ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN
UIT TOÀN DIỆN • SÁNG TẠO • PHỤNG SỰ

BÀI GIẢNG
XÁC SUẤT THỐNG KÊ

Giảng viên: TS. PHÙNG MINH ĐỨC

(Bộ môn Toán Lý)

Chương 3: Luật số lớn và định lý giới hạn trung tâm

3.1 Các dạng hội tụ của dãy các biến ngẫu nhiên

3.2 Luật số lớn và định lý giới hạn trung tâm

3.3 Các công thức xấp xỉ

3.1 Các dạng hội tụ của dãy các biến ngẫu nhiên

Giả sử chúng ta muốn tìm hiểu một biến ngẫu nhiên X, nhưng ta không thể quan sát X một cách trực tiếp.

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: thu được $X_1, X_2, X_3, \ldots$.

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: thu được $X_1, X_2, X_3, \ldots$. 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.

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: thu được $X_1, X_2, X_3, \ldots$. 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.

Câu hỏi.

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 3.1.1

Xét phép thử ngẫu nhiên: Gieo một đồng xu.

Ví dụ 3.1.1

Xét phép thử ngẫu nhiên: Gieo một đồng xu. Không gian mẫu $\Omega = \{S, N\}$ (S là xuất hiện mặt sấp, N là xuất hiện mặt ngửa).

Ví dụ 3.1.1

Xét phép thử ngẫu nhiên: 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 $X_1, X_2, X_3, \ldots$ như sau

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

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.

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)

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)

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)

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)

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.

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".

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.

Định nghĩa 3.1.1 (Hội tụ theo xác suất)

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 $X_n \xrightarrow{P} X$,

Định nghĩa 3.1.1 (Hội tụ theo xác suất)

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

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

Định nghĩa 3.1.1 (Hội tụ theo xác suất)

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 3.1.2

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

Định nghĩa 3.1.1 (Hội tụ theo xác suất)

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 3.1.2

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

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

Định nghĩa 3.1.1 (Hội tụ theo xác suất)

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 3.1.2

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 $X_n \xrightarrow{P} 0$.

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

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

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

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

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

$$\lim_{n\to\infty} P(|X_n - 0| \geq \varepsilon) = \lim_{n\to\infty} P(X_n \geq \varepsilon) \text{ vì } X_n \geq 0$$
$$= \lim_{n\to\infty} \left(\int_{\varepsilon}^{\infty} f_{X_n}(t)dt \right)$$

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

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

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

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

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

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

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

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

Mệnh đề 3.1.1

Cho $\{X_n\}_{n\geq 1}$ hội tụ theo xác suất đến μ_1 và $\{Y_n\}_{n\geq 1}$ hội tụ theo xác suất đến μ_2. Khi đó

1. $\{X_n + Y_n\}_{n\geq 1}$ hội tụ theo xác suất đến $\mu_1 + \mu_2$.
2. $\{X_nY_n\}_{n\geq 1}$ hội tụ theo xác suất đến $\mu_1\mu_2$.
3. $\{X_n/Y_n\}_{n\geq 1}$ hội tụ theo xác suất đến μ_1/μ_2 nếu $\mu_2 \neq 0$.
4. $\{\sqrt{X_n}\}_{n\geq 1}$ hội tụ theo xác suất đến $\sqrt{\mu_1}$ nếu $P(X_n \geq 0) = 1$ với mọi n.

Định nghĩa 3.1.2 (Hội tụ theo phân phối)

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 $X_n \xrightarrow{d} X$,

Định nghĩa 3.1.2 (Hội tụ theo phân phối)

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.

Định nghĩa 3.1.2 (Hội tụ theo phân phối)

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 3.1.3

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

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

Định nghĩa 3.1.2 (Hội tụ theo phân phối)

Dãy các biến ngẫu nhiên $X_1, X_2, \dots$ 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 3.1.3

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

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

Khi đó $X_n \xrightarrow{d} 1 - e^{-x}$.

Giải. Với mọi $x \leq 0$, ta có

$$F_{X_n}(x) = F_X(x) = 0, n = 1, 2, \ldots.$$

Giải. Với mọi $x \leq 0$, ta có

$$F_{X_n}(x) = F_X(x) = 0, n = 1, 2, \ldots.$$

Xét $x \geq 0$, ta có

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

Giải. Với mọi $x \leq 0$, ta có

$$F_{X_n}(x) = F_X(x) = 0, n = 1, 2, \ldots.$$

Xét $x \geq 0$, ta có

$$\lim_{n\to\infty} F_{X_n}(x) = \lim_{n\to\infty} \left(1 - \left(1 - \frac{1}{n}\right)^{nx}\right)$$

Giải. Với mọi $x \leq 0$, ta có

$$F_{X_n}(x) = F_X(x) = 0, n = 1, 2, \ldots .$$

Xét $x \geq 0$, ta có

$$\begin{aligned}
\lim_{n\to\infty} F_{X_n}(x) &= \lim_{n\to\infty} \left(1 - \left(1 - \frac{1}{n}\right)^{nx}\right) \\
&= 1 - \lim_{n\to\infty} \left(1 - \frac{1}{n}\right)^{nx}
\end{aligned}$$

Giải. Với mọi $x \leq 0$, ta có

$$F_{X_n}(x) = F_X(x) = 0, n = 1, 2, \ldots.$$

Xét $x \geq 0$, ta có

$$\begin{aligned}
\lim_{n\to\infty} F_{X_n}(x) &= \lim_{n\to\infty} \left(1 - \left(1 - \frac{1}{n}\right)^{nx}\right) \\
&= 1 - \lim_{n\to\infty} \left(1 - \frac{1}{n}\right)^{nx} \\
&= 1 - e^{-x} = F_X(x)
\end{aligned}$$

Giải. Với mọi $x \leq 0$, ta có

$$F_{X_n}(x) = F_X(x) = 0, n = 1, 2, \ldots .$$

Xét $x \geq 0$, 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 3.1.1

Cho dãy các biến ngẫu nhiên $X_1, X_2, X_3, \ldots$ và biến ngẫu nhiên X.

Định lý 3.1.1

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.

Định lý 3.1.1

Cho dãy các biến ngẫu nhiên $X_1, X_2, X_3, \dots$ 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 $k = 1, 2, \dots$.

Định lý 3.1.1

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 3.1.4

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.

Định lý 3.1.1

Cho dãy các biến ngẫu nhiên $X_1, X_2, X_3, \dots$ 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 $k = 1, 2, \dots$.

Ví dụ 3.1.4

Cho dãy các biến ngẫu nhiên $X_1, X_2, X_3, \dots$ có phân phối nhị thức

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

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 3.1.1, ta chứng minh $\lim\limits_{n\to\infty} P(X_n = k) = P(X = k)$

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

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

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

Giải. Theo Định lý 3.1.1, 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

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

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

Giải. Theo Định lý 3.1.1, 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

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

Hội tụ hầu chắc chắn

Một dãy các biến ngẫu nhiên $X_1, X_2, \ldots$ xác định trên cùng một không gian mẫu Ω được gọi là hội tụ hầu chắc chắn về một biến ngẫu nhiên X, ký hiệu $X_n \stackrel{a.s}{\to} X$, nếu

$$P(\{s \in \Omega \mid \lim_{n\to\infty} X_n(s) = X(s)\}) = 1.$$

3.2 Luật số lớn và Định lý giới hạn trung tâm

Định lý 3.2.1 (Luật số lớn yếu-weak law of large numbers)

Cho $X_1, X_2, \dots, X_n$ là các biến ngẫu nhiên độc lập và phân phối giống nhau với $E(X_i) = \mu < \infty$ với mọi i.

Định lý 3.2.1 (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 phân phối giống nhau với $E(X_i) = \mu < \infty$ với mọi i. Đặt

$$\overline{X}_n = \frac{X_1 + X_2 + \ldots + X_n}{n}.$$

Khi đó $\overline{X}_n$ hội tụ theo xác suất đến μ.

Định lý 3.2.1 (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 phân phối giống nhau với $E(X_i) = \mu < \infty$ với mọi i. Đặt

$$\overline{X}_n = \frac{X_1 + X_2 + \ldots + X_n}{n}.$$

Khi đó $\overline{X}_n$ hội tụ theo xác suất đến μ.

Luật số lớn mạnh-strong law of large numbers)

Cho $X_1, X_2, \ldots, X_n$ là các biến ngẫu nhiên độc lập và phân phối giống nhau với $E(X_i) = \mu < \infty$ với mọi i. Đặt

$$\overline{X}_n = \frac{X_1 + X_2 + \ldots + X_n}{n}.$$

Khi đó $\overline{X}_n$ hội tụ theo hầu chắc chắn đến μ.

Định lý 3.2.2 (Định lý giới hạn trung tâm)

Cho $X_1, X_2, \ldots, X_n$ là các biến ngẫu nhiên độc lập và phân phối giống nhau với kỳ vọng $E(X_i) = \mu$ và phương sai $V(X_i) = \sigma^2 < \infty$ với mọi i.

Định lý 3.2.2 (Định lý giới hạn trung tâm)

Cho $X_1, X_2, \ldots, X_n$ là các biến ngẫu nhiên độc lập và phân phối giống nhau với kỳ vọng $E(X_i) = \mu$ và phương sai $V(X_i) = \sigma^2 < \infty$ với mọi i. Đặt

$$S_n = X_1 + X_2 + \ldots + X_n \text{ và } Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}.$$

Định lý 3.2.2 (Định lý giới hạn trung tâm)

Cho $X_1, X_2, \ldots, X_n$ là các biến ngẫu nhiên độc lập và phân phối giống nhau với kỳ vọng $E(X_i) = \mu$ và phương sai $V(X_i) = \sigma^2 < \infty$ với mọi i. Đặt

$$S_n = X_1 + X_2 + \ldots + X_n \text{ và } Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}.$$

Khi đó $Z_n \xrightarrow{d} Z \sim N(0; 1)$.

Định lý 3.2.2 (Định lý giới hạn trung tâm)

Cho $X_1, X_2, \ldots, X_n$ là các biến ngẫu nhiên độc lập và phân phối giống nhau với kỳ vọng $E(X_i) = \mu$ và phương sai $V(X_i) = \sigma^2 < \infty$ với mọi i. Đặt

$$S_n = X_1 + X_2 + \ldots + X_n \text{ và } Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}.$$

Khi đó $Z_n \xrightarrow{d} Z \sim N(0;1)$.

Định lý 3.2.3

Cho $X_1, X_2, \ldots, X_n$ là các biến ngẫu nhiên độc lập và phân phối giống nhau với kỳ vọng $E(X_i) = \mu$ và phương sai $V(X_i) = \sigma^2 < \infty$ với mọi i.

Định lý 3.2.2 (Định lý giới hạn trung tâm)

Cho $X_1, X_2, \ldots, X_n$ là các biến ngẫu nhiên độc lập và phân phối giống nhau với kỳ vọng $E(X_i) = \mu$ và phương sai $V(X_i) = \sigma^2 < \infty$ với mọi i. Đặt

$$S_n = X_1 + X_2 + \ldots + X_n \text{ và } Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}.$$

Khi đó $Z_n \xrightarrow{d} Z \sim N(0;1)$.

Định lý 3.2.3

Cho $X_1, X_2, \ldots, X_n$ là các biến ngẫu nhiên độc lập và phân phối giống nhau với kỳ vọng $E(X_i) = \mu$ và phương sai $V(X_i) = \sigma^2 < \infty$ với mọi i. Đặt

$$\overline{X}_n = \frac{X_1 + X_2 + \ldots + X_n}{n} \text{ và } Y_n = \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}.$$

Định lý 3.2.2 (Định lý giới hạn trung tâm)

Cho $X_1, X_2, \ldots, X_n$ là các biến ngẫu nhiên độc lập và phân phối giống nhau với kỳ vọng $E(X_i) = \mu$ và phương sai $V(X_i) = \sigma^2 < \infty$ với mọi i. Đặt

$$S_n = X_1 + X_2 + \ldots + X_n \text{ và } Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}.$$

Khi đó $Z_n \xrightarrow{d} Z \sim N(0;1)$.

Định lý 3.2.3

Cho $X_1, X_2, \ldots, X_n$ là các biến ngẫu nhiên độc lập và phân phối giống nhau với kỳ vọng $E(X_i) = \mu$ và phương sai $V(X_i) = \sigma^2 < \infty$ với mọi i. Đặt

$$\overline{X}_n = \frac{X_1 + X_2 + \ldots + X_n}{n} \text{ và } Y_n = \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}.$$

Khi đó $Y_n \xrightarrow{d} Z \sim N(0;1)$.

Nhận xét 3.2.1

1. Nếu các biến ngẫu nhiên X_i độc lập và **đều có phân phối chuẩn** thì tổng S_n và S_n/n cũng có phân phối chuẩn.

Nhận xét 3.2.1

1. Nếu các biến ngẫu nhiên X_i độc lập và **đều có phân phối chuẩn** thì tổng S_n và S_n/n cũng có phân phối chuẩn.
2. Cho X_i là các biến ngẫu nhiên độc lập và **có phân phối giống nhau** với cùng trung bình μ và phương sai σ^2. Khi đó, với n đủ lớn (thông thường $n \geq 30$)

Nhận xét 3.2.1

1. Nếu các biến ngẫu nhiên X_i độc lập và **đều có phân phối chuẩn** thì tổng S_n và S_n/n cũng có phân phối chuẩn.
2. Cho X_i là các biến ngẫu nhiên độc lập và **có phân phối giống nhau** với cùng trung bình μ và phương sai σ^2. Khi đó, với n đủ lớn (thông thường $n \geq 30$)

$$S_n = X_1 + X_2 + \ldots + X_n \approx N(n\mu; n\sigma^2)$$

Nhận xét 3.2.1

1. Nếu các biến ngẫu nhiên X_i độc lập và **đều có phân phối chuẩn** thì tổng S_n và S_n/n cũng có phân phối chuẩn.
2. Cho X_i là các biến ngẫu nhiên độc lập và **có phân phối giống nhau** với cùng trung bình μ và phương sai σ^2. Khi đó, với n đủ lớn (thông thường $n \geq 30$)

$$S_n = X_1 + X_2 + \ldots + X_n \approx N(n\mu; n\sigma^2)$$

và

$$\overline{X}_n = \frac{X_1 + X_2 + \ldots + X_n}{n} \approx N(\mu; \sigma^2/n).$$

Ví dụ 3.2.1

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.

Ví dụ 3.2.1

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.

Giải. Đặt X_i là thời gian xem tivi của bé thứ i.

Ví dụ 3.2.1

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.

Giải. Đặt X_i là thời gian xem tivi của bé thứ i. Thời gian xem trung bình của 20 bé là $Y = \dfrac{X_1 + \ldots + X_{20}}{20}$.

Ví dụ 3.2.1

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.

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, $Y \sim N(\mu; \sigma^2/n)$.

Ví dụ 3.2.1

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.

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, $Y \sim N(\mu; \sigma^2/n)$. 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 3.2.2

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. Hỏi ổ cứng đó có khả năng lưu được 300 hình này không?

Ví dụ 3.2.2

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. Hỏi ổ cứng đó có khả năng lưu được 300 hình này không?

Giải. Đặt X_i là dung lượng của hình ảnh thứ i

Ví dụ 3.2.2

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. Hỏi ổ cứng đó có khả năng lưu được 300 hình này không?

Giải. Đặt X_i là dung lượng của hình ảnh thứ i và $S = \sum_{i=1}^{300} X_i$.

Ví dụ 3.2.2

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. Hỏi ổ cứng đó có khả năng lưu được 300 hình này không?

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 $V(X_i) = \sigma^2 = 0,5^2$. Áp dụng Đinh lý giới hạn trung tâm, $S \approx N(n\mu; n\sigma^2)$ ta có

Ví dụ 3.2.2

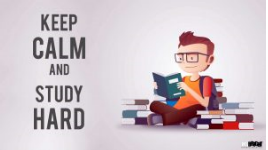
Một đĩa cứng có dung lượng trống là 330 megabyte. Cho 300 hình ảnh độc lập, kích thước trung bình mỗi ảnh là 1 megabyte với độ lệch chuẩn là 0,5 megabyte. Hỏi ổ cứng đó có khả năng lưu được 300 hình này không?

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 $V(X_i) = \sigma^2 = 0,5^2$. Áp dụng Đinh lý giới hạn trung tâm, $S \approx N(n\mu; n\sigma^2)$ ta có

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

Vì xác suất này rất cao nên đĩa có khả năng lưu được 300 ảnh.

Bài tập 1. Giả sử thời gian chơi thể thao (tính bằng giờ) của các sinh viên ở một trường đại học có phân phối chuẩn với trung bình là 2 giờ và độ lệch chuẩn là 0,5 giờ. Chọn ngẫu nhiên 50 sinh viên. Xác suất thời gian chơi thể thao trung bình của 50 sinh viên này từ 1,8 giờ đến 2,3 giờ là bao nhiêu?



Bài tập 2. Giả sử thời gian tự học ở nhà của các sinh viên trường UIT có phân phối giống nhau với trung bình 3 giờ và độ lệch chuẩn 1,15 giờ. Chọn ngẫu nhiên 75 sinh viên. Tính xác suất 75 sinh viên này có tổng thời gian tự học ở nhà ít hơn 200 giờ.

3.3 Các công thức xấp xỉ

- **Xấp xỉ phân phối siêu bội bằng phân phối nhị thức:**
  - Cho biến ngẫu nhiên X có phân phối siêu bội $X \sim H(N, M, n)$.

- **Xấp xỉ phân phối siêu bội bằng phân phối nhị thức:**
  - Cho biến ngẫu nhiên X có phân phối siêu bội $X \sim H(N, M, n)$.
  - Nếu $N \to \infty$ thì $H(N, M, n) \xrightarrow{P} B(n, \frac{M}{N})$.

- **Xấp xỉ phân phối siêu bội bằng phân phối nhị thức:**
  - Cho biến ngẫu nhiên X có phân phối siêu bội $X \sim H(N, M, n)$.
  - Nếu $N \to \infty$ thì $H(N, M, n) \xrightarrow{P} B(n, \frac{M}{N})$.
  - Khi $N \geq 20n$, ta có thể dùng phân phối nhị thức để xấp xỉ cho phân phối siêu bội.

Ví dụ 3.3.1 Một vườn lan có 10 000 cây lan sắp nở hoa trong đó có 1000 cây hoa màu đỏ. Chọn ngẫu nhiên 20 cây lan. Tính xác suất có 5 cây lan màu đỏ trong 20 cây.

Ví dụ 3.3.1 Một vườn lan có 10 000 cây lan sắp nở hoa trong đó có 1000 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 $N = 10000, M = 1000, n = 20$ và $X \sim H(10000; 1000; 20)$. Vì $N = 10000 > 20n$ nên ta có thể dùng phân phối nhị thức để xấp xỉ cho phân phối siêu bội.

Ví dụ 3.3.1 Một vườn lan có 10 000 cây lan sắp nở hoa trong đó có 1000 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 $N = 10000, M = 1000, n = 20$ và $X \sim H(10000; 1000; 20)$. Vì $N = 10000 > 20n$ nên ta có thể dùng phân phối nhị thức để xấp xỉ cho phân phối siêu bội. Ta có $\frac{M}{N} = \frac{1000}{10000} = 0,1$ và $X \approx B(20; 0,1)$.

Ví dụ 3.3.1 Một vườn lan có 10 000 cây lan sắp nở hoa trong đó có 1000 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 $N = 10000, M = 1000, n = 20$ và $X \sim H(10000; 1000; 20)$. Vì $N = 10000 > 20n$ nên ta có thể dùng phân phối nhị thức để xấp xỉ cho phân phối siêu bội. Ta có $\frac{M}{N} = \frac{1000}{10000} = 0,1$ và $X \approx B(20; 0,1)$. Như vậy

$$P(X = 5) \approx C_{20}^{5}.(0,1)^5.(0,9)^{15} = 0,0319.$$

Ví dụ 3.3.1 Một vườn lan có 10 000 cây lan sắp nở hoa trong đó có 1000 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 $N = 10000, M = 1000, n = 20$ và $X \sim H(10000; 1000; 20)$. Vì $N = 10000 > 20n$ nên ta có thể dùng phân phối nhị thức để xấp xỉ cho phân phối siêu bội. Ta có $\frac{M}{N} = \frac{1000}{10000} = 0,1$ và $X \approx B(20; 0,1)$. Như vậy

$$P(X = 5) \approx C_{20}^{5}.(0,1)^5.(0,9)^{15} = 0,0319.$$

Nhận xét. Nếu ta dùng phân phối siêu bội thì

Ví dụ 3.3.1 Một vườn lan có 10 000 cây lan sắp nở hoa trong đó có 1000 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 $N = 10000, M = 1000, n = 20$ và $X \sim H(10000; 1000; 20)$. Vì $N = 10000 > 20n$ nên ta có thể dùng phân phối nhị thức để xấp xỉ cho phân phối siêu bội. Ta có $\frac{M}{N} = \frac{1000}{10000} = 0,1$ và $X \approx B(20; 0,1)$. Như vậy

$$P(X = 5) \approx C_{20}^{5}.(0,1)^5.(0,9)^{15} = 0,0319.$$

Nhận xét. Nếu ta dùng phân phối siêu bội thì

$$P(X = 5) = \frac{C_{1000}^{5} C_{9000}^{15}}{C_{10000}^{20}} = 0,0318.$$

- **Xấp xỉ phân phối nhị thức bằng phân phối Poisson:**
  - Cho X là một biến ngẫu nhiên có phân phối nhị thức $X \sim B(n; p)$.

- **Xấp xỉ phân phối nhị thức bằng phân phối Poisson:**
  - Cho X là một biến ngẫu nhiên có phân phối nhị thức $X \sim B(n;p)$.
  - Theo Ví dụ 3.1.4, ta thấy $B(n;p) \xrightarrow{d} P(np)$.

- **Xấp xỉ phân phối nhị thức bằng phân phối Poisson:**
  - Cho X là một biến ngẫu nhiên có phân phối nhị thức $X \sim B(n;p)$.
  - Theo Ví dụ 3.1.4, ta thấy $B(n;p) \xrightarrow{d} P(np)$.
  - **Nếu** $n \geq 30$ **và** $p \leq 0,05$ thì ta dùng **phân phối Poisson** để xấp xỉ cho phân phân phối nhị thức: $X \approx P(np)$.

- **Xấp xỉ phân phối nhị thức bằng phân phối Poisson:**
  - Cho X là một biến ngẫu nhiên có phân phối nhị thức $X \sim B(n; p)$.
  - Theo Ví dụ 3.1.4, ta thấy $B(n; p) \xrightarrow{d} P(np)$.
  - **Nếu** $n \geq 30$ **và** $p \leq 0,05$ thì ta dùng **phân phối Poisson** để xấp xỉ cho phân phân phối nhị thức: $X \approx P(np)$.

Ví dụ 3.3.2 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.

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

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

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 $X \sim B(200; 0,03)$.

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

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

• **Xấp xỉ phân phối nhị thức bằng phân phối chuẩn:** Cho X là biến ngẫu nhiên có phân phối nhị thức $X \sim B(n;p)$.

• **Xấp xỉ phân phối nhị thức bằng phân phối chuẩn:** Cho X là biến ngẫu nhiên có phân phối nhị thức $X \sim B(n; p)$.

- Khi n lớn ta có thể xấp xỉ

$$P(X = k) \approx \frac{1}{\sqrt{np(1-p)}} f\Big(\frac{k - np}{\sqrt{np(1-p)}}\Big),$$

ở đó $f(x) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{\frac{-x^2}{2}}$ là hàm mật độ của phân bố chuẩn tắc.

• **Xấp xỉ phân phối nhị thức bằng phân phối chuẩn:** Cho X là biến ngẫu nhiên có phân phối nhị thức $X \sim B(n;p)$.

- Khi n lớn ta có thể xấp xỉ

$$P(X=k) \approx \frac{1}{\sqrt{np(1-p)}} f\Big(\frac{k-np}{\sqrt{np(1-p)}}\Big),$$

ở đó $f(x) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{\frac{-x^2}{2}}$ là hàm mật độ của phân bố chuẩn tắc.

- Khi $0.05 \leq p \leq 0.95$ và n lớn, ta có thể xấp xỉ cho phân phối Nhị thức bởi phân phối chuẩn

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

Xấp xỉ này tốt khi $np > 5$ và $n(1-p) > 5$ hoặc khi $np(1-p) > 20$.

- Kỹ thuật *hiệu chỉnh liên tục* (continuity correction): Để tăng độ chính xác khi xấp xỉ phân phối nhị thức bằng phân phối chuẩn, ta mở rộng khoảng 0,5 đơn vị theo mỗi hướng. Việc này được gọi là *hiệu chỉnh liên tục* và không làm thay đổi xác suất của biến cố.

• Kỹ thuật *hiệu chỉnh liên tục* (continuity correction): Để tăng độ chính xác khi xấp xỉ phân phối nhị thức bằng phân phối chuẩn, ta mở rộng khoảng 0,5 đơn vị theo mỗi hướng. Việc này được gọi là *hiệu chỉnh liên tục* và không làm thay đổi xác suất của biến cố.

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

Ví dụ 3.3.3 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?

Ví dụ 3.3.3 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 $\sigma = \sqrt{np(1-p)} = \sqrt{1350.0,75.0,25} = 15,91$.

Ví dụ 3.3.3 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. Áp dụng công thức $P(X = k) \approx \dfrac{1}{\sqrt{np(1-p)}} f\left(\dfrac{k - np}{\sqrt{np(1-p)}}\right)$,

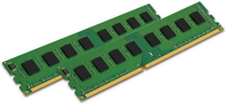
ở đó $f(x) = \dfrac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}$ là hàm mật độ của phân bố chuẩn tắc.

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

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

Giải các bài toán sau bằng 2 cách.

Bài tập 3. Một công ty sản xuất các RAM máy tính cho biết tỉ lệ RAM bị lỗi là 0,5%. Kiểm tra ngẫu nhiên 100 RAM máy tính của công ty này. Tính xác suất



a. Không có RAM bị lỗi.

b. Có nhiều hơn 1 RAM bị lỗi.

Bài tập 4. Một nhà máy xếp 1000 cái bánh quy vào mỗi hộp trong đó có 50 cái bánh loại đặc biệt. Chọn ngẫu nhiên 40 cái bánh. Tìm xác suất có ít nhất 2 cái bánh quy loạ

Bài tập 5. Giả sử một công ty chuyên sản xuất chip bán dẫn sản xuất 50 chip bị lỗi trong số 1000. Lấy ngẫu nhiên 100 chip (không thay thế). Tính xác suất có ít nhất 1 chip bị lỗi trong mẫu.