§2 CÁC ĐỊNH NGHĨA XÁC SUẤT

2.1 Định nghĩa cổ điển

a/ <u>ĐN 1</u>. Giả sử một phép thử có n phép thử *đồng khả năng*, trong đó có m phép thử thuận lợi cho A

$$\rightarrow \quad P(A) = \frac{\text{Số kết cục thuận lợi cho A}}{\text{Tổng số kết cục}} = \frac{m}{n}. \qquad (1)$$

* *Ưu điểm*: đơn giản và trực quan

* *Nhược điểm*: số kết cục hữu hạn và chúng đồng khả năng.

+ <u>Thí dụ 1</u>. Hộp bi có 3 bi đỏ và 5 trắng. lấy ngẫu nhiên ra 3 bi. Tính XS:

a) trong 3 bi có đúng 1 trắng;

b) 3 bi cùng màu.

Giải:

a) $n = C_8^3 = 56$, $m = C_3^2.5 = 15 \quad \rightarrow \quad$ XS $= 15/56$.

b) $n = 56$, $m = C_3^3 + C_5^3 = 9 \quad \rightarrow \quad$ XS $= 9/56$.

+ **<u>Thí dụ 2</u>**. Tính XS khi gieo 4 đồng tiền xuất hiện đúng 2 sấp.

Giải: $n = 16\ (= 2^4)$, $m = C_4^2 = 6 \quad \rightarrow \quad$ XS $= 3/8$.

b/ <u>Tính chất</u> (1) $\quad 0 \leq P(A) \leq 1$;

(2) $\quad P(V) = 0$; $P(U) = 1$;

(3) Nếu A, B xung khắc thì $P(A+B) = \mathrm{P}(A) + P(B)$;

(4) $P(\bar{A}) = 1 - P(A)$.

+ **<u>Thí dụ 3</u>**. Một nhóm có 6 cặp nam nữ, trong đó có 3 cặp là vợ chồng. Chọn ngẫu nhiên ra 3 người. Tính XS:

a) trong 3 người có ít nhất 1 nữ;

b) trong 3 người không có cặp vợ chồng nào.

Giải:

a) Gọi A - trong 3 người có ít nhất 1 nữ $\rightarrow$ $P(A) = 1 - P(\bar{A})$

$$= 1 - C_6^3 \,/\, C_{12}^3 = 1 - 1/11 = 5/11.$$

b) Gọi A - trong 3 người không có cặp vợ chồng nào

$$P(A) = 1 - P(\bar{A}) = 1 - 3.10/\, C_{12}^3 = 1 - 3/22 = 19/22.$$

+ **<u>Thí dụ 4</u>**. Cậu bé có 6 bi đỏ và 4 trắng. Một hôm cậu thấy mất 1 bi (mà không biết màu). Tính XS để lấy ngẫu nhiên ra 1 bi thì đó là bi đỏ.

Giải: $n = 9.10$; $m = 6.5 + 4.6$ $\rightarrow$ XS = 0,6

(chú ý thông tin không đổi thì XS không đổi!).

Để khắc phục hạn chế của ĐN1 về sự hữu hạn kết cục, người ta đưa ra *định nghĩa hình học* của xác suất: Giả sử tập vô hạn các kết cục đồng khả năng của phép thử có thể biểu thị bởi một miền hình học G, còn tập các kết cục thuận lợi cho A bởi một miền con S nào đó, S ⊆ G. Khi đó

b/ <u>ĐN 2</u>. $$P(A) = \frac{\text{Độ đo } S}{\text{Độ đo } G}\,. \qquad (2)$$

+ **<u>Thí dụ 5</u>**. Đường dây điện thoại ngầm nối tổng đài với một trạm dài 1 km. Tính XS để dây bị đứt tại nơi cách tổng đài không quá 100 m.

Giải: Có thể coi nơi đứt dây là đồng khả năng ở một điểm bất kỳ trên toàn chiều dài của dây. Từ đó theo (2)

$$P(A) = 100/1000 = 0{,}1$$

(chú ý sự kiện có XS = 0 vẫn có thể xảy ra!)

+ **<u>Thí dụ 6</u>**. (*coi như bài tập*) Hai người hẹn gặp nhau ở quán từ 8 giờ đến 9 giờ với quy ước ai đến trước sẽ đợi người đến sau 15 phút, sau đó bỏ đi. Tính XS hai người đó gặp được nhau, biết rằng thời điểm đến quán của mỗi người là đồng khả năng trong khoảng thời gian nói trên.

2.1 Định nghĩa thống kê (theo tần suất)

Điều kiện ĐN cổ điển của XS rất khó được đảm bảo, vì vậy tính khả thi của nó rất hạn chế.

Người ta tiến hành loạt n phép thử cùng loại và nếu sự kiện A nào đó xuất hiện trong m phép thử thì tỷ số m/n được gọi là *tần suất* xuất hiện A. Trên cơ sở quan sát khi n đủ lớn người ta thấy tần suất này rất ổn định, thay đổi ít và dao động xung quanh một hằng số nào đó. Từ đó có thể cho rằng hằng số đó là một xấp xỉ tốt cho XS khi số phép thử tăng cao. Nhưng do hằng số đó chưa biết, nên trong thực tế người ta lấy ngay tần suất đó làm XS xuất hiện sự kiện. Cách tiếp cận như vậy được gọi là *định nghĩa thống kê* của XS (về mặt toán học có thể hiểu như là giới hạn của tần suất khi số phép thử tăng vô hạn) và có tính thực tiễn cao.

Có thể tham khảo hai trường hợp cổ điển sau đây. Đầu tiên là tần suất xuất hiện mặt sấp khi gieo một đồng tiền nhiều lần

Người thí nghiệm	*Số lần gieo*	*Số lần sấp*	*Tần suất*
Buýt-phông	4040	2048	0,5080
Piếc-xơn	12000	6019	0,5016
Piếc-xơn	24000	12012	0,5005

Một thí dụ khác khi xác định XS phân rã của một nguyên tử chất phóng xạ *Ra* sau 100 năm là 0,04184 (độ chính xác tới 5 chữ số sau dấu phẩy), số lượng nguyên tử tham gia thí nghiệm rất lớn (cỡ 10^{23} – 10^{24}).

2.3 Định nghĩa tiên đề

Xét không gian các sự kiện sơ cấp Ω và xác định hệ thống A các tập con của Ω, các phần tử của A được coi là các sự kiện ngẫu nhiên. Yêu cầu A là *đại số Bull*

(i) A chứa Ω;

(ii) Nếu A và $B \in$ A, thì $\bar{A}, \bar{B}, A+B, AB \in$ A.

Khi có thêm điều kiện

(iii) Nếu $A_1, A_2, \ldots, A_n, \ldots \in$ A, thì $A_1+A_2+\ldots+A_n+\ldots$ và $A_1A_2\ldots A_n\ldots$ cũng $\in$ A,

thì ta có một *trường Borell hay* $\sigma -$ *đại số*.

<u>**Định nghĩa**</u>. Ta gọi xác suất trên (Ω,A) là một hàm số xác định trên A có giá trị trong [0, 1] và thoả mãn 3 tiên đề:

(T_1) $P(\Omega) = 1$;

(T_2) $P(A+B) = P(A) + P(B)$, A, B xung khắc;

(T_3) Nếu dãy $\{A_n\}$ có tính chất $A_j \Rightarrow A_i$, $\forall\ i \le j$ và

$$A_1A_2\ldots A_n\ldots = V, \text{ thì } P(A_n) \to 0 \text{ khi } n \to \infty.$$

Tuy nhiên cách chọn XS theo định nghĩa trên không duy nhất!

Có thể thay thế tiên đề 2 và 3 bằng *tiên đề công mở rộng*

(T_+) Nếu dãy $\{A_n\}$ có tính chất xung khắc từng đôi và

$A = \sum_{n=1}^{\infty} A_n \in$ A, thì

$P(A) = P(A_1) + P(A_2) + \ldots + P(A_n) + \ldots = \sum_{n=1}^{\infty} P(A_n).$

- $P(.)$ là độ đo không âm, trực chuẩn, cộng tính.
- Tổ hợp {Ω, A, P} thường được gọi là *không gian xác suất*.

TÀI LIỆU THAM KHẢO

1. Barnes J. W. *Statistical analysis for engineers and scientists.* McGraw – Hill, 2008.
2. Monfort A. *Cours de statistique mathématique*. Economica, Paris, 1982.
3. Tống Đình Quỳ. *Giáo trình xác suất thống kê*. NXB Bách khoa Hà Nội, 2018 (*tái bản*).

BÀI TẬP

1. Một lô hàng có 18 sản phẩm, trong đó có 12 sản phẩm loại A, 5 loại B và 1 loại C. Chọn ngẫu nhiên ra 3 sản phẩm. Tính XS để trong 3 sản phẩm đó:
   a) có 2 loại A và 1 loại B;
   b) có ít nhất 1 sản phẩm loại A;
   c) có cả 3 loại sản phẩm.

2. Một khách sạn có 5 phòng đơn, trong khi đó có 8 khách đến thuê phòng (5 nam, 3 nữ). Chọn ngẫu nhiên 5 người, tính xác suất để trong đó:
   a) có 3 nam, 2 nữ;
   b) cả 5 người cùng giới;
   c) có ít nhất 1 nữ.

3. Một công ty cần tuyển 2 nhân viên. Có 6 người đến thi tuyển, trong đó có 4 nữ và 2 nam (giả sử khả năng được tuyển của mỗi người là như nhau).
   a) Trong 4 nữ có một cô tên Hoa, tính xác suất để Hoa được chọn.
   b) Tính xác suất để có 2 nữ đươc chọn.

4. Đội A và đội B đều có 3 người cùng tham gia một cuộc thi chạy, giả sử khả năng mỗi người là như nhau. Tính xác suất để 3 người đội A về nhất, nhì, ba.

5. Trong tuần ở một thành phố có 7 tai nạn giao thông. Tính xác suất để mỗi ngày trong tuần có đúng 1 tai nạn.

6. Ở một vùng có 40% thích xem bóng đá. Chọn ngẫu nhiên 12 người. Tính xác suất để trong đó có đúng 5 người thích xem bóng đá.

7. Trong ga có một đoàn tầu 4 toa. 4 hành khách lên tầu, mỗi người chọn ngẫu nhiên một toa. Tính xác suất trong đó:
   a) trong 1 toa có 3 người, 1 toa có 1 người;
   b) 1 toa có ít nhất 2 người;
   c) toa thứ nhất có đúng 3 người lên.

8. Chọn ngẫu nhiên một vé số có 5 chữ số. Tính xác suất:
   a) trong số của vé không có chữ số 1 và không có chữ số 5;
   b) trong số vé có chữ số 5 và chữ số chẵn.

9. Xếp ngẫu nhiên 5 người ngồi quanh một chiếc bàn tròn 5 ghế.
   a) Tính xác suất để 2 người định trước được ngồi cạnh nhau.
   b) Tính như vậy cho trường hợp bàn dài.

10. Một người bỏ ngẫu nhiên 3 lá thư vào 3 chiếc phong bì đã ghi địa chỉ. Tính xác suất để ít nhất có một lá thư bỏ đúng phong bì của nó.

11. Ba nữ nhân viên phục vụ A, B và C thay nhau rửa đĩa chén và giả sử ba người "khéo léo" như nhau. Trong một tháng có 4 chén bị vỡ. Tính xác suất để:
    a) A đánh vỡ 3 chén và B đánh vỡ 1 chén;
    b) một trong 3 người đánh vỡ 3 chén;
    c) một trong 3 người đánh vỡ cả 4 chén.

12. Xếp ngẫu nhiên 8 quyển sách vào 2 ngăn kéo. Tính xác suất:
    a) ngăn kéo nào cũng có sách;
    b) ngăn kéo thứ nhất có 2 quyển sách, còn ngăn thứ hai có 6 quyển.

§3 CỘNG VÀ NHÂN XÁC SUẤT

3.1 Xác suất có điều kiện

a/ Khái niệm. Thực ra mọi XS $P(A)$ đều là XS có điều kiện (phép thử). Tuy nhiên nếu có thêm điều kiện khác, chẳng hạn B, thì ta có khái niệm mới: *xác suất của A với điều kiện B,* ký hiệu là $P(A|B)$ và gọi chung là *xác suất có điều kiện*. Bằng trực giác dễ thấy $P(A|B)$ tỷ lệ với $P(AB)$, tức là $P(A|B) = k\,P(AB)$, k là một hằng số > 0. Nếu đặt $A = B$, ta có $P(B|B) = 1$ và $k\,P(A|B)$ =1. Từ đó

* **ĐN 1**. Giả sử trong một phép thử ta có $P(B) > 0$, khi đó

$$P(A|B) = \frac{P(AB)}{P(B)}. \qquad (1)$$

Để ý rằng nói chung $P(A|B) \neq P(A)$ và XS có điều kiện có các tính chất như các XS không điều kiện. Ta cũng có thể tính $P(A|B)$ bằng cách dùng ĐN cổ điển trong bộ điều kiện mới.

Thí dụ 1. Gieo một con xúc xắc. Ký hiệu A – xuất hiện mặt lục, B – xuất hiện mặt chẵn, tính $P(A|B)$.

Giải: Do có điều kiện B, ta tưởng tượng gieo con xúc xắc chỉ có 3 mặt đồng khả năng (mặt có số chấm chẵn). Từ đó dùng ĐN cổ điển ở tiết trước ta có $P(A|B) = 1/3$. Mặt khác dễ thấy $P(A) = 1/6$, $P(B) = 1/2$ và do $AB = A$, từ đó theo ĐN (1) ở trên

$$P(A|B) = \frac{P(A)}{P(B)} = 1/3; \qquad P(B|A) = \frac{P(A)}{P(A)} = 1.$$

<u>**Thí dụ 2**</u>. Rút lần lượt 2 con bài từ bộ bài tú lơ khơ 52 con. Tính XS con bài thứ hai là át, biết rằng con thứ nhất cũng là át.

Giải: Nếu ký hiệu A_k – sự kiện con bài thứ k là át, $k = 1;2$, dễ dàng tính được $P(A_2|A_1) = 3/51 = 1/17$, tương đương với sự kiện nếu biết A_1, việc tính XS có điều kiện đưa về tính trường hợp chỉ còn 51 con bài với 3 con át trong đó.

* <u>**ĐN 2**</u>. Ta nói rằng A và B *độc lập* (*độc lập thống kê*), nếu

$$P(A|B) = P(A) \text{ hoặc } P(B|A) = P(B). \qquad (2)$$

Như vậy nếu A, B độc lập thì việc có sự kiện này không làm thay đổi XS của sự kiện kia. Tuy nhiên việc tính các XS trong (2) thực tiễn rất khó, thậm chí là không thể, vì vậy trên thực tế ta phải thừa nhận nhiều sự kiện độc lập trong các bài tập sau này.

Chú ý:

- Độc lập là khái niệm tương hỗ.

- Kết quả bắn của hai xạ thủ được coi là độc lập, thậm chí kết quả bắn của hai lần bắn khác nhau của cùng một xạ thủ cũng được coi là độc lập v.v…
- Nếu cặp (A, B) gồm 2 sự kiện độc lập thì ta có 3 cặp các sự kiện độc lập là $(A, \bar{B})$; $(B, \bar{A})$ và $(\bar{A}, \bar{B})$.

Biểu thức tương đương với (2), có để ý đến (1) là, nếu A, B độc lập

$$P(AB) = P(A)P(B). \tag{3}$$

* **<u>ĐN 3</u>**. Ta nói bộ sự kiện $A_1, A_2, \ldots, A_n$ *độc lập trong tổng thể* nếu

$$P(A_{i_1}, A_{i_2}, \ldots, A_{i_k}) = P(A_{i_1})P(A_{i_2}) \ldots P(A_{i_k}) \tag{4}$$

với mọi dãy $\{i_1, i_2, \ldots, i_k\}$ gồm các số nguyên khác nhau lấy từ tập $\{1, 2, \ldots, n\}$ và $k = 2, 3, \ldots, n$.

<u>Thí dụ 3</u>. Gieo 2 lần một đồng tiền và ta có 4 kết cục đồng khả năng $\Omega = \{SS, SN, NS, NN\}$.

Các sự kiện $A = SS + SN$, $B = SS + NS$, $C = SS + NN$ độc lập từng đôi vì

$P(A) = P(B) = P(C) = 1/2$, còn $P(AB) = P(BC) = P(AC) = 1/4$

thoả mãn (3). Tuy nhiên chúng không độc lập tổng thể do (4) không được thoả mãn

$$P(ABC) = 1/4 \neq P(A)P(B)P(C) = 1/8.$$

Rõ ràng tính độc lập tổng thể kéo theo độc lập từng đôi (do (3) là trường hợp riêng của (4) với $k = 2$), nhưng ngược lại nói chung không đúng.

3.2 Công thức cộng và nhân xác suất

1/ **Công thức nhân xác suất**

$$P(AB) = P(A)P(B|A) = P(B)P(A|B). \tag{5}$$

Kết quả này được suy ra trực tiếp từ (1). Còn từ (5) có thể dẫn ra các hệ quả quan trọng:

(i) Nếu A, B độc lập thì $P(AB) = P(A)P(B)$ (xem (3))

(ii) Mở rộng cho tích n sự kiện

$$P(A_1A_2\ldots A_n) = P(A_1)P(A_2|A_1)\ldots P(A_n|A_1A_2\ldots A_{n-1}). \tag{6}$$

Trường hợp $n = 3$

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

(iii) Nếu $A_1, A_2, \ldots, A_n$ độc lập trong tổng thể, thì

$$P\left(\prod_{i=1}^{n} A_i\right) = \prod_{i=1}^{n} P(A_i).$$

2/ **<u>Công thức cộng xác suất</u>**

$$P(A+B) = P(A) + P(B) - P(AB). \qquad (7)$$

Công thức (6) có thể được minh hoạ dễ dàng bằng sơ đồ Venn và là cơ sở để dẫn ra các hệ quả sau:

(i) Nếu A, B xung khắc thì $P(A+B) = P(A) + P(B)$.

(ii) Mở rộng cho tổng n sự kiện

$$P\left(\sum_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) - \sum_{i<j} P(A_i A_j)$$

$$+ \sum_{i<j<k} P(A_i A_j A_k) - \ldots + (-1)^{n-1}\, P(A_1 A_2 \ldots A_n). \qquad (8)$$

Trường hợp $n = 3$

$$P(A+B+C) = P(A) + P(B) + P(C)$$

$$- P(AB) - P(AC) - P(BC) + P(ABC).$$

(iii) Nếu A_1, A_2, ..., A_n xung khắc từng đôi, thì

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

Các công thức (5)-(8) cho ta công cụ hiệu quả để tính XS các sự kiện phức tạp qua XS các sự kiện đơn giản hơn.

<u>Thí dụ 4</u>. Hai cọc bài được lấy từ một bộ bài tú lơ khơ, cọc bài thứ nhất gồm 4 con át, cọc thứ hai gồm 4 con ka. Rút ngẫu nhiên từ mỗi cọc bài ra một con bài, tính các xác suất:

a) cả 2 con là cơ;

b) có ít nhất 1 con cơ trong 2 con bài.

Cũng các câu hỏi như vậy nhưng trong điều kiện khác: trộn 2 cọc bài thành một và từ đó rút ra 2 con bài.

Giải: Đặt A – con bài thứ nhất là cơ, B – con bài thứ hai là cơ.

Trong trường hợp hai cọc bài riêng rẽ, dễ thấy A, B độc lập và:

a) XS cần tìm là $P(AB)$, dùng (3) ta có

$$P(AB) = P(A)P(B) = \frac{1}{4} \cdot \frac{1}{4} = \frac{1}{16} .$$

b) Sự kiện ta quan tâm là $P(A + B)$, theo (7)

$$P(A + B) = P(A) + P(B) - P(AB) = \frac{1}{4} + \frac{1}{4} - \frac{1}{16} = \frac{7}{16} .$$

Trong trường hợp hai cọc bài trộn thành một, ta vẫn dùng các sự kiện A, B; tuy nhiên chúng không còn độc lập, mặc dù XS $P(A) = P(B) = 2/8 = 1/4$ do vai trò hai con bài ngang nhau. Từ đó:

a) Dùng công thức (5)

$$P(AB) = P(A)P(B|A) = \frac{1}{4} \cdot \frac{1}{7} = \frac{1}{28} .$$

b) Một lần nữa theo (7)

$$P(A+B) = P(A) + P(B) - P(AB) = \frac{1}{4} + \frac{1}{4} - \frac{1}{28} = \frac{13}{28}.$$

Tất nhiên trong trường hợp này ta có thể dùng định nghĩa cổ điển để tính các XS tương ứng.

<u>Thí dụ 5</u>. Ba xe ô tô của một công ty có XS sự cố trong tháng tương ứng là 0,05; 0,02 và 0,1. Tính các XS:

a) có đúng 2 xe bị sự cố trong tháng;

b) có ít nhất 1 xe không bị sự cố trong tháng.

Giải: Đặt A_k là sự kiện xe thứ k bị sự cố trong tháng ($k = 1, 2, 3$) và $P(A_1) = 0{,}05$; $P(A_2) = 0{,}02$; $P(A_3) = 0{,}1$.

a) Nếu gọi A là sự kiện có đúng 2 xe bị sự cố trong tháng

$$A = A_1A_2\bar{A}_3 + A_1\bar{A}_2A_3 + \bar{A}_1A_2A_3.$$

Dùng tính xung khắc của ba số hạng và tính độc lập của các thừa số trong các số hạng tích, ta có

$$P(A) = P(A_1)P(A_2)P(\bar{A}_3) + P(A_1)P(\bar{A}_2)P(A_3) + P(\bar{A}_1)P(A_2)P(A_3)$$

$$= 0{,}05.0{,}02.(1-0{,}1) + 0{,}05.(1-0{,}02).0{,}1 + (1-0{,}05).0{,}02.0{,}1$$

$$= 0{,}0077.$$

b) Nếu gọi B - sự kiện có ít nhất 1 xe không bị sự cố trong tháng

$$P(B) = 1 - P(\bar{B}) = 1 - P(A_1A_2A_3)$$

$$= 1 - 0{,}05.0{,}02.0{,}1 = 0{,}9999.$$

<u>Thí dụ 6</u>. Trong thời gian có dịch ở một vùng cứ 100 người mắc dịch thì có 10 người phải cấp cứu. XS gặp một người bị cấp cứu vì mắc dịch ở vùng đó là 0,06. Tìm tỷ lệ mắc bệnh dịch của vùng đó.

Giải: Đặt A – gặp người mắc dịch, B – gặp người bị cấp cứu và ta phải tìm $P(A)$. Từ đầu bài ta có $P(B|A) = 10/100 = 0{,}1$; $P(AB) = 0{,}06$. Theo (1) $P(B|A) = P(AB)/P(A)$, từ đó suy ra

$$P(A) = P(AB)/P(B|A) = 0{,}06/0{,}1 = 0{,}6.$$

<u>Thí dụ 7</u>. XS trúng đích của mỗi lần bắn là 0,4. Hỏi cần bắn bao nhiêu phát để XS có ít nhất một viên trúng sẽ lớn hơn 0,95?

Giải: Gọi số lần bắn thoả mãn yêu cầu của bài toán là n. Đặt A là sự kiện có ít nhất 1 lần bắn trúng và dễ thấy

$$P(A) = 1 - P(\bar{A}) = 1 - (1 - 0{,}4)^n.$$

Theo yêu cầu đầu bài $P(A) > 0{,}95$, suy ra

$$1 - 0{,}6^n > 0{,}95 \quad \Rightarrow \quad 0{,}6^n < 0{,}05 \ \Rightarrow \ n > \frac{ln 0{,}05}{ln 0{,}6} \approx 5{,}8647$$

$$\Rightarrow \quad n > 6.$$

<u>Thí dụ 8</u>. Theo thống kê xác suất để hai ngày liên tiếp có mưa ở một thành phố vào mùa hè là 0,5; còn không mưa là 0,3. Biết các sự kiện có một ngày mưa, một ngày không mưa là đồng khả năng. Tính XS để ngày thứ hai có mưa, biết rằng ngày đầu không mưa.

Giải: Đặt A_1 – ngày thứ nhất có mưa, A_2 – ngày thứ hai có mưa; theo đầu bài ta có $P(A_1A_2) = 0{,}5$; $P(\bar{A}_1\bar{A}_2) = 0{,}3$; còn hai sự kiện một ngày mưa, một ngày không mưa là đồng khả năng, nên dễ dàng tính được $P(A_1\bar{A}_2) = P(\bar{A}_1A_2) = 0{,}1$. Mặt khác

$$P(\bar{A}_1) = P(\bar{A}_1(A_2 + \bar{A}_2)) = P(\bar{A}_1A_2) + P(\bar{A}_1\bar{A}_2) = 0{,}4.$$

$$P(A_2|\bar{A}_1) = P(\bar{A}_1A_2)/P(\bar{A}_1) = 1/4 = 0{,}25.$$

<u>**Thí dụ 9**</u>. Một người viết n lá thư cho n người khác nhau, và bỏ ngẫu nhiên vào n phong bì đã có sẵn địa chỉ. Tìm XS để có ít nhất một lá thư được bỏ vào đúng phong bì.

Giải: Đặt A_i – sự kiện lá thư thứ i được bỏ đúng phong bì ($i = 1, 2, \ldots, n$), A - sự kiện cần tính XS, ta có $A = A_1 + A_2 + \ldots + A_n$. Do các A_i không xung khắc, nên ta sẽ sử dụng công thức (8). Dễ thấy:

$$P(A_i) = \frac{1}{n} = \frac{(n-1)!}{n!};$$

$$P(A_iA_j) = P(A_i)P(A_j|A_i) = \frac{1}{n}.\frac{1}{n-1} = \frac{(n-2)!}{n!};$$

$$P(A_iA_jA_k) = \ldots = \frac{(n-3)!}{n!}; \ldots$$

$$P(A_1A_2\ldots A_n) = \frac{1}{n!}.$$

Từ đó thay vào (8)

$$P\left(\sum_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) - \sum_{i<j} P(A_iA_j)$$

$$+ \sum_{i<j<k} P(A_iA_jA_k) - \ldots + (-1)^{n-1} P(A_1A_2\ldots A_n)$$

$$= C_n^1 \frac{(n-1)!}{n!} - C_n^2 \frac{(n-2)!}{n!} + C_n^3 \frac{(n-3)!}{n!} - \ldots + (-1)^{n-1} \frac{1}{n!}$$

$$= 1 - \frac{1}{2!} + \frac{1}{3!} - \ldots + (-1)^{n-1} \frac{1}{n!}.$$

Khi n khá lớn XS cần tìm $\approx 1 - 1/e$.

3.3 Công thức Béc-nu-li (Bernoulli)

Xét *lược đồ Béc-nu-li*:

- dãy n phép thử giống nhau, độc lập;
- trong mỗi phép thử có $P(A) = p$ (đặt $q = 1 - p$).

Ta quan tâm đến sự kiện "*A xuất hiện đúng k lần trong lược đồ trên*", và XS xuất hiện sự kiện đó đượcký hiệu là $P_n(k)$. Do mỗi kết cục của phép thử (dãy Béc-nu-li) là tích của n sự kiện con A hoặc $\bar{A}$, nên dễ dàng chứng minh được *công thức Béc-nu-li*

$$P_n(k) = C_n^k p^k q^{n-k}. \qquad (9)$$

Công thức (9) là dạng thuận tiện để tính toán hơn so với các công thức cộng và nhân XS, vì vậy nó có ý nghĩa thực tiễn lớn.

<u>Thí dụ 10</u>. Trong phân xưởng có 5 máy hoạt động, XS để mỗi máy bị hỏng trong ca đều bằng 0,1. Tính XS để trong ca đó có đúng 2 máy hỏng.

Giải: Ta có lược đồ Béc-nu-li, với $n = 5$, $p = 0{,}1$ và $k = 2$, áp dụng (9) ta có

$$P_5(2) = C_5^2 0{,}1^2 . 0{,}9^3 = 0{,}0729.$$

Tất nhiên ta có thể giải bằng các công thức cộng và nhân XS, nhưng sẽ dài dòng hơn.

<u>Thí dụ 11</u>. XS để chữa khỏi bệnh A bằng một loại thuốc là 0,8. Có thể kết luận cứ 5 người bị bệnh A dùng thuốc trên thì 4 người khỏi bệnh không?

Giải: Ở đây ta có lược đồ Béc-nu-li với $n = 5$, $p = 0{,}8$ và $k = 4$. Vậy ta có

$$P_5(4) = C_5^4 0{,}8^4 . 0{,}2^1 = 0{,}4096.$$

Từ đó kết luận trên không thể coi là đúng.

Nhiều khi ta cần tính XS để trong dãy *n* phép thử Béc-nu-li sự kiện *A* xuất hiện với số lần từ k_1 đến k_2; dễ thấy XS cần tìm, ký hiệu là $P_n(k_1; k_2)$, sẽ là

$$P_n(k_1; k_2) = \sum_{k=k_1}^{k_2} P_n(k) = \sum_{k=k_1}^{k_2} C_n^k\, p^k q^{n-k}. \qquad (10)$$

Nhận xét: Khi *n* và *k* khá lớn, việc tính toán XS theo (9)-(10) rất cồng kềnh và khó khăn. Vì vậy ta có thể tính xấp xỉ theo các cách sau đây:

(i) Nếu *n* rất lớn, trong khi *p* rất nhỏ, XS trong (9) có thể được tính gần đúng bằng (*xấp xỉ Poison*)

$$P_n(k) \approx \frac{(np)^k}{k!} e^{-np}. \qquad (11)$$

(ii) Nếu *n* rất lớn, trong khi *p* không quá bé và quá lớn (quá gần 1), ta có *xấp xỉ chuẩn* của (9)

$$P_n(k) \approx \frac{\varphi(x_k)}{\sqrt{npq}}, \; x_k = \frac{k-np}{\sqrt{npq}}, \qquad (12)$$

trong đó $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ là hàm Gauss.

(iii) Nếu *n* rất lớn, trong khi *p* không quá bé và quá lớn, thì XS trong (10) có thể xấp xỉ bằng

$$P_n(k_1;k_2) \approx \phi(x_2) - \phi(x_1)\,,\ x_j = \frac{k_j - np}{\sqrt{npq}}\,,\ j = 1;2, \quad (13)$$

trong đó $\phi(x) = \frac{1}{\sqrt{2\pi}}\int_0^x e^{-t^2/2}\,dt$ là hàm Laplace.

<u>Thí dụ 12</u>. Xác suất sản xuất ra phế phẩm của một máy là 0,005. Tìm XS để trong 800 sản phẩm của máy đó có đúng 3 phế phẩm.

Giải: Ở đây có thể dùng xấp xỉ Poisson theo (11) với $np = 4$

$$P_{800}(3) \approx \frac{4^3}{3!}e^{-4} = 0{,}1954.$$

<u>Thí dụ 13</u>. Xác suất ném trúng rổ của một cầu thủ là 0,8. Tìm XS để trong 100 lần ném thì cầu thủ đó:

a) ném trúng 75 lần;

b) ném trúng không ít hơn 75 lần.

Giải: Việc tính XS theo (9)-(10) khá phức tạp. Ta sẽ tính xấp xỉ theo (12) và (13):

a) $$P_{100}(75) \approx \frac{\varphi\left(\frac{75-100.0{,}8}{\sqrt{100.0{,}8.0{,}2}}\right)}{\sqrt{100.0{,}8.0{,}2}} = \frac{\varphi(-1{,}25)}{4} = 0{,}04565.$$

b) $$P_{100}(75;100) \approx \phi(5) - \phi(-1{,}25) = 0{,}8943.$$

PHỤ LỤC

1. Bảng hàm Gauss $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$

x	0	1	2	3	4	5	6	7	8	9
0.0	0,3989	3989	3989	3986	3986	3984	3982	3980	3977	3973
0.1	3970	3965	3961	3956	3951	3945	3939	3932	3925	3918
0.2	3910	3902	3894	3885	3876	3867	3857	3847	3836	3825
0.3	3814	3802	3790	3778	3765	3752	3739	3726	3712	3697
0.4	3683	3668	9653	3637	3621	3605	3589	3572	3555	3538
0.5	3521	3503	3485	3467	3448	3929	3410	3391	3372	3352
0.6	3332	3312	3292	3271	3251	3230	3209	3187	3166	3144
0.7	3123	3101	3079	3056	3034	3011	2989	2966	2943	2920
0.8	2897	2874	2850	2827	2803	2780	2756	2732	2709	2685
0.9	2661	2637	2613	2589	2565	2541	2516	2492	2468	2444
1.0	0,2420	2396	2371	2347	2323	2299	2275	2251	2227	2203
1.1	2179	2155	2131	2107	2083	2059	2036	2012	1989	1965
1.2	1942	1919	1895	1872	1849	1826	1804	1781	1758	1736
1.3	1714	1691	1669	1647	1626	1604	1582	1561	1539	1518
1.4	1497	1476	1456	1435	1415	1394	1374	1354	1334	1315
1.5	1295	1276	1257	1238	1219	1200	1182	1163	1145	1127
1.6	1109	1092	1074	1057	1040	1023	1006	0989	0973	0957
1.7	0940	0925	0909	0893	0878	0863	0848	0833	0818	0804
1.8	0790	0775	0761	0748	0734	0721	0707	0694	0681	0669
1.9	0656	0644	0632	0620	0608	0596	0584	0573	0562	0551

2.0	0,0540	0529	0519	0508	0498	0488	0478	0468	0459	0449
2.1	0440	0431	0422	0413	0404	0396	0388	0379	0371	0363
2.2	0355	0347	0339	0332	0325	0317	0310	0303	0297	0290
2.3	0283	0277	0270	0264	0258	0252	0246	0241	0235	0229
2.4	0224	0219	0213	0208	0203	0198	0194	0189	0184	0180
2.5	0175	0171	0167	0163	0158	0154	0151	0147	0143	0139
2.6	0136	0132	0129	0126	0122	0119	0116	0113	0110	0107
2.7	0104	0101	0099	0096	0093	0091	0088	0086	0084	0081
2.8	0079	0077	0075	0073	0071	0069	0067	0065	0063	0061
2.9	0060	0058	0056	0055	0053	0051	0050	0048	0047	0046
3.0	0,0044	0043	0042	0040	0039	0038	0037	0036	0035	0034
3.1	0033	0032	0031	0030	0029	0028	0027	0026	0025	0025
3.2	0024	0023	0022	0022	0021	0020	0020	0019	0018	0018
3.3	0017	0017	0016	0016	0015	0015	0014	0014	0013	0013
3.4	0012	0012	0012	0011	0011	0010	0010	0010	0009	0009
3.5	0009	0008	0008	0008	0008	0007	0007	0007	0007	0006
3.6	0006	0006	0006	0006	0006	0005	0005	0005	0005	0004
3.7	0004	0004	0004	0004	0004	0004	0003	0003	0003	0003
3.8	0003	0003	0003	0003	0003	0002	0002	0002	0002	0002
3.9	0002	0002	0002	0002	0002	0002	0002	0002	0001	0001
x	0	1	2	3	4	5	6	7	8	9

2 Bảng hàm Laplace $\phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-t^2/2}$

X	0	1	2	3	4	5	6	7	8	9
0.0	0,0000	00399	00798	01197	01595	01994	02392	02790	03188	03586
0.1	03983	04380	04776	05172	05567	05962	06356	06749	07142	07535
0.2	07926	08317	08706	09095	09483	09871	10257	10642	11026	11409
0.3	11791	12172	12556	12930	13307	13683	14058	14431	14803	15173
0.4	15542	15910	16276	16640	17003	17364	17724	18082	18439	18793
0.5	19146	19497	19847	20194	20194	20884	21226	21566	21904	22240
0.6	22575	22907	23237	23565	23891	24215	24537	24857	25175	25490
0.7	25804	26115	26424	26730	27035	27337	27637	27935	28230	28524
0.8	28814	29103	29389	29673	29955	30234	30511	30785	31057	31327
0.9	31594	31859	32121	32881	32639	32894	33147	33398	33646	33891
1.0	34134	34375	34614	34850	35083	35314	35543	35769	35993	36214
1.1	36433	36650	36864	37076	37286	37493	37698	37900	38100	38298
1.2	38493	38686	38877	39065	39251	39435	39617	39796	39973	40147
1.3	40320	40490	40658	40824	40988	41149	41309	41466	41621	41774
1.4	41924	42073	42220	42364	42507	42647	42786	42922	43056	43189
1.5	43319	43448	43574	43699	43822	43943	44062	44179	44295	44408
1.6	44520	44630	44738	44815	44950	45053	45154	45254	45352	45449
1.7	45543	45637	45728	45818	45907	45994	46080	46164	46246	46327
1.8	46407	46485	46562	46638	46712	46784	46856	46926	46995	47062
1.9	47128	47193	47257	47320	47381	47441	47500	47558	47615	47670
2.0	47725	47778	47831	47882	47932	47982	48030	48077	48124	48169
2.1	48214	48257	48300	48341	48382	48422	48461	48500	48537	48574
2.2	48610	48645	48679	48713	48745	48778	48809	48840	48870	48899
2.3	48928	48956	48983	49010	49036	49061	49086	49111	49134	49158
2.4	49180	49202	49224	49245	49266	49285	49305	49324	49343	49361
2.5	49379	49396	49413	49430	49446	49261	49477	49492	49506	49520
2.6	49534	49547	49560	49573	49585	49598	49609	49621	49632	49643
2.7	49653	49664	49674	49683	49693	49702	49711	49720	49728	49736
2.8	49744	49752	49760	49767	49774	49781	49788	49795	49801	49807
2.9	49813	49819	49825	49831	49836	49841	49846	49851	49856	49861
3.0	0,49865		3,1	49903	3,2	49931	3,3	49952	3,4	49966
3.5	49977		3,6	49984	3,7	49989	3,8	49993	3,9	49995
4.0	499968									
4.5	499997									
5.0	49999997									

BÀI TẬP

1. Gieo một con xúc xắc và đặt A là sự kiện xuất hiện mặt chẵn, B – mặt có số chấm là bội số của 3.
   a) Hai sự kiện trên có xung khắc không? Tại sao?
   b) Hai sự kiện trên có độc lập không? Tại sao?

2. Trong hộp có n quả bóng bàn mới. Người ta lấy ra ngẫu nhiên k quả để chơi ($k < n/2$) sau đó bỏ trở lại vào hộp. Tính XS để lần sau lấy ra ngẫu nhiên k quả lại được k quả mới.

3. Túi I đựng 2 bi trắng, 4 bi đỏ; túi II đựng 3 bi trắng 4 bi đỏ. Rút hú hoạ từ mỗi túi ra hai viên bi. Tính các XS để:
   a) rút được hai bi trắng;
   b) số bi trắng được rút từ mỗi túi bằng nhau;
   c) số bi trắng được rút từ túi I nhiều hơn từ túi II.

4. Lô hàng có 12 sản phẩm. Mỗi lần kiểm tra chất lượng lấy ngẫu nhiên ra 4 sản phẩm, sau khi kiểm tra xong trả lại vào lô. Tính XS để sau 3 lần kiểm tra thì tất cả các sản phẩm của lô hàng đều được kiểm tra.

5. Ba cầu thủ mỗi người ném hai quả bóng vào rổ, XS trúng rổ của từng người trong mỗi lần ném tương ứng là 0,6; 0,7 và 0,8.
   a) Tính XS có đúng 3 quả bóng trúng rổ.
   b) Tính XS số bóng trúng rổ của 3 người bằng nhau.
   c) Biết có đúng 2 quả trúng rổ, tính XS để cả hai quả đó là của cầu thủ thứ hai.

6. Một phòng máy có 3 máy tính, XS hỏng trong một ngày của mỗi máy tương ứng là 0,01; 0,02 và 0,03.
   a) Tính XS để có ít nhất 2 máy hỏng trong ngày.
   b) Biết trong ngày có ít nhất 1 máy hỏng, tính XS để trong số máy hỏng có máy thứ ba.

7. Có 3 thùng: thùng I đựng 1 bi trắng, 4 bi đỏ; thùng II đựng 2 bi trắng 3 đỏ; thùng III đựng 3 bi trắng 2 đỏ. Rút ngẫu nhiên từ mỗi thùng ra 1 viên bi.
   a) Tính XS để trong 3 bi có 2 bi trắng 1 bi đỏ.
   b) Biết trong 3 bi đó có ít nhất một bi đỏ, tính các XS để trong số các viên bi đỏ có viên bi của thùng I.

8. Một lô có 50 sản phẩm, trong đó có 5 phế phẩm. Lấy ngẫu nhiên từ lô ra 10 sản phẩm đem kiểm tra: nếu trong 10 sản phẩm nếu có nhiều nhất 1 phế phẩm thì lô được xếp đạt chất lượng, ngược lại (có nhiều hơn 1 phế phẩm) lô bị xếp không đạt chất lượng. Tính XS để lô hàng được xếp là đạt chất lượng.

9. Cho hai sự kiện A, B, trong đó $P(A) = 0{,}4$ và $P(B) = 0{,}7$. Xác định giá trị lớn nhất và nhỏ nhất của $P\,(AB)$ và $P(A + B)$, và cho thí dụ các sự kiện đạt được các giá trị đó.

10. Theo thống kê trong các gia đình có 2 con thì XS cả hai con là trai bằng 0,27, XS cả hai là gái bằng 0,23; các sự kiện có một trai, một gái là đồng khả năng. Tính XS trong một gia đình hai con được chọn ngẫu nhiên có con thứ hai là trai, biết rằng con thứ nhất là gái.

11. Một bài thi trắc nghiệm gồm 12 câu hỏi, mỗi câu hỏi cho 5 câu trả lời với 1 câu trả lời đúng. Nếu chọn câu trả lời đúng thì được 4 điểm và chọn mỗi câu trả lời sai bị trừ đi 1 điểm. Một học sinh kém làm bài bằng cách chọn hú hoạ một câu trả lời cho mỗi câu hỏi. Tính XS để:
    a) học sinh đó được 13 điểm;
    b) học sinh đó bị điểm âm.

12. Một người say rượu đi 8 bước, mỗi bước anh ta hoặc tiến lên phía trước một mét hoặc lùi lại phía sau một mét với XS như nhau. Tính XS để sau 8 bước:
    a) anh ta trở lại điểm xuất phát;

b) anh ta cách điểm xuất phát hơn 4 mét.

13. Một gia đình có 6 con, XS sinh con trai là 0,52 ở mỗi lần sinh. Tính XS để trong 6 con đó:
   a) có đúng 3 con trai;
   b) có không quá 3 con trai;
   c) có nhiều nhất 4 con trai.

14. Một công nhân đứng máy 1000 ống sợi, XS mỗi ống sợi bị đứt trong vòng một giờ là 0,005. Tính XS để trong vòng 1 giờ:
   a) có 40 ống sợi bị đứt;
   b) có không quá 40 ống sợi bị đứt.

15. Máy tính có n bộ phận. XS máy hỏng trong khoảng thời gian T của bộ phận thứ k bằng p_k ($k = 1, 2, \ldots, n$). Biết nếu chỉ cần một bộ phận hỏng thì máy ngừng làm việc. Tính XS để máy tính đó ngừng làm việc trong khoảng thời gian T.

§4 CÔNG THỨC BAYES

4.1 Nhóm đầy đủ

*** <u>Định nghĩa</u>.** Nhóm sự kiện $A_1, A_2, \ldots, A_n (n > 1)$ của một phép thử được gọi là (hay tạo thành) một *nhóm đầy đủ*, nếu

(i) $A_iA_j = V, \forall\ i \neq j$ (xung khắc từng đôi);

(ii) $A_1 + A_2 + \ldots + A_n = \text{U}$.

Theo định nghĩa này khi thực hiện phép thử trên chỉ có thể xuất hiện một và chỉ một trong số *n* sự kiện $A_1, A_2, \ldots, A_n$. Ngoài ra để ý

$$P(A_1) + P(A_2) + \ldots + P(A_n) = 1.$$

*** <u>Thí dụ 1</u>**. Gieo một con xúc xắc (phép thử). Ký hiệu A_i – sự kiện xuất hiện mặt *i* chấm ($i = \overline{1,6}$), ta có một nhóm đầy đủ là $\{A_i, i = \overline{1,6}\}$. Để ý nhóm đầy đủ này bao gồm tất cả các sự kiện sơ cấp của phép thử. Ở đây có thể tạo ra nhiều nhóm đầy đủ khác, chẳng hạn nếu đặt $A = A_6$ thì $\bar{A} = A_1 + A_2 + \ldots + A_5$ và $\{A, \bar{A}\}$ tạo nên một nhóm đầy đủ; hoặc nếu đặt $A_c = A_2 + A_4 + A_6$ thì nhóm $\{A_c, A_1, A_3, A_5\}$ cũng là một nhóm đầy đủ.

Tổng quát hoá tập các sự kiện tạo nên một phân hoạch của không gian Ω là một nhóm đầy đủ. Còn nhóm $\{A, \bar{A}\}$, với A – sự kiện bất kỳ, tạo ra nhóm đầy đủ bé nhất (chỉ có 2 phần tử).

* **<u>Thí dụ 2</u>**. Đặt A là sự kiện bóng thứ nhất tốt, B là sự kiện bóng thứ hai tốt. Khi đó bộ $\{AB, A\bar{B}, \bar{A}B, \bar{A}\bar{B}\}$ tạo nên nhóm đầy đủ. Nhóm này mô tả mọi phương án (về chất lượng tốt xấu) của cả hai bóng đang xét.

4.2 Công thức xác suất đầy đủ

Giả sử ta có nhóm đầy đủ $\{A_1, A_2, \dots, A_n\}$, đồng thời xét một sự kiện H nào đó. Cũng giả thiết rằng ta biết các thông tin (tính được) qua các XS $P(A_i)$ và $P(H|A_i)$, $i = \overline{1, n}$. Bây giờ ta đi tính $P(H)$. Từ điều kiện (ii) của nhóm đầy đủ ta có

$$H = A_1H + A_2H + \dots + A_nH = \sum_{i=1}^{n} A_iH.$$

Từ đó do xung khắc từng đôi ($A_iHA_jH = A_iA_jH = V$)

$$P(H) = P\left(\sum_{i=1}^{n} A_i H\right) = \sum_{i=1}^{n} P(A_iH)$$

và áp dụng công thức nhân XS

$$P(H) = \sum_{i=1}^{n} P(A_i)P(H|A_i). \qquad (1)$$

Công thức (1) có tên gọi là *công thức xác suất đầy đủ* (hay có tên khác là *công thức xác suất toàn phần*).

Giải: Đặt A_1, A_2 và A_3 là sản phẩm chọn ra do máy I, II và III tương ứng. Rõ ràng $\{A_i, i = \overline{1,3}\}$ là một nhóm đầy đủ và

*** <u>Thí dụ 3</u>**. Một cậu bé có 6 bi đỏ và 4 bi trắng. Một hôm cậu thấy mất 1 viên bi. Tính XS để khi lấy hú họa ra một viên bi thì ta được bi đỏ.

Giải: Vấn đề ở đây là viên bi mất không biết màu. Ta lập một nhóm đầy đủ là rõ viên bi mất có màu nào: Đặt A_1- viên bi bị mất màu đỏ, A_2 – viên bi mất màu trắng. Rõ ràng $\{A_1, A_2\}$ là một nhóm đầy đủ. Từ đó nếu gọi H là sự kiện lấy được bi đỏ trong số bi chưa mất thì theo công thức XS đầy đủ (1)

$$P(H) = P(A_1)P(H|A_1) + P(A_2)P(H|A_2)$$

$$= \frac{6}{10}\cdot\frac{5}{9} + \frac{4}{10}\cdot\frac{6}{9} = 0{,}6.$$

*** <u>Thí dụ 4</u>**. Một tổ có 3 máy sản xuất cùng một loại sản phẩm với tỷ lệ phế phẩm tương ứng 0,5%, 1% và 2%. Cũng biết rằng máy I sản xuất 30%, máy II – 25% và máy III – 45% sản phẩm của tổ. Chọn hú hoạ ra một sản phẩm, tìm xác suất đó là phế phẩm.

Giải: Đặt A_1, A_2 và A_3 là sản phẩm chọn ra do máy I, II và III tương ứng. Rõ ràng $\{A_i, i = \overline{1,3}\}$ là một nhóm đầy đủ và

$$P(A_1) = 0{,}30;\ P(A_2) = 0{,}25;\ P(A_3) = 0{,}45.$$

Gọi H sự kiện rút được phế phẩm, áp dụng công thức XSĐĐ (1) và để ý

$$P(H|A_1) = 0{,}5\%;\ P(H|A_2) = 1\%;\ P(H|A_3) = 2\%,$$

ta có

$$P(H) = \sum_{i=1}^{3} P(A_i)P(H|A_i)$$

$$= 0{,}30.0{,}5\% + 0{,}25.1\% + 0{,}45.2\% = 1{,}3\%.$$

XS này có ý nghĩa như là tỷ lệ phế phẩm chung của cả ba máy.

*** <u>Thí dụ 5</u>**. Có 3 hộp bi: hộp I đựng 3 bi đen và 5 bi trắng, hộp II đựng 2 bi đen và 4 bi trắng, hộp III đựng 3 bi đen và 4 bi trắng. Lấy hú hoạ 2 bi từ hộp I và 1 bi từ hộp II bỏ sang hộp III rồi trộn đều, sau đó lấy hú họ một viên bi từ hộp III này. Tính XS để viên bi lấy được là đen.

Giải: Đặt A_1, A_2 và A_3 là sự kiện viên bi lấy ra thuộc hộp I, II và

III tương ứng. Khi đó A_1, A_2 và A_3 lập thành một nhóm đầy đủ và:

$$P(A_1) = 2/10 = 0{,}2;\ P(A_2) = 0{,}1;\ P(A_3) = 0{,}7.$$

Gọi H là sự kiện lấy được bi đen, ta có:

$$P(HA_1) = \frac{C_3^1 C_5^1}{C_8^2}\,\frac{C_1^1}{C_{10}^1} + \frac{C_3^2}{C_8^2}\,\frac{C_2^1}{C_{10}^1} = 3/40,$$

$$P(HA_2) = \frac{C_2^1}{C_6^1}\,\frac{C_1^1}{C_{10}^1} = 1/30,$$

$$P(HA_3) = \frac{C_3^1}{C_{10}^1} = 3/10.$$

Bây giờ áp dụng công thức trung gian để dẫn ra công thức XS đầy đủ (1) (để ý $H = HU = H(A_1 + A_2 + A_3)$)

$$\begin{aligned} P(H) &= P(HA_1) + P(HA_2) + P(HA_3) \\ &= 3/40 + 1/30 + 3/10 \\ &= 49/120 = 0{,}4083. \end{aligned}$$

*** <u>Thí dụ 6</u>**. Ba khẩu pháo bắn mỗi khẩu một phát vào một mục tiêu với XS trúng đích tương ứng là 0,7; 0,8 và 0,9. Xác suất để mục tiêu bị tiêu diệt tương ứng nếu trúng 1 viên là 0,4, trúng 2 viên là 0,6, trúng 3 viên là 1. Tìm XS mục tiêu bị tiêu diệt.

Giải: Đặt A_i, $i = \overline{0,3}$, là sự kiện số đạn trúng mục tiêu bằng *i*. Gọi K_i, $i = \overline{1,3}$, sự kiện khẩu pháo thứ *i* bắn trúng mục tiêu. Khi đó:

$$A_0 = \overline{K}_1\overline{K}_2\overline{K}_3;\ P(A_0) = 0{,}3.0{,}2.0{,}1 = 0{,}006;$$

$$A_1 = K_1\overline{K}_2\overline{K}_3 + K_2\overline{K}_1\overline{K}_3 + K_3\overline{K}_1\overline{K}_2;$$

$$P(A_1) = 0{,}7.0{,}2.0{,}1+0{,}8.0{,}3.0{,}1+0{,}9.0{,}3.0{,}2 = 0{,}092;$$

$$A_2 = K_1K_2\overline{K}_3+K_1\overline{K}_2K_3+\overline{K}_1K_2K_3;$$

$$P(A_2) = 0{,}7.0{,}8.0{,}1+0{,}7.0{,}2.0{,}9+0{,}3.0{,}8.0{,}9 = 0{,}398;$$

$$A_3 = K_1K_2K_3;\ P(A_3) = 0{,}7.0{,}8.0{,}9 = 0{,}504.$$

Dễ thấy $\{A_i, i = \overline{0,3}\}$ lập thành một nhóm đầy đủ. Gọi H là sự kiện mục tiêu bị tiêu diệt, khi đó:

$$P(H|A_0) = 0;\ P(H|A_1) = 0{,}4;\ P(H|A_2) = 0{,}6;\ P(H|A_3) = 1;$$

$$P(H) = \sum_{i=0}^{3} P(A_i)P(H|A_i)$$

$$= 0{,}006.0 + 0{,}092.0{,}4 + 0{,}398.0{,}6 + 0{,}504.1 = 0{,}7796.$$

4.3 Công thức Bayes

Giả sử ta có một nhóm đầy đủ $\{A_1, A_2, \ldots, A_n\}$ và một sự kiện H nào đó. Nhiều khi ta muốn xác định XS $P(A_i|H)$, trong đó i là một số nào đó thuộc tập $\{1, 2, \ldots, n\}$. Từ công thức nhân XS

$$P(A_i H) = P(A_i)P(H|A_i) = P(H)P(A_i|H),$$

ta có $$P(A_i|H) = \frac{P(A_i)P(H|A_i)}{P(H)} \qquad (2)$$

và thay (1) vào công thức (2)

$$P(A_i|H) = \frac{P(A_i)P(H|A_i)}{\sum_{i=1}^{n} P(A_i)P(H|A_i)}\,. \qquad (3)$$

Công thức (3) có tên gọi là *công thức Bayes*.

Chú ý:

- $P(A_i)$, $i = \overline{1,n}$, thường được gọi là XS tiên nghiệm (*a priori probability*);
- $P(A_i|H)$, $i = \overline{1,n}$, được tính sau khi đã có kết quả thí nghiệm nào đó thể hiện bởi sự xuất hiện *H*, thường được gọi là XS hậu nghiệm (*a postpriori probability*);
- A_i, *i* là một hằng số nào đó thuộc tập {1, 2, …, *n*}, trong XS $P(A_i|H)$ phải là thành viên của nhóm đầy đủ.

Như vậy sự kiện cần tính XS trong $P(A_i|H)$ đã gợi ý cách chọn nhóm đầy đủ sao cho sự kiện quan tâm A_i là thành viên của nhóm. Trong trường hợp không tìm được nhóm đầy đủ như vậy, hoặc tìm được nhưng không phù hợp để tính được, nên dùng công thức

(2), tuy nhiên việc tính $P(H)$ sẽ khó hơn là dùng công thức (1).

* **<u>Thí dụ 7</u>**. Ba xạ thủ mỗi người bắn một phát vào bia với XS trúng đích tương ứng là 0,6; 0,8 và 0,9. Người báo bia thông báo có 2 viên trúng, tính XS để trong số phát trúng có của anh thứ nhất.

Giải: Đầu tiên ta sử dụng công thức Bayes để giải bài toán. Đặt $A_i, i = \overline{1,3}$, *là sự kiện* xạ thủ thứ *i* bắn trúng bia, H – có hai người bắn trúng, khi đó ta cần tìm $P(A_1|H)$. Chọn nhóm đầy đủ đơn giản $\{A_1, \bar{A}_1\}$, khi đó theo (1)

$$P(H) = P(A_1)P(H|A_1) + P(\bar{A}_1)P(H|\bar{A}_1)$$

$$= P(A_1)P(A_2\bar{A}_3 + \bar{A}_2A_3) + P(\bar{A}_1)P(A_2A_3)$$

$$= 0{,}6.(0{,}8.0{,}1 + 0{,}2.0{,}9) + 0{,}4.0{,}8.0{,}9 = 0{,}444.$$

Từ đó theo công thức Bayes

$$P(A_1|H) = \frac{P(A_1)P(H|A_1)}{\sum_{i=1}^{2} P(A_i)P(H|A_i)} = 0{,}156/0{,}444 = 0{,}3514.$$

Để ý ta có thể dùng công thức (2) mà không cần nhóm đầy đủ, khi đó $P(H)$ có thể được tính trực tiếp nhờ phân tích sự kiện

$$H = A_1A_2\bar{A}_3 + A_1\bar{A}_2A_3 + \bar{A}_1A_2A_3.$$

* **<u>Thí dụ 8</u>**. Một nhà máy có 3 phân xưởng sản xuất cùng một loại

sản phẩm với tỷ lệ phế phẩm tương ứng là 1,2 %; 1% và 0,8%. Biết phân xưởng I sản xuất 36%, phân xưởng II – 34%, phân xưởng III – 30% sản phẩm của nhà máy.

a) Lấy hú hoạ từ kho chung ra một sản phẩm thì đó là phế phẩm, tính XS sản phẩm đó thuộc phân xưởng II.

b) Phế phẩm đó có khả năng thuộc phân xưởng nào lớn nhất?

Giải:

a) Gọi H là sự kiện lấy được phế phẩm, ta cũng đã biết $P(H)$ chính là tỷ lệ phế phẩm chung của nhà máy. Đặt $A_i, i = \overline{1,3}$, là sự kiện sản phẩm do phân xưởng I sản xuất và chúng lập thành một nhóm đầy đủ. Theo dữ kiện đầu bài:

$$P(A_1) = 0,36;\ P(A_2) = 0,34;\ P(A_3) = 0,30;$$

$$P(H|A_1) = 0,012;\ P(H|A_2) = 0,01;\ P(H|A_3) = 0,009.$$

Khi đó theo công thức Bayes (3)

$$P(A_2|H) = \frac{P(A_2)P(H|A_2)}{\sum_{i=1}^{3} P(A_i)P(H|A_i)}$$

$$= \frac{0,34.0,01}{0,36.0,012+0,34.0,01+0,3.0,009} = \frac{0,0034}{0,01042} = 0,3263.$$

b) Ta phải so sánh các XS $P(A_i|H), , i = \overline{1,3}$,

$$P(A_1|H) = \frac{P(A_1)P(H|A_1)}{\sum_{i=1}^{3} P(A_i)P(H|A_i)} = 0{,}4146;$$

$$P(A_3|H) = \frac{P(A_3)P(H|A_3)}{\sum_{i=1}^{3} P(A_i)P(H|A_i)} = 0{,}2591.$$

Vậy khả năng phế phẩm đó thuộc phân xưởng I là lớn nhất.

*** <u>Thí dụ 9</u>**. Tại một phòng khám chuyên khoa tỷ lệ người đến khám có bệnh là 80%. Theo thống kê biết rằng nếu chẩn đoán có bệnh thì đúng tới 95%, còn nếu chẩn đoán không bệnh thì chỉ đúng 85%.

a) Tính XS chẩn đoán đúng.

b) Biết có một trường hợp chẩn đoán đúng, tính XS để người được chẩn đoán đó có bệnh.

Giải:

Gọi H là sự kiện chẩn đoán đúng, vậy $\overline{H}$ - sự kiện chẩn đoán sai; đặt A – người khám có bệnh, $\bar{A}$ - người khám không bệnh, B - chẩn đoán có bệnh, $\bar{B}$ - chẩn đoán không bệnh. Đầu bài đã cho: $P(H|B) = 95\%$, $P(H|\bar{B}) = 85\%$ và $P(A) = 0{,}8$.

a) Ở đây có thể dùng 2 nhóm đầy đủ là $\{A, \bar{A}\}$ và $\{B, \bar{B}\}$. Ta thử sử dụng công thức XS đầy đủ

$$P(H) = P(A)P(H|A) + P(\bar{A})P(H|\bar{A})$$

hoặc $$P(H) = P(B)P(H|B) + P(\bar{B})P(H|\bar{B}).$$

Ở công thức thứ nhất ta không biết các XS có điều kiện, còn trong công thức thứ hai các XS của B và $\bar{B}$ không biết, vì vậy không tính được. Ta khai thác công thức sau

$$P(A) = P(B)P(A|B) + P(\bar{B})P(A|\bar{B}). \qquad (*)$$

Từ ý nghĩa thực tế ta có:

$$P(A|B) = P(H|B) = 0{,}95;$$

$$P(A|\bar{B}) = P(\bar{H}|\bar{B}) = 1 - P(H|\bar{B}) = 0{,}15$$

và đặt $P(B) = x \Rightarrow P(\bar{B}) = 1 - x$, thay vào (*) ta tìm được $x = P(B) = 13/16$. Bây giờ có thể sử dụng công thức XS đầy đủ thứ hai

$$P(H) = P(B)P(H|B) + P(\bar{B})P(H|\bar{B}) = \frac{13}{16}.0{,}95 + \frac{3}{16}.0{,}85$$

$$= 0{,}93125.$$

b) XS cần tìm là $P(A|H)$. Áp dụng công thức (2)

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

Nhưng ở đây $P(H|A)$ không biết, tuy nhiên từ ý nghĩ thực tế

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

Và thay vào công thức trên

$$P(A|H) = \frac{P(B)P(A|B)}{P(H)} = \frac{0{,}8125.0{,}95}{0{,}93125} = 0{,}8289.$$

BÀI TẬP

1. Có 3 xạ thủ cùng bắn vào bia, XS trúng bia của các xạ thủ tương ứng là 0,9; 0,8 và 0,7. Gọi ngẫu nhiên một xạ thủ ra bắn một viên đạn và anh ta bắn trúng. Tìm XS anh ta là người thứ nhất.

2. Có 3 máy cùng sản xuất một loại linh kiện với tỷ lệ phế phẩm tương ứng là 0,02; 0,03 và 0,04. Chọn một máy để sản xuất 1 linh kiện, biết XS để chọn máy tương ứng là 0,5; 0,3 và 0,2.
   a) Tính XS để linh kiện được sản xuất ra là tốt.
   b) Biết linh kiện được sản xuất ra là tốt, tìm XS để linh kiện này do máy thứ hai sản xuất.

3. Có hai hộp: hộp I đựng 2 bi đen, 3 bi trắng, hộp II đựng 2 bi đen, 2 bi trắng. Lấy hú hoạ hai viên bi từ hộp I bỏ vào hộp II rồi rút ngẫu nhiên từ hộp II này một viên bi.
   a) Tính XS để viên bi rút ra là bi đen.
   b) Biết viên bi rút ra là đen, tính XS đó là viên bi của hộp I.

4. Một phòng máy có 30 máy tính, trong đó 14 máy có XS hỏng trong một ngày của mỗi máy là 0,1; 10 máy có XS hỏng mỗi máy là 0,2 và 6 máy có XS hỏng mỗi máy là 0,03. Giao hú hoạ cho 2 sinh viên sử dụng 2 máy trong một ngày. Tính XS để 2 máy đều không hỏng trong ngày.

5. Ba người máy độc lập chẩn đoán một bệnh nhân. Sau khi làm việc có đúng một người máy cho kết quả chẩn đoán đúng. Tìm XS đó là người máy thứ ba, biết rằng XS chẩn đoán đúng của từng người máy tương ứng là 0,8; 0,9 và 0,7.

6. Theo thống kê ở Anh năm 1993 thì có 5% cha và con cùng mắt đen, 7,9% cha mắt đen con mắt xanh, 8,9% cha xanh con đen và 78,2% cả hai cha con có cùng màu mắt xanh.
   a) Hãy tìm XS để khi cha mắt xanh thì con cũng xanh.
   b) Tìm XS để khi cha mắt đen mà con không đen.

7. Một cặp sinh đôi cùng trứng thì bao giờ cũng cùng giới. Trong trường hợp sinh đôi khác trứng thì hai trường hợp sinh cùng giới và khác giới là đồng khả năng. Biết XS sinh đôi cùng trứng là p. Một người sinh một cặp sinh đôi cùng giới, tìm XS để cặp trẻ đó cùng trứng.

8. Trong một kỳ thi 20 học sinh, có 8 học sinh xuất sắc, 6 khá, 4 trung bình và 2 trung bình yếu. Biết trong 40 vấn đề thi học sinh xuất sắc trả lời được tất cả, học sinh khá – 35 vấn đề, học sinh trung bình – 25 vấn đề và học sinh yếu trả lời được 20 vấn đề. Một học sinh đã làm được bài thi, trong đó có 3 vấn đề, tìm XS để anh ta thuộc nhóm học sinh trung bình.

9. Một chiếc máy bay có thể xuất hiện ở vị trí A với XS 2/3 và ở vị trí B với XS 1/3. Có 3 phương án bố trí 4 khẩu pháo phòng không: (1) 3 khẩu đặt tại A và 1 khẩu đặt tại B, (2) 2 khẩu tại A và 2 khẩu tại B, (3) 1 khẩu tại A và 3 khẩu tại B. Hỏi phương án nào tốt nhất, biết rằng XS bắn hạ máy bay của mỗi khẩu pháo là 0,7.

10. Trong một kho rượu số lượng chai rượu loại A và loại B bằng nhau. Người ta chọn ngẫu nhiên một chai rượu và đưa cho 5 người sành rượu nếm thử để xác định loại rượu, giả sử mỗi người đó có XS đoán trúng là 75%. Sau

khi nếm thử có 4 người kết luận chai rượu loại A và 1 người kết luận chai rượu loại B, tính XS để chai rượu được chọn là loại A.

11. Biết rằng một người có nhóm máu AB có thể nhận máu của bất kỳ nhóm máu nào, còn nếu người có một trong 3 nhóm máu còn lại (hoặc A hoặc B hoặc O) thì chỉ có thể nhận máu của người cùng nhóm hoặc có nhóm máu O. Theo thống kê tỷ lệ người có nhóm máu O, A, B và AB tương ứng là 33,7%, 37,5%, 20,9% và 7,9%.
    a) Chọn ngẫu nhiên một người cần tiếp máu và một người cho máu, tính XS để sự truyền máu có thể thực hiện được.
    b) Chọn ngẫu nhiên một người cần tiếp máu và hai người cho máu, tính XS để sự truyền máu thực hiện được.

12. Một bệnh nhân bị nghi có thể mắc một trong 3 bệnh A, B, C với các XS tương ứng là 0,3; 0,4 và 0,3. Người đó đến khám bệnh ở 4 bác sỹ: bác sỹ I chẩn đoán mắc bệnh A, bác sỹ II – mắc bệnh B, bác sỹ III – mắc bệnh C và bác sỹ IV – măc bệnh A, cho biết XS chẩn đoán đúng của mỗi bác sỹ là 0,6 và chẩn đoán nhầm mắc hai bệnh khác có XS 0,2 và 0,2. Hỏi sau khi khám bệnh XS mắc bệnh A, B, C sẽ là bao nhiêu?

§1 BIẾN NGẪU NHIÊN RỜI RẠC

1.1 Khái niệm biến số ngẫu nhiên

*** Khái niệm**

Biến ngẫu nhiên là biến số phụ thuộc vào yếu tố ngẫu nhiên nào đó. Ta sẽ ký hiệu biến ngẫu nhiên là X, Y, ..., X_1, X_2,...

Trong toán học người ta đưa ra định nghĩa

$$X = X(\omega): \Omega \longmapsto \mathbb{R}.$$

*** <u>Thí dụ 1</u>**. Gieo một con xúc xắc (phép thử) và gọi X là số chấm xuất hiện. Rõ ràng biến ngẫu nhiên X phụ thuộc vào kết cục của phép thử và lấy giá trị trong tập {1, 2, ..., 6}.

*** <u>Thí dụ 2</u>**. Chiều cao của một sinh viên là một biến ngẫu nhiên. Phụ thuộc vào độ chính xác của phép đo, ta có giá trị của nó nằm trong một khoảng số nào đó, chẳng hạn (1,69; 1,71m).

Biến ngẫu nhiên
rời rạc

Biến ngẫu nhiên
liên tục

*** Phân loại**

(i) Biến ngẫu nhiên được gọi là *rời rạc*, nếu tập giá trị của nó là một tập hữu hạn $\{x_1, x_2, \dots, x_n\}$ hoặc vô hạn đếm được các phần tử (*tập đếm được*).

(ii) Biến ngẫu nhiên được gọi là *liên tục*, nếu tập giá trị của nó lấp kín một khoảng (đoạn) hoặc một số khoảng trên trục số hoặc cũng có thể là cả trục số (*tập không đếm được*).

1.2 Bảng phân phối xác suất (Bảng PPXS)

Ta ký hiệu $p_i = P(X = x_i), i = 1, 2, \dots, n, \dots$

*** Định nghĩa.** *Bảng PPXS của một biến ngẫu nhiên rời rạc X* được xác định như sau

$X = x$	x_1	x_2	...	x_n	...
$p(x)$	p_1	p_2	...	p_n	...

Chú ý tập giá trị của X được sắp xếp theo thứ tự tăng dần, còn hàm $p(x) = P(X = x)$.

*** Thí dụ 3**. Gọi X là số chấm xuất hiện khi gieo một con xúc xắc (xem thí dụ 1). Khi đó bảng PPXS của X sẽ là

$X = x$	1	2	3	...	6
$p(x)$	1/6	1/6	1/6	...	1/6

Để ý các XS $p(x)$ bằng nhau trên tập giá trị {1, 2, ..., 6} của X (đồng khả năng). Ta nói rằng X có phân phối đều trên tập {1, 2, ..., 6} và ký hiệu $X \sim \mathcal{U}_{\{1,2,\ldots,6\}}$ (*uniform distribution of probability*). Người ta cũng coi XS để X nhận các giá trị không có trong bảng bằng 0.

*** Thí dụ 4**. Một xạ thủ có 3 viên đạn. Anh ta được yêu cầu bắn từng viên cho đến khi trúng thì dừng bắn. Tìm PPXS của số đạn cần bắn, biết rằng XS bắn trúng mỗi lần bắn là 0,4.

Giải: Rõ ràng số đạn cần bắn, ký hiệu là X, là biến ngẫu nhiên rời rạc và có 3 giá trị 1, 2, 3.

$X = 1$ $\Leftrightarrow$ sự kiện phát đạn đầu trúng, $p_1 = P(X=1) = 0{,}4$;

$X = 2$ $\Leftrightarrow$ phát đầu trượt, phát thứ hai trúng và dễ dàng tính được $p_2 = 0{,}6.0{,}4 = 0{,}24$;

$X = 3$ $\Leftrightarrow$ hai phát đạn đầu trượt, $p_3 = 0{,}36$. Từ đó

$X = x$	1	2	3
$p(x)$	0,4	0,24	0,36

*** Thí dụ 5**. Lấy lại thí dụ 4 nhưng xạ thủ được yêu cầu bắn hết đạn.

Tìm PPXS của số đạn trúng.

Giải: Nếu gọi X là số đạn trúng thì tập giá trị của nó {0, 1, 2, 3}, ở đây ta có lược đồ Bernoulli với n =3, p =0,4 và $k = \overline{0,3}$. Sử dụng công thức Bernoulli $p(k) = C_n^k\, p^k q^{n-k}$

$X = x$	0	1	2	3
$p(x)$	0,064	0,288	0,432	0,216

Trong trường hợp tổng quát (n, p tuỳ ý) ta gọi PPXS của X là *phân phối nhị thức* (*binomial distribution*) và ký hiệu là $X \sim \mathcal{B}(n, p)$. Ở đây $X \sim \mathcal{B}(3; 0{,}4)$.

Hàm số $p(x) = P(X = x)$, $x \in$ tập giá trị của X, được gọi là *hàm xác suất* (*probability function*) của X; nó có hai tính chất quan trọng:

*** Tính chất**

(i) $p(x) > 0 \ \forall x$;

(ii) $\sum_{\forall x} p(x) = 1$.

*** <u>Thí dụ 6</u>**. Biến ngẫu nhiên X được coi là tuân theo *luật PPXS Poisson*, ký hiệu là $X \sim \mathcal{P}(\lambda)$, λ là tham số nào đó, nếu hàm xác suất của nó được xác định như sau

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

Chú ý hàm của các biến ngẫu nhiên rời rạc tiếp tục là biến ngẫu nhiên rời rạc. Việc xác định PPXS của biến hàm trong trường hợp này tương đối đơn giản.

* **Thí dụ 7**. Cho hai biến X và Y có bảng PPXS tương ứng:

x	-1	0	1
$p(x)$	0,3	0,4	0,3

y	1	2
$p(y)$	0,3	0,7

Lập bảng PPXS của: a) X^2; b) $X+Y$.

Giải:

a) Biến $Z = X^2$ chỉ có hai giá trị 0 và 1. Từ đó ta có bảng PPXS

z	0	1
$p(z)$	0,4	0,6

b) Biến $Z = X + Y$ có các giá trị 0, 1, 2 và 3. Để ý rằng

$$P(Z = z_k) = \sum_{\forall i,j:\, z_k = x_i + y_j} P(\{X = x_i\}.\{Y = y_j\}).$$

Tích trong tổng được hiểu là sự kiện xảy ra đồng thời $\{X = x_i\}$ và $\{Y = y_j\}$. Nếu thêm giả thiết hai sự kiện trên độc lập thì

$$P(\{X = x_i\}.\{Y = y_j\}) = P(X = x_i)P(Y = y_j).$$

Từ đó bảng PPXS của Z

z	0	1	2	3
$p(z)$	0,09	0,33	0,37	0,21

(chẳng hạn $P(Z = 2) = P(0;2) + P(1;1) = 0{,}4.0{,}7 + 0{,}3.0{,}3$).

1.3 Hàm phân phối xác suất (*hàm phân phối*)

Có một cách khác để đặc trưng cho phân phối XS.

*** Định nghĩa.** *Hàm phân phối xác suất* (*probability distribution function*) của một biến ngẫu nhiên X, ký hiệu là $F(x)$, được xác định như sau

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

Ý nghĩa: $F(x)$ phản ánh độ tập trung XS ở bên trái điểm x

$\Rightarrow$ *hàm phân phối tích luỹ (xác suất tích luỹ)*

$\Rightarrow$ *hàm phân phối tích phân.*

*** <u>Thí dụ 8</u>**. Từ bảng PPXS của thí dụ 4 ta có $F(x) = \sum_{x_i < x} p(x_i) \Rightarrow$

$$F(x) = \begin{cases} 0; & x < 1, \\ 0{,}4; & 1 < x \leq 2, \\ 0{,}64; & 2 < x \leq 3, \\ 1; & x > 3. \end{cases}$$

Nếu vẽ đồ thị, ta có dạng bậc thang. Để ý là tại mỗi nơi chuyển bậc thì ta có điểm gián đoạn loại 2. Hàm phân phối cho biến liên tục có vai trò quan trọng trong nghiên cứu xác suất.

*** Tính chất**

(i) $1 \geq F(x) \geq 0$;

(ii) $F(x)$ là hàm không giảm $\Leftrightarrow \forall\, x_2 > x_1 \Rightarrow F(x_2) \geq F(x_1)$;

(iii) $P(\alpha \leq X < \beta) = F(\beta) - F(\alpha)$;

(iv) $F(+\infty) = 1$; $F(-\infty) = 0$.

Việc chứng minh các tính chất này khá đơn giản dùng định nghĩa hàm phân phối.

Hạn chế: + khó biết nơi nào tập trung nhiều XS hơn,

+ sau này ta thấy không thể biểu diễn $F(x)$ dưới dạng hàm sơ cấp trong nhiều trường hợp.

1.4 Các số đặc trưng cơ bản

A. Kỳ vọng

Ta xét *phép tính kỳ vọng* trong trường hợp biến rời rạc. Giả thiết rằng X có hàm XS $p(x)$. Khi đó kỳ vọng (*Expectation value*) của $Z = g(X)$ được định nghĩa như sau

*** Định nghĩa 1.** Kỳ vọng của $Z = g(X)$ với $p(x)$ là hàm XS đã cho của biến X được tính theo

$$EZ = \sum_{\forall x} g(x)p(x). \tag{1}$$

Trường hợp riêng khi $Z = X$, ta có

*** Định nghĩa 2.** *Kỳ vọng của X*, với $p(x)$ là hàm XS đã cho của biến X, được tính theo

$$EX = \sum_{\forall x} xp(x) = \sum_{\forall i} x_i p_i. \tag{2}$$

*** Tính chất**

(i) $E(c) = c$, $c = const$;

(ii) $E(cX) = cEX$;

(iii) $E(X + Y) = EX + EY$;

(iv) $E(XY) = EX.EY$, X và Y độc lập.

Chú ý: tính chất (ii) và (iii) được gọi là *tính tuyến tính* của phép toán, vốn quen thuộc trong phép tính đạo hàm hay tích phân.

Ý nghĩa thực tế: kỳ vọng là *giá trị trung bình* (*mean value*) của biến ngẫu nhiên hoặc đóng vai trò *định vị* của biến.

*** <u>Thí dụ 9</u>.** Xét lại thí dụ 3 và cần tính kỳ vọng của X.

Giải: Do $X \sim \mathcal{U}_{\{1,2,\dots,6\}}$ nên sử dụng (2)

$$EX = \frac{1}{6}(1 + 2 + \dots + 6) = 3{,}5.$$

Nhận xét rằng nếu X có phân phối đều thì EX nhận giá trị chính giữa tập giá trị, ở đây $EX = (1 + 6)/2$.

Nếu muốn tìm kỳ vọng của tổng số chấm xuất hiện khi gieo 3 con xúc xắc, sủ dụng tính chất (iii) ta có kỳ vọng bằng $3.3{,}5 = 10{,}5$.

* **Thí dụ 10**. Một nhóm có 6 nam và 3 nữ. Chọn ngẫu nhiên ra 3 người. Hỏi trung bình chọn được bao nhiêu nam trong 3 người?

Giải: Gọi X – số nam được chọn, dễ thấy với $k = 0, 1, 2, 3$

$$P(X = k) = \frac{C_6^k C_3^{3-k}}{C_9^3}.$$

Thay vào (2) và thực hiện tính toán

$$p(0) = 1/84; p(1) = 3/14; p(2) = 15/28; p(3) = 5/21$$

$$\Rightarrow \qquad EX = 0.\frac{1}{84} + 1.\frac{3}{14} + 2.\frac{15}{28} + 3.\frac{5}{21} = 2.$$

* **Thí dụ 11**. Một trạm y tế có 1 bác sỹ, 2 y sỹ và 6 điều dưỡng. Chọn ngẫu nhiên từng người một cho đến khi được 1 điều dưỡng thì dừng chọn. Tính (theo cách chọn trên):

a) Số người trung bình được chọn.

b) Số y sỹ trung bình được chọn.

Giải:

a) Gọi X – số người được chọn theo cách của đầu bài và X nhận các giá trị 1, 2, 3 và 4. Từ đó, nếu đặt A_i - chọn được người không phải là điều dưỡng ở lần chọn thứ i, i =1, 2, 3, 4:

$$P(X=1) = \mathrm{P}(\bar{A}_1) = 6/9 = 2/3;$$

$$P(X=2) = P(A_1\bar{A}_2) = 3.6/(9.8) = 1/4;$$

$$P(X=3) = P(A_1A_2\bar{A}_3) = 3.2.6/(9.8.7) = 1/14;$$

$$P(X=4) = P(A_1A_2A_3\bar{A}_4) = 3.2.1.6/(9.8.7.6) = 1/84.$$

Có thể kiểm tra dễ dàng $\sum_{\forall x} p(x) = 1$. Áp dụng (2) ta có

$EX = \sum_{\forall x} xp(x) = 1.\frac{2}{3} + 2.\frac{1}{4} + 3.\frac{1}{14} + 4.\frac{1}{84} = \frac{10}{7} \approx 1{,}4286.$

b) Gọi Y là số y sỹ được chọn và ta cần tính EY. Việc lập bảng PPXS cho Y khá phức tạp. Tuy nhiên nếu sử dụng tính chất của kỳ vọng thì đơn giản hơn nhiều. Gọi Z – số bác sỹ được chọn theo cách của đầu bài. Ta có:

$$Y + Z + 1 = X \quad \Rightarrow \quad \boldsymbol{E}Y + EZ + 1 = EX = 10/7$$

$$\Rightarrow \quad \boldsymbol{E}Y + EZ = 3/7;$$

mặt khác theo quy tắc tam xuất

$$EY/EZ = 2/1 \quad \Rightarrow \quad \boldsymbol{E}Y = 2/7 \approx 0{,}2857.$$

B. Phương sai

Dùng phép tính kỳ vọng ở mục trước ta có thể đưa ra khái niệm *phương sai* (*Variance value*).

*** Định nghĩa.** Phương sai của biến ngẫu nhiên X, ký hiệu là VX, được xác định như sau

$$VX = E[(X - EX)^2]. \tag{3}$$

Nhận xét: $X - EX$ là độ lệch của X so với trung bình của nó

⇒ Phương sai – trung bình của bình phương độ lệch

⇒ Đặc trưng cho độ phân tán của X quanh trung bình

độ bất định; độ rủi ro ...

Chú ý: Công thức (3) có dạng tương đương

$$VX = E(X^2) - (EX)^2.$$

Cách tính: $VX = \sum_{\forall x}(x - EX)^2\, p(x) = \sum_{\forall i}(x_i - EX)^2 p_i$

$= \sum_{\forall x} x^2\, p(x) - (EX)^2 = \sum_{\forall i} {x_i}^2 p_i - (EX)^2.$

*** Tính chất**

(i) $V(c) = 0$, $c = const$;

(ii) $V(cX) = c^2 VX$;

(iii) $V(X + Y) = VX + VY$, với X và Y độc lập.

* **<u>Thí dụ 12</u>**. Tính phương sai của $X \sim \mathcal{U}_{\{1,2,\dots,6\}}$ (xem thí dụ 3).

Giải: Ta đã có trong thí dụ 3 $EX = 3{,}5$. Bây giờ ta phải tính

$$E(X^2) = \frac{1}{6}(1 + 2^2 + 3^2 + \dots + 6^2) = \frac{91}{6}$$

$$\Rightarrow VX = E(X^2) - (EX)^2 = \frac{91}{6} - 3{,}5^2 = \frac{35}{12} \approx 2{,}9167.$$

* **<u>Thí dụ 13</u>. Hai xạ thủ A và B bắn loạt 100 viên mỗi người vào bia có các vòng tròn từ 5 đến 10 điểm. Tần suất bắn đạt các điểm của mỗi xạ thủ được cho trong bảng sau:**

Điểm X	5	6	7	8	9	10
A	0,01	0,03	0,05	0,12	0,39	0,40
B	0,02	0,04	0,07	0,10	0,25	0,52

Bạn có thể đánh giá gì về kết quả bắn của hai xạ thủ?

Giải: Có thể coi tần suất là xác suất xấp xỉ. Từ đó tính được nếu gọi A và B tương ứng là điểm của A và B:

$$EA = 9{,}05;\ EB = 9{,}08;\ E(A^2) = 83{,}05;\ E(B^2) = 84{,}02$$

$$\Rightarrow \qquad VA = 1{,}1475;\ VB = 1{,}5736.$$

Ta thấy điểm trung bình của A ($EA = 9{,}05$) bé hơn của B ($EB =$

9,08), tuy nhiên phương sai của A $VA = 1{,}1475$ nhỏ hơn $VB = 1{,}5736$ hay kết quả của A ổn định hơn B.

BÀI TẬP

1. Có 4 xạ thủ cùng bắn mỗi người một phát vào bia, XS trúng bia của các xạ thủ tương ứng là 0,9; 0,8; 0,7 và 0,6.
   a) Tìm PPXS của số đạn trúng. Sau đó lập hàm phân phối tương ứng và vẽ đồ thị của hàm.
   b) Tính xác suất để có không quá 2 người bắn trúng.

2. Một lô gồm 100 sản phẩm có 10 phế phẩm. Lấy ngẫu nhiên ra 5 sản phẩm (không hoàn lại). Lập bảng PPXS của số phế phẩm trong 5 sản phẩm đó.

3. Biến ngẫu nhiên X có bảng PPXS sau, λ là tham số,

x	1	2	3	4	5
$p(x)$	2λ	2λ	3λ	λ	2λ

   a) Xác định tham số λ.
   b) Lập hàm phân phối của X và vẽ đồ thị.
   c) Tìm giá trị k nhỏ nhất sao cho $P(X \leq k) > 0{,}5$.

4. Cho biến ngẫu nhiên X có bảng PPXS

x	1	3	5	7	9
$p(x)$	0,1	0,2	0,3	0,3	0,1

   Tìm PPXS của $Y = \min\{X, 6\}$.

5. Trọng lượng (g) của một chất trong 200 gói cho kết quả

Trọng lượng	47	48	49	50	51
Số gói	17	36	98	35	14

Tìm PPXS của trọng lượng trên, sau đó tính kỳ vọng và phương sai của trọng lượng đó.

6. Chữa bệnh cho 40 người, người ta thống kê được thời gian khỏi bệnh như sau:
   - có 12 người khỏi sau 7 ngày điều trị,
   - có 15 người khỏi sau 8 ngày điều trị,
   - có 13 người khỏi sau 9 ngày điều trị.

   Tính kỳ vọng và phương sai của thời gian điều trị khỏi bệnh của nhóm bệnh nhân trên.

7. Một người có một chùm chìa khoá 7 chiếc, trong đó chỉ có hai chiếc mở được cửa. Người đó thử ngẫu nhiên từng chiếc cho đến khi mở được cửa. Gọi X là số lần thử, tìm PPXS của X và tính EX, VX.

8. Trong một chiếc hộp có 4 tấm thẻ được đánh số từ 1 đến 4. Chọn ngẫu nhiên ra 2 tấm thẻ và gọi X là tổng của hai số trên hai thẻ. Tìm PPXS của X.

9. Một người chọn ngẫu từng sản phẩm từ một lô hàng (có số lượng sản phẩm rất lớn) cho đến khi được một phế phẩm thì dừng. Gọi X là số lần cần chọn. Tìm PPXS của X, biết tỷ lệ phế phẩm của lô hàng là 10%. Giả sử có 100 người cùng chọn như vậy, hỏi có khoảng bao nhiêu người phải chọn ít nhất 3 lần?

10. Một xạ thủ có 5 viên đạn. Anh ta được yêu bắn cho đến khi có 2 viên đạn trúng thì dừng. Gọi X là số đạn phải bắn, biết XS bắn trúng mỗi lần là 0,4. Tìm PPXS của X, sau đó tìm kỳ vọng và phương sai của X.

11. Có 10 sản phẩm, trong đó có 2 phế phẩm. Lấy lần lượt ra 3 sản phẩm (không hoàn lại).
    a) Gọi X là số chính phẩm trong 3 sản phẩm đó. Tìm PPXS của X, sau đó tìm kỳ vọng và phương sai của X.

b) Gọi Y là số phế phẩm trong 3 sản phẩm lấy ra, tìm EY và VY.

12. Một người bỏ ra x\$ để tham gia trò chơi tung con xúc xắc 3 lần. Nếu cả 3 lần đều xuất hiện mặt lục thì thu về 36\$, nếu 2 lần xuất hiện mặt lục thì thu về 2,8\$, còn nếu 1 lần xuất hiện mặt lục thì thu về 0,4\$.
a) Tìm x sao cho trò chơi là vô thưởng vô phạt.
b) x bằng bao nhiêu thì trung bình mỗi lần chơi sẽ mất 1\$?

13. Hai cầu thủ thay nhau ném bóng vào rổ cho đến khi nào trúng rổ thì dừng ném, biết rằng XS ném trúng mỗi lần của từng người tương ứng là 0,4 và 0,6. Tìm PPXS của:
a) số lần ném của cầu thủ thứ nhất;
b) số lần ném của cả hai cầu thủ.

14. XS bắn trúng của một khẩu súng là p. Tiến hành bắn liên tiếp trong điều kiện như nhau cho đến khi trúng thì dừng bắn. Tìm số đạn trung bình phải bắn.

§2 BIẾN NGẪU NHIÊN LIÊN TỤC

2.1 Biến ngẫu nhiên liên tục và hàm phân phối

* *Khái niệm*

Khái niệm *biến ngẫu nhiên liên tục* đã được đưa ra ở tiểu mục 1.1. Ta nhắc lại các khái niệm biến liên tục và *hàm phân phối XS*.

- Biến ngẫu nhiên được gọi là *liên tục*, nếu tập giá trị của nó lấp kín một khoảng (đoạn) hoặc một số khoảng trên trục số hoặc cũng có thể là cả trục số (tập không đếm được).

- *Hàm phân phối xác suất* (*probability distribution function*) của một biến ngẫu nhiên X, ký hiệu là $F(x)$, được xác định như sau

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

Ý nghĩa: $F(x)$ phản ánh độ tập trung XS ở bên trái điểm x

$\Rightarrow$ *hàm phân phối tích luỹ (xác suất tích luỹ)*

$\Rightarrow$ *hàm phân phối tích phân.*

- *Tính chất*

(i) $1 \geq F(x) \geq 0$;

(ii) $F(x)$ là hàm không giảm $\Leftrightarrow \forall\, x_2 > x_1 \Rightarrow F(x_2) \geq F(x_1)$;

(iii) $P(\alpha \leq X < \beta) = F(\beta) - F(\alpha)$;

(iv) $F(+\infty) = 1$; $F(-\infty) = 0$.

Hạn chế: + khó biết nơi nào tập trung nhiều XS hơn,

+ sau này ta thấy không thể biểu diễn $F(x)$ dưới dạng hàm sơ cấp trong một số trường hợp.

*** Thí dụ 1**. Cho hàm phân phối của biến X liên tục có dạng

$$F(x) = \begin{cases} 0; & x \leq a, \\ c(x-a); & a < x \leq b, \\ 1; & x > b. \end{cases}$$

Tìm hằng số c để hàm $F(x)$ liên tục.

Giải: Dùng kiến thức của giải tích toán ta có ngay $c = \dfrac{1}{b-a}$. Dễ dàng kiểm chứng hàm $F(x)$ này thoả mãn các tính chất (i) – (iv) ở trên.

Từ tính chất (iii) nếu X là biến ngẫu nhiên liên tục

$$P(X = \alpha) = 0, \forall \alpha \in \mathbb{R}.$$

Trong trường hợp này:

$$P(\alpha \leq X < \beta) = P(\alpha < X < \beta) = P(\alpha < X \leq \beta) = P(\alpha \leq X \leq \beta).$$

*** Thí dụ 2**. Cho hàm phân phối của biến X liên tục có dạng

$$F(x) = \begin{cases} 0; & x \leq 2, \\ \alpha(x-2)^2; & 2 < x \leq 4, \\ 1; & x > 4. \end{cases}$$

Xác định hằng số α và tính $P(2 < X < 3)$.

Giải: Do $F(x)$ liên tục nên tại $x = 4$ phải có $\alpha(x-2)^2 = 1$, từ đó $\alpha = 1/4$. Sử dụng tính chất (iii) của hàm $F(x)$ ta có

$$P(2 < X < 3) = F(3) - F(2) = 1/4.$$

2.2 Hàm mật độ

Từ hạn chế của hàm PP ta không xác định được XS sẽ tập trung nhiều ở chỗ nào trên trục số $\Rightarrow$

* ***Định nghĩa.*** *Hàm mật độ xác suất* (*probability density function*), ký hiệu là $f(x)$, của biến ngẫu nhiên X có hàm phân phối $F(x)$, được xác định từ biểu diễn sau

$$F(x) = \int_{-\infty}^{x} f(t)dt\,. \tag{1}$$

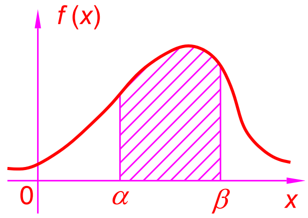
*** Tính chất:**

(i) $f(x) \geq 0$;

(ii) Nếu $F(x)$ khả vi $\Rightarrow$ $f(x) = F'(x)$;

(iii) $P(\alpha < X < \beta) = \int_{\alpha}^{\beta} f(x)dx$;

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



$$P(\alpha < X < \beta)$$

*** <u>Thí dụ 3</u>**. Xét hàm $F(x)$ trong thí dụ 1. Tìm hàm mật độ của biến ngẫu nhiên X tương ứng.

Giải: Sử dụng tính chất (ii) của hàm $F(x)$, ta có

$$f(x) = \begin{cases} \frac{1}{b-a}, & x \in (a;b) \\ 0, & x \notin (a;b) \end{cases}.$$

Để ý rằng $f(x)$ luôn nhận giá trị hằng (khác 0) trên $(a;b)$, ta nói rằng X có *phân phối đều* (*liên tục*) *trên* $(a;b)$ và ký hiệu

$$X \sim \mathcal{U}(a;b).$$

$$X \sim \mathcal{U}(0;1) \quad \Leftrightarrow \quad f(x) = \begin{cases} 1, x \in (0;1) \\ 0, x \notin (0;1) \end{cases},$$

$$\Leftrightarrow \quad F(x) = \begin{cases} 0; & x \le 0, \\ x; & 0 < x \le 1, \\ 1; & x > 1. \end{cases}$$

khi $a = 0, b = 1$, hàm $f(x) \equiv 1$ trên $(0;1)$; trong các ngôn ngữ lập trình bậc cao đã xây dựng lệnh máy để tạo ra các số ngẫu nhiên có phân phối đều trên $(0;1)$ để tính toán mô phỏng (*simulation*).

*** <u>Thí dụ 4</u>**. Cho hàm mật độ của biến ngẫu nhiên X có dạng

$$f(x) = \begin{cases} a\cos x; & x \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right) \\ 0; & x \notin \left(-\frac{\pi}{2}, \frac{\pi}{2}\right) \end{cases}.$$

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

b) Tính XS $P(\frac{\pi}{4} < X < \pi)$.

Giải: a) Dùng tính chất (iv) của hàm mật độ

$$\int_{-\infty}^{+\infty} f(x)dx = a\int_{-\pi/2}^{\pi/2} \cos x \, dx = 2a = 1$$

$$\Longrightarrow \qquad a = 1/2.$$

Dựa vào (1) ta có:

với $x \le -\frac{\pi}{2}$, $\int_{-\infty}^{x} f(x)dx = 0$;

với $x > \frac{\pi}{2}$, $\int_{-\infty}^{x} f(x)dx = \int_{-\pi/2}^{\pi/2} f(x)dx = 1$;

với $-\frac{\pi}{2} < x \le \frac{\pi}{2}$, $\int_{-\infty}^{x} f(x)dx =$

$$= \int_{-\infty}^{-\pi/2} f(x)dx + \int_{-\pi/2}^{x} f(x)dx = \frac{1}{2}(\sin x + 1).$$

Từ đó

$$F(x) = \begin{cases} 0; & x \le -\frac{\pi}{2}, \\ \frac{1}{2}(\sin x + 1); & -\frac{\pi}{2} < x \le \frac{\pi}{2}, \\ 1; & x > \frac{\pi}{2}. \end{cases}$$

b) Theo tính chất (iii) của hàm phân phối

$$P\left(\frac{\pi}{4} < X < \pi\right) = F(\pi) - F\left(\frac{\pi}{4}\right) = 1 - \frac{\sqrt{2}}{4}.$$

*** <u>Thí dụ 5</u>**. Cho XS phân rã của một nguyên tử chất phóng xạ trong khoảng thời gian đủ bé dt là λdt (giả sử việc phân rã đó không phụ thuộc vào quá khứ). Hãy xác định:

a) XS để nguyên tử đó phân rã trong khoảng thời gian t,

b) hàm mật độ XS của thời điểm phân rã của nguyên tử.

Giải:

a) XS không phân rã của một nguyên tử chất phóng xạ trong khoảng thời gian dt dễ thấy là $1 - \lambda dt$. Chia khoảng thời gian t thành t/dt các khoảng con có độ dài dt; từ đó XS để nguyên tử không phân rã trong khoảng thời gian đó xấp xỉ là (do có giả thiết độc lập) $(1 - \lambda dt)^{t/dt}$. Lấy giới hạn khi $dt \to 0$, ta có XS cần tìm

$$(1 - \lambda dt)^{t/dt} \to 1 - e^{-\lambda t}$$

(để ý $e^{-\lambda t}$ là XS nguyên tử không phân rã trong khoảng thời gian t).

b) Gọi T là thời điểm phân rã của nguyên tử và $f(t)$ là hàm mật độ của T. XS để nguyên tử phân rã trong khoảng thời gian từ t đến $t + dt$ sẽ bằng XS không phân rã trong khoảng thời gian t trước đó nhân với XS phân rã trong khoảng thời gian dt, từ đó

$$P(t < T < t + dt) = f(t)dt = e^{-\lambda t}\lambda dt.$$

Vậy ta có
$$f(t) = \begin{cases} 0, & t \leq 0; \\ \lambda e^{-\lambda t}, & t > 0. \end{cases}$$

Đây chính là hàm mật độ của biến ngẫu nhiên tuân theo *luật phân phối mũ*, ký hiệu $T \sim \mathcal{E}(\lambda)$, λ là tham số.

2.3 Các số đặc trưng cơ bản

A. Kỳ vọng

Ta xét *phép tính kỳ vọng* trong trường hợp biến liên tục. Giả thiết rằng X có hàm mật độ XS $f(x)$. Khi đó kỳ vọng của $Z = g(X)$ được định nghĩa như sau

*** Định nghĩa 1.** Kỳ vọng của $Z = g(X)$, với $f(x)$ là hàm mật độ đã cho của biến X, được tính theo công thức

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

Trường hợp riêng khi $Z = X$, ta có

*** Định nghĩa 2.** *Kỳ vọng của X*, với $f(x)$ là hàm mật độ XS đã cho của biến *X*, được tính như sau

$$EX = \int_{-\infty}^{+\infty} xf(x)dx. \qquad (2)$$

*** Tính chất**

(i) $E(c) = c,\ c = const$;

(ii) $E(cX) = cEX$;

(iii) $E(X + Y) = EX + EY$;

(iv) $E(XY) = EX.EY$, *X* và *Y* độc lập.

Chú ý: tính chất (ii) và (iii) được gọi là *tính tuyến tính* của phép toán, vốn quen thuộc trong phép tính đạo hàm hay tích phân.

Ý nghĩa thực tế: kỳ vọng là *giá trị trung bình* (*mean value*) của biến ngẫu nhiên hoặc đóng vai trò *định vị* của biến.

*** Thí dụ 6**. Xét lại thí dụ 3 và cần tính kỳ vọng của *X*.

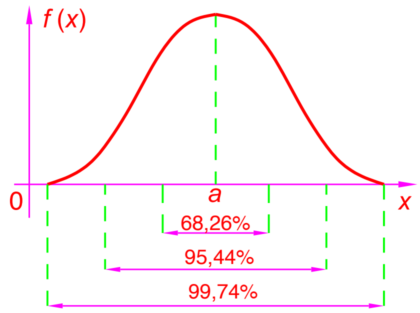
Giải: Do $X \sim \mathcal{U}(a, b)$ nên sử dụng (2)

$$EX = \int_{-\infty}^{+\infty} xf(x)dx = \frac{1}{b-a}\int_a^b xdx = \frac{a+b}{2}.$$

*** Thí dụ 7**. Biến ngẫu nhiên *X* được cho là tuân theo *luật phân phối chuẩn* (*normal distribution*) hay *luật PP Gauss*, ký hiệu là

$X \sim \mathcal{N}(\mu, \sigma^2)$, nếu hàm mật độ của nó dạng

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



Nếu $X \sim \mathcal{N}(0; 1)$, thì hàm Gauss $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ chính là mật độ của biến X. Hàm Laplace $\phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-t^2/2}\, dt$ - tích phân của X và ta có (*tích phân Euler-Poisson*)

$$\int_0^{+\infty} e^{-x^2/2}\, dx = \sqrt{2\pi}/2.$$

Dễ dàng tính được $EX = \mu$. Để ý hàm e^{-x^2} không có nguyên hàm dạng hàm sơ cấp nên không viết được hàm PP dưới dạng các hàm sơ cấp. Có thể chứng minh, nếu $X \sim \mathcal{N}(\mu, \sigma^2)$, bằng đổi biến $y = (t-\mu)/\sigma$ hay $t = \mu + \sigma y$

$$P(\alpha < X < \beta) = \frac{1}{\sigma\sqrt{2\pi}} \int_\alpha^\beta e^{-\frac{(t-\mu)^2}{2\sigma^2}}\, dt = \frac{1}{\sqrt{2\pi}} \int_{\frac{\alpha-\mu}{\sigma}}^{\frac{\beta-\mu}{\sigma}} e^{-\frac{y^2}{2}}\, dy$$

$$= \phi\left(\frac{\beta-\mu}{\sigma}\right) - \phi\left(\frac{\alpha-\mu}{\sigma}\right). \qquad (3)$$

*** <u>Thí dụ 8</u>**. Gọi X là độ dài một chi tiết do một máy tự động sản xuất ra và giả sử $X \sim \mathcal{N}(\mu, \sigma^2)$, với μ được coi là độ dài quy định (trung bình), $\sigma = 2$. Tính XS để độ dài lệch so với quy định không vượt quá $\varepsilon = 3$.

Giải: Ta cần tính $P(|X - \mu| < \varepsilon) = P(\mu - \varepsilon < X < \mu + \varepsilon)$, áp dụng công thức (3)

$$\Rightarrow \quad P(\mu - \varepsilon < X < \mu + \varepsilon)$$

$$= \phi\left(\frac{\varepsilon}{\sigma}\right) - \phi\left(\frac{-\varepsilon}{\sigma}\right) = 2\phi\left(\frac{\varepsilon}{\sigma}\right) = 2\phi(1{,}5) = 2.0{,}4332 = 0{,}8664.$$

Nếu coi ε là dung sai cho phép, ý nghĩa thực tế của kết quả 86,64% chính là tỷ lệ chính phẩm của máy đã cho.

*** <u>Thí dụ 9</u>**. Cho biến ngẫu nhiên X có hàm mật độ

$$f(x) = \begin{cases} kx^2; & 1 \geq x \geq 0 \\ 0; & x \text{ khác} \end{cases}.$$

a) Tìm hằng số k.

b) Xét biến ngẫu nhiên $Y = 2\sqrt{X}$. Tính *các XS*: $P(0{,}5 < Y < 1{,}5)$ và $P(Y > 1)$.

Giải:

a) Ta có theo tính chất (iv) của hàm mật độ

$$k\int_0^1 x^2\,dx = k/3 = 1 \;\Rightarrow\; k = 3.$$

b) Từ đó $P(0{,}5 < Y < 1{,}5) = P(1/16 < X < 9/16)$

$$= 3\int_{1/16}^{9/16} x^2\,dx = 91/512;$$

$$P(Y > 1) = P(X > 1/4) = 3\int_{0{,}25}^{1} x^2\,dx = 63/64.$$

B. Phương sai

*** Định nghĩa.** Phương sai của biến ngẫu nhiên *X*, ký hiệu là *VX*, được xác định như sau

$$VX = E[(X - EX)^2]. \qquad (4)$$

Nhận xét: $X - EX$ là độ lệch của X so với trung bình của nó

⇒ Phương sai – trung bình của bình phương độ lệch

⇒ Đặc trưng cho độ phân tán của X quanh trung bình

độ bất định; độ rủi ro ...

Chú ý: Công thức (4) có dạng tương đương

$$VX = E(X^2) - (EX)^2.$$

Cách tính: $VX = \int_{-\infty}^{+\infty} (x - EX)^2 f(x)dx$

$$= \int_{-\infty}^{+\infty} x^2 f(x)dx - (EX)^2.$$

*** Tính chất**

(i) $V(c) = 0$, $c = const$;

(ii) $V(cX) = c^2 VX$;

(iii) $V(X + Y) = VX + VY$, với X và Y độc lập.

*** <u>Thí dụ 10</u>**. Tính phương sai của $X \sim \mathcal{U}(a, b)$ (xem thí dụ 3).

Giải: Ta đã có trong thí dụ 3 $EX = (a + b)/2$. Bây giờ ta phải tính

$$E(X^2) = \int_{-\infty}^{+\infty} x^2 f(x)dx = \frac{1}{b-a}\int_a^b x^2 dx = \frac{b^3 - a^3}{3(b-a)} = \frac{b^2 + ab + a^2}{3},$$

Từ đó

$$\Rightarrow VX = E(X^2) - (EX)^2 = \frac{b^2 + ab + a^2}{3} - \frac{(a+b)^2}{4} = \frac{(a-b)^2}{12}.$$

*** <u>Thí dụ 11</u>**. Cho biến ngẫu nhiên X có hàm mật độ

$$f(x) = \begin{cases} ae^{-x^2}; & x \geq 0 \\ 0; & x < 0 \end{cases}.$$

Xác định hằng số a; sau đó tính EX và VX.

Giải: Tìm a bằng cách dùng tính chất (iv) của hàm mật độ

$$\int_{-\infty}^{+\infty} f(x)dx = 1 = a\int_0^{+\infty} e^{-x^2}dx$$

$$= \frac{a}{\sqrt{2}}\int_0^{+\infty} e^{-t^2/2}dt = \frac{a}{\sqrt{2}}\,\frac{\sqrt{2\pi}}{2} = \frac{a\sqrt{\pi}}{2} \Rightarrow a = \frac{2}{\sqrt{\pi}}.$$

Bây giờ ta tính kỳ vọng và phương sai của X:

$$EX = \frac{2}{\sqrt{\pi}}\int_0^{+\infty} xe^{-x^2}dx = \frac{1}{\sqrt{\pi}}\int_0^{+\infty} e^{-x^2}d(x^2) = \frac{1}{\sqrt{\pi}},$$

$$VX = \frac{2}{\sqrt{\pi}}\int_0^{+\infty} x^2e^{-x^2}dx - (EX)^2 = \frac{\sqrt{2}}{2} - \frac{1}{\pi} = \frac{\pi\sqrt{2}-2}{2\pi}.$$

* **<u>Thí dụ 12</u>**. Chặt ngẫu nhiên một đoạn dây dài $1m$ thành hai đoạn con và lấy hai đoạn con đó làm hai cạnh của một hình chữ nhật. Tính diện tích trung bình của hình chữ nhật, sau đó tính phương sai của diện tích đó.

Giải: Gọi X là độ dài của một đoạn con, diện tích hình chữ nhật sẽ là $S = X(1-X)$. Ta tìm hàm phân phối của S, để ý $X \sim \mathcal{U}(0;1)$ và $0 < S \leq 0{,}25$,

$$F(s) = P(S < s) = P[X(1-X) < s] = P(X^2 - X + s > 0).$$

Với $0 < s \leq 0{,}25$

$$P(S < s) = P(\{x_1 < X\} \vee \{X > x_2\}),\ x_{1;2} = \frac{1 \pm \sqrt{1-4s}}{2}.$$

$$\Rightarrow F(s) = P(\{x_1 < X) + P(X > x_2) = F_X(x_1) + 1 - F_X(x_2).$$

Để ý hàm PP của X với $x \in (0;1)$ nhận giá trị $F_X(x) = x$, đồng thời do $0 < x_{1;2} < 1$, nên

$$F(s)=\begin{cases} 0, & s\le 0;\\ 1-\sqrt{1-4s}, & 0<s\le 0{,}25;\\ 1, & s>0{,}25.\end{cases}$$

$$\Leftrightarrow f(s)=\begin{cases} 2\sqrt{1-4s}, & s\in(0;0{,}25);\\ 0, & s\notin(0;0{,}25).\end{cases}$$

Từ đó:

$$ES=\int_{-\infty}^{+\infty} sf(s)ds=2\int_0^{0,25} s\sqrt{1-4s}\,ds=1/30\ (m);$$

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

$$=2\int_0^{0,25} s^2\sqrt{1-4s}\,ds - 1/900 = 1/840 - 1/900 = 1/12600.$$

BÀI TẬP

1. Cho biên độ dao động của thành tầu thuỷ là biến ngẫu nhiên X tuân theo *luật phân phối Releу* có hàm PPXS
$$F(x) = \begin{cases} 1 - e^{-x^2/2\sigma^2}; & x \geq 0 \\ 0; & x < 0 \end{cases},$$
trong đó σ là tham số đã biết. Tính xác suất để biên độ dao động trên lớn hơn trị trung bình của nó.

2. Cho biến ngẫu nhiên X có hàm mật độ
$$f(x) = \begin{cases} k(1+x)^{-3}; & x \geq 0 \\ 0; & x < 0 \end{cases}.$$
Tìm k, hàm PPXS và sau đó tính EX.

3. Cho biến ngẫu nhiên X tuân theo *luật phân phối Lô-ga chuẩn* với hàm mật độ
$$f(x) = \begin{cases} \frac{1}{\sigma x\sqrt{2\pi}} \exp\left[-\frac{(\ln x-\alpha)^2}{2\sigma^2}\right]; & x \geq 0 \\ 0; & x < 0 \end{cases},$$
trong đó α, σ là tham số đã biết. Tính EX.

4. Thời gian đi từ nhà đến trường của một sinh viên A là biến ngẫu nhiên T có phân phối chuẩn. Theo thống kê biết rằng 65% số ngày đến trường của A mất hơn 20 phút và 8% số ngày mất hơn 30 phút.
   a) Tính trị trung bình và phương sai của thời gian đến trường của A.
   b) Giả sử A xuất phát ở nhà trước giờ vào học 25 phút, tính xác suất để A muộn học.
   c) A cần xuất phát trước giờ học bao nhiêu phút để XS bị muộn học của A bé hơn 0,02?

Cho: $F(-0{,}385)$ =0,35; $F(0{,}695)$= 0,51; $F(1{,}405)$=0,92; $F(2{,}054)$=0,98; $F(x)$ là hàm PP của biến $X \sim \mathcal{N}(0,1)$.

Chú ý: Trong (13) tiết 4 chương I có thể hiệu chỉnh x_j như sau:

$$x_2 = \frac{k_2 - np + 0,5}{\sqrt{npq}}; \qquad x_1 = \frac{k_1 - np - 0,5}{\sqrt{npq}}.$$

5. Cho biến ngẫu nhiên X có hàm mật độ

$$f(x) = \frac{k}{e^x + e^{-x}}, x \in \mathbb{R}.$$

   a) Tìm hằng số k và hàm PPXS.
   b) Phải quan sát X bao nhiêu lần để thấy có ít nhất một lần X có giá trị rơi vào khoảng $(\ln\frac{1}{\sqrt{3}}, \ln\sqrt{3})$ với XS 0,9?

6. Cho biến ngẫu nhiên có phân phối mũ với tham số $\lambda = 1$. Xét biến $Y = 2X^2$. Tính: a) $P(2 < Y < 18)$, b) $P(Y < 4)$.

7. Lãi suất đầu tư vào một dự án được coi là biến ngẫu nhiên tuân theo luật chuẩn. Theo đánh giá của các chuyên gia thì với XS 0,1587 dự án cho lãi suất lớn hơn 20% và với XS 0,0288 cho lãi suất lớn hơn 25%. Vậy khả năng đầu tư không bị lỗ là bao nhiêu?

8. Trong một kỳ thi điểm số trung bình của các sinh viên là 80 và phương sai là 100. Giả sử điểm thi đó tuân theo luật phân phối chuẩn.
   a) Nếu giáo viên muốn 25% sinh viên đạt điểm A (nhóm điểm cao nhất) thì điểm số thấp nhất của nhóm phải là bao nhiêu?
   b) Nếu chọn ngẫu nhiên 50 sinh viên, tính XS trong đó có nhiều hơn 10 sinh viên đạt điểm A (điểm A ở câu a).

9. Cho $X \sim \mathcal{N}(0,1)$ và xét $Y = a + bX + cX^2$, a, b, c là các hằng số. Tính EY và VY.

10. Chặt ngẫu nhiên một đoạn dây có độ dài a và dùng một đoạn con làm bán kính của một mặt tròn. Tìm trị trung bình và phương sai của diện tích mặt tròn đó.

11. Cho biến ngẫu nhiên X có hàm mật độ

$$f(x) = ke^{-\lambda|x|}, x \in \mathbb{R}.$$

a) Tìm k và hàm PPXS của X.

b) Tính kỳ vọng và phương sai của X.

12. Cho một phần tư mặt tròn tâm O(0,0) bán kính bằng a, ký hiệu là OAB, với toạ độ tương ứng A(a,0) và B(0,a). Trên đoạn OA lấy ngẫu nhiên một điểm C, dựng một đường thẳng đi qua C vuông góc với OA và cắt cung tròn tại điểm D. Tính kỳ vọng và phương sai của độ dài đoạn CD.

§3 CÁC SỐ ĐẶC TRƯNG CỦA BIẾN NGẪU NHIÊN

3.1 Kỳ vọng (*Expectation value*)

*** Định nghĩa.** *Kỳ vọng của X*, với $f(x)$ là hàm mật độ XS đã cho của X liên tục hoặc $p(x)$ là hàm xác suất của X rời rạc, được tính như sau:

$$EX = \sum_{\forall x} xp(x),\ X \text{ rời rạc;} \qquad (1a)$$

hoặc $$EX = \int_{-\infty}^{+\infty} xf(x)dx,\ X \text{ liên tục.} \qquad (1b)$$

*** Tính chất**

(i) $E(c) = c,\ c = const$;

(ii) $E(cX) = cEX$;

(iii) $E(X + Y) = EX + EY$;

(iv) Nếu X và Y độc lập $\Longrightarrow E(XY) = EX.EY$.

Chú ý: tính chất (ii) và (iii) được gọi là *tính tuyến tính* của phép toán, vốn quen thuộc trong phép tính đạo hàm hay tích phân.

Ý nghĩa thực tế: kỳ vọng là *giá trị trung bình* (*mean value*) của biến ngẫu nhiên và đóng vai trò *định vị* biến.

*** <u>Thí dụ 1</u>**. Cho biến ngẫu nhiên $X \sim \mathcal{B}(n, p)$. Tính kỳ vọng của X.

Giải. Sử dụng công thức Bernoulli ta có

$$EX = \sum_{\forall x} xp(x) = \sum_{x=0}^{n} xC_n^x p^x q^{n-x} = np.$$

Việc tính tổng ở trên không đơn giản. Ta có cách làm khác dễ hơn như sau: Do $X = X_1 + X_2 + \dots X_n$, trong đó $X_i \sim \mathcal{B}(1,p)$ (*PPXS Bernoulli*) với $p(0) = P(X = 0) = q$ và $p(1) = p$ và bảng PPXS

$X = x$	0	1
$p(x)$	q	p

Dễ thấy $EX_i = p, \forall i \quad \Rightarrow \quad EX = \sum_{i=1}^{n} EX_i = np$ (dùng tính chất (iii) ở trên).

* **<u>Thí dụ 2</u>**. Cho biến ngẫu nhiên $X \sim \mathcal{E}(\lambda)$. Tính kỳ vọng của X.

Giải. $$EX = \int_0^{+\infty} x\,\lambda e^{-\lambda x} dx = \frac{1}{\lambda}.$$

* **<u>Thí dụ 3</u>**. Một người mua 10000 đồng một số đề. Tính số tiền thắng trung bình trong lần chơi đó.

Giải. Gọi X là số tiền thắng trong lần chơi, rõ ràng

$X = x$	0	$700000đ$
$p(x)$	99%	1%

và số tiền thắng trung bình $EX = 7000$ đồng.

*** Thí dụ 4**. Một người hàng ngày đi bộ từ nhà đến nơi làm việc trên quãng đường dài 600 m với vận tốc đều $V\, m/s$. Biết thời gian đi bộ của người đó là một biến ngẫu nhiên có phân bố đều trong khoảng từ 6 phút đến 10 phút. Tìm kỳ vọng của X.

Giải. Gọi T là thời gian đi bộ ở trên, rõ ràng $T \sim \mathcal{U}(6; 10)$ và

$V = \dfrac{600}{60T} = \dfrac{10}{T}\,(m/s)$. Để ý $f_T(x) = \frac{1}{4}$, $6 < t < 10$, từ đó

$$EV = \frac{1}{4}\int_6^{10} \frac{10dt}{t} = \frac{5}{2}\ln\frac{10}{6} \approx 1{,}277\ (m/s).$$

3.2 Phương sai (*Variance value*)

*** Định nghĩa 1.** Phương sai của biến ngẫu nhiên X, ký hiệu là VX, được xác định như sau

$$VX = \mathrm{var}(X) = E[(X - EX)^2]. \qquad (2)$$

Nhận xét: $X - EX$ là độ lệch của X so với trung bình của nó

⇒ Phương sai – trung bình của bình phương độ lệch

⇒ Đặc trưng cho độ phân tán của X quanh trung bình

độ bất định; dung sai; độ rủi ro …

Chú ý: Công thức (2) có dạng tương đương

$$VX = E(X^2) - (EX)^2.$$

Cách tính:

- Với X rời rạc

$$VX = \sum_{\forall x}(x - EX)^2\, p(x) = \sum_{\forall x} x^2\, p(x) - (EX)^2;$$

- Với X liên tục

$$VX = \int_{-\infty}^{+\infty}(x - EX)^2 f(x)dx = \int_{-\infty}^{+\infty} x^2 f(x)dx - (EX)^2.$$

*** Tính chất**

(i) $V(c) = 0$, $c = const$;

(ii) $V(cX) = c^2 VX$;

(iii) Nếu X và Y độc lập $\Longrightarrow V(X + Y) = VX + VY$.

*** <u>Thí dụ 5</u>**. Tính phương sai của biến ngẫu nhiên $X \sim \mathcal{B}(n, p)$ (xem thí dụ 1).

Giải: Ta đã có trong thí dụ 3 $EX = np$. Có thể tính trực tiếp

$$VX = \sum_{\forall x}(x - EX)^2\, p(x) = \sum_{x=0}^{n}(x - EX)^2\, C_n^x p^x q^{n-x}.$$

Tuy nhiên ta có thể dùng cách tiếp cận ở thí dụ 1. Từ ý nghĩa thực tế ta có các X_i độc lập và dễ thấy $E(X_i^2) = p$, $VX_i = pq$, nên

$$\Rightarrow VX = \sum_{i=1}^{n} VX_i = npq.$$

*** Thí dụ 6**. Tìm kỳ vọng của biến ngẫu nhiên X trong thí dụ 4.

Giải. Từ thí dụ 4 ta có

$$EV = \frac{1}{4}\int_6^{10} \frac{100\,dt}{t^2} = \frac{5}{3}$$

$$\Rightarrow \operatorname{var}(V) = E(V^2) - (EV)^2 \approx 0{,}0358.$$

Để ý VX là một số không âm, tuy nhiên về mặt vật lý nó không cùng thứ nguyên với X, vì vậy ta đưa ra khái niệm sau:

*** Định nghĩa 2.** *Độ lệch chuẩn* (*standard deviation*) của biến ngẫu nhiên X, ký hiệu là $\sigma(X)$, được xác định bằng

$$\sigma(X) = \sqrt{VX} \qquad (3)$$

Từ (3) ta có thể ký hiệu phương sai là $\sigma^2(X)$ hoặc đơn giản σ^2.

Chú ý: $\sigma(cX) = |c|\,\sigma(X)$

và nếu X và Y độc lập thì $\sigma(X+Y) = \sqrt{\sigma^2(X) + \sigma^2(Y)}$.

Một hệ quả quan trọng của tính chất (iii) của VX: Nếu ta có n biến ngẫu nhiên độc lập $X_i, i = \overline{1;n}$ và $VX_i = \sigma^2\ \forall i = \overline{1;n}$, thì

$$V\bar{X} = V\left(\frac{X_1 + X_2 + \ldots + X_n}{n}\right) = \frac{\sigma^2}{n}.$$

3.2 Một số đặc số khác

1. Mốt (*mode*)

Mốt là giá trị của biến ngẫu nhiên có khả năng xuất hiện lớn nhất (trong một lân cận nào đó).

- Biến rời rạc: mốt là giá trị có XS lớn nhất.
- Biến liên tục: giá trị làm hàm mật độ đạt *max*.

$\Rightarrow$ Biến ngẫu nhiên có thể có nhiều mốt.

*** <u>Thí dụ 7</u>**. Tìm mốt của biến $X \sim \mathcal{B}(n, p)$

Giải. Mốt của X là phần nguyên của số thực $(n+1)p$. Nếu số này là số nguyên thì ta có hai mốt là $(n+1)p$ và $(n+1)p-1$. Trong ứng dụng mốt còn có tên gọi là *số lần xuất hiện chắc nhất* (của A trong lược đồ Bernoulli tương ứng).

*** <u>Thí dụ 8</u>**. Tìm mốt của biến X có *phân phối Weibull* với hàm mật độ

$$f(x) = \begin{cases} \frac{x}{2}\, e^{-x^2/4}; & x \geq 0 \\ 0; & x < 0 \end{cases}.$$

Giải. Mốt của X là nghiệm của phương trình

$$f'(x) = \frac{1}{2}\ e^{-x^2/4} - \frac{x^2}{4}\ e^{-x^2/4} = 0.$$

Từ đó mốt là nghiệm của $1 - x^2/2 = 0$, nhưng do $x > 0$ suy ra

$\mathrm{mod}(X) = \sqrt{2} \approx 1{,}414.$

2. Trung vị (median)

Nếu ký hiệu trung vị là med$X$, thì

$$P(X < \ medX) = P(X \geq \ medX) = 1/2.$$

Có nghĩa là để tìm trung vị ta phải phải giải phương trình

$$F(x) = 1/2.$$

Chú ý cả mốt và trung vị đều có ý nghĩa định vị biến, trong một số trường hợp còn hay hơn kỳ vọng. Chẳng hạn xét tập điểm số của bốn lần kiểm tra giữa kỳ {2, 2, 2, 10}. Trung bình ở đây bằng 4, có vẻ mốt cho đánh giá chính xác hơn?

*** <u>Thí dụ 9</u>**. Tìm trung vị của biến X có *phân phối Weibull*.

Giải. Rõ ràng trung vị là nghiệm của phương trình

$$\int_0^{\mathrm{med}X} f(x)dx = 0{,}5 \text{ hay } 1 - e^{-(\mathrm{med}X)^2/4} = 0{,}5;$$

$$\Rightarrow \quad \mathrm{med}X = 1{,}665.$$

Nói chung 3 đặc trưng kỳ vọng, mốt và trung vị không bằng nhau; ở đây với biến X có phân phối Weibull ta thấy $EX = 1{,}772$; mod$X = 1{,}414$; med$X = 1{,}665$. Trong trường hợp phân phối đối xứng (và đồ thị mật độ chỉ có 1 đỉnh) thì cả ba đặc trưng trùng nhau.

3. Phân vị (percentile)

Ta gọi x_α là *phân vị* α của X, nếu $P(X < x_\alpha) = \alpha \Leftrightarrow F(x_\alpha) = \alpha$. Số thực α được gọi là bậc của phân vị. Trong xác suất người ta quan tâm nhiều đến các bậc phân vị 25%, 50% (trung vị) và 75% (còn gọi là các tứ phân vị); còn trong thống kê hay sử dụng các phân vị có bậc lớn hoặc bé (chẳng hạn 90%, 95%, 99%, ... hoặc 0,5%, 1%, 5%, ...).

*** Thí dụ 10**. Tìm các phân vị 25%, 50%, 95%, 99% của biến ngẫu nhiên $X \sim \mathcal{N}(0; 1)$.

Giải. Sử dụng bảng hàm Laplace ta sẽ có:

$F(x_{0,25}) = 0{,}25 \Rightarrow \phi(-x_{0,25}) = 0{,}25 \Rightarrow x_{0,25} = -0{,}675;$

$F(x_{0,5}) = 0{,}5 \Rightarrow x_{0,5} = 0 = \mathrm{mod}(X);$

$F(x_{0,95}) = 0{,}95 \Rightarrow \phi(x_{0,95}) = 0{,}45 \Rightarrow x_{0,95} = 1{,}645;$

$F(x_{0,99}) = 0{,}99 \;\Rightarrow\; \phi(x_{0,99}) = 0{,}49 \;\Rightarrow\; x_{0,95} = 2{,}325.$

4. Mô men (moment)

Mô men cấp k đối với a của biến ngẫu nhiên X, ký hiệu $\nu_k(a)$ là một số thực được xác định như sau

$$\nu_k(a) = E[(X-a)^k].$$

Nếu $a = 0$, ký hiệu $\nu_k = \nu_k(0) = E(X^k)$, là *mô men gốc cấp k*.

Nếu $a = EX$, ký hiệu $\mu_k = \mu_k(EX) = E[(X-EX)^k]$, là *mô men trung tâm cấp k*.

$$\Rightarrow \quad EX = \nu_1;\, VX = \mu_2.$$

Có thể thấy quan hệ giữa hai loại mô men:

$$\mu_2 = \nu_2 - \nu_1^2 = \sigma^2,$$

$$\mu_3 = \nu_3 - 3\nu_1\nu_2 + 2\nu_1^3,$$

$$\mu_4 = \nu_4 - 4\nu_1\nu_3 + 6\nu_1^2\,\nu_2 - 3\nu_1^4, \ldots$$

Một số mô men cho ta đặc trưng về hình dạng phân phối XS:

- *Hệ số bất đối xứng* (*skewness*) là tỷ số $\beta_1 = \mu_3/\sigma^3$; nếu $\beta_1 = 0$ ta có đường cong mật độ đối xứng, còn nếu nó khác không thì phân phối bất đối xứng.
- *Hệ số nhọn* (*kurtosis* hay *peakedness*) là tỷ số $\beta_2 = \mu_4/\sigma^4 - 3$;

nếu β_2 càng lớn thì đường cong mật độ có đỉnh càng nhọn;

đường cong mật độ chuẩn có $\beta_2 = 0$.

* **Thí dụ 11**. Cho biến ngẫu nhiên $X \sim \mathcal{E}(0{,}5)$. Tìm các hệ số bất đối xứng và hệ số nhọn.

Giải. Dễ dàng tìm được $EX = \nu_1 = 2$; $\sigma^2 = \mu_2 = 4$. Từ đó suy ra:

$$\beta_1 = \mu_3/\sigma^3 = 16/8 = 2;$$

$$\beta_2 = \mu_4/\sigma^4 - 3 = 208/16 - 3 = 10.$$

Chú ý khi tính các tích phân ta sử dụng công thức

$$\int_0^\infty x^n e^{-ax}\,dx = \frac{\Gamma(n+1)}{a^{n+1}} = \frac{n!}{a^{n+1}}.$$

* **Thí dụ 12**. Cho biến ngẫu nhiên X có hàm mật độ

$$f(x) = Ae^{\frac{1}{4} + x - x^2}, x \in \mathbb{R}.$$

Tìm A, sau đó tính các mô men gốc cấp k, $k = 1, 2, 3, 4$.

Giải. Dùng tính chất (iv) của hàm mật độ, ta có

$$1 = A\int_{-\infty}^{+\infty} e^{\frac{1}{4} + x - x^2} dx = A\int_{-\infty}^{+\infty} e^{\frac{1}{2} - (x - \frac{1}{2})^2}\,dx$$

$$= A\sqrt{e\pi}\,\frac{1}{\sqrt{\pi}}\int_{-\infty}^{+\infty} e^{-\frac{1}{2.(\frac{1}{2})}(x - \frac{1}{2})^2}\,dx = A\sqrt{e\pi}$$

$$\Rightarrow \qquad A = \frac{1}{\sqrt{e\pi}}.$$

Mặt khác, có thể thấy rằng $X = \sqrt{e}Z$, với $Z \sim \mathcal{N}\left(\frac{1}{2}, \frac{1}{2}\right)$, từ đó:

$$\nu_1 = \sqrt{e}/2;\ \nu_2 = \mu_2 + \nu_1^2 = 3e/4;$$

$$\nu_3 = 3\nu_2\nu_1 - 2\nu_1^3 = 7e\sqrt{e}/8;$$

$$\nu_4 = \mu_4 + 4\nu_1\nu_3 - 6\nu_1^2\,\nu_2 + 3\nu_1^4 = \frac{1}{8}(8 - 11e^2).$$

Chú ý: Đối với biến $X \sim \mathcal{N}(\mu, \sigma^2)$ ta có:

$$EX = \mu;\ VX = \sigma^2;\ \mathrm{mod}X = \mathrm{med}X = EX;\ \mu_3 = 0;\ \mu_4 = 3\sigma^4.$$

BÀI TẬP

1. Tìm kỳ vọng và phương sai của biến ngẫu nhiên X có:
   a) phân phối Poisson $\mathcal{P}(\lambda)$ với $\lambda = 4$;
   b) phân phối mũ $\mathcal{E}(\lambda)$ với với $\lambda = 0{,}4$.

2. Cho X và Y là hai biến ngẫu nhiên X độc lập.
   a) Giả sử $X \sim \mathcal{B}(1; 0{,}2)$ và $Y \sim \mathcal{B}(2; 0{,}2)$, lập bảng PPXS của $Z = X + Y$, sau đó tính EZ và VZ.
   b) Giả sử $X \sim \mathcal{B}(1; 0{,}5)$ và $Y \sim \mathcal{B}(2; 0{,}2)$, lập bảng PPXS của $Z = X + Y$; Z có phân phối nhị thức không?

3. Một người cho thuê 3 xe con. Anh ta hàng ngày phải nộp thuế 8$ cho 1 xe (dù xe có được thuê hay không), mỗi xe được cho thuê với giá 20$. Giả sử số yêu cầu thuê xe trong một ngày tuân theo luật Poisson với $\lambda = 2{,}8$.
   a) Tìm PPXS của số tiền anh ta thu được trong 1 ngày, sau đó tính số tiền trung bình thu được trong ngày.
   b) Giải bài toán trong trường hợp có 4 xe. So sánh nên có 3 hay 4 xe.

4. Gieo một đồng tiền cho đến khi xuất hiện mặt ngửa thì dừng. XS xuất hiện mặt ngửa là p. Gọi X là số lần gieo cần thiết.
   a) Tính trị trung bình X.
   b) Tìm PPXS của X với điều kiện trong n lần gieo đầu tiên chỉ có đúng 1 lần xuất hiện mặt ngửa.

5. Cho hàm mật độ của X
   $$f(x) = \frac{1}{\sqrt{2\pi}} x^{0,5} e^{-0,5x}, x > 0.$$
   a) Tìm trị trung bình, phương sai và mốt của X.
   b) Tính các hệ số β_1 và β_2.
   Gợi ý: Dùng $\int_0^{+\infty} x^{s-1} e^{-px} dx = \Gamma(s)/p^s$, $p>0$.

6. Một tổ có 5 công nhân với XS mắc bệnh nghề nghiệp A sau 10 năm làm việc là 0,3.
   a) Lập bảng PPXS của số người mắc bệnh nghề nghiệp A sau 10 năm công tác.
   b) Tìm số người chắc nhất mắc bệnh.

7. Một người đi làm từ nhà đến cơ quan phải qua 3 ngã tư, XS gặp đèn đỏ ở các ngã tư tương ứng là 0,2; 0,4 và 0,5. Gọi X là số đèn đỏ mà người đó gặp trong một lần đi làm
   a) Lập bảng PPXS và hàm phân phối của X. Tính EX, VX.
   b) Hỏi thời gian trung bình người đó phải dừng vì đèn đỏ là bao nhiêu, biết rằng mỗi khi gặp đèn đỏ phải đợi khoảng 3 phút?

8. Tuổi thọ của một loại côn trùng là biến ngẫu nhiên X (đơn vị tháng) với hàm mật độ
$$f(x) = \begin{cases} kx^2(4-x); & x \in [0;4]; \\ 0; & x \notin [0;4]. \end{cases}$$
   a) Tìm k và mod$X$.
   b) Tính XS để côn trùng chết trước khi nó tròn 1 tháng tuổi.

9. Cho biến ngẫu nhiên X có hàm mật độ
$$f(x) = \begin{cases} kx^2e^{-2x}; & x \geq 0; \\ 0; & x < 0. \end{cases}$$
   a) Tìm k và hàm phân phối của X.
   b) Tìm kỳ vọng, phương sai và mốt của $\boldsymbol{X}$.

10. Cho X là biến ngẫu nhiên có phân phối mũ với tham số $\lambda = 2$, tính kỳ vọng và độ lệch chuẩn của $Y = e^{-X}$.

11. Một nhà máy bán một loại sản phẩm với giá 1$/1 sản phẩm. Trọng lượng một sản phẩm là biến ngẫu nhiên có phân phối chuẩn với kỳ vọng μ *kg* và độ lệch chuẩn 1 *kg*. Giá thành một sản phẩm là $c = 0{,}05\mu + 0{,}3$. Nếu trọng lượng bé hơn 8 *kg* thì phải loại bỏ vì không bán được. Hãy xác định μ để lợi nhuận của nhà máy là lớn nhất.

12. Đối với luật phân phối đối xứng có thể lấy độ lệch trung bình θ xác định từ điều kiện

$$P(|X - EX| < \theta) = 1/2$$

làm thước đo độ phân tán của biến ngẫu nhiên. Hãy tìm phương sai và độ lệch trung bình của biến ngẫu nhiên X có hàm mật độ

$$f(x) = \begin{cases} \dfrac{1}{\pi\sqrt{a^2 - x^2}}, & |x| \le a; \\ 0, & |x| > a. \end{cases}$$

§4 MỘT SỐ PHÂN PHỐI XÁC SUẤT THÔNG DỤNG

4.1 Phân phối đều

1. Phân phối đều rời rạc

*** Định nghĩa 1.** Biến ngẫu nhiên X được gọi là tuân theo *luật PP đều rời rạc*, ký hiệu là $X \sim \mathcal{U}_{\{1,2,\dots,n\}}$, nếu nó có bảng PPXS

x	1	2	...	n
$p(x)$	$1/n$	$1/n$	...	$1/n$

Như vậy hàm xác suất của X có dạng $p(x) = 1/n, x = \overline{1,n}$. Ta có thể mở rộng tập giá trị $\{x_1, x_2, \dots, x_n\}$ và $p(x_i) = 1/n, i = \overline{1,n}$; trong trường hợp này ta ký hiệu $X \sim \mathcal{U}_{\{x_1,x_2,\dots,x_n\}}$.

Dễ dàng, nếu $X \sim \mathcal{U}_{\{1,2,\dots,n\}}$, ta có:

$$EX = \frac{n+1}{2} ; VX = \frac{n^2-1}{12} .$$

2. Phân phối đều liên tục

*** Định nghĩa 2.** Biến ngẫu nhiên X được gọi là tuân theo *luật PP đều liên tục* trên (a, b), ký hiệu là $X \sim \mathcal{U}(a, b)$, nếu nó có hàm mật độ

$$f(x) = \begin{cases} \frac{1}{b-a}; \ x \in (a,b), \\ \quad 0; \ x \notin (a,b). \end{cases}$$

Bằng tính toán đơn giản ta tính được

$$EX = \frac{a+b}{2}\,; VX = \frac{(b-a)^2}{12}\,.$$

Phân phối đều $\mathcal{U}(0;1)$ có vai trò quan trọng trong tính toán mô phỏng và nếu $X \sim \mathcal{U}(0;1)$ thì:

$$f(x) = \begin{cases} 1; \ x \in (0,1), \\ 0; \ x \notin (0,1). \end{cases} \qquad F(x) = \begin{cases} 0; \quad x \le 0, \\ x; 0 < x \le 1, \\ 1; \quad x > 1. \end{cases}$$

4.2 Phân phối nhị thức

1. Phân phối Bernoulli

*** Định nghĩa 1.** Biến ngẫu nhiên X được gọi là tuân theo *luật PP Bernoulli*, ký hiệu $X \sim \mathcal{B}(1;p)$, nếu hàm xác suất của nó có dạng

$$p(x) = p^x(1-p)^{1-x}, x = 0 \text{ hoặc } 1.$$

Bảng PPXS của biến $X \sim \mathcal{B}(1;p)$ là

x	0	1
$p(x)$	$q = 1 - p$	p

Để ý mọi phép thử chỉ có 2 kết cục đều có thể mô hình hoá bằng phân phối này. Dễ dàng có được nếu $X \sim \mathcal{B}(1;p)$

$$EX = p\,; VX = pq.$$

2. Phân phối nhị thức

*** Định nghĩa 2.** Biến ngẫu nhiên X được gọi là tuân theo *luật PP nhị thức*, ký hiệu là $X \sim \mathcal{B}(n; p)$, nếu hàm xác suất của nó có dạng

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

Biến ngẫu nhiên này liên quan chặt chẽ đến khái niệm lược đồ và công thức Bernoulli đã nói đến ở chương I. Cần nhắc lại các điều kiện của lược đồ Bernoulli:

- dãy n phép thử giống nhau và độc lập;
- trong mỗi phép thử, sự kiện quan tâm xuất hiện với XS p.

Dễ thấy PP Bernoulli là một trường hợp riêng của PP nhị thức.

*** <u>Thí dụ 1</u>**. Cho biến ngẫu nhiên $X \sim \mathcal{B}(5; 0{,}25)$. Lập bảng PP của X, tính EX, sau đó tìm: a) $P(X > 3)$; b) $P(X \leq 4)$.

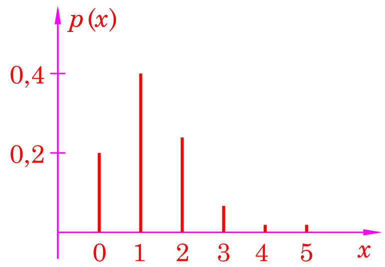
Giải. Sử dụng công thức Bernoulli trong định nghĩa, ta có

$$EX = \sum_{\forall x} x p(x) = \sum_{x=0}^{n} x \, C_n^x \, p^x q^{n-x} = np = 1{,}25,$$

còn $p(x)$ cho trong bảng PPXS của X

x	0	1	2	3	4	5
$p(x)$	0,2373	0,3955	0,2637	0,0879	0,0146	0,0010

Đồ thị của hàm XS



Bây giờ ta có ngay: a) $P(X > 3) = p(4) + p(5) = 0{,}0156$ và

b) $P(X \leq 4) = 1 - P(X > 4) = 1 - p(5) = 0{,}999$.

Mặt khác ở §2 chương II ta đã biết $X = \sum_{i=1}^{n} X_i$, X_i độc lập cùng

PP $\sim \mathcal{B}(1; p)$, suy ra:

$$EX = np\ ;VX = npq.$$

Chú ý rằng khi n khá lớn mức độ đối xứng (đối với kỳ vọng) của hàm XS càng rõ rệt. Ngoài ra có thể chứng minh hai kết quả sau:

(1) Nếu $X \sim \mathcal{B}(n; p)$ thì $Y = n - X \sim \mathcal{B}(n; 1 - p)$;

(2) Nếu $X_1 \sim \mathcal{B}(n_1; p)$, $X_2 \sim \mathcal{B}(n_2; p)$, thì $X_1 + X_2 \sim \mathcal{B}(n_1 + n_2; p)$.

4.3 Phân phối Poisson

**** Định nghĩa.*** Biến ngẫu nhiên X được gọi là tuân theo *luật PP Poisson*, ký hiệu là $X \sim \mathcal{P}(\lambda)$, nếu hàm xác suất của nó có dạng

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

Phân phối Poisson có nhiều ứng dụng thực tế trong lý thuyết phục vụ công cộng, kiểm tra chất lượng sản phẩm…

Có thể chứng minh rằng $C_n^x\, p^x q^{n-x}$, khi $n \to +\infty$, $p \to 0$ sao cho $np \to \lambda = const$, có giới hạn $\frac{\lambda^x e^{-\lambda}}{x!}$. Trong thực hành, khi n khá lớn và p đủ bé (trong thực tế $n > 50;\ p < 0{,}1$), thì $(\lambda = np)$

$$P_n(x) = C_n^x\, p^x q^{n-x} \approx \frac{\lambda^x e^{-\lambda}}{x!}.$$

* **<u>Thí dụ 2</u>**. Người ta vận chuyển 5000 chai rượu vào kho với XS mỗi chai bị vỡ là 0,0004. Tính XS để khi vận chuyển có không quá 1 chai bị vỡ.

Giải. Có thể dùng công thức Bernoulli để tính, nhưng ở đây $n = 5000$ khá lớn trong khi $p = 0{,}0004$ quá bé. Nếu gọi X là số chai bị vỡ khi vận chuyển, dễ thấy X có phân phối xấp xỉ Poisson với $\lambda \approx np = 2$, từ đó

$$P(0 \le X \le 1) = \frac{2^0 e^{-2}}{0!} + \frac{2^1 e^{-2}}{1!} = \frac{3}{e^2} \approx 0{,}406.$$

Các đặc trưng của $X \sim \mathcal{P}(\lambda)$: $EX = \lambda;\ VX = \lambda.$

Để ý giá trị $X = 1$ chính là mod$X$. Người ta đã chứng minh $\lambda - 1 \le$ mod$X \le \lambda$: nếu λ không nguyên thì mod$X$ là số nguyên nằm giữa λ và $\lambda - 1$; còn nếu λ nguyên thì ta có hai mốt là λ và $\lambda - 1$. Trong thí dụ 2, mốt của X là các giá trị 1 và 2 (xác suất $= 0,2707$ cho cả hai trường hợp).

4.4 Các phân phối rời rạc khác

1. Phân phối siêu bội

Một trong giả thiết của PP nhị thức là sự độc lập của các phép thử thành viên. Một trường hợp cổ điển là giả sử ta có N sản phẩm, trong đó có một tỷ lệ phế phẩm p, nếu ta chọn không hoàn lại ra n sản phẩm và gọi $X = m$ là số phế phẩm được rút ra thì $P(X = m)$ sẽ không thể tính theo công thức Bernoulli.

**** Định nghĩa 1.*** Biến ngẫu nhiên X được gọi là tuân theo *luật PP siêu bội*, ký hiệu là $X \sim \mathcal{H}(N,n,p)$ (với p là tỷ lệ phế phẩm lúc ban đầu $\Rightarrow Np$ số phế phẩm ban đầu), nếu hàm XS của nó có dạng

$$p(x) = \frac{C_{Np}^{x} C_{N-Np}^{n-x}}{C_N^n}, x = \overline{1,n}.$$

Để ý nếu đặt $q = 1 - p$, thì công thức trên có thể viết lại

$$p(x) = \frac{C_{Np}^{x} C_{Nq}^{n-x}}{C_N^n}, x = \overline{1,n}.$$

Khi N rất lớn, p sẽ ít thay đổi (coi là hằng số xác định) và phân phối siêu bội có thể xấp xỉ bằng phân phối nhị thức.

*** <u>Thí dụ 3</u>**. Trong một hộp đèn 15 bóng có 5 bóng kém chất lượng. Chọn ngẫu nhiên ra 10 bóng (không hoàn lại), hãy lập bảng PPXS của số bóng kém chất lượng trong mẫu chọn ra.

Giải. Gọi X là số bóng kém chất lượng trong mẫu, rõ ràng

x	0	1	2	3	4	5
$p(x)$	0,00033	0,01665	0,14985	0,39960	0,34965	0,08392

Trong thực hành khi $N > 10n$ mới xấp xỉ bằng PP nhị thức.

Có thể tính được các đặc trưng của $X \sim \mathcal{H}(N,n,p)$:

$$EX = np;\ VX = npq\,\frac{N-n}{N-1}\,.$$

Ngoài ra khi $N \to +\infty$ sao cho $n/N \to 0$, ta sẽ có

$$\lim_{\frac{n}{N}\to 0} \frac{C_{Np}^{x} C_{Nq}^{n-x}}{C_{N}^{n}} = C_n^x\, p^x q^{n-x}.$$

Trên cơ sở lược đồ Bernoulli ta có thể đưa ra hai PP khác.

2. Phân phối hình học

**** Định nghĩa 2.*** Biến ngẫu nhiên X được gọi là tuân theo *luật PP học*, ký hiệu là $X \sim \mathcal{G}(p)$, nếu hàm XS của nó có dạng

$$p(x) = p(1-p)^x, x = 0, 1, 2, \ldots$$

Nếu đặt A là sự kiện trong dãy phép thử Bernoulli với $p = P(A)$, thì X là số lần không xuất hiện trước lần xuất hiện đầu tiên của A.

Dễ dàng chứng minh khi $X \sim \mathcal{G}(p)$

$$EX = \frac{1-p}{p} = \frac{q}{p}; VX = \frac{q}{p^2}.$$

3. Phân phối nhị thức âm

**** Định nghĩa 3.*** Biến ngẫu nhiên X được gọi là tuân theo *luật PP nhị thức âm*, ký hiệu là $X \sim \mathcal{NB}(r, p)$, nếu hàm XS của nó có dạng

$$p(x) = C_{r+x+1}^{x} \, p^r (1-p)^x, x = 0, 1, 2, \ldots$$

Ý nghĩa của X chính là số lần không xuất hiện trước lần xuất hiện thứ r ($r > 0$) của một sự kiện A trong dãy phép thử Bernoulli. So sánh với định nghĩa 2 ở trên ta thấy PP hình học là trường hợp riêng của PP nhị thức âm khi $r = 1$.

Cũng có thể tính được nếu $X \sim \mathcal{NB}(r, p)$: với $q = 1 - p$

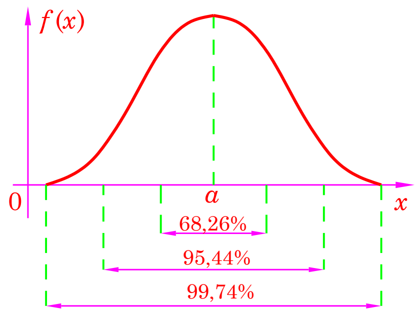
$$EX = \frac{rq}{p}; VX = \frac{rq}{p^2}.$$

4.5 Phân phối chuẩn

**** Định nghĩa.*** Biến ngẫu nhiên X được gọi là tuân theo *luật PP chuẩn* (hay *luật Gauss*), ký hiệu là $X \sim \mathcal{N}(\mu, \sigma^2)$, nếu hàm mật độ của nó có dạng

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

Đồ thị của $f(x)$



Ta đã biết từ §2 và §3 $EX = \mu; VX = \sigma^2$ (σ là độ lệch chuẩn).

Đường cong mật độ chuẩn có dạng chuông, nên trong ứng dụng người ta còn gọi là *PP dạng chuông*. Trong đồ thị $\mu = a$, giá trị này vừa là trị trung bình, vừa là mốt và trung vị của X. Còn

$$P(|X - \mu| < \sigma) = 0{,}6826; P(|X - \mu| < 2\sigma) = 0{,}9544;$$

$$P(|X - \mu| < 3\sigma) = 0{,}9974 \ (\textit{quy tắc } 3\sigma).$$

*** <u>Thí dụ 4</u>**. Độ dài một chi tiết máy giả sử tuân theo luật chuẩn với trị trung bình 20 *cm* và độ lệch chuẩn là 0,5 *cm*. Hãy tính XS để với một chi tiết loại trên được chọn ngẫu nhiên thì độ dài của nó:

a) lớn hơn 20 cm; b) bé hơn 19,5 cm; c) lớn hơn 21,5 cm.

Giải. Gọi X là độ dài chi tiết máy ở trên $\Rightarrow X \sim \mathcal{N}(20; 0{,}5^2)$

a) Do tính đối xứng của PP qua kỳ vọng nên $P(X > 20) = 0{,}5$.

b) Do $P(19{,}5 < X < 20{,}5) = 68{,}28\% \Rightarrow$ XS nằm ngoài khoảng bằng 31,74%. Do tính đối xứng $P(X < 19{,}5) = 15{,}87\%$ (và cũng bằng $P(X > 20{,}5)$).

c) Do cùng các lý do như trên và dùng quy tắc 3σ $P(X > 20{,}5)$ $= (1-99{,}74\%)/2 = 0{,}0013$ (XS khá bé).

Trong trường hợp tổng quát ta có, nếu $X \sim \mathcal{N}(\mu, \sigma^2)$,

$$P(\alpha < X < \beta) = \phi\left(\frac{\beta-\mu}{\sigma}\right) - \phi\left(\frac{\alpha-\mu}{\sigma}\right),$$

với $\phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-t^2/2}\, dt$ là hàm Laplace, và $\forall \varepsilon > 0$

$$P(|X-\mu| < \varepsilon) = 2\,\phi\left(\frac{\varepsilon}{\sigma}\right).$$

Để ý hàm phân phối của X sẽ thoả mãn $F(x) = \phi(x) + 0{,}5$. Đồng thời bằng phép biến đổi $Z = \dfrac{X-\mu}{\sigma}$ ta có thể đưa $X \sim \mathcal{N}(\mu, \sigma^2)$ về $Z \sim \mathcal{N}(0; 1)$ với hàm mật độ (hàm Gauss)

$$\varphi(x) = \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}, x \in \mathbb{R}.$$

Tổng của n biến ngẫu nhiên độc lập cùng phân phối chuẩn là một biến ngẫu nhiên chuẩn (ta sẽ thấy chặt chẽ hơn ở chương III).

Từ đó nếu $X_i \sim \mathcal{N}(\mu, \sigma^2)\ \forall i = \overline{1, n}$ và độc lập thì

$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n} = \mathcal{N}\left(\mu; \frac{\sigma^2}{n}\right).$$

và $$Z = \frac{\bar{X} - \mu}{\sigma}\sqrt{n} \sim \mathcal{N}(0; 1) \qquad (\textit{PP chuẩn chuẩn tắc}).$$

Cuối cùng PP chuẩn có thể được dùng để xấp xỉ khá tốt cho một số PP rời rạc. Về mặt lý thuyết có thể chứng minh nếu $X \sim \mathcal{B}(n; p)$

$$\frac{X - np}{\sqrt{npq}} \xrightarrow[n\to\infty]{L} Z \sim \mathcal{N}(0; 1) \qquad (\textit{hội tụ theo luật PP}).$$

Chính sự kiện này cho phép ta xấp xỉ phân phối nhị thức với n khá lớn và $np \geq 5$ (khi $p \leq 0{,}5$) *hoặc* $n(1 - p) \geq 0{,}5$ (*khi* $p \geq 0{,}5$) bằng phân phối chuẩn: Nếu $X \sim \mathcal{B}(n; p)$, thì

$$P(X = \alpha) = \frac{\varphi\left(\frac{\alpha - np}{\sqrt{npq}}\right)}{\sqrt{npq}};$$

$$P(\alpha \leq X \leq \beta) = \phi\left(\frac{\beta - np}{\sqrt{npq}}\right) - \phi\left(\frac{\alpha - np}{\sqrt{npq}}\right).$$

*** <u>Thí dụ 5</u>**. Cho $X \sim \mathcal{B}(20; 0{,}4)$, tính $P(4 \leq X \leq 13)$.

Giải. Theo công thức ở trên

$$P(4 \leq X \leq 13) = \phi\left(\frac{13 - 8}{\sqrt{4{,}8}}\right) - \phi\left(\frac{4 - 8}{\sqrt{4{,}8}}\right) = \phi(2{,}28) + \phi(1{,}83)$$

$$= 0{,}4884 + 0{,}4664 = 0{,}9548.$$

Nhưng do n =20 chưa thật lớn, trong thực hành người ta hiệu chỉnh công thức xấp xỉ trên như sau

$$P(\alpha \le X \le \beta) = \phi\left(\frac{\beta+0{,}5-np}{\sqrt{npq}}\right) - \phi\left(\frac{\alpha-0{,}5-np}{\sqrt{npq}}\right).$$

Từ đó $\quad P(4 \le X \le 13) = \phi(2{,}51) + \phi(1{,}60) = 0{,}9743.$

Để ý kết quả đúng của XS này là 0,978.

Người ta cũng chứng minh được rằng, nếu $X \sim \mathcal{P}(\lambda)$ thì

$$\frac{X-\lambda}{\sqrt{\lambda}} \xrightarrow[\lambda\to\infty]{L} Z \sim \mathcal{N}(0;1) \qquad (\textit{hội tụ theo luật PP}).$$

4.6 Phân phối mũ

**** Định nghĩa.*** Biến ngẫu nhiên X được gọi là tuân theo *luật PP mũ*, ký hiệu là $X \sim \mathcal{E}(\lambda)$, nếu hàm mật độ của nó có dạng

$$f(x) = \begin{cases} \lambda e^{-\lambda x}; & x > 0, \\ 0; & x \le 0. \end{cases}$$

Dễ dàng tính được:

$$EX = \frac{1}{\lambda}\,; VX = \frac{1}{\lambda^2}\,;$$

và hàm PPXS $\quad F(x) = 1 - e^{-\lambda x}, x > 0.$

*** <u>Thí dụ 6</u>**. Thời gian hoạt động của một bóng đèn là biến ngẫu nhiên X có PP mũ với kỳ vọng là 500. Tìm XS để thời gian hoạt động của bóng đèn không bé hơn 1000 giờ.

Giải. Vì $EX = \frac{1}{\lambda} = 500$ nên $\lambda = 0{,}002$. Vậy XS phải tìm là

$P(X \geq 1000) = 1 - P(X < 1000) = e^{-0{,}002.1000} = e^{-2} \approx 0{,}1353.$

4.7 Các phân phối liên tục khác

1. Phân phối χ^2

Nhiều PP liên tục được cảm sinh trực tiếp bởi PP chuẩn. Các PP này cũng rất quan trọng và hay được dùng trong thống kê.

**** Định nghĩa 1.*** Biến ngẫu nhiên X được gọi là tuân theo *luật PP χ^2 với n bậc tự do*, ký hiệu là $X \sim \chi^2(n)$, nếu hàm mật độ có dạng

$$f(x) = \frac{x^{\frac{n}{2}-1} e^{-\frac{x}{2}}}{2^{\frac{n}{2}}\,\Gamma\left(\frac{n}{2}\right)}, x > 0, n > 0,$$

trong đó hàm ga-ma đã quen thuộc trong giải tích toán

$$\Gamma(x) = \int_0^{+\infty} t^{x-1} e^{-t} dt, x > 0$$

với các tính chất $\forall$ i *nguyên*

(i) $\Gamma(i+1) = i!\ (i > 0)$;

(ii) $\Gamma\left(\frac{i}{2}\right) = \left(\frac{i}{2} - 1\right)\left(\frac{i}{2} - 2\right) \ldots \frac{3}{2} . \frac{1}{2}\sqrt{\pi}$ (i lẻ > 2);

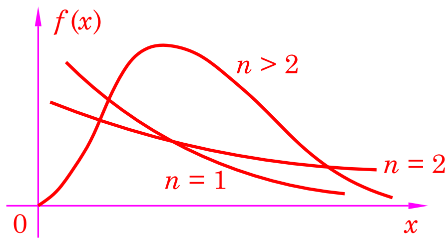
(iii) $\Gamma(x) = (x-1)\Gamma(x-1), x \in \mathbb{R}$.

Tuy nhiên cách định nghĩa này khá phức tạp và khó cho phép nhận biết PP rõ ràng trong thực hành.

*** Định nghĩa 2.** Cho n biến ngẫu nhiên độc lập $X_i \sim \mathcal{N}(0; 1)$, $i = \overline{1, n}$. Khi đó biến ngẫu nhiên

$$U_n = \sum_{i=1}^{n} {X_i}^2 \sim \chi^2(n).$$

Đồ thị của hàm mật độ:



Các đặc trưng:

$$EU_n = n; VU_n = 2n.$$

Phân phối χ^2 có một số tính chất quan trọng:

(i) Nếu $X \sim \chi^2(n)$, $Y \sim \chi^2(m)$ và độc lập $\Rightarrow X + Y \sim \chi^2(n+m)$.

(ii) $\frac{U_n - n}{\sqrt{2n}} \xrightarrow[n\to\infty]{L} Z \sim \mathcal{N}(0; 1)$ (*hội tụ theo luật PP*).

Trong thống kê ta dùng một hệ quả quan trọng của tính chất (ii):

Nếu ta có n biến ngẫu nhiên độc lập $X_i \sim \mathcal{N}(\mu, \sigma^2)$, $i = \overline{1, n}$, và

$$\bar{X} = \frac{1}{n}(X_1 + X_2 + \cdots + X_n) \Rightarrow \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \bar{X})^2 \sim \chi^2(n-1).$$

2. Phân phối Student (phân phối t(n))

**** Định nghĩa 1.*** Biến ngẫu nhiên X được gọi là tuân theo *luật PP Student với* n *bậc tự do,* ký hiệu là $X \sim t(n)$, nếu hàm mật độ có dạng

$$f(x) = \frac{1}{\sqrt{n\pi}} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{x^2}{n}\right)^{-\left[\frac{n+1}{2}\right]}, n > 0.$$

*** Định nghĩa 2.** Cho X và Y là hai biến ngẫu nhiên độc lập tuân theo luật $\mathcal{N}(0;1)$ và $\chi^2(n)$ tương ứng. Khi đó biến ngẫu nhiên

$$T_n = \frac{X}{\sqrt{Y/n}} \sim t(n).$$

Đồ thị của PP $t(n)$ có dạng rất giống với đường cong chuẩn $\mathcal{N}(0;1)$. Các số đặc trưng của T_n:

$$ET_n = 0 \ (n > 1); VT_n = \frac{n}{n-2} \ (n > 2).$$

PP Student có tính chất

$$T_n \xrightarrow[n\to\infty]{L} Z \sim \mathcal{N}(0;1) \qquad \text{(}\textit{hội tụ theo luật PP}\text{).}$$

Trong thực hành, khi $n \geq 30$, đồ thị của mật độ phân phối $t(n)$ đã rất gần với đồ thị mật độ chuẩn $\mathcal{N}(0;1)$. Chú ý khi $n = 1$, ta có T_1 tuân theo *PP Cauchy* với hàm mật độ $f(x) = \frac{1}{\pi(1+x^2)}$, là PPXS không có mô men nào.

3. Phân phối Fisher - Snedecor (phân phối $\mathcal{F}(n,m)$)

**** Định nghĩa 1.*** Biến ngẫu nhiên X được gọi là tuân theo *luật PP Fisher - Snedecor với n và m bậc tự do,* ký hiệu là $X \sim \mathcal{F}(n,m)$, nếu hàm mật độ có dạng

$$f(x) = \frac{n^{\frac{n}{2}}\, m^{\frac{m}{2}}\, \Gamma\left(\frac{n+m}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{m}{2}\right)}\, x^{\frac{n-2}{2}}(m+nx)^{-\frac{n+m}{2}};\ x,m,n>0.$$

Ta đưa ra một định nghĩa khác giúp nhận dạng biến có PP $\mathcal{F}(n,m)$ dễ dàng hơn.

*** Định nghĩa 2.** Cho X và Y là hai biến ngẫu nhiên độc lập tuân theo luật $\chi^2(n)$ và $\chi^2(m)$ tương ứng. Khi đó biến ngẫu nhiên

$$U = \frac{X/n}{Y/m} \sim \mathcal{F}(n,m).$$

Đồ thị của hàm mật độ PP $\mathcal{F}(n,m)$ có dạng gần giống với mật độ của PP χ^2. Các số đặc trưng:

$$EX = \frac{m}{m-2}\ (m>2);\ VX = \frac{2m^2(n+m-2)}{n(m-4)(m-2)^2}\ (m>4).$$

Nếu $n = 1$ ta thấy $(T_m)^2 = \mathcal{F}(1,m)$.

3. Phân phối gamma

**** Định nghĩa 1.*** Biến ngẫu nhiên X được gọi là tuân theo *luật PP gamma,* ký hiệu là $X \sim \gamma(r, \lambda)$, nếu hàm mật độ có dạng

$$f(x) = \frac{\lambda}{\Gamma(r)} e^{-\lambda x} x^{r-1}; x, r, \lambda > 0.$$

Các số đặc trưng của $X \sim \gamma(r, \lambda)$:

$$EX = \frac{r}{\lambda}; VX = \frac{r}{\lambda^2}.$$

Một số tính chất quan trọng của PP gamma:

(i) Nếu $X \sim \gamma(p, \lambda)$, $Y \sim \gamma(q, \lambda)$ và độc lập $\Rightarrow X + Y \sim \gamma(p + q, \lambda)$.

(ii) Nếu $X \sim \gamma(r, \lambda)$ thì $\frac{X - r}{\sqrt{r}} \xrightarrow[r \to \infty]{L} Z \sim \mathcal{N}(0; 1)$

(*hội tụ theo luật PP*).

Để ý nếu $r = 1$ ta có phân phối mũ $\mathcal{E}(\lambda)$.

BÀI TẬP

1. Một dây chuyền sản xuất mỗi ngày 2000 chi tiết máy, biết XS để một chi tiết bị lỗi là 5%. Người ta lấy ngẫu nhiên ra một mẫu gồm 1% của số chi tiết được sản xuất trong ngày nào đó. Tính XS để trong mẫu đó số chi tiết bị lỗi không quá 10%.

2. Thống kê liên quan đến xe ô tô tư nhân ở Gotham City cho thấy XS để một xe vượt chuẩn Federal EPA là 0,45. Hỏi một thanh tra viên của thành phố này cần kiểm tra bao nhiêu xe trước khi với XS lớn hơn 0,95 anh ta tìm được 3 xe vượt chuẩn EPA?

3. Giả sử một thiết bị ngừng hoạt động do môi trường nhiệt độ cao và cho biết XS để một công tắc điện tử kích hoạt thiết bị dự phòng là 0,6. Nếu các công tắc hoạt động độc lập và chỉ kích hoạt mỗi lần một thiết bị, thì cần lắp song song bao nhiêu công tắc để XS kích hoạt thành công đạt ít nhất 95%?

4. XS để bán được 1 chiếc máy giặt mới trong vòng 1 ngày là 0,006. Tìm XS để trong ngày tiêu thụ được không quá 9 chiếc từ lô hàng 1000 chiếc.

5. Một quyển sách 500 trang có 500 lỗi. Tính XS để:
   a) trong một trang nào đó có không ít hơn 3 lỗi;
   b) ít nhất 1 trang chứa nhiều hơn 4 lỗi.

6. Trung bình người bạch tạng chiếm 1% cư dân; đánh giá khả năng trong 200 cư dân được chọn ngẫu nhiên có không ít hơn 4 người bạch tạng.

7. (Bài toán Banach) Một người có trong túi 2 bao diêm, mỗi bao có n que. Mỗi khi cần diêm anh ta rút hú hoạ ra một bao. Tìm XS sao cho khi người đó lần đầu rút phải bao rỗng thì trong bao kia còn đúng k que.

8. Cho biến ngẫu nhiên X có kỳ vọng μ và độ lệch chuẩn σ. Hãy tính $P(|X - \pi| < 3\sigma)$ cho các trường hợp:
   a) X có PP Poisson với tham số $\lambda = 0{,}09$;
   b) X có PP đều trên $(0, 1)$;
   c) X có PP mũ với tham số λ.

9. Tìm XS để một biến ngẫu nhiên có PP $t(14)$ se nhận giá trị: a) lớn hơn 2,145; b) lớn hơn – 2,145.

10. Tìm XS để một biến ngẫu nhiên có PP $\chi^2(12)$ lấy giá trị lớn hơn 23,337. Biến $X \sim \chi^2(17)$ sẽ nhận các giá trị nào để XS lớn hơn 0,05?

11. Cho 2 biến ngẫu nhiên $X \sim \mathcal{N}(0; 1)$ và $Y \sim \chi^2(17)$. Tính XS $P(X > \sqrt{Y} + 10)$.

12. Sức bền thiết kế của một dây chịu lực cầu dây văng có PP $\chi^2(9)$. Sau khi cầu xây xong ứng suất lực dây đó phải chịu có PP $\chi^2(4)$. Hỏi dây chịu lực có đủ sức bền cần thiết hay không?

§1 LUẬT PPXS CỦA BIẾN NGẪU NHIÊN NHIỀU CHIỀU

1.1 Khái niệm cơ sở

*** Khái niệm biến ngẫu nhiên nhiều chiều**

Thực tế yêu cầu xét *đồng thời* nhiều thuộc tính có quan hệ với nhau của đối tượng mà ta quan tâm, ký hiệu $X_1, X_2, \ldots, X_n$.

Ta có thể dùng công cụ véc tơ của giải tích toán

$$\boldsymbol{X} = (X_1\ X_2\ \ldots\ X_n)^T \in \mathbb{R}^n$$

và coi $\boldsymbol{X}$ là một *biến ngẫu nhiên n chiều* hay *véc tơ ngẫu nhiên* (n chiều).

Để cho đơn giản, ta chỉ xét trường hợp $n = 2$ và $\boldsymbol{X} = (X\ \ Y)^T$, trong đó X, Y là các biến ngẫu nhiên 1 chiều.

*** Phân loại biến ngẫu nhiên hai chiều**

- Biến $\boldsymbol{X}$ được coi là *rời rạc* nếu cả 2 thành phần là biến rời rạc.
- Biến $\boldsymbol{X}$ được coi là *liên tục* nếu 2 thành phần là biến liên tục.
- Biến $\boldsymbol{X}$ có dạng hỗn hợp nếu 1 thành phần rời rạc và thành phần kia liên tục mà ta không xét ở đây.

Ta mở rộng khái niệm hàm PPXS cho biến ngẫu nhiên 2 chiều. Ký hiệu 2 sự kiện $A = \{X < x\}$ và $B = \{Y < y\}$ }.

*** Định nghĩa.** *Hàm phân phối xác suất của biến 2 chiều* $\boldsymbol{X} =$ $(X \;\; Y)^T$ được xác định như sau

$$F(x, y) = P(AB) = P(X < x, Y < y); x, y \in \mathbb{R}. (1)$$

Người ta còn dùng thuật ngữ *hàm phân phối xác suất đồng thời của 2 biến* X, Y.

Mở rộng các tính chất của hàm PPXS của chương 2, ta có

*** Tính chất của $\boldsymbol{F(x,y)}$**

(i) $1 \geq F(x, y) \geq 0$;

(ii) $F(x, y)$ là hàm không giảm theo từng đối số;

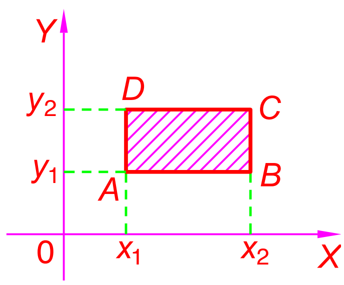
(iii) Với $x_1 \leq x_2, y_1 \leq y_2$

$$P(x_1 \leq X \leq x_2, y_1 \leq Y \leq y_2)$$

$$= F(x_2, y_2) - F(x_2, y_1) - F(x_1, y_2) + F(x_1, y_1);$$

(iv) $F(+\infty, +\infty) = 1$; $F(-\infty, y_1) = F(x_1, -\infty) = 0$.

Tính chất (iii) có ý nghĩa là XS để điểm ngẫu nhiên (X, Y) rơi vào miền chữ nhật ABCD.



*** Phân phối biên**

Để ý rằng

$$F(x,+\infty) = P(X < x, Y < +\infty) = P(X < x) = F_1(x);$$

$$F(+\infty,y) = P(X < +\infty, Y < y) = P(Y < y) = F_2(y);$$

là các phân phối riêng của từng thành phần X, Y tương ứng; chúng được gọi là các *phân phối biên* của biến hai chiều $\boldsymbol{X}$. Đó cũng chính là các phân phối một chiều đã biết ở chương II.

*** Khái niệm độc lập**

Hai biến X, Y được gọi là *độc lập* nếu (dùng định nghĩa ở chương I: $P(AB) = P(A)P(B) \Leftrightarrow A, B$ độc lập)

$$F(x,y) = F_1(x)F_2(y).$$

1.2 PPXS của biến ngẫu nhiên hai chiều rời rạc

*** Định nghĩa.** *Bảng phân phối xác suất của biến ngẫu nhiên 2 chiều* $\boldsymbol{X} = (X \;\; Y)^T$ là

$x \setminus y$	y_1	y_2	$\cdots$	y_j	$\cdots$	y_m	$\sum_j$
x_1	p_{11}	p_{12}	$\cdots$	p_{1j}	$\cdots$	p_{1m}	$p_1(x_1)$
x_1	p_{21}	p_{22}	$\cdots$	p_{2j}	$\cdots$	p_{2m}	$p_1(x_2)$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
x_i	p_{i1}	p_{i2}	$\cdots$	p_{ij}	$\cdots$	p_{im}	$p_1(x_i)$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
x_n	p_{n1}	p_{n2}	$\cdots$	p_{nj}	$\cdots$	p_{nm}	$p_1(x_n)$
$\sum_i$	$p_2(y_1)$	$p_2(y_2)$	$\cdots$	$p_2(y_j)$	$\cdots$	$p_2(y_m)$	1

trong đó $p_{ij} = P(X = x_i, Y = y_j)$ là *xác suất đồng thời* để X và Y nhận các giá trị tương ứng.

Tương tự trường hợp 1 chiều ta có thể xác định *hàm xác suất đồng thời* $p(x, y)$ sao cho $p(x_i, y_j) = p_{ij};\ i = \overline{1, n}, j = \overline{1, m}$.

*** Tính chất của $\boldsymbol{p(x, y)}$**

(i) $p_{ij} \geq 0, i = \overline{1, n}, j = \overline{1, m}$ hay $p(x, y) \geq 0\ \forall\ x, y$;

(ii) $\sum_{i=1}^{n} \sum_{j=1}^{m} p_{ij} = 1$ hay $\sum_{\forall x, y} p(x, y) = 1$.

*** Hàm phân phối xác suất đồng thời**

$$F(x, y) = \sum_{x_i < x} \sum_{y_j < y} p_{ij}.$$

*** Hàm xác suất biên**

$$P(X = x_i) = p_1(x_i) = \sum_j p_{ij},\ i = \overline{1, n};$$

$$P(Y = y_j) = p_2(y_j) = \sum_i p_{ij},\ j = \overline{1, m}.$$

*** Thí dụ 1**. Từ bảng PPXS đồng thời của X và Y

$x \backslash y$	1	2	3
1	0,10	0,25	0,15
2	0,15	0,00	0,35

Tìm các PP biên của X và Y, sau đó tính $F(2, 3)$.

Giải. Lấy tổng hàng và tổng cột của bảng PP đồng thời, ta có các bảng PP biên tương ứng:

x	1	2
$p_1(x)$	0,50	0,50

y	0	1	2
$p_2(y)$	0,25	0,25	0,50

Việc tính $F(2, 3)$ dựa vào công thức hàm PP đồng thời

$$F(2,3) = \sum_{x_i<2} \sum_{y_j<3} p_{ij} = p_{11} + p_{12} = 0{,}35.$$

*** Khái niệm độc lập**

Hai biến rời rạc X và Y được gọi là độc lập, nếu với mọi cặp giá trị x_i, y_j, ta luôn có

$$p_{ij} = p_1(x_i)\, p_2(y_j);\ i = \overline{1,n}, j = \overline{1,m}.$$

Trong thí dụ 1 ta thấy $p_{11} = 0{,}1 \neq p_1(1)\, p_2(1) = 0{,}125 \Rightarrow$ hai biến X, Y không độc lập.

*** Phân phối có điều kiện**

Từ chương I ta có

$$P(X = x_i|Y = y_k) = \frac{P(X = x_i; Y = y_k)}{P(Y = y_k)}, i = \overline{1, n}.$$

Công thức này cho phép ta xác định XS của phân phối có điều kiện của X, biết Y nhận một giá trị cụ thể y_k.

*** <u>Thí dụ 2</u>**. Tìm phân phối có điều kiện của X, biết $Y = 1$ trong bài toán thí dụ 1.

Giải. Theo công thức XS có điều kiện ở trên

$$P(X = 1|Y = 1) = \frac{P(X=1; Y=1)}{P(Y=1)}, = \frac{p_{11}}{p_2(1)} = \frac{0,1}{0,25} = 0,4;$$

$$P(X = 2|Y = 1) = \frac{p_{21}}{p_2(1)} = \frac{0,15}{0,25} = 0,6.$$

Từ đó bảng PP có điều kiện của X biết $Y = 1$ là

x	1	2
$p_1(x\|Y = 1)$	0,4	0,6

Có thể tổng quát hoá PP có điều kiện với bộ điều kiện C_Y nào đó

$$P(X = x|C_Y) = \frac{P(X = x; C_Y)}{P(C_Y)}.$$

Chẳng hạn nếu ta biết $C_Y = \{ y_1 \le Y \le y_2\}$, thì

$$P(X = x| y_1 \le Y \le y_2) = \frac{P(X = x;\ y_1 \le Y \le y_2)}{P(y_1 \le Y \le y_2)}.$$

1.3 PPXS của biến ngẫu nhiên hai chiều liên tục

Ta đưa ra định nghĩa của hàm mật độ hai chiều:

*** Định nghĩa.** Nếu hàm phân phối $F(x, y)$ của biến ngẫu nhiên hai chiều $\boldsymbol{X} = (X \;\; Y)^T$ được biểu diễn dưới dạng

$$F(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u, v) du dv,$$

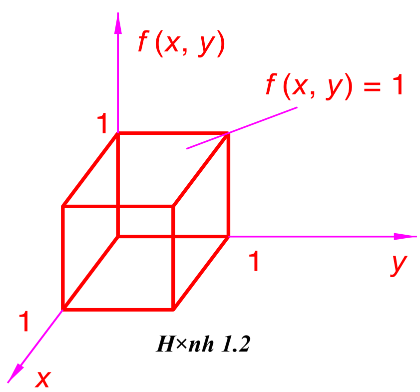
thì hàm $f(x, y)$ được gọi là *hàm mật độ của biến* $\boldsymbol{X}$ hay *hàm mật độ đồng thời của X và Y*.

Về mặt ý nghĩa hình học hàm $z = f(x, y)$ vẽ ra một mặt cong trong $\mathbb{R}^3$ và được gọi là mặt mật độ PPXS. Nếu $F(x, y)$ khả vi theo cả hai biến thì

$$f(x, y) = \frac{\partial^2 F(x,y)}{\partial x \partial y}\,.$$

*** <u>Thí dụ 3</u>.** Cho hàm mật độ đồng thời của X và Y là $f(x, y) = 1$, với $0 \leq x, y \leq 1$. Vẽ hàm $f(x, y)$ và tính hàm PP đồng thời $F(x, y)$.

Giải. Mặt cong mật độ ở đây là phần mặt phẳng trong mặt $z = 1$, tức là $f(x, y) \neq 0$ với các (x, y) thuộc hình vuông $[0;1] \times [0;1]$. Hàm PP đồng thời $F(x, y)$ được



H×nh 1.2

tính sau đây :

$$F(x,y) = \begin{cases} 0, & x \leq 0 \text{ hoặc } y \leq 0; \\ xy, & 0 < x \leq 1,\ 0 < y \leq 1; \\ x, & 0 < x \leq 1,\ y > 1; \\ y, & x > 1,\ 0 < y \leq 1; \\ 1, & x > 1,\ \ y > 1. \end{cases}$$

Hàm PP có dạng rất phức tạp, nên người ta hay sử dụng hàm mật độ. Đây là thí dụ về *PP đều hai chiều*, tổng quát hoá PP đều liên tục đã xét ở chương II.

*** Tính chất của $\boldsymbol{f(x,y)}$**

(i) $f(x,y) \geq 0$;

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

(iii) $P(\boldsymbol{X} \in \mathcal{D}) = \iint_{\mathcal{D}} f(x,y)\, dxdy$.

Ở tính chất (iii) $\mathcal{D}$ là một miền nào đó trong mặt xOy và tích phân cho thể tích một hộp chữ nhật cong có đáy trên nằm trong mặt $z = f(x,y)$, đáy dưới là miền $\mathcal{D}$ (thuộc mặt $z = 0$).

*** Hàm mật độ biên**

$$f_1(x) = \int_{-\infty}^{+\infty} f(x,y)dy, \quad f_2(y) = \int_{-\infty}^{+\infty} f(x,y)dx.$$

Để ý $f_1(x) = \dfrac{\partial F_1}{\partial x}$ và là mật độ của biến X, tương tự đối với $f_2(y)$.

*** <u>Thí dụ 4</u>**. Tìm các hàm mật độ biên của biến $\boldsymbol{X} = (X,\ Y)^T$ có hàm mật độ của X và Y là

$$f(x,y) = \frac{1}{\pi^2(1+x^2)(1+y^2)}\,,\ x, y \in \mathbb{R}.$$

Giải. Từ công thức mật độ biên

$$f_1(x) = \int_{-\infty}^{+\infty} f(x,y)dy = \int_{-\infty}^{+\infty} \frac{1}{\pi^2(1+x^2)(1+y^2)}\ dy = \frac{1}{\pi(1+x^2)}\,.$$

Do tính đối xứng ta có $f_2(y) = \dfrac{1}{\pi(1+y^2)}$.

*** Khái niệm độc lập**

Hai biến ngẫu nhiên X và Y được coi là *độc lập*, nếu

$$f(x,y) = f_1(x) f_2(y).$$

*** Hàm mật độ có điều kiện**

$$\varphi(x|y) = \frac{f(x,y)}{f_2(y)} = \frac{f(x,y)}{\int_{-\infty}^{+\infty} f(x,y)dx}\,.$$

Tương tự $\psi(y|x)$ hàm mật độ có điều kiện của Y đối với $X = x$ và nó sẽ bằng $f(x,y)/\ f_1(x)$. Các mật độ có điều kiện có tính chất như các mật độ không điều kiện.

*** <u>Thí dụ 5</u>**. Cho hàm mật độ đồng thời của X và Y

$$f(x,y) = \begin{cases} x + y, & 0 \le x, y \le 1; \\ 0, & \text{với } x, y \text{ khác.} \end{cases}$$

Xác định các hàm mật độ có điều kiện.

Giải. Để dùng được công thức định nghĩa, ta phải tính các hàm mật độ biên $f_1(x), f_2(y)$:

$$f_1(x) = \int_0^1 (x + y)\, dy = x + 0{,}5,\ 0 \le x \le 1;$$

tương tự $f_2(y) = y + 0{,}5,\ 0 \le y \le 1$. Từ đó với $0 \le y \le 1$

$$\varphi(x|y) = \frac{f(x{,}y)}{f_2(y)} = \begin{cases} \dfrac{x + y}{y + 0{,}5}, 0 \le x \le 1; \\ 0, x \notin [0, 1]; \end{cases}$$

và với $0 \le x \le 1$

$$\psi(y|x) = \begin{cases} \dfrac{x + y}{x + 0{,}5}, 0 \le y \le 1 \\ 0, y \notin [0, 1]. \end{cases}$$

Để ý $\varphi(x|y)$ là hàm của cả x và y. Ngoài ra

$$f(x,y) = f_1(x)\, \psi(y|x) = f_2(y)\, \varphi(x|y);$$

và điều kiện X, Y độc lập cho thấy

$$\varphi(x|y) = f_1(x); \psi(y|x) = f_2(y).$$

Có thể viết công thức tổng quát hơn

$$\varphi(x|y_1 \le Y \le y_2) = \frac{\int_{y_1}^{y_2} f(x{,}y)dy}{\int_{-\infty}^{+\infty} dx \int_{y_1}^{y_2} f(x{,}y)dy}\,.$$

Cuối cùng ta viết các công thức cho *hàm phân phối có điều kiện*

$$\phi(x|y) = \frac{\int_{-\infty}^{x} f(u,y)du}{\int_{-\infty}^{+\infty} f(x,y)dx};$$

$$\phi(x|y_1 \le Y \le y_2) = \frac{\int_{-\infty}^{x} du \int_{y_1}^{y_2} f(u,y)dy}{\int_{-\infty}^{+\infty} dx \int_{y_1}^{y_2} f(x,y)dy}.$$

*** <u>Thí dụ 6</u>**. Lấy hàm mật độ của thí dụ 5, hãy tính các hàm mật độ có điều kiện của X, biết a) $Y = 0{,}5$; b) $Y \in [0{,}5; 0{,}75]$.

Giải. a) Trong trường hợp biết $Y = 0{,}5$

$$\varphi(x|Y = 0{,}5) = x + 0{,}5,\ 0 \le x \le 1.$$

b) Trường hợp biết $Y \in [0{,}5; 0{,}75]$, với $0 \le x \le 1$

$$\varphi(x|0{,}5 \le Y \le 0{,}75) = \frac{\int_{0,5}^{0,75}(x+y)dy}{\int_0^1 dx \int_{0,5}^{0,75}(x+y)dy} = \frac{8x+5}{9}.$$

Để ý nếu X ∉ [0;1], cả hai mật độ có điều kiện đều bằng 0.

*** <u>Thí dụ 7</u>**. Cho $X,\ Y$ có hàm mật độ đồng thời

$$f(x,y) = \begin{cases} \frac{3}{\pi}\left(1 - \sqrt{x^2+y^2}\right), & x^2+y^2 \le R^2; \\ 0, & x^2+y^2 > R^2. \end{cases}$$

Tính $P(X^2 + Y^2 \le 1/4)$.

Giải. Gọi $\mathcal{D}$ là mặt tròn $x^2 + y^2 \leq 1/4$. Khi đó

$$P(X^2 + Y^2 \leq 1/4) = \frac{3}{\pi}\iint_{\mathcal{D}}\left(1 - \sqrt{x^2 + y^2}\right)dxdy$$

$$= \frac{3}{\pi}\int_0^{2\pi}\int_0^{1/2}(1 - r)\, rdrd\varphi = 1/2.$$

*** <u>Thí dụ 8</u>**. Cho biến 2 chiều $\boldsymbol{X} = (X,\ Y)^T$ có PP đều trên một mặt tròn $\mathcal{D}$: $x^2 + y^2 \leq R^2$. Tìm các hàm mật độ (đồng thời, biên, có điều kiện).

Giải. Biến 2 chiều có PP đều trên $\mathcal{D}$: $f(x, y) = c$, $(x, y) \in \mathcal{D}$.

Từ đó $$c\iint_{\mathcal{D}} dxdy = 1 \Rightarrow c = \frac{1}{\pi R^2}$$

và $$f(x,y) = \begin{cases} \dfrac{1}{\pi R^2}, & x^2 + y^2 \leq R^2; \\ 0, & x^2 + y^2 > R^2. \end{cases}$$

Do tính đối xứng, ta chỉ tìm $f_2(y)$ và $\varphi(x|y)$:

$$f_2(y) = \int_{-\infty}^{+\infty} f(x,y)dx$$

$$= \begin{cases} \dfrac{1}{\pi R^2}\displaystyle\int_{-\sqrt{R^2-y^2}}^{\sqrt{R^2-y^2}} dx\,, & |y| \leq R; \\ 0, & |y| > R; \end{cases}$$

$$\Rightarrow \quad f_2(y) = \begin{cases} \dfrac{2\sqrt{R^2 - y^2}}{\pi R^2}, & |y| \leq R; \\ 0, & |y| > R; \end{cases}$$

$$\varphi(x|y) = \frac{f(x,y)}{f_2(y)} = \begin{cases} \dfrac{1}{2\sqrt{R^2 - y^2}}, & x^2 + y^2 \le R^2; \\ 0, & x^2 + y^2 > R^2. \end{cases}$$

$$\Longrightarrow \qquad f_1(x) = \begin{cases} \dfrac{2\sqrt{R^2 - x^2}}{\pi R^2}, & |x| \le R; \\ 0, & |x| > R; \end{cases}$$

$$\psi(y|x) = \frac{f(x,y)}{f_1(x)} = \begin{cases} \dfrac{1}{2\sqrt{R^2 - x^2}}, & x^2 + y^2 \le R^2; \\ 0, & x^2 + y^2 > R^2. \end{cases}$$

BÀI TẬP

1. Cho X và Y là hai biến ngẫu nhiên độc lập có PPXS:

x	0	1	2	3
$p(x)$	0,3	0,1	0,4	0,2

y	0	1	2
$p(y)$	0,2	0,5	0,3

a) Tìm PPXS đồng thời của X và Y.
b) Tính $P(X > Y)$.

2. Cho X và Y có bảng PPXS đồng thời

x \ y	1	2	3
1	0,12	0,15	0,03
2	0,28	0,35	0,07

a) X và Y có độc lập không?
b) Tìm luật PPXS của $Z = XY$, sau đó tính EZ.

3. Hai biến ngẫu nhiên X và Y có bảng PPXS đồng thời

x \ y	1	2	3
1	0,20	0,05	3λ
2	λ	0,20	0,10
3	0,05	0,00	0,25

a) Xác định tham số λ.
b) Lập các bảng phân phối biên.
c) Lập bảng PP có điều kiện $P(X = x|\, Y = 3)$.

3. Bảng PPXS đồng thời của số lỗi vẽ màu X và số lỗi đúc Y của một loại sản phẩm nhựa ở một công ty được cho bởi

x \ y	0	1	2
0	0,59	0,06	0,03
1	0,10	0,05	0,01
2	0,06	0,05	0,01
3	0,02	0,02	0,00

Hai biến có độc lập không? Tính XS để tổng số lỗi vẽ màu và lỗi đúc lớn hơn 4. Nếu ta biết trên sản phẩm có 2 lỗi vẽ màu thì XS để không có lỗi đúc bằng bao nhiêu?

4. Cho luật PP của một biến 2 chiều như sau (X, Y)

$x \backslash y$	2	3	5
1	0,1	0	0,1
4	0,2	0,5	0,1

Tìm luật PPXS của các hàm $X+Y$ và XY, sau đó tính các kỳ vọng và phương sai.

5. Hai máy tự động hoạt động độc lập, với XS để từng máy sản xuất ra sản phẩm tốt tương ứng là p_1 và p_2. Giả sử mỗi máy làm được 2 sản phẩm và gọi X và Y là số sản phẩm tốt của các máy tương ứng. Tìm PPXS đồng thời của hai biến X, Y.

6. Biến ngẫu nhiên 2 chiều có hàm mật độ đồng thời

$$f(x,y) = a(x^2 + y^2), \text{ nếu } x^2 + y^2 \leq 4.$$

Tìm hệ số a, sau đó tìm các mật độ biên và mật độ có điều kiện.

7. Cho hàm mật độ đồng thời của X và Y

$$f(x,y) = a\, e^{-\frac{1}{2}(x^2+2xy+5y^2)}.$$

Xác định hằng số a, sau đó tìm các mật độ biên và mật độ có điều kiện.

8. Cho X và Y là hai biến ngẫu nhiên có phân phối đều trên $(0; 2)$. Tính XS đồng thời $P(XY \leq 1, Y \leq 2X, X \leq 2Y)$.

9. Cho X và Y là hai biến ngẫu nhiên có hàm mật độ đồng thời $f(x,y) = a\left(1 - \sqrt{x^2 + y^2}\right)$, nếu $x^2 + y^2 \leq 1$. Xác định a, sau đó tính $P(X^2 + Y^2 \leq 1/4)$.

10. Một điểm A rơi ngẫu nhiên vào một hình vuông $\mathcal{D}$ có độ dài cạnh bằng 1 và giả sử (X, Y) là toạ độ của A. Biết rằng hàm mật độ độ đồng thời của X và Y là $f(x, y) = 1$, nếu $(x, y) \in \mathcal{D}$. Tính XS để khoảng cách từ A đến cạnh gần nhất của hình vuông không vượt quá 0,3.

11. Lấy ngẫu nhiên một điểm trong tam giác 3 đỉnh O(0,0), A(0,1) và B(4,0). Gọi X và Y là toạ độ của điểm đó.
    a) Tìm hàm mật độ đồng thời của X và Y.
    b) Tìm các mật độ biên.
    c) X và Y có độc lập không?

12. A và B đến cửa hàng cùng một thời điểm. Giả sử thời gian mua hàng của A có phân phối đều trong (10;20) và thời gian mua hàng của B có phân phối đều trong (15;25). Tính XS để A kết thúc mua trước B.

§2 CÁC SỐ ĐẶC TRƯNG CỦA BIẾN NGẪU NHIÊN HAI CHIỀU

2.1 Các số đặc trưng biên

Ta đã biết các *đặc trưng cơ bản* của X và Y ở chương II. Chúng có thể tính được trực tiếp từ các khái niệm mới của PP đồng thời. Do tính tương tự ta viết chỉ viết các công thức cho biến X.

Nếu X rời rạc, *kỳ vọng biên* và *phương sai biên* là:

$$EX = \sum_i \sum_j x_i p\left(x_i, y_j\right) = \sum_i x_i\, p_1(x_i);$$

$$VX = \sum_i \sum_j (x_i - EX)^2 p\left(x_i, y_j\right) = \sum_i \sum_j {x_i}^2 p\left(x_i, y_j\right) - (EX)^2$$

$$= \sum_i (x_i - EX)^2\, p_1(x_i)$$

Nếu X liên tục:

$$EX = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x f(x,y)dxdy = \int_{-\infty}^{+\infty} x f_1(x)dx;$$

$$VX = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x - EX)^2 f(x,y)dxdy$$

$$= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x^2 f(x,y)dxdy - (EX)^2$$

$$= \int_{-\infty}^{+\infty} (x - EX)^2 f_1(x)dx.$$

Ta có thể viết công thức tổng quát hơn, chẳng hạn nếu (X, Y) có

PP đã biết và ta cũng có một hàm $Z = g(X, Y)$, khi đó

$E[g(X, Y)] = \sum_i \sum_j g(x_i, y_j) p\left(x_i, x_j\right)$ (biến rời rạc);

hoặc $= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x, y) f(x, y) dxdy$ (biến liên tục).

Dễ thấy VX chính là $E[g(X, Y)]$ với $g(X, Y) = (x - EX)^2$.

Ngoài ra ta còn có một số đặc trưng quan trọng là các *độ lệch chuẩn biên*:

$$\sigma_X = \sqrt{VX}, \sigma_Y = \sqrt{VY}.$$

2.2 Hiệp phương sai và hệ số tương quan

*** Định nghĩa 1**. *Hiệp phương sai* (*covariance*) *của hai biến X và Y*, ký hiệu là μ_{XY}, được xác định như sau

$$\mu_{XY} = E[(X - EX)(Y - EY)] = E(XY) - EX.EY.$$

Phụ thuộc vào (X, Y) rời rạc hay liên tục, ta có các công thức tính:

$$\mu_{XY} = \sum_i \sum_j x_i y_j p\left(x_i, y_j\right) - EX.EY;$$

$$\mu_{XY} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xyf(x, y) dxdy - EX.EY.$$

Dễ thấy $VX = \mu_{XX}$. Trong chừng mực nào đó, hiệp phương sai được dùng là độ đo quan hệ giữa hai biến: *nếu hiệp phương sai dương* ⇔ *hai biến có khuynh hướng đồng biến* và *nếu chúng có* *khuynh hướng nghịch biến* ⇔ *hiệp phương sai âm*.

*** Định nghĩa 2**. Nếu $\mu_{XY} = 0$ ta nói rằng hai biến X và Y *không tương quan*.

Rõ ràng nếu X, Y độc lập ⇒ chúng không tương quan; điều ngược lại nói chung không đúng.

Nhiều khi để đơn giản các ký hiệu, người ta tập hợp các hiệp phương sai và phương sai của một véc tơ ngẫu nhiên vào một ma trận được gọi là *ma trận hiệp phương sai* (*covariance matrix*); trong trường hợp biến 2 chiều $\boldsymbol{X} = (X, Y)$ đó là ma trận

$$\Gamma = \begin{bmatrix} VX & \mu_{XY} \\ \mu_{XY} & VY \end{bmatrix}.$$

Do hiệp phương sai có nhiều hạn chế (khó xác định miền biến thiên để so sánh quan hệ hai biến, thứ nguyên phức tạp,...), người ta đưa ra khái niệm dưới đây:

*** Định nghĩa 3**. *Hệ số tương quan của hai biến X, Y* (*correlation coefficient*), ký hiệu là ρ_{XY}, được xác định như sau

$$\rho_{XY} = \frac{\mu_{XY}}{\sigma_X \sigma_Y}.$$

Hệ số tương quan có tính chất quan trọng $|\rho_{XY}| \le 1$ và:

- Nếu $\rho_{XY} = \pm 1 \Rightarrow \exists\, a, b: Y = aX + b$ (quan hệ tuyến tính);
- Nếu $\rho_{XY} = 0 \Leftrightarrow X, Y$ không tương quan;
- Nếu $0 < |\rho_{XY}| \le 1 \Rightarrow X, Y$ phụ thuộc;
- Nếu $\rho_{XY} > 0$, ta có tương quan dương (khuynh đồng biến);
- Nếu $\rho_{XY} < 0$, tương quan âm (khuynh nghịch biến)...

*** <u>Thí dụ 1</u>**. Từ bảng PPXS đồng thời của X và Y

$x \backslash y$	1	2	3
1	0,10	0,25	0,15
2	0,15	0,00	0,35

hãy tính hiệp phương sai và hệ số tương quan của chúng.

Giải. Đầu tiên ta tính các đặc số biên: $EX = 1{,}5$; $VX = 0{,}25$;

$EY = 2{,}25$; $VY = 0{,}6845$; sau đó tính $E(XY) = 3{,}45$. Từ đó:

$$\mu_{XY} = E(XY) - EX.EY = 3{,}45 - 1{,}5.2{,}25 = 0{,}075;$$

$$\rho_{XY} = \frac{\mu_{XY}}{\sigma_X \sigma_Y} = \frac{0{,}075}{\sqrt{0{,}25.0{,}6845}} \approx 0{,}1813.$$

*** <u>Thí dụ 2</u>**. Cho biến ngẫu nhiên 2 chiều (X, Y) có mật độ đồng thời

$$f(x, y) = \begin{cases} \dfrac{1}{2\pi}, & 4x^2 + y^2 \le 4; \\ 0, & 4x^2 + y^2 > 4. \end{cases}$$

Chứng tỏ X, Y phụ thuộc và tính hiệp phương sai μ_{XY}.

Giải. Trước hết ta tính các mật độ biên:

$$f_1(x) = \int_{-\infty}^{+\infty} f(x,y)dy = \begin{cases} \frac{2}{\pi}\sqrt{1-x^2}, |x| \le 1; \\ \qquad 0, \quad |x| > 1. \end{cases}$$

$$f_2(y) = \int_{-\infty}^{+\infty} f(x,y)dx = \begin{cases} \frac{1}{2\pi}\sqrt{4-y^2}, |y| \le 2; \\ \qquad 0, \quad |y| > 2. \end{cases}$$

Vì $f(x,y) \ne f_1(x)\, f_2(y)$ nên X và Y phụ thuộc. Do $f_1(x), f_2(y)$ là các hàm chẵn nên $EX = EY = 0$, từ đó

$$\mu_{XY} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xyf(x,y)dxdy$$

$$= \frac{1}{2\pi}\int_{-1}^{1} xdx \int_{-2\sqrt{1-x^2}}^{2\sqrt{1-x^2}} ydy = 0.$$

Rõ ràng X và Y không tương quan, nhưng vẫn phụ thuộc.

2.3 Các số đặc trưng có điều kiện

Dùng các khái niệm xác suất có điều kiện và mật độ có điều kiện ở tiết trước ta có thể định nghĩa các đặc trưng có điều kiện, chẳng hạn *kỳ vọng có điều kiện* của X với điều kiện $Y = y$:

$$E(X|y_k) = \sum_i x_i P(X = x_i | Y = y_k) \ (X \text{ rời rạc}),$$

$$E(X|y) = \int_{-\infty}^{+\infty} x\varphi(x|y)\, dx \ (X \text{ liên tục}).$$

Tương tự có thể định nghĩa $E(Y|x)$ và các *phương sai có điều kiện* tương ứng, chẳng hạn:

$$V(X|y_k) = \sum_i x_i^2 P(X = x_i|Y = y_k) - [E(X|y_k)]^2 \text{ (}X\text{ rời rạc),}$$

$$V(X|y) = \int_{-\infty}^{+\infty} x^2\varphi(x|y)\,dx - [E(X|y)]^2 \text{ (}X\text{ liên tục).}$$

Kỳ vọng có điều kiện $E(Y|x) = E(Y|X = x)$ là hàm của x, trong thống kê người ta gọi là *hàm hồi quy* của Y đối với X và đồ thị của hàm trên mặt phẳng toạ độ Đề-các được gọi là đường hồi quy.

Để ý kỳ vọng có điều kiện $E(Y|X)$, cũng như các đặc trưng có điều kiện khác, là biến ngẫu nhiên và đến lượt mình có thể có các đặc số tương ứng.

*** <u>Thí dụ 3</u>**. Cho bảng PP đồng thời của X và Y

x \ y	1	2	3
2	0,10	0,25	0,15
4	0,15	0,00	0,35

Tính các kỳ vọng và phương sai có điều kiện $E(X|1)$, $E(Y|4)$, $V(X|1)$, $V(Y|4)$.

Giải. Dễ thấy:

$$P(X = 2|Y = 1) = \frac{p_{11}}{p_2(1)} = \frac{0{,}10}{0{,}25} = 0{,}4;$$

$$P(X = 4|Y = 1) = \frac{p_{21}}{p_2(1)} = \frac{0{,}15}{0{,}25} = 0{,}6.$$

Từ đó $\qquad E(X|1) = 2.0{,}4 + 4.0{,}6 = 3{,}2;$

$$V(X|1) = 2^2.0{,}4 + 4^2.0{,}6 - 3{,}2^2 = 0{,}96.$$

Tiếp theo

$$P(Y = 1|X = 4) = P(Y = 1|4) = \frac{p_{21}}{p_1(2)} = \frac{0{,}15}{0{,}5} = 0{,}3;$$

$$P(Y = 2|4) = \frac{p_{22}}{p_1(2)} = \frac{0}{0{,}5} = 0;$$

$$P(Y = 3|4) = \frac{p_{32}}{p_1(2)} = \frac{0{,}35}{0{,}5} = 0{,}7.$$

Và ta có: $\qquad E(Y|4) = 1.0{,}3 + 3.0{,}7 = 2{,}4;$

$$V(Y|4) = 1.0{,}3 + 3^2.0{,}7 - 2{,}4^2 = 0{,}84.$$

*** <u>Thí dụ 4</u>**. Cho X và Y có hàm mật độ đồng thời

$$f(x,y) = \begin{cases} \frac{1}{x}, & 0 < y \le x \le 1; \\ 0, & \text{nếu trái lại.} \end{cases}$$

Tìm hàm mật độ có điều kiện $\psi(y|x)$ và tính $P(X^2 + Y^2 \le 1)$.

Giải. Đầu tiên ta tìm hàm mật độ biên

$$f_1(x) = \begin{cases} 1, & 0 < x \le 1; \\ 0, & \text{nếu trái lại.} \end{cases}$$

Từ đó

$$\psi(y|x) = \frac{f(x,y)}{f_1(x)} = \begin{cases} \frac{1}{x}, & 0 < y \leq x \leq 1; \\ 0, & \text{nếu trái lại}; \end{cases}$$

(PP có điều kiện là phân phối đều trên $(0, x)$).

Theo tính chất (iii) của hàm mật độ

$$P(X^2 + Y^2 \leq 1) = \iint_{\mathcal{D}} \frac{dx}{x},$$

với $\mathcal{D}$ là miền $\{(x, y): 0 < y \leq x \leq 1,\ x^2 + y^2 \leq 1\}$. Dùng toạ độ cực ta có

$$P(X^2 + Y^2 \leq 1) = \int_0^1 dr \int_0^{\pi/4} \frac{r d\varphi}{r\, cos\varphi} = \ln\left(\left|tan \frac{3\pi}{8}\right|\right) = \ln\left(1 + \sqrt{2}\right).$$

*** <u>Thí dụ 5</u>**. Điểm ngẫu nhiên (X, Y) rơi đồng khả năng vào miền e-líp có trục đối xứng nằm trên trục toạ độ và độ dài các bán trục là a và b.

a) Xác định mật độ XS của từng toạ độ và mật độ XS có điều kiện.

b) Tính hiệp phương sai của X và Y.

Giải. a) Dễ thấy hàm mật độ đồng thời của X và Y

$$f(x, y) = \begin{cases} \frac{1}{\pi ab}, & \frac{x^2}{a^2} + \frac{y^2}{b^2} \leq 1; \\ 0, & \frac{x^2}{a^2} + \frac{y^2}{b^2} > 1. \end{cases}$$

Rõ ràng $f_1(x) \neq 0$ chỉ nếu $|x| \leq a$, khi đó

$$f_1(x) = \frac{2\sqrt{a^2-x^2}}{\pi a^2}, |x| \leq a; \text{tương tự } f_2(y) = \frac{2\sqrt{b^2-y^2}}{\pi b^2}, |y| \leq b.$$

Từ đó:

$$\varphi(x|y) = \frac{f(x,y)}{f_2(y)} = \begin{cases} \dfrac{b}{2a\sqrt{b^2-y^2}}, & \dfrac{x^2}{a^2}+\dfrac{y^2}{b^2} \leq 1; \\ 0, & \dfrac{x^2}{a^2}+\dfrac{y^2}{b^2} > 1. \end{cases}$$

$$\psi(y|x) = \frac{f(x,y)}{f_1(x)} = \begin{cases} \dfrac{a}{2b\sqrt{a^2-x^2}}, & \dfrac{x^2}{a^2}+\dfrac{y^2}{b^2} \leq 1; \\ 0, & \dfrac{x^2}{a^2}+\dfrac{y^2}{b^2} > 1. \end{cases}$$

b) Do miền lấy tích phân đối xứng nên $EX = EY = 0$, nên

$$\mu_{XY} = \iint_{\mathbb{R}^2} xyf(x,y)\,dxdy,$$

với để ý hàm $f(x,y)$ chỉ khác 0 trong e-líp $\dfrac{x^2}{a^2}+\dfrac{y^2}{b^2} \leq 1$. Để tính tích phân ta đổi biến sang toạ độ cực suy rộng $\dfrac{1}{\pi ab}$

$$x = arcos\varphi, y = brsin\varphi,$$

$$\mu_{XY} = \int_0^{2\pi}\int_0^1 abr^2\,cos\varphi sin\varphi\,\frac{1}{\pi ab}\,abr\,drd\varphi = 0.$$

Như vậy hai biến X và Y không tương quan, nhưng chúng phụ thuộc vì $f(x,y) \neq f_1(x)\,f_2(y)$.

*** Tính chất của kỳ vọng có điều kiện**

(i) Với mọi hàm $g(.)$ liên tục $E[g(X)Y|X] = g(X)E(Y|X)$;

(ii) $E(X_1 + X_2|X) = E(X_1|X) + E(X_2|X)$;

(iii) Nếu X, Y độc lập $E(Y|X) = EY$;

(iv) $E[E(Y|X)] = EY$.

2.4 Phân phối chuẩn hai chiều

Ta dùng các ký hiệu rút gọn:

$$\mu_X = EX, \mu_Y = EY, \sigma_X^2 = VX, \sigma_Y^2 = VY, \rho = \rho_{XY}.$$

Từ đó có thể xác định *hàm mật độ chuẩn đồng thời* của hai biến X và Y, ký hiệu là $\mathcal{N}(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho)$, có dạng

$$f(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \times$$

$$\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2 - 2\rho\frac{x-\mu_X}{\sigma_X}.\frac{y-\mu_Y}{\sigma_Y}\right]\right\}.$$

Để ý $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$, $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$ và nếu X và Y không tương quan ($\rho = 0$), thì giả thiết chuẩn cho phép kết luận chúng độc lập (vì có thể chứng minh dễ dàng $f(x,y) = f_1(x)f_2(y)$).

Sử dụng khái niệm véc tơ – ma trận

$$\boldsymbol{x} = \begin{pmatrix} x \\ y \end{pmatrix}, \boldsymbol{\mu} = \begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}, \boldsymbol{\Gamma} = \begin{pmatrix} \sigma_X^2 & \rho \\ \rho & \sigma_Y^2 \end{pmatrix},$$

ta có biểu diễn

$$f(\boldsymbol{x}) = f(x, y) = \frac{1}{2\pi\sqrt{det\,\boldsymbol{\Gamma}}} \exp\left\{-\frac{1}{2}\,(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{\Gamma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\}.$$

Ở đây det là ký hiệu định thức ma trận, T – phép chuyển vị. Biến chuẩn 2 chiều trong trường hợp này ký hiệu $\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Gamma})$.

Trường hợp tổng quát nếu $\boldsymbol{X}$ là véc tơ n chiều và $\sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Gamma})$, khi đó hàm mật độ có dạng

$$f(\boldsymbol{x}) = (2\pi)^{-n/2}\,(det\,\boldsymbol{\Gamma})^{-1/2} \exp\left\{-\frac{1}{2}\,(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{\Gamma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\}.$$

*** <u>Thí dụ 6</u>**. Cho biến 2 chiều $(X, Y) \sim \mathcal{N}(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho)$. Tính các kỳ vọng và phương sai có điều kiện.

Giải. Dễ dàng tìm được $f_2(y)$ từ mật độ đồng thời ở trên

$$f_2(y) = \frac{1}{\sigma_Y\sqrt{2\pi}}\, e^{-\frac{1}{2\sigma_Y^2}(x-\mu_Y)^2},$$

từ đó

$$\varphi(x|y) = \frac{f(x,y)}{f_2(y)}$$

$$= \frac{1}{\sqrt{2\pi}\sigma_X\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2\sigma_X^2(1-\rho^2)}\left[x-\mu_X-\rho\frac{\sigma_X}{\sigma_Y}(y-\mu_Y)\right]^2\right\}.$$

Biểu thức trên chính là hàm mật độ của phân phối chuẩn

$$\mathcal{N}\left(\mu_X + \rho\frac{\sigma_X}{\sigma_Y}(y - \mu_Y);\ \sigma_X^2(1 - \rho^2)\right),$$

từ đó:

$$E(X|y) = \mu_X + \rho\ \frac{\sigma_X}{\sigma_Y}\ (y - \mu_Y); V(X|y) = \sigma_X^2(1 - \rho^2).$$

Tương tự:

$$E(Y|x) = \mu_Y + \rho\ \frac{\sigma_Y}{\sigma_X}\ (x - \mu_X); V(Y|x) = \sigma_Y^2(1 - \rho^2).$$

BÀI TẬP

1. Cho X và Y là toạ độ của một điểm ngẫu nhiên có PP đều trong hình chữ nhật giới hạn bởi các đường $x = a, x = b, y = c, y = d$ $(b > a, d > c)$. Tìm hàm mật độ đồng thời của X và Y, sau đó tính kỳ vọng của các toạ độ của điểm ngẫu nhiên đó.

2. Cho X và Y có bảng PPXS

x \ y	1	2	3	4
1	0,16	0,08	0,10	0,20
2	0,08	0,10	0,14	0,14

Tìm các kỳ vọng biên, kỳ vọng có điều kiện và các phương sai tương ứng.

3. Cho X và Y có hàm mật độ đồng thời

$$f(x,y) = \frac{A}{\pi^2(16+x^2)(25+y^2)}.$$

a) Xác định A.
b) X và Y có độc lập không?
c) Tìm các đặc trưng biên của X và Y.

4. Xác định XS rơi của điểm (X, Y) vào miền $\{1 \le x \le 2, 1 \le y \le 2\}$, nếu hàm PP của (X, Y) có dạng $(a > 0)$

$$F(x,y) = \begin{cases} 1 - a^{-x^2} - a^{-2y^2} + a^{-(x^2+2y^2)}, & x \ge 0, y \ge 0; \\ 0, & x < 0 \text{ hay } y < 0. \end{cases}$$

Sau đó tính các đặc trưng có điều kiện.

5. Bảng PPXS đồng thời của số lỗi vẽ màu X và số lỗi đúc Y của một loại sản phẩm nhựa ở một công ty được cho bởi

x \ y	0	1	2
0	0,59	0,06	0,03
1	0,10	0,05	0,01
2	0,06	0,05	0,01
3	0,02	0,02	0,00

Tính các kỳ vọng biên và ma trận tương quan của (X, Y).

6. Cho luật PP của một biến 2 chiều như sau (X, Y)

x \ y	2	3	5
1	0,1	0	0,1
4	0,2	0,5	0,1

Tìm luật PPXS của các hàm $X+Y$ và XY, sau đó tính các kỳ vọng và phương sai.

7. Biến ngẫu nhiên 2 chiều (X, Y) có PP chuẩn trong một hình vuông có cạnh a. Các đường chéo của hình vuông trùng với các trục toạ độ.
a) Xác định mật độ của (X, Y).
b) Tính các mật độ biên và mật độ có điều kiện.
c) Tính ma trận tương quan của (X, Y).
d) X và Y có phụ thuộc không?

8. Biến ngẫu nhiên 2 chiều có hàm mật độ đồng thời

$$f(x, y) = a(x^2 + y^2), \text{ nếu } x^2 + y^2 \leq 4.$$

Tìm hệ số a, sau đó tìm các đặc trưng biên và đặc trưng có điều kiện.

9. Cho hàm mật độ đồng thời của X và Y

$$f(x, y) = a\, e^{-\frac{1}{2}\left(x^2+2xy+5y^2\right)}.$$

Xác định hằng số a, sau đó tìm các đặc trưng biên và đặc trưng có điều kiện.

10. Cho X và Y là hai biến ngẫu nhiên có hàm mật độ đồng thời $f(x, y) = 1/2x^2y$, nếu $x \geq 1$ và $x \geq y \geq 1/x$. Tìm các kỳ vọng có điều kiện.

11. Cho X và Y là hai biến ngẫu nhiên: Y có phân phối đều trong (0;10); còn hàm mật độ có điều kiện $\varphi(x|y) = 1/y$, với $0 < x < y < 10$. Tính
    a) $E(X|Y = y)$; b) EX.