Mục lục

Lời mở đầu

Xác suất thống kê là môn học có nhiều ứng dụng trong thực tế. Lịch sử đã chứng minh rằng nhờ vào trực quan mà các nhà toán học,vật lí,... đã tìm ra rất nhiều các lí thuyết, định lí, định luật. Tuy nhiên trực quan không phải lúc nào cũng đúng. Để giúp các bạn biết thêm về những ứng dụng của xác suất thống kê , đồng thời tránh được những quan điểm sai lầm do trực quan mang lại, tôi đã hoàn thành báo cáo này. Báo cáo trình bày những lí thuyết chung nhất về "đường dẫn" và các định nghĩa,định lí, định luật liên quan đến đường dẫn. Mặc dù đã hết sức cố gắng ,song báo cáo không thể tránh khỏi những thiếu sót. Vì vậy rất mong nhận được những ý kiến phản hồi cũng như đóng góp để báo cáo hoàn thiện hơn.

Hà Nội, tháng 6 năm 2017

Danh mục các ký hiệu

1. Card : Lực lượng (Bản số) ;

2. P(A) : Xác suất của biến cố A;

3. $\begin{pmatrix} n \\ x \end{pmatrix} = \dfrac{n\,(n-1)\ldots(n-x+1)}{x!}$;

BÁO CÁO HỌC THUẬT CẤP BỘ MÔN

Tên đề tài

ỨNG DỤNG CỦA DU ĐỘNG NGẪU NHIÊN

Chủ nhiệm đề tài: Ths. Nguyễn Thế Lâm

Chương 1.

Du động ngẫu nhiên đối xứng

1. Đường dẫn và nguyên lí phản xạ

1.1. Đường dẫn

Xét $n = p + q$ kí hiệu $\varepsilon_1, \varepsilon_2, ..., \varepsilon_n$, mỗi kí hiệu thay thế cho $+1$ hoặc -1. Giả sử rằng có p dấu $+1$ và q dấu -1 .
Tổng riêng :

$$s_k = \varepsilon_1 + \varepsilon_2 + ... + \varepsilon_k$$

biểu diễn sự khác biệt giữa số dấu $+1$ và dấu -1 xảy ra tại thời điểm k
Do đó :

$$s_k - s_{k-1} = \varepsilon_k = \pm 1, s_0 = 0, s_n = p - q \quad (1.1.1),$$

với $k = 1, 2, ..., n$.
Xét hệ tọa độ vuông góc t, x (trục t là nằm ngang và trục x là thẳng đứng). Mỗi bộ sắp thứ tự $(\varepsilon_1, \varepsilon_2, ..., \varepsilon_n)$ sẽ được biểu diễn bởi một đường đa giác mà cạnh thứ k của nó có độ dốc ε_k, và đỉnh thứ k cuả nó có tung độ S_k. Những đường như thế sẽ được gọi là đường dẫn. Cho $n > 0$ và x là các số nguyên. Một đường dẫn $(S_1, S_2, ..., S_n)$ từ điểm gốc tới điểm (n, x) là một đường đa giác mà các đỉnh của nó có các hoành độ là $0, 1, ..., n$ và các tung độ là $S_0, S_1, ..., S_n$ thỏa mãn (1.1.1) với $S_n = x$, n: độ dài của đường dẫn. Có 2^n đường dẫn độ dài là n. Nếu có p dấu $+1$ và q dấu -1 thì

$$n = p + q, x = p - q \quad (1.1.2).$$

Một đường dẫn từ điểm gốc tới một điểm tùy ý (n, x) chỉ tồn tại nếu n và x có dạng (1.1.2). Trong trường hợp này p điểm $\varepsilon_k > 0$ có thể được chọn từ $n = p + q$

Hình 1

có sẵn theo :

$$N_{n,x} = \begin{pmatrix} p+q \\ p \end{pmatrix} = \begin{pmatrix} p+q \\ q \end{pmatrix} \quad (1.1.3)$$

cách khác nhau.
Quy ước $N_{n,x} = 0$ nếu n và x không có dạng (1.1.2). Với quy ước này ta sẽ có $N_{n,x}$ đường dẫn khác nhau từ điểm gốc tới điểm tùy ý (n, x).

1.2. Nguyên lí phản xạ

Cho $A = (a, \alpha)$ và $B = (b, \beta)$ là những điểm nguyên thuộc góc phần tư thứ nhất : $b > a \geq 0, \alpha > 0, \beta > 0$. Điểm đối xứng của A qua trục t là $A' = (a, -\alpha)$ (Hình 1). (Nguyên lí phản xạ). Số đường dẫn từ A tới B mà chạm hoặc cắt trục x bằng số đường dẫn từ A' tới B.

Chứng minh. Xét 1 đường dẫn $(s_a = \alpha, s_{a+1}, ..., s_b = \beta)$ từ A tới B thỏa mãn có 1 hoặc nhiều đỉnh nằm trên trục t. Gọi t là hoành độ của đỉnh đầu tiên nằm trên trục t (xem Hình 1), tức là chọn t để $s_a > 0, ..., s_{t-1} > 0, s_t = 0$. Khi đó $(-s_a, -s_{a+1}, ..., -s_{t-1}, s_t = 0, s_{t+1}, s_{t+2}, ..., s_b)$ là đường dẫn từ A' tới B và có $T = (t, 0)$ là đỉnh đầu tiên của nó mà nằm trên trục t. Các đoạn AT và $A'T$ là đối xứng với nhau. Do đó tồn tại một mối tương ứng $1-1$ giữa số đường dẫn từ A' tới B và số đường dẫn như thế từ A tới B (đường dẫn mà có 1 đỉnh ở trên trục x). □

(Định lí bỏ phiếu). Cho n và x là những số nguyên dương. Có chính xác $\frac{x}{n}.N_{n,x}$ đường dẫn $(s_1, ..., s_n = x)$ từ điểm gốc tới điểm (n, x) sao cho $s_1 > 0, ..., s_n > 0$.

Chứng minh. Ta có : Số đường dẫn từ điểm gốc tới điểm (n, x) sao cho $s_1 > 0, ..., s_n > 0$ bằng số đường dẫn từ điểm $(1, 1)$ tới điểm (n, x) mà không chạm và cũng không cắt trục t.
Theo bổ đề trên số đường dẫn như thế bằng :

$$N_{n-1.x-1} - N_{n-1.x+1} = \begin{pmatrix} p+q-1 \\ p-1 \end{pmatrix} - \begin{pmatrix} p+q-1 \\ p \end{pmatrix}$$

$$= \begin{pmatrix} p+q \\ p \end{pmatrix} . \frac{p-q}{p+q} = N_{n,x} \frac{p-q}{p+q} = \frac{x}{n} . N_{n,x}$$

□

2. Sự trở về gốc

Một đường dẫn $(s_1, ..., s_n)$ được xem như sự ghi lại của việc tung 1 đồng xu n lần liên tiếp . Dãy các tổng riêng $s_1, ..., s_n$ sẽ được đánh dấu như các điểm trên trục thẳng đứng x. Chúng sẽ được gọi là các vị trí của 1 hạt mà thực hiện 1 du động ngẫu nhiên. Hạt sẽ chuyển động theo bậc đơn vị, lên hoặc xuống trên một đường thẳng. Một đường dẫn biểu diễn sự ghi lại của 1 chuyển động như thế.
Ta kí hiệu các bước nhảy là $X_1, X_2, ..., X_n$ và các vị trí của hạt là $S_1, S_2, ..., S_n$

$$\Rightarrow S_n = X_1 + ... + X_n, S_0 = 0 \quad (1.2.1),$$

trong đó X_i là các đại lượng ngẫu nhiên độc lập.

$$P\{X_i = 1\} = P\{X_i = -1\} = \frac{1}{2}.$$

Từ 1 đường dẫn cụ thể bất kì có thể đọc được các giá trị tương ứng của $X_1, X_2, ..., X_n$. Đường dẫn $(1, 2, 1, 0, -1, 0)$ có :$X_1 = X_2 = X_6 = 1, X_3 = X_4 = X_5 = -1$. Kí hiệu biến cố "Tại thời điểm n hạt là ở điểm r " là $\{S_n = r\}$. Kí hiệu xác suất của biến cố này là $p_{n,r}$. Số $N_{n,r}$ đường dẫn từ điểm gốc tới điểm (n, r) được cho bởi (1.1.3) và do đó ta có định lí sau:

Định lý 2.2. *Xác suất để " tại thời điểm n hạt là ở điểm r" được cho bởi công thức sau:*

$$p_{n,r} = p\{s_n = r\} = \binom{n}{\frac{n+r}{2}} 2^{-n} \quad (1.2.2).$$

Chứng minh. Theo công thức (1.1.3) ta có : Số đường dẫn từ điểm gốc tới điểm (n, r) là :

$$N_{n,r} = \binom{p+q}{p} = \binom{p+q}{q},$$

trong đó : $\begin{cases} n = p + q \\ r = p - q \end{cases} \Rightarrow p = \dfrac{n+r}{2} \Rightarrow N_{n,r} = \dbinom{n}{\frac{n+r}{2}}.$

Vậy $p_{n,r} = \dfrac{1}{2^n} \dbinom{n}{\frac{n+r}{2}} = \dbinom{n}{\frac{n+r}{2}} 2^{-n}.$ □

Sự trở về (điểm) gốc xảy ra tại thời điểm k nếu $S_k = 0$. Trong định nghĩa trên k phải chẵn . Cho k=2v. Khi đó xác suất của " Sự trở về gốc xảy ra tại thời điểm $k = 2v$" là : $p_{2v,0} = u_{2v} = \begin{pmatrix} 2v \\ v \end{pmatrix} 2^{-2v}$ (1.2.3). Xác suất để sự trở về gốc không xảy ra tại các thời điểm nhỏ hơn hoặc bằng $2n$ là giống như xác suất để sự trở về gốc xảy ra tại thời điểm $2n$. Tức là:

$$P\{S_1 \neq 0, ..., S_{2n} \neq 0\} = P\{S_{2n} = 0\} = u_{2n} \quad (1.2.4).$$

Trong công thức (1.2.4) biến cố ở vế trái xảy ra với mọi $S_j > 0$ hoặc với mọi $S_j < 0$. Khả năng xảy ra 2 điều ngẫu nhiên này là như nhau, do đó chúng ta có thể viết lại (1.2.4) dưới dạng:

$$P\{S_1 > 0, ..., S_{2n} > 0\} = \frac{1}{2} u_{2n} \quad (1.2.5).$$

Chứng minh. Xét tất cả các giá trị có thể của S_{2n}. Rõ ràng là :

$$P\{S_1 > 0, ..., S_{2n} > 0\} = \sum_{r=1}^{\infty} P\{S_1 > 0, ..., S_{2n-1} > 0, S_{2n} = 2r\} \quad (1.2.6).$$

Với $P\{S_1 > 0, S_2 > 0, ..., S_{2n-1} > 0, S_{2n} = 2r\} = 0$ nếu $r > n$.
Theo định lí bỏ phiếu thì số đường dẫn thỏa mãn điều kiện được chỉ ra ở vế phải bằng :

$$N_{2n-1,2r-1} - N_{2n-1,2r+1}$$

$$\Rightarrow P\{S_1 > 0, S_2 > 0, ..., S_{2n-1} > 0, S_{2n} > 0\} = \sum_{r=1}^{\infty} \frac{1}{2}(p_{2n-1,2r-1} - p_{2n-1,2r+1})$$

$$= \frac{1}{2}(p_{2n-1,1} - p_{2n-1,3} + p_{2n-1,3} - p_{2n-1,5} + ...) = \frac{1}{2} p_{2n-1,1} = \frac{1}{2} u_{2n}. \quad \square$$

Bổ đề 1 có thể được phát biểu lại theo cách sau :

$$P\{S_1 \geq 0, ..., S_{2n} \geq 0\} = u_{2n} \quad (1.2.7).$$

Thật vậy, một đường dẫn độ dài $2n$ với tất cả các đỉnh ở phía trên trục x sẽ đi qua điểm (1.1). Coi điểm này như điểm gốc mới chúng ta sẽ được 1 đường dẫn độ dài $2n - 1$ với tất cả các đỉnh ở phía trên hoặc nằm trên trục x mới . Điều này chỉ ra rằng :

$$P\{S_1 > 0, , ..., S_{2n-1} > 0, S_{2n} > 0\} = \frac{1}{2} P\{S_1 \geq 0, , ..., S_{2n-1} \geq 0\} \quad (1.2.8).$$

nhưng S_{2n-1} là 1 số lẻ, do đó $S_{2n-1} \geq 0$ chỉ ra $S_{2n} \geq 0$. Xác suất ở vế phải trong công thức (1.2.8) do đó giống như (1.2.7) và bởi vậy (1.2.7) là đúng.

3. Sự trở về gốc lần đầu

Sự trở về gốc lần đầu tiên xảy ra tại thời điểm $2v$ nếu $S_1 \neq 0, ..., S_{2v-1} \neq 0$ nhưng $S_{2v} = 0$. Kí hiệu xác suất cho biến cố "Sự trở về gốc lần đầu xảy ra tại thời điểm $2v$" là f_{2v}. Theo định nghĩa $\Rightarrow f_0 = 0$.

3.1. Mối liên hệ giữa f_{2n} và u_{2n}

Xác suất để sự trở về gốc lần đầu xảy ra tại thời điểm $2n$ được cho bởi:

$$f_{2n} = u_{2n-2} - u_{2n}, \quad n = 1, 2, ... \quad (1.3.1),$$

hoặc

$$f_{2n} = \frac{1}{2n-1} u_{2n} \quad (1.3.2).$$

Chứng minh. Xét 1 đường dẫn $(S_1, ..., S_{2n}$ với $S_{2n} = 0$. Có 2 trường hợp xảy ra :
Trường hợp 1: Sự trở về gốc lần đầu xảy ra tại thời điểm $2n$.
Trường hợp 2: Sự trở về gốc lần đầu xảy ra tại thời điểm $2k < 2n$ và sau $(2n-2k)$ đơn vị thời gian nữa thì sự trở về gốc mới xảy ra tại thời điểm $2n$.
Dễ thấy xác suất trong trường hợp 2 bằng $f_{2k}u_{2n-2k}$. Bởi vì có $2^{2k}f_{2k}$ đường dẫn độ dài $2k$ mà kết thúc với sự trở về gốc lần đầu và $2^{2n-2k}u_{2n-2k}$ đường dẫn từ điểm $(2k, 0)$ tới điểm $(2n, 0)$.
Điều này chỉ ra rằng :

$$u_{2n} = f_2 u_{2n-2} + f_4 u_{2n-4} + ... + f_{2n} u_0, \quad n \leq 1 \quad (1.3.3).$$

Định nghĩa (1.3.1) có thể phát biểu lại như sau :" Sự trở về gốc lần đầu xảy ra tại thời điểm $2n$ nếu các điều kiện $S_1 \neq 0, ..., S_{2k} \neq 0$ là đúng với $k = n - 1$, nhưng không đúng với $k = n$. Theo (1.2.4) $\Rightarrow$

$$f_{2n} = u_{2n-2} - u_{2n}, \quad n = 1, 2, ... \quad (1.3.1).$$

Biến đổi (1.3.1) ta được $f_{2n} = \dfrac{1}{2n-1} u_{2n} \quad (1.3.2)$. □

4. Sự viếng thăm cuối cùng và sự dẫn đầu

Sự viếng thăm cuối cùng tới điểm gốc xảy ra tại thời điểm $2k$ $(2k \leq 2n)$ nếu : $S_{2k} = 0; S_{2k+1} \neq 0, \ldots, S_{2n} \neq 0$.

Định lý 4.3. *(Luật arcsin cho sự viếng thăm cuối cùng) Xác suất để sự viếng thăm cuối cùng tới điểm gốc xảy ra tại thời điểm $2k(2k \le 2n)$ được cho bởi :*

$$\alpha_{2k,2n} = u_{2k}u_{2n-2k}, k = \overline{0,n} \quad (1.4.1).$$

Chứng minh. Xét các đường dẫn thỏa mãn điều kiện :

$$S_{2k} = 0; S_{2k+1} \neq 0, \ldots, S_{2n} \neq 0. \quad (*)$$

$2k$ đỉnh đầu tiên có thể được chọn theo $2^{2k}u_{2k}$ cách khác nhau. Coi điểm $(2k, 0)$ như là điểm gốc mới. Theo (1.2.4) $\Rightarrow$ $(2n - 2k)$ đỉnh tiếp theo có thể được chọn theo $2^{2n-2k}u_{2n-2k}$ cách $\Rightarrow$ xác suất để sự viếng thăm cuối cùng tới điểm gốc xảy ra tại thời điểm $2k(2k \le 2n)$ là :

$$\frac{1}{2^{2n}}.2^{2k}u_{2k}2^{2n-2k}u_{2n-2k} = u_{2k}u_{2n-2k}.$$

□

Sự phân bố xác suất mà gắn liền với giá trị $\alpha_{2k,2n}$ tại thời điểm $2k$ được gọi là phân bố arcsin rời rạc cấp n. Sự phân bố là đối xứng, tức là :$\alpha_{2k,2n} = \alpha_{2n-2k,2n}$.

Định lý 4.4. *(Luật arcsin rời rạc cho thời gian tạm trú) Xác suất để trong khoảng thời gian từ 0 đến $2n$,hạt dành $2k$ đơn vị thời gian ở phía dương và $2n - 2k$ đơn vị thời gian ở phía âm là bằng $\alpha_{2k,2n}$.*

Chứng minh. Xét các đường dẫn với độ dài cố định là $2n$, kí hiệu $b_{2k,2n}$ là xác suất để có đúng $2k$ cạnh nằm ở phía trên trục t.Ta phải chứng minh:

$$b_{2k,2v} = \alpha_{2k,2v} \quad (1.4.2).$$

(1.2.7) khẳng định rằng $b_{2v,2v} = u_{2v}$. Do tính đối xứng nên ta cũng có $b_{0,2v} = u_{2v}$. Do đó ta chỉ cần chứng minh (1.4.2) với $1 \le k \le v - 1$. Giả sử rằng có đúng $2k$ trong số $2n$ đơn vị thời gian hạt ở phần dương với $1 \le k \le v - 1$. Trong trường hợp này sự trở về gốc lần đầu phải xảy ra tại thời điểm $2r$ nào đó $(2r < 2n)$. Có 2 khả năng :

Khả năng 1: Phần đường dẫn cho tới thời điểm $2r$ là ở phía dương. Trong trường hợp này $r \le k \le n - 1$, và phần đường dẫn ở bên phải đỉnh $(2r, 0)$ có đúng $2k - 2r$ cạnh ở phía trên trục (t). Hiển nhiên số đường dẫn như vậy bằng :

$$\frac{1}{2}.2^{2r}f_{2r}.2^{2n-2r}.b_{2k-2r,2n-2r}.$$

Khả năng 2: Phần đường dẫn cho tới thời điểm $2r$ là ở phía âm. Trong trường hợp này, phần (đường dẫn) ở bên phải đỉnh $(2r, 0)$ có đúng $2k$ cạnh ở phía trên trục (t), do đó $n - r \geq k$. Số đường dẫn như thế bằng :

$$\frac{1}{2}.2^{2r} f_{2r}.2^{2n-2r}.b_{2k,2n-2r}$$

$$\Rightarrow b_{2k,2n} = \frac{1}{2}\sum_{r=1}^{k} f_{2r} b_{2k-2r,2n-2r} + \frac{1}{2}\sum_{r=1}^{n-k} f_{2r} b_{2k,2n-2r} \quad (1 \leq k \leq n-1) \quad (1.4.3).$$

Ta sẽ chứng minh (1.4.2) bằng phương pháp quy nạp.
+ Với $v = 1 \Rightarrow$ (1.4.2) đúng.
+ Giả sử (1.4.2) đúng với $v \leq n - 1$. Do đó từ (1.4.3)

$$\begin{aligned}\Rightarrow b_{2k,2n} &= \frac{1}{2}u_{2n-2k}\sum_{r=1}^{k} f_{2r}u_{2k-2r} + \frac{1}{2}u_{2k}\sum_{r=1}^{n-k} f_{2r}u_{2n-2k-2r} \\ &= \frac{1}{2}u_{2n-2k}.u_{2k} + \frac{1}{2}u_{2k}.u_{2n-2k} \\ &= u_{2n-2k}.u_{2k} = \alpha_{2k,2n}.\end{aligned}$$

$\Rightarrow$ (1.4.2) đúng với $v = n$. □

5. Sự đổi dấu

Một sự thay đổi về dấu được nói là xảy ra tại thời điểm n nếu S_{n-1} và S_{n+1} là trái dấu. Tức là nếu đường dẫn cắt trục (t). Trong trường hợp này $S_n = 0$ và do đó n phải là 1 số nguyên dương và chẵn.

Định lý 5.5. *Xác suất để tới thời điểm $2n+1$ đã xảy ra r lần sự thay đổi về dấu là $\xi_{r,2n+1} = 2p_{2n+1,2r+1}$.*
Nói cách khác:

$$\xi_{r,2n+1} = 2P\left\{S_{2n+1} = 2r+1\right\}, r = 0, 1, ... \quad (1.5.1).$$

Chứng minh. Trước tiên chúng ta diễn đạt lại định lí theo 1 dạng thuận tiện hơn. Nếu bước nhảy đầu tiên tới điểm $(1, 1)$, chúng ta sẽ coi điểm này như điểm gốc của 1 hệ tọa độ mới. Khi đó đường dẫn cắt trục nằm ngang (t) của hệ tọa độ cũ ta coi như cắt ở mức -1. Trường hợp $S_1 = -1$ được làm tương tự.
Với cách làm như vậy thì định lí 1 sẽ tương đương với mệnh đề sau đây:"Xác suất để tới thời điểm $2n$, mức -1 được cắt r lần bằng $2p_{2n+1,2r+1}$.

+) Xét trường hợp $r = 0$. Khi đó đường dẫn sẽ không cắt ở mức -1. Do đó nó cũng sẽ không chạm (hoặc cắt) ở mức $-2 \Rightarrow S_{2n}$ phải là 1 số nguyên chẵn và không âm. Với $k \geq 0$, từ bổ đề phản xạ $\Rightarrow$ số đường dẫn từ điểm $(0,0)$ tới điểm $(2n, 2k)$ mà chạm mức -2 bằng số đường dẫn từ điểm $(0,0)$ tới điểm $(2n, 2k+4)$. Xác suất để tới điểm $(2n, 2k)$ đường dẫn vẫn chưa chạm mức -2 bằng $p_{2n,2k} - p_{2n,2k+4}$. Xác suất để tới thời điểm $2n$ đường dẫn vẫn chưa chạm mức -2 là:

$$P = \sum_{k \geq 0} (p_{2n,2k} - p_{2n,2k+4})$$

$$= p_{2n,0} - p_{2n,4} + p_{2n,2} - p_{2n,6} + p_{2n,4} - p_{2n,8} + ...$$

$$= p_{2n,0} + p_{2n,2}.$$

Mặt khác $p_{2n+1,1} = \frac{1}{2}(p_{2n,0} + p_{2n,2}) \quad (1.5.2)$.
(do mỗi đường dẫn mà đi qua điểm $(2n+1, 1)$ thì sẽ phải đi qua điểm $(2n, 0)$ hoặc $(2n, 2)$)
$P = 2p_{2n+1,1} \Rightarrow$ mệnh đề đúng với $r = 0$.
+) Xét trường hợp $r = 1$
Một đường dẫn mà cắt mức -1 tại thời điểm $2v - 1$ có thể được phân tích thành 2 đường dẫn : Đường dẫn thứ 1 từ điểm $(0,0)$ tới điểm $(2v, -2)$ và đường dẫn thứ 2 có độ dài $2n - 2v$ mà bắt đầu tại điểm $(2v, -2)$. Đường dẫn thứ 2 chúng ta có thể áp dụng kết quả với $r = 0$ nhưng với $S_1 = -1$ mà không phải là 1.
Ta có : Số đường dẫn có độ dài $2n - 2v$ với điểm gốc là $(2v, -2)$ và không cắt mức -1 chính là bằng số đường dẫn từ điểm $(2v, -2)$ tới điểm $(2n+1, -3)$. Nhưng mỗi đường dẫn như vậy mà kết hợp với đường dẫn thứ 1 ta sẽ được 1 đường dẫn từ điểm $(0,0)$ tới điểm $(2n+1, -3)$. Do đó số đường dẫn độ dài $2n$ mà cắt mức -1 sẽ bằng số đường dẫn từ điểm gốc tới điểm $(2n+1, -3)$ và bằng $2^{2n+1}.p_{2n+1,3} \Rightarrow$ mệnh đề đúng với $r = 1$.
Khẳng định mệnh đề đúng với r tùy ý được suy ra bằng phương pháp quy nạp. $\square$

Hệ quả: Kí hiệu $\xi_{r,n}$ là xác suất để trong n phép thử sẽ xảy ra r lần sự thay đổi về dấu . Khi đó $\xi_{r,n}$ sẽ giảm khi r tăng.
Tức là :

$$\xi_{0,n} \geq \xi_{1,n} \geq \xi_{2,n} \geq ...$$

6. Các giá trị lớn nhất và sự đi qua lần đầu tiên

Ta nói giá trị lớn nhất của một đường dẫn là nhỏ hơn r nếu tất cả các đỉnh của đường dẫn này đều nằm ở phía dưới đường thẳng $x = r$,tức là đường dẫn thỏa mãn điều kiện sau :

$$S_0 < r, S_1 < r, ..., S_n < r \quad (1.6.1).$$

Giá trị lớn nhất sẽ ≥ 0 bởi vì $S_0 = 0$. Cho điểm $A(n,k), k \leq r$. Một đường dẫn từ $O(0,0)$ tới A mà chạm hoặc cắt đường thẳng $x = r$ nếu nó vi phạm điều kiện (1.6.1). Theo nguyên lí phản xạ, số đường dẫn như thế sẽ bằng số đường dẫn từ điểm gốc tới điểm $A'(n, 2r-k)$. (A' là điểm đối xứng của A qua đường thẳng $x = r$).
Từ nhận xét trên ta có bổ đề sau đây:
Bổ đề 1: Cho $k \geq r$. Xác suất để 1 đường dẫn độ dài n (từ điểm gốc) tới điểm $A(n,k)$ và có 1 giá trị lớn nhất $\leq r$ bằng : $p_{n,2r-k} = P\{S_n = 2r-k\}$.

Định lý 6.6. *Xác suất để giá trị lớn nhất của* 1 *đường dẫn độ dài* n *bằng* $r \geq 0$ *là* : $p_{n,r} + p_{n,r+1} = \begin{cases} p_{n,r}, p_{n,r} > 0 \\ p_{n,r+1}, p_{n,r+1} > 0 \end{cases}$

Chứng minh. Xác suất để giá trị lớn nhất bằng r là :$p_{n,2r-k} - p_{n,2r+2-k}$
Xác suất để 1 đường dẫn tùy ý độ dài n có 1 giá trị lớn nhất bằng r là :

$$\sum_{k \leq r} (p_{n,2r-k} - p_{n,2r+2-k}) = p_{n,r} + p_{n,r+1}.$$

Nếu n và r cùng chẵn hoặc cùng lẻ $\Rightarrow P_{n,r+1} = 0$. □

Với $r = 0$ và độ dài của đường dẫn là một số chẵn $(2n)$ ta có:

$$P\{S_1 \leq 0, S_2 \leq 0, ..., S_{2n} \leq 0\} = u_{2n} \quad (1.6.2).$$

Công thức (1.6.2) tương đương với công thức (1.2.7). Do đó chúng ta thấy rằng định lí 1 chính là sự khái quát của bổ đề (1.2.4). Một sự đi qua điểm r lần đầu tiên $(r > 0)$ là diễn ra tại thời điểm n nếu :

$$S_1 < r, ..., S_{n-1} < r, S_n = r \quad (1.6.3).$$

Hình 2

Định lý 6.7. *Xác suất để sự đi qua điểm r lần đầu tiên xảy ra tại thời điểm n là*

$$\phi_{r,n} = \frac{1}{2}\left[p_{n-1,r-1} - p_{n-1,r+1}\right] \quad (1.6.4)$$

$$= \frac{1}{2}\left[\begin{pmatrix} n-1 \\ \dfrac{n+r-2}{2} \end{pmatrix} - \begin{pmatrix} n-1 \\ \dfrac{n+r}{2} \end{pmatrix}\right] = \frac{r}{n}\begin{pmatrix} n \\ \dfrac{n+r}{2} \end{pmatrix} 2^{-n} \quad (1.6.5).$$

Chứng minh. Dễ thấy 1 đường dẫn thỏa mãn điều kiện (1.6.3) phải đi qua điểm $(n-1, r-1)$ và giá trị lớn nhất của nó cho đến thời điểm $n-1$ phải là $r-1$. Chúng ta thấy xác suất cho biến cố này chính là $p_{n-1,r-1} - p_{n-1,r+1}$. □

(Xác suất để sự đi qua điểm r lần đầu tiên xảy ra trước thời điểm N là : $\sum\limits_{n \le N} \phi_{n,r}$.

Định lý 6.8. *Xác suất để (hạt) trở về điểm gốc lần thứ r xảy ra tại thời điểm n là : $\phi_{r,n-r}$.*

Chứng minh. Xét 1 đường dẫn từ điểm gốc tới điểm $(n, 0)$ mà có tất cả các cạnh nằm ở phía dưới trục (t) và có đúng $r-1$ đỉnh thuộc trục (t). Đường dẫn thỏa mãn các điều kiện nói trên được gọi là đường dẫn đại diện (hình 2 chỉ ra 1 đường như thế với $n = 20$ và $r = 5$). Một đường dẫn đại diện bao gồm r đường dẫn nhỏ . Mỗi đường dẫn nhỏ có các điểm cuối thuộc trục (t). Bằng việc gán các dấu khác nhau cho các đỉnh trong từng đường dẫn nhỏ (tức là bằng việc lấy đối xứng từng đường dẫn nhỏ qua trục (t)) ta sẽ được 2^r đường dẫn khác nhau. Mỗi đường dẫn này đều trở về điểm gốc lần thứ r tại thời điểm n. Do đó định lí được phát biểu lại như sau :" Số đường dẫn đại diện độ dài n bằng số đường dẫn độ dài $n-r$ mà kết thúc với 1 sự đi qua điểm r lần đầu tiên". Điều này là đúng, bởi vì nếu trong 1 đường đại diện ta bỏ đi r cạnh thỏa mãn "các điểm cuối ở bên trái của các cạnh này là nằm trên trục (t)" ta sẽ được 1 đường dẫn độ dài $n-r$ mà kết thúc với "sự đi qua điểm r lần đầu tiên ". Ngược lại nếu ta chèn r cạnh với độ dốc âm mà bắt đầu tại điểm gốc và $r-1$ đỉnh ta sẽ được một đường dẫn đại diện độ dài n (xem Hình 2). □

7. Đối ngẫu. Vị trí của giá trị lớn nhất

Mỗi đường dẫn tương ứng với 1 dãy hữu hạn các dấu $+1$ và -1, và nếu đảo ngược thứ tự của các số hạng ta sẽ được 1 đường dẫn mới. Về mặt hình học, đường dẫn mới thu được bằng việc quay đường dẫn cũ 180^0 quanh điểm (n, S_n) (đường dẫn cũ đi từ điểm $O(0,0)$ tới điểm (n, S_n)), và coi điểm (n, S_n) như là điểm gốc của 1 hệ tọa độ mới.

Theo trên, suy ra với mỗi lớp đường dẫn C tương ứng sẽ có 1 lớp đường dẫn mới C' mà $CardC = CardC'$. Nếu các bước nhảy của du động ngẫu nhiên ban đầu là $X_1, X_2, ..., X_n$ thì các bước nhảy của du động ngẫu nhiên mới là :

$X_1^* = X_n, ..., X_n^* = X_1 \quad (1.7.1)$.

Và các đỉnh của du động ngẫu nhiên mới được xác định bởi các tổng riêng:

$S_k^* = X_1^* + ... + X_k^* = S_n - S_{n-k} \quad (1.7.2)$.

$S_0^* = 0, S_n^* = S_n$. Du động ngẫu nhiên mới được xác định bởi các công thức (1.7.1) và (1.7.2) được gọi là du động ngẫu nhiên đối ngẫu của du động ngẫu nhiên ban đầu. Với mỗi biến cố A được xác định trong du động ngẫu nhiên ban đầu tương ứng sẽ có 1 biến cố B trong du động ngẫu nhiên đối ngẫu mà $P(A) = P(B)$.

Điểm (n, S_n) của đường dẫn được gọi là điểm cuối (điểm kết thúc). **Các ví dụ**:

Ví dụ 1: Các thời điểm đi qua đầu tiên

Các biến cố :$A = \left\{S_j^* > 0, j = 1, 2, ..., n\right\} \quad (1.7.3)$

và $B = \{S_n > S_j, j = 0, 1, ..., n-1\} \quad (1.7.4)$

là đối ngẫu của nhau (S_j^* được xác định trong (1.7.2)).

Biến cố B có nghĩa là :" Một chuyến thăm tới điểm cuối không thể xảy ra trước thời điểm n ".

Từ (1.2.5) $\Rightarrow P(A) = \dfrac{1}{2}u_{2v}\,(n = 2v > 0)$.

Với $n = 2v+1$ ta cũng có $P(A) = \dfrac{1}{2}u_{2v}\,(n = 2v > 0)$. Bởi vì $S_{2v}^* > 0 \Rightarrow S_{2v+1}^* > 0$.

Do $P(A) = P(B)$ nên ta có : Xác suất để sự đi qua 1 điểm dương đầu tiên mà diễn ra tại thời điểm n bằng $\dfrac{1}{2}u_{2v}$ với $v = \dfrac{n}{2}$, hoặc $v = \dfrac{n-1}{2}$. (Điều này đúng với $n = 1$ nhưng lại sai với $n = 0$).

Ví dụ 2 :Tính liên tục

Xét biến cố B với $S_n = r$. Biến cố đối ngẫu bao gồm các đường dẫn từ điểm gốc tới điểm (n, r) với tất cả các đỉnh nằm ở phía trên trục (t). Số đường dẫn như vậy được suy ra từ bổ đề phản xạ (với $A = (1,1)$ và $B = (n, r)$)

Ví dụ 3:Giá trị lớn nhất tại điểm cuối

Cho cặp đối ngẫu :

$$C = \left\{ S_j^* \geq 0, j = \overline{1,n} \right\}$$
$$C' = \left\{ S_n \geq S_j, j = \overline{0, n-1} \right\}$$

Biến cố C' xảy ra khi S_n đạt giá trị lớn nhất, thậm chí giá trị lớn nhất này đạt được trước thời điểm n.
Từ (1.2.7) suy ra : $P(C') = P(C) = u_{2v}$ với $v = \dfrac{n}{2}$ hoặc $v = \dfrac{n+1}{2}$.
Dễ thấy $P(C') = P(C) = 2P(A)$.

Ví dụ 4: Định lý arcsin cho chuyến thăm đầu tiên tới điểm cuối

Xét biến cố D :" Chuyến thăm đầu tiên tới điểm cuối xảy ra tại thời điểm $2k$ ". Tức $S_{2k} = S_{2v}, S_j \neq S_{2v}$ với $j \leq 2k.D$ là đối ngẫu của biến cố D' :" Chuyến thăm cuối cùng tới điểm gốc xảy ra tại thời điểm $2k$" .
Theo 1.4 suy ra $P(D) = P(D') = \alpha_{2k,2v} = u_{2k}u_{2v-2k}, \quad (k = 0, ..., v)$.

Ví dụ 5: Định lý arcsin cho vị trí của giá trị lớn nhất

Cho $n = 2v$. Giá trị lớn nhất đầu tiên xảy ra tại thời điểm k nếu :
$S_0 < S_k, ..., S_{k-1} < S_k \quad (1.7.5a)$
và $S_{k+1} \leq S_k, ..., S_{2v} \leq S_k \quad (1.7.5b)$
Theo ví dụ 1, xác suất của $(1.7.5a)$ là $\dfrac{1}{2}u_{2p} \quad (k = 2p, p \neq 0)$.
Xác suất của $(1.7.5b)$ chính là xác suất để một đường dẫn độ dài $2v - k$ với tất cả các đỉnh nằm ở phía dưới hoặc nằm trên trục (t). Theo ví dụ 3 xác suất này bằng $u_{2v-2p} \quad (k = 2p)$.
Do đó nếu $0 \leq k \leq 2v$ thì xác suất để trong dãy $S_0, ..., S_{2v}$ giá trị lớn nhất đầu tiên xảy ra tại thời điểm $k = 2p$ là : $\dfrac{1}{2}u_{2p}u_{2v-2p}$.
Với $k = 0$ và $k = 2v$ thì xác suất lần lượt là u_{2v} và $\dfrac{1}{2}u_{2v}$.

Chương 2.

Các ví dụ ứng dụng

1. Ứng dụng của đường dẫn trong thực tế

1.1. Định lí bỏ phiếu

Mệnh đề:" Giả sử rằng trong 1 đợt bỏ phiếu kín, người ứng cử P được p phiếu bầu và người ứng cử Q được q phiếu bầu $(p > q)$. Khi đó xác suất để số phiếu bầu cho P lớn hơn số phiếu bầu cho Q là :

$$(p - q) \,/\, (p + q) \,.\text{”}$$

Toàn bộ số phiếu đã bầu có thể được biểu diễn bởi 1 đường dẫn độ dài $p + q$ trong đó $\varepsilon_k = +1$ nếu phiếu thứ k là bầu cho P; ngược lại mỗi đường dẫn từ điểm gốc tới điểm $(p + q, p - q)$ có thể được hiểu như 1 sự ghi lại đợt bỏ phiếu với tổng $p + q$ phiếu đã bầu.Người ứng cử P dẫn đầu về số phiếu nếu và chỉ nếu $S_1 > 0, ..., S_n > 0$, nghĩa là tất cả các đỉnh đều nằm ở phía trên trục (t).

1.2. Kiểm tra thứ tự hạng của Galton

Giả sử rằng một đại lượng (ví dụ như chiều cao của cây) được đo bằng một trong r đối tượng xử lí và cũng được đo bằng một trong r đối tượng điều khiển. Kí hiệu các kết quả đo lần lượt là $a_1, ..., a_r$ và $b_1, ..., b_r$. Giả sử rằng $a_1 > a_2 > ...$, $b_1 > b_2 > ...$ và không có 2 quan trắc nào trùng nhau. Ta sẽ kết hợp 2 dãy thành 1 dãy gồm $n = 2r$ số được sắp xếp theo thứ tự giảm dần. Một xử lí thành công nếu tất cả a đều đứng trước b, trái lại một xử lí không tốt sẽ dẫn đến một vị trí ngẫu nhiên của a và b. Do đó hiệu quả xử lí sẽ căn cứ vào số các số ađứng trước b mà có cùng hạng với b, tức là :căn cứ vào số các chỉ số dưới k mà $a_k > b_k$. Galton đã nghĩ ra ý tưởng trên vào năm 1876. Để minh họa ý tưởng trên ta sẽ xét đường dẫn độ dài $2n$ nối điểm gốc tới điểm $(2r, 0)$. Trong đó $\varepsilon_k = +1$ hoặc -1 nếu số hạng thứ k của dãy kết hợp là a hoặc b. Biến cố $a_k > b_k$ xảy ra nếu và chỉ nếu s_{2k-1} chứa ít nhất k dấu $+1$, tức là nếu $s_{2k-1} > 0 \Rightarrow s_{2k} \geq 0$. Do đó các cạnh thứ $2k-1$ và $2k$ phải ở phía trên trục (t). Như vậy bất đẳng thức $a_k > b_k$ đúng v lần nếu và chỉ nếu $2v$ cạnh nằm ở phía trên trục t

1.3. Trò chơi tung đồng xu

Trò chơi tung đồng xu n lần liên tiếp có thể được biểu diễn bởi 1 đường dẫn độ dài n trong đó $\varepsilon_k = +1$ nếu phép thử thứ k đồng xu xuất hiện mặt ngửa. Các tổng riêng S_k biểu diễn sự khác biệt giữa số mặt sấp và số mặt ngửa tại phép thử thứ k.

2. Minh họa sự viếng thăm cuối cùng và sự dẫn đầu

Ví dụ 1. Chúng ta biết phân bố Arcsin rời rạc cấp n là đối xứng .Tức là : $\alpha_{2k,2n} = \alpha_{2n-2k,2n}$

+) Với $n = 2$ có 3 giá trị $\frac{3}{8}, \frac{2}{8}, \frac{3}{8}$

+)Với $n = 10$ hãy xem bảng 1 sau đây (các số hạng trung tâm luôn nhỏ nhất)

Bảng 1

Phân bố Arcsin rời rạc cấp 10

	k = 0 k = 10	k = 1 k = 9	k = 2 k = 8	k = 3 k = 7	k = 4 k = 6	k = 5
$\alpha_{2k,20}$	0.1762	0.0927	0.0736	0.0655	0.0617	0.0606

Ví dụ 2. Trò chơi tung đồng xu được kéo dài 365 ngày. Bảng sau đây sẽ đưa ra xác suất p và thời gian t_p tương ứng sao cho với xác suất p thì người chơi sẽ dành chiến thắng trong khoảng thời gian $< t_p$.

p	t_p	p	t_p
0.9	153.95 ngày	0.3	19.89 ngày
0.8	126.10 ngày	0.2	8.93 ngày
0,7	99.65 ngày	0.1	2.24 ngày
0,6	75.23 ngày	0.05	13.5 giờ
0.5	53.45 ngày	0.02	2.16 giờ
0.4	34.85 ngày	0.01	32,4 phút

Ví dụ 3. Cho đường dẫn (1,2,1,0,1,0,-1,-2). Khi đó sự viếng thăm cuối cùng tới điểm gốc xảy ra tại thời điểm $k = 6$ và trong khoảng thời gian từ 0 đến 8 hạt dành 6 đơn vị thời gian ở phía dương và 2 đơn vị thời gian ở phía âm...

3. Minh họa sự đổi dấu

Ví dụ 1. Xác suất x_r để trong 99 phép thử (lần tung) xảy ra r lần sự thay đổi về dấu được cho trong bảng dưới đây :

r	x_r	r	x_r
0	0.1592	7	0.0517
1	0.1529	8	0.0375
2	0.1412	9	0.0260
3	0.1252	10	0.0174
4	0.1066	11	0.0111
5	0.0873	12	0.0068
6	0.0686	13	0.0040

Ví dụ 2.Xác suất x_r để trong 10000 phép thử xảy ra r lần sự thay đổi về dấu là:

r	x_r
10	0.0156
20	0.0146
30	0.0130

Xác suất để trong 10000 phép thử (lần tung), vị trí dẫn đầu thay đổi nhiều nhất 10 lần là khoảng 0.0174.

Ví dụ 3. Đường dẫn (-1,0,1,0,-1,0,1,0) tới thời điểm $k = 7$ đã xảy ra 3 lần đổi dấu

: Lần thứ nhất tại thời điểm $k = 2$, lần thứ 2 tại thời điểm $k = 4$, lần thứ 3 tại thời điểm $k = 6$.

4. Minh họa thực nghiệm

Hình 3 biểu diễn kết quả của 1 thí nghiệm thu được trên máy tính. Kết quả này mô phỏng việc tung 1 đồng xu 10,000 lần. Đường trên cùng là đồ thị biểu diễn 550 phép thử đầu tiên. Hai đường tiếp theo biểu diễn sự ghi lại toàn bộ 10,000 phép thử với tỉ lệ theo phương ngang là 1:10, tỉ lệ theo phương thẳng đứng là đúng như hình vẽ.
Nhìn vào đồ thị chúng ta sẽ ngạc nhiên bởi khoảng cách giữa 2 lần liên tiếp mà đường dẫn cắt trục (t). Nếu chúng ta quan sát đồ thị theo hướng ngược lại. (tức là quan sát từ lần tung thứ 10,000 cho đến lần tung thứ 1). Khi đó du động ngẫu nhiên có đặc điểm sau: Bắt đầu từ điểm gốc.
Đường dẫn ở

Phần âm	Phần dương
7804 bước đầu tiên	8 bước tiếp theo
2 bước tiếp theo	54 bước tiếp theo
30 bước tiếp theo	2 bước tiếp theo
48 bước tiếp theo	6 bước tiếp theo
2046 bước tiếp theo	
Tổng : 9930 bước	Tổng : 70 bước
Tỉ lệ thời gian : 0.993	Tỉ lệ thời gian : 0.007

Xác suất để trong 10,000 nghìn lần tung đồng xu , có trên 9930 lần "sự dẫn đầu " ở 1 phía và ít hơn 70 lần ở phía còn lại là lớn hơn 1/10. Xác suất để xảy ra sự cân bằng chỉ là 0.072. Ghi lại ban đầu trong hình 3 chứa 78 lần sự thay đổi về dấu và 64 lần trở về điểm gốc. Trong khi đó chuỗi đảo ngược chỉ có 8 lần thay đổi về dấu và 6 lần trở về điểm gốc. Xác suất để xảy ra nhiều nhất 8 lần sự thay đổi về dấu >0.14. Trong khi đó xác suất để xảy ra ít nhất 78 lần đổi dấu là khoảng 0.12.

5. Minh họa giá trị lớn nhất và sự đi qua đầu tiên

Ví dụ 1.
a). Đường dẫn $(1, 2, 3, 4, 3, 2, 1, 0)$ có giá trị lớn nhất là 4 xảy ra tại thời điểm $k = 4$.
b) Đường dẫn $(0, 1, 2, 1, 0, -1, 0, 1, 2, 1)$ có giá trị lớn nhất đầu tiên là 2 và giá trị lớn nhất đầu tiên này xảy ra tại thời điểm $k = 3$.

Ví dụ 2. Cho đường dẫn $(-1, 0, 1, 2, 1, 0)$. Khi đó sự đi qua điểm 1 lần đầu tiên là diễn ra tại thời điểm $k = 3$ và sự đi qua điểm 1 lần thứ 2 là diễn ra tại thời điểm $k = 5$.

Ví dụ 3. Cho đường dẫn $(-1, 0, 1, 2, 3, 4)$. Khi đó sự đi qua điểm 4 lần đầu tiên là diễn ra tại thời điểm $n = 6$.

KẾT LUẬN

Nội dung được trình bày trong báo cáo là những kiến thức cơ sở, rất dễ hiểu. Mỗi một bài trong chương đều có định nghĩa, định lý và chứng minh khá chặt chẽ. Những độc giả mới bắt đầu học xác suất hoàn toàn có thể đọc và hiểu được báo cáo này. Mặt khác nó cũng là tài liệu tham khảo tốt cho những ai muốn nghiên cứu sâu về lĩnh vực xác suất thống kê. Do thời gian có hạn nên những nội dung được trình bày trong báo cáo vẫn còn mang tính chất khái quát. Các ví dụ ứng dụng hầu hết đều là những ví dụ hết sức đơn giản. Vì vậy nếu bạn đọc muốn tìm hiểu sâu hơn về báo cáo có thể tham khảo thêm các tài liệu được trích ở cuối báo cáo này, các bạn có thể nghiên cứu thêm phần " Các thời điểm đi qua đầu tiên (Chương XI)", "Du động ngẫu nhiên(Chương XIV) " trong [3].

TÀI LIỆU THAM KHẢO

[1] Đào Hữu Hồ, Xác suất thống kê, Nhà xuất bản Đại Học Quốc Gia Hà Nội (Bản in lần thứ V- 1999, lần thứ VII- 2003).

[2] Đặng Hùng Thắng, Mở đầu về lý thuyết xác suất và ứng dụng, NXB Giáo dục, 1998.

[3] William Feller, An introduction to Probability theory and its applications, Newyork. London. Sydney, Volume I, Third edition.