Chương 3. Các định lý về giới hạn

Nguyễn Minh Trí

Trường Đại học Công nghệ Thông tin

Mục tiêu

- Biết sự hội tụ của các dãy biến ngẫu nhiên.
- Biết hai định lý quan trọng: luật số lớn và định lý giới hạn trung tâm.
- Áp dụng hai định lý trên trong thực tế.
- Sử dụng các công thức xấp xỉ khi tính xác suất.

3.1 Hội tụ theo xác suất và hội tụ theo phân phối

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

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

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

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

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

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

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

3.1 Hội tụ theo xác suất và hội tụ theo phân phối

Định nghĩa 3.2 Dãy các biến ngẫu nhiên $X_1, X_2, \ldots$ được gọi là **hội tụ theo phân phối** đến biến ngẫu nhiên X, kí hiệu $X_n \xrightarrow{d} X$, nếu

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

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

Ví dụ 3.3 Cho một dãy các biến ngẫu nhiên $X_1, X_2, \ldots$ với hàm phân phối xác suất của X_n là

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

Chứng minh rằng X_n hội tụ theo phân phối đến $X \sim \mathrm{Exp}(1)$.
Nhắc lại: Nếu $X \sim \mathrm{Exp}(1)$ thì hàm phân phối của X là

$$F_X(x) = \begin{cases} 1 - e^{-x}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

Giải. Với mọi $x \leq 0$, ta có
$$F_{X_n}(x) = F_X(x) = 0$$

với mọi $n = 1, 2, \ldots$.
Xét $x > 0$, ta có
$$\begin{aligned} \lim_{n\to\infty} F_{X_n}(x) &= \lim_{n\to\infty}\left(1 - \left(1 - \frac{1}{n}\right)^{nx}\right) \\ &= 1 - \lim_{n\to\infty}\left(1 - \frac{1}{n}\right)^{nx} \\ &= 1 - e^{-x} = F_X(x) \end{aligned}$$

Như vậy $X_n \xrightarrow{d} X$.

Định lý 3.4 Cho dãy các biến ngẫu nhiên $X_1, X_2, X_3, \ldots$ và biến ngẫu nhiên X. Giả sử rằng X và X_n (với mọi n) có tập giá trị là các số nguyên không âm. Khi đó $X_n \xrightarrow{d} X$ nếu và chỉ nếu

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

với $k = 1, 2, \ldots$.

Ví dụ 3.5 Cho dãy các biến ngẫu nhiên $X_1, X_2, X_3, \ldots$ có phân phối nhị thức

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

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

Giải. Theo Định lý 3.4, ta chứng minh $\lim\limits_{n\to\infty} P(X_n = k) = P(X = k)$.

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

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

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

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

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

Do đó

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

Định nghĩa 3.6 Dãy các biến ngẫu nhiên $X_1, X_2, \ldots$ được gọi là **hội tụ theo xác suất** đến biến ngẫu nhiên X, kí hiệu $X_n \xrightarrow{P} X$, nếu với mọi $\varepsilon > 0$, ta có

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

Ví dụ 3.7 Cho một dãy các biến ngẫu nhiên $X_1, X_2, \ldots$ trong đó X_n có hàm mật độ xác suất như sau

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

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

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

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

Ví dụ 3.8 Cho một không gian mẫu $\Omega = \{1, 2, 3, 4\}$. Cho dãy các biến ngẫu nhiên $X_1, X_2, \ldots$ xác định như sau

$$X_n(1) = X_n(2) = 1; X_n(3) = X_n(4) = 0, \quad n = 1, 2, \ldots$$

và biến ngẫu nhiên X xác định như sau

$$X(1) = X(2) = 0; X(3) = X(4) = 1.$$

Khi đó

$$|X_n(\omega) - X(\omega)| = 1$$

với mọi $\omega \in \Omega$.Nếu chọn $\varepsilon = 0,5$ thì

$$P(|X_n(\omega) - X(\omega)| \geq \varepsilon) = 1.$$

Do đó X_n không hội tụ theo xác suất về X khi $n \to \infty$.

Xét các hàm phân phối

$$F_{X_n}(x) = \begin{cases} 0, & x < 0 \\ \frac{1}{2}, & 0 \leq x < 1 \\ 1, & x \geq 1 \end{cases} \qquad F_X(x) = \begin{cases} 0, & x < 0 \\ \frac{1}{2}, & 0 \leq x < 1 \\ 1, & x \geq 1 \end{cases}$$

Ta thấy $F_{X_n}(x) = F_X(x)$ với mọi $x \in \mathbb{R}$. Do đó

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

với mọi x sao cho $F_X(x)$ liên tục tại x. Như vậy $X_n \xrightarrow{d} X$.

Định lý 3.9 Cho $X_1, X_2, \ldots$ là một dãy các biến ngẫu nhiên và X là một biến ngẫu nhiên. Nếu $X_n \xrightarrow{P} X$ thì $X_n \xrightarrow{d} X$.

Chiều ngược lại của Định lý 3.9 nói chung là không đúng.

Trường hợp đặc biệt: Nếu $X_n \xrightarrow{d} c$ với c là một hằng số thì $X_n \xrightarrow{P} c$.

Định lý 3.10 Cho $X_1, X_2, \ldots$ là một dãy các biến ngẫu nhiên và X là một biến ngẫu nhiên. Đặt $M_n(t)$ và $M(t)$ tương ứng là các hàm sinh moment của X_n và X xác định trên $(-c, c)$ với $c > 0$. Khi đó

1. Nếu $X_n \xrightarrow{d} X$ thì $\lim_{n\to\infty} M_n(t) = M(t)$ với mọi $t \in (-c, c)$.
2. Nếu $\lim_{n\to\infty} M_n(t) = M(t)$ với mọi $t \in (-c, c)$ thì

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

   với mọi điểm liên tục x của $F(x)$. Do đó $X_n \xrightarrow{d} X$.

3.2 Luật số lớn và Định lý giới hạn trung tâm

3.2 Luật số lớn và Định lý giới hạn trung tâm

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

Ví dụ 3.12

1. Nếu $X_1, X_2, \ldots, X_n$ là các biến ngẫu nhiên độc lập có phân phối nhị thức $B(1, p)$ thì $E(X_i) = p$ và do đó $\overline{X}_n \xrightarrow{P} p$.
2. Nếu $X_1, X_2, \ldots, X_n$ là các biến ngẫu nhiên độc lập có phân phối Poisson $P(\lambda)$ thì $E(X_i) = \lambda$ và do đó $\overline{X}_n \xrightarrow{P} \lambda$.
3. Nếu $X_1, X_2, \ldots, X_n$ là các biến ngẫu nhiên độc lập có phân phối chuẩn $N(\mu, \sigma^2)$ thì $E(X_i) = \mu$ và do đó $\overline{X}_n \xrightarrow{P} \mu$.

Định lý 3.13 (Định lý giới hạn trung tâm-Central Limit Theorem) Cho $X_1, X_2, \ldots, X_n$ là các biến ngẫu nhiên độc lập và phân phối giống nhau với kỳ vọng $\mu < \infty$ và phương sai $\sigma^2 < \infty$ với mọi i và đặt $\overline{X}_n = \dfrac{X_1 + X_2 + \ldots + X_n}{n}$. Khi đó $\overline{X}_n \xrightarrow{d} Z$ trong đó $Z \sim N(\mu, \dfrac{\sigma^2}{n})$.

Nhận xét: Cho $X_1, X_2, \ldots, X_n$ là các biến ngẫu nhiên độc lập và phân phối giống nhau với kỳ vọng $\mu < \infty$ và phương sai $\sigma^2 < \infty$ với mọi i và đặt $S_n = X_1 + X_2 + \ldots + X_n$. Khi đó $S_n \xrightarrow{d} Z$ trong đó $Z \sim N(n\mu; n\sigma^2)$.

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

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

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

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

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

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

Ví dụ 3.16 Gieo đồng thời 3 con xúc xắc 1000 lần độc lập với nhau. Đặt X_i là số số nguyên tố xuất hiện ở mặt trên của 3 con xúc xắc khi gieo lần thứ i với $1 \le i \le 1000$. Tính

$$P(X_1 + X_2 + \cdots + X_{1000} > 1550).$$

Giải. Với mỗi $1 \le i \le 1000$, ta có

$$\begin{aligned}\mu = E(X_i) &= 0.P(X_i = 0) + 1.P(X_i = 1) + 2.P(X_i = 2) + 3.P(X_i = 3)\\ &= 1.\frac{C_3^1.3.3.3}{6.6.6} + 2.\frac{C_3^2.3.3.3}{6.6.6} + 3.\frac{C_3^3.3.3.3}{6.6.6}\\ &= 1\frac{3}{8} + 2\frac{3}{8} + 3\frac{1}{8} = 1,5\end{aligned},$$

$$\sigma^2 = \mathrm{Var}(X_i) = E(X_i^2) - E(X_i)^2$$

trong đó

$$E(X_i^2) = 1^2\frac{3}{8} + 2^2\frac{3}{8} + 3^2\frac{1}{8} = 3$$

Do đó

$$\sigma^2 = E(X_i^2) - E(X_i)^2 = 3 - 1,5^2 = 0,75.$$

Theo Định lý giới hạn trung tâm, $S = X_1 + X_2 + \cdots + X_{1000}$ có phân phối chuẩn $N(1000\mu, 1000\sigma^2)$. Khi đó

$$\begin{aligned} P(S > 1550) &= P\left(Z > \frac{1550 - 1000.1,5}{\sqrt{1000.0,75}}\right) \\ &= P(Z > 1,83) \\ &= 1 - \Phi(1,83) = 1 - 0,9664 = 0,0336. \end{aligned}$$

3.3 Các công thức xấp xỉ

3.3 Các công thức xấp xỉ

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

- Cho biến ngẫu nhiên X có phân phối siêu bội $X \sim H(N, k, n)$. Giả sử khi $N, k \to \infty$ thì $\dfrac{k}{N} = p$ luôn là một hằng số. Khi đó

$$
\begin{aligned}
P(X = x) &= \frac{C_k^x C_{N-k}^{n-x}}{C_N^n} \\
&= \frac{k!}{(k-x)!x!} \cdot \frac{(N-k)!}{(n-x)!(N-k-n+x)!} \cdot \frac{n!(N-n)!}{N!} \\
&= C_n^x \cdot \frac{k(k-1)\dots(k-x+1)}{N(N-1)\dots(N-x+1)} \cdot \frac{(N-k)(N-k-1)\dots(N-k-(n-x)+1)}{(N-x)(N-x-1)\dots(N-n+1)}
\end{aligned}
$$

Ta thấy

$$\frac{k(k-1)\dots(k-x+1)}{N(N-1)\dots(N-x+1)} \approx \frac{k^x}{N^x} = p^x$$

và

$$\frac{(N-k)(N-k-1)\dots(N-k-(n-x)+1)}{(N-x)(N-x-1)\dots(N-n+1)} \approx \frac{(N-k)^{n-x}}{N^{n-x}} = \left(\frac{N-k}{N}\right)^{n-x} = (1-p)^{n-x}$$

Do đó

$$\lim_{N,k\to\infty} P(X=x) = C_n^x p^x (1-p)^{n-x}$$

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

Ví dụ 3.17 Một vườn lan có 1 000 cây lan sắp nở hoa trong đó có 200 cây hoa màu đỏ. Chọn ngẫu nhiên 20 cây lan. Tính xác suất có 5 cây lan màu đỏ trong 20 cây.

Giải. Gọi X là số cây lan đỏ trong 20 cây. Khi đó $N = 1000, k = 200, n = 20$ và $X \sim H(1000; 200; 20)$. Vì $N = 1000 > 20n$ nên ta có thể dùng phân phối nhị thức để xấp xỉ cho phân phối siêu bội. Ta có $\dfrac{k}{N} = \dfrac{200}{1000} = 0,2$ và X xấp xỉ với phân phối nhị thức $B(20; 0,2)$. Như vậy

$$P(X = 5) \approx C_{20}^{5}.(0,2)^5.(0,8)^{15} = 0,1746.$$

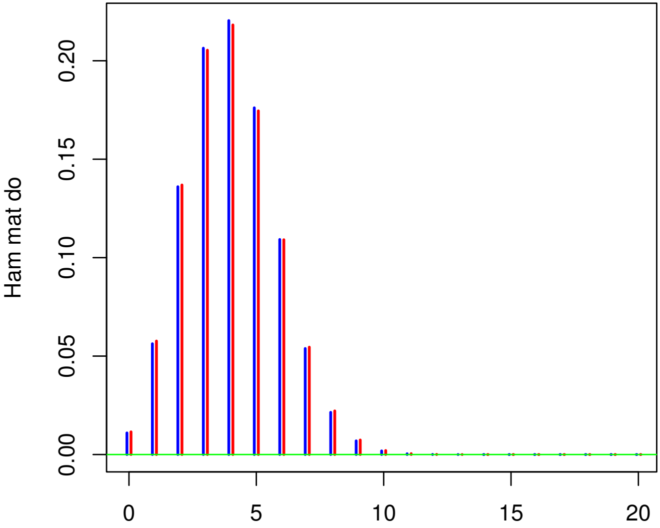
Nhận xét. Nếu dùng phân phối siêu bội thì

$$P(X = 5) = \frac{C_{200}^{5} C_{800}^{15}}{C_{1000}^{20}} = 0,1761.$$

Tính xác suất bằng R

1. Giả sử $X \sim B(20; 0,2)$ và tính $P(X = 5)$.
   - R code: `pbinom(5, 20, 0.2)-pbinom(4, 20, 0.2)`
   - Kết quả: `0.1745595`
2. Giả sử $X \sim H(1000; 200; 20)$ và tính $P(X = 5)$.
   - R code: `phyper(5,200,800,20)-phyper(4,200,800,20)`
   - Kết quả: `0.1761055`

B(20,0.5) (red) và H(1000,200,20) (blue)



Ví dụ 3.18 Một hộp chứa 20 sản phẩm trong đó có 4 sản phẩm có lỗi. Lấy ngẫu nhiên 10 sản phẩm. Tính xác suất có nhiều nhất 2 sản phẩm có lỗi.

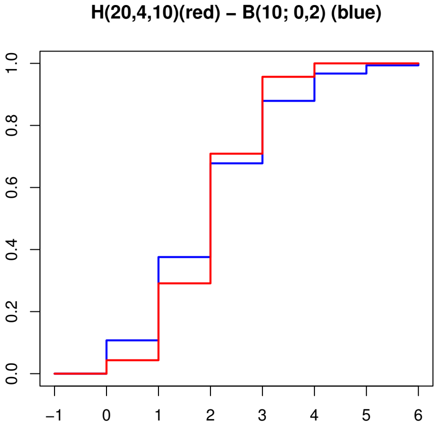
Giải. (Dùng phân phối siêu bội) Gọi X là số sản phẩm lỗi trong 10 sản phẩm. Khi đó $X \sim H(20, 4, 10)$ và do đó

$$\begin{aligned} P(X \leq 2) &= P(X = 0) + P(X = 1) + P(X = 2) \\ &= \frac{C_{16}^{10}}{C_{20}^{10}} + \frac{C_4^1 C_{16}^9}{C_{20}^{10}} + \frac{C_4^2 C_{16}^8}{C_{20}^{10}} = 0,709 \end{aligned}$$

(Dùng phân phối nhị thức) Đặt $p = \dfrac{4}{20} = 0,2$ là tỉ lệ sản phẩm lỗi. Gọi X là số sản phẩm lỗi trong 10 sản phẩm. Khi đó $X \sim B(10; 0,2)$ và do đó

$$\begin{aligned} P(X \leq 2) &= P(X = 0) + P(X = 1) + P(X = 2) \\ &= C_{10}^0 0,8^{10} + C_{10}^1 0,2^1 0,8^9 + C_{10}^2 0,2^2 0,8^8 = 0,6778 \end{aligned}$$

Chú ý: Trong ví dụ trên, ta không nên dùng phân phối nhị thức để xấp xỉ cho phân phối siêu bội vì $N = 20$ và $n = 10$ không thỏa mãn $N \geq 20n$.



3.3.2 Xấp xỉ phân phối nhị thức bằng phân phối Poisson

- Cho biến ngẫu nhiên X có phân phối nhị thức $B(n; p)$.
- Theo Ví dụ 3.5, ta đã có $B(n; p) \xrightarrow{d} P(np)$.
- **Nếu** $n \geq 30$ **và** $p \leq 0,05$ thì ta dùng **phân phối Poisson** để xấp xỉ cho phân phân phối nhị thức. Ta thấy $X \approx P(np)$.

Ví dụ 3.19 Trong một dây chuyền sản xuất chip máy tính, xác suất một chip bị lỗi là 0,005. Chọn ngẫu nhiên 400 chip để kiểm tra.

a. Tính xác suất có 1 chip bị lỗi.

b. Tính xác suất có nhiều hơn 3 chip bị lỗi.

Giải. Đặt X là số chip bị lỗi trong 400 chip. Ta thấy $X \sim B(400; 0,005)$. Vì $n = 400$ và $p = 0,005$ nên ta có thể dùng xấp xỉ phân phối Poisson. Khi đó $X \approx P(\lambda)$ với $\lambda = np = 2$.

a. $P(X = 1) = e^{-2}1^1 = 0,271.$

b. $$P(X \leq 3) = \sum_{k=0}^{3} \frac{e^{-2}2^k}{k!} = 0,857.$$

Suy ra $P(X > 3) = 1 - P(X \leq 3) = 0,143.$

Tính xác suất bằng R

a. Giả sử $X \sim P(2)$ và tính $P(X = 1)$.

- R code: `ppois(1, 2)-ppois(0, 2)`
- Kết quả: `0.2706706`

b. Giả sử $X \sim P(2)$ và tính $P(X > 3)$.

- R code: `1 - ppois(3, 2)`
- Kết quả: `0.1428765`

Ví dụ 3.20 Xác suất một máy tính đã được cài chương trình diệt virus bị nhiễm virus là 0,03. Chọn ngẫu nhiên 200 máy tính đã được cài đặt chương trình diệt virus. Tính xác suất có nhiều nhất 6 máy bị nhiễm virus.

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

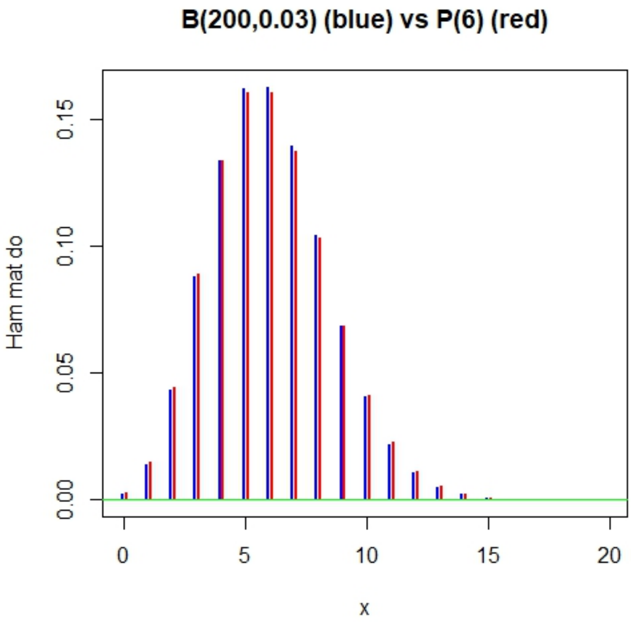
$$P(X \leq 6) = \sum_{k=0}^{6} C_{200}^{k}(0,03)^k.(0,97)^{200-k} = 0,6063.$$

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

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

Tính xác suất bằng R

1. Giả sử $X \sim B(200; 0,03)$ và tính $P(X \leq 6)$.
   - R code: `pbinom(6, 200, 0.03)`
   - Kết quả: `0.6063152`
2. Giả sử $X \sim P(6)$ và tính $P(X \leq 6)$.
   - R code: `ppois(6, 6)`
   - Kết quả: `0.6063028`



B(200,0.03) (blue) vs P(6) (red)
Ham mat do
0.00
0.05
0.10
0.15
0
5
10
15
20
x

3.3.3 Xấp xỉ phân phối nhị thức bằng phân phối chuẩn

- Cho $X_1, X_2, \ldots, X_n$ là n biến ngẫu nhiên có phân phối nhị thức $B(1;p)$. Khi đó biến ngẫu nhiên $X = \sum_{i=1}^{n} X_i$ có phân phối nhị thức $B(n;p)$
- Vì $E(X_i) = p$ và $\mathrm{Var}(X_i) = p(1-p)$ với $i = 1, 2, \ldots, n$ nên theo Định lý giới hạn trung tâm $X \sim N(np; np(1-p))$.
- **Khi** $np \geq 5$ **và** $n(1-p) \geq 5$ thì ta có thể dùng phân phối chuẩn xấp xỉ cho phân phối nhị thức. Ta có

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

Kỹ thuật *hiệu chỉnh liên tục* (continuity correction)

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

- Nếu X có phân phối nhị thức thì

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

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

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

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

Ví dụ 3.21 Một loại virus máy tính mới tấn công một thư mục bao gồm 1350 tệp. Mỗi tệp bị hỏng với xác suất 0,75 độc lập với các tệp khác.

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

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

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

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

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

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

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

Tính xác suất bằng R

1. Giả sử $X \sim B(1350; 0,75)$ và tính $P(X = 1000), P(1000 \leq X \leq 1020)$

a. R code: `pbinom(1000, 1350, 0.75) - pbinom(999, 1350, 0.75)`

- Kết quả: `0.01823763`

b. R code: `pbinom(1020, 1350, 0.75) - pbinom(999, 1350, 0.75)`

- Kết quả: `0.4846322`

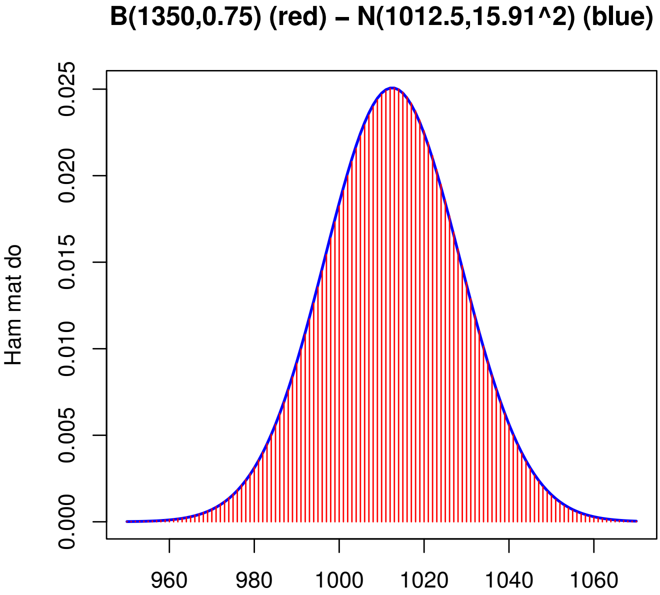
2. Giả sử $X \sim N(1012.5; 15,91^2)$ và tính $P(999,5 < X < 1000.5), P(999.5 < X < 1020.5)$

a. R code: `pnorm(1000.5, 1012.5, 15.91) - pnorm(999.5, 1012.5, 15.91)`

- Kết quả: `0.01841504`

b. R code: `pnorm(1020.5, 1012.5, 15.91) - pnorm(999.5, 1012.5, 15.91)`

- Kết quả: `0.4855208`


B(1350,0.75) (red) – N(1012.5,15.91^2) (blue)
Ham mat do
0.000
0.005
0.010
0.015
0.020
0.025
960
980
1000
1020
1040
1060

3.3.4 Xấp xỉ phân phối Poisson bằng phân phối chuẩn

- Cho biến ngẫu nhiên $X \sim P(\lambda)$ và $\mu = \lambda, \sigma^2 = \lambda$.
- Nếu λ đủ lớn thì $X \approx N(\lambda, \lambda)$.
- Nếu $\lambda \geq 20$ thì ta có thể dùng phân phối chuẩn để xấp xỉ cho phân phối Poisson.
- **Chú ý:** Dùng hiệu chỉnh liên tục khi chuyển từ phân phối Poisson sang phân phối chuẩn.

Ví dụ 3.22 Số lượng người truy cập vào một trang web trên Internet trong một ngày tuân theo phân phối Poisson. Trung bình có 30 người truy cập trong một ngày.

a. Tính xác suất có ít nhất 25 người truy cập trong một ngày.

b. Tính xác suất có nhiều hơn 32 người truy cập trong một ngày.

Giải. Đặt X là số người truy cập vào trang web trong một ngày. Ta có X có phân phối Poisson $X \sim P(\lambda)$ với $\lambda = 30$. Vì $\lambda > 20$ nên $X \approx N(30; 30)$.

a. Xác suất có ít nhất 25 người truy cập trong một ngày.

$$\begin{aligned} P(X \geq 25) &= P(X > 24,5) \\ &= P\left(\frac{X - 30}{\sqrt{30}} > \frac{24,5 - 30}{\sqrt{30}}\right) \\ &= P(Z > -1) \\ &= 1 - \Phi(-1) = 1 - 0,1587 = 0,8423 \end{aligned}$$

b. Xác suất có nhiều hơn 32 người truy cập trong một ngày.

$$\begin{aligned} P(X > 32) &= P(X > 32,5) \\ &= P\left(\frac{X-30}{\sqrt{30}} > \frac{32,5-30}{\sqrt{30}}\right) \\ &= P(Z > 0,46) \\ &= 1 - \Phi(0,46) = 1 - 0,6772 = 0,3228. \end{aligned}$$

Tính xác suất bằng R

1. Giả sử $X \sim P(30)$ và tính $P(X \geq 25), P(X > 32)$

   a. R code: `1-ppois(24,30)` $(P(X \geq 25) = 1 - P(X \leq 24))$

   - Kết quả: `0.842758`

   b. R code: `1-ppois(32,30)` $(P(X > 32) = 1 - P(X \leq 32))$

   - Kết quả: `0.3154588`

2. Giả sử $X \sim N(30; 30)$ và tính $P(X > 24,5), P(X > 32,5)$

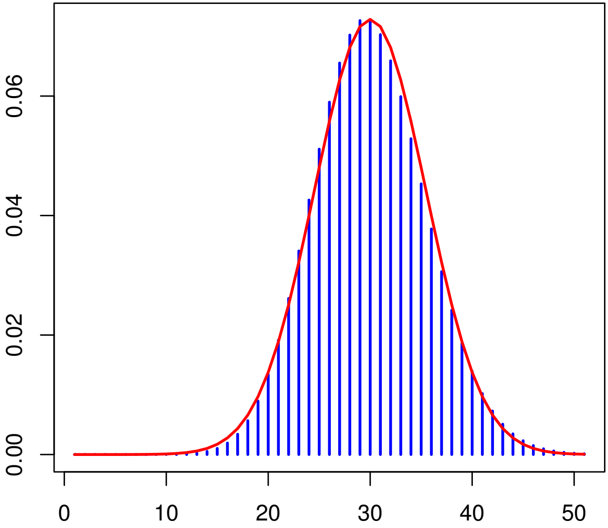
   a. R code: `1-pnorm(24.5, 30, sqrt(30))`

   - Kết quả: `0.8423488`

   b. R code: `1- pnorm(32.5, 30, sqrt(30))`

   - Kết quả: `0.3240384`

P(30) – N(30,sqrt(30))



Bài tập

Bài 3.1 Theo một cuộc khảo sát, thời gian xem tivi trung bình của các em bé từ 2 đến 5 tuổi là 25 giờ mỗi tuần. Giả sử thời gian xem tivi có phân phối chuẩn với độ lệch chuẩn là 3 giờ. Chọn ngẫu nhiên 20 em bé từ 2 đến 5 tuổi. Tính xác suất thời gian xem tivi trung bình của 20 bé lớn hơn 26,3 giờ. (Đáp số: 0,0262)

Bài 3.2 Một cuộc nghiên cứu về stress được thực hiện đối với các sinh viên một trường đại học. Giả sử điểm stress là một biến ngẫu nhiên liên tục có phân phối đều $X \sim U(1;5)$. Chọn ngẫu nhiên 75 sinh viên của trường đó. Đặt $\overline{X}$ là điểm stress trung bình của nhóm sinh viên này.

a. Tìm k sao cho $P(\overline{X} < k) = 0,9$ (Đáp số: $k = 3,2$)

b. Tính xác suất tổng điểm stress của 75 sinh viên này nhỏ hơn 200. (Đáp số: 0)

Bài 3.3 Bạn đang đứng chờ một thang máy có tải trọng 650 kg. Khi thang máy đến, đã có 10 người trong đó. Giả sử bạn nặng 70 kg và cân nặng của người có phân phối chuẩn với trung bình 58 kg và độ lệch chuẩn là 0,9 kg. Tính xác suất khi bạn bước vào thang máy bị vượt tải trọng. (Đáp số: 0,5)

Bài tập

Bài 3.4 Giả sử xác suất bị lỗi trong một lần giao dịch của một ngân hàng là 0,0003. Nếu có 10000 giao dịch được thực hiện thì xác suất có hơn 6 lỗi là bao nhiêu? (Đáp số: 0,0335)

Bài 3.5 Một hộp chứa 10000 viên bi trong đó có 500 bi đỏ. Lấy ngẫu nhiên 10 bi (không thay thế). Xác suất trong 10 bi lấy ra có nhiều nhất 1 bi bỏ là bao nhiêu? Yêu cầu: Giải bằng hai cách khác nhau. (Đáp số: 0,9139 (phân phối siêu bội), 0,9138 (phân phối nhị thức))

Bài 3.6 Theo một cuộc khảo sát, người ta thấy rằng có 25% người lái xe buồn ngủ khi đang lái xe. Chọn ngẫu nhiên 200 người lái xe. Dùng xấp xỉ phân phối chuẩn để tính xác suất có 62 người buồn ngủ khi đang lái xe. (Đáp số: 0,94)

Bài 3.7 Người ta thấy rằng xác suất những người sống đến 65 tuổi ở một thị trấn nọ là 0,8. Chọn ngẫu nhiên 500 người. Dùng xấp xỉ phân phối chuẩn, tính xác suất có từ 375 đến 425 người sống đến 65 tuổi. (Đáp số: 0,9956)