ĐẠI HỌC THÁI NGUYÊN
TRƯỜNG ĐẠI HỌC KHOA HỌC

BÙI THỊ LAN ANH

VỀ LUẬT SỐ LỚN VÀ MỘT SỐ ỨNG 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 viết tắt

$\mathcal{A}, \mathcal{F}$	σ-đại số
$\mathcal{B}(X)$	σ-đại số Borel của X
$\mathbb{P}$	Độ đo xác suất
$(\Omega, \mathcal{F}, P)$	Không gian xác suất
Leb	Độ đo Lebesgue
ξ	Hàm $\mathcal{F}$-đo được hay biến ngẫu nhiên (BNN) hay đại lượng ngẫu nhiên (ĐLNN)
F_ξ	Hàm phân phối của ξ
f_ξ	Hàm mật độ của ξ
$\mathbb{E}(\xi)$	Kỳ vọng của ξ
$D(\xi)$	Phương sai của ξ
1_A	Hàm chỉ tiêu của tập A
L^2	Họ các biến ngẫu nhiên bình phương khả tích
C	Phép thử ngẫu nhiên
$n(A)$	Số lần xảy ra phép thử A
$\xrightarrow{P}$	Hội tụ theo xác suất
L_1	Không gian các hàm đo được
$\mathcal{L}(X)$	Luật của X
φ_F	Phép biến đổi Fourier của F
S_n	Tổng riêng của dãy biến ngẫu nhiên
$\varphi_n^*(t)$	Hàm đặc trưng của S_n
h.c.c	Hầu chắc chắn

Mở đầu

P.S. Laplace (1812) đã từng nói "Phần lớn những vấn đề quan trọng của cuộc sống thực ra chỉ là những bài toán xác suất". Xác suất hay giải tích ngẫu nhiên là một nhánh của toán học nghiên cứu các hiện tượng ngẫu nhiên, nghĩa là nghiên cứu các hiện tượng không thể nói trước nó xảy ra hay không xảy ra khi thực hiện một lần quan sát, nhưng nếu tiến hành quan sát nhiều lần ta có thể rút ra kết luận khoa học về hiện tượng đó. Do yếu tố ngẫu nhiên mà lý thuyết xác suất ứng dụng, từ văn học, vật lý, thị trường chứng khoán, dự báo thời tiết, kinh tế y học,... Luật số lớn là một kết quả của xác suất có nhiều ứng dụng và là cơ sở cho nhiều hiện tượng thống kê và thực tiễn.

Trong thực tế, để xác định giá trị của một biến ngẫu nhiên nào đó người ta thường tiến hành n lần quan sát (đo đạc) một cách độc lập và lấy trung bình cộng các kết quả đo ấy làm giá trị ước lượng cho giá trị cần biết. Câu hỏi đặt ra là cơ sở nào để khẳng định việc xấp xỉ này là chấp nhận được (hay nói cách khác là sai số của việc xấp xỉ là bao nhiêu và cần thực hiện tối thiểu bao nhiêu phép thử để thu được sai số không vượt quá một giá trị cho trước).

Câu trả lời là luật số lớn, hay cụ thể hơn là bất đẳng thức Chebysev. Tuy nhiên, một vấn đề đặt ra là: Nếu áp dụng bất đẳng thức Chebysev trong bài toán chọn cỡ mẫu thì số phép thử quá lớn. Lý do cơ bản là đánh giá sai số trong bài toán cỡ mẫu dựa trên bất đẳng thức Chebysev chưa đủ chặt, từ đó tốc độ hội tụ thu được chưa thật chính xác. Vậy tốc độ hội tụ của luật số lớn như thế nào? Bất đẳng thức martingale sẽ giải quyết vấn đề đó. Nói cách khác, chọn cỡ mẫu theo bất đẳng thức ước lượng mũ là tối ưu hơn theo bất đẳng thức Chebysev.

Trong thực tế chúng ta thường chỉ gặp một số biến ngẫu nhiên có phân phối đặc biệt như phân phối nhị thức, phân phối Poisson, phân phối mũ, ... Đặc biệt là các biến ngẫu nhiên có phân phối chuẩn hoặc xấp xỉ chuẩn. Ví dụ: Xét một quần thể. Chọn ngẫu nhiên một nhóm gồm n cá thể (mẫu có kích thước n). Giả sử X là số cả thể có đặc tính A trong mẫu. Khi đó, người ta thấy rằng nếu n lớn thì X có phân phối xấp xỉ chuẩn. Cơ sở lí thuyết của hiện tượng đó chính là định lý giới hạn trung tâm. Áp dụng định lý giới hạn trung tâm cho bài toán chọn cỡ mẫu ta thấy cách chọn cỡ mẫu theo bất đẳng thức ước lượng mũ và theo định lý giới hạn trung tâm là tương đương.

Với mục đích hệ thống lại các kiến thức của xác suất, trình bày lại luật số lớn, bất đẳng thức Chebysev, bất đẳng thức Martingale, định lý giới hạn trung tâm và áp dụng cho bài toán chọn cỡ mẫu trong thống kê, dưới sự hướng dẫn của TS. Trần Xuân Quý và TS. Đỗ Thị Phương Quỳnh, tôi chọn đề tài "*Về luật số lớn và một số ứng dụng*" để làm đề tài luận văn thạc sĩ.

Ngoài phần Mở đầu, Kết luận và Tài liệu tham khảo, luận văn gồm hai chương.

Chương 1. Một số kiến thức chuẩn bị. Chương này nhắc lại một số định nghĩa cơ bản trong lý thuyết xác suất như biến cố và xác suất, biến ngẫu nhiên, xác suất có điều kiện, ba dạng hội tụ của dãy biến ngẫu nhiên.

Chương 2. Về luật số lớn và một số áp dụng. Chương này trình bày về các luật số lớn, bất đẳng thức Chebysev, bất đẳng thức martingale và định lý giới hạn trung tâm, so sánh ba cách chọn cỡ mẫu và ưu điểm, nhược điểm của từng cách chọn cỡ mẫu.

Luận văn được hoàn thành tại trường Đại học Khoa học - Đại học Thái Nguyên dưới sự hướng dẫn của TS. Trần Xuân Quý, Trường Đại học Khoa học - Đại học Thái Nguyên.

Tôi xin bày tỏ lòng biết ơn sâu sắc nhất tới - TS. Trần Xuân Quý và TS. Đỗ Thị Phương Quỳnh, người đã định hướng chọn đề tài và tận tình hướng dẫn để tôi hoàn thành luận văn này.

Tôi xin bày tỏ lòng biết ơn chân thành tới Phòng Đào tạo, các thầy cô giáo dạy cao học chuyên ngành Toán ứng dụng, trường Đại học Khoa học - Đại học Thái Nguyên đã giúp đỡ tôi trong suốt quá trình học tập và hoàn thành luận văn tốt nghiệp.

Tôi xin được gửi lời cảm ơn chân thành tới gia đình, bạn bè, người thân đã luôn động viên, cổ vũ, tạo mọi điều kiện thuận lợi cho tôi trong quá trình học tập và hoàn thành luận văn.

Thái Nguyên, ngày 05 tháng 11 năm 2020

Tác giả

Bùi Thị Lan Anh

Chương 1

Một số kiến thức chuẩn bị

Trong chương này chúng ta nhắc lại một vài định nghĩa cơ bản về lý thuyết xác suất. Cụ thể ta nhắc lại một số vấn đề sau:

(i). Không gian xác suất, $\sigma-$ đại số và độ đo;

(ii). Biến ngẫu nhiên và hàm phân phối của chúng;

(iii). Kì vọng và phương sai;

(iv). $\sigma-$ đại số sinh bởi biến ngẫu nhiên;

(v). Tính độc lập, xác suất có điều kiện;

(vi). Các dạng hội tụ của dãy các biến ngẫu nhiên.

1.1 Biến cố và xác suất

Định nghĩa 1.1.1. Cho Ω là tập khác rỗng. Một $\sigma-$ đại số $\mathcal{F}$ trên Ω là họ các tập hợp của Ω sao cho

1) Tập $\emptyset \in \mathcal{F}$;

2) Nếu $A \in \mathcal{F}$ thì phần bù $\overline{A} \in \mathcal{F}$;

3) Nếu $A_1, A_2, \ldots$ là dãy đếm được các tập hợp trong $\mathcal{F}$ thì hợp của chúng $A_1 \cup A_2 \cup \cdots$ cũng thuộc $\mathcal{F}$.

Ví dụ 1.1.2. $\mathbb{R}$ được định nghĩa là tập hợp các số thực. Họ các tập Borel $\mathcal{F} = \mathcal{B}(\mathbb{R})$ là $\sigma-$ đại số trên $\mathbb{R}$ trong đó $\mathcal{B}(\mathbb{R})$ là $\sigma-$ đại số chứa tất cả các đoạn trên $\mathbb{R}$.

Định nghĩa 1.1.3. Cho $\mathcal{F}$ là một $\sigma-$ đại số trên Ω. Độ đo xác suất P là hàm $P : \mathcal{F} \longrightarrow [0, 1]$ sao cho

1) $\mathbb{P}(\Omega) = 1$;

2) Nếu $A_1, A_2, \ldots$ là tập rời nhau từng đôi một (nghĩa là $A_i \cap A_j = \emptyset$ với $i \neq j) \subset \mathcal{F}$ thì

$$\mathbb{P}(A_1 \cup A_2 \cup \ldots) = \mathbb{P}(A_1) + \mathbb{P}(A_2) + \cdots .$$

$(\Omega, \mathcal{F}, P)$ được gọi là không gian xác suất. Tập hợp thuộc $\mathcal{F}$ được gọi là biến cố. Biến cố A xảy ra hầu chắc chắn khi $\mathbb{P}(A) = 1$.

Ví dụ 1.1.4. Chúng ta đưa ra khoảng cách có độ dài bằng một đơn vị $\Omega = [0, 1]$ với $\sigma-$ đại số $\mathcal{F} = \mathcal{B}([0, 1])$ là tập hợp các tập Borel $B \subset [0, 1]$ và độ đo Lebesgue $P = Leb$ trên $[0, 1]$. Khi đó $(\Omega, \mathcal{F}, P)$ là một không gian xác suất. Nhắc lại rằng Leb là độ đo duy nhất được định nghĩa trên tập Borel sao cho với bất kì $[a, b]$

$$Leb[a, b] = b - a.$$

Định lý 1.1.5. Nếu $A_1, A_2, \ldots$ là dãy tăng các biến cố, nghĩa là

$$A_1 \subset A_2 \subset \cdots$$

thì $\mathbb{P}(A_1 \cup A_2 \cup \cdots) = \lim_{n\to\infty} \mathbb{P}(A_n)$.
Tương tự, nếu $A_1, A_2, \ldots$ là dãy giảm các biến cố, nghĩa là

$$A_1 \supset A_2 \supset \cdots$$

thì $\mathbb{P}(A_1 \cap A_2 \cap \ldots) = \lim_{n\to\infty} \mathbb{P}(A_n)$.

Chứng minh. Nếu $A_1 \subset A_2 \subset \ldots$ thì

$$A_1 \cup A_2 \cup \ldots = A_1 \cup (A_2 \setminus A_1) \cup (A_3 \setminus A_2) \cup \ldots .$$

Trong đó, các tập $A_1, A_2 \setminus A_1, A_3 \setminus A_2, \ldots$ rời nhau từng đôi một. Do đó, theo định nghĩa của độ đo xác suất

$$\begin{aligned}
\mathbb{P}(A_1 \cup A_2 \cup \cdots) &= \mathbb{P}(A_1 \cup (A_2 \setminus A_1) \cup (A_3 \setminus A_2) \cup \cdots) \\
&= \mathbb{P}(A_1) + \mathbb{P}(A_2 \setminus A_1) + \mathbb{P}(A_3 \setminus A_2) + \cdots \\
&= \lim_{n\to\infty} \mathbb{P}(A_n).
\end{aligned}$$

Ta có

$$\begin{aligned}\mathbb{P}(A_1 \cup A_2 \cup \cdots + A_n) &= \mathbb{P}(A_1 \cup (A_2 \setminus A_1) \cup (A_3 \setminus A_2) \cup \cdots + \mathbb{P}(A_n \setminus A_{n-1})\\ &= \mathbb{P}(A_1) + \mathbb{P}(A_2) - \mathbb{P}(A_1) + \cdots + \mathbb{P}(A_n) - \mathbb{P}(A_{n-1})\\ &= \mathbb{P}(A_n).\end{aligned}$$

Nếu $A_1 \supset A_2 \supset \cdots$ thì

$$\mathbb{P}(A_1 \cap A_2 \cap \cdots) = \lim_{n\to\infty} \mathbb{P}(A_n).$$

Áp dụng luật De Morgan ta có

$$\Omega \setminus (A_1 \cap A_2 \cap \cdots) = (\Omega \setminus A_1) \cup (\Omega \setminus A_2) \cup \cdots .$$

□

Bổ đề 1.1.6 (Borel- Cantelli)**.** Cho $A_1, A_2, \ldots$ là dãy các biến cố sao cho

$$\mathbb{P}(A_1) + \mathbb{P}(A_2) + \cdots < \infty$$

và đặt $B_n = A_n \cup A_{n+1} \cup \cdots$ thì

$$\mathbb{P}(B_1 \cap B_2 \cap \cdots) = 0.$$

Chứng minh. Vì B_n là dãy giảm các biến cố, theo kết quả Định lý 1.1.5 suy ra rằng

$$\begin{aligned}\mathbb{P}(B_1 \cap B_2 \cap \cdots) &= \lim_{n\to\infty} \mathbb{P}(B_n) = \lim_{n\to\infty} \mathbb{P}(A_n \cup A_{n+1} \cup \cdots)\\ &\le \lim_{n\to\infty} [\mathbb{P}(A_n) + \mathbb{P}(A_{n+1}) + \cdots] = 0.\end{aligned}$$

Đẳng thức cuối cùng đúng do chuỗi $\sum_{n=1}^{\infty} \mathbb{P}(A_n)$ là hội tụ. Bất đẳng thức trên đúng do tính chất cộng tính dưới

$$\mathbb{P}(A_n \cup A_{n+1} \cup \ldots) \le \mathbb{P}(A_n) + \mathbb{P}(A_{n+1}) + \cdots .$$

Suy ra $\mathbb{P}(B_1 \cap B_2 \cap \cdots) = 0$. □

1.2 Biến ngẫu nhiên

Định nghĩa 1.2.1. Nếu $\mathcal{F}$ là $\sigma-$ đại số trên Ω thì hàm $\xi : \Omega \longrightarrow \mathbb{R}$ được gọi là $\mathcal{F}-$ đo được nếu $\{\xi \in B\} \in \mathcal{F}$ với mỗi tập Borel $B \in \mathcal{B}(\mathbb{R})$. Nếu $(\Omega, \mathcal{F}, P)$ là không gian xác suất thì hàm ξ được gọi là biến ngẫu nhiên.

Chú ý 1.2.2. Để cho ngắn gọn, ta ký hiệu $\{\xi \in B\}$ thay vì viết

$$\{\omega \in \Omega : \xi(\omega) \in B\}.$$

Định nghĩa 1.2.3. $\sigma-$ đại số $\sigma(\xi)$ sinh bởi biến ngẫu nhiên $\xi : \Omega \longrightarrow \mathbb{R}$ được định nghĩa là lớp tất cả các tập có dạng $\{\omega \in \Omega : \xi(\omega) \in B\}$, trong đó B là tập Borel trong $\mathbb{R}$.

Định nghĩa 1.2.4. $\sigma-$ đại số $\sigma(\{\xi_i : i \in I\})$ sinh bởi họ các biến ngẫu nhiên $\{\xi_i : i \in I\}$ được định nghĩa là $\sigma-$ đại số nhỏ nhất chứa tất cả biến cố có dạng $\{\omega \in \Omega : \xi_i(\omega) \in B\}$ trong đó B là tập Borel trong $\mathbb{R}$ và $i \in I$.

Nhận xét 1.2.5. Ta gọi $f : \mathbb{R} \longrightarrow \mathbb{R}$ là hàm Borel nếu nghịch ảnh $f^{-1}(B)$ với mọi tập Borel B trong $\mathbb{R}$ là tập Borel. Nếu f là hàm Borel và ξ là biến ngẫu nhiên thì $f(\xi)$ là $\sigma(\xi)-$ đo được.

Thật vậy, nếu B là tập Borel trong $\mathbb{R}$ và $f : \mathbb{R} \longrightarrow \mathbb{R}$ là hàm Borel thì $f^{-1}(B)$ cũng là tập Borel. Do đó

$$\{f(\xi) \in B\} = \{\xi \in f^{-1}(B)\}$$

thuộc $\sigma-$ đại số $\sigma(\xi)$ sinh bởi ξ. Vậy $f(\xi)$ là $\sigma(\xi)-$ đo được.

Bổ đề 1.2.6. (Doob - Dynkin)
Cho ξ là biến ngẫu nhiên. Khi đó mỗi biến ngẫu nhiên $\sigma(\xi)-$ đo được η có thể viết $\eta = f(\xi)$ với $f : \mathbb{R} \longrightarrow \mathbb{R}$ là hàm Borel.

Định nghĩa 1.2.7. Giả sử $\xi : \Omega \longrightarrow \mathbb{R}$ là biến ngẫu nhiên, xác định độ đo xác suất như sau

$$\mathbb{P}_\xi(B) = \mathbb{P}\{\omega \in \Omega : \xi(\omega) \in B\}.$$

Trên $\mathbb{R}$ xác định $\sigma-$ đại số của tập Borel $B \in \mathcal{B}(\mathbb{R})$. Ta định nghĩa F_ξ là hàm phân phối của ξ, ký hiệu $F_\xi : \mathbb{R} \longrightarrow [0, 1]$ xác định bởi

$$F_\xi(x) = \mathbb{P}\{\omega \in \Omega : \xi(\omega) \leq x\}.$$

Nhận xét dưới đây cho ta biết được một số tính chất của hàm phân phối của biến ngẫu nhiên.

Nhận xét 1.2.8. Hàm phân phối F_ξ là không giảm, liên tục phải và thỏa mãn

$$\lim_{x\to-\infty} F_\xi(x) = 0, \quad \lim_{x\to+\infty} F_\xi(x) = 1.$$

Thật vậy, nếu $x \leq y$ thì $\{\omega \in \Omega : \xi(\omega) \leq x\} \subset \{\omega \in \Omega : \xi(\omega) \leq y\}$. Do đó

$$F_\xi(x) = \mathbb{P}\{\omega \in \Omega : \xi(\omega) \leq x\} \leq \mathbb{P}\{\omega \in \Omega : \xi(\omega) \leq y\} = F_\xi(y).$$

Điều đó có nghĩa là F_ξ không giảm.

Tiếp theo, ta lấy dãy bất kỳ $x_1 \geq x_2 \geq \ldots$ và đặt

$$\lim_{n\to\infty} x_n = x.$$

Khi đó, ta có dãy biến cố

$$\{\omega \in \Omega : \xi(\omega) \leq x_1\} \supset \{\omega \in \Omega : \xi(\omega) \leq x_2\} \supset \cdots .$$

Lấy giao ta được

$$\{\omega \in \Omega : \xi(\omega) \leq x\} = \{\omega \in \Omega : \xi(\omega) \leq x_1\} \cap \{\omega \in \Omega : \xi(\omega) \leq x_2\} \cap \cdots .$$

Từ Định lý 1.1.5 ta suy ra rằng

$$F_\xi(x) = P\{\omega \in \Omega : \xi(\omega) \leq x\} = \lim_{n\to\infty} \mathbb{P}\{\omega \in \Omega : \xi(\omega) \leq x_n\} = \lim_{n\to\infty} F_\xi(x_n).$$

Điều đó chứng tỏ rằng F_ξ là liên tục giảm. Do các biến cố

$$\{\omega \in \Omega : \xi(\omega) \leq -1\} \supset \{\omega \in \Omega : \xi(\omega) \leq -2\} \supset \cdots$$

là dãy giảm với giao bằng $\emptyset$ và

$$\{\omega \in \Omega : \xi(\omega) \leq 1\} \subset \{\omega \in \Omega : \xi(\omega) \leq 2\} \subset \cdots$$

là dãy tăng với hợp bằng Ω. Theo Định lý 1.1.5 ta có

$$\lim_{x\to\infty} F_\xi(x) = \lim_{n\to\infty} F_\xi(-n) = \lim_{n\to\infty} \mathbb{P}\{\omega \in \Omega : \xi(\omega) \leq -n\} = \mathbb{P}(\emptyset) = 0,$$

$$\lim_{x\to\infty} F_\xi(x) = \lim_{n\to\infty} F_\xi(n) = \lim_{n\to\infty} \mathbb{P}\{\omega \in \Omega : \xi(\omega) \leq n\} = \mathbb{P}(\Omega) = 1.$$

Vì F_ξ là hàm không giảm.

Định nghĩa 1.2.9. Nếu hàm Borel $f_\xi : \mathbb{R} \longrightarrow \mathbb{R}$ sao cho với bất kỳ tập Borel $B \in \mathbb{R}$

$$\mathbb{P}\{\omega \in \Omega : \xi(\omega) \in B\} = \int_B f_\xi(x)dx$$