California Polytechnic State University, San Luis Obispo

JAY L. DEVORE

Tài liệu môn học

Probability and Statistics for Engineering and Sciences

XÁC SUẤT THỐNG KÊ ỨNG DỤNG

Biên soạn:
Chương 2,3 - Nguyễn Ngọc Tứ

Bộ môn Toán - ĐH SPKT, Tp. Hồ Chí Minh - Năm 2017

Mục lục

1 TỔNG QUAN VÀ THỐNG KÊ MÔ TẢ **3**

1.1 Tổng thể, mẫu và qui trình 3

1.2 Phương pháp trực quan và biểu đồ trong Thống kê mô tả 3

1.3 Các số đo đặc trưng vị trí 3

2 PHÉP TÍNH XÁC SUẤT **4**

2.1 Không gian mẫu và biến cố 4

2.2 Các tiên đề và tính chất của xác suất 6

2.3 Giải tích tổ hợp 7

2.4 Xác suất có điều kiện 9

2.5 Sự độc lập 11

3 BIẾN NGẪU NHIÊN RỜI RẠC VÀ PHÂN PHỐI XÁC SUẤT **13**

3.1 Biến ngẫu nhiên 13

3.2 Phân phối xác suất của biến ngẫu nhiên rời rạc 14

3.3 Kỳ vọng và phương sai 17

3.4 Phân phối nhị thức 18

3.5 Phân phối siêu bội và phân phối nhị thức âm 21

3.6 Phân phối Poisson 23

4 BIẾN NGẪU NHIÊN LIÊN TỤC VÀ PHÂN PHỐI XÁC SUẤT **25**

4.1 Hàm mật độ xác suất 25

4.2 Hàm phân phối tích lũy và các số đặc trưng 26

4.3 Phân phối chuẩn 30

4.4 Phân phối mũ và Gamma 33

4.5 Một số phân phối liên tục khác 36

MỞ ĐẦU

Chương 1

TỔNG QUAN VÀ THỐNG KÊ MÔ TẢ

Giới thiệu

1.1 Tổng thể, mẫu và qui trình

1.2 Phương pháp trực quan và biểu đồ trong Thống kê mô tả

1.3 Các số đo đặc trưng vị trí

Chương 2

PHÉP TÍNH XÁC SUẤT

Giới thiệu

2.1 Không gian mẫu và biến cố

Không gian mẫu của một phép thử

Một **phép thử** là một hành động hay một quá trình mà cho kết quả một cách ngẫu nhiên.

Định nghĩa 2.1.1. **Không gian mẫu** của một phép thử là tập hợp tất cả các kết quả có thể xảy ra của phép thử, thông thường được kí hiệu là $\Omega, \mathcal{S}$.

Ví dụ 2.1 Một phép thử đơn giản nhất là chỉ cho kết quả có hai kết quả. Chẳng hạn, tung một đồng xu thì chỉ có hai khả năng xảy ra là xuất hiện mặt sấp H hoặc mặt ngữa T. Khi đó, ta có không gian mẫu của phép thử tung đồng xu là $\Omega = \{H, T\}$. Một trường hợp khác trong thực tế là giới tính của một trẻ sơ sinh mà có không gian mẫu là $\Omega = \{M, F\}$.

Ví dụ 2.2 Nếu ta tung đồng xu ba lần và ghi lại các kết quả xảy ra của mỗi lần thì một kết quả của cả ba lần tung là dãy H hoặc T. Vì thế,

$$\Omega = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}.$$

Ví dụ 2.3 Gieo hai con xúc xắc (6 mặt) đồng thời thì ta có 36 khả năng xảy ra với không gian mẫu là

$$\Omega = \{(i, j),\ \text{với } i, j = 1, \ldots, 6\}.$$

Định nghĩa 2.1.2. Một **biến cố** là một tập hợp của một số kết quả trong không gian mẫu Ω. Một biến cố **đơn** nếu nó chỉ chứa một kết quả và một biến cố **kép** nếu nó chứa nhiều hơn một kết quả.

Ví dụ 2.4 Xét một phép thử về việc chuyển hướng của ba phương tiện giao thông tại một ngã ba, nghĩa là xe sẽ rẽ trái (L) hoặc phải (R) tại ngã ba. Không gian mẫu sẽ gồm 8 trường hợp

$$\Omega = \{LLL, LLR, LRL, LRR, RLL, RLR, RRL, RRR\}.$$

Do đó, ta có được 8 biến cố đơn và một số biến cố kép, chẳng hạn

$$A = \{RLL, LRL, LLR\} = \{\text{biến cố chỉ một xe rẽ phải}\}$$
$$B = \{LLL, RLL, LRL, LLR\} = \{\text{biến cố nhiều nhất một xe rẽ phải}\}$$
$$C = \{LLL, RRR\} = \{\text{biến cố cả 3 xe rẽ cùng hướng}\}$$

Ví dụ 2.5 Tiếp tục ví dụ 2.1, ta có 36 biến cố đơn $E_1 = (1,1), \ldots, E_{36} = (6,6)$. Ví dụ một số biến cố kép

$$\begin{aligned} A &= \{(1,1),(2,2),(3,3),(4,4),(5,5),(6,6)\} \\ &= \{\text{biến cố 2 xúc xắc xuất hiện mặt giống nhau}\} \\ B &= \{(1,3),(2,2),(3,1)\} = \{\text{biến cố tổng 2 mặt bằng 4}\}. \end{aligned}$$

Mối quan hệ của lý thuyết tập hợp

Định nghĩa 2.1.3. i) Phần bù của biến cố A, kí hiệu A', là tập hợp tất cả các kết quả trong không gian mẫu Ω mà không chứa trong biến cố A.

ii) Hợp của hai biến cố A và B, kí hiệu $A \cup B$, đọc là "A hoặc B", là biến cố chứa tất cả các biến cố của A hoặc của B, hoặc đồng thời của cả 2 biến cố, (nghĩa là chỉ cần ít nhất một biến cố xảy ra.)

iii) Giao của hai biến cố A và B, kí hiệu $A \cap B$, đọc là "A và B", là biến cố chứa tất cả các kết quả của cả 2 biến cố A và B.

iv) Biến cố rỗng, kí hiệu $\emptyset$, là biến cố không chứa kết quả nào của phép thử. Khi $A \cap B = \emptyset$ thì ta gọi A và B là hai biến cố rời nhau.

Ví dụ 2.6 Tiếp tục ví dụ 2.1, ta có các biến cố sau $A = \{1,2,3,4\}, B = \{3,4,5,6\}, C = \{1,3,5\}, D = \{6\}$. Khi đó,

$$A' = \{5,6\}, A \cup B = \{1,2,3,4,5,6\}, A \cup C = \{1,2,3,4,5\}, A \cap B = \{3,4\}$$

$$A \cap C = \{1,3\}, (A \cap C)' = \{1,3\}, A \cap D = \emptyset, C \cap D = \emptyset.$$

2.2 Các tiên đề và tính chất của xác suất

Cho một phép thử và một không gian mẫu tương ứng Ω, xác suất của một biến cố A là độ đo chính xác khả năng xảy ra của biến cố A đó.

Tiên đề 2.2.1. *1. **Tính không âm:** $P(A) \geq 0$, với biến cố A bất kì.*

*2. **Tính cộng tính:** Nếu A và B là hai biến cố rời nhau thì*

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

Hơn thế nữa, nếu $A_1, A_2, \ldots$ là tập hợp vô hạn các biến cố rời nhau thì

$$P(A_1 \cup A_2 \cup A_3 \ldots) = \sum_{j=1}^{\infty} P(A_i).$$

*3. **Tính chuẩn hóa:** $P(\Omega) = 1$.*

Mệnh đề 2.2.2. *1. $P(\emptyset) = 0$.*

2. Cho biến cố A bất kì, $P(A) + P(A') = 1$.

3. $P(A) \leq 1$.

4. Cho ba biến cố A, B và C bất kì

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(A \cap B). \\ P(A \cup B \cup C) &= P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) \\ &\quad - P(B \cap C) + P(A \cap B \cap C). \end{aligned}$$

Ví dụ 2.7 Trong một thị trấn, 60% hộ gia đình sử dụng dịch vụ Internet của công ty **A**, 80% hộ gia đình sử dụng dịch vụ truyền hình cáp cũng của công ty **A**, và 50% hộ gia đình sử dụng cả hai dịch vụ của công ty này. Chọn ngẫu nhiên một hộ gia đình, tính xác suất để hộ gia đình này sử dụng ít nhất một dịch vụ của công ty **A** và xác suất để hộ gia đình này chỉ sử dụng duy nhất một loại dịch vụ?

Đặt $A = \{\text{sử dụng Internet}\}$ và $B = \{\text{sử dụng truyền hình cáp}\}$, ta có

$$P(A) = 0.6, P(B) = 0.8 \text{ và } P(A \cap B) = 0.5.$$

Khi đó,

$$\begin{aligned}
&P(\text{biến cố sử dụng ít nhất một loại dịch vụ}) \\
&\quad = P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.6 + 0.8 - 0.5 = 0.9. \\
&P(\text{sử dụng chỉ một dịch vụ}) \\
&\quad = P(A \cup B) - P(A \cap B) = 0.9 - 0.5 = 0.4.
\end{aligned}$$

Định nghĩa 2.2.3. Nếu một không gian mẫu có n kết quả có thể xảy ra như nhau (nghĩa là mỗi biến cố đơn có xác suất như nhau) thì xác suất của một biến cố A bất kì xảy ra được cho bởi

$$P(A) = \frac{\text{Số lượng kết quả chứa trong A}}{n} = \frac{n(A)}{n}.$$

Ví dụ 2.8 Nam có 6 sách tiểu thuyết và 6 sách khoa học giả tưởng chưa đọc. Ba quyển sách đầu tiên của mỗi loại là sách bìa cứng và 3 sách còn lại là bìa mềm. Nam chọn ngẫu nhiên một sách tiểu thuyết và một sách khoa học giả tưởng mang đi du lịch. Tính xác suất Nam chọn được cả hai sách bìa mềm?

Giả sử ta đánh số $1, \ldots, 6$ chọn sách tiểu thuyết và tương tự cho sách khoa học giả tưởng. Khi đó, ta có 36 khả năng chọn được hai quyển sách với xác suất như nhau. Và số cách chọn được hai sách bìa mềm là 9 cách như sau $\{(4,4), (4,5), (4,6), (5,4), (5,5), (5,6), (6,4), (6,5), (6,6)\}$. Vậy xác suất của biến cố A chọn hai sách bìa mềm là

$$P(A) = \frac{n(A)}{n} = \frac{9}{36}.$$

2.3 Giải tích tổ hợp

Mệnh đề 2.3.1. ***Quy tắc nhân:*** *Giả sử ta có một tập hợp có thứ tự gồm có k phần tử và có n_1 cách chọn cho phần tử đầu tiên; với mỗi cách chọn phần tử đầu tiên, ta lại có n_2 cách chọn phần tử thứ hai;...; với mỗi cách chọn phần tử thứ $k-1$, ta lại có n_k cách chọn phần tử thứ k. Do đó có $n_1 n_2 \ldots n_k$ cách chọn k phần tử.*

Ví dụ 2.9 Một người chủ nhà yêu cầu cung cấp dịch vụ điện và nước. Có 12 đơn vị dịch vụ về nước và 9 đơn vị về điện. Hỏi có bao nhiêu cách để người chủ nhà chọn dịch vụ điện và nước?

Có 12 cách chọn dịch vụ về nước và với mỗi cách chọn dịch vụ về nước thì người chủ nhà lại có 9 cách chọn dịch vụ về điện. Vì thế người chủ nhà có $12 \times 9 = 108$ cách.

Định nghĩa 2.3.2. Giả sử ta xét cách chọn k phần tử từ một tập hợp có n phần tử khác nhau. Nếu thứ tự chọn lựa có kể đến thì cách chọn lựa này gọi là một **hoán vị**; ngược lại nếu không kể đến thứ tự chọn lựa thì ta gọi là một **tổ hợp**, kí hiệu là $C_{k,n}$ hoặc $\begin{pmatrix} n \\ k \end{pmatrix}$.

Giả sử ta có n phần tử khác nhau, cho k là một số dương và $k \leq n$. Ta đếm số cách khác nhau từ việc chọn k phần tử từ tập hợp n phần tử và sắp xếp thành một dãy, nghĩa là số lượng dãy của k phần tử phân biệt. Khi đó, ta có n cách chọn phần tử đầu tiên. Với cách chọn phần tử đầu tiên, ta chỉ còn $n-1$ cách chọn phần tử thứ hai, v.v.... Khi ta chọn phần tử thứ k thì ta đã chọn được $k-1$ phần tử, vì thế ta sẽ có $n-(k-1)$ cách chọn phần tử cuối cùng. Do đó, theo quy tắc nhân, số lượng các dãy có thể xảy ra mà ta gọi là $k-$**hoán vị** hay còn gọi là **chỉnh hợp** chập k của n, là

$$\begin{aligned} n(n-1)\dots(n-k+1) &= \frac{n(n-1)\dots(n-k+1)(n-k)\dots 2\cdot 1}{(n-k)\dots 2\cdot 1} \\ &= \frac{n!}{(n-k)!} := P_{k,n} \end{aligned}$$

Đặc biệt, khi $k=n$, số lượng dãy có thể mà ta gọi là **hoán vị** là

$$n\cdot(n-1)\cdot(n-2)\dots 2\cdot 1 = n!$$

Để đếm số lượng tổ hợp, ta chú ý rằng việc chọn một $k-$hoán vị chính là việc chọn lần đầu tiên của một tổ hợp của k phần tử và sau đó sắp xếp lại chúng có thứ tự. Vì thế, ta có $k!$ cách chọn có thứ tự của k phần tử đã chọn trước và ta chỉ ra rằng số lượng của $k-$hoán vị bằng với số lượng của tổ hợp nhân với $k!$. Do đó, số lượng của tổ hợp là

$$\begin{pmatrix} n \\ k \end{pmatrix} = \frac{n!}{k!(n-k)!}.$$

Ví dụ 2.10 Có mười trợ giảng để chấm thi cho môn giải tích tại một trường đại học. Bài kiểm tra đầu tiên có bốn câu hỏi và giáo sư muốn chọn trợ giảng khác nhau chấm điểm cho từng câu hỏi (mỗi trợ giảng chỉ chấm một câu hỏi). Hỏi có bao nhiêu cách chọn trợ giảng để chấm thi.

Số cách chọn là số lượng của 4-hoán vị của tập hợp 10 phần tử:

$$P_{k,n} = \frac{10!}{(10-4)!} = 5040.$$

Ví dụ 2.11 Một danh sách bài hát trong iPod thì chứa 100 bài hát và có 10 bài hát của nhóm Beatle. Giả sử danh sách nhạc này phát ngẫu nhiên. Tính xác suất để bài hát của nhóm Beatle xuất hiện ở lần iPod phát ra bài hát thứ năm?

Để bài hát của nhóm Beatle xuất hiện ở lần thứ năm thì bốn lần đầu không xuất hiện bài hát của nhóm Beatle. Do đó, số cách để xuất hiện bài hát của nhóm Beatle ở lần thứ năm là $90 \cdot 89 \cdot 88 \cdot 87 \cdot 10$. Và số cách xuất hiện năm bài hát bất kì là $100 \cdot 99 \cdot 98 \cdot 97 \cdot 96$. Vì thế, ta có xác suất xuất hiện bài hát của nhóm Beatle ở lần thứ 5 là

$$\frac{90 \cdot 89 \cdot 88 \cdot 87 \cdot 10}{100 \cdot 99 \cdot 98 \cdot 97 \cdot 96} =$$

2.4 Xác suất có điều kiện

Định nghĩa 2.4.1. Cho trước hai biến cố A và B sao cho $P(B) > 0$. Khi đó, xác suất của biến cố A với điều kiện biến cố B xảy ra trước được xác định bởi

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \tag{2.1}$$

Ví dụ 2.12 Người ta khảo sát những người đi mua điện thoại ở một cửa hàng điện tử thì nhận thấy rằng có 60% mua thêm thẻ nhớ, 40% mua thêm sạc dự phòng và 30% mua cả hai. Chọn một người đi mua hàng bất kỳ, và gọi

A={mua thêm bộ nhớ}

B={mua thêm sạc}

C={mua thêm cả hai}

Khi đó, $P(A) = 0.6, P(B) = 0.4, P(C) = 0.3$ Giả sử, người này đã mua sạc dự phòng thì xác suất người này cũng mua thẻ nhớ là

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0.3}{0.4} = 0.75$$

Nghĩa là, trong tất cả những người mua sạc dự phòng có 75% trong số đó mua thêm thẻ nhớ. Tượng tự,

$$P(\text{sạc}|\text{thẻ nhớ}) = P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{0.3}{0.6} = 0.5$$

Lưu ý rằng: $P(A|B) \neq P(A); P(B|A) \neq P(B)$.

Công thức nhân xác suất: Từ công thức (2.1), ta có

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

Mở rộng cho n biến cố $A_1, A_2, \ldots, A_n$, ta có

$$P(A_1A_2 \ldots A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1A_2) \ldots P(A_n|A_1 \ldots A_{n-1})$$

Ví dụ 2.13 Một công ty sản xuất máy nghe nhạc cho DVD gồm 3 loại khác nhau: 50% loại I, 30% loại II, 20% loại III. Công ty bảo hành một năm cho các sản phẩm của mình. Qua khảo sát công ty thu thập được số liệu cần bảo hành trong một năm cho 3 loại sản phẩm là: 25% cho loại I, 20% cho loại II và 10% cho loại III.

1. Tính xác suất người mua sản phẩm loại I phải bảo hành trong năm đầu tiên?
2. Người mua hàng chọn ngẫu nhiên một sản phẩm từ 3 loại sản phẩm, tính xác suất sản phẩm này phải bảo hành?
3. Một người đến bảo hành sản phẩm, tính xác suất sản phẩm này thuộc loại nào có khả năng nhất?

Gọi $A_i = \{$sản phẩm loại i, với $i = 1, 2, 3\}$. Gọi $B = \{$sản phẩm phải bảo hành$\}$. Khi đó, ta có:

$P(A_1) = 0.5, P(A_2) = 0.3, P(A_3) = 0.2, P(B|A_1) = 0.25, P(B|A_2) = 0.2, P(B|A_3) = 0.1.$

Do đó, người mua sản phẩm loại I phải bảo hành trong năm đầu tiên là

$$P(A_1B) = P(B|A_1)P(A_1) = 0.25 \cdot 0.5 = 0.125.$$

và một người mua bất kì một trong ba sản phẩm phải bảo hành là

$$P(B) = P(B|A_1) + P(B|A_2) + P(B|A_3) = 0.125 + 0.06 + 0.02 = 0.205$$

Khi đó,

$$P(A_1|B) = \frac{P(A_1B)}{P(B)} = \frac{0.125}{0.205} = 0.61$$
$$P(A_2|B) = \frac{P(A_2B)}{P(B)} = \frac{0.06}{0.205} = 0.29$$
$$P(A_1|B) = \frac{P(A_1B)}{P(B)} = \frac{0.02}{0.205} = 0.1$$

Công thức toàn phần

Cho $A_1, A_2, \ldots, A_k$ là họ các biến cố đầy đủ và đôi một rời nhau. Khi đó, với biến cố B bất kì, ta có

$$\begin{aligned} P(B) &= P(B|A_1)P(A_1) + \ldots + P(B|A_k)P(A_k) \\ &= \sum_{i=1}^{k} P(B|A_i)P(A_i) \end{aligned} \tag{2.2}$$

Định lý 2.4.2. ***Công thức Bayes***

Cho $A_1, A_2, \ldots, A_k$ là họ các biến cố đầy đủ và đôi một rời nhau với xác suất tiên nghiệm $P(A_i), i = 1, \ldots, k$. Khi đó, với biến cố B bất kì với điều kiện $P(B) > 0$, ta có xác suất hậu nghiệm của biến cố A_j với điều kiện biến cố B đã xảy ra là

$$P(A_j|B) = \frac{P(A_jB)}{P(B)} = \frac{P(B|A_j)P(A_j)}{\sum_{i=1}^{k} P(B|A_i)P(A_i)}, \qquad j = 1, \ldots, k. \tag{2.3}$$

Ví dụ 2.14 *Tỉ lệ của một bệnh hiếm.* Thống kê về một loại bệnh hiếm gặp cho thấy trong 1000 người lớn thì chỉ có 1 người mắc bệnh. Để kiểm tra điều này người ta tiến hành xét nghiệm cho kết quả chính xác 99% nếu một người bĩ nhiễm bệnh, và sai số 2% đối với một người không nhiễm bệnh. Chọn ngẫu nhiên một người và xét nghiệm cho kết quả dương tính, tính xác suất người này có bệnh thật sự?

Gọi A_1 = người có bệnh, A_2 = người không bệnh và B = xét nghiệm dương tính. Khi đó, ta có $P(A_1) = 0.001, P(A_2) = 0.999, P(B|A_1) = 0.99$ và $P(B|A_2) = 0.02$. Vì thế, xác suất cần tìm là

$$P(A_1|B) = \frac{P(A_1B)}{P(B)} = \frac{P(B|A_1)P(A_1)}{P(B|A_1)P(A_1) + P(B|A_2)P(A_2)} = 0.047$$

2.5 Sự độc lập

Định nghĩa 2.5.1. Hai biến cố A và B gọi là độc lập nếu $P(A|B) = P(A)$; ngược lại gọi là phụ thuộc.

Mệnh đề 2.5.2. *A và B độc lập nếu và chỉ nếu*

$$P(AB) = P(A) \cdot P(B).$$

Ví dụ 2.15 Một công ty sản xuất máy giặt và máy sấy có tỉ lệ bảo hành sản phẩm trong 1 năm lần lượt là 30% và 10%. Một người mua cả hai sản phẩm này, tính xác suất cả hai máy đều phải bảo hành?

Gọi A là biến cố bảo hành máy giặt và B là biến cố bảo hành máy sấy. Ta có: $P(A) = 0.3, P(B) = 0.1$. Khi đó, xác suất để bảo hành 2 máy là

$$P(AB) = P(A) \cdot P(B) = 0.03.$$

Định nghĩa 2.5.3. Các biến cố $A_1, \ldots, A_n$ gọi là độc lập nếu

$$P(\cap_{i \in S} A_i) = \prod_{i \in S} P(A_i), \quad \text{với S là tập con của } \{1, 2, \ldots, n\}.$$

Chương 3

BIẾN NGẪU NHIÊN RỜI RẠC VÀ PHÂN PHỐI XÁC SUẤT

3.1 Biến ngẫu nhiên

Định nghĩa 3.1.1. Xét một phép thử bất kì với không gian mẫu Ω cho trước, môt **biến ngẫu nhiên** là một quy luật kết hợp giữa một số và một kết quả của không gian mẫu. Theo ngôn ngữ toán học, một biến ngẫu nhiên là một hàm số có miền xác định là không gian mẫu và miền giá trị là tập số thực. Thông thường, ta hay kí kiệu biến ngẫu nhiên bằng chữ in hoa như X, Y, Z.

Ví dụ 3.1 Giả sử tung một đồng xu cân đối, đồng chất. Ta có không gian mẫu là $\Omega = \{H, T\}$. Khi đó, ta định nghĩa một biến ngẫu nhiên X bởi

$$X(H) = 1, X(T) = 0.$$

Định nghĩa 3.1.2. Bất kì một biến ngẫu nhiên nào chỉ nhận các giá trị 0 và 1 thì đều được gọi là **biến ngẫu nhiên Bernoulli**.

Ví dụ 3.2 Xét thí nghiệm kiểm tra 9 pin chỉ dừng lại khi một pin được chấp nhận. Ta có không gian mẫu là $\Omega = \{S, FS, FFS, \ldots\}$. Khi đó, ta định nghĩa biến ngẫu nhiên X là số lần thí nghiệm cho đến khi một pin được chấp nhận.

$$X(S) = 1, X(FS) = 2, X(FFS) = 3, \ldots, X(FFFFFFFFS) = 9.$$

Định nghĩa 3.1.3. Một biến ngẫu nhiên **rời rạc** nếu tập giá trị của nó là tập hữu hạn hay *vô hạn đếm được*.

Một biến ngẫu nhiên gọi là **liên tục** nếu thỏa mãn hai điều kiện sau:

1. Tập giá trị của nó là tập hợp chứa tất cả các giá trị trên một khoảng hay là hợp rời nhau của các khoảng giá trị.

2. Xác suất của một điểm bất kì luôn là 0, ($P(X = c) = 0, \forall c \in \mathbb{R}$).

3.2 Phân phối xác suất của biến ngẫu nhiên rời rạc

Định nghĩa 3.2.1. Phân phối xác suất hay **hàm khối lượng xác suất** của một biến ngẫu nhiên rời rạc X nhận giá trị x bất kì được xác định bởi

$$p(x) = P(X = x) = P(\forall x \in \Omega : X(s) = x).$$

Lưu ý: $p(x) \geq 0$ và $\sum_{\text{tất cả giá trị của x}} p(x) = 1$.

Ví dụ 3.3 Sáu kiện hàng được vận chuyển đến người tiêu dùng. Biết rằng trong các kiện hàng này có một số sản phẩm bị lỗi được cho trong bảng sau

Kiện hàng	1	2	3	4	5	6
Số sản phẩm lỗi	0	2	0	1	2	0

Gọi X là số sản phẩm lỗi khi kiểm tra một kiện hàng. X nhận ba giá trị là 0, 1, và 2. Khi đó, ta có

$$p(0) = P(X = 0) = P(\text{kiện 1 hoặc kiện 3 hoặc kiện 6}) = \frac{3}{6}$$
$$p(1) = P(X = 1) = P(\text{kiện 4}) = \frac{1}{6}$$
$$p(2) = P(X = 2) = P(\text{kiện 2 hoặc kiện 5}) = \frac{2}{6}$$

Ví dụ 3.4 Một nhóm gồm 5 người tham gia hiến máu nhân đạo là a, b, c, d, e trong đó chỉ 2 mẫu máu a và b thuộc nhóm máu O+. Giả sử người ta lấy lần lượt từng mẫu máu cho đến khi được mẫu nhóm O+ thì dừng lại. Gọi Y là số lần chọn mẫu máu để chọn được nhóm O+. Khi đó, ta có hàm khối lượng xác suất của Y là

$p(1) = P(Y = 1) = P(\text{chọn được mẫu a hoặc b}) = \dfrac{2}{5} = 0.4$

$p(2) = P(Y = 2) = P(\text{chọn được mẫu c, d hoặc e đầu, sau đó được a hoặc b})$
$\quad = \dfrac{3}{5} \cdot \dfrac{2}{4} = 0.3$

$p(3) = P(Y = 3) = P(\text{chọn được mẫu c, d hoặc e 2 lần đầu, sau đó được a hoặc b})$
$\quad = \dfrac{3}{5} \cdot \dfrac{2}{4} \cdot \dfrac{2}{3} = 0.2$

$p(4) = P(Y = 4) = P(\text{chọn được mẫu c, d hoặc e trong ba lần đầu, sau đó được a hoặc b})$
$\quad = \dfrac{3}{5} \cdot \dfrac{2}{4} \cdot \dfrac{1}{3} = 0.1$

$p(y) = 0$ nếu $y \neq 1, 2, 3, 4.$

Bảng phân phối xác suất của Y là

y	1	2	3	4
p(y)	0.4	0.3	0.2	0.1

Định nghĩa 3.2.2. Hàm phân phối tích lũy (cdf) $F(x)$ của một biến ngẫu nhiên rời rạc X với hàm khối lượng xác suất $p(x)$ được xác định như sau:

$$F(x) = P(X \leq x) = \sum_{y \leq x} p(y), \forall x \in \mathbb{R}. \tag{3.1}$$

Ví dụ 3.5 Một cửa hàng bán thẻ nhớ có các loại sau: 1 GB, 2 GB, 4 GB, 8 GB hoặc 16 GB. Gọi biến ngẫu nhiên Y là số lượng thẻ nhớ được bán được cho trong bảng phân phối xác suất sau:

y	1	2	4	8	16
p(y)	0.05	0.1	0.35	0.4	0.1

Khi đó, $F(y)$ được xác định như sau

$$\begin{aligned}
F(1) &= P(Y \leq 1) = P(Y = 1) = p(1) = 0.05 \\
F(2) &= P(Y \leq 2) = P(Y = 1, 2) = p(1) + p(2) = 0.15 \\
F(3) &= P(Y \leq 3) = P(Y = 1, 2, 4) = p(1) + p(2) + p(4) = 0.5 \\
F(8) &= P(Y \leq 8) = P(Y = 1, 2, 4, 8) = p(1) + p(2) + p(4) + p(8) = 0.9 \\
F(16) &= P(Y \leq 16) = 1.
\end{aligned}$$

Cho giá trị y bất kì, ta xác định giá trị của $F(y)$ bằng với giá trị của F tại giá trị gần nhất với biến ngẫu nhiên Y về phía bên trái của y. Chẳng hạn,

$$F(2.7) = P(Y \leq 2.7) = P(Y \leq 2) = F(2) = 0.15$$
$$F(7.999) = P(Y \leq 7.999) = P(Y \leq 4) = F(4) = 0.5$$

Nếu y nhỏ hơn 1 thì $F(y) = 0$, và nếu y lớn hơn 16 thì $F(y) = 1$. Khi đó, cdf là

$$F(y) = \begin{cases} 0 & y < 1 \\ 0.05 & 1 \leq y < 2 \\ 0.15 & 2 \leq y < 4 \\ 0.5 & 4 \leq y < 8 \\ 0.9 & 8 \leq y < 16 \\ 1 & 16 \leq y \end{cases}$$

Mệnh đề 3.2.3. *Cho hai số thực a và b bất kì với $a \leq b$*

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

trong đó, $a-$ là giá trị lớn nhất có thể của X mà bé hẳn hơn a. Đặc biệt, nếu các giá trị của X là số nguyên và a và b cũng là số nguyên thì

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

Nếu $a = b$ thì $P(X = a) = F(a) - F(a-1)$.

Ví dụ 3.6 Gọi X là số ngày nghỉ ốm trong năm của một nhân viên trong một công ty. Giả sử số ngày phép tối đa của một người là 14. Khi đó, X có thể nhận các giá trị $0, 1, \ldots, 14$. Giả sử $F(0) = 0.58, F(1) = 0.72, F(2) = 0.76, F(3) = 0.81, F(4) = 0.88$, và $F(5) = 0.94$,

$$P(2 \leq X \leq 5) = F(5) - F(1) = 0.22$$

và

$$P(X = 3) = F(3) - F(2) = 0.05.$$

3.3 Kỳ vọng và phương sai

Định nghĩa 3.3.1. Giả sử X là biến ngẫu nhiên rời rạc có tập giá trị D và hàm khối lượng xác suất $p(x)$. **Giá trị trung bình** hay còn gọi là **kì vọng** của biến ngẫu nhiên X, kí hiệu $E(X)$ hoặc μ_X hoặc đơn giản là μ được xác định bởi

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

Ví dụ 3.7 Từ ví dụ 3.5, ta tìm được kì vọng của biến ngẫu nhiên Y là

$$\mu_Y = 1 \cdot 0.05 + 2 \cdot 0.1 + 4 \cdot 0.35 + 8 \cdot 0.4 + 16 \cdot 0.1 = 6.45$$

Mệnh đề 3.3.2. *Giả sử X là biến ngẫu nhiên rời rạc có tập giá trị D và hàm khối lượng xác suất $p(x)$. Khi đó, kỳ vọng của hàm $h(X)$, kí hiệu $E[h(X)]$ được cho bởi*

$$E[h(X)] = \sum_{x \in D} h(x) \cdot p(x)$$

Ví dụ 3.8 Một cửa hàng vi tính nhập 3 laptop về với giá 500$ mỗi máy. Họ sẽ bán mỗi máy này cho khách hàng giá 1000$. Nhưng nếu không bán được thì nhà sản xuất sẽ mua lại mỗi máy với giá 200$. Gọi X là số máy tính cửa hàng đã bán và giả sử rằng $p(0) = 0.1, p(1) = 0.2, p(2) = 0.3, p(3) = 0.4$. Gọi $h(X)$ là lợi nhuận thu được của cửa hàng. Khi đó, ta có

$$h(X) = \text{thu nhập} - \text{chi phí} = 1000X + (3 - X) \cdot 200 - 1500 = 800X - 900.$$

Vì thế, ta có kì vọng lợi nhuận thu được là

$$E[h(X)] = h(0) \cdot p(0) + h(1) \cdot p(1) + h(2) \cdot p(2) + h(3) \cdot p(3) = 700.$$

Nhân xét. Giả sử $h(X)$ là một hàm tuyến tính nghĩa là $h(X) = aX + b$. Khi đó, ta có

$$E[h(X)] = E(aX + b) = a \cdot E(X) + b$$

Định nghĩa 3.3.3. Cho X là biến ngẫu nhiên có hàm khối lượng xác suất $p(x)$. Khi đó, **phương sai** của X, kí kiệu $V(X)$ hoặc σ_X^2, hoặc đơn giản là σ^2 được xác định bởi

$$V(X) = \sum_D (x - EX) \cdot p(x) = E[(X - EX)^2] = E(X^2) - (EX)^2$$

Độ lệch chuẩn của X có công thức là

$$\sigma_X = \sqrt{V(X)}$$

Ví dụ 3.9 Một thư viện giới hạn số lần kiểm tra video cho một người là 6 lần trong cùng một thời điểm. Giả sử một người muốn kiểm tra video và gọi X là số lần kiểm tra. Ta có bảng phân phối xác suất của X

x	1	2	3	4	5	6
$p(x)$	0.3	0.25	0.15	0.05	0.1	0.15

Ta tìm được kì vọng là $EX = 2.85$. Khi đó, ta tính được phương sai của X là

$$\begin{aligned} V(X) = \sigma^2 &= \sum_{x=1}^{6}(x-2.85)^2 \cdot p(x) \\ &= (1-2.85)^2 \cdot 0.3 + \ldots + (6-2.85)^2 \cdot 0.15 = 3.2275 \end{aligned}$$

Độ lệch chuẩn của X là $\sigma = \sqrt{3.2275} = 1.8$

Một cách tính phương sai khác:

$$E(X^2) = \sum_{x=1}^{6} x^2 \cdot p(x) = 1^2 \cdot 0.3 + \ldots + 6^2 \cdot 0.15 = 11.35.$$

Khi đó, $\sigma^2 = 11.35 - 2.85^2 = 3.2275$.

Mệnh đề 3.3.4.

$$V(aX+b) = a^2 \cdot \sigma_X^2 \quad \text{và} \quad \sigma_{aX+b} = |a| \cdot \sigma_X.$$

3.4 Phân phối nhị thức

Định nghĩa 3.4.1. Một phép thử gọi là **phép thử nhị thức** nếu thỏa mãn 2 điều kiện sau đây:

1. Phép thử này gồm n phép thử nhỏ hơn và chúng độc lập với nhau.
2. Mỗi phép thử nhỏ hơn này chỉ có hai khả năng xảy ra, ta kí kiệu là S và F, và xác suất xảy ra mỗi biến cố S là như nhau và ta kí hiệu là p.

Ví dụ 3.10 Tung một đồng xu n lần liên tiếp nhau một cách độc lập. Ta kí hiệu H là biến cố xuất hiện mặt sấp và T là biến cố xuất hiện mặt ngữa. Khi đó, ta thấy phép thử này thỏa mãn 2 điều kiện trên nên nó là một phép thử nhị thức.

Định nghĩa 3.4.2. Biến ngẫu nhiên nhị thức X kết hợp với một phép thử nhị thức chứa n phép thử được xác định bởi

$$X = \text{ số lượng của S trong } n \text{ phép thử.}$$

Chẳng hạn, cho $n = 3$. Khi đó, ta có 8 kết quả có thể

$$SSS \quad SSF \quad SFS \quad SFF \quad FSS \quad FSF \quad FFS \quad FFF$$

Từ định nghĩa của X, ta có $X(SSF) = 2, X(SFF) = 1, \ldots$. Giá trị của X trong n phép thử có thể nhận là $x = 0, 1, 2, \ldots, n$. Chúng ta thường kí hiệu $X \sim \text{Bin}(n, p)$ để chỉ biến ngẫu nhiên X là biến ngẫu nhiên nhị thức thực hiện n phép thử với xác suất thành công p.

Kí hiệu.
Hàm khối lượng xác suất của một biến ngẫu nhiên nhị thức X phụ thuộc vào hai tham số n và p nên ta kí hiệu hàm khối lượng xác suất của X là $b(x; n, p)$.

Chẳng hạn, với $n = 4$ và ta tính xác suất của $SSFS$

$$\begin{aligned} P(SSFS) &= P(S) \cdot P(S) \cdot P(F) \cdot P(S) \quad \text{(các phép thử độc lập)} \\ &= p \cdot p \cdot (1-p) \cdot p \quad \text{(P(S) = p)} \\ &= p^3 \cdot (1-p) \end{aligned}$$

Định lý 3.4.3.

$$b(x; n, p) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & x = 0, 1, 2 \ldots, n \\ 0 & \textit{nếu khác} \end{cases}$$

Ví dụ 3.11 Sáu người uống coca chọn ngẫu nhiên hai loại coca S và F. Gọi $p = P(\text{chọn coca S}) = 0.5$. Khi đó, ta có

$$X = \text{ số người chọn coca S}, X \sim \text{Bin}(6, 0.5).$$

Chẳng hạn, ta tính

$$P(X = 3) = b(3; 6, 0.5) = \binom{6}{3} 0.5^3 0.5^3 = 200.5^6 = 0.313$$

Xác suất ít nhất ba người chọn coca S

$$P(X \geq 3) = \sum_{x=3}^{6} b(x; 6, 0.5) = \sum_{x=3}^{6} \binom{6}{x} 0.5^x 0.5^{6-x} = 0.656$$

và xác suất nhiều nhất một người chọn S

$$P(X \leq 1) = \sum_{x=0}^{1} b(x; 6, 0.5) = 0.109$$

Kí hiệu.
Cho $X \sim \mathrm{Bin}(n, p)$, hàm phân phối tích lũy của X là

$$B(x; n, p) = P(X \leq x) = \sum_{y=0}^{x} b(y; n, p) \quad x = 0, 1, \ldots, n$$

Ví dụ 3.12 Một nhà xuất bản thống kê rằng 20% số lượng bản in ấn bị mắc lỗi. Giả sử gọi X là số sách bị lỗi trong 15 sách chọn ngẫu nhiên. Khi đó, X có phân phối nhị thức với $n = 15$ và $p = 0.2$

1. Xác suất nhiều nhất 8 sách bị lỗi

$$P(X \leq 8) = \sum_{y=0}^{8} b(y; 15, 0.2) = B(8; 15, 0.2) = .999$$

2. Xác suất có chính xác 8 sách bị lỗi

$$P(X = 8) = P(X \leq 8) - P(X \leq 7) = B(8; 15, 0.2) - B(7; 15, 0.2) = 0.003$$

3. Xác suất ít nhất 8 sách bị lỗi

$$P(X \geq 8) = 1 - P(X \leq 7) = 1 - B(7; 15, 0.2) = 1 - 0.996 = 0.004$$

4. Xác suất có từ 4 đến 7 sách bị lỗi

$$P(4 \leq X \leq 7) = P(X \leq 7) - P(X \leq 3) = 0.996 - 0.648 = 0.348$$

Mệnh đề 3.4.4. ***Kì vọng và phương sai của X.***
Cho $X \sim B(n, p)$. Khi đó $EX = n \cdot p$, $\quad V(X) = np(1 - p) = npq$ (với $q = 1 - p$).

3.5 Phân phối siêu bội và phân phối nhị thức âm

PHÂN PHỐI SIÊU BỘI

Một biến ngẫu nhiên có phân phối siêu bội nếu thỏa mãn các điều kiện sau:

1. Tập hợp mẫu chứa N phần tử.
2. Mỗi phần tử có hai khả năng xảy ra là xảy ra (S) hoặc không xảy ra (F) và có M lần xảy ra trong tập hợp mẫu.
3. Một mẫu có n phần tử được chọn một cách không lặp lại sao cho mỗi lần chọn tập n phần tử này đều có khả năng như nhau.

Xét biến ngẫu nhiên X là số lần xảy ra (S) trong mẫu có kích thước n. Phân phối xác suất của X phụ thuộc vào tham số n, M, và N. Vì thế, ta có $P(X = x) = h(x; n, M, N)$.

Ví dụ 3.13 Một trung tâm bảo hành điện máy nhận được 20 đơn hàng trong đó có 8 yêu cầu sửa chữa máy in và 12 yêu cầu thay hộp mực. Sau khi sửa xong, trung tâm chọn ngẫu nhiên 5 đơn hàng và kiểm tra sự hài lòng của khách hàng về dịch vụ của mình. Tính xác suất có $x(x = 0, 1, 2, 3, 4, 5)$ đơn hàng yêu cầu thay hộp mực.

Ta có, tập hợp mẫu là $N = 20$, mẫu kiểm tra là $n = 5$ và số lần xảy ra với S (thay mực) là $M = 12$ và $N - M = 8$. Khi đó, chẳng hạn cho $x = 2$, ta có

$$P(X = x) = h(x; 5, 12, 20) = \frac{\text{số kết quả xảy ra khi X =2}}{\text{số kết quả có thể xảy ra}} = \frac{\begin{pmatrix} 12 \\ 2 \end{pmatrix}\begin{pmatrix} 8 \\ 3 \end{pmatrix}}{\begin{pmatrix} 20 \\ 5 \end{pmatrix}}$$

Mệnh đề 3.5.1. *Nếu X là số lần xảy ra (S) trong mẫu ngẫu nhiên có kích thước n được rút từ một tập hợp cho trước có M (S) và $(N - M)$ (F). Khi đó, hàm phân phối xác suất của X gọi là* ***phân phối siêu bội*** *được cho bởi*

$$P(X = x) = h(x; n, M, N) = \frac{\begin{pmatrix} M \\ x \end{pmatrix}\begin{pmatrix} N-M \\ n-x \end{pmatrix}}{\begin{pmatrix} N \\ n \end{pmatrix}} \tag{3.2}$$

với x là số nguyên thỏa mãn $\max(0, n - N + M) \leq x \leq \min(n, M)$.

Mệnh đề 3.5.2. *Kì vọng và phương sai của biến ngẫu nhiên X có phân phối siêu bội*

$$E(X) = n \cdot \frac{M}{N} \quad V(X) = \frac{N-M}{N-1} \cdot n \cdot \frac{M}{N} \cdot \left(1 - \frac{M}{N}\right)$$

Ví dụ 3.14 Năm con hổ được bắt trong một vùng được đeo định vị và sau đó thả lại tự nhiên. Sau đó, người ta bắt ngẫu nhiên 10 con hổ trong vùng đó kiểm tra xem có bao nhiêu con được định vị. Gọi X là số hổ được gắn định vị trong 10 con hổ. Giả sử rằng trong vùng này chỉ còn 25 con hổ. Tính
$a) X = 2 \quad b) X \leq 2 \quad c) EX, V(X)$?
Ta có: $n = 10, M = 5, N = 25$. Do đó,

$$P(X = x) = h(x; 10, 5, 25) = \frac{\begin{pmatrix} 5 \\ x \end{pmatrix}\begin{pmatrix} 20 \\ 10-x \end{pmatrix}}{\begin{pmatrix} 25 \\ 10 \end{pmatrix}}, \quad x = 0, 1, 2, 3, 4, 5$$

a) $P(X = 2) = h(2; 10, 5, 25) = \dfrac{\begin{pmatrix} 5 \\ 2 \end{pmatrix}\begin{pmatrix} 20 \\ 8 \end{pmatrix}}{\begin{pmatrix} 25 \\ 10 \end{pmatrix}} = 0.385$

b) $P(X \leq 2) = P(X = 0, 1, 2) = \sum_{x=0}^{2} h(x; 10, 5, 25) = 0.699$

c) $EX = 10 \cdot \dfrac{5}{25} = 2, \quad V(X) = \dfrac{15}{24} \cdot 10 \cdot \dfrac{5}{25}\left(1 - \dfrac{5}{25}\right) = 1$

PHÂN PHỐI NHỊ THỨC ÂM

Một biến ngẫu nhiên có phân phối nhị thức âm nếu thỏa mãn các điều kiện sau:

1. Thí nghiệm là dãy các phép thử độc lập.
2. Mỗi phép thử chỉ chứa 2 khả nẳng là thành công (S) hoặc thất bại (F).
3. Khả năng xảy ra S trong mọi phép thử là như nhau và bằng p.
4. Thí nghiệm tiến hành các phép thử sao cho đạt được r lần S, với r là số dương cụ thể.

Giá trị của biến ngẫu nhiên X là số lần F xảy ra trước khi đạt tới lần thứ r thành công (S). X gọi là **biến ngẫu nhiên nhị thức âm** vì số lần thành công là cố định trong khi số lần thực hiện phép thử là ngẫu nhiên.

Mệnh đề 3.5.3. *Hàm khối lượng xác suất của biến ngẫu nhiên nhị thức âm X với tham số r = số lần thành công (S) và $p = P(S)$ là*

$$nb(x; r, p) = \begin{pmatrix} x+r-1 \\ r-1 \end{pmatrix} p^r(1-p)^x \quad x = 0, 1, 2, \ldots$$

Ví dụ 3.15 Một bác sĩ nhi khoa muốn tìm 5 cặp vợ chồng tham gia chương trình ăn uống mới cho đứa con đầu tiên của những cặp vợ chồng này. Gọi $p = P$(chọn ngẫu nhiên được cặp vợ chồng đồng ý tham gia). Cho $p = 0.2$, hãy tìm xác suất để 15 cặp vợ chồng được hỏi tham gia chương trình biết rằng sẽ có 5 cặp đồng ý? Nghĩa là với $S = \{$đồng ý tham gia$\}$, tính xác suất có 10 cặp từ chối trước khi có được cặp thứ 5 đồng ý tham gia.

Ta có: $r = 5, p = 0.2, x = 10$. Khi đó

$$nb(10; 5, 0.2) = \begin{pmatrix} 14 \\ 4 \end{pmatrix} 0.2^5 0.8^{10} = 0.034$$

Xác suất để nhiều nhất có 10 cặp từ chối là

$$P(X \leq 10) = \sum_{x=0}^{10} nb(x; 5, 0.2) = 0.2^5 \sum_{x=0}^{10} \begin{pmatrix} x+4 \\ 4 \end{pmatrix} 0.8^x = 0.164$$

Mệnh đề 3.5.4. *Cho X là biến ngẫu nhiên nhị thức âm với hàm khối lượng xác suất $nb(x; r, p)$. Khi đó, ta có*

$$EX = \frac{r(1-p)}{p} \quad V(X) = \frac{r(1-p)}{p^2}$$

3.6 Phân phối Poisson

Định nghĩa 3.6.1. Một biến ngẫu nhiên rời rạc X gọi là có **phân phối Poisson** với tham số $\mu (\mu > 0)$ nếu hàm mật độ của X có dạng

$$p(x; \mu) = \frac{e^{-\mu} \cdot \mu^x}{x!} \quad x = 0, 1, 2, 3 \ldots$$

Ví dụ 3.16 Gọi X là số lượng chuột bị bắt bởi một cái bẫy trong một khoảng thời gian nhất định. Giả sử X có phân phối Poisson với $\mu = 4.5$. Khi đó, xác suất để chính xác 5 con chuột dính bẫy là

$$P(X = 5) = \frac{e^{-4.5} \cdot 4.5^5}{5!} = 0.1708$$

Xác suất mà nhiều nhất 5 con chuột dính bẫy là

$$P(X \leq 5) = \sum_{x=0}^{5} \frac{e^{-4.5} \cdot 4.5^x}{x!} = e^{-4.5} \left[1 + 4.5 + \frac{4.5^2}{2!} + \ldots + \frac{4.5^5}{5!} \right] = 0.7029$$

Phân phối Poisson như là một giới hạn

Mệnh đề 3.6.2. *Giả sử một biến ngẫu nhiên X có hàm khối lượng xác suất nhị thức là $b(x; n, p)$. Ta cho $n \to \infty$ và $p \to 0$ sao cho $np \to \mu$ với $\mu > 0$. Khi đó, $b(x; n, p) \to p(x, \mu)$.*

Dựa vào mệnh đề này, ta chú ý rằng bất kì phép thử nhị thức có n đủ lớn và p đủ nhỏ thì $b(x; n, p) \simeq p(x, \mu)$ với $\mu = np$. Ở đây, ta có thể áp dụng quy luật này khi $n > 50$ và $np < 5$.

Ví dụ 3.17 Một nhà xuất bản chuẩn bị phát hành một quyển sách tiểu thuyết với yêu cầu sai sót lỗi đánh máy trong mỗi trang sách có ít nhất một lỗi với xác suất là 0.005 và sai sót mỗi trang là độc lập nhau. Giả sử quyển sách này có 400 trang. Tính xác suất có chỉ có chính xác một trang có lỗi? nhiều nhất ba trang có lỗi?

Giả sử S là một trang chứa ít nhất một lỗi và F là một trang lỗi. Khi đó, X là số lượng trang chứa ít nhất một lỗi và X cũng là biến ngẫu nhiên nhị thức với $n = 400, p = 0.005$, vì thế $np = 2$. Khi đó,

$$P(X = 1) = b(1; 400, 0.005) \simeq p(1; 2) = \frac{e^{-2} 2^1}{1!} = 0.270671$$

Giá trị X nếu tính theo phân phối nhị thức là $b(1; 400, 0.005) = 0.270669$, vì thế xấp xỉ này rất tốt. Tương tự,

$$\begin{aligned} P(X \leq 3) &\simeq \sum_{x=0}^{3} p(x, 2) = \sum_{x=0}^{3} \frac{e^{-2} 2^x}{x!} \\ &= 0.135335 + 0.270671 + 0.270671 + 0.180447 = 0.8571 \end{aligned}$$

và gần bằng với giá trị nhị thức là $P(X \leq 3) = 0.8576$.

Mệnh đề 3.6.3. *Nếu X có phân phối Poisson với tham số μ thì $EX = V(X) = \mu$.*

Chương 4

BIẾN NGẪU NHIÊN LIÊN TỤC VÀ PHÂN PHỐI XÁC SUẤT

4.1 Hàm mật độ xác suất

Nhắc lại: Một biến ngẫu nhiên X gọi là liên tục nếu thỏa mãn:

(1) các giá trị có thể có của X nằm trong một khoảng số thực hoặc là hợp của các khoảng rời nhau;

(2) $P(X = c) = 0$ với bất kì số c là một giá trị của X.

Ví dụ 4.1

(1) Đo độ sau của một hồ nước tại một vị trí chọn ngẫu nhiên. Khi đó, gọi $X =$ là độ sâu tại một vị trí thì X là biến ngẫu nhiên liên tục. Khoảng giá trị của X nằm trong vị trí thấp nhất và sâu nhất của hồ nước.

(2) Một hợp chất hóa học được chọn ngẫu nhiên và ta đo mức pH X của nó. Khi đó, X là một biến ngẫu nhiên liên tục vì bất kì mức pH nào cũng nằm trong khoảng 0 và 14.

Định nghĩa 4.1.1. Cho X là một biến ngẫu nhiên liên tục. Khi đó, hàm mật độ xác suất của X là hàm số $f(x)$ sao cho với bất kì hai giá trị a, b nào và $a \leq b$ thì

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

Hơn thế nữa, $f(x)$ phải thỏa mãn hai điều kiện sau

1. $f(x) \geq 0$ với mọi x.
2. $\int_{-\infty}^{\infty} f(x)dx = 1$.

Ví dụ 4.2 Cho X là biến ngẫu nhiên liên tục có hàm mật độ xác suất là

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

Rõ ràng, $f(x) \geq 0$ và $\int_{-\infty}^{\infty} f(x)dx = \int_{0}^{360} f(x)dx = 1$. Chẳng hạn, ta tính

$$P(90 \leq X \leq 180) = \int_{90}^{180} \frac{1}{360}dx = 0.25$$

Định nghĩa 4.1.2. Một biến ngẫu nhiên liên tục X gọi là một **phân phối đều** trên một khoảng $[A, B]$ nếu hàm mật độ của X là

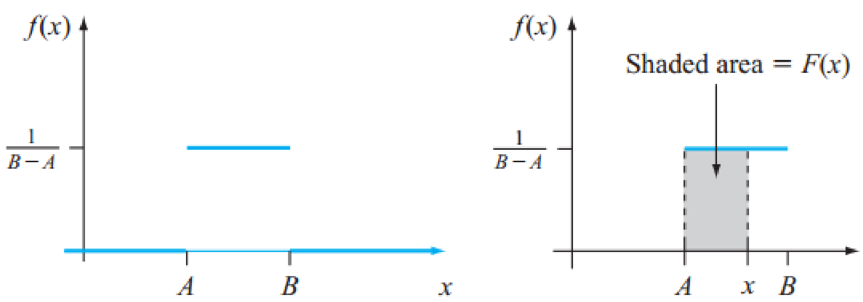
$$f(x; A, B) = \begin{cases} \dfrac{1}{B-A} & A \leq x \leq B \\ 0 & \text{nếu khác.} \end{cases}$$

4.2 Hàm phân phối tích lũy và các số đặc trưng

Định nghĩa 4.2.1. Hàm phân phối tích lũy $F(x)$ của biến ngẫu nhiên liên tục X được xác định bởi: với mọi số thực x thì

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(y)dy.$$

Ví dụ 4.3 Cho X là biến ngẫu nhiên có phân phối đều trên khoảng [A, B]. Hàm mật độ của X chỉ ra trong hình phía dưới.



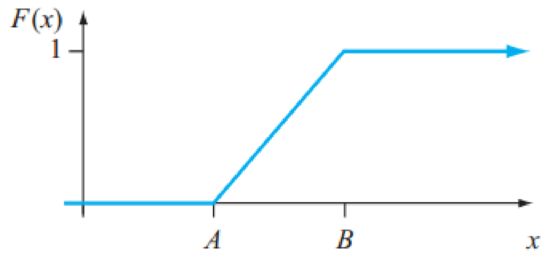
Nếu $x < A, F(x) = 0$ bởi vì không có diện tích trong đồ thị của hàm mật độ phía bên trái $x < A$. Nếu $x \geq B, F(x) = 1$ bởi vì toàn bộ diện tích tích lũy phía bên trái của $x > B$. Cuối cùng, nếu $A \leq x \leq B$,

$$F(x) = \int_{-\infty}^{x} f(y)dy = \int_{A}^{x} \frac{1}{B-A}dy = \frac{x-A}{B-A}$$

Khi đó, ta có hàm phân phối tích lũy là

$$F(x) = \begin{cases} 0 & x < A \\ \dfrac{x-A}{B-A} & A \leq x < B \\ 1 & x \geq B \end{cases}$$

Đồ thị của hàm $F(x)$ như sau.



Mệnh đề 4.2.2. *Cho X là một biến ngẫu nhiên liên tục có hàm mật độ $f(x)$ và hàm phân phối $F(x)$. Khi đó, với mọi số thực a, ta có*

$$P(X > a) = 1 - F(a),$$

và cho hai số a và b với $a < b$, ta có

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

Ví dụ 4.4 Cho X là biến ngẫu nhiên liên tục có hàm mật độ là

$$f(x) = \begin{cases} \dfrac{1}{8} + \dfrac{3}{8}x & 0 \leq x \leq 2 \\ 0 & \text{nếu khác} \end{cases}$$

Xét giá trị x bất kì trong đoạn $[0, 2]$, ta có

$$F(x) = \int_{-\infty}^{x} f(y)dy = \int_{0}^{x} \left(\frac{1}{8} + \frac{3}{8}y\right) dy = \frac{x}{8} + \frac{3}{16}x^2$$

Khi đó,

$$F(x) = \begin{cases} 0 & x < 0 \\ \dfrac{x}{8} + \dfrac{3}{16}x^2 & 0 \le x \le 2 \\ 1 & x > 2 \end{cases}$$

Chẳng hạn, ta tính

$$\begin{aligned} P(1 \le X \le 1.5) &= F(1.5) - F(1) \\ &= \left[\frac{1}{8}1.5 + \frac{3}{16}1.5^2\right] - \left[\frac{1}{8}1 + \frac{3}{16}1^2\right] \\ &= \frac{19}{64} = 0.297 \end{aligned}$$

và

$$P(X > 1) = 1 - P(X \le 1) = 1 - F(1) = 1 - \left[\frac{1}{8}1 + \frac{3}{16}1^2\right]$$
$$= \frac{11}{16} = 0.688$$

Mệnh đề 4.2.3. *Nếu X là biến ngẫu nhiên liên tục với hàm mật độ $f(x)$ và hàm phân phối $F(x)$ thì với mọi x sao cho đạo hàm $F'(x)$ tồn tại và $F'(x) = f(x)$.*

Phân vị của phân phối liên tục

Định nghĩa 4.2.4. Cho p là một số thuộc khoảng $[0, 1]$. **Phân vị thứ (100p)** của một hàm phân phối của một biến ngẫu nhiên liên tục X, kí hiệu $\eta(p)$, được xác định bởi

$$p = F(\eta(p)) = \int_{-\infty}^{\eta(p)} f(y)dy \tag{4.1}$$

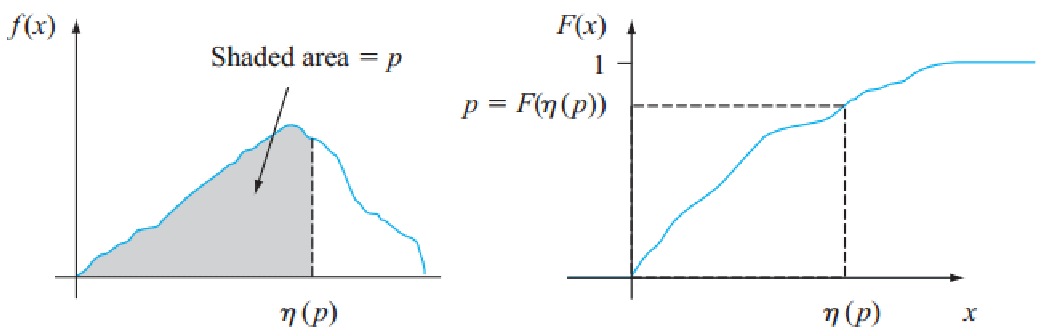


Figure 4.10 The (100*p*)th percentile of a continuous distribution

Ví dụ 4.5 Cho X là biến ngẫu nhiên liên tục có hàm mật độ

$$f(x) = \begin{cases} \frac{3}{2}(1-x^2) & 0 \le x \le 1, \\ 0 & \text{nếu khác.} \end{cases}$$

Khi đó, cho x thuộc khoảng $[0,1]$ ta có,

$$F(x) = \int_0^x \frac{3}{2}(1-y^2)dy = \frac{3}{2}\left(x - \frac{x^3}{3}\right)$$

Phân vị thứ $(100p)$ của phân phối này thỏa mãn biểu thức sau

$$p = F(\eta(p)) = \frac{3}{2}\left(\eta(p) - \frac{\eta(p)^3}{3}\right)$$

nghĩa là,

$$(\eta(p))^3 - 3\eta(p) + 2p = 0$$

Chẳng hạn, phân vị thứ 50, tức là $p = 0.5$ và ta có phương trình $\eta^3 - 3\eta + 1 = 0$; với nghiệm tương ứng là $\eta = \eta(0.5) = 0.347$.

Định nghĩa 4.2.5. Trung vị của một phân phối liên tục, kí hiệu $\widetilde{\mu}$, là phân vị thứ 50, sao cho $\widetilde{\mu}$ thỏa mãn $0.5 = F(\mu)$. Nghĩa là, trung vị là giá trị chia phân phối thành 2 phần bằng nhau.

Kì vọng

Định nghĩa 4.2.6. Kì vọng hay **giá trị trung bình** của một biến ngẫu nhiên liên tục X với hàm mật độ $f(x)$ là

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

Ví dụ 4.6 Cho X là biến ngẫu nhiên liên tục có hàm mật độ

$$f(x) = \begin{cases} \frac{3}{2}(1-x^2) & 0 \le x \le 1, \\ 0 & \text{nếu khác.} \end{cases}$$

Khi đó,

$$E(X) = \int_{-\infty}^{\infty} x \cdot f(x)dx = \int_0^1 x \cdot \frac{3}{2}(1-x^2)dx = \frac{3}{8}$$

Mệnh đề 4.2.7. *Cho X là biến ngẫu nhiên liên tục và $h(X)$ là một hàm bất kì của X. Khi đó,*

$$E[h(X)] = \mu_{h(X)} = \int_{-\infty}^{\infty} h(x) \cdot f(x)dx$$

Định nghĩa 4.2.8. Phương sai của một biến ngẫu nhiên liên tục X với hàm khối lượng $f(x)$ và kì vọng μ là

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

Độ lệch chuẩn của X là $\sigma_X = \sqrt{Var(X)}$.

Ví dụ 4.7 Cho X là biến ngẫu nhiên liên tục có hàm mật độ

$$f(x) = \begin{cases} \frac{3}{2}(1-x^2) & 0 \leq x \leq 1, \\ 0 & \text{nếu khác.} \end{cases}$$

Khi đó, ta có

$$E(X^2) = \int_{-\infty}^{\infty} x^2 \cdot f(x)dx = \int_0^1 x^2 \cdot \frac{3}{2}(1-x^2) = \frac{1}{5}$$

và

$$Var(X) = \frac{1}{5} - \left(\frac{3}{8}\right)^2 = \frac{19}{320} \quad \text{và} \quad \sigma_X = 0.244$$

4.3 Phân phối chuẩn

Định nghĩa 4.3.1. Một biến ngẫu nhiên liên tục X gọi là có phân phối chuẩn với hai tham số μ và σ, trong đó $-\infty < \mu < \infty$ và $0 < \sigma$, nếu hàm mật độ của X là

$$f(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/(2\sigma^2)}, \quad -\infty < x < \infty$$

Định nghĩa 4.3.2. Phân phối chuẩn với hai tham số $\mu = 0, \sigma = 1$ gọi là **phân phối chuẩn tắc**. Một biến ngẫu nhiên Z có phân phối chuẩn tắc với hàm mật độ xác suất là

$$f(z;0,1) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \quad -\infty < z < \infty$$

Hàm phân phối của Z là $P(Z \leq z) = \int_{-\infty}^{z} f(y;0,1)dy = \Phi(z)$.

Ví dụ 4.8 Cho Z là biến ngẫu nhiên có phân phối chuẩn tắc. Tính các xác suất sau: $a)P(Z \leq 1.25)$, $b)P(Z > 1.25)$ $c)P(Z \leq -1.25)$ $d)P(-0.38 \leq Z \leq 1.25)$

a. $P(Z \leq 1.25) = \Phi(1.25)$, tra bảng ta được $\Phi(1.25) = 0.8944$. Do đó, $P(Z \leq 1.25) = 0.8944$

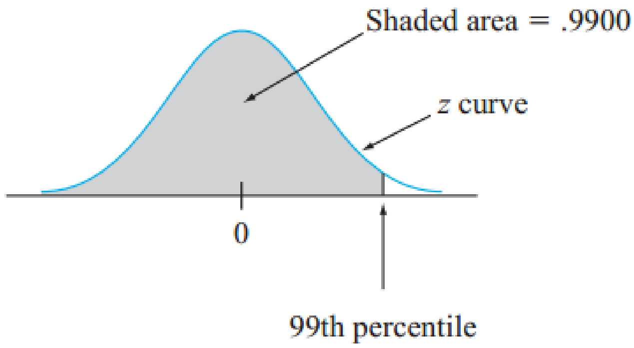
b. $P(Z > 1.25) = 1 - P(Z \leq 1.25) = 0.1056$

c. $P(Z \leq -1.25) = \Phi(-1.25) = 0.1056$

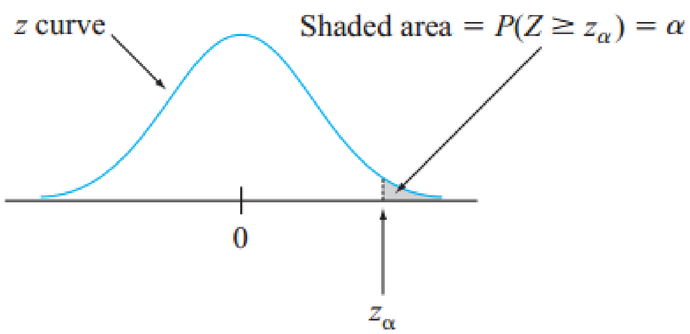
d. $P(-0.38 \leq Z \leq 1.25) = P(Z \leq 1.25) - P(Z \leq -0.38) = \Phi(1.25) - \Phi(0.38) = 0.8944 - 0.3520 = 0.5424$.

Phân vị của phân phối chuẩn

Ví dụ 4.9 Phân vị thứ 99 của phân phối chuẩn là giá trị nằm trên trục hoành sao cho diện tích dưới đường cong z nằm bên trái của giá trị này là 0.99.



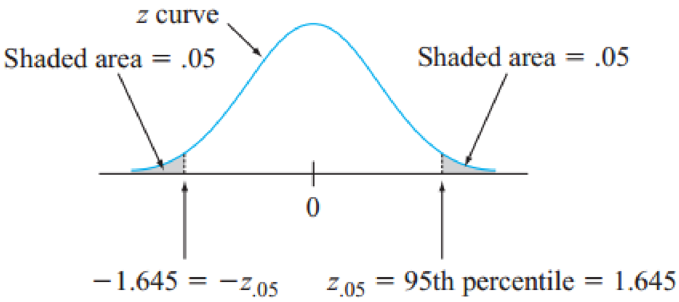
Ký hiệu. z_α biểu thị giá trị trên trục z mà α là diện tích vùng dưới đường cong z (đường cong của phân phối chuẩn) nằm bên phải z_α như hình dưới đây. Khi đó, z_α chính là phân vị thứ $100(1-\alpha)$ của phân phối chuẩn.



Ta có bảng phân vị của phân phối chuẩn

Percentile	90	95	97.5	99	99.5	99.9	99.95
α (tail area)	.1	.05	.025	.01	.005	.001	.0005
$z_\alpha = 100(1-\alpha)$th percentile	1.28	1.645	1.96	2.33	2.58	3.08	3.27

Ví dụ 4.10 $z_{0.05}$ là phân vị thứ $100(1-0.05) = 95$ của phân phối chuẩn vì thế $z_{0.05} = 1.645$. Do tính đối xứng của phân phối chuẩn, diện tích dưới đường cong của phân phối chuẩn nằm bên trái của $-z_{0.05}$ cũng bằng với diện tích dưới đường cong của phân phối chuẩn nằm bên phải của $z_{0.05}$ (xem hình bên dưới).



Mệnh đề 4.3.3. *Cho X là biến ngẫu nhiên có phân phối chuẩn với trung bình μ và độ lệch chuẩn σ. Khi đó,*

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

có phân phối chuẩn tắc. Do đó,

$$P(a \le X \le b) = P\left(\frac{a-\mu}{\sigma} \le Z \le \frac{b-\mu}{\sigma}\right) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

$$P(X \le a) = \Phi\left(\frac{a-\mu}{\sigma}\right) \quad P(X \ge b) = 1 - \Phi\left(\frac{b-\mu}{\sigma}\right)$$

Ví dụ 4.11 Cho X là biến ngẫu nhiên có phân phối chuẩn với giá trị trung bình 1.25 và độ lệch chuẩn 0.46. Tính xác suất: a) $P(1 \le X \le 1.75)$ b) $P(X > 2)$

a) Ta có

$$1 \le X \le 1.75$$

nếu và chỉ nếu

$$\frac{1-1.25}{0.46} \le \frac{X-1.25}{0.46} \le \frac{1.75-1.25}{0.46}$$

Khi đó,

$$\begin{aligned} P(1 \le X \le 1.75) &= P\left(\frac{1-1.25}{0.46} \le Z \le \frac{1.75-1.25}{0.46}\right) \\ &= P(-0.54 \le Z \le 1.09) = \Phi(1.09) - \Phi(-0.54) \\ &= 0.8621 - 0.2946 = 0.5675 \end{aligned}$$

b) $P(X > 2) = P\left(Z > \dfrac{2-1.25}{0.46}\right) = P(Z > 1.63) = 1 - \Phi(1.63) = 0.0516$

Xấp xỉ phân phối nhị thức

Mệnh đề 4.3.4. *Cho X là biến ngẫu nhiên có phân phối nhị thức với n phép thử với xác suất thành công p. Khi đó nếu đồ thị của phân phối nhị thức không quá lệch (không quá mất đối xứng) thì ta có thể xấp xỉ X như một phân phối chuẩn với $\mu = np$ và $\sigma = \sqrt{np(1-p)}$. Đặc biệt, nếu x là một giá trị của biến ngẫu nhiên X thì*

$$P(X \le x) = B(x; n, p) \simeq \Phi\left(\frac{x + 0.5 - np}{\sqrt{np(1-p)}}\right)$$

Trong các bài tập cụ thể, tính xấp xỉ này được áp dụng nếu $np \ge 10$ và $n(1-p) \ge 10$ thỏa mãn. Bởi vì đây là điều kiện cần để phân phối nhị thức có tính chất đối xứng.

Ví dụ 4.12 Giả sử một trường đại học có 25% sinh viên được nhận hỗ trợ tài chính. Gọi X là số lượng sinh viên được nhận hỗ trợ trong 50 sinh viên sao cho $p = 0.25$. Khi đó, ta có $\mu = 12.5$ và $\sigma = 3.06$. Vì $np = 50 \cdot 0.25 = 12.5 \ge 10$ và $n(1-p) = 37.5 \ge 10$ nên có thể xấp xỉ phân phối nhị thức. Xác suất mà nhiều nhất 10 sinh viên nhận hỗ trợ

$$P(X \le 10) = B(10; 50, 0.25) \simeq \Phi\left(\frac{10 + 0.5 - 12.5}{3.06}\right) = \Phi(-0.65) = 0.2578$$

Xác suất có từ 5 đến 15 sinh viên được nhận hỗ trợ tài chính

$$\begin{aligned} P(5 \le X \le 15) &= B(15; 50, 0.25) - B(4; 50, 0.25) \\ &\simeq \Phi\left(\frac{15.5 - 12.5}{3.06}\right) - \Phi\left(\frac{4.5 - 12.5}{3.06}\right) = 0.8320 \end{aligned}$$

4.4 Phân phối mũ và Gamma

Phân phối mũ

Định nghĩa 4.4.1. Một biến ngẫu nhiên X gọi là có phân phối mũ với tham số $\lambda(\lambda > 0)$ nếu hàm mật độ xác suất của X là

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0, \\ 0 & \text{nếu khác.} \end{cases}$$

Hàm phân phối của X là

$$F(x;\lambda) = \begin{cases} 1 - e^{-\lambda x} & x \geq 0, \\ 0 & x < 0 \end{cases}$$

Giá trị trung bình và phương sai của X lần lượt là $\mu = \dfrac{1}{\lambda}; \sigma^2 = \dfrac{1}{\lambda^2}$

Ví dụ 4.13 Cho X là biến ngẫu nhiên có phân phối mũ với giá trị trung bình là 6. Tính xác suất: a) $P(X \leq 10)$ b) $P(5 \leq X \leq 10)$

Ta có: $6 = EX = \dfrac{1}{\lambda}$, do đó $\lambda = 0.1667$.

a) $P(X \leq 10) = F(10; 0.1667) = 1 - e^{-(0.1667)(10)} = 1 - 0.189 = 0.811$

b) $P(5 \leq X \leq 10) = F(10; 0.1667) - F(5; 0.1667) = (1 - e^{-1.667}) - (1 - e^{-8.335}) = 0.246$

Phân phối Gamma

Định nghĩa 4.4.2. Cho $\alpha > 0$, khi đó **hàm gamma** $\Gamma(\alpha)$ được xác định bởi

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

Một số tính chất quan trọng của hàm gamma:

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

Đặt

$$f(x;\alpha) = \begin{cases} \dfrac{x^{\alpha-1} e^{-x}}{\Gamma(\alpha)} & x \geq 0, \\ 0 & \text{nếu khác.} \end{cases} \tag{4.2}$$

Khi đó, ta thấy $f(x;\alpha) \geq 0$ và $\int_0^\infty f(x;\alpha)dx = \Gamma(\alpha)/\Gamma(\alpha) = 1$, vì thế $f(x;\alpha)$ thỏa mãn hai tính chất cơ bản của hàm mật độ xác suất.

Định nghĩa 4.4.3. Một biến ngẫu nhiên liên tục X gọi là có **phân phối gamma** nếu hàm mật độ của X là

$$f(x;\alpha,\beta) = \begin{cases} \dfrac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x} & x \geq 0, \\ 0 & \text{nếu khác.} \end{cases} \tag{4.3}$$

trong đó α và β là hai tham số dương.
Hàm gamma chuẩn tắc khi $\beta = 1$ và hàm mật độ của X lúc đó là (4.2)

Kì vọng và phương sai của biến ngẫu nhiên X có hàm phân phối gamma $f(x;\alpha,\beta)$ là

$$E(X) = \mu = \alpha\beta \quad Var(X) = \sigma^2 = \alpha\beta^2$$

Hàm phân phối của biến ngẫu nhiên X có phân phối gammma chuẩn tắc là

$$F(x;\alpha) = \int_0^x \frac{y^{\alpha-1}e^{-y}}{\Gamma(\alpha)} dy \quad x > 0$$

còn được gọi là **hàm gamma không đầy đủ**.
Ví dụ 4.14 Cho X là biến ngẫu nhiên có phân phối gamma chuẩn tắc với $\alpha = 2$. Ta có

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

khi X là biến ngẫu nhiên liên tục. Do đó,

$$P(3 \leq X \leq 5) = F(5;2) - F(3;2) = 0.960 - 0.801 = 0.159$$

và

$$P(X > 4) = 1 - P(X \leq 4) = 1 - F(4;2) = 1 - 0.908 = 0.092$$

Mệnh đề 4.4.4. *Cho X là biến ngẫu nhiên có hàm phân phối gamma với tham số α và β. Khi đó, với mọi $x > 0$, hàm phân phối của X là*

$$P(X \leq x) = F(x;\alpha,\beta) = F\left(\frac{x}{\beta};\alpha\right)$$

với $F(\bullet;\alpha)$ là hàm gamma không đầy đủ.

Ví dụ 4.15 Giả sử vòng đời X của một con chuột đực tính theo tuần sống trong môi trường có bức xạ 240rads có hàm phân phối gamma với $\alpha = 8$ và $\beta = 15$. Khi đó, thời gian sống sót trung bình là $EX = 8 \cdot 15 = 120$ tuần, trong khi $Var(X) =$

$8 \cdot 15^2 = 1800$ và $\sigma_X = \sqrt{1800} = 42.43$ tuần. Xác suất mà con chuột sống từ 60 đến 129 tuần là

$$\begin{aligned} P(60 \leq X \leq 120) &= P(X \leq 120) - P(X \leq 60) \\ &= F(120/15; 8) - F(60/15; 8) \\ &= F(8; 8) - F(4; 8) = 0.547 - 0.051 = 0.496 \end{aligned}$$

Xác suất con chuột sống ít nhất 30 tuần

$$\begin{aligned} P(X \geq 30) &= 1 - P(X < 30) = 1 - P(X \leq 30) \\ &= 1 - F(30/15; 8) = 0.999 \end{aligned}$$

Phân phối Chi-bình phương

Định nghĩa 4.4.5. Cho ν là một số thực dương. Khi đó, một biến ngẫu nhiên X được gọi là **phân phối Chi-bình phương** với tham số ν nếu hàm mật độ xác suất của X là hàm mật độ gamma có tham số $\alpha = \nu/2$ và $\beta = 2$. Hàm mật độ của một biến ngẫu nhiên có phân phối Chi-bình phương là

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

Tham số ν gọi là **bậc tự do** của X. Ta dùng kí hiệu χ^2 thường dùng thay thế "Chi-bình phương".

4.5 Một số phân phối liên tục khác

Phân phối Weibull

Định nghĩa 4.5.1. Một biến ngẫu nhiên X gọi là có **phân phối Weibull** với tham số α and β $(\alpha > 0, \beta > 0)$ nếu hàm mật độ của X là

$$f(x;\alpha,\beta) = \begin{cases} \dfrac{\alpha}{\beta^\alpha} x^{\alpha-1} e^{-(x/\beta)^\alpha} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Kì vọng và phương sai của X

$$\mu = \beta\Gamma\left(1 + \frac{1}{\alpha}\right) \quad \sigma^2 = \beta^2 \left[\Gamma\left(1 + \frac{2}{\alpha}\right) - \left(\Gamma\left(1 + \frac{1}{\alpha}\right)\right)^2\right]$$

Hàm phân phối của X

$$F(x;\alpha,\beta)=\begin{cases}0 & x<0\\ 1-e^{-(x/\beta)^{\alpha}} & x\geq 0\end{cases}$$

Ví dụ 4.16 Cho X là biến ngẫu nhiên có phân phối Weibull với $\alpha=2,\beta=10$. Tính xác suất:

$$P(X\leq 10)=F(10;2,10)=1-e^{-(10/10)^2}=0.632$$

Phân phối lô-ga chuẩn

Định nghĩa 4.5.2. Một biến ngẫu nhiên không âm X gọi là có **phân phối lô-ga chuẩn** nếu biến ngẫu nhiên $Y=\ln(X)$ có phân phối chuẩn. Khi đó, hàm mật độ của X là

$$f(x;\mu,\sigma)=\begin{cases}\dfrac{1}{\sqrt{2\pi}\sigma x}e^{-[ln(x)-\mu]^2/(2\sigma^2)} & x\geq 0\\ 0 & x<0\end{cases}$$

Kì vọng và phương sai của X là

$$E(X)=e^{\mu+\sigma^2/2}\quad Var(X)=e^{2\mu+\sigma^2}\cdot(e^{\sigma^2}-1)$$

Vì $ln(X)$ có phân phối chuẩn nên hàm phân phối của X có thể biểu diễn qua hàm phân phối $\Phi(z)$ của biến ngẫu nhiên chuẩn tắc Z.

$$\begin{aligned}F(x;\mu,\sigma)&=P(X\leq x)=P[\ln(X)\leq\ln(x)]\\&=P\left(Z\leq\frac{\ln(x)-\mu}{\sigma}\right)=\Phi\left(\frac{\ln(x)-\mu}{\sigma}\right)\quad x\geq 0\end{aligned}$$

Ví dụ 4.17 Cho X là biến ngẫu nhiên có phân phối lô-ga chuẩn với $\mu=0.353,\sigma=0.754$. Khi đó, ta tính được kì vọng và phương sai của X

$$\begin{aligned}EX&=e^{0.353+0.754^2/2}=1.891\\Var(X)&=e^{2\cdot 0.353+0.754^2}\cdot(e^{0.754^2}-1)=2.7387\end{aligned}$$

Khi đó,

$$\begin{aligned}P(1\leq X\leq 2)&=P\big(\ln(1)\leq\ln(X)\leq\ln(2)\big)=P\big(0\leq\ln(X)\leq 0.693\big)\\&=P\left(\frac{0-0.353}{0.754}\leq Z\leq\frac{0.693-0.353}{0.754}\right)=\Phi(0.47)-\Phi(-0.45)=0.354\end{aligned}$$

Phân phối Beta

Định nghĩa 4.5.3. Một biến ngẫu nhiên X có **phân phối beta** với tham số $\alpha >, \beta > 0$, A, B nếu hàm mật độ của X có dạng

$$f(x;\alpha,\beta,A,B) = \begin{cases} \dfrac{1}{B-A} \cdot \dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\cdot\Gamma(\beta)} \left(\dfrac{x-A}{B-A}\right)^{\alpha-1} \left(\dfrac{B-x}{B-A}\right)^{\beta-1} & A \le x \le B \\ 0 & \text{nếu khác} \end{cases}$$

Trường hợp, $A = 0, B = 1$, ta gọi X có **phân phối beta chuẩn**.

Kì vọng và phương sai của X

$$\mu = A + (B-A)\cdot\frac{\alpha}{\alpha+\beta} \qquad \sigma^2 = \frac{(B-A)^2\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

Ví dụ 4.18 Cho X là biến ngẫu nhiên có phân phối beta với các tham số sau $A = 2, B = 5, \alpha = 2, \beta = 3$. Khi đó,

$$\begin{aligned} P(X \le 3) &= \int_2^3 \frac{1}{3}\cdot\frac{4!}{1!2!}\cdot\frac{x-2}{3}\left(\frac{5-x}{3}\right)dx \\ &= \frac{4}{27}\int_2^3 (x-2)(5-x)^2 dx = 0.407 \end{aligned}$$

BÀI TẬP

1. Cho X là biến ngẫu nhiên có hàm mật độ

$$f(x) = \begin{cases} 0.075x + 0.2 & 3 \le x \le 5 \\ 0 & \text{nếu khác} \end{cases}$$

a. Vẽ đồ thị và kiểm tra tổng diện tích dưới hàm mật độ có bằng 1 không.
b. Tính $P(X \le 4)$ và so sánh với $P(X < 4)$.
c. Tính $P(3.5 \le X \le 4.5)$ và $P(4.5 < X)$.

3. Cho X là biến ngẫu nhiên có hàm mật độ

$$f(x) = \begin{cases} 0.09375(4 - x^2) & -2 \le x \le 2 \\ 0 & \text{nếu khác} \end{cases}$$

a. Vẽ đồ thị $f(x)$.
b. Tính $P(X < 0)$.
c. Tính $P(-1 < X < 1)$.
d. Tính $P(X < -0.5$ hoặc $X > 0.5)$.

11. Cho X là biến ngẫu nhiên có hàm phân phối

$$F(x) = \begin{cases} 0 & x < 0 \\ \dfrac{x^2}{4} & 0 \leq x \leq 2 \\ 1 & 2 \leq x \end{cases}$$

Tính:
a. $P(X \leq 1)$
b. $P(0.5 \leq X \leq 1)$
c. $P(X > 1.5)$
d. Tìm trung vị $\widetilde{\mu}$ (tức là giải $0.5 = F(\widetilde{\mu})$
e. Tìm $F'(x)$ để lấy được hàm mật độ $f(x)$.
f. $EX, V(X), \sigma_X$.

15. Cho X là biến ngẫu nhiên có hàm mật độ

$$f(x) = \begin{cases} 90x^8(1-x) & 0 \leq x \leq 1 \\ 0 & \text{nếu khác} \end{cases}$$

a. Vẽ đồ thị $f(x)$. Sau đó tính $F(x)$ và vẽ nó.
b. $P(X \leq 0.5)$ là gì?
c. Từ câu (a), tính $P(0.25 < X \leq 0.5)$
d. Phân vị thứ 75 của phân phối là bao nhiêu?
e. Tính $E(X), \sigma_X$.
f. Tính xác suất X lớn hơn 1 độ lệch chuẩn từ giá trị trung bình?

19. Cho X là biến ngẫu nhiên có hàm phân phối

$$F(x) = \begin{cases} 0 & x \leq 0 \\ \dfrac{x}{4}\left[1 + \ln\left(\dfrac{4}{x}\right)\right] & 0 < x \leq 4 \\ 1 & x > 4 \end{cases}$$

a. $P(X \leq 1)$?
b. $P(1 \leq X \leq 3)$
c. Tìm hàm mật độ của X.

29. Trong mỗi trường hợp sau, xác định giá trị của hằng số c
a. $\Phi(c) = 0.9838$ **b.** $P(0 \leq Z \leq c) = 0.291$
c. $P(c \leq Z) = 0.121$ **d.**$P(-c \leq Z \leq c) = 0.668$
e. $P(c \leq |Z|) = 0.016$.

31. Xác định giá trị z_α khi
a. $\alpha = 0.0055$ **b.** $\alpha = 0.663$ **c.** $\alpha = 0.09$

33. Một xe máy có vận tốc tuân theo phân phối chuẩn với giá trị trung bình là 46.8 km/h và độ lệch chuẩn 1.75 km/h.
a. Tính xác suất vận tốc tối đa nhiều nhất là 50km/h.
b. Tính xác suất vận tốc tối đa ít nhất là 50km/h.
c. Tính xác suất vận tốc tối đa khác giá trị trung bình nhiều nhất là 1.5 độ lệch chuẩn.

35. Đường kính (inch) của một số cây có phân phối chuẩn với $\mu = 8.8$ và $\sigma = 2.8$.
a. Tính xác suất chọn được cây có đường kính ít nhất 10 inch? vượt qua 10 inch?
b. Tính xác suất chọn được cây có đường kính vượt qua 20 inch?
c. Tính xác suất chọn được cây có đường kính nằm trong khoảng từ 5 inch đến 10 inch?
d. Tìm giá trị c sao cho khoảng $(8.8 - c, 8.8 + c)$ bao gồm 98% của tất cả các giá trị đường kính?
e. Nếu bốn cây được chọn ngẫu nhiên, tính xác suất co ít nhất một cây có đường kính vượt qua 10 inch?

37. Giả sử nồng độ máu chloride (mmol/L) co phân phối chuẩn với trung bình 104 và độ lệch chuẩn 5.
a. Tính xác suất nồng độ máu chloride bằng 105? ít hơn 105? nhiều nhất 105?
b. Tính xác suất nồng độ máu chloride khác với giá trị trung bình 1 độ lệch chuẩn? Xác suất này có phụ thuộc vào μ và σ không?
c. Làm thế nào để đặc trưng nhất 0.1% nồng độ chloride?

41. Thiết bị tự động mở dù của một dù nhảy quân sự là 200 m trên mặt đất. Giả sử độ cao mở dù thực sự có phân phối chuẩn với giá trị trung bình 200 m và

độ lệch chuẩn 30 m. Thiết bị sẽ bị thiệt hại nếu dù mở ra ở độ cao dưới 100 m. Tính xác suất có ít nhất một trong năm chiếc dù rơi độc lập bị thiệt hại?

45. Một máy sản xuất vòng bi ban đầu thiết lập để đường kính trung bình của vòng bi là 0.500 inch. Một vòng bi được chấp nhận nếu đường kính của nó lệch khoảng 0,004 inch so với giá trị mục tiêu này. Tuy nhiên, giả sử thiết lập đã thay đổi trong quá trình sản xuất, để vòng bi có đường kính phân bố chuẩn với giá trị trung bình là 0.499 inch và độ lệch chuẩn 0.002 inch. Tính tỷ lệ phần trăm của vòng bi sản xuất ra sẽ không được chấp nhận được?

49. Xem xét các em bé sinh ra trong phạm vi "chuẩn" từ 37-43 tuần tuổi thai. Dữ liệu mở rộng hỗ trợ giả định rằng đối với những đứa trẻ sinh ra ở Hoa Kỳ, trọng lượng khi sinh thường phân bố với mức trung bình là 3432 g và độ lệch tiêu chuẩn 482 g.
a. Tính xác suất sinh một đứa trẻ được chọn ngẫu nhiên vượt quá 4000 g? giữa 3000 đến 4000 g?
b. Tính xác suất sinh một đứa trẻ được chọn ngẫu nhiên dưới 2000 g hoặc lớn hơn 5000 g.
c. Tính xác suất là trọng lượng sinh của một em bé được chọn vượt quá 7 lb?
d. Làm thế nào đặc trưng nhất 0,1% của trọng lượng sinh em bé?

53. Cho X là biến ngẫu nhiên có phân phối nhị thức với các tham số $n = 25$ và p. Tính mỗi xác suất sau bằng phương pháp xấp xỉ chuẩn trong các trường hợp $p = 0.5, 0.6, 0.8$.
a. $P(15 \leq X \leq 20)$.
b. $P(X \leq 15)$.
c. $P(20 \leq X)$.

55. Giả sử chỉ có 75% số người lái xe có đeo dây an toàn. Một mẫu ngẫu nhiên của 500 tài xế được chọn. Tính xác suất
a. Giữa 360 và 400 các tài xế đeo dây an toàn?
b. Ít hơn 400 người đeo dây an toàn?