Giáo trình Lý thuyết Xác suất

TS. Phạm Việt Hùng

12/2020

Mục lục

Chương 1

Không gian xác suất

1.1 Giới thiệu

Một trong những nguồn gốc của Toán học đó là việc mô hình hóa các sự vật, hiện tượng trong tự nhiên và đời sống. Ta biết rằng các số tự nhiên sinh ra từ việc đếm; dãy số Fibonaci mô hình số cánh hoa hay sự tăng trưởng của đàn thỏ; các phương trình vi phân hay phương trình dạo hàm riêng mô tả sự tăng trưởng của dân số, sự phân rã của phóng xạ, quá trình truyền âm, truyền nhiệt, truyền sóng hay chuyển động của chất lỏng. Các mô hình như trên được gọi là mô hình **tất định** vì nếu cho trước một giá trị ban đầu, ta hoàn toàn có thể xác định chính xác giá trị tại thời điểm nào đó sau này. Ngoài ra, ta thấy rằng còn có nhiều sự vật, hiện tượng có tính **ngẫu nhiên**, với cùng một hành động ban đầu, ta không biết chắc chắn kết quả là gì, ví dụ như khi tung đồng xu hoặc xúc xắc. Những sự vật , hiện tượng có tính ngẫu nhiên như vậy là khởi đầu cho việc nghiên cứu Lý thuyết Xác suất trong Toán học.

Với khởi nguồn như trên, việc nghiên cứu xác suất đã sớm được quan tâm, đặc biệt cho các trò chơi có tính may rủi, đỏ đen và cá cược. Trong bài thơ "Bà lão[1]" vào thế kỉ 13, các tính toán xác suất khi tung xúc xắc đã được nhắc đến. Vào thế kỉ 15, Cardano đã viết trọn vẹn một cuốn sách "Các trò chơi may rủi" để trình bày các nghiên cứu về xác suất. Trong các thư từ trao đổi năm 1654, hai nhà toán học nổi tiếng Blaise Pascal và Pierre de Fermat đã thảo luận về một câu hỏi được nhà quý tộc và cũng là tay đánh bạc khét tiếng Chevalier de Méré gửi đến hỏi Pascal như sau:

"Tôi thường cá cược tiền của mình cho việc trong bốn lần tung đồng xu thì sẽ có ít nhất một lần ra được mặt Lục. Theo tôi, xác suất để xảy ra điều này bằng bốn lần xác suất ra

[1]Tên nguyên bản là "De Vetula", viết bằng tiếng Latin

mặt Lục của một lần tung, kết quả là $4/6 = 2/3$. *Như vậy tôi có lợi thế khi chơi và thực sự tôi đã kiếm được tiền thắng cược. Tuy nhiên, khi tôi đặt cược rằng trong 24 lần tung một cặp xúc xắc sẽ có ít nhất một lần tung ra cả hai mặt Lục, thì tôi lại bị mất tiền, dù theo tính toán của tôi, xác suất xảy ra điều này là* $24/6^2 = 2/3$ *đúng bằng cách chơi đầu tiên. Tại sao lại vậy?"*
Pascal và Fermat đã thống nhất rằng các kết quả của de Méré đều SAI! Kết quả đúng sau khi làm tròn lần lượt là 0.5177 và 0.4914, khá là gần xác suất của một trò chơi công bằng. Để de Méré có thể nhận ra mình thắng hay mất tiền, chứng tỏ ông ta phải chơi trò cá cược này rất nhiều lần!

Lý thuyết xác suất tiếp tục được nghiên cứu và hoàn thiện trong các cuốn sách "Van Rekeningh in Spelen van Gluck [2]" của Huyghens năm 1657, "Ars conjectandi" của Jacod Bernoulli năm 1713, "Doctrine of chances" của de Moivre năm 1718. Nhiều nhà toán học đã quan tâm và phát triển Lý thuyết xác suất như Laplace, Euler, Gauss, Lagrange, Legendre, Poisson, ...

Tuy nhiên, đến thế kỉ 19, khi đã được phát triển ở mức cao với nhiều kết quả sâu sắc, thì Lý thuyết xác suất gặp phải vấn đề giống như Toán học đương đại là sự thiếu hụt một nền tảng cơ sở vững chắc. Điều này được ví von giống như ta xây một lâu đài nguy nga tráng lệ trên nền cát, có thể bị sụp đổ bất cứ lúc nào. Để tìm hiểu sự thiếu hụt này ta tìm hiểu ví dụ sau:

Trong các bài toán xác suất cổ điển, số tình huống là hữu hạn nên ta có thể định nghĩa xác suất một cách trực quan và dễ dàng. Tuy nhiên, vấn đề sẽ nảy sinh khi tập các tình huống là vô hạn không đếm được. Cụ thể, ta lấy ra ngẫu nhiên một điểm trên đoạn thẳng đơn vị $[0, 1]$, tính xác suất để điểm chọn ra thuộc vào một tập nào đó. Một cách trực quan, nếu tập đang xét là một đoạn thẳng hoặc hợp của một số đoạn thẳng thì xác suất sẽ bằng độ dài của tập đó. Câu hỏi đặt ra là có thể tính xác suất cho tập tổng quát bất kì hay không? Nếu có thì trong trường hợp này khái niệm xác suất là một mở rộng của khái nhiệm độ dài đoạn thẳng, hay là độ đo Lebesgue một chiều. Nhưng ta biết rằng trong Lý thuyết độ đo thì không phải tập con nào của đoạn thẳng $[0, 1]$ cũng là tập đo được và tính được độ dài!

Đến nửa đầu thế kỉ 20, dựa trên Lý thuyết độ đo, Kolmogorov đã xây dựng hệ tiên đề cho không gian xác suất $(\Omega, \mathcal{F}, \mathbb{P})$ mà ta sẽ tìm hiểu ở mục sau. Đây được coi là dấu mốc cho sự bắt đầu của **Lý thuyết xác suất hiện đại**.

[2] Dịch sang tiếng Latin là "De ratiociniis in ludo aleae" và sang tiếng Anh là "On Reasoning in Games of Chance"

1.2 Không gian xác suất $(\Omega, \mathcal{F}, \mathbb{P})$

1.2.1 Định nghĩa

Xét Ω là một tập khác rỗng. Ta gọi Ω là *không gian mẫu* gồm tất cả các tình huống có thể xảy ra. Kí hiệu 2^Ω tập tất cả các tập con của Ω .

Định nghĩa 1.2.1. *Một tập $\mathcal{F} \subset 2^\Omega$ gồm một họ các tập con của Ω được gọi là một σ-*đại số *(hoặc σ-trường) nếu*

1. *$\emptyset \in \mathcal{F}$ và $\Omega \in \mathcal{F}$;*
2. *Nếu $A \in \mathcal{F}$ thì phần bù của A, kí hiệu A^c hoặc $\overline{A} = \Omega \backslash A$, cũng thuộc $\mathcal{F}$;*
3. *$\mathcal{F}$ đóng đối với phép hợp đếm được: tức là, với mọi dãy $A_i, i = 1, 2, \dots$ các phần tử của $\mathcal{F}$, ta có $\cup_{i\geq 1} A_i$ cũng thuộc $\mathcal{F}$.*

Ta có nhận xét là một σ-đại số đóng với phép giao đếm được. Tức là, với mọi dãy $A_i, i = 1, 2, \dots$ các phần tử của $\mathcal{F}$, ta có $\cap_{i\geq 1} A_i$ cũng thuộc $\mathcal{F}$. Tính chất này dễ dàng suy ra từ tính chất 3 và quy tắc De Morgan.

Ta có một số ví dụ cơ bản như sau về các σ-đại số trừu tượng.

Ví dụ 1.2.2.
1. *$\mathcal{F} = \{\emptyset, \Omega\}$: σ-đại số tầm thường.*
2. *Với mỗi tập con A của Ω, ta có σ-đại số $\sigma(A) = \{\emptyset, A, A^c, \Omega\}$. Đây là σ-đại số bé nhất chứa tập A.*
3. *$\mathcal{F} = 2^\Omega$: σ-đại số tuyệt đối.*

Mở rộng ví dụ 2 ở trên, ta có định nghĩa như sau.

Định nghĩa 1.2.3. *Giả sử $\mathcal{C} \subset 2^\Omega$,* σ-đại số sinh bởi $\mathcal{C}$, *kí hiệu là $\sigma(\mathcal{C})$ là σ-đại số bé nhất chứa $\mathcal{C}$.*

Chú ý là $\sigma(\mathcal{C})$ luôn tồn tại vì 2^Ω là một σ-đại số và giao của mỗi họ bất kì các σ-đại số cũng là một σ-đại số. Từ đây, ta có thể định nghĩa σ-đại số Borel trên $\mathbb{R}^d$ như sau. Đây là σ-đại số quan trọng và ta sẽ gặp lại trong Chương 2.

Định nghĩa 1.2.4. $\mathcal{B}(\mathbb{R}^d)$ *là σ-đại số Borel trên $\mathbb{R}^d$ là σ-đại số sinh bởi các tập có dạng $\prod_{i=1}^d (-\infty, a_i]$ với $a_i \in \mathbb{Q}$ với mọi $i = 1, \dots d$. Tương tự, $\mathcal{B}([0,1]^d)$ là σ-đại số Borel trên $\mathbb{R}^d$ là σ-đại số sinh bởi các tập có dạng $\prod_{i=1}^d (0, a_i]$ với $a_i \in \mathbb{Q}$ và $a_i \leq 1$ với mọi $i = 1, \dots d$.*

Theo đề xuất của Kolmogorov, ta có định nghĩa về một *không gian xác suất* như sau.

Định nghĩa 1.2.5. *Một bộ ba* $(\Omega, \mathcal{F}, \mathbb{P})$ *được gọi là một* không gian xác suất *gồm có không gian mẫu* Ω*, một* σ*-đại số* mf *và ánh xạ* $\mathbb{P} : \mathcal{A} \to [0, 1]$ *được gọi là một* độ đo xác suất *thỏa mãn*

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

(2) với mọi dãy gồm đếm được các tập con (A_i) *của* $\mathcal{F}$ *và đôi một rời nhau (tức là* $A_m \cap A_n = \emptyset \, \forall m \neq n$*) ta có*

$$\mathbb{P}\Big(\cup_{i=1}^{\infty} A_i \Big) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$$

Từ định nghĩa về không gian xác suất như trên, ta thấy rằng ta chỉ có thể gán giá trị - *xác suất* cho các tập A thuộc vào σ-đại số đang xét mf. Để phân biệt các tập đặc biệt như vậy, ta có định nghĩa cơ bản sau đây.

Định nghĩa 1.2.6. *Trong không gian xác suất* $(\Omega, \mathcal{F}, \mathbb{P})$*, một tập* $A \in \mathcal{F}$ *được gọi là một* biến cố *và giá trị* $\mathbb{P}(A)$ *được gọi là* xác suất của biến cố *A.*

Trong các bài toán xác suất, ta thường định nghĩa biến cố thông qua tập các tình huống có tính chất đặc biệt. Do đó đôi khi ta có thể đồng nhất giữa biến cố và tính chất. Với cách hiểu này, ta có thể diễn giải ý nghĩa của các phép toán trên tập hợp (biến cố) như sau:

- Biến cố đối $\overline{A}$: tập các tình huống không có tính chất A.
- Biến cố giao $A \cap B$: tập các tình huống có cả hai tính chất A và B.
- Biến cố hợp $A \cup B$: tập các tình huống thỏa mãn ít nhất một trong hai tính chất A và B.

Ta có một số ví dụ cho không gian xác suất như sau:

Ví dụ 1.2.7. *1. Tung ngẫu nhiên một đồng xu cân bằng. Có hai khả năng xảy ra là ngửa (H, head) hoặc sấp (T, tail). Khi đó,* $\Omega = \{H, T\}, \mathcal{F} = 2^{\Omega}$ *và*

$$\mathbb{P}(\{H\}) = \mathbb{P}(\{T\}) = 1/2.$$

Với đồng xu không cân bằng, thì tồn tại $p \in [0, 1]$ *sao cho*

$$\mathbb{P}(\{H\}) = p, \ \mathbb{P}(\{T\}) = 1 - p.$$

2. *Tung ngẫu nhiên một viên xúc xắc cân đối. Khi đó,* $\Omega = \{1, 2, 3, 4, 5, 6\}, \mathcal{F} = 2^\Omega$ *và với mọi biến cố* $A \in \mathcal{F}$ *thì* $\mathbb{P}(A) = \frac{|A|}{|\Omega|}$, *với* $|.|$ *là lực lượng của tập hợp. Ví dụ, xét biến cố* A *ra mặt chẵn thì* $A = \{2, 4, 6\}$, *và* $\mathbb{P}(A) = 3/6 = 0.5$.

3. *Không gian xác suất* $([0, 1], \mathcal{B}([0, 1]), \lambda)$ *với* λ *là độ đo Lebesgue một chiều xây dựng từ độ dài các đoạn thẳng.*

4. *Xét* $\Omega = \mathbb{N}$ *là tập các số tự nhiên và* $\mathcal{F} = 2^\Omega$. *Ta xây dựng độ đo xác suất trên* $(\mathbb{N}, 2^\mathbb{N})$ *từ một dãy các số không âm* $\{p_i\}_{i=0,1,...}$ *thỏa mãn* $\sum_{i=0}^{\infty} p_i = 1$ *như sau*

$$\mathbb{P}(A) = \sum_{i \in A} p_i, \ \forall A \in 2^\mathbb{N}.$$

Từ tính chất của σ-đại số, ta có kết quả sau.

Mệnh đề 1.2.8. *Trong không gian xác suất* $(\Omega, \mathcal{F}, \mathbb{P})$, *cho một dãy các biến cố* $A_1, A_2, \ldots$. *Khi đó hai tập được xác định như sau*

$$\limsup A_n = \cap_{n=1}^{\infty} \left(\cup_{m \geq n} A_m\right), \text{ và } \liminf A_n = \cup_{n=1}^{\infty} \left(\cap_{m \geq n} A_m\right)$$

cũng là các biến cố.

Ta có ý nghĩa của $\limsup A_n$ là các tình huống thuộc vô hạn biến cố, hay nói cách khác, các tính chất A_n xuất hiện vô hạn lần; còn ý nghĩa của $\liminf A_n$ là các tình huống sẽ luôn xuất hiện từ một thời điểm nào đó. Theo quy tắc DeMorgan,

$$\overline{\limsup A_n} = \liminf \overline{A}_n.$$

1.2.2 Một số tính chất của độ đo xác suất

Một độ đo xác suất có một số tính chất như sau

Định lý 1.2.9. *Nếu* $\mathbb{P}$ *là độ đo xác suất trên* $(\Omega, \mathcal{F})$ *thì*

1. $\mathbb{P}(\emptyset) = 0$;

2. $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$ *với mọi* $A \in \mathcal{F}$;

3. *Với hai biến cố rời nhau* $A, B \in \mathcal{F}$ *thì*

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B).$$

4. *Với hai biến cố bất kì $A, B \in \mathcal{F}$ thì*

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B).$$

5. *Tính đơn điệu: nếu $A, B \in \mathcal{F}$ và $A \subset B$ thì $\mathbb{P}(A) \leq \mathbb{P}(B)$;*

6. *Tính liên tục: với mọi dãy biến cố tăng $A_1 \subset A_2 \subset \ldots \subset A_n \subset \ldots$ thì*

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_n\right) = \lim_{n\to\infty} \mathbb{P}(A_n).$$

Và với mọi dãy giảm các biến cố $A_1 \supset A_2 \supset \ldots \supset A_n \supset \ldots$ thì

$$\mathbb{P}\left(\bigcap_{i=1}^{\infty} A_n\right) = \lim_{n\to\infty} \mathbb{P}(A_n).$$

Mở rộng tính chất thứ 4 của Định lý 1.2.9, ta có *Nguyên lý bù trừ* (Inclusion-Exclusion Principle) như sau:

Định lý 1.2.10. *Cho họ n biến cố $A_1, A_2, \ldots, A_n$. Khi đó,*

$$\begin{aligned}\mathbb{P}(A_1 \cup A_2 \cup \ldots \cup A_n) = &\sum_{i=1}^{n} \mathbb{P}(A_i) - \sum_{i\neq j=1}^{n} \mathbb{P}(A_i \cap A_j) + \ldots \\ &+(-1)^k \sum_{i_1\neq i_2\neq \ldots i_k=1}^{n} \mathbb{P}(A_{i_1} \cap A_{i_2) \ldots \cap A_{i_k}}) + \ldots + (-1)^n \mathbb{P}(\bigcap_{i=1}^{n} A_i).\end{aligned}$$

Từ tính chất thứ 4 của Định lý 1.2.9, ta thấy rằng $\mathbb{P}(A \cup B) \leq \mathbb{P}(A) + \mathbb{P}(B)$. Mở rộng cho họ nhiều biến cố, ta có bất đẳng thức Boolean hay còn gọi là chặn trên hợp (Union bound) như sau.

Định lý 1.2.11. *Cho họ biến cố $A_1, A_2, \ldots$. Khi đó,*

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_n\right) \leq \sum_{i=1}^{\infty} \mathbb{P}(A_n).$$

Tính chất thứ hai mà độ đo xác suất $\mathbb{P}$ cần thỏa mãn trong hệ tiên đề Kolmogorov như trong Định nghĩa 1.2.5 được gọi là tính σ-cộng tính. Để kiểm tra điều kiện σ-cộng tính, ta thường kiểm tra tính cộng tính và tính liên tục như trong định lý sau.

Định lý 1.2.12. *Giả sử $\mathcal{F}$ là một σ-đại số và $\mathbb{P} : \mathcal{F} \to [0, 1]$ thỏa mãn $\mathbb{P}(\Omega) = 1$ và cộng tính theo nghĩa là với mọi hai biến cố rời nhau $A, B \in \mathcal{F}$ thì*

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B).$$

. Khi đó các khẳng định sau là tương đương:

(i) $\mathbb{P}$ *là* σ*-cộng tính.*

(ii) Nếu $A_n \in \mathcal{F}$ *và* $A_1 \supset A_2 \supset \ldots \supset A_n \supset \ldots$ *và* $\bigcap\limits_{i=1}^{\infty} A_n = \emptyset$ *thì* $\mathbb{P}(A_n) \downarrow 0$*;*

(iii) Nếu $A_n \in \mathcal{F}$ *và* $A_1 \supset A_2 \supset \ldots \supset A_n \supset \ldots$ *thì*

$$\mathbb{P}\left(\bigcap_{i=1}^{\infty} A_n\right) = \lim_{n\to\infty} \mathbb{P}(A_n).$$

(iv) Nếu $A_n \in \mathcal{F}$ *và* $A_1 \subset A_2 \subset \ldots \subset A_n \subset \ldots$ *và* $\bigcup\limits_{i=1}^{\infty} A_n = \Omega$ *thì* $\mathbb{P}(A_n) \uparrow 1$*;*

(v) Nếu $A_n \in \mathcal{F}$ *và* $A_1 \subset A_2 \subset \ldots \subset A_n \subset \ldots$ *thì*

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_n\right) = \lim_{n\to\infty} \mathbb{P}(A_n).$$

Trong phần cuối của Mục này, chúng ta tìm hiểu về cấu trúc của một σ-đại số và tính duy nhất của độ đo xác suất xác định trên *sm*-đại số cảm sinh. Đây là các kết quả cổ điển trong Lý thuyết độ đo.

Định nghĩa 1.2.13. *Một họ* $\mathcal{P}$ *các tập con của* Ω *được gọi là* π-hệ *nếu họ khác rỗng và đóng với phép giao hữu hạn, tức là nếu* $A, B \in \mathcal{P}$ *thì* $A \cap B \in$
$mathcalP$*. Một họ* $\mathcal{L}$ *các tập con của* Ω *được gọi là* λ-hệ *nếu thỏa mãn các điều kiện*

i. $\Omega \in \mathcal{L}$*.*

ii. Nếu $A, B \in \mathcal{L}$ *và* $A \subset B$ *thì* $B \setminus A \in \mathcal{L}$*.*

iii. Với mọi dãy các tập con tăng dần $A_n \uparrow A$ *thì* $A \in \mathcal{L}$*.*

Mệnh đề 1.2.14. *Một họ* $\mathcal{F}$ *gồm các tập con của* Ω *là một* σ*-đại số khi và chỉ khi* $\mathcal{F}$ *đồng thời là* π*-hệ và* λ*-hệ.*

Định lý 1.2.15 (Định lý $\pi - \lambda$ của Dynkin)**.** *Nếu* $\mathcal{P} \subset \mathcal{L}$ *và* $\mathcal{P}$ *là một* π*-hệ, còn* $\mathcal{L}$ *là một* λ*-hệ, thì* $\sigma(\mathcal{P}) \subset \mathcal{L}$*.*

Chứng minh. Gọi $\mathcal{L}(\mathcal{P})$ là một λ-hệ bé nhất chứa $\mathcal{P}$. Ta sẽ chứng minh rằng $\mathcal{L}(\mathcal{P})$ cũng là một π-hệ. Từ đó suy ra $\mathcal{L}(\mathcal{P})$ là một σ-đại số và ta có

$$mathcalP \subset \sigma(\mathcal{P}) \subset \mathcal{L}(\mathcal{P}) \subset \mathcal{L}.$$

Để chứng minh $\mathcal{L}(\mathcal{P})$ cũng là một π-hệ, ta xét một tập $A \in \mathcal{L}(\mathcal{P})$ và đặt

$$L_A = \{B \in \mathcal{L}(\mathcal{P}) :\ B \cap A \in \mathcal{L}(\mathcal{P})\}.$$

Ta có thể kiểm tra L_A cũng là một λ-hệ, nên theo tính tối tiểu của $\mathcal{L}(\mathcal{P})$, suy ra $\mathcal{L}(\mathcal{P}) = L_A$. Từ định nghĩa của L_A thì với mọi $B \in \mathcal{L}(\mathcal{P}) = L_A$, ta đều có $B \cap A \in \mathcal{L}(\mathcal{P})$. ĐPCM. $\square$

Định lý 1.2.16. *Xét một họ các tập con $\mathcal{P}$. Nếu như có hai độ đo xác suất μ_1 và μ_2 cùng xác định trên không gian $(\Omega, \sigma(\mathcal{P}))$ và $\mu_1(A) = \mu_2(A)\ \forall A \in \mathcal{P}$ thì $\mu_1 = \mu_2$ trên cả $\sigma(\mathcal{P})$.*

Chứng minh. Đặt

$$\mathcal{L} = \{A \in \sigma(\mathcal{P}) :\ \mu_1(A) = \mu_2(A)\}.$$

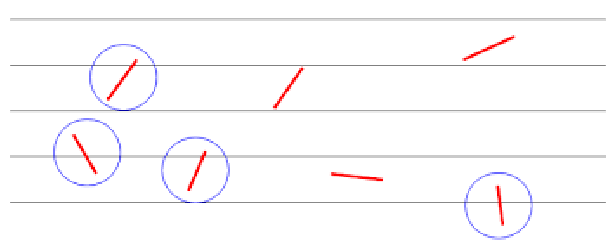
Ta thấy rằng $\mathcal{P} \subset \mathcal{L}$ và $\Omega \in \mathcal{L}$. Ta có thể kiểm tra $\mathcal{L}$ là một λ-hệ, nên theo Định lý $\pi - \lambda$ của Dynkin thì $\sigma(\mathcal{P}) \subset \mathcal{L}$. ĐPCM. □

1.2.3 Một số ví dụ

Chúng ta đã định nghĩa về không gian xác suất một cách trừu tượng và tổng quát. Tuy nhiên với các bài toán cụ thể thì việc xây dựng một không gian xác suất cụ thể và thích hợp là rất quan trọng. Ta cùng xét các ví dụ sau.

Ví dụ đầu tiên là bài toán que tăm Buffon, được đề xuất bởi một nhà quý tộc người Pháp có tên là Georges-Louis Leclerc, Bá tước vùng Buffon vào năm 1777.

Ví dụ 1.2.17 (Bài toán que tăm Buffon)**.** *Thả ngẫu nhiên một que tăm độ dài l lên một tờ giấy (mặt phẳng) được chia thành các dải song song cỡ $h \geq l$ bởi các đường kẻ song song. Hãy tính xác suất để que tăm cắt vào các đường kẻ song song.*



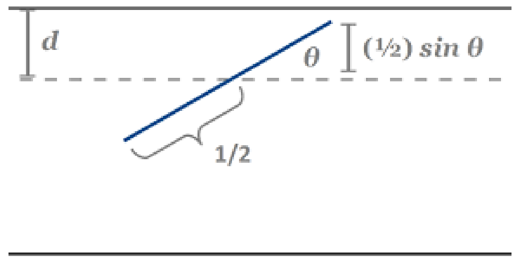
Hình 1.1: Một số minh họa mô phỏng khi tung que tăm

Ta xác định vị trí của que tăm qua hai tham số (d, θ) với (xem hình 1.2.3)

+ $d \in [0, h/2]$ là khoảng cách từ tâm que tăm (trung điểm) đến đường kẻ gần nhất.

+ $\theta \in [0, \pi]$ là góc giữa que tăm và đường kẻ ngang.

Khi đó ta có không gian mẫu $\Omega = \{(d, \theta) :\ d \in [0, h/2], \theta \in [0, \pi]\}$ có dạng hình chữ nhật. Trên không gian mẫu này, ta xét σ-đại số Borel và độ đo xác suất theo tỉ lệ diện tích.



Hình 1.2: Cách xác định vị trí của que tăm

Xét biến cố que tăm cắt vào đường kẻ song song. Đường song song này chính là đường kẻ gần tâm que tăm nhất vì $l \leq h$. Que tăm cắt đường kẻ này khi

$$d \leq \frac{l}{2}\sin\theta.$$

Do đó có thể biểu diễn biến cố cần tìm là

$$\{(d,\theta):\ d \in [0,h/2], \theta \in [0,\pi], d \leq \frac{l}{2}\sin\theta\}.$$

Từ đây ta tính được xác suất

$$\frac{1}{h\pi/2}\int_0^\pi \left(\int_0^{\frac{l}{2}\sin\theta} dx\right) d\theta = \frac{2l}{h\pi}.$$

Trong trường hợp đặc biệt, khi $l = h$, ta có xác suất cần tính là $2/\pi$. Ta sẽ trở lại kết quả này trong mục Luật số lớn ở Chương 3, với ứng dụng để sấp xỉ cho số π.

Ví dụ 1.2.18. *Cho một que có độ dài đơn vị. Ta bẻ ngẫu nhiên que đã cho thành ba đoạn. Tính xác suất để ba đoạn con trên có thể nối tạo thành hình tam giác.*

Gọi x, y và z lần lượt là độ dài các đoạn con được bẻ ra theo thứ tự từ trái sang phải. Khi đó, ta xác định không gian mẫu

$$\Omega = \{(x,y,z):\ x+y+z = 1,\ x,y,z > 0\},$$

hay Ω chính là miền tam giác có ba đỉnh $A(0,0,1)$, $B(0,1,0)$ và $C(1,0,0)$. Xét σ-đại số cảm sinh từ các tập mở. Khi đó biến cố ba đoạn con trên có thể nối tạo thành hình tam giác được biểu diễn thành

$$M = \{(x,y,z) \in \Omega:\ x+y > 1/2,\ y+z > 1/2,\ z+x > 1/2\},$$

hay là miền trong tam giác có ba đỉnh là trung điểm các cạnh của tam giác ABC.

Với độ đo xác suất là tỉ lệ diện tích, ta tính được xác suất của biến cố cần tính là $\mathbb{P}(M) = 1/4$.

Ví dụ 1.2.19. *Cho một que AB có độ dài đơn vị. Chọn ngẫu nhiên một điểm và bẻ que đã cho thành hai đoạn. Xét đoạn con chứa đầu mút B, tương tự như trước, lại chọn ngẫu nhiên một điểm trên đoạn con này và bẻ thành hai đoạn con nhỏ hơn. Như vậy từ que AB ban đầu ta đã chia thành ba đoạn con. Tính xác suất để ba đoạn con trên có thể nối tạo thành hình tam giác.*

Ví dụ này bạn đọc có thể trở lại sau khi đã tìm hiểu về phân phối đều trên đoạn thẳng ở Chương 2. Xét X là độ dài đoạn con chưa đầu mút A tương ứng với vị trí của điểm chọn đầu tiên. Khi đó, X theo phân phối đều trên đoạn đơn vị AB. Tiếp theo xét Y là độ dài đoạn con ở giữa được bẻ ra từ đoạn con độ dài $1 - X$, cũng theo phân phối đều. Ta biểu diễn biến cố ba đoạn con có thể nối tạo thành hình tam giác thành

$$N = \{(X, Y) : 0 < X < 1/2,\ 0 < Y < 1/2,\ X + Y > 1/2\}.$$

Do đó xác suất của biến cố trên bằng

$$\int_0^{1/2} \left(\frac{1}{1-x} \int_{1/2-x}^{1/2} dy \right) dx = \int_0^{1/2} \frac{x}{1-x} dx = \log 2 - 1/2 \neq 1/4!!!$$

1.3 Xác suất điều kiện

Trong thực tế ta thấy rằng *xác suất sẽ bị điều chỉnh với các luồng thông tin thêm*. Ta xét ví dụ trong môn bóng đá như sau. Xét một trận đấu giữa đội nhà bạn ủng hộ và đội khách. 24 giờ trước trận đấu, tỉ lệ thắng của đội nhà là 0.6. Tuy nhiên, 2 giờ trước trận đấu, một cầu thủ quan trọng của đội nhà bị chấn thương khi khởi động, lúc này xác suất thắng giảm xuống còn 0.45. Kết thúc hiệp một của trận đấu, đội nhà đang dẫn 2-0, lúc này xác suất thắng tăng lên là 0.8. Và khi trận đấu kết thúc với kết quả 3-1, đội nhà thắng lợi xác suất 1.

Để tổng quát, ta có dịnh nghĩa xác suất điều kiện như sau.

Định nghĩa 1.3.1. *Giả sử A và B là hai biến cố bất kì, trong đó $\mathbb{P}(B) > 0$. Khi đó xác suất điều kiện của biến cố A khi đã biết biến cố Bxảy ra là*

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

Ta xét thêm ví dụ như sau. Hai bạn An và Bình tham gia trò chơi như sau: các bạn tung một đồng xu ba lần, nếu trong ba lần số lần ra mặt ngửa nhiều hơn thì bạn An thắng và được phần thưởng, ngược lại nếu số lần ra mặt sấp nhiều hơn thì Bình thắng và được phần thưởng. Ta thấy rằng trò chơi là công bằng xác suất 50-50, với không gian mẫu

$$\Omega = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\},$$

và các biến cố thuận lợi

$$A = \{HHH, HHT, HTH, THH\},$$

$$B = \{HTT, THT, TTH, TTT\}.$$

Bây giờ, ta giả sử đã tung đồng xu một lần và ra mặt ngửa. Biến cố nầy được biểu diễn là

$$E = \{HHH, HHT, HTH, HTT\}.$$

Khi đó,

$$A \cap E = \{HHH, HHT, HTH\},$$

$$B \cap E = \{HTT\}.$$

Và ta có xác suất thắng của từng bạn với điều kiện tung lần đầu ra mặt ngửa như sau:

$$\mathbb{P}(A|E) = 3/4,\ \mathbb{P}(B|E) = 1/4.$$

Nhận xét: ta có thể hiểu một cách đơn giản là *việc lấy điều kiện làm thay đổi không gian mẫu.*

Từ định nghĩa xác suất điều kiện ta suy ra công thức nhân xác suất sau

$$\mathbb{P}(AB) = \mathbb{P}(A|B)\mathbb{P}(B).$$

Mở rộng công thức nhân xác suất trên, ta có công thức sau.

Định lý 1.3.2. *Cho n biến cố $A_1, \dots, A_n$ sao cho $\mathbb{P}(A_1 \cap \dots \cap A_{n-1}) > 0$. Ta có*

$$\mathbb{P}(A_1 \cap \dots \cap A_n) = \mathbb{P}(A_n|A_1 \cap \dots \cap A_{n-1})\mathbb{P}(A_{n-1}|A_1 \cap \dots \cap A_{n-2}) \dots \mathbb{P}(A_2|A_1)\mathbb{P}(A_1).$$

Ta xét ví dụ thực tế quen thuộc sau.

Ví dụ 1.3.3. *Trong hộp có n lá thăm, trong đó chỉ có 1 lá trúng thưởng. Một nhóm n người chơi lần lượt lấy ra một cách ngẫu nhiên (không hoàn lại) từng lá thăm từ hộp cho đến khi có người đầu tiên lấy được lá thăm trúng thưởng thì dừng lại. Tính xác suất trúng thưởng của từng người chơi?*

Gọi A_1 là biến cố người thứ nhất rút trúng thưởng, ta có $\mathbb{P}(A_1) = 1/n$.

Xét biến cố người thứ hai trúng thưởng. Ta thấy rằng để người thứ hai được rút thì người thứ nhất phải rút không trúng lá trúng thưởng, và người thứ hai rút trúng trong $n-1$ lá phiếu còn lại sau khi đã bỏ đi lá phiếu của người thứ nhất. Vậy xác suất để người thứ hai trúng thưởng là

$$\mathbb{P}(A_2 \cap \overline{A_1}) = \mathbb{P}(A_2|\overline{A_1})\mathbb{P}(\overline{A_1}) = \frac{1}{n-1}\frac{n-1}{n} = 1/n.$$

Tương tự, xác suất trúng thưởng cho những người còn lại là

$$\mathbb{P}(A_3 \cap \overline{A_2} \cap \overline{A_1}) = \mathbb{P}(A_3|\overline{A_2} \cap \overline{A_1})\mathbb{P}(\overline{A_2}|\overline{A_1}\mathbb{P}(\overline{A_1}) = \frac{1}{n-2}\frac{n-2}{n-1}\frac{n-1}{n} = 1/n.$$

$$\mathbb{P}(A_k \cap \overline{A_{k-1}} \cap \ldots \cap \overline{A_1}) = \mathbb{P}(A_k|\overline{A_{k-1}} \cap \ldots \cap \overline{A_1}) \ldots \mathbb{P}(\overline{A_2}|\overline{A_1}\mathbb{P}(\overline{A_1}) = 1/n.$$

Như vậy , cách rút thăm trúng thưởng này là công bằng.

Tiếp theo ta phát biểu hai công thức *Xác suất toàn phần* và *Công thức Bayes*. Trước tiên, ta có định nghĩa về một họ biến cố đầy đủ như sau.

Định nghĩa 1.3.4. *Hệ các biến cố $\{A_1, \ldots, A_n\}$ được gọi là* đầy đủ *nếu nó là một phân hoạch của Ω trong $\mathcal{F}$, tức là*

1. $A_i \cap A_j = \emptyset$ *với mọi* $i \neq j$;
2. $\cup_{i=1}^n A_i = \Omega$.

Định lý 1.3.5. *Xét $\{A_1, \ldots, A_n\}$ là một hệ đầy đủ các biến cố với $\mathbb{P}(A_i) > 0$ với mọi i.*

1. *Với biến cố B bất kì, ta có công thức xác suất toàn phần*

$$\mathbb{P}(B) = \sum_{i=1}^n \mathbb{P}(B|A_i)\mathbb{P}(A_i).$$

2. *Với biến cố B có xác suất dương, ta có công thức Bayes*

$$\mathbb{P}(A_k|B) = \frac{\mathbb{P}(B|A_k)\mathbb{P}(A_k)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B|A_k)\mathbb{P}(A_k)}{\sum_{i=1}^n \mathbb{P}(B|A_i)\mathbb{P}(A_i)}.$$

Ta có một số ví dụ áp dụng công thức Xác suất toàn phần và công thức Bayes như sau.

Ví dụ 1.3.6. *Ta xét lại Ví dụ 1.3.3. Kết quả rút thăm sẽ thay đổi thế nào nếu trong n lá thăm có 2 lá trúng thưởng?*

Ta vẫn gọi A_1 là biến cố người thứ nhất rút trúng thưởng, ta có $\mathbb{P}(A_1) = 2/n$.

Tuy nhiên tình huống sẽ thay đổi với biến cố người thứ hai trúng thưởng. Nếu như chỉ có đúng một lá thăm trúng thưởng thì người thứ hai được rút và có thể rút trúng thưởng chỉ khi người thứ nhất phải rút không trúng lá trúng thưởng. Còn khi có 2 lá trúng thưởng thì dù người thứ nhất có rút trúng thưởng hay không thì người thứ hai vẫn có cơ hội rút. Cơ hội rút trúng của người thứ hai phụ thuộc vào lần đầu người thứ nhất có rút trúng hay không. Theo công thức Xác suất toàn phần, xác suất để người thứ hai trúng thưởng là

$$\mathbb{P}(A_2) = \mathbb{P}(A_2|A_1)\mathbb{P}(A_1) + \mathbb{P}(A_2|\overline{A_1})\mathbb{P}(\overline{A_1}) = \frac{1}{n-1}\frac{2}{n} + \frac{2}{n-1}\frac{n-2}{n} = \frac{4n-6}{(n-1)n}.$$

Ví dụ 1.3.7. *Một phương pháp test cho kết quả dương tính với* 98% *với những người thực sự bị bệnh và cho kết quả dương tính giả* 1% *với những người không mắc bệnh. Biết rằng tỉ lệ người nhiễm bệnh trong một cộng đồng nào đó là* 0, 1%.*Tính xác suất để người đó thực sự bị bệnh khi có kết quả test dương tính.*

Ta có

$$\mathbb{P}(+|B) = 98\%,\ \mathbb{P}(+|\overline{B}) = 1\%,\ \mathbb{P}(B) = 0,1\%.$$

Vậy

$$\mathbb{P}(B|+) = \frac{\mathbb{P}(+|B)\mathbb{P}(B)}{\mathbb{P}(+|B)\mathbb{P}(B) + \mathbb{P}(+|\overline{B})\mathbb{P}(overlineB)} \approx 0.09.$$

Có nghĩa là nếu bạn có kết quả test dương tính thì chỉ có 9% xác suất bạn thực sự bị bệnh.

Ví dụ 1.3.8. *Xét một đồng xu có xác suất ra mặt ngửa trong mỗi lần tung là p. Goi π_n là xác suất để sau n lần tung đồng xu thì số lần ra mặt ngửa là chẵn. Chứng minh rằng*

$$\pi_{n+1} = (1-p)\pi_n + p(1-\pi_n).$$

Gọi A_n là biến cố sau n lần tung đồng xu thì số lần ra mặt ngửa là chẵn, ta có $\pi_n = \mathbb{P}(A_n)$. Theo công thức Xác suất toàn phần và lấy điều kiện theo trạng thái ngửa hoặc sấp của lần tung đầu tiên, ta có

$$\mathbb{P}(A_{n+1}) = \mathbb{P}(A_{n+1}|H)\mathbb{P}(H) + \mathbb{P}(A_{n+1}|T)\mathbb{P}(T).$$

Với điều kiện lần đầu tung ra mặt ngửa thì để trong cả $n+1$ lần tung có số lần ra mặt ngửa là chẵn, ta phải có là trong n lần tung từ lần thứ 2 đến lần thứ $n+1$ có tổng cộng lẻ lần ra mặt ngửa. Và tương tự nếu lần đầu ra sấp thì trong n lần tung sau có chẵn lần ra mặt ngửa. Vậy ta có

$$\mathbb{P}(A_{n+1}) = \mathbb{P}(\overline{A}_n)\mathbb{P}(H) + \mathbb{P}(A_n)\mathbb{P}(T),$$

suy ra ĐPCM.

Từ ví dụ trên, ta thấy rằng các công thức Xác suất toàn phần và Bayes rất hữu dụng để đưa ra các quan hệ dạng truy hồi.

1.4 Tính độc lập của biến cố

Tính độc lập là tính chất quan trọng của Lý thuyết xác suất. Có thể nói rằng định nghĩa về tính độc lập là điểm khởi đầu để Lý thuyết xác suất tách khỏi Lý thuyết độ đo để thành một ngành riêng. Theo Durett,

"Measure theory ends and probability begins with the definition of independence."

Trước tiên, ta có định nghĩa về hai biến cố độc lập như sau.

Định nghĩa 1.4.1. *Hai biến cố A và B được gọi là* độc lập *với nhau nếu*

$$\mathbb{P}(AB) = \mathbb{P}(A)\mathbb{P}(B).$$

Để tìm hiểu kĩ hơn về công thức trên, ta giả sử $\mathbb{P}(B) > 0$ và xét xác suất điều kiện của biến cố A khi biến cố B đã xảy ra. Ở mục trước, ta thấy rằng khi lấy xác suất điều kiện thì có thể làm thay đổi đi xác suất ban đầu. Tuy nhiên, trong tình huống hai biến cố A và B độc lập, ta có

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(A)\mathbb{P}(B)}{\mathbb{P}(B)} = \mathbb{P}(A),$$

tức là xác suất để A xảy ra không phụ thuộc vào việc B đã xảy ra hay không.

Ta xét một số ví dụ về hai biến cố độc lập như sau:

+ Nếu $\mathbb{P}(A) = 0$ thì biến cố A độc lập với mọi biến cố $B \in \mathcal{F}$. Thật vậy, vì $0 \le \mathbb{P}(A\cap B) \le \mathbb{P}(A) = 0$, nên $\mathbb{P}(A\cap B) = 0 = \mathbb{P}(A)\mathbb{P}(B)$.

+ Tung một đồng xu hai lần. Gọi A là biến cố lần tung đầu ra mặt ngửa, và B là biến cố lần tung sau ra mặt sấp. Khi đó, hai biến cố A và B độc lập, vì

$$\mathbb{P}(A) = \mathbb{P}(B) = 1/2,\ \mathbb{P}(A\cap B) = \mathbb{P}(\{HT\}) = 1/4.$$

Ta có mở rộng khái niệm độc lập cho một họ các biến cố như sau

Định nghĩa 1.4.2. *1. Một họ (có thể vô hạn) các biến cố $(A_i)_{i\in I}$ được gọi là* độc lập từng đôi *nếu*

$$\mathbb{P}(A_iA_j) = \mathbb{P}(A_i)\mathbb{P}(A_j) \text{ với mọi } i \neq j \in I.$$

2. Một họ (có thể vô hạn) các biến cố $(A_i)_{i\in I}$ được gọi là độc lập trong toàn thể *(gọi tắt là độc lập) nếu*

$$\mathbb{P}(A_{i_1}\dots A_{i_n}) = \mathbb{P}(A_{i_1})\dots\mathbb{P}(A_{i_n})$$

với mọi tập con gồm hữu hạn các phần tử phân biệt $\{i_1,\dots,i_n\}$ của I.

Ta có ví dụ về họ biến cố độc lập mở rộng từ ví dụ ở trên như sau: tung đồng xu đủ nhiều lần, gọi A_k là biến cố chỉ trạng thái (ngửa hoặc sấp) của lần tung thứ k.

Ví dụ 1.4.3. *Một họ biến cố độc lập từng đôi chưa chắc đã độc lập toàn thể.*

Thật vậy, tung một viên xúc xắc ba lần. Xét các biến cố $A_{i,j}$ sao cho ở lần tung thứ i và thứ j ra mặt giống nhau. Khi đó,

$$\mathbb{P}(A_{1,2}) = \mathbb{P}(A_{2,3}) = \mathbb{P}(A_{1,3}) = 1/6;$$

biến cố giao $A_{1,2}\cap A_{2,3}$ thể hiện cả ba lần tung đều có mặt như nhau nên

$$\mathbb{P}(A_{1,2}\cap A_{2,3}) = \frac{6}{6^3} = 1/36 = \mathbb{P}(A_{1,2}).\mathbb{P}(A_{2,3}),$$

suy ra hai biến cố $A_{1,2}$ và $A_{2,3}$ độc lập với nhau, tương tự với các cặp còn lại, ta suy ra họ ba biến cố $\{A_{1,2}, A_{2,3}, A_{1,3}\}$ độc lập theo từng đôi; nhưng họ cả ba biến cố lại không độc lập vì

$$\mathbb{P}(A_{1,2}\cap A_{2,3}\cap A_{1,3}) = \frac{6}{6^3} = 1/36 \neq \mathbb{P}(A_{1,2}).\mathbb{P}(A_{2,3})\mathbb{P}(A_{1,3}).$$

Kết quả sau được suy ra trực tiếp từ định nghĩa của họ biến cố độc lập và thường được sử dụng.

Mệnh đề 1.4.4. *Cho họ hữu hạn các biến cố $A_1,\dots A_n$ độc lập. Khi đó,*

$$\mathbb{P}(A_1\cap\dots\cap A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2)\dots\mathbb{P}(A_n).$$

Ta cũng thường dùng các kết quả sau.

Mệnh đề 1.4.5. *1. Nếu hai biến cố A và B độc lập thì các căp biến cố $(A,\overline{B})$, $(\overline{A},B)$ và $(\overline{A},\overline{B})$ cũng là các cặp biến cố độc lập.*

2. Nếu họ ba biến cố A,B,C độc lập thì biến cố A độc lập với biến cố hợp $B\cup C$.

Chứng minh. 1. Ta xét cặp biến cố A và $\overline{B}$. Ta có

$$\mathbb{P}(A \cap \overline{B}) = \mathbb{P}(A) - \mathbb{P}(A \cap B) = \mathbb{P}(A)_{\mathbb{P}}(A)\mathbb{P}(B) = \mathbb{P}(A)\mathbb{P}(\overline{B}.$$

Suy ra hai biến cố A và $\overline{B}$ độc lập. Tương tự cho các cặp biến cố còn lại.

2. Ta có:

$$\begin{aligned}\mathbb{P}(A \cap (B \cup C)) =& \mathbb{P}((A \cap B) \cup (A \cap C)) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap C) - \mathbb{P}(A \cap B \cap C)) \\ =& \mathbb{P}(A)[\mathbb{P}(B) + \mathbb{P}(C) - \mathbb{P}(B \cap C)] = \mathbb{P}(A)\mathbb{P}(B \cup C)\end{aligned}$$

Suy ra biến cố A độc lập với biến cố $B \cup C$.

□

1.5 Phương pháp xác suất

Trong mục này, chúng ta tìm hiểu về *Phương pháp xác suất* để giải quyết các bài toán trong các lĩnh vực khác. Những ứng dụng đầu tiên của phương pháp này dựa trên nguyên lý sau.

Nguyên lý 1: Nếu biến cố A có xác suất dương thì biến cố này không rỗng, tức là tồn tại cấu hình $\omega \in \Omega$ sao cho $\omega \in A$.

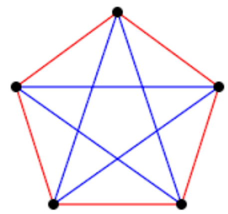
Ta sẽ áp dụng nguyên lý trên để tìm chặn dưới cho chỉ số Ramsey $R(k,k)$. Trước tiên ta có định nghĩa tổng quát cho chỉ số này như sau.

Định nghĩa 1.5.1. *Chỉ số Ramsey $R(m,n)$ là số đỉnh N nhỏ nhất sao cho khi ta tô màu xanh đỏ các cạnh nối các đỉnh của đồ thị gồm N đỉnh một cách bất kỳ thì luôn có thể tìm ra m đỉnh mà các cạnh nối giữa chúng tô xanh hoặc n đỉnh mà các cạnh nối giữa chúng tô đỏ.*

Các giá trị tường minh của chỉ số Ramsey đã tìm được là: $R(3,3) = 6, \$; R(3,4) = 9,\ R(4,4) = 18$. Thông thường, để tìm chặn dưới cho chỉ số Ramsey, ta sẽ chỉ ra một ví dụ với cách tô màu cụ thể để không thỏa mãn yêu cầu. Ví dụ, để chứng minh $R(3,3) > 5$ ta có cách tô màu các cạnh của đồ thị gồm năm đỉnh sao cho không có tam giác có ba cạnh cùng màu như sau.

Tuy nhiên, với sốk lớn thì việc xây dựng cách tô tường minh là rất khó khăn. Erdos vào những năm 1950 đã đề xuất phương pháp xác suất để chứng minh sự tồn tại một cách tô màu như vậy

Định lý 1.5.2. *Ta có $R(k,k) \geq \lfloor 2^{k/2} \rfloor$.*



Hình 1.3: Đồ thị 5 đỉnh, các cạnh biên tô đỏ, các đường chéo tô xanh.

Chứng minh. Thật vậy, ta sẽ chỉ ra rằng với $N = \lfloor 2^{k/2} \rfloor$ thì tồn tại một cách tô màu xanh hoặc đỏ các cạnh của một đồ thị cỡ N đỉnh sao cho không tồn tại k đỉnh mà tất cả các cạnh nối giữa k đỉnh này được tô cùng màu (xanh hoặc đỏ).

Với một cạnh bất kỳ, ta tô màu nó một cách ngẫu nhiên một trong hai màu xanh và đỏ với cùng xác suất $1/2$, và cách tô màu các cạnh là độc lập với nhau. Ta xét xác suất của biến cố A: tồn tại k đỉnh mà các cạnh nối cùng màu với nhau. Dễ thấy $A = A_d \cup A_x$, trong đó A_d và A_x lần lượt là các biến cố tồn tại k đỉnh mà các cạnh nối cùng màu đỏ hoặc xanh. Với A_x ta có

$$A_x = \bigcup_{1 \le i_1 < i_2 < \ldots < i_k < N} A_x(i_1, i_2, \ldots, i_k),$$

trong đó $A_x(i_1, i_2, \ldots, i_k)$ là biến cố mà các cạnh nối giữa các đỉnh $i_1, i_2, \ldots, i_k$ được tô xanh. Do với k đỉnh ta có C_k^2 cạnh nối và việc tô màu các cạnh là độc lập với nhau nên

$$\mathbb{P}\left(A_x(i_1, i_2, \ldots, i_k)\right) = \left(\frac{1}{2}\right)^{C_k^2}.$$

Suy ra

$$\mathbb{P}(A_x) \le \sum_{1 \le i_1 < i_2 < \ldots < i_k < N} \mathbb{P}\left(A_x(i_1, i_2, \ldots, i_k)\right) = C_N^k \left(\frac{1}{2}\right)^{C_k^2}.$$

Tương tự, ta cũng có

$$\mathbb{P}(A_d) \le \sum_{1 \le i_1 < i_2 < \ldots < i_k < N} \mathbb{P}\left(A_d(i_1, i_2, \ldots, i_k)\right) = C_N^k \left(\frac{1}{2}\right)^{C_k^2}.$$

Vậy nên

$$\mathbb{P}(A) \le \mathbb{P}(A_x) + \mathbb{P}(A_d) \le 2.C_N^k \left(\frac{1}{2}\right)^{C_k^2}.$$

Ta thấy rằng khi $N = \lfloor 2^{k/2} \rfloor$ thì

$$2.C_N^k \left(\frac{1}{2}\right)^{C_k^2} < 1.$$

Khi đó $\mathbb{P}(A) < 1$ hay là xác suất để có k đỉnh mà các cạnh nối cùng màu là nhỏ hơn 1, suy ra không phải mọi cách tô màu trên đồ thị $N = \lfloor 2^{k/2} \rfloor$ đỉnh đều thỏa mãn tính chất trên và do đó sẽ có một cách tô màu nào đó mà không thể tìm được k đỉnh có các cạnh nối cùng màu. □

Ví dụ trên là một trong những ví dụ khởi đầu cho lý thuyết Đồ thị ngẫu nhiên. Dạng đồ thị trên còn được gọi là mô hình Erdos-Renyi. Cũng xin lưu ý độc giả thêm rằng, câu hỏi về việc cải thiện chặn dưới ch chỉ số Ramsey $R(k,k)$ theo dạng $R(k,k) \geq C.a^k$ với $a > \sqrt{2}$ đến nay vẫn là câu hỏi mở.

Ta xét thêm một số ví dụ minh họa cho nguyên lý của phương pháp xác suất như sau.

Ví dụ 1.5.3 (Kì thi Toán quốc tế IMO 1987)**.** *Chứng minh rằng ta có thể tô các số thuộc tập $\{1, 2, \ldots, 1987\}$ bởi một trong bốn màu sao cho mọi dãy cấp số cộng gồm 10 số trong tập thì không cùng màu với nhau.*

Chứng minh. Tô màu các số một cách ngẫu nhiên bởi một trong bốn màu với xác suất như nhau sao cho việc tô màu trên các cạnh là độc lập với nhau. Xét biến cố A thỏa mãn tồn tại một dãy cấp số cộng gồm 10 số trong tập được tô cùng màu.

Lấy ra 10 số xác định lập thành một cấp số cộng. Xác suất để 10 số này được tô cùng màu là $1/4^9$.

Bây giờ, ta đi đếm số các dãy cấp số cộng gồm 10 số có thể. Một dãy hoàn toàn xác định bởi u và d với u là số đầu tiên trong dãy và d là công sai. Chú ý là $1 < u + 9d \leq 1987$ nên $1 \leq d \leq 220$ và với mỗi d thì có $1987 - 9d$ số u tương ứng nên số dãy cấp số cộng thỏa mãn là

$$\sum_{d=1}^{220}(1987 - 9d) = 218350.$$

Do đó, dễ thấy

$$\mathbb{P}(A) \leq \frac{218350}{4^9} < 1.$$

Từ đó suy ra ĐPCM. □

Ví dụ 1.5.4. *Xét $A_1, \ldots, A_n$ là tập con của một tập cho trước sao cho với mỗi chỉ số i thì tập con A_i có đúng k phần tử. Chứng minh rằng nếu $n < 2^{k-1}$ thì sẽ tồn tại một cách tô màu các phần tử trong tập bởi mộ trong hai màu xanh và đỏ sao cho trong các tập con A_i, không có tập con nào có các phần tử cùng màu.*

Chứng minh. Với một phần tử, ta tô màu nó một cách ngẫu nhiên một trong hai màu xanh và đỏ với cùng xác suất 1/2. Việc tô màu các phần tử là độc lập với nhau.

Dễ thấy xác suất để một tập con A_i có các phần tử tô cùng màu là $2^{-(k-1)}$.

Ta có n tập con nên theo bất đẳng thức chặn hợp, xác suất để tồn tại tập con trong số n tập con có các phần tử tô cùng màu sẽ không quá $n.2^{-(k-1)} < 1$. Suy ra ĐPCM. □

1.6 Bổ đề Lovasz địa phương

Trong các ví du của Mục 1.5, ta thấy rằng ý tưởng chính là ta muốn tránh các *biến cố xấu* dạng A_i. Sử dụng bất đẳng thức chặn hợp

$$\mathbb{P}(A_1 \cup \ldots A_n) \leq \sum_{i=1}^{n} \mathbb{P}(A_i),$$

và nếu như

$$\sum_{i=1}^{n} \mathbb{P}(A_i) < 1, \tag{1.1}$$

thì ta có

$$\mathbb{P}(\overline{A}_1 \cap \ldots \cap \overline{A}_n) > 0,$$

nên theo Nguyên lý 1 của Phương pháp xác suất phải tồn tại tình huống (cấu hình) tránh được tất cả các tính chất A_i.

Câu hỏi đặt ra là liệu khi bất đẳng thức (1.1) không còn đúng (ví du như khi $n \geq 2^{k-1}$ trong Ví dụ 1.5.4) thì liệu có cách nào để xử lý hay không? Ta xét một ví dụ đơn giản khi họ biến cố A_1, λ, A_n là độc lập. Khi đó họ các biến cố đối $\overline{A}_1, \ldots, \overline{A}_n$ cũng độc lập. Nên ta có

$$\mathbb{P}(\overline{A}_1 \cap \ldots \overline{A}_n) = \mathbb{P}(\overline{A}_1) \ldots \mathbb{P}(\overline{A}_n) > 0,$$

nếu như không có biến cố A_i nào đủ lớn một cách tầm thường, theo nghĩa $\mathbb{P}(A_i) = 1$.

Tuy nhiên điều kiện họ tất cả biến cố đang xét độc lập là một điều kiện tương đối ngặt. Trong thực tế, ta hay gặp hơn các dạng bài toán mà mỗi cá thể chỉ tương tác với một nhóm nhỏ xung quanh và độc lập với đa số cá thể khác trong quần thể. Đó là khởi nguồn chính của Bổ đề Lovasz địa phương. Trước khi phát biểu cụ thể Bổ đề này, ta có định nghĩa sau.

Định nghĩa 1.6.1. *Một biến cố A_i được gọi là độc lập với một họ biến cố $\{A_j, j \in J\}$ nếu với mọi bộ hữu hạn các chỉ số $j_1, \ldots, j_k \in J$ thì*

$$\mathbb{P}(A_i \cap A_{j_1} \cap \ldots \cap A_{j_k}) = \mathbb{P}(A_i)\mathbb{P}(A_{j_1} \cap \ldots \cap A_{j_k}).$$

Ta có Bổ đề Lovasz địa phương theo hai dạng: Tổng quát và Đối xứng như sau.

Định lý 1.6.2 (Bổ đề Lovasz địa phương tổng quát)**.** *Cho họ biến cố $A_1, \ldots, A_n$. Với mỗi biến cố A_i, kí hiệu J_i là tập lớn nhất gồm chỉ số $j \in \{1, \ldots, n\}$ sao cho A_i độc lập với họ biến cố $\{A_j,\, j \in J_i\}$. Giả sử tồn tại một bộ số thực $x_1, \ldots, x_n$ với $0 \le x_i < 1$ sao cho*

$$\mathbb{P}(A_i) \le x_i \prod_{j \notin J_i, j \neq i} (1 - x_j)\, \forall 1 \le i \le n.$$

Khi đó

$$\mathbb{P}(\overline{A}_1 \cap \ldots \overline{A}_n) \ge \prod_{i=1}^{n} (1 - x_i) > 0.$$

Chứng minh. Trước tiên, ta sẽ chứng minh khẳng định sau: Với mọi tập con $S \subset \{1, \ldots, n\}$ thỏa mãn $|S| = s < n$ và mọi chỉ số $i \notin S$, ta đều có

$$\mathbb{P}\left(A_i \mid \underset{j \in S}{\cap} \overline{A}_j\right) \le x_i.$$

Ta sẽ chứng minh khẳng định trên bằng quy nạp theo s. Dễ thấy khẳng định đúng với $s = 0$. Giả sử khẳng định đã đúng với mọi $s' < s$ ta sẽ chứng minh cho s. Ta chia tập S làm hai phần S_1 và S_2 như sau

$$S_1 = \{j \in S : j \in J_i\}, \;\; S_2 = S \setminus S_1.$$

Theo công thức nhân xác suất ta có

$$\mathbb{P}\left(A_i \mid \underset{j \in S}{\cap} \overline{A}_j\right) = \frac{\mathbb{P}\left[A_i \cap \left(\underset{j \in S_2}{\cap} \overline{A}_j\right) \mid \underset{j \in S_1}{\cap} \overline{A}_j\right]}{\mathbb{P}\left[\underset{j \in S_2}{\cap} \overline{A}_j \mid \underset{j \in S_1}{\cap} \overline{A}_j\right]}$$

Ta có đánh giá cho tử số như sau

$$\mathbb{P}\left[A_i \cap \left(\underset{j \in S_2}{\cap} \overline{A}_j\right) \mid \underset{j \in S_1}{\cap} \overline{A}_j\right] \le \mathbb{P}\left[A_i \mid \underset{j \in S_1}{\cap} \overline{A}_j\right] = \mathbb{P}(A_i) \le\le x_i \prod_{j \notin J_i, j \neq i} (1 - x_j).$$

Bây giờ ta sẽ đanh giá mẫu số. Đặt $S_2 = \{j_1, \ldots, j_r\}$. Nếu $r = 0$ thì mẫu số bằng 1, và có thể suy ra khẳng định ban đầu. Nếu $r > 0$, sử dụng công thức nhân xác suất và giả thiết quy nạp ta có

$$
\begin{aligned}
&\mathbb{P}\left[\overline{A}_{j_1} \cap \ldots \cap \overline{A}_{j_r} \mid \bigcap_{k\in S_1} \overline{A}_k\right] \\
=&\mathbb{P}\left[\overline{A}_{j_1} \mid \bigcap_{k\in S_1} \overline{A}_k\right] \mathbb{P}\left[\overline{A}_{j_2} \mid \overline{A}_{j_1} \cap \bigcap_{k\in S_1} \overline{A}_k\right] \ldots \mathbb{P}\left[\overline{A}_{j_r} \mid \overline{A}_{j_1} \cap \ldots \cap \overline{A}_{j_{r-1}} \cap \bigcap_{k\in S_1} \overline{A}_k\right] \\
=&\left(1 - \mathbb{P}\left[A_{j_1} \mid \bigcap_{j\in S_1} \overline{A}_j\right]\right) \ldots \left(1 - \mathbb{P}\left[A_{j_r} \mid \overline{A}_{j_1} \cap \ldots \cap \overline{A}_{j_{r-1}} \cap \bigcap_{k\in S_1} \overline{A}_k\right]\right) \\
\geq&(1 - x_{j_1})(1 - x_{j_2}) \ldots (1 - x_{j_r}) \geq \prod_{j\notin J_i, j\neq i} (1 - x_j).
\end{aligned}
$$

Thay lại các đánh gí cho tử số và mẫu số vào công thức, ta chứng minh xong khẳng định.

Cuối cùng, để hoàn thiện chứng minh định lý, ta tiếp tục sử dụng công thức nhân xác suất và khẳng định ở trên để có

$$
\begin{aligned}
\mathbb{P}(\overline{A}_1 \cap \ldots \overline{A}_n) =&\mathbb{P}(\overline{A}_1)\mathbb{P}(\overline{A}_2 \mid \overline{A}_1) \ldots \mathbb{P}(\overline{A}_n \mid \overline{A}_1 \cap \ldots \cap \overline{A}_{n-1}) \\
=&(1 - \mathbb{P}(A - 1))(1 - \mathbb{P}(A_2 \mid \overline{A}_1)) \ldots \big(1 - \mathbb{P}(A_n \mid \overline{A}_1 \cap \ldots \cap \overline{A}_{n-1})\big) \\
\geq&(1 - x_1)(1 - x_2) \ldots (1 - x_n) > 0.
\end{aligned}
$$

□

Định lý 1.6.3 (Bổ đề Lovasz địa phương đối xứng)**.** *Cho họ biến cố $A_1, \ldots, A_n$. Giả sử rằng $\mathbb{P}(A_i) \leq p$ với mọi $1 \leq i \leq n$. Giả sử thêm rằng tồn tại số nguyên dương d sao cho với mỗi biến cố A_i thì tồn tại ít nhất $n - 1 - d$ biến cố trong các biến cố còn lại tạo thành một họ biến cố độc lập với A_i, hay nói cách khác, A_i phụ thuộc với không quá d biến cố khác. Nếu như*

$$e.p.(d + 1) \leq 1,$$

thì ta có

$$\mathbb{P}(\overline{A}_1 \cap \ldots \overline{A}_n) > 0.$$

Chứng minh. Nếu $d = 0$ thì kết quả là hiển nhiên vì họ biến cố đã cho độc lập. Trong trường hợp $d > 0$, ta áp dụng kết quả của Bổ đề Lovasz địa phương tổng quát với

$$x_i = \frac{1}{d+1} < 1.$$

và sử dụng bất đẳng thức $(1 - 1/(d + 1))^d > 1/e$ với $d \geq 1$ để kiểm tra

$$\mathbb{P}(A_i) = p \leq \frac{1}{e.(d+1)} \leq \frac{1}{d+1}\left(1 - \frac{1}{d+1}\right)^d \leq x_i \prod_{j\notin J_i, j\neq i} (1 - x_j).$$

□

Ta áp dụng Bổ đề Lovasz địa phương cho ví dụ tương tự Ví dụ 1.5.4 như sau. Chú ý rằng bài toán này là cách phát biểu khác cho bài toán k-SAT nổi tiếng trong lĩnh vực Khoa học máy tính.

Ví dụ 1.6.4. *Cho một số hữu hạn tập con của một tập cho trước sao cho mỗi tập con có ít nhất k phần tử và mỗi tập con có giao khác rỗng với không quá d tập con khác. Nếu $e.(d+1) \leq 2^{k-1}$ thì sẽ tồn tại một cách tô màu các phần tử trong tập bởi mộ trong hai màu xanh và đỏ sao cho trong các tập con đang xét không có tập con nào có các phần tử cùng màu.*

Chứng minh. Ta tô màu các phần tử một cách ngẫu nhiên và độc lập với nhau hai màu xanh và đỏ với xác suất đều nhau. Gọi A_i là biến cố chỉ tập con thứ i có các phần tử được tô cùng màu. Vì mỗi tập con có ít nhất k phần tử nên

$$\mathbb{P}(A_i) \leq 2^{-(k-1)} = p.$$

Từ giả thiết đã cho, ta có $e.p(d+1) \leq 1$ và mỗi biến cố A_i phụ thuộc với không quá d biến cố khác. Do đó, áp dụng Bổ đề Lovasz địa phương, ta có

$$\mathbb{P}(\overline{A}_1 \cap \ldots \overline{A}_n) > 0.$$

Theo Nguyên lý 1 của Phương pháp xác suất, tồn tại cách tô màu thỏa mãn. □

Trong ví dụ trên, có thể thấy rằng ta không cần quan tâm đến số lượng tập con như ở Ví dụ 1.5.4 mà chỉ quan tâm đến điều kiện $e.(d+1) \leq 2^{k-1}$. Có thể thấy rằng điều kiện này không quá chặt.

Áp dụng Bổ đề Lovasz địa phương ta có thể đưa ra một chặn dưới tốt hơn cho chỉ số Ramsey $R(k,k)$ so với chặn dưới ở Định lý 1.5.2 như sau.

Định lý 1.6.5. *Ta có $R(k,k) > (\sqrt{2}/e)(1+o(1))k2^{k/2}$.*

Chứng minh. Ta sẽ chứng minh rằng nếu

$$e\binom{k}{2}\binom{n-2}{k-2}2^{1-\binom{k}{2}} < 1, \tag{1.2}$$

thì $R(k,k) > n$. Thật vậy, ta sẽ sử dụng Bổ đề Lovasz địa phương cho các biến cố A_i là biến cố mà các cạnh nối gữa các đỉnh $i_1, \ldots, i_k$ (cố định) được tô cùng màu. Vì có $\binom{k}{2}$ cạnh nối giữa k đỉnh nên

$$\mathbb{P}(A_i) = 2^{1-\binom{k}{2}} = p.$$

Tiếp theo, với mỗi biến cố A_i ta xét số các biến cố có phụ thuộc với A_i. Để biến cố A_j phụ thuộc với A_i thì phải có chung cạnh, hay nói cách khác phải có chung ít nhất hai đỉnh. Do đó, số lượng biến cố A_j như thế là nhỏ hơn hẳn

$$\binom{k}{2}\binom{n-2}{k-2} > d.$$

Ta viết lại bất đẳng thức (1.2) thành $e.p.(d+1) < 1$. Áp dụng Bổ đề Lovasz địa phương và Nguyên lý 1 của Phương pháp xác suất, ta có cách tô màu sao cho không có k đỉnh mà các cạnh nối được tô cùng màu. Nên ta có chặn dưới $R(k,k) > n$. Thay vào (1.2), ta suy ra chặn dưới tường minh như trong phát biểu của định lý. □

Ta xét tiếp ví dụ sau.

Ví dụ 1.6.6 (Kì thi Toán Quốc tế IMO 2014)**.** *Cho n đường thẳng trên mặt phẳng ở vị trí tổng quát theo nghĩa sao cho không có hai đường nào song song và không có ba đường nào đồng quy. Họ n đường thẳng như thế chia mặt phẳng thành các miền. Ta xét các miền bị chặn, có diện tích hữu hạn. Chứng minh rằng tồn tại số $c > 0$ sao cho ta có thể chọn ra $c\sqrt{n}$ đường thẳng để tô xanh sao cho không có miền bị chặn nào bị tô xanh cả biên.*

Chứng minh. Ta sẽ chứng minh với $c = (6e)^{-1/2}$. Thật vậy, chia n đường thẳng thành $c\sqrt{n}$ nhóm, mỗi nhóm gồm cỡ $\sqrt{n}/c$ đường thẳng. Bây giờ, từ mỗi nhóm ta lấy ra ngẫu nhiên một đường thẳng để tô xanh và việc chọn đường giữa các nhóm là độc lập với nhau. Gọi các miền bị chặn được tạo ra là $R_1, \dots, R_k$. Với mỗi miền R_i ta chọn trước cố định ba đường biên, và xét biến cố A_i thỏa mãn ba đường biên này đều được chọn ra để tô xanh. Khi đó

$$\mathbb{P}(A_i) \leq (cn^{-1/2})^3 = p,$$

chú ý rằng $\mathbb{P}(A_i)$ có thể bằng 0 nếu như hai đường biên cố định sẵn nằm trong cùng một nhóm.

Bây giờ, với mỗi biến cố A_i ta đếm số biến cố phụ thuộc với A_i. Ta thấy rằng A_i liên quan đến ba đường thẳng biên được chọn ra của miền R_i, và mỗi đường thẳng tạo ra không quá $2n-2$ miền bị chặn do ta chỉ có $n-1$ đoạn thẳng trên đường thẳng này (bỏ đi hai tia hai bên), và mỗi đoạn tạo thành không quá hai miền bị chặn tương ứng với hai nửa mặt phẳng. Do đó, số biến cố phụ thuộc với A_i là không quá

$$3\frac{\sqrt{n}}{c}(2n-2) > d.$$

Với cách chọn $c = (6e)^{-1/2}$, ta có

$$e.p.(d+1) \leq e.(cn^{-1/2})^3.3\frac{\sqrt{n}}{c}(2n-2) \leq 1.$$

Do đó áp dụng Bổ đề Lovasz địa phương, ta suy ra có cách chon $c\sqrt{n}$ đường để tô xanh sao cho không có miền bị chặn nào được tô xanh cả ba đường biên chọn trước tương ứng, và suy ra không có miền bị chặn nào được tô xanh kín các đường biên. □

Bổ đề Lovasz địa phương là một công cụ rất mạnh của Phương pháp xác suất. Bổ đề này còn được dùng để thiết kế các thuật toán ngẫu nhiên. Bạn đọc có thể tìm hiểu thêm trong cuốn sách [2]

Chương 2

Biến ngẫu nhiên và Vector ngẫu nhiên

2.1 Biến ngẫu nhiên

Sau khi đã tìm hiểu về không gian mẫu gồm các tình huống có thể xảy ra và xác suất để xảy ra như trong Chương 1, điều tự nhiên ta quan tâm tiếp theo là giá trị của từng tình huống. Một ví dụ ta hay gặp là trong các trò chơi may rủi, xổ số với phần thưởng tương ứng với các giải. Đây là khởi nguồn cho khái niệm về biến ngẫu nhiên. Ta có định nghĩa một cách chính xác như sau.

Định nghĩa 2.1.1. *Hàm thực $X : \Omega \to \mathbb{R}$ được gọi là $\mathcal{F}$-đo được hay* biến ngẫu nhiên *nếu*

$$X^{-1}(B) := \{w : X(w) \in B\} \in \mathcal{F} \quad \text{với mọi } B \in \mathcal{B}(\mathbb{R}).$$

Để kiểm tra một hàm $X : \Omega \to \mathbb{R}$ có phải là biến ngẫu nhiên, ta có kết quả sau.

Định lý 2.1.2. *Giả sử $X : \Omega \to \mathbb{R}$. Khi đó các khẳng định sau là tương đương*

1. *X là biến ngẫu nhiên.*
2. *$\{w : X(w) \leq a\} \in \mathcal{F}$ với mọi $a \in \mathbb{R}$.*
3. *$\{w : X(w) < a\} \in \mathcal{F}$ với mọi $a \in \mathbb{R}$.*
4. *$\{w : X(w) \in [a, b]\} \in \mathcal{F}$ với mọi $a, b \in \mathbb{R}$.*
5. *$\{w : X(w) \in (a, b)\} \in \mathcal{F}$ với mọi $a, b \in \mathbb{R}$.*

Ta có ví dụ về hai dạng biến ngẫu nhiên sẽ được sử dụng nhiều ở các phần sau.

Ví dụ 2.1.3. *Với mỗi tập $A \in \mathcal{F}$ ta có thể định nghĩa biến ngẫu nhiên* hàm chỉ báo $\mathbb{I}_A$ *như sau*

$$\mathbb{I}_A(\omega) = \begin{cases} 1 \text{ nếu } \omega \in A, \\ 0 \text{ nếu } \omega \notin A. \end{cases}$$

Tổng quát hơn, xét họ hữu hạn các biến cố $A_1, \ldots, A_n$ và các số thực $x_1, \ldots, x_n$ với mọi i, ta có biến ngẫu nhiên đơn giản

$$X(\omega) = \sum_{i=1}^{n} x_i \mathbb{I}_{A_i}(\omega).$$

Vai trò của biến ngẫu nhiên được thể hiện qua kết quả sau về cấu trúc của biến ngẫu nhiên.

Định lý 2.1.4. *Nếu X là biến ngẫu nhiên không âm thì tồn tại dãy biến ngẫu nhiên đơn giản X_n sao cho $X_n \uparrow X$. Còn khi X là biến ngẫu nhiên tổng quát thì tồn tại dãy biến ngẫu nhiên đơn giản X_n sao cho $X_n(\omega) \to X(\omega), \forall \omega \in \Omega$.*

Chứng minh. Giả sử $X \geq 0$. Với mỗi $n \geq 1$, đặt

$$X_n(\omega) = \begin{cases} \frac{k}{2^n} \text{ nếu } \frac{k}{2^n} \leq X(\omega) < \frac{k+1}{2^n}, \forall k \in \{0, 1, \ldots, n2^n - 1\} \\ n \text{ nếu } X(\omega) \geq n. \end{cases}$$

Khi đó (X_n) là dãy biến ngẫu nhiên đơn giản và $X_n \uparrow X$.

Trong trường hợp tổng quát, ta viết $X = Xk - X^-$. Do X^+ và X^- là hai biến ngẫu nhiên không âm nên ta có thể sấp xỉ qua dãy các biến ngẫu nhiên không âm đơn giản. Từ đó ta có dãy biến ngẫu nhiên không âm sấp xỉ cho X. □

Khi ta tác động hàm vào một họ các biến ngẫu nhiên sẽ được một biến ngẫu nhiên mới

Định nghĩa 2.1.5. *Hàm $\varphi : \mathbb{R}^d \to \mathbb{R}$ được gọi là* hàm Borel *nếu $\varphi^{-1}(B) \in \mathcal{B}(\mathbb{R}^d)$ với mọi $B \in \mathcal{B}(\mathbb{R})$.*

Định lý 2.1.6. *Giả sử $X_1, \ldots, X_d$ là các biến ngẫu nhiên cùng xác định trên không gian đo $(\Omega, \mathcal{F})$ và $\varphi : \mathbb{R}^d \to \mathbb{R}$ là hàm Borel. Khi đó $Y = \varphi(X_1, \ldots, X_d)$ cũng là một biến ngẫu nhiên.*

Hệ quả 2.1.7. *Cho hai biến ngẫu nhiênX và Y. Khi đó $X \pm Y, XY, X \wedge Y, X \vee Y, |X|, X^+ := X \vee 0, X^- = (-X) \vee 0$ và X/Y (nếu $Y \neq 0$) cũng là các biến ngẫu nhiên.*

Như vậy ta thấy rằng tập các biến ngẫu nhiên là đóng kín đối với các phép toán số học thông thường, và là một không gian vector.

Định lý 2.1.8. *Giả sử $(X_n)_{n\geq 1}$ là dãy biến ngẫu nhiên. Khi đó các đại lượng $\sup_n X_n$, $\inf_n X_n$, $\limsup_n X_n$, $\liminf_n X_n$ cũng là các biến ngẫu nhiên nếu chúng là hữu hạn.*

Trở lại với định nghĩa của biến ngẫu nhiên, ta thấy rằng để thỏa mãn tính đo được, có đôi khi cấu trúc topo của $\mathcal{F}$ là quá lớn và không thực sự cần thiết. Điều đó dẫn đến định nghĩa sau.

Định nghĩa 2.1.9. *Cho biến ngẫu nhiên $X : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Ta định nghĩa σ-đại số sinh bởi X, kí hiệu là $\sigma(X)$, là σ-đại số sinh bởi các tập*

$$X^{-1}(B) := \{w : X(w) \in B\} \in \mathcal{A} \quad \text{với mọi } B \in \mathcal{B}(\mathbb{R}).$$

Dễ thấy rằng $\sigma(X)$ là một σ-đại số con của $\mathcal{F}$ và là σ-đại số bé nhất trên Ω để X vẫn thỏa mãn tính đo được. $\sigma(X)$ sẽ chứa đựng các thông tin về biến ngẫu nhiên X, ví dụ như $\sigma(X) = \{\emptyset, \Omega\}$ là σ-đại số tầm thường nếu như X bằng hằng số, hay $\sigma(X) = \{\emptyset, \Omega, A, \overline{A}\}$ nếu như X là hàm chỉ báo $\mathbb{I}_A$, hay X chỉ nhận hai giá trị phân biệt.

Định lý 2.1.10. *Giả sử X là biến ngẫu nhiên trên $(\Omega, \mathcal{F})$ và Y là ánh xạ từ Ω vào $\mathbb{R}$. Khi đó Y là $\sigma(X)$-đo được khi và chỉ khi tồn tại hàm Borel $\varphi : \mathbb{R} \to \mathbb{R}$ sao cho $Y = \varphi(X)$.*

Chứng minh. Điều kiện đủ là hiển nhiên. Ta chứng minh điều kiện cần. Trước tiên giả sử Y là biến ngẫu nhiên đơn giản nhận các giá trị là $y_1, y_2, \ldots, y_n$. Do Y là $\sigma(X)$-đo được nên các tập $A_i = \{w : Y(w) = y_i\} \in \sigma(X)$. Theo định nghĩa của $\sigma(X)$, tồn tại dãy $B_i \in \mathcal{B}(\mathbb{R})$ sao cho $A_i = X^{-1}(B_i)$, $\forall i = 1, \ldots, n$. Suy ra

$$Y = \sum_{i=1}^{n} y_i \mathbb{I}_{A_i} = \sum_{i=1}^{n} y_i \mathbb{I}_{B_i} \circ X = \varphi \circ X,$$

với hàm $\varphi = \sum_{i=1}^{n} y_i \mathbb{I}_{B_i}$ là hàm đo được.

Trong trường hợp tổng quát, theo Định lí 2.1.4 tồn tại dãy hàm rời rạc Y_n là $\sigma(X)$-đo được và hội tụ đến Y. Vậy nên tồn tại các hàm Borel φ_n sao cho $Y_n = \varphi_n(X)$. Xét hàm $\varphi : \mathbb{R} \to \mathbb{R}$ như sau

$$\varphi(x) = \begin{cases} \lim\limits_{n\infty} \varphi_n(x) \text{ nếu giới hạn tồn tại,} \\ 0 \text{ nếu ngược lại.} \end{cases}$$

Ta có hàm φ như trên là hàm Borel. Với mọi $\omega \in \Omega$, do $\varphi_n(X(\omega)) = Y_n(\omega) \to Y(\omega)$ nên

$$X(\omega) \in \{x \in \mathbb{R} : \exists \lim_{n\infty} \varphi_n(x)\}.$$

Suy ra $Y = \varphi(X)$. □

Để kết thúc mục này chúng ta đưa ra định nghĩa *hầu chắc chắn* (almost surely). Đây là khái niệm quan trọng trong Lý thuyết xác suất, và là cách gọi khác cho khái niệm *hầu khắp nơi* trong Lý thuyết độ đo.

Định nghĩa 2.1.11. *Một biến cố hay một tính chất được gọi là xảy ra hầu chắc chắn nếu có xác suất bằng 1.*

Định nghĩa 2.1.12. *Cho hai biến ngẫu nhiên X và Y cùng xác định trên một không gian xác suất $(\Omega, \mathcal{F}, \mathbb{P})$. Hai biến ngẫu nhiên được gọi là bằng nhau hầu chắc chắn, kí hiệu $X \overset{a.s}{=\!=} Y$, nếu*

$$\mathbb{P}(\omega \in \Omega : X(\omega) = Y(\omega)) = 1.$$

2.2 Phân phối của biến ngẫu nhiên

2.2.1 Định nghĩa

Định nghĩa 2.2.1. *Cho biến ngẫu nhiên $X : (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathbb{R}, mB(\mathbb{R}))$. Khi đó ta có thể định nghĩa một độ đo μ_X trên $(\mathbb{R}, mB(\mathbb{R}))$, được gọi là phân phối của X, như sau*

$$\mu_X(B) = \mathbb{P}(X^{-1}(B)),\ \forall B \in mB(\mathbb{R}).$$

Có thể nói rằng khái niệm phân phối như trên là khái niệm trọng tâm và căn bản của biến ngẫu nhiên. Nó cho phép ta quên đi không gian xác suất $(\Omega, \mathcal{F}, \mathbb{P})$ trừu tượng để làm việc trên một không gian quen thuộc và có nhiều tính chất đẹp là tập số thực. Do đó trong các phát biểu về sau, ta chủ yếu phát biểu biến ngẫu nhiên kèm theo phân phối và thường không nhắc đến không gian xác suất, trừ trường hợp cần thiết.

Định nghĩa 2.2.2. *Cho hai biến ngẫu nhiên X và Y có thể xác định trên hai không gian xác suất khác nhau. Hai biến ngẫu nhiên được gọi là* có cùng phân phối*, kí hiệu $X \overset{d}{=\!=} Y$, nếu*

$$\mu_X(B) = \mu_Y(B),\ \forall B \in mB(\mathbb{R}).$$

Tiếp theo ta có định nghĩa về *hàm phân phối* của một biến ngẫu nhiên như sau.

Định nghĩa 2.2.3. *Hàm phân phối biến ngẫu nhiên X được xác định bởi*

$$F_X(x) = \mathbb{P}[X < x], \quad \forall x \in \mathbb{R}.$$

Tương tự như Định lý 1.2.16, ta thấy rằng hàm phân phối xác định duy nhất độ đo phân phối từ kết quả sau.

Định lý 2.2.4. *Cho hai độ đo μ_1 và μ_2 trên $(\mathbb{R}, mB(\mathbb{R}))$ sao cho*

$$\mu_1[(-\infty, x)] = \mu_2[(-\infty, x)],\ \forall x \in \mathbb{R}.$$

Khi đó, $\mu_1 = \mu_2$ theo nghĩa

$$\mu_1(B) = \mu_2(B),\ \forall B \in mB(\mathbb{R}).$$

Từ định lý trên, để nghiên cứu về phân phối của biến ngẫu nhiên, ta sẽ tập trung tìm hiểu về hàm phân phối. Hàm phân phối có các tính chất như sau.

Mệnh đề 2.2.5.

1. *F đơn điệu tăng;*
2. *F liên tục trái và có giới hạn phải tại mọi điểm;*
3. $\lim_{x\to-\infty} F(x) = 0,\ \lim_{x\to+\infty} F(x) = 1.$
4. *Nếu $x \le y$ thì $F(y) - F(x) = \mathbb{P}[x \le X < y]$;*
5. $\mathbb{P}(X = x) = \mu_X(\{x\}) = \lim\limits_{x_n \downarrow x} F_X(x_n) - F_X(x).$

Mệnh đề 2.2.6. *Cho hàm số $F : \mathbb{R} \to [0, 1]$ thỏa mãn ba tính chất 1.-3. của Mệnh đề trên. Khi đó tồn tại một biến ngẫu nhiên X sao cho F là hàm phân phối của biến ngẫu nhiên đó, hay là $F(x) = \mu_X[(-\infty, x)]$, với mọi $x \in \mathbb{R}$.*

Tiếp theo, chúng ta tìm hiểu các phân phối quen thuộc được chia theo hai dạng: rời rạc và liên tục.

2.2.2 Biến ngẫu nhiên rời rạc

Mở rộng khái niệm về biến ngẫu nhiên đơn giản, ta có định nghĩa sau về biến ngẫu nhiên rời rạc.

Định nghĩa 2.2.7. *Biến ngẫu nhiên X được gọi là có phân phối rời rạc nếu tập các giá trị mà X có thể nhận được có lực lượng không quá đếm được. Gọi tập giá trị này là $\{x_1, x_2, \ldots\}$. Khi đó, phân phối của X được cho bởi* bảng phân phối *như sau*

X	x_1	x_2	$\ldots$	x_n	$\ldots$
$\mathbb{P}$	p_1	p_2	$\ldots$	p_n	$\ldots$

trong đó $p_i = \mathbb{P}(X = x_i)$ *chính là xác suất tương ứng để* X *nhận giá trị* x_i.

Chú ý rằng các xác suất p_i là các số không âm và có tổng $\sum p_i = 1$. Khi đó, hàm phân phối của X có biểu diễn

$$F_X(a) = \mathbb{P}[X < a] = \sum_{x_i < a} p_i.$$

Theo biểu diễn đồ thị ta thấy hàm phân phối F_X có dạng *hàm bậc thang*. Ta cũng có thể dùng tính chất này của hàm phân phối để định nghĩa cho biến ngẫu nhiên rời rạc.

Sau đây ta liệt kê một số phân phối liên tục tuyệt đối thường gặp nhất.

Phân phối đều trên tập giá trị hữu hạn

Cho biến ngẫu nhiên X chỉ có thể nhận giá trị trong tập hữu hạn là $\{x_1, \ldots, x_n\}$. Khi đó ta nói X có phân phối đều trên tập giá trị nếu

$$\mathbb{P}(X = x_i) = \frac{1}{n},\ \forall i = 1, \ldots, n.$$

Đây là phân phối ta hay dùng trong các bài toán thống kê thực nghiệm.

Phân phối Rademacher

Biến ngẫu nhiên X có phân phối Rademacher nếu X chỉ nhận một trong hai giá trị 1 và -1 với xác suất bằng nhau $1/2$.

Phân phối Bernoulli $Ber(p)$

Cho $p \in [0, 1]$. Biến ngẫu nhiên X được gọi là có phân phối Bernoullivới tham số p, kí hiệu là $X \sim Ber(p)$), nếu X chỉ nhận một trong hai giá trị 1 và 0 với xác suất tương ứng là p và $1 - p$. Ta có bảng phân phối

X	0	1
$\mathbb{P}$	$1-p$	p

Phân phối Bernoulli $Ber(p)$ tương ứng với việc tung đồng xu, nếu ra ngửa thì kí hiệu giá trị 1 với xác suất p, còn lại nếu ra sấp thì kí hiệu giá trị 0 với xác suất $1 - p$. Ngoài ra, gọi $A \subset \Omega$ là biến cố $X = 1$ thì ta có biễu diễn $X = \mathbb{I}_A$ là hàm chỉ báo của biến cố A.

Phân phối nhị thức $B(n, p)$

Cho $p \in [0, 1]$ là xác suất thành công của một lần làm phép thử. Ta thực hiện phép thử n lần khác nhau. Khi đó, biến ngẫu nhiên X đếm số lần thành công trong số n lần làm phép thử sẽ có phân phối nhị thức với hai tham số p và n, kí hiệu $X \sim B(n, p)$. Cụ thể, X nhận các giá trị $0, 1, \ldots, n$ với xác suất tương ứng

$$\mathbb{P}(X = i) = \binom{n}{i} p^i (1-p)^{n-i}, \ \forall i = 0, 1, \ldots, n.$$

Phân phối Poisson $Poi(\lambda)$

Giả sử $\lambda > 0$. Biến ngẫu nhiên X được gọi là có *phân phối Poisson* với tham số λ nếu X nhận các giá trị nguyên không âm $0, 1, \ldots$ với xác suất tương ứng

$$pr(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \ \forall k = 0, 1, \ldots.$$

Phân phối Poisson biểu diễn số lần quan sát thấy một hiện tượng theo dõi trong một đơn vị thời gian. Tham số λ tương ứng với tần suất xuất hiện của hiện tượng, λ càng lớn thì ta càng có cơ hội quan sát được nhiều lần.

Phân phối hình học $Geo(p)$

Cho $p \in [0, 1]$ là xác suất thành công của một lần làm phép thử. Ta liên tiếp thực hiện phép thử nhiều lần cho đến khi thành công thì dừng lại. Khi đó, biến ngẫu nhiên X đếm số lần thực hiện phép thử cho đến khi dừng lại có phân phối hình học với tham số p, kí hiệu $X \sim Geo(p)$. Cụ thể, X nhận các giá trị nguyên dương $1, 2, \ldots$ với xác suất tương ứng

$$pr(X = k) = p(1-p)^{k-1}, \ \forall k = 02, 1, \ldots.$$

Mệnh đề 2.2.8. *Nếu biến ngẫu nhiên X có phân phối hình học $Geo(p)$ thì X có tính chất* thiếu trí nhớ *theo nghĩa*

$$\mathbb{P}(X \geq k + n \mid X \geq n) = \mathbb{P}(X \geq k), \ \forall k, n \geq 0.$$

2.2.3 Biến ngẫu nhiên liên tục

Định nghĩa 2.2.9. *Biến ngẫu nhiên X được gọi là có phân phối liên tục tuyệt đối với* hàm mật độ f_X *nếu*

$$F_X(a) = \mathbb{P}[X < a] = \int_{-\infty}^{a} f_X(x) dx, \quad \forall a \in \mathbb{R}.$$

Ta có thể thấy rằng hàm mật độ là đạo hàm của hàm phân phối. Hàm mật độ thỏa mãn các tính chất sau đây.

1. $f(x) \geq 0$ với mọi $x \in \mathbb{R}$ và $\int_{-\infty}^{+\infty} f(x)dx = 1$.
2. $\mathbb{P}[a < X < b] = \int_a^b f(x)dx$ với mọi số thực $a < b$. Hơn nữa, với mọi tập Borel $A \in \mathcal{B}(\mathbb{R})$ ta có
$$\mathbb{P}[X \in A] = \int_A f(x)dx. \tag{2.1}$$

Ta có thể suy ra rằng X có phân phối liên tục tuyệt đối thì $\mathbb{P}[X = a] = 0$ với mọi $a \in \mathbb{R}$. Ta gọi tính chất này là *không nhân* (nonatomic). Đây là một sự khác biệt giữa phân phối liên tục và phân phối rời rạc.

Sau đây ta liệt kê một số phân phối liên tục tuyệt đối thường gặp nhất.

Phân phối đều $U[a, b]$

Với mỗi cặp số thực $a < b$, biến ngẫu nhiên X được gọi là có *phân phối đều* trên đoạn $[a, b]$, kí hiệu là $X \sim U[a, b]$, nếu X nhận giá trị thuộc đoạn $[a, b]$ và có hàm mật độ

$$f_X(x) = \frac{1}{b-a}\mathbb{I}_{[a,b]}(x).$$

Hàm phân phối của X là

$$F_X(x) = \begin{cases} 0 & \text{nếu } x < a, \\ \frac{x-a}{b-a} & \text{nếu } a \leq x \leq b, \\ 1 & \text{nếu } x > b. \end{cases}$$

Phân phối mũ $Exp(\lambda)$

Giả sử $\lambda > 0$. Biến ngẫu nhiên X được gọi là có *phân phối mũ* với tham số λ, kí hiệu là $X \sim Exp(\lambda)$, nếu X nhận giá trị trên tập $(0, \infty)$ có hàm mật độ

$$f_X(x) = \lambda e^{-\lambda x}\mathbb{I}_{(0,\infty)}(x).$$

Hàm phân phối của X là

$$F_X(x) = (1 - e^{-\lambda x})\mathbb{I}_{(0,\infty)}(x).$$

Phân phối mũ thường dùng để mô hình hóa cho tuổi thọ của người, sinh vật, máy móc,... Ở đây tham số λ biểu diễn cho tuổi thọ bình quân.

Tương tự như phân phối hình học, phân phối mũ cũng thỏa mãn tính chất thiếu trí nhớ.

Mệnh đề 2.2.10. *Nếu biến ngẫu nhiên X có phân phối mũ $Exp(\lambda)$) thì*

$$\mathbb{P}(X \geq t + s \mid X \geq s) = \mathbb{P}(X \geq s),\ \forall t, s > 0.$$

Phân phối chuẩn $\mathcal{N}(a, \sigma^2)$

Biến ngẫu nhiên X được gọi là có phân phối chuẩn với tham số a, σ^2 ($\sigma > 0$), kí hiệu là $X \sim \mathcal{N}(a, \sigma^2)$, nếu hàm mật độ của nó có dạng

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-a)^2}{2\sigma^2}}, \quad x \in \mathbb{R}.$$

Khi $a = 0$ và $\sigma^2 = 1$ thì phân phối chuẩn $\mathcal{N}(0, 1)$ được gọi là *phân phối chuẩn tắc.*

Mệnh đề 2.2.11.

1. *Nếu $X \sim \mathcal{N}(0, 1)$ có phân phối chuẩn tắc thì $a + \sigma X \sim \mathcal{N}(a, \sigma^2)$.*
2. *Ngược lại, nếu $Y \sim \mathcal{N}(a, \sigma^2)$ thì $\dfrac{Y_a}{\sigma}$ có phân phối chuẩn tắc ($\sim \mathcal{N}(0, 1)$).*

Phân phối Gamma $\mathcal{G}(\alpha, p)$

Biến ngẫu nhiên X được gọi là có phân phối Gamma với tham số α, p ($\alpha, p > 0$), kí hiệu là $X \sim \mathcal{G}(\alpha, p)$, nếu hàm mật độ của nó có dạng

$$f_X(x) = \frac{\alpha^p x^{p-1} e^{-\alpha x}}{\Gamma(p)} \mathbb{I}_{(0,\infty)}(x),$$

trong đó $\Gamma(p) = \int_0^\infty x^{p-1} e^{-x} dx$.

Phân phối mũ là trường hợp đặc biệt của phân phối Gamma khi $p = 1$. Phân phối $\mathcal{G}(\frac{1}{2}, \frac{n}{2})$ được gọi là *phân phối khi bình phương* với n bậc tự do (viết gọi là $\chi^2(n)$).

2.3 Kì vọng và moment của biến ngẫu nhiên

2.3.1 Giới thiệu

Trong mục này ta sẽ tìm hiểu về kì vọng và moment của biến ngẫu nhiên. Đây là các đại lượng quan trọng của một biến ngẫu nhiên. Trước tiên, chúng ta sẽ trình bày cách xây dựng kì vọng. Như chúng ta đã biết biến ngẫu nhiên X là hàm số đo được trên không gian đo được $(\Omega, \mathcal{F})$ với độ đo xác suất $\mathbb{P}$, nên ta có thể định nghĩa tích phân dạng Lebesgue của X trên không gian này. Một cách hình thức ta có định nghĩa của kì vọng như sau.

Định nghĩa 2.3.1. Kì vọng *của biến ngẫu nhiên* $X : (\Omega, \mathcal{F}) \to (\mathbb{R}\mathcal{B}(\mathbb{R}))$, *kí hiệu là* $\mathbb{E}X$, *được xác định như sau*

$$\mathbb{E}X = \int_\Omega X(\omega) d\mathbb{P}.$$

Cách xây dựng kì vọng một cách chính xác sẽ được tình bày chi tiết trong mục sau. Cách xây dựng giống như cách định nghĩa tích phân Lebesgue cho hàm đo được thông qua các bước sau:

- Bước 1: định nghĩa cho biến ngẫu nhiên (hàm) đơn giản.
- Bước 2: định nghĩa cho biến ngẫu nhiên (hàm) không âm là giới hạn của dãy các kì vọng các biến ngẫu nhiên không âm đơn giản.
- Bước 3: định nghĩa cho biến ngẫu nhiên (hàm) tổng quát dựa vào phân tích $X = X^+ - X^-$.

Sau khi đã định nghĩa được kì vọng cả biến ngẫu nhiên X. Ta có thể xác định moment và các đại lượng khác của X như là kì vọng của một biến ngẫu nhiên mới được xác định bằng cách tác động hàm vào biến ngẫu nhiên X đã cho. Cụ thể, ta có các định nghĩa sau.

Định nghĩa 2.3.2. *Cho biến ngẫu nhiên X. Ta có các đại lượng sau*

1. *Với mọi* $\in \mathbb{N}$, moment bậc k *của* X *là* $\mathbb{E}X^k$.
2. *Với mọi* $\in \mathbb{N}$, moment quy tâm bậc k *của* X *là* $\mathbb{E}(X - \mathbb{E}X)^k$.
3. *Đặc biệt, moment quy tâm bậc* 2 *còn được gọi là* phương sai *(variance, dispersion) của* X,
$$Var(X) = \mathbb{E}(X - \mathbb{E}X)^2. \tag{2.2}$$
Ta còn có thể kí hiệu phương sai là DX.
4. Độ lệch chuẩn *của* X *bằng căn của phương sai của* X, $\sigma_X = \sqrt{\mathbb{E}(X - \mathbb{E}X)^2}$.
5. Độ bất đối xứng *của* X *là* $\mathbb{E}\left[\left(\frac{X - \mathbb{E}X}{\sigma_X}\right)^3\right]$. *item* Độ nhọn *của* X *là* $\mathbb{E}\left[\left(\frac{X - \mathbb{E}X}{\sigma_X}\right)^4\right]$.

Ta chú ý rằng với độ đo xác suất thì $\mathbb{P}(\Omega) = 1$, tức là tổng trọng bằng 1. Do đó, từ định nghĩa (2.3.1) của kì vọng, ta có nhận xét rằng

Kì vọng chính là giá trị trung bình!

Ta sẽ tìm hiểu kĩ hơn nhận xét này trong Mục 2.8. Cũng với cách hiểu như trên, từ định nghĩa (2.2) của phương sai, thì phương sai là đại lượng đo mức độ tập trung quanh giá trị trung bình.

2.3.2 Xây dựng kì vọng

• Trước hết ta xây dựng kì vọng của biến ngẫu nhiên đơn giản.

Định nghĩa 2.3.3. *Giả sử X là biến ngẫu nhiên đơn giản có biểu diễn*

$$X = \sum_{i=1}^{n} a_i \mathbb{I}_{A_i} \tag{2.3}$$

với $a_i \in \mathbb{R}$ và $A_i \in \mathcal{F}$ với mọi $i = 1, \ldots, n$. Kì vọng *của X (hay* tích phân *của X đối với độ đo $\mathbb{P}$) là*

$$\mathbb{E}X := \sum_{i=1}^{n} a_i \mathbb{P}(A_i).$$

Kí hiệu tập các biến ngẫu nhiên đơn giản là $\mathcal{L}_s = \mathcal{L}_s(\Omega, \mathcal{F}, \mathbb{P})$. Ta có $\mathcal{L}_s$ là một không gian vector tuyến tính.

Thật vậy, xét hai biến ngẫu nhiên X và Y là thuộc $\mathcal{L}_s$. Khi đó ta có thể biểu diễn cả X và Y ở dạng (2.3) với cùng một phân hoạch $\{A_i, 1 \leq i \leq n\}$ của Ω trong $\mathcal{A}$

$$X = \sum_{i=1}^{n} a_i \mathbb{I}_{A_i}, \text{ và } Y = \sum_{i=1}^{n} b_i \mathbb{I}_{A_i}.$$

Ta có với mọi $\alpha, \beta \in \mathbb{R}$,

$$\alpha X + \beta Y = \sum_{i=1}^{n} (\alpha a_i + \beta b_i) \mathbb{I}_{A_i}.$$

Vậy nên $\alpha X + \beta Y$ cũng thuộc $\mathcal{L}_s$. Ngoài ra, từ biểu diễn trên, ta có

$$\mathbb{E}(\alpha X + \beta Y) = \alpha \mathbb{E}X + \beta \mathbb{E}Y.$$

Như vậy toán tử kì vọng $\mathbb{E}$ có tính chất *tuyến tính*. Ta cũng thấy rằng nếu $X \leq Y$, tức là $a_i \leq b_i$ với mọi $i = 1, \ldots, n$ thì ta cũng có $\mathbb{E}X \leq \mathbb{E}Y$.

• Tiếp theo ta xây dựng kì vọng của biến ngẫu nhiên không âm. Với mỗi biến ngẫu nhiên không âm X ta đặt

$$\mathbb{E}X = \sup\{\mathbb{E}Y : Y \text{ đơn giản và } 0 \leq Y \leq X\}. \tag{2.4}$$

Dễ thấy supremum trên luôn tồn tại và thuộc $[0, \infty]$. Để đơn giản hóa cách tính kì vọng của biến ngẫu nhiên không âm, ta có mệnh đề sau.

Mệnh đề 2.3.4. *Giả sử X là biến ngẫu nhiên không âm và $(X_n)_{n \geq 1}$ là dãy biến ngẫu nhiên không âm đơn giản tăng dần tới X. Khi đó $\mathbb{E}X_n \uparrow \mathbb{E}X$.*

Chứng minh. Ta có $(\mathbb{E}X_n)_{n\geq 1}$ là dãy tăng và bị chặn trên bởi $\mathbb{E}X$. Từ Định nghĩa 2.4, suy ra $(\mathbb{E}X_n)_{n\geq 1}$ hội tụ tới $a \leq \mathbb{E}X$. Nếu $a = \infty$ thì ta đã chứng minh xong. Do đó tiếp theo ta chỉ xét $a < \infty$. Để chứng tỏ $a = \mathbb{E}X$ ta chỉ cần chỉ ra rằng với mọi biến ngẫu nhiên đơn giản Y thỏa mãn $0 \leq Y \leq X$ ta đều có $\mathbb{E}Y \leq a$.

Thật vậy, giả sử Y nhận m giá trị khác nhau $y_1, \dots, y_m$. Đặt $A_k = \{w : Y(w) = y_k\}$. Với mỗi $\epsilon \in (0, 1]$, xét dãy biến ngẫu nhiên $Y_{n,\epsilon} = (1-\epsilon)Y\mathbb{I}_{\{(1-\epsilon)Y \leq X_n\}}$. Ta có $Y_{n,\epsilon}$ là biến ngẫu nhiên đơn giản, $Y_{n,\epsilon} \leq X_n$ nên

$$\mathbb{E}Y_{n,\epsilon} \leq \mathbb{E}X_n \leq a \text{ với mọi } n. \tag{2.5}$$

Lại do $Y \leq \lim_n X_n$ nên với mọi $w \in \Omega$, tồn tại $n = n(w)$ sao cho $(1-\epsilon)Y(w) \leq X_n(w)$, tức là $A_k \cap \{w : (1-\epsilon)Y(w) \leq X_n(w)\} \to A_k$ khi $n \to \infty$. Theo Định lí **??**,

$$\begin{aligned}\mathbb{E}Y_{n,\epsilon} &= (1-\epsilon)\sum_{k=1}^{m} y_k \mathbb{P}\Big(A_k \cap [(1-\epsilon)Y \leq X_n]\Big) \\ &\to (1-\epsilon)\sum_{k=1}^{m} y_k \mathbb{P}(A_k) = (1-\epsilon)\mathbb{E}Y, \quad \text{khi } n \to \infty.\end{aligned}$$

Kết hợp với (2.5) ta có $(1-\epsilon)\mathbb{E}Y \leq a$ với mọi $\epsilon \in (0, 1]$, tức là $\mathbb{E}Y \leq a$. □

• Cuối cùng, ta định nghĩa kì vọng cho biến ngẫu nhiên bất kì. Như đã giới thiệu ở trên, ta có biểu diễn

$$X = X^+ - X^-.$$

Do X^+ và X^- đều là các biến ngẫu nhiên không âm nên ta có thể định nghĩa được kì vọng của chúng. Trong trường hợp có ít nhất một trong hai kì vọng $\mathbb{E}X^+$ và $\mathbb{E}X^-$ là hữu hạn thì ta có thể định nghĩa

$$\mathbb{E}X = \mathbb{E}X^+ - \mathbb{E}X^-.$$

2.3.3 Không gian $\mathcal{L}^p(\Omega, \mathcal{A}, \mathbb{P})$

Trong trường hợp cả hai biến ngẫu nhiên X^+ và X^- đều có kì vọng hữu hạn, thì thay vào công thức ta có kì vọng của X cũng hữu hạn. Ta gọi các biến ngẫu nhiên có kì vọng hữu hạn như vậy là biến ngẫu nhiên khả tích theo định nghĩa như sau.

Định nghĩa 2.3.5. *Biến ngẫu nhiên X được gọi là* khả tích *nếu* $\mathbb{E}(|X|) < \infty$. *Khi đó, kì vọng của X là*

$$\mathbb{E}X = \mathbb{E}(X^+) - \mathbb{E}(X^-). \tag{2.6}$$

Định nghĩa 2.3.6. *Trên không gian xác suất $(\Omega, \mathcal{F}, \mathbb{P})$, ta định nghĩa các không gian hàm*

1. *$\mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ (viết tắt là $\mathcal{L}^1$) gồm các biến ngẫu nhiên X khả tích, hay là thỏa mãn $\mathbb{E}(|X|) < \infty$.*

2. *Một cách tổng quát, với số $p > 0$ cho trước, $\mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P})$ (viết tắt là $\mathcal{L}^p$) gồm các biến ngẫu nhiên X khả tích bậc p, hay là thỏa mãn $\mathbb{E}(|X|^p) < \infty$.*

Định lý 2.3.7. *1. $\mathcal{L}^1$ là không gian vector và kì vọng là kì vọng là toán tử tuyến tính trên $\mathcal{L}^1$, tức là với mọi $X, Y \in \mathcal{L}^1$ và $\alpha, \beta \in \mathbb{R}$ ta có $\alpha X + \beta Y \in \mathcal{L}^1$ và*

$$\mathbb{E}(\alpha X + \beta Y) = \alpha \mathbb{E}X + \beta \mathbb{E}Y.$$

2. Nếu $0 \le X \le Y$ và $Y \in \mathcal{L}^1$ thì $X \in \mathcal{L}^1$ và $\mathbb{E}X \le \mathbb{E}Y$.

Ta có hệ quả như sau.

Hệ quả 2.3.8. *1. Nếu $Y \in \mathcal{L}^1$ và biến ngẫu nhiên X thỏa mãn $|X| \le Y$ thì $X \in \mathcal{L}^1$.*

2. Nếu $X \ge 0$ h.c.c và $\mathbb{E}(X) < \infty$ thì $X < \infty$ h.c.c.

3. Nếu $\mathbb{E}(|X|) = 0$ thì $X = 0$ h.c.c.

4. Nếu $VarX = 0$ thì $X = \mathbb{E}X$ h.c.c.

Định lý 2.3.9. *Nếu hai biến ngẫu nhiên khả tích X và Y bằng nhau hầu chắc chắn thì $\mathbb{E}X = \mathbb{E}Y$.*

Chú ý rằng quan hệ bằng nhau hầu chắc chắn là quan hệ tương đương, nên ta có thể xét không gian thương L^1 của không gian $\mathcal{L}^1$.

Nếu biến ngẫu nhiên X khả tích bình phương, tức $X \in \mathcal{L}^2$, ta có định nghĩa phương sai của X như trong công thức (2.2). Từ tính tuyến tính của kì vọng, ta có các tính chất sau của phương sai.

Mệnh đề 2.3.10. *Cho biến ngẫu nhiên $X \in \mathcal{L}^2$, khi đó*

1. *$DX = \mathbb{E}X^2 - (\mathbb{E}X)^2$.*

2. *Với mọi số thực $a, b \in \mathbb{R}$ thì $D(aX + b) = a^2 DX$.*

2.3.4 Định lí giới hạn

Định lý 2.3.11 (Định lí hội tụ đơn điệu). *Giả sử dãy biến ngẫu nhiên $(X_n)_{n\geq 1}$ là không âm và tăng hcc tới X, khi đó $\lim_{n\to\infty} \mathbb{E}X_n = \mathbb{E}X$ (kể cả khi $\mathbb{E}X = \infty$).*

Chứng minh. Với mỗi n gọi $(Y_{n,k})_{k\geq 1}$ là dãy biến ngẫu nhiên đơn giản tăng tới X_n và đặt $Z_k = \max_{n\leq k} Y_{n,k}$. Khi đó $(Z_k)_{k\geq 1}$ là dãy không giảm các biến ngẫu nhiên đơn giản không âm vậy nên tồn tại $Z = \lim_{k\to\infty} Z_k$. Lại có

$$Y_{n,k} \leq Z_k \leq X \quad \forall n \leq k$$

nên cho $k \to \infty$ ta được

$$X_n \leq Z \leq X \quad hcc.$$

Tiếp theo cho $n \to \infty$ ta được $Z = X$ hcc. Do kì vọng là toán tử dương, ta có

$$\mathbb{E}Y_{n,k} \leq \mathbb{E}Z_k \leq \mathbb{E}X_k \quad \forall n \leq k.$$

Cố định n và cho $k \to \infty$, theo Mệnh đề 2.3.4 ta được

$$\mathbb{E}X_n \leq \mathbb{E}Z \leq \lim_{k\to\infty} \mathbb{E}X_k.$$

Lại cho $n \to \infty$ ta được

$$\lim_{n\to\infty} \mathbb{E}X_n \leq \mathbb{E}Z \leq \lim_{k\to\infty} \mathbb{E}X_k.$$

Do biểu thức tận cùng bên trái và phải của công thức trên là bằng nhau, $X = Z$ hcc, ta được điều phải chứng minh. □

Hệ quả 2.3.12. *Cho biến ngẫu nhiên khả tích $X \in L^1$. Khi đó*

$$\lim_{M\to\infty} \mathbb{E}(|X|\mathbb{I}_{|X|>M}) = 0.$$

Chứng minh. Do $X \in L^1$ nên $\mathbb{E}|X| < \infty$. Xét dãy biến ngẫu nhiên

$$X_M = |X|\mathbb{I}_{|X|\leq M}.$$

Dễ thấy dãy này không âm và tăng h.c.c đến $|X|$. Áp dụng Định lý hội tụ đơn điệu, ta có

$$\mathbb{E}X_n = \mathbb{E}(|X|\mathbb{I}_{|X|\leq M}) \to \mathbb{E}|X|.$$

Suy ra $\mathbb{E}|X| - \mathbb{E}X_n = \mathbb{E}(|X|\mathbb{I}_{|X|>M}) \to 0.$ □

Từ đây ta có tiêu chuẩn kiểm tra tính khả tích của biến ngẫu nhiên như sau

Định lý 2.3.13. *Biến ngẫu nhiên X khả tích khi và chỉ khi $\mathbb{P}(|X| = \infty) = 0$ và với mọi $\epsilon > 0$ tồn tại $\delta_\epsilon > 0$ sao cho với mọi biến cố $A \in \mathcal{F}$ thỏa mãn $\mathbb{P}(A) \leq \delta_\epsilon$ ta đều có $\mathbb{E}(|X|\mathbb{I}_A) \leq \epsilon$.*

Chứng minh. Ta có

$$\mathbb{E}(|X|\mathbb{I}_A) = \mathbb{E}(|XbI_{|X|>M})|\mathbb{I}_A) + \mathbb{E}(|X|\mathbb{I}_{|X|\leq M})\mathbb{I}_A) \leq bE(|X|\mathbb{I}_{|X|>M})) + M\mathbb{P}(A).$$

Do đó khi X khả tích ta có thể chọn M đủ lớn sao cho $bE(|XbI_{|X|>M})|) \leq \epsilon/2$ và khi cố định M xong ta chọn $\delta_\epsilon = \epsilon/(2M)$.

Để chứng minh chiều ngược lại ta xét dãy biến cố

$$A_n = \{\omega \in \omega : |X(\omega)| > n\}.$$

Đây là dãy biến cố giảm nên ta có

$$\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(\bigcap_{n\in\mathbb{N}} A_n) = \mathbb{P}(|X| = \infty) = 0.$$

Do đó cố định ϵ, tồn tại n để $\mathbb{P}(A_n) \leq \delta_\epsilon$. Suy ra

$$\mathbb{E}|X| = \mathbb{E}(|X|\mathbb{I}_{|X|\leq M}) + \mathbb{E}(|X|\mathbb{I}_{A_n}) \leq n + \epsilon < \infty.$$

□

Định lý 2.3.14 (Bổ đề Fatou)**.** *Giả sử dãy biến ngẫu nhiên (X_n) thỏa mãn $X_n \geq Y$ h.c.c với mọi n và $Y \in L^1$ nào đó. Khi đó*

$$\mathbb{E}(\liminf_{n\to\infty} X_n) \leq \liminf_{n\to\infty} \mathbb{E}X_n.$$

Chứng minh. Trước hết ta chứng minh Định lí cho trường hợp $Y = 0$. Đặt $Y_n = \inf_{k\geq n} X_k$. Ta có (Y_n) là dãy biến ngẫu nhiên không giảm và

$$\lim_{n\to\infty} Y_n = \liminf_{n\to\infty} X_n.$$

Do $X_n \geq Y_n$, ta có $\mathbb{E}X_n \geq \mathbb{E}Y_n$. Kết hợp với định lí hội tụ đơn điệu cho dãy Y_n, ta được

$$\liminf_{n\to\infty} \mathbb{E}X_n \geq \lim_{n\to\infty} \mathbb{E}Y_n = \mathbb{E}(\lim_{n\to\infty} Y_n) = \mathbb{E}(\liminf_{n\to\infty} X_n).$$

Trường hợp tổng quát suy ra từ việc áp dụng kết quả vừa chứng minh cho dãy biến ngẫu nhiên không âm $\hat{X}_n := X_n - Y$. □

Định lý 2.3.15 (Định lí hội tụ bị chặn Lebesgue)**.** *Giả sử dãy biến ngẫu nhiên X_n hội tụ hcc đến biến ngẫu nhiên X và tồn tại $Y \in L^1$ sao cho $|X_n| \le Y$ hcc. Khi đó $X, X_n \in L^1$ và*

$$\lim_{n\to\infty} \mathbb{E}(|X_n - X|) = 0.$$

Chứng minh. Vì $|X| \le Y$ nên $X \in L^1$. Đặt $Z_n = |X_n - X|$. Vì $Z_n \ge 0$ và $-Z_n \ge -2Y$, áp dụng Bổ đề Fatou cho Z_n và $-Z_n$, ta được

$$0 = \mathbb{E}(\liminf_{n\to\infty} Z_n) \le \liminf_{n\to\infty} \mathbb{E}Z_n \le \limsup_{n\to\infty} \mathbb{E}Z_n = -\liminf_{n\to\infty} \mathbb{E}(-Z_n) \le -\mathbb{E}(\liminf_{n\to\infty}(-Z_n)) = 0.$$

Vậy $\lim_{n\to\infty} \mathbb{E}Z_n = 0$ hay $\lim_{n\to\infty} \mathbb{E}(|X_n - X|) = 0$. □

2.3.5 Một số bất đẳng thức

Trong mục này chúng ta trình bày một số bất đẳng thức liên quan đến kì vọng của biến ngẫu nhiên. Đầu tiên là ba bất đẳng thức Markov, Chebyshev và dạng mũ. Đây là ba bất đẳng thức thông dụng, sẽ được sử dụng nhiều ở Chương 3.

Định lý 2.3.16. *1. (Bất đẳng thức Markov) Giả sử $X \in L^1$, khi đó với mọi $a > 0$,*

$$\mathbb{P}(|X| \ge a) \le \frac{\mathbb{E}(|X|)}{a}.$$

2. (Bất đẳng thức Chebyshev) Giả sử $X \in L^2$, khi đó với mọi $a > 0$,

$$\mathbb{P}(|X - \mathbb{E}X| \ge a) \le \frac{DX}{a^2}.$$

3. Cho trước số thực $a \in \mathbb{R}$, khi đó với mọi $\theta > 0$,

$$\mathbb{P}(X > a) \le e^{-\theta a}\mathbb{E}(e^{\theta X}).$$

Chứng minh. 1) Do $a\mathbb{I}_{\{|X|\ge a\}}(w) \le |X(w)|\mathbb{I}_{\{|X|\ge a\}}(w) \le |X(w)|$ với mọi $w \in \Omega$. Lấy kì vọng hai vế ta được $a\mathbb{P}(|X| \ge a) \le \mathbb{E}(|X|)$.

2) Áp dụng Bất đẳng thức Markov, ta có

$$\mathbb{P}(|X - \mathbb{E}X| \ge a) = \mathbb{P}(|X - \mathbb{E}X|^2 \ge a^2) \le \frac{DX}{a^2}.$$

3) Ta có các biểu diễn biên cố như sau

$$\{X > a\} = \{\theta X > \theta a\} = \{e^{\theta X} > e^{\theta a}\}.$$

Áp dụng bất đẳng thức Markov cho biến ngẫu nhiên $e^{\theta X}$ ta có ĐPCM. □

Tiếp theo ta có các bất đẳng thức Cauchy-Schwarz, Holder, Lyapunov, Minkowski và Jensen. Trước tiên ta có định nghĩa chuẩn bậc p như sau.

Định nghĩa 2.3.17. *Cho biến ngẫu nhiên $X \in \mathcal{L}^p(\omega, \mathcal{F}, \mathbb{P})$. Khi đó, chuẩn bậc p của X được định nghĩa như sau*

$$\|X\|_p = (\mathbb{E}|X|^p)^{1/p}.$$

Định lý 2.3.18 (Bất đẳng thức Cauchy-Schwarz)**.** *Cho hai biến ngẫu nhiên $X, Y \in \mathcal{L}^2$. Khi đó*

$$\mathbb{E}|XY| \leq \sqrt{\mathbb{E}X^2.\mathbb{E}Y^2} = \|X\|_2\|Y\|_2.$$

Chứng minh. Sử dụng bất đẳng thức sơ cấp $2|ab| \leq a^2 + b^2$, thay

$$a = \frac{X}{\|X\|_2}, b = \frac{Y}{\|Y\|_2},$$

khi hai chuẩn này dương và hữu hạn, và lấy kì vọng hai vế. □

Từ bất đẳng thức Cauchy-Schwarz, xét $Y(\omega) = 1\ \forall \omega \in \Omega$, ta có hệ quả sau.

Hệ quả 2.3.19. *Ta có*

1. $(\mathbb{E}|X|)^2 \leq \mathbb{E}X^2$.
2. *Nếu $X \in \mathcal{L}^2$ thì $X \in \mathcal{L}^1$, hay không gian $\mathcal{L}^2$ là không gian con của không gian $\mathcal{L}^1$.*
3. *$\mathcal{L}^2$ là một không gian vector: nếu $X, Y \in \mathcal{L}^2$ và $\alpha, \beta \in \mathbb{R}$ thì $\alpha X + \beta Y \in \mathcal{L}^2$.*

Định lý 2.3.20 (Bất đẳng thức Holder)**.** *Cho hai số thực $p, q > 1$ sao cho*

$$\frac{1}{p} + \frac{1}{q} = 1,$$

và hai biến ngẫu nhiên $X \in \mathcal{L}^p$ và $Y \in \mathcal{L}^q$. Khi đó

$$\mathbb{E}|XY| \leq \|X\|_p\|Y\|_q.$$

Chứng minh. Sử dụng bất đẳng thức Young

$$\frac{|a|^p}{p} + \frac{|b|^q}{q} \geq |ab|,$$

thay

$$a = \frac{X}{\|X\|_2}, b = \frac{Y}{\|Y\|_2},$$

khi hai chuẩn này dương và hữu hạn, và lấy kì vọng hai vế. □

Với $p = q = 2$, ta suy ra lại bất đẳng thức Cauchy-Schwarz. Ngoài ra cho $Y = 1$ h.c.c, $X = |Z|^s$ và $p = t/s$ với hai số thực dương $0 < s < t$, từ bất đẳng thức Holder ta có

$$\mathbb{E}|Z|^s = \mathbb{E}|X| \le \|X\|_p = (\mathbb{E}|Z|^t)^{s/t},$$

với điều kiện $Z \in \mathcal{L}^t$. Viết lại bất đẳng thức trên ta có

Định lý 2.3.21 (Bất đẳng thức Lyapunov). *Cho biến ngẫu nhiên $Z \in \mathcal{L}^t$. Khi đó, với mọi $0 < s < t$,*

$$\|X\|_s \le \|X\|_t.$$

Suy ra $\mathcal{L}^t$ là không gian con của $\mathcal{L}^s$.

Định lý 2.3.22 (Bất đẳng thức Minkowski). *Cho số thực $p \ge 1$ và hai biến ngẫu nhiên $X, Y \in \mathcal{L}^p$. Khi đó $X + Y \in \mathcal{L}^p$ và*

$$\|X + Y\|_p \le \|X\|_p + \|Y\|_p.$$

Chứng minh. Với $p = 1$ dễ thấy kết quả đúng. Với $p > 1$, áp dụng bất đẳng thức sơ cấp

$$(a + b)^p \le 2^{p-1}(a^p + b^p),\ \forall p \ge 1,\ a, b > 0,$$

ta suy ra $X + Y \in \mathcal{L}^p$. Ngoài ra, ta có

$$\begin{aligned}
\mathbb{E}|X+Y|^p =& \mathbb{E}(|X+Y||X+Y|^{p-1}) \\
\le& \mathbb{E}(|X||X+Y|^{p-1}) + \mathbb{E}(|Y||X+Y|^{p-1}) \\
\le& (\mathbb{E}|X|^p)^{1/p} \left(\mathbb{E}|X+Y|^{(p-1)q}\right)^{1/q} + (\mathbb{E}|Y|^p)^{1/p} \left(\mathbb{E}|X+Y|^{(p-1)q}\right)^{1/q} \\
=& (\|X\|_p + \|Y\|_p) \left(\mathbb{E}|X+Y|^p\right)^{1/q},
\end{aligned}$$

với dòng thứ ba áp dụng bất đẳng thức Holder và dòng cuối suy ra từ đẳng thức $(p-1)q = p$. Từ đây suy ra ĐPCM. □

Ta thấy rằng với $p \ge 1$ thì $\mathcal{L}^p$ là không gian vector. Khi $0 < p < 1$, khẳng định này vẫn đúng từ nhận xét

$$(a + b)^p \le a^p + b^p,\ \forall p < 1,\ a, b > 0.$$

Định lý 2.3.23 (Bất đẳng thức Jensen). *Cho hàm $\varphi : \mathbb{R} \to \mathbb{R}$ là hàm lồi theo nghĩa*

$$\varphi(\lambda x + (1-\lambda)y) \le \lambda\varphi(x) + (1-\lambda)\varphi(y),\ \forall x, y \in \mathbb{R},\ \lambda \in [0, 1].$$

Cho biến ngẫu nhiên X sao cho cả X và $\varphi(X)$ đều khả tích. Khi đó

$$\mathbb{E}(\varphi(X)) \ge \varphi(\mathbb{E}X).$$

Chứng minh. Theo tính chất của hàm lồi, với mọi $x_0 \in \mathbb{R}$ tồn tại số thực k_0 sao cho

$$\varphi(x) \ge \varphi(x_0) + k_0(x - x_0), \forall x \in \mathbb{R}.$$

Thay x_0 bởi $\mathbb{E}X$, $x = X$ và lấy kì vọng hai vế, ta có ĐPCM. □

2.3.6 Kì vọng của biến ngẫu nhiên có phân phối liên tục tuyệt đối

Trước tiên, ta sẽ nhắc lại cách tính kì vọng cho biến ngẫu nhiên có phân phối rời rạc như sau.

Mệnh đề 2.3.24. *Cho biến ngẫu nhiên X có phân phối rời rạc với bảng phân phối*

$$\begin{array}{c|ccccc} X & x_1 & x_2 & \dots & x_n & \dots \\ \hline \mathbb{P} & p_1 & p_2 & \dots & p_n & \dots \end{array}.$$

Khi đó, nếu X khả tích, ta có

$$\mathbb{E}X = \sum_i x_i.p_i.$$

Một cách tổng quát, cho hàm Borel $\varphi : \mathbb{R} \to \mathbb{R}$, nếu biến ngẫu nhiên rời rạc $\varphi(X)$ khả tích, ta có

$$\mathbb{E}[\varphi(X)] = \sum_i \varphi(x_i).p_i.$$

Với biến ngẫu nhiên có phân phối liên tục, ta có cách tính kì vọng thông qua hàm mật độ như sau.

Định lý 2.3.25. *Giả sử biến ngẫu nhiên X có hàm mật độ f và $h : \mathbb{R} \to \mathbb{R}$ là một hàm Borel. Khi đó nếu $\mathbb{E}(|h(X)|) < \infty$ hoặc h là không âm thì*

$$\mathbb{E}(h(X)) = \int_{\mathbb{R}} h(x)f(x)dx.$$

Chứng minh. Trước hết ta xét trường hợp $h \geq 0$. Khi đó tồn tại dãy các hàm Borel đơn giản, không âm (h_n) tăng dần tới h. Giả sử $h_n = \sum_{i=1}^{k_n} a_i^n \mathbb{I}_{A_i^n}$ với $a_i^n \in \mathbb{R}^+$ và $A_i^n \in \mathcal{B}(\mathbb{R})$ với mọi i. Theo định lí hội tụ đơn điệu

$$\mathbb{E}(h(X)) = \mathbb{E}(\lim_n h_n(X)) = \lim_n \mathbb{E}(h_n(X)) = \lim_n \sum_{i=1}^{k_n} h_n(a_i^n)\mathbb{P}[X \in A_i^n].$$

Áp dụng tính chất (2.1) và định lí hội tụ đơn điệu, ta được

$$\mathbb{E}(h(X)) = \lim_n \sum_{i=1}^{k_n} h_n(a_i^n) \int_{A_i^n} f(x)dx = \lim_n \int f(x)h_n(x)dx = \int f(x)h(x)dx.$$

Vậy khi h không âm ta luôn có $\mathbb{E}(h(X)) = \int h(x)f(x)dx$.

Trong trường hợp tổng quát, áp dụng kết quả trên cho h^+ và h^- ta được điều phải chứng minh. $\square$

2.3.7 Một số ví dụ

Trong phần này, chúng ta sẽ tìm hiểu về giá trị cụ thể của kì vọng, phương sai, hay moment của một số dạng phân phối quen thuộc.

Phân phối Rademacher

Ta có $\mathbb{E}X = 0,\ DX = 1$, và

$$\mathbb{E}X^n = \begin{cases} 1 \text{ nếu } n \text{ chẵn,} \\ 0 \text{ nếu } n \text{ lẻ.} \end{cases} .$$

Phân phối Bernoulli $Ber(p)$

Ta có $\mathbb{E}X = \mathbb{E}X^n = p$, và $DX = p(1-p)$.

Phân phối nhị thức $B(n, p)$

Ta có $\mathbb{E}X = n.p$ và $DX = n.p(1-p)$.

Phân phối Poisson $Poi(\lambda)$

Ta có $\mathbb{E}X = DX = \lambda$.

Phân phối hình học $Geo(p)$

Ta có $\mathbb{E}X = \dfrac{1}{p}$ và $DX = \dfrac{1-p}{p^2}$.

Phân phối đều $U[a, b]$

Ta có $\mathbb{E}X = \dfrac{a+b}{2}$, tổng quát

$$\mathbb{E}X^n = \frac{b^{n+1} - a^{n+1}}{(n+1)(b-a)},$$

và $DX = \dfrac{(b-a)^2}{12}$.

Phân phối mũ $Exp(\lambda)$

Ta có $\mathbb{E}X = \lambda$ và $DX = \lambda^2$.

Phân phối chuẩn $\mathcal{N}(a, \sigma^2)$

Với biến ngẫu nhiên $X \sim \mathcal{N}(0,1)$ có phân phối chuẩn tắc, ta có $\mathbb{E}X = 0$, $DX = 1$ và

$$\mathbb{E}X^n = \begin{cases} \dfrac{(2K)!}{2^K K!} = 1.3.\ldots.(2K-1) \text{ nếu } n = 2K \text{ chẵn}, \\ 0 \text{ nếu } n \text{ lẻ}. \end{cases}.$$

Với biến ngẫu nhiên $X \sim \mathcal{N}(a, \sigma^2)$, ta có $\mathbb{E}X = a$, $DX = \sigma^2$.

Phân phối Gamma $\mathcal{G}(\alpha, p)$

Ta có $\mathbb{E}X = \dfrac{p}{\alpha}$ và $DX = \dfrac{p}{p\alpha^2}$.

2.4 Phần tử ngẫu nhiên và Vector ngãu nhiên

2.4.1 Định nghĩa

Tổng quát hóa khái niệm biến ngẫu nhiên ta có khái niệm phần tử ngẫu nhiên như sau.

Định nghĩa 2.4.1. *Giả sử* $(E, \mathcal{E})$ *là một không gian đo. Ánh xạ* $X : \Omega \to E$ *được gọi là* $\mathcal{F}/\mathcal{E}$-đo được *hay là một* phần tử ngẫu nhiên *nếu* $X^{-1}(B) \in \mathcal{A}$ *với mọi* $B \in \mathcal{E}$. *Hàm tập*

$$\mathbb{P}^X(B) = \mathbb{P}(X^{-1}(B)), \quad B \in \mathcal{E},$$

được gọi là phân phối xác suất *của* X *trên* $(E, \mathcal{E})$.

Thông thường, ta sẽ chọn các không gian $(E, \mathcal{E})$ có các cấu trúc topo tương đối đẹp. Một cách tổng quát ta thường chọn E là không gian Banach với topo sinh bởi chuẩn. Ngoài ra, ta có thể xét một số ví dụ như sau:

- Cho $E = Mat_d(\mathbb{R})$ hoặc $Mat_d(\mathbb{C})$, tức là tập các ma trận vuông với phần tử thực (hoặc phức), ta có *Ma trận ngẫu nhiên.*

- Cho $E = \mathbb{R}^{\mathbb{N}}$,ta có *Quá trình ngẫu nhiên rời rạc.*

- Cho $E = C[0, T]$, tức là tập các hàm liên tục trên đoạn $[0, T]$, ta có *Quá trình ngẫu nhiên liên tục.*

- Cho $E = \mathbb{R}^d$, ta có *Vector ngẫu nhiên.*

Trong phần tiếp theo, ta sẽ tìm hiểu kĩ hơn về vector ngẫu nhiên. Ta có định nghĩa chính xác như sau.

Định nghĩa 2.4.2. *Một vector ngẫu nhiên là một ánh xạ* $X : (\Omega, \mathcal{F}) \to (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ *thỏa mãn tính đo được. Với mọi tập Borel* $B \in \mathcal{B}(\mathbb{R}^d)$*, ta có thể xác định giá trị*

$$\mu_X(B) = \mathbb{P}(\omega \in \Omega : X(\omega) \in B) = \mathbb{P}(X^{-1}(B)),$$

sinh ra độ đo μ_X *trên* $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$*. Độ đo này còn được gọi là phân phối xác suất của vector ngẫu nhiên* X*.*

Về mặt ý nghĩa, một vector ngẫu nhiên là cách đánh giá không gian Ω theo nhiều góc độ. Xét ví dụ Ω là các học sinh trong một lớp, và ta có các biến ngẫu nhiên là điểm các môn khác nhau như Toán, Lý, Hóa, Tiếng Anh, ...

Cũng giống như đã xây dựng cho biến ngẫu nhiên, ta thấy rằng Borel σ-đại số $\mathcal{B}(\mathbb{R}^d)$ sinh bởi các tập có dạng tích

$$(-\infty, x_1) \times \ldots \times (-\infty, x_d),$$

nên phân phối xác suất μ_X sẽ xác định duy nhất nếu ta biết giá trị của độ đo tại các tập tích trên. Từ đó mở rộng khái niệm hàm phân phối của biến ngẫu nhiên, ta có khái niệm *hàm phân phối đồng thời* (gọi tắt là hàm phân phối) của vector ngẫu nhiên như sau.

Định nghĩa 2.4.3. *Cho vector ngẫu nhiên* X*, hàm* $F_X : \mathbb{R}^d \to \mathbb{R}$ *cho bởi*

$$F_X(x_1, \ldots, x_d) = \mathbb{P}(X^{-1}((-\infty, x_1) \times \ldots \times (-\infty, x_d))),$$

được gọi là hàm phân phối của X*.*

Mệnh đề 2.4.4. *Với vector ngẫu nhiên* X*, ta xác định* X_i *là thành phần thứ* i *của* X*. Khi đó* X_i *là biến ngẫu nhiên với mọi* $i = 1, 2, \ldots, n$*.*

Định nghĩa 2.4.5. *Cho vector ngẫu nhiên* X*.*

1. *Các phân phối của các thành phần* X_i *được gọi là phân phối biên duyên (marginal) của* X*.*

2. *Với biểu diễn* $X = (X_1, \ldots, X_d)$*, ta viết lại hàm phân phối đồng thời như sau*

$$F_X(x) = \mathbb{P}[X < x] = \mathbb{P}[X_1 < x_1, \ldots, X_d < x_d], \quad x \in \mathbb{R}^d,$$

trong đó kí hiệu $x < y$ *với* $x, y \in \mathbb{R}^d$ *được hiểu là* $x_i < y_i$ *với mọi* $i = 1, \ldots, d$*.*

Hàm phân phối đồng thời F_X có các tính chất sau:

1. $0 \le F_X(x) \le 1$ với mọi $x \in \mathbb{R}^d$.

2. $\lim_{x_k \to -\infty} F_X(x) = 0$ với mọi $k = 1, \ldots, d$.

3. $\lim_{x_1 \to +\infty, \ldots, x_d \to +\infty} F_X(x) = 1$.

4. Hàm F_X liên tục trái.

Ta cũng có định nghĩa về hàm mật độ đồng thời như sau.

Định nghĩa 2.4.6. *Vector ngẫu nhiên X được gọi là có* phân phối liên tục tuyệt đối với hàm mật độ $f : \mathbb{R}^d \to \mathbb{R}_{\geq 0}$ *nếu*

$$F_X(x) = \int_{u<x} f(u)du, \qquad \text{với mọi } x \in \mathbb{R}^d.$$

Từ định nghĩa này ta có ngay

$$\mathbb{P}[X \in B] = \int_B f(x)dx, \qquad \text{với mọi } B \in \mathcal{B}(\mathbb{R}^d).$$

Đặc biệt, ta có

$$\mathbb{P}[X_1 \in B_1] = \mathbb{P}[X \in B_1 \times \mathbb{R}^{d-1}] = \int_{B_1} \Big(\int_{\mathbb{R}^{d-1}} f(x_1, \ldots, x_d) dx_2 \ldots dx_d \Big) dx_1 \text{ với mọi } B_1 \in \mathcal{B}(\mathbb{R}^d).$$

Do đó nếu $X = (X_1, \ldots, X_d)$ là vector ngẫu nhiên có phân phối liên tục tuyệt đối với hàm mật độ f thì biến ngẫu nhiên X_1 cũng có phân phối liên tục tuyệt đối với hàm mật độ là

$$f_{X_1}(x_1) = \int_{\mathbb{R}^{d-1}} f(x_1, x_2, \ldots, x_d) dx_2 \ldots dx_d, \quad \text{với mọi } x_1 \in \mathbb{R}. \tag{2.7}$$

Tương tự như Định lí 2.3.25, ta có kết quả sau.

Định lý 2.4.7. *Giả sử $X = (X_1, \ldots, X_d)$ là vector ngẫu nhiên có phân phối liên tục tuyệt đối với hàm mật độ f và $\varphi : \mathbb{R}^d \to \mathbb{R}$ là hàm đo được. Khi đó*

$$\mathbb{E}[\varphi(X)] = \int_{\mathbb{R}^d} \varphi(x) f(x) dx$$

nếu φ không âm hoặc $\int_{\mathbb{R}^d} |\varphi(x)| f(x) dx < \infty$.

Tiếp theo ta có hai ví dụ về vector ngẫu nhiên với hai dạng phân phối liên tục và rời rạc như sau.

Phân phối chuẩn

Giả sử $a = (a_1, \dots, a_d)$ là vector d-chiều và $M = (m_{i,j})_{i,j=1}^d$ là ma trận vuông cấp d, đối xứng và xác định dương. Ta nói vector ngẫu nhiên $X = (X_1, \dots, X_d)$ có phân phối chuẩn $\mathcal{N}(a, M)$ nếu mật độ của X có dạng

$$p(x) = \frac{1\sqrt{\det M}}{(2\pi)^{d/2}\sqrt{\det M}} \exp\Big\{ -\frac{1}{2}(x-a)M^{-1}(x-a)^*\Big\}.$$

Phân phối đa thức

vector ngẫu nhiên d-chiều X được gọi là có phân phối đa thức với tham số $n, p_1, \dots, p_d$, kí hiệu là $X \sim MUT(n; p_1, \dots, p_d)$, với $n \in \mathbb{N}^*$ và $p_1, \dots, p_d \geq 0$ nếu

$$\mathbb{P}[X_1 = k_1, \dots, X_d = k_d] = \frac{n!}{k_1!k_2!\dots k_{d+1}!} p_1^{k_1} p_2^{k_2} \dots p_{d+1}^{k_{d+1}},$$

trong đó $p_{d+1} = 1 - (p_1 + \dots + p_d),\ 0 \leq k_i \leq n$ và $k_{d+1} = n - (k_1 + \dots + k_d) \geq 0$.

2.4.2 Phân phối của hàm của vector ngẫu nhiên

Định lí sau là hệ quả trực tiếp của Định lí 2.4.7 và công thức đổi biến số Jacobi trong giải tích.

Định lý 2.4.8. *Giả sử vector ngẫu nhiên $X = (X_1, \dots, X_d)$ có mật độ f_X. Giả sử $g : \mathbb{R}^d \to \mathbb{R}^d$ là đơn ánh khả vi liên tục với ma trận Jacobi*

$$J_g(x) = \left(\frac{\partial g_i}{\partial x_j}(x)\right)_{i,j=1,\dots,d}$$

không suy biến với mọi $x \in \mathbb{R}^d$. Khi đó vector ngẫu nhiên $Y = g(X)$ có mật độ

$$f_Y(y) = \begin{cases} f_X(g^{-1}(y))|\det J_{g^{-1}}(y)| & \textit{nếu } y \in g(\mathbb{R}^d) \\ 0 & \textit{nếu } y \notin g(\mathbb{R}^d). \end{cases}$$

Sử dụng định lý trên, ta có các vi dụ sau về phân phối của hàm của hai biến ngẫu nhiên chuẩn tắc độc lập.

Ví dụ 2.4.9. *1. Cho X và Y là hai biến ngẫu nhiên độc lập với cùng phân phối chuẩn tắc $N(0,1)$. Khi đó $U = X + Y$ và $V = X - Y$ cũng là hai biến ngẫu nhiên độc lập với cùng phân phối chuẩn $N(0,2)$.*

Thật vậy, đặt $g(x,y) = (x+y, x-y)$ *ta có* $g^{-1}(u,v) = ((u+v)/2, (u-v)/2)$ *và*

$$J_{g^{-1}}(u,v) = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & -1/2 \end{pmatrix}.$$

Áp dụng Định lí 2.4.8 ta được

$$f_{(U,V)}(u,v) = f_{(X,Y)}\Big(\frac{u+v}{2}, \frac{u-v}{2}\Big)\Big|\det J_{g^{-1}}(u,v)\Big|.$$

Áp dụng Hệ quả 2.5.5, ta được

$$f_{(U,V)}(u,v) = \frac{1}{2} f_X\Big(\frac{u+v}{2}\Big) f_Y\Big(\frac{u-v}{2}\Big) = \frac{1}{2}\frac{e^{-(u+v)^2/8}}{\sqrt{2\pi}}\frac{e^{-(u-v)^2/8}}{\sqrt{2\pi}} = \frac{1}{4\pi}e^{-(u^2+v^2)/4}.$$

Áp dụng công thức (2.7) *ta có mật độ của* U *là*

$$f_U(u) = \int_{\mathbb{R}} f_{(U,V)}(u,v)dv = \frac{e^{-u^2/4}}{\sqrt{4\pi}}.$$

Tương tự mật độ của V *là*

$$f_V(v) = \int_{\mathbb{R}} f_{(U,V)}(u,v)du = \frac{e^{-v^2/4}}{\sqrt{4\pi}}.$$

Do đó U *và* V *cùng có phân phối chuẩn* $N(0,2)$ *và hơn nữa*

$$f_{(U,V)}(u,v) = f_U(u)f_V(v), \quad \text{với mọi } u, v \in \mathbb{R},$$

vậy nên theo Hệ quả 2.5.5 X *độc lập với* Y.

2. *Tiếp theo, xét biến ngẫu nhiên*

$$Z = \begin{cases} X/Y & \text{nếu } Y \neq 0 \\ 0 & \text{nếu } Y = 0. \end{cases}$$

Để tìm phân phối của Z, *ta đặt* $g(x,y) = (x/y, y)$, *ta có* $g^{-1}(z,t) = (zt, t)$ *và*

$$J_{g^{-1}}(z,t) = \begin{pmatrix} t & z \\ 0 & 1 \end{pmatrix}.$$

Áp dụng Định lí 2.4.8 và Hệ quả 2.5.5 ta được

$$f_{(Z,Y)}(z,t) = f_{(X,Y)}(zt,t)|t| = |t|\frac{e^{t^2(z^2+1)/2}}{2\pi}.$$

Do đó, hàm mật độ của Z *là*

$$f_Z(z) = \int_{\mathbb{R}} |t|\frac{e^{t^2(z^2+1)/2}}{2\pi}dt = \frac{1}{\pi(1+z^2)}.$$

biến ngẫu nhiên Z *được gọi là có phân phối Cauchy.*

3. *Xét biến ngẫu nhiên $\chi^2 = X^2 + Y^2$. Với mọi tập Borel $A \in \mathcal{B}([0,\infty))$, áp dụng Định lí 2.4.7 ta có*

$$\mathbb{P}(\chi^2 \in A) = \mathbb{E}(\mathbb{I}_A(\chi^2)) = \int_{\mathbb{R}^2} \mathbb{I}_A(x^2+y^2) f_{(X,Y)}(x,y)dxdy.$$

Sử dụng công thức tọa độ cực $x = r\cos\varphi,\ y = r\sin\varphi$ ta được

$$\mathbb{P}(\chi^2 \in A) = \int_0^\infty dr \int_0^{2\pi} \mathbb{I}_A(r^2) r \frac{e^{-r^2/2}}{2\pi} d\varphi = \int_0^\infty \mathbb{I}_A(r^2) r e^{-r^2/2} dr = \int_A \frac{e^{-z/2}}{2} dz,$$

trong đó đẳng thức cuối cùng có được từ phép đổi biến $z = r^2$. Vậy χ^2 có phân phối Gamma với hàm mật độ là

$$f_{\chi^2}(z) = \begin{cases} \frac{e^{-z/2}}{2} & \text{nếu } z > 0 \\ 0 & \text{nếu } z \le 0. \end{cases}$$

2.5 Biến ngẫu nhiên độc lập

Định nghĩa 2.5.1. 1. *Các σ-đại số con $(\mathcal{A}_i)_{i\in I}$ của $\mathcal{F}$ được gọi là* độc lập *nếu với mọi tập con hữu hạn J của I và với mọi $A_i \in \mathcal{A}_i$ với $i \in J$, ta có*

$$\mathbb{P}(\cap_{i\in J} A_i) = \prod_{i\in J} \mathbb{P}(A_i).$$

2. *Các biến ngẫu nhiên $(X_i)_{i\in I}$ được gọi là* độc lập *nếu các σ-đại số sinh bởi các biến ngẫu nhiên này $(\sigma(X_i))_{i\in I}$ là độc lập.*

Ta có ví dụ đơn giản sau về họ các biến ngẫu nhiên độc lập

Ví dụ 2.5.2. *Cho họ các biến cố độc lập $\{A_i,\, i \in I\}$. Khi đó họ các biến ngẫu nhiên hàm chỉ báo $\{\mathbb{I}_{A_i},\, i \in I\}$ là các biến ngẫu nhiên độc lập.*

Cũng từ việc xét các biến ngẫu nhiên chỉ báo, ta thấy rằng, một họ các biến ngẫu nhiên đôi một độc lập chưa chắc đã là họ độc lập toàn thể.

Sử dụng Định lý 1.2.15 của Dynkin về$\pi - \lambda$ hệ, ta có tiêu chuẩn kiểm tra tính độc lập của các σ-đại số như sau.

Định lý 2.5.3. *Với mỗi $i = 1, \ldots, n$, xét một π-hệ $\mathcal{C}_i$ gồm các biến cố thuộc $\mathcal{F}$ sao cho các hệ $\{\mathcal{C}_i, 1 \le i \le n\}$ là độc lập. Khi đó các σ-đại số $\{\sigma(\mathcal{C}_i), 1 \le i \le n\}$ là họ các σ-đại số độc lập.*

Chứng minh. Ta trình bày chứng minh cho $n = 2$. Cố định biến cố $A_2 \in \mathcal{C}_2$ và xét họ

$$\mathcal{L} = \{A \in \mathcal{F} : \mathbb{P}(A \cap A_2)\mathbb{P}(A)\mathbb{P}(A_2)\}.$$

Ta có $\mathcal{C}_1 \subset \mathcal{L}$ và $\mathcal{L}$ là một λ-hệ. Nên theo Định lý Dynkin, $\sigma(\mathcal{C}_1) \subset \mathcal{L}$, suy ra σ-đại số $\sigma(\mathcal{C}_1)$ độc lập với hệ $\mathcal{C}_2$. Ta tiếp tục cố định $A_1 \in \sigma(\mathcal{C}_1)$ và làm tương tự như trên, suy ra ĐPCM □

Hệ quả 2.5.4. *Các biến ngẫu nhiên $X_1, \ldots, X_n$ độc lập nếu như với mọi $x_1, \ldots, x_n \in \mathbb{R}$ đều có*

$$\mathbb{P}(X_1 < x_1, \ldots, X_n < x_n) = pr(X_1 < x1)\mathbb{P}(X_2 < x_2) \ldots \mathbb{P}(X_< x_n).$$

Hay nói cách khác, các biến ngẫu nhiên $X_1, \ldots, X_n$ độc lập nếu như hàm phân phối đồng thời có thể biểu diễn thành tích các hàm phân phối thành phần.

Hệ quả 2.5.5. *Giả sử $(X_1, \ldots, X_n)$ là vector ngẫu nhiên có phân phối liên tục tuyệt đối. Khi đó các biến ngẫu nhiên thành phần $X_1, X_2, \ldots, X_n$ là độc lập khi và chỉ khi*

$$f_{(X_1,\ldots,X_n)}(x_1, \ldots, x_n) = f_{X_1}(x_1)f_{X_2}(x_2) \ldots f_{X_n}(x_n), \quad \text{với mọi } x_1, \ldots x_n \in \mathbb{R}. \tag{2.8}$$

Ta cũng có thể kiểm tra tính độc lâp của hai biến ngẫu nhiên qua kì vọng như sau.

Định lý 2.5.6. *Giả sử X và Y là hai biến ngẫu nhiên bất kì. Các khẳng định sau là tương đương.*

(i) X *độc lập với* Y*;*

(ii) $F_{X,Y}(x, y) = F_X(x)F_Y(y)$ *với mọi* $x, y \in \mathbb{R}$*;*

(iii) $f(X)$ *và* $g(Y)$ *là độc lập với mọi hàm Borel* $f, g : \mathbb{R} \to \mathbb{R}$*;*

(iv) $\mathbb{E}[f(X)g(Y)] = \mathbb{E}[f(X)]\mathbb{E}[g(Y)]$ *với mọi hàm Borel* $f, g : \mathbb{R} \to \mathbb{R}$ *cùng bị chặn hoặc cùng không âm.*

Chứng minh. Ta đã có $(i) \Leftrightarrow (ii)$.

$(i) \Rightarrow (iii)$: Với mọi $A, B \in \mathcal{B}(\mathbb{R})$, ta có $f^{-1}(A), g^{-1}(B) \in \mathcal{B}(\mathbb{R})$ nên

$$\begin{aligned}&\mathbb{P}(f(X) \in A, g(Y) \in B) = \mathbb{P}(X \in f^{-1}(A), Y \in g^{-1}(B)) \\ &= \mathbb{P}(X \in f^{-1}(A))\mathbb{P}(Y \in g^{-1}(B)) = \mathbb{P}(f(X) \in A)\mathbb{P}(g(Y) \in B).\end{aligned}$$

Vậy nên $f(X)$ độc lập với $g(Y)$.

$(iii) \Rightarrow (i)$: Ta chỉ cần chọn $f(x) = g(x) = x$.

$(i) \Rightarrow (iv)$: Do (i) tương đương với (iii) ta chỉ cần chứng tỏ

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y) \text{ với mọi biến ngẫu nhiên khả tích hoặc không âm } X \text{ và } Y.$$

Trước hết ta giả sử X và Y không âm. Theo Định lí 2.1.4 tồn tại dãy biến ngẫu nhiên đơn giản $X_n = \sum_{i=1}^{k_n} a_i \mathbb{I}_{A_i}$ tăng đến X và dãy biến ngẫu nhiên đơn giản $Y_n = \sum_{j=1}^{l_n} b_j \mathbb{I}_{B_j}$ tăng đến Y với $A_i \in \sigma(X)$ và $B_i \in \sigma(Y)$. Áp dụng định lí hội tụ đơn điệu, ta có

$$\begin{aligned}
\mathbb{E}(XY) = \lim_{n\to\infty} \mathbb{E}(X_nY_n) &= \lim_{n\to\infty} \sum_{i=1}^{k_n}\sum_{j=1}^{l_n} a_ib_j\mathbb{P}(A_iB_j) = \lim_{n\to\infty} \sum_{i=1}^{k_n}\sum_{j=1}^{l_n} a_ib_j\mathbb{P}(A_i)\mathbb{P}(B_j) \\
&= \lim_{n\to\infty} \Big(\sum_{i=1}^{k_n} a_i\mathbb{P}(A_i)\Big)\Big(\sum_{j=1}^{l_n} b_j\mathbb{P}(B_j)\Big) = \lim_{n\to\infty} \mathbb{E}(X_n)\mathbb{E}(Y_n) = \mathbb{E}(X)\mathbb{E}(Y).
\end{aligned}$$

Trong trường hợp tổng quát, ta viết $X = X^+ - X^-$ và $Y = Y^+ - Y^-$. Lại do (iii), ta có $X^{\pm}$ độc lập với $Y^{\pm}$, do đó

$$\begin{aligned}
\mathbb{E}(XY) &= \mathbb{E}(X^+Y^+) + \mathbb{E}(X^-Y^-) - \mathbb{E}(X^+Y^-) - \mathbb{E}(X^-Y^+) \\
&= \mathbb{E}(X^+)\mathbb{E}(Y^+) + \mathbb{E}(X^-)\mathbb{E}(Y^-) - \mathbb{E}(X^+)\mathbb{E}(Y^-) - \mathbb{E}(X^-)\mathbb{E}(Y^+) = \mathbb{E}(X)\mathbb{E}(Y).
\end{aligned}$$

$(iv) \Rightarrow (i)$: Chọn $f = \mathbb{I}_{(-\infty,x)}$ và $g = \mathbb{I}_{(-\infty,y)}$.

□

Từ định lý trên, ta có hai kết quả đáng chú ý sau.

Định lý 2.5.7. *Cho các biến ngẫu nhiên độc lập $X_1, \ldots, X_n$ và các hàm Borel*

$$f_i : \mathbb{R} \to \mathbb{R},\ 1 \le i \le n.$$

Khi đó các biến ngẫu nhiên $f1(X1), \ldots, f_n(X_n)$ độc lập.

Ý nghĩa của định lý trên là tính độc lập vẫn được giữ nguyên qua phép biến đổi hàm. Định lý trên vẫn đúng nếu các hàm f_i là hàm nhiều biến và tác động lên các nhóm rời nhau.

Định lý 2.5.8. *Xét các biến ngẫu nhiên độc lập và khả tích $X_1, \ldots, X_n$. Khi đó*

$$\mathbb{E}(X_1X_2\ldots X_n) = \mathsf{E}X_1bEX2\ldots\mathbb{E}X_n,$$

hay ta còn nói kì vọng của tích bằng tích các kì vọng.

2.5.1 Không gian tích

Trong mục này, ta sẽ trình bày mối liên hệ giữa khái niệm độc lập của các biến ngẫu nhiên với khái niệm không gian tích và độ đo tích trong Lý thuyết độ đo. Từ đó ta có thể xây dựng các ví dụ không tầm thường về các biến ngẫu nhiên độc lập. Trước tiên, ta nhắc lại định lý sau về độ đo tích.

Định lý 2.5.9. *Cho hai không gian đo $(\Omega_1, \mathcal{F}_1, \mu_1)$ và $(\Omega_2, \mathcal{F}_2, \mu_2)$. Khi đó trên không gian tích $\Omega_1 \times \Omega_2$ cùng với σ- đại số*

$$\mathcal{F}_1 \otimes \mathcal{F}_2 = \sigma(\{A_1 \times A_2 : \ A1 \in \mathcal{F}_1, A_2 \in \mathcal{F}_2\}),$$

sẽ tồn tại một độ đo μ sao cho $\mu(A_1 \times A_2) = \mu_1 \times \mu_2(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2)$ với mọi $A1 \in \mathcal{F}_1, A_2 \in \mathcal{F}_2$. Độ đo này được gọi là độ đo tích.

Bằng cách quy nạp ta có thể định nghĩa độ đo tích trên không gian tích củan không gian đo cho trước.

Bây giờ dựa trên khái niệm không gian tích ta có thể xây dựng các biến ngẫu nhiên độc lập như sau.

Ví dụ 2.5.10. *Cho hai không gian xác suất $(\Omega_1, \mathcal{F}_1, \mathbb{P}_1)$ và $(\Omega_2, \mathcal{F}_2, \mathbb{P}_2)$. Ta có thể thấy không gian tích*

$$(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2, \mu_1 \times \mu_2)$$

cũng là không gian xác suất vì $\mu_1 \times \mu_2(\Omega_1 \times \Omega_2) = 1$.

Xét hai biến ngẫu nhiên X_1 và X_2 lần lượt xác định trên hai không gian xác suất đã cho ban đầu. Khi đó, ta có thể xác định hai biến ngẫu nhiên Y_1 và Y_2 trên không gian xác suất tích như sau

$$\begin{aligned} Y_i : \ &\Omega_1 \times \Omega_2 \to \mathbb{R} \\ &(\omega_1, \omega_2) \mapsto Y_i(\omega_1, \omega_2) = X_i(\omega_i). \end{aligned}$$

Ta có

$$\begin{aligned} \mathbb{P}(Y_1 < x_1, Y_2 < x_2) =& \mathbb{P}((\omega_1, \omega_2) \in \Omega_1 \times \Omega_2 : \ X_1(\omega_1) < x_1, X_2(\omega_1) < x_2) \\ =& \mathbb{P}_1 \times \mathbb{P}_2(\{\omega_1 \in \Omega_1 : \ X_1(\omega_1) < x_1\} \times \{\omega_2 \in \Omega_2 : \ X_2(\omega_2) < x_2\}) \\ =& \mathbb{P}_1(\{\omega_1 \in \Omega_1 : \ X_1(\omega_1) < x_1\})\mathbb{P}_2(\{\omega_2 \in \Omega_2 : \ X_2(\omega_2) < x_2\}) \\ =& \mathbb{P}(Y_1 < x_1)\mathbb{P}(Y_2 < x_2). \end{aligned}$$

Suy ra hai biến ngẫu nhiên Y_1 và Y_2 cùng xác định trên không gian xác suất $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2, \mu_1 \times \mu_2)$ là hai biến ngẫu nhiên độc lập.

2.5.2 Tổng hai biến ngẫu nhiên độc lập

Trong mục này chúng ta tìm hiểu về phân phối của tổng của hai biến ngẫu nhiên độc lập. Đầu tiên ta có sự liên hệ với khái niệm tích chập (convolution) của độ đo.

Định nghĩa 2.5.11. *Cho hai phân phối xác suất μ_X và μ_Y trên $\mathbb{R}$ của hai biến ngẫu nhiên độc lập X và Y. Khi đó ta xác định tích chập của hai độ đo, kí hiệu là $\mu_X * \mu_Y$, là phân phối xác suất trên $\mathbb{R}$ của biến ngẫu nhiên tổng $X+Y$ như sau*

$$\mu_X * \mu_Y(A) = \mu_X \times \mu_Y(\{(x,y) : x+y \in A\}), \ \forall A \in \mathcal{B}(\mathbb{R}).$$

Định lý 2.5.12. *1. Nếu hai biến ngẫu nhiên độc lập X và Y chỉ nhận giá trị nguyên thì*

$$\mathbb{P}(X+Y=n) = \sum_{k\in\mathbb{Z}} \mathbb{P}(X=k)\mathbb{P}(Y=n-k).$$

2. Nếu hai biến ngẫu nhiên độc lập X và Y đều có hàm mật độ thì biến ngẫu nhiên tổng $X+Y$ cũng có hàm mật độ được xác đinh như sau

$$f_{X+Y}(u) = f_X * f_Y(u) = \int_{\mathbb{R}} f_X(t) f_Y(u-t)dt.$$

Đặc biệt ta có các ví dụ sau.

Ví dụ 2.5.13. *1. Cho hai biến ngẫu nhiên độc lập có phân phối nhị thức $X \sim B(n,p)$ và $Y \sim B(m,p)$. Chú ý là tham số p ở hai phân phối là bằng nhau. Khi đó, từ ý nghĩa của phân phối nhị thức đếm số lần thành công trong các lần thực hiện phép thử, ta có biến ngẫu nhiên tổng cũng có phân phối nhị thức $X+Y \sim B(n+m,p)$.*

2. Giả sử X và Y là hai biến ngẫu nhiên độc lập có phân phối Poisson với tham số lần lượt là λ và μ $(\lambda, \mu > 0)$. Khi đó $X+Y$ có phân phối Poisson với tham số $\lambda+\mu$.

Thật vậy, với mọi số tự nhiên n,

$$\begin{aligned}\mathbb{P}(X+Y=n) &= \sum_{k=0}^{n} \mathbb{P}(X=k, Y=n-k) = \sum_{k=0}^{n} \mathbb{P}(X=k)\mathbb{P}(Y=n-k) \\ &= \sum_{k=0}^{n} \frac{e^{-\lambda}\lambda^k}{k!}\frac{e^{-mu}\mu^{n-k}}{(n-k)!} = \frac{e^{-\lambda-\mu}}{n!}\sum_{k=0}^{n} C_n^k \lambda^k \mu^{n-k} = \frac{e^{-\lambda-\mu}(\lambda+\mu)^n}{n!}.\end{aligned}$$

3. Cho X và Y là hai biến ngẫu nhiên độc lập với cùng phân phối chuẩn tắc $N(0,1)$. Khi đó $X+Y$ có phân phối chuẩn $N(0,2)$.

Thật vậy,

$$f_{X+Y}(u) = \int_{\mathbb{R}} f_X(t) f_Y(u-t)dt = \int_{\mathbb{R}} e^{-t^2/2}.e^{-(u-t)^2/2}/(2\pi)dt = \frac{e^{-t^2/4}}{\sqrt{4\pi}}.$$

4. *Cho X và Y là hai biến ngẫu nhiên độc lập với phân phối Gamma lần lượt là $\mathcal{G}(\alpha, p)$ và $\mathcal{G}(\lambda, p)$. Khi đó $X+Y$ cũng có phân phối Gamma $\mathcal{G}(\alpha+\lambda, p)$.*

2.6 Hệ số tương quan

Định nghĩa 2.6.1. 1. Hiệp phương sai *của hai biến ngẫu nhiên $X, Y \in L^2$ là*

$$cov(X,Y) = \mathbb{E}[(X-\mathbb{E}X)(Y-\mathbb{E}Y)].$$

2. Hệ số tương quan *của hai biến ngẫu nhiên $X, Y \in L^2$ là*

$$\rho(X,Y) = \frac{cov(X,Y)}{\sqrt{DXDY}}.$$

3. Ma trận tương quan *của vector ngẫu nhiên $X = (X_1, \ldots, X_n)$với các thành phần là các biến ngẫu nhiên bình phương khả tích là*

$$\Sigma_X = (\Sigma_{ij})_{i,j=1,\ldots,n} = (cov(X_i, X_j))_{i,j=1,\ldots,n}.$$

Mệnh đề 2.6.2. *Xét các biến ngẫu nhiên bình phương khả tích X, Y, Z. Ta có*

1. $Cov(X,X) = DX$.

2. $Cov(aX, bY) = ab.Cov(X,Y)$ *với mọi* $a, b \in \mathbb{R}$.

3. $Cov(X, Y+Z) = Cov(X,Y) + Cov(X,Z)$.

4. *Nếu X và Y độc lập thì $Cov(X,Y) = 0$. Ngược lại thì không đúng.*

Ta thấy rằng hiệp phương sai, hay hệ số tương quan là đại lượng được đưa ra để đo mức độ phụ thuộc lẫn nhau của hai biến ngẫu nhiên.

Định nghĩa 2.6.3. 1. *Nếu $cov(X,Y) = 0$, ta nói hai biến ngẫu nhiên X và Y là* không tương quan.

2. *Nếu $cov(X,Y) > 0$, ta nói hai biến ngẫu nhiên X và Y là* tương quan dương. *Khi đó hai biến ngẫu nhiên X và Y có xu hướng thay đổi giá trị cùng chiều nhau so với giá trị trung bình.*

3. *Nếu $cov(X,Y) < 0$, ta nói hai biến ngẫu nhiên X và Y là* tương quan âm. *Khi đó hai biến ngẫu nhiên X và Y có xu hướng thay đổi giá trị ngược chiều nhau so với giá trị trung bình.*

Từ bất đẳng thức Cauchy-Schwarz, ta có hệ quả sau

Mệnh đề 2.6.4. *Cho hai biến ngẫu nhiên $X, Y \in \mathcal{L}^2$.*

1. *Ta có $|\rho(X, Y)| < 1$.*
2. *Nếu $\rho(X, Y) = 1$, thì tồn tại hai hằng số $c > 0$ và a sao cho $Y = cX + a$ h.c.c.*
3. *Nếu $\rho(X, Y) = -1$, thì tồn tại hai hằng số $c > 0$ và a sao cho $Y = cX + a$ h.c.c.*

Mệnh đề 2.6.5. *Ta có*

$$Var\,(X_1 + \ldots + X_n) = \sum_{i,j=1}^{n} Cov(X_i, X_j).$$

Tổng quát, với mọi bộ số thực $a_1, \ldots, a_n \in \mathbb{R}$,

$$Var\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i,j=1}^{n} a_i a_j . Cov(X_i, X_j) = a^T \Sigma a \geq 0,$$

với Σ là ma trận tương quan của vector ngẫu nhiên $X = (X_1, \ldots, X_n)$.

Hệ quả 2.6.6. *Ma trận tương quan của vector ngẫu nhiên $X = (X_1, \ldots, X_n)$ là ma trận đối xứng, nửa xác định dương.*

Mệnh đề 2.6.7. *Giả sử $(X_n)_{n\geq 1}$ là một dãy các biến ngẫu nhiên đôi một không tương quan. Khi đó*

$$D(X_1 + \ldots + X_n) = D(X_1) + \ldots + D(X_n).$$

Mệnh đề 2.6.8. *Cho vector ngẫu nhiên có phân phối chuẩn $X \sim \mathcal{N}(a, \Sigma)$; Khi đó, $a = (a_1, \ldots, a_n)$ là vector kì vọng của các biến ngẫu nhiên thành phần và Σ chính là ma trận tương quan của vector.*

2.7 Tính tuyến tính của kì vọng

Về bản chất thì kì vọng là giá trị của tích phân Lebesgue trên không gian Ω nên kì vọng có tính tuyến tính. Đây là tính chất quan trọng, cho phép ta tính được giá trị cụ thể của kì vọng của một biến ngẫu nhiên mà không cần nghiên cứu kĩ về phân phối của biến ngẫu nhiên này, đặc biệt khi phân phối này rất phức tạp. Ta xét một số ví dụ như sau.

Ví dụ 2.7.1. *Cho các biến ngẫu dương $X_1, \ldots, X_n$ độc lập cùng phân phối và luôn nhận giá trị dương. Khi đó*

$$\mathbb{E}\frac{X_1}{X_1 + \ldots + X_n} = \frac{1}{n}.$$

Chứng minh. Ta thấy rằng, các biến ngẫu nhiên

$$\frac{X_1}{X_1 + \ldots + X_n}, \frac{X_2}{X_1 + \ldots + X_n}{}', \ldots, \frac{X_n}{X_1 + \ldots + X_n},$$

là các biến ngẫu nhiên có cùng phân phối. Ngoài ra dễ thấy là các biến ngẫu nhiên này bị chặn (nhận giá trị dương, và luôn nhỏ hơn 1) nên khả tích. Vậy ta có

$$\mathbb{E}\frac{X_1}{X_1 + \ldots + X_n} = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\frac{X_i}{X_1 + \ldots + X_n} = \frac{1}{n}\mathbb{E}\frac{X_1 + \ldots + X_n}{X_1 + \ldots + X_n} = \frac{1}{n}.$$

□

Điều thú vị trong ví dụ trên là ta có công thức chung cho nhiều dạng phân phối của các biến ngẫu nhiên X_i. Các tính chất đúng cho một lớp tổng quát các dạng phân phối như vậy được gọi là *tính phổ dụng*. Ngoài ra, ta có thể chú ý thêm rằng nói chung việc tính được chính xác hàm phân phối của biến ngẫu nhiên $X_1/(X_1 + \ldots + X_n)$ là rất phức tạp.

Ví dụ 2.7.2. *Gọi $p_n(k)$ là số các hoán vị từ 1 đến n có đúng k điểm cố định. Hãy tính*

a. $\sum_{k=1}^{n} kp_n(k)$.

b. $\sum_{k=1}^{n} k^2 p_n(k)$.

Chứng minh. Chọn ra ngẫu nhiên các hoán vị theo phân phối đều. Gọi biến ngẫu nhiên X là số các điểm bất động trong hoán vị lấy ra. Khi đó ta có bảng phân phối của X như sau:

X	0	1	$\ldots$	n
$\mathbb{P}$	$p_n(0)/n!$	$p_n(1)/n!$	$\ldots$	$p_n(n)/n!$

a. Theo định nghĩa của kì vọng, ta có

$$\mathbb{E}X = \sum_{k=0}^{n} kp_n(k)/n! = \sum_{k=1}^{n} kp_n(k)/n!.$$

Biễu diễn X thành tổng các biến ngẫu nhiên chỉ báo cho sự kiện phần tử thứ i có cố định hay không ta có

$$X = \sum_{i=1}^{n} I_{i \text{ cố định}}.$$

Khi đó,

$$\mathbb{E}X = \sum_{i=1}^{n} \mathbb{P}(i \text{ cố định}) = \sum_{k=1}^{n} \frac{(n-1)!}{n!} = 1.$$

Vậy $\sum_{k=1}^{n} kp_n(k) = n!$.

b. Ta thấy rằng

$$\mathbb{E}X^2 = \sum_{k=0}^{n} k^2 p_n(k)/n! = \sum_{k=1}^{n} k^2 p_n(k)/n!.$$

Khai triển X^2 ta có

$$X^2 = \left(\sum_{i=1}^{n} I_{i \text{ cố định}}\right)^2 = \sum_{i=1}^{n} I_{i \text{ cố định}} + \sum_{i \neq j=1}^{n} I_{i \text{ cố định}} I_{j \text{ cố định}}.$$

Suy ra

$$\begin{aligned}\mathbb{E}X^2 &= \sum_{i=1}^{n} \mathbb{P}(i \text{ cố định}) + \sum_{i \neq j=1}^{n} \mathbb{P}(i, j \text{ đều cố định}) \\ &= \sum_{i=1}^{n} \frac{(n-1)!}{n!} + \sum_{i \neq j=1}^{n} \frac{(n-2)!}{n!} = 1 + 1 = 2.\end{aligned}$$

Kết luận $\sum_{k=1}^{n} k^2 p_n(k) = 2.n!$.

□

Ta có nhận xét là có thể tính được giá trị chính xác của các đại lượng $p_n(k)$ từ Nguyên lý bù trừ và thay lại vào công thức tổng để tính toán tiếp. Tuy nhiên việc tính tổng khá là phức tạp. Việc chia xuống cho $n!$ và chuyển về kì vọng của biến ngẫu nhiên X như trên cho ta thấy bản chất của công thức tổng trong đề bài.

Trong ví dụ cuối cùng của mục này, chúng ta trở lại với bài toán Buffon đã được trình bày ở Chương 1. Lời giải sau được tham khảo từ cuốn sách *Proofs from THE BOOK* [1].

Ví dụ 2.7.3. *Thả ngẫu nhiên một que tăm độ dài l lên một tờ giấy (mặt phẳng) được chia thành các dải song song cỡ h bởi các đường kẻ song song. Gọi X_l là biến ngẫu nhiên chỉ số giao điểm của que tăm với các đường kẻ. Khi đó,*

$$\mathbb{E}X_l = \frac{2l}{\pi h}.$$

Chứng minh. Ý tưởng chứng minh như sau. Ta thấy rằng ta có thể bẻ đoạn thẳng độ dài l ban đầu thành hai đoạn thẳng có độ dài l_1 và l_2. Với xác suất bằng 1 (trừ đi trường hợp vị trí bẻ nằm trên các đường kẻ) thì số giao điểm của que ban đầu với các đường kẻ bằng với số giao điểm trên hai đoạn con. Do đó, theo tính tuyến tính của kì vọng

$$\mathbb{E}X_l = \mathbb{E}X_{l_1} + \mathbb{E}X_{l_2}, \tag{2.9}$$

và tổng quát

$$\mathbb{E}X_l = \mathbb{E}X_{l_1} + \mathbb{E}X_{l_2} + \ldots + \mathbb{E}X_{l_k},$$

nếu $l_1 + \ldots + l_k = l$. Đặt $\varphi(l) = \mathbb{E}X_l$. Từ (2.9), ta suy ra φ là hàm cộng tính theo nghĩa

$$\varphi(x+y) = \varphi(x) + \varphi(y).$$

Hơn nữa, do $\varphi(x)$ không âm, nên φ là hàm không giảm. Từ lý thuyết phương trình hàm, ta suy ra φ là hàm tuyến tính, tức là tồn tại hằng số $c \in \mathbb{R}$ sao cho

$$\varphi(l) = c.l, \ \forall l \geq 0.$$

Để tìm hằng số c, ta có chú ý như sau. Từ công thức (2.9), ta cũng thấy rằng vị trí tương quan của hai đoạn con sẽ không ảnh hưởng đến kết quả tính kì vọng. Tức là thay vì đoạn thẳng độ dài l ban đầu, ta có thể xét đường gấp khúc cũng có độ dài l, và lấy giới hạn thì ta có thể xét đường cong trơn có độ dài l.

Xét đường cong trơn đặc biệt là hình tròn có đường kính h. Đường cong này có độ dài πh. Điều đặc biệt là ở mọi vị trí tương đối thì hình tròn này luôn có đúng hai giao điểm với các đường kẻ song song, tức là biến ngẫu nhiên số giao điểm trong tình huống này nhận giá trị hằng số bằng 2. Suy ra

$$2 = \mathbb{E}X_{\pi h} = c(\pi h).$$

Vậy ta có $c = 2/(\pi h)$ và dẫn đến ĐPCM. □

Đặc biệt, khi $l < h$ thì trừ trường hợp có xác suất bằng 0 là que tăm thuộc đường kẻ song song, còn lại chỉ có khai khả năng que tăm có cắt các đường kẻ hay không với số giao điểm tương ứng là 0 và 1. Khi đó, kì vọng $\mathbb{E}X_l$ trở thành xác suất để que tăm cắt vào các đường kẻ song song và bằng $\dfrac{2l}{h\pi}$. Ta thu lại được kết quả đã trình bày ở Chương 1.

2.8 Nguyên lý thứ hai của Phương pháp xác suất

Như đã trình bày ở phần trước, ý nghĩa của kì vọng của biến ngẫu nhiên X chính là *giá trị trung bình* của hàm số X trên cả không gian Ω. Từ đó ta có nguyên lý thứ hai của Phương pháp xác suất như sau.

Nguyên lý 2:

1. *Nếu* $\mathbb{E}X \geq a$, *thì tồn tại tình huống* $\omega \in \Omega$ *sao cho* $X(\omega) \geq a$.
2. *Nếu* $\mathbb{E}X \leq a$, *thì tồn tại tình huống* $\omega \in \Omega$ *sao cho* $X(\omega) \leq a$.
3. *Nếu* $\mathbb{E}X = a$, *thì tồn tại tình huống* $\omega \in \Omega$ *sao cho* $X(\omega)$ *không lớn hơn (hoặc không nhỏ hơn)* a.

Ta có một số ví dụ minh họa cho Nguyên lý thứ hai này như sau. Ví dụ đầu tiên cũng được đưa ra bởi Erdos.

Ví dụ 2.8.1. *Chứng minh rằng từ mỗi tập gồm* n *số nguyên dương phân biệt có thể lấy ra một tập con ít nhất* $n/3$ *phần tử và có tính tổng tự do: không tồn tại 3 phần tử* a, b, c *sao cho* $a + b = c$.

Chứng minh. Xét một số nguyên tố $p = 3k + 2$ đủ lớn (lớn hơn ba lần các phần tử trong tập đang xét). Xét tập $C = \{k + 1, k + 2, \ldots, 2k + 1\}$. Dễ thấy tập C có tính tổng tự do khi lấy tổng theo modulo p.

Xét tập đang xét là $A = \{a_1, a_2, \ldots, a_n\}$. Lấy ngẫu nhiên phần tử x trên tập $\{1, 2, \ldots, p-1\}$ với phân phối đều. Xác định biến ngẫu nhiên X là số phần tử thuộc A sao cho khi nhân với x thì thuộc vào C hay là

$$X = \text{card}\{xA \cap C\} = \text{card}\{a \in A : \ a.x \in C\}.$$

Ta sẽ tính kì vọng của X. Ta có thể biểu diễn

$$X = \sum_{i=1}^{n} I_{x.a_i \in C}.$$

Do đó,

$$\mathbb{E}X = \sum_{i=1}^{n} \mathbb{E}I_{x.a_i \in C} = \sum_{i=1}^{n} \mathbb{P}(x.a_i \in C).$$

Chú ý là p là số nguyên tố nên với mọi số a_i thì tập $\{a_i.x : \ x = 1, 2 \ldots, p-1\}$ tạo thành hệ thặng dư thu gọn đầy đủ. Do đó

$$\mathbb{P}(x.a_i \in C) = \frac{|C|}{p-1} > \frac{1}{3}.$$

Suy ra, $\mathbb{E}X > n/3$. Vậy tồn tại một số x nào đó sao cho có ít nhất $n/3$ phần tử thuộc A mà nhân với x sẽ thuộc vào C. Đây là $n/3$ phần tử tạo nên tập con cần tìm. Thật vậy, nếu tập con này không có tính tổng tự do, tức là tồn tại 3 phần tử a, b, c sao cho $a+b=c$, thì $ax+bx=cx(modp)$, điều này mâu thuẫn với việc 3 phần tử ax, bx và cx đều thuộc C và C có tính tổng tự do. □

Ví dụ 2.8.2 (Định lý Turan)**.** *Xét đồ thị $G=(V,E)$ sao cho với mỗi đỉnh $v \in V$ thì có bậc là d_v. Chứng minh rằng ta có thể chọn ra một tập con có ít nhất $\sum_{v\in V} \frac{1}{d_v+1}$ đỉnh sao cho không có hai đỉnh nào được nối với nhau.*

Chứng minh. Xét một thứ tự ngẫu nhiên các đỉnh trong đồ thị. Xét biến ngẫu nhiên X chỉ số đỉnh đứng trước tất cả đỉnh kề với nó trong thứ tự sắp xếp. Ta có

$$X = \sum_{v\in V} I_{v \text{ đứng trước các đỉnh kề}}.$$

Do đó,

$$\mathbb{E}X = \sum_{v\in V} \mathbb{P}(v \text{ đứng trước các đỉnh kề}).$$

Xét một đỉnh v cố định. Ta thấy rằng để v đứng trước các đỉnh kề với nó thì ta chọn ra d_v+1 vị trí, sau đó cho v vào vị trí đầu tiên và các đỉnh kề còn lại chọn vị trí tùy ý trong d_v vị trí còn lại, còn các đỉnh không kề với v chọn vị trí tùy ý ngoài d_v+1 vị trí đã chọn. Khi đó,

$$\mathbb{P}(v \text{ đứng trước các đỉnh kề}) = \frac{C_{|V|}^{d_v+1}(d_v)!(|V|-d_v-1)!}{(|V|)!} = \frac{1}{d_v+1}.$$

Suy ra

$$\mathbb{E}X = \sum_{v\in V} \frac{1}{d_v+1}.$$

Tức là tồn tại một thứ tự mà có ít nhất $\sum_{v\in V} \frac{1}{d_v+1}$ đỉnh đứng trước các đỉnh kề với nó. Các đỉnh này tạo thành tập con thỏa mãn yêu cầu. □

Ví dụ 2.8.3. *Cho n vector $V_1, V_2, \ldots, v_n \in \mathbb{R}^n$ sao cho $|v_i| \le 1, \forall i = 1, 2 \ldots, n$. Lấy các số $p_1, p_2, \ldots, p_n \in [0,1]$ tùy ý và đặt*

$$w = p_1v_1 + p_2v_2 + \ldots + p_nv_n.$$

Chứng minh rằng tồn tại $\epsilon_1, \epsilon_2, \ldots, \epsilon_n \in \{0,1\}$ để với $v = \epsilon_1v_1 + \ldots \epsilon_nv_n$ thì

$$|w-v| \le \frac{\sqrt{n}}{2}.$$

Chứng minh. Chọn các ϵ_i một cách ngẫu nhiên và độc lập với nhau sao cho

$$\mathbb{P}(\epsilon_i = 1) = p_i,\ \mathbb{P}(\epsilon_i = 0) = 1 - p_i\ \forall i = 1, 2 \ldots, n.$$

Khi đó ta có biến ngẫu nhiên bình phương độ dài

$$X = |w - v|^2.$$

Ta có

$$X = \left|\sum_{i=1}^{n}(p_i - \epsilon_i)v_i\right|^2 = \sum_{i=1}^{n}\sum_{j=1}^{n}(p_i - \epsilon_i)(p_j - \epsilon_j)\langle v_i, v_j\rangle,$$

suy ra

$$\mathsf{E}X = \sum_{i=1}^{n}\sum_{j=1}^{n}\langle v_i, v_j\rangle \mathsf{E}[(p_i - \epsilon_i)(p_j - \epsilon_j)].$$

Chú ý là khi $i \neq j$ thì

$$\mathsf{E}[(p_i - \epsilon_i)(p_j - \epsilon_j)] = \mathsf{E}(p_i - \epsilon_i).\mathsf{E}(p_j - \epsilon_j) = 0,$$

còn khi $i = j$ thì

$$\mathsf{E}(p_i - \epsilon_i)^2 = p_i(p_i - 1)^2 + (1 - p_i)(p_i - 0)^2 = p_i(1 - p_i) \leq \frac{1}{4}.$$

Do đó $\mathsf{E}X \leq n/4$. Suy ra tồn tại một cách chọn các số ϵ_i sao cho bình phương của độ dài của $|w - v|$ không lớn hơn $n/4$. ĐPCM. □

Ví dụ 2.8.4. *Trong hình vuông đơn vị có một đường gấp khúc có độ dài 100. Chứng minh rằng tồn tại một đường thẳng ngang hoặc dọc (song song với cạnh của hình vuông) sao cho cắt đường gấp khúc tại ít nhất 50 điểm.*

Chứng minh. Đường gấp khúc tạo bởi n đoạn thẳng có độ dài lần lượt là $l_1, l_2, \ldots, l_n$. Theo giả thiết ta có $l_1 + l_2 + \ldots + l_n = 100$.

Lấy ngẫu nhiên (theo phân phối đều) một điểm trên cạnh đáy của hình vuông đơn vị. Gọi X là biến ngẫu nhiên chỉ số lần đường thẳng đứng qua điểm được lấy cắt đường gấp khúc đang xét. Tương tự, lấy ngẫu nhiên (theo phân phối đều) một điểm trên cạnh bên của hình vuông đơn vị và gọi Y là biến ngẫu nhiên chỉ số lần đường ngang qua điểm được lấy cắt đường gấp khúc đang xét. Ngoài ra, với mỗi đoạn thẳng thứ i, ta xác định hai biến ngẫu nhiên tương ứng là X_i và Y_i.

Chú ý rằng các đoạn thẳng được chọn có thể bao gồm hoặc không các điểm đầu mút. Khi đó, ta có:

$$X = X_1 + X_2 + \ldots + X_n, \quad Y = Y_1 + Y_2 + \ldots + Y_n.$$

Bây giờ, ta xét riêng từng cặp X_i và Y_i. Theo định nghĩa, hai biến ngẫu nhiên này chỉ nhận giá trị 0 hoặc 1 vì một đường thẳng chỉ cắt một đoạn thẳng tại nhiều nhất là một điểm. Giả dụ với X_i, biến ngẫu nhiên này nhận giá trị 1 khi và chỉ khi điểm lấy ra thuộc vào hình chiếu của đoạn thẳng thứ i trên cạnh đáy của hình vuông. Do đó, kì vọng $\mathbb{E}X_i$ sẽ bằng xác suất chọn phải điểm thuộc vào hình chiếu trên cạnh đáy, hay chính là độ dài của hình chiếu trên cạnh đáy (do hình vuông có cạnh đơn vị). Tương tự, kì vọng $\mathbb{E}Y_i$ sẽ bằng độ dài của hình chiếu của đoạn thẳng thứ i trên cạnh bên. Dễ thấy

$$\mathbb{E}X_i + \mathbb{E}Y_i \geq l_i.$$

Suy ra

$$\mathbb{E}X + \mathbb{E}Y = \mathbb{E}X_1 + \ldots + \mathbb{E}X_n + \mathbb{E}Y_1 + \ldots + \mathbb{E}Y_n \geq l_1 + \ldots + l_n = 100.$$

Do đó, trong hai kì vọng $\mathbb{E}X$ và $\mathbb{E}Y$ có một số không nhỏ hơn 50. Theo nguyên lý thứ hai, suy ra đpcm. □

2.9 Bài toán phân bổ danh mục đầu tư

Trong mục này chúng ta cùng tìm hiểu về Lý thuyết phân bổ danh mục đầu tư do Markowitz đề xuất từ những năm 1950. Đây là công trình giúp Markowitz giành giải Nobel kinh tế năm 1990 cùng Miller và Sharpe.

Trong mô hình này, ta xét trên thị trường có n loại tài sản để đầu tư, tạm gọi là cổ phiếu. Xét biến ngẫu nhiên lãi suất R_i của cổ phiếu thứ i. Biến ngẫu nhiên R_i có hai đại lượng là

$$\mathbb{E}R_i = r_i,\ DR_i = \sigma_i^2.$$

Ngoài ra giữa các biến ngẫu nhiên R_i, ta có thêm thông tin về hệ số tương quan ρ_{ij}.

Câu hỏi ta thường gặp trong tài chính là: với lượng tiền vốn ban đầu, ta nên phân chia tiền vào các cổ phiếu như thế nào để thỏa mãn yêu cầu có thể của mình?

Yêu cầu ở đây là gì? Một cách tự nhiên, ta sẽ nghĩ đến đó là **ít rủi ro**. Điều này dẫn đến một câu hỏi quan trọng trong Lý thuyết Toán tài chính đó là làm thế nào để mô hình hóa rủi ro bằng các công cụ Toán học. Có nhiều cách để mô hình rủi ro, nhưng trong mô hình Markowitz, ta sẽ giới thiệu một cách rất đơn giản và thú vị.

Định nghĩa 2.9.1. *Mức rủi ro của cổ phiếu thứ i được định nghĩa bằng độ lệch chuẩn (hoặc phương sai) của lãi suất R_i của cổ phiếu đó. Một cách tổng quát, với một cách phân chia danh mục đầu tư, rủi ro của của danh mục là độ lệch chuẩn (hoặc phương sai) của lãi suất của danh mục.*

Cụ thể hơn ta có định nghĩa như sau.

Định nghĩa 2.9.2. *Một danh mục đầu tư là cách phân bổ tài sản vào các cổ phiếu R_i, tương ứng với mức trọng tài sản w_i, thỏa mãn*

$$w_1 + w_2 + \ldots + w_n = 1.$$

Trong mô hình tự nhiên, ta yêu cầu $w_i \in [0, 1]$, tức là ta đầu tư hoàn toàn bằng vốn sẵn có.

Ngoài ra có thể xét đến mô hình bán khống*, cho phép trọng $w_i < 0$.*

Bán khống là hình thức có thể tồn tại trên thị trường, dựa theo quy định của Quỹ. Hình thức của bán khống như sau: ta vay tạm một số cổ phiếu R_i của Quỹ để bán đi và thu về một lượng tiền; lượng tiền này cùng với tiền vốn ban đầu (hoặc các tài sản có trọng $w_i > 0$) được dùng làm việc bảo đảm cho việc vay nợ; trong tình huống xấu nhất khi giá trị bảo đảm tương ứng với giá trị để mua lại lượng cổ phiếu R_i đã bán đi để trả cho Quỹ thì Quỹ sẽ lập tức thực hiện giao dịch và đóng tài khoản.

Theo tính tuyến tính của kì vọng ta có

Mệnh đề 2.9.3. *Lãi suất của danh mục đầu tư là*

$$R_V = w_1 R_1 + \ldots + w_n R_n.$$

Khi đó, kì vọng lợi nhuận và rủi ro của danh mục là

$$\mathbb{E}R_V = w_1 r_1 + \ldots + w_n r_n,$$

$$DR_V = \sigma_V^2 = \sum_{i,j=1}^{n} w_i w_j . \sigma_i \sigma_j \rho_{ij}.$$

Tiếp theo ta sẽ xét đến hai câu hỏi sau:

bullet Tìm danh mục đầu tư ít rủi ro nhất.

bullet Trong các danh mục có cùng mức lợi nhuận kì vọng, tìm danh mục có ít rủi ro nhất.

Sử dụng phương pháp nhân tử Lagrange trong Lý thuyết tối ưu, ta có các kết quả sau.

Mệnh đề 2.9.4. *Danh mục đầu tư có phương sai nhỏ nhất (ít rủi ro nhất) tương ứng với tỉ trọng*

$$w = \frac{u\Sigma^{-1}}{u\Sigma^{-1}u^T},$$

với $u = (1, \ldots, 1)$ và Σ là ma trận tương quan của vector ngẫu nhiên $(R_1 \ldots, R_n)$.

Mệnh đề 2.9.5. *Trong số các danh mục đầu tư có kì vọng lợi nhuận μ đã xác định trước, danh mục đầu tư có phương sai nhỏ nhất (ít rủi ro nhất) tương ứng với tỉ trọng*

$$w = \frac{\begin{vmatrix} 1 & u\Sigma^{-1}m^T \\ \mu & m\Sigma^{-1}m^T \end{vmatrix} u\Sigma^{-1} + \begin{vmatrix} u\Sigma^{-1}u^T & 1 \\ m\Sigma^{-1}u^T & \mu \end{vmatrix} m\Sigma^{-1}}{\begin{vmatrix} u\Sigma^{-1}u^T & u\Sigma^{-1}m^T \\ m\Sigma^{-1}u^T & m\Sigma^{-1}m^T \end{vmatrix}},$$

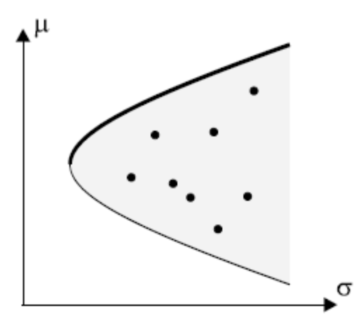
với u, Σ như phát biểu ở Mệnh đề trên, và $m = (r_1, \ldots, r_n)$ là vector các kì vọng lợi nhuận của các cổ phiếu.

Trên quan điểm khách quan của nhà đầu tư, giữa hai danh mục theo các tỉ trọng w và w', nếu như

$$\mu_w \geq \mu_{w'}; \quad \sigma_w \leq \sigma_{w'},$$

thì ta nên chọn danh mục theo tỉ trọng w, vì *w hiệu quả hơn w'*. Ta có định nghĩa về *đường biên hiệu quả* (efficient frontier) như sau.

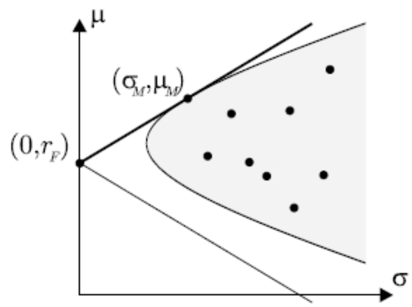
Định nghĩa 2.9.6. *Trên mặt phẳng tọa độ (σ, μ) biểu diễn độ lệch chuẩn và kì vọng lợi nhuận của các danh mục đầu tư có thể, đường biên hiệu quả là tập hợp các điểm tương ứng với các cách đầu tư sao cho không có cách đầu tư nào hiệu quả hơn.*



Ở hình trên, đường tô đậm là đường biên hiệu quả. Ta thấy rằng điểm cực biên trái của đường biên hiệu quả tương ứng với tỉ trọng đầu tư ở Mệnh đề 2.9.4. Và các điểm trên đường biên hiệu quả đều tương ứng với tỉ trọng đầu tư ở Mệnh đề 2.9.5, khi ta xét theo đường kẻ ngang. Chú ý ta không xét các điểm biên ở dưới không tô đậm, vì theo các đường thẳng đứng, ta có các cách đầu tư cùng mức rủi ro nhưng có kì vọng lợi nhuận cao hơn.

Mệnh đề 2.9.7. *Đường biên hiệu quả có tính lồi. Có nghĩa là với ba điểm $(\sigma_1, \mu_1), (\sigma_2, \mu_2)$ và (σ_3, μ_3) trên đường biên hiệu quả sao cho $\mu_2 = p\mu_1 + (1-p)\mu_3$ với $p \in [0, 1]$ thì $\sigma_2 \leq p\sigma_1 + (1-p)\sigma_3$.*

Bây giờ ta xét thêm yếu tố cổ phiếu không rủi ro (có phương sai bằng 0), ví dụ như lãi suất gửi ngân hàng. Điểm này có tọa độ $(0, r_f)$. Xét nửa đường thẳng L xuất phát từ $(0, r_f)$ và tiếp xúc với đường biên hiệu quả, với điểm tiếp xúc tương ứng (σ_M, μ_M) gọi là điểm Markowitz.



Mệnh đề 2.9.8. *Nửa đường thẳng L là đường biên hiệu quả cho các danh mục đầu tư chứa cả cổ phiếu không rủi ro.*

Chứng minh. Xét một cách đầu tư hiệu quả có cả cổ phiếu không rủi ro. Khi bỏ phần không rủi ro này, ta được danh mục gồm các cổ phiếu rủi ro. Chú ý là cổ phiếu không rủi ro chỉ làm thay đổi kì vọng chứ không ảnh hưởng đến phương sai. Do đó phần danh mục con gồm các cổ phiếu rủi ro là danh mục hiệu quả, nằm trên biên hiệu quả. Khi đó danh mục ban đầu có dạng $(1-p, pw)$, với $1-p$ là tỉ trọng cho cổ phiếu không rủi ro và phần p tiền còn lại phân bổ theo tỉ trọng w.

Suy ra là mọi điểm trên nửa đường thẳng L đều tương ứng với danh mục đầu tư, có phần rủi ro tương ứng với điểm Markowitz.

Cuối cùng, để chứng minh L là đường biên hiệu quả, giả sử có điểm nằm bên trái L tương ứng với danh mục nào đó. Bỏ đi phần rủi ro, ta vẫn được một điểm bên trái L, suy ra bên trái biên hiệu quả. Điều này cho thấy có danh mục hiệu quả hơn các danh mục trên biên hiệu quả (mâu thuẫn). □

Chương 3

Sự hội tụ của dãy biến ngẫu nhiên

3.1 Các dạng hội tụ của dãy biến ngẫu nhiên theo nghĩa hàm số

Theo định nghĩa, một biến ngẫu nhiên là một hàm số đo được từ $(\Omega, \mathcal{F})$ vào $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Do đó, với cách nhìn một dãy các biến ngẫu nhiên là dãy hàm số, ta có các dạng hội tụ như sau.

Định nghĩa 3.1.1. *Giả sử $(X_n)_{n\geq 1}$ là dãy biến ngẫu nhiên xác định trên không gian xác suất $(\Omega, \mathcal{F}, \mathbb{P})$. Dãy (X_n) được gọi là*

- hội tụ hầu chắc chắn *đến biến ngẫu nhiên X, kí hiệu là $X_n \xrightarrow{h.c.c} X$ hay $\lim_n X_n = X$ h.c.c hoặc $X_n \xrightarrow{a.s} X$, nếu*

$$\mathbb{P}\Big(w : \lim_{n\to\infty} X_n(w) = X(w)\Big) = 1;$$

- hội tụ theo xác suất *đến biến ngẫu nhiên X, kí hiệu là $X_n \xrightarrow{\mathbb{P}} X$, nếu với mọi $\epsilon > 0$,*

$$\lim_{n\to\infty} \mathbb{P}(|X_n - X| > \epsilon) = 0;$$

- hội tụ theo trung bình bậc p $(p > 0)$ *đến biến ngẫu nhiên X, kí hiệu là $X_n \xrightarrow{L^p} X$ nếu $\mathbb{E}(|X_n|^p) < \infty$ với mọi n, $\mathbb{E}(|X|^p) < \infty$ và*

$$\lim_{n\to\infty} \mathbb{E}(|X_n - X|^p) = 0.$$

Ta có các ví dụ như sau.

Ví dụ 3.1.2. *Xét một dãy các biến ngẫu nhiên*

$$\mathbb{P}(X_n = 0) = 1 - \frac{1}{n^2} \quad \text{and} \mathbb{P}(X_n = n) = \frac{1}{n^2}.$$

Khi đó X_n hội tụ về 0 theo cả ba dạng là xác suất, L^p với $0 < p < 2$ và hầu chắc chắn.

Chứng minh. • Với mọi $\epsilon > 0$, ta có

$$0 \leq \mathbb{P}(|X_n - 0| > \epsilon) \leq \mathbb{P}(X_n = n) = \frac{1}{n^2}.$$

Suy ra

$$\lim_{n\to\infty} \mathbb{P}(|X_n - 0| > \epsilon) = 0.$$

Vậy ta có $X_n \xrightarrow{\mathbb{P}} 0$.

• Với mọi $p \in (0, 2)$, ta có

$$\mathbb{E}(|X_n - 0|^p) = n^{p-2} \to 0.$$

Do đó $X_n \xrightarrow{L^p} 0$. Ngoài ra, dễ thấy khi $p \leq 2$ thì $X_n \overset{L^p}{\not\longrightarrow} 0$.

• Để chứng minh hội tụ hầu chắc chắn, ta có thể sử dụng *Bổ đề Borel-Cantelli* được phát biểu như sau.

Bổ đề 3.1.3 (Borel-Cantelli)**.** *Cho dãy biến cố A_n trong không gian xác suất $\{\Omega, \mathcal{F}, \mathbb{P}\}$. Xét biến cố* $\limsup A_n = \cap_{n=1}^{\infty} (\cup_{m\geq n} A_m)$.

1. *Nếu $\Sigma_{n=1}^{\infty}\mathbb{P}(A_n) < \infty$, thì* $\mathbb{P}(\limsup A_n) = 0$.

2. *Nếu $\Sigma_{n=1}^{\infty}\mathbb{P}(A_n) = \infty$ và họ các biến cốA_n độc lập thì* $\mathbb{P}(\limsup A_n) = 1$.

Chứng minh. 1. Từ định nghĩa của $\limsup A_n$, ta thấy rằng $\forall i \in \mathbb{N}$,

$$\limsup A_n \subset \cup_{m\geq i} A_m.$$

Suy ra $\mathbb{P}(\limsup A_n) \leq \mathbb{P}(\cup_{m\geq i} A_m) \leq \Sigma_{m=i}^{\infty}\mathbb{P}(A_m)$. Vì ta có chuỗi hội tụ $\Sigma_{n=1}^{\infty}\mathbb{P}(A_n) < \infty$, nên tổng $\Sigma_{m=i}^{\infty}\mathbb{P}(A_m)$ có thể bé tùy ý với i đủ lớn. Vậy $\mathbb{P}(\limsup A_n) = 0$.

2. Ta có

$$\begin{aligned}
1 - \mathbb{P}(\limsup A_n) &= \mathbb{P}\left(\overline{\cap_{n=1}^{\infty} \cup_{m\geq n} A_m}\right) \\
&= \mathbb{P}\left(\cup_{n=1}^{\infty} \overline{\cup_{m\geq n} A_m}\right) \\
&= \mathbb{P}\left(\cup_{n=1}^{\infty} \cap_{m\geq n} \overline{A_m}\right)
\end{aligned}$$

Để chứng minh $\mathbb{P}(\limsup A_n) = 1$, tức là $1 - \mathbb{P}(\limsup A_n) = 0$, ta sẽ chứng minh $\mathbb{P}\left(\cap_{m\geq n}\overline{A_m}\right) = 0$ với mọi n. Thật vậy, do các biến cố độc lập nên

$$\begin{aligned}\mathbb{P}\left(\cap_{m\geq n}\overline{A_m}\right) &= \Pi_{m\geq n}\mathbb{P}(\overline{A_m})\\ &= \Pi_{m\geq n}[1-\mathbb{P}(A_m)]\\ &\leq \Pi_{m\geq n}e^{-\mathbb{P}(A_m)} = e^{-\Sigma_{m\geq n}\mathbb{P}(A_m)} = e^{-\infty} = 0.\end{aligned}$$

Ta có ĐPCM.

□

Trở lại ví dụ đang xét, xét các biến cố $A_n = \{X_n \neq 0\} = \{X_n = n\}$. Ta có

$$\Sigma_{n=1}^{\infty}\mathbb{P}(A_n) = \Sigma_{n=1}^{\infty}\frac{1}{n^2} < \infty.$$

Nên theo Bổ đề Borel-Cantelli, $\mathbb{P}(\limsup A_n) = 0$. Điều đó có nghĩa là với xác suất bằng 0 thì các biến cố A_n xuất hiện vô hạn lần. Hay là hầu chắc chắn, ta có A_n xuất hiện hữu hạn lần. Tức là với xác suất bằng 1 thì với mỗi $\omega \in \Omega$, sẽ tồn tại n_ω sao cho với mọi $n > n_\omega$ thì $\omega \notin A_n$, hay là $X_n(\omega) = 0$. Vậy $X_n \xrightarrow{a.s.} X$. □

Áp dụng phần hai của Bổ đề Borel-Cantelli, ta có kết quả sau.

Ví dụ 3.1.4. *Cho dãy các biến ngẫu nhiên X_n độc lập có phân phối*

$$\mathbb{P}(X_n = 0) = 1 - \frac{1}{n} \quad \text{and} \mathbb{P}(X_n = n) = \frac{1}{n}.$$

Khi đó X_n không hội tụ về 0 hầu chắc chắn.

Ta có tiêu chuẩn kiểm tra hội tụ hầu chắc chắn như sau.

Mệnh đề 3.1.5. *Dãy biến ngẫu nhiên X_n hội tụ hầu chắc chắn đến X khi và chỉ khi với mọi $\epsilon > 0$*

$$\mathbb{P}\left(\limsup |X_n - X| > \epsilon\}\right) = 1 - \mathbb{P}\left(\liminf |X_n - X| \leq \epsilon\}\right) = 0.$$

Chứng minh. Ở chiều suy ra, với xác suất bằng 1, ta có $X_n(\omega) \to X$, suy ra tồn tại $n\omega$ sao cho với mọi $n > n_\omega$ thì $|X_n(\omega) - X| \leq \epsilon$, hay là $\omega \in \liminf |X_n - X| \leq \epsilon\}$. Nên ta có $\mathbb{P}\left(\liminf |X_n - X| \leq \epsilon\}\right) = 1$.

Với chiều ngược lại, xét các $\epsilon_k = 1/k$ với $k \in \mathbb{N}$. Theo giả thiết

$$\mathbb{P}\left(\limsup |X_n - X| > 1/k\}\right) = 0,$$

suy ra

$$\mathbb{P}\left(\bigcup_{k\geq 1} \limsup |X_n - X| > 1/k\}\right) = 0.$$

Lấy phần bù ta có

$$\mathbb{P}\left(\bigcap_{k\geq 1} \liminf |X_n - X| \leq 1/k\}\right) = 1.$$

Với ω bất kì thuộc biến cố giao trên, ta có $\omega \in \liminf |X_n - X| \leq 1/k$ với mọi $k \in \mathbb{N}$, tức là luôn có $n_{k,\omega}$ sao cho khi $n > n_{k,\omega}$ thì

$$||X_n(\omega) - X(\omega)| \leq 1/k.$$

Suy ra $X_n \to X$ h.c.c. □

Hệ quả 3.1.6. *Nếu chuỗi*

$$\sum_{n=1}^{\infty} \mathbb{P}(|X_n - X| > \epsilon) < \infty,$$

với mọi $\epsilon > 0$, *thì* $X_n \xrightarrow{h.c.c} X$.

Hệ quả 3.1.7. *Nếu như tồn tại dãy số dương* $\epsilon_n \downarrow 0$ *sao cho*

$$\sum_{n=1}^{\infty} \mathbb{P}(|X_n - X| > \epsilon_n) < \infty,$$

thì $X_n \xrightarrow{h.c.c} X$.

Mệnh đề 3.1.8. *Giả sử* $(X_n)_{n\geq 1}$ *là dãy biến ngẫu nhiên.*

1. *Nếu* $X_n \xrightarrow{L^p} X$ *thì* $X_n \xrightarrow{\mathbb{P}} X$ *với mọi* $p > 0$.

2. *Nếu* $X_n \xrightarrow{a.s} X$ *thì* $X_n \xrightarrow{\mathbb{P}} X$.

Chứng minh. 1) Giả sử $X_n \xrightarrow{L^p} X$. Với mọi $\epsilon > 0$, theo bất đẳng thức Markov, ta có

$$\mathbb{P}(|X_n - X| \geq \epsilon) = \mathbb{P}(|X_n - X|^p \geq \epsilon^p) \leq \frac{\mathbb{E}[|X_n - X|^p}{\epsilon^p} \to 0 \text{ khi } n \to \infty.$$

2) Ta có

$$\mathbb{P}(|X_n - X| > \epsilon) \leq \mathbb{P}(\cup_{m\geq n}\{|X_m - X| > \epsilon\}).$$

Do đó,

$$\limsup_{n\to} \mathbb{P}(|X_n - X| > \epsilon) \le \lim_{n\to} \mathbb{P}(\cup_{m\ge n}\{|X_m - X| > \epsilon\}) = \mathbb{P}(\limsup\{|X_m - X| > \epsilon\}) = 0.$$

Suy ra ĐPCM. □

Trong phần trước, ta đã tìm hiểu sự liên hệ của hai dạng hội tụ: hầu chắc chắn và theo xác suất thông qua việc lấy lim sup của dãy biến cố $\{|X_m - X| > \epsilon\}$. Tiếp theo ta sẽ tìm hiểu về sự liên hệ giữa hai dạng: L^p và xác suất. Ta thấy rằng hội tụ trên L^p thực chất là sự hội tụ theo khoảng cách $\mathbb{E}|X_n - X|^p$ trên không gian $\mathcal{L}^p$. Với cấu trúc tương tự, ta xây dựng khoảng cách sau cho hội tụ theo xác suất.

Định nghĩa 3.1.9. *Với hai biến ngẫu nhiên X và Y bất kì, ta kí hiệu*

$$d_{\mathbb{P}}(X, Y) = \mathbb{E}\Big[\frac{|X - Y|}{|X - Y| + 1}\Big].$$

Mệnh đề 3.1.10. *Ta có các mệnh đề sau tương đương.*

1. *Dãy biến ngẫu nhiên $(X_n)_{n\ge 1}$ hội tụ theo xác suất đến biến ngẫu nhiên X.*
2. *$d_{\mathbb{P}}(X_n, X) \to 0$.*
3. *$\mathbb{E}(|X_n - X| \wedge 1) \to 0$.*

Chứng minh. 1. ⇒ 2. Giả sử $X_n \xrightarrow{\mathbb{P}} X$. Với mọi $\epsilon > 0$ và $w \in \Omega$, do tính đồng biến của hàm $x \mapsto \frac{x}{x+1}$ trên đoạn $[0, \infty)$, ta có

$$\frac{|X_n(w) - X(w)|}{|X_n(w) - X(w)| + 1} \le \frac{\epsilon}{\epsilon + 1}\mathbb{I}_{\{|X_n - X| \le \epsilon\}}(w) + \mathbb{I}_{\{|X_n - X| > \epsilon\}}(w).$$

Lấy kì vọng hai vế ta được

$$d_{\mathbb{P}}(X_n, X) \le \frac{\epsilon}{\epsilon + 1}\mathbb{P}(|X_n - X| \le \epsilon) + \mathbb{P}(|X_n - X| > \epsilon).$$

Do đó

$$\limsup_{n\to\infty} d_{\mathbb{P}}(X_n, X) \le \epsilon + \limsup_{n\to\infty} \mathbb{P}(|X_n - X| > \epsilon) = \epsilon \text{ với mọi } \epsilon > 0.$$

2. ⇒ 1. Với mọi $\epsilon > 0$, theo bất đẳng thức Markov, ta có

$$\begin{aligned}\mathbb{P}(|X_n - X| \ge \epsilon) &= \mathbb{P}\Big(\frac{|X_n - X|}{|X_n - X| + 1} \ge \frac{\epsilon}{\epsilon + 1}\Big) \\ &\le \frac{\epsilon + 1}{\epsilon}\mathbb{E}\Big[\frac{|X_n - X|}{|X_n - X| + 1}\Big] \to 0.\end{aligned}$$

2. ⇔ 3. Ta có thể suy ra từ quan hệ $\frac{x\wedge 1}{2} \le \frac{x}{x+1} \le x \wedge 1$ với $x \ge 0$. □

Ta tiếp tục có mối liên hệ giữa hội tụ hầu chắc chắn và hội tụ theo xác suất, cũng như một tiêu chuẩn để kiểm tra hội tụ theo xác suất như sau.

Mệnh đề 3.1.11. *1. Giả sử $X_n \xrightarrow{\mathbb{P}} X$. Khi đó tồn tại một dãy con $(n_k)_{k\geq 1}$ của dãy số tự nhiên sao cho $X_{n_k} \xrightarrow{h.c.c} X$.*

2. Ngược lại, giả sử từ mọi dãy con $(n_k)_{k\geq 1}$ của dãy số tự nhiên đều có thể lấy ra một dãy con $(m_k)_{k\geq 1}$ sao cho $X_{m_k} \xrightarrow{h.c.c} X$ thì $X_n \xrightarrow{\mathbb{P}} X$.

Chứng minh. 1) Giả sử $X_n \xrightarrow{\mathbb{P}} X$, khi đó $d_{\mathbb{P}}(X_n, X) \to 0$. Chọn dãy n_k tăng sao cho $d_{\mathbb{P}}(X_{n_k}, X) < 2^{-k}$. Áp dụng Định lí hội tụ đơn điệu 2.3.11 ta có

$$\mathbb{E}\Big[\sum_{k=1}^{\infty} \frac{|X_{n_k} - X|}{|X_{n_k} - X| + 1}\Big] = \sum_{k=1}^{\infty} d_{\mathbb{P}}(X_{n_k}, X) < \sum_{k=1}^{\infty} 2^{-k} < \infty.$$

Vậy nên $\sum_{k=1}^{\infty} \frac{|X_{n_k}-X|}{|X_{n_k}-X|+1} < \infty$ h.c.c. Suy ra $\frac{|X_{n_k}-X|}{|X_{n_k}-X|+1} \overset{h.c.c}{\to} 0$. Điều này cũng có nghĩa là $X_{n_k} \xrightarrow{h.c.c} X$.

2) Ngược lại, giả sử mọi dãy con của dãy $(X_n)_{n\geq 1}$ đều chứa một dãy con khác hội tụ hầu chắc chắn đến X. Nếu $X_n \overset{\mathbb{P}}{\not\to} X$ thì tồn tại $\epsilon > 0$ và dãy con $(n_k)_{k\geq 1}$ của dãy số tự nhiên sao cho $d_{\mathbb{P}}(X_{n_k}, X) > \epsilon$. Mặt khác, theo giả thiết tồn tại dãy con $(m_k)_{k\geq 1}$ của dãy $(n_k)_{k\geq 1}$ sao cho $X_{m_k} \overset{h.c.c}{\to} X$. Theo định lí hội tụ bị chặn, $\lim_{k\to\infty} d_{\mathbb{P}}(X_{m_k}, X) = 0$. Điều này mâu thuẫn với $d_{\mathbb{P}}(X_{m_k}, X) > \epsilon$ với mọi k. Do đó $X_n \xrightarrow{\mathbb{P}} X$. □

Kết quả sau chứng tỏ sự hội tụ bảo toàn qua tác động hàm liên tục.

Mệnh đề 3.1.12. *Giả sử $f : \mathbb{R}^2 \to \mathbb{R}$ liên tục. Khi đó*

1. nếu $X_n \xrightarrow{h.c.c} X$ và $Y_n \xrightarrow{h.c.c} Y$ thì $f(X_n, Y_n) \xrightarrow{h.c.c} f(X, Y)$;

2. nếu $X_n \xrightarrow{\mathbb{P}} X$ và $Y_n \xrightarrow{\mathbb{P}} Y$ thì $f(X_n, Y_n) \xrightarrow{\mathbb{P}} f(X, Y)$.

Chứng minh. 1) Kí hiệu $A = \{w \in \Omega : \lim_{n\to\infty} X_n(w) = X(w)\} \cap \{w \in \Omega : \lim_{n\to\infty} Y_n(w) = Y(w)\}$. Ta có $\mathbb{P}(A) = 1$ và với mọi $w \in A$, do f liên tục nên $\lim_{n\to\infty} f(X_n(w), Y_n(w)) = f(X(w), Y(w))$. Vậy $f(X_n) \xrightarrow{h.c.c} f(X)$.

Khẳng định 2) suy ra từ 1) và phần hai của Mệnh đề 3.1.11. □

3.1.1 Tính khả tích đều của dãy biến ngẫu nhiên

Khái niệm khả tích đều là một khái niệm quan trọng trong Lý thuyết độ đo để đảm bảo một dãy hàm hội tụ hầu chắc chắn thì sẽ hội tụ theo trung bình bậc p. Trong Lý thuyết xác suất, ta sẽ có một kết quả mạnh hơn đó là một dãy biến ngẫu nhiên hội tụ theo xác suất và thỏa mãn điều kiện khả tích đều thì sẽ hội tụ theo trung bình bậc p. Ta có định nghĩa về *khả tich đều* như sau.

Định nghĩa 3.1.13. *Một họ các biến ngẫu nhiên $\{X_t,\, t \in T\}$ cùng xác định trên không gian xác suất $(\Omega, \mathcal{F}, \mathbb{P})$ được gọi là khả tích đều nếu thỏa mãn hai điều kiện sau*

1. $\sup\limits_{t\in T}\mathbb{E}|X_t| < \infty$,
2. *với mọi $\epsilon > 0$, tồn tại $\delta_\epsilon > 0$ sao cho với bất cứ biến cố $A \in \mathcal{F}$ thỏa mãn $\mathbb{P}(A) \le \delta_\epsilon$ thì*
$$\mathbb{E}(|X_t|\mathbb{I}_A) < \epsilon,\ \forall t \in T.$$

Ta thấy rằng điều kiện textitkhả tích giống như trong Định lý 2.3.13 cho một biến ngẫu nhiên. Tính *đều* có nghĩa là điều kiện đúng cho mọi biến ngẫu nhiên trong họ.

Mệnh đề 3.1.14. *Họ các biến ngẫu nhiên $\{X_t,\, t \in T\}$ khả tích đều khi và chỉ khi*

$$\lim_{n\to\infty}\sup_{t\in T}\mathbb{E}(|X_t|\mathbb{I}_{|X_t|\ge n}) = 0.$$

Chứng minh. ($\Rightarrow$). Theo giả thiết, tồn tại $C > 0$ sao cho $\mathbb{E}|X_t| < C$ với mọi $t \in T$. Từ bất đẳng thức Markov ta có,

$$\mathbb{P}(|X_t| \ge n) \le \frac{C}{n}.$$

Với mọi $\epsilon > 0$, ta có δ_ϵ như trong giả thiết. Với n đủ lớn sao cho $C/n < \delta_\epsilon$, do biến cố $A_{n,t} = \{|X_t| \ge n\}$ có xác suất $\mathbb{P}(A_{n,t}) < \delta_\epsilon$ nên

$$\sup_{t\in T}\mathbb{E}(|X_t|\mathbb{I}_{|X_t|\ge n}) < \epsilon.$$

Do đó

$$\lim_{n\to\infty}\sup_{t\in T}\mathbb{E}(|X_t|\mathbb{I}_{|X_t|\ge n}) = 0.$$

($\Leftarrow$). Cố định $\epsilon > 0$, với mọi biến cố $A \in \mathcal{F}$, ta có

$$\begin{aligned}\mathbb{E}(|X_t|\mathbb{I}_A) =& \mathbb{E}[|X_t|\mathbb{I}_A(\mathbb{I}_{|X_t|<n} + \mathbb{I}_{|X_t|\ge n})]\\ \le& n\mathbb{P}(A) + \mathbb{E}[|X_t|\mathbb{I}_{|X_t|\ge n}].\end{aligned}$$

Theo giả thiết, tồn tại $n = n_\epsilon$ đủ lớn sao cho với mọi $t \in T$

$$\mathbb{E}[|X_t|\mathbb{I}_{|X_t|\ge n}] \le \epsilon/2.$$

Khi đó, với các biến cố A có xác suất $\mathbb{P}(A) < \epsilon/(2n)$, ta có

$$\mathbb{E}(|X_t|\mathbb{I}_A) < \epsilon.$$

Ngoài ra, với ϵ và n_ϵ như trên thì với mọi $t \in T$

$$\mathbb{E}(|X_t|) = \mathbb{E}[|X_t|\mathbb{I}_{|X_t|<n_\epsilon}] + \mathbb{E}[|X_t|\mathbb{I}_{|X_t|\geq n_\epsilon}] \leq n_\epsilon + \epsilon/2.$$

□

Từ đó, ta có cách kiểm tra khả tích đều đơn giản hơn theo kết quả sau.

Mệnh đề 3.1.15. *Xét một họ các biến ngẫu nhiên $\{X_t,\, t \in T\}$ khả tích bậc $p > 1$ sao cho tồn tại $B > 0$ để $\mathbb{E}|X_t|^p < B$ với mọi $t \in T$. Khi đó, họ các biến ngẫu nhiên đã cho khả tích đều.*

Chứng minh. Ta có

$$\mathbb{E}(|X_t|\mathbb{I}_{|X_t|\geq n}) \leq \mathbb{E}\left(\frac{|X_t|^p}{n^{p-1}}|X_t|\mathbb{I}_{|X_t|\geq n}\right) \leq \frac{\mathbb{E}|X_t|^p}{n^{p-1}} \leq \frac{B}{n^{p-1}}.$$

□

Mệnh đề 3.1.16. *Xét một họ các biến ngẫu nhiên $\{X_t,\, t \in T\}$ khả tích đều và biến ngẫu nhiên $X \in \mathcal{L}^1$. Khi đó họ biến ngẫu nhiên $\{X_t + X,\, t \in T\}$ khả tích đều.*

Chứng minh. Vì $\mathbb{E}|X_t + X| \leq \mathbb{E}|X_t| + \mathbb{E}|X|$ nên $\sup_{t\in T}\mathbb{E}|X_t + X| < \infty$.

Để kiểm tra với mọi $\epsilon > 0$, tồn tại $\delta_\epsilon > 0$ sao cho với bất cứ biến cố $A \in \mathcal{F}$ thỏa mãn $\mathbb{P}(A) \leq \delta_\epsilon$ thì

$$\mathbb{E}(|X_t + X|\mathbb{I}_A) < \epsilon,\ \forall t \in T;$$

ta có nhận xét

$$\mathbb{E}(|X_t + X|\mathbb{I}_A) \leq \mathbb{E}(|X_t|\mathbb{I}_A) + \mathbb{E}(|X|\mathbb{I}_A),$$

và dùng giả thiết khả tích đều của họ $\{X_t,\, t \in T\}$ cùng điều kiện khả tích của X như trong Định lý 2.3.13. □

Kết quả chính trong mục này được phát biểu như sau.

Định lý 3.1.17. *Dãy biến ngẫu nhiên X_n hội tụ về biến ngẫu nhiên X theo dạng L^1 khi và chỉ khi $X_n \xrightarrow{\mathbb{P}} X$ và họ các biến ngẫu nhiên $\{X_n,\, n \in \mathbb{N}\}$ khả tích đều.*

Chứng minh. ($\Rightarrow$). Vì $X_n \xrightarrow{L^1} X$ nên $X_n \xrightarrow{\mathbb{P}} X$. Ngoài ra ta có

$$\mathbb{E}|X_n| \to \mathbb{E}|X| < \infty,$$

nên $\sup\limits_{n\in\mathbb{N}} \mathbb{E}|X_n| < \infty$. Lại có, do $\mathbb{E}|X_n - X| \to 0$, nên với mọi $\epsilon > 0$ thì $\mathbb{E}|X_n - X| < \epsilon/2$ với $n \geq n_\epsilon$ đủ lớn. Do X khả tích nên theo Định lý 2.3.13 tồn tại $\delta > 0$ sao cho với mọi biến cố A thỏa mãn $\mathbb{P}(A) < \delta$ thì

$$\mathbb{E}(|X|\mathbb{I}_A) < \epsilon/2.$$

Từ đây suy ra với $n \geq n_\epsilon$ thì

$$\mathbb{E}(|X_n|\mathbb{I}_A) \leq \mathbb{E}(|X_n - X|\mathbb{I}_A) + \mathbb{E}(|X|\mathbb{I}_A) \leq \mathbb{E}|X_n - X| + \mathbb{E}(|X|\mathbb{I}_A) < \epsilon.$$

Với $n \leq n_\epsilon$, ta có hữu hạn biến ngẫu nhiên khả tích $X_1, \ldots, X_{n_\epsilon}$ nên áp dụng điều kiện khả tích trong Định lý 2.3.13 cho các biến ngẫu nhiên này và chọn $\delta_0 = \min\{\delta, \delta_1, \ldots, \delta_{n_\epsilon}\}$,thì

$$\mathbb{E}(|X_n|\mathbb{I}_A) < \epsilon.$$

Tức là với δ_0 chọn như trên thì với mọi biến cố A thỏa mãn $\mathbb{P}(A) < \delta_0$ thì

$$\mathbb{E}(|X_n|\mathbb{I}_A) < \epsilon,\ \forall n \in \mathbb{N}.$$

($\Leftarrow$). Theo Định lý 3.1.11, vì $X_n \xrightarrow{\mathbb{P}} X$ nên tồn tại dãy con X_{n_k} hội tụ về X hầu chắc chắn. Dãy con X_{n_k} thỏa mãn tính khả tích đều, nên theo Bổ đề Fatou, ta suy ra $X \in \mathcal{L}^1$.

Xét dãy $Y_n = X_n - X$. Suy ra dãy biến ngẫu nhiênY_n này khả tích đều. Ta sẽ chứng minh $\mathbb{E}|Y_n| = \mathbb{E}|X_n - X| \to 0$. Theo tính khả tích đểu, tồn tại $\delta_\epsilon > 0$ sao cho với bất cứ biến cố $A \in \mathcal{F}$ thỏa mãn $\mathbb{P}(A) \leq \delta_\epsilon$ thì

$$\mathbb{E}(|Y_n|\mathbb{I}_A) < \epsilon/2,\ \forall n \in T\mathbb{N}.$$

Lại do $Y_n \xrightarrow{\mathbb{P}} 0$, nên với mọi $\epsilon > 0$, với $n \geq n_\epsilon$ đủ lớn thì $\mathbb{P}(|Y_n| > \epsilon/2) < \delta_\epsilon$. Với $n \geq n_\epsilon$ ta có

$$\mathbb{E}|Y_n| = \mathbb{E}(|Y_n|\mathbb{I}_{|Y_n|>\epsilon/2})\mathbb{E}(|Y_n|\mathbb{I}_{|Y_n|\leq\epsilon/2}) \leq \epsilon.$$

Suy ra ĐPCM.

$\square$

3.2 Luật số lớn

Trong mục này ta xét $(X_n)_{n\geq1}$ là dãy biến ngẫu nhiên xác định trên không gian xác suất $(\Omega, \mathcal{F}, \mathbb{P})$ và kí hiệu tổng thành phần $S_n = X_1 + \ldots + X_n$.

3.2.1 Luật yếu số lớn

Trong ba dạng hội tụ ta đã biết là hầu chắc chắn, L^p và theo xác suất thì dạng hội tụ theo xác suất là dạng yếu nhất. Từ đó ta có định nghĩa về *luật yếu số lớn* như sau.

Định nghĩa 3.2.1. *Ta có luật yếu số lớn nếu tồn tại hai dãy số a_n và $b_n > 0$ sao cho*

$$\frac{S_n - a_n}{b_n} \xrightarrow{\mathbb{P}} 0.$$

Trước tiên ta có kết quả cơ bản như sau.

Định lý 3.2.2. *Giả sử $(X_n)_{n\geq 1}$ là dãy biến ngẫu nhiên thỏa mãn*

$$\lim_{n\to\infty} \frac{D(S_n)}{n^2} = 0.$$

Khi đó dãy biến ngẫu nhiên $\frac{S_n - \mathbb{E}S_n}{n}$ hội tụ về 0 theo dạng L^2 và dạng xác suất $\xrightarrow{\mathbb{P}} 0$, khi $n \to \infty$.

Chứng minh. Ta chỉ cần chứng minh sự hội tụ theo dạng L^2 là đủ. Thật vậy, ta có

$$\mathbb{E}\left(\frac{S_n - \mathbb{E}S_n}{n}\right)^2 = \frac{D(S_n)}{n^2} \to 0.$$

□

Hệ quả 3.2.3. *Giả sử $(X_n)_{n\geq 1}$ là dãy các biến ngẫu nhiên đôi một không tương quan và thỏa mãn*

$$\lim_{n\to\infty} \frac{D(X_1) + \ldots + D(X_n)}{n^2} = 0.$$

Khi đó

$$\frac{S_n - \mathbb{E}S_n}{n} \xrightarrow{\mathbb{P}} 0, \text{ khi } n \to \infty.$$

Chứng minh. Ta có $DS_n = D(X_1) + \ldots + D(X_n)$. □

Hệ quả 3.2.4. *Giả sử $(X_n)_{n\geq 1}$ là dãy các biến ngẫu nhiên đôi một tương quan âm (tức là $Cov(X_i, X_j) < 0$) và thỏa mãn*

$$\lim_{n\to\infty} \frac{D(X_1) + \ldots + D(X_n)}{n^2} = 0.$$

Khi đó

$$\frac{S_n - \mathbb{E}S_n}{n} \xrightarrow{\mathbb{P}} 0, \text{ khi } n \to \infty.$$

Chứng minh. Ta có $0 \leq DS_n < D(X_1) + \ldots + D(X_n)$. □

Hệ quả 3.2.5. *Giả sử $(X_n)_{n\geq 1}$ là dãy các biến ngẫu nhiên độc lập và cùng phân phối khả tích bình phương. Khi đó*

$$\frac{S_n}{n} \xrightarrow{\mathbb{P}} \mathbb{E}X_1, \ \textit{khi } n \to \infty.$$

Chứng minh. Ta có $DS_n <= n.D(X_1))$ và $\mathbb{E}S_n = n.\mathbb{E}X_1$. □

Ta có kết quả mạnh hơn như sau.

Định lý 3.2.6. *Giả sử $(X_n)_{n\geq 1}$ là dãy các biến ngẫu nhiên độc lập và cùng phân phối khả tích. Khi đó*

$$\frac{S_n}{n} \xrightarrow{L^1} \mathbb{E}X_1, \ \textit{khi } n \to \infty.$$

Chứng minh. Cố định $B > 0$. Ta sử dụng kĩ thuật làm tròn ("truncation") bằng cách biểu diễn

$$X_k = X_k\mathbb{I}_{|X_k|\leq B} + X_k\mathbb{I}_{|X_k|>B} = X_k^B + X_k^T,$$

với thành phần đầu tiên là biến ngẫu nhiên bị chặn, hành phần còn lại là biến ngẫu nhiên đuôi. Ta có

$$\mu = \mathbb{E}X_k = \mathbb{E}X_k^B + \mathbb{E}X_k^T = \mu_B + \mu_T.$$

Khi đó

$$\begin{aligned}\mathbb{E}\left|\frac{S_n}{n} - \mu\right| =& \frac{1}{n}\mathbb{E}|\sum_{i=1}^n (X_i - \mu)| \leq \frac{1}{n}\mathbb{E}|\sum_{i=1}^n (X_i^B - \mu_B)| + \frac{1}{n}\mathbb{E}|\sum_{i=1}^n (X_i^T - \mu_T)| \\ \leq& \frac{1}{n}\left(\mathbb{E}|\sum_{i=1}^n (X_i^B - \mu_B)|^2\right)^{1/2} + \frac{1}{n}\mathbb{E}|\sum_{i=1}^n (X_i^T - \mu_T)|.\end{aligned}$$

Ta thấy rằng hạng tử đầu tiến về 0 như trong chứng minh Định lý 3.2.2, còn hạng tử cuối bị chặn bởi $\dfrac{2n}{n}\mathbb{E}|X_1\mathbb{I}_{|X_1|>B}|$.

Do X_1 khả tích nên $\mathbb{E}|X_1\mathbb{I}_{|X_1|>B}| \to 0$ khi $B \to \infty$. Nên với mọi $\epsilon > 0$, ta có thể chọn B đủ lớn để $\mathbb{E}|X_1\mathbb{I}_{|X_1|>B}| < \epsilon/4$ và n_B sao cho với mọi $n > n_B$ thì

$$\frac{1}{n}\left(\mathbb{E}|\sum_{i=1}^n (X_i^B - \mu_B)|^2\right)^{1/2} < \epsilon/2.$$

Suy ra

$$\mathbb{E}\left|\frac{S_n}{n} - \mu\right| < \epsilon.$$

Ta có ĐPCM. □

3.2.2 Luật mạnh số lớn

Định nghĩa 3.2.7. *Ta có* luật mạnh số lớn *nếu tồn tại hai dãy số a_n và $b_n > 0$ sao cho*

$$\frac{S_n - a_n}{b_n} \xrightarrow{h.c.c} 0.$$

Ta có kết quả đầu tiên về luật mạnh số lớn như sau.

Định lý 3.2.8. *Giả sử $(X_n)_{n\geq 1}$ là dãy các biến ngẫu nhiên đôi một không tương quan và $\sup_n D(X_n^2) \leq \sigma^2 < \infty$. Khi đó*

$$\lim_{n\to\infty} \frac{S_n - \mathbb{E}S_n}{n} = 0 \quad h.c.c \text{ và trong } L^2.$$

Chứng minh. Trước hết ta giả sử $\mathbb{E}(X_n) = 0$. Đặt $Y_n = S_n/n$. Ta có $\mathbb{E}(Y_n) = 0$ và theo Mệnh đề 2.6.7 ta có

$$\mathbb{E}(Y_n^2) = \frac{1}{n^2}\sum_{i=1}^{n} DX_i \leq \frac{\sigma^2}{n}.$$

Vậy $Y_n \stackrel{L^2}{\to} 0$. Hơn nữa

$$\sum_{n=1}^{\infty} \mathbb{E}(Y_{n^2}^2) \leq \sum_{n=1}^{\infty} \frac{\sigma^2}{n^2} < \infty.$$

Áp dụng Định lí hội tụ đơn điệu 2.3.11, ta có

$$\mathbb{E}\Big[\sum_{n=1}^{\infty} Y_{n^2}^2\Big] = \sum_{n=1}^{\infty} \mathbb{E}(Y_{n^2}^2) < \infty,$$

vậy nên chuỗi $\sum_{n=1}^{\infty} Y_{n^2}^2 < \infty$ hầu chắc chắn. Từ đây suy ra

$$Y_{n^2} \stackrel{h.c.c}{\to} 0. \tag{3.1}$$

Với mỗi $n \in \mathbb{N}$, kí hiệu $p(n)$ là phần nguyên của $\sqrt{n}$. Khi đó, từ

$$Y_n - \frac{p(n)^2}{n} Y_{p(n)^2} = \frac{1}{n} \sum_{j=p(n)^2+1}^{n} X_j,$$

ta có

$$\mathbb{E}\Big[\Big(Y_n - \frac{p(n)^2}{n} Y_{p(n)^2}\Big)^2\Big] \leq \frac{n - p(n)^2}{n^2}\sigma^2 \leq \frac{2p(n)+1}{n^2}\sigma^2 \leq \frac{2\sqrt{n}+1}{n^2}\sigma^2 \leq \frac{3}{n^{3/2}}\sigma^2,$$

vì $n \leq (p(n)+1)^2$ và $p(n) \leq \sqrt{n}$. Lập luận tương tự như trên ta có từ

$$\sum_{n=1}^{\infty} \mathbb{E}\Big[\Big(Y_n - \frac{p(n)^2}{n} Y_{p(n)^2}\Big)^2\Big] \leq \sum_{n=1}^{\infty} \frac{3}{n^{3/2}}\sigma^2 < \infty,$$

suy ra

$$Y_n - \frac{p(n)^2}{n} Y_{p(n)^2} \overset{h.c.c}{\to} 0.$$

Điều này kết hợp với (3.1) và nhận xét $\frac{p(n)^2}{n} \to 1$ kéo theo $Y_n \overset{h.c.c}{\to} 0$.

Nếu $\mathbb{E}(X_n) \neq 0$, ta đặt $Z_n = X_n - \mathbb{E}(X_n)$. Khi đó $(Z_n)_{n\geq 1}$ là dãy các biến ngẫu nhiên đôi một không tương quan, có kì vọng bằng không và phương sai bị chặn đều. Do đó

$$\frac{S_n - \mathbb{E}S_n}{n} = \frac{Z_1 + \ldots + Z_n}{n} \overset{h.c.c}{\to} 0.$$

$\square$

Luật mạnh số lớn vẫn đúng cho dãy biến ngẫu độc lập cùng phân phối khả tích.

Định lý 3.2.9. *Nếu $(X_n)_{n\geq 1}$ là dãy biến ngẫu nhiên độc lập, cùng phân phối khả tích. Khi đó*

$$\frac{S_n}{n} \xrightarrow{h.c.c} \mathbb{E}(X_1).$$

Sau đây ta xét một số ứng dụng của luật số lớn. Tính chất quan trọng của luật số lớn đó là khi ta lấy trung bình của nhiều biến ngẫu nhiên thì *tính ngẫu nhiên* lại giảm đi, đến giới hạn thì chỉ còn hằng số tất định. Thông thường ta sẽ sử dụng luật số lớn cho các biến ngẫu nhiên độc lập cùng phân phối, và khi đó hằng số giới hạn chính là kì vọng. Một trong những ứng dụng của luật số lớn là bài toán thống kê đưa ra xấp xỉ cho kì vọng.

i. Để đo chiều cao hoặc cân nặng trung bình, ví dụ của thanh niên Việt Nam, việc cân đo cho tất cả mọi người là không khả thi, khi đó ta chỉ cần lấy trên một mẫu số lượng đủ lớn và lấy giá trị trung bình.

ii. Tương tự, để kiểm tra tính cân đối của đồng xu (hoặc xúc xắc), ta có thể thử tung nhiều lần và tính tỉ lệ thực nghiệm.

iii. Với bài toán Buffon, trong công thức kết quả của xác suất có xuất hiện số π, từ đó ta có cách xấp xỉ số π theo cách thực nghiệm là thực hiện tung que tăm nhiều lần và tính tỉ lệ que tăm cắt vào dải. Chú ý là cách làm thực nghệm này tỏ ra hiệu quả hơn cách xấp xỉ số π cổ điển là tính chu vi đa giác đều.

iv. Ta có một kết quả rất thú vị là với dãy số ngẫu nhiên thì ta có bảng phân phối của chữ số hàng cao nhất theo phân phối Benford như sau

X	1	2	...	9
$\mathbb{P}$	$\log_{10} 2$	$\log_{10}(3/2)$	...	$\log_{10}(10/9)$

Do đó, một cách kiểm tra bảng dữ liệu số đang có là tự nhiên hay nhân tạo đó là tính tỉ lệ xuất hiện các chữ số đầu tiên xem có gần với phân phối Benford hay không.

Ta có hai ví dụ rất thú vị sẽ được trình bày tiếp theo.

3.2.3 Phương pháp Monte Carlo

Phương pháp Monte Carlo là phương pháp xấp xỉ số sử dụng máy tính. Ta xét ví dụ sau.

Ví dụ 3.2.10. *Giả sử f là một hàm đo được và khả tích trên đoạn $[0,1]$. Hãy tính xấp xỉ giá trị $I = \int_0^1 |f(x)|dx$.*

Nói chung trong phần lớn các ứng dụng thực tế, giá trị của $I = \int_0^1 f(x)dx$ không thể được xác định một cách chính xác bằng các phương pháp giải tích nên người ta phải tìm cách xấp xỉ I. Khi hàm f đủ trơn, I có thể được xấp xỉ khá tốt bằng các phương pháp vét cạn truyền thống như hình chữ nhật hoặc hình thang cân. Ví dụ khi f khả vi liên tục đến cấp 2 thì ta có công thức hình bình hành sau

$$I \approx \frac{f(t_0^n) + 2f(t_1^n) + \ldots + 2f(t_{n-1}^n) + f(t_n^n)}{2n},$$

trong đó $t_i^n = \frac{i}{n},\ i = 0, 1, \ldots, n$.

Sai số của các phép xấp xỉ trên thường dựa vào chuẩn sup của đạo hàm. Do đó, khi f không đủ trơn thì phương pháp trên thường tỏ ra kém hiệu quả.

Trong trường hợp này ta có thể sử dụng phương pháp Monte Carlo như sau. Ta biểu diễn giá trị I thành kì vọng của một biến ngẫu nhiên. Cụ thể, xét $U \sim U[0,1]$ có phân phối đều trên đoạn $[0,1]$, ta có nhận xét quan trọng là

$$I = \int_0^1 |f(x)|dx = \mathbb{E}f(U).$$

Do đó, với dãy biến ngẫu nhiên độc lập cùng phân phối $U_i \sim U[0,1]$, ta sinh được dãy biến độc lập cùng phân phối $f(U_i)$ khả tích. Khi đó, từ luật mạnh số lớn ta có

$$I_n = \frac{f(U_1) + \ldots + f(U_n)}{n} \xrightarrow{h.c.c} \mathbb{E}f(U) = I$$

Ý tưởng của phương pháp Monte-Carlo là ta sẽ dùng máy tinh (hoặc bảng số) để sinh ra mẫu ngẫu nhiên $u_1, \ldots, u_n$, và tính trung bình thực nghiệm

$$\frac{f(u_1) + \ldots + f(u_n)}{n} \approx I.$$

Để đánh giá sai số của ước lượng, ta giả thiết thêm điều kiện khả tích bình phương

$$\int_0^1 |f(x)|^2 dx < \infty. \tag{3.2}$$

Khi đó, bình phương sai số của ước lượng là

$$\mathbb{E}[(I_n - I)^2] = \mathbb{E}[(I_n - \mathbb{E}[I_n])^2] = \frac{1}{n} Df(U_1) \leq \frac{1}{n} \int_0^1 |f(x)|^2 dx.$$

Từ đó suy ra sai số của ước lượng chỉ phụ thuộc vào cỡ mẫu n mà không phụ thuộc vào độ trơn của f. Tổng quát hơn, ta cũng có thể sử dụng phương pháp Monte Carlo để tính các tích phân bội nhiều lớp. Trong tình huống này, nếu theo cách chia hình truyền thống thì việc tính toán rất phức tạp và sai số tương đối lớn, điều này cho thấy ưu thế vượt trội của phương pháp Monte Carlo.

Trên đây là những ý tưởng sơ khai ban đầu về phương pháp Monte Carlo. Ngày nay, các nhà nghiên cứu tập trung cải thiện phương pháp để giảm khối lượng và thời gian tính toán của máy tính, cũng như cải thiện sai số, ví dụ như phương pháp Monte Carlo Markov Chain (MCMC).

3.2.4 Định lý xấp xỉ Weierstrass

Ta sẽ sử dụng luật số lớn để chứng minh kết quả nổi tiếng sau trong Lý thuyết giải tích.

Định lý 3.2.11. *Cho hàm liên tục $f : [0,1] \to \mathbb{R}$. Khi đó, hàm f có thể xấp xỉ đều bởi các đa thức, theo nghĩa là với mọi $\epsilon > 0$ đều tồn tại một đa thức $P(x)$ sao cho $\|f - P\| = \max_{x \in [0,1]} |f(x) - P(x)| < \epsilon$.*

Chứng minh. Ta xét *đa thức Bernstein* như sau

$$B_n(x) = \sum_{i=0}^{n} \binom{n}{i} x^i (1-x)^{n-i} f\left(i/n\right).$$

Trước tiên, ta sẽ chứng minh dãy B_n hội tụ điểm tới f, tức là với mọi $x \in [0,1]$ thì $B_n(x) \to f(x)$. Thật vậy, xét biến ngẫu nhiên $S_n \sim B(n,x)$ có phân phối nhị thức. Khi đó, S_n có thể biểu diễn thành tổng của n biến ngẫu độc lập có cùng phân phối Bernoulli $Ber(x)$ với x là xác suất thành công của phép thử. Theo luật mạnh số lớn,

$$\frac{S_n}{n} \xrightarrow{h.c.c} x.$$

Vì hàm f liên tục nên ta cũng có $f(S_n/n) \xrightarrow{h.c.c} f(x)$. Ngoài ra, hàm f liên tục trên tập compact nên có giá trị tuyệt đối bị chặn bởi M. Theo Định lý hội tụ bị chặn, $\mathbb{E}f(S_n/n) \to f(x)$. Từ bảng phân phối của S_n ta có $\mathbb{E}f(S_n/n) = B_n(x)$.

Tiếp theo ta sẽ chứng minh tính hội tụ đều. Ta sử dụng tính chất hàm liên tục trên tập compact thì liên tục đều, tức là với mọi $\epsilon > 0$ tồn tại $\delta > 0$ sao cho $|f(x) - f(y)| < \epsilon$ với mọi $|x - y| < \delta$. Cố định ϵ và δ, sử dụng bất đẳng thức Chebyshev

$$\mathbb{P}\left(\left|\frac{S_n}{n} - x\right| \geq \delta\right) \leq \frac{x(1-x)}{n\delta^2},$$

ta có

$$\mathbb{P}\left(|f(S_n/n) - f(x)| \geq \epsilon\right) \leq \frac{x(1-x)}{n\delta^2}.$$

Suy ra

$$\begin{aligned}\mathbb{E}|f(S_n/n) - f(x)| =& \mathbb{E}\left[|f(S_n/n) - f(x)|(\mathbb{I}_{|f(S_n/n)-f(x)|\geq\epsilon} + \mathbb{I}_{|f(S_n/n)-f(x)|<\epsilon})\right] \\ \leq& \epsilon + \frac{2Mx(1-x)}{n\delta^2}.\end{aligned}$$

Khi đó, với mọi $x \in [0, 1]$, ta có

$$|B_n(x) - f(x)| = |\mathbb{E}f(S_n/n) - f(x)| \leq \mathbb{E}|f(S_n/n) - f(x)| \leq \epsilon + \frac{2Mx(1-x)}{n\delta^2}.$$

□

3.3 Hội tụ yếu

Với cách nhìn biến ngẫu nhiên là một hàm số, ta có các dạng hội tụ như hầu chắc chắn, L^p và theo xác suất. Ngoài ra, ta có thể nhìn biến ngẫu nhiên như là một độ đo, thông qua phân phối xác suất của biến ngẫu nhiên trên $(\mathbb{R}\mathcal{B}(\mathbb{R}))$. Với cách nhìn này, ta không cần quan tâm đến không gian xác suất nền $(\Omega, \mathcal{F}, \mathbb{P})$. Và ta sẽ hiểu sự hội tụ của các biến ngẫu nhiên thông qua sự hội tụ của các độ đo. Ta có hai hướng tiếp cận cho sự hội tụ của độ đo như sau.

Định nghĩa 3.3.1. *Dãy biến ngẫu nhiên (X_n) được gọi là* hội tụ yếu *đến biến ngẫu nhiên X, kí hiệu là $X_n \xrightarrow{w} X$, nếu với mọi hàm $f : \mathbb{R} \to \mathbb{R}$ liên tục và bị chặn, ta có*

$$\lim_{n\to\infty} \mathbb{E}[f(X_n)] = \mathbb{E}[f(X)].$$

Định nghĩa 3.3.2. *Dãy biến ngẫu nhiên (X_n) được gọi là* hội tụ theo phân phối *đến biến ngẫu nhiên X, kí hiệu là $X_n \xrightarrow{d} X$, nếu hàm phân phối $F_{X_n}(t)$ hội tụ đến hàm phân phối $F_X(t)$ tại mọi điểm liên tục của hàm F_X.*

Ta có các ví dụ sau đây về hội tụ yếu và hội tụ theo phân phối

Ví dụ 3.3.3. *Xét biến ngẫu nhiên X_n có phân phối đều với tập giá trị $\{1/n, 2 : n, \dots, 1\}$, tức là ta có bảng phân phối như sau:*

X	$1/n$	$2/n$	$\dots$	1
$\mathbb{P}$	$1/n$	$1/n$	$\dots$	$1/n$

Khi đó, với mọi hàm f liên tục thì

$$\mathbb{E}f(X_n) = \frac{1}{n}\sum_{i=1}^{n} f(i/n) = \int_0^1 f(x)dx = \mathbb{E}f(U[0,1]).$$

Suy ra ta có dãy X_n hội tụ yếu đến biến ngẫu nhiên có phân phối đều $U[0,1]$.

Ví dụ 3.3.4. *Xét dãy biến ngẫu nhiên $X_n \sim U[0, 1/n]$. Ta có thể kiểm tra X_n hội tụ theo phân phối về biến ngẫu nhiên $X = 0$ h.c.c.*

Ví dụ 3.3.5. *Xét dãy biến ngẫu nhiên $X_n \sim B(n, p_n)$. Nếu $n.p_n \to \lambda > 0$, ta có thể kiểm tra X_n hội tụ theo phân phối về biến ngẫu nhiên $X \sim Poi(\lambda)$. Thật vậy, ta chỉ cần kiểm tra với mọi $k \in \mathbb{N}$ thì*

$$\mathbb{P}(X_n = k) = \binom{n}{k} p^k (1-p)^{n-k} \to e^{-\lambda}\frac{\lambda^k}{k!} = \mathbb{P}(Poi(\lambda) = k).$$

Từ định nghĩa hội tụ yếu, ta có ngay kết quả sau.

Mệnh đề 3.3.6. *Nếu $X_n \xrightarrow{w} X$ thì với mọi hàm h liên tục, ta có $h(X_n) \xrightarrow{w} h(X)$.*

Ta có các kết quả sau về mối liên hệ giữa hội tụ yếu và dạng hội tụ theo xác suất.

Mệnh đề 3.3.7. *Giả sử $(X_n)_{n\geq 1}$ và X là các biến ngẫu nhiên xác định trên cùng không gian xác suất $(\Omega, \mathcal{F}, \mathbb{P})$. Nếu $X_n \xrightarrow{\mathbb{P}} X$ thì $X_n \xrightarrow{w} X$.*

Chứng minh. Ta chứng minh bằng phản chứng. Giả sử $X_n \xrightarrow{\mathbb{P}} X$ nhưng $X_n \overset{w}{\not\longrightarrow} X$. Khi đó tồn tại một hàm liên tục bị chặn f, hằng số $\epsilon > 0$ và dãy con $(n_k)_{k\geq 1}$ sao cho

$$|\mathbb{E}(f(X_{n_k})) - \mathbb{E}(f(X))| > \epsilon \quad \text{với mọi} \quad k \geq 1. \tag{3.3}$$

Theo Mệnh đề 3.1.11, tồn tại dãy con $(m_k)_{k\geq 1}$ của dãy $(n_k)_{k\geq 1}$ sao cho $X_{m_k} \xrightarrow{h.c.c} X$. Do f liên tục nên $f(X_{m_k}) \xrightarrow{h.c.c} f(X)$. Lại do f bị chặn nên áp dụng định lí hội tụ bị chặn Lebesgue, $\mathbb{E}(f(X_{m_k})) \to \mathbb{E}(f(X))$. Điều này mâu thuẫn với (3.3). Vậy ta được điều phải chứng minh. $\square$

Mệnh đề 3.3.8. *Giả sử $(X_n)_{n\geq 1}$ và X là các biến ngẫu nhiên xác định trên cùng không gian xác suất $(\Omega, \mathcal{F}, \mathbb{P})$. Nếu $X_n \xrightarrow{w} X$ và $X = const$ h.c.c thì $X_n \xrightarrow{\mathbb{P}} X$.*

Chứng minh. Giả sử $X \equiv a$ h.c.c. Xét hàm liên tục và bị chặn $f(x) = \frac{|x-a|}{|x-a|+1}$. Do $X_n \xrightarrow{w} a$ nên $\mathbb{E}(f(X_n)) \to f(a) = 0$. Theo Mệnh đề 3.1.10 ta có $X_n \xrightarrow{\mathbb{P}} a$. □

Kết quả sau chỉ ra rằng hai dạng hội tụ yếu và theo phân phối thực chất là một.

Định lý 3.3.9. *Ta có $X_n \xrightarrow{w} X$ khi và chỉ khi $X_n \xrightarrow{d} X$.*

Để chứng minh Định lý trên, ta sử dụng Định lý biểu diễn Skorokhod như sau.

Định lý 3.3.10. *Nếu như $F_{X_n}(x) \to F_X(x)$ tại mọi điểm liên tục của F_X thì sẽ tồn tại dãy biến ngẫu nhiên $Y_n \stackrel{d}{=} X_n$ và $Y \stackrel{d}{=} X$ sao cho $Y_n \xrightarrow{h.c.c} Y$.*

Chứng minh. Ý tưởng chứng minh dựa trên việc xây dựng các biến ngẫu Y_n trên không gian $([0,1], \mathcal{B}([0,1]), \lambda)$. □

Ta có chứng minh của Định lý 3.3.9 như sau.

Chứng minh. $(\Rightarrow)$. Xét dãy hàm liên tục bị chặn như sau

$$g_{x,\epsilon} = \Big\{ 1, \text{ nếu } u \leq x0, \text{ nếu } u \geq x + \epsilon(x+\epsilon-u)/\epsilon, \text{ nếu } x < u < x+\epsilon.$$

Dãy hàm như trên là xấp xỉ cho hàm $\mathbb{I}_{(-\infty,x]}$. Ta có

$$\limsup^{n\to\infty} \mathbb{P}(X_n \leq x) \leq \limsup^{n\to\infty} \mathbb{E} g_{x,\epsilon}(X_n) = \mathbb{E} g_{x,\epsilon}(X) \leq \mathbb{P}(X \leq x+\epsilon.$$

Cho $\epsilon \to 0$ và sử dụng tính liên tục tại điểm x ta có

$$\limsup^{n\to\infty} \mathbb{P}(X_n \leq x) \leq \mathbb{P}(X \leq x).$$

Ta cũng có

$$\liminf^{n\to\infty} \mathbb{P}(X_n \leq x) \geq \liminf^{n\to\infty} \mathbb{E} g_{x-\epsilon,\epsilon}(X_n) = \mathbb{E} g_{x-\epsilon,\epsilon}(X) \geq \mathbb{P}(X \leq x-\epsilon,$$

nên lại cho $\epsilon \to 0$ và sử dụng tính liên tục tại điểm x ta có

$$\limsup^{n\to\infty} \mathbb{P}(X_n \leq x) \geq \mathbb{P}(X \leq x).$$

Ta chứng minh xong chiều suy ra.

$(\Leftarrow)$. Áp dụng định lý biểu diễn Skorokhod và định lý hội tụ bị chặn. □

Định lý 3.3.11 (Định lý Slutsky)**.** *Cho hai dãy biến ngẫu nhiên $\{X_n\}$ và $\{Y_n\}$ xác định trên cùng một không gian xác suất sao cho $X_n \xrightarrow{d} X$ và $Y_n \xrightarrow{d} c = const$. Khi đó*

$$X_n + Y_n \xrightarrow{d} X + c, \quad \text{và} \quad X_n Y_n \xrightarrow{d} cX.$$

3.3.1 Hàm đặc trưng

Trong phần này chúng ta tìm hiểu về hàm đực trưng. Đây là công cụ mạnh để nghiên cứu về sự hội tụ yếu.

Định nghĩa 3.3.12. *1. Hàm đặc trưng của biến ngẫu nhiên X được xác định là ánh xạ $\varphi_X : \mathbb{R} \to \mathbb{C}$ cho bởi*

$$\varphi_X(t) = \mathbb{E}[e^{itX}] = \int_{\mathbb{R}} e^{itx} dF_X(x).$$

2. Hàm đặc trưng của véc tơ ngẫu nhiên $X = (X_1, \dots, X_n)$ là ánh xạ $\varphi_X : \mathbb{R}^d \to \mathbb{C}$ cho bởi

$$\varphi_X(t_1, \dots, t_n) = \mathbb{E}\Big[\exp\Big(i\sum_{j=1}^n t_j X_j\Big)\Big].$$

Với các phân phối thường gặp, ta tính được các hàm đặc trnwg như sau

Ví dụ 3.3.13. *1. Với X có phân phối Rademacher, thì $\varphi_X(t) = \cos t$.*

2. Với $X \sim Poi(\lambda)$, thì $\varphi_X(t) = \exp(\lambda(e^{it} - 1))$.

3. Với $X \sim Exp(\lambda)$, thì $\varphi_X(t) = \dfrac{\lambda}{\lambda - it}$.

4. Với $X \sim U([-K, K])$, thì $\varphi_X(t) = \dfrac{\sin Kt}{Kt}$.

5. Với X có phân phối chuẩn $N(0,1)$. Khi đó

$$\varphi_X(t) = \int_{-\infty}^{\infty} e^{itx}\frac{e^{-x^2/2}}{\sqrt{2\pi}}dx = \int_{-\infty}^{\infty}\frac{\cos tx}{\sqrt{2\pi}}e^{-x^2/2}dx + \int_{-\infty}^{\infty}\frac{\sin tx}{\sqrt{2\pi}}e^{-x^2/2}dx.$$

Do hàm $x \mapsto e^{-x^2/2}\sin tx$ là hàm khả tích và lẻ nên tích phân thứ hai bằng 0. Theo Định li 3.3.19 ta có thể đạo hàm hai vế của đẳng thức trên theo t và thu được

$$\varphi_X'(t) = -\int_{-\infty}^{\infty}\frac{\sin tx}{\sqrt{2\pi}}xe^{-x^2/2}dx.$$

Áp dụng công thức tích phân từng phần, ta được

$$\varphi_X'(t) = -\int_{-\infty}^{\infty}\frac{\cos tx}{\sqrt{2\pi}}te^{-x^2/2}dx = -t\varphi_X(t).$$

Giải phương trình vi phân $\frac{\varphi_X'}{\varphi_X} = -t$ với điều kiện $\varphi_X(0) = 1$, ta được

$$\varphi_X(t) = e^{-t^2/2}.$$

Nếu X có phân phối chuẩn $N(a, \sigma^2)$ thì

$$\varphi_X(t) = \frac{1}{\sqrt{2\pi\sigma^2}} \int e^{itx} e^{-\frac{(x-a)^2}{2\sigma^2}} dx.$$

Đổi biến $y = (x-a)/\sigma$, ta được

$$\varphi_X(t) = \frac{e^{ita}}{\sqrt{2\pi}} \int e^{it\sigma y} e^{-y^2/2} dy = e^{ita - t^2\sigma^2/2}.$$

Từ định nghĩa của hàm đặc trưng, ta có các tính chất sau.

Định lý 3.3.14. *Cho biến ngẫu nhiên X. Khi đó hàm đặc trưng $\varphi_X(t)$ thỏa mãn các tính chất*

1. $\varphi_X(0) = 1$.
2. $|\varphi_X(t)| \le \mathbb{E}|e^{itX}| = 1$.
3. $\varphi_X(-t) = \varphi_{-X}(t) = \overline{\varphi_X(t)}$. *Suy ra nếu X có phân phối đối xứng, tức là $X \stackrel{d}{=} -X$, thì hàm đặc trưng $\varphi_X(t)$ nhận giá trị thực.*
4. $\varphi_X(t)$ *là hàm liên tục đều.*
$$|\varphi_X(t+h) - \varphi_X(t)| \le \mathbb{E}|e^{ihX} - 1| \xrightarrow{h \to 0} 0.$$
5. $\varphi_{aX+b}(t) = e^{itb}\varphi_X(at)$ *với mọi $a, b \in \mathbb{R}$.*
6. *Nếu X và Y độc lập thì $\varphi_{X+Y}(t) = \varphi_X(t).\varphi_Y(t)$.*

Trong trường hợp khi X là biến ngẫu nhiên liên tục có hàm mật độ f_X, từ định nghĩa của hàm đặc trưng, ta có

$$\varphi_X(t) = \int_{\mathbb{R}}^{itx} f_X(x)dx = \hat{f}_X(t).$$

Hay nói cách khác hàm đặc trưng chính là hàm biến đổi Fourier của hàm mật độ, hay trong trường hợp tổng quát là của phân phối xác suất của X. Theo Lý thuyết giải tích Fourier, ta có công thức ngược để tính f_X từ hàm đặc trưng như sau

$$f_X(x) \stackrel{h.c.c}{=} \int_{\mathbb{R}} e^{2\pi itx} \hat{f}_X(t) dt.$$

Với độ đo tổng quát, ta có kết quả sau

Định lý 3.3.15 (Công thức ngược Levy)**.** *Với $a < b$ và φ là hàm đặc trưng của độ đo μ thì*

$$\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} \varphi(t) \left(\frac{e^{-ita} - e^{-itb}}{it} \right) dt = \mu(a,b) + \frac{1}{2}\mu\{a,b\}.$$

Ta có hệ quả như sau.

Hệ quả 3.3.16. *1. Nếu $\varphi_X(t) = \varphi_Y(t)$ với mọi $t \in \mathbb{R}$ thì $X \stackrel{d}{=} Y$.*

2. Nếu $\varphi_X(t) \in L^1(\mathbb{R})$ thì biến ngẫu nhiên X có hàm mật độ f_X xác định bởi

$$f_X(x) = \frac{1}{2\pi} \int_{\mathbb{R}} \varphi_X(t) e^{-itx} dt.$$

Từ đây ta có nhận xét là thuật ngữ "đặc trưng" để diễn đạt cho việc hàm đặc trưng xác định duy nhất phân phối tương ứng. Ta có thể tổng quát cho hàm đặc trưng của vector ngẫu nhiên.

Ví dụ 3.3.17. *Giả sử X và Y là hai biến ngẫu nhiên độc lập có phân phối Poisson với tham số lần lượt là μ và λ. Khi đó $X + Y$ có hàm đặc trưng là*

$$\varphi_{X+Y}(t) = \mathbb{E}(e^{it(X+Y)}) = \mathbb{E}(e^{itX})\mathbb{E}(e^{itY}) = e^{(\lambda+\mu)(e^{it}-1)}.$$

Do đó $X + Y$ có phân phối Poisson với tham số $\mu + \lambda$.

Mệnh đề 3.3.18. *Một hàm đặc trưng có tính chất nửa xác định dương, theo nghĩa với mọi bộ số $(t_1, \ldots, t_n) \in \mathbb{R}^n$ và $(z_1, \ldots, z_n) \in \mathbb{C}^n$ thì*

$$\sum_{i,j} \overline{z}_i z_j f(t_i - t_j) \geq 0.$$

Và ngược lại, một hàm có tính chất nửa xác định dương, liên tục tại 0 và có giá trị $\varphi(0) = 1$ thì sẽ là hàm dặc trưng của một phân phối xác suất trên $\mathbb{R}$.

Tiếp theo ta tìm hiểu về sự liên hệ giữa các moment của biến ngẫu nhiên X với tính khả vi của hàm đặc trưng $\varphi_X(t)$.

Định lý 3.3.19. *Giả sử biến ngẫu nhiên X có $\mathbb{E}(|X|^n) < \infty$ với số tự nhiên n nào đó. Khi đó hàm đặc trưng φ_X có đạo hàm liên tục đến cấp n tại mọi điểm và*

$$\varphi^{(k)}(t) = i^k \mathbb{E}(X^k e^{itX}), \tag{3.4}$$

$$\mathbb{E}(X^k) = \frac{\varphi^{(k)}(0)}{i^k}, \tag{3.5}$$

$$\varphi_X(t) = \sum_{k=0}^{n} \frac{(it)^k}{k!} \mathbb{E}(X^k) + \frac{(it)^n}{n!} \alpha_n(t), \tag{3.6}$$

trong đó $|\alpha_n(t)| \leq 2\mathbb{E}(|X^n|)$ *và* $\alpha_n(t) \to 0$ *khi* $t \to 0$.

Chứng minh. Do $\mathbb{E}(|X|^n) < \infty$ nên với mọi $k = 1, \ldots, n$ ta có $\mathbb{E}(|X|^k) < \infty$, và do đó

$$\sup_t \int |(ix)^k e^{itx}| dF_X(x) \leq \int |x|^k dF_X(x) < \infty.$$

Theo định lí Lebesgue, ta có thể lấy đạo hàm dưới dấu tích phân để thu được (3.4). (3.5) là hệ quả của (3.4) khi cho $t = 0$. Mặt khác, áp dụng khai triển Taylor cho hàm e^x tại 0, ta được

$$\begin{aligned}\mathbb{E}(e^{itx}) &= \mathbb{E}\Big(\sum_{k=0}^{n-1} \frac{(itX)^k}{k!} + \frac{(itX)^n}{n!} e^{i\theta X}\Big)\\ &= \sum_{k=0}^{n-1} \frac{(it)^k}{k!} \mathbb{E}(X^k) + \frac{(it)^n}{n!} \Big(\mathbb{E}(X^n) + \alpha_n(t)\Big),\end{aligned}$$

trong đó $|\theta| \leq 1,\ \alpha_n(t) = \mathbb{E}\big(X^n(e^{itX} - 1)\big)$. Ta có $|\alpha_n(t)| \leq 2\mathbb{E}(|X|^n)$ và áp dụng định lí hội tụ bị chặn, ta có $\alpha_n(t) \to 0$ khi $t \to 0$. □

Khi ta có các thông tin về tính liên tục của hàm đặc trưng tại 0, ta có thể tính được moment của X.

Mệnh đề 3.3.20. *Nếu*

$$\limsup_{h \to 0} \frac{\varphi(h) - 2\varphi(0) + \varphi(-h)}{h^2} > -\infty$$

(thỏa mãn khi $\varphi''(0)$ *tồn tại), thì* $\mathbb{E}X^2 < \infty$.

Tổng quát hơn ta có

Định lý 3.3.21. *Nếu* $\varphi^{(2m)}(0)$ *tồn tại và hữu hạn với số nguyên dương* m *nào đó thì* $\mathbb{E}(X^{2m}) < \infty$.

Hàm đặc trưng có thể được sử dụng để kiểm tra tính độc lập của các biến ngẫu nhiên.

Định lý 3.3.22. *Các biến ngẫu nhiên* $X_1, \ldots, X_n$ *là độc lập với nhau khi và chỉ khi*

$$\varphi_{(X_1,\ldots,X_n)}(t_1, \ldots, t_n) = \varphi_{X_1}(t_1) \ldots \varphi_{X_n}(t_n) \quad \text{với mọi} \quad (t_1, \ldots, t_n) \in \mathbb{R}^n.$$

Cuối cùng, ta đưa ra định lý chính của phần này cho phép kiểm tra sự hội tụ yếu thông qua sự hội tụ của hàm đặc trưng.

Định lý 3.3.23 (Định lý liên tục Levy)**.** *Giả sử $(F_n)_{n\geq 1}$ là dãy hàm phân phối xác suất với $(\varphi_n)_{n\geq 1}$ là dãy hàm đặc trưng tương ứng,*

$$\varphi_n(t) = \int_{\mathbb{R}} e^{itx} dF_n(x).$$

1. *Nếu $F_n \overset{w}{\to} F$ với F là hàm phân phối xác suất nào đó thì (φ_n) hội tụ điểm đến hàm đặc trưng φ của F.*
2. *Giả sử $\varphi_n(t) \to \varphi(t),\ t \in \mathbb{R}$. Khi đó hai mệnh đề sau là tương đương*
   (a) *$\varphi(t)$ là hàm đặc trưng và $F_n \overset{w}{\to} F$ với F là hàm phân phối xác suất tương ứng với φ;*
   (b) *φ liên tục tại $t = 0$.*

Trong ví dụ sau đây, ta chứng minh lại luật yếu số lớn dựa trên phương pháp hàm đặc trưng.

Ví dụ 3.3.24 (Luật yếu số lớn)**.** *Giả sử $(X_k)_{k\geq 1}$ là dãy biến ngẫu nhiên độc lập cùng phân phối với kì vọng chung là a hữu hạn. Khi đó*

$$\frac{1}{n}(X_1 + \ldots + X_n) \xrightarrow{\mathbb{P}} a.$$

Thật vậy, đặt $S_n = X_1 + \ldots + X_n$ và gọi φ là hàm đặc trưng của X_k. Khi đó

$$\varphi_{S_n/n}(t) = \varphi_{S_n}(t/n) = [\varphi(t/n)]^n.$$

Theo Định lí 3.3.19,

$$\varphi(t/n) = 1 + \frac{ita}{n} + \frac{t}{n}\alpha(t/n),$$

trong đó $\alpha(t) \to 0$ khi $t \to 0$. Do đó

$$\varphi_{S_n/n}(t) = \Big(1 + \frac{ita}{n} + \frac{t}{n}\alpha(t/n)\Big)^n \to e^{ita}$$

khi $n \to \infty$. Lưu ý rằng nếu biến ngẫu nhiên $X \equiv a$ thì hàm đặc trưng của X là $\varphi_X(t) = e^{ita}$. Do vậy, $S/n \overset{w}{\to} a$. Áp dụng Mệnh đề 3.3.8 ta được điều phải chứng minh.

3.4 Định lí giới hạn trung tâm

Dựa trên công cụ hàm đặc trưng, chúng ta có thể chứng minh Định lý giới hạn trung tâm dạng cổ điển. Đây là một trong những kết quả quan trọng và đẹp đẽ nhất của Lý thuyết xác suất.

Định lý 3.4.1. *Giả sử $(X_n)_{n\geq 1}$ là dãy biến ngẫu nhiên độc lập và cùng phân phối với $\mathbb{E}(X_n) = \mu$ và $DX_n = \sigma^2 \in (0, \infty)$. Đặt $S_n = X_1 + \ldots + X_n$. Khi đó dãy biến ngẫu nhiên $Y_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}$ hội tụ theo phân phối tới biến ngẫu nhiên có phân phối chuẩn tắc $N(0,1)$.*

Chứng minh. Gọi φ là hàm đặc trưng của biến ngẫu nhiên $X_n - \mu$. Do $(X_n)_{n\geq 1}$ là dãy biến ngẫu nhiên có cùng phân phối nên φ không phụ thuộc vào n. Do các biến ngẫu nhiên $(X_j)_{j\geq 1}$ là độc lập nên

$$\varphi_{Y_n}(t) = \mathbb{E}\exp\Big(it\sum_{j=1}^{n}\frac{X_j-\mu}{\sigma\sqrt{n}}\Big) = \prod_{j=1}^{n}\mathbb{E}\exp\Big(it\frac{X_j-\mu}{\sigma\sqrt{n}}\Big) = \Big(\varphi\big(\frac{t}{\sigma\sqrt{n}}\big)\Big)^n.$$

Do $\mathbb{E}(X_j - \mu) = 0$ và $\mathbb{E}((X_j-\mu)^2) = \sigma^2$ nên theo Định lí 3.3.19, φ có đạo hàm cấp hai liên tục và hơn nữa

$$\varphi(t) = 1 - \frac{\sigma^2 t^2}{2} + t^2\alpha(t),$$

trong đó $\alpha(t) \to 0$ khi $t \to 0$. Do đó sử dụng khai triển $\ln(1+x) = x + o(x)$ khi $x \to 0$, ta được

$$\ln\varphi_{Y_n}(t) = n\ln\Big(1 - \frac{t^2}{2n} + \frac{t^2}{n\sigma^2}\alpha\big(\frac{t}{\sigma\sqrt{n}}\big)\Big) \to -\frac{t^2}{2},$$

tức là $\varphi_{Y_n}(t) \to e^{-t^2/2}$ khi $n \to \infty$. Áp dụng Định lí 3.3.23 ta có ĐPCM. □

Một trong những ứng dụng của Định lý giá trị trung tâm nói riêng và sự hội tụ yếu nói chung là dùng để xấp xỉ cho các xác suất. Các xấp xỉ này thường được dùng để đưa ra các khoảng tin cậy hoặc trong các test kiểm định thống kê trong Lý thuyết thống kê. Để minh họa ta xét việc xấp xỉ của phân phối nhị thức qua phân phối chuẩn tắc.

Ví dụ 3.4.2. *Ta biết rằng biến ngẫu nhiên S_n với phân phối nhị thức $B(n,p)$ có thể biểu diễn thành tổng của n biến ngẫu nhiên độc lập cùng phân phối Bernoulli $Ber(p)$. Khi n đủ lớn, theo định lý giới hạn trung tâm ta có thể xấp xỉ phân phối của $(S_n - np)/\sqrt{np(1-p)}$ bởi phân phối chuẩn tắc $\mathcal{N}(0,1)$.*

Để tính xác suất $a \leq S_n \leq b$ ta có công thức chính xác như sau

$$\Sigma_{i=a}^{b} C_n^i p^i (1-p)^{n-i}.$$

Tuy nhiên khi n rất lớn thì nói chung việc tính C_n^i là không khả thi vì vượt quá khả năng lưu trữ của máy tính. Sau đó chúng ta lại phải nhân với một số rất nhỏ để ra kết quả xác *suất không quá 1. Trong thực tế, ta có thể xấp xỉ xác suất trên theo cách sau*

$$\begin{aligned}\mathbb{P}(a \le S_n \le b) =& \mathbb{P}\left(\frac{S_n - np}{\sqrt{np(1-p)}} \in \left[\frac{a-np}{\sqrt{np(1-p)}}, \frac{b-np}{\sqrt{np(1-p)}}\right]\right) \\ \cong& \mathbb{P}\left(\mathcal{N}(0,1) \in \left[\frac{a-np}{\sqrt{np(1-p)}}, \frac{b-np}{\sqrt{np(1-p)}}\right]\right).\end{aligned}$$

Chú ý rằng với xác suất tương ứng với phân phối chuẩn tắc ta có thể xấp xỉ trên máy tính một cách dễ dàng hoặc sử dụng bảng phân phối có sẵn.

Mở rộng Định lý giới hạn trung tâm cổ điển bằng cách cho phép các biến ngẫu nhiên có thể không cùng phân phối, ta có định lý Lindeberg như sau.

Định lý 3.4.3 (Định lý Lindeberg)**.** *Cho dãy biến ngẫu nhiên $(X_n)_{n\ge 1}$ độc lập với hương sai hữu hạn. Denote $S_n = X_1 + \ldots + X_n,\ B_n^2 = DX_1 + \ldots + DX_n$. Giả sử rằng*

$$L_n(\epsilon) := \frac{1}{B_n^2}\sum_{k=1}^{n} \mathbb{E}\Big((X_k - \mathbb{E}X_k)^2 \mathbb{I}_{\{|X_k - \mathbb{E}X_k| > \epsilon B_n\}}\Big) \to 0,\ \forall \epsilon > 0. \tag{3.7}$$

Khi đó $\frac{S_n - \mathbb{E}S_n}{B_n} \xrightarrow{w} N(0,1)$.

3.5 Phương pháp moment

Như ta đã trình bày ở trên, Định lý giới hạn trung tâm là một ví dụ quan trọng của sự hội tụ yếu và ta hiểu hội tụ yếu là sự hội tụ của các độ đo. Trước khi phương pháp hàm đặc trưng ra đời, một phương pháp hay dùng để kiểm tra sự hội tụ yếu đó là *phương pháp moment*. Ý tưởng chính của phương pháp moment đó là nếu ta tính được moment thứ k của biến ngẫu nhiên (phân phối) giới hạn X và có thể kiểm tra được với mọi $k \in \mathbb{N}$

$$\mathbb{E}X_n^k \to \mathbb{E}X^k,$$

thì liệu ta có thể kết luận X_n hội tụ yếu về X?

Định lý 3.5.1 (Điều kiện Carleman)**.** *Cho biến ngẫu nhiên X có phân phối xác suất μ_X với moment hữu hạn tại mọi bậc, tức là $M_k = \mathbb{E}X^k < \infty, \forall k \in \mathbb{N}$. Khi đó nếu điều kiện sau được thỏa mãn*

$$\sum_k M_{2k}^{1/2k} < \infty,$$

thì phân phối μ_X xác định duy nhất qua các moment.

Ngoài ra, nếu ta có dãy biến ngẫu nhiên có moment hữu hạn tại mọi cấp và với mọi $k \in \mathbb{N}$

$$\mathbb{E}X_n^k \to \mathbb{E}X^k,$$

thì $X_n \xrightarrow{d} X$.

Sử dụng phương pháp moment, ta có định lý Erdos-Kac như sau.

Định lý 3.5.2 (Định lý Erdos-Kac)**.** *Cho biến ngẫu nhiên* X_n *có phân phối đều trong tập* $\{1, 2, \ldots, n\}$*. Đặt* Y_n *là số các ước nguyên tố của* X_n*. Khi đó*

$$\frac{Y_n - \log\log n}{\sqrt{\log\log n}} \xrightarrow{d} \mathcal{N}(0, 1).$$

Tài liệu tham khảo

[1] M. Aigner, G. Ziegler, *Proofs from THE BOOK*, 4th edition, Springer, 2010.

[2] N. Alon, J. Spencer, *Probabilistic method*, 4th edition, Wiley series, 2016.

[3] A. Dembo, *Probability Theory: STAT310/MATH230*, Lecture notes Stanford University, 2020.

[4] Nguyễn Tiến Dũng, Đỗ Đức Thái, *Nhập môn Toán tài chính*, Nhà xuất bản Đại học Cần Thơ, 2013.

[5] P. Jung, *Probability Theory Course Notes*, KAIST, 2019.

[6] Ngô Hoàng Long, *Lý thuyết Xác suất và Thống kê*, giáo trình Đại học Sư phạm Hà Nội.

[7] J. Jacod, P. Protter, *Probability Essential*, 2nd edition, Springer, 2004.

[8] Trần Hùng Thao, *Nhập môn Toán tài chính*, Nhà xuất bản ĐHQG Hà Nội, 2009.

[9] Nguyễn Duy Tiến, Vũ Viết Yên, *Lý thuyết xác suất*, Nhà xuất bản Giáo dục, 2013.