

Phương pháp xác suất trong tổ hợp

Huỳnh Xuân Tín

Trường THPT Chuyên Lương Văn Chánh - Phú Yên

Các số Ramsey $R(k, l)$ được chỉ ra là luôn tồn tại với mọi $k, l \in \mathbb{N}$, nhưng chỉ rất ít trong các số đó là được biết giá trị chính xác. Năm 1947, P. Erdős đã đưa ra một chứng minh cho cận dưới của số Ramsey dạng đối xứng bằng một phương pháp mới lúc bấy giờ: phương pháp xác suất. Gần đây, phương pháp này đã phát triển mạnh mẽ và trở thành một công cụ hữu hiệu để giải quyết các bài toán tổ hợp. Cơ sở của phương pháp xác suất có thể được diễn tả như sau: để chứng minh sự tồn tại của một cấu trúc tổ hợp thỏa tính chất nào đó, ta xây dựng một không gian xác suất thích hợp rồi chỉ ra rằng một phần tử với tính chất đã cho được chọn ngẫu nhiên trong không gian đó có xác suất dương. Trong tài liệu này, chúng tôi cũng đề cập đến một số ứng dụng của phương pháp xác suất trong tổ hợp, đặc biệt là chứng minh bài toán tồn tại.

Nội dung chính là xem xét một số ứng dụng của phương pháp xác suất trong các bài toán tổ hợp và đồ thị theo hai hướng cơ bản: dựa vào định nghĩa xác suất và dựa vào tính chất của kỳ vọng. Ngoài các bài toán về tổ hợp và đồ thị, tác giả cũng đã đưa thêm các bài toán mà có thể ứng dụng phương pháp này trong các lĩnh vực khác.

1 Định nghĩa xác suất với bài toán tổ hợp, đồ thị

Trước tiên ta cần mở rộng khái niệm đồ thị.

Một *siêu đồ thị* là một cặp $H = (V, E)$, ở đây V là tập hữu hạn các phần tử được gọi là các *đỉnh* và E là họ các tập con của V gọi là các *cạnh*. H được gọi là *n -siêu đồ thị đều* nếu mỗi cạnh của nó chứa đúng n đỉnh. Ta nói rằng H thỏa tính chất B hoặc 2-tô màu được nếu có một cách 2-tô màu cho các đỉnh trong V sao cho không có cạnh nào cùng màu. Ký hiệu $m(n)$ là số cạnh nhỏ nhất của một n -siêu đồ thị đều không có tính chất B . Ta đi xác định cận dưới cho $m(n)$.

Bài toán 1. *Mỗi n -siêu đồ thị đều với ít hơn 2^{n-1} cạnh có tính chất B . Do đó $m(n) \geq 2^{n-1}$.*

Giải. Đặt $H = (V, E)$ là một n -siêu đồ thị đều với ít hơn 2^{n-1} cạnh. Tô màu ngẫu nhiên cho V bằng 2 màu (mỗi màu có xác suất được chọn là $\frac{1}{2}$). Với mỗi cạnh $e \in E$, đặt A_e là biến cố chỉ e có cùng màu. Khi đó

$$\mathbb{P}[A_e] = \frac{2}{2^n} = 2^{1-n}.$$

Do đó, xác suất để có ít nhất một cạnh trong E cùng màu là

$$\mathbb{P}\left[\bigcup_{e \in E} A_e\right] \leq \sum_{e \in E} \mathbb{P}[A_e] < 1,$$

tức $1 - \mathbb{P}\left[\bigcup_{e \in E} A_e\right] > 0$. Điều này nghĩa là tồn tại một cách 2-tô màu cho V sao cho không có cạnh cùng màu.

Bài toán 2. Cho $G = (V, E)$ là đồ thị hai mảng n đỉnh với một tập $S(v)$ chứa nhiều hơn $\log_2 n$ màu gắn với mỗi đỉnh $v \in V$. Chứng minh rằng có một cách tô màu thích hợp cho G mà mỗi đỉnh v được tô một màu từ tập màu $S(v)$ của nó.

Giải. Do G là đồ thị hai mảng nên tập V có thể phân hoạch thành hai tập rời nhau V_1 và V_2 sao cho mỗi cạnh trong G có một đỉnh trong V_1 và một đỉnh trong V_2 , đặt $S = \bigcup_{v \in V} S(v)$ là tập tất cả các màu có thể. Xét phân hoạch ngẫu

nhien $S = S_1 \cup S_2$, trong đó mỗi màu được chọn ngẫu nhiên, độc lập cho vào S_1 hoặc S_2 với xác suất bằng nhau (và bằng $\frac{1}{2}$). Ta sẽ chứng minh tồn tại một phân hoạch của S sao cho tất cả các đỉnh trong V_i , $i = 1, 2$, có thể được tô màu bằng các màu trong S_i , $i = 1, 2$.

Lấy $v \in V_i$, $i = 1, 2$, khi đó xác suất để không có màu nào trong tập màu $S(v)$ nằm trong S_i xác định bởi:

$$\mathbb{P}[S(v) \cap S_i = \emptyset] = \left(\frac{1}{2}\right)^{|S(v)|} < \frac{1}{n}, \text{ (do } |S(v)| > \log_2 n \text{)}.$$

Vậy

$$\mathbb{P}\left[\bigcup_{v \in V_i} \{S(v) \cap S_i = \emptyset\}\right] < \frac{|V_i|}{n},$$

do đó, xác suất để có ít nhất một đỉnh không thể được tô bằng một màu bất kì trong tập màu của đỉnh đó sẽ bị chặn bởi $\frac{|V_1|}{n} + \frac{|V_2|}{n} = 1$.

Vậy có một phân hoạch $S = S_1 \cup S_2$ sao cho tất cả các đỉnh trong V_i có thể được tô bằng các màu trong S_i , $i = 1, 2$.

Bài toán 3. Cho m, n là hai số nguyên dương lớn hơn 1, và $m > 2.013 \log_2 n$. Khi đó, ta có thể tô màu mỗi cạnh của $K_{n,n}$ là đỏ hoặc xanh sao cho không có đồ thị con $K_{m,m}$ có cùng màu cạnh được tạo thành.

Giải. Đồ thị $K_{m,m}$ có $2m$ đỉnh và m^2 cạnh, do đó số cách để 2-tô màu cạnh cho đồ thị con $K_{m,m}$ của đồ thị $K_{n,n}$ là 2^{m^2} , và trong các cách tô màu đó chỉ có

2 kết quả thuận lợi để được $K_{m,m}$ cùng màu. Suy ra, xác suất để được $K_{m,m}$ cùng màu cạnh là

$$\frac{2}{2^{m^2}} = 2^{1-m^2}.$$

Đồ thị $K_{n,n}$ có $(C_n^m)^2$ đồ thị con $K_{m,m}$, mỗi đồ thị con $K_{m,m}$ có khả năng cùng màu cạnh như nhau. Do đó, xác suất để có ít nhất một đồ thị con $K_{m,m}$ cùng

màu luôn nhỏ hơn hoặc bằng $(C_n^m)^2 \cdot 2^{1-m^2}$.

Vậy để chứng minh yêu cầu của bài toán, ta chỉ cần chứng minh

$$(C_n^m)^2 \cdot 2^{1-m^2} < 1.$$

+ Vì $m > 2.013 \log_2 n > 2$, nên ta có

$$2(C_n^m)^2 = 2 \left[\frac{(n-m+1)(n-m+2) \cdots (n-1)n}{m!} \right]^2 < n^{2m}.$$

+ Vì $m > 2.013 \log_2 n > 2 \log_2 n$, nên suy ra

$$n^{2m} < (2^{\frac{m}{2}})^{2m}.$$

Từ 2 điều trên, suy ra

$$2(C_n^m)^2 < n^{2m} < (2^{\frac{m}{2}})^{2m} = 2^{m^2}.$$

Bản chất của số 2.013 trong điều kiện $m > 2.013 \log_2 n$ đó là nó lớn hơn 2. Bởi vậy, bất kì số $2 + \varepsilon$, với $\varepsilon > 0$ nào đó đều có thể được.

Bài toán 4. (Định lý Erdős - Ko - Rado, [10]) Nếu $|X| = n$, $n \geq 2k$ và $\mathcal{F}$ là họ giao nhau các k -tập con của X , tức là $\forall A, B \in \mathcal{F}, A \cap B \neq \emptyset$, thì ta có $|\mathcal{F}| \leq C_{n-1}^{k-1}$.

Giải. Ta cần bổ đề sau:

Bổ đề 1. Xét $X = \{0, 1, \dots, n-1\}$, và với $0 \leq s < n$, ta định nghĩa $A_s = \{s, s+1, \dots, s+k-1\} \subseteq X$ với phép cộng modulo n . Khi đó, với $n \geq 2k$, thì bất kì họ giao nhau $\mathcal{F}$ các k -tập con của X đều chứa nhiều nhất k tập A_s .

Chứng minh. Nếu $A_i \in \mathcal{F}$, thì bất kì tập $A_s \in \mathcal{F}$ nào đó khác A_i phải là 1 trong số các tập $A_{i-k+1}, \dots, A_{i-1}$ hoặc $A_{i+1}, \dots, A_{i+k-1}$. Có $2k-2$ tập như thế, các tập này có thể được chia thành $(k-1)$ cặp có dạng (A_s, A_{s+k}) . Vì $n \geq 2k$, $A_s \cap A_{s+k} = \emptyset$ và chỉ có một tập trong mỗi cặp là có thể xuất hiện trong $\mathcal{F}$, nên ta có điều phải chứng minh. $\square$

Giả sử $X = \{0, 1, \dots, n-1\}$ và $\mathcal{F}$ là họ giao nhau các k -tập con của X . Với một hoán vị $\sigma : X \rightarrow X$, ta định nghĩa

$$\sigma(A_s) = \{\sigma(s), \sigma(s+1), \dots, \sigma(s+k-1)\},$$

với phép cộng modulo n . Các tập $\sigma(A_s)$ chính là các tập con trong bộ đề 1 với các phần tử được gán nhãn lại bởi hoán vị σ , do đó, theo bộ đề trên thì có nhiều nhất k trong số n tập này nằm trong $\mathcal{F}$. Do đó, nếu chọn s độc lập, ngẫu nhiên và đều, thì

$$\mathbb{P}[\sigma(A_s) \in \mathcal{F}] \leq \frac{k}{n}.$$

Nhưng việc chọn $\sigma(A_s)$ này tương đương với việc chọn ngẫu nhiên một k -tập con của X . Bởi vậy

$$\mathbb{P}[\sigma(A_s) \in \mathcal{F}] = \frac{|\mathcal{F}|}{C_n^k},$$

và

$$|\mathcal{F}| = C_n^k \cdot \mathbb{P}[\sigma(A_s) \in \mathcal{F}] \leq C_n^k \cdot \frac{k}{n} = C_{n-1}^{k-1}.$$

Bài toán 5. Cho $A_1, A_2, \dots, A_n$ và $B_1, B_2, \dots, B_n$ là các tập con phân biệt của $\mathbb{N}$ sao cho

- $|A_i| = k$ và $|B_i| = l, \forall 1 \leq i \leq n$,
- với mỗi $i, A_i \cap B_i = \emptyset$, và
- với mỗi $i \neq j, A_i \cap B_j \neq \emptyset$.

Chứng minh $n \leq C_{k+l}^k$.

Giải. Đặt $X = \bigcup_{i=1}^n (A_i \cup B_i)$ và xét một cách sắp thứ tự các thành phần trong X một cách ngẫu nhiên (có tất cả $|X|!$ cách sắp thứ tự có xác suất như nhau). Đặt U_i là biến cố chỉ mỗi phần tử của A_i đứng trước mỗi phần tử của B_i . Ta có

$$\mathbb{P}[U_i] = \frac{C_{|X|}^{k+l} \cdot k! \cdot l! \cdot (|X| - k - l)!}{|X|!} = \frac{1}{C_{k+l}^k}, \quad 1 \leq i \leq k.$$

Ta cần chú ý rằng U_i và U_j không xảy ra đồng thời với $i \neq j$. Thật vậy, giả sử U_i và U_j xảy ra đồng thời. Không mất tính tổng quát ta giả sử phần tử cuối cùng của A_i không đứng sau phần tử cuối cùng của A_j . Nhưng trong trường hợp này, tất cả các phần tử của A_i đều đứng trước tất cả các phần tử của B_j . Điều này mâu thuẫn với giả thiết $A_i \cap B_j \neq \emptyset$.

Vậy

$$1 \geq \mathbb{P}\left[\bigcup_{i=1}^n U_i\right] = \sum_{i=1}^n \mathbb{P}[U_i] = \frac{n}{C_{k+l}^k}.$$

Bài toán 6. Cho $A_1, A_2, \dots, A_n$ và $B_1, B_2, \dots, B_n$ là các tập con phân biệt của $\mathbb{N}$ sao cho

- $|A_i| = r$ và $|B_i| = s, \forall 1 \leq i \leq n$,
- với mỗi $i, A_i \cap B_i = \emptyset$, và

- với mỗi $i \neq j$, $(A_i \cap B_j) \cup (A_j \cap B_i) \neq \emptyset$.

Chứng minh $n \leq \frac{(r+s)^{r+s}}{r^r \cdot s^s}$.

Bài toán 7. Chứng minh rằng giữa 2^{100} người, không nhất thiết phải có 200 người đôi một quen nhau hoặc 200 người đôi một không quen nhau.

Giải. Ta sẽ cho một cặp hai người bất kì quen nhau hoặc đôi một không quen nhau bằng cách tung một đồng xu đối xứng. Trong một nhóm gồm 200 người, xác suất để họ đôi một quen nhau hoặc đôi một không quen nhau là: $2 \cdot 2^{-C_{200}^2} = 2^{-19899}$.

Vì có $C_{2^{100}}^{200}$ cách chọn ra 200 người, xác suất tồn tại 200 người đôi một quen nhau hoặc đôi một không quen nhau nhiều nhất bằng:

$$C_{2^{100}}^{200} \cdot 2^{-19899} < \frac{(2^{100})^{200}}{200!} \cdot 2^{-19899} = \frac{2^{101}}{200!} < 1.$$

Từ đây suy ra xác suất tồn tại 200 người đôi một không quen nhau hoặc đôi một quen nhau không lớn hơn 0. Nói cách khác, không nhất thiết phải có 200 người đôi một quen nhau hoặc đôi một không quen nhau. Bài toán được chứng minh. $\square$

Ta thấy ở đây phương pháp tổng quát để xây dựng ví dụ ngẫu nhiên: Nếu xác suất tồn tại ví dụ ta cần là dương thì tồn tại ví dụ đó.

Bài toán 8. Trong mỗi ô của bảng 100×100 , ta viết một trong các số nguyên $1, 2, \dots, 5000$. Hơn nữa, mỗi một số nguyên trong bảng xuất hiện đúng hai lần. Chứng minh ta có thể chọn được 100 ô của bảng thỏa mãn ba điều kiện sau:

- (1) Mỗi một hàng được chọn đúng một ô.
- (2) Mỗi một cột được chọn đúng một ô.
- (3) Các số trong các ô được chọn đôi một khác nhau.

Giải. Chọn hoán vị ngẫu nhiên $(a_1, \dots, a_{100})$ của $\{1, \dots, 100\}$ và chọn ô thứ a_i trong hàng thứ i . Cách chọn như vậy thỏa mãn (1) và (2). Với mỗi $j = 1, 2, \dots, 5000$, xác suất để chọn hai ô có cùng số j là 0 nếu hai ô này cùng hàng hoặc cùng cột và là $\frac{1}{100} \cdot \frac{1}{99}$ trong trường hợp ngược lại. Do đó xác suất để cách chọn này thỏa (3) ít nhất là

$$1 - 5000 \cdot \frac{1}{100 \cdot 99} > 0$$

và ta có điều phải chứng minh.

Tiếp theo ta dùng tính chất của xác suất để giải toán.

Bài toán 9. Cho p, q là các số thực dương sao cho $p + q = 1$. Chứng minh rằng: $p + pq + pq^2 + \dots = 1$

Giải. Xét thí nghiệm tung đồng xu với xác suất ra mặt ngửa là p và mặt sấp là q . Ta thực hiện cho đến khi ra được mặt ngửa. Gọi X là số lần tung, khi đó: $P(X = n) = pq^{n-1}$.

Vế trái của đẳng thức bằng $P(X = 1) + P(X = 2) + \dots + P(X = n) + \dots$ và dĩ nhiên bằng 1. $\square$

Bài toán 10. [IMO Shortlist 2006] Cho S là tập hữu hạn các điểm trên mặt phẳng sao cho không có ba điểm nào thẳng hàng. Với mỗi đa giác lồi P với các điểm thuộc S , gọi $a(P)$ là số các điểm của P và $b(P)$ là số các điểm của S nằm ngoài P . Chứng minh rằng với mọi số thực x , ta có đẳng thức:

$$\sum_P x^{a(P)} (1 - x)^{b(P)} = 1$$

trong đó tổng được tính theo tất cả các đa giác lồi có đỉnh thuộc S . (Chú ý: đoạn thẳng, một điểm và tập rỗng được coi là đa giác lồi với 2, 1, 0 đỉnh tương ứng)

Giải. Ta tô màu một cách ngẫu nhiên các điểm bằng màu đen và màu trắng, trong đó các điểm được tô màu đen với xác suất x . Với mỗi đa giác lồi P , gọi EP là biến cố tất cả các đỉnh nằm trên chu vi của P có màu đen và tất cả các đỉnh nằm ngoài P có màu trắng.

Các biến cố này đôi một xung khắc nhau, như thế vế trái là xác suất của sự kiện chắc chắn xảy ra: ta chỉ cần xét bao lồi của tất cả các điểm màu đen. $\square$

Để tính xác suất của một biến cố theo định nghĩa cổ điển ta thường phải giải quyết hai bài toán tổ hợp: tính số kết quả thuận lợi và tính số các kết quả có thể. Thông thường bài toán sau đơn giản hơn bài toán trước. Và chính điều này tạo ra một ứng dụng thú vị của xác suất: Nếu ta tính được số các kết quả có thể và xác suất thì sẽ tính được số kết quả thuận lợi.

Bài toán 11. Trong số cách chọn ra ba đỉnh của hình lập phương đơn vị, có bao nhiêu cách chọn thỏa mãn điều kiện ba đỉnh được chọn là ba đỉnh của một tam giác đều.

2 Phép chứng minh tồn tại sử dụng kỳ vọng

Một điều hiển nhiên rằng giá trị trung bình của một tập các số không bao giờ vượt quá số lớn nhất trong tập này. Điều này cũng đúng cho kỳ vọng.

Định lý 1. ([4]) Cho $X : \Omega \rightarrow \mathbb{R}$ là một biến ngẫu nhiên sao cho tập hợp $S = \{X(u)/u \in \Omega\}$ là hữu hạn, và đặt j là phần tử lớn nhất của S . Khi đó ta có $j \geq \mathbb{E}[X]$.

Chứng minh. Theo định nghĩa của $\mathbb{E}[X]$, ta có

$$\mathbb{E}[X] = \sum_{i \in S} i \cdot \mathbb{P}[X = i] \leq j \cdot \sum_{i \in S} \mathbb{P}[X = i] = j.$$

□

Sau đây sẽ là một số ứng dụng của định lý 1 trong việc giải quyết các bài toán tồn tại.

Bài toán 12. (Szele 1943, [4]) *Chứng minh rằng có một đồ thị đầy đủ có hướng n đỉnh chứa ít nhất $\frac{n!}{2^{n-1}}$ đường Hamilton. Kết luận gì về số vòng Hamilton?*

Giải. Lấy một đồ thị đầy đủ K_n và định hướng mỗi cạnh của nó một cách ngẫu nhiên để được một đồ thị đầy đủ có hướng T . Nếu p là một xích Hamilton trong K_n , thì đặt $X_p(T) = 1$ nếu p trở thành một đường Hamilton trong T và đặt $X_p(T) = 0$ nếu ngược lại. Vì p có $n - 1$ cạnh, nên

$$\mathbb{E}[X_p] = \frac{1}{2^{n-1}}.$$

Đặt $X = \sum_p X_p$, p chạy khắp $n!$ xích Hamilton trong K_n , thì X chính là số đường Hamilton trong T . Theo tính chất của kỳ vọng ta có

$$\mathbb{E}[X] = n! \cdot \mathbb{E}[X_p] = \frac{n!}{2^{n-1}},$$

và áp dụng định lý 1 ta có điều phải chứng minh.

Với vòng Hamilton thì chỉ khác với đường Hamilton là chúng có thêm 1 cạnh có hướng. Do đó, tồn tại một đồ thị đầy đủ có hướng n đỉnh với ít nhất $\frac{n!}{2^n}$ vòng Hamilton.

Bài toán 13. *Cho G là một đồ thị đơn với tập đỉnh $[n]$, và có m cạnh. Khi đó G chứa một đồ thị con 2 mảng với ít nhất $\frac{m}{2}$ cạnh.*

Giải. Đặt $G = (V, E)$, và chọn ngẫu nhiên một tập con $T \subseteq V$ (ở đây, các biến cố $x \in T$ là độc lập lẫn nhau với xác suất $\frac{1}{2}$). Với một cạnh e cho trước, gọi X_e là biến chỉ của biến cố có đúng một đỉnh của cạnh e nằm trong T . Khi đó

$$\mathbb{E}[X_e] = \frac{1}{2}.$$

Nếu ký hiệu X là số cạnh có đúng một đỉnh nằm trong T , thì

$$\mathbb{E}[X] = \sum_{e \in E} \mathbb{E}[X_e] = \frac{m}{2}.$$

Vậy với bất kì $T \subseteq V$, tồn tại ít nhất $\frac{m}{2}$ cạnh có một đỉnh trong T và một đỉnh trong $V \setminus T$, tạo thành một đồ thị hai mảng.

Tiếp theo là một bài toán được liên hệ tới một vấn đề nổi tiếng trong lý thuyết phức tạp, có thể gọi là "*vấn đề ở giữa*".

Bài toán 14. (*Iran Team Selection Test 2008/6*) Giả sử 799 đối thủ tham gia vào một giải đấu, trong đó mỗi cặp đối thủ thi đấu với nhau đúng một lần. Chứng minh rằng tồn tại 2 tập rời nhau A và B gồm 7 đối thủ sao cho mỗi đối thủ trong A đều thắng mỗi đối thủ trong B .

Giải. Giải đấu này có thể diễn tả bằng một đồ thị đầy đủ có hướng G có tập đỉnh E gồm 799 điểm (trong mặt phẳng hoặc trong không gian) tương ứng với 799 đối thủ, hai đỉnh x, y bất kì được nối với nhau bằng một cung từ x đến y nếu đối thủ x thắng đối thủ y .

Gọi A là 7-tập con của E được chọn ngẫu nhiên. Đặt X là số đỉnh có cung đi vào từ các đỉnh trong A . Ta chứng minh $\mathbb{E}[X] \geq 7$.
Gọi X_i là biến chỉ của biến cố đỉnh i có cung đi vào từ các đỉnh trong A , suy ra

$$\mathbb{E}[X_i] = \frac{C_{d_i^-}^7}{C_{799}^7}.$$

Theo tính chất tuyến tính của kỳ vọng, ta có

$$\mathbb{E}[X] = \sum_{i=1}^{799} \mathbb{E}[X_i] = \sum_i \frac{C_{d_i^-}^7}{C_{799}^7}.$$

Mà

$$\sum_i d_i^- = C_{799}^2 = \frac{798 \cdot 799}{2} = 399 \cdot 799,$$

điều này nghĩa là bậc vào trung bình của một đỉnh i bất kì đúng bằng 399, và áp dụng tính lồi của hàm $f(x) = C_x^k$ trên $[k, +\infty)$, ta được:

$$\mathbb{E}[X] = \sum_i \frac{C_{d_i^-}^7}{C_{799}^7} \geq 799 \cdot \frac{C_{399}^7}{C_{799}^7} \approx 6.025.$$

Vì X là một số nguyên nên ta có điều phải chứng minh

Bài toán 15. (*Russia 1996/4*) Tại Duma có 1600 đại biểu, thành lập 16000 ủy ban. Mỗi ủy ban gồm 80 đại biểu. Chứng minh rằng ta có thể tìm được 2 ủy ban có ít nhất 4 thành viên chung.

Giải. Chọn một cặp ủy ban ngẫu nhiên trong số C_{16000}^2 cặp. Đặt X là số người thuộc vào cả 2 ủy ban được chọn. Chú ý rằng

$$X = X_1 + X_2 + \cdots + X_{1600},$$

ở đây X_i là biến chỉ của biến cố đại biểu thứ i thuộc vào cả 2 ủy ban được chọn.

Theo tính chất tuyến tính của kỳ vọng, ta có

$$\mathbb{E}[X] = \mathbb{E}[X_1] + \mathbb{E}[X_2] + \cdots + \mathbb{E}[X_{1600}].$$

Để tính $\mathbb{E}[X_i]$, đặt n_i là số ủy ban có người thứ i tham gia, khi đó

$$\mathbb{E}[X_i] = \mathbb{P}[\text{Đại biểu thứ } i \text{ thuộc vào 2 ủy ban được chọn}] = \frac{C_{n_i}^2}{C_{16000}^2}.$$

Mà $\sum_i n_i = 16000 \cdot 80$, do đó giá trị trung bình của các n_i là

$$\bar{n} = \frac{16000 \cdot 80}{1600} = 800,$$

và theo tính lồi của hàm $f(x) = C_x^k$ trên $[k, +\infty)$, suy ra

$$\mathbb{E}[X] \geq 1600 \cdot \frac{C_{\bar{n}}^2}{C_{16000}^2} = 1600 \cdot \frac{800 \cdot 799}{16000 \cdot 15999} \approx 3.995.$$

Do X luôn là số nguyên, nên ta có điều phải chứng minh

3 Một số bài tập tương tự

Bài toán 16. (*MOP Test 2007/7/1*) Trong một ma trận $n \times n$, mỗi số $1, 2, \dots, n$ xuất hiện đúng n lần. Chứng minh rằng có một dòng hoặc cột chứa ít nhất $\sqrt{n}$ số phân biệt.

Bài toán 17. (*IMO Shortlist 1999/C4*) Đặt A là tập gồm n thặng dư mod n^2 . Chỉ ra rằng có một tập B gồm n thặng dư mod n^2 sao cho ít nhất một nửa thặng dư mod n^2 có thể viết dưới dạng $a + b$ với $a \in A$ và $b \in B$.

Bài toán 18. (*Taiwan 1997/9*) Với $n \geq k \geq 3$, đặt $X = \{1, 2, \dots, n\}$ và $\mathcal{F}_k$ là họ các k -tập con của X sao cho bất kỳ hai tập con trong $\mathcal{F}_k$ đều có nhiều nhất $k - 2$ phần tử chung. Chứng minh rằng tồn tại một tập con M_k của X với ít nhất $\lfloor \log_2 n \rfloor + 1$ phần tử mà không chứa tập con nào trong $\mathcal{F}_k$.

Bài toán 19. (Định lý sperner, [9]) Chứng minh rằng: Nếu $\mathcal{F}$ là họ các tập con của $[n]$ sao cho không tồn tại 2 tập $A, B \in \mathcal{F}$ thỏa mãn $A \subset B$, thì $|\mathcal{F}| \leq C_n^{\lfloor n/2 \rfloor}$.

Bài toán 20. Cho đồ thị $G = (V, E)$ có n đỉnh và $\frac{nd}{2}$ cạnh, $d \geq 1$. Chứng minh rằng có một tập con U gồm các đỉnh đôi một không kề nhau có lực lượng $\geq \frac{n}{2d}$.

Bài toán 21. Cho đồ thị vô hướng $G = (V, E)$. Một tập $U \subseteq V$ được gọi là tập bao quát trong G nếu mỗi đỉnh $v \in V \setminus U$ có ít nhất một đỉnh kề của nó nằm trong U . Chứng minh rằng: Nếu G có n đỉnh, với bậc nhỏ nhất $\delta > 1$, thì G có một tập bao quát với nhiều nhất $n \cdot \frac{1 + \ln(\delta + 1)}{\delta + 1}$ đỉnh.

Bài toán 22. Cho $L = (L_1, L_2, \dots, L_k)$ gồm các bộ ba thứ tự $L_i = (a_i, b_i, c_i)$ sao cho với i bất kỳ, các số a_i, b_i, c_i là các phần tử khác nhau của $[n]$. Tuy nhiên, các ký hiệu đó với các chỉ số i và j khác nhau có thể ký hiệu cho cùng một số. Đặt $p = p_1 p_2 \dots p_n$ là một n -hoán vị. Ta nói rằng p thỏa mãn L_i nếu phần tử b_i ở

giữa a_i và c_i trong p (không quan tâm đến thứ tự của 3 phần tử này trong p là $a_i b_i c_i$ hay $c_i b_i a_i$).

Chứng minh rằng tồn tại một n -hoán vị p thỏa mãn ít nhất $\frac{1}{3}$ của tất cả L_i trong một L cho trước.

Bài toán 23. Cho đồ thị $G = (V, E)$. Một tập $U \subseteq V$ được gọi là độc lập trong G nếu không tồn tại cạnh trong U . Chứng minh: nếu G có n đỉnh và ký hiệu d_v là bậc của đỉnh v , $v \in V$, thì G có một tập độc lập có lực lượng ít nhất $\sum_v \frac{1}{d_v+1}$. Hơn nữa, nếu G có e cạnh thì G có một tập độc lập có lực lượng ít nhất $\frac{n^2}{2e+n}$.

Bài toán 24. Cho G là một đồ thị đơn. Nếu G có $2n$ đỉnh và e cạnh thì nó chứa một đồ thị con hai mảng với ít nhất $\frac{en}{2n-1}$ cạnh. Nếu G có $2n+1$ đỉnh và e cạnh thì nó chứa một đồ thị con hai mảng với ít nhất $\frac{e(n+1)}{2n+1}$ cạnh.

Bài toán 25. Giả sử $n \geq 4$ và H là một n -siêu đồ thị đều với ít nhất $\frac{4^{n-1}}{3^n}$ cạnh. Chứng minh có một cách tô màu cho các đỉnh của H bằng 4 màu sao cho mỗi cạnh đều có 4 màu.

Bài toán 26. (Austrian-Polish Competition 1997/8) Cho $|X| = n$. Tìm số lớn nhất các tập con khác nhau của X sao cho mỗi tập con này có 3 phần tử và không có 2 tập con nào rời nhau.

Bài toán 27. Đặt $p = p_1 p_2 \cdots p_n$ là một n -hoán vị. Chỉ số i được gọi là chỉ số vượt của p nếu $p_i > i$. Trung bình một n -hoán vị có bao nhiêu chỉ số vượt?

Bài toán 28. Cho G là một đồ thị có $n \geq 10$ đỉnh và giả sử rằng: nếu ta thêm vào G bất kỳ một cạnh nào đó không nằm trong G thì số đồ thị K_{10} của G tăng lên. Chứng minh rằng số cạnh trong G ít nhất là $8n - 36$.

Bài toán 29. Một giải đấu T với n đối thủ chính là một sự định hướng cho các cạnh của K_n . Ta nói rằng T thỏa mãn tính chất S_k nếu với bất kỳ k -tập của n đối thủ, thì luôn có một đối thủ thắng tất cả các đối thủ còn lại. Chứng minh rằng: nếu $C_n^k \cdot (1 - 2^{-k})^{n-k} < 1$, thì có một giải đấu trên K_n với tính chất S_k .

4 Một số bài toán Đại số, Số học

Bài toán 30. [Bulgaria MO 1984] Cho $x_i, y_i (i = 1, 2, \dots, n)$ là $2n$ số thực dương sao cho $x_i + y_i = 1$. Chứng minh rằng với mọi số nguyên dương m , ta đều có: $(1 - x_1 x_2 \dots x_n)^m + (1 - y_1^m)(1 - y_2^m) \dots (1 - y_n^m) \geq 1$

Giải. Xét thí nghiệm xác suất: Cho $c_1, c_2, \dots, c_n$ là các đồng xu sao cho với mỗi i , xác suất để c_i ra mặt ngửa là x_i . Ta tung các xu này một cách độc lập m lần. Khi đó $(1 - x_1 x_2 \dots x_n)^m$ là xác suất $P(A)$ của biến cố "với mỗi một trong m lần tung, có ít nhất một đồng xu ra mặt ngửa".

Chú ý rằng $A = B \cup C$, trong đó B là biến cố "tồn tại một đồng xu ra mặt ngửa ở mỗi một trong m lần tung" và C là biến cố "có ít nhất một đồng xu

ra mặt ngửa ở mỗi lần tung, nhưng đồng xu này không giống nhau qua mỗi lần tung". Hơn nữa, $B \cap C = \emptyset$, do đó: $P(A) = P(B) + P(C)$

Mặt khác, $(1 - y_1^m)(1 - y_2^m) \dots (1 - y_n^m)$ là xác suất của biến cố mỗi một đồng xu không ra mặt ngửa trong m lần tung bằng $P(\overline{B})$, trong đó $\overline{B}$ là biến cố đối của biến cố B

Như vậy về trái của bất đẳng thức đã cho là:

$$P(A) + P(\overline{B}) = P(B) + P(\overline{B}) + P(C) = 1 + P(C) \geq 1$$

Chú ý đẳng thức xảy ra khi và chỉ khi $n = 1$. □

Bài toán 31. [MOP Test 2008] Giả sử a, b, c là các số thực dương sao cho với mọi n nguyên dương thì: $[an] + [bn] = [cn]$. Chứng minh rằng ít nhất một trong ba số a, b, c nguyên.

Giải. Giả sử không có số nào trong ba số a, b, c là số nguyên. Chia hai vế cho n và cho n dần đến vô cùng ta được $a + b = c$.

$$\text{Từ đó suy ra } \{an\} + \{bn\} = \{cn\} \quad (*)$$

Nếu x vô tỷ thì $\{xn\}$ phân bố đều trên đoạn $[0; 1]$. Nói riêng nếu ta chọn n một cách ngẫu nhiên trong $\{1, 2, \dots, N\}$ thì $E(\{xn\}) \rightarrow \frac{1}{2}$ khi $N \rightarrow \infty$. Mặt khác, nếu x là số hữu tỷ có dạng tối giản $\frac{p}{q}$ thì $\{xn\}$ có kỳ vọng tiến đến $\frac{q-1}{2q} = \frac{1}{2} - \frac{1}{2q}$. Như vậy nó nằm trong khoảng $[\frac{1}{4}; \frac{1}{2}]$. Tóm lại, với mọi số không nguyên x , ta có $E(\{xn\}) \rightarrow t$ trong đó $t \in [\frac{1}{4}; \frac{1}{2}]$. Lấy kỳ vọng hai vế của (*) và n dần đến vô cùng, ta thấy rằng cách duy nhất để có đẳng thức là $E(\{an\})$ và $E(\{bn\})$ phải tiến đến $\frac{1}{4}$.

Nhưng cách duy nhất để có kỳ vọng $\frac{1}{4}$ là khi a, b hữu tỷ, còn cách duy nhất để có kỳ vọng $\frac{1}{2}$ là c vô tỷ. Nhưng do $a + b = c$ nên ta không thể hai số hữu tỷ cộng ra số vô tỷ. Mâu thuẫn. □

Bài toán 32. Cho p, q là các số không âm có tổng bằng 1. Chứng minh rằng: $(1 - p^m)^n + (1 - q^n)^m \geq 1$

Giải. Xét một ma trận ngẫu nhiên kích thước $m \times n$ trong đó mỗi một ô được điền số 1 với xác suất p và số 0 với xác suất là q .

Xác suất để một dòng có ít nhất một số 1 bằng $1 - q^n$. Khi đó xác suất để ma trận trên có mỗi dòng ít nhất một số 1 là $(1 - q^n)^m$.

Xác suất để một cột có ít nhất một số 0 bằng $1 - p^m$. Khi đó xác suất để ma trận trên có mỗi cột ít nhất một số 0 là $(1 - p^m)^n$.

Theo cách điền số vào ma trận như trên thì ta có điều phải chứng minh.

Bài toán 33. [Bay Area MO 2004] Cho n số thực không đồng thời bằng 0 có tổng bằng 0. Chứng minh rằng tồn tại một cách đánh số các số này là $a_1, a_2, \dots, a_n$ sao cho: $a_1a_2 + a_2a_3 + \dots + a_{n-1}a_n + a_na_1 < 0$

Hướng dẫn Giả sử $a_1, a_2, \dots, a_n$ là hoán vị ngẫu nhiên của các số ban đầu. Giá trị kỳ vọng của biểu thức vế trái bằng bao nhiêu?

Tài liệu

- [1] Trần Nam Dũng, Võ Quốc Bá Cẩn, Nguyễn Phúc Lữ (2012), *Các phương pháp giải toán qua các kì thi olympic*.
- [2] Đặng Huy Ruận (2004), *Lý thuyết đồ thị và ứng dụng*, NXB Khoa học và kỹ thuật, Hà Nội.
- [3] N. Alon, J. Spencer (2000), *The Probabilistic Method*, A Wiley-Interscience Publication. John Wiley and Sons, Inc.
- [4] M. Bóna (2002), *A Walk Through Combinatorics An Introduction to Enumeration and Graph Theory*, World Scientific Publishing Co. , Inc.
- [5] M. Bóna, G. Tóth (1996), "A Ramsey-type problem on right - angles in space", *Discrete Math* (**150**), 1-3, 61-67.
- [6] M. Bóna (1993), "A Euclidean Ramsey theorem", *Discrete math* (**122**), 349-352.
- [7] A. Engel (1998), *Problem - Solving Strategies*, Springer.
- [8] P. Erdős, R. L. Graham, P. Montgomery, B. L. Rothschild, J. Spencer and E. G. Straus (1973), "Euclidean Ramsey theorems I", *Journal of combinatorial Theory, Series A* (**14**), 341-363.
- [9] <http://www.princeton.edu/~ploh/docs/math/mop2008/prob-comb-soln.pdf>, 10/12/2008.
- [10] <http://kam.mff.cuni.cz/~matousek/prob-ln.ps.gz>, 12/10/2008.
- [11] <http://www.math.ust.hk/~mabfchen/Math391I/pigeonhole.pdf>, 6/11/2008.
- [12] Jole Spencer, *sample chapters of the probabilistic method*, Internet.
- [13] Po-Shen Loh, *probabilistic method in combinatorics*, Internet.