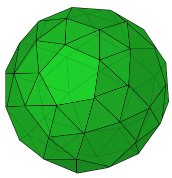
Projects in Mathematics and Applications

XÁC SUẤT VÀ PHÂN PHỐI

The-Anh Vu-Le*
Manh Pham Nguyen†
Linh Tran‡

23/06/2018



* University of Sciences, VNU, Ho Chi Minh City.
† Minerva University, USA.
‡ National University of Singapore, Singapore.

Mục lục

1 Xác suất cơ bản

1.1 Phép thử và biến cố

Bài giảng này sẽ không đề cập đến các quy tắc đếm và tổ hợp. Chúng ta bắt đầu với khái niệm phép thử ngẫu nhiên.

Định nghĩa 1.1 (Phép thử ngẫu nhiên)**. Phép thử** là một khái niệm cơ bản không có định nghĩa trong toán học (tương tự *tập hợp* hay *điểm*). Phép thử có thể hiểu là việc thực hiện một hành động và quan sát kết quả của nó.

Phép thử ngẫu nhiên là phép thử mà kết quả của nó không thể đoán trước được.

Ví dụ 1.2. Các phép thử sau là ngẫu nhiên:

- ⋆ Tung đồng xu có 2 mặt khác nhau và quan sát mặt nằm trên.
- ⋆ Tung xúc xắc 6 mặt và quan sát mặt trên.
- ⋆ Đếm số xe chạy qua một trạm thu phí trong một ngày bất kì.

Ví dụ 1.3. Các phép thử sau không phải là ngẫu nhiên:

- ⋆ Dùng quỳ tím kiểm tra tính acid của một dung dịch không rõ thành phần.
- ⋆ Đọc đơn đăng ký tham gia trại hè PiMA của một bạn học sinh và quyết định xem có nhận bạn không.

Định nghĩa 1.4 (Biến cố)**.** Một **biến cố** là một tập hợp các kết quả có thể quan sát được. Các tập hợp có một phần tử được gọi là **biến cố sơ cấp** vì các biến cố này đơn giản nhất. Biến cố gồm từ hai kết quả được gọi là **biến cố phức hợp**.

Ví dụ 1.5.

- ⋆ Khi tung đồng xu, các biến cố sơ cấp là Sấp và Ngửa.
- ⋆ Khi tung xúc xắc, các biến cố sơ cấp là $\{1\}, \{2\}, \ldots, \{6\}$; các biến cố phức hợp bao gồm $\{1, 2, 3\}$, {"Số chẵn"}, {"Số chia hết cho 3"}, ...

Định nghĩa 1.6 (Không gian mẫu)**. Không gian mẫu** là tập hợp hợp thành từ tất cả các biến cố sơ cấp (tập hợp gồm tất cả các kết quả có thể xảy ra), ký hiệu là Ω.

Biến cố tương ứng với Ω là **biến cố chắc chắn**. Biến cố ứng với $\varnothing$ là **biến cố rỗng**.

Nhận xét 1.7. Mọi biến cố đều là con của biến cố chắc chắn Ω. Biến cố rỗng là con của mọi biến cố.

Ví dụ 1.8. Các phép thử sau có không gian mẫu là một tập hợp hữu hạn:

- ⋆ Khi tung đồng xu, $\Omega = \{\text{Sấp}, \text{Ngửa}\}$.
- ⋆ Khi tung xúc xắc, $\Omega = \{1, 2, 3, 4, 5, 6\}$.

Ví dụ 1.9. Các phép thử sau có không gian mẫu là một tập hợp vô hạn:

- ⋆ Vô hạn đếm được: Chọn một số hữu tỉ ngẫu nhiên trong đoạn $[0, 1]$.

$\star$ Vô hạn không đếm được: Chọn một điểm ngẫu nhiên trong hình vuông $[0,1] \times [0,1]$.

Ta có thể nêu ra một biến cố bằng cách liệt kê các kết quả trong nó, nhưng thông thường ta mô tả biến cố bằng *tính chất chung* của các kết quả trong nó.

Ví dụ 1.10. Khi tung xúc xắc, biến cố $A = \{2, 4, 6\}$ tương đương với "Tung được mặt có số chẵn".

Ví dụ 1.11. Khi chọn một điểm ngẫu nhiên trên hình vuông $[-1,1] \times [-1,1]$, biến cố $A = [0,1] \times [0,1]$ tương đương với "Chọn được điểm nằm ở góc phần tư thứ nhất".

1.2 Phép toán trên biến cố

Vì các biến cố thực chất là các tập hợp, các phép toán trên chúng cũng chính là các phép toán tập hợp quen thuộc.

Định nghĩa 1.12. Các phép toán trên biến cố được định nghĩa như sau:

1. Phép hợp (cộng): $A \cup B \equiv A + B := \{\omega : \omega \in A \vee \omega \in B\}$.
2. Phép giao (nhân): $A \cap B \equiv AB := \{\omega : \omega \in A \wedge \omega \in B\}$. Nếu $AB = \varnothing$, ta nói A và B **xung khắc**.
3. Phép trừ (hiệu): $A \setminus B := \{\omega : \omega \in A \wedge \omega \notin B\}$.
4. Phép bù (đối): $A^C \equiv \overline{A} := \{\omega : \omega \notin A\}$. A^C được gọi là **phần bù** của A và là một biến cố xung khắc với A.

Tính chất 1.13 (Các tính chất cơ bản của phép toán trên biến cố)**.** *Với mọi biến cố A, B, C cùng không gian mẫu Ω, ta có:*

1. *Giao hoán: $A + B = B + A$ và $AB = BA$.*
2. *Kết hợp: $(A + B) + C = A + (B + C)$ và $A(BC) = (AB)C$.*
3. *De Morgan: $\overline{A + B} = \overline{A} \,.\, \overline{B}$, và $\overline{AB} = \overline{A} + \overline{B}$.*

1.3 Xác suất của biến cố

Định nghĩa 1.14 (Xác suất)**.** Lặp lại một phép thử vô hạn lần. Với mỗi $n \in \mathbb{N}$ và biến cố A, gọi $\sigma(A, n)$ là số lần biến cố A xảy ra trong n lần thử đầu tiên. Khi đó ta thừa nhận tồn tại giới hạn sau:

$$\mathrm{P}(A) = \lim_{n \to \infty} \frac{\sigma(A, n)}{n}.$$

Giá trị $\mathrm{P}(A)$ đo **khả năng xảy ra** của A, và được gọi là **xác suất** của A.

Tính chất 1.15. *Hàm xác suất P thỏa mãn các tính chất sau:*

1. $P(\Omega) = 1;\ P(\varnothing) = 0;\ 0 \leq P(A) \leq 1, \quad \forall A \subseteq \Omega.$
2. $P(A + B) = P(A) + P(B) - P(AB).$

3. *$P(A+B) \leq P(A)+P(B)$, nếu $AB = \varnothing$ thì xảy ra đẳng thức.*
4. *$P(B) \leq P(A), \quad \forall A, B : B \subseteq A$.*
5. *$P(A+B+C) = P(A)+P(B)+P(C)-P(AB)-P(BC)-P(BA)+P(ABC)$.*
6. *Union bound: $P(A_1 + A_2 + \cdots + A_n) \leq P(A_1) + P(A_2) + \cdots + P(A_n)$, nếu $A_iA_j = \varnothing, \quad \forall i \neq j$ thì xảy ra đẳng thức.*

Ví dụ 1.16. Giả sử không gian mẫu có hữu hạn các phần tử $\Omega = \{1, 2, \dots, N\}$ và các biến cố sơ cấp có xác suất đều nhau: $\mathrm{P}(1) = \mathrm{P}(2) = \dots = \mathrm{P}(N) = \frac{1}{N}$, khi đó:

$$\mathrm{P}(A) = \frac{|A|}{N}.$$

Nhận xét 1.17. Giả thiết mọi biến cố sơ cấp đều có cùng xác suất xảy ra khá thông dụng và thường được sử dụng trong các bài toán tính xác suất rời rạc.

Ví dụ 1.18. Tung 2 con xúc xắc độc lập nhau. Tính xác suất để tổng trên 2 mặt bằng 7.

- $\star$ Tổng số kết quả có thể xảy ra: $6 \times 6 = 36$.
- $\star$ Tổng số kết quả sao cho tổng 2 mặt là 7: $|\{(1,6),(2,5),\cdots,(6,1)\}| = 6$. Xác suất để tổng 2 mặt bằng 7 là $\frac{6}{36} = \frac{1}{6}$.

1.4 Hệ biến cố đầy đủ

Trong các bài toán tính xác suất phức tạp, nhiều khi phải chia thành nhiều trường hợp nhỏ tách biệt nhau để đơn giản hóa tính toán, sau đó mới tổng hợp lại. Khái niệm **Hệ biến cố đầy đủ** chính là công cụ để chia trường hợp trong các bài toán đó.

Định nghĩa 1.19. Một hệ hữu hạn các biến cố $\{A_1, A_2, \dots, A_n\}$ được gọi là **đầy đủ** nếu

1. $A_iA_j = \varnothing, \quad \forall i \neq j$.
2. $A_1 + A_2 + \cdots + A_n = \Omega$.

Ví dụ 1.20. Các hệ biến cố sau là đầy đủ:

- $\star$ Khi tung xúc xắc: $\{\{1\},\{2\},\{3\},\{4\},\{5\},\{6\}\}$.
- $\star$ Khi tung xúc xắc: "Xuất hiện mặt lẻ", "Xuất hiện mặt chẵn".

Nhận xét 1.21. Trong mọi phép thử, với mọi biến cố A, $\{A, \overline{A}\}$ là một hệ biến cố đầy đủ.

Tính chất 1.22. *Một số tính chất của hệ đầy đủ:*

1. *Với mọi hệ biến cố đầy đủ $\{B_1, B_2, \dots, B_n\}$ và biến cố A, ta có:*

$$AB_1 + AB_2 + \cdots + AB_n = A$$

2. *$P(B_1) + P(B_2) + \cdots + P(B_n) = 1$.*
3. *$P(AB_1) + P(AB_2) + \cdots + P(AB_n) = P(A)$.*

4. *Dựa theo nhận xét 1.21, ta có:*

   a) $P(B) + P(\overline{B}) = 1, \quad \forall B \subseteq \Omega.$

   b) $P(AB) + P(A\overline{B}) = P(A), \quad \forall A, B \subseteq \Omega.$

1.5 Biến cố độc lập - Công thức nhân

Để bắt đầu tìm hiểu các biến cố phức hợp, là giao của nhiều biến cố, ta xem xét trường hợp đơn giản nhất trước. Ta đưa ra khái niệm **biến cố độc lập**, và cách sử dụng sự độc lập để tính xác suất.

Định nghĩa 1.23 (Biến cố độc lập)**.** 2 biến cố A và B được gọi là **độc lập** nếu và chỉ nếu:

$$\mathrm{P}(AB) = \mathrm{P}(A)\mathrm{P}(B).$$

Trong thực tế, nếu 2 biến cố xảy ra từ 2 phép thử độc lập nhau (kết quả của phép này không ảnh hưởng đến phép kia) thì hiển nhiên chúng là độc lập.

Đẳng thức ở định nghĩa 1.23 còn được gọi là **công thức nhân xác suất**. Để tìm hiểu ứng dụng của công thức nhân, ta xem xét ví dụ sau.

Ví dụ 1.24. Tung một con xúc xắc 2 lần. Tính xác suất để tổng 2 mặt là 7.

Rõ ràng với mọi số nguyên $n \in X = \{1, 2, \dots, 6\}$, $7 - n \in X$. Do đó với mỗi $n \in X$, ta có:

1. Xác suất lần đầu tung được n: $\frac{1}{6}$.
2. Xác suất lần thứ hai tung được $7 - n$: $\frac{1}{6}$.
3. Do 2 lần tung là độc lập, xác suất tung được bộ $(n, 7 - n)$ là: $\frac{1}{6} \times \frac{1}{6} = \frac{1}{36}$.
4. Dùng công thức cộng, do các biến cố $(n, 7 - n)$ là xung khắc, ta có:

$$\mathrm{P}(\text{Tổng 2 mặt bằng 7}) = \sum_{n=1}^{6} \mathrm{P}\left((n, 7 - n)\right) = \frac{6}{36} = \frac{1}{6}.$$

Lưu ý rằng 2 biến cố độc lập không nhất thiết phải thuộc 2 phép thử độc lập. Điều kiện độc lập chỉ dựa hoàn toàn vào công thức nhân ở định nghĩa 1.23. Ta xem xét ví dụ sau.

Ví dụ 1.25. Tung một con xúc xắc 2 lần. Đặt A_n là biến cố lần đầu tung được n, với $n \in \{1, 2, \dots, 6\}$, và B là biến cố 2 lần tung có tổng bằng 7. Khi đó A và B độc lập.

Chứng minh. Dành cho phần **Bài tập**. □

Bài tập

Bài tập 1.1. Khi nào thì có các đẳng thức sau:

1. $A + B = \overline{A}$.

2. $AB = \overline{A}$.

3. $A + B = AB$.

Hai biến cố A và $\overline{A+B}$ có xung khắc không?

Bài tập 1.2. Cho A và B là hai biến cố ngẫu nhiên. Tìm biến cố X từ hệ thức:

$$\overline{X+A} + \overline{X+\overline{A}} = B.$$

Bài tập 1.3. Có 3 giải thưởng đánh số 1, 2 và 3. Một hộp có 4 mảnh giấy kích thước như nhau: (1) thắng giải 1, (2) thắng giải 2, (3) thắng giải 3, (4) thắng giải 1, 2 và 3. Một tờ giấy được chọn ngẫu nhiên. Gọi A_i là biến cố "thắng giải i" (i = 1..3). Chứng minh rằng A_i độc lập A_j, $\forall i \neq j$. Chứng minh $P(A_1A_2A_3) \neq P(A_1)P(A_2)P(A_3)$.

Bài tập 1.4. Tỷ lệ một người bị bệnh tim trong một vùng dân cư là 9%, bị bệnh huyết áp là 12%, bị cả hai bệnh là 7%. Chọn ngẫu nhiên một người trong vùng. Tính xác suất để người đó:

1. Không bị bệnh tim hoặc không bị bệnh huyết áp.

2. Bị bệnh tim nhưng không bị bệnh huyết áp.

3. Không bị bệnh tim nhưng bị bệnh huyết áp.

Bài tập 1.5. Founder của PiMA viết n lá thư và bỏ vào n phong bì đã viết sẵn địa chỉ. Tính xác suất để có ít nhất 1 lá thư được bỏ đúng phong bì của nó.

Bài tập 1.6. Một hộp đựng 10 phiếu rút thăm trong đó có 2 phiếu trúng thưởng. 10 người lần lượt lên rút mỗi người 1 lá thăm cho đến khi cả 2 phiếu trúng thưởng đều đã đượt rút trúng. Tính xác suất trúng thưởng của mỗi người.

2 Biến ngẫu nhiên

2.1 Khái niệm

Trong một phép thử ngẫu nhiên, đầu ra của nó có thể là số hoặc không. Ví dụ, phép thử tung xúc xắc và xem số nút cho kết quả là các số từ 1 đến 6 nhưng phép thử chọn một học sinh trong lớp và xác định giới tính thì cho kết quả không phải số là Nam hoặc Nữ. Tuy nhiên, người ta muốn đầu ra của mọi phép thử đều gắn với một đại lượng đo đạc được, hay còn gọi là thuộc tính số. Do đó, khái niệm biến ngẫu nhiên ra đời như một quy tắc để liên hệ kết quả của một phép thử với một con số cụ thể.

Định nghĩa 2.1. Biến ngẫu nhiên là một hàm thực hiện ánh xạ từ không gian mẫu $\mathcal{S}$ đến tập hợp các số thực $\mathbb{R}$.

Ví dụ 2.2. Khi chọn ngẫu nhiên một người trong lớp thì người đó có thể là Nam (M) hay Nữ (F). Với không gian mẫu $\mathcal{S} = \{M, F\}$, ta định nghĩa một biến ngẫu nhiên X như sau:

$$X(M) = 1 \quad , \quad X(F) = 0.$$

Biến ngẫu nhiên này gắn kết quả người được chọn là Nam với số 1 và nữ với số 0.

Biến ngẫu nhiên này được định nghĩa bằng cách gán trực tiếp từng kết quả với một con số. Cách định nghĩa này chỉ nên được sử dụng khi lực lượng của không gian mẫu là không lớn.

Ví dụ 2.3. Khi chọn ngẫu nhiên một ngày trong năm 2018, ta có 365 kết quả khả dĩ, nếu liệt kê hết toàn bộ tất cả các ngày cùng với giá trị của nó thì thật phiền phức. Thay vào đó, ta có thể định nghĩa một biến ngẫu nhiên Y như sau:

$$Y = \begin{cases} 1 & \text{nếu là ngày lễ} \\ 0 & \text{nếu là ngày thường} \end{cases}$$

Với định nghĩa này, $Y(01/06) = 1$ do 01/06 là ngày quốc tế thiếu nhi, còn $Y(03/06) = 0$ do đó chỉ là ngày bình thường (sinh nhật của tác giả).

Trong hai ví dụ trên, các giá trị của biến ngẫu nhiên chỉ là 0 và 1. Các biến ngẫu nhiên thuộc loại này thường xuyên xuất hiện nên được ưu ái đặt tên riêng theo người đầu tiên nghiên cứu nó, nhà Toán học Thụy Sĩ **Jacob Bernoulli**.

Định nghĩa 2.4 (Biến Bernoulli)**.** Một biến ngẫu nhiên mà miền giá trị của nó chỉ gồm 0 và 1 gọi là một **biến ngẫu nhiên Bernoulli**.

Cùng một không gian mẫu có thể cho nhiều biến ngẫu nhiên khác nhau.

Ví dụ 2.5. Trên địa bàn TPHCM có nhiều bến xe buýt lớn, ta xét hai trạm xe buýt là bến xe miền Đông và bến xe miền Tây. Định nghĩa các biến ngẫu nhiên như sau:

$$\begin{aligned} X &= \text{tổng số xe buýt đang chờ ở cả hai bến,} \\ Y &= \text{khác biệt về số xe đang chờ giữa hai bến,} \\ Z &= \text{số lượng xe buýt ở bến có số lượng xe buýt đang chờ lớn hơn.} \end{aligned}$$

Giả sử phép thử này được thực hiện cho kết quả là $\omega = (5, 8)$ (tức 2 xe ở bến xe miền Đông, 3 xe ở bến xe miền Tây) thì $X(\omega) = 5 + 8 = 13$, ta nói **giá trị quan sát** (hay giá trị) của biến ngẫu nhiên X là $x = 13$. Tương tự có giá trị của Y là $y = |5 - 8| = 3$ và Z là $z = max(5, 8) = 8$.

Trong các ví dụ trên, các biến ngẫu nhiên chỉ có một miền giá trị hữu hạn. Ta xét các trường hợp biến ngẫu nhiên có thể không phải như vậy.

Ví dụ 2.6. Xét phép thử tung một đồng xu cho đến khi xuất hiện mặt ngửa thì dừng lại. Gọi mặt ngửa là H, mặt sấp là T. Ta có không gian mẫu $\mathcal{S} = \{H, TH, TTH, ...\}$ có vô hạn phần tử. Định nghĩa một biến ngẫu nhiên X là:

$$X = \text{số lần tung cho đến khi phép thử dừng lại.}$$

Khi đó, $X(H) = 1$, $X(TH) = 2, \ldots$. Bất kì số tự nhiên khác 0 nào cũng có thể là giá trị của X, miền giá trị của X là vô hạn.

Ví dụ 2.7. Xét phép thử đo độ pH trong dịch vị dạ dày của một người bình thường. Ta định nghĩa biến ngẫu nhiên X:

$$X = \text{độ pH trong dịch vị dạ dày.}$$

Người bình thường có độ pH dạ dày từ 1.5 đến 3.5. Ta có miền giá trị của X là $\{x \in \mathbb{R} \mid 1.5 \leq x \leq 3.5\}$.

2.2 Phân loại

Có thể chia các biến ngẫu nhiên thành 2 loại phụ thuộc vào miền giá trị của nó.

Định nghĩa 2.8. Biến ngẫu nhiên được xem là **rời rạc** khi mà miền giá trị của nó là một tập hữu hạn hoặc vô hạn đếm được.

Ví dụ 2.9. Các ví dụ ở trên trừ ví dụ cuối cùng (về độ pH) đều là các biến ngẫu nhiên rời rạc. Có thể thấy trong ví dụ tung đồng xu, miền giá trị không là tập hữu hạn mà là một tập vô hạn đếm được (trùng với tập số nguyên dương).

Định nghĩa 2.10. Biến ngẫu nhiên được xem là **liên tục** khi cả hai điều sau thỏa:

- ⋆ Miền giá trị là một đoạn liên tục trên trục số (có thể tiến đến vô cùng) hoặc một tập hợp các đoạn như vậy (ví dụ: $[-\infty, 1] \cup [5, 10]$).
- ⋆ Không có giá trị nào cho xác suất dương, nói cách khác, $P(X = c) = 0, \quad \forall c$.

Ví dụ 2.11. Lượng mưa trong một tháng ở một khu vực, ví dụ Huế, có thể nhận bất kì giá trị nào trong khoảng $[25, 1200]$ (mm). Do vậy, đây là một biến ngẫu nhiên liên tục.

Để làm việc với các biến ngẫu nhiên rời rạc cần các công cụ của toán rời rạc như là phép tổng hay sai phân. Đối với các biến ngẫu nhiên liên tục, cần các công cụ như tích phân và đạo hàm.

3 Biến ngẫu nhiên rời rạc

3.1 Phân phối xác suất

Tưởng tượng như bạn có một ấm trà và nhiều ly, bạn cần chia hết trà vào các ly nhỏ. Nếu có một ly thì tất nhiên ly đó sẽ có toàn bộ trà. Nếu có từ 2 ly trở lên thì sẽ có vô số cách chia khác nhau. Việc phân phối xác suất cũng giống như vậy, với biến ngẫu nhiên X có thể nhận nhiều giá trị khác nhau.

Một phân phối xác suất của biến ngẫu nhiên X cho biết xác suất tổng 1 sẽ được phân phối thế nào cho các giá trị khác nhau của X. Gọi $\mathrm{p}(x)$ là xác suất gán với giá trị x, ta có:

$$\mathrm{p}(x) = \text{xác suất X có giá trị x} = \mathrm{P}(X = x).$$

Ví dụ 3.1. Việc tung đồng xu hoàn hảo cho các kết quả là $\mathcal{S} = \{H, T\}$. Ta định nghĩa một biến ngẫu nhiên X như trên: $X(H) = 1$, $X(T) = 0$. Có $\mathrm{P}(0) = \mathrm{P}(X = 0) = \mathrm{P}(\text{"xuất hiện mặt sấp"}) = 0.5$.

3.2 Hàm khối xác suất

Định nghĩa 3.2. Hàm khối xác suất *(probability mass function - pmf)* của một biến ngẫu nhiên rời rạc được xác định cho mọi số x là $\mathrm{p}(x) = \mathrm{P}(X = x) = \mathrm{P}\left(\forall \omega \in \mathcal{S} : X(\omega) = x\right)$.

Nói bằng ngôn ngữ tự nhiên, với mọi giá trị của biến ngẫu nhiên, hàm khối xác suất cho biết xác suất xảy ra của kết quả của phép thử liên hệ với giá trị đó. Dễ nhận thấy có hai điều kiện của hàm phân phối xác suất:

$$0 \leq \mathrm{p}(x) \leq 1 \text{ và } \sum_x \mathrm{p}(x) = 1.$$

Ví dụ 3.3. Một xe hàng chở 10 kiện hàng, mỗi kiện hàng có 50 sản phẩm. Xe hàng được kiểm tra chất lượng sản phẩm và cho kết quả: 3 kiện hàng có 5 sản phẩm lỗi, 6 kiện hàng có 2 sản phẩm lỗi, chỉ có 1 kiện hàng không có sản phẩm lỗi.
Một kiện hàng sẽ được giao đến cho người dùng. Gọi biến ngẫu nhiên X là lượng sản phẩm lỗi có trong kiện hàng được giao. Ta có:

$$\begin{aligned}
&\mathrm{P}(0) = \mathrm{P}(X = 0) = \frac{1}{10} = 0.1, \\
&\mathrm{P}(2) = \mathrm{P}(X = 2) = \frac{6}{10} = 0.6, \\
&\mathrm{P}(5) = \mathrm{P}(X = 5) = \frac{3}{10} = 0.3, \\
&\mathrm{p}(x) = 0, \quad x \neq 0, 2, 5.
\end{aligned}$$

Một cách biểu diễn khác là

$$\mathrm{p}(x) = \begin{cases} 0.1 & x = 0, \\ 0.6 & x = 2, \\ 0.3 & x = 5, \\ 0 & x \neq 0, 2, 5. \end{cases}$$

Bảng phân phối xác suất

x	0	2	5
P(X=x)	0.1	0.6	0.3

Hàm khối xác suất $p(x)$ có thể phụ thuộc vào một đại lượng với giá trị xác định cho trước, gọi là **tham số**. Khi thay đổi giá trị tham số thì chúng ta có một phân phối xác suất khác nhau, tập hợp các phân phối xác suất này gọi là một **họ** các phân phối xác suất.

Ví dụ 3.4. Khi tung đồng xu không chuẩn có xác suất xuất hiện mặt ngửa là α. Gọi X là biến ngẫu nhiên thỏa $X(H) = 1$, $X(T) = 0$, ta có:

$$\mathrm{P}(x, \alpha) = \begin{cases} 1 - \alpha & \text{khi } x = 0 \\ \alpha & \text{khi } x = 1 \end{cases}.$$

Mỗi giá trị α khác nhau sẽ cho một hàm khối xác suất khác nhau.

3.3 Hàm phân phối tích lũy

Định nghĩa 3.5. Hàm phân phối tích lũy *(cumulative distribution function - cdf)* của một biến ngẫu nhiên rời rạc với hàm khối $p(x)$ được định nghĩa với mọi x là

$$F(x) = \mathrm{P}(X \leq x) = \sum_{y:y \leq x} \mathrm{p}(y).$$

Nói cách khác, hàm phân phối tích lũy sẽ cho biết xác suất để biến ngẫu nhiên X có giá trị không quá x.

Ví dụ 3.6. Gọi X là số giường trống trong phòng cấp cứu của bệnh viện vào lúc 2 giờ sáng. Giả sử có bảng phân phối xác suất như sau:

x	0	1	2	3	4
p(x)	0.2	0.25	0.3	0.15	0.1

Xác suất để có nhiều nhất 2 giường còn trống là

$$F(2) = \mathrm{P}(X \leq 2) = \mathrm{P}(0) + \mathrm{P}(1) + \mathrm{P}(2) = 0.75.$$

Cần chú ý rằng với biến ngẫu nhiên rời rạc thì $F(2.7) = F(2)$ (do $X \leq 2.7 \iff X \leq 2$). Ngoài ra, cũng cần lưu ý rằng $\mathrm{P}(X < 2) = \mathrm{P}(X \leq 2) - \mathrm{P}(X = 2) = F(2) - \mathrm{P}(2) \leq F(2)$.

Từ ví dụ trên, ta biết thêm rằng:

$$F(x) = F(\lfloor x \rfloor) \text{ và } \mathrm{P}(X < x) = F(x) - \mathrm{p}(x) \leq F(x).$$

Ví dụ 3.7. Một cửa tiệm photo bán phao thi với các mức điểm khác nhau từ 6 đến 10. Gọi X là biến ngẫu nhiên mức điểm đạt được với phao thi đã mua. Có bảng phân phối:

x	6	7	8	9	10
p(x)	0.05	0.1	0.35	0.4	0.1

Ta có hàm phân phối tích lũy như sau:

$$F(x) = \begin{cases} 0 & \text{khi} & x < 6 \\ 0.05 & \text{khi} & 6 \leq x < 7 \\ 0.15 & \text{khi} & 7 \leq x < 8 \\ 0.5 & \text{khi} & 8 \leq x < 9 \\ 0.9 & \text{khi} & 9 \leq x < 10 \\ 1 & \text{khi} & 10 \leq x \end{cases}.$$

Sử dụng hàm phân phối tích lũy ta cũng có thể tìm được xác suất biến ngẫu nhiên X có giá trị rơi vào một khoảng $[a, b]$ nào đó. Có:

$$\mathrm{P}(a \leq X \leq b) = \mathrm{P}(X \leq b) - \mathrm{P}(X < a) = F(b) - F(a^-).$$

Trong đó, a^- là giá trị nguyên lớn nhất của X bé hơn a. Ta có tính chất sau cho các biến rời rạc nhận giá trị nguyên:

Định lý 3.8. *Cho X là một biến ngẫu nhiên nhận giá trị nguyên với hàm khối $p(x)$ và hàm tích lũy $F(x)$. Với mọi số nguyên $a \leq b$, ta có*

$$P(a \leq X \leq b) = F(b) - F(a-1),$$

và

$$p(a) = F(a) - F(a-1).$$

3.4 Đặc trưng

3.4.1 Kì vọng

Giả sử tổ chức của bạn có dự định đến một trường học để gây quỹ cho hoạt động sắp tới. Do thời gian có hạn, bạn chỉ có thể bước vào đúng một phòng học ngẫu nhiên để giới thiệu về tổ chức của mình với mọi người trong phòng đó. Bạn muốn biết mình sẽ được ("kỳ vọng") nói chuyện với bao nhiêu học sinh. Đó là lúc cần dùng đến giá trị kì vọng.

Định nghĩa 3.9. **Giá trị kì vọng** hay **giá trị trung bình** của biến ngẫu nhiên rời rạc X với tập giá trị có thể D, kí hiệu $\mathrm{E}(X)$ hay μ_X hay μ, là

$$\mathrm{E}(X) = \sum_{x \in D} x \cdot \mathrm{p}(x).$$

Ví dụ 3.10. Tiếp tục câu chuyện lúc nãy, giả sử tại thời điểm đó, thống kê số học sinh trong từng phòng như sau:

Phòng	A	B	C	D	E
Số lượng	32	28	32	35	20

Gọi X là biến ngẫu nhiên số người trong phòng được chọn. Bạn lập được bảng phân phối xác suất:

x	20	28	32	35
p(x)	0.2	0.2	0.4	0.2

Từ đó, số lượng học sinh bạn "kì vọng" sẽ gặp được:

$$\mathrm{E}(X) = 20 \cdot 0.2 + 28 \cdot 0.2 + 32 \cdot 0.4 + 35 \cdot 0.2 = 29.4.$$

Tất nhiên điều này không có nghĩa là bạn bước vào phòng và sẽ gặp được 29 và 2/5 người. Bạn sẽ gặp 16, 28, 32 hoặc 35 người với xác suất như trên. Giá trị này chỉ mang ý nghĩa trung bình, nếu lặp lại việc này đủ nhiều lần (giả dụ có 1000 tổ chức cùng làm) thì trung bình của mọi lần thử sẽ tiến gần 29.4.

Đôi lúc, chúng ta không chỉ quan tâm tới kì vọng của biến ngẫu nhiên mà còn quan tâm tới kì vọng của hàm số liên quan tới biến ngẫu nhiên đó. Lúc này:

$$E[h(X)] = \sum_D h(x) \cdot \mathrm{p}(x).$$

Ví dụ 3.11. Tiếp tục với ví dụ trên, bạn bắt đầu suy nghĩ về chi phí cho công tác marketing này. Giả sử để đến được trường như vậy bạn phải bỏ ra số tiền di chuyển là 50.000đ và khi gây quỹ bạn thu được một số tiền cố định a từ mỗi người bạn giới thiệu. Giá trị a cần là bao nhiêu để bạn vừa đủ thu hồi vốn mà vẫn gây quỹ được 100.000đ cho tổ chức?

Gọi $h(X)$ là số tiền lời (ngàn đồng), rõ ràng $h(X) = aX - 50$. Có:

$$E[h(X)] = \sum_D (ax - 50)\mathrm{p}(x) = a \sum_D x\mathrm{p}(x) - 50 \sum_D \mathrm{p}(x) = a\mathrm{E}(X) - 50.$$

Ta cần $E[h(X)] \geq 100$ tức $29.4a - 50 \geq 100$ vậy $a \geq 5.1$. Nói cách khác, chỉ cần tận thu mỗi người 5.100đ.

Từ ví dụ trên cũng có thể thấy:

$$E[aX + b] = a\mathrm{E}(X) + b.$$

3.4.2 Phương sai và độ lệch chuẩn

Điểm thi cuối kì môn "Những nguyên lý cơ bản của Chủ nghĩa Marx - Lenin" của hai lớp A và B được cho như sau:

x	1	2	8	10
$\mathrm{p}_A(x)$	0.2	0.35	0.2	0.25

x	4	5	6	9
$\mathrm{p}_B(x)$	0.6	0.1	0.2	0.1

Cả hai lớp A và B đều có điểm trung bình cả lớp là 5.0. Liệu có thể xem cả hai lớp là như nhau? Lúc này ta cần một thông tin khác để nhận xét gọi là phương sai (và độ lệch chuẩn). Nếu mà giá trị kì vọng xác định vị trí mà các giá trị của biến ngẫu nhiên X xoay quanh thì phương sai giúp xác định mức độ phân tán của các giá trị.

Định nghĩa 3.12. Phương sai của biến ngẫu nhiên X, kí hiệu $\mathrm{Var}(X)$ hay σ_X^2 hay σ^2 là

$$\mathrm{Var}(X) = \sum_D (x - \mu)^2 \cdot \mathrm{p}(x) = E[(X - \mu)^2].$$

Độ lệch chuẩn của X là:

$$\sigma_X = \sqrt{\sigma_X^2}.$$

Có thể thấy phương sai là giá trị kì vọng của bình phương khoảng cách giữa mọi giá trị của X với giá trị trung bình. Nếu như các giá trị của X càng tập trung về giá trị trung bình thì phương sai càng nhỏ, càng phân tán ra xa thì phương sai càng lớn.

Ví dụ 3.13. Xét điểm thi hai lớp nói trên:

$$\sigma_A^2 = 0.2(1-5)^2 + 0.35(2-5)^2 + 0.2(8-5)^2 + 0.25(10-5)^2 = 14.4 = 3.8^2.$$

Tương tự có:

$$\sigma_B^2 = 0.6(4-5)^2 + 0.1(5-5)^2 + 0.2(6-5)^2 + 0.1(9-5)^2 = 2.4 = 1.5^2.$$

Rõ ràng là lớp B có mức điểm tập trung hơn lớp A.

Thông thường, chúng ta mong muốn phương sai thấp trong các đo đạc hoặc dự đoán. Tuy nhiên, phương sai lớn cũng không thật sự có nghĩa là xấu, tùy vào từng trường hợp. Ví dụ bạn có hai phương pháp làm giàu, phương pháp có phương sai nhỏ hơn sẽ giúp bạn đảm bảo ở một mức tiền chấp nhân được còn phương pháp có phương sai lớn hơn lại mang tính "được ăn cả ngã về không", cho bạn có cơ hội có lợi nhuận cao nhưng đổi lại cũng tăng khả năng lỗ vốn.

Phương sai có một cách tính khác:

$$\begin{aligned}\mathrm{Var}(X) &= \textstyle\sum_D (x-\mu)^2 \cdot \mathrm{p}(x) \\ &= \textstyle\sum_D x^2\mathrm{p}(x) - 2\mu \sum_D x\mathrm{p}(x) + \mu^2 \sum_D \mathrm{p}(x) \\ &= \mathrm{E}(X^2) - \mu^2 = \mathrm{E}(X^2) - \mathrm{E}^2(X).\end{aligned}$$

Cũng như trên, nếu ta quan tâm đến phương sai của một hàm biến ngẫu nhiên:

$$\mathrm{Var}[h(X)] = \sum_D \left(h(x) - \mathrm{E}[h(X)]\right)^2 \mathrm{p}(x).$$

Trường hợp $h(X) = aX + b$ ta có:

$$\mathrm{Var}[ax + b - (a\mu + b)] = \mathrm{Var}[a(x-\mu)] = a^2 \sum_D (x-\mu)^2 \mathrm{p}(x) = a^2 \cdot \mathrm{Var}(X).$$

3.4.3 Mode

Định nghĩa 3.14. Mốt hay **Mode** của biến ngẫu nhiên X, kí hiệu Mode(X) là (các) giá trị của biến ngẫu nhiên X mà có xác suất lớn nhất.

$$\mathrm{Mode}(X) = \arg\max_x \mathrm{p}(x).$$

4 Một số phân phối rời rạc

4.1 Phân phối Bernoulli

Xét một thí nghiệm ngẫu nhiên có hai kết quả $\Omega = \{\omega, \bar{\omega}\}$ với xác suất cho kết quả ω là p. Gọi X là biến ngẫu nhiên thỏa $X(\omega) = 1$, $X(\bar{\omega}) = 0$. Ta có hàm mật độ

$$\mathrm{p}(x) = \begin{cases} p & \text{khi } x = 1 \\ 1 - p & \text{khi } x = 0 \\ 0 & \text{khi } x \text{ khác} \end{cases}.$$

Định nghĩa 4.1. Ta nói X có **phân phối Bernoulli** khi X có hàm khối xác suất như trên. Kí hiệu: $X \sim B(1, p)$ trong đó $p \in (0, 1)$.

Ví dụ 4.2. Chọn một người bất kì trên đường và xem giới tính (trên chứng minh thư) của người đó. Có hai khả năng là nam (M) hoặc nữ (F), với việc gặp nam có xác suất là $p = 0.8$. Gọi X là biến ngẫu nhiên thỏa $X(M) = 1$, $X(F) = 0$ ta có:

$$\mathrm{p}(x) = \begin{cases} 0.8 & \text{khi } x = 1 \\ 0.2 & \text{khi } x = 0 \\ 0 & \text{khi } x \text{ khác} \end{cases}.$$

Vậy $X \sim B(1, 0.8)$.

Ví dụ 4.3. Tung một cục xúc xắc, đặt

$$X = \begin{cases} 1 & \text{khi mặt chẵn} \\ 0 & \text{khi mặt lẻ} \end{cases}.$$

Ta có $X \sim B(1, 0.5)$.

Đặc trưng của phân phối Bernoulli:

- ⋆ Kì vọng: $\mathrm{E}(X) = p$.
- ⋆ Phương sai: $\mathrm{Var}(X) = p(1 - p)$.
- ⋆ Mode: $\mathrm{Mode}(X) = \lfloor p + 0.5 \rfloor$.

4.2 Phân phối nhị thức

Xét một thí nghiệm ngẫu nhiên có hai kết quả $\Omega = \{\omega, \bar{\omega}\}$ với xác suất cho kết quả ω là p. Ta lặp lại thí nghiệm này n lần độc lập và quan sát số lần xuất hiện kết quả ω trong n lần quan sát đó. Gọi X là số lần ω xuất hiện.

Ở mỗi thí nghiệm, xác suất có ω xuất hiện là p và $\bar{\omega}$ xuất hiện là $1 - p$. Số trường hợp mà có x thí nghiệm cho ω bằng số cách chọn x trong n thí nghiệm tức C_n^x. Do các thí nghiệm là độc lập nên xác suất để có x lần xuất hiện ω là $\mathrm{P}(X = x) = C_n^x p^x (1 - p)^{n - x}$.

Định nghĩa 4.4. Cho biến ngẫu nhiên X nhận các giá trị $0, 1, 2, \ldots, n$. X có **phân phối nhị thức**, kí hiệu $X \sim B(n, p)$ $(0 < p < 1)$ khi hàm khối xác suất có dạng:

$$\mathrm{p}(x) = \begin{cases} C_n^x p^x (1-p)^{n-x} & \text{khi } x = 0, 1, 2, \ldots, n \\ 0 & \text{khi } x \text{ khác} \end{cases}.$$

Ví dụ 4.5. Xác suất để một trang sách có lỗi in là 0.001. Ta kiểm tra một cuốn sách 20 trang. Gọi X là số trang lỗi tìm được. Ta có: $X \sim B(20, 0.001)$. Hàm khối xác suất:

$$\mathrm{p}(x) = \mathrm{P}(X = x) = \begin{cases} C_{20}^x 0.001^x 0.999^{20-x} & x = 0, 1, 2, ..., 20 \\ 0 & \text{x khác} \end{cases}$$

Xác suất để tìm thấy đúng 5 trang lỗi là:

$$\mathrm{P}(X = 5) = C_{20}^5 0.001^5 0.999^{15} = 1.53 \times 10^{-11}.$$

Xác suất để không quá 1 trang lỗi là:

$$F(1) = \mathrm{P}(X \leq 1) = \mathrm{P}(X = 0) + \mathrm{P}(X = 1) = 0.9998.$$

Đặc trưng của phân phối nhị thức:

- ⋆ Kì vọng: $\mu = np$.
- ⋆ Phương sai: $\sigma^2 = np(1-p)$.
- ⋆ Mode: các số nguyên thỏa $np + p - 1 \leq \mathrm{Mode}(X) \leq np + p$.

4.3 Phân phối Poisson

Bạn đã bao giờ ghi chú lại số email mà mình nhận được trong một ngày? Hay số khách mà gia đình bạn tiếp trong một tuần? Hay số người đi ra vào một bảo tàng trong một giờ? Những biến ngẫu nhiên này đều được mô hình hóa bởi một phân phối đặc biệt gọi là phân phối Poisson.

Định nghĩa 4.6. Cho biến ngẫu nhiên rời rạc X nhận giá trị $0, 1, 2, \ldots$ X có **phân phối Poisson**, kí hiệu $X \sim \mathrm{P}(\lambda)$, khi hàm khối xác suất có dạng:

$$\mathrm{P}(X = x) = e^{-\lambda} \frac{\lambda^x}{x!}.$$

Ví dụ 4.7. Số ca sinh đôi sinh ra trong một tháng tại bệnh viện Từ Dũ là một biến ngẫu nhiên theo phân phối Poisson với tham số $\lambda = 0.5$. Vậy xác suất để trong một tháng có 2 ca sinh đôi là $\mathrm{P}(X = 2) = e^{-0.5} \dfrac{0.5^2}{2!} = 7.58\%$.

Ở đây, λ đóng vai trò là giá trị kì vọng của biến ngẫu nhiên.

Ví dụ 4.8. Một trung tâm bưu điện nhận trung bình 150 cuộc điện thoại trong một giờ. Vậy trung bình một phút nhận 2.5 cuộc gọi. Gọi X là số cuộc gọi trung tâm nhận trong một phút, có $X \sim \mathrm{P}(2.5)$. Xác suất để trung tâm không nhận cuộc gọi nào trong một phút là $\mathrm{P}(X = 0) = e^{-2.5} = 8.2\%$.

Điều đặc biệt của các hiện tượng nêu ở trên là việc chúng ta sẽ thực hiện phép thử rất nhiều lần (ghi số email mình nhận trong một ngày, thực hiện suốt một năm chẳng hạn) và xác suất xảy ra của nó khá là nhỏ (hiện nay chắc cũng không nhiều người đi bảo tàng nữa nhỉ). Phân phối Poisson chính là giới hạn của phân phối nhị thức khi mà n (hay số lần thử) lớn và p (hay xác suất xảy ra) nhỏ.

Phân phối nhị thức $B(n, p)$ có thể được xấp xỉ bởi phân phối Poisson $\mathrm{P}(np)$ (tức $\lambda = np$) khi n lớn (thường là hơn 20) và p nhỏ (dưới 0.01).

Ví dụ 4.9. Giả sử xác suất tử vong của bệnh sốt xuất huyết là 0.007. Ta tính xác suất để có 5 người chết do sốt xuất huyết trong một nhóm 400 người. Gọi X là biến ngẫu nhiên số người chết do sốt xuất huyết trong nhóm 400 người. Ta có $X \sim B(400, 0.007)$. Do $n = 400$ lớn và $p = 0.007$ nhỏ, có thể xấp xỉ bằng phân phối Poisson với $\lambda = 400 \times 0.007 = 2.8$, có $X \sim \mathrm{P}(2.8)$. Từ đó tìm được $\mathrm{P}(X = 5) = 8.72\%$.

Ví dụ 4.10. Tỉ lệ lọ thuốc hỏng một lô rất nhiều lọ thuốc là 0.05. Ta lấy ngẫu nhiên 20 lọ. Gọi X là số lọ hỏng trong 20 lọ lấy ra. Tỉ lệ trong một lô rất nhiều thứ sẽ xấp xỉ bằng xác suất, tức xác suất một lọ thuốc hỏng là 0.05. Có $X \sim B(20, 0.05)$ và $\mathrm{P}(X = 2) = 18.87\%$. Lúc này thử xấp xỉ bằng phân phối Poisson, $X \sim 1$, có $\mathrm{P}(X = 2) = 18.39\%$. Có thể thấy khi n không đủ lớn và p không đủ nhỏ thì xấp xỉ không được chính xác lắm.

Đặc trưng của phân phối Poisson:

- ⋆ Kì vọng: $\mathrm{E}(X) = \lambda$
- ⋆ Phương sai: $\mathrm{Var}(X) = \lambda$
- ⋆ Mode: các số nguyên thỏa $\lambda - 1 \leq \mathrm{Mode}(X) \leq \lambda$

4.4 Phân phối hình học

Xét một thí nghiệm có hai kết quả $\Omega = \{\omega, \bar{\omega}\}$ với xác suất cho kết quả ω là p. Ta lặp lại thí nghiệm này nhiều lần độc lập cho đến khi kết quả ω xuất hiện thì dừng lại.

Xác suất để có x lần thực hiện thí nghiệm là xác suất để $x - 1$ lần đầu xuất hiện $\bar{\omega}$ và 1 lần cuối xuất hiện ω tức $(1 - p)^{x-1}p$.

Định nghĩa 4.11. Cho biến ngẫu nhiên X nhận các giá trị $0, 1, 2, \ldots$. X có **phân phối hình học**, kí hiệu $X \sim Geo(p)$, khi hàm khối xác suất có dạng:

$$\mathrm{P}(X = x) = (1 - p)^{x-1}p.$$

Ví dụ 4.12. Một xạ thủ có xác suất bắn trúng mục tiêu là $p = 0.7$. Anh bắn mục tiêu cho đến khi trúng thì dừng lại. Gọi X là số lần bắn trước khi dừng lại. Ta có $X \sim Geo(0.7)$. Xác suất để anh dừng lại sau 4 lần bắng là $\mathrm{P}(X = 4) = 0.3^3 \cdot 0.7 = 1.89\%$.

Đặc trưng của phân phối hình học:

- ⋆ Kì vọng: $\mathrm{E}(X) = \dfrac{1}{p}$.
- ⋆ Phương sai: $\mathrm{Var}(X) = \dfrac{1 - p}{p^2}$.

⋆ Mode: $\text{Mode}(X) = 1$.

4.5 Phân phối siêu hình học

Xét một quần thể có kích thước N gồm K phần tử có tính chất a và $N - K$ phần tử không có tính chất a. Một thí nghiệm bốc từ quần thể đó có hai kết quả là ω nếu bốc được phần tử có tính chất a và $\bar{\omega}$ nếu không. Lặp lại thí nghiệm này n lần, mỗi lần bốc không trả phần tử bốc được về lại quần thể.

Xác suất để có k lần xuất hiện ω (bốc phần tử có tính chất a) được tính như sau: chọn k phần tử từ K phần tử có tính chất a, chọn $n - k$ phần tử từ $N - K$ phần tử không có tính chất a; không gian mẫu là cách chọn n phần tử từ N phần tử của quần thể.

Định nghĩa 4.13. Cho biến ngẫu nhiên X nhận các giá trị $0, 1, 2, \ldots, n$. X có **phân phối siêu hình học**, kí hiệu $X \sim H(N, K, n)$, khi hàm khối xác suất có dạng:

$$\mathrm{P}(X = k) = \frac{C_K^k C_{N-K}^{n-k}}{C_N^n}.$$

Ví dụ 4.14. Một lọ đựng 20 quả bóng trong đó có 12 quả bóng xanh, còn lại là bóng đỏ. Bốc từ trong lọ không hoàn lại 10 quả bóng. Gọi X là số bóng xanh bốc được, ta có $X \sim H(20, 12, 10)$. Xác suất để bốc được 7 quả bóng xanh là $\mathrm{P}(X = 7) = \dfrac{C_{20}^7 C_8^3}{C_{20}^{10}}$.

Đặc trưng của phân phối siêu hình học:

⋆ Kì vọng: $\mathrm{E}(X) = n\dfrac{K}{N}$.

⋆ Phương sai: $\mathrm{Var}(X) = n\dfrac{K}{N}\dfrac{N-K}{N}\dfrac{N-n}{N-1}$.

Bài tập

Bài tập 4.1. Chứng minh công thức cho các giá trị kỳ vọng, phương sai và mode của tất cả các phân phối xác suất trên.

Bài tập 4.2. Hai xúc xắc chuẩn được tung độc lập. Gọi M là biến ngẫu nhiên số lớn hơn trong hai viên xúc xắc, ví dụ $M(1, 5) = 5$, $M(3, 3) = 3$.

1. Cho biết hàm khối xác suất của M.
2. Cho biết hàm phân phối tích lũy của M.
3. Vẽ đồ thị cho hai hàm trên.

Bài tập 4.3. Giả sử bạn đang đọc báo Tuổi Trẻ và ghi nhận lại mọi con số bạn đọc được - thu nhập của một CEO, số lít bia mà người Việt Nam uống, số tiền tham nhũng của một quan chức, số tiền thuế mà một công ty đã trốn,. . . . Chỉ quan tâm đến chữ số đầu tiên, tức xét một biến ngẫu nhiên có thể nhận giá trị $1, 2, 3, \ldots, 9$. Số liệu thực nghiệm cho thấy biến ngẫu nhiên này tuân theo Luật Benford (luật chữ số thứ nhất):

$$p(x) = \log\left(\frac{x+1}{x}\right), \quad x = 1, 2, \ldots, 9.$$

1. Không tính xác suất của từng giá trị, chứng minh $p(x)$ là một hàm khối xác suất hợp lệ.
2. Lập bảng xác suất từng giá trị.
3. Tìm trung bình, độ lệch chuẩn và mode của biến ngẫu nhiên.
4. Tìm hàm phân phối tích lũy của biến ngẫu nhiên trên.

Bài tập 4.4. Xét trò chơi: tung một con xúc xắc 3 lần, số tiền nhận được sẽ gấp 2 lần số lần tung ra mặt 6 nút (ví dụ: 2 lần nhận 4 đồng). Mỗi lần chơi phải mất A đồng. Hỏi:

1. A là bao nhiêu thì người chơi về lâu dài huề vốn (gọi là trò chơi công bằng)?
2. A là bao nhiêu thì trung bình một lần người chơi mất 1 đồng.

Bài tập 4.5. Chứng minh lại các công thức về kì vọng và phương sai của các phân phối đã nêu ở trên.

Bài tập 4.6. Cho X và Y là hai đại lượng ngẫu nhiên độc lập.

1. Giả sử $X \sim B(1, 0.2), Y \sim B(2, 0.2)$. Lập bảng phân phối xác suất của $X + Y$.
2. Giả sử $X \sim B(1, 0.5), Y \sim B(2, 0.2)$. Lập bảng phân phối xác suất của $X + Y$.
3. Trong 2 trường hợp trên, trường hợp nào $X + Y$ có phân phối nhị thức? Xác định phân phối đó, tính kì vọng và độ lệch chuẩn.

Bài tập 4.7. Các sản phẩm được sản xuất trong một dây chuyền. Để thực hiện kiểm tra chất lượng, mỗi giờ người ta rút ngẫu nhiên không hoàn lại 10 sản phẩm từ một hộp có 25 sản phẩm.
Qui trình sản xuất được báo cáo là đạt yêu cầu nếu có không quá 1 sản phẩm là thứ phẩm.

1. Nếu tất cả hộp được kiểm tra đều chứa chính xác 2 thứ phẩm, xác suất qui trình sản xuất được báo cáo đạt yêu cầu ít nhất 7 lần trong một ngày làm việc 8 giờ là bao nhiêu?
2. Sử dụng phân phối Poisson để xấp xỉ xác suất tính được trong câu trên.
3. Biết rằng lần kiểm tra chất lượng cuối cùng trong câu (1), quy trình sản xuất được báo đạt yêu cầu. Hỏi xác suất mẫu 10 sản phẩm tương ứng không chứa thứ phẩm là bao nhiêu?

Bài tập 4.8. Trong hộp có 20 sản phẩm, mỗi sản phẩm đều có thể là chính phẩm hay thứ phẩm với xác suất như nhau 0.5. Lấy ngẫu nhiên 8 sản phẩm có hoàn lại thì được toàn chính phẩm. Tính xác suất hộp có toàn chính phẩm.

Bài tập 4.9. Cho X và Y là hai biến ngẫu nhiên độc lập tuân theo phân phối Poisson với tham số lần lượt là λ_1, λ_2. Chứng minh $X + Y \sim P(\lambda_1 + \lambda_2)$.

5 Biến ngẫu nhiên liên tục

5.1 Phân phối xác suất

Trở lại với ví dụ về ấm trà, lần này, tưởng tượng bạn có số ly nhiều đến vô hạn. Lúc này thì mặc cho bạn chia trà như thế nào, lượng trà trong mỗi ly cũng sẽ tiến gần về 0. Trường hợp này giống như phân phối xác suất của biến ngẫu nhiên liên tục vậy. Trong trường hợp liên tục, biến ngẫu nhiên X có thể nhận vô hạn các giá trị, có thể là trong một khoảng $[a, b]$ trên trục số, có thể là hội của nhiều khoảng như vậy. Với mọi giá trị x, như ly trà vậy, xác suất để biến ngẫu nhiên nhận giá trị x đều là 0, hay $\mathrm{P}(X = x) = 0, \quad \forall x$.

Ví dụ 5.1. Gọi X là độ sâu của một dòng sông tại một điểm được chọn ngẫu nhiên trên mặt nước. Giả sử độ sâu tối đa của dòng sông là 20.7, vậy X nhận giá trị trong khoảng $[0, 20.7]$. Trường hợp ta "rời rạc hóa" biến ngẫu nhiên X bằng cách chỉ đo chính xác đến phần nguyên độ sâu, thì X trở thành biến ngẫu nhiên rời rạc nhận các giá trị nguyên từ 0 đến 20. Nếu biểu diễn bằng cách hình chữ nhật như hình ta có tổng diện tích (tương ứng với tổng xác suất) sẽ là 1.
Một cách rời rạc hóa khác, chính xác hơn là làm tròn đến 0.5 độ sâu. Lúc này X là biến ngẫu nhiên rời rạc nhận các giá trị bán nguyên từ 0 đến 20.5. Tuy có nhiều hình chữ nhật hơn nhưng tổng diện tích vẫn là 1.
Cứ tăng độ chính xác phép đo như vậy, lúc này ta có X là một biến ngẫu nhiên liên tục với đồ thị phân phối nhìn giống như đồ thị của một hàm số. Tất nhiên, diện tích phần bên dưới đồ thị cũng là 1.

5.2 Hàm mật độ xác suất

Do xác suất cho mọi giá trị của biến ngẫu nhiên X đều bằng 0 nên các biến ngẫu nhiên liên tục không có hàm khối xác suất như rời rạc. Ta định nghĩa một hàm khác cho biến ngẫu nhiên liên tục.

Định nghĩa 5.2. Gọi X là một biến ngẫu nhiên liên tục. **Hàm mật độ xác suất** (probability density function - pdf) của X là một hàm $f(x)$ sao cho với bất kì a, b sao cho $a \leq b$ ta có

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

Nói cách khác, xác suất để X nhận một giá trị trong khoảng $[a, b]$ là diện tích phần dưới đồ thị giới hạn trong đoạn này của hàm mật độ. Hàm này có hai tính chất:

$$0 \leq f(x) \leq 1 \text{ và } \int_{-\infty}^{\infty} f(x)dx = 1.$$

Ví dụ 5.3. Gọi X là góc quay (độ) theo chiều kim đồng hồ của kim giây tại một thời điểm nào đó trong ngày. X là một biến ngẫu nhiên liên tục với hàm mật độ:

$$f(x) = \begin{cases} \dfrac{1}{360} & 0 \leq x < 360 \\ 0 & \text{x khác} \end{cases}$$

Ví dụ 5.4. Gọi X là biến ngẫu nhiên có hàm mật độ có dạng:

$$f(x) = \begin{cases} ax & \text{khi } x \in [0,1] \\ 0 & \text{khi } x \text{ khác} \end{cases}.$$

Để đây là hàm mật độ, ta có:

$$\int_{-\infty}^{\infty} f(x)dx = 1 \iff 1 = a\int_0^1 xdx = \frac{a}{2}[(1)^2 - (0)^2] = \frac{a}{2} \iff a = 2.$$

Xác suất để biến ngẫu nhiên X nhận giá trị trong đoạn $[0.3, 0.6]$ là

$$\mathrm{P}(0.3 \leq X \leq 0.6) = \int_{0.3}^{0.6} 2xdx = 0.6^2 - 0.3^2 = 0.27.$$

5.3 Hàm phân phối tích lũy

Tương tự như trường hợp rời rạc, ta cũng định nghĩa hàm phân phối tích lũy cho biến ngẫu nhiên liên tục.

Định nghĩa 5.5. Hàm phân phối tích lũy $F(x)$ cho biến ngẫu nhiên liên tục X với mọi giá trị x là

$$F(x) = \mathrm{P}(X \leq x) = \int_{-\infty}^{x} f(t)dt.$$

Nói cách khác, $F(x)$ là phần diện tích bên dưới đồ thị nằm về bên trái của x.

Từ định nghĩa trên, ta suy ra được tính chất tương tự với định lý 3.8 như sau

Định lý 5.6. *Cho X là một biến ngẫu nhiên liên tục với hàm mật độ $f(x)$ và hàm tích lũy $F(x)$. Khi đó:*

$$f(x) = F'(x), \quad \forall x \in \mathbb{R}$$

Định lý trên đặc biệt hữu dụng trong việc tìm hàm phân phối xác suất của các biến ngẫu nhiên liên tục. Khác với biến rời rạc, hàm mật độ xác suất (khái niệm tương đương với hàm khối xác suất) không thể biểu thị được dưới dạng xác suất của một biến cố $X = x$ nào đó. Do đó việc dùng các công thức xác suất của biến cố để chứng minh các công thức liên quan tới hàm mật độ của biến liên tục sẽ không chặt chẽ. Với định lý 5.6, ta có thể dùng những công thức xác suất để chứng minh công thức cho hàm tích lũy, và đạo hàm để được công thức tương ứng cho hàm mật độ.

Phương pháp này sẽ được sử dụng nhiều khi chứng minh các công thức liên quan tới đa biến ngẫu nhiên liên tục trong phần 7.

Ví dụ 5.7. Gọi X là tải trọng (theo N) trên một chiếc cầu. Hàm mật độ xác suất của X được cho như sau:

$$f(x) = \begin{cases} \frac{1}{8} + \frac{3}{8}x & \text{khi } x \in [0,2] \\ 0 & \text{khi } x \text{ khác} \end{cases}.$$

Có thể xác định hàm phân phối tích lũy như sau. Xét mọi giá trị x trong khoảng $[0, 2]$:

$$F(x) = \int_{-\infty}^{x} f(t)dt = \int_0^x \left(\frac{1}{8} + \frac{3}{8}t\right) dt = \frac{1}{8}x + \frac{3}{16}x^2.$$

Vậy ta có hàm phân phối tích lũy:

$$F(x) = \begin{cases} 0 & \text{khi } x < 0 \\ \frac{1}{8}x + \frac{3}{16}x^2 & \text{khi } 0 \le x \le 2 \\ 1 & \text{khi } 2 < x \end{cases}.$$

Cầu chỉ có thể chịu được tải trọng không quá 1.5. Xác suất để cầu sập là:

$$\mathrm{P}(X > 1.5) = 1 - \mathrm{P}(X \le 1.5) = 1 - F(1.5) = 1 - \left(\frac{1}{8} \cdot 1.5 + \frac{3}{16} \cdot 1.5^2\right) = 39\%.$$

Không giống với biến ngẫu nhiên rời rạc, trong trường hợp biến ngẫu nhiên liên tục, ta có

$$\mathrm{P}(X < x) = \mathrm{P}(X \le x) - \mathrm{P}(X = x) = F(x) - 0 = F(x).$$

Do đó ta có:

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

5.4 Đặc trưng

Tương tự biến ngẫu nhiên rời rạc, các biến ngẫu nhiên liên tục cũng có các đặc trưng như kỳ vọng, phương sai, độ lệch chuẩn.

Định nghĩa 5.8. **Giá trị kì vọng** hay **giá trị trung bình** của biến ngẫu nhiên liên tục X, kí hiệu $\mathrm{E}(X)$ hay μ_X hay μ, là

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

Định nghĩa 5.9. **Phương sai** của biến ngẫu nhiên X, kí hiệu $\mathrm{Var}(X)$ hay σ_X^2 hay σ^2 là

$$\mathrm{Var}(X) = \int_{-\infty}^{\infty} (x - \mu)^2 \cdot f(x) dx = E[(X - \mu)^2].$$

Độ lệch chuẩn của X là:

$$\sigma_X = \sqrt{\sigma_X^2}.$$

Định nghĩa 5.10. **Mốt** hay **Mode** của biến ngẫu nhiên X, kí hiệu $\mathrm{Mode}(X)$ là (các) giá trị của biến ngẫu nhiên X mà cho hàm mật độ xác suất lớn nhất.

$$\mathrm{Mode}(X) = \arg\max_x f(x).$$

Các tính chất khác mà biến ngẫu nhiên rời rạc có cũng xuất hiện trong biến ngẫu nhiên liên tục.

Ví dụ 5.11. Hai loài đang tranh giành một lượng thức ăn nhất định. Gọi X là biến ngẫu nhiên tỷ lệ thức ăn mà loài 1 đang chiếm. Giả sử X có hàm mật độ xác suất:

$$f(x) = \begin{cases} 1 & \text{khi } x \in [0, 1] \\ 0 & \text{khi } x \text{ khác} \end{cases}.$$

Tỷ lệ thức ăn loài 1 kỳ vọng chiếm được là:

$$\mathrm{E}(X) = \int_0^1 x dx = 0.5.$$

Lượng lương thực mà loài chiếm nhiều hơn chiếm được là:

$$h(X) = max(X, 1-X) = \begin{cases} 1-X & \text{khi } 0 \leq X < 0.5 \\ X & \text{khi } 0.5 \leq X \leq 1 \end{cases}.$$

Tỷ lệ thức ăn loài chiếm được nhiều hơn kỳ vọng chiếm được là:

$$\begin{aligned} E[h(x)] &= \int_{-\infty}^{\infty} \max(x, 1-x) f(x) dx = \int_0^1 \max(x, 1-x) dx \\ &= \int_0^{0.5} (1-x) dx + \int_{0.5}^1 x dx = \frac{3}{4}. \end{aligned}$$

Một đặc trưng cần nói đến của biến ngẫu nhiên liên tục là phân vị.

Định nghĩa 5.12. Cho p là một số trong đoạn $[0, 1]$. **Phân vị thứ 100p** của phân phối của biến ngẫu nhiên liên tục X, kí hiệu $\eta(p)$, thỏa

$$p = F(\eta(p)).$$

Nói cách khác, phân vị thứ $100p$ của phân phối chính là giá trị η sao cho diện tích phần bên trái η là p.
Lưu ý: hàm phân phối tích lũy là hàm tăng từ 0 tới 1, do đó giá trị $\eta(p)$ luôn tồn tại với mọi $p \in (0, 1)$.

Ví dụ 5.13. Lượng vải bán được (theo m^2) của một xí nghiệp trong một tuần xác định là một biến ngẫu nhiên X với hàm mật độ

$$f(x) = \begin{cases} 1.5(1-x^2) & \text{khi } 0 \leq x \leq 1 \\ 0 & \text{khi } x \text{ khác} \end{cases}$$

Hàm phân phối trong khoảng $[0, 1]$ là:

$$F(x) = \int_0^x 1.5(1-t^2) dt = 0.5x(3-x^2).$$

Trung vị, hay phân vị thứ 50, của phân phối này là η thỏa:

$$0.5 = F(\eta) = 0.5\eta(3-\eta^2) \iff \eta^3 - 3\eta + 1 = 0 \iff \eta = 0.347.$$

Vậy trong thời gian dài, 50% số tuần sẽ có lượng vải bán ra ít hơn $0.347m^2$.

Bài tập

Bài tập 5.1. Tuổi thọ của một loài côn trùng là một biến ngẫu nhiên liên tục X (đơn vị: tháng) có hàm mật độ:

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

1. Tìm hằng số k
2. Tìm $F(x)$
3. Tìm $E(X), \mathrm{Var}(X), \mathrm{Mode}(X)$
4. Tìm xác suất côn trùng chết trước 1 tháng tuổi

Bài tập 5.2. Một họ các hàm mật độ xác suất được sử dụng để xấp xỉ phân phối của thu nhập và số lượng người trong thành phố là họ Pareto. Họ này gồm 2 tham số, k và θ đều > 0 và có hàm mật độ xác suất:

$$f(x; k, \theta) = \begin{cases} \dfrac{k\theta^k}{x^{k+1}} & x \geq \theta \\ 0 & x < \theta \end{cases}.$$

1. Phác họa đồ thị của họ hàm này, thay đổi các tham số
2. Kiểm chứng các hàm trong họ này thỏa mãn yêu cầu của một hàm mật độ xác suất.
3. Gọi X là biến ngẫu nhiên có hàm mật độ xác suất trên, với $b > \theta$ xác định, tìm $P(X \leq b)$
4. Với $\theta < a < b$, tìm $P(a \leq X \leq b)$

6 Một số phân phối liên tục

6.1 Phân phối đều

Định nghĩa 6.1. Biến ngẫu nhiên liên tục X được gọi là có **phân phối đều** trên đoạn $[a, b]$, ký hiệu $X \sim U(a, b)$, nếu hàm mật độ xác suất có dạng

$$f(x) = \begin{cases} \dfrac{1}{b-a} & \text{khi } x \in [a, b] \\ 0 & \text{khi } x \text{ khác} \end{cases}.$$

Từ định nghĩa trên, ta có hàm phân phối xác suất của $X \sim U(a, b)$:

$$F(x) = \begin{cases} 0 & \text{khi } x < a \\ \dfrac{x-a}{b-a} & \text{khi } x \in [a, b] \\ 1 & \text{khi } x > b \end{cases}.$$

Tính chất 6.2. *Đặc trưng của phân phối đều:*

- ⋆ *Kì vọng:* $E(X) = \dfrac{a+b}{2}$.
- ⋆ *Phương sai:* $Var(X) = \dfrac{(a-b)^2}{12}$.

Ví dụ 6.3. Tại một trạm xe buýt khoảng cách giữa các chuyến liên tiếp của một tuyến xe buýt T là 15 phút. Chuyến đầu tiên đến trạm lúc 7 giờ sáng. Một hành khách tới trạm xe buýt vào một thời điểm có phân phối đều từ 7 giờ tới 7 giờ 30 để đi tuyến xe buýt T. Gọi X là thời điểm hành khách đến trạm tính từ 7 giờ, ta có $X \sim U(0, 30)$. Xác suất để anh ta phải đợi không quá k phút ($k \leq 15$) là:

$$\begin{aligned} \mathrm{P}(15-k \leq X \leq 15 \cup 30-k \leq X \leq 30) &= F(15) - F(15-k) + F(30) - F(30-k) \\ &= 0.5 - (0.5 - \frac{k}{30}) + 1 - (1 - \frac{k}{30}) = \frac{k}{15} \end{aligned}.$$

Ví dụ 6.4. Cho X là một biến ngẫu nhiên liên tục theo phân phối đều trên đoạn $[a, b]$. Xác suất để X nhận giá trị nằm trong khoảng 1 độ lệch chuẩn so với kì vọng là:

$$\begin{aligned} \mathrm{P}(\mu - \sigma \leq X \leq \mu + \sigma) &= F(\mu + \sigma) - F(\mu - \sigma) \\ &= \frac{(\mu + \sigma - a) - (\mu - \sigma - a)}{b-a} = \frac{2\sigma}{b-a} \\ &= 2\frac{b-a}{2\sqrt{3}}\frac{1}{b-a} = \frac{1}{\sqrt{3}} = 57.74\% \end{aligned}.$$

6.2 Phân phối chuẩn

Định nghĩa 6.5. Cho biến ngẫu nhiên liên tục X với hai tham số μ và $\sigma > 0$. X được gọi là có **phân phối chuẩn** tương ứng với tham số μ, σ^2, ký hiệu $X \sim N(\mu, \sigma^2)$, nếu hàm mật độ xác suất có dạng

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right].$$

Câu hỏi đặt ra: **Hàm mật độ xác suất trên có thỏa mãn tính chất cơ bản của hàm phân phối xác suất ở định nghĩa 5.2 không?** Định lý sau khẳng định là có.

Định lý 6.6. *Với mọi $\sigma, \mu \in \mathbb{R}$, ta có đẳng thức sau:*

$$\int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx = 1.$$

Nói cách khác hàm mật độ xác suất của phân phối chuẩn là hợp lệ.

Điều tiện lợi của phân phối chuẩn là các tham số của nó cũng chính là kỳ vọng và phương sai.

Tính chất 6.7. *Cho $X \sim N(\mu, \sigma^2)$, khi đó $E(X) = \mu$ và $Var(X) = \sigma^2$.*

Các phân phối chuẩn với tham số khác nhau tuy có hàm mật độ xác suất khác nhau nhưng có thể đưa về giống như nhau qua một phép biến đổi đơn giản.

Tính chất 6.8. *Nếu $X \sim N(\mu, \sigma^2)$ thì với mọi $a, b \in \mathbb{R}, a \neq 0$, ta có*

$$aX + b \sim N(a\mu + b, a^2\sigma^2).$$

Tính chất 6.8 cho ta một phương pháp để tính các xác suất của mọi phân phối chuẩn bất kì mà chỉ cần biết các giá trị xác suất của một phân phối chuẩn duy nhất. Người ta quy ước sử dụng một dạng phân phối chuẩn để làm "cơ sở" cho mọi phân phối chuẩn khác.

Định nghĩa 6.9 (Phân phối chuẩn tắc)**.** Biến ngẫu nhiên liên tục Z có **phân phối chuẩn tắc**, kí hiệu $Z \sim N(0,1)$ nếu hàm mật độ xác suất có dạng:

$$f(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right).$$

Từ định nghĩa trên, ta định nghĩa hàm phân phối xác suất của $Z \sim N(0,1)$:

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx.$$

Tích phân này không có công thức từ các hàm sơ cấp, do đó thông thường người ta sử dụng **bảng tra** để tìm giá trị của $\Phi(z)$. Một số tính chất của $\Phi(z)$:

Tính chất 6.10.

1. $\Phi(z) + \Phi(-z) = 1, \quad \forall z \in \mathbb{R}.$
2. $P(-a \leq Z \leq a) = 2\Phi(a) - 1, \quad \forall a \in \mathbb{R}.$
3. *Chúng ta có thể dễ dàng đưa một phân phối chuẩn bất kì về lại phân phối chuẩn hóa. Nếu $X \sim N(\mu, \sigma^2)$ thì $\dfrac{X-\mu}{\sigma} \sim N(0,1)$. Từ đó, cũng có:*

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

Ví dụ 6.11. Chỉ số thông minh (IQ), được đo bằng bài kiểm tra IQ Stanford-Binet, có phân phối chuẩn trong một tổng thể nào đó. IQ trung bình là 100 điểm, và độ lệch chuẩn là 16 điểm. Gọi X là IQ của một người, ta có $X \sim N(100, 16^2)$. Phần trăm số người trong tổng thể có IQ trên 140 là:

$$\mathrm{P}(X > 140) = 1 - \mathrm{P}(X \leq 140) = 1 - \Phi\left(\frac{140 - 100}{16}\right) = 0.62\%.$$

Ví dụ 6.12. Cho X là một biến ngẫu nhiên liên tục theo phân phối chuẩn, $X \sim N(\mu, \sigma^2)$. Xác suất để X nhận giá trị nằm trong khoảng 1 độ lệch chuẩn so với kì vọng là:

$$\mathrm{P}(\mu - \sigma \leq X \leq \mu + \sigma) = \Phi(1) - \Phi(-1) = 2\Phi(1) - 1 = 68.26\%.$$

Giống như phân phối Poisson, phân phối chuẩn cũng có thể xấp xỉ phân phối nhị thức, nhưng trong trường hợp p không quá gần 0 hoặc 1. Khi n lớn, ta có thể xấp xỉ phân phối $B(n, p)$ bằng phân phối chuẩn $N(np, np(1-p))$.

Ngoài ra, do sử dụng một phân phối liên tục để xấp xỉ một phân phối rời rạc, để việc xấp xỉ được chính xác hơn, người ta còn đề xuất phép hiệu chỉnh liên tục:

$$\mathrm{P}(X \leq x) = \mathrm{P}(X < x + 0.5).$$

$$\mathrm{P}(X < x) = \mathrm{P}(X < x - 0.5).$$

Ví dụ 6.13. Điểm thi của một học sinh học môn Tư tưởng Hồ Chi Minh có xác suất dưới trung bình là 0.8. Xét một lớp có 64 học sinh. Gọi X là số học sinh dưới trung bình, có $X \sim B(64, 0.8)$. Ta muốn biết xác suất không quá 50 học sinh rớt môn là bao nhiêu.

1. Tính theo phân phối nhị thức ta có $\mathrm{P}(X \leq 50) = 40.19\%$.
2. Tính xấp xỉ bằng phân phối chuẩn $N(51.2, 10.24)$, không dùng hiệu chỉnh liên tục có $\mathrm{P}(X \leq 50) = \Phi(-0.375) = 35.57\%$.
3. Sử dụng phép hiệu chỉnh liên tục cho $\mathrm{P}(X < 50.5) = \Phi(-0.22) = 41.29\%$.

Có thể thấy phép hiệu chỉnh liên tục cho xấp xỉ tốt hơn đáng kể.

Z-Score và ứng dụng của phân phối chuẩn

Phân phối chuẩn có rất nhiều ứng dụng trong lĩnh vực Machine Learning nói riêng và Xác Suất thống kê nói chung. Một ứng dụng tiêu biểu là phân phối chuẩn thường được sử dụng là thước đo cho để kiểm tra xem một giá trị nào đó có phải giá trị của sự kiện hay không. Nếu ta có một người cân nặng 72kg, và ta đã có một số dữ liệu cân nặng của nam giới, ta có thể tính khả năng cân nặng đó là của nam hay nữ dựa vào **Z-Score**. Phương pháp này rất hiệu quả đối với phần lớn các dạng phân phối. Thông tin này kết hợp với nhiều thông tin khác như chiều cao, kích thước chân,. . . có thể cho ta dự đoán với độ chính xác cao. Trong Machine Learning, khi ta có mẫu thử đủ lớn (thông thường trên 40), trung bình và độ lệch chuẩn của mẫu thử (tức ta lấy ngẫu nhiên mẫu thử nhiều lần) sẽ gần bằng trung bình và độ lệch chuẩn thật sự và được xem là vậy.

Phần lớn sự kiện ngoài đời sống có phân phối gần bằng phân phối chuẩn. Vì vậy, ta có thể **giả thiết phân phối ban đầu** (initial value) cho kết quả cho Machine Learning trước khi train là phân phối chuẩn. Chú ý rằng phân phối chuẩn hiệu quả nhất khi sử dụng cho hàm liên tục.

Z-Score là phương pháp để tính có độ lệch chuẩn mà một giá trị nào đó cách giá trị trung bình của phân phối chuẩn bao nhiêu. Z-score có công dụng tính khả năng một giá trị nào đó có thuộc sự kiện hay không, do nếu cách *quá xa* giá trị trung bình, giá trị đó khó lòng thuộc sự kiện.

Định nghĩa 6.14. Khi tiến hành thống kê một số liệu X trên một cộng đồng C, gọi μ là giá trị trung bình và σ^2 là phương sai, tức là σ là độ lệch chuẩn. Khi đó với một cá thể bất kì có giá trị x, giá trị

$$z = Z(x) = \frac{x-\mu}{\sigma}.$$

được gọi là **Z-Score** của giá trị x.

Khi đó xác suất giá trị x xuất hiện là $\Phi(-|Z|)$. Có thể thấy, khoảng cách giữa x và μ càng lớn thì $|Z|$ càng lớn, do đó $\Phi(-|Z|)$ càng nhỏ.

Có thể thấy, công thức của Z-Score cũng gần giống như công thức biến đổi một phân phối chuẩn sang phân phối chuẩn tắc. Tuy nhiên, điểm khác biệt là 2 giá trị μ và σ chỉ là giá trị trung bình và độ lệch chuẩn **thực nghiệm** của cộng đồng C.

Hiển nhiên ta không thể khẳng định chắc chắn phân phối của X trên C là phân phối chuẩn, cũng không thể khẳng định phân phối này có kỳ vọng μ và độ lệch chuẩn σ. Ta chỉ đang *xấp xỉ* phân phối trên bằng phân phối chuẩn, và *xấp xỉ* kỳ vọng và độ lệch chuẩn thực sự bằng giá trị thực nghiệm của chúng. Ta hy vọng rằng cách xấp xỉ này sẽ mang lại một kết quả xấp xỉ đủ gần cho xác suất giá trị x xuất hiện.

Trong thực tế, xấp xỉ trên khá chuẩn xác và được dùng nhiều trong thống kê để kiểm tra các giả thuyết hoặc phát hiện việc làm giả dữ liệu. Ta xem xét ví dụ sau:

Ví dụ 6.15. Điểm trung bình cả năm của lớp 10A5 của trường trung học ABC là 8.8 với độ lệch chuẩn 0.2. Để điểm trung bình của cả khối trên 9.5 có thật sự khả thi hay không?

Giả sử phân phối điểm cả năm của cả khối là phân phối chuẩn và phân phối điểm của lớp 10A1 không quá lệch qua một bên so với cả khối. Z-score của số điểm này là:

$$Z(8.8) = \frac{8.8-9.5}{0.2} = -3.5.$$

Ta có $\Phi(3.5) < 0.0002$, xác suất quá thấp nên ta kết luận điểm trung bình cả khối của trường ABC gần như không thể trên 9.5.

Bài tập

Bài tập 6.1. Tính chiều cao của 6 người bất kì mà bạn biết, tính phương sai, độ lệch chuẩn. Sử dụng Z-score để tính xác suất hai đầu xem chiều cao trung bình của người Việt Nam có khả thi hay không dựa vào dữ liệu thu thập được.

Bài tập 6.2. Tạo phân phối chuẩn cho chiều cao của người Việt Nam, dựa vào chiều cao trung bình và độ lệch chuẩn 12cm. Tạo những giá trị ngẫu nhiên dựa vào phân phối chuẩn ấy (có thể sử dụng code hoặc tìm kiếm trên mạng). Kiểm tra xem với bao nhiêu mẫu thử thì giá trị trung bình và độ lệch chuẩn của mẫu sẽ gần như bằng giá trị thật.

Bài tập 6.3. Entropy H của một biến ngẫu nhiên liên tục X được định nghĩa là $H = E[-\ln f_X(X)]$ với f_X là hàm mật độ xác suất của biến ngẫu nhiên X và ln là logarit tự nhiên. Tính entropy của biến ngẫu nhiên theo phân phối chuẩn và $\mu = 0, \sigma^2 = 2$.

7 Đa biến ngẫu nhiên

Ở trên, chúng ta đã nói về trường hợp 1 biến ngẫu nhiên. Tuy nhiên, trong thực tế, người ta thường quan tâm đến 2 hay nhiều biến ngẫu nhiên. Lấy ví dụ: gọi X, Y lần lượt là các biến ngẫu nhiên chỉ cân nặng (kg) và chiều cao (m) của một người. Việc nghiên cứu nhiều biến ngẫu nhiên, bên cạnh xác suất từng biến, xác suất hợp, xác suất biên,... còn liên quan đến việc nghiên cứu sự độc lập và mức độ tương quan giữa 2 biến ngẫu nhiên.

Trong khuôn khổ bài giảng, chúng ta nói đến trường hợp 2 biến ngẫu nhiên, việc mở rộng ra n biến ngẫu nhiên được thực hiện tương tự.

7.1 Phân phối kết hợp

Tương tự như với trường hợp 1 biến ngẫu nhiên, các cặp biến ngẫu nhiên sẽ có những xác suất ứng với những cặp giá trị chúng có thể nhận. Những giá trị xác suất này được thể hiện bằng **hàm khối xác suất kết hợp** đối với các cặp biến rời rạc và **hàm mật độ xác suất kết hợp** đối với các cặp biến liên tục.

Tập hợp các cặp giá trị có thể nhận được của cặp biến ngẫu nhiên và xác suất tương ứng của chúng tạo thành **phân phối kết hợp** của cặp biến ngẫu nhiên đó.

7.1.1 Hàm khối xác suất kết hợp

Định nghĩa 7.1. Gọi X và Y là hai biến ngẫu nhiên rời rạc xác định trên không gian mẫu S của một phép thử. **Hàm khối xác suất kết hợp** $\mathrm{p}(x, y)$ được định nghĩa với mỗi cặp số (x, y) là:

$$\mathrm{p}(x, y) = P(X = x \text{ và } Y = y).$$

Ta có điều kiện tương tự:

$$\mathrm{p}(x, y) \in [0, 1] \text{ và } \sum_x \sum_y \mathrm{p}(x, y) = 1.$$

Định lý 7.2. *Cho 2 biến ngẫu nhiên rời rạc X, Y với miền giá trị D_X, D_Y và hàm khối kết hợp $p(x, y)$. Với mỗi tập $A \subseteq D_X \times D_Y$, ta có*

$$P[(X, Y) \in A] = \sum_{(x,y) \in A}\sum p(x, y).$$

7.1.2 Hàm mật độ xác suất hợp

Định nghĩa 7.3. Gọi X và Y là hai biến ngẫu nhiên liên tục. **Hàm mật độ xác suất hợp** $f(x, y)$ cho hai biến ngẫu nhiên này thỏa mãn

$$f(x, y) \in [0, 1] \text{ và } \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)dxdy = 1.$$

Định lý 7.4. *Cho X và Y là 2 biến ngẫu nhiên liên tục có hàm mật độ kết hợp $f(x, y)$. Với mỗi tập $A \subseteq \mathbb{R}^2$, ta có:*

$$P[(X, Y) \in A] = \iint_A f(x, y)dxdy.$$

7.1.3 Hàm phân phối tích lũy hợp

Định nghĩa 7.5. Hàm phân phối tích lũy hợp được định nghĩa là:

$$F_{X,Y}(x,y) = P(X \leq x, Y \leq y), \quad \forall x, y \in \mathbb{R}.$$

Nói cách khác, hàm phân phối tích lũy hợp là giao của phần bên trái hàm tích lũy của riêng biến X và biến Y.

Định lý 7.6. *Cho 2 biến ngẫu nhiên X và Y, ta có:*

1. *Nếu X, Y rời rạc: $F_{X,Y}(x_0, y_0) = \sum_{x \leq x_0} \sum_{y \leq y_0} p(x,y)$.*
2. *Nếu X, Y liên tục: $F_{X,Y}(x_0, y_0) = \int_{-\infty}^{y_0} \int_{-\infty}^{x_0} f(x,y)dxdy$.*

Các định lý sau là mở rộng của định lý 3.8 và định lý 5.6:

Định lý 7.7. *Nếu X, Y là 2 biến ngẫu nhiên nhận giá trị nguyên, ta có:*

$$p(x,y) = F_{XY}(x,y) - F_{XY}(x-1,y) - F_{XY}(x,y-1) + F_{XY}(x-1,y-1), \quad \forall x, y \in \mathbb{Z}.$$

Định lý 7.8. *Nếu X, Y là 2 biến ngẫu nhiên liên tục, ta có:*

$$f(x_0, y_0) = \frac{d^2}{dxdy} F_{X,Y}(x_0, y_0), \quad \forall x_0, y_0 \in \mathbb{R}.$$

7.2 Phân phối biên

Một câu hỏi tự nhiên khi xem xét các cặp biến ngẫu nhiên với phân phối kết hợp là: **Nếu ta chỉ quan tâm tới giá trị của một biến thì phân phối xác suất của nó sẽ như thế nào?** Mỗi biến ngẫu nhiên X đều có một phân phối xác suất p_X tương ứng. Nếu ta chỉ được cho phân phối kết hợp p_{XY} giữa X với một biến ngẫu nhiên Y nào đó, phân phối p_X được gọi là **phân phối biên** (**marginal distribution**) của X trong cặp (X, Y).

Tương ứng với các trường hợp biến ngẫu nhiên rời rạc và liên tục, ta cũng có các khái niệm hàm khối xác suất biên và hàm mật độ xác suất biên.

Định lý 7.9. *Giả sử (X, Y) là một cặp biến ngẫu nhiên rời rạc có hàm khối kết hợp $p(x,y)$. Khi đó hàm khối xác suất của X, như một biến ngẫu nhiên riêng lẻ, được cho bởi:*

$$p_X(x) = \sum_y p(x,y), \qquad \text{với mọi giá trị của } x.$$

Định nghĩa 7.10. Khi (X, Y) là một cặp biến ngẫu nhiên rời rạc, hàm khối xác suất của X được gọi là **Hàm khối xác suất biên của X**.

Định lý 7.11. *Giả sử (X, Y) là một cặp biến ngẫu nhiên liên tục có hàm mật độ kết hợp $f(x,y)$. Khi đó hàm mật độ xác suất của X, như một biến ngẫu nhiên riêng lẻ, được cho bởi:*

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy, \quad \text{với } -\infty < x < \infty.$$

Định nghĩa 7.12. Khi (X, Y) là một cặp biến ngẫu nhiên liên tục, hàm mật độ xác suất của X được gọi là **Hàm mật độ xác suất biên của X**.

Để xem xét một cách trực quan hơn khái niệm phân phối xác suất biên và cách xác định nó, ta có thể hình dung một chương trình máy tính xuất ra một cặp giá trị ngẫu nhiên (x, y) theo phân phối kết hợp p (hoặc f) ở mỗi lần chạy. Để đơn giản hóa, ta giả sử rằng phân phối kết hợp là rời rạc.

Để tìm xác suất $X = x_0$ với một x_0 bất kỳ, ta chạy chương trình N lần với N rất lớn, đếm số lần xuất hiện cặp (x_0, y) với y bất kỳ, lấy số lần đó chia cho N.

Gọi số lần xuất hiện cặp (x, y) trong N lần thử là $N(x, y)$ và số lần thỏa mãn $x = x_0$ là $N(x_0)$, ta có:

$$N(x_0) = \sum_y N(x_0, y) \implies \frac{N(x_0)}{N} = \sum_y \frac{N(x_0, y)}{N}.$$

Cho $N \to \infty$, ta được:

$$\mathrm{p}_X(x_0) = \sum_y \mathrm{p}(x_0, y).$$

Chứng minh chi tiết của định lý 7.9 và định lý 7.11 sẽ được dành cho bạn đọc.

Khái niệm xác suất biên giúp ta có cái nhìn nhiều chiều về đa biến ngẫu nhiên, ta không chỉ làm việc với chúng như những bộ các biến ngẫu nhiên phụ thuộc lẫn nhau, mà còn có thể tách riêng chúng ra để tiện lấy các phần thông tin đáng quan tâm. Điển hình là định lý sau:

Định lý 7.13. *Cho 2 biến ngẫu nhiên độc lập X và Y.*

1. *Nếu X, Y rời rạc với miền giá trị D_X và D_Y, ta có:*

$$p_{XY}(x, y) = p_X(x)p_Y(y) \quad \forall x \in D_X, y \in D_Y$$

2. *Nếu X, Y liên tục thì ta có:*

$$f(x, y) = f_X(x)f_Y(y), \quad \forall x, y \in \mathbb{R}$$

Nói nôm na, định lý trên cho thấy các cặp biến ngẫu nhiên độc lập có thể được tách rời hoàn toàn. Chứng minh sẽ được dành cho bạn đọc.

7.3 Một số tính chất đặc trưng

Phần này sẽ giới thiệu hai tính chất đặc trưng của phân phối kết hợp như kỳ vọng và *hiệp phương sai*. Đại lượng đầu tiên được tìm hiểu là kỳ vọng.

Định lý 7.14. *Nếu X, Y là 2 biến ngẫu nhiên rời rạc có miền giá trị lần lượt là D_X, D_Y, với mọi hàm số $h : D_X \times D_Y \to \mathbb{R}$, ta có:*

$$E(h(X, Y)) = \sum_{x \in D_X} \sum_{y \in D_Y} h(x, y)p(x, y).$$

Định lý 7.15. *Nếu X, Y là 2 biến ngẫu nhiên liên tục, với mọi hàm số $h : \mathbb{R}^2 \to \mathbb{R}$, ta có:*

$$E(h(X, Y)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x, y)f(x, y)dxdy.$$

Hai định lý quan trọng trên cho phép ta tính toán các giá trị kỳ vọng của một hàm số theo X, Y mà không cần xem xét đến $h(X, Y)$ như một biến ngẫu nhiên và tính toán phân phối của nó.

Tiếp theo, ta đến với khái niệm *hiệp phương sai*, một khái niệm đặc trưng ở đa biến ngẫu nhiên thay cho khái niệm phương sai.

Định nghĩa 7.16. Cho 2 biến ngẫu nhiên X, Y, **hiệp phương sai** của X và Y được định nghĩa bởi công thức:

$$\mathrm{Cov}(X, Y) = \mathrm{E}\left[(X - \mathrm{E}(X))(Y - \mathrm{E}(Y))\right].$$

Tính chất 7.17. *Hiệp phương sai của X và Y còn có thể được tính bởi $Cov(X, Y) = E(XY) - E(X)E(Y)$.*

Hiệp phương sai của cặp biến ngẫu nhiên (X, Y) đo *độ phân tán của tương đối giữa chúng*. Nói nôm na, hiệp phương sai đo độ phân tán của X tương ứng với khi Y nằm gần và xa kỳ vọng của nó, và ngược lại.

Hiểu một cách giản lược, hiệp phương sai đo mức độ ảnh hưởng lẫn nhau giữa X và Y. Có thể thấy 2 đại lượng $X - \mathrm{E}(X)$ và $Y - \mathrm{E}(Y)$ đều có kỳ vọng là 0. Tuy nhiên, nếu X và Y có xu hướng tăng cùng nhau, việc Y nằm trên kỳ vọng sẽ dẫn đến khả năng X nằm trên kỳ vọng cao, do đó $\mathrm{Cov}(X, Y) > 0$.

Tương tự, việc X và Y có xu hướng *xung khắc*, tức là biến này tăng thì biến kia giảm, sẽ dẫn đến $\mathrm{Cov}(X, Y) < 0$.

Hướng suy luận trực giác trên dẫn đến kết quả sau cho 2 biến độc lập, mà chứng minh chặt chẽ sẽ dành cho bạn đọc:

Định lý 7.18. *Nếu X và Y độc lập thì $Cov(X, Y) = 0$, hay $E(XY) = E(X)E(Y)$.*

Lưu ý. Định lý trên **không** có chiều ngược lại. Hai biến ngẫu nhiên có hiệp phương sai bằng 0 không nhất thiết phải độc lập.

Bài tập

Bài tập 7.1. Chứng minh định lý 7.7 và định lý 7.8.

Bài tập 7.2.

(a) Chứng minh định lý 7.9. **Gợi ý:** tính chất 1.22 của hệ biến cố đầy đủ có thể mở rộng cho vô hạn biến cố.

(b) Chứng minh định lý 7.11. **Gợi ý:** sử dụng hàm tích lũy kết hợp, định lý 7.4 và định lý 7.8.

Bài tập 7.3. Cho 2 biến ngẫu nhiên X và Y.

(a) Giả sử X, Y rời rạc với hàm khối $\mathrm{p}(x, y)$. Chứng minh

$$\mathrm{E}(X) = \sum_x x\mathrm{p}(x, y)$$

bằng 2 cách: dùng định lý 7.14 của kỳ vọng hợp và dùng định lý 7.9 của phân phối biên.

(b) Giả sử X, Y liên tục với hàm mật độ $f(x, y)$. Chứng minh

$$\mathrm{E}(X) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f(x, y) \mathrm{d}x\mathrm{d}y$$

bằng 2 cách: dùng định lý 7.15 của kỳ vọng hợp và dùng định lý 7.11 của phân phối biên.

Bài tập 7.4. Chứng minh định lý 7.13 cho trường hợp 2 biến liên tục. **Gợi ý:** sử dụng định lý 7.8 và công thức nhân xác suất của 2 biến cố độc lập.

Bài tập 7.5. Chứng minh định lý 7.18. **Gợi ý:** dùng công thức nhân xác suất và các định lý định lý 7.15, định lý 7.14.

Bài tập 7.6. Cho 2 biến ngẫu nhiên X, Y, chứng minh

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X, Y)$$

từ đó suy ra $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$ nếu chúng độc lập.

Bài tập 7.7. An và Bình hẹn gặp nhau tại một thời điểm giữa 5 giờ và 6 giờ chiều để đi ăn tối. Gọi X là thời điểm An tới, Y là thời điểm Bình tới. Giả sử X và Y độc lập và đều tuân theo phân phối đều trên đoạn $[5, 6]$.

1. Cho biết hàm mật độ xác suất hợp của X và Y.
2. Cho biết xác suất cả hai cùng đến tại một thời điểm giữa 5:15 và 5:45.
3. Nếu người đến trước chỉ đợi 10 phút trước khi bỏ đi ăn chỗ khác, xác suất hai người cùng ăn tối là bao nhiêu?

Bài tập 7.8. Hai bộ phận của một máy tính có hàm mật độ xác suất kết hợp cho tuổi đời X và Y là:

$$f(x, y) = \begin{cases} x \exp\left(-x(1 + y)\right) & \text{với } x \geq 0, y \geq 0 \\ 0 & \text{với } x, y khc \end{cases}.$$

1. Tính xác suất tuổi đời X của bộ phận đầu tiên vượt quá 3.
2. Cho biết hàm mật độ xác suất biên của X và Y. Hai biến ngẫu nhiên này có độc lập hay không? Giải thích.
3. Tính xác suất tuổi đời của ít nhất 1 bộ phận vượt quá 3.

Bài tập 7.9. Trung muốn cắt một tấm bìa hình vuông với mỗi cạnh có độ dài L. Tuy nhiên, do vụng về, Trung cắt nhầm thành tấm bìa hình chữ nhật với một cặp cạnh độ dài X và cặp cạnh còn lại độ dài Y. Giả sử X và Y độc lập và đều tuân theo phân phối đều trên đoạn $[L - A, L + A]$ (với $0 < A < L$). Chứng minh $\mathrm{E}(XY) = \mathrm{E}(X)\mathrm{E}(Y)$ và tính diện tích trung bình của hình chữ nhật.

8 Xác suất có điều kiện, Công thức Bayes, Xác suất tiền-hậu nghiệm

8.1 Xác suất có điều kiện, công thức Bayes

Giả thiết các biến cố độc lập là lý tưởng và thường khó đạt được trong thực tế. Để nghiên cứu sự tương tác, phụ thuộc lẫn nhau giữa các biến cố, ta đặt câu hỏi: Xác suất của một biến cố thay đổi như thế nào, nếu một biến cố khác đã xảy ra? Khái niệm **xác suất có điều kiện** ra đời để trả lời câu hỏi này.

Định nghĩa 8.1 (Xác suất có điều kiện)**.** Cho một phép thử ngẫu nhiên và 2 biến cố A, B. Khi đó ta ký hiệu $\mathrm{P}(A|B)$ là **xác suất của** A **với điều kiện** B, hay còn gọi là xác suất của A, biết trước B. $\mathrm{P}(A|B)$ đo khả năng A xảy ra khi B *đã xảy ra*.

Công thức sau là một trong những công thức quan trọng nhất trong xác suất, và được đặt theo tên người đã tìm ra nó, Thomas Bayes.

Định lý 8.2 (Định lý Bayes)**.** *Cho một phép thử và 2 biến cố A, B, sao cho $P(B) > 0$. Khi đó:*

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

Nếu bỏ đi điều kiện $P(B) > 0$, ta có thể viết lại:

$$P(A|B)P(B) = P(B|A)P(A) = P(AB).$$

Nhận xét 8.3. Nếu A và B độc lập và có xác suất lớn hơn 0, công thức Bayes dẫn đến $\mathrm{P}(A|B) = \mathrm{P}(A)$ và $\mathrm{P}(B|A) = \mathrm{P}(B)$. Điều này mang ý nghĩa 2 biến cố độc lập thì việc B xảy ra không ảnh hưởng gì đến A và ngược lại.

8.2 Công thức xác suất đầy đủ

Trong thực tế, khi thực hiện các thống kê phức tạp trên các cộng đồng lớn, người ta chia cộng đồng đó thành nhiều nhóm nhỏ có cùng tính chất nào đó, và thực hiện thống kê trên từng nhóm nhỏ đó. Việc làm này vừa làm tăng tốc độ thống kê (vì công việc được chia nhỏ và thực hiện song song), vừa đảm bảo các mẫu lấy được có tính đa dạng và bao quát (mỗi nhóm đều có đại diện).

Để tổng hợp kết quả thống kê ở các nhóm nhỏ lại, ta cần một phương pháp để chuyển đổi nhóm các xác suất trên các không gian mẫu nhỏ thành một xác suất trên không gian mẫu lớn - là hợp của các không gian nhỏ đó. Định lý Bayes và cùng hệ biến cố đầy đủ cho phép ta làm điều này.

Định lý 8.4 (Định lý xác suất đầy đủ)**.** *Cho $B_1, B_2, \ldots, B_n$ là một hệ biến cố đầy đủ và A là một biến cố bất kì của một phép thử. Khi đó:*

$$P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \cdots + P(A|B_n)P(B_n).$$

Khi cần tính xác suất một người có tính chất A, giả sử cộng đồng được chia thành các nhóm nhỏ $B_1, B_2, \ldots, B_n$. Giả sử thông qua lấy mẫu trên các nhóm B_i, ta biết được các xác suất $\mathrm{P}(A|B_i)$. Khi đó, bằng việc ước lượng tỉ lệ thành viên của các nhóm B_i, từ đó suy $\mathrm{P}(B_i)$ khi chọn một người ngẫu nhiên, ta có thể tính được $\mathrm{P}(A)$.

Công thức Bayes còn được dùng nhiều trong các bài toán tính xác suất phức tạp, gồm nhiều phép thử lần lượt, với kết quả của phép thử sau phụ thuộc vào phép thử trước.

Một ví dụ nổi tiếng là bài toán Monty Hall. Bài toán được phát biểu như sau:

Ví dụ 8.5 (Bài toán Monty Hall)**.** Trong trò chơi Let's Make a Deal, người chơi sẽ phải chọn một trong 3 cánh cửa đóng và nhận phần thưởng đằng sau cánh cửa đó. Có 2 cửa có con dê và 1 cửa có chiếc Ferrari đằng sau.

Anh Trung chọn một cánh cửa, nhưng trước khi anh kịp mở, người dẫn chương trình Thế Anh đột nhiên mở một cánh cửa khác và chỉ cho anh Trung con dê sau đó. Thế Anh khuyên Trung chọn cánh cửa còn lại chưa được chọn. Hỏi anh Trung có nên đổi cửa?

Lời giải. Thoạt nhìn có thể giải như sau:

Không mất tính tổng quát giả sử Trung ban đầu chọn cửa số 1, Thế Anh mở cửa số 3 chứa dê và khuyên Trung chọn cửa số 2.

Gọi X_i, D_i lần lượt là các biến cố cửa thứ i có xe và có dê, với $i = 1, 2, 3$. Ta có:

$$\begin{aligned} \mathrm{P}(X_1) = \frac{1}{3}, \mathrm{P}(D_3) = \frac{2}{3}, \\ \mathrm{P}(D_3|X_1) = 1 \text{ (vì cửa 1 có xe thì 2 cửa còn lại chắc chắn có dê).} \end{aligned}$$

Khi đó:

$$\mathrm{P}(X_1|D_3) = \frac{\mathrm{P}(D_3|X_1)\mathrm{P}(X_1)}{\mathrm{P}(D_3)} = \frac{1 \cdot (1/3)}{2/3} = \frac{1}{2}.$$

Như vậy việc đổi cửa và không đổi cửa là như nhau!

Tuy nhiên, khi tính $\mathrm{P}(D_3|X_1)$ và $\mathrm{P}(D_3)$, ta đã bỏ qua một dữ kiện quan trọng: Thế Anh chỉ mở cửa số 3 vì Trung chọn cửa số 1, do đó thực ra việc mở cửa số 3 và cho Trung xem con dê không hẳn là ngẫu nhiên. Để giải quyết đúng, gọi O_i là biến cố Thế Anh mở cửa i. Khi đó:

$$\begin{aligned} \mathrm{P}(O_3|X_1) &= \frac{1}{2} \text{ vì Thế Anh có thể chọn 1 trong 2 cửa có dê để mở,} \\ \mathrm{P}(O_3|X_2) &= 1 \text{ vì chỉ còn mỗi cửa 3 chưa được chọn và có dê.} \end{aligned}$$

Hiển nhiên $\mathrm{P}(O_3|X_3) = 0$, khi đó:

$$\mathrm{P}(O_3) = \mathrm{P}(O_3|X_1)\mathrm{P}(X_1) + \mathrm{P}(O_3|X_2)\mathrm{P}(X_2) = \frac{1}{2} \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} = \frac{1}{2}.$$

Do vậy

$$\mathrm{P}(X_1|O_3) = \frac{\mathrm{P}(O_3|X_1)\mathrm{P}(X_1)}{\mathrm{P}(O_3)} = \frac{(1/2) \cdot (1/3)}{1/2} = \frac{1}{3}.$$

Như vậy, việc đổi cửa sẽ dẫn đến tỉ lệ thành công $\frac{2}{3}$!

$\square$

Người đầu tiên giải chính xác bài toán này là Marilyn vos Savant (1990). Có hàng chục nghìn người đã gửi thư phản đối bà khi đó, trong đó có nhiều nhà toán học nổi tiếng như Paul Erdos. Chỉ khi vos Savant chứng minh bằng cách liệt kê các trường hợp, lời giải của bà mới dần được chấp nhận. Ví dụ trên cho thấy các kết quả toán học đôi khi có thể đi ngược lại trực giác thông thường.

8.3 Xác suất tiền nghiệm, hậu nghiệm

Công thức Bayes là một trong những công thức quan trọng nhất trong XSTK vì nó giúp chúng ta tính được **mức độ đáng tin** của một giả thuyết khi quan sát được một số thông tin. Nhớ lại công thức Bayes:

$$\mathrm{P}(A \mid B) = \frac{\mathrm{P}(B \mid A)\mathrm{P}(A)}{\mathrm{P}(B)}.$$

Trong phần trước, ta chỉ xem A và B như 2 biến cố, vai trò của chúng khá giống nhau. Khi dùng công thức Bayes để kiểm tra giả thuyết, vai trò của chúng được tách biệt rõ ràng:

- ⋆ A là một **giả thuyết** về một thống kê nào đó, ví dụ như "Hơn 80% số trại sinh PiMA học giỏi toán".
- ⋆ B là một **kết quả quan sát thực tế** về thống kê đó, ví dụ như "Trong 5 trại sinh được chọn ngẫu nhiên, có 3 bạn giỏi toán".

Khi đó, ta có các khái niệm sau:

Định nghĩa 8.6 (Xác suất tiền nghiệm, hậu nghiệm)**.** Trong công thức Bayes:

$$\mathrm{P}(A \mid B) = \frac{\mathrm{P}(B \mid A)\mathrm{P}(A)}{\mathrm{P}(B)}.$$

Khi A là giả thuyết và B là kết quả quan sát được, ta có:

- ⋆ $\mathrm{P}(A)$ là **Xác suất tiền nghiệm** (**Prior probability**) của A, mang ý nghĩa là *niềm tin ban đầu* của người quan sát về khả năng A đúng.
- ⋆ $\mathrm{P}(A \mid B)$ là **Xác suất hậu nghiệm** (**Posterior probability**) của A, mang ý nghĩa là *niềm tin sau khi quan sát* về khẳ năng A đúng.

Như vậy, trong bài toán kiểm tra giả thuyết, ta bắt đầu với một niềm tin cơ bản về khả năng giả thuyết A là đúng, và sau khi có được thông tin từ các quan sát (có thể là kết quả của các phép thử), ta cập nhật niềm tin của mình về A. Nếu ban đầu ta cho rằng khả năng A xảy ra là cao nhưng các ví dụ quan sát được đều không ủng hộ A, niềm tin về A của ta sẽ giảm, điều này tương ứng với việc xác suất hậu nghiệm bé hơn xác suất tiền nghiệm nếu quan sát B chống lại A.

Ví dụ 8.7. Gọi A là giả thuyết "Hơn 80% số trại sinh PiMA học giỏi toán", tức là ít nhất 23 bạn giỏi toán trên 28 bạn. Trung ban đầu tin rằng khả năng A xảy ra là 90%. Sau khi Trung hỏi 5 bạn trại sinh ngẫu nhiên và đánh giá có 3 bạn giỏi toán, gọi biến cố này là B, tức là "Có không quá 3 bạn giỏi toán trong 5 bạn được chọn", ta có:

$$\mathrm{P}(B \mid A) \leq \frac{\sum_{i=0}^{3} C_{23}^{i} C_{5}^{5-i}}{C_{28}^{5}} \approx 0.207.$$

Nếu kết quả khảo sát các năm trước cho thấy $\mathrm{P}(B) \approx 0.4$, ta có:

$$\mathrm{P}(A \mid B) = \frac{\mathrm{P}(B \mid A)\mathrm{P}(A)}{\mathrm{P}(B)} \leq \frac{0.207 \cdot 0.9}{0.4} \approx 0.466$$

Như vậy sau khi quan sát, niềm tin của Trung đã giảm khoảng một nửa và thấp hơn 50%.

Tuy nhiên trong ví dụ trên, ta đã phải sử dụng việc khảo sát các năm trước để dự đoán $\mathrm{P}(B)$ cho trại năm nay. Thực chất ta có thể tính $\mathrm{P}(B)$ bằng Định lý xác suất đầy đủ như sau:

$$\mathrm{P}(B) = \sum_{i=0}^{28} \mathrm{P}(B|X=i)\mathrm{P}(X=i).$$

trong đó X là biến ngẫu nhiên chỉ số trại sinh giỏi toán, và A lúc này là biến cố $\{X \geq 23\}$.

Để tính chính xác $\mathrm{P}(B)$, $\mathrm{P}(A)$ là chưa đủ, ta cần $\mathrm{P}(X=k), \quad \forall k = 1, \ldots, 28$, tức là cần cả phân phối xác suất của X. Nhu cầu này dẫn đên khái niệm **phân phối tiền nghiệm** và **phân phối hậu nghiệm** sau đây.

8.4 Phân phối tiền nghiệm, hậu nghiệm

Bài toán đặt ra trong phần này chi tiết hơn bài toán kiểm tra giả thuyết ở phần trước, và giúp giải quyết được nhiều vấn đề hơn.

Cho biến ngẫu nhiên X và biến ngẫu nhiên Y phụ thuộc vào X. Ban đầu ta có một ước lượng sơ bộ về phân phối của X và từng phân phối của Y tương ứng với từng giá trị của X. Sau một phép thử, ta thu được giá trị $Y = y_0$. Khi đó ta có thể tính toán phân phối của X trong phép thử vừa rồi bằng công thức Bayes.

Định nghĩa 8.8 (Phân phối tiền nghiệm, hậu nghiệm)**.** Với 2 biến ngẫu nhiên X, Y như trên, nếu chúng là rời rạc, ta có công thức:

$$\mathrm{P}(X = x \mid Y = y) = \frac{\mathrm{P}(Y = y \mid X = x)\mathrm{P}(X = x)}{\mathrm{P}(Y = y)}, \quad \forall x, y$$

Nếu chúng là liên tục, ta có:

$$p_{X|Y}(x \mid y) = \frac{p_{Y|X}(y|x)p_X(x)}{p_Y(y)}, \quad \forall x, y,$$

trong đó các hàm $p_{X|Y}, p_{Y|X}, p_Y, p_X$ lần lượt là hàm mật độ xác suất biên của X (biết Y), Y biết X, hàm mật độ xác suất của Y và X. Khi đó:

- ⋆ Phân phối ứng với p_X là **Phân phối tiền nghiệm** (**Prior distribution**) của X, biểu thị niềm tin ban đầu về phân phối của X khi chưa có quan sát gì.

⋆ Phân phối ứng với $p_{X|Y}$ là **Phân phối hậu nghiệm** (**Posterior distribution**) của Y, biểu thị niềm tin về phân phối của X sau khi quan sát được Y.

Như đã nhận xét, có thể tính p_Y bằng tổng (rời rạc) hoặc tích phân (liên tục) của các tích $p_{Y|X}p_X$. Ta có:

Tính chất 8.9. *Với 2 biến ngẫu nhiên X, Y như trên, nếu chúng là rời rạc:*

$$P(X = x \mid Y = y) = \frac{P(Y = y \mid X = x)P(X = x)}{\sum_{x,y} P(Y = y \mid X = x)P(X = x)}, \quad \forall x, y.$$

Nếu chúng là liên tục:

$$p_{X|Y}(x \mid y) = \frac{p_{Y|X}(y|x)p_X(x)}{\int_x p_{Y|X}(y|x)p_X(x)dx} \quad \forall x, y$$

Ví dụ 8.10. Quay trở lại ví dụ về số trại sinh PiMA giỏi toán. Giả sử ta có thông tin về phân phối tiền nghiệm như sau: Anh Trung tin rằng xác suất để một bạn trại sinh bất kì giỏi toán là λ, với $\lambda \geq 0.8$, trong khi phân phối tiền nghiệm của λ là $\mathrm{U}(0, 1)$. Gọi số bạn giỏi toán trong trại là x, khi đó

$$\begin{aligned} p(\lambda \mid x) &= \frac{p(x \mid \lambda)p(\lambda)}{\int_0^1 p(x \mid \lambda)p(\lambda)} = \frac{C_{28}^x \lambda^x (1-\lambda)^{28-x}}{\int_0^1 C_{28}^x \lambda^x (1-\lambda)^{28-x} \mathrm{d}\lambda} \\ &= \frac{\lambda^x (1-\lambda)^{28-x}}{\int_0^1 \lambda^x (1-\lambda)^{28-x} \mathrm{d}\lambda} \end{aligned}$$

Hàm số ở mẫu số thực chất là hàm Beta[1] của x và $28 - x$, ta có:

$$\int_0^1 \lambda^x (1-\lambda)^{28-x} \mathrm{d}\lambda = \frac{1}{29C_{28}^x}$$

Do đó:

$$p(\lambda \mid x) = 29C_{28}^x \lambda^x (1-\lambda)^{28-x}$$

Giả sử thống kê cho thấy chỉ có 18 bạn giỏi toán, khi đó:

$$p(\lambda \mid x = 18) = 29C_{28}^{14} \lambda^{18} (1-\lambda)^{10}$$

Xác suất để $\lambda \geq 0.8$ là:

$$\mathrm{P}(\lambda \geq 0.8 \mid x = 18) = 29C_{28}^{14} \int_{0.8}^1 \lambda^{18} (1-\lambda)^{10} \approx 0.06$$

Như vậy chỉ có khoảng 6% khả năng niềm tin của anh Trung là đúng.

Bài tập

Bài tập 8.1. Có 3 hộp kín, một hộp có 2 quả bóng đen, một hộp có 2 quả trắng, hộp còn lại có 1 quả mỗi màu. Rút ngẫu nhiên một quả trong một hộp ra và được màu trắng. Tính xác suất quả còn lại trong hộp đó là màu trắng.

1 https://en.wikipedia.org/wiki/Betafunction

Bài tập 8.2. Một trung tâm chẩn đoán bệnh dùng một phép kiểm định T để kiểm tra một bệnh D nọ. Biết trong một quần thể 300 triệu người, có 1000 người mắc bệnh D. Phép kiểm định T có thể xác định chính xác 99% người mắc bệnh và 99.9% người không mắc bệnh. Một người đến trung tâm thực hiện phép kiểm định T, kết quả trả về dương tính. Xác suất để người đó thật sự mắc bệnh là bao nhiêu?

Bài tập 8.3. Hộp I có 10 linh kiện, trong đó có 3 cái bị hỏng. Hộp II có 15 linh kiện, trong đó có 4 cái bị hỏng. Lấy ngẫu nhiên từ mỗi hộp ra 1 linh kiện.

1. Tính xác suất cả 2 linh kiện lấy ra đều hỏng
2. Số linh kiện còn lại trong 2 hộp đem bỏ vào hộp III. Từ hộp III lấy ra ngẫu nhiên 1 linh kiện. Tính xác suất linh kiện này bị hỏng.
3. Biết linh kiện tự hộp III bị hỏng. Tính xác suất để 2 linh kiện lấy ra ban đầu từ hộp I và II bị hỏng.

9 Ước lượng phân phối xác suất từ thực nghiệm

Một trong những bài toán quan trọng nhất trong ngành XSTK được phát biểu đại khái như sau: Cho n dữ liệu $x_1, x_2, \ldots, x_n$ được sinh ra từ một phân phối xác suất nào đó. **Làm thế nào để tìm được phân phối xác suất đó từ những giá trị sinh ra này?**

Trên thực tế, việc khôi phục lại chính xác hàm mật độ của phân phối chưa biết là điều không thể. Tuy nhiên,cùng với một số giả thiết về dạng của hàm mật độ, ta có thể khôi phục lại một số thông tin của phân phối này như kỳ vọng, độ lệch chuẩn một cách tương đối chính xác.

Để bạn đọc hình dung rõ hơn, ta có thể bắt đầu với ví dụ sau:

Ví dụ 9.1. Cầu thủ M tập sút penalty và đã sút 100 quả, trong đó vào 25 quả. Ước lượng xác suất sút vào một quả bất kỳ của cầu thủ M.

Ở ví dụ trên, một ước lượng tự nhiên là lấy tổng số quả sút vào chia cho tổng số quá đã sút: $25/100 = 0.25$. Như vậy ở một quả bất kì thì cầu thủ M có xác suất sút vào khoảng 0.25. Thực ra đây là một ước lượng khá chính xác! Để tìm hiểu cơ sở toán học đằng sau các ước lượng này, người ta sử dụng 2 phương pháp chính:

1. Khi không có thông tin gì ngoài các mẫu thực nghiệm: **Ước lượng hợp lý cực đại** (**Maximum Likelihood Estimation**).
2. Khi có *thông tin tiền nghiệm* về phân phối xác suất: **Ước lượng cực đại hậu nghiệm** (**Maximum a Posteriori Estimation**).

9.1 Bài toán tìm phân phối xác suất

Trước hết ta bắt đầu với khái niệm *tham số* của một họ các phân phối xác suất cùng cấu trúc. Một số ví dụ cho các họ phân phối cùng cấu trúc là phân phối Bernoulli, phân phối nhị phân, phân phối chuẩn. Có thể thấy nếu chỉ biết một phân phối $\mathcal{D}$ nào đó thuộc 1 họ, ta chưa thể xác định toàn bộ $\mathcal{D}$, mà phải biết thêm một số hằng số ứng với $\mathcal{D}$ nữa. Các hằng số đặc trưng cho $\mathcal{D}$ chính là các **tham số** của $\mathcal{D}$. Ở trong một họ phân phối, mỗi bộ tham số sẽ cho ra một phân phối cụ thể tương ứng.

Ví dụ 9.2. Một số họ phân phối cùng cấu trúc và các tham số:

1. Họ các phân phối Bernoulli $\mathrm{B}(1, \lambda)$. Cần biết thêm xác suất thành công λ để hoàn toàn xác định một phân phối Bernoulli, do đó λ là tham số duy nhất của họ phân phối này.
2. Họ các phân phối nhị thức $\mathrm{B}(n, p)$. Cần biết thêm n và p để hoàn toàn xác định $\mathrm{B}(n, p)$, do đó họ này có 2 tham số là n và p.
3. Họ các phân phối chuẩn $\mathrm{N}(\mu, \sigma^2)$. Cần biết thêm μ và σ để hoàn toàn xác định $\mathrm{N}(\mu, \sigma^2)$, do đó họ này có 2 tham số là μ và σ.

Như vừa đề cập, **bài toán tìm phân phối xác suất** được phát biểu như sau:

Định nghĩa 9.3 (Bài toán tìm phân phối xác suất)**.** Cho một phân phối xác suất chưa biết $\mathcal{D}$, một biến ngẫu nhiên $X \sim \mathcal{D}$ và $x_1, x_2, \ldots, x_n$ là n giá trị thu thập được từ X. Hãy khôi phục $\mathcal{D}$ từ n giá trị này.

Trong thực tế, ta vẫn có quá ít thông tin để xác định $\mathcal{D}$ vì cấu trúc của nó không được cố định. Điều tốt nhất ta có thể làm là cố định cấu trúc của $\mathcal{D}$ và ước lượng các tham số trong cấu trúc đó để *khả năng $\mathcal{D}$ với các tham số này sinh ra được* $x_1, x_2, \ldots, x_n$ là cao nhất. Khả năng này được gọi là **độ hợp lý** (**Likelihood**) của $\mathcal{D}$, và phương pháp tìm các tham số sao cho độ hợp lý của $\mathcal{D}$ cao nhất được gọi là phương pháp **ước lượng hợp lý cực đại** (**maximum likelihood estimation - MLE**).

9.2 Ước lượng hợp lý cực đại

Định nghĩa 9.4 (Maximum Likelihood Estimation)**.** Cho một họ phân phối $\mathcal{D}(\theta)$ với tham số θ, một biến ngẫu nhiên $X \sim \mathcal{D}(\theta)$ và n giá trị $x_1, x_2, \ldots, x_n$ thu được từ X. Khi đó **Ước lượng hợp lý cực đại** của θ là giá trị của θ_0 sao cho xác suất

$$\mathrm{P}\left((X_1 = x_1)\,(X_2 = x_2)\cdots(X_n = x_n) \mid \theta = \theta_0\right).$$

là lớn nhất, trong đó $X_i \sim \mathcal{D}(\theta_0)$ là giá trị của X trong lần thử thứ i.

Biểu thức ở trên là xác suất để kết quả n lần thử ra được $x_1, x_2, \ldots, x_n$, và được gọi là **độ hợp lý** của $\mathcal{D}(\theta_0)$.

Để rút gọn ký hiệu, ta thường viết độ hợp lý theo dạng

$$\mathrm{P}(x_1, x_2, \ldots, x_n \mid \theta).$$

Nói cách khác, ước lượng hợp lý cực đại cho θ là giá trị sau:

$$\theta^* = \underset{\theta}{\operatorname{argmax}}\ \mathrm{P}(x_1, x_2, ..., x_n \mid \theta). \tag{1}$$

Với dữ liệu $x_1, x_2, \ldots, x_n$ cho trước, có thể thấy biểu thức $\mathrm{P}(x_1, x_2, \ldots, x_n \mid \theta)$ là một hàm số theo biến θ, do đó việc tìm MLE chính là tìm cực đại của hàm số. MLE có công dụng tính cho ta giá trị hoặc bộ giá trị khả thi nhất của tham số dựa vào những dữ liệu đã có đó.

Ở dạng xác suất kết hợp, ta thường chưa thể tìm cực đại của $\mathrm{P}(x_1, x_2, \ldots, x_n \mid \theta)$ một cách dễ dàng. Một ý tưởng đơn giản với giả thiết việc thu thập dữ liệu là độc lập giúp biến đổi $\mathrm{P}(x_1, x_2, \ldots, x_n \mid \theta)$ về dạng dễ xử lý hơn nhiều.

9.2.1 Giả thiết độc lập

Ta có thể giả thiết việc thu thập các giá trị $x_1, x_2, \ldots, x_n$ là độc lập - một giả thiết khá hợp lý trong thực tế.

Ví dụ 9.5. Giả sử ta cần dùng MLE để ước lượng tham số λ của phân phối Bernoulli $\mathrm{B}(1, \lambda)$ biết $x_1, x_2, \ldots, x_n \in \{0, 1\}$ là n giá trị lấy mẫu từ phân phối này. Khi đó mỗi lần lấy mẫu tương đương với việc tung 1 đồng xu có xác suất mặt ngửa λ, việc tung đồng xu 2 lần khác nhau hiển nhiên là độc lập với nhau, do đó giả thiết độc lập hợp lý trong trường hợp này.

Với giả thiết độc lập trên, các biến cố $(X_i = x_i)$ là độc lập, do đó ta có thể sử dụng công thức nhân như sau:

$$\mathrm{P}(x_1, ..., x_n \mid \theta) = \mathrm{P}(x_1 \mid \theta) \cdot \mathrm{P}(x_2 \mid \theta) \cdot \ldots \cdot \mathrm{P}(x_n \mid \theta) = \prod_{i=1}^{n} \mathrm{P}(x_i \mid \theta). \tag{2}$$

Với giả sử này, bài toán (1) có thể được giải quyết bằng phương trình sau:

$$\theta^* = \underset{\theta}{\operatorname{argmax}} \prod_{i=1}^{n} \mathrm{P}(x_i \mid \theta). \tag{3}$$

Việc tìm công thức cho $\mathrm{P}(x_i \mid \theta)$ thường khá đơn giản vì đó chính là hàm mật độ xác suất của $\mathcal{D}(\theta)$, từ đó có thể tìm được công thức cho $\mathrm{P}(x_1, x_2, \ldots, x_n \mid \theta)$ và tìm được θ^*.

Ví dụ 9.6. Cầu thủ M sút penalty vào m lần trên tổng số n lần, với $0 < m < n$. Ước lượng xác suất cầu thủ M sút penalty vào mỗi lần.

Lời giải. Đặt một biến ngẫu nhiên X sao cho $X = 1$ khi cầu thủ M sút vào và $X = 0$ nếu ngược lại. Khi đó $X \sim \mathrm{B}(1, \lambda)$ với λ là xác suất thành công mỗi lần sút, cũng là số cần ước lượng. Khi đó:

$$\mathrm{P}(0 \mid \lambda) = 1 - \lambda, \quad \mathrm{P}(1 \mid \lambda) = \lambda.$$

Với m lần sút vào $n - m$ lần hỏng, ta có

$$\mathrm{P}(x_1, x_2, \ldots, x_n \mid \lambda) = \prod_{i=1}^{n} \mathrm{P}(x_i \mid \lambda) = \lambda^m (1 - \lambda)^{n-m}.$$

Có thể chứng minh được rằng, giá trị $\lambda^* = \dfrac{m}{n}$ chính là giá trị tại cực đại của hàm số $\lambda^m (1 - \lambda)^{n-m}$. Do đó MLE của λ là $\dfrac{m}{n}$ - khá hợp với việc ước lượng theo trực giác. □

9.2.2 Log-likelihood

Thông thường, việc tối ưu hoá tích sẽ phức tạp hơn nhiều so với tổng. Ngoài ra, do xác suất nhỏ hơn 1, khi n đủ lớn thì vế phải (3) sẽ trở nên rất nhỏ, và có nguy cơ bị máy tính làm tròn thành 0, hoặc làm sai số tương đối cao. Hơn nữa, việc giữ nguyên dạng

$$\prod_{i=1}^{n} \mathrm{P}(x_i \mid \theta),$$

khiến việc tính đạo hàm theo θ để tìm cực đại rất khó. Việc sử dụng hàm log của hàm độ hợp lý giúp khắc phục 2 vấn đề trên. Ta đưa bài toán (3) thành bài toán Maximum Log-likelihood:

$$\theta^* = \underset{\theta}{\operatorname{argmax}} \prod_{i=1}^{n} \mathrm{P}(x_i \mid \theta) = \underset{\theta}{\operatorname{argmax}} \sum_{i=1}^{n} \log(\mathrm{P}(x_i \mid \theta)). \tag{4}$$

Ví dụ 9.7. Quay trở lại ví dụ 9.6, giả sử ta thấy cầu thủ M sút vào 600 quả trên tổng số 1200 quả. Để ước lượng xác suất sút vào λ, ta cần tìm cực đại của hàm số

$$f(\lambda) = \lambda^{600}(1-\lambda)^{600}.$$

Tuy nhiên, máy tính sẽ không thể tìm cực đại của hàm số trên vì lý do

$$\forall \lambda \in [0,1] : \lambda^{600}(1-\lambda)^{600} \le \frac{1}{4^{600}} \approx 0.0 \text{ trên đa số máy tính hiện nay.}$$

Nhưng nếu ta đi tìm cực đại của hàm $\log(f(\lambda))$, tình hình sẽ thay đổi:

$$l(\lambda) = \log(f(\lambda)) = 600(\log(\lambda) + \log(1-\lambda)).$$

Đa số máy tính hiện nay có thể tìm cực trị của hàm $l(\lambda)$ bằng phương pháp Newton (tương tự gradient descent). Đáp án chính là $\lambda^* = \frac{1}{2} = \frac{600}{1200}$.

9.3 Ước lượng cực đại hậu nghiệm

Trong thực tế, khi có quá ít điểm dữ liệu và việc thu thập không lý tưởng, MLE có thể gặp tình trạng **overfit**, tức là kết quả ước lượng không chính xác vì bị ảnh hưởng nặng bởi bộ dữ liệu không đủ tốt. Ta xem xét ví dụ sau:

Ví dụ 9.8. Khi đổ xúc xắc 7 lần liên tiếp, xác suất không đổ đượcc mặt số 1 là $(5/6)^7 \approx 0.27$. Như vậy hoàn toàn có thể mặt số 1 sẽ không xảy ra lần nào. Trên lý thuyết, xác suất đổ được số 1 là $1/6$, tuy nhiên MLE sẽ cho rằng xác suất xuất hiện số 1 trong một lần đổ là 0.

Có thể nói MLE đã bị ảnh hưởng với hiện tượng 7 lần liên tiếp không hiện số 1 và cho ra một ước lượng sai, đây gọi là **overfitting**. Một phần lý do là vì số lượng lần đổ chỉ là 7, không đủ để ước lượng xác suất ra các mặt.

Khi có quá ít dữ liệu, việc chỉ ước lượng dựa vào dữ liệu sẽ dễ dẫn đến overfit. Ta cần giả thiết một vài thông tin có sẵn về tham số θ, và sẽ ước lượng dựa và cả dữ liệu lẫn giả thiết có sẵn này để làm giảm sự phụ thuộc vào dữ liệu. **Ước lượng cực đại hậu nghiệm** (**Maximum a Posteriori - MAP**) giúp ta làm điều đó.

Định nghĩa 9.9 (Maximum a Posteriori Estimation)**.** Cho một họ phân phối $\mathcal{D}(\theta)$ với tham số θ, trong đó θ là một biến ngẫu nhiên có phân phối tiền nghiệm $\theta \sim \mathcal{P}$.

Cho một biến ngẫu nhiên $X \sim \mathcal{D}(\theta)$ và n giá trị $x_1, x_2, \ldots, x_n$ thu được từ X. Khi đó **Ước lượng cực đại hậu nghiệm** của θ là giá trị của θ_0 sao cho xác suất

$$\mathrm{P}\left(\theta = \theta_0 \mid (X_1 = x_1)(X_2 = x_2)\cdots(X_n = x_n)\right),$$

là lớn nhất, trong đó $X_i \sim \mathcal{D}(\theta_0)$ là giá trị của X trong lần thử thứ i.

Biểu thức ở trên là xác suất để kết quả n lần thử ra được $x_1, x_2, \ldots, x_n$, và được gọi là **độ hợp lý hậu nghiệm** của $\mathcal{D}(\theta_0)$.

Tương tự MLE, để rút gọn ký hiệu, ta thường viết độ hợp lý hậu nghiệm theo dạng

$$\mathrm{P}(\theta \mid x_1, x_2, \ldots, x_n).$$

Nói cách khác, ước lượng cực đại hậu nghiệm cho θ là giá trị sau:

$$\theta^* = \underset{\theta \sim \mathcal{P}}{\operatorname{argmax}} \mathrm{P}(\theta \mid x_1, x_2, \ldots, x_n). \tag{5}$$

Từ đó sử dụng công thức Bayes, ta có:

$$\begin{aligned} \theta^* &= \underset{\theta \sim \mathcal{P}}{\operatorname{argmax}} \mathrm{P}(\theta \mid x_1, x_2, \ldots, x_n) \\ &= \underset{\theta \sim \mathcal{P}}{\operatorname{argmax}} \frac{\mathrm{P}(x_1, x_2, \ldots, x_n \mid \theta)\mathrm{P}(\theta)}{\mathrm{P}(x_1, x_2, \ldots, x_n)} \\ &= \underset{\theta \sim \mathcal{P}}{\operatorname{argmax}} \mathrm{P}(x_1, x_2, \ldots, x_n \mid \theta)\mathrm{P}(\theta). \end{aligned} \tag{6}$$

trong đó $\mathrm{P}(\theta)$ được tính dựa vào phân phối tiền nghiệm $\mathcal{P}$, còn phân phối dữ liệu $\mathrm{P}(x_1, x_2, \ldots, x_n)$ tuy không biết trước, có thể được bỏ đi vì nó là như nhau đối với mọi giá trị của θ.

Có thể thấy, hàm cần tìm cực đại trong MAP chính là tích của hàm độ hợp lý và hàm mật độ xác suất tiền nghiệm tại θ. Sự có mặt của hàm mật độ tiền nghiệm $\mathrm{P}(\theta)$ mang ý nghĩa *độ hợp lý sẵn có* của θ: khi $\mathrm{P}(\theta_0)$ quá nhỏ, tức là khả năng $\theta = \theta_0$ *ban đầu* quá nhỏ, thì khả năng $\theta = \theta_0$ *sau khi thu thập dữ liệu* cũng nhỏ theo, giúp tránh việc chọn phải một giá trị θ_0 "không hợp lý" như MLE.

Ví dụ 9.10. Trong ví dụ 9.8, nếu ta giả thiết phân phối tiền nghiệm của λ, xác suất đổ được 1, là phân phối đều trong khoảng $[1/12, 3/12]$, nghĩa là xác suất $\lambda = 0$ là 0, MAP sẽ ước lượng θ dựa vào

$$\begin{aligned} \theta^* &= \underset{\theta \sim U\left(\frac{1}{12}, \frac{3}{12}\right)}{\operatorname{argmax}} \mathrm{P}(x_1, x_2, \ldots, x_7 \mid \theta)\mathrm{P}(\theta) \\ &= \underset{\theta \in \left[\frac{1}{12}, \frac{3}{12}\right]}{\operatorname{argmax}} \mathrm{P}(x_1 \mid \theta)\mathrm{P}(x_2 \mid \theta) \cdots \mathrm{P}(x_7 \mid \theta) \cdot 6 \\ &= \underset{\theta \in \left[\frac{1}{12}, \frac{3}{12}\right]}{\operatorname{argmax}} (1-\theta)^7 = \frac{1}{12}. \end{aligned} \tag{7}$$

Như vậy MAP sẽ ước lượng rằng con xúc xắc bị lệch, và xác suất cho mặt số 1 chỉ là 1/12, một kết quả có vẻ hợp lý hơn MLE rất nhiều.

9.3.1 Giả thiết độc lập và hàm log

Tương tự như MLE, ta có thể giả thiết một cách hợp lý là việc thu thập các điểm dữ liệu khác nhau là độc lập, và có thể sử dụng hàm log để tránh việc tích của quá nhiều phần tử bé hơn 1 hội tụ về 0 quá nhanh. Kết hợp lại, ta có

$$\begin{aligned} \theta^* &= \underset{\theta}{\operatorname{argmax}} \mathrm{P}(x_1, x_2, \ldots, x_7 \mid \theta)\mathrm{P}(\theta) \\ &= \underset{\theta}{\operatorname{argmax}} \mathrm{P}(x_1 \mid \theta)\mathrm{P}(x_2 \mid \theta) \cdots \mathrm{P}(x_n \mid \theta) \cdot \mathrm{P}(\theta) \\ &= \underset{\theta}{\operatorname{argmax}} \left[\sum_{i=1}^{n} \log\left(\mathrm{P}(x_i \mid \theta)\right) + \log\left(\mathrm{P}(\theta)\right)\right]. \end{aligned} \tag{8}$$

Ví dụ 9.11. Cho $x_1, x_2, \ldots, x_n$ là n giá trị thu được từ $\mathrm{N}(\mu, 1)$ với phân phối tiền nghiệm $\mu \sim \mathrm{N}(0, 1)$. Tìm ước lượng MAP của μ ứng với n điểm dữ liệu này.

Lời giải. Gọi ước lượng MAP là μ^*, ta có:

$$\begin{aligned}
\mu^* &= \underset{\theta}{\operatorname{argmax}} \left[\sum_{i=1}^{n} \log\left(\mathrm{P}(x_i \mid \mu)\right) + \log\left(\mathrm{P}(\mu)\right)\right] \\
&= \underset{\theta}{\operatorname{argmax}} \left[\sum_{i=1}^{n} \left(-\frac{(x_i-\mu)^2}{2}\right) - \frac{\mu^2}{2}\right] \\
&= \underset{\theta}{\operatorname{argmin}} \left[\sum_{i=1}^{n} \frac{(x_i-\mu)^2}{2} + \frac{\mu^2}{2}\right].
\end{aligned}$$

Biểu thức cuối cùng là một hàm bậc hai theo μ, ta có thể dễ dàng tìm được

$$\mu^* = \frac{x_1 + x_2 + \cdots + x_n}{n+1}$$

□

Nhận xét 9.12. Nếu bỏ qua phân phối tiền nghiệm $\mathbf{N}(0, 1)$ của μ, chỉ sử dụng MLE thì sẽ được kết quả

$$\mu^* = \frac{x_1 + x_2 + \cdots + x_n}{n}.$$

9.3.2 Chọn phân phối tiền nghiệm cho θ

Có thể thấy, việc chọn phân phối tiền nghiệm cho θ là rất quan trọng để MAP cho kết quả chính xác. Trong các bài tập lý thuyết về MAP, phân phối tiền nghiệm thường được cho trước, tuy nhiên trong thực tế không có một phương pháp chặt chẽ nào để tìm phân phối này. Trong thực nghiệm, một phân phối tiền nghiệm phổ biến là **phân phối beta**, có hàm mật độ như sau:

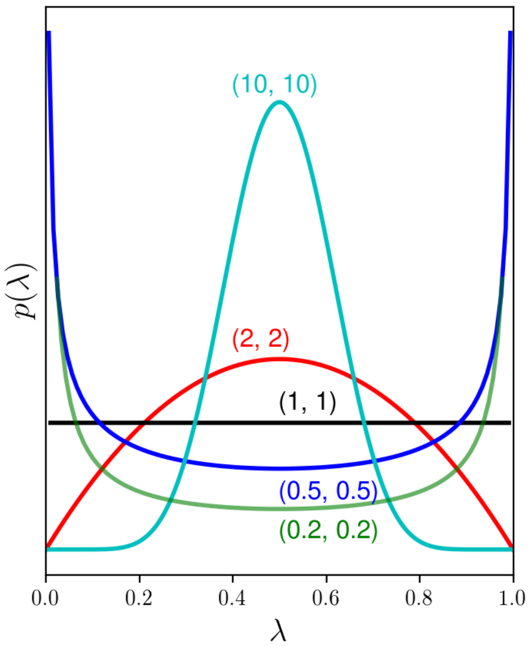
$$p(\theta) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta^{a-1}(1-\theta)^{b-1},$$

trong đó a và b là các tham số. Đối với thuật toán MAP, a và b đóng vai trò là các *siêu tham số* (các tham số không thể tối ưu hóa được, phải được cố định từ trước). Khi chưa có kiến thức gì về phân phối tiền nghiệm, người ta thường chọn $a = b = 1$.

9.4 Mô hình phân loại Naive Bayes

Naive Bayes là một mô hình Machine Learning dùng để phân loại dữ liệu bằng cách sử dụng **Maximum a Posteriori Estimation**. Naive Bayes có ưu điểm là lý thuyết đơn giản, trực quan và đưa ra dự đoán khá chính xác trong thực tế. Những ý tưởng của Naive Bayes cũng là nền tảng cho các mô hình ML phức tạp hơn sau này.

Trước tiên ta phát biểu bài toán phân loại như sau:



Hình 1: Đồ thị hàm mật độ xác suất của Beta distribution khi $\alpha = \beta$ và nhận các giá trị khác nhau. Khi cả hai giá trị này lớn, xác suất để λ gần 0.5 sẽ cao hơn.

Định nghĩa 9.13 (Bài toán phân loại)**.** Cho một **hàm mục tiêu** chưa biết $c : \mathbb{R}^d \to C$, trong đó $\mathbb{R}^d$ là không gian đầu vào và $C = \{c_1, c_2, \ldots, c_m\}$ là m loại (**class**), và n điểm dữ liệu huấn luyện (training data)

$$D = \{\langle \mathbf{x}_1, c(\mathbf{x}_1)\rangle, \{\langle \mathbf{x}_2, c(\mathbf{x}_2)\rangle, \ldots, \{\langle \mathbf{x}_n, c(\mathbf{x}_n)\rangle\}.$$

trong đó mỗi $\mathbf{x}_i \in \mathbb{R}^d$, $c(\mathbf{x}_i) \in C$.

Nói cách khác, ta biết được giá trị của c tại n điểm dữ liệu. Với một điểm dữ liệu $\mathbf{x}$ chưa có phân loại, hãy ước lượng một giá trị phân loại $c(\mathbf{x}) \in C$ phù hợp nhất.

Ý tưởng cơ bản của Naive Bayes chính là tìm một class $c \in C$ sao cho **khả năng x thuộc class** c **là lớn nhất**, biết phân loại của các điểm $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$. Rõ ràng **MAP** chính là giải pháp phù hợp để hiện thực hóa ý tưởng này. **Thuật toán Naive Bayes** sẽ sử dụng MAP để làm điều đó.

Cho bài toán phân loại với các class $C = \{c_1, c_2, \ldots, c_m\}$, các tập dữ liệu huấn luyện

$$D = \{\langle \mathbf{x}_1, c_1\rangle, \{\langle \mathbf{x}_2, c_2\rangle, \ldots, \{\langle \mathbf{x}_n, c_n\rangle\}.$$

Khi đó một điểm dữ liệu mới $\mathbf{x} = (x_1, x_2, \ldots, x_d)$ sẽ được phân loại theo công thức

$$\begin{aligned} c^*(\mathbf{x}) &= \underset{c \in C}{\operatorname{argmax}}\ \mathrm{P}\left(c(\mathbf{x}) = c \mid x_1, x_2, \ldots, x_d\right) \\ &= \underset{c \in C}{\operatorname{argmax}}\ \left[\mathrm{P}\left(x_1, x_2, \ldots, x_d \mid c\right) \mathrm{P}(c)\right]. \end{aligned}$$

Đến đây để biến đổi tiếp, ta sử dụng giả sử độc lập của các **thuộc tính** $x_1, x_2, \ldots, x_d$ (**naive Bayes assumption**). **Lưu ý**, các thuộc tính này thường ít khi độc lập trong thực tế, do đó một số phương pháp giảm tính phụ thuộc giữa các thuộc tính như **Principal Component Analysis** nên được dùng trước để biến đổi dữ liệu đầu vào.

Giả sử trường hợp lý tưởng, dữ liệu đã được xử lý để các thuộc tính tương đối độc lập, ta có:

$$\begin{aligned} c^*(\mathbf{x}) &= \operatorname*{argmax}_{c \in C} \left[\prod_{i=1}^{n} \mathrm{P}\left(x_i \mid c\right) \cdot \mathrm{P}(c) \right] \\ &= \operatorname*{argmax}_{c \in C} \left[\sum_{i=1}^{n} \log\left(\mathrm{P}\left(x_i \mid c\right)\right) + \log(\mathrm{P}(c)) \right]. \end{aligned}$$

Để tiếp tục việc phân loại, cần thiết phải ước lượng được các xác suất $\mathrm{P}(x_i \mid c)$ và $\mathrm{P}(c)$. Thực chất việc làm này chính là thiết lập các phân phối tiền nghiệm cho các thuộc tính x_i và ứng viên c. Đây là lúc ta sử dụng thông tin từ tập dữ liệu huấn luyện $D = \{\langle \mathbf{x}_i, c(\mathbf{x}_i) \rangle\}_{i=1}^{n}$.

Một phương pháp ước lượng trực quan là *thống kê* từ chính tập D. Cụ thể, để ước lượng $\mathrm{P}(c)$, ta có thể dùng tỉ lệ các điểm $\mathbf{x}_k$ thuộc class c trong D.

$$\mathrm{P}(c) \approx \frac{|\{1 \le k \le n : c\left(\mathbf{x}_k\right) = c\}|}{n}.$$

Để ước lượng $\mathrm{P}(x_i \mid c)$, có thể dùng tỉ lệ các điểm $\mathbf{x}_k$ có thuộc tính i gần bằng x_i trong các điểm thuộc class c:

$$\mathrm{P}(x_i \mid c) \approx \frac{|\{1 \le k \le n : (c\left(\mathbf{x}_k\right) = c) \wedge (\mathbf{x}_k(i) \approx x_i)\}|}{|\{1 \le k \le n : c\left(\mathbf{x}_k\right) = c\}|}.$$

Trên đây chỉ là một trong những ý tưởng đơn giản nhất để ước lượng $\mathrm{P}(x_i \mid c)$ và $\mathrm{P}(c)$. Chúng có nhiều nhược điểm cần khắc phục để trở nên hữu dụng trong thực tế, sẽ được làm rõ trong project về Naive Bayes.

Bây giờ giả sử ta đã có được một phương pháp tốt để ước lượng các phân phối tiền nghiệm cho x_i và c, và các thuộc tính x_i tương đối độc lập, ta có thể xây dựng quy trình huấn luyện và phân loại của Naive Bayes.

Thuật toán Naive Bayes

Huấn luyện: Với mỗi c_j, $j = 1, 2, \ldots, m$:

1. Ước lượng $\mathrm{P}(c_j)$ từ D.
2. Với mỗi giá trị thuộc tính x_i, $i = 1, 2, \ldots, d$:
   a) Ước lượng $\mathrm{P}\left(x_i \mid c_j\right)$ từ D.

Phân loại: Tìm class c^* thỏa mãn

$$c^* = \operatorname*{argmax}_{c \in C} \left[\sum_{i=1}^{n} \log\left(\mathrm{P}\left(x_i \mid c\right)\right) + \log(\mathrm{P}(c)) \right],$$

và gán class c^* cho điểm dữ liệu $\mathbf{x} = (x_1, x_2, \ldots, x_d)$.

Phụ lục

Bảng 1 có các giá trị của hàm phân phối xác suất Φ của phân phối chuẩn tắc. Để biết giá trị $\Phi(x)$ với $x \in [0, 3.909]$, làm theo các bước:

1. Tìm hàng tương ứng với chữ số hàng đơn vị và chữ số đầu tiên sau dấu thập phân của x.
2. Tìm cột tương ứng với chữ số thứ 2 sau dấu thập phân của x.
3. Ô tương ứng với hàng và cột vừa tìm được chính là giá trị $\Phi(x)$.

Ví dụ 9.14. Tìm giá trị $\Phi(1.54)$.

Đầu tiên, tìm hàng ứng với số 1.5, sau đó tìm cột ứng với số 0.04. Giá trị trong ô là 0.9382, vậy $\Phi(1.54) = 0.9382$.

Để tính $\Phi(x)$ với $x < 0$, nhớ lại công thức

$$\Phi(x) + \Phi(-x) = 1.$$

Như vậy chỉ cần dùng bảng để tra $\Phi(x)$, sau đó tính $\Phi(-x) = 1 - \Phi(x)$.

Ví dụ 9.15. Tìm giá trị $\Phi(-2.45)$.

Đầu tiên, tính $\Phi(2.45)$ bằng cách tìm hàng 2.4, cột 0.05 để được giá trị 0.9929. Do vậy $\Phi(-2.45) = 1 - 0.9929 = 0.0071$. .

Tài liệu

[1] Jay L. Devore. *Probability and Statistics for Engineering and the Sciences*. Cengage Learning. 1982.

[2] Nguyễn Văn Thìn. *Bài giảng môn Xác suất Thống kê*. Trường Đại học Khoa học Tự nhiên, ĐHQG-HCM. 2017.

x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7703	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Bảng 1: Bảng giá trị hàm phân phối xác suất của phân phối chuẩn tắc