4.1 SỰ HỘI TỤ CỦA DÃY CÁC BIẾN NGẪU NHIÊN

4.1.1 Hội tụ theo xác suất

Xét dãy biến ngẫu nhiên $\{X_n\}_{n=1}^{\infty}$ và biến ngẫu nhiên X trong cùng một phép thử.

Ta nói rằng dãy các biến ngẫu nhiên $\{X_n\}_{n=1}^{\infty}$ hội tụ theo xác suất về biến ngẫu nhiên X, ký hiệu $X_n \xrightarrow[n\to\infty]{P} X$ nếu

$$\forall \varepsilon > 0 \quad \lim_{n\to\infty} P\left\{\left|X_n - X\right| > \varepsilon\right\} = 0$$

Như vậy dãy các biến ngẫu nhiên $\{X_n\}_{n=1}^{\infty}$ hội tụ theo xác suất về biến ngẫu nhiên X thì với n đủ lớn, thực tế ta có thể coi rằng, X_n không khác mấy so với X

4.1.2. Hội tụ theo phân bố

Dãy các biến ngẫu nhiên $\{X_n\}_{n=1}^{\infty}$ được gọi là hội tụ theo phân bố về biến ngẫu nhiên X nếu dãy các hàm phân bố xác suất $\{F_{X_n}(x)\}_{n=1}^{\infty}$ hội tụ về hàm phân bố xác suất $F_X(x)$.

Tức là với mọi $x \in R$

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

Trường hợp dãy các biến ngẫu nhiên rời rạc $\{X_n\}_{n=1}^{\infty}$ và biến ngẫu nhiên rời rạc X có cùng tập giá trị $R = \{c_1, c_2, \dots\}$ thì hội tụ theo phân bố tương đương với điều kiện $\forall c_k \in R$

$$\lim_{n\to\infty} P\{X_n = c_k\} = P\{X = c_k\}$$

4.2. LUẬT SỐ LỚN

4.2.1. Bất đẳng thức Markov

Cho Y là biến ngẫu nhiên không âm có kỳ vọng hữu hạn. Khi đó với mọi $a > 0$ ta có

$$P\{Y > a\} \le \frac{\mathrm{E}Y}{a}$$

4.2.2 Bất đẳng thức Trêbưsép

Giả sử X là biến ngẫu nhiên có kỳ vọng và phương sai hữu hạn, khi đó với mọi $\varepsilon > 0$ ta có

$$P\{|X - \mathrm{E}X| > \varepsilon\} \le \frac{\mathrm{D}X}{\varepsilon^2} \qquad P\{|X - \mathrm{E}X| \le \varepsilon\} \ge 1 - \frac{\mathrm{D}X}{\varepsilon^2}$$

4.2.3 Luật số lớn Trêbưsép

Định lý: Giả sử là dãy các biến ngẫu nhiên độc lập $X_1, X_2, \ldots$, có các kỳ vọng hữu hạn và phương sai đều bị chặn trên bởi hằng số C($\mathrm{D}X_i \le C;\ \forall i = 1, 2, \ldots$). Khi đó

$$\lim_{n\to\infty} P\left(\left| \frac{X_1 + \cdots + X_n}{n} - \frac{\mathrm{E}\, X_1 + \cdots + \mathrm{E}\, X_n}{n} \right| > \varepsilon \right) = 0$$

Hệ quả 1: Giả sử $X_1, X_2, \ldots$ là dãy các biến ngẫu nhiên độc lập có cùng kỳ vọng μ và phương sai đều bị chặn trên bởi hằng số C ($\mathrm{D}X_i \le C;\ \forall i = 1, 2, \ldots$). Khi đó

$$\boxed{\frac{X_1 + \cdots + X_n}{n} \xrightarrow[n\to\infty]{P} \mu}$$

Hệ quả 2: Giả sử $X_1, X_2, \ldots$ là dãy các biến ngẫu nhiên độc lập có cùng phân bố, có kỳ vọng μ và phương sai σ^2. Khi đó

$$\frac{X_1 + \cdots + X_n}{n} \xrightarrow[n\to\infty]{P} \mu$$

4.3. LUẬT SỐ LỚN BERNOULLI

Xét phép thử ngẫu nhiên **C** và A là một biến cố liên quan đến phép thử **C**. Tiến hành n lần độc lập phép thử **C** và gọi k_n là tần số xuất hiện biến cố A trong n phép thử đó.

$f_n = \dfrac{k_n}{n}$ được gọi là tần suất xuất hiện của A trong n phép thử

Định lý: Tần suất f_n hội tụ theo xác suất về xác suất p của biến cố A.

Nghĩa là với mọi $\varepsilon > 0$

$$\lim_{n\to\infty} P\left\{\left| f_n - p\right| < \varepsilon\right\} = 1$$

4.3. ĐỊNH LÝ GIỚI HẠN TRUNG TÂM

Giả sử $X_1, X_2, \ldots$ là dãy các biến ngẫu nhiên độc lập có cùng phân bố, có kỳ vọng μ và phương sai σ^2

$$\mathrm{E}\left[\frac{X_1+\cdots+X_n}{n}\right]=\frac{n\mu}{n}=\mu \qquad \mathrm{D}\left[\frac{X_1+\cdots+X_n}{n}\right]=\frac{n\sigma^2}{n^2}=\frac{\sigma^2}{n}$$

$$S_n=\frac{X_1+\cdots+X_n-n\mu}{\sigma\sqrt{n}}=\frac{\dfrac{X_1+\cdots+X_n}{n}-\mu}{\dfrac{\sigma}{\sqrt{n}}} \Rightarrow \mathrm{E}\,S_n=0, \mathrm{D}\,S_n=1$$

Khi đó dãy biến ngẫu nhiên S_n hội tụ theo phân bố về phân bố chuẩn tắc $\mathrm{N}(0;1)$, nghĩa là:

Với mọi $x \in \mathbb{R}$, $\lim\limits_{n\to\infty} P\{S_n < x\} = \Phi(x)$

Áp dụng định lý giới hạn trung tâm cho dãy các biến ngẫu nhiên độc lập X_1, X_2, ... có cùng phân bố Bernoulli tham số p ta được định lý Moivre –Laplace

Giả sử X_1, X_2, $\ldots$ là dãy các biến ngẫu nhiên độc lập có cùng phân bố Bernoulli tham số p, khi đó

Với mọi $x \in \mathrm{R}$:

$$\lim_{n\to\infty} P\left\{\frac{X_1+\cdots+X_n-np}{\sqrt{npq}}<x\right\}=\Phi(x)$$

4.4. XẤP XỈ PHÂN BỐ NHỊ THỨC

4.4.1. Xấp xỉ phân bố nhị thức bằng phân bố chuẩn

Giả sử $X_1, X_2, \dots, X_n$ là dãy các biến ngẫu nhiên độc lập có cùng phân bố Bernoulli tham số p, khi đó

$$U_n = X_1 + X_2 + \cdots + X_n \sim \mathcal{B}(n, p)$$

Mặc dù ta đã biết công thức tính xác suất

$$P\{U_n = k\} = \frac{n!}{k!(n-k)!} p^k q^{n-k}$$

Tuy nhiên khi n khá lớn ta không thể áp dụng công thức này để tính mà cần đến công thức xấp xỉ

Định lý giới hạn địa phương

Giả sử X là biến ngẫu nhiên có phân bố nhị thức **B**$(n; p)$, khi đó

$$P\{X = k\} = \varphi\left(\frac{k-np}{\sqrt{npq}}\right)\frac{1}{\sqrt{nqp}}\left(1+\varepsilon_{n,k}\right)$$

trong đó $\left|\varepsilon_{n,k}\right| < \frac{C}{\sqrt{n}}$ với C là hằng số

Do đó khi n đủ lớn ta có thể xấp xỉ

$$P\{X = k\} \approx \frac{1}{\sqrt{nqp}}\varphi\left(\frac{k-np}{\sqrt{npq}}\right) = \frac{1}{\sqrt{nqp}}\frac{1}{\sqrt{2\pi}}e^{-\frac{(k-np)^2}{2npq}}$$

Áp dụng định lý Moivre-Laplace ta có công thức xấp xỉ giá trị của hàm phân bố xác suất nhị thức

$$P\left\{U_n < x\right\} = P\left\{\frac{U_n - np}{\sqrt{npq}} < \frac{x - np}{\sqrt{npq}}\right\} \approx \Phi\left(\frac{x - np}{\sqrt{npq}}\right)$$

$$P\left\{a \le U_n \le b\right\} \approx \Phi\left(\frac{b - np}{\sqrt{npq}}\right) - \Phi\left(\frac{a - np}{\sqrt{npq}}\right)$$

Người ta thấy rằng xấp xỉ là tốt khi np và nq lớn hơn 5 hoặc khi npq lớn hơn 20

Khi $a = b = k$, với $0 \le k \le n$, vế trái của công thức trên sẽ là $P\{U_n = k\} \neq 0$, trong khi đó vế phải bằng 0.

Điều này xảy ra vì ta đã dùng hàm phân bố liên tục để xấp xỉ phân bố rời rạc, vì vậy để xấp xỉ tốt hơn người ta thường sử dụng công thức có dạng sau

$$P\{U_n \le x\} \approx \Phi\left(\frac{x + 1/2 - np}{\sqrt{npq}}\right)$$

$$P\{a \le U_n \le b\} \approx \Phi\left(\frac{b + 1/2 - np}{\sqrt{npq}}\right) - \Phi\left(\frac{a - 1/2 - np}{\sqrt{npq}}\right)$$

Ví dụ: Giả sử U_n là một biến ngẫu nhiên có phân bố nhị thức với tham số $n = 36$ và $p = 0,5$

Ta có các kết quả tính xác suất $P\{U_n \le 21\}$

$$P\{U_n \le 21\} = \sum_{k=0}^{21} \frac{36!}{k!(36-k)!}(0,5)^{36} = 0,8785$$

$$P\{U_n \le 21\} \approx \Phi\left(\frac{21-np}{\sqrt{npq}}\right) = \Phi\left(\frac{21-18}{3}\right) = \Phi(1) = 0,8413$$

$$P\{U_n \le 21\} \approx \Phi\left(\frac{21,5-np}{\sqrt{npq}}\right) = \Phi\left(\frac{21,5-18}{3}\right) = \Phi(1,17) = 0,879$$

4.4.2. Xấp xỉ phân bố nhị thức bằng phân bố Poisson

Khi điều kiện xấp xỉ phân bố nhị thức bằng phân bố chuẩn không thỏa mãn (np hoặc nq nhỏ hơn 5, và npq nhỏ hơn 20), ta có thể xấp xỉ phân bố nhị thức bằng phân bố Poisson

Khi $n > 50$ và $p < 0,1$ người ta có thể xấp xỉ phân bố nhị thức **B**$(n;p)$ với phân bố Poisson **P**(np) tham số $\lambda = np$

$$P\{U_n = k\} \approx e^{-np}\frac{(np)^k}{k!}$$