ĐẠI HỌC THÁI NGUYÊN

TRƯỜNG ĐẠI HỌC KHOA HỌC

TẠ THỊ THẮM

ĐỘ LỆCH LỚN CHO DÃY CÁC BIẾN NGẪU NHIÊN BERNOULLI VÀ ÁP DỤNG

LUẬN VĂN THẠC SĨ TOÁN HỌC

THÁI NGUYÊN - 2021

ĐẠI HỌC THÁI NGUYÊN
TRƯỜNG ĐẠI HỌC KHOA HỌC

TẠ THỊ THẮM

ĐỘ LỆCH LỚN CHO DÃY CÁC BIẾN NGẪU NHIÊN BERNOULLI VÀ ÁP DỤNG

Chuyên ngành: Toán ứng dụng
Mã số : 8 46 01 12

LUẬN VĂN THẠC SĨ TOÁN HỌC

NGƯỜI HƯỚNG DẪN KHOA HỌC
1. TS. Trần Xuân Quý
2. TS. Đỗ Thị Phương Quỳnh

THÁI NGUYÊN - 2021

Mục lục

Danh sách kí hiệu

$\mathbb{N}$	Tập hợp các số tự nhiên
$\mathbb{R}$	Tập hợp các số thực
$\mathbb{R}^+$	Tập hợp các số thực dương
$n!$	đọc là n giai thừa, ký hiệu cho tích $1.2.\ldots.n$
max	Cực đại
min	Cực tiểu
sup	Cận trên lớn nhất
$(\Omega, \mathcal{A}, \mathbb{P})$	Không gian xác suất
2^X	Họ các tập hợp con khác rỗng của X
$\sum a_i$	$a_1 + a_2 + \ldots + a_n$
$\mathbb{P}$	Độ đo xác suất
ξ_i	Biến ngẫu nhiên ξ_i
$\{\xi \geqslant \epsilon\}$	Biến cố $\{\omega \in \Omega : \xi(\omega) \geqslant \epsilon\}$
$\mathbb{E}(X)$	Kỳ vọng hay giá trị trung bình của biến ngẫu nhiên X
p-lim	Giới hạn của sự hội tụ theo xác suất
h.c.c.	Hầu chắc chắn
$\exp(x)$	e^x
$\binom{n}{k}$	Số tổ hợp chập k của n, hay $\binom{n}{k} = C_n^k = \frac{n!}{k!(n-k)!}$.

Mở đầu

Lý thuyết xác suất thống kê là một bộ phận của toán học, nghiên cứu các hiện tượng ngẫu nhiên và ứng dụng chúng vào thực tế. Là hiện tượng ngẫu nhiên nên không thể nói trước nó xảy ra hay không xảy ra khi thực hiện các quan sát. Tuy nhiên, nếu tiến hành quan sát khá nhiều lần một hiện tượng ngẫu nhiên trong các phép thử như nhau, ta có thể rút ra được những kết luận khoa học về hiện tượng này. Trong lý thuyết xác suất, độ lệch lớn liên quan tới dáng điệu đuôi của dãy các phân phối xác suất, nó cho phép đánh giá tốc độ hội tụ của họ các biến cố với xác suất lớn. Luật số lớn là một phần của Lý thuyết xác suất và thống kê. Trong thực tế, những hiện tượng ngẫu nhiên do rất nhiều nguyên nhân ngẫu nhiên gây ra. Việc tìm hiểu kiện để những hiện tượng như vậy xảy ra theo một quy luật nào đó là ý nghĩa của nội dung "luật số lớn". Với khuôn khổ đề tài luận văn thạc sĩ, tác giả tập trung trình bày về "Độ lệch lớn cho dãy các biến ngẫu nhiên Bernoulli và áp dụng".

Mục đích của đề tài luận văn là trình bày về ý nghĩa thực nghiệm của luật số lớn, các địnhlý giới hạn, độ lệch lớn cho dãy biến ngẫu nhiên Bernoulli và đánh giá xác suất thành công trong dãy Bernoulli. Ngoài phần Mở đầu, phần Kết luận, luận văn trình bày hai chương, gồm các nội dung:

Chương 1. Luật số lớn cho dãy biến ngẫu nhiên có phân phối Bernoulli.

Chương 2. Các định lý giới hạn, độ lệch lớn, bài toán đánh giá xác suất thành công cho dãy biến ngẫu nhiên có phân phối Bernoulli.

Trong quá trình học tập và nghiên cứu tại trường Đại học Khoa học, Đại học Thái Nguyên, em luôn nhận được sự quan tâm giúp đỡ và động viên của các thầy cô trong Ban Giám hiệu, Phòng Đào tạo và Khoa Toán -Tin. Qua đây, em gửi lời tri ân tới tập thể các thầy cô giảng viên của trường Đại học Khoa học – Đại học Thái Nguyên nói chung và Khoa Toán - Tin nói riêng, đã truyền thụ cho em nhiều kiến thức khoa học quý báu trong thời gian em được là học viên của trường. Tác giả xin chân thành cảm ơn Ban Giám hiệu trường THPT Nam Khoái Châu, Hưng Yên, cùng toàn thể các anh chị em đồng nghiệp đã tạo điều kiện tốt nhất cho tác giả trong thời gian đi học Cao học; cảm ơn các anh chị em học viên lớp Cao học Toán K13 và bạn bè đồng nghiệp đã trao đổi, động viên và khích lệ tác giả trong quá trình học tập và làm luận văn tại trường Đại học Khoa học, Đại học Thái Nguyên. Đặc biệt em xin được bày tỏ lòng biết ơn sâu sắc tới giáo viên hướng dẫn - TS. Trần Xuân Quý và TS. Đỗ Thị Phương Quỳnh đã luôn quan tâm ân cần chỉ bảo, động viên khích lệ, giúp đỡ tận tình và góp ý sâu sắc cho em trong suốt quá trình học tập cũng như thực hiện đề tài. Chặng đường vừa qua sẽ là những kỉ niệm đáng nhớ và đầy ý nghĩa đối với các anh chị em học viên lớp K13 nói chung và với bản thân em nói riêng. Dấu ấn ấy hiển nhiên không thể thiếu sự hỗ trợ, sẻ chia đầy yêu thương của tất cả người thân trong gia đình. Xin chân thành cảm ơn tất cả những người thân yêu đã giúp đỡ, đồng hành cùng em trên chặng đường vừa qua.

Một lần nữa, em xin trân trọng cảm ơn!

Thái Nguyên, ngày 25 tháng 01 năm 2021

Học viên

Tạ Thị Thắm

Chương 1

Luật số lớn cho dãy biến ngẫu nhiên có phân phối Bernoulli

1.1 Một số khái niệm và kết quả liên quan

Trong mục này sẽ giới thiệu về dãy biến ngẫu nhiên có phân phối Bernoulli, bất đẳng thức Chebysev được sử dụng cho các đánh giá trong chương sau. Cho bộ ba $(\Omega, \mathcal{A}, \mathbb{P})$ với

$$\Omega = \{\omega : \omega = (a_1, \ldots, a_n)\};$$
$$A = \{A : A \subset \Omega\};$$
$$p(\omega) = p^{\sum a_i} \cdot q^{n-\sum a_i};$$

gọi là sơ đồ Bernoulli.

Trong phần này và phần tiếp theo ta nghiên cứu một số tính chất giới hạn của sơ đồ Bernoulli.

Xét dãy biến ngẫu nhiên $\{\xi_1; \ldots; \xi_n\}$ xác định như sau:

$$\xi_i(\omega) := a_i; i := 1; \ldots; n; \omega = (a_1, \ldots, a_n).$$

Các biến ngẫu nhiên Bernoulli $\xi_i(\omega)_i$ độc lập với nhau và có cùng phân phối xác suất:

$$\mathbb{P}\{\xi_i = 1\} := p;$$

$$\mathbb{P}\{\xi_i = 0\} := q; i = 1, \ldots, n.$$

Khi đó, biến ngẫu nhiên ξ_i được coi như kết quả của phép thử thứ i (hoặc lần thứ i). Ta đặt $S_0 := 0$;

$$S_k(\omega) := \xi_1 + \cdots + \xi_k; \text{ với } k = 1, \ldots, n.$$

Từ đó suy ra

$$\begin{aligned}\mathbb{E}S_n &= \mathbb{E}(\xi_1 + \cdots + \xi_n)\\ &= \mathbb{E}(\xi_1) + \cdots + \mathbb{E}(\xi_n);\end{aligned}$$

$$\Rightarrow \mathbb{E}\frac{S_n}{n} = p. \tag{1.1}$$

Mặt khác, giá trị trung bình của số lần thử thành công (nghĩa là $\frac{S_n}{n}$) trùng với xác suất thành công p.

Trước hết, chúng ta lưu ý, ta không thể cho rằng: với $\varepsilon > 0$ đủ nhỏ và với n đủ lớn thì độ lệch của $\frac{S_n}{n}$ với p là bé hơn $\varepsilon, \forall\omega$. Nghĩa là

$$\left|\frac{S_n(\omega)}{n} - p\right| < \varepsilon, \forall\omega \in \Omega. \tag{1.2}$$

Thực vậy, với $0 < p < 1$:

$$\begin{aligned}\mathbb{P}\Big\{\frac{S_n}{n} := 1\Big\} &= \mathbb{P}\{\xi_1 = 1, \ldots, \xi_n = 1\} := p^n;\\ \mathbb{P}\Big\{\frac{S_n}{n} := 0\Big\} &= \mathbb{P}\{\xi_1 = 0, \ldots, \xi_n = 0\} := q^n.\end{aligned}$$

Từ đó cho thấy rằng, (1.2) không thỏa mãn với ε đủ nhỏ.

Tuy nhiên, với n đủ lớn ta sẽ thu được xác suất của biến cố $\frac{S_n}{n} = 1$ và $\frac{S_n}{n} = 0$ là nhỏ. Do đó, một cách tự nhiên cho rằng xác suất của biến cố

$$\left|\frac{S_n}{n} - p\right| > \varepsilon$$

cũng sẽ nhỏ với n đủ lớn.

Từ đó, ta đánh giá xác suất của biến cố

$$\Big\{\omega : \left|\frac{S_n}{n} - p\right| > \varepsilon\Big\}.$$

Với mục đích đó, ta phải sử dụng bất đẳng thức sau:

Mệnh đề 1.1.1 (Bất đẳng thức Chebysev)**.** Cho $(\Omega, \mathcal{A}, \mathbb{P})$ là không gian xác suất và $\xi = \xi(\omega)$ là biến ngẫu nhiên không âm thì

$$\mathbb{P}\{\xi \geqslant \varepsilon\} \leqslant \frac{\mathbb{E}\xi}{\varepsilon},\ \forall\varepsilon > 0. \tag{1.3}$$

Chứng minh. Ta có

$$\xi = \xi \cdot \mathbb{I}\{\xi \geqslant \varepsilon\} + \xi \cdot \mathbb{I}\{\xi < \varepsilon\} \geqslant \xi \cdot \mathbb{I}\{\xi \geqslant \varepsilon\} \geqslant \xi \cdot \mathbb{I}\{\xi \geqslant \varepsilon\},$$

trong đó $\mathbb{I}(A)$ là hàm chỉ tiêu của A. Lấy kỳ vọng toán hai vế:

$$\mathbb{E}\xi \geqslant \varepsilon \cdot \mathbb{P}\{\xi \geqslant \varepsilon\}.$$

□

Hệ quả 1.1.2. Nếu ξ là biến ngẫu nhiên bất kỳ, với ε, ta có

$$\begin{aligned}
&\mathbb{P}\{|\xi| \geqslant \varepsilon\} \leqslant \frac{\mathbb{E}|\xi|}{\varepsilon}.\\
&\mathbb{P}\{|\xi| \geqslant \varepsilon\} = \mathbb{P}\{|\xi|^2 \geqslant \varepsilon^2\} \leqslant \frac{\mathbb{E}|\xi|^2}{\varepsilon^2}.\\
&\mathbb{P}\{|\xi - \mathbb{E}\xi| \geqslant \varepsilon\} \leqslant \frac{\mathbb{D}\xi}{\varepsilon^2}.
\end{aligned} \tag{1.4}$$

Trong bất đẳng thức cuối cùng của (1.4) thay $\xi = \frac{S_n}{n}$, ta được

$$\mathbb{P}\Big\{\Big|\frac{S_n}{n} - p\Big| \geqslant \varepsilon\Big\} \leqslant \frac{\mathbb{D}(\frac{S_n}{n})}{\varepsilon^2} = \frac{pq}{n\varepsilon^2}.$$

Do đó

$$\mathbb{P}\Big\{\Big|\frac{S_n}{n} - p\Big| \geqslant \varepsilon\Big\}\frac{pq}{n\varepsilon^2} \leqslant \frac{1}{4n\varepsilon^2}. \tag{1.5}$$

Từ đó cho thấy rằng: với n đủ lớn, xác suất để $\frac{S_n}{n}$ lệch so với xác suất p lớn hơn ε là rất nhỏ.

Với $n \geqslant 1$ và $0 \leqslant k \leqslant n$, ta viết

$$\mathbb{P}_n(k) = \binom{n}{k} \cdot p^k \cdot q^{(n-k)}$$

thì

$$\mathbb{P}\Big\{\Big|\frac{S_n}{n} - p\Big| \geqslant \varepsilon\Big\} = \sum_{\{k:|\frac{k}{n}-p|\geqslant\varepsilon\}} \mathbb{P}_n(k).$$

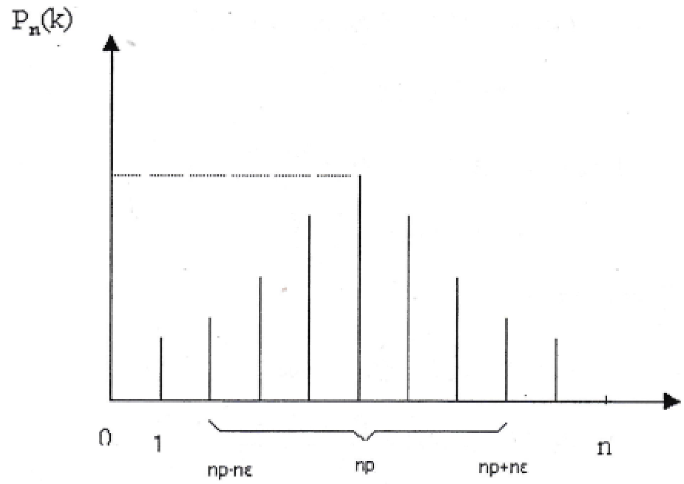
Từ đây ta thu được

$$\sum_{\{k:|\frac{k}{n}-p|\geqslant\varepsilon\}} \mathbb{P}_n(k) \leqslant \frac{pq}{n\varepsilon^2} \leqslant \frac{1}{4n\varepsilon^2} \tag{1.6}$$

Từ (1.6) ta thấy

$$\sum_{\{k:|\frac{k}{n}-p|\geqslant\varepsilon\}} \mathbb{P}_n(k) \to 0,\ (n \to \infty). \tag{1.7}$$

Ta có thể phân tích hình vẽ trên như sau.



Hình 1.1

Ta cùng biểu diễn phân phối nhị thức $\{\mathbb{P}_n(k), 0 \leqslant k \leqslant n\}$ như Hình 1.1. Khi n tăng thì hình vẽ được mở rộng và trở nên bằng phẳng hơn, tại thời điểm

$$\sum_{\{k:|\frac{k}{n}-p|\geqslant\varepsilon\}} \mathbb{P}_n(k) \to 1$$

với $np - n\varepsilon \leqslant k \leqslant np + n\varepsilon$.

Xét dãy các biến ngẫu nhiên: $S_0, \ldots, S_n$ như những đường riêng không ổn định thì (1.7) được giải thích như sau:

Khi n đủ lớn, điểm S_n xác định vị trí của phần tử tại thời điểm n nằm trong khoảng $[n(p-\varepsilon), n(p+\varepsilon)]$ với xác suất tương đối lớn (xấp xỉ 1). Xem Hình 1.2.

Ta viết lại dưới dạng sau:

$$\mathbb{P}\Big\{\Big|\frac{S_n}{n} - p\Big| \geqslant \varepsilon\Big\} \to 0, \quad (n \to \infty). \tag{1.8}$$

Ta thấy, biểu thức (1.8) thực sự được xác định chỉ nếu $\mathbb{P}$ là xác suất trên không gian $(\Omega, \mathcal{A})$, ở trên đó có nhiều dãy vô hạn các biến ngẫu nhiên độc lập có phân phối Bernoulli: $\xi_1, \xi_2, \ldots$ được định nghĩa. Các không gian như thế được xây dựng và (1.8) được xác định theo ý nghĩa xác suất hoàn toàn chặt chẽ.

Theo trên, ta thấy ngoài công cụ giải tích thì sử dụng ngôn ngữ của lý thuyết xác suất, ta chỉ chứng minh được như sau:

Cho $(\Omega^{(n)}, \mathcal{A}^{(n)}, \mathbb{P}^{(n)}), n \geqslant 1$ là dãy có phân phối Bernoulli sao cho

$$\Omega^{(n)} = \{\omega^{(n)} : \omega^{(n)} = (a_1^{(n)}, \ldots, a - n^{(n)}); a_i^{(n)} := 0, 1\},$$