Lý thuyết xác suất và thống kê toán học

Phạm Đình Tùng

Đại Học Khoa Học Tự Nhiên

15/1/2010

Tài liệu

Tài liệu bắt buộc

1. Đặng Hùng Thắng, Mở đầu về lý thuyết xác suất và các ứng dụng, NXB Giáo Dục 2005.
2. Đặng Hùng Thắng, Bài tập xác suất, NXB Giáo Dục 2008.
3. Đào Hữu Hồ, Xác suất thống kê, NXB ĐHQG Hà Nội 2004.

Tài liệu tham khảo

1. Nguyễn Viết Phú, Nguyễn Duy Tiến, Cơ sở lý thuyết xác suất, NXB ĐHQG Hà Nội 2004.
2. Đặng Hùng Thắng,Thống kê ứng dụng, NXB Giáo Dục 2008
3. Đào Hữu Hồ, Hướng dẫn giải các bài toán xác suất thống kê , NXB ĐHQG Hà Nội 2007.

Lý thuyết xác suất I

1. Biến cố và xác suất của biến cố
   - Phép thử ngẫu nhiên và không gian mẫu
   - Biến cố và quan hệ giữa các biến cố
   - Xác suất của biến cố và các quy tắc tính
   - Xác suất có điều kiện
   - Công thức xác suất đầy đủ và công thức Bayes
   - Phép thử lặp và công thức Bernoulli
2. Đại lượng ngẫu nhiên rời rạc
   - Phân bố xác suất và hàm phân bố
   - Các đại lượng đặc trưng của đại lượng ngẫu nhiên rời rạc
   - Phân bố đồng thời và hệ số tương quan
   - Hàm của đại lượng ngẫu nhiên
   - Một số phân bố rời rạc thường gặp

Lý thuyết xác suất II

3. Đại lượng ngẫu nhiên liên tục
   - Hàm mật độ và hàm phân bố xác suất
   - Các đặc trưng của đại lượng ngẫu nhiên liên tục
   - Hàm của đại lượng ngẫu nhiên
   - Một số phân phối liên tục thường gặp
4. Luật số lớn và các định lý giới hạn
   - Hội tụ theo xác suất của dãy đại lượng ngẫu nhiên
   - Luật số lớn
   - Định lý giới hạn trung tâm tổng quát và các dạng đặc biệt

Thống kê ứng dụng

Phần I

Lý thuyết xác suất

Phép thử ngẫu nhiên và không gian mẫu

Định nghĩa

- Trong thực tế ta gặp rất nhiều hành động mà không biết trước được kết quả. Tất cả những hành động đó là các phép thử ngẫu nhiên. Phép thử ngẫu nhiên thường được ký hiệu là ξ .
- Tập hợp tất cả các kết quả của ξ được kí hiệu là Ω. Khi đó Ω được gọi là không gian mẫu của phép thử ξ.

Ví dụ

- Phép thử ξ: thực hiện tung một con xúc xắc lên, sau đó quan sát mặt xuất hiện của con xúc sắc.
- Không gian mẫu $\Omega = \{$'mặt 1','mặt 2','mặt 3','mặt 4','mặt 5','mặt 6' $\}$.

Biến cố và quan hệ giữa các biến cố

Biến cố

- Biến cố là kết quả của phép thử ngẫu nhiên.
- Ký hiệu của biến cố là các chữ cái in hoa như : A,B,C, ...
- Ví dụ : A='mặt 1', ...

Phân loại biến cố

- Biến cố không thể xảy ra, kí hiệu: $\emptyset$.
- Biến cố chắc chắn xảy ra, kí hiệu: Ω.
- Biến cố ngẫu nhiên là biến cố có thể xảy ra hoặc không.
- Biến cố sơ cấp là biến cố không thể phân chia thành các biến cố nhỏ hơn.

Quan hệ giữa các biến cố

- Hợp hai biến cố A và B là biến cố xảy ra nếu có ít nhất một trong hai biến cố A hoặc B xảy ra. Kí hiệu $A \cup B$ hay A+B.
- Giao hai biến cố A và B là biến cố xảy ra nếu cả hai biến cố A và B cùng xảy ra. Kí hiệu $A \cap B$ hay AB.
- Biến cố A được gọi là kéo theo B nếu A xảy ra thì B xảy ra. Kí hiệu $A \subset B$.
- Biến cố đối của biến cố A là $\bar{A} = \Omega \setminus A$.
- Biến cố xung khắc: A và B là hai biến cố xung khắc nếu $A \cap B = \emptyset$.
- Biến cố độc lập : A và B là hai biến cố độc lập khi và chỉ khi A xảy ra không ảnh hưởng đến việc B xảy ra và ngược lại.

Ví dụ

Tung con xúc sắc, đặt các biến cố sau
A=" xuất hiện mặt 1"; B="xuất hiện mặt chẵn". Khi đó

- $A \cup B$="Xuất hiện mặt chẵn hoặc mặt 1".
- $A \cap B = \emptyset$, hay A và B là hai biến cố xung khắc.
- $\bar{A}$ = "không xuất hiện mặt 1".
- $\bar{A} \cap B = B$.
- B là biến cố kéo theo với $\bar{A}$.

Ví dụ.

Ba sinh viên A, B, C cùng thi môn xác suất. Gọi A là biến cố: " Sinh viên A thi đỗ"; B là biến cố: " Sinh viên B thi đỗ"; C là biến cố: " Sinh viên C thi đỗ"

1. Hãy mô tả các biến cố sau:

$$ABC; A \cup B \cup C; \bar{A}\bar{B}\bar{C}; \bar{A}BC;$$

2. Hãy biểu diễn các biến cố sau theo 3 biến cố A, B, C
   D : " Có ít nhất hai sinh viên thi đỗ ."
   E : "Có nhiều nhất 1 sinh viên thi đỗ."
   F : "Có duy nhất sinh viên A thi đỗ."

Lời giải:

1. Ta có:
   ABC: "Cả 3 sinh viên thi đỗ."
   $A \cup B \cup C$: "Có ít nhất 1 sinh viên thi đỗ".
   $\bar{A}\bar{B}\bar{C}$: "Cả 3 sinh viên đều thi trượt".
   $\bar{A}BC$: "Chỉ có sinh viên A thi trượt".
2. $D = AB \cup BC \cup CA = AB\bar{C} \cup A\bar{B}C \cup \bar{A}BC \cup ABC.$
   $E = \bar{A}\bar{B} \cup \bar{B}\bar{C} \cup \bar{C}\bar{A} = A\bar{B}\bar{C} \cup \bar{A}\bar{B}C \cup \bar{A}B\bar{C} \cup \bar{A}\bar{B}\bar{C}.$
   $F = A\bar{B}\bar{C}.$

Xác suất của biến cố và các quy tắc tính

Định nghĩa xác suất của biến cố

A là một biến cố, xác suất để xảy ra biến cố A là một con số thể hiện khả năng xảy ra A hoặc tỉ lệ xuất hiện A trong một tập hợp các kết quả của phép thử ngẫu nhiên. Kí hiệu là P(A).

Tính chất

- $P(\emptyset) = 0, P(\Omega) = 1.$.
- $0 \leq P(A) \leq 1.$
- $P(A) + P(\bar{A}) = 1.$

Ví dụ : Tung một đồng xu, gọi A="xuất hiện mặt sấp". Khi đó P(A)=0,5 hay khả năng xuất hiện mặt sấp là 50 %.

Quy tắc tính xác suất

- Cho hai biến cố bất kỳ A, B

$$P(A \cup B) = P(A) + P(B) - P(A \cap B),$$

- Nếu $A \cap B = \emptyset$ thì

$$P(A \cup B) = P(A) + P(B),$$

- Nếu A và B độc lập thì

$$P(AB) = P(A)P(B)$$

Chú ý: Việc khái quát các công thức trên trong trường hợp ba biến cố trở lên rất đơn giản nhờ phép quy nạp.

Ví dụ

Trong một vùng dân cư tỉ lệ người mắc bệnh tim là 9%, mắc bệnh huyết áp là 12% và mắc cả hai bệnh là 7%. Chọn ngẫu nhiên một người trong vùng. Tính xác suất để người đó không mắc cả bệnh tim và bệnh huyết áp.

Lời giải: A="Người đó mắc bệnh tim"; B="Người đó mắc bệnh huyết áp". Theo giả thiết ta có P(A)=0,09; P(B)=0,12 và P(AB)=0,07.
H="người đó không mắc cả bệnh tim và huyết áp", suy ra $\bar{H}$="Người đó mắc bệnh tim hoặc huyết áp"$=A \cup B$.
Khi đó

$$P(\bar{H})=P(A)+P(B)-P(AB)=0,14.$$

Do đó

$$P(H)=1-P(\bar{H})=0,86.$$

Công thức xác suất cổ điển

Thực hiện phép thử ngẫu nhiên ξ. Gọi A là một biến cố trong không gian mẫu Ω. Nếu như lực lượng của Ω là hữu hạn ($|\Omega| < \infty$) và các kết quả là đồng khả năng thì

$$P(A) = \frac{\text{số } \textit{biến có sơ cấp có trong } A}{\text{tổng số biến cố sơ cấp trong } \Omega}$$

- Điều kiện ($|\Omega| < \infty$) là để cho phép chia thực hiện được.
- Điều kiện các kết quả đồng khả năng đảm bảo tính đúng đắn của công thức.

Ví dụ

Trong một nhóm gồm 10 người có 5 nam và 5 nữ. Tiến hành chọn lấy 4 người. Tính xác suất để :
a) Có 1 nam 3 nữ.
b) Có ít nhất 2 nam.

Lời giải: Số biến cố sơ cấp khi chọn lấy 4 người là C_{10}^4.
a. Gọi A="có 1 nam 3 nữ". Số biến cố sơ cấp trong A là : $C_5^1.C_5^3$.
Do đó $P(A) = \frac{C_5^1.C_5^3}{C_{10}^4} = \frac{5.10}{210} = \frac{5}{21}$.

b. Gọi B="có ít nhất hai nam", B_0="không có nam trong 4 người". Suy ra $\bar{B} = B_0 \cup A, B_0 \cap A = \emptyset$.
Từ đó ta có : $P(\bar{B}) = P(B_0) + P(A) = \frac{C_5^4}{C_{10}^4} + \frac{5}{21} = 55/210$ và

$$P(B) = 1 - P(\bar{B}) = \frac{155}{210}.$$

Công thức xác suất bằng tần suất

Thực hiện phép thử ngẫu nhiên ξ n lần, k là số lần xuất hiện A trong n phép thử. Kí hiệu

$$f(n) = \frac{k}{n}$$

là tần suất xuất hiện A trong n phép thử. Khi $n \to \infty$ thì f(n) tiến đến một giới hạn không đổi chính là xác suất xuất hiện A. Kí hiệu

$$P(A) = \lim_{n \to \infty} f(n)$$

Xác suất có điều kiện

Định nghĩa xác suất có điều kiện

Cho A , B là hai biến cố. Xác suất để B xảy ra trong điều kiện biết rằng A đã xảy ra được gọi là xác suất có điều kiện của B với điều kiện A và được kí hiệu là $P(B|A)$.

Ví dụ :

Công thức tính xác suất có điều kiện

Cho hai biến cố A và B, trong đó $P(A) \neq 0$. Khi đó :

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

Chú ý:

1. Nếu P(A)=0 thì P(B|A) vẫn tồn tại.
2. Xác suất có điều kiện P(B|A) có thể tính trực tiếp từ bài toán bằng công thức xác suất cổ điển.

Ví dụ

Gieo đồng thời hai con xúc xắc cân đối. Tính xác suất để tổng số chấm xuất hiện trên hai con không nhỏ hơn 10 biết rằng ít nhất một con đã ra 5 chấm.

Lời giải :

- Cách 1: A="ít nhất một con ra 5 chấm", B="Tổng số chấm xuất hiện trên hai con không nhỏ hơn 10", AB="có ít nhất một con ra chấm 5 và tổng số chấm là không nhỏ hơn 10". Khi đó $P(B|A) = \frac{P(AB)}{P(A)} = \frac{3/36}{11/36} = \frac{3}{11}$.
- Cách 2: Sử dụng công thức xác suất cổ điển. Khi A xảy ra tức là các kết quả nhận được là { (1,5); (2,5); (3,5); (4,5);(5,5); (6,5); (5,1); (5,2); (5,3); (5,4); (5,6) }. Số kết quả để B xảy ra là { (5,6); (6,5); (5,5) }. Khi đó theo công thức xác suất cổ điển : $P(B|A) = \frac{3}{11}$.

Quy tắc nhân tổng quát

- Với hai biến cố bất kỳ, ta luôn có: $P(AB) = P(A|B)P(B)$. Khi A, B độc lập thì $P(A|B) = P(A)$ suy ra $P(AB) = P(A)P(B)$.
- Với n biến cố bất kỳ $A_1, \cdots, A_n$, ta có : $P(A_1 \cdots A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1A_2) \cdots P(A_n|A_1 \cdots A_{n-1})$.
- Với n biến cố độc lập $A_1, \cdots, A_n$, ta có: $P(A_1 \cdots A_n) = P(A_1)P(A_2)P(A_3) \cdots P(A_n)$.

Ví dụ

Một thủ kho có chùm chìa khóa gồm 9 chìa có bề ngoài giống hệt nhau trong đó chỉ có 2 chiếc mở được. Anh ta thử ngẫu nhiên từng chìa(chìa nào không đúng thì bỏ ra). Tính xác suất để mở được cửa ở lần thử thứ ba.

Lời giải : A_i=" chọn được chìa đúng lần thử thứ i", i=1,..,9. Khi đó, xác suất để mở được cửa ở lần thử thứ 3 là :
$P(\bar{A}_1\bar{A}_2A_3) = P(\bar{A}_1)P(\bar{A}_2|\bar{A}_1)P(A_3|\bar{A}_1\bar{A}_2)$. Dễ thấy
$P(\bar{A}_1) = \frac{7}{9}; P(\bar{A}_2|\bar{A}_1) = \frac{6}{8}; P(A_3|\bar{A}_1\bar{A}_2) = \frac{2}{7}$. Thay vào ta thu được
$P(\bar{A}_1\bar{A}_2A_3) = \frac{7}{9}.\frac{6}{8}.\frac{2}{7} = \frac{1}{6}$.

Công thức xác suất đầy đủ và công thức Bayes

Hệ đầy đủ

Hệ các biến cố $B_1, \cdots, B_n$ được gọi là hệ đầy đủ các biến cố nếu chúng đôi một xung khắc với nhau và hợp của chúng là biến cố chắc chắn. Nghĩa là :

- $B_iB_j = \emptyset$ với $i \neq j$.
- $\Omega = B_1 \cup B_2 \cup \cdots \cup B_n$.

Công thức xác suất đầy đủ

Nếu $B_1, \cdots, B_n$ là hệ biến cố đầy đủ, thì với mỗi biến cố A ta có

$$P(A) = \sum_{i=1}^{n} P(B_i)P(A|B_i)$$

Chứng minh

Các biến cố $AB_1, \cdots, AB_n$ là các biến cố xung khắc từng đôi một và hợp của chúng là biên cố A. Do đó $P(A) = \sum_{i=1}^{n} P(AB_i)$. Theo quy tắc nhân tổng quát $P(AB_i) = P(A|B_i)P(B_i)$. Thay vào ta được

$$P(A) = \sum_{i=1}^{n} P(B_i)P(A|B_i).$$

Ví dụ

Trong nhà máy có ba phân xưởng A, B,C tương ứng làm ra 25%, 35% và 40% tổng số sản phẩm của nhà máy. Biết rằng, xác suất làm ra sản phẩm hỏng của phân xưởng A,B,C tương ứng là 0,01; 0,02 và 0,025. Chọn ngẫu nhiên một sản phẩm của nhà máy. Tính xác suất để đó là một sản phẩm hỏng.

Lời giải: A="sản phẩm phân xưởng A", B="sản phẩm phân xưởng B", C="sản phẩm phân xưởng C" và H="sản phẩm đó là hỏng". Khi đó A,B,C lập thành hệ đầy đủ với $P(A) = 0,25; P(B) = 0,35; P(C) = 0,4$; và các xác suất có điều kiện từ giả thiết là :
$P(H|A) = 0,01; P(H|B) = 0,02; P(H|C) = 0,025$. Khi đó, áp dụng công thức xác suất đầy đủ ta có:

$$P(H) = P(A)P(H|A) + P(B)P(H|B) + P(C)P(H|C) = 0,0195$$

Vậy xác suất để sản phẩm đó hỏng là 1,95%.

Công thức Bayes

Nếu $B_1, \cdots, B_n$ là hệ biến cố đầy đủ và A là biến cố với P(A)>0 thì với mỗi k=1,2,... ,n

$$P(B_k|A) = \frac{P(B_k)P(A|B_k)}{\sum_{i=1}^{n} P(B_i)P(A|B_i)}$$

Ví dụ

Trong nhà máy có ba phân xưởng A, B,C tương ứng làm ra 25%, 35% và 40% tổng số sản phẩm của nhà máy. Biết rằng, xác suất làm ra sản phẩm hỏng của phân xưởng A,B,C tương ứng là 0,01; 0,02 và 0,025. Chọn ngẫu nhiên một sản phẩm của nhà máy thì nhận được sản phẩm hỏng. Tính xác suất để sản phẩm hỏng đó là của phân xưởng A.

Lời giải: A="sản phẩm phân xưởng A", B="sản phẩm phân xưởng B", C="sản phẩm phân xưởng C" và H="sản phẩm đó là hỏng". Khi đó A,B,C lập thành hệ đầy đủ với
$P(A) = 0,25; P(B) = 0,35; P(C) = 0,4;$ và các xác suất có điều kiện từ giả thiết là :
$P(H|A) = 0,01; P(H|B) = 0,02; P(H|C) = 0,025$. Khi đó, áp dụng công thức xác suất đầy đủ ta có:
$P(H) = P(A)P(H|A) + P(B)P(H|B) + P(C)P(H|C) = 0,0195$.
Sử dụng công thức Bayes ta được
$P(A|H) = \frac{P(H|A)P(A)}{P(H)} = \frac{0{,}01.0{,}25}{0{,}0195} = 0,0128$. Vậy xác suất để sản phẩm hỏng đó là của phân xưởng A là 1,28%.

Ví dụ

Có hai chuồng thỏ: chuồng một có 3 thỏ trắng và 2 thỏ nâu, chuồng hai có 6 thỏ trắng và 4 thỏ nâu. Bắt ngẫu nhiên 4 con thỏ ở chuồng một vào chuồng hai rồi sau đó bắt ngẫu nhiên một con thỏ ở chuồng hai ra được con thỏ nâu. Tính xác suất để con thỏ nâu bắt ra là của chính chuồng hai từ ban đầu chứ không phải là do chuồng một nhảy sang .

Đây là bài toán về công thức dạng Bayes nhưng không phải với biến cố trong hệ đầy đủ mà là biến cố khác. Do đó cách giải quyết bài toán sẽ bao gồm hai bước :

- Bước 1: Ta xác định các biến cố và xác suất cần tìm là xác suất có điều kiện.
- Bước 2: Tính các xác suất trong công thức xác suất có điều kiện bằng công thức xác suất đầy đủ.

Lời giải: Gọi A="bắt được thỏ nâu từ chuồng hai", B="con thỏ nâu bắt ra là của chuồng hai".
Bước 1: Xác suất cần tìm là : $P(B|A) = \frac{P(BA)}{P(A)}$.
Bước 2: Gọi B_i="Bắt được i con thỏ nâu trong 4 con từ chuồng một" (i=1,2). Khi đó B_1, B_2 lập thành hệ đầy đủ với
$P(B_1) = \frac{2}{5}; P(B_2) = \frac{3}{5}$. Ngoài ra ta dễ dàng có:
$P(A|B_1) = \frac{5}{14}; P(A|B_2) = \frac{6}{14}; P(BA|B_1) = \frac{4}{14}; P(BA|B_2) = \frac{4}{14}$.
Khi đó, áp dụng công thức xác suất đầy đủ ta được
$P(A) = P(B_1)P(A|B_1) + P(B_2)P(A|B_2) = 0,4$.
$P(BA) = P(B_1)P(BA|B_1) + P(B_2)P(BA|B_2) = 2/7$.
Từ đó : $P(B|A) = \frac{2/7}{0,4} = 0,7142$.

Phép thử lặp và công thức Bernoulli

Dãy phép thử lặp Bernoulli

Xét phép thử ngẫu nhiên ξ, A là biến cố của phép thử ξ với P(A)=p (0< p < 1). Thực hiện n phép thử ngẫu nhiên ξ một cách độc lập trong các điều kiện giống nhau, khi đó ta gọi dãy phép thử lặp nói trên là dãy n phép thử lặp Bernoulli với tham số p.

Công thức Bernoulli

H_k="xuất hiện k lần A trong dãy n phép thử lặp Bernoulli". Khi đó

$$P(H_k) = C_n^k p^k (1-p)^{n-k}$$

Ví dụ

Xác suất thành công của một thí nghiệm sinh hóa là 40 %. Một nhóm gồm 9 sinh viên tiến hành cùng một thí nghiệm trên độc lập với nhau. Tìm xác suất để :

a) Có đúng 6 thí nghiệm thành công.
b) Có ít nhất một thí nghiệm thành công.
c) Có ít nhất 8 thí nghiệm thành công.

Lời giải:
Phép thử ξ là tiến hành thí nghiệm, A =" thí nghiệm thành công". Khi đó, ta nhận được dãy phép thử lặp Bernoulli với p=P(A)= 0,4 và n=9.

a) $P(H_6) = C_9^6(0,4)^6(0,6^3) = 0.0743.$
b) B="có ít nhất một thí nghiệm thành công". Khi đó $\bar{B} = H_0$ và

$$P(B) = 1 - P(\bar{B}) = 1 - P(H_0) = 1 - (0,6)^9 = 0,9899$$

c) C="có ít nhất 8 thí nghiệm thành công". Khi đó $C = H_8 \cup H_9$ và $H_8 \cap H_9 = \emptyset$, suy ra

$$P(C) = P(H_8) + P(H_9) = C_9^8(0,4)^8(0,6) + (0,4)^9 = 0,0038$$

Đại lượng ngẫu nhiên rời rạc

Khái niệm đại lượng ngẫu nhiên

X là đại lượng(biến) ngẫu nhiên nếu các giá trị của X là không dự báo trước được và khả năng để X nhận được một giá trị nào đó là con số xác suất. Ta sử dụng X, Y, Z để ký hiệu cho các đại lượng ngẫu nhiên(ĐLNN).

Khái niệm đại lượng ngẫu nhiên rời rạc

X là đại lượng(biến) ngẫu nhiên rời rạc nếu tập các giá trị nhận được của X là hữu hạn hoặc đếm được(đánh số được). Tập giá trị của X kí hiệu là $X(\Omega)$.

Các ví dụ

1. Gọi X là biến thể hiện số chấm xuất hiện của con xúc xắc sau khi tung. Khi đó $X(\Omega) = \{1, 2, 3, 4, 5, 6\}$ và P(X=1) =P(X=2)=P(X=3)=P(X=4)=P(X=5)=P(X=6)= 1/6.
2. Thực hiện n phép thử lặp Bernoulli với tham số p=P(A). Gọi X là số lần xuất hiện A trong n phép thử. Khi đó X là ĐLNN rời rạc với $X(\Omega) = \{0, 1, 2, \cdots, n\}$ và $P(X = k) = C_n^k p^k (1-p)^{n-k}$ với k=0,1,...,n .
3. Thực hiện chọn ngẫu nhiên 3 người trong nhóm có 4 nam và 6 nữ. Gọi X là số người nữ trong 3 người. Khi đó, X là ĐLNN rời rạc nhận các giá trị 0,1,2,3 và $P(X = 0) = \frac{C_4^3 C_6^0}{C_{10}^3}$......

Phân bố xác suất và hàm phân bố

Phân bố xác suất

X là đại lượng ngẫu nhiên rời rạc, phân bố xác suất của X là bảng :

X	x_1	x_2	$\cdots$
P	p_1	p_2	$\cdots$

trong đó $X(\Omega) = \{x_1, x_2, \cdots\}$ và $p_i = P(X = x_i)$.

Hàm phân bố

Hàm $F(x) = P(X < x) = \sum_{x_i < x} P(X = x_i)$ với mọi $x \in \mathbb{R}$ được gọi là hàm phân bố xác suất của ĐLNN rời rạc X. Hàm này có đồ thị dạng bậc thang.

Ví dụ

Có hai hộp bi: hộp I có 4 trắng và 2 đỏ, hộp II có 2 trắng và 2 đỏ. Lấy ngẫu nhiên 3 bi từ hộp I sang hộp II. Gọi X là số bi đỏ trong hộp II. Tìm bảng phân bố xác suất của X và hàm phân bố của X.

Lời giải : Ta sẽ có $X(\Omega) = \{2, 3, 4\}$.

Lấy 3 bi trắng sang hộp II: $P(X = 2) = \frac{C_4^3}{C_6^3} = \frac{4}{20}$

Lấy 2 bi trắng 1 đỏ sang hộp II: $P(X = 3) = \frac{C_4^2 C_2^1}{C_6^3} = \frac{12}{20}$

Lấy 1 bi trắng 2 đỏ sang hộp II: $P(X = 4) = \frac{C_4^1 C_2^2}{C_6^3} = \frac{4}{20}$

Do đó bảng phân bố xác suất của X là :

X	2	3	4
P	1/5	3/5	1/5

Hàm phân bố của X là :

$$F(x) = \begin{cases} 0 & \text{nếu } x \leq 2 \\ 1/5 & \text{nếu} 2 < x \leq 3 \\ 4/5 & \text{nếu } 3 < x \leq 4 \\ 1 & \text{nếu} 4 < x \end{cases}$$

Các đại lượng đặc trưng của đại lượng ngẫu nhiên rời rạc

Cho X là biến ngẫu nhiên rời rạc với bảng xác suất sau đây:

X	x_1	x_2	$\cdots$
P	p_1	p_2	$\cdots$

trong đó $p_i = P(X = x_i)$ và $\sum_i p_i = 1$.

Kỳ vọng

Kỳ vọng hay giá trị trung bình X là con số sau đây, ký hiệu bởi EX hoặc μ:

$$\mu = EX = \sum_i x_i p_i$$

Phương sai

Phương sai hay độ phân tán của X so với EX được ký hiệu bởi DX là:

$$DX = E(X - EX)^2 = EX^2 - (EX)^2 = \sum_i x_i^2 p_i - (EX)^2$$

Giá trị $\sigma = \sqrt{DX}$ được gọi là độ lệch tiêu chuẩn của X.

Mode

Giá trị x_{i_0} được gọi là mode của X, kí hiệu là modX nếu nó là giá trị có xác suất lớn nhất, tức là

$$P(X = x_{i_0}) \geq P(X = x_i), \forall i.$$

Moment

Moment cấp k của ĐLNN X là số :

$$m_k = EX^k = \sum_i x_i^k p_i.$$

Số $\alpha_k = E(X - EX)^k$ được gọi là moment quy tâm cấp k.

Hệ số bất đối xứng và hệ số nhọn

Hệ số bất đối xứng của X là : $S = \frac{\alpha_3}{\sigma^3}$.

Hệ số nhọn của X là : $E = \frac{\alpha_4}{\alpha_2^2} - 3$.

Ví dụ

Cho biến ngẫu nhiên rời rạc X với bảng phân bố xác suất:

X	0	1	2
P	1/4	1/2	1/4

Hãy tính các giá trị đặc trưng của X.

Lời giải:
Kỳ vọng: $EX = 0.\frac{1}{4} + 1.\frac{1}{2} + 2.\frac{1}{4} = 1.$
Phương sai: $DX = 0^2.\frac{1}{4} + 1^2.\frac{1}{2} + 2^2.\frac{1}{4} - (1)^2 = \frac{1}{2}.$
Mode: $ModX = 1$ do P(X=1) là lớn nhất.
Moment cấp hai : $EX^2 = 0^2.\frac{1}{4} + 1^2.\frac{1}{2} + 2^2.\frac{1}{4} = \frac{3}{2}.$

Phân bố đồng thời và hệ số tương quan

Phân bố đồng thời

Cho X và Y là hai biến ngẫu nhiên rời rạc với $X(\Omega) = \{x_1; x_2; \cdots, x_m\}$ và $Y(\Omega) = \{y_1; y_2; \cdots; y_n\}$. Khi đó phân bố đồng thời của X và Y được xác định bởi bảng sau:

X \ Y	y_1	...	y_j	...	y_n
x_1	p_{11}	...	p_{1j}	...	p_{1n}
...	...	...	...	...	...
x_i	p_{i1}	...	p_{ij}	...	p_{in}
...	...	...	...	...	...
x_m	p_{m1}	...	p_{mj}	...	p_{mn}

Với $p_{ij} = P(X = x_i; Y = y_j)$ và $\sum p_{ij} = 1$.
Từ bảng phân bố xác suất đồng thời của X và Y, ta nhận được phân bố xác suất của X là: $P(X = x_i) = \sum_{j=1}^{n} p_{ij}, i = 1, 2, \cdots, m$
Phân bố xác suất của Y là: $P(Y = y_j) = \sum_{i=1}^{m} p_{ij}, j = 1, 2, \cdots, n.$
Giá trị đặc trưng của XY là :
$EXY = \sum_{i,j} x_i y_j p_{ij}$

Ví dụ

Tung ba đồng xu A, B, C cân đối đồng chất một cách đồng thời. X là thể hiện số mặt ngửa xuất hiện trên đồng xu A và B, Y là số mặt ngửa trên ba đồng xu. Hãy tìm phân phối đồng thời của X và Y.

Hệ số tương quan

Định nghĩa

X và Y là hai biến ngẫu nhiên rời rạc với phân bố đồng thời $p_{ij} = P(X = x_i; Y = y_j), i = 1, \cdots, m; j = 1, \cdots, n$. Khi đó, hệ số tương quan giữa X và Y là :

$$\rho = \frac{E(X - EX)(Y - EY)}{\sqrt{DXDY}} = \frac{EXY - EXEY}{\sqrt{DXDY}}.$$

Chú ý:

- Ta ký hiệu $Cov(X, Y) = EXY - EXEY$
- ρ đo mức độ phụ thuộc tuyến tính giữa X và Y.Khi đó $|\rho| \leq 1$, với $|\rho| = 1$ thì P(Y=aX +b)=1, với $\rho = 0$ thì X và Y được gọi là **không tương quan**. Khi $|\rho|$ càng gần 1 thì mức độ phụ thuộc tuyến tính giữa chúng càng chặt.

Đại lượng ngẫu nhiên độc lập

Hai đại lượng ngẫu nhiên X và Y được gọi là độc lập nếu việc biết một thông tin về giá trị của X không có ảnh hưởng gì đến phân bố xác suất của Y và ngược lại. Nói cách khác :

$$P(Y = y_j | X = x_i) = P(Y = y_j)$$
$$P(X = x_i | Y = y_j) = P(X = x_i)$$

hay tương đương với :

$$P(X = x_i; Y = y_j) = P(X = x_i)P(Y = y_j).$$

Ví dụ

Cho phân bố đồng thời của X, Y như sau:

X \ Y	1	2	3
1	0.12	0.15	0.03
2	0.28	0.35	0.07

Kiểm tra xem X và Y có độc lập không?

Quan hệ giữa độc lập và không tương quan

Hai ĐLNN X và Y độc lập thì X và Y là không tương quan.

Nhận xét :Chứng minh nhận xét này rất dễ dàng. Điều ngược lại là không đúng. Xem phản ví dụ:

X \ Y	-1	0	1
-1	1/5	2/15	1/5
1	2/15	1/5	2/15

Từ bảng trên ta có được
$EXY = (-1)(-1)\frac{1}{5} + (-1)(1)\frac{1}{5} + (-1)(1)\frac{2}{15} + (1)(1)\frac{2}{15} = 0.$
P(X=-1)=8/15; P(X=1)=7/15, suy ra EX =-1/15.
P(Y=-1)=P(Y=1)=1/3; P(Y=0)=1/3, suy ra EY = 0.
Từ đó ta có cov(X,Y)=EXY - EX EY =0. Suy ra $\rho = 0$ hay là X và Y không tương quan.
Nhưng P(X=1)P(Y=1)=7/45 $\neq$ 2/15=P(X=1,Y=1). Suy ra X và Y không độc lập. Như vậy phản ví dụ này cho ta kết quả X và Y không độc lập nhưng X và Y không tương quan.

Hàm của ĐLNN

Hàm của ĐLNN

Cho X và Y là ĐLNN rời rạc với phân bố xác suất đồng thời như bảng trên. ĐLNN Z=g(X,Y) là hàm của ĐLNN X và Y với phân bố sau:

- $Z(\Omega) = \{g(x_i, y_j) : i = \overline{1, m}; j = \overline{1, n}\} = \{z_1, \cdots, z_s\}$
- $P(Z = z_k) = \sum_{(i,j):g(x_i,y_j)=z_k} p_{ij}, k = 1, \cdots, s.$

Kỳ vọng của hàm ĐLNN Z=g(X,Y)

$$EZ = \sum_{i=1}^{m} \sum_{j=1}^{n} g(x_i, y_j) P(X = x_i; Y = y_j)$$

Các kết quả quan trọng

X, Y là các ĐLNN, C là hằng số. Khi đó :

1. E(C)=C , D(C)=0.
2. E(CX) =C EX, $D(CX) = C^2 DX$
3. E(X+Y)=EX + EY, D(X+Y)=DX - 2 cov(X,Y) +DY.
4. X và Y độc lập thì D(X+Y)=DX + DY.

Ví dụ

Cho ĐLNN X, Y có phân bố xác suất đồng thời như sau:

X \ Y	-1	0	1
-1	1/5	2/15	1/5
1	2/15	1/5	2/15

Tìm phân bố của Z =X+Y.

Lời giải : Z=X+Y, khi đó :
Z(Ω)={ -2,-1,0,1,2 }.
P(Z=-2)=P(X=-1, Y=-1)=1/5; P(Z=-1)=P(X=-1,Y=0)=2/15;
P(Z=0)=P(X=-1,Y=1)+P(X=1,Y=-1)=1/3;
P(Z=1)=P(X=1,Y=0)=1/5;P(Z=2)=P(X=1,Y=1)=2/15.
Viết gọn lại:

Z	-2	-1	0	1	2
P	1/5	2/15	1/3	1/5	2/15

Phân bố nhị thức

Định nghĩa

ĐLNN rời rạc X được gọi là có phân bố nhị thức với tham số p (0<p<1) và n, kí hiệu là $X \sim B(n,p)$, nếu:

- $X(\Omega) = \{0, 1, \cdots, n\}$
- $P(X = k) = C_n^k p^k (1-p)^{n-k}$ với $k = 0, 1, \cdots, n$.

Các đại lượng đặc trưng của phân bố nhị thức

- EX = np
- DX = np(1-p)
- ModX = [(n+1)p]

Các ví dụ

- Thực hiện dãy n phép thử lặp Bernoulli với tham số p=P(A) (0<p<1). X là ĐLNN thể hiện số lần xuất hiện A trong n phép thử. Khi đó $X \sim B(n,p)$.
- Tỷ lệ phế phẩm trong một lô sản phẩm là 3%. Lấy ngẫu nhiên 100 sản phẩm để kiểm tra. Hãy tính xác suất để :
  a) Có 3 phế phẩm.
  b) Có ít nhất 2 phế phẩm.

Lời giải : Gọi X là số phế phẩm trong 100 sản phẩm. Khi đó $X \sim B(100, 0.03)$. Suy ra,
a) $P(X = 3) = C_{100}^{3}(0.03)^3 0.97^{97} = 0.2247$
b) $P(X \geq 2) = 1 - P(X < 2) = 1 - (0.97)^{100} - 100(0.03)(0.97)^{99}$.

Phân bố Poisson

Định nghĩa

ĐLNN rời rạc X được gọi là có phân bố Poisson với tham số $\lambda > 0$, kí hiệu là $X \sim Poisson(\lambda)$, nếu:

- $X(\Omega) = \{0, 1, \cdots\}$
- $P(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}$ với $k = 0, 1, \cdots$ và quy ước $0! = 1$.

Các đại lượng đặc trưng của phân bố Poisson

- EX $= \lambda$.
- DX$= \lambda$.
- ModX $= [\lambda]$.

Ví dụ

X là ĐLNN có phân bố Poisson với tham số $\lambda = 2$.
a) Tính các đại lượng đặc trưng của X.
b) Tính $P(X \geq 1)$.

Lời giải :
a) ĐLNN X có phân bố Poisson với tham số 2, suy ra EX =DX = modX =2.
b) $P(X \geq 1) = 1 - P(X = 0) = 1 - e^{-2}$.

Chú ý:

- ĐLNN thể hiện số hiện tượng xuất hiện trong khoảng thời gian (t_1, t_2) có phân bố Poisson với tham số $\lambda > 0$ được xác định như sau: $\lambda = c(t_2 - t_1)$, trong đó c là cường độ xuất hiện hiện tượng. Từ công thức xác định λ ta thấy rằng tham số chỉ phụ thuộc vào độ dài khoảng thời gian chứ không phụ thuộc vào mốc thời gian.
- Tổng của hai ĐLNN độc lập có phân bố Poisson với hai tham số λ_1, λ_2 là một ĐLNN có phân bố Poisson với tham số $\lambda_1 + \lambda_2$.

Ví dụ : Trong một trạm điện thoại, trung bình có 5 cuộc điện gọi đi trong vòng 5 phút. Khi đó, hãy tính xác suất
a) Có chính xác 3 cuộc điện thoại gọi đi trong vòng 2 phút.
b) Có ít nhất hai cuộc điện thoại gọi đi trong vòng 1 phút.

Lời giải: Giả thiết bài cho ta sẽ có được cường độ xuất hiện cuộc điện gọi đi là c= 1 cuộc/ phút.

a) Gọi X là đại lượng thể hiện số cuộc điện thoại gọi đi trong vòng 2 phút. Khi đó X là ĐLNN có phân bố Poisson với tham số $\lambda = c.2 = 2$. Do đó, xác suất để có chính xác 3 cuộc điện thoại gọi đi trong vòng 2 phút là : $P(X = 3) = e^{-2}\frac{2^3}{3!}$.

b) Gọi Y là đại lượng thể hiện số cuộc điện thoại gọi đi trong vòng 1 phút. Khi đó Y là ĐLNN có phân bố Poisson với tham số $\lambda = c.1 = 1$. Do đó, xác suất để có ít nhất hai cuộc điện thoại gọi đi trong vòng 1 phút là :

$P(X \geq 2) = 1 - P(X = 0) - P(X = 1) = 1 - \frac{2}{e}$.

Đại lượng ngẫu nhiên(ĐLNN) liên tục

Khái niệm

X là đại lượng(biến) ngẫu nhiên liên tục nếu $P(X = a) = 0$ và tập các giá trị nhận được của X là khoảng.

Ví dụ

Gọi X là biến thể hiện chiều cao của con người trưởng thành. Khi đó giá trị của X nhận được sẽ là cả khoảng (1,3). Việc xuất hiện đúng một người có chiều cao là $a \in (1, 3)$ là không có, hay P(X=a)=0. Do đó X là đại lượng (biến) ngẫu nhiên liên tục

Hàm mật độ và hàm phân bố xác suất

Hàm mật độ xác suất

Hàm số f(x) xác định trên $\mathbb{R}$ được gọi là hàm mật độ xác suất của biến ngẫu nhiên liên tục X nếu như thỏa mãn :

- $f(x) \geq 0$
- $\int_{-\infty}^{\infty} f(x)dx = 1.$

Khi đó, với mọi $a, b \in \mathbb{R}$ thì $P(a < X < b) = \int_a^b f(x)dx.$

Hàm phân bố xác suất

Hàm $F(x) = P(X < x) = \int_{-\infty}^{x} f(t)dt$ với mọi $x \in \mathbb{R}$ được gọi là hàm phân bố (phân phối) xác suất của X

Ví dụ

Cho biến ngẫu nhiên liên tục X có hàm mật độ

$$f(x) = \begin{cases} cx^3 & \text{nếu } 0 \le x \le 3 \\ 0 & \text{nếu } x \notin [0,3] \end{cases}$$

a) Tìm c.
b) Tìm P(1<X<2).
c) Tìm hàm phân bố của X.

Lời giải :

a) Do $\int_{-\infty}^{\infty} f(x)dx = 1$, suy ra $\int_0^3 cx^3 dx = 1$. Từ đó $\left.\frac{cx^4}{4}\right|_0^3 = 1$, suy ra c=4/81.

b) Tính xác suất $P(1 < X < 2) = \int_1^2 cx^3 dx = \left.\frac{cx^4}{4}\right|_1^2 = \frac{15c}{4} = \frac{15}{81}$

c) Hàm phân bố $F(x) = P(X < x) = \int_{-\infty}^{x} f(t)dt$

- Nếu $x \le 0$ thì F(x)=0
- Nếu $0 < x < 3$ thì
  $F(x) = \int_{-\infty}^{0} f(t)dt + \int_0^x f(t)dt = 0 + \int_0^x ct^3 dt = x^4/81$
- Nếu $x \ge 3$ thì
  $F(x) = \int_{-\infty}^{0} f(t)dt + \int_0^3 f(t)dt + \int_3^x f(t)dt = 0 + 1 + 0 = 1$

Vậy

$$F(x) = \begin{cases} 0 & \text{nếu } x \le 0 \\ 15x^4/81 & \text{nếu } 0 < x < 3 \\ 1 & \text{nếu } x \ge 3 \end{cases}$$

Tính chất

- $0 \leq F(x) \leq 1$.
- $P(a \leq X < b) = F(b) - F(a)$.
- F(x) là hàm không giảm liên tục trái và có giới hạn từ bên phải.
- $\lim_{x \to +\infty} F(x) = 1, \lim_{x \to -\infty} F(x) = 0$. Để ngắn gọn ta viết $F(-\infty) = 0; F(+\infty) = 1$.
- Nếu f(x) là hàm liên tục thì F(x) là khả vi liên tục và $F(x)' = f(x)$.

Các đặc trưng của ĐLNN liên tục

Cho X là ĐLNN liên tục với hàm mật độ f(x). Khi đó các đại lượng đặc trưng của X là :

1. Kỳ vọng : $\mu = EX = \int_{-\infty}^{+\infty} xf(x)dx$.
2. Phương sai:
   $DX = E(X - EX)^2 = EX^2 - (EX)^2 = \int_{-\infty}^{+\infty} x^2 f(x)dx - \mu^2$.
3. Độ lệch tiêu chuẩn: $\sigma = \sqrt{DX}$.
4. Mode: ModX $=$ c với $f(c) = max_x f(x)$.
5. Median (Trung vị): MedianX=m nếu P(X<m)=P(X>m) (hoặc F(m)=1/2 với F(x) liên tục).
6. Moment cấp k: $EX^k = \int_{-\infty}^{+\infty} x^k f(x)dx$.

Ví dụ

Cho ĐLNN liên tục X với hàm mật độ

$$f(x) = \begin{cases} \frac{4}{81}x^3 & \text{với } 0 \le x \le 3 \\ 0 & \text{với } x \notin [0,3]. \end{cases}$$

Tính các đại lượng đặc trưng của X?

Lời giải:

1. Kỳ vọng : $\mu = EX = \int_{-\infty}^{+\infty} xf(x)dx = \int_0^3 x\frac{4}{81}x^3dx = \frac{12}{5}$.

2. Phương sai: $EX^2 = \int_{-\infty}^{+\infty} x^2f(x)dx = \int_0^3 x^2\frac{4}{81}x^3dx = 6$.

Suy ra $DX = EX^2 - \mu^2 = \frac{6}{25}$ và $\sigma = \sqrt{\frac{6}{25}} = \frac{\sqrt{6}}{5}$.

3. Mode : Với $x \notin [0,3]$ thì f(x)=0, suy ra giá trị lớn nhất hàm f đạt được nằm trên [0,3].
Mặt khác, f(0)=0, $f(3) = \frac{4}{3}$ và hàm f(x) là hàm tăng từ 0 tới 3.
Suy ra giá trị lớn nhất của hàm f đạt được tại x=3.
Do đó ModX =3.
4. Median: Giả sử m=MedianX. Khi đó :

$$P(X < m) = \frac{1}{2} \Leftrightarrow \int_{-\infty}^{m} f(x)dx = \frac{1}{2}$$

$$\Leftrightarrow \begin{cases} m \in [0,3] \\ \int_0^m \frac{4}{81}x^3 dx = \frac{1}{2} \end{cases}$$

Giải phương trình thứ hai ta được $\frac{m^4}{81} = \frac{1}{2} \Leftrightarrow m = \sqrt[4]{\frac{81}{2}}$. Giá trị này thỏa mãn điều kiện.
Do đó $MedianX = \sqrt[4]{\frac{81}{2}}$.

Hàm của ĐLNN

Định nghĩa

Giả sử X là một ĐLNN và g(x) là một hàm đã cho trước. Xét ĐLNN Y xác định bởi

$$Y = g(X)$$

Khi đó Y là hàm của ĐLNN X.

Câu hỏi: Cho hàm mật độ (phân bố) của X, tìm hàm mật độ (phân bố) của Y?

Ví dụ

X là ĐLNN liên tục có hàm phân bố

$$F_X(x) = \begin{cases} 0 & \text{nếu } x \leq 0 \\ \frac{x}{2} & \text{nếu } 0<x<2 \\ 1 & \text{nếu } x \geq 2 \end{cases}$$

Tìm hàm phân bố của Y=2X +1?

Lời giải: Trước hết ta có:

$$P(Y < y) = P(2X + 1 < y) = P(2X < y - 1)$$
$$= P(X < \frac{y-1}{2}) = F_X(\frac{y-1}{2}).$$

Ta sẽ xét các trường hợp của y:

1. Nếu $\frac{y-1}{2} \leq 0$ thì P(Y<y)=0.
2. Nếu $0 < \frac{y-1}{2} < 2$ thì $P(Y < y) = \frac{y-1}{4}$.
3. Nếu $2 \leq \frac{y-1}{2}$ thì P(Y<y)=1.

Kết luận : Hàm phân bố của Y là

$$F_Y(y) = \begin{cases} 0 & \text{nếu } y \leq 1 \\ \frac{y-1}{4} & \text{nếu } 1<y<5 \\ 1 & \text{nếu } y \geq 5 \end{cases}$$

Ví dụ

X là ĐLNN liên tục có hàm mật độ

$$f(x) = \begin{cases} 0 & \text{nếu } x \notin [0,2] \\ \frac{1}{2} & \text{nếu } x \in [0,2] \end{cases}$$

Tìm hàm phân bố của Y=2X +1?

Lời giải:
Trước tiên ta tìm hàm phân bố của X như sau:

$$F_X(x) = P(X < x) = \int_{-\infty}^{x} f(t)dt \quad .$$

Sau đó ta thực hiện như ví dụ trên.

Phân phối chuẩn

Phân phối chuẩn tắc

ĐLNN liên tục Z có phân bố chuẩn tắc, kí hiệu $Z \sim N(0,1)$, nếu hàm mật độ của X có dạng:

$$f(x) = \frac{1}{\sqrt{2\pi}} exp\{-\frac{x^2}{2}\}.$$

Hàm phân bố của ĐLNN Z là $\Phi(x) = P(Z < x) = \int_{-\infty}^{x} f(t)dt$.

Một số kết quả quan trọng của Z

1. EZ=0; DZ=1; ModZ=0; MedianZ=0.
2. $\Phi(x) + \Phi(-x) = 1$. Hàm Φ không thể biểu diễn được qua các hàm sơ cấp nhưng ta có thể tính được các giá trị hàm Φ .

Phân phối chuẩn

ĐLNN liên tục X gọi là ĐLNN có phân phối chuẩn với tham số μ và $\sigma > 0$, kí hiệu $X \sim N(\mu; \sigma^2)$, nếu hàm mật độ của X có dạng :

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} exp\{-\frac{(x-\mu)^2}{2\sigma^2}\}.$$

Các kết quả quan trọng

1. $EX = \mu; DX = \sigma^2; ModX = MedianX = \mu.$
2. Khi $\mu = 0; \sigma = 1$ thì phân bố chuẩn chính là phân bố chuẩn tắc.
3. $X = \mu + \sigma Z.$
4. $P(X < a) = \Phi(\frac{a-\mu}{\sigma}); P(a < X < b) = \Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma}).$

Ví dụ

X là ĐLNN có phân bố chuẩn với kì vọng $\mu = 2100$ và độ lệch tiêu chuẩn $\sigma = 200$. Hãy tìm :
a. P(X>2400).
b. P(1700<X<2200).
c. Tìm a để P(X>a)=0.03

Ví dụ

Trọng lượng của một gói đường (đóng bằng máy tự động) có phân bố chuẩn. Trong 1000 gói đường có 70 gói có trọng lượng lớn hơn 1015g. Hãy ước lượng xem có bao nhiêu gói đường có trọng lượng ít hơn 1008g, biết rằng trọng lượng trung bình của 1000 gói đường là 1012g.

Phân phối Mũ

ĐLNN X được gọi là có phân bố mũ với tham số $\lambda > 0$ nếu nó có hàm mật độ như sau

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{nếu x>0} \\ 0 & \text{nếu x<0} \end{cases}$$

Hàm phân bố của X là :

$$F(x) = \begin{cases} 1 - e^{-\lambda x} & \text{nếu x>0} \\ 0 & \text{nếu } x \leq 0 \end{cases}$$

Các giá trị đặc trưng:

$$EX = \frac{1}{\lambda}; DX = \frac{1}{\lambda^2}.$$

Phân phối Đều

ĐLNN liên tục X có phân bố đều trên đoạn [a,b] nếu X có thể nhận biết bất kì giá trị nào trên [a,b] với xác suất như nhau và không nhận giá trị nào bên ngoài [a,b].

Hàm mật độ của X là :

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{nếu } x \in [a,b] \\ 0 & \text{nếu } x \notin [a,b] \end{cases}$$

Hàm phân bố của X là :

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

Giá trị đặc trưng của X: $EX = \frac{a+b}{2}$; $DX = \frac{(b-a)^2}{12}$; $MedianX = \frac{a+b}{2}$.

Một số phân phối khác

Phân bố student

ĐLNN liên tục X gọi là có phân bố student với n bậc tự do nếu hàm mật độ của nó có dạng

$$f(x) = c(1 + \frac{x^2}{n})^{\frac{n+1}{2}}.$$

Phân phối χ^2 (Khi bình phương)

ĐLNN liên tục X gọi là có phân bố χ^2 với n bậc tự do nếu hàm mật độ của nó có dạng

$$f(x) = \begin{cases} 0 & \text{nếu } x \leq 0 \\ cx^{\frac{n}{2}-1}e^{-\frac{x}{2}} & \text{nếu x>0} \end{cases}$$

Trong hai phân bố trên c là hằng số chuẩn hóa.

Phần II

Thống kê ứng dụng