CHƯƠNG IV: LUẬT SỐ LỚN VÀ ĐỊNH LÝ GIỚI HẠN

Trong chương này ta nghiên cứu luật số lớn đó là sự hội tụ theo xác suất của dãy các biến ngẫu nhiên và định lý giới hạn trung tâm: khảo sát sự hội tụ theo phân bố xác suất cúa dãy các biến ngẫu nhiên.

4.1 SỰ HỘI TỤ CỦA DÃY CÁC BIẾN NGẪU NHIÊN

4.1.1 Hội tụ theo xác suất

Định nghĩa 4.1: *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*:

$$\forall \varepsilon > 0 \quad \lim_{n\to\infty} P\{|X_n - X| > \varepsilon\} = 0. \tag{4.1}$$

Như vậy dãy các biến ngẫu nhiên $X_1, X_2, ...$ hội tụ theo xác suất về biến ngẫu nhiên X thì với n đủ lớn, thực tế gần như chắc chắn ta có thể coi rằng X_n không khác mấy so với X .

4.1.2 Hội tụ theo phân bố

Định nghĩa 4.2: D*ãy 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* $x \in \mathrm{R}$:

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x). \tag{4.2}$$

Trường hợp dãy các biến ngẫu nhiên rời rạc $\{X_n\}_{n=1}^{\infty}$ và biến ngẫu nhiên rời rạc X có cùng miền giá trị $R = \{c_1, c_2, ...\}$ thì dãy $\{X_n\}_{n=1}^{\infty}$ hội tụ theo phân bố về biến ngẫu nhiên X khi và chỉ khi với mọi $c_k \in R$:

$$\lim_{n\to\infty} P\{X_n = c_k\} = P\{X = c_k\}. \tag{4.3}$$

4.2 LUẬT SỐ LỚN

4.2.1 Bất đẳng thức Markov

Định lý 4.1: Cho Y là biến ngẫu nhiên không âm có kỳ vọng hữu hạn. Khi đó với mọi $a > 0$ ta có bất đẳng thức:

$$P\{Y \ge a\} \le \frac{\mathrm{E}Y}{a}. \tag{4.4}$$

Chứng minh:

a) Trường hợp Y rời rạc có tập giá trị $R_Y = \{y_1, y_2, ...\}$.

Đặt $R_1 = \{y_i \in R_Y, y_i < a\}$; $R_2 = \{y_i \in R_Y, y_i \ge a\}$.

$$\mathrm{E}Y = \sum_{y_i \in R_Y} y_i P\{Y = y_i\} \ (*)$$

- Trường hợp tổng (*) có hữu hạn các số hạng thì đương nhiên có thể thay đổi thứ tự lấy tổng
- Trường hợp tổng (*) có vô hạn số hạng thì đây là tổng của một chuỗi số dương do đó cũng có thể thay đổi thứ tự, vì vậy có thể viết lại:

$$\mathrm{E}Y = \sum_{y_i \in R_Y} y_i P\{Y = y_i\} = \sum_{y_i \in R_1} y_i P\{Y = y_i\} + \sum_{y_i \in R_2} y_i P\{Y = y_i\}$$

$$\ge \sum_{y_i \in R_2} y_i P\{Y = y_i\} \ge a \sum_{y_i \in R_2} P\{Y = y_i\} = aP\{Y \ge a\}.$$

Suy ra $\quad P\{Y \ge a\} \le \dfrac{\mathrm{E}Y}{a}$.

b) Giả sử Y liên tục có hàm mật độ xác suất $f_Y(y)$, ta có

$$\mathrm{E}Y = \int_0^{+\infty} y f_Y(y)dy = \int_0^{a} y f_Y(y)dy + \int_a^{+\infty} y f_Y(y)dy \ge \int_a^{+\infty} y f_Y(y)dy \ge a \int_a^{+\infty} f_Y(y)dy = aP\{Y \ge a\}.$$

Suy ra $\quad P\{Y \ge a\} \le \dfrac{\mathrm{E}Y}{a}$.

4.2.2 Bất đẳng thức Trêbưsép

Định lý 4.2: Giả sử X là biến ngẫu nhiên có kỳ vọng và phương sai hữu hạn, khi đó với mọi $\varepsilon > 0$ ta có:

$$P\{|X - \mathrm{E}X| \ge \varepsilon\} \le \frac{\mathrm{D}X}{\varepsilon^2}. \tag{4.5}$$

$$P\{|X - \mathrm{E}X| < \varepsilon\} \ge 1 - \frac{\mathrm{D}X}{\varepsilon^2}. \tag{4.6}$$

Chứng minh: Áp dụng công thức (4.4) cho biến ngẫu nhiên $Y = (X - \mathrm{E}X)^2$ và $a = \varepsilon^2$ ta có:

$$P\{|X - \mathrm{E}X| \ge \varepsilon\} = P\{Y \ge \varepsilon^2\} \le \frac{\mathrm{E}Y}{\varepsilon^2} = \frac{\mathrm{E}(X - \mathrm{E}X)^2}{\varepsilon^2} = \frac{\mathrm{D}X}{\varepsilon^2}.$$

Từ (4.5), áp dụng quy tắc xác suất biến cố đối ta được bất đẳng thức (4.6).

Bất đẳng thức (4.5)-(4.6) được gọi là bất đẳng thức Trêbưsép.

Bất đẳng thức Trêbưsép có nhiều ứng dụng. Trước hết nó cho phép ta đánh giá cận trên hoặc cận dưới của xác suất để biến ngẫu nhiên X nhận giá trị sai lệch so với kỳ vọng $\mathrm{E}\,X$ không quá ε. Bất đẳng thức Trêbưsép có ý nghĩa to lớn về mặt lý thuyết, nó được sử dụng để chứng minh các định lý của luật số lớn.

Ví dụ 4.1: Một cửa hàng muốn ước lượng nhanh chóng sai số của số vải bán ra trong một tháng của mình. Số vải của mỗi khách hàng được làm tròn bởi số nguyên gần nhất (ví dụ trong sổ ghi 195,6 *m* thì làm tròn là 196*m*). Ký hiệu X_i là sai số giữa số mét vải thực bán và số mét vải đã làm tròn của khách hàng thứ *i*.

Giải: Các sai số $X_1, X_2, ..., X_n$ là các biến ngẫu nhiên độc lập có phân bố đều trên đoạn $[-0,5;0,5]$. Khi đó $\mathrm{E}X_i = 0, \mathrm{D}X_i = \frac{1}{12}$.

Sai số tổng cộng trong cả tháng là $S = X_1 + \cdots + X_n$ (trong đó n là số khách hàng mua hàng trong tháng). Ta có:

$$\mathrm{E}S = \sum_{i=1}^{n} \mathrm{E}X_i = 0,\ \mathrm{D}S = \sum_{i=1}^{n} \mathrm{D}X_i = \frac{n}{12}.$$

Theo bất đẳng thức Trêbưsép, xác suất để sai số vượt quá ε mét sẽ được đánh giá bởi:

$$P\{|S| \ge \varepsilon\} \le \frac{\mathrm{D}S}{\varepsilon^2} = \frac{n}{12\varepsilon^2}.$$

Giả sử có $n = 10^4$ khách hàng trong tháng. Để xác suất $P\{|S| > \varepsilon\}$ bé hơn 0,01 ta phải có

$$\frac{n}{12\varepsilon^2} \le 0,01 \text{ hay } \varepsilon \ge \sqrt{\frac{n}{12 \cdot 0,01}} = 288,67.$$

Vậy ta có thể kết luận: Với xác suất 0,99 sai số giữa số vải thực bán với số vải đã tính tròn không vượt quá 289 *m*, nếu số khách hàng là 1 vạn.

4.2.3 Luật số lớn Trêbưsép

Định lý 4.3: Giả sử $X_1, X_2, ...$ là dãy các biến ngẫu nhiên độc lập, có các kỳ vọng hữu hạn và tất cả các phương sai bị chặn trên bởi hằng số C ($\mathrm{D}\,X_i \le C;\ \forall i = 1,2,...$). Khi đó

$$\lim_{n\to\infty} P\left(\left|\frac{X_1 + \cdots + X_n}{n} - \frac{\mathrm{E}\,X_1 + \cdots + \mathrm{E}\,X_n}{n}\right| > \varepsilon\right) = 0 \qquad \textbf{(4.7)}$$

Chứng minh: Xét biến ngẫu nhiên $S_n = \frac{X_1 + \cdots + X_n}{n}$. Từ giả thiết độc lập của dãy các biến ngẫu nhiên $X_1, X_2, ...$ ta suy ra $\mathrm{E}\,S_n = \frac{\mathrm{E}\,X_1 + \cdots + \mathrm{E}\,X_n}{n}$; $\mathrm{D}\,S_n = \frac{\mathrm{D}\,X_1 + \cdots + \mathrm{D}\,X_n}{n^2} \le \frac{C}{n}$.

Áp dụng bất đẳng thức Trêbưsép (4.5) cho biến ngẫu nhiên S_n ta có:

$$P\left(\left|\frac{X_1+\cdots+X_n}{n}-\frac{\mathrm{E}\,X_1+\cdots+\mathrm{E}\,X_n}{n}\right|>\varepsilon\right)\le\frac{C}{n\varepsilon^2}\xrightarrow[n\to\infty]{}0.$$

Hệ quả 4.1: Giả sử $X_1, X_2,...$ là dãy các biến ngẫu nhiên độc lập có cùng có kỳ vọng μ và tất cả các phương sai bị chặn trên bởi hằng số C ($\mathrm{D}\,X_i \le C;\ \forall i=1,2,...$). Khi đó

$$\frac{X_1+\cdots+X_n}{n}\xrightarrow[n\to\infty]{P}\mu \tag{4.8}$$

Hệ quả 4.2: Giả sử $X_1, X_2,...$ là dãy các biến ngẫu nhiên độc lập có cùng phân bố, có kỳ vọng μ và phương sai σ^2. Khi đó

$$\frac{X_1+\cdots+X_n}{n}\xrightarrow[n\to\infty]{P}\mu \tag{4.9}$$

Định lý Trêbưsép chứng tỏ rằng trung bình số học của các biến ngẫu nhiên độc lập hội tụ theo xác suất về trung bình số học của kỳ vọng tương ứng của nó. Nói cách khác có sự ổn định của trung bình số học của một số lớn các biến ngẫu nhiên xung quanh trung bình số học của các kỳ vọng của các biến ngẫu nhiên ấy. Như vậy mặc dù từng biến ngẫu nhiên độc lập có thể nhận giá trị khác nhiều so với kỳ vọng của chúng, song trung bình số học của một số lớn các biến ngẫu nhiên lại nhận giá trị gần bằng trung bình số học các kỳ vọng của chúng với xác suất rất lớn. Điều đó cho phép dự đoán giá trị trung bình số học của các biến ngẫu nhiên. Chẳng hạn, gieo một con xúc xắc cân đối. Giả sử X là số nốt xuất hiện ở mặt trên con xúc xắc. Ta có $\mathrm{E}\,X=3,5$. Một nhà thống kê đã gieo một con xúc xắc cân đối 1 triệu lần (nhờ sự trợ giúp của máy vi tính) và ghi lại số nốt xuất hiện ở mặt trên con xúc xắc. Số trung bình của 1 triệu lần gieo được tìm thấy là

$$\frac{x_1+\cdots+x_{10^6}}{10^6}\approx 3,500867\approx 3,5.$$

Định lý Trêbưsép có ứng dụng rộng rãi trong nhiều lĩnh vực, chẳng hạn nó chính là cơ sở cho phương pháp đo lường trong vật lý. Để xác định giá trị của một đại lượng vật lý nào đó người ta thường tiến hành đo n lần độc lập và lấy trung bình số học của các kết quả đo làm giá trị thực của đại lượng cần đo. Thật vậy, giả sử xem kết quả của n lần đo là các biến ngẫu nhiên $X_1, X_2,..., X_n$. Ta thấy rằng các biến ngẫu nhiên này độc lập, có cùng kỳ vọng bằng chính giá trị thực của đại lượng vật lý (giả sử không có sai số hệ thống), các phương sai của chúng đều bị chặn trên bởi bình phương của độ chính xác của thiết bị đo. Do đó theo định lý Trêbưsép ta có thể cho rằng trung bình số học của các kết quả đo sẽ sai lệch rất ít so với giá trị thực của đại lượng vật lý với xác suất gần như bằng một.

Định lý Trêbưsép còn là cơ sở cho lý thuyết mẫu, ứng dụng trong thống kê.

4.2.4 Luật số lớn Bernoulli

Xét phép thử ngẫu nhiên $\mathcal{C}$ và A là một biến cố liên quan đến phép thử $\mathcal{C}$. Thực hiện n lần phép thử $\mathcal{C}$ một cách độc lập 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 4.4: Tần suất f_n hội tụ theo xác suất về xác suất p của biến cố A, nghĩa là với mọi $\varepsilon > 0$

$$\lim_{n\to\infty} P\{|f_n - p| < \varepsilon\} = 1 \qquad \textbf{(4.10)}$$

Chứng minh: Xét dãy các biến ngẫu nhiên $X_1, X_2, ..., X_n$ xác định như sau:

$$X_k = \begin{cases} 1 & \text{nÕu } A \text{ x¶y ra ë phÐp thö thø } k \\ 0 & \text{nÕu } A \text{ kh«ng x¶y ra ë phÐp thö thø } k \end{cases}$$

Như vậy dãy các biến ngẫu nhiên $X_1, X_2, ..., X_n$ độc lập có cùng phân bố Bernoulli tham số p (công thức (2.9)). $\mathrm{E}X_k = p,\ \mathrm{D}X_k = p(1-p)$.

Ta có $\dfrac{X_1 + \cdots + X_n}{n} = \dfrac{k_n}{n} = f_n$.

Vậy theo hệ quả 4.2 của định lý 4.3 suy ra f_n hội tụ theo xác suất về p .

Định lý Bernoulli chỉ ra rằng tần suất xuất hiện của biến cố trong n phép thử độc lập sẽ hội tụ theo xác suất về xác suất của biến cố đó khi số lần thử tăng lên vô hạn. Chính vì vậy định lý Bernoulli là cơ sở lý thuyết của định nghĩa thống kê về xác suất.

Ở thế kỷ 18, nhà toán học Pháp Buffon gieo một đồng tiền 4040 lần và ghi được 2048 lần xuất hiện mặt ngửa, tần suất là 0,507. Một nhà thống kê người Anh gieo đồng tiền 12000 lần và thu được 6019 lần xuất hiện mặt ngửa, tần suất tương ứng 0,5016. Trong một thí nghiệm khác, ông ta gieo 24000 lần và thu được 12012 lần xuất hiện mặt ngửa, tần suất tương ứng là 0,5005. Như vây ta thấy rằng khi số phép thử tăng lên thì tần suất tương ứng sẽ càng gần 0,5.

Ví dụ 4.2: Giả sử p là tỉ lệ cử tri sẽ bầu cho ứng cử viên A. Để ước lượng trước tỉ lệ này người ta phỏng vấn ngẫu nhiên n cử tri. Có thể coi kết quả bầu của cử tri thứ i là biến ngẫu nhiên X_i có phân bố Bernoulli tham số p (X_i nhận giá trị 1 nếu cử tri thứ i bầu A và nhận giá trị 0 trong trường hợp ngược lại). Các biến ngẫu nhiên X_i, $i = 1,...,n$ độc lập và có cùng phân bố Bernoulli tham số p với kỳ vọng $\mathrm{E}\,X_i = p$ phương sai $\mathrm{D}\,X_i = p(1-p)$.

Tần số $\qquad k_n = X_1 + \cdots + X_n$

Áp dụng bất đẳng thức Trêbưsép ta có

$$P\{|k_n - p| \ge \varepsilon\} \le \frac{p(1-p)}{n\varepsilon^2}.$$

Tham số p chưa biết nhưng theo bất đẳng thức Cauchy ta có $p(1-p) \le 1/4$. Vậy

$$P\left\{\left|k_n - p\right| \geq \varepsilon\right\} \leq \frac{1}{4n\varepsilon^2} \text{ cũng có nghĩa là } P\left\{\left|k_n - p\right| < \varepsilon\right\} \geq 1 - \frac{1}{4n\varepsilon^2}$$

Chẳng hạn, nếu $\varepsilon = 0,1$ và $n = 100$ thì $P\left\{\left|k_{100} - p\right| \geq 0,1\right\} \leq \dfrac{1}{4 \cdot 100 \cdot (0,1)^2} = 0,25$

Nói cách khác, nếu phỏng vấn 100 cử tri và lấy kết quả này để ước lượng cho tỉ lệ cử tri sẽ bỏ phiếu cho ứng cử viên A thì sai số vượt quá 0,1 có xác suất nhỏ hơn 0,25.

Nếu muốn ước lượng tin cậy hơn (chẳng hạn xác suất lớn hơn 95%) và chính xác hơn (với sai số 0,01) thì

$$P\left\{\left|k_n - p\right| \geq 0,01\right\} \leq \frac{1}{4n(0,01)^2} \text{ cũng có nghĩa là } P\left\{\left|k_n - p\right| < 0,01\right\} \geq 1 - \frac{1}{4n(0,01)^2}$$

Vậy số cử tri phải phỏng vấn thỏa mãn

$$1 - \frac{1}{4n(0,01)^2} \geq 0,95 \Rightarrow \frac{1}{4n(0,01)^2} \leq 1 - 0,95 = 0,05 \Rightarrow n \geq 50.000 .$$

Như vậy để ước lượng với độ tin cậy cao và độ chính xác cao thì cần phải lấy mẫu với số lượng lớn. Tuy nhiên ở đây ta chỉ dựa vào bất đẳng thức Trêbusép để giải quyết bài toán, trong chương 6 ta sẽ nghiên cứu về bài toán ước lượng này và bằng phương pháp khác ta sẽ chỉ ra số cử tri phải phỏng vấn nhỏ hơn kết quả trên.

4.3 ĐỊNH LÝ GIỚI HẠN TRUNG TÂM

Giả sử $X_1, X_2, ...$ là dãy các biến ngẫu nhiên độc lập có cùng phân bố, có kỳ vọng μ và phương sai σ^2. Theo công thức (2.61), (2.71) ta có

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

Do đó

$$S_n = \frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} = \frac{X_1 + \cdots + X_n - \mathrm{E}\left[X_1 + \cdots + X_n\right]}{\sqrt{\mathrm{D}\left[X_1 + \cdots + X_n\right]}} \text{ có } \mathrm{E}\,S_n = 0 \text{ và } \mathrm{D}\,S_n = 1.$$

Định lý 4.5: Giả sử $X_1, X_2, ...$ là dãy các biến ngẫu nhiên độc lập có cùng phân bố, có kỳ vọng μ và phương sai σ^2. Khi đó dãy biến ngẫu nhiên $S_n = \dfrac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}$ hội tụ theo phân bố về phân bố chuẩn tắc $\mathbf{N}(0;1)$, tức là:

$$\text{Với mọi } x \in \mathrm{R},\ \lim_{n \to \infty} F_{S_n}(x) = \Phi(x) \qquad \textbf{(4.11)}$$

$\Phi(x)$ là hàm phân bố xác suất của phân bố chuẩn tắc $\mathbf{N}(0;1)$.

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 (công thức (2.9)) ta được định lý Moivre –Laplace:

Định lý 4.6 (Moivre –Laplace): Đối với dãy các biến ngẫu nhiên $X_1, X_2,...$ độc lập có cùng phân bố Bernoulli tham số p thì:

$$\text{Với mọi } x \in \mathrm{R}, \lim_{n\to\infty} P\left\{\frac{X_1+\cdots+X_n-np}{\sqrt{npq}} \le x\right\} = \Phi(x). \quad \textbf{(4.12)}$$

4.4 XẤP XỈ PHÂN BỐ NHỊ THỨC

4.4.1 Xấp xỉ phân bố nhị thức bằng phân bố chuẩn

Giả sử $X_1, X_2,..., X_n$ độc lập có cùng phân bố Bernoulli tham số p. Theo công thức (2.17) và (2.20) ta có $U_n = X_1 + X_2 + \cdots + X_n \sim \mathcal{B}(n, p)$. Công thức (2.21) cho phép tính xác suất $P\{U_n = k\} = C_n^k p^k q^{n-k} = \frac{n!}{k!(n-k)!} p^k q^{n-k}$. Tuy nhiên khi n khá lớn ta không thể áp dụng công thức này để tính mà cần đến công thức xấp xỉ (vì sẽ tràn bộ nhớ khi tính toán có sử dụng máy tính).

Định lý 4.7 (*định lý giới hạn địa phương*): Giả sử X là biến ngẫu nhiên có phân bố nhị thức $\mathcal{B}(n, p)$. Đặt $x_k = \frac{k-np}{\sqrt{npq}}$, khi đó

$$P_n(k\,; p) = P\{X = k\} = \varphi\left(\frac{k-np}{\sqrt{npq}}\right)\frac{1}{\sqrt{nqp}}\left(1+\varepsilon_{n,k}\right). \quad \textbf{(4.13)}$$

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

Vì vậy khi n đủ lớn ta có thể xấp xỉ :

$$P\{X = k\} \approx \varphi\left(\frac{k-np}{\sqrt{npq}}\right)\frac{1}{\sqrt{nqp}} = \frac{1}{\sqrt{nqp}}\frac{1}{\sqrt{2\pi}} e^{\frac{-(k-np)^2}{2npq}}. \quad \textbf{(4.14)}$$

Để tính xấp xỉ giá trị của hàm phân bố xác suất nhị thức ta có thể áp dụng định lý Moivre-Laplace - công thức (4.12) như sau

$$P\{U_n \le x\} = P\left\{\frac{U_n - np}{\sqrt{npq}} \le \frac{x-np}{\sqrt{npq}}\right\} \approx \Phi\left(\frac{x-np}{\sqrt{npq}}\right) \quad \textbf{(4.15)}$$

$$P\{a \le U_n \le b\} \approx \Phi\left(\frac{b-np}{\sqrt{npq}}\right) - \Phi\left(\frac{a-np}{\sqrt{npq}}\right) \quad \textbf{(4.16)}$$

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

Khi $a = b = k$, $0 \le k \le n$, vế trái của công thức (4.16) sẽ là $P\{U_n = k\} \ne 0$, 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

$$P\{U_n \le x\} \approx \Phi\left(\frac{x+0,5-np}{\sqrt{npq}}\right); P\{a \le U_n \le b\} \approx \Phi\left(\frac{b+0,5-np}{\sqrt{npq}}\right) - \Phi\left(\frac{a-0,5-np}{\sqrt{npq}}\right) \quad \textbf{(4.17)}$$

Ví dụ 4.3: Giả sử U_n là một biến ngẫu nhiên có phân bố nhị thức với tham số $n = 36$ và $p = 0,5$. Xác suất U_n nhận giá trị ≤ 21 được tính chính xác theo công thức là

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

Áp dụng công thức (4.15) ta được

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

Áp dụng công thức (4.17) ta được

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

Như vậy sử dụng công thức 4.17 có kết quả gần với giá trị chính xác.

$$P\{U_n = 19\} \approx \Phi\left(\frac{19,5-18}{3}\right) - \Phi\left(\frac{18,5-18}{3}\right) = 0,6915 - 0,5675 = 0,124.$$

Giá trị này cũng khá gần với giá trị chính xác

$$\frac{36!}{19!(36-19)!}(0,5)^{36} = 0,1251.$$

Ví dụ 4.4: Gieo 3200 lần một đồng xu cân đối và đồng chất. Gọi X là số lần xuất hiện mặt sấp trong 3200 lần gieo đó.

a) Tìm số lần xuất hiện mặt sấp có khả năng nhất. Tính xác suất tương ứng.

b) Tính xác suất $P\{1610 \le X \le 1650\}$.

Giải: a) Ta có: $n = 3200$, $p = 0,5$ $\Rightarrow$ $(n+1)p = 1600,5$. Vậy số lần xuất hiện mặt sấp có khả năng nhất là 1600 với xác suất tương ứng

$$P_{3200}(1600; 0,5) = \frac{3200!}{160011600!} 0,5^{3200} \text{ (không thể tính được vì tràn bộ nhớ).}$$

Mặt khác nếu tính gần đúng ta có: $x_m = \dfrac{1600 - 3200 \cdot 0,5}{\sqrt{3200 \cdot 0,5 \cdot 0,5}} = 0$. Do đó

$$P_{3200}(1600; 0,5) \approx \frac{1}{\sqrt{3200 \cdot 0,5 \cdot 0,5}} \varphi(0) = \frac{1}{40\sqrt{\pi}} \approx 0,014.$$

b) $P\{10 \le X - 1600 \le 50\} = P\left\{\dfrac{10}{20\sqrt{2}} \le \dfrac{X-1600}{20\sqrt{2}} \le \dfrac{50}{20\sqrt{2}}\right\} \approx \Phi(1,7678) - \Phi(0,3536)$

Tra bảng ta có $\Phi(1,7678) - \Phi(0,3536) \approx 0,9616 - 0,6406 = 0,321$.

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 (điều kiện: np và nq lớn hơn 5 hoặc khi $npq > 20$ không thỏa mãn), ta có thể xấp xỉ phân bố nhị thức bằng phân bố Poisson.

Định lý 4.8: Cho $X_1, X_2, ...$ là dãy các biến ngẫu nhiên có cùng phân bố nhị thức, ở đó với mỗi n, X_n có phân bố nhị thức $\mathcal{B}(n, p_n)$. Nếu tồn tại giới hạn $\lim\limits_{n\to\infty} np_n = \lambda > 0$ thì X_n hội tụ theo phân bố về biến ngẫu nhiên X có phân bố Poisson tham số λ .

Trong thực tế khi $n > 50$ và $p < 0{,}1$ người ta có thể xấp xỉ $\mathcal{B}(n; p)$ với phân bố Poisson $\mathcal{P}(np)$ tham số $\lambda = np$:

$$P\{X_n = k\} \approx e^{-np} \frac{(np)^k}{k!} . \qquad \textbf{(4.18)}$$

Ví dụ 4.5: Giả sử xác suất để làm ra mỗi đinh ốc không đúng quy cách là $p = 0{,}015$. Người ta xếp đinh ốc vào từng hộp, mỗi hộp 100 chiếc.

a) Tính tỉ lệ hộp chứa toàn đinh ốc đúng quy cách.

b) Cần phải xếp ít nhất bao nhiêu đinh ốc trong mỗi hộp để tỉ lệ hộp chứa 100 đinh ốc tốt tối thiểu là 80%.

Giải: a) Nếu gọi X là số đinh ốc không đúng quy cách trong hộp chứa 100 đinh ốc thì $X \sim \mathcal{B}(n; p)$ với $n = 100,\ p = 0{,}015$.

Tính gần đúng: $P\{X = 0\} \approx e^{-np} \dfrac{(np)^0}{0!} = e^{-1{,}5} = 0{,}2231$

Vậy có khoảng 22,3% số hộp chứa 100 đinh ốc tốt.

b) Giả sử mỗi hộp chứa $100 + k$ đinh ốc, k là số tự nhiên. Gọi X là số đinh ốc không đúng quy cách trong mỗi hộp chứa $100 + k$ đinh ốc. $X \sim \mathcal{B}(n; p)$ với $n = 100 + k,\ p = 0{,}015$. Ta phải xác định k nhỏ nhất để

$$P\{X \le k\} = \sum_{i=0}^{k} C_n^i (0{,}015)^i (0{,}985)^{n-i} \ge 0{,}8 .$$

Dùng công thức xấp xỉ $P\{X = i\} = C_n^i (0{,}015)^i (0{,}985)^{n-i} \approx e^{-\lambda} \dfrac{\lambda^i}{i!}$

ở đây $\lambda = np = (100 + k)(0{,}015) \approx 1{,}5 + 0{,}015k \approx 1{,}5$ (vì $0{,}015k$ nhỏ).

Vậy cần tìm k nhỏ nhất để

$$e^{-1{,}5}\left\{1 + \frac{1{,}5}{1!} + \frac{1{,}5^2}{2!} + \cdots + \frac{1{,}5^k}{k!}\right\} \ge 0{,}8 \quad \Leftrightarrow \quad 1 + \frac{1{,}5}{1!} + \frac{1{,}5^2}{2!} + \cdots + \frac{1{,}5^k}{k!} \ge 0{,}8e^{1{,}5} = 3{,}5853$$

Thử với $k = 1, 2, ...$, ta thấy $k = 2$ bất đẳng thức trên được thoả mãn. Như vậy dùng xấp xỉ

Poisson ta có thể kết luận mỗi hộp cần đóng 102 chiếc đinh ốc. Khi đó xác suất để có ít nhất 100 đinh ốc tốt trong hộp 102 chiếc là 0,8022.

TÓM TẮT

Luật số lớn Trêbưsép xét sự hội tụ theo xác suất của trung bình cọng của dãy các biến ngẫu nhiên độc lập có kỳ vọng và phương sai hữu hạn dựa vào bất đẳng thức Trêbưsép. Khi các biến ngẫu nhiên của dãy cùng có phân bố Bernoulli thì ta có luật số lớn Bernoulli.

Định lý giới hạn trung tâm xét sự hội tụ theo phân bố của tổng các số hạng của dãy các biến ngẫu nhiên về phân bố chuẩn tắc. Trường hợp các biến ngẫu nhiên của dãy cùng có phân bố Bernoulli thì ta có định lý Moivre-Laplace.

Áp dụng định lý giới hạn trung tâm ta có thể tính gần đúng xác suất của phân bố nhị thức khi tham số n khá lớn, công thức (4.13) - (4.17). Khi tham số n lớn nhưng p bé thì xấp xỉ phân bố nhị thức $\mathcal{B}(n;p)$ với phân bố Poisson $\mathcal{P}(np)$ tham số $\lambda = np$, công thức (4.18).

CÂU HỎI ÔN TẬP VÀ BÀI TẬP

4.1 Luật số lớn kết luận về sự hội tụ theo xác suất của trung bình cộng các biến ngẫu nhiên độc lập về trung bình cộng của kỳ vọng của chúng nếu các phương sai của các biến ngẫu nhiên này bị chặn.

Đúng ☐ Sai ☐.

4.2 Giả sử $\{X_n\}$ là dãy các biến ngẫu nhiên có kỳ vọng bằng nhau và phương sai dần tới 0, khi đó dãy sẽ hội tụ theo xác suất đến kỳ vọng chung của dãy biến ngẫu nhiên trên.

Đúng ☐ Sai ☐.

4.3 Bất đẳng thức Trêbưsép chỉ đúng đối với các biến ngẫu nhiên rời rạc.

Đúng ☐ Sai ☐.

4.4 Bất đẳng thức Trêbưsép chỉ đúng đối với các biến ngẫu nhiên nhận giá trị dương.

Đúng ☐ Sai ☐.

4.5 Luật số lớn Bernoulli là một trường hợp đặc biệt của luật số lớn Trêbưsép khi dãy các biến ngẫu nhiên được có cùng phân bố Bernoulli A(p) .

Đúng ☐ Sai ☐.

4.6 Luật số lớn Bernoulli là cơ sở lý thuyết của định nghĩa thống kê về xác suất.

Đúng ☐ Sai ☐.

4.7 Tổng của dãy các biến ngẫu nhiên độc lập có cùng phân bố với kỳ vọng và phương sai hữu hạn tiệm cận phân bố chuẩn.

Đúng ☐ Sai ☐.

4.8 Luật số lớn xét sự hội tụ theo xác suất còn định lý giới hạn trung tâm xét sự hội tụ theo phân bố của dãy các biến ngẫu nhiên.

Đúng ☐ Sai ☐.

4.9 Có 10 máy hoạt động độc lập với nhau. Xác suất để trong ca làm việc mỗi máy bị hỏng là 0,05. Dựa vào bất đẳng thức Trêbưsép hãy đánh giá xác suất của sự sai lệch giữa số máy hỏng và số máy hỏng trung bình.

a) Nhỏ hơn 2.

b) Lớn hơn 2

4.10 Cho $X_1, X_2, ..., X_{12}$ là các biến ngẫu nhiên độc lập với $\mathrm{E}X_i = 16$, $\mathrm{D}X_i = 1$ $(i = \overline{1,12})$. Sử dụng bất đẳng thức Trêbưsép để tìm hai hằng số a, b sao cho

$$P\left\{a \le \sum_{i=1}^{12} X_i \le b\right\} \ge 0{,}99.$$

4.11 Cho $X_1, X_2, ..., X_{10000}$ là các biến ngẫu nhiên độc lập có phân bố đều trong đoạn $\left[-\frac{1}{2}, \frac{1}{2}\right]$.

Chứng minh rằng $P\left\{\left|\sum_{i=1}^{10000} X_i\right| \ge 500\right\} \le \frac{1}{300}$.

4.12 Xét biến ngẫu nhiên liên tục X có hàm mật độ $f_X(x)$ và giả sử $X_1, X_2, ..., X_n$ là các biến ngẫu nhiên độc lập có cùng phân bố với X. Ta gọi $(X_1, X_2, ..., X_n)$ là một mẫu ngẫu nhiên kích thước n của X. Trung bình mẫu được định nghĩa là $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Ký hiệu $\mathrm{E}\,X = \mu$, $\mathrm{D}\,X = \sigma^2$. Cần phải lấy mẫu kích thước bao nhiêu để $P\left\{\left|\overline{X} - \mu\right| \le \frac{\sigma}{10}\right\} \ge 0{,}95$.

4.13 Gieo một con xúc xắc cân đối n lần một cách độc lập. Gọi S là số lần xuất hiện mặt lục. Chứng minh rằng $P\left\{\frac{n}{6} - \sqrt{n} < S < \frac{n}{6} + \sqrt{n}\right\} \ge \frac{31}{36}$.

4.14 Giả sử tiền điện của một gia đình phải trả trong 1 tháng là một biến ngẫu nhiên với trung bình 16USD và độ lệch tiêu chuẩn 1USD. Sử dụng bất đẳng thức Trêbưsép, hãy xác định số M nhỏ nhất để với xác suất 0,99 số tiền điện phải trả trong 1 năm (12 tháng) không vượt quá M.

4.15 Cho dãy các biến ngẫu nhiên độc lập $\{X_n\}$ xác định như sau:

X_n	-2^n	0	2^n
P	$2^{-(2n+1)}$	$1 - 2^{-2n}$	$2^{-(2n+1)}$

Dãy $\{X_n\}$ thỏa mãn luật số lớn Trêbưsép không?

4.16 Cho dãy các biến ngẫu nhiên độc lập $\{X_n\}$ xác định như sau:

X_n	$-na$	0	na
P	$\frac{1}{2n^2}$	$1-\frac{1}{n^2}$	$\frac{1}{2n^2}$

trong đó a là một hàng số. Dãy $\{X_n\}$ thỏa mãn luật số lớn Trêbưsép không?

4.17 Cho dãy các biến ngẫu nhiên độc lập $\{X_n\}$ xác định như sau:

X_n	$-a$	a
P	$\frac{n+1}{2n+1}$	$\frac{n}{2n+1}$

trong đó a là một hàng số. Dãy $\{X_n\}$ thỏa mãn luật số lớn Trêbưsép không?

4.18 Xác suất chậm tầu của mỗi hành khách là 0,007. Dùng bất đẳng thức Trêbưsép hãy đánh giá xác suất để trong 20.000 hành khách có từ 100 đến 180 người chậm tầu.

4.19 Phải kiểm tra bao nhiêu chi tiết để với xác suất không nhỏ hơn 0,98 có thể hy vọng rằng sai lệch giữa tần suất xuất hiện chi tiết tốt và xác suất để chi tiết là tốt bằng 0,95 sẽ không vượt quá 0,01.

4.20 Một xí nghiệp sản xuất máy tính có xác suất làm ra sản phẩm phế phẩm là 0,02. Chọn ngẫu nhiên 250 máy tính để kiểm tra. Tính xác suất để:

a) Có đúng hai máy phế phẩm;

b) Có không quá hai máy phế phẩm.

4.21 Một nhà nghỉ có 1000 khách. Nhà ăn phục vụ bữa trưa làm hai đợt liên tiếp. Số chỗ ngồi của nhà ăn phải ít nhất là bao nhiêu để xác suất của biến cố "không đủ chỗ cho người đến ăn" bé hơn 1%?

4.22 Một trường đại học có chỉ tiêu tuyển sinh là 300.

a) Giả sử có 325 người dự thi và xác suất thi đỗ của mỗi người là 90%. Tính xác suất để số người trúng tuyển không vượt quá chỉ tiêu.

b) Cần cho phép tối đa bao nhiêu người dự thi (xác suất đỗ của họ vẫn là 90%) để biến cố "số người trúng tuyển không vượt quá chỉ tiêu tuyển sinh" có xác suất lớn hơn 0,99.

4.23 Một kênh truyền có nhiễu với xác suất lỗi trên mỗi ký số là $0,01$.

a) Tính xác suất có nhiều hơn một lỗi trong 10 ký số nhận được;

b) Tính kết quả của ý a) bằng cách sử dụng công thức xấp xỉ Poisson.

4.24 Giả sử X là biến ngẫu nhiên có phân bố Poisson tham số λ. Áp dụng định lý giới hạn trung tâm hãy xây dựng công thức xấp xỉ

$$P\{X \le x\} \approx \Phi\left(\frac{x-\lambda}{\sqrt{\lambda}}\right).$$