XÁC SUẤT

Để lập phương án xây đập thủy điện, chúng ta phải có các thông tin về lưu lượng nước của các dòng sông chảy vào hồ chứa nước. Thí dụ nếu chúng ta biết chắc chắn lưu lượng tổng hợp các con sông vào giữa tháng 7 hằng năm chính xác là *x*, chúng ta có thể dùng *x* như là một biến số trong các tính toán của công trình.

Tuy nhiên con số *x* này lại thay đổi hằng năm, có năm rất nhỏ vì những cơn hạn hán lịch sử, nhưng có năm lại cực kỳ lớn trong những cơn lũ lịch sử. Các số liệu bất thường này nhiều khi làm cho trung bình cộng của *x* trong một số năm quan sát quá khác biệt với số *x* trong đại đa số các năm thông thường.

Để giải quyết việc này, chúng ta dùng lý thuyết độ đo. Đặt Ω là tập hợp các năm chúng ta lấy dử liệu, thường được mô hình toán học như là $\{1, 2, 3, \ldots, m\}$, nếu số năm quan sát là m. Chọn $\mathcal{M} = \mathcal{P}(\Omega)$. Đặt $\mu(k)$ là một số thực dương dựa vào dử liệu lịch sử hàng trăm năm (có thể nhiều hơn rất nhiều số phần tử của Ω) với mọi k trong Ω, sao cho $\mu(\Omega) = 1$. Gọi X là ánh xạ từ Ω, với $X(k)$ chính là giá trị của x tương ứng với năm "k" .

Cho hai số thực α và β. Giả sử $\mu(\{k: \alpha < X(k) \leq \beta\}) > 0.9$. Chúng ta có thể nói : khả năng lưu lượng giữa tháng bảy hằng năm là giữa hai số α và β, điều này chính xác đến 90%. Sau đó dùng các phương trình ngẫu nhiên để tính toán tiếp !

Vì thế, trong ngành xác suất thống kê, ta gọi các hàm số đo được là các biến số ngẫu nhiên. Các biến số ngẫu nhiên thường được ký là X, Y, …(vừa có ý nghĩa biến số x, y, . . . , vừa có ý nghĩa là chúng không phải là số thực).

Ta thấy việc lựa chọn đô đo μ trên không gian lấy mẩu Ω rất quan trọng. Nếu làm theo kiểu bình quân cào bằng sẽ đưa đến các kết luận không chính xác trong các bài toán thực tế ngoài đời. Vì độ đo μ trong các trường hợp này tương ứng với các xác xuất của các dử liệu, nên chúng ta thường ký hiệu nó là P (probability).

Định nghĩa. Cho $(\Omega, \mathfrak{M}, P)$ là một không gian đo được với đo đo dương P. Ta nói $(\Omega, \mathfrak{M}, P)$ là không gian xác suất nếu $P(\Omega) = 1$.

Lúc đó Ω được gọi là không gian mẫu (sample space), các tập $A \in \mathfrak{M}$ được gọi là các biến cố, và độ đo $P(A)$ được gọi là xác suất của biến cố A. Một ánh xạ đo được X từ Ω vào $\mathbb{R}$, được gọi là một biến số ngẫu nhiên.

Tuy có nền tảng chung là lý thuyết độ đo, nhưng xác suất thống kê có những đặc điểm khác hẳn giải tích và cơ học. Thí dụ sau đây cho chúng ta thấy phần nào sự khác nhau này.

Qua một số khảo sát, xác suất từ một trứng cá hồi trở thành một con cá hồi trưởng thành là 10^{-4}, nghĩa là cứ 10.000 trứng cá hồi có khả năng có một con sống đến lúc trưởng thành . Vậy xác suất để có 3 cá hồi trưởng thành khi ta nuôi 40.000 trứng cá hồi là bao nhiêu? Có phải là 100% ?

Câu trả lời : không phải 100% !

Như vậy xác xuất thống kê không đơn thuần là chuyển đổi tên gọi của giải tích !

Trong các trường hợp Ω là một tập có n phần tử, chúng ta có thể chọn $\mathcal{M} = \mathcal{P}(\Omega)$, tập hợp tất cả các tập con của Ω, và đặt

$$P(A) = \frac{|\mathrm{A}|}{|\Omega|} \qquad \forall\ A \subset \Omega,$$

ở đây $|B|$ là số phần tử của B.

Lúc đó (Ω, $\mathcal{M}$, P) là một không gian xác suất.

Cách đặt này đưa đến việc phải tính cho đúng số phần tử của A, để trả lời câu hỏi xác suất $P(A)$ của A là bao nhiêu. Nhiều lúc, người ta hiểu lầm việc tính số phần tử của các tập con A là công việc chủ yếu của môn xác suất thống kê !

Cho một tập hợp hữu hạn U và n tập hợp con rời nhau của U : $E_1, \ldots, E_n$, sao cho $U = \bigcup_{k=1}^{n} E_k$. Đặt Ω là tập $\{1, \ldots, n\}$, $\mathfrak{M} = \mathcal{P}(\Omega)$ và

$$P(A) = \frac{\left|\bigcup_{i \in A} E_i\right|}{|U|} \qquad \forall\ A \subset \Omega.$$

Lúc đó $(\Omega, \mathfrak{M}, P)$ là một không gian xác suất.

Như vậy, khởi đầu từ không gian lấy mẫu U, nhưng sau khi mô hình, ta lại làm việc trên không gian xác suất $(\Omega, \mathfrak{M}, P)$, với độ đo P không còn tính bình quân.

Nếu chúng ta khảo sát xác suất về các chiều dài các trẻ sơ sinh ở một vùng nào đó, Ω không còn là một tập hữu hạn nữa, nó là một khoảng đóng $[a, b]$.

Định nghĩa. Cho hai biến cố A và B trong một không gian xác suất $(\Omega, \mathfrak{M}, P)$, ta nói A và B độc lập với nhau nếu $P(A \cap B) = P(A)\, P(B)$

Khái niệm độc lập này có ý nghĩa như sau

$$\frac{P(A \cap B)}{P(B)} = \frac{P(A)}{P(\Omega)}$$

Vậy tỉ trọng của biến cố A đối với toàn cục (Ω) bằng tỉ trọng của biến cố $A \cap B$ đối với biến cố B.

Định nghĩa. Cho m biến cố $A_1, \ldots A_m$ trong một không gian xác suất $(\Omega, \mathfrak{M}, P)$, ta nói $A_1, \ldots A_m$ độc lập với nhau nếu $P(A_1 \cap \ldots \cap A_m) = P(A_1) \ldots P(A_m)$

Thí dụ . Úp 52 lá bài lên một mặt bàn và chọn ngẫu nhiên trong đó một lá bài. Gọi *A* là biến cố khi ta chọn đúng một quân bài có hình (bồi, đầm, già , *K, Q, J*) , *B* là biến cố khi ta chọn đúng một quân bài già (*K*) và *C* là biến cố khi ta chọn đúng một quân bài cơ (♥). Ta có

$$P(A) = \frac{12}{52} = \frac{3}{13}, \qquad P(B) = \frac{4}{52} = \frac{1}{13}, \qquad P(C) = \frac{13}{52} = \frac{1}{4},$$

$$P(A \cap C) = \frac{3}{52}, \qquad P(B \cap C) = \frac{1}{52}, \qquad P(A \cap B) = P(B) = \frac{1}{13}$$

Vậy *B* và *C* độc lập với nhau, *A* và *C* độc lập với nhau, nhưng *A* và *B* không độc lập với nhau

Tính chất độc lập có vai trò quan trọng trong các qui luật số lớn và giới hạn của các tiến trình ngẫu nhiên.

Thí dụ. Cho I và J là hai khoảng trong $[0,1]$. Đặt $\Omega = [0,1]\times[0,1]$ và P là độ đo Lebesgue trên Ω. Đặt $A = I\times[0,1]$ và $B = [0,1]\times J$. Lúc đó

$$P(A\cap B) = P(I\times J) = P(A)\,P(B)\,.$$

Vậy A và B độc lập với nhau.

Định nghĩa. Cho m biến số ngẫu nhiên $X_1, \ldots X_m$ trong một không gian xác suất $(\Omega, \mathfrak{M}, \mathrm{P})$, ta nói $X_1, \ldots X_m$ độc lập với nhau nếu $X_1^{-1}(U_1), \ldots X_m^{-1}(U_m)$ độc lập với nhau với mọi $U_1, \ldots, U_m$ mở trong $\mathbb{R}$.

Khái niệm độc lập của các biến số ngẫu nhiên có vai trò rất quan trọng trong việc tiếp cận thực tế bằng xác suất thống kê. Tuy nhiên chúng ta cần một số mô hình toán học để sữ dụng chúng.

Cho m không gian xác suất $(\Omega_1, \mathfrak{M}_1, \mathrm{P}_1)$, . . ., $(\Omega_m, \mathfrak{M}_m, \mathrm{P}_m)$, và Y_1 , ..., Y_m lần lượt là các biến số ngẫu nhiên trên các không gian xác suất đó. Giả sử các biến số ngẫu nhiên này không liên quan với nhau.

Chúng ta sẽ mô hình toán học sự không liên quan này như sau. Đặt $\Omega = \Omega_1 \times \ldots \times \Omega_m$. Dùng lý thuyết độ đo ta xây dựng được một không gian xác suất $(\Omega, \mathfrak{M}, \mathrm{P})$ có các tính chất sau :

- $\mathfrak{M}$ là σ- đại số nhỏ nhất chứa $A_1 \times \ldots \times A_m$ nếu $A_k \in \mathfrak{M}_k$ với mọi $k \in \{1, \ldots, m\}$.
- Nếu $A_k \in \mathfrak{M}_k$ với mọi $k \in \{1, \ldots, m\}$, ta có $\mathrm{P}(A_1 \times \ldots \times A_m) = \mathrm{P}_1(A_1) \times \ldots \times \mathrm{P}_m(A_m)$

Đặt $X_k(\omega_1, \ldots, \omega_k, \ldots, \omega_m) = Y_k(\omega_k) \ \forall \ (\omega_1, \ldots, \omega_m) \in \Omega$.

Bài toán 3.1. $X_1, \ldots X_m$ là các biến số ngẫu nhiên độc lập với nhau trong không gian xác xuất $(\Omega, \mathfrak{M}, P)$.

Cho một họ không gian xác suất $\{(\Omega_i, \mathfrak{M}_i, P_i) : i \in I\}$. và Y_i là một biến số ngẫu nhiên trên $\{(\Omega_i, \mathfrak{M}_i, P_i)$. Giả sử các biến số ngẫu nhiên này không liên quan với nhau. Đặt $\Omega = \prod_{i \in I} \Omega_i$. Dùng lý thuyết độ đo ta xây dựng được một không gian xác suất $(\Omega, \mathfrak{M}, P)$ có các tính chất sau :

- $\mathfrak{M}$ là σ- đại số nhỏ nhất chứa $(\prod_{k \in J} A_k) \times (\prod_{i \in I \setminus J} \Omega_i)$ nếu $A_k \in \mathfrak{M}_k$ với mọi $k \in J$, và J là tập con hữu hạn của I
- Nếu $A_k \in \mathfrak{M}_k$ với mọi $k \in J$, ta có

$$P((\prod_{k \in J} A_k) \times (\prod_{i \in I \setminus J} \Omega_i)) = \prod_{k \in J} P_k(A_k)$$

Đặt $X_k((\omega_i)) = Y_k(\omega_k) \qquad \forall\ (\omega_i) \in \Omega.$

Bài toán 3.2. Cho J là một tập con hữu hạn trong I. Chứng minh $\{X_k : k \in J\}$ là các biến số ngẫu nhiên độc lập với nhau trong không gian xác xuất $(\Omega, \mathfrak{M}, P)$.

Như vậy khái niệm độc lập của các biến số ngẫu nhiên cho phép chúng ta mô hình toán học chặc chẽ sự không liên quan với nhau của một họ biến số ngẫu nhiên xác định trên các không gian ngẫu nhiên khác nhau. Hơn nữa khái niệm này đưa các biến số ngẫu nhiên đó về các biến số ngẫu nhiên trong cùng một không gian ngẫu nhiên.

Xin xem thêm chi tiết sự thiết lập $(\Omega, \mathfrak{M}, P)$ trong sách "A course in probability theory" của K.L. Chung (57-64 pp). (http://gigapedia.com/items/26256/a-course-in-probability-theory-revised)

Sau đây chúng ta quan sát một hiện tượng cho thấy cần phát triển thêm khái niệm mới trong xác suất.

Giả sử như chúng ta có một đồng xu, có một mặt có hình ảnh và một mặt có số, và khi tung đồng xu này, khả năng đồng xu rơi xống và mặt phía trên là mặt hình hoặc số như nhau. Như vậy nếu chúng ta tung 2n lần, thì biến cố *A "có n lần được mặt hình"* phải có xác xuất tiến dần đến đâu khi *n* tiến ra vô hạn?

Do tính "vô tư" của đồng xu, chúng ta nghĩ ngay là giới hạn này là 1.

Thực ra xác suất $P(A)$ lại tiến về 0 khi n tiến ra vô hạn.

Thật vậy, tập hợp tất cả các trường hợp tung $2n$ lần đồng xu có k lần mặt hình tương ứng với số đơn ánh từ $\{1, \ldots, k\}$ vào $\{1, \ldots, 2n\}$ và chính là

$$\frac{(2n)!}{k!(2n-k)!}$$

Đặt $a_n = P(A)$, ta có

$$a_n = 2^{-2n} \frac{(2n)!}{n!n!}$$

$$\frac{a_n}{a_{n-1}} = [2^{-2n-2} \frac{(2n-2)!}{(n-1)!(n-1)!}]^{-1}[2^{-2n} \frac{(2n)!}{(n)!(n)!}] = \frac{2n-1}{2n} = 1 - \frac{1}{2n}$$

$$a_n = (1 - \frac{1}{2n})a_{n-1} = a_1 \prod_{k=2}^{n} (1 - \frac{1}{2k}) = \frac{1}{2} \prod_{k=2}^{n} (1 - \frac{1}{2k})$$

Dùng định lý giá trị trung bình cho [ln(1-*t*) - ln1], ta có

$$\ln a_n = \ln\frac{1}{2} + \sum_{k=2}^{n} \ln\left(1-\frac{1}{2n}\right) \le \ln\frac{1}{2} - \sum_{k=2}^{n} \frac{1}{2n} \to -\infty$$

Vậy $P(A) = a_n$ tiến về 0 khi n tiến ra vô cùng. Điều này cho thấy dựa vào tính "vô tư" của đồng xu để cho rằng biến cố *A "có n lần được mặt hình"* phải có xác xuất tiến dần đến 1 khi *n* tiến ra vô hạn là không đúng!

Vậy chúng ta diễn tã thế nào về tác động của tính "vô tư" của đồng xu vào thực tế?

Để trả lời vấn đề này, chúng ta cần những định lý hội của xác suất.

Cho ω là một lượt tung $2n$ đồng xu, ta đặt $h(\omega)$ là số lần tung cho mặt hình, và $q_n(\omega) = 2^{-2n} h(\omega)$. Nếu ω có số lần tung có mặt hình là n ta có $q_n(\omega) = 2^{-1}$, đây là tỉ lệ tương ứng với tính "vô tư" của đồng xu. Ta có kết quả sau.

Với mọi số thực dương ε , ta có

$$\lim_{n\to\infty} P(\{\omega : | q_n(\omega) - \tfrac{1}{2} | < \varepsilon\}) = 1$$

Điều này có thể lý giải như sau : với bất kỳ số dương ε, xác suất các lượt tung đồng xu có tỉ lệ mặt hình gần với tính "vô tư" của đồng xu (với sai biệt là ε) tiến về 1 khi n tiến ra vô cùng.

Nay ta mô hình hóa bài toán trên

Đặt Ω_n là tập hợp tất cả các lượt tung đồng xu $2n$ lần, $\mathcal{M}_n = \mathcal{P}(\Omega_n)$ và $P_n(A)$ là số phần tử của A với mọi A trong $\mathcal{M}_n$. Đặt $\Omega = \prod_{n\in\mathbb{N}} \Omega_n$, $\mathcal{M}$ và P như trong phần trước.

Cho ω là một lượt tung $2n$ đồng xu, ta đặt $h(\omega)$ là số lần tung cho mặt hình trong ω, và $q_n(\omega) = (2n)^{-1}h(\omega)$. Với mọi $\omega = (\omega_m) \in \Omega$ và $n \in \mathbb{N}$, ta đặt $X(\omega) = 2^{-1}$ và

$$X_n(\omega) = q_n(\omega_n) .$$

Kết quả bên trên viết thành

$$\lim_{n\to\infty} P_n(\{\omega_n \in \Omega_n : | q_n(\omega_n) - \tfrac{1}{2} | < \varepsilon\}) = 1$$

$$\lim_{n\to\infty} P(\{\omega \in \Omega :| X_n(\omega) - X(\omega) | < \varepsilon\}) = 1$$

Định nghĩa. Cho một dãy biến số ngẫu nhiên $\{X_m\}$ và một biến số ngẫu nhiên X trong một không gian xác suất $(\Omega, \mathfrak{M}, P)$, ta nói dãy $\{X_m\}$ hội tụ xác suất về X, nếu với mọi số thực dương ε, ta có

$$\lim_{n\to\infty} P(\{\omega\in\Omega : | X_n(\omega) - X(\omega) | > \varepsilon\}) = 0$$

Nếu X là một ánh xạ khả tích trong không gian xác xuất $(\Omega, \mathfrak{M}, P)$, ta gọi tích phân của X là kỳ vọng và ký hiệu là $E(X)$, nghĩa là

$$E(X) = \int_{\Omega} X dP$$

Kỳ vọng có ý nghĩa giá trị trung bình của biến số ngẫu nhiên X. Nếu $\Omega = \{1, 2, \ldots, n\}$ là một mô hình toán học trong một khảo sát nào đó . Lúc đó

$$E(X) = \int_{\Omega} X dP = \sum_{k=1}^{n} X(k) P(\{k\})$$

chính là trị giá trung bình của X theo độ đo xác suất P.

Ta thường ký hiệu $E(X)$ là μ .

Nếu $\Omega = \{1, 2, \ldots, n\}$. Ta coi các giá trị lấy mẫu như là một vectơ $(X(1), \ldots, X(n))$, và sự sai biệt giữa μ và các giá trị này như vectơ $(X(1) - \mu, \ldots, X(n) - \mu)$. Tương tự như khoảng cách trong các không gian Euclide, ta có thể định độ sai lệch chung của trị giá trung bình μ và các giá trị $X(1), \ldots, X(n)$ trong độ đo xác suất P như là

$$\sqrt{\sum_{k=1}^{n} (X(k) - \mu)^2 P(\{k\})}$$

Lúc đó $\sigma = \sqrt{\sum_{k=1}^{n} (X(k) - \mu)^2 P(\{k\})}$ được gọi là phương sai của biến ngẫu nhiên X.

Bài toán 3.3. Trong trường hợp trên, chứng minh

$$\sigma^2 = \sum_{k=1}^{n}(X(k)-\mu)^2 P(k) = [\sum_{k=1}^{n} X^2(k)P(k)] - \mu^2$$

Trong các trường hợp tổng quát của Ω, ta có thể định độ sai lệch chung của trị giá trung bình μ và các giá trị của X trong độ đo xác suất P như là phương sai

$$\sigma = \sqrt{\int_{\Omega}(X-\mu)^2 dP}$$

Nếu X là một biến ngẫu nhiên trong không gian xác suất $(\Omega, \mathfrak{M}, P)$. Chúng ta rất khó tìm trị giá chung cho $X(\omega)$ với mọi $\omega \in \Omega$, và thực ra ta chỉ cần tìm hai số thực α và β sao cho xác suất $P(\{\omega : \alpha < X(\omega) \leq \beta \})$ lớn hơn một số cho sẵn (thí dụ 0,93). Vì thế khái niệm sau đây có vai trò rất quan trọng trong xác suất.

Định nghĩa. Cho X là một biến ngẫu nhiên trong không gian xác suất $(\Omega, \mathfrak{M}, P)$. Hàm phân phối của biến số ngẫu nhiên X được xác định như sau

$$F(t) = P(\{\omega \in \Omega : \ X(\omega) \leq t\}) \qquad \forall \ t \in \mathbb{R}.$$

Định nghĩa. Cho X là một biến ngẫu nhiên trong không gian xác suất $(\Omega, \mathfrak{M}, P)$. Hàm phân phối của biến số ngẫu nhiên X được xác định như sau

$$F(t) = P(\{\omega \in \Omega : \ X(\omega) \le t\}) \qquad \forall\ t \in \mathbb{R}.$$

Bài toán 3.4. $F(t) \le F(s)$ nếu $t < s$.

H.D. Để ý $\{\omega \in \Omega : X(\omega) < t\} \subset \{\omega \in \Omega : X(\omega) < s\}$

Bài toán 3.5. $\lim_{t \to \infty} F(t) = 1$

H.D : Đặt $f_n(\omega) = 1$ nếu $X(\omega) < n$ và $f_n(\omega) = 0$ nếu $X(\omega) \ge n$. Chứng minh $\{f_n\}$ là dãy hàm tăng và hội tụ về hàm $f(\omega) = 1$ với mọi $\omega \in \Omega$. Dùng định lý hội tụ đơn điệu của Lebesgue. Áp dụng thêm bài toán 3.1.

Bài toán 3.6. $\lim_{t \to -\infty} F(t) = 0$

H.D : Đặt $f_n(\omega) = 1$ nếu $X(\omega) < -n$ và $f_n(\omega) = 0$ nếu $X(\omega) \geq -n$. Chứng minh $\{f_n\}$ là dãy hàm giãm và hội tụ về hàm $f(\omega) = 0$ với mọi $\omega \in \Omega$. Dùng định lý hội tụ bị chặn của Lebesgue. Áp dụng thêm bài toán 3.1.

Bài toán 3.7. $\lim_{t \to s, t < s} F(t) = F(s)$

H.D : Đặt $f_n(\omega) = 1$ nếu $X(\omega) < s - n^{-1}$ và $f_n(\omega) = 0$ nếu $X(\omega) \geq s - n^{-1}$. Chứng minh $\{f_n\}$ là dãy hàm tăng và hội tụ về hàm f, với $f(\omega) = 1$ nếu $X(\omega) < s$ và $f(\omega) = 0$ nếu $X(\omega) \geq s$. Dùng định lý hội tụ đơn điệu của Lebesgue. Áp dụng thêm bài toán 3.1.

Định nghĩa. Cho $A_1, A_2, \ldots, A_m$ là m tập đo được rời nhau trong một một không gian xác suất $(\Omega, \mathfrak{M}, P)$ và $c_1, c_2, \ldots, c_m$ là m số thực. Giả sử $c_1 < c_2 < \ldots < c_m$ và

$$\Omega = \bigcup_{k=1}^{m} A_k$$

và

$$X(\omega) = \sum_{k=1}^{m} c_k \chi_{A_k}(\omega) \qquad \forall\ \omega \in \Omega.$$

Đặt $p(c_k) = P(A_k)$ với mọi $k \in \{1, \ldots, m\}$ và $p(t) = 0$ khi $t \in \mathbb{R} \setminus \{c_1, c_2, \ldots, c_m\}$. Lúc đó p_k được gọi là hàm mật độ của biến số ngẫu nhiên X.

Bài toán 3.8. Lúc đó, chứng minh hàm phân bố F của X được tính theo công thức $F(t) = \sum_{c_k < t} p(c_k)$

Định nghĩa. Cho X là một biến ngẫu nhiên trong không gian xác suất $(\Omega, \mathcal{M}, P)$. Giả sử có hàm số p không ậm, liên tục và khả tích (Lebesgue) trên $\mathbb{R}$ sao cho hàm phân phối của biến số ngẫu nhiên X được xác định theo công thức

$$F(t) = \int_{-\infty}^{t} p(s)ds \qquad \forall\ t \in \mathbb{R}.$$

p được gọi là hàm mật độ của biến ngẫu nhiên X.

Trong nhiều bài toán, chúng ta không cần nói đến không gian xác suất $(\Omega, \mathcal{M}, P)$ trên đó biến số ngẫu nhiên X được xác định, mà chỉ đề cập đến các yếu tố sau của X: μ, σ, hàm phân phối và hàm mật độ của X.

Khi chúng ta đưa một số biến cố ngẫu nhiên trên các không gian xác suất khác nhau và không liên quan với nhau về một họ biến ngẫu nhiên độc lập trên cùng một không gian xác suất, các μ,σ và các hàm phân phối và hàm mật độ của chúng không thay đổi.

Bài toán 3.8. Cho X là một biến ngẫu nhiên không âm trong không gian xác suất $(\Omega,\mathcal{M}, P)$. Chứng minh

$$P(\{\omega \in \Omega : X(\omega) \geq \alpha\}) \leq \frac{1}{\alpha} E[X] \qquad \forall\ \alpha > 0.$$

H.D. Đặt $A = \{\omega\in\Omega: X(\omega) \geq \alpha\}$. Chứng minh

$$\alpha\chi_A(\omega) \leq X(\omega) \qquad \forall\ \omega \in \Omega.$$

Bài toán 3.9. (Markov) Cho X là một biến ngẫu nhiên không âm trong không gian xác suất $(\Omega, \mathfrak{M}, P)$. Chứng minh

$$P(\{\omega \in \Omega : X(\omega) \geq \alpha\}) \leq \frac{1}{\alpha^r} E[X^r] \qquad \forall \ \alpha > 0, r > 0.$$

H.D. Đặt $A = \{\omega \in \Omega: X(\omega) \geq \alpha\}$. Chứng minh

$$\alpha^r \chi_A(\omega) \leq X^r(\omega) \qquad \forall \ \omega \in \Omega.$$

Bài toán 3.10. (Chebyshev) Cho Y là một biến ngẫu nhiên trong không gian xác suất $(\Omega, \mathfrak{M}, P)$ và $\mu = E[Y]$ và phương sai σ. Chứng minh

$$P(\{\omega \in \Omega : |Y(\omega) - \mu| \geq \alpha\}) \leq \frac{\sigma^2}{\alpha^2} \qquad \forall \ \alpha > 0.$$

H.D. Đặt $X(\omega) = |Y(\omega) - \mu|$ với mọi $\omega \in \Omega$. Áp dụng bài toán 3.9.

Bài toán 3.11. Cho X là một biến số ngẫu nhiên không âm trên $(\Omega, \mathfrak{M}, P)$ với hàm mật độ p. Chứng minh

$$E(X) = \int_0^{\infty} sp(s)ds$$

H.D. Giải bài toán cho các biến ngẫu nhiên đơn trước. Chọn một dãy biến ngẫu nhiên đơn không âm $\{X_m\}$ đơn điệu tăng hội tụ về X. Đặt p_m là hàm mật độ của X_m. Chứng minh $\{X_m{}^q\}$ đơn điệu tăng hội tụ về X^q, và $\{p_m\}$ đơn điệu tăng và hội tụ hầu hết mọi nơi về p. Dùng định lý hội tụ đơn điệu của Lebesgue.

Bài toán 3.12. Cho X là một biến số ngẫu nhiên khả tích trên $(\Omega, \mathfrak{M}, P)$ với hàm mật độ p. Chứng minh

$$E(X) = \int_0^\infty sp(s)ds$$

H.D. Để ý $X = X^+ - X^-$, với $X^+ = \max\{X,0\}$ và $X^- = \max\{-X,0\}$. Đặt p, p_1 và p_2 lần lượt là hàm mật độ của X, X^+ và X^-. Chứng minh $p = p_1 + p_2$. Áp dụng bài toán 3.11.

Bài toán 3.13. Cho một số nguyên dương m và X là một biến số ngẫu nhiên khả tích trên $(\Omega, \mathfrak{M}, P)$ với hàm mật độ p. Chứng minh

$$\int_\Omega X^m dP = \int_0^\infty s^m p(s) ds$$

H.D. Giải bài toán cho các biến ngẫu nhiên đơn trước. Sau đó xét các biến ngẫu nhiên X không âm. Áp dụng phương pháp của bài toán 3.12, để ý $X^+ . X^- = 0$.

Các dạng hàm phân bố thông dụng

•Phân bố nhị thức .Giả sử chúng ta tiến hành các thử nghiệm, trong mỗi thử nghiệm có n lần thử, trong mỗi lần thử chỉ có hai trường hợp : thành công hoặc thất bại.. Giả sử dựa vào một số dự báo được kiểm định từ trước, ta biết có một số thực q trong (0,1), để cho khả năng thành công của mỗi lần thử là q. Lúc đó khả năng thất bại của mỗi lần thử là (1-q). Đặt Ω là tập hợp các thử nghiệm, $\mathfrak{M} = \mathcal{P}(\Omega)$. Cho $\omega \in \Omega$, nếu có k lần thành công trong thử nghiệm ω, ta đặt $P(\omega) = q^k(1-q)^{n-k}$. Vì số các thử nghiệm có k lần thành công là $\dfrac{n!}{k!(n-k)!}$

$$P(\Omega) = \sum_{k=0}^{n} \frac{n!}{k!(n-k)!} q^k (1-q)^{n-k} = [q + (1-q)]^n = 1$$

Vậy $(\Omega,\mathfrak{M},P)$ là không gian xác xuất. Đặt $X(\omega)$ là số lần thành công trong lượt thí nghiệm $\omega\in\Omega$. Ta thấy X là một biến ngẫu nhiên trên $(\Omega,\mathfrak{M},P)$ và có hàm mật độ và hàm phân bố của X như sau

$$p(k)=\begin{cases} q^k(1-q)^{n-k} & \forall\ k\in\{1,\cdots,n\},\\ 0 & \forall\ k\in\mathbb{R}\setminus\{1,\cdots,n\}.\end{cases}$$

$$F_X(t)=\sum_{k<t}\frac{n!}{k!(n-k)!}q^k(1-q)^{n-k} \qquad \forall\ t\in[0,\infty).$$

Đây là dạng phân phối nhị thức B(n,q) .

Thí dụ. Theo các nghiên cứu khoa học, chỉ có 0.01% trứng cá hồi có thể phát triển thành một con các trưởng thành. Nay chúng ta nuôi 40.000 trứng cá hồi, hỏi có bao nhiêu khả năng chúng ta thu được ít nhất ba con cá trưởng thành?

Ta xem một lần nuôi 40.000 trứng cá hồi như là một thử nghiệm. Đặt Ω là tập các lần thử nghiệm như vậy, và $X(\omega)$ là số lần thành công trong lượt thí nghiệm $\omega\in\Omega$. Ta thấy X có hàm phân bố nhị thức $B(n,q)$ với $n = 40000$ và $q = 0{,}0001$. và có hàm mật độ là

$$p_n(j) = P(\{X = j\}) = \frac{n!}{j!(n-j)!}(q)^j(1-q)^{n-j} \qquad \forall\ j \in \{1,\cdots,n\}$$

Vậy xác suất để có ít nhất ba con cá trưởng thành từ 40000 trứng là

$$P(\{X \geq 3\}) = 1 - p_n(0) - p_n(1) - p_n(2)$$

$$= 1 - \sum_{j=0}^{2} \frac{40000!}{j!(40000-j)!}(0,0001)^j (0,9999)^{40000-j}$$

Xác suất này không bằng 1 như chúng ta tưởng. Như vậy với kiến thức của thông kê, chúng ta sẽ lập các dự án khả thi và có kết quả hơn.

Thực ra, nếu tính bằng máy tính, xác suất trên lại bằng 1, vì 0,0001 quá nhỏ và 40.000 quá lớn! Để tính xấp xĩ các số này chúng ta dùng Poisson’s limit law.

- Phân bố Poisson.

Cho một số thực dương α, ta thấy

$$\sum_{k=0}^{\infty}\frac{\alpha^k e^{-\alpha}}{k!} = e^{-\alpha}\sum_{k=0}^{\infty}\frac{\alpha^k}{k!} = e^{-\alpha}e^{\alpha} = 1$$

Một biến số ngẫu nhiên X có hàm mật độ

$$p(t) = \begin{cases} \dfrac{\alpha^k e^{-\alpha}}{k!} & \forall\ t \in \{0,1,2,3,\cdots\}, \\ 0 & \forall\ t \in \mathbb{R} \setminus \{0,1,2,3,\cdots\}. \end{cases}$$

được gọi là biến số ngẫu nhiên có phân phối Poisson với tham số α và ký hiệu là $X \sim P(\alpha)$.

Phân bố nhị thức B(n,q) rất tiện ích trong các vấn đề chỉ có hai trạng thái (sai và đúng, sinh thêm và chết đi, . . .), nhưng chúng ta sẽ gặp khó khăn trong tính toán khi dùng luật phân bố này khi q quá gần với 0 và n quá lớn. Lúc đó chúng ta có thể dùng phần bố Poisson P(np) để tính xấp xĩ các xác suất trong phân bố B(n,q) như sau.

Định lý (Luật Poisson). Cho α là một số thực dương. Đặt p_n là hàm mật độ của một biến số ngẫu nhiên có phân phối nhị phân $B(n,n^{-1}\alpha)$, và p là hàm mật độ của biến số ngẫu nhiên có phân phối Poisson $P(\alpha)$. Lúc đó, với mọi số nguyên dương k, ta có $\lim_{n\to\infty} p_n(k) = p(k)$

Với $n = 40000$, $q = 0{,}0001$ và $\alpha = nq = 4$, xấp xĩ $B(n,q)$ bằng $P(\alpha)$, ta có thể tính xác xuất có ít nhất 3 con cá hồi sống sót đến tuổi trưởng thành trong từng đợt nuôi 40 ngàn trứng như sau .

$$P(\{X \geq 3\}) = 1 - p_n(0) - p_n(1) - p_n(2)$$

$$\simeq 1 - p(0) - p(1) - p(2) = 1 - e^{-4}\sum_{j0}^{2}\frac{4^{-j}}{j!} \simeq 0{,}7619.$$

Nếu cho ta tính mức rủi ro để bảo hiểm các sản xuất nuôi cá hồi, ta không thể tính mức chi phí bảo hiểm quá nhỏ !

- Phân phối chuẩn

Cho một số thực μ và một số thực dương σ, đặt

$$f(t) = \frac{1}{\sigma\sqrt{2\pi}} \exp \frac{-(t-\mu)^2}{2\sigma^2} \qquad \forall\ t \in \mathbb{R}.$$

Ta thấy

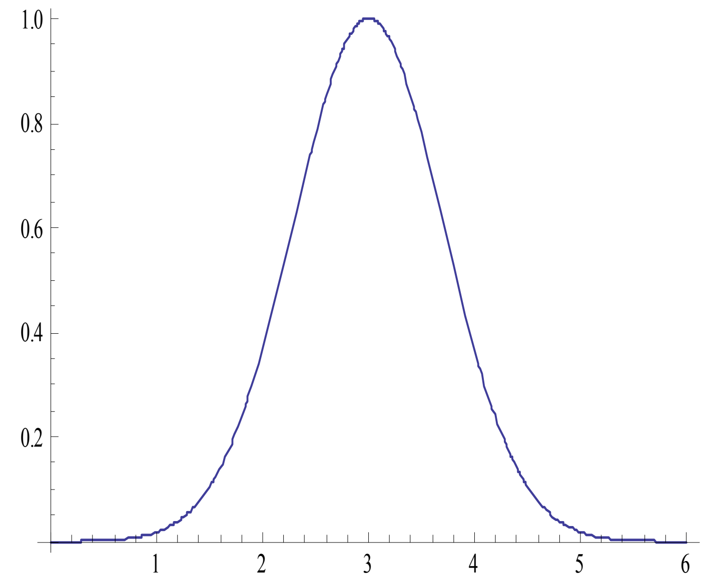
$$\int_{-\infty}^{\infty} f(t)dt = 1$$

Các biến ngẫu nhiên có hàm mật độ dạng hàm *f* được gọi là biến số ngẫu nhiên có phân phối chuẩn $N(\mu, \sigma^2)$.

Bài toán 3.14. Cho X là một biến ngẫu nhiên có phân bố chuẩn $N(\mu,\sigma^2)$. Chứng minh $E(X) = \mu$ và phương sai của X là σ.

H.D. Dùng các bài toán 3.11 và 3.12.

Khi biến số ngẫu nhiên diễn tả những sự kiện có định hướng về một số nào đó : khi sản xuất một sợi dây cáp điện, đường kính của dây được kiểm soát liên tục để luôn luôn bằng một cở μ.



Cho X là biến số ngẫu nhiên trên $[0,10]$ chỉ kích cở của sợi dây dài 10 mét. Lúc đó X có phân bố chuẩn $N(\mu,\sigma^2)$, σ phản ánh độ đồng đều của sợi dây.

Hàm phân bố của biến ngẫu nhiên có phân bố $N(0,1)$ có dạng

$$F(t)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t}\exp\frac{-s^2}{2}ds \qquad \forall\ t\in\mathbb{R}.$$

Chúng ta có thêm một cách tính xấp xĩ cho các biến số có phân bố nhị thức $B(n,p)$ như sau.

Định lý (DeMoivre-Laplace). Cho a và b là các số dương, $q\in(0,1)$. Cho X_n là biến số ngẫu nhiên có phân bố nhị thức $N(n,q)$. Ta có

$$\lim_{n\to\infty}P(\{a\le\frac{S_n-nq}{\sqrt{nq(1-q)}}\le b\})=\Phi(b)-\Phi(a).$$