TẠP CHÍ KHOA HỌC
Khoa học Tự nhiên và Công nghệ, Số 16 (6/2019) tr.38 - 48

HỘI TỤ THEO XÁC SUẤT CỦA ĐẠI LƯỢNG NGẪU NHIÊN VÀ MỘT SỐ ỨNG DỤNG

Đặng Kim Phương, Nguyễn Thanh Lâm
Trường Đại học Tây Bắc

Tóm tắt: *Trong khuôn khổ của bài viết này chúng tôi sẽ trình bày về tính ổn định của trung bình cộng một số lớn các đại lượng ngẫu nhiên theo nghĩa hội tụ theo xác suất và ứng dụng các định lý về luật yếu số lớn để giải một số bài toán trong xác suất thống kê, trong đó có những bài toán thống kê có ý nghĩa trong nghiên cứu khoa học thực nghiệm.*

Từ khóa: *Đại lượng ngẫu nhiên, Hội tụ theo xác suất, Luật yếu số lớn, Kỳ vọng, Hội tụ hầu chắc chắn.*

1. Mở đầu

Nghiên cứu về lý thuyết xác suất ta biết: Một biến cố ngẫu nhiên có thể xảy ra và có thể không xảy ra đối với một phép thử. Đại lượng ngẫu nhiên có thể lấy một trong các giá trị có thể có của nó. Nhưng khi xét một số lớn những biến cố ngẫu nhiên hay đại lượng ngẫu nhiên, ta có thể thu được kết luận nào đó mà thực tế có thể xem là chắc chắn. Trong lý thuyết xác suất người ta gọi những định lý mà nội dung của nó khẳng định sự hội tụ theo xác suất của một dãy các đại lượng ngẫu nhiên tới hằng số, hay tính ổn định của trung bình cộng một số lớn các đại lượng ngẫu nhiên theo nghĩa hội tụ theo xác suất là những định lý của luật yếu số lớn. Luật yếu số lớn có ý nghĩa quan trọng trong lý thuyết xác suất và trong thực tiễn. Nghiên cứu những điều kiện để một dãy các đại lượng ngẫu nhiên tuân theo luật yếu số lớn là vấn đề hết sức quan trọng. Vậy điều kiện để một dãy các đại lượng ngẫu nhiên tuân theo luật yếu số lớn là gì? Việc ứng dụng các định lý về luật yếu số lớn để giải một số bài toán trong xác suất thống kê, trong đó có bài toán thống kê trong nghiên cứu khoa học thực nghiệm được thực hiện như thế nào? Thông qua cơ sở lý luận và thực tiễn bài viết sẽ làm sáng tỏ thêm vấn đề này.

2. Hội tụ theo xác suất của dãy đại lượng ngẫu nhiên và một số ứng dụng

Chúng tôi nhắc lại một số khái niệm và kết quả sau trong [2] và [4].

Định nghĩa 2.1 Dãy các đại lượng ngẫu nhiên $X_n, n \geq 1$ được gọi là hội tụ theo xác suất tới đại lượng ngẫu nhiên X khi $n \to \infty$ nếu với mọi $\varepsilon > 0$

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

Ký hiệu: $X_n \xrightarrow{P} X,\ n \to \infty.$

Nhận xét 2.1: Ta có $X_n \xrightarrow{P} X,\ n \to \infty$ nghĩa là: $\forall \varepsilon > 0, \forall \delta > 0, \exists N = N(\varepsilon, \delta),$ sao cho

$$P\ \left|X_n - X\right| \geq \varepsilon\ < \delta, \forall n > N.$$

Ngày nhận bài: 11/03/2019. Ngày nhận đăng: 05/05/2019.
Liên lạc: Đặng Kim Phương, e-mail: dangkimphuongtbu@gmail.com

Hay nói cách khác, với $\forall \varepsilon > 0$:

$$\lim_{n\to\infty} P\left(\left|X_n - X\right| < \varepsilon\right) = 1.$$

Nghĩa là: $\forall \varepsilon > 0, \forall \delta : 0 < \delta < 1, \exists N = N(\varepsilon, \delta)$, sao cho $P\left(\left|X_n - X\right| \le \varepsilon\right) > 1 - \delta, \forall n > N$.

Định nghĩa 2.2 Dãy các đại lượng ngẫu nhiên $\left(X_n, n \ge 1\right)$ có kỳ vọng được gọi là tuân theo luật yếu số lớn nếu với mọi $\varepsilon > 0$ cho trước tùy ý ta có

$$\lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} EX_i\right| \ge \varepsilon\right\} = 0.$$

Nhận xét 2: Các *hệ thức sau là tương đương với định nghĩa:*

i) $\left(X_n, n \ge 1\right)$ tuân theo luật yếu số lớn.

ii) $\forall \varepsilon > 0, \lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} EX_i\right| < \varepsilon\right\} = 1.$

iii) $\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} EX_i \xrightarrow{P} 0, n \to \infty.$

Định lý 2.1 (Bất đẳng thức Chebyshev) *Nếu X là đại lượng ngẫu nhiên có phương sai DX hữu hạn thì với $\varepsilon > 0$ cho trước tùy ý ta có:*

$$P\left(\left|X - EX\right| \ge \varepsilon\right) \le \frac{DX}{\varepsilon^2}.$$

Bất đẳng thức tương đương $P\left(\left|X - EX\right| < \varepsilon\right) \ge 1 - \frac{DX}{\varepsilon^2}; \forall \varepsilon > 0.$

Bất đẳng thức Chebyshev được ứng dụng vào thực tiễn trong việc đánh giá cận trên và cận dưới xác suất để đại lượng ngẫu nhiên X nhận giá trị sai lệch so với kỳ vọng EX không quá một số $\varepsilon > 0$ cho trước mà không cần biết qui luật phân phối xác suất của X, từ đó lý giải cho các sai số trong đo lường vật lý.

Ví dụ 2.1 Kiểm tra ngẫu nhiên 10.000 máy tính trong một kho máy tính. Xác suất một máy tính lấy ra kiểm tra không đảm bảo chất lượng là 0,3. Tìm xác suất để độ lệch tuyệt đối giữa tần suất xuất hiện máy tính không đảm bảo chất lượng trong 10.000 máy tính trên so với xác suất để một máy tính lấy ra kiểm tra không đảm bảo chất lượng không quá 0,01.

Gọi A là biến cố "máy tính lấy ra kiểm tra là không đảm bảo chất lượng"

Đặt X_k là số lần xuất hiện A ở lần kiểm tra thứ k, vậy $\left(X_k, k \ge 1\right)$ là các đại lượng ngẫu nhiên độc lập, cùng phân phối:

X_k	0	1
P	0,7	0,3

Ta có: $EX_k = 0,3$, $DX_k = 0,21$, $m = \sum_{k=1}^{n} X_k$, $n = 10.000$

$$P\left\{\left|\frac{m}{n} - 0,3\right| < 0,01\right\} = P\left\{\left|\frac{m}{n} - E\left(\frac{m}{n}\right)\right| < 0,01\right\} \ge 1 - \frac{0,21}{n.(0,01)^2} = 1 - 0,3.0,7 = 0,79.$$

Ví dụ 2.2 Thu nhập trung bình hàng năm của dân cư một vùng là 20 triệu đồng và độ lệch chuẩn là 1,5 triệu. Hãy xác định khoảng thu nhập trung bình hàng năm của ít nhất 90% dân cư của vùng đó.

Gọi X là mức thu nhập hàng năm của dân cư thì ta có: $EX = 2; DX = 1,5$. Theo bất đẳng thức Chebyshev ta có với $\varepsilon > 0$

$$P\left(|X - EX| < \varepsilon\right) \geq 1 - \frac{DX}{\varepsilon^2} \Leftrightarrow P\left(|X - 20| < \varepsilon\right) \geq 1 - \frac{1,5}{\varepsilon^2}$$

Ta có $1 - \frac{1,5}{\varepsilon^2} = 0,9 \Rightarrow \varepsilon = 3,872$. Vậy ít nhất 90% dân cư của vùng đó có mức thu nhập nằm trong khoảng $(16,128; 23,872)$ triệu đồng.

Định lý 2.2 (Định lý Markov) [4] *Nếu* $(X_n, n \geq 1)$ *là dãy các đại lượng ngẫu nhiên thỏa mãn điều kiện*

$$\frac{1}{n^2} D\left(\sum_{i=1}^{n} X_i\right) \to 0,\ n \to \infty \text{ thì } \frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} EX_i \xrightarrow{P} 0,\ n \to \infty.$$

Chứng minh. Đặt $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ ta có $E\overline{X}_n = \frac{\mu_1 + \mu_2 + \ldots + \mu_n}{n}$, $D\overline{X}_n = \frac{1}{n^2} D\left(\sum_{i=1}^{n} X_i\right)$.

Áp dụng bất đẳng thức Chebyshev đối với $\overline{X}_n$ ta có với mọi $\varepsilon > 0$:

$$P\left(|\overline{X}_n - E\overline{X}_n| \geq \varepsilon\right) \leq \frac{1}{n^2\varepsilon^2} D\left(\sum_{i=1}^{n} X_i\right) \to 0,\ n \to \infty.$$

Do đó $\lim_{n\to\infty} P\left(|\overline{X}_n - E\overline{X}_n| \geq \varepsilon\right) = 0, \forall \varepsilon > 0$.

Vậy $\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} EX_i \xrightarrow{P} 0,\ n \to \infty$.

Ví dụ 2.3 Cho $(X_k, k \geq 1)$ là dãy các đại lượng ngẫu nhiên độc lập xác định như sau:

$$P(X_k = \sqrt{k}) = \frac{1}{2\sqrt{k}},\ P(X_k = -\sqrt{k}) = \frac{1}{2\sqrt{k}},\ P(X_k = 0) = 1 - \frac{1}{\sqrt{k}}.$$

Chứng minh rằng dãy $(X_k, k \geq 1)$ tuân theo luật yếu số lớn.

Với mỗi $k = 1,2,\ldots$ ta có

$$EX_k = -\sqrt{k}.\frac{1}{2\sqrt{k}} + 0.(1 - \frac{1}{\sqrt{k}}) + \sqrt{k}.\frac{1}{2\sqrt{k}} = 0$$

$$EX_k^2 = (-\sqrt{k})^2.\frac{1}{2\sqrt{k}} + 0.(1 - \frac{1}{\sqrt{k}}) + (\sqrt{k})^2.\frac{1}{2\sqrt{k}} = \sqrt{k}$$

$$DX_k = \sqrt{k}$$

Ta có

$$\frac{1}{n^2} D\left(\sum_{k=1}^{n} X_k\right) = \frac{1}{n^2}\sum_{k=1}^{n} DX_k = \frac{1}{n^2}\sum_{k=1}^{n}\sqrt{k} < \frac{1}{n^2}(\sqrt{n} + \sqrt{n} + \ldots + \sqrt{n}) = \frac{1}{\sqrt{n}} \to 0,\ n \to \infty \text{ nên}$$

$\lim_{n\to\infty} \frac{1}{n^2} D\left(\sum_{k=1}^{n} X_k\right) = 0$. Vậy theo định lý Markov ta có $\frac{1}{n}\sum_{k=1}^{n} X_k - \frac{1}{n}\sum_{k=1}^{n} EX_k \xrightarrow{P} 0,\ n \to \infty$ hay dãy $(X_k, k \geq 1)$ tuân theo luật yếu số lớn.

Định lý 2.3 (Định lý Chebyshev) [2] *Giả sử* $(X_n, n \geq 1)$ *là dãy các đại lượng ngẫu nhiên độc lập, có các phương sai bị chặn đều, nghĩa là tồn tại* $C > 0$ *sao cho* $DX_i \leq C$ *với mọi* $i = 1,2,\ldots$ *Khi đó*

$$\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} EX_i \xrightarrow{P} 0\,, n \to \infty.$$

Chứng minh. Đặt $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i\,,\ a_n = \frac{1}{n}\sum_{i=1}^{n} EX_i,$ ta có:

$$\overline{X}_n - a_n = \frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} a_i = \frac{1}{n}\sum_{i=1}^{n}(X_i - a_i)$$

$$E(\overline{X}_n - a_n) = 0$$

$$D(\overline{X}_n - a_n) = \frac{1}{n^2}\sum_{i=1}^{n} DX_i \le \frac{C}{n}$$

Áp dụng bất đẳng thức Chebyshev đối với $\overline{X}_n - a_n$, ta có với mọi $\varepsilon > 0$:

$$P\left(\left|\overline{X}_n - a_n\right| \ge \varepsilon\right) \le \frac{C}{n\varepsilon^2} \to 0, n \to \infty.$$

Do đó $\lim_{n\to\infty} P\left(\left|\overline{X}_n - a_n\right| \ge \varepsilon\right) = 0, \forall \varepsilon > 0.$

Vậy $\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} EX_i \xrightarrow{P} 0, n \to \infty$ hay $\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} \frac{1}{n}\sum_{i=1}^{n} EX_i, n \to \infty.$

Định lý Chebyshev chứng minh sự hội tụ theo xác suất của trung bình số học của một số lớn các đại lượng ngẫu nhiên về trung bình số học của các kỳ vọng tương ứng, tức là nó chứng minh tính ổn định của giá trị trung bình.

Hệ quả 2.1 *Giả sử* $\{X_n, n \ge 1\}$ *là dãy các đại lượng ngẫu nhiên độc lập, cùng phân phối, có kỳ vọng là* μ *và các phương sai* σ^2*. Khi đó*

$$\frac{1}{n}\sum_{i=1}^{n} X_i - \mu \xrightarrow{P} 0\,, n \to \infty.$$

Như vậy, mặc dù từng đại lượng ngẫu nhiên X_i có thể nhận giá trị khác nhiều so với kỳ vọng $EX_i,$ nhưng trung bình cộng của một số rất lớn các đại lượng ngẫu nhiên X_i lại nhận giá trị rất gần với kỳ vọng với xác suất gần 1 tùy ý. Mặt khác ta có

$$\frac{1}{n}\sum_{i=1}^{n} X_i - \mu \xrightarrow{P} 0, n \to \infty$$

tức là $\lim_{n\to\infty} P\left(\left|X_n - X\right| < \varepsilon\right) = 1, \forall \varepsilon > 0$ điều đó chứng tỏ rằng

$\forall \varepsilon > 0, \forall \delta : 0 < \delta < 1, \exists N = N(\varepsilon, \delta)$ sao cho $P\left(\left|X_n - X\right| \le \varepsilon\right) > 1 - \delta, \forall n > N$. Nếu ε, δ được chọn nhỏ đến mức: Sự khác biệt nhỏ thua ε được coi như đồng nhất, biến cố có xác suất lớn hơn $1 - \delta$ coi là luôn xuất hiện thì mặc dù $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ có tính chất ngẫu nhiên ta có thể đồng nhất nó với hằng số μ. Điều đó thể hiện rõ sự ổn định của trung bình số học của các đại lượng ngẫu nhiên độc lập cùng phân phối có phương sai hữu hạn, chính sự ổn định này được lấy làm cơ sở cho định nghĩa xác suất theo tần suất.

Định lý 2.4 (Định lý Poisson) [2] *Giả sử trong một dãy phép thử độc lập, xác suất để một biến cố* A *nào đó xảy ra trong phép thử thứ* k *bằng* p_k*. Ký hiệu* Y_n *là số lần xuất hiện biến cố* A *trong* n *phép thử đầu. Khi đó*

$$\frac{1}{n}Y_n - \frac{1}{n}\sum_{k=1}^{n} p_k \xrightarrow{P} 0.$$

Chứng minh. Định nghĩa dãy đại lượng ngẫu nhiên $X_1, X_2,...X_n,...$ như sau:

X_k nhận giá trị là 1 nếu A xảy ra ở phép thử thứ k và X_k nhận giá trị là 0 nếu A không xảy ra ở phép thử thứ k, với $k = 1,2,...$ Ta có: $X_1, X_2,...X_n,...$ là các đại lượng ngẫu nhiên độc lập, $Y_n = \sum_{k=1}^{n} X_k$. Với mỗi $k = 1,2,...$ thì:

$$P(X_k = 1) = p_k, P(X_k = 0) = 1 - p_k$$
$$EX_k = EX_k^2 = p_k$$
$$DX_k = EX_k^2 - (EX_k)^2 = p_k(1 - p_k) \le \frac{1}{4}$$

Theo định lý Chebyshev ta có $\frac{1}{n}Y_n - \frac{1}{n}\sum_{k=1}^{n} p_k \xrightarrow{P} 0$.

Hệ quả 2.2 (Định lý Bernoulli) [4] *Giả sử S_n là số lần xuất hiện biến cố A trong dãy n phép thử Bernoulli và $P(A) = p$ không đổi trong mỗi phép thử, thì $\frac{S_n}{n} \xrightarrow{P} p, n \to \infty$.*

Định lý Bernoulli đã khẳng định rằng: Với n đủ lớn thì tần suất và xác suất xuất hiện biến cố A trong dãy n phép thử Bernoulli sẽ sai khác một lượng không quá một số $\varepsilon > 0$ cho trước tùy ý với xác suất gần 1. Kết luận trên cho ta xác định xấp xỉ (ước lượng) xác suất chưa biết của một biến cố A qua tần suất $\frac{S_n}{n}$ của A trong dãy phép thử Bernoulli.

Luật yếu số lớn đã được ứng dụng rất nhiều trong lý thuyết xác suất và là cơ sở khoa học trong nghiên cứu khoa học thực nghiệm. Thông qua một số ví dụ sau đây chúng ta sẽ thấy rõ hơn về vấn đề này.

Ví dụ 2.4 Cho $X_n, n \ge 1$ là dãy các đại lượng ngẫu nhiên độc lập xác định như sau:

$$P\left[X_k = \pm\sqrt{\ln k}\right] = \frac{1}{2}.$$

Dãy các đại lương ngẫu nhiên $X_n, n \ge 1$ có tuân theo luật yếu số lớn không?

Bảng phân phối xác suất của X_k :

X_k	$-\sqrt{\ln k}$	$+\sqrt{\ln k}$
P	$\frac{1}{2}$	$\frac{1}{2}$

Ta có $EX_k = 0, DX_k = \ln k$. Với mọi $\varepsilon > 0$:

$$P\left\{\left|\frac{1}{n}\sum_{k=1}^{n} X_k - \frac{1}{n}\sum_{k=1}^{n} EX_k\right| \ge \varepsilon\right\} = P\left\{\left|\frac{1}{n}\sum_{k=1}^{n} X_k\right| \ge \varepsilon\right\} \le \frac{1}{n^2\varepsilon^2}\sum_{k=1}^{n} \ln k \ (*)$$

Ta lại có: $\sum_{k=1}^{n} \ln k \le \sum_{k=1}^{n}\int_k^{k+1} \ln x dx = \int_1^{n+1} \ln x \le \ln(n+1)[n+1-1] = n\ln(n+1)$ (vì $\int_k^{k+1} \ln x dx \ge \int_k^{k+1} \ln k\, dx = \ln k$) Thay vào (*) có: $P\left\{\left|\frac{1}{n}\sum_{k=1}^{n} X_k\right| \ge \varepsilon\right\} \le \frac{1}{n^2\varepsilon^2}.n.\ln(n+1) \to 0; n \to \infty$.

Vậy dãy các đại lượng ngẫu nhiên $X_n, n \geq 1$ tuân theo luật yếu số lớn.

Ví dụ 2.5 Cho $X_k, k \geq 1$ là dãy các đại lượng ngẫu nhiên độc lập xác định như sau:

$$P(X_k = 2^k) = \frac{1}{2},\ P(X_k = -2^k) = \frac{1}{2}.$$

Chứng minh rằng dãy $X_k, k \geq 1$ không tuân theo luật yếu số lớn.

Với mỗi $k = 1, 2, \ldots$ ta có $EX_k = 0$. Xét biến cố $A_n = \{X_n = 2^n, X_{n-1} = 2^{n-1}\}$ có xác suất $P(A_n) = \frac{1}{4}$. Nếu biến cố A_n xảy ra thì với $n \geq 1$ ta có

$$X_1 + X_2 + \ldots + X_{n-1} + X_n = X_1 + X_2 + \ldots + X_{n-2} + 2^{n-1} + 2^n \geq -(2 + 2^2 + \ldots + 2^{n-2}) + 2^{n-1} + 2^n$$
$$= -2^{n-1} + 2 + 2^{n-1} + 2^n = 2^n + 2 \geq n.$$

Vậy $P\left\{\frac{X_1 + X_2 + \ldots + X_n}{n} \geq 1\right\} \geq P(A_n) = \frac{1}{4},\ \forall n \geq 1,$ do đó không xảy ra giới hạn

$\lim\limits_{n \to \infty} P\left\{\frac{X_1 + X_2 + \ldots + X_n}{n} \geq 1\right\} = 0$. Vậy dãy $X_k, k \geq 1$ không tuân theo luật yếu số lớn.

Ví dụ 2.6 Một cửa hàng bán vải muốn ước lượng nhanh chóng số vải bán ra trong một tháng (số vải của mỗi khách hàng mua được làm tròn đến số nguyên gần nhất). Ký hiệu X_i là sai số giữa số mét vải thực bán và số mét vải đã tính tròn của khách hàng thứ i. Với xác suất 0,99 hãy ước lượng sai số giữa số mét vải thực sự bán và số mét vải đã tính tròn trong một tháng của cửa hàng.

Các sai số $X_i, i = \overline{1, n}$ là các đại lượng ngẫu nhiên độc lập có phân phối đều trên đoạn $[-0{,}5 ; 0{,}5]$ nên với mỗi $i = 1, 2, \ldots, n$ thì $EX_i = 0 ;\ DX_i = \frac{1}{12}$.

Sai số tổng cộng trong cả tháng là $S = X_1 + X_2 + \ldots + X_n$ (n là số khách hàng mua vải trong tháng) và: $ES = \sum\limits_{i=1}^{n} X_i = 0 ;\ DS = \sum\limits_{i=1}^{n} DX_i = \frac{n}{12}$.

Theo bất đẳng thức Chebyshev, xác suất để sai số không vượt quá ε mét được đánh giá bởi $P\{|S| < \varepsilon\} \geq 1 - \frac{DS}{\varepsilon^2} = 1 - \frac{n}{12\varepsilon^2}$. Với xác suất $P\{|S| < \varepsilon\} = 0{,}99$ để sai số không vượt quá ε mét thì $\frac{n}{12\varepsilon^2} \geq 0{,}01 \Leftrightarrow \varepsilon \leq \sqrt{\frac{n}{12.0{,}01}}$.

Giả sử số khách hàng đến mua trong một tháng là $n = 10000$ thì $\varepsilon \leq \sqrt{\frac{10000}{12.0{,}01}} = 288{,}67$. Vậy ta có thể kết luận rằng, với xác suất 0,99 sai số giữa số mét vải thực bán và số mét vải đã tính tròn trong một tháng của cửa hàng không vượt quá 289 mét nếu số khách hàng đến mua là 1 vạn.

Ví dụ 2.7 Xác định số phép thử Bernoulli tối thiểu để ước lượng xác suất p của biến cố A với độ chính xác 0,1 và với độ tin cậy 0,95.

Yêu cầu trên được diễn đạt như sau: Xác định n nhỏ nhất sao cho

$$P\left\{\left|\frac{S_n}{n} - p\right| \geq 0{,}1\right\} \leq 1 - 0{,}95 = 0{,}05.$$

Từ hệ quả 2.2 ta chỉ biết $\frac{S_n}{n}$ hội tụ theo xác suất tới xác suất $p = P(A)$ chưa biết. Tuy nhiên trong dãy n phép thử Bernoulli thì $E\left(\frac{S_n}{n}\right) = p$, $D\left(\frac{S_n}{n}\right) = \frac{1}{n}p(1-p)$.

Ta có $P\left\{\left|\frac{S_n}{n} - p\right| \geq 0,1\right\} \leq \frac{1}{0,1^2 n}p(1-p) = \frac{100p(1-p)}{n}$.

Vì $p(1-p) \leq \frac{1}{4}$ nên để $P\left\{\left|\frac{S_n}{n} - p\right| \geq 0,1\right\} \leq 0,05$ thì $\frac{100}{4n} \leq 0,05$ hay $n \geq 500$. Vậy $n_{\min} = 500$.

Theo lý thuyết thống kê, nếu $(X_1, X_2, ..., X_n)$ là mẫu ngẫu nhiên của dấu hiệu nghiên cứu X có kỳ vọng $EX = a$ phương sai $DX = \sigma^2$ thì $X_1, X_2, ..., X_n$ là các đại lượng ngẫu nhiên độc lập cùng phân phối và

$$EX_1 = EX_2 = ... = EX_n = a; DX_1 = DX_2 = ... = DX_n = \sigma^2.$$

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \Rightarrow E\overline{X} = \frac{1}{n}\sum_{i=1}^{n} EX_i = \frac{na}{n} = a.$$

Theo định lý chebyshev ta có $\lim_{n\to\infty} P\left|\overline{X} - a\right| \geq \varepsilon = 0, \forall \varepsilon > 0$. Đây là cơ sở khoa học để khi nghiên cứu về khoa học thực nghiệm, có thể ước lượng trung bình của tổng thể thông qua giá trị trung bình mẫu, ước lượng tỷ lệ của tổng thể thông qua tần suất.

Ví dụ 2.8 Để đánh giá mức thu nhập (triệu đồng/ tháng) của người lao động trong thời kỳ hội nhập kinh tế quốc tế của một nhà máy, tiến hành khảo sát thu nhập của 100 người lao động. Kết quả thu được như sau:

Tổng thu nhập	$\leq$ 10	(10 - 12]	(12 - 14]	(14 - 16]	(16 - 18]	(18 - 20]	> 20
Số người	8	10	12	15	20	20	15

Giả sử thu nhập của người lao động là đại lượng ngẫu nhiên có phân phối chuẩn dạng tổng quát $N(a;\sigma^2)$. Với độ tin cậy 0,95 có thể khẳng định thu nhập trung bình của người lao động trong thời kỳ hội nhập kinh tế quốc tế của nhà máy là bao nhiêu?

Lập bảng số liệu:

STT	Khoảng	Tần số (n_i)	X_i	$u_i = \frac{X_i - X_0}{h}; i = 1,2,3,4,5,6.$	$n_i u_i$	$n_i u_i^2$
1	$\leq$ 10	8	9	-3	-24	72
2	(10 - 12]	10	11	-2	-20	40
3	(12 - 14]	12	13	-1	-12	12
4	(14 - 16]	15	15	0	0	0
5	(16 - 18]	20	17	1	20	20
6	(18 - 20]	20	19	2	40	80
7	> 20	15	21	3	45	135
$\sum$		100			49	359

Ta có trung bình mẫu $\overline{X} = X_0 + h\overline{u}$.

Tính:

$$\bar{u} = \frac{1}{n}\sum_{i=1}^{6} n_i u_i = \frac{49}{100} = 0,49$$

$$\bar{X} = X_0 + h\bar{u} = 15 + 2.0,49 = 15,98$$

Nếu kích thước của mẫu đã đủ lớn thì theo định lý Chebyshev ta có thể kết luận thu nhập trung bình của người lao động trong nhà máy là 15,98 triệu đồng/ tháng. Nhưng thực tế ta chưa biết được kích thước của mẫu đã thỏa mãn là đủ lớn hay chưa? Nếu kết luận thu nhập trung bình của người lao động trong nhà máy là 15,98 triệu đồng/ tháng thì có thể xảy ra sai số rất lớn, do đó căn cứ vào kết quả tính $\bar{X}$ của mẫu tìm khoảng ước lượng thu nhập trung bình a của tổng thể:
Tính:

$$S_n^{*2}(X) = h^2\left[\frac{1}{n}\sum_{i=1}^{6} n_i u_i^2 - \bar{u}^2\right] = 4\left(\frac{359}{100} - 0,49^2\right) = 13,39$$

$$S_n^{*2}(X) = \frac{n}{n-1}S_n^2(X) = \frac{100}{99}.13,39 = 13,52$$

$$S_n^{*}(X) = 3,67$$

Khoảng ước lượng thu nhập trung bình của người lao động trong nhà máy là:

$$\bar{X} - x_\alpha \frac{S_n^*(X)}{\sqrt{n}} < a < \bar{X} + x_\alpha \frac{S_n^*(X)}{\sqrt{n}}$$

$$15,98 - 1,96\frac{3,67}{10} < a < 15,98 + 1,96\frac{3,67}{10}$$

$$15,27 < a < 16,69$$

Kết quả này cho biết thu nhập trung bình của người lao động trong nhà máy sẽ nằm trong khoảng từ 15,27 đến 16,69 (triệu đồng/ tháng) nên có thể dự đoán thu nhập trung bình của người lao động trong nhà máy là 16,6 (triệu đồng/ tháng).

Kiểm định giả thiết:

$$\begin{cases} H_0 : a = 16,6 \\ K : a \neq 16,6 \end{cases}$$

Với mức ý nghĩa $\alpha = 0,05$.
Ta có

$$|T| = \frac{|\bar{X} - a_0|\sqrt{n}}{S_n^*(X)} = \frac{|15,98 - 16,6|\sqrt{100}}{3,67} = 1,68$$

Tra bảng phân phối chuẩn $N(0,1)$ ta có $x_\alpha = 1,96$. Do $|T| < x_\alpha$ nên giả thiết $H_0 : a = 16,6$ được chấp nhận ở mức ý nghĩa $\alpha = 0,05$. Với kết quả này ta kết luận, thu nhập trung bình của người lao động trong thời kỳ hội nhập kinh tế quốc tế của nhà máy là 16,6 triệu đồng/ tháng.

Ví dụ 2.9 Để kiểm tra chất lượng sản phẩm do một nhà máy sản xuất, người ta đã sử dụng phương pháp chọn mẫu như sau: Từ mỗi lô sản phẩm chọn ngẫu nhiên lần lượt ra n sản phẩm để kiểm tra. Với a là một số cho trước, nếu số sản phẩm không đảm bảo chất lượng nhỏ hơn hoặc bằng a thì lô sản phẩm được chấp nhận, nếu số sản phẩm không đảm bảo chất lượng lớn hơn a thì không chấp nhận lô sản phẩm.

1. Từ một lô sản phẩm do nhà máy sản xuất chọn ngẫu nhiên lần lượt ra 100 sản phẩm để kiểm tra, thấy có 5 sản phẩm không đảm bảo chất lượng. Có thể khẳng định tỷ lệ sản phẩm không đảm bảo chất lượng của lô sản phẩm là bao nhiêu?

2. Giả sử có một lô hàng của nhà máy trong đó có 5% sản phẩm không đảm bảo chất lượng. Với $n = 5 ; a = 1$ thì xác suất để lô hàng đó được chấp nhận là bao nhiêu? không được chấp nhận là bao nhiêu?

Theo kết quả thống kê, tần suất của sản phẩm không đảm bảo chất lượng trong mẫu là $\frac{5}{100}$. Theo luật số lớn, tần suất xuất hiện sản phẩm không đảm bảo chất lượng trong mẫu sẽ được coi là tỷ lệ sản phẩm không đảm bào chất lượng của lô sản phẩm nếu kích thước mẫu n đủ lớn. Nhưng ta chưa biết kích thước mẫu $n = 100$ như vậy là đã đủ lớn hay chưa? Do đó để khẳng định tỷ lệ sản phẩm không đảm bảo chất lượng (p) của lô sản phẩm là bao nhiêu? Ta thực hiện như sau:

Với độ tin cậy 0,95 xác định khoảng ước lượng tỷ lệ sản phẩm không đảm bảo chất lượng của lô hàng:

$$\hat{p} - x_\alpha\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \hat{p} + x_\alpha\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Ta có:

$$\hat{p} = \frac{5}{100} = 0,05 \; ; \; x_\alpha = 1,96$$

nên

$$0,05 - 1,96\sqrt{\frac{0,05.0,95}{100}} < p < 0,05 + 1,96\sqrt{\frac{0,05.0,95}{100}}$$
$$0,0073 < p < 0,0927$$

Kết quả này cho biết, tỷ lệ sản phẩm không đảm bảo chất lượng của lô hàng nằm trong khoảng từ 0,73% đến 9,27%.

Với mức ý nghĩa $\alpha = 0,05$ kiểm định giả thiết

$$\begin{cases} H_0 : p = 0,0925 \\ K : p \neq 0,0925 \end{cases}$$

Tính

$$\left|Z\right| = \frac{\left|X - np_0\right|}{\sqrt{np_0(1-p_0)}} = \frac{\left|5 - 100.0,0925\right|}{\sqrt{100.0,0925.0,9075}} = 1,47.$$

Bởi vì $\left|Z\right| < 1,96$ nên giả thiết $H_0 = 0,0925$ được chấp nhận ở mức ý nghĩa $\alpha = 0,05$. Do đó có thể khẳng định tỷ lệ sản phẩm không đảm bảo chất lượng của lô hàng là 9,25%.

Phân tích bài toán:

- Việc chọn một mẫu $n = 5$ là thực hiện 5 lần thử giống nhau
- Số lượng sản phẩm trong lô hàng lớn nên các lần thử độc lập với nhau
- Mỗi lần thử chỉ có 2 khả năng có thể xảy ra: Sản phẩm lấy ra là đảm bảo chất lượng hoặc không đảm bảo chất lượng
- Lô hàng có 5% sản phẩm không đảm bảo chất lượng, nên theo luật yếu số lớn xác suất để 1 sản phẩm của lô hàng chọn ra là sản phẩm không đảm bảo chất lượng luôn bằng 0,05

Do đó việc chọn mẫu này là thực hiện một dãy phép thử Bernoulli.

Ta có

$$P_5(k) = C_5^k 0,05^k.0,95^{5-k} \; ; \; k = 0,1,2,3,4,5.$$

Vậy

P (Chấp nhận) = $P_5(0) + P_5(1) = C_5^0 0,05^0.0,95^5 + C_5^0 0,05^1.0,95^4 = 0,98.$

P (Không chấp nhận) = 1- P (Chấp nhận) = $1 - 0,98 = 0,02.$

Ví dụ 2.10 Trong bài viết "*Tòa án tối cao đồng ý về quyền của Quốc hội trong việc điều tiết truyền hình cáp*" của David Savage (1994) ghi nhận rằng 60% các hộ gia đình ở Hoa Kỳ có truyền hình cáp. Giả sử con số 60% là chính xác, nếu điều tra 4 hộ gia đình ở Hoa Kỳ về việc họ có hay không có truyền hình cáp thì:

1. Xác suất để cả 4 hộ gia đình đều sử dụng truyền hình cáp là bao nhiêu?
2. Xác suất để ít nhất 1 hộ gia đình sử dụng truyền hình cáp là bao nhiêu?
3. Nếu khảo sát 50 hộ gia đình thì trung bình sẽ có bao nhiêu hộ gia đình sử dụng truyền hình cáp?

Theo kết quả thống kê, có 60% hộ gia đình ở Hoa Kỳ sử dụng truyền hình cáp, nên theo luật yếu số lớn xác suất để một hộ gia đình ở Hoa Kỳ có sử dụng truyền hình cáp là 0,6. Phân tích tương tự ví dụ 2.9 ta thấy, việc điều tra 4 hộ gia đình là thực hiện một dãy phép thử Bernoulli.
Ta có

$$P_4(k) = C_4^k 0,6^k.0,4^{4-k} \ ; k = 0,1,2,3,4.$$

Vậy:

$$P(k = 4) = C_4^4 0,6^4.0,4^0 = 0,1296$$
$$P(k \geq 1) = 1 - P(k = 0) = 1 - C_4^0 0,6^0.0,4^4 = 0,0256$$

Nếu khảo sát 50 hộ gia đình thì trung bình sẽ có $np = 50.0,6 = 30$ hộ gia đình có sử dụng truyền hình cáp.

Ví dụ 2.11 Trong sự nhìn nhận toàn diện về nhau của người Nhật và người Mỹ, người Nhật đã cho thấy niềm tự hào to lớn trong chất lượng các sản phẩm của họ. Tuy thế, họ vẫn cho rằng Hoa Kỳ sẽ đóng một vai trò lớn hơn cả về sự lãnh đạo lẫn quyền lực kinh tế so với người Nhật trong những năm tới. Cụ thể là, 71% số người Nhật cho rằng sản phẩm của họ tốt hơn sản phẩm của người Mỹ, và 42% cho rằng Hoa Kỳ sẽ là cường quốc kinh tế số một thế giới trong thế kỷ tới ("How the Japanese See Themselves...and US" (1990)). Lựa chọn ngẫu nhiên 50 công dân Nhật để thăm dò ý kiến.

1. Tìm phân phối xác suất của số người Nhật cho rằng sản phẩm của họ tốt hơn sản phẩm của người Mỹ.
2. Tìm phân phối xác suất của số người Nhật cho rằng Hoa Kỳ sẽ là cường quốc kinh tế số một thế giới trong thế kỷ tới.
3. Tìm trung bình và độ lệch chuẩn của số người Nhật cho rằng Hoa Kỳ sẽ là cường quốc kinh tế số một thế giới trong thế kỷ tới.

Theo kết quả thống kê:

+) 71% số người Nhật cho rằng sản phẩm của họ tốt hơn sản phẩm của người Mỹ nên theo luật yếu số lớn, xác suất để một công dân Nhật khi được thăm dò ý kiến sẽ cho rằng "sản phẩm của họ tốt hơn sản phẩm của người Mỹ" là 0,71.

+) 42% số người Nhật cho rằng Hoa Kỳ sẽ là cường quốc kinh tế số một thế giới trong thế kỷ tới nên theo luật yếu số lớn, xác suất để một công dân Nhật khi được thăm dò ý kiến sẽ cho rằng ''Hoa Kỳ sẽ là cường quốc kinh tế số một thế giới trong thế kỷ tới'' là 0,42.

Phân tích tương tự các ví dụ trên ta thấy, việc lựa chọn ngẫu nhiên 50 công dân Nhật để thăm dò ý kiến là thực hiện một dãy phép thử Bernoulli.

Gọi X là số người Nhật cho rằng sản phẩm của họ tốt hơn sản phẩm của người Mỹ thì X là đại lượng ngẫu nhiên có phân phối nhị thức với tham số $(n,p) = (50;0,71)$. Do đó phân phối xác suất của X là

$$P(X = k) = C_{50}^k 0,71^k.0,29^{50-k} \ ; k = \overline{0;50}.$$

Gọi Y là số người Nhật cho rằng Hoa Kỳ sẽ là cường quốc kinh tế số một thế giới trong thế kỷ tới thì Y là đại lượng ngẫu nhiên có phân phối nhị thức với tham số $(n,p)=(50;0,42)$. Do đó phân phối xác suất của Y là

$$P(Y=k)=C_{50}^{k}0,42^{k}.0,58^{50-k}\ ;\ k=\overline{0;50}.$$

Trung bình của số người Nhật cho rằng Hoa Kỳ sẽ là cường quốc kinh tế số một thế giới trong thế kỷ tới là $EY=n.p=50.0,42=21$ (người).

Xác định độ lệch chuẩn của số người Nhật cho rằng Hoa Kỳ sẽ là cường quốc kinh tế số một thế giới trong thế kỷ tới:

Ta có $DY=npq=50.0,42.0,58=12,18$ nên độ lệch chuẩn là $\sqrt{DY}=\sqrt{12,18}=3,49.$

TÀI LIỆU THAM KHẢO

[1] Đinh Văn Gắng (2003), Lý thuyết xác suất và thống kê, Nxb Giáo dục.

[2] Nguyễn Văn Hộ (2005), Xác suất thống kê, Nxb Giáo dục.

[3] Phạm Văn Kiều (2011), Giáo trình xác suất và thống kê, Nxb Giáo dục.

[4] Đặng Hùng Thắng (2011), Mở đầu về lý thuyết xác suất và các ứng dụng, Nxb Giáo dục.

CONVERGENCE IN PROBABILITY OF RANDOM VARIABLES AND SOME APPLICATIONS

Dang Kim Phuong, Nguyen Thanh Lam

Tay Bac University

Abstract: *In this article, we present the stability of the arithmetic mean of a large number of random variables in terms of convergence in probability and the applications of large number weak law theorems to solve some problems in statistical probability, including meaningful ones in experimental research.*

Keywords: *Random variables, Convergence in probability, Weak law of large numbers, Expected Value*