Chương 3: BIẾN NGẪU NHIÊN LIÊN TỤC

Trần Văn Long

28th April 2020

Mục lục

3.1 Khái niệm biến ngẫu nhiên liên tục

Khái niệm

- ☞ Trong nhiều bài toán trong kỹ thuật, chúng ta cần dùng những biến ngẫu nhiên để mô hình hóa cho một số đo nào đó về nguồn dữ liệu. Chẳng hạn như đo độ dày của lớp cản quang tráng lên một thanh bán dẫn, đo nhiệt độ trong một phản ứng hóa học, đo tuổi thọ của một con chíp điện tử, v.v.
- ☞ Trong những bài toán này, chúng ta sẽ sử dụng đến những biến ngẫu nhiên mà tập giá trị có thể nhận của nó lấp đầy những khoảng số thực. Ta gọi đó là những biến ngẫu nhiên liên tục.

Ví dụ 3.1

Xét X là biến ngẫu nhiên chỉ thời gian kết nối của một máy điện thoại gọi đến tổng đài thông tin 1080. Khi đó biến X nhận các giá trị thực không âm x, $x \geqslant 0$.

Trong ví dụ này nếu ta lấy đơn vị thời gian là giây thì tập giá trị có thể nhận của biến X có thể coi là tập số rời rạc, X bằng 1 giây, 2 giây, 3 giây v.v. Do giới hạn của các đồng hồ đo thời gian cho nên người ta đã làm tròn thời gian kết nối của các cuộc gọi đến một số nguyên dương gần nhất. Nhưng nếu ta vẫn coi biến X là biến ngẫu nhiên rời rạc thì sẽ bất tiện cho việc tìm luật phân phối của biến X. Vì vậy trong ví dụ này, chúng ta sẽ coi tập giá trị có thể nhận của biến X lấp đầy một khoảng số thực (a, b) nào đó. Tức ta có X là biến ngẫu nhiên liên tục.

3.1 Khái niệm biến ngẫu nhiên liên tục

Khái niệm

Nhắc lại rằng luật phân phối xác suất được hiểu là cách đưa thông tin về một đại lượng ngẫu nhiên X sao cho ta có thể xác định được hoàn toàn các giá trị xác suất $\mathbb{P}(X = a), \mathbb{P}(X < a), \mathbb{P}(a < X < b)$ v.v..

Hai cách đưa ra luật phân phối xác suất cho đại lượng ngẫu nhiên liên tục:

- Sử dụng hàm mật độ xác suất.
- Sử dụng hàm phân phối xác suất.

Các luật phân phối liên tục được sử dụng trong thực tế

- Luật phân phối đều trên một đoạn.
- Luật phân phối chuẩn (phân phối Gauss).
- Luật phân phối mũ.
- Luật phân phối Erlang và phân phối Gamma.
- Luật phân phối Weibull.
- Luật phân phối logarit chuẩn. Luật phân phối Beta.

Cho X là một đại lượng ngẫu nhiên liên tục. Nếu $a < b$ thì ta gọi tỷ số

$$\frac{\mathbb{P}(a \leqslant X \leqslant b)}{b-a}$$

là mật độ trung bình trên đoạn $[a, b]$.
Khi thu hẹp đoạn [a, b] về điểm x thì mật độ trung bình sẽ tiến tới một giới hạn

$$f(x) = \lim_{\substack{a \to x, b \to x \\ a \leqslant x \leqslant b}} \frac{\mathbb{P}(a \leqslant X \leqslant b)}{b-a}$$

và $f(x)$ được gọi là mật độ xác suất của X tại x.

Một cách tiếp cận khác là định nghĩa sau đây

Định nghĩa 3.1

Cho X là biến ngẫu nhiên liên tục, hàm mật độ xác suất của biến X là hàm $f(x)$ thỏa mãn

1. $f(x) \geqslant 0$ với mọi $-\infty < x < \infty$,
2. $\int\limits_{-\infty}^{\infty} f(x)dx = 1$,
3. $\mathbb{P}(a \leqslant X \leqslant b) = \int\limits_{a}^{b} f(x)dx$ bằng diện tích của hình thang cong giới hạn bởi trục hoành, hàm số $f(x)$ và hai đường thẳng $x = a$, $x = b$.

Ví dụ 3.2

Một xưởng cơ khí sản xuất ra đai ốc để phục vụ cho việc xây dựng đường ray xe lửa. Theo tiêu chuẩn quy định các đai ốc có đường kính $12,5$ (mm), nhưng người ta thấy hầu hết các đai ốc bị lỗi đều có đường kính lớn hơn $12,5$ (mm). Gọi X là biến ngẫu nhiên chỉ đường kính của các đai ốc. Từ các mẫu thực nghiệm, các kỹ sư xây dựng các biểu đồ tần suất và mô phỏng hàm mật độ xác suất của biến ngẫu nhiên X là

$$f(x) = \begin{cases} 20e^{-20(x-12,5)}, & \text{nếu } x \geqslant 12,5 \\ 0 & \text{nếu } x < 12,5. \end{cases}$$

Nếu một đai ốc có đường kính lớn hơn $12,6$ (mm) sẽ bị loại ngay. Hãy tính tỷ lệ các đai ốc bị loại và đưa ra nhận xét về chất lượng của xưởng cơ khí.

Giải: Ta có tỷ lệ đai ốc bị loại được tính theo xác suất

$$\mathbb{P}(X > 12,6) = \int_{12,6}^{\infty} 20e^{-20(x-12,5)}dx$$

$$= -e^{-20(x-12,5)}\Big|_{12,6}^{\infty} = \lim_{M\to\infty} -e^{-20(x-12,5)}\Big|_{12,6}^{M}$$

$$= e^{-20.0,1} - \lim_{M\to\infty} e^{-20(M-12,5)} = e^{-2} = 0,135.$$

Vậy tỷ lệ đai ốc bị loại xấp xỉ $13,5\%$, đây là một tỷ lệ khá cao. Vì vậy xưởng cơ khí cần phải nâng cao chất lượng để sản xuất ra những đai ốc đạt tiêu chuẩn.

Nếu X là đại lượng ngẫu nhiên liên tục thì ta quy ước

$$\mathbb{P}(X = a) = 0 \quad \text{với mọi} \quad x \in \mathbb{R}.$$

Ta có công thức

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

Thay đổi giá trị của hàm $f(x)$ tại một số hữu hạn điểm không gây ảnh hưởng tới luật phân phối xác suất của X.
Nếu hàm số $f(x)$ thỏa mãn hai điều kiện 1., 2. trong Định nghĩa 3.1 thì $f(x)$ là hàm mật độ xác suất của một đại lượng ngẫu nhiên X nào đấy.

Định nghĩa 3.2

Hàm phân phối xác suất của biến ngẫu nhiên liên tục X là

$$F(x) = \mathbb{P}(X \leqslant x) = \int_{-\infty}^{x} f(t)dt, \qquad (1)$$

với $x \in \mathbb{R}$.

Tính chất của hàm phân phối xác suất

- $0 \leqslant F(x) \leqslant 1$,
- $F(-\infty) = 0, F(+\infty) = 1$,
- Nếu $x < y$ thì $F(x) \leqslant F(y)$,
- $F(x)$ là hàm số liên tục trên trục số thực.

Mối liên hệ giữa hàm mật độ và hàm phân phối xác suất.
Nếu hàm mật độ $f(x)$ liên tục tại điểm x nào đó thì hàm phân phối xác suất $F(x)$ có đạo hàm tại điểm x đó và

$$F'(x) = f(x). \tag{3.3}$$

Ví dụ 3.3

Xét lại biến ngẫu nhiên X trong Ví dụ 3.2 với hàm mật độ xác suất

$$f(x) = \begin{cases} 20e^{-20(x-12,5)}, & \text{nếu } x \geqslant 12,5 \\ 0 & \text{nếu } x < 12,5. \end{cases}$$

a) Tìm hàm phân phối xác suất của biến ngẫu nhiên X.
b) Tìm tỷ lệ đai ốc có đường kính không vượt quá 12,55 (mm).

Giải. a) Trước tiên ta đi tìm hàm phân phối xác suất bằng cách xét hai trường hợp

Nếu $x < 12,5$ thì $F(x) = \int\limits_{-\infty}^{x} 0dt = 0.$

Nếu $x \geqslant 12,5$ thì

$$F(x) = \int\limits_{-\infty}^{x} f(t)dt = \int\limits_{-\infty}^{12,5} f(t)dt + \int\limits_{12,5}^{x} f(t)dt$$

$$= \int\limits_{12,5}^{x} 20e^{-20(t-12,5)}dt = 1 - e^{-20(x-12,5)}.$$

Vậy ta có hàm phân phối của biến ngẫu nhiên X như sau

$$F(x) = \begin{cases} 1 - e^{-20(x-12,5)} & \text{nếu } x \geqslant 12,5 \\ 0 & \text{nếu } x < 12,5. \end{cases}$$

b) Tỷ lệ đai ốc có đường kính không vượt quá 12,55 mm là

$$\mathbb{P}(X \leqslant 12,5) = F(12,55) = 1 - e^{-1} = 0,6321.$$

Ví dụ 3.4

Một đơn vị tổ chức các gói thầu xây dựng và đưa ra giá mời thầu là b. Căn cứ vào các kết quả đấu thầu đã thực hiện, các nhà thống kê đã xây dựng hàm mật độ cho giá thắng thầu (giá thấp nhất) như sau

$$f(x) = \begin{cases} \frac{a}{b} & \text{nếu } x \in [\frac{2}{5}b, 2b] \\ 0 & \text{nếu } x \notin [\frac{2}{5}b, 2b]. \end{cases}$$

a) Hãy xác định tham số a để hàm $f(x)$ là hàm mật độ xác suất.
b) Tìm hàm phân phối xác suất.
c) Hãy tính xác suất để một công ty xây dựng bỏ giá thấp hơn giá mời thầu mà vẫn trúng thầu.

Giải: a) Gọi X là giá thắng thầu. Ta đi tìm hàm mật độ của biến ngẫu nhiên X.

Vì giá mời thầu $b > 0$ nên để $f(x) \geqslant 0$ với mọi $-\infty < x < \infty$ thì $a \geqslant 0$. Hơn nữa

$$\int_{-\infty}^{\infty} f(x)dx = \int_{\frac{2}{5}b}^{2b} \frac{a}{b}dx = \frac{8}{5}a.$$

Từ $\frac{8}{5}a = 1$, ta tìm được $a = \frac{5}{8}$. Do đó hàm mật độ của biến X là

$$f(x) = \begin{cases} \frac{5}{8b} & \text{nếu } x \in [\frac{2}{5}b, 2b], \\ 0 & \text{nếu } x \notin [\frac{2}{5}b, 2b]. \end{cases}$$

b) Với $x \in [\frac{2}{5}b, 2b]$ ta có

$$F(x) = \int_{\frac{2}{5}b}^{x} \frac{5}{8b}dt = \frac{5t}{8b}\Big|_{2b/5}^{x} = \frac{5x}{8b} - \frac{1}{4}.$$

Vậy hàm phân phối xác suất cần tìm là

$$F(x) = \begin{cases} 0 & \text{nếu } x < \frac{2}{5}b \\ \frac{5x}{8b} - \frac{1}{4} & \text{nếu } \frac{2}{5}b \leqslant x < 2b \\ 1 & \text{nếu } x \geqslant 2b. \end{cases}$$

c) Ta cần tính xác suất

$$\mathbb{P}(X \leqslant b) = F(b) = \frac{5}{8} - \frac{1}{4} = \frac{3}{8} = 0,375.$$

Vậy xác suất để công ty xây dựng bỏ giá thấp hơn giá mời thầu mà vẫn trúng thầu là $37,5\%$.

Đối với biến ngẫu nhiên liên tục thì kỳ vọng và phương sai được định nghĩa thông qua tích phân của hàm mật độ xác suất.

Định nghĩa

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

- Kỳ vọng hay giá trị trung bình của X là
$$\mu = \mathbb{E}[X] = \int_{-\infty}^{\infty} x f(x) dx. \tag{3.5}$$
- Phương sai của X là
$$\sigma^2 = \mathbb{V}[X] = \int_{-\infty}^{\infty} (x-\mu)^2 f(x) dx. \tag{3.6}$$
- Độ lệch chuẩn của biến X là $\sigma = \sqrt{\sigma^2}$.

Tính chất của kỳ vọng và phương sai: Giống với tính chất của các đại lượng ngẫu nhiên rời rạc.

Ví dụ 3.5

Xét X là biến ngẫu nhiên chỉ đường kính của đinh ốc trong Ví dụ 3.2 với hàm mật độ xác suất

$$f(x) = \begin{cases} 20e^{-20(x-12,5)}, & \text{nếu } x \geqslant 12,5 \\ 0 & \text{nếu } x < 12,5. \end{cases}$$

Hãy tính kỳ vọng của X.

Giải: Kỳ vọng, hay đường kính trung bình của đinh ốc là

$$\mathbb{E}[X] = \int\limits_{-\infty}^{\infty} xf(x)dx = 20 \int\limits_{12,5}^{\infty} xe^{-20(x-12,5)}dx = -\int\limits_{12,5}^{\infty} xd\left(e^{-20(x-12,5)}\right)$$

$$= -xe^{-20(x-12,5)}\Big|_{12,5}^{\infty} + \int\limits_{12,5}^{\infty} e^{-20(x-12,5)}dx =$$

$$= 12,5 - \frac{1}{20} \int_{12,5}^{\infty} d\left(e^{-20(x-12,5)}\right)$$

$$= 12,5 - \frac{e^{-20(x-12,5)}}{20}\Bigg|_{12,5}^{\infty} = 12,55.$$

Phương sai của biến ngẫu nhiên X được tính theo công thức

$$\mathbb{V}[X] = 20 \int_{12,5}^{\infty} (x - 12,55)^2 e^{-20(x-12,5)} dx.$$

Việc tính phương sai $\mathbb{V}[X]$ trong ví dụ này là khá phức tạp, ta phải sử dụng công thức tích phân từng phần hai lần và ta thu được $\mathbb{V}[X] = 0.0025$. Ta có thể tính phương sai $\mathbb{V}[X]$ theo một công thức gọn hơn.

Công thức tính kỳ vọng

Cho X là đại lượng ngẫu nhiên liên tục với hàm mật độ xác suất $f(x)$ và $h(x)$ là hàm số xác định trên toàn bộ trục số. Khi đó $Y = h(X)$ cũng là một biến ngẫu nhiên có giá trị trung bình là

$$\mathbb{E}[Y] = \mathbb{E}[h(X)] = \int_{-\infty}^{\infty} h(x)f(x)dx. \tag{3.7}$$

Khai triển công thức (2.24) ta có

$$\mathbb{V}[X] = \int_{-\infty}^{\infty} (x^2 - 2\mu x + \mu^2)f(x)dx = \int_{-\infty}^{\infty} x^2 f(x)dx - \mu^2.$$

Chú ý rằng

$$\mathbb{E}[X^2] = \int_{-\infty}^{\infty} x^2 f(x)dx,$$

vì vậy ta có thể tính phương sai của biến ngẫu nhiên liên tục X theo công

thức rút gọn sau đây:

Công thức phương sai

$$\mathbb{V}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2. \tag{3.8}$$

Ví dụ 3.6

Độ dày X của lớp cản quang tráng lên một tấm bán dẫn là biến ngẫu nhiên có đơn vị là micromet (μm) và có hàm mật độ xác suất là $f(x) = 600x^{-2}$ với $100\ (\mu m) < x < 120\ (\mu m)$. Ta có thể coi $f(x)$ bằng 0 khi $x \notin [100, 120]\ (\mu m)$. Tính độ dày trung bình và độ lệch chuẩn của lớp cản quang.

Giải: Độ dày trung bình của lớp cản quang tráng lên một tấm bán dẫn là

$$\mathbb{E}[X] = \int_{100}^{120} x600x^{-2}dx = 600\int_{100}^{120}\frac{dx}{x} = 600\ln(x)\Big|_{100}^{120} = 109,3929\ \mu m.$$

Để tính độ lệch chuẩn của X ta đi tính

$$\mathbb{E}[X^2] = \int_{100}^{120} x^2 600x^{-2}dx = 600\int_{100}^{120} dx = 600x\Big|_{100}^{120} = 12000\ \mu m^2,$$

vậy phương sai của X là

$$\mathbb{V}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = 12000 - 109,3929^2 = 33,1860\ \mu m^2.$$

Do đó độ lệch chuẩn của X là $\sigma = \sqrt{33,1860} = 5,7607\ (\mu m)$.
Nếu chi phí cho mỗi μm lớp cản quang tráng lên một tấm bán dẫn là 10000 VND, thì chi phí trung bình cho việc tráng lớp cản quang lên một tấm bán dẫn sẽ là $109,3929.10000 = 1093929$ VND.

Trong các biến ngẫu nhiên liên tục thì biến ngẫu nhiên có phân phối đều là biến đơn giản nhất và cũng có nhiều ứng dụng nhất.

Định nghĩa 3.3

Biến ngẫu nhiên liên tục X được gọi là tuân theo luật phân phối đều trên đoạn $[a,b]$ nếu X có hàm mật độ xác suất là

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{nếu } x \in [a,b] \\ 0 & \text{nếu } x \notin [a,b]. \end{cases} \tag{3.9}$$

Ta ký hiệu biến ngẫu nhiên X có phân phối đều trên đoạn $[a,b]$ là $X \sim \mathscr{U}([a,b])$.

Giá trị trung bình của biến X là

$$\mathbb{E}[X] = \int_a^b \frac{x}{b-a}dx = \frac{x^2}{2(b-a)}\Big|_a^b = \frac{a+b}{2}.$$

3.5 Một số phân phối liên tục

Phân phối đều

Phương sai của biến X là

$$\mathbb{V}[X] = \int_a^b \frac{\left(x - \frac{a+b}{2}\right)^2}{b-a} dx = \frac{\left(x - \frac{a+b}{2}\right)^3}{3(b-a)} \Bigg|_a^b = \frac{(b-a)^2}{12}.$$

Hàm phân phối xác suất của biến ngẫu nhiên có phân phối đều được xác định từ công thức tích phân

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

- Nếu $x < a$ thì $F(x) = 0$.
- Nếu $a \leqslant x \leqslant b$ thì
$$F(x) = \int_a^x \frac{1}{b-a} dt = \frac{x-a}{b-a}.$$
- Nếu $x > b$ thì $(x) = \displaystyle\int_a^b \frac{1}{b-a} dt = 1.$

3.5 Một số phân phối liên tục

Phân phối đều

Vậy hàm phân phối xác suất của $X \sim \mathscr{U}[a, b]$ là

$$F(x) = \begin{cases} 0 & x < a \\ \dfrac{x-a}{b-a} & a \leqslant x < b \\ 1 & x \geqslant b. \end{cases}$$

Ví dụ 3.7

Thể tích của một loại dầu gội đầu được đóng vào lọ tuân theo luật phân phối đều, có giá trị biến thiên trong khoảng từ 374 (ml) đến 380 (ml). Tiêu chuẩn yêu cầu là một lọ dầu được đóng phải có thể tích 375 (ml).

a) Hãy tính giá trị trung bình và phương sai của thể tích dầu gội đầu được đóng chai.

b) Tính tỷ lệ lọ dầu có thể tích nhỏ hơn thể tích tiêu chuẩn.

c) Hãy tìm thể tích a (ml) của một chai dầu gội mà có 95% số lọ dầu có thể tích lớn hơn a.

Ví dụ 3.7 (tiếp)

d) Giá 1ml dầu gội là 40 VND. Hãy tính hao phí trung bình cho 1ml dầu gội do đóng vượt thể tích tiêu chuẩn.

Giải:a) Gọi X là biến ngẫu nhiên chỉ thể tích của một lọ dầu gội đầu. Biến ngẫu nhiên X tuân theo phân phối đều trên đoạn $[374, 380]$ (ml). Hàm mật độ xác suất của X là

$$f(x) = \begin{cases} \frac{1}{6} & \text{nếu } x \in [374, 380] \\ 0 & \text{nếu } x \notin [374, 380]. \end{cases}$$

Thể tích trung bình của một lọ dầu gội là

$$\mathbb{E}[X] = \frac{374 + 380}{2} = 377 \ (ml).$$

Phương sai của thể tích một lọ dầu là

$$\mathbb{V}[X] = \frac{(380 - 374)^2}{12} = 3 \ (ml^2).$$

3.5 Một số phân phối liên tục

Phân phối đều

b) Ta có

$$\mathbb{P}(X < 375) = \int\limits_{-\infty}^{375} \frac{1}{6}dx = \int\limits_{374}^{375} \frac{1}{6}dx = \frac{1}{6} = 0,1667.$$

Vậy tỷ lệ chai dầu có thể tích nhỏ hơn thể tích tiêu chuẩn là $16,67\%$.

c) Ta có giá trị a phải nằm trong miền giá trị $[374, 380]$ (ml) của biến X.

$$\mathbb{P}(X > a) = 1 - \mathbb{P}(X \leqslant a) = 1 - \int\limits_{374}^{a} \frac{1}{6}dx = 1 - \frac{a - 374}{6} = 0,95.$$

Giải phương trình này ta tìm được $a = 374,3$ (ml).

d) Xác suất để một chai dầu vượt tiêu chuẩn về thể tích là

$$\mathbb{P}(X > 375) = 1 - \mathbb{P}(X \leqslant 375) = 1 - 0,1667 = 0,8333.$$

Như vậy trung bình khi đóng một chai dầu nhà máy phải chi phí thêm $40.0,8333 = 33,332$ VND cho 1 ml do lỗi đóng vượt thể tích quy định

3.5 Một số phân phối liên tục

Phân phối chuẩn

- Có thể nói biến ngẫu nhiên có phân phối chuẩn là biến ngẫu nhiên quan trọng nhất và được sử dụng để mô tả cho nhiều hiện tượng xảy ra trong tự nhiên cũng như trong kỹ thuật.
- Các phép đo vật lý về các hiện tượng khí tượng học như nghiên cứu lượng mưa, các số đo của các sản phẩm được sản suất hàng loạt như trọng lượng của bao xi măng được đóng gói bằng máy tự động, các sai số trong các phép đo của các dụng cụ đo lường, v.v. đều được mô hình hóa thành các biến ngẫu nhiên có phân phối chuẩn.
- Những biến ngẫu nhiên là tổng hoặc giá trị trung bình của các kết quả của nhiều phép đo đạc trong các phép thử ngẫu nhiên cũng thường được mô hình hóa dưới dạng các biến ngẫu nhiên có phân phối xấp xỉ phân phối chuẩn.

Định nghĩa 3.4

Biến ngẫu nhiên X được gọi là tuân theo luật phân phối chuẩn với hai tham số μ và $\sigma > 0$, ký hiệu là $X \sim \mathcal{N}(\mu, \sigma^2)$, nếu hàm mật độ xác suất của nó có dạng

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}. \tag{3.11}$$

Giá trị trung bình và phương sai của biến chuẩn $\mathcal{N}(\mu, \sigma^2)$

$$\mathbb{E}[X] = \mu, \qquad \mathbb{V}[X] = \sigma^2. \tag{3.12}$$

Định nghĩa 3.5

Biến ngẫu nhiên Z tuân theo luật phân phối chuẩn với $\mu = 0$ và $\sigma = 1$ được gọi là biến ngẫu nhiên chuẩn tắc và ta viết $Z \sim \mathcal{N}(0,1)$. Hàm mật độ và phân phối xác suất của biến chuẩn tắc được xác định bởi

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

$$\Phi(z) = \mathbb{P}(Z \leqslant z) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{z} e^{-\frac{t^2}{2}} dt.$$

Hàm mật độ của biến chuẩn tắc là hàm đối xứng qua trục tung. Giá trị hàm phân phối của biến chuẩn tắc $\Phi(z)$ chính là diện tích của hình phẳng giới hạn bởi trục hoành, hàm mật độ và phần bên trái đường thẳng $x = z$. Các giá trị của hàm $\Phi(z)$ được cho trong bảng phân phối

Ví dụ 3.8

Cho biến ngẫu nhiên Z có phân phối chuẩn tắc. Tính các xác suất
a) $\mathbb{P}(Z \leqslant 1,53)$.
b) $\mathbb{P}(-1,43 \leqslant Z \leqslant 0,72)$.
c) $\mathbb{P}(Z > 1,26)$.
d) $\mathbb{P}(Z \leqslant -4,2)$.

Giải: a) Do Z là chuẩn tắc nên ta có
$P(Z \leqslant 1,53) = \Phi(1,53) = 0,9370$.
b) $P(-1,43 \leqslant Z \leqslant 0,72) = \Phi(0,72) - \Phi(-1,43)$
$= 0,7642 - 0,0764 = 0,6878$.
c) $P(Z > 1,26) = 1 - P(Z \leqslant 1,26) = 1 - 0,8961 = 0,1039$.
d) $P(Z \leqslant -4,2) < P(Z \leqslant -3,99) = 0,000033$, giá trị xác suất này rất nhỏ nên ta có thể coi $P(Z \leqslant -4,2) = 0$.

Để tính các xác suất liên quan đến một biến chuẩn $X \sim \mathcal{N}(a, \sigma^2)$ bất kỳ chúng ta sử dụng phép đổi biến đưa X về biến chuẩn tắc.

Đổi biến

Nếu X là biến ngẫu nhiên tuân theo luật phân phối chuẩn với hai tham số μ và σ thì biến ngẫu nhiên

$$Z = \frac{X - \mu}{\sigma} \sim \mathcal{N}(0, 1). \tag{3.15}$$

tuân theo luật phân phối chuẩn tắc với $\mathbb{E}[Z] = 0$ và $\mathbb{V}[Z] = 1$.

Quy tắc 3σ. Cho biến ngẫu nhiên $X \sim \mathcal{N}(a, \sigma^2)$. Với số dương k, ta tính xác suất để biến ngẫu nhiên X nhận giá trị trong khoảng $(\mu - k\sigma, \mu + k\sigma)$. Ta có

$$\mathbb{P}(\mu - k\sigma < X < \mu + k\sigma) = \mathbb{P}\Big(-k < \frac{X-\mu}{\sigma} < k\Big)$$
$$= \mathbb{P}(-k < Z < k) = \Phi(k) - \Phi(-k).$$

Từ bảng giá trị của hàm phân phối chuẩn tắc với $k = 1, 2, 3$ ta tính được

$$\mathbb{P}(|X-\mu| < \sigma) = \Phi(1) - \Phi(-1) = 0,8413 - 0,1587 = 0,6826;$$
$$\mathbb{P}(|X-\mu| < 2\sigma) = \Phi(2) - \Phi(-2) = 0,9772 - 0,0228 = 0,9544;$$
$$\mathbb{P}(|X-\mu| < 3\sigma) = \Phi(3) - \Phi(-3) = 0,9987 - 0,0013 = 0,9974.$$

Ví dụ 3.9

Cường độ nén của một mẫu bê tông được mô hình hóa bởi một phân bố chuẩn với trung bình là 1000 (kg/cm^2) và độ lệch chuẩn là 100 (kg/cm^2).
a) Tính xác suất để một mẫu bê tông tùy ý có cường độ nén nhỏ hơn 1250 (kg/cm^2).
b) Tính xác suất để một mẫu bê tông tùy ý có cường độ nén nằm trong khoảng từ 800 (kg/cm^2) đến 900 (kg/cm^2).
c) Hãy tìm cường độ a (kg/cm^2) để có hơn 95% số mẫu bê tông có cường độ nén lớn hơn a.

Giải: a) Gọi X là biến ngẫu nhiên chỉ cường độ nén của mẫu bê tông, thì X tuân theo phân phối chuẩn với $\mu = 1000$ (kg/cm^2) và $\sigma = 100$ (kg/cm^2).

Khi đó $Z = \dfrac{X - 1000}{100}$ có phân phối chuẩn tắc $\mathcal{N}(0, 1)$.

Xác suất để một mẫu bê tông tùy ý có cường độ nén nhỏ hơn 1250 (kg/cm^2) được tính theo công thức

$$\begin{aligned}\mathbb{P}(X < 1250) &= P\Big(\frac{X-1000}{100} < \frac{1250-1000}{100}\Big) = \mathbb{P}(Z < 2,5)\\ &= \Phi(2,5) = 0,9938.\end{aligned}$$

b) Xác suất để một mẫu xi măng tùy ý có cường độ nén nằm trong khoảng $(1800, 1900)$ (kg/cm^2) là

$$\mathbb{P}(800 < X < 900) = \mathbb{P}\Big(\frac{800-1000}{100} < \frac{X-1000}{100} < \frac{900-1000}{100}\Big)$$
$$= \mathbb{P}(-2 < Z < -1) = \Phi(-1) - \Phi(-2) = 0,1587 - 0,0228 = 0,1359.$$

c) Ta đi tìm cường độ nén a_0 sao cho

$$
\begin{aligned}
&\mathbb{P}(X > a_0) = 0,95\\
\Leftrightarrow\ &\mathbb{P}(X \leqslant a_0) = 0,05\\
\Leftrightarrow\ &\mathbb{P}\Big(\frac{X-1000}{100} \leq \frac{a_0-1000}{100}\Big) = 0,05\\
\Leftrightarrow\ &\frac{a_0-1000}{100} = -1,645\\
\Leftrightarrow\ &a_0 = 1164,5
\end{aligned}
$$

Vậy với $a \leqslant 1164,5$ (kg/cm^2) thì có

$$\mathbb{P}(X > a) \geqslant \mathbb{P}(X > 1164,5) = 0,95;$$

tức là có hơn 95% số mẫu bê tông có cường độ nén lớn hơn a_0.

Xấp xỉ phân phối chuẩn cho phân phối nhị thức.

Nếu X là biến nhị thức với hai tham số n và p và nếu $np \geqslant 5$, $n(1-p) \geqslant 5$, thì biến ngẫu nhiên

$$Z = \frac{X - np}{\sqrt{np(1-p)}} \tag{3.16}$$

có phân phối xấp xỉ phân phối chuẩn tắc $\mathcal{N}(0,1)$.

Để tính xấp xỉ xác suất trong phân phối nhị thức bởi một phân phối chuẩn tắc đạt độ chính xác cao ta cần một hiệu chỉnh liên tục như sau:

$$\mathbb{P}(X \leqslant x) = \mathbb{P}(X \leqslant x + 0,5) \approx P\left(Z \leqslant \frac{x + 0,5 - np}{\sqrt{np(1-p)}}\right) \tag{3.17}$$

và

$$\mathbb{P}(X \geqslant x) = \mathbb{P}(X \geqslant x - 0,5) \approx \mathbb{P}\left(Z \geqslant \frac{x - 0,5 - np}{\sqrt{np(1-p)}}\right). \tag{3.18}$$

3.5 Một số phân phối liên tục

Phân phối chuẩn

Ví dụ 3.11

Độ tin cậy của một loại cầu chì là $0,98$, đó là xác suất để khi một cầu chì được chọn ra nó sẽ hoạt động tốt theo những tiêu chuẩn được thiết kế. Người ta kiểm tra ngẫu nhiên một lô hàng có 1000 cầu chì. Tính xác suất để có ít nhất 27 cầu chì bị lỗi.

Giải: Xác suất để một cầu chì bị lỗi là $p = 1 - 0,98 = 0,02$. Gọi X là biến ngẫu nhiên chỉ số cầu chì bị lỗi trong lô hàng 1000 cầu chì thì X tuân theo phân phối nhị thức với các tham số $n = 1000$ và $p = 0,02$.
Xác suất để có ít nhất 27 cầu chì bị lỗi là

$$\mathbb{P}(X \geqslant 27) = 1 - \mathbb{P}(X \leqslant 26) = 1 - \sum_{k=0}^{26} C_{1000}^{k}\ 0,02^{k}.0,98^{1000-k}.$$

Việc tính xác suất này khá khó khăn, vì vậy ta sẽ dùng phân phối xấp xỉ phân phối chuẩn để tính. Giá trị trung bình và độ lệch chuẩn của X

lần lượt là $\mu = np = 20$, $\sigma = \sqrt{np(1-p)} = 4,4272$

Ta có $np = 20 \geqslant 5$ và $n(1-p) = 980 \geqslant 5$ nên biến ngẫu nhiên $Z = \frac{X-20}{4{,}4272}$ có phân phối xấp xỉ phân phối chuẩn tắc $\mathcal{N}(0,1)$. Để tính $\mathbb{P}(X \geqslant 27)$ ta dùng công thức hiệu chỉnh liên tục (3.18)

$$\mathbb{P}(X \geqslant 27) \approx P\left(Z \geqslant \frac{27 - 0,5 - 20}{4,4272}\right) = 1 - \mathbb{P}(Z < 1,4682) = 0,0708.$$

Ví dụ 3.12

Các hãng hàng không thường chấp thuận bán số lượng vé vượt quá số ghế của mỗi máy bay để giảm thiểu thiệt hại do những khách đã đặt mua vé nhưng lại hủy vé. Các con số thống kê của các hãng hàng không cho thấy có khoảng 2% số khách sẽ hủy vé. Một máy bay có 197 ghế nhưng đã bán cho 200 khách, mỗi khách 1 vé. Hãy tính xác suất để mỗi khách đã mua vé và đến sân bay đều được nhận 1 vé của chuyến bay đó.

Giải: Xác suất để một khách đã đặt mua vé và đến sân bay làm thủ tục nhận vé là $p = 1 - 0,02 = 0,98$. Gọi X là đại lượng ngẫu nhiên chỉ số

3.5 Một số phân phối liên tục

Phân phối chuẩn

Như vậy X có phân phối nhị thức với $n = 200$ và $p = 0,98$ có trung bình $\mu = np = 196$ và độ lệch chuẩn $\sigma = \sqrt{np(1-p)} = 1,9799$. Xác suất để mỗi khách đã mua vé và đến sân bay đều được nhận 1 vé của chuyến bay đó là

$$\begin{aligned}\mathbb{P}(X \leqslant 197) &= 1 - \mathbb{P}(X > 197) = 1 - \mathbb{P}(X = 198) - \mathbb{P}(X = 199) - \mathbb{P}(X = \\ &= 1 - C_{200}^{198} 0,98^{198} 0,02^{2} - C_{200}^{199} 0,98^{198} 0,02^{1} - C_{200}^{200} 0,98^{20} \\ &= 1 - 0,1458 - 0,0718 - 0,0176 = 0,7648.\end{aligned}$$

Bây giờ ta dùng phân phối chuẩn để xấp xỉ cho biến nhị thức X. Áp dụng công thức hiệu chỉnh liên tục (3.17) ta có

$$\mathbb{P}(X \leqslant 197) \approx \mathbb{P}\left(Z \leqslant \frac{197 + 0,5 - 196}{1,9799}\right) = \mathbb{P}(Z \leqslant 0,76) = 0,7764.$$

Trong bài toán này chúng ta thấy giá trị xấp xỉ cho $\mathbb{P}(X \leqslant 197)$ không được tốt do $n(1-p) = 4 < 5$.

Xấp xỉ phân phối chuẩn cho phân phối Poisson.

Nếu X là biến ngẫu nhiên Poisson với tham số λ thì biến ngẫu nhiên

$$Z = \frac{X - \lambda}{\sqrt{\lambda}} \tag{3.19}$$

có phân phối xấp xỉ phân phối chuẩn tắc $\mathcal{N}(0, 1)$.

Tương tự như trên ta cũng có công thức hiệu chỉnh liên tục và công thức xấp xỉ sẽ đạt độ chính xác cao nếu $\lambda \geqslant 5$.

Ví dụ 3.13

Số hạt bụi Amiang trong $1(m^2)$ chất bụi bám lên bề mặt tường ở một nhà máy tuân theo phân phối Poisson với trung bình là 1000 hạt/$1m^2$. Người ta kiểm tra số hạt bụi Amiang trên 1 (m^2) của một tường của nhà máy. Tính xác suất để có ít hơn hoặc bằng 950 hạt bụi Amiang.

Giải: Gọi X là biến ngẫu nhiên chỉ số hạt bụi Amiang trên 1 (m^2) tường của nhà máy thì X tuân theo phân phối Poisson với tham số $\lambda = 1000$. Như vậy ta cần tính xác suất $\mathbb{P}(X \leqslant 950.)$ Nếu sử dụng phân phối Poisson thì ta có

$$\mathbb{P}(X \leqslant 950) = \sum_{x=0}^{950} \frac{e^{-1000}1000^x}{x!}.$$

Việc tính xác suất này sẽ rất khó khăn, nhưng do $\lambda = 1000 > 5$, nên ta sẽ xấp xỉ phân phối Poisson của biến X bằng phân phối chuẩn, tức là biến

$$Z = \frac{X - 1000}{\sqrt{1000}}$$

có phân phối xấp xỉ phân phối chuẩn tắc. Khi đó

$$\mathbb{P}(X \leqslant 950) \approx \mathbb{P}\left(Z \leqslant \frac{950 + 0,5 - 1000}{\sqrt{1000}}\right) = \mathbb{P}(Z \leqslant -1,57) = 0,0582.$$

Bây giờ chúng ta trở lại Ví dụ 2.22, với biến ngẫu nhiên Poisson chỉ số vết nứt dọc theo chiều dài của sợi dây đồng. Trong thực tế nhiều khi người ta cũng quan tâm đến các biến ngẫu nhiên chỉ khoảng cách giữa hai vết nứt. Gọi X là biến ngẫu nhiên chỉ khoảng cách từ một điểm bất kỳ trên sợi dây đến vết nứt đầu tiên được phát hiện tính từ điểm đó và gọi N là biến ngẫu nhiên chỉ số vết nứt trong x mm dây đồng. Nếu trung bình trong 1 (mm) sợi dây có λ vết nứt thì N là biến Poisson với giá trị trung bình là λx. Khi đó

$$\mathbb{P}(X > x) = \mathbb{P}(N = 0) = \frac{e^{-\lambda x}(\lambda x)^0}{0!} = e^{-\lambda x}.$$

Do đó

$$F(x) = \mathbb{P}(X \leqslant x) = 1 - e^{-\lambda x}, \ x > 0$$

là hàm phân phối tích lũy của biến X. Lấy đạo hàm hàm phân phối tích lũy ta được hàm mật độ xác suất của biến X như sau

$$f(x) = F'(x) = \lambda e^{-\lambda x}, \ x > 0.$$

Định nghĩa 3.6

Trong quá trình Poisson với số trung bình các sự kiện xuất hiện trong một đơn vị thời gian là $\lambda > 0$, biến X chỉ khoảng cách giữa hai lần liên tiếp xuất hiện sự kiện được gọi là biến ngẫu nhiên tuân theo luật phân phối mũ với tham số λ. Hàm mật độ xác suất của X là

$$f(x) = \lambda e^{-\lambda x},\ x > 0.$$

Ký hiệu biến ngẫu nhiên X có phân phối mũ với tham số λ là $X \sim Exp(\lambda)$.

Kỳ vọng và phương sai của biến mũ

Nếu X là biến ngẫu nhiên có phân phối mũ với tham số $\lambda > 0$ thì

$$\mathbb{E}[X] = \frac{1}{\lambda} \text{ và } \mathbb{V}[X] = \frac{1}{\lambda^2}. \tag{3.20}$$

Ví dụ 3.14

Số người đăng nhập (sử dụng Username và Password) vào sử dụng máy tính ở một hệ thống mạng máy tính của một trường đại học được mô hình hóa thành quá trình Poisson với trung bình 1 giờ có 25 lần đăng nhập được thực hiện.
a) Tính xác suất để không có đăng nhập nào được thực hiện trong khoảng thời gian 6 phút.
b) Tính xác suất để lần đăng nhập đầu tiên được thực hiện trong khoảng thời gian từ 3 đến 5 phút tiếp theo.

Giải: a) Trước tiên ta thống nhất đơn vị thời gian là giờ. Như vậy ta cần tính xác suất để không có đăng nhập nào được thực hiện trong khoảng thời gian $6/60 = 0,1$ giờ. Nếu ta gọi X là biến ngẫu nhiên chỉ khoảng thời gian từ lúc bắt đầu cho đến khi có log-on đầu tiên thì X là biến ngẫu nhiên tuân theo phân phối mũ với tham số $\lambda = 25$.

Xác suất cần tính là

$$\mathbb{P}(X > 0,1) = \int\limits_{0,1}^{+\infty} 25e^{-25x}dx = -e^{-25x}\Big|_{0,1}^{+\infty} = e^{-25.0,1} = 0,082.$$

Nếu sử dụng hàm phân phối xác suất thì ta có thể tính xác suất trên đơn giản hơn như sau:

$$\mathbb{P}(X > 0,1) = 1 - \mathbb{P}(X \leqslant 0,1) = 1 - F(0,1) = e^{-25.0,1} = 0,082.$$

b) Bây giờ ta chọn một thời điểm bất kỳ để bắt đầu tính giờ, quy về cùng đơn vị thời gian là giờ thì khi đó xác suất để đăng nhập đầu tiên được thực hiện trong khoảng thời gian từ 3 đến 5 phút tiếp theo là

$$\mathbb{P}(0,05 < X < 0,083) = \int\limits_{0,05}^{0,083} 25e^{-25x}dx = -e^{-25x}\Big|_{0,05}^{0,083} = 0,1609.$$

Hoặc ta dùng hàm phân phối xác suất thì

$$\mathbb{P}(0,05 < X < 0,083) = F(0,083) - F(0,05) = e^{-25.0,05} - e^{-25.0,083} = 0,$$

Ví dụ 3.15

Khoảng thời gian giữa hai khách hàng liên tiếp đến thanh toán ở một quầy thu ngân trong một siêu thị là một biến ngẫu nhiên tuân theo luật phân phối mũ với thời gian trung bình là 1,4 phút.
a) Chọn một thời điểm bất kỳ, tính xác suất để trong vòng 30 giây kể từ một thời điểm đó thì có 1 khách đến thanh toán.
b) Cho biết không có khách nào đến thanh toán trong khoảng thời gian 3 phút, hãy tính xác suất để trong vòng 30 giây tiếp theo có một khách đến quầy thanh toán.

Giải: a) Gọi X là biến ngẫu nhiên chỉ khoảng thời gian giữa hai khách hàng liên tiếp đến thanh toán ở quầy thu ngân, thì X tuân theo phân phối mũ với $\mathbb{E}[X] = \frac{1}{\lambda} = 1,4$. Vậy tham số của phân phối mũ là $\lambda = 1/1,4$ và hàm phân phối tích lũy của biến X là $F(x) = 1 - e^{-x/1,4}, \ x > 0$.

Xác suất để trong khoảng 30 giây (0,5 phút) có 1 khách đến thanh toán là
$$\mathbb{P}(X < 0,5) = F(0,5) = 1 - e^{-0,5/1,4} = 0,30.$$

b) Vì quầy thanh toán đã "đợi" được 3 phút mà không có khách nào đến thanh toán, xác suất để trong vòng 0,5 phút tiếp theo có 1 khách đến thanh toán là xác suất có điều kiện

$$\mathbb{P}(X < 3,5|X > 3) = \frac{\mathbb{P}((X < 3,5) \cap (X > 3))}{\mathbb{P}(X > 3)} = \frac{\mathbb{P}(3 < X < 3,5)}{1 - \mathbb{P}(X \leqslant 3)}$$
$$= \frac{F(3,5) - F(3)}{1 - F(3)} = \frac{e^{-3/1,4} - e^{-3,5/1,4}}{1 - e^{-3/1,4}} = 0,30.$$

Mặc dù quầy thanh toán đã "đợi" được 3 phút, xác suất để trong vòng 30 giây kế tiếp có 1 khách đến thanh toán vẫn không thay đổi (bằng xác suất tính được ở phần (a). Người ta gọi đó là tính chất "không có trí nhớ" của biến ngẫu nhiên tuân theo luật phân phối mũ.

Hàm sống sót của phân phối mũ là $\overline{F}(x) = 1 - F(x) = e^{-\lambda x}$, hay

$$\mathbb{P}(X > x) = e^{-\lambda x}$$

giảm theo hàm số mũ. Phân phối mũ là biến ngẫu nhiên không nhớ.

Tính chất không nhớ của biến mũ

Giả sử X là biến ngẫu nhiên có phân phối mũ với tham số $\lambda > 0$, thì ta có

$$\mathbb{P}(X > s + t | X > s) = \mathbb{P}(X > t), \text{ với } s, t > 0. \tag{3.21}$$

3.5 Một số phân phối liên tục

Phân phối mũ

- Phân phối mũ thường được sử dụng để mô hình hóa toán học cho thời điểm bị hỏng một thiết bị điện tử nào đó.
- Ta mô tả tuổi thọ của một loại bóng đèn bán dẫn bởi phân phối mũ với tuổi thọ trung bình là 50.000 giờ. Tính chất không nhớ của phân phối mũ nói rằng bóng đèn không bị hao hụt về tuổi thọ. Nghĩa là, khi xem xét về thời gian hoạt động của bóng đèn thì xác suất để bóng đèn hoạt động 1000 giờ tiếp theo cũng bằng xác suất để bóng đèn hoạt động 1000 giờ đầu tiên.
- Nếu mô tả tuổi thọ T của thiết bị theo các tác động mạnh ngẫu nhiên thì ta xấp xỉ bởi phân phối mũ.
- Nếu mô tả tuổi thọ T của thiết bị có tính đến sự hao mòn cơ học thì mô tả xác suất $\mathbb{P}(T < t + \Delta t | T > t)$ là hàm tăng theo t. Trong thực hành với đặc tính này người ta thường sử dụng phân phối Weibull (được trình bày ở phần sau trong chương này) để mô tả tuổi thọ của thiết bị.

Phân phối mũ dùng để mô tả khoảng thời gian xảy ra sự kiện đầu tiên trong quá trình Poisson. Tổng quát hóa phân phối mũ, ta xét biến ngẫu nhiên là khoảng thời gian xảy ra r sự kiện đầu tiên trong quá trình Poisson. Ta xem xét ví dụ dưới đây.

Ví dụ 3.16

Số các lỗi gây ra do bộ xử lý trung tâm (CPU) của một máy tính lớn được mô hình hóa thành một quá trình Poisson. Thông thường các lỗi này là do các mạch bán dẫn trong CPU gây ra. Giả thiết rằng khi một lỗi gây ra sẽ được sửa chữa ngay lập tức và số trung bình các lỗi được sửa trong một giờ là $0,0001$. Gọi X là biến chỉ thời gian đến khi có 4 lỗi xẩy ra trong hệ thống CPU. Tính xác suất để X vượt quá 40000 giờ.

Giải: Gọi N là biến ngẫu nhiên chỉ số lỗi do CPU gây ra trong 40000 giờ. Ta có N là biến Poisson với giá trị trung bình

$$\lambda = 40000.0,0001 = 4 \text{ lỗi}/40000 \text{ giờ}.$$

3.5 Một số phân phối liên tục

Phân phối Erlang

Như vậy
$$\mathbb{P}(X > 40000) = \mathbb{P}(N \leqslant 3) = \sum_{k=0}^{3} \frac{e^{-4}4^k}{k!} = 0,433.$$

Ta có thể tổng quát hóa như sau: Gọi X là biến ngẫu nhiên chỉ thời gian cho đến khi sự kiện thứ r với $(r = 1, 2, 3, \ldots)$ trong quá trình Poisson xẩy ra thì
$$\mathbb{P}(X > x) = \sum_{k=0}^{r-1} \frac{e^{-\lambda x}(\lambda x)^k}{k!}. \tag{3.22}$$

Hơn nữa nếu $F(x)$ là hàm phân phối xác suất của biến ngẫu nhiên X thì
$$\mathbb{P}(X > x) = 1 - \mathbb{P}(X \leqslant x) = 1 - F(x).$$

Như vậy từ công thức (3.22) ta có
$$F(x) = 1 - \sum_{k=0}^{r-1} \frac{e^{-\lambda x}(\lambda x)^k}{k!}.$$

Lấy đạo hàm hai vế đẳng thức này ta nhận được hàm mật độ của biến X. Biến ngẫu nhiên X trên gọi là có luật phân phối Erlang

Định nghĩa

Biến ngẫu nhiên X được gọi là có phân phối Erlang với tham số r nguyên dương và $\lambda > 0$, ký hiệu $X \sim Erlang(r, \lambda)$, nếu X có hàm mật độ

$$f(x) = \frac{\lambda^r x^{r-1} e^{-\lambda x}}{(r-1)!} \text{ với } x > 0. \tag{3.23}$$

Giá trị kỳ vọng và phương sai của phân phối Erlang

$$\mathbb{E}X = \frac{r}{\lambda}, \qquad \mathbb{V}X = \frac{r}{\lambda^2}.$$

Khi $r = 1$ thì phân phối Erlang trở thành phân phối mũ. Để thuận tiện ta tổng quát hóa phân phối Erlang đối với số thực dương r.

Ta cần tổng quát hóa khái niệm giai thừa cho các số thực dương và hàm đó gọi là hàm Gamma.

Định nghĩa

Hàm Gamma $\Gamma(r)$ xác định bởi công thức

$$\Gamma(r) = \int_{0}^{+\infty} x^{r-1} e^{-x} dx, \text{ với } r > 0. \tag{3.24}$$

Sử dụng công thức tích phân từng phần ta tính được $\Gamma(r+1) = r\Gamma(r)$. Ta có $\Gamma(1) = 1$ và với r là số nguyên dương ta có

$$\Gamma(r+1) = r!$$

và với $r = \frac{1}{2}$ ta có công thức

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

Phân phối Gamma là trường hợp tổng quát của phân phối Erlang với hệ số r thực. Ta định nghĩa phân phối Gamma như sau:

Định nghĩa

Biến ngẫu nhiên X có hàm mật độ

$$f(x) = \frac{\lambda^r x^{r-1} e^{-\lambda x}}{\Gamma(r)}, \text{ với } x > 0 \tag{3.25}$$

gọi là phân phối Gamma với tham số $\lambda > 0, r > 0$. Ký hiệu là $X \sim Gamma(\lambda, r)$.

- Các tham số λ gọi là tham số đồng dạng và r gọi là tham số hình học (hình dạng).
- Nếu r là số nguyên dương thì phân phối Gamma chính là phân phối Erlang.
- Khi tham số $\lambda = \frac{1}{2}$ và $r = \frac{n}{2}$ thì phân phối Gamma được gọi là phân phối khi bình phương χ^2 với n bậc tự do. Phân phối khi bình phương được sử dụng nhiều trong thống kê toán học.

Kỳ vọng phương sai của phân phối Gamma

Nếu $X \sim Gamma(\lambda, r)$ thì

$$\mu = \mathbb{E}X = \frac{r}{\lambda}, \quad \sigma^2 = \mathbb{V}X = \frac{r}{\lambda^2}.$$

Phân phối Weibull thường được sử dụng để mô hình hóa tuổi thọ của nhiều hệ thống vật lý khác nhau. Các tham số của phân phối Weibull được sử dụng mềm dẻo đối với các hệ thống khác nhau.

Định nghĩa

Biến ngẫu nhiên X có hàm mật độ

$$f(x) = \frac{\alpha}{\beta}\left(\frac{x}{\beta}\right)^{\alpha-1} \exp\left[-\left(\frac{x}{\beta}\right)^{\alpha}\right], x > 0 \qquad (3.26)$$

được gọi là phân phối Weibull với các tham số đồng dạng $\beta > 0$ và tham số hình học $\alpha > 0$. Ký hiệu là $X \sim Weibull(\alpha, \beta)$.

- ☞ Khi $\alpha = 1$, thì phân phối Weibull trở thành phân phối mũ.
- ☞ Khi $\alpha = 2$, thì phân phối Weibull trở thành phân phối Rayleigh.

Định nghĩa

Biến ngẫu nhiên X có hàm mật độ

$$f(x) = \frac{x}{\sigma^2} \exp\left(-\frac{x^2}{2\sigma^2}\right), x > 0 \tag{3.27}$$

được gọi là phân phối Rayleigh với tham số $\sigma > 0$. Ký hiệu là $X \sim Rayleigh(\sigma^2)$.

Hàm phân phối tích lũy của phân phối Weibull là

$$F(x) = 1 - \exp\left[-\left(\frac{x}{\beta}\right)^{\alpha}\right].$$

Các giá trị kỳ vọng và phương sai được xác định bởi công thức sau:

$$\mu = \mathbb{E}X = \beta\,\Gamma(1+\frac{1}{\alpha}),\ \sigma^2 = \mathbb{V}X = \beta^2\,\Gamma(1+\frac{2}{\alpha}) - \beta^2\left[\Gamma(1+\frac{1}{\alpha})\right]^2.$$

Ví dụ 3.17

Thời gian hoạt động của CPU máy tính được mô tả bởi phân phối Weibull với tham số $\alpha = 2$ và $\beta = 900$ (giờ).

a) Xác định thời gian hoạt động trung bình của CPU.

b) Tính xác suất để CPU bị hỏng trước 500 giờ.

Giải: a) Theo công thức tính giá trị trung bình của phân phối Weibull ta có thời gian hoạt động trung bình của CPU là

$$\mu = \mathbb{E}[X] = \beta\Gamma(1 + \frac{1}{\alpha}) = 900\frac{1}{2}\Gamma(\frac{1}{2}) = 450\sqrt{\pi} = 797,6 \text{ giờ.}$$

b) Xác suất để CPU bị hỏng trước 500 giờ là

$$\mathbb{P}(X < 500) = F(500) = 1 - \exp\left[-\left(\frac{500}{900}\right)^2\right] = 1 - \exp(-0,3086) = 0,2$$

Các biến ngẫu nhiên trong một hệ thống có thể có mối liên hệ theo hàm số mũ như $X = e^Y$. Nếu biến ngẫu nhiên Y có phân phối chuẩn thì biến ngẫu nhiên X được gọi là có phân phối logarit chuẩn. Giả sử biến ngẫu nhiên $Y \sim \mathcal{N}(\mu, \sigma^2)$ thì hàm phân phối của X được xác định như sau:

$$\begin{aligned} F(x) &= \mathbb{P}(X \leqslant x) = \mathbb{P}(e^Y \leqslant x) = \mathbb{P}(Y \leqslant \ln(x)) \\ &= \mathbb{P}\Big(\frac{Y-\mu}{\sigma} \leqslant \frac{\ln(x)-\mu}{\sigma}\Big) \\ &= \mathbb{P}\Big(Z \leqslant \frac{\ln(x)-\mu}{\sigma}\Big) = \Phi\Big(\frac{\ln(x)-\mu}{\sigma}\Big) \end{aligned}$$

với $x > 0$ và $Z = \dfrac{Y-\mu}{\sigma}$ có phân phối chuẩn tắc. Lấy đạo hàm của hàm phân phối trên ta được hàm mật độ của biến ngẫu nhiên X.

Định nghĩa

Cho biến ngẫu nhiên Y có phân phối chuẩn $\mathcal{N}(\mu, \sigma^2)$. Khi đó biến ngẫu nhiên $X = e^Y$ có phân phối logarit chuẩn, ký hiệu là $X \sim \ln \mathcal{N}(\mu, \sigma^2)$, với hàm mật độ

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}}\, e^{-\dfrac{(\ln(x)-\mu)^2}{2\sigma^2}}, x > 0.$$

Giá trị kỳ vọng và phương sai của phân phối logarit chuẩn

$$\mathbb{E}X = e^{\mu+\sigma^2/2}, \qquad \mathbb{V}X = e^{2(\mu+\sigma^2)} - e^{2\mu+\sigma^2}.$$

Ví dụ 3.18

Tuổi thọ của một loại đèn laze bán dẫn có phân phối logarit chuẩn với các tham số $\mu = 10$ (giờ) và $\sigma = 1,5$ (giờ).
a) Tính tuổi thọ trung bình của bóng đèn laze.
b) Tính xác suất để tuổi thọ của bóng đèn hơn 10000 (giờ).

Giải: a) Gọi X là biến chỉ tuổi thọ của bóng đèn. Khi đó X có phân phối logarit chuẩn với tuổi thọ trung bình

$$\mathbb{E}[X] = e^{\mu+\sigma^2/2} = e^{10+1,5^2/2} = e^{11,125} = 67846,3 \text{ giờ.}$$

b) Ta cần tính

$$\begin{aligned}\mathbb{P}(X > 10000) &= 1 - \mathbb{P}(X \leqslant 10000) = 1 - \Phi\Big(\frac{\ln(10000) - \mu}{\sigma}\Big) \\ &= 1 - \Phi\Big(\frac{\ln(10000) - 10}{1,5}\Big) \\ &= 1 - \Phi(-0,52) = 1 - 0,3015 = 0,6985. \qquad \square\end{aligned}$$

- Các phân phối xác suất trên một khoảng hữu hạn được sử dụng nhiều trong các mô hình về xác suất.
- Không mất tính tổng quát ta xét các phân phối xác suất trên đoạn $[0,1]$.
- Phân phối Beta được dùng để mô tả nhiều dạng phân phối xác suất trên đoạn $[0,1]$.

Định nghĩa

Biến ngẫu nhiên X với hàm mật độ

$$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}, x \in [0,1] \tag{3.28}$$

có phân phối Beta với tham số $\alpha > 0, \beta > 0$ và ký hiệu là $X \sim Beta(\alpha,\beta)$.

Giá trị kỳ vọng và phương sai của phân phối Beta

$$\mathbb{E}[X] = \frac{\alpha}{\alpha + \beta}, \qquad \mathbb{V}[X] = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

- Nếu $\alpha = \beta$ thì phân phối đối xứng qua đường thẳng $x = 0,5$.
- Nếu $\alpha = \beta = 1$ thì phân phối trên là phân phối đều trên đoạn $[0,1]$. Nếu $\alpha > \beta$ thì phân phối lệch về phía $x = 1$, và nếu $\alpha < \beta$ thì phân phối lệch về phía $x = 0$.