thể dùng phân phối nhị thức để xấp xỉ cho phân phối siêu bội. Ta có $\dfrac{M}{N} = \dfrac{1000}{10000} = 0,1$ và X xấp xỉ với phân phối nhị thức B$(20; 0,1)$. Như vậy

$$P(X = 5) \approx C_{20}^{5}.(0,1)^5.(0,9)^{15} = 0,0319.$$

Nhận xét. Nếu dùng phân phối siêu bội thì

$$P(X = 5) = \frac{C_{1000}^{5} C_{9000}^{15}}{C_{10000}^{20}} = 0,0318.$$

3.3.2 Xấp xỉ phân phối nhị thức bằng phân phối Poisson

- Cho X là một biến ngẫu nhiên có phân phối nhị thức $X \sim B(n; p)$.
- Theo Ví dụ 3.1.8, ta thấy $B(n; p) \xrightarrow{d} P(np)$.
- **Nếu** $n \geq 30$ **và** $p \leq 0,05$ thì ta dùng **phân phối Poisson** để xấp xỉ cho phân phân phối nhị thức. Ta thấy $X \approx P(np)$.

Ví dụ 3.19 Xác suất một máy tính đã được cài chương trình diệt virus bị nhiễm virus là 0,03. Chọn ngẫu nhiên 200 máy tính đã được cài đặt chương trình diệt virus. Tính xác suất có nhiều nhất 6 máy bị nhiễm virus.

Giải. Cách 1. (Dùng phân phối nhị thức) Đặt X là số máy tính bị nhiễm virus. Khi đó $X \sim B(200; 0,03)$. Do đó

$$P(X \leq 6) = \sum_{k=0}^{6} C_{200}^{k}(0,03)^k.(0,97)^{200-k} = 0,6063.$$

Cách 2. (Dùng xấp xỉ phân phối Poisson) Đặt X là số máy tính bị nhiễm virus. Khi đó $X \sim B(200; 0,03)$. Vì $n = 200$ và $p = 0,03 < 0,05$ nên ta có thể dùng phân phối Poisson để xấp xỉ cho X. Đặt $\lambda = np = 6$, khi đó $X \approx P(\lambda)$. Như vậy

$$P(X \leq 6) \approx \sum_{k=0}^{6} \frac{e^{-6}.6^k}{k!} = 0,606$$

(Dùng bảng A3)

3.3.3 Xấp xỉ phân phối nhị thức bằng phân phối chuẩn

- Cho X là biến ngẫu nhiên có phân phối nhị thức $X \sim B(n; p)$.
- Nếu p nhỏ thì ta có thể dùng phân phối Poisson để xấp xỉ phân phối Nhị thức.
- **Khi** $np \geq 5$ **và** $n(1-p) \geq 5$ thì ta có thể dùng phân phối chuẩn xấp xỉ cho phân phối Nhị thức. Ta có

$$B(n; p) \approx N(np; np(1-p)).$$

Kỹ thuật hiệu chỉnh liên tục (continuity correction) Chúng ta có thể xấp xỉ một phân phối rời rạc (trong trường hợp này là phân phối nhị thức) bằng phân phối liên tục (phân phối chuẩn). Chú ý rằng xác suất $P(X = x)$ có thể dương nếu X là rời rạc, trong khi $P(X = x) = 0$ nếu X liên tục.

- Nếu X có phân phối nhị thức thì

$$P(X = x) = P(x - 0,5 < X < x + 0,5).$$

- Mở rộng khoảng 0,5 đơn vị theo mỗi hướng, sau đó sử dụng xấp xỉ phân phối chuẩn. Việc này được gọi là *hiệu chỉnh liên tục* và không làm thay đổi xác suất của biến cố.
- Khi xấp xỉ một phân phối rời rạc với một phân phối liên tục, ta nên sử dụng một hiệu chỉnh liên tục.

Khi hiệu chỉnh liên tục, ta sử dụng bảng sau để xấp xỉ phân phối nhị thức bằng phân phối chuẩn

Phân phối nhị thức	Phân phối chuẩn
$P(X = a)$	$P(a - 0,5 < X < a + 0,5)$
$P(X < a)$	$P(X < a - 0,5)$
$P(X > a)$	$P(X > a + 0,5)$
$P(X \leq a)$	$P(X < a + 0,5)$
$P(X \geq a)$	$P(X > a - 0,5)$
$P(a \leq X \leq b)$	$P(a - 0,5 < X < b + 0,5)$

Ví dụ 3.20 Một loại virus máy tính mới tấn công một thư mục bao gồm 1350 tệp. Mỗi tệp bị hỏng với xác suất 0,75 độc lập với các tệp khác.

a. Xác suất có ít hơn 1000 tệp bị hỏng là bao nhiêu?

b. Xác suất có từ 1000 đến 1020 tệp bị hỏng là bao nhiêu?

Giải. Gọi X là số tệp bị hỏng trong 1350 tệp. Ta có $X \sim B(n; p)$ với $n = 1350$ và $p = 0,75$. Vì $np = 1350.0,75 = 1012,5$ và $n(1-p) = 1350.0,25 = 337,5$ nên có thể dùng xấp xỉ phân phối chuẩn $X \approx N(np; np(1-p))$ với $\mu = np = 1012,5$ và $\sigma = \sqrt{np(1-p)} = \sqrt{1350.0,75.0,25} = 15,91$.

a. Dùng hiệu chỉnh liên tục và xấp xỉ phân phối chuẩn để tính

$$\begin{aligned} P(X = 1000) &= P(999,5 < X < 1000,5) \\ &= P(\frac{999,5 - 1012,5}{15,91} < Z < \frac{1000,5 - 1012,5}{15,91}) \\ &= P(-0,82 < Z < -0,75) \\ &= \Phi(-0,75) - \Phi(-0,82) = 0,2266 - 0,2061 = 0,0205 \end{aligned}$$

b. Dùng hiệu chỉnh liên tục và xấp xỉ phân phối chuẩn

$$\begin{aligned} P(1000 \leq X \leq 1020) &= P(999,5 < X < 1020,5) \\ &\approx P(\frac{999,5 - 1012,5}{15,91} < Z < \frac{1020,5 - 1012,5}{15,91}) \\ &= P(-0,82 < Z < 0,5) \\ &= \Phi(0,5) - \Phi(-0,82) = 0,6915 - 0,2061 = 0,4854 \end{aligned}$$

Ví dụ 3.21 Một công ty sản xuất các RAM máy tính cho biết tỉ lệ RAM bị lỗi là 0,5%. Kiểm tra ngẫu nhiên 100 RAM máy tính của công ty này. Tính xác suất

a. Không có RAM bị lỗi.

b. Có nhiều hơn 1 RAM bị lỗi.

Giải. ....

Ví dụ 3.22 Giả sử một công ty chuyên sản xuất chip bán dẫn sản xuất 50 chip bị lỗi trong số 1000. Lấy ngẫu nhiên 100 chip (không thay thế). Tính xác suất có ít nhất 1 chip bị lỗi trong mẫu.

Giải.

. . . .

. . . .

. . . .

. . . .

. . . .

. . . .

. . . .

. . . .

. . . .

. . . .

. . . .

Ví dụ 3.23 Một nhà máy xếp 1000 cái bánh quy vào mỗi hộp trong đó có 50 cái bánh loại đặc biệt. Chọn ngẫu nhiên 40 cái bánh. Tìm xác suất có ít nhất 2 cái bánh quy loại đặc biệt.

Giải. . . . .

. . . .

. . . .

. . . .

. . . .

. . . .

. . . .

. . . .

. . . .

BÀI TẬP

Bài 3.1 Mức lương trung bình của giáo viên ở New Jersey là 52174\$. Giả sử mức lương có phân phối chuẩn với độ lệch chuẩn 7500\$.
a. Xác suất một giáo viên được chọn ngẫu nhiên kiếm được ít hơn 50000\$ mỗi năm là bao nhiêu?
b. Chọn ngẫu nhiên 100 giáo viên. Xác suất lương trung bình của các giáo viên này ít hơn 50000\$ mỗi năm là bao nhiêu?
Bài 3.2 Giả sử điểm bài thi SAT có phân phối chuẩn với trung bình là 1518 và độ lệch chuẩn là 325.
a. Một bài thi SAT được chọn ngẫu nhiên. Tìm xác suất để điểm đó nằm trong khoảng 1440 đến 1480.
b. Chọn 16 bài thi SAT ngẫu nhiên. Tìm xác suất để chúng có điểm trung bình từ 1440 đến 1480.
Bài 3.3 a. Một hộp chứa 8 bi đỏ và 8 bi xanh, người ta lấy ngẫu nhiên 4 bi (không thay thế). Xác suất để lấy được 2 bi đỏ và 2 bi xanh là bao nhiêu? (0,431)
b. Một hộp chứa 50.000 viên bi đỏ và 50.000 viên bi xanh, người ta lấy ngẫu nhiên 4 viên bi (không thay thế). Hãy ước lượng xác suất để lấy được 2 bi đỏ và 2 bi xanh. (0,375)
Bài 3.4 a. Cho một hộp chứa 40 sản phẩm, trong đó có 4 sản phẩm bị lỗi. Nếu chọn ngẫu nhiên 5 sản phẩm thì xác suất nó chứa không quá 1 sản phẩm lỗi là bao nhiêu?
b. Nếu lấy ngẫu nhiên gồm 5 sản phẩm được chọn từ một kho chứa 4.000.000 sản phẩm, trong đó có 400.000 sản phẩm có lỗi, thì xác suất nó chứa không quá 1 sản phẩm lỗi là bao nhiêu?

Bài 3.5 Giả sử rằng xác suất một người bị phản ứng xấu khi tiêm một loại vaccine nào đó là 0,001. Chọn ngẫu nhiên 1000 người được tiêm vaccine.
a. Tính xác suất có ít nhất 2 người sẽ bị phản ứng xấu. (0,264)
b. Tính xác suất có ít nhất 96% người không bị phản ứng xấu.
Bài 3.6 Xác suất một người bị dị ứng khi tiêm một loại vaccine đã cho là 0,001. Tính xác suất để trong số 2000 người tiêm
a. Có chính xác 3 người bị dị ứng. (0,18)
b. Có nhiều hơn 2 người bị dị ứng. (0,323)
Bài 3.7 Một bài báo cho thấy rằng cứ 200 người thì có 1 người mang gen khiếm khuyết gây ra bệnh ung thư ruột kết di truyền. Chọn ngẫu nhiên 1000 người, tính xác suất
a. Từ 5 đến 8 người mang gen này. (0,492)
b. Ít nhất 8 người mang gen này. (0,133)
Bài 3.8 Xác suất đánh trúng bóng của một vận động viên bóng chày là 0,32 (32%). Tìm xác suất để người chơi đó đánh trúng bóng nhiều nhất 26 lần trong 100 lần ném bóng. (0,119)
Bài 3.9 Sáu phần trăm số táo trong một lô hàng lớn bị hư hỏng. Trước khi nhận mỗi lô hàng, người quản lý kiểm tra chất lượng chọn ngẫu nhiên 100 quả táo. Nếu có nhiều hơn 3 quả táo bị hư thì lô hàng bị từ chối. Xác suất lô hàng này bị từ chối là bao nhiêu? (0,8531)
Bài 3.10 Tỷ lệ người Mỹ từ 25 tuổi trở lên có ít nhất một bằng đại học là 53,1%. Khảo sát ngẫu nhiên gồm 300 người Mỹ từ 25 tuổi trở lên, xác suất để hơn 175 người có ít nhất một bằng đại học là bao nhiêu? (0,0301 (TI: 0,0304))
Bài 3.11 Tỉ lệ sản phẩm bị lỗi do một máy sản xuất là 10%. Tìm xác suất trong 400 sản phẩm được chọn ngẫu nhiên do máy này sản xuất
a. Có nhiều nhất là 30 sản phẩm lỗi. (0,0567)
b. Có từ 30 đến 50 sản phẩm lỗi. (0,9198)
c. Có không ít hơn 65 sản phẩm lỗi. (0,0079)

Bài 3.12 Người ta thấy rằng trong số 3000 con chip của một công ty sản xuất có 2 con chip bị lỗi. Lấy ngẫu nhiên 12000 chip loại này để kiểm tra.
a. Tính xác suất có tối đa 2 chip bị lỗi.
b. Tính xác suất có ít nhất 96% chip không bị lỗi.
Bài 3.13 Khoảng 10% dân số thuận tay trái. Sử dụng phân phối chuẩn để tính gần đúng xác suất mà trong một lớp gồm 150 học sinh
a. có ít nhất 25 người trong số họ thuận tay trái.
b. có từ 15 đến 20 thuận tay trái.

Index