PHƯƠNG PHÁP TOÁN
TRONG MACHINE LEARNING

Chương 3. Giải tích Ma trận

Chương này bắt đầu với một số một số vấn đề ứng dụng của đạo hàm của hàm số một biến và nhiều biến số. Phần tiếp theo là phép tính vi phân đối với dạng hàm mà biến đầu vào có thể là véc tơ hay nói chung là ma trận, còn giá trị đầu ra có thể là số hoặc véc tơ.

3.1 Xấp xỉ hàm số một biến số

3.1.1. Đạo hàm cấp cao: Xét hàm số $y = f(x)$

+ Đạo hàm (đạo hàm cấp 1) : $y' = f'(x) = \lim\limits_{t\to x} \frac{f(t)-f(x)}{t-x}$

+ Đạo hàm cấp n : $y^{(n)} = f^{(n)}(x) = \left\{f^{(\boldsymbol{n-1})}(x)\right\}'$ (quy ước: $f^{(0)}(x) = f(x)$)

Ví dụ 1. Với $f(x) = \sin x : f'(x) = (\sin x)' = \cos x = \sin\left(x + \frac{\pi}{2}\right)$;

$$f^{(2)}(x) = (\sin x)^{(2)} = \left(\sin(x + \frac{\pi}{2})\right)' = \sin\left(x + 2.\frac{\pi}{2}\right)$$

$$f^{(n)}(x) = (\sin x)^{(n)} = \sin\left(x + n.\frac{\pi}{2}\right)$$

3.1.2. Đạo hàm cấp cao của một số hàm cơ bản:

+ $(\sin x)^{(n)} = \sin\left(x + n.\frac{\pi}{2}\right)$, $\quad (\cos x)^{(n)} = \cos\left(x + n.\frac{\pi}{2}\right)$, $\quad \forall n = 0,1,2,\ldots$

+ $\left(x^k\right)^{(n)} = \begin{cases} k(k-1)\ldots(k-n-1).x^{k-n}, \text{nếu } n \le k \\ 0, \text{nếu } n > k \end{cases}$, $\quad \forall n = 0,1,2,\ldots$

+ $\left(a^{b.x}\right)^{(n)} = (b.\ln a)^n.a^{b.x}$, $\left(\frac{1}{x+a}\right)^{(n)} = (-1)^n.\frac{n!}{(x+a)^{n+1}}$, $\quad \forall n = 0,1,2,\ldots$

+ $(\ln(x+a))^{(n)} = (-1)^{n-1}.\frac{(n-1)!}{(x+a)^n}, \forall n = 1,2,\ldots$

3.1.3. Công thức Taylor: Tồn tại điểm c bao hàm giữa x và t sao cho:

$$f(x) = f(t) + \frac{f'(t)}{1!}(x-t) + \frac{f''(t)}{2!}.(x-t)^2 + \cdots + \frac{f^{(n)}(t)}{n!}.(x-t)^n + \frac{f^{(n+1)}(c)}{(n+1)!}.(x-t)^{n+1} \quad \textbf{(3.1)}$$

Chú ý:

+ (3.1) gọi là *công thức khai triển Taylor của hàm $f(x)$ tại t đến cấp n*

Biểu thức $R_n(x) = \frac{f^{(n+1)}(c)}{(n+1)!}.(x-t)^{n+1}$ gọi là *phần dư thứ n của khai triển Taylor*

+ Với t = 0, (3.1) trở thành:

$$f(x) = f(0) + \frac{f'(0)}{1!}.x + \frac{f''(0)}{2!}.x^2 + \cdots + \frac{f^{(n)}(0)}{n!}.x^n + \frac{f^{(n+1)}(c)}{(n+1)!}.x^{n+1} \quad (3.2)$$

(3.2) gọi là công thức khai triển Mac Laurin.

3.1.4 Xấp xỉ hàm số bởi đa thức

Khi hàm $f(x)$ có các đạo hàm bị chặn thì $R_n(x) = o\left((x-t)^{n+1}\right)\ (n \to +\infty)$, tức là phần dư sẽ dần về 0 (vô cùng bé) khi $n \to +\infty$. Khi đó từ (3.1) có công thức xấp xỷ:

$$f(x) \approx f(t) + \frac{f'(t)}{1!}(x-t) + \frac{f''(t)}{2!}.(x-t)^2 + \cdots + \frac{f^{(n)}(t)}{n!}.(x-t)^n \quad (3.3)$$

+ Với n = 1, khi x khá gần t, ta có công thức xấp xỷ:

$$f(x) \approx f(t) + \frac{f'(t)}{1!}(x-t), \text{ hay: } f'(t) \approx \frac{f(x)-f(t)}{x-t} \quad (3.4)$$

Khi đó (3.2) trở thành:

$$f(x) \approx f(0) + \frac{f'(0)}{1!}.x + \frac{f''(0)}{2!}.x^2 + \cdots + \frac{f^{(n)}(0)}{n!}.x^n \quad (3.5)$$

+ Các công thức (3.3), (3.5) là các công thức tính xấp xỷ giá trị hàm $f(x)$ bởi đa thức.

Ví dụ 2. Khi n đủ lớn, từ ((3.5) ta có: $e^x \approx 1 + x + \frac{1}{2!}.x^2 + \cdots + \frac{1}{n!}.x^n$

Đặc biệt: $e \approx 1 + 1 + \frac{1}{2!} + \cdots + \frac{1}{n!}$

3.2. Hàm nhiều biến số và đạo hàm riêng của hàm nhiều biến số

3.2.1 Khái niệm về hàm nhiều biến số

Trong thực tế, thường gặp một đại lượng Y phụ thuộc vào nhiều đại lượng khác nhau $\mathrm{X}_1, \mathrm{X}_2, \dots, \mathrm{X}_\mathrm{n}$ mà với một giá trị đã cho của $X_1, X_2, \dots, X_n$ có một và chỉ một giá trị tương ứng của Y. Ta nói Y là một hàm của các biến $X_1, X_2, \dots, X_n$ (hàm n biến).

- Theo ngôn ngữ toán học: một hàm n biến số là một ánh xạ

$$\mathrm{f}: \mathrm{D} \ni \mathrm{x} = (\mathrm{x}_1, \mathrm{x}_2, \dots, \mathrm{x}_n) \to \mathrm{y} = \mathrm{f}(\mathrm{x}_1, \mathrm{x}_2, \dots, \mathrm{x}_n) \in \mathbb{R}\ (\mathrm{D} \subset \mathbb{R}^\mathrm{n}, \mathrm{D}\text{ gọi là tập xác định })$$

Ví dụ 3: Với hình chữ nhật có số đo chiều rộng là x, số đo chiều dài là y thì khi chiều rộng và chiều dài thay đổi, diện tích S = x . y là hàm của 2 biến x, y. Với hình hộp chữ nhật có các kích thước x, y, z thì khi x, y, z thay đổi thể tích V = x . y . z là hàm của 3 biến x, y, z.

Ví dụ 4: Với ma trận vuông $\mathrm{A} = \left(\mathrm{x}_{\mathrm{ij}}\right)_{\mathrm{n}\times\mathrm{n}}$ thì

+ $\det(\mathrm{A})$ là hàm của n^2 biến $\mathrm{x}_{\mathrm{ij}}, \mathrm{i} = \overline{1,\mathrm{n}}, \mathrm{j} = \overline{1,\mathrm{n}}$

+ $\mathrm{trace}(\mathrm{A}) = \mathrm{x}_{11} + \mathrm{x}_{22} + \cdots + \mathrm{x}_{\mathrm{nn}}$ là hàm của n biến $\mathrm{x}_{11}, \mathrm{x}_{22}, \dots, \mathrm{x}_{\mathrm{nn}}$

+ Các chuẩn, các khoảng cách trong không gian R^n là các hàm n biến.

3.2.3. Đạo hàm riêng của hàm nhiều biến. Cho hàm n biến: $\mathrm{u} = \mathrm{f}(\mathrm{x}_1, \mathrm{x}_2, \dots, \mathrm{x}_n)$

- Nếu chỉ cho x_i thay đổi còn các biến khác cố định (hằng số) thì u chỉ là hàm của một biến x_i và đạo hàm của hàm u với biến là x_i được ký hiệu là $\frac{\partial \mathrm{u}}{\partial \mathrm{x}_\mathrm{i}}$ (hoặc $\mathrm{u}'_{\mathrm{x}_\mathrm{i}}$) và gọi là đạo hàm riêng (cấp 1) của hàm $\mathrm{u} = \mathrm{f}(\mathrm{x}_1, \mathrm{x}_2, \dots, \mathrm{x}_n)$ theo biến x_i. Khi đó $\frac{\partial \mathrm{u}}{\partial \mathrm{x}_\mathrm{i}}$ cũng là hàm n biến.

- Đạo hàm riêng của $\frac{\partial \mathrm{u}}{\partial \mathrm{x}_\mathrm{i}}$ theo biến x_j là $\frac{\partial}{\partial \mathrm{x}_\mathrm{j}}\left(\frac{\partial \mathrm{u}}{\partial \mathrm{x}_\mathrm{i}}\right)$ được gọi là đạo hàm riêng cấp 2 của hàm u lần lượt theo các biến x_i và x_j và được kí hiệu là: $\frac{\partial^2 \mathrm{u}}{\partial \mathrm{x}_\mathrm{i} \partial \mathrm{x}_\mathrm{j}}$. Đặc biệt kí hiệu : $\frac{\partial^2 \mathrm{u}}{\partial \mathrm{x}_\mathrm{i}^2} = \frac{\partial^2 \mathrm{u}}{\partial \mathrm{x}_\mathrm{i} \partial \mathrm{x}_\mathrm{i}}$ là đạo hàm cấp 2 liên tiếp 2 lần theo cùng một biến x_i và lưu ý đ.với các hàm thường gặp thì $\frac{\partial^2 \mathrm{u}}{\partial \mathrm{x}_\mathrm{j} \partial \mathrm{x}_\mathrm{i}} = \frac{\partial^2 \mathrm{u}}{\partial \mathrm{x}_\mathrm{i} \partial \mathrm{x}_\mathrm{j}}$

Ví dụ 5: Tính các đạo hàm riêng cấp 2 của hàm $u = 2x.e^{-yz} - x^2.z$, ta có:

+ Các đạo hàm riêng cấp 1: $\frac{\partial u}{\partial x} = 2e^{-yz} - 2xz; \frac{\partial u}{\partial y} = -2xze^{-yz};\ \frac{\partial u}{\partial z} = -2xye^{-yz} - x^2$ + Các đạo hàm riêng cấp 2: $\frac{\partial^2 u}{\partial x \partial y} = \frac{\partial}{\partial y}(2e^{-yz} - 2xz) = -2z.e^{-yz}; \frac{\partial^2 u}{\partial x^2} = \frac{\partial}{\partial x}(2e^{-yz} - 2xz) = -2z, \ldots$

3.2.4. Cực trị của hàm nhiều biến

Cho hàm n biến: $u = f(x_1, x_2, \ldots, x_n)$ có tập xác định D. Nói rằng hàm u đạt cực đại (cực tiểu) tại $a \in D$, nếu ∃ hình cầu mở $B(r,a) \subset D$ sao cho: $f(x) \le f(a)\ \big(f(x) \ge f(a)\big), \forall x \in B(r,a)$

a. Điều kiện cần của cực trị: Giả sử hàm u = $f(x)$ đạt cực trị tại a. khi đó nếu $f(x)$ có tất cả các đạo hàm riêng tại a thì $\frac{\partial u}{\partial x_i} = 0, \forall i = 1,2, \ldots, n$

b. Điều kiện đủ của cực trị: Với hàm $u = f(x_1, x_2, \ldots, x_n)$ có các đạo hàm riêng cấp 2, ma trận:

$H_u(a) = \left(\frac{\partial^2 u(a)}{\partial x_i \partial x_i}\right)_{n \times n}$ gọi là ma trận Hessian của hàm u tại điểm a

Mệnh đề: Giả sử a là nghiệm của hệ ph.trình : $\frac{\partial u}{\partial x_i} = 0, \forall i = \overline{1,n}$. Khi đó gọi a là điểm dừng của u và:

(1) Nếu ma trận Hessian $H_u(a)$ là ma trận xác định dương thì u đạt cực tiểu tại a

(2) Nếu ma trận Hessian $H_u(a)$ bán xác định dương, hoặc bán xác định âm nhưng $\exists b \in D, b \neq 0$ sao cho: $(b).H_u(a).[b] = 0$, thì chưa thể kết luận cực trị tại a

(3) Nếu $\exists b, c \in D$ mà: $(b).H_u(a).[b] > 0$ và: $(c).H_u(a).[c] < 0$ thì u không đạt cực trị tại a.

Hệ quả (cực trị hàm 2 biến): Cho hàm 2 biến $z = f(x,y)$ có các đạo hàm riêng cấp 2.

Đặt: $A = \frac{\partial^2 u}{\partial x^2}$, $B = \frac{\partial^2 u}{\partial x \partial y}$, $C = \frac{\partial^2 u}{\partial y^2}, \Delta = A.C - B^2$

Xét hệ phương trình: $\begin{cases} \frac{\partial u}{\partial x} = 0 \\ \frac{\partial u}{\partial y} = 0 \end{cases}$ (*)

Khi đó:
(a) Nếu hệ (*) vô nghiệm thì hàm u không có cực trị
(b) Khi (*) có nghiệm a:
- Nếu tại a có: $\Delta(\mathrm{a}) < 0$, thì u không đạt cực trị tại a
- Nếu tại a có: $\Delta(\mathrm{a}) > 0$, thì u đạt cực trị tại a, cụ thể: với A(a) > 0 thì u đạt cực tiểu tại b, với A(a) < 0 thì u đạt cực đại tại a.
- Nếu $\Delta(\mathrm{a}) = 0$ thì chưa thể kết luận về cực trị tại a

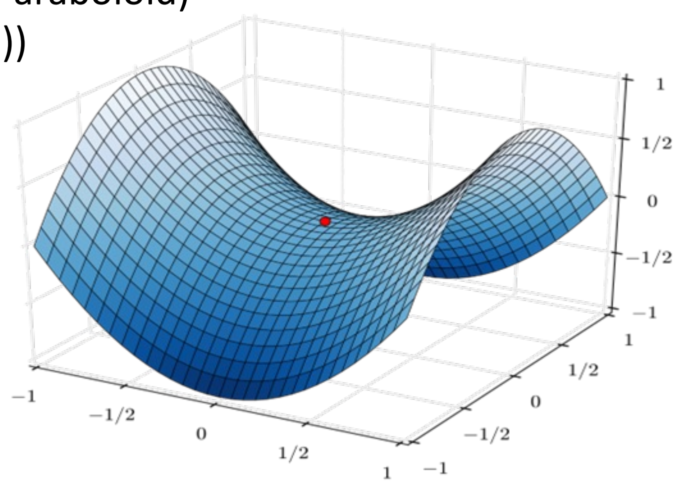
Ví dụ 5: Tìm cực trị của hàm $\mathrm{u} = \mathrm{x}^2 + \mathrm{y}^2 + \mathrm{z}^2 - 2\mathrm{z}$

Dễ thấy điểm dừng là a = (0, 0, 1) duy nhất và ma trận Hessian $\mathrm{H_u}(\mathrm{a}) = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$ là ma trận xác định dương. Vậy hàm u đạt cực trị duy nhất (và là cực tiểu) tại a = (0, 0, 1) và $\mathrm{u}_{\text{cực tiểu}} = -2$

Ví dụ 6: Tìm cực trị của hàm $\mathrm{z} = \mathrm{x}^2 - \mathrm{y}^2$(hàm Hyperbolic - Paraboloid)
(Đồ thi của hàm này còn được gọi là mặt yên ngựa (hình bên))
- Dễ thấy điểm dừng duy nhất của hàm này là a = (0, 0)
- Tính $\mathrm{A} = \dfrac{\partial^2 \mathrm{z}}{\partial \mathrm{x}^2} = 2, \quad \mathrm{B} = \dfrac{\partial^2 \mathrm{z}}{\partial \mathrm{x} \partial \mathrm{y}} = 0$,

$$\mathrm{C} = \frac{\partial^2 \mathrm{z}}{\partial \mathrm{y}^2} = -2, \Delta = \mathrm{A}.\mathrm{C} - \mathrm{B}^2 = -4 < 0$$

Vậy hàm này không có cực trị. Điểm dừng (0, 0) không phải là điểm cực trị. Điểm này trên mặt yên ngựa được gọi là điểm yên ngựa hay điểm minimax



3.3. Đạo hàm của hàm có giá trị là số vô hướng

3.3.1. Đạo hàm của hàm $f: \mathbb{R}^n \ni x = (x_1, x_2, \dots, x_n) \to f(x) \in \mathbb{R}$

a. *Đạo hàm bậc nhất* của f, kí hiệu $\boldsymbol{\nabla f(x)} \triangleq \begin{bmatrix} \dfrac{\partial f(x)}{\partial x_1} \\ \dfrac{\partial f(x)}{\partial x_2} \\ \vdots \\ \dfrac{\partial f(x)}{\partial x_n} \end{bmatrix} \in \mathbb{R}^n$ (3.6)

b. *Đạo hàm bậc 2*, kí hiệu $\boldsymbol{\nabla^2 f(x)} \triangleq \begin{bmatrix} \dfrac{\partial^2 f(x)}{\partial x_1^2} & \dfrac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \cdots & \dfrac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \dfrac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \dfrac{\partial^2 f(x)}{\partial x_2^2} & \cdots & \dfrac{\partial^2 f(x)}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \cdots & \vdots \\ \dfrac{\partial^2 f(x)}{\partial x_n \partial x_1} & \dfrac{\partial^2 f(x)}{\partial x_n \partial x_2} & \cdots & \dfrac{\partial^2 f(x)}{\partial x_n^2} \end{bmatrix} \in \mathbb{R}^{n\times n}$ (3.7)

Lưu ý: (1) Ánh xạ $f: \mathbb{R}^n \to \mathbb{R}$ có đạo hàm bậc 1 $\boldsymbol{\nabla f(x)}$ là một ma trận trong $\mathbb{R}^{n\times 1}$, và có đạo hàm bậc hai $\boldsymbol{\nabla^2 f(x)}$ là một ma trận trong $\mathbb{R}^{n\times n}$

(2) Đạo hàm cấp 2: $\nabla_x^2 f(x)$ của hàm số nhiều biến số còn gọi là Hessian là m.trận đối xứng.

VD7. Cho hàm $f(x) = f(x_1, x_2) = x_1^2 - 2x_1x_2 + 2x_2^2$

Có: $\nabla_x f(x) = \begin{bmatrix} 2x_1 - 2x_2 \\ 4x_2 - 2x_1 \end{bmatrix}$; $\nabla_x^2 f(x) = \begin{bmatrix} 2 & -2 \\ -2 & 4 \end{bmatrix}$

3.3.2. Đạo hàm của hàm (biến ma trận) $f: \mathbb{R}^{m\times n} \ni X = \left(x_{ij}\right)_{m\times n} \to f(X) \in \mathbb{R}$

Đạo hàm (bậc 1), kí hiệu: $\nabla_X f(X) \triangleq \begin{bmatrix} \frac{\partial f(X)}{\partial x_{11}} & \frac{\partial f(X)}{\partial x_{12}} & \cdots & \frac{\partial f(X)}{\partial x_{1n}} \\ \frac{\partial f(X)}{\partial x_{21}} & \frac{\partial f(X)}{\partial x_{22}} & \cdots & \frac{\partial f(X)}{\partial x_{2n}} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial f(X)}{\partial x_{m1}} & \frac{\partial f(X)}{\partial x_{m2}} & \cdots & \frac{\partial f(X)}{\partial x_{mn}} \end{bmatrix} \in \mathbb{R}^{m\times n}$ (3.8)

Lưu ý: Đạo hàm $\nabla_X f(X)$ của hàm) f: $\mathbb{R}^{m\times n} \to \mathbb{R}$ là một ma trận trong $\mathbb{R}^{m\times n}$

VD 8. Xét hàm: $f: \mathbb{R}^{2\times 3} \ni X = \begin{bmatrix} x & y & z \\ u & v & t \end{bmatrix} \to f(X) = \|X\|_F^2 \in \mathbb{R}$

Có: $f(X) = \|X\|_F^2 = x^2 + y^2 + z^2 + u^2 + v^2 + t^2$

Vậy đạo hàm: $\nabla_X f(X) = \begin{bmatrix} 2x & 2y & 2z \\ 2u & 2v & 2t \end{bmatrix} = 2.X$

3.4. Đạo hàm của hàm có giá trị véc tơ (đầu ra là véc tơ): $f: \mathbb{R}^m \to \mathbb{R}^n$**:**

$$f(x) = f(x_1, x_2, \ldots, x_m) = (f_1(x), f_2(x), \ldots, f_n(x))^T$$

Đạo hàm của f, kí hiệu: $\nabla f(x) \triangleq \left[\frac{\partial f_j(x)}{\partial x_i}\right]_{m\times n}$ (i: chỉ số hàng, j: chỉ số cột) **(3.9)**

Lưu ý:

(1) Đạo hàm $\nabla f(x)$ của hàm $f: \mathbb{R}^m \to \mathbb{R}^n$ **là một ma trận trong** $\mathbb{R}^{m\times n}$

(2) Đ/với $f: \mathbb{R} \to \mathbb{R}^n$: $f(x) = (f_1(x), f_2(x), \ldots, f_n(x))^T$, người ta còn quan tâm đến đạo hàm bậc 2:

$$\nabla^2 f(x) \triangleq \left(\frac{d^2 f_1(x)}{dx^2}, \frac{d^2 f_2(x)}{dx^2}, \ldots \frac{d^2 f_n(x)}{dx^2}\right) \quad \textbf{(3.10)}$$

(3) $f: \mathbb{R}^m \to \mathbb{R}^n$, $\nabla f(x) \triangleq \left[\frac{\partial f_j(x)}{\partial x_i}\right]_{m\times n} = ([\nabla f_1(x)]\ \ [\nabla f_2(x)] \ldots [\nabla f_n(x)])$

VD 9. Xét $f(x, y) = (x + y, x.y, x - y) \Rightarrow \nabla f(x) = \begin{bmatrix} 1 & y & 1 \\ 1 & x & -1 \end{bmatrix}$

3.5. Các tính chất quan trọng của đạo hàm

3.5.1. Đạo hàm của tích (Qui tắc tích: Product Rule**):** Giả sử f(X), g(X) có cùng đầu vào là biến ma trận và có cùng chiều. Khi đó:

$$\boldsymbol{\nabla}\left(\boldsymbol{f}(\boldsymbol{X})^{\mathrm{T}}.\mathrm{g}(\mathrm{X})\right) = \left(\boldsymbol{\nabla f}(\boldsymbol{X})\right).\mathrm{g}(\mathrm{X}) + \left(\boldsymbol{\nabla}\mathrm{g}(\boldsymbol{X})\right).\mathrm{f}(\mathrm{X}) \qquad \textbf{(3.11)}$$

Lưu ý: (1) Trong (3.11) véc tơ được viết theo cột.

(2) Đ/v các hàm một biến số thông thường: u(x), v(x), có :

$$\{\boldsymbol{u}(\boldsymbol{x}).\boldsymbol{v}(\boldsymbol{x})\}' = \boldsymbol{u}'(\boldsymbol{x}).\boldsymbol{v}(\boldsymbol{x}) + \boldsymbol{v}'(\boldsymbol{x}).\boldsymbol{u}(\boldsymbol{x}) \qquad \textbf{(3.12)}$$

Trong (3.12) các tích có thể giao hoán, nhưng trong (3.11) thì không thể.

VD 10. Tìm $\nabla_{\mathrm{X}}\{\boldsymbol{f}(\boldsymbol{X})^{\mathrm{T}}.\mathrm{g}(\mathrm{X})\}$: $\mathrm{f}(\mathrm{x},\mathrm{y}) = (\mathrm{x},\mathrm{x}+\mathrm{y},\mathrm{y})$; $\mathrm{g}(\mathrm{x},\mathrm{y}) = (\mathrm{x},\mathrm{x}-\mathrm{y},\mathrm{y}), \forall\mathrm{X} = (\mathrm{x},\mathrm{y})^{\mathrm{T}} \in \mathbb{R}^2$

- Dùng quy tắc tích, có: $\boldsymbol{\nabla f}(\boldsymbol{X}) = \begin{bmatrix}1 & 1 & 0\\ 0 & 1 & 1\end{bmatrix}, \boldsymbol{\nabla}\mathrm{g}(\boldsymbol{X}) = \begin{bmatrix}1 & 1 & 0\\ 0 & -1 & 1\end{bmatrix}$

$$\boldsymbol{\nabla}\left(\boldsymbol{f}(\boldsymbol{X})^{\mathrm{T}}.\mathrm{g}(\mathrm{X})\right) = \begin{bmatrix}1 & 1 & 0\\ 0 & 1 & 1\end{bmatrix}.\begin{bmatrix}\mathrm{x}\\ \mathrm{x}-\mathrm{y}\\ \mathrm{y}\end{bmatrix} + \begin{bmatrix}1 & 1 & 0\\ 0 & -1 & 1\end{bmatrix}.\begin{bmatrix}\mathrm{x}\\ \mathrm{x}+\mathrm{y}\\ \mathrm{y}\end{bmatrix} = \begin{bmatrix}4\mathrm{x}\\ 0\end{bmatrix}$$

- Tính trực tiếp: $\mathrm{f}^{\mathrm{T}}.\mathrm{g}\colon \mathbb{R}^2 \to \mathbb{R}$, $\boldsymbol{f}(\boldsymbol{X})^{\mathrm{T}}.\mathrm{g}(\mathrm{X}) = [\mathrm{x}\ \ \mathrm{x}+\mathrm{y}\ \ \mathrm{y}].\begin{bmatrix}\mathrm{x}\\ \mathrm{x}-\mathrm{y}\\ \mathrm{y}\end{bmatrix} = 2\mathrm{x}^2 \Rightarrow \boldsymbol{\nabla}\left(\boldsymbol{f}(\boldsymbol{X})^{\mathrm{T}}.\mathrm{g}(\mathrm{X})\right) = \begin{bmatrix}4\mathrm{x}\\ 0\end{bmatrix}$

3.5.2. Quy tắc chuỗi (Chain Rule): Nếu $\mathrm{u}(\mathrm{X}) = \mathrm{g}(\mathrm{f}(\mathrm{X}))$ thì:

$$\boldsymbol{\nabla_X}\mathbf{u}(\boldsymbol{X}) = \boldsymbol{\nabla_X g}\left(\boldsymbol{f}(\boldsymbol{X})\right) = \boldsymbol{\nabla_X f}(\boldsymbol{X}).\boldsymbol{\nabla_f g}(\boldsymbol{f}) \qquad \textbf{(3.13)}$$

Chú ý: (3.13) tương tự như quy tắc đạo hàm hàm số hợp một biến u(x) = g(f(x):

u’(x) =(g(f(x))’ = g’(f).f’(x) = f’(x).g’(f)

tuy nhiên trong vế phải (3.13) không thể thay đổi thứ tự nhân 2 ma trận

VD 11. Cho f(x, y, z) = (y, z, x), g(x, y, z) = (x + z, y – z). Tìm $\boldsymbol{\nabla_X g}\left(\boldsymbol{f}(\boldsymbol{X})\right)$

Giải: Có $\mathrm{f}\colon \mathbb{R}^3 \to \mathbb{R}^3$, $\mathrm{g}\colon \mathbb{R}^3 \to \mathbb{R}^2$; $\mathrm{u} = \mathrm{g}\circ\mathrm{f}\colon \mathbb{R}^3 \to \mathbb{R}^2$

- Dùng qui tắc chuỗi: có $\boldsymbol{\nabla_X f(X)} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \boldsymbol{g(f)} = \mathrm{g(f_1, f_2, f_3) = (f_1 + f_3, f_2 - f_3)}$,

$$\boldsymbol{\nabla_f g(f)} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & -1 \end{bmatrix} \Rightarrow \boldsymbol{\nabla_X g\big(f(X)\big)} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} . \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$$

- Tính trực tiếp: Có $\boldsymbol{g\big(f(X)\big)} = \boldsymbol{g}(\mathrm{y, z, x}) = (\mathrm{y + x, z - x}) \Rightarrow \boldsymbol{\nabla_X g\big(f(X)\big)} = \begin{bmatrix} 1 & -1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$

3.6. Đạo hàm của một số hàm thường gặp.

3.6.1. Hàm: **f**: $\mathbb{R}^{\boldsymbol{n}} \to \mathbb{R}$: $\boldsymbol{f(x) = a^T . x}, \quad \boldsymbol{x} \in \mathbb{R}^{\boldsymbol{n}}$, a là véc tơ cho trước $\in \mathbb{R}^{\mathrm{n}}$

Có:$\boldsymbol{f(x) = a^T . x = \sum_{i=1}^{n} a_i x_i \Rightarrow \nabla f(x) = (a_1\ a_2 \dots a_n)^T = a}$

NX. Vì $\mathrm{x}^{\boldsymbol{T}}.\mathrm{a} = \boldsymbol{a^T . x}$, nên: $\boldsymbol{\nabla}\big(\mathrm{x}^{\boldsymbol{T}}.\mathrm{a}\big) = \boldsymbol{\nabla}\big(\boldsymbol{a^T . x}\big) = \mathrm{a}$

3.6.2. Hàm: $\boldsymbol{f(x)} = \mathrm{A}.\boldsymbol{x}, \ \mathrm{x} \in \mathbb{R}^{\boldsymbol{n}} (\mathrm{A} = \big(\mathrm{a_{ij}}\big)_{\mathrm{m \times n}}$ cho trước)

Ta có: f: $\mathbb{R}^{\mathrm{n}} \to \mathbb{R}^{\mathrm{m}}, \boldsymbol{f(x)} = \big[\sum_{\mathrm{j=1}}^{\mathrm{n}} \mathrm{a_{1j}.x_j} \ \ \sum_{\mathrm{i=1}}^{\mathrm{n}} \mathrm{a_{2j}.x_j} \ \dots . \sum_{\mathrm{j=1}}^{\mathrm{n}} \mathrm{a_{mj}.x_j} \ \ \big]^{\mathrm{T}}$

$$\boldsymbol{\nabla f(x)} = [\boldsymbol{\nabla}\mathrm{f_1(x)} \ \ \boldsymbol{\nabla}\mathrm{f_2(x)} \ \dots \ \boldsymbol{\nabla}\mathrm{f_m(x)}] = \begin{bmatrix} \mathrm{a_{11}} & \mathrm{a_{21}} & \cdots & \mathrm{a_{m1}} \\ \mathrm{a_{12}} & \mathrm{a_{22}} & \cdots & \mathrm{a_{m2}} \\ \vdots & \vdots & \cdots & \vdots \\ \mathrm{a_{1n}} & \mathrm{a_{2n}} & \cdots & \mathrm{a_{mn}} \end{bmatrix} = \mathrm{A^T}$$

Đặc biệt: $\qquad \boldsymbol{\nabla(x)} = \boldsymbol{\nabla}(\mathrm{I_n}.\boldsymbol{x}) = \mathrm{I_n}$

3.6.3. Hàm: $\boldsymbol{f(x)} = \mathrm{x^T.A.}\boldsymbol{x}, \ \mathrm{x} \in \mathbb{R}^{\boldsymbol{n}} (\mathrm{A} = \big(\mathrm{a_{ij}}\big)_{\mathrm{n \times n}}$ cho trước)

- Tìm theo quy tắc tích: $\boldsymbol{\nabla f(x)} = \boldsymbol{\nabla}\big(\mathrm{x^T}.(\mathrm{A}.\boldsymbol{x})\big) = \boldsymbol{\nabla}(\mathrm{x}).(\mathrm{Ax}) + \boldsymbol{\nabla}(\mathrm{Ax}).\mathrm{x} = \mathrm{Ax} + \mathrm{A^T.x} = \big(\boldsymbol{A + A^T}\big).\boldsymbol{x}$

- Tìm trực tiếp: Có **f**: $\mathbb{R}^{\boldsymbol{n}} \to \mathbb{R}$: $\boldsymbol{f(x)} = \mathrm{x^T.A.}\boldsymbol{x} = \boldsymbol{\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}.x_i.x_j}$

$$\boldsymbol{\nabla f(x)} = \Big(\sum_{\mathrm{i=1}}^{\mathrm{n}} (\mathrm{a_{i1} + a_{1i}})\mathrm{x_i} \quad \sum_{\mathrm{i=1}}^{\mathrm{n}} (\mathrm{a_{i2} + a_{2i}})\mathrm{x_i} \quad \cdots \quad \sum_{\mathrm{i=1}}^{\mathrm{n}} (\mathrm{a_{in} + a_{ni}})\mathrm{x_i}\Big)^{\boldsymbol{T}} = \big(\boldsymbol{A + A^T}\big).\boldsymbol{x}$$

3.6.4. Hàm $\boldsymbol{f}(\boldsymbol{x}) = \|\boldsymbol{A}.\boldsymbol{x} - \boldsymbol{b}\|_{\mathbf{2}}^{\mathbf{2}}$ $(\mathrm{A} = (\mathrm{a_{ij}})_{\mathrm{m\times n}}, \mathrm{b} \in \mathbb{R}^{\mathrm{m}}, \mathrm{x} \in \mathbb{R}^{\mathrm{n}})$

Có: $\boldsymbol{f}(\boldsymbol{x}) = \mathrm{g}(\mathrm{h}(\mathrm{x}))(\mathrm{h}(\mathrm{x}) = \mathrm{Ax} - \mathrm{b}, \mathrm{g}(\mathrm{x}) = \|\mathrm{x}\|_{\mathbf{2}}^{\mathbf{2}})$. Từ các vd trên và q.tắc chuỗi

$\nabla f(x) = \nabla g(h(x)) = (\nabla h(x))^T.\nabla_h g(h) = A^T.2.(A.x - b) = 2.A^T.(A.x - b)$

3.6.5. Hàm: $\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{a}^{\mathrm{T}}.\boldsymbol{x}.\boldsymbol{x}^{\mathrm{T}}.\boldsymbol{b}$, $\mathrm{x} \in \mathbb{R}^{\boldsymbol{n}}$ ($\boldsymbol{a}.\boldsymbol{b} \in \mathbb{R}^{\boldsymbol{n}}$cho trước)

Có: $f(x) = (a^T.x).(x^T.b) = (x^Ta)^T.(x^T.b)$

$\nabla f(x) = \nabla\{(x^Ta)^T.(x^T.b)\} = \nabla(x^Ta).(x^T.b) + \nabla(x^T.b).\ (x^Ta)$

$= a.(x^T.b) + b.(x^T.a) = a.x^Tb + b.a^T.x = a.b^Tx + b.a^T.x = (a.b^T + b.a^T).x$

3.6.6. $\boldsymbol{f}(\mathrm{X}) = \boldsymbol{trace}(\boldsymbol{A}.\boldsymbol{X})$ $(\mathrm{A} = (\boldsymbol{a_{ij}})_{\boldsymbol{n\times m}} \in \mathbb{R}^{\boldsymbol{n\times m}}, \boldsymbol{X} = (\boldsymbol{x_{ij}})_{\boldsymbol{m\times n}} \in \mathbb{R}^{\mathrm{m}\times\boldsymbol{n}})$

Có $f: \mathbb{R}^{m\times n} \to \mathbb{R}$; $A.X = (c_{ij})_{n\times n}$ với: $c_{ii} = \sum_{k=1}^{m} a_{ik}.x_{ki}$; $f(X) = \sum_{i=1}^{n} c_{ii}$

$$\nabla_X f(X) \triangleq \begin{bmatrix} \dfrac{\partial f(X)}{\partial x_{11}} & \dfrac{\partial f(X)}{\partial x_{12}} & \cdots & \dfrac{\partial f(X)}{\partial x_{1n}} \\ \dfrac{\partial f(X)}{\partial x_{21}} & \dfrac{\partial f(X)}{\partial x_{22}} & \cdots & \dfrac{\partial f(X)}{\partial x_{2n}} \\ \vdots & \vdots & \cdots & \vdots \\ \dfrac{\partial f(X)}{\partial x_{m1}} & \dfrac{\partial f(X)}{\partial x_{m2}} & \cdots & \dfrac{\partial f(X)}{\partial x_{mn}} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{n2} \\ \vdots & \vdots & \cdots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{nn} \end{bmatrix} = A^T$$

3.6.7.Hàm $\boldsymbol{f}(\mathrm{X}) = \boldsymbol{a}^{\mathrm{T}}.\mathrm{X}.\boldsymbol{b},\ \mathrm{X} = \left(\boldsymbol{x_{ij}}\right)_{\boldsymbol{m\times n}} \in \mathbb{R}^{\mathrm{m\times n}}$ (a $\in \mathbb{R}^{\boldsymbol{m}}$, b $\in \mathbb{R}^{\boldsymbol{n}}$)

Có: : f: $\mathbb{R}^{\mathrm{m\times n}} \to \mathbb{R}$; $\boldsymbol{f}(\boldsymbol{X}) = \sum_{\boldsymbol{j=1}}^{\boldsymbol{n}} \sum_{\boldsymbol{i=1}}^{\boldsymbol{m}} \boldsymbol{a_i b_j x_{ij}}$

$$\nabla_X f(X) \triangleq \begin{bmatrix} \frac{\partial f(X)}{\partial x_{11}} & \frac{\partial f(X)}{\partial x_{12}} & \dots & \frac{\partial f(X)}{\partial x_{1n}} \\ \frac{\partial f(X)}{\partial x_{21}} & \frac{\partial f(X)}{\partial x_{22}} & \dots & \frac{\partial f(X)}{\partial x_{2n}} \\ \vdots & \vdots & \dots & \vdots \\ \frac{\partial f(X)}{\partial x_{m1}} & \frac{\partial f(X)}{\partial x_{m2}} & \dots & \frac{\partial f(X)}{\partial x_{mn}} \end{bmatrix} = \begin{bmatrix} a_1b_1 & a_1b_2 & \cdots & a_1b_n \\ a_2b_1 & a_2b_2 & \cdots & a_2b_n \\ \vdots & \vdots & \cdots & \vdots \\ a_mb_1 & a_mb_2 & \cdots & a_mb_n \end{bmatrix} = a.b^T$$

3.6.8. Hàm $\boldsymbol{f}(\mathrm{X}) = \|\mathrm{X}\|_{\mathrm{F}}^{\boldsymbol{2}}$ ($\mathrm{X} = \left(\boldsymbol{x_{ij}}\right)_{\mathrm{n}\times\boldsymbol{n}} \in \mathbb{R}^{\mathrm{n\times n}}$)

Có f: $\mathbb{R}^{\mathrm{n\times n}} \to \mathbb{R}$: $\boldsymbol{f}(\mathrm{X}) = \sum_{\mathrm{j=1}}^{\mathrm{n}} \sum_{\mathrm{i=1}}^{\mathrm{n}} \mathrm{x}_{\mathrm{ij}}^{2}$;

$$\nabla_X f(X) = \begin{bmatrix} \frac{\partial f(X)}{\partial x_{11}} & \frac{\partial f(X)}{\partial x_{12}} & \dots & \frac{\partial f(X)}{\partial x_{1n}} \\ \frac{\partial f(X)}{\partial x_{21}} & \frac{\partial f(X)}{\partial x_{22}} & \dots & \frac{\partial f(X)}{\partial x_{2n}} \\ \vdots & \vdots & \dots & \vdots \\ \frac{\partial f(X)}{\partial x_{\mathrm{n}1}} & \frac{\partial f(X)}{\partial x_{\mathrm{n}2}} & \dots & \frac{\partial f(X)}{\partial x_{\mathrm{n}n}} \end{bmatrix} = \begin{bmatrix} 2\mathrm{x}_{11} & 2\mathrm{x}_{12} & \cdots & 2\mathrm{x}_{1\mathrm{n}} \\ 2\mathrm{x}_{21} & 2\mathrm{x}_{22} & \cdots & 2\mathrm{x}_{2\mathrm{n}} \\ \vdots & \vdots & \cdots & \vdots \\ 2\mathrm{x}_{\mathrm{n}1} & 2\mathrm{x}_{\mathrm{n}2} & \cdots & 2\mathrm{x}_{\mathrm{nn}} \end{bmatrix} = 2.\mathrm{X}$$

3.6.9. Hàm $\boldsymbol{f(x) = trace\left(X^T.A.X\right)}$, $\left(\mathrm{X} = \left(\boldsymbol{x_{ij}}\right)_{\boldsymbol{m\times n}} \in \mathbb{R}^{\mathrm{m\times n}}, \mathrm{A} \in \mathbb{R}^{\mathrm{m\times m}}\right)$

Có: f: $\mathbb{R}^{\mathrm{m\times n}} \to \mathbb{R}$. Kí hiệu các cột của X là: $\mathrm{x}_{\mathbf{1}}\ \mathrm{x}_{\mathbf{2}} \dots \mathrm{x}_{\boldsymbol{n}}$; $\mathrm{X} = [\mathrm{x}_{\mathbf{1}}\ \ \mathrm{x}_{\mathbf{2}}\ \ \dots\ \ \mathrm{x}_{\boldsymbol{n}}]$

Khi đó: $\boldsymbol{X^T.A.X} = [\mathrm{x}_{\mathbf{1}}\ \ \mathrm{x}_{\mathbf{2}}\ \ \dots\ \ \mathrm{x}_{\boldsymbol{n}}]^{\boldsymbol{T}}.\boldsymbol{A}.[\mathrm{x}_{\mathbf{1}}\ \ \mathrm{x}_{\mathbf{2}}\ \ \dots\ \ \mathrm{x}_{\boldsymbol{n}}] = \left(\mathrm{x_i^T.A.x_j}\right)_{\mathrm{n\times n}}$

$$\boldsymbol{f(x) = trace\left(X^T.A.X\right)} = \sum_{\mathrm{i=1}}^{\mathrm{n}} \mathrm{x_i^T.A.x_i}$$

Vì thế theo k.quả 3.6.3, ta có:

$$\nabla_X f(X) = \textstyle\sum_{\mathrm{i=1}}^{\mathrm{n}} \boldsymbol{\nabla}(\mathrm{x_i^T.A.x_i}) = \sum_{\mathrm{i=1}}^{\mathrm{n}}\left(\mathrm{A + A^T}\right).\mathrm{x_i} = \left(\mathrm{A + A^T}\right).\mathrm{X}$$

NX: $\nabla \boldsymbol{trace\left(X^T.X\right)} = \boldsymbol{\nabla trace\left(X^T.\mathrm{I}.X\right)} = 2\mathrm{I.X} = 2\mathrm{X}$ (là k.quả 3.4.8)

3.6.10. Hàm $\boldsymbol{f(X) = detX}$ $\left(\mathrm{X} = \left(\boldsymbol{x_{ij}}\right)_{\mathrm{n}\times\boldsymbol{n}} \in \mathbb{R}^{\mathrm{n\times n}}, \mathrm{X\ khả\ nghịch}\right)$

Khai triển theo hàng i có $\mathrm{detX} = \sum_{\mathrm{j}} \boldsymbol{x_{ij}}.\mathrm{X}_{\boldsymbol{ij}}$ ($\mathrm{X}_{\boldsymbol{is}}$ là phần phụ đại số của $\mathrm{x}_{\boldsymbol{ij}}$ trong X nên không chứa $\mathrm{x}_{\boldsymbol{ij}}, \forall \mathrm{s} = 1,2,\dots,\mathrm{n}$), từ đó: $\frac{\boldsymbol{\partial}(\mathrm{detX})}{\boldsymbol{\partial}\mathrm{x}_{\boldsymbol{ij}}} = \mathrm{X}_{\boldsymbol{ij}}$. Vì thế:

$$\nabla_{\mathrm{X}} f(X) = \nabla_{\mathrm{X}}\mathrm{detX} = \left(\frac{\boldsymbol{\partial}(\mathrm{detX})}{\boldsymbol{\partial}\mathrm{x}_{\boldsymbol{ij}}}\right)_{\mathrm{n\times n}} = \left(\mathrm{X}_{\boldsymbol{ij}}\right)_{\mathrm{n\times n}} = (\mathrm{X}^*)^{\mathrm{T}} = (\mathrm{detX}).\mathrm{X}^{-1}$$

3.6.11. (BT) Hàm $\boldsymbol{f}(\mathrm{X}) = \|\boldsymbol{A}.\mathrm{X} - \mathrm{B}\|_{\mathrm{F}}^{\mathbf{2}}$ $\left((\mathrm{X} \in \mathbb{R}^{\mathrm{m\times n}}, \mathrm{A} \in \mathbb{R}^{\mathrm{k\times m}}, \mathrm{B} \in \mathbb{R}^{\mathrm{k\times n}}\right)$

Bằng cách tương tự như trong 3.6.8, ta có:

$$\boldsymbol{\nabla f(X) = \nabla\|A.X - B\|_F^2 = 2A^T(A.X - B)}$$

Bảng các đạo hàm cơ bản: Tổng hợp các kết quả trên, ta có bảng các đạo hàm cơ bản:

Dạng hàm f	Đạo hàm ∇f	Dạng hàm f	Đạo hàm ∇f
$\mathrm{x}, (\mathrm{x} \in \mathbb{R}^{\mathrm{n}})$	$\nabla \mathrm{f}(\mathrm{x}) = \mathrm{I}$	$trace(X)$, $\mathrm{X} \in \mathbb{R}^{\mathrm{n} \times n}$	$\nabla \mathrm{f}(X) = \mathrm{I}$
$\boldsymbol{a}^{\boldsymbol{T}}.\boldsymbol{x}$, $(\boldsymbol{x} \in \mathbb{R}^{\boldsymbol{n}})$	$\boldsymbol{\nabla f}(\boldsymbol{x}) = \mathrm{a}$	$\boldsymbol{trace}(\boldsymbol{A}.\boldsymbol{X})$	$\nabla_X f(X) = A^T$
$\mathrm{x}^{\mathrm{T}}.\mathrm{A}.\boldsymbol{x}$	$\boldsymbol{\nabla f}(\boldsymbol{x}) = (\boldsymbol{A} + \boldsymbol{A}^{\boldsymbol{T}}).\boldsymbol{x}$	$\boldsymbol{trace}(\boldsymbol{X}^{\boldsymbol{T}}.\boldsymbol{A}.\boldsymbol{X})$	$(\boldsymbol{A} + \boldsymbol{A}^{\boldsymbol{T}}).\boldsymbol{X}$
$\mathrm{x}^{\mathrm{T}}.\boldsymbol{x} = \lVert \boldsymbol{x} \rVert_{\boldsymbol{2}}^{\boldsymbol{2}}$	$\boldsymbol{\nabla f}(\boldsymbol{x}) = \boldsymbol{2}.\boldsymbol{x}$	$\boldsymbol{trace}(\boldsymbol{X}^{\boldsymbol{T}}.\boldsymbol{X})$	$\boldsymbol{2}.\boldsymbol{X}$
$\lVert \boldsymbol{A}.\boldsymbol{x} - \boldsymbol{b} \rVert_{\boldsymbol{2}}^{\boldsymbol{2}}$	$2.A^T.(A.x - b)$	$\lVert \boldsymbol{A}.\mathrm{X} - \mathrm{B} \rVert_{\mathrm{F}}^{\boldsymbol{2}}$	$2.A^T.(A.X - B)$
$\boldsymbol{a}^{\mathrm{T}}.\boldsymbol{x}.\boldsymbol{x}^{\mathrm{T}}.\boldsymbol{b}$	$(a.b^T + b.a^T).x$	$\boldsymbol{trace}(\boldsymbol{A}^{\boldsymbol{T}}.\boldsymbol{X}.\boldsymbol{B})$	$\nabla \mathrm{f}(X) = \mathrm{A}.B^T$
$\boldsymbol{a}^{\mathrm{T}}.\boldsymbol{x}^{\mathrm{T}}.\boldsymbol{x}.\boldsymbol{b}$	$2.a^T.b.x$	$\boldsymbol{a}^{\mathrm{T}}.\boldsymbol{X}.\boldsymbol{b}$ Det(X)	$\mathrm{a}.b^T$ $\nabla \mathrm{deX} = \begin{cases} 0, \text{nếu detX} = 0 \\ (\mathrm{detX}).\mathrm{X}^{-1}, \mathrm{detX} \neq 0 \end{cases}$

3.7. Kiểm tra đạo hàm.

Việc tính toán tìm đạo hàm của hàm nhiều biến là khá phức tạp và dễ mắc lỗi. Trong thực nghiệm để kiểm tra sự chính xác của việc tính toán đạo hàm, thường dựa trên định nghĩa đạo hàm của hàm một biến.

3.7.1. Xấp xỉ đạo hàm của hàm một biến số.

- Với $\delta > 0$ đủ nhỏ ta có khai triển Taylor:

$$\mathrm{f}(\mathrm{x}+\delta) = \mathrm{f}(\mathrm{x}) + \mathrm{f}'(\mathrm{x}).\delta + \alpha(\delta^2) \;(\alpha(\delta^2) \text{ là VCB cấp cao hơn } \delta^2) \qquad (3.14)$$

$$\mathrm{f}(\mathrm{x}-\delta) = \mathrm{f}(\mathrm{x}) - \mathrm{f}'(\mathrm{x}).\delta + \beta(\delta^2) \;(\beta(\delta^2) \text{ là VCB cấp cao hơn } \delta^2) \qquad (3.15)$$

Từ (3.10) có: $\mathrm{f}'(\mathrm{x}) = \frac{\mathrm{f}(\mathrm{x}+\delta)-\mathrm{f}(\mathrm{x})}{\delta} + \mathrm{o}(\delta)$ ($\mathrm{o}(\delta)$ là VCB cấp cao hơn δ)

hay có xấp xỉ: $\mathrm{f}'(\mathrm{x}) \approx \frac{\mathrm{f}(\mathrm{x}+\delta)-\mathrm{f}(\mathrm{x})}{\delta}$, với sai số là $\mathrm{o}(\delta)$ (3.16)

Từ (3.11) có: $\mathrm{f}'(\mathrm{x}) = \frac{\mathrm{f}(\mathrm{x})-\mathrm{f}(\mathrm{x}-\delta)}{\delta} + \mathrm{o}(\delta)$ ($\mathrm{o}(\delta)$ là VCB cấp cao hơn δ)

hay có xấp xỉ: $\mathrm{f}'(\mathrm{x}) \approx \frac{\mathrm{f}(\mathrm{x})-\mathrm{f}(\mathrm{x}-\delta)}{\delta}$, với sai số là $\mathrm{o}(\delta)$ (3.17)

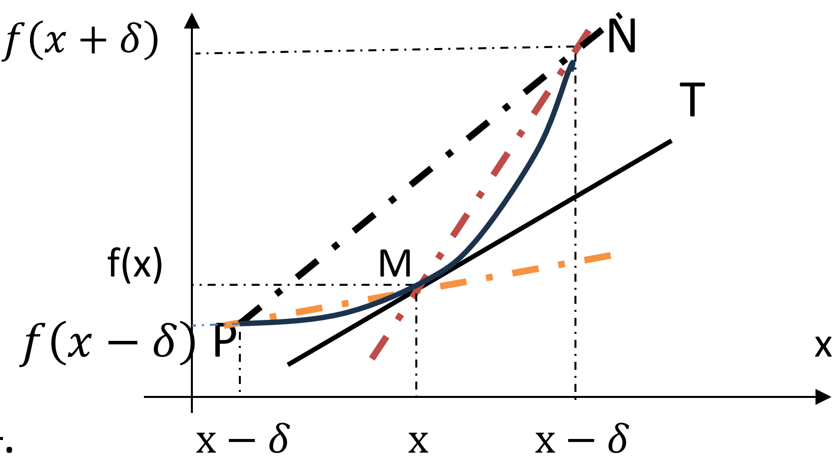
Từ (3.10), (3.11) có: $\mathrm{f}'(\mathrm{x}) = \frac{\mathrm{f}(\mathrm{x}+\delta)-\mathrm{f}(\mathrm{x}-\delta)}{2\delta} + \mathrm{o}(\delta^2)$ ($\mathrm{o}(\delta^2)$ là VCB cấp cao hơn δ^2)

nên có xấp xỉ : $\mathrm{f}'(\mathrm{x}) \approx \frac{\mathrm{f}(\mathrm{x}+\delta)-\mathrm{f}(\mathrm{x}-\delta)}{2\delta}$, với sai số là $\mathrm{o}(\delta^2)$ (3.18)

Cách tính xấp xỉ này được gọi là Numerical Gradient. Công thức xấp xỉ (3.18) được sử dụng phổ biến để tính Numerical Gradient cho hàm số một biến số do có độ chính xác cao hơn: Khi δ đủ bé thì sai số $\mathrm{o}(\delta^2) \ll \mathrm{o}(\delta)$.

- Dựa vào ý nghĩa hình học của đạo hàm: Với hàm số y = f(x), đạo hàm f'(x) là hệ số góc tiếp tuyến MT đồ thị tại điểm M(x, f(x)), $\frac{\mathrm{f}(\mathrm{x}+\delta)-\mathrm{f}(\mathrm{x})}{\delta}$ là hệ số góc của cát tuyến bên phải MN với N$(\mathrm{x}+\delta, \mathrm{f}(\mathrm{x}+\delta))$; $\frac{\mathrm{f}(\mathrm{x})-\mathrm{f}(\mathrm{x}-\delta)}{\delta}$ là hệ số góc của cát tuyến

bên trái MP với $P(x-\delta, f(x-\delta))$.
Rõ ràng là so với các cát tuyến MN, MP cát tuyến PN (có hệ số góc $\frac{f(x+\delta)-f(x-\delta)}{2\delta}$) có phương gần với tiếp tuyến MT hơn, tức là: $\frac{f(x+\delta)-f(x-\delta)}{2\delta}$ xấp xỉ cho f’(x) chính xác hơn là $\frac{f(x+\delta)-f(x)}{\delta}$ hay $\frac{f(x)-f(x-\delta)}{\delta}$.



3.7.2. Xấp xỉ đạo hàm của hàm nhiều biến số.

a. Khai triển Taylor cho hàm nhiều biến:

Giả sử $f: \mathbb{R}^n \to \mathbb{R}, f(x) = f(x_1, x_2, \dots, x_n)$ là hàm số n biến số. Với số dương δ đủ nhỏ, ta có công thức khai triển:

$$f(x+\delta.y) = f(x) + \delta. < \nabla f(x), y > +o(\delta) \qquad (3.19)$$

$$f(x+\delta.y) = f(x) + \delta. < \nabla f(x), y > +\frac{\delta^2}{2}.y^T.\nabla^2 f(x).y + o(\delta^2) \qquad (3.20)$$

Khai triển Taylor là cơ sở lý thuyết cho rất nhiều thuật toán tối ưu bằng cách xấp xỉ, đặc biệt là *Gradient descent* và *Newton step*

b. Xấp xỉ đạo hàm của hàm nhiều biến. Đ/v hàm nhiều biến $f(x) = f(x_1, \dots, x_n)$ xấp xỉ (3.18) được áp dụng cho từng biến (coi các biến khác là cố định) trong đạo hàm: $\boldsymbol{\nabla f(x)} \triangleq \left(\frac{\partial f(x)}{\partial x_1}, \frac{\partial f(x)}{\partial x_2}, \dots, \frac{\partial f(x)}{\partial x_n}\right)^T$.

Chương 4. Những nội dung cần thiết về xác suất

4.1.Xác suất và các công thức xác suất quan trọng (*ôn tập*)

4.1.1. Các khái niệm về xác suất

a. Phép thử và biến cố

- Một công việc sắp phải giải quyết là một phép thử
- Mỗi kết cục có thể của công việc là một biến cố sơ câp, kí hiệu: $\omega, \mathrm{a}, \mathrm{b}, \mathrm{c}$, ...
- Tập hợp tất cả các kết cục (biến cố sơ cấp) của phép thử gọi là không gian các biến cố sơ cấp, kí hiệu Ω
- Mỗi tập con của Ω bao gồm một số nào đó các biến cố sơ cấp gọi là một biến cố.
- Với ω là kết cục của phép thử. $\mathrm{A} \subset \Omega$:

Nói A xảy ra (không xảy ra) nếu $\omega \in \mathrm{A}$ ($\omega \notin \mathrm{A}$).

- $\mathrm{A} = \phi$: biến cố không thể; $\mathrm{A} \subset \Omega$: biến cố chắc chắn;
- $\phi \neq \mathrm{A} \neq \Omega$: Biến cố ngẫu nhiên.

b. Quan hệ giữa các biến cố:

b1- Quan hệ kéo theo: Nói rằng biến cố A kéo theo biến cố B, nếu như A xảy ra thì B cũng xảy ra. Khi đó ta viết: $\mathrm{A} \subset$ B: A tập con của B .

Nếu: $\mathrm{A} \subset$ B đồng thời: $\mathrm{B} \subset \mathrm{A}$ thì ta viết: $\mathrm{A} =$ B hay $\mathrm{A} \sim$ B hoặc $\mathrm{A} \Leftrightarrow$ B .

b2- Quan hệ xung khắc: A và B được gọi là xung khắc nhau nếu có ít nhất một biến cố trong chúng không xảy ra, tức là: $\mathrm{A}.\mathrm{B} = \emptyset$

• Họ các biến cố $\{A_\alpha\colon\ \alpha \in I\}$ được gọi là họ xung khắc từng đôi nếu hai biến cố bất kỳ trong chúng xung khắc nhau, tức là: $A_\alpha . A_\beta = \emptyset\colon\ \forall \alpha, \beta \in I, \alpha \neq \beta$

• Họ các biến cố $\{A_\alpha\colon\ \alpha \in I\}$ được gọi là một hệ đầy đủ nếu nhất thiết có một và chỉ một biến cố trong họ xảy ra, tức là: $\begin{cases} A_\alpha . A_\beta = \emptyset\colon\ \forall \alpha, \beta \in I, \alpha \neq \beta \\ \bigcup_{\alpha \in I} A_\alpha = \Omega \end{cases}$.

b3- Quan hệ đối lập: Hai biến cố A và B được gọi là đối lập nhau nếu {A, B} là một hệ đầy đủ, tức là: $A \cap B = \emptyset$ và $A \cup B = \Omega$

- Biến cố đối lập với A được ký hiệu là $\overline{\boldsymbol{A}}$**.** Ta có: $\overline{\boldsymbol{A}} = \boldsymbol{\Omega} \setminus \boldsymbol{A}$.

b4- Quan hệ độc lập: Hai biến cố được gọi là độc lập nhau nếu biến cố này xảy ra hay không không ảnh hưởng tới khả năng xảy ra hay không của biến cố kia và ngược lại.

c. Khái niệm về xác suất. *Xác suất của biến cố là đại lượng bằng số, đặc trưng cho khả năng xuất hiện của biến cố và được quy ước bao hàm giữa 0 và 1 sao cho biến cố chắc chắn có xác suất bằng 1.*

Xác suất của biến cố A được ký hiệu là P(A) .

d- Các tính chất của xác suất:

Tính chất 1: $0 \leq P(A) \leq 1$

Tính chất 2: $\mathrm{P}(\Omega) = 1$

Tính chất 3: Nếu $\mathrm{A} \subset$ B thì $\mathrm{P(A)} \leq \mathrm{P(B)}$

Tính chất 4: Nếu $\{A_j\colon j = 1,2, \dots\}$ xung khắc từng đôi thì: $P\left(\bigcup_j A_j\right) = \sum_j P(A_j)$

• ***Hệ quả:***

a/ $P(A \cup B) = P(A) + P(B) - P(A.B)$ (Công thức cộng)

b/ $\boldsymbol{P(\overline{A}) = 1 - P(A)}$, hay: $\boldsymbol{P(A) = 1 - P(\overline{A})}$ (Công thức biến cố đối)

c/ $P(\emptyset) = 0$

4.1.2. Các công thức định nghĩa xác suất

***a*- Công thức định nghĩa xác suất theo quan điểm cổ điển**

Định nghĩa: *Giả sử phép thử gồm n kết cục đồng khả năng (tức là khả năng xảy ra của chúng là như nhau), trong đó có m kết cục thuộc A(mà ta gọi là m kết cục thuận lợi cho A), Khi đó*: $\boldsymbol{P(A) = \frac{m}{n}}$ **(1)**

b – Công thức đ.nghĩa x.suất theo q.điểm hình học:

Xét tập $A \subset \mathbb{R}^k, k = 1, 2, 3$. Đặt: m(A) là độ dài của A nếu k = 1 , là diện tích của A nếu k = 2, là thể tích của A nếu k = 3. m(A) gọi là độ đo của A trong $\mathbb{R}^k$.

Đ.N: *Giả sử phép thử có các kết cục đồng khả năng và k.gian các biến cố sơ cấp* $\Omega \subset \mathbb{R}^k$*, có* $0 < m(\Omega) < +\infty$*. Khi đó*: $\boldsymbol{P(A) = \frac{m(A)}{m(\Omega)}}$ (2)

c- ĐN xác suất theo quan điểm thống kê: *Giả sử trong phép thử ta q/sát biến cố A và giả sử phép thử có thể được lặp lại nhiều lần. Gọi m(A) là số lần xuất hiện A trong n lần thử, khi đó f(A) = m(A)/n được gọi là tần suất xuất hiện A trong n lần thử. Rõ ràng là khi n thay đổi thì f(A) cũng thay đổi, nhưng khi n tăng lên đủ lớn, người ta thấy rằng f(A) dao động bé dần xung quanh một số cố định. Số cố định đó được gọi là xác suất của biến cố và ký hiệu là P(A).*

Nx: Đ/n này không chỉ ra một c.thức tính đúng x.suất của biến cố, nhưng nó chỉ ra rằng khi số lần thử n đủ lớn thì ta có c.thức tính gần đúng x.suất:

$$P(A) \approx f(A) \qquad (3)$$

4.1.3. Xác suất có đ.kiện- Các c.thức xác suất quan trọng

a- Xác suất có điều kiện: *Xác suất của biến cố B xét trong đ/kiện biến cố A xảy ra được ký hiệu là $P(B|A)$ và được tính theo công thức :*

$$\boldsymbol{P(B|A)} = \frac{\boldsymbol{P(A.B)}}{\boldsymbol{P(A)}} \text{ (nếu } P(A) \neq 0\text{)} \qquad (4)$$

Gọi $\boldsymbol{P(B|A)}$ là x.s có đ/kiện của biến cố B với đ/k biến cố A.

b. Các công thức xác suất quan trọng

b1. *Công thức nhân*: $\boldsymbol{P(A.B) = P(A).P(B|A)}$(nếu $P(A) \neq 0$) (5)

Hệ quả: A và B độc lập nhau ⇔ $\boldsymbol{P(A.B) = P(A).P(B)}$ (6)

Tính độc lập: Họ các biến cố $\{A_\alpha:\ \alpha \in I\}$ được gọi là:

• *Độc lập từng đôi*, nếu hai biến cố bất kỳ trong họ độc lập nhau, tức là:
$P\left(A_\alpha A_\beta\right) = P(A_\alpha).P\left(A_\beta\right), \forall \alpha, \beta \in I, \alpha \neq \beta$.

• *Độc lập nhau* (hay để nhấn mạnh, gọi là *độc lập trong toàn bộ*), nếu với mọi họ con gồm một số hữu hạn các biến cố: $A_{\alpha_1}, A_{\alpha_2}, \dots, A_{\alpha_n}$, đều có:

$$P(A_{\alpha_1}.A_{\alpha_2} \dots A_{\alpha_n}) = P(A_{\alpha_1}).P(A_{\alpha_2}) \dots P(A_{\alpha_n})$$

NX: Họ độc lập nhau $\not\rightleftarrows$ độc lập từng đôi.

VD1*:* Có 4 thùng hàng bề ngoài giống nhau, biết: 1 thùng chứa toàn sp loại 1, 1 thùng chứa toàn sp loại 2, một thùng chứa toàn sp loại 3, một thùng chứa cả ba loại 1, 2, 3. Lấy ra 1 thùng. Gọi A_i là biến cố: thùng lấy ra có chứa sp loại i, i = 1, 2, 3. Ta chỉ ra: $\{A_1, A_2, A_3\}$ độc lập từng đôi nhưng không độc lập nhau.

b2. Công thức xác suất đầy đủ: *Giả sử $A_1, A_2, \dots, A_n, \dots$ là hệ đầy đủ các biến cố có $P(A_j) > 0, \forall j$. Khi đó với mọi biến cố A , ta đều có:*

$$P(A) = \sum_j P(A_j) . P\left(A \middle| A_j\right) \qquad (7)$$

Bài toán*:* Một tập hợp Ω được chia thành k tập con không giao nhau: với số lượng các phần tử tương ứng theo tỷ lệ: $n_1 : n_2 : \dots : n_k$. Biết tỷ lệ các phần tử có tính chất A trong các tập con này tương ứng là: $p_1 : p_2 : \dots : p_k$. Hãy tính tỷ lệ các phần tử có tính chất A trong toàn bộ Ω

Giải. Phép thử: chọn ngẫu nhiên 1 p.tử từ tổng thể Ω, gọi: $\mathrm{A_j}$: *"p.tử được chọn thuộc tập thứ j", j = 1, 2,..., k. Ta có $\{\mathrm{A}_1, \mathrm{A}_2 \dots \mathrm{A}_k\}$* là hệ đầy đủ có xác suất $\mathrm{P}\left(\mathrm{A_j}\right) = \frac{\mathrm{n_j}}{\mathrm{N}}\ (\mathrm{N} = n_1 + n_2 + \cdots + n_k), \mathrm{j} = 1,2, \dots, \mathrm{k}.$

Gọi A: "p.tử được chọn có t/c A", thì tỉ lệ t/c A trong tập con thứ j là: $\mathrm{P}\left(\mathrm{A}\middle|\mathrm{A_j}\right) = \mathrm{p_j}$

Áp dụng c.thức xs đầy đủ cho A theo hệ đầy đủ $\{\mathrm{A}_1, \mathrm{A}_2 \dots \mathrm{A}_k\}$, ta có tỷ lệ t/c A trong tổng thể là: $P(A) = \sum_j P(A_j) . P\left(A\middle|A_j\right) = P(A) = \sum_j \frac{\mathrm{n_j}}{\mathrm{N}} . \mathrm{p_j}$

b3. Công thức Bayes. Với các giả thiết của đ.lý 1 và A là biến cố có P(A) > 0 . Khi đó ta có:

$$\boldsymbol{P(A_k|A) = \frac{P(A_k).P\left(A\middle|A_k\right)}{P(A)} = \frac{P(A_k).P\left(A\middle|A_k\right)}{\sum_j P(A_j).P\left(A\middle|A_j\right)}, \forall k.} \qquad (8)$$

Công thức xs toàn phần và công thức Bayes được ứng dụng rộng rãi trong Machin Learning

VD 2: Cơ sở A có 3 p/xưởng cùng s/x một loại s/p với sản lượng tương ứng tỷ lệ 4:5:4 và tỷ lệ loại 1 là 0,6; 0,7, 0,75. Chọn ngẫu nhiên 1 sp từ c.sở này thì gặp sp không đạt loại 1. Hỏi rằng khi đó sp này có nhiều khả năng nhất thuộc về p.xưởng nào?

b4. Công thức Bernoulli. Giả sử trong phép thử ta q.sát biến cố A, gọi là sự kiện " Thành công ", với xác suất thành công p = P(A) và giả sử rằng phép thử có thể được lặp lại nhiều lần. Khi đó x.suất để có đúng k t.công trong n lần thử là:

$$P_k(n,p) = C_n^k . p^k . (1-p)^{n-k}, (k = 0, , 1, \dots, n) \qquad (9)$$

b/ Một số chú ý:

Mô hình lặp lại n lần phép thử nói trên được gọi là m.h Bernoulli. Đây là một m.h khá phổ biến trong thực tế. Trong m.h Bernoulli, ta thường q.tâm tới số t.công x.hiện với x.suất lớn nhất, gọi là **số thành công có khả năng nhất**.

Mệnh đề: Số thành công có khả năng nhất trong m.hình *Bernoulli* là k^* được x.định như sau: $k^* = \begin{cases} [(n+1)p], & \text{nếu } (n+1)p \notin \mathbb{N} \\ \{(n+1)p - 1\} \text{ và } (n+1)p, & \text{nếu } (n+1)p \in \mathbb{N} \end{cases}$

VD3: Một cuộc thi trắc nghiệm gồm 10 câu hỏi, mỗi câu có kèm theo 5 p.án trả lời nhưng trong đó chỉ có 1 p.án đúng. Với mỗi câu, nếu trả lời đúng thì được 10 điểm, nếu trả lời sai thì bị trừ 2 điểm. Một thí sinh không nắm được n.dung bài thi nên làm bài theo cách chọn ngẫu nhiên một p.án trả lời cho mỗi câu hỏi. Hỏi rằng điểm bài thi của thí sinh này có nhiều k.năng nhất là bao nhiêu?

4.2. Đại lượng ngẫu nhiên và phân phối xác suất. (ôn tập)

4.2.1. Các k/n về đại lượng ngẫu nhiên

a. Đ/n: *Đ.lượng ngẫu nhiên (hay biến n.nhiên) là đ.lượng bằng số l.quan đến một phép thử nào đó mà g.trị của nó phụ thuộc vào k quả của phép thử.*

b. Ký hiệu: Thường dùng các ký hiệu: $\xi, \eta, \zeta, X, Y, Z, \ldots$

c. Phân loại: - Biến *n.nhiên rời rạc:* có tập g.trị có thể đánh số thứ tự.

Nếu tập giá trị của biến ngẫu nhiên X chỉ có một số c, tức là: P(X = c) = 1, thì nói X suy biến thành hằng số và viết X = c.

- *Biến n.nhiên l.tục*: có tập g.trị lấp đầy một khoảng số thực nào đó.

Vd 4: Một người đi từ nhà đến c.quan phải qua 5 ngã tư có đèn xanh, đèn đỏ. Gọi X là số ngã tư người này gặp đèn đỏ khi đi từ nhà đến c.quan. X là biến n.nhiên rời rạc.

Vd 5: Tung một chất điểm lên phạm vi một hình tròn tâm I, bán kính 1m. Đó là một phép thử, mỗi điểm của hình tròn là một k.cục của nó. Gọi Y là k.cách từ điểm rơi đến tâm I. Rõ ràng Y là đ.lượng bằng số, có g.trị lấp đầy khoảng [0, 1]. Vậy Y là một biến ngẫu nhiên l.tục.

4.2.2. Phân phối xác suất của đại lượng ngẫu nhiên

a/ Hàm phân phối xác suất: Giả sử X là một biến n.nhiên. Với mỗi số thực *x*, x.suất P(X $\leq$ x) là một số x.định. Khi *x* thay đổi, ta có hàm số: $F(x) = P(X \leq x)$

Ta gọi F(x) là hàm phân phối x.suất của biến ngẫu nhiên X . Nếu cần phải phân biệt, ta dùng ký hiệu: $F_X(x)$

a2/ Các tính chất: Giả sử F(x) là hàm phân phối xác suất của biến ngẫu nhiên X , Khi đó:

T.chất 1: $0 \leq F(x) \leq 1, \forall x$

T.chất 2: $F(x)$ là hàm đơn điệu không giảm, tức là: $F(x) \leq F(x'), \forall x < x'$

T.chất 3: $\boldsymbol{P(a < X \leq b) = F(b) - F(a)}, \forall a, b: a < b$

T.chất 4: $F(-\infty) = \lim\limits_{x \to -\infty} F(x) = 0; F(+\infty) = \lim\limits_{x \to +\infty} F(x) = 1;$

T.chất 5: $F(x)$ là hàm liên tục bên phải, tức là: $\lim\limits_{\substack{x \to t \\ x > t}} F(x) = F(t), \forall t;$

T.chất 6: Nếu $F(x)$ liên tục tại t thì: $P(X = t) = 0.$

* Nói hai biến ngẫu nhiên ξ_1, ξ_2 độc lập nhau nếu mọi biến cố chỉ liên quan đến $\xi_1$thì độc lập với mọi biến cố chỉ liên quan đến ξ_2, tức là:

$$P\{(\xi_1 \leq x).(\xi_2 \leq y)\} = P(\xi_1 \leq x).P(\xi_2 \leq y), \forall x, y \in \mathbb{R}.$$

* Nói các biến ngẫu nhiên $\xi_1, \xi_2, \dots, \xi_n$ độc lập nhau nếu với mọi giá trị x_1, x_2,..., $x_n \in \mathbb{R}$ thì họ các biến cố $\{\xi_j \leq x_j : j = 1,2, \dots, n\}$ là độc lập.

* Dãy các biến ngẫu nhiên $\xi_1, \xi_2, \dots, \xi_n$, ... độc lập nhau nếu với mọi số tự nhiên n thì các biến ngẫu nhiên $\xi_1, \xi_2, \dots, \xi_n$ đ.lập nhau.

b/ Hàm mật độ x.suất của biến ngẫu nhiên l.tục. Giả sử F(x) là hàm p.phối x.suất của biến n.nhiên ξ. Người ta chỉ ra rằng: ξ là l.tục $\Leftrightarrow$ $\exists$f(x) sao cho có biểu diễn: $F(x) = \int_{-\infty}^{x} f(t)dt, \ \forall x \in \mathbb{R}$. Khi đó $f(x)$ gọi là *hàm mật độ x.suất của ξ.*

Các tính chất:

T/c *1:* $f(x) = \mathrm{F}'(\mathrm{x}), \forall x \in C(F)$.
T/c *2:* $f(x) \geq 0, \forall x$
T/c *3:* $\boldsymbol{P(a \leq \xi \leq b) = P(a < \xi < b) = P(a \leq \xi < b) = P(a < \xi \leq b)}$
$\boldsymbol{= F(b) - F(a) = \int_a^b f(t)dt}$. Đặc biệt ta có: $\int_{-\infty}^{+\infty} \boldsymbol{f(x)dx = 1}$

c/ Bảng p.phối xác xuất của đ.lượng n.nhiên rời rạc:

Giả sử X là đ.lượng n.nhiên rời rạc có tập g.trị: $\boldsymbol{x_1, x_2, \ldots, x_n}, \ldots$. Đặt: $\boldsymbol{p_j} = \boldsymbol{P(X = x_j), j = 1, 2}, \ldots$ thì ta có $\boldsymbol{p_1 + p_2 + \cdots + p_n + \cdots = 1}$.

Bảng sau gọi là bảng p.phối x.suất của đại lượng ngẫu nhiên rời rạc:

X	x_1	x_2	$\cdots$	x_n	$\cdots$	Σ
P	p_1	p_2	$\cdots$	p_n	$\cdots$	1

(*)

Chú ý: Có nhiều tài liệu đ/n hàm p.phối của X là: $\mathrm{F}(\mathrm{x}) = \mathrm{P}(\mathrm{X} < \mathrm{x})$, khi đó $\mathrm{F}(\mathrm{x})$ là liên tục trái và trong các tinh chất trên, dấu < và dấu $\leq$ được hoan đổi cho nhau.

4.2.3 Các số đặc trưng quan trọng của biến ngẫu nhiên

a. Kỳ vọng: Kỳ vọng của biến ngẫu nhiên X (còn gọi là giá trị trung bình của X) là số được ký hiệu EX và được xác định như sau:

$$\mathrm{EX} = \begin{cases} \sum_j \boldsymbol{p_j . x_j}, \text{nếu X rời rạc có bảng phân phối xs } (*) \\ \int_{-\infty}^{+\infty} \boldsymbol{x . f(x)dx}, \text{nếu X liên tục có hàm mật độ xs f(x)} \end{cases} \quad \textit{(1)}$$

Các tính chất của kỳ vọng

T/c *1:* Nếu X = c thì: EX = c.

T/c *2:* Nếu $X \geq 0$ thì: $EX \geq 0$

T/c *3:* Với 2 biến ngẫu nhiên X, Y và ∀ a,b, ta có: $E(a.X + b.Y) = a.EX + b.EY$

T/c *4:* Nếu X và Y là các biến ngẫu nhiên độc lập nhau thì: $E(X.Y) = EX.EY$

T/c 5: Với $\varphi(x)$ là một hàm l.tục và X là một biến n.nhiên, khi đó $\varphi(X)$ là một biến n.nhiên có kỳ vọng được x.định như sau:

$$E\varphi(X) = \begin{cases} \sum_j p_j.\varphi(x_j)\text{, nếu X có bảng phân phối xs } (*) \\ \int_{-\infty}^{+\infty} \varphi(x).f(x)dx\text{, nếu X có hàmmaatj độ xs là f(x)} \end{cases} \quad (2)$$

Đặc biệt ta có: $$EX^2 = \begin{cases} \sum_j p_j.x_j^2\text{, nếu X có bảng phân phối xs } (*) \\ \int_{-\infty}^{+\infty} x^2.f(x)dx\text{, nếu X có hàmmaatj độ xs là f(x)} \end{cases} \quad (3)$$

b. Phương sai và độ lệch chuẩn. Phương sai của biến n.nhiên X là số được ký hiệu là DX (hoặc VarX) và x.định bởi hệ thức: $DX = E(X - EX)^2$ (4)

Các tính chất:

T/c 1: $DX \geq 0; DX = 0 \Leftrightarrow X = const$. Gọi $\sigma = \sqrt{DX}$ là độ lệch chuẩn của X.

T/c 2: Với mọi hằng số a, ta có: $D(aX) = a^2 DX$

T/c 3: $$DX = EX^2 - (EX)^2 \quad (4a)$$

T/c 4: Nếu X và Y độc lập nhau thì: $D(X \pm Y) = DX + DY$

T/c 5: ∀a ∈ ℝ, có: $E(X-a)^2 \geq E(X-EX)^2$; $E(X-a)^2 = E(X-EX)^2 \Leftrightarrow a = EX$

Ý nghĩa và cách dùng kỳ vọng và phương sai:

* Kỳ vọng hay g.trị t.bình của biến n.nhiên là hằng số xấp xỉ tốt nhất cho các g.trị

của biến n.nhiên, theo nghĩa sai số bình phương t.bình nhỏ nhất.
* Phương sai (hoăc độ lệch chuẩn) là số đánh giá sự tập trung hay p.tán các g.trị của biến n.nhiên x.quanh kỳ vọng của nó: P.sai (hoặc độ lệch chuẩn) càng bé thì các g.trị của biến n.nhiên tập trung càng gần kỳ vọng; p.sai (hoặc độ lệch chuẩn) càng lớn thì các g.trị của biến n.nhiên p.tán càng xa x.quanh kỳ vọng của nó
- Trong q.lý và k.doanh, kỳ vọng thường được dùng với tên gọi là số b.quân, p.sai thường được dùng để đ.giá độ rủi ro của các q.định.

c. Phân vị, trung vị và mốt . Xét biến n.nhiên X có hàm p.phối x.suất F(x) .

- Phân vị: Phân vị mức α của biến n.nhiên X là g.trị q_α thỏa mãn đ.kiện:

$$F(q_\alpha - 0) \leq \alpha \leq F(q_\alpha) \qquad (5)$$

- ***Trung vị (****median):* Trung vị của X, ký hiệu là MedX, đó là p.vị mức 0,5 của X. Vậy trung vị là vị trí phân chia p.phối x.suất thành hai nửa đều nhau.
- ***Mốt****(mode):*Mốt của biến X, ký hiệu: ModX, là số được x.định như sau:
* Nếu X rời rạc thì ModX là g.trị của X có x.suất lớn nhất.
* Nếu X l.tục thì ModX là g.trị của X mà tại đó hàm mật độ x.suất có trị lớn nhất. Theo đó ModX là giá trị có nhiều khả năng nhất của X.

d. Hệ số phân tán: Hệ số phân tán của đại lượng ngẫu nhiên X, ký hiệu V(X), được xác định theo hệ thức: $V(X) = \frac{\sqrt{DX}}{|EX|}.100(\%)$ (6)

- Hệ số phân tán thường được dùng để so sánh sự phân tán của 2 biến ngẫu nhiên khác nhau về quy mô hoặc đơn vị đo.

VD 5*:* Mức lãi (triệu đ) khi đầu tư 300 triệu đồng vào d.án A là biến n.nhiên X(triệu đồng) có bảng p.phối x.suất:

X	- 15	- 10	0	10	20	30	Σ
P	0,05	0,05	0,1	0,25	0,45	0,1	1

a/ Tính mức lãi b.quân khi đầu tư 300 triệu đ vào d.án A.
b/ Tính các đặc trưng: DX, ModX.
c/ Tính x.s để có mức lãi ít nhất là 20 triệu đồng khi đầu tư 300 triệu đồng vào dự án A.
d/ G.sử khi đầu tư 300 triệu đồng vào dự án B có mức tiền lãi là Y và có mức lãi bình quân là EY = EX, nhưng DY < DX. Khi đó nên chọn đầu tư vào dự án nào hơn? Tại sao?

VD *6*: Nhu cầu hàng năm về một mặt hàng A ở đ.p B là biến ngẫu nhiên Y(ngàn tấn) có hàm mật độ x.suất là: $f(x) = \begin{cases} kx(30-x), & x \in [0;30] \\ 0, & x \notin [0;30] \end{cases}$

a/ Xác định hằng số k và cho biết nhu cầu bq hàng năm về mặt hàng A ở đp B.
b/ Tính các số đặc trưng: DY, ModY.
c/ Tính xác suất để trong một năm nhu cầu về mặt hàng A ở địa phương B không dưới 25 ngàn tấn.

4.3. Các mô hình phân phối xác suất quan trọng

4.3.1. Phân phối nhị thức

a/ ĐN: Phân phối nhị thức với các tham số n và p ($p > 0$, $n \in \mathbb{N}^*$) ký hiệu B(n, p), là phân phối xs của biến ngẫu nhiên rời rạc X có tập giá trị: {0, 1,..., n} với xs:

$$\mathrm{p_k} = \mathrm{P(X = k)} = \mathrm{C_n^k}.p^k.(1-p)^{n-k}, k = 0,1,\dots,n$$

Khi đó ta viết: $X \sim B(n;p)$ để chỉ rằng: X là biến ngẫu nhiên có phân phối $B(n;p)$

b. Chú ý: **(1)** ***Mô hình nhị thức:*** Từ mh *Bernoulli* lặp lại n lần một phép thử (với xác suất thành công trong mỗi lần thử là p), nếu gọi X là số thành công trong n lần thử thì theo công thức *Bernoulli* , ta có: $p_k = P(X = k) = \mathrm{C_n^k}.p^k.(1-p)^{n-k}, k = 0,1,\dots,n.$

Như vậy mh *Bernoulli* cho ta một biến ngẫu nhiên X = số thành công trong n lần thử, có phân phối nhị thức. MH nhị thức là mh rất phổ biến trong thực tế.

(2). P.phối B (1, p) còn gọi là p.phối 0-1 tham số p, hay phân phối Bernoulli.

(3). Nếu $X \sim B(n;p)$ thì: EX = n.p; DX = np(1 – p) (6)

(4). Nếu $X_1, X_2, \dots, X_k$ là các biến ngẫu nhiên độc lập nhau và có cùng phân phối B(n, p) thì $X = X_1 + X_2 + \cdots + X_k$ là biến ngẫu nhiên có p.phối $B(k.n, p)$

VD 7*:* Tỷ lệ s.phẩm đạt t.chuẩn x.khẩu của một cơ sở sản xuất là 60%. Tìm x.suất để có ít nhất 4 s.phẩm đạt t.chuẩn x.khẩu trong 5 s.phẩm được lấy ngẫu nhiên từ cơ sở này

4.3.2. **Phân phối siêu bội. (tham khảo)** *Xét tập hợp Ω gồm N phần tử, trong đó có M p.tử có t/chất A*

a. ĐN: *Từ Ω lấy ngẫu nhiên (không hoàn lại) n phần tử. Gọi X là số phần tử có t/chất A trong n phần tử được lấy ra. Khi đó phân phối xs của X được gọi là phân phối siêu bội với các tham số n, M, N hay viết gọn là* $X \sim H(n, M, N)$.

b. Chú ý: (1)Mô hình siêu bội: *Đ/n trên cũng đã chỉ ra m.h p.phối siêu bội. Tuy nhiên cần phân biệt m.h siêu bội với mh nhị thức. Giả sử X là số p.tử có t.chất A trong n phần tử được lấy ra từ Ω. Khi đó:*

- Nếu n phần tử lấy có hoàn lại hoặc lấy ra từ một tập Ω có số lượng p.tử rất lớn, lớn hơn rất nhiều so với số n (chẳng hạn Ω là toàn bộ s.phẩm của một cơ sở đang s.xuất, hoặc từ một kho hàng,... và nói chung trong thực hành khi $n \leq 0{,}05.N$), thì X có phân phối nhị thức.
- Nếu n p.tử lấy không hoàn lại từ Ω có số lượng p.tử không lớn hơn nhiều so với số n (chẳng hạn Ω là một nhóm người, một lô hàng,...), thì X có p.phối siêu bội.
Từ đó cho thấy mh siêu bội cũng là một mh đơn giản nhưng khá phổ biến. Phân phối siêu bội còn được gọi là phân phối siêu hình học.

(2) Bảng phân phối xác suất: *Giả sử $X \sim H(n, M, N)$, khi đó X có giá trị nguyên, không âm, đồng thời số phần tử có (hay không có) t/chất A trong n phần tử lấy ra không thể vượt quá số phần tử có (hay không có) t/chất A trong toàn Ω và cũng không thể vượt quá n, từ đó suy ra bảng phân phối xác suất của X:*

X	k_0	k_0+1	$\cdots$	K	Σ
P	p_{k_0}	p_{k_0+1}	$\cdots$	p_K	1

trong đó $k_0 = max\{0, n+M-N\}, K = min\{n, M\}, p_k = \mathrm{C}_{\mathrm{M}}^{\mathrm{k}}\mathrm{C}_{\mathrm{N-M}}^{\mathrm{n-k}}/\mathrm{C}_{\mathrm{N}}^{\mathrm{n}}$

(3) Với $X \sim H(n, M, N)$, ta có: $EX = np; DX = np(1-p).\frac{N-n}{N-1}$ $\left(p = \frac{M}{N}\right)$

4.3.3. Phân phối phân loại (*Categorical distribution*). Trong nhiều trường hợp, đầu ra của một biến ngẫu nhiên rời rạc X có thể là một trong k giá trị khác nhau (k > 1), chẳng hạn một bức ảnh có thể chứa hình một người, một con thú cưng hay một vật dụng,.... Các đầu ra của nó có thể được mô tả bởi một phần tử trong tập hợp k phần tử: {1, 2,..., k}, khi đó p.phối phân loại chính là phân phối của biến ngẫu nhiên rời rạc gồm k giá trị, là tổng quát hóa của p.phối Bernoulli. Với k đầu ra, p.phối categorical được mô tả bởi tham véc tơ Λ gồm k tham số k tham số: $\Lambda = (\lambda_1, \lambda_2, \dots, \lambda_\mathrm{k})$, trong đó mỗi tham số λ_j là xác suất để đầu ra nhận giá trị j: $\lambda_\mathrm{j} = \mathrm{P}(\mathrm{X} = \mathrm{j}), \mathrm{j} = \overline{1, k}$ $(\lambda_1 + \lambda_2 + \dots + \lambda_\mathrm{k} = 1)$.

Nhiều trường hợp, biểu diễn đầu ra thứ j là (X = j) bởi một véc tơ đơn vị k chiều $\mathrm{e_j} = (\mathrm{x_1}, \mathrm{x_2}, \dots, \mathrm{x_k}), \left(\mathrm{x_i} = \begin{cases} 1, \text{nếu } \mathrm{i} = \mathrm{j} \\ 0, \text{nếu } \mathrm{i} = \mathrm{j} \end{cases}\right), \mathrm{j} = 1,2, \dots, \mathrm{k}$. Khi đó có biểu diễn:

$$\mathrm{P(X \coloneqq e_j)} = \lambda_\mathrm{j} = \prod_{\mathrm{i}=1}^{\mathrm{k}} \lambda_\mathrm{i}^{\mathrm{x_i}}$$

VD 8: Một lô hàng gồm 10 s/p, trong đó có 7 s/p loại A và 3 s/p loại B. Mỗi s/p loại A có giá 100 ngàn đồng, mỗi s/p loại B có giá 80 ngàn đồng. Lấy ngẫu nhiên 4 s/p từ lô hàng. Gọi X là tổng giá trị của 4 s/p lấy ra.

(a) Lập bảng p.phối xs của X; (b)Tính xs để tổng giá trị của 4 s/p lấy ra $\geq$370 ngàn đồng.

4.3.3. Phân phối Poisson

a. ĐN: Phân phối Poisson với tham số λ $(\lambda > 0)$, ký hiệu là P_λ , là p.phối xs của biến rời rạc X có tập giá trị là $\mathbb{N}$ với xs: $p_k = P(X = k) = \frac{\lambda^k.e^{-\lambda}}{k!}$, $k = 0,1,2,\dots$

Ký hiệu: $X \sim P_\lambda$ để chỉ X là biến ngẫu nhiên có p.phối Poisson với tham số λ

b. Chú ý: **(1)** Nếu $X \sim P_\lambda$ thì: $\mathrm{EX} = \mathrm{DX} = \lambda$

(2) ***Mô hình Poisson:*** *Số lần x.hiện biến cố n.nhiên A nào đó trong khoảng t.gian T là biến ng.nhiên có phân phối Poisson*. Như vậy phân phối Poisson khá phổ biến trong nhiều l.vực khác nhau. Xét trong một khoảng t.gian T nào đó thì: Số k.hàng đến một h.thống d.vụ, số tín hiệu nhận được ở một máy thu sóng, số trẻ chào đời ở một b.viện phụ sản,… là các biến n.nhiên có p.phối Poisson. P.phối Poisson được ứ.dụng rộng rãi đối với nhiều q.trình có l.quan đến số q.sát đ.với một đơn vị t.gian hoặc k.gian, chẳng hạn trong các lĩnh vực quản lý, kinh doanh: k.tra c.lượng s.phẩm, lý thuyết quản trị dự trữ, lý thuyết sắp hàng, các hệ phục vụ đám đông,....

(3) Trong mh Poisson, Với $X \sim P_\lambda$, thì: $ModX = \begin{cases} \lambda \text{ và } \lambda - 1, \text{nếu } \lambda \in \mathbb{N}^* \\ [\lambda], \qquad\qquad \text{nếu } \lambda \notin \mathbb{N}^* \end{cases}$

(4). Nếu $X_1, X_2, \dots, X_n$ là các biến ngẫu nhiên độc lập nhau và có cùng p.phối $P(\lambda)$ thì $X = X_1 + X_2 + \cdots + X_n$ là biến ngẫu nhiên có p.phối $P(n.\lambda)$

VD 9*:* Mỗi khách hàng vào hệ thống dịch vụ A có thể chọn một trong 3 mức phí dịch vụ: 100 ngàn đồng, 150 ngàn đồng, 200 ngàn đồng. Lượng khách vào chọn các mức phí dịch vụ này tương ứng theo tỷ lệ: 3: 5: 2. Được biết trong vòng một giờ làm việc, bình quân có 30 khách vào đây.

a/ Tìm xs để trong vòng 1 giờ làm việc, có từ 28 đến 32 khách vào hệ dịch vụ A.

b/ Trong số 25 khách vào hệ dịch vụ A, số khách chọn mức phí 200 ngàn đồng có nhiều khả năng nhất là bao nhiêu?

c/ Tìm xác suất để trong 30 khách vào hệ dịch vụ A, lượng khách chọn mức phí dịch vụ 200 ngàn đồng gấp 5 lần lượng khách chọn 2 mức phí còn lại.

4.3.4. Phân phối chuẩn

a. ĐN: Phân phối chuẩn với các tham số μ, σ^2, ký hiệu $N(\mu, \sigma^2)$, là phân phối xs của biến ngẫu nhiên liên tục có hàm mật độ xs: $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, x \in \mathbb{R}$

Ký hiệu $\xi \sim N(\mu, \sigma^2)$ để chỉ ξ là biến n.nhiên có p.phối chuẩn với các tham số μ, σ^2.

b. Chú ý

b1. Nếu $\xi \sim N(\mu, \sigma^2)$ thì: $\mathrm{E}\xi = \mu, D\xi = \sigma^2$

b2. $\xi \sim N(\mu, \sigma^2) \Leftrightarrow \xi_0 = \frac{\xi - \mu}{\sigma} \sim N(0, 1)$

b3. P.phối N(0;1) gọi là p.phối chuẩn chính tắc, có hàm mật độ x.s là:

$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}, x \in \mathbb{R}$ và có hàm p.phối x.suất là: $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}t^2} dt$

* Phân vị mức 1- α của p.phối chuẩn, ký hiệu là $u(\alpha)$ (hay z_α), còn gọi là g.trị tới hạn mức α của p.phối chuẩn và được x.định bởi hệ thức: $\Phi(u(\alpha)) = 1 - \alpha$

* G.trị của hàm $\Phi(x)$ được cho bởi bảng, với $x \geq 0, lưu\ ý\ l$à: $\Phi(-x) = 1 - \Phi(x)$.

* Hàm: $\Phi_0(x) = \int_0^x \varphi(t)dt = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{1}{2}t^2} dt$, được gọi là *hàm Laplace*. Bảng g.trị hàm Laplace có thể thay cho bảng g.trị hàm phân phối chuẩn, với chú ý:

$$\Phi(x) = \Phi_0(x) + 0{,}5; \Phi_0(-x) = -\Phi_0(x)$$

b4. $Với\ X \sim N(\mu, \sigma^2), ta\ c$ó: $\mathrm{P}(\alpha \leq \mathrm{X} \leq \beta) = \Phi\left(\frac{\beta-\mu}{\sigma}\right) - \Phi\left(\frac{\alpha-\mu}{\sigma}\right)$;

$$\mathrm{P}(X \geq \alpha) = 1 - \Phi\left(\frac{\alpha-\mu}{\sigma}\right);\ \ \mathrm{P}(\mathrm{X} \leq \beta) = \Phi\left(\frac{\beta-\mu}{\sigma}\right).$$

b5. *Quy tắc " Ba xích- ma": Giả sử $X \sim N(\mu, \sigma^2)$, khi đó hầu hết (với xác suất 99,74%) các g.trị của biến này sẽ rơi vào khoảng từ: $\mu - 3\sigma$ đến $\mu + 3\sigma$.*

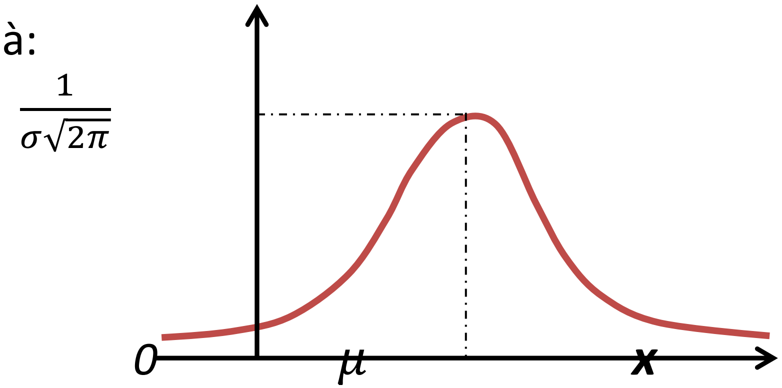
b6. Nếu $X_1, X_2, \dots, X_n$ là các biến n.nhiên đ.lập nhau và có cùng p.phối $N(\mu, \sigma^2)$ thì: $X = X_1 + X_2 + \cdots + X_n \sim N(n\mu, n\sigma^2)$

b6. *Sự xấp xỉ p.phối chuẩn:* Giả sử $X_1, X_2, \dots, X_n$ là các biến ngẫu nhiên độc lập nhau và có cùng p.phối với $EX_j = \mu, DX_j = \sigma^2 < +\infty$. Khi đó với n đủ lớn ta có:

(1). $\boldsymbol{\bar{X}_n = \frac{1}{n}\sum_{j=1}^{n} X_j \approx \mu}$, tức là: $\lim\limits_{n\to\infty} \boldsymbol{P(|\bar{X}_n - \mu| \geq \varepsilon) = 0},\ \forall \boldsymbol{\varepsilon > 0}$

(2). $\boldsymbol{\bar{X}_n = \frac{1}{n}\sum_{j=1}^{n} X_j}$ có p.phối xấp xỉ chuẩn $N(\mu, \frac{\sigma^2}{n})$

b7. Đồ thị hàm mật độ chuẩn p.ánh một thực tế là: trong nhiều tr.hợp, đ/v một biến ngẫu nhiên o đang q.sát thì các g.trị của nó thường t.trung nhiều x.quanh g.trị t.bình (có q.sát lớn hơn thì cũng có q.sát bé hơn), và tỷ lệ các q.sát càng xa g.trị t.bình càng ít dần.



Trong thực tế nhiều biến ng.nhiên tuân theo luật chuẩn hoặc gần chuẩn như: trọng lượng và chiều cao của người trưởng thành, lực chịu đựng của thanh vật liệu, các sai số đo đạc, sai số q.sát, độ bền của máy móc, v.v...

Vd 10: *T*rọng lượng của một loài g.súc 1 năm tuổi là biến n.nhiên X(kg) có p.phối $\mathrm{N}(\mu = 70, \sigma = 4)$. Như vậy trọng lượng t.bình của chúng là 70 kg và có tới 99,74% số lượng loài g.súc này có tr.lượng trong khoảng từ 58 kg đến 82 kg.

Vd 11. T.gian thanh toán hóa đơn tại một c.ty là biến ng.nhiên X(đơn vị tính: ngày) có p.phối chuẩn N(18, 16). Hãy cho biết: (a). Tỷ lệ hóa đơn có t.gian thanh toán từ 12 ngày đến 18 ngày. (b). Tỷ lệ hóa đơn có t.gian thanh toán trước 8 ngày.
(c). Tỷ lệ hóa đơn có t.gian thanh toán quá 30 ngày.

***4.3.5.* Phân phối đều (ôn tập)**

a. Đ.n: P.phối đều trên đoạn [a, b] là p.phối của biến ngẫu nhiên X liên tục tập trung trên [a, b] có hàm mật độ: $f(x) = \begin{cases} \frac{1}{b-a}, x \in [a, b] \\ 0,\ \ x \notin [a, b] \end{cases}$ (Khi đó viết $X \sim U_{[a,\,b]}$)

b. Chú ý: (1). Nếu ξ *có hàm p.phối* $\mathrm{F(x)}$ l.tục thì $\mathrm{F}(\xi)$ *có p.phối đều trên* $[0,1]$.
Tính chất này được ứng dụng trong lý thuyết mô phỏng.

(2) Nếu $X \sim U_{[a,\, b]}$ thì: $EX = \frac{a+b}{2}$; $DX = \frac{(b-a)^2}{12}$ và hàm p.phối x.s của X là:

$$F(x) = \begin{cases} 0, \text{nếu } x \le a \\ \frac{x-a}{b-a}, \text{nếu } x \in (a,b) \\ 1, \text{nếu } x \ge b \end{cases}$$

4.3.6. Phân phối mũ (ôn tập)

a. Đ.n: P.phối mũ với tham số λ $(\lambda > 0)$ là p.phối xs của biến ngẫu nhiên l.tục X có hàm mật độ: $f(x) = \begin{cases} \lambda e^{-\lambda x}, x > 0 \\ 0, \quad\quad x \le 0 \end{cases}$, hay có hàm p.phối: $\mathrm{F}(x) = \begin{cases} 1 - e^{-\lambda x}, x > 0 \\ 0, \quad\quad x \le 0 \end{cases}$

b. Chú ý: **(1).** Trong thực tế nhiều biến ngẫu nhiên tuân theo luật mũ: T.gian phục vụ của một hệ phục vụ, t.gian chờ của một k.hàng để được p.vụ, tuổi thọ của thiết bị, máy móc, sp, của một loài sinh vật,...là các biến n.nhiên có p.phối mũ.

(2) Nếu X có phân phối mũ với tham số λ thì : $EX = \frac{1}{\lambda}$; $DX = \frac{1}{\lambda^2}$

VD 12*:* Tuổi thọ t.bình của thiết bị A là 4 năm. Một cửa hàng có bán loại thiết bị này, nếu bán được 1 thiết bị thì lập tức họ có số tiền lời là 180.000đ, nhưng nếu trong t.gian bảo hành thiết bị đó bị hỏng các hạng mục bảo hành thì b.quân họ phải bỏ ra 300.000đ cho bảo hành. Biết tuổi thọ của loại thiết bị này là biến ngẫu nhiên có phân phối mũ.

(1) Nếu t.gian bảo hành là 1 năm thì b.quân với một thiết bị bán ra cửa hàng lãi được bao nhiêu? **(2).** Nếu muốn cửa hàng có tiền lãi b.quân trên một thiết bị bán ra $\ge$ 120 ngàn đồng thì cần phải đưa ra t.gian bảo hành tối đa là bao lâu?

4.3.7. Phân phối Chi-bình phương và phân phối Student (ôn tập)

a. ĐN: **(1).** Phân phối khi-bình phương với n bậc tự do là phân phối của biến ngẫu nhiên: $\chi^2(\mathrm{Chi-square}) = \mathrm{X}_1^2 + \mathrm{X}_2^2 + \cdots + \mathrm{X}_n^2$, trong đó $X_1, X_2, \dots, X_n \sim \mathrm{N}(0,1)$ và độc lập nhau. Ký hiệu $\chi^2 \sim \chi_{\mathrm{n}}^2$ để chỉ χ^2 là biến ngẫu nhiên có phân phối khi-bình phương với n bậc tự do.

(2). Phân phối Student với n bậc tự do là p.phối xs của biến ngẫu nhiên:
$T = \frac{X}{\sqrt{Y/n}}$, trong đó X và Y độc lập nhau và $X \sim N(0,1), Y \sim \chi^2_{\mathrm{n}}$.

• Ký hiệu $\chi^2_{\mathrm{n}}(\alpha)$ là phân vị mức $1 - \mathrm{\alpha}$ của biến ngẫu nhiên có p.phối khi-bình phương với n bậc tự do (hay giá trị tới hạn mức α của p.phối khi-bình phương).
Hàm phân phối xs của T dần về hàm phân phối chuẩn chính tắc khi $n \to \infty$
Trong thực hành, với $n \geq 30$, có thể coi T có phân phối xấp xỉ chuẩn N(0; 1).
• Ký hiệu $t_n(\alpha)$ là phân vị mức $1 - \alpha$ của biến ngẫu nhiên có phân phối Student với n bậc tự do, còn gọi là giá trị tới hạn mức α của phân phối Student (với chú ý: $t_n(\alpha) = -t_n(1-\alpha)$).
• Nếu $X_1, X_2, \dots, X_n$ độc lập nhau và có cùng phân phối χ^2 với m bậc tự do thì $X_1 + X_2 + \cdots + X_n$ có phân phối χ^2 với n.m bậc tự do.
• Nếu $X_1, X_2, \dots, X_n$ độc lập nhau và có cùng phân phối $N(a, \sigma^2)$ thì $t = \frac{(\bar{X}_n - a)\sqrt{n-1}}{S(\xi)}$ có phân phối Student với (n – 1) bậc tự do ($\bar{X}_n = \frac{1}{n}(X_1 + X_2 + \cdots + X_n)$).
Các phân phối này có nhiều ứng dụng trong các mô hình ươc lượng và kiểm định.

4.3.8. Phân phối Beta (*Beta distribution*). Phân phối Beta là một phân phối liên tục được định nghĩa trên một biến ngẫu nhiên λ ∈ [0,1]. Phân phối này phù hợp với miệc mô tả sự *biến động* của tham số λ trong phân phối Bernoulli. Phân phối Beta được mô tả bởi hai tham số dương α, β ∈ (0,∞). Hàm mật độ xác suất của nó là: $\mathrm{p}(\lambda) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha).\Gamma(\beta)}.\lambda^{\alpha-1}.(1-\lambda)^{\beta-1}$(viết gọn là: $\mathrm{p}(\lambda) = \mathrm{Beta}_{\lambda}(\alpha,\beta)$) trong đó $\Gamma(.)$ là hàm Gamma: $\Gamma(\mathrm{x}) = \int_0^{+\infty} \mathrm{t}^{\mathrm{x}-1}\mathrm{e}^{-\mathrm{t}}\mathrm{dt}$

4.4. Một số bất đẳng thức quan trọng

4.4.1. Bất đẳng thức Markov: Với biến ngẫu nhiên $X \geq 0$, ta có:

$$P(X \geq u) \leq \frac{EX}{u},\ \forall u > 0 \qquad (1)$$

C/m: Có $EX = E\left(X.I_{[X<u]}\right) + E\left(X.I_{[X\geq u]}\right) \geq E\left(X.I_{[X\geq u]}\right) \geq u.P(X \geq u)$**.** Từ đó suy ra (1)

NX. BĐT cho một cận trên của xs $P(X \geq u)$, hay cho một cận dưới của xs $P(X < u)$

VD 13. Giả sử một trang trên PowerPoint bình quân chứa 400 từ tiếng Việt cỡ chữ 18. Như vậy số trang có từ 600 từ trở lên chiếm tỷ lệ không quá 2/3.

4.4.2. Bất đẳng thức Chebyshev: Với biến ngẫu nhiên X bất kì có phương sai hữu hạn, ta có:

$$P(|X - EX| \geq u) \leq \frac{Var(X)}{u^2},\ \forall u > 0 \qquad (2)$$

C/m: $Var(X) = E\{(X - EX)^2\} = E\left((X - EX)^2.I_{[|X-EX|\geq u]}\right) + E\left((X - EX)^2.I_{[|X-EX|<u]}\right)$

$\geq E\left((X - EX)^2.I_{[|X-EX|\geq u]}\right) \geq u^2.P(|X - EX| \geq u)$ **.** Từ đây suy ra (2).

NX. (1). Tổng quát hơn, ta có: $P(|X - EX| \geq u) \leq \inf\limits_{q>0} \frac{E(|X-EX|)^q}{u^q},\ \forall u > 0$ (3)

(Đối với biến X có moment mọi cấp)

(2). BĐT (2) cho một cận trên của xs $P(|X - EX| \geq u)$ là: $\frac{Var(X)}{u^2}$, hay cho một cận dưới của xs $P(|X - EX| < u)$ ***là:*** $1 - \frac{Var(X)}{u^2}$**.**

VD 14. Lượng khách đến một hệ dịch vụ trong một ngày làm việc là biến ngẫu nhiên có độ lệch chuẩn là 8 và bình quân một ngày làm việc có 100 khách đến.

Như vậy theo BĐT Chebyshev, lượng khách từ 80 đến 120 trong mỗi ngày làm việc chiếm tỷ lệ không dưới 84%.

4.4.3. Bất đẳng thức Chernoff: Với biến ngẫu nhiên X với hàm sinh moment $\mathrm{M_X(s) = Ee^{s.X}}$, ta có: $\mathrm{P(X \geq u) \leq \inf_s\{e^{-s.u}.M_X(s)\},\ \forall u \in \mathbb{R}}$ (4)

C/m: Trong BĐT, thay X bởi $\mathrm{e^{s.X}}$, ta nhận được (4).

4.4.4. Bất đẳng thức Jensen: Nếu f(x) là hàm lồi thì với mọi biến ngẫu nhiên X, ta có: $\mathrm{E[f(X)] \geq f[E(X)]}$ (5)

Chú ý: (1) Tập D trong $\mathbb{R}^{\mathrm{n}}$ được gọi là một tập lồi, nếu mọi đoạn thẳng nối 2 điểm x, y bất kỳ trong D đều nằm trọn trong D, tức là: $\forall \mathrm{x}, \mathrm{y} \in \mathrm{D}, \forall \lambda \in [0, 1]$, thì:

$$\lambda \mathrm{x} + (1 - \lambda)\mathrm{y} \in \mathrm{D}$$

(2) Nói x là một tổ hợp lồi của các điểm $\mathrm{x^1, x^2, \dots, x^k \in D}$, nếu x có biểu diễn: $\mathrm{x} = \sum_{\mathrm{i=1}}^{\mathrm{k}} \lambda_{\mathrm{i}}\mathrm{x^i}, \lambda_1, \lambda_2, \dots, \lambda_{\mathrm{k}} \geq 0 : \lambda_1 + \lambda_2 + \cdots + \lambda_{\mathrm{k}} = 1$

(3) Tập D là lồi ⟺ D chứa mọi tổ hợp lồi các điểm của nó.

(4) Hàm f: $\mathrm{D} \longrightarrow \mathbb{R}$ là hàm lồi (convex) trên D, nếu D là tập lồi và:

$$\mathrm{f}(\lambda \mathrm{x} + (1 - \lambda)\mathrm{y}) \leq \lambda \mathrm{f(x)} + (1 - \lambda)\mathrm{f(y)}, \forall \mathrm{x}, \mathrm{y} \in \mathrm{D}, \forall \lambda \in [0, 1]$$

(Dây chắn cung đồ thị không nằm dưới cung đồ thị)

4.5. Véc tơ ngẫu nhiên

4.5.1. Các khái niệm về véc tơ ngẫu nhiên

Trong thực tế, trên cùng một đối tượng q.sát, nhiều trường .hợp ta phải xét đồng thời nhiều tiêu chuẩn khác nhau. Điều này có nghĩa là trong cùng một phép thử, ta thường phải xét đồng thời n biến n.nhiên khác nhau: $X_1, X_2, \dots, X_n$. Ta gọi véc tơ: $\mathrm{X} = (X_1, X_2, \dots, X_n)$ là véc tơ n.nhiên n chiều.

Với véc tơ n.nhiên: $\mathrm{X} = (X_1, X_2, \dots, X_n)$, gọi hàm n biến:

$$F_X(x_1, \dots, x_n) = P(X_1 < x_1, \dots, X_n < x_n), x_1, x_2, \dots, x_n \in \mathbb{R}$$

là hàm p.phối x.suất của véc tơ n.nhiên n chiều đã cho, hay còn gọi là hàm p.phối x.suất đồng thời của các biến ngẫu nhiên. Ký hiệu $X^{(k)}$ là véc tơ n.nhiên (n – 1) chiều, thu được từ bằng cách bỏ đi t.phần thứ k.

T/c 1: $0 \le F_X(x_1, x_2, \dots, x_n) \le 1$

T/c 2: $F_X(x_1, x_2, \dots, x_n)$ đ.điệu kh.giảm theo từng biến số

T/c 3: $\lim\limits_{x_k \to -\infty} F_X(x_1, x_2, \dots, x_n) = 0, \forall k = 1,2, \dots, n;$

$$\lim_{x_k \to +\infty} F_X(x_1, x_2, \dots, x_n) = F_{X^{(k)}}(x_1, \dots, x_{k-1}, x_{k+1}, \dots, x_n)$$

4.5.2. Một số đặc trưng của véc tơ ngẫu nhiên

Cho véc tơ n.nhiên: $\mathrm{X} = (X_1, X_2, \dots, X_n)$

a. Véc tơ kỳ vọng: Gọi: $\mathrm{EX} = (EX_1, EX_2, \dots, EX_n)$ là véc tơ kỳ vọng của véc tơ n.nhiên n chiều đã cho.

b. Hiệp phương sai và ma trận hiệp phương sai

- $\boldsymbol{cov(X_i, X_j) = E(X_i . X_j) - EX_i . EX_j}$: là covarian (hay hiệp p.sai) của 2 biến X_i, X_j
- $\boldsymbol{\Lambda = \left(cov(X_i, X_j)\right)_{n \times n}}$ là m.trận mô men hay m.trận hiệp p.sai của véc tơ X.

c. Hệ số tương quan và ma trận tương quan

- $\boldsymbol{\rho(X_i, X_j) = \frac{cov(X_i, X_j)}{\sqrt{DX_i}\sqrt{DX_j}}}$ gọi là hệ số t.quan giữa X_i, X_j
- $\boldsymbol{R = \left(\rho(X_i, X_j)\right)_{n \times n}}$ gọi là m.trận t.quan của véc tơ n.nhiên X .

T/c 1: $cov(X_i, X_j) = cov(X_j, X_i)$, vì vậy m.trận hiệp p.sai, m.trận t.quan của véc tơ n.nhiên là các m.trận đ.xứng.
T/c 2: $\rho(aX_i + b, pX_j + q) = \rho(X_i, X_j), \forall a, b, p, q : a.p > 0$
• Nói X_i, X_j không tương quan nếu $\rho(X_i, X_j) = 0$,
khi đó cũng có: $E(X_i.X_j) = EX_i.EX_j; D(X_i \pm X_j) = DX_i + DX_j$
T/c 3: Hai biến n.nhiên đ.lập nhau thì không t.quan.
T/c 4: $\boldsymbol{-1 \leq \rho(X_i, X_j) \leq 1}$, và $|\rho(X_i, X_j)| = 1 \Leftrightarrow \exists \mathrm{a, b, c}$: $aX_i + bX_j = c$
Hệ số t.quan p.ánh x.thế của t.quan và mức độ t.quan tuyến tính giữa 2 biến
* $\rho > 0$: x.thế t.quan thuận: biến này tăng nói chung kéo theo biến kia cũng tăng và ngược lại; $\rho < 0$: x.thế t.quan nghịch: biến này tăng nói chung kéo theo biến kia giảm và ngược lại.
* $|\rho| \uparrow 1$: sự p.thuộc t.quan t.tính giữa 2 biến càng chặt, $|\rho| \downarrow 0$: sự p.thuộc t.quan t.tính giữa hai biến càng lỏng lẻo.
4.5.3. K.n về p.phối có đ.kiện và t.bình có đ.kiện. Giả sử trên mỗi cá thể, ta q.sát 2 t.chuẩn về số lượng là X, Y, tức là xét véc tơ n.nhiên 2 chiều:(X, Y). Khi đó trong tập hợp Ω_x gồm các cá thể có cùng t.chuẩn $X = x$, ta q.sát t.chuẩn Y thì g.trị của t.chuẩn Y (nói chung p.thuộc vào x), ký hiệu là Y_x cũng là một biến n.nhiên.
* *Gọi p.phối x.suất của biến Y_x là p.phối có đ.kiện của biến Y với đ.kiện $X = x$.*
* *Kỳ vọng (g.trị t.bình) của biến Y_x là EY_x được gọi là kỳ vọng có đ.kiện (t.bình có đ.kiện) của biến Y với đ.kiện $X = x$ và được ký hiệu là: $EY_x = E(Y|X = x) = f(x)$*

Khi đó biến n.nhiên $f(X)$ được gọi là kỳ vọng có đ.kiện (hay t.bình có đ.kiện) của biến Y với đ.kiện biến X và được ký hiệu là $f(X) = E(Y|X)$.

* Hàm biến thực: $f(x) = E(Y|X = x)$ được gọi là hàm hồi quy của Y theo X, có ý nghĩa q.trọng trong các bài toán thống kê để n/c sự p.thuộc giữa các biến q.sát.

Hoàn toàn tương tự, ta có k/n về kỳ vọng có đ.kiện của một biến n.nhiên với đ.kiện một véc tơ n.nhiên và hàm h.quy nhiều chiều, bằng cách thay biến n.nhiên X bởi véc tơ n.nhiên X.

4.5.4. Một số phân phối nhiều chiều.

a. Phân phối đa thức k chiều

a1. ĐN. Xét dãy *n* phép thử độc lập , trong mỗi phép thử có một và chỉ một trong k sự kiện $A_1, A_2, \dots, A_k$ xảy ra với xác suất tương ứng $p_1, p_2, \dots, p_k, (p_1 + p_2 + \cdots + p_k = 1)$. Gọi X_i là số lần xuất hiện sự kiện A_i $(i = 1,2, \dots, k)$ trong *n* lần thử. Khi đó luật phân phối của véc tơ ngẫu nhiên $X = (X_1, X_2, \dots, X_k)$ được gọi là phân phối đa thức *k* chiều với các tham số $n, p_1, p_2, \dots, p_{k-1}$.

a2. Phân phối của véc tơ ngẫu nhiên có phân phối đa thức *k* chiều.

Giả sử $X = (X_1, X_2, \dots, X_k)$ là véc tơ có phân phối đa thức k chiều. Ta có $X_i \sim B(n, p_i)(i = 1,2, \dots, k)$ và: $X_1 + X_2 + \cdots + X_k = n$, vì thế $X = (X_1, X_2, \dots, X_k)$ là véc tơ rời rạc k chiều, mỗi giá trị của nó là một điểm k chiều có tọa độ nguyên không âm: $(r_1, r_2, \dots, r_k)$, sao cho: $r_1 + r_2 + \cdots + r_k = n$. Mỗi kết cục thuận lợi cho biến cố $(X_1 = r_1, X_2 = r_2, \dots, X_k = r_k)$ là một dãy gồm n sự kiện liên kết với nhau bởi phép giao, trong đó: A_1 xuất hiện r_1 lần, A_2 xuất hiện r_2 lần, ,…, A_k xuất hiện r_k lần.

. Như vậy số kết cục thuận lợi cho biến cố $(X_1 = r_1, X_2 = r_2, \dots, X_k = r_k)$ chính là số cách chọn r_1 vị trí cho A_1 $(có\ C_n^{r_1}\ cách)$, r_2 vị trí cho A_2 $(có\ C_{n-r_1}^{r_2} cách)$,...,$r_k$ vị trí cho A_k $(có\ C_{n-r_1-r_2-\cdots r_{k-2}}^{r_{k-1}} cách)$ trong dãy n vị trí nói trên, tức là bằng: $C_n^{r_1}.C_{n-r_1}^{r_2}.C_{n-r_1-r_2}^{r_3}\dots..C_{n-r_1-r_2-\cdots r_{k-2}}^{r_{k-1}} = \frac{n!}{r_1!r_2!\dots r_k!}$

Mặt khác do các lần thử độc lập nhau nên mỗi kết cục này đều có xác suất:

$$p_1^{r_1}.p_2^{r_2}\dots..p_k^{r_k}$$

Từ đó nhận được p.phối x.suất của véc tơ ngẫu nhiên có p.phối đa thức k chiều:

$$P(X_1 = r_1, X_2 = r_2, \dots, X_k = r_k) = \frac{n!}{r_1!r_2!\dots r_k!}.p_1^{r_1}.p_2^{r_2}\dots..p_k^{r_k} \qquad (*)$$

$\forall(r_1, r_2, \dots, r_k)$ có tọa độ nguyên, không âm mà: $r_1 + r_2 + \cdots + r_k = n$.

Nhận xét:

- Đối với véc tơ ngẫu nhiên $X = (X_1, X_2, \dots, X_k)$ có phân phối đa thức *k* chiều, với các tham số $n, p_1, p_2, \dots, p_{k-1}$, ta có véc tơ kỳ vọng:

$$EX = (EX_1, EX_2, \dots, EX_k) = (np_1, np_2, \dots, np_k)$$

- Phân phối đa thức với *k* = 2, là phân phối đồng thời của 2 biến nhị thức X_1, $n - X_1$, trong đó $X_1 \sim B(n, p_1)$, $(n - X_1) \sim B(n, 1 - p_1)$.
- Nói chung tham số n được xác định trước, nên phân phối đa thức được xác định bởi k – 1 tham số trong k tham số $p_1, p_2, \dots, p_{k-1}$.

VD 15. Mỗi khách hàng sử dụng dịch vụ A có thể chọn một trong 3 mức phí phục vụ: 100 ngàn đồng, 150 ngàn đồng và 200 ngàn đồng. Được biết lượng khách chọn các mức phí này tương ứng theo tỷ lệ: 5: 3: 2. Có 3 khách hàng vào sử dụng dịch vụ này và sự lựa chọn của họ là độc lập với nhau, Tìm xác suất để có ít nhất 2 khách chọn mức phí 200 ngàn đồng.

Giải: Gọi $X_1, X_2,\ X_3$ lần lượt là số khách chọn mức phí 100 ngàn đồng, 150 ngàn đồng, 200 ngàn đồng trong số 3 khách nói trên. Khi đó véc tơ $X = (X_1, X_2, X_3)$ là véc tơ ngẫu nhiên có phân phối đa thức 3 chiều. Ký hiệu *S* là biến cố trong số 3 khách hàng có ít nhất 2 khách chọn mức phí 200 ngàn đồng, ta có biểu diễn:

$$S = (X_1 = 0, X_2 = 1, X_3 = 2) \cup (X_1 = 1, X_2 = 0, X_3 = 2) \cup (X_1 = 0, X_2 = 0, X_3 = 3)$$

Từ đó suy ra xác suất cần tính:

$$P(S) = \frac{3!}{0!\,1!\,2!}.0{,}3.\,0{,}2^2 + \frac{3!}{1!\,0!\,2!}.0{,}5.\,0{,}2^2 + \frac{3!}{0!\,0!\,3!}.0{,}2^3 = 0{,}104$$

b. Phân phối chuẩn n chiều: Phân phối của véc tơ ngẫu nhiên $\mathrm{X} = (\mathrm{X}_1, \mathrm{X}_2, \dots, \mathrm{X}_\mathrm{n})$ có hàm mật độ xác suất:

$$f(x) = \mathrm{f}(\mathrm{x}_1, \mathrm{x}_2, \dots, \mathrm{x}_\mathrm{n}) = \frac{1}{(2\pi)^{n/2}.\,\|\Lambda\|^{1/2}} exp\left\{\frac{1}{2}.(x-\mu)^T \Lambda^{-1}.(x-\mu)\right\}$$

Trong đó $\mu = \mathrm{EX}$; Λ là ma trận hiệp phương sai của véc tơ ngẫu nhiên X

Khi đó ta viết: $\mathrm{X} \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Lambda})$ hoặc $f(x) = \mathrm{Norm}_\mathrm{x}[\boldsymbol{\mu}, \boldsymbol{\Lambda}]$

c. Phân phối Dirichlet. Phân phối Dirichlet là trường hợp tổng quát của phân phối Beta khi dùng để mô tả các tham số của phân phối Categorical. Phân phối Dirichlet được định nghĩa trên k biến liên tục không âm: $\lambda_1, \lambda_2, \dots, \lambda_k$ có tổng: $\lambda_1 + \lambda_2 + \dots + \lambda_k = 1$. Hàm mật độ của phân phối Dirichlet là:

$$p(\lambda_1, \lambda_2, \dots, \lambda_k) = \frac{\Gamma(\alpha_1 + \alpha_2 + \dots + \alpha_k)}{\Gamma(\alpha_1).\Gamma(\alpha_2)\dots.\Gamma(\alpha_k)} . \lambda_1^{\alpha_1 - 1} \lambda_2^{\alpha_2 - 1} \dots . \lambda_k^{\alpha_k - 1}$$

4.5.5. Véc tơ ngẫu nhiên hai chiều. Xét véc tơ ngẫu nhiên 2 chiều (X, Y)

a. Véc tơ ngẫu nhiên rời rạc 2 chiều (X, Y): X và Y đều là các biến rời rạc. Giả sử tập giá trị của X là: $\{x_1, x_2, \dots, x_m\}$***, tập giá trị của Y là:*** $y_1, y_2, \dots, y_n$. Đặt:

$$p_{ij} = P(X = x_i, Y = y_j); p_{*j} = \sum_i p_{ij} = P\left(Y = y_j\right); p_{i*} = \sum_j p_{ij} = P(X = x_i)$$

Khi đó: $\sum_i \sum_j p_{ij} = \sum_j p_{*j} = \sum_i p_{i*} = 1$ và ta có các bảng phân phối xác suất của véc tơ ngẫu nhiên rời rạc 2 chiều (X, Y), bảng phân phối biên, bảng phân phối có điều kiện:

X \ Y	y_1	y_2	$\cdots$	y_n	p_{i*}
x_1	p_{11}	p_{12}	$\cdots$	p_{1n}	p_{1*}
x_2	p_{21}	p_{22}	$\cdots$	p_{2n}	p_{2*}
$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\vdots$
x_m	p_{m1}	p_{m2}	$\cdots$	p_{mn}	p_{m*}
p_{*j}	p_{*1}	p_{*2}	$\cdots$	p_{*n}	1

Bảng p.phối (đồng thời) của véc tơ ngẫu nhiên rời rạc 2 chiều (X, Y)

X	x_1	x_2	$\cdots$	x_m	Σ
p_{i*}	p_{1*}	p_{2*}	$\cdots$	p_{m*}	1
Bảng p.phối biên của X					

Y	y_1	y_2	$\cdots$	y_n	Σ
p_{*j}	p_{*1}	p_{*2}	$\cdots$	p_{*n}	1
Bảng p.phối biên của Y					

Y_{x_i}	y_1	y_2	$\cdots$	y_n	Σ
P_{x_i}	p_{i1}/p_{i*}	p_{i2}/p_{i*}	$\cdots$	p_{in}/p_{i*}	1
Bảng phân phối có điều kiện của Y với đ/kiện $X \coloneqq x_i$					

Các công thức tính: $EX = \sum_{i,j} p_{ij} . x_i = \sum_i p_{i*} . x_i; EY = \sum_{i,j} p_{ij} . y_j = \sum_j p_{*j} . y_j$

$$E(X.Y) = \sum_{i,j} p_{ij} . x_i . y_j; EX^2 = \sum_i p_{i*} . x_i^2; EY^2 = \sum_i p_{*j} . y_j^2$$

$$E(Y|X = x_i) = E\left(Y_{x_i}\right) = \frac{1}{p_{i*}} \sum_j p_{ij} . y_j$$

b. Véc tơ ngẫu nhiên 2 chiều (X, Y) với X rời rạc, Y liên tục: Giả sử X có tập giá trị $\{x_1, x_2, \dots, x_m\}$**,** Y có hàm mật độ xs $f(y)$**.** Khi đó: $f(x_i, y)$ là hàm mật độ xs của Y ứng với $X = x_i$, (hàm mật độ xs của biến Y_{x_i}) và $\sum_i \int_{-\infty}^{+\infty} f(x_i, y)\, dy = \int_{-\infty}^{+\infty} \sum_i f(x_i, y)\, dy = 1$

$$E(X.Y) = \int_{-\infty}^{+\infty} \sum_i x_i . y\, f(x_i, y)\; dy = \sum_i x_i \int_{-\infty}^{+\infty} y . f(x_i, y) dy$$

***c.* Véc tơ ngẫu nhiên 2 chiều (X, Y) liên tục, với hàm mật độ** $f(x, y)$**.** Khi đó:

X có hàm mật độ xs: $p(x) = \int_{-\infty}^{+\infty} f(x, y)\, dy$, Y có hàm mật độ xs: $q(y) = \int_{-\infty}^{+\infty} f(x, y)\, dx$***;***

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y)\, dxdy = \int_{-\infty}^{+\infty} p(x)\, dx = \int_{-\infty}^{+\infty} q(y)\, dy = 1;$$

$$E(X.Y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x . y f(x, y)\, dxdy$$

Vd 16: Một người cân nhắc giữa đầu tư vào cổ phiếu hay trái phiếu. Lãi suất X(%) của cổ phiếu và l.suất Y(%) của trái phiếu là véc tơ n.nhiên rời rạc có bảng p.phối x.suất như ở bên:

X \ Y	6	8	10
-10	0	0	0,1
0	0	0,1	0,1
10	0,1	0,3	0
20	0,1	0,2	0

a. Nếu đ.tư toàn bộ số tiền vào c.phiếu thì l.suất kỳ vọng và độ lệch t.chuẩn của l.suất là bao nhiêu?

b. Nếu đ.tư toàn bộ số tiền vào t.phiếu thì l.suất kỳ vọng và độ lệch t.chuẩn của l.suất là bao nhiêu?

c. Nên đ.tư vào c.phiếu và t.phiếu theo tỷ lệ như thế nào để tổng l.suất kỳ vọng lớn nhất?

d. Nếu muốn đ.tư để mức rủi ro về l.suất là nhỏ nhất thì nên đ.tư theo tỷ lệ nào?

Giải. a. Nếu đầu tư toàn bộ số tiền vào cổ phiếu, thì mức lãi X(%) là biến ngẫu nhiên có bảng p.phối xs:

X	-10	0	10	20	Σ
P	0,1	0,2	0,4	0,3	1

Khi đó thì l.suất kỳ vọng là:

EX = -10.0,1 + 0.0,2 + 10.0,4 + 20.0,3 = 9(%)

Độ lệch tiêu chuẩn của lãi suất là: se(X) = $\sqrt{\mathrm{VarX}} = \sqrt{\mathrm{EX}^2 - (\mathrm{EX})^2}$

Tính $\mathrm{EX}^2 = 100.0{,}1$ + 0.0,2 + 100.0,4 + 400.0,3 = 170. Vậy se(X) = 9,434(%)

b. Nếu đầu tư toàn bộ số tiền vào trái phiếu, thì mức lãi Y(%) là biến ngẫu nhiên có bảng p.phối xs:

Y	6	8	10	Σ
P	0,2	0,6	0,2	1

Khi đó thì l.suất kỳ vọng là: EY = 6.0,2 + 8.0,6 + 10.0,2 = 8(%)

Độ lệch tiêu chuẩn của lãi suất là: se(Y) = $\sqrt{\mathrm{VarY}} = \sqrt{\mathrm{EY}^2 - (\mathrm{EY})^2}$

Tính $\mathrm{EY}^2 = 36.0{,}2$ + 64.0,6 + 100.0,2 = 65,6. Vậy se(Y) = 24,3311

c. Gọi s là tỉ lệ vốn đầu tư vào cổ phiếu thì (1 – s) là tỉ lệ đầu tư vào trái phiếu. Khi đó tổng lãi suất kì vọng là: E[s.X + (1-s)Y] = s.EX + (1 – s)EY = 9s + 8.(1 – s) = s + 8

Vậy kỳ vọng tổng lãi suất lớn nhất là $\max\limits_{0\le s\le 1}\{\mathrm{E}[\mathrm{s.X} + (1-\mathrm{s}).\mathrm{Y}]\} = 9(\%)$, đạt được khi s = 1, tức là khi đầu tư 100% vốn vào cổ phiếu.

d. Khi đầu tư vốn theo tỉ lệ s cho cổ phiếu và (1 – s) cho trái phiếu thì tổng lãi suất là s.X + (1 – s).Y và mức rủi ro sẽ là:

Var{s.X + (1 – s).Y} = $\mathrm{E}[\mathrm{s.X} + (1-\mathrm{s}).\mathrm{Y}]^2 - \left\{\mathrm{E}\left[[\mathrm{s.X} + (1-\mathrm{s}).\mathrm{Y}]\right]\right\}^2$

Có: $\left\{\mathrm{E}\left[[\mathrm{s.X} + (1-\mathrm{s}).\mathrm{Y}]\right]\right\}^2 = \{s+8\}^2 = s^2 + 16s + 64$

$$\mathrm{E}[\mathrm{s.X} + (1-\mathrm{s}).\mathrm{Y}]^2 = \mathrm{s}^2\mathrm{EX}^2 + (1-\mathrm{s})^2\mathrm{EY}^2 + 2s(1-\mathrm{s})\mathrm{E}(\mathrm{X.Y})$$
$$= 170\mathrm{s}^2 + 65{,}6(1-\mathrm{s})^2 + (2s - 2s^2).\mathrm{E}(\mathrm{X.Y})$$

Tính: $\mathrm{E}(\mathrm{X.Y}) = \sum_j \sum_i p_{ij}x_iy_j = -100.0{,}1 + 60.0{,}1 + 80.0{,}3 + 120.0{,}1 + 160.0{,}2 = 64$

$\mathrm{E}[\mathrm{s.X} + (1-\mathrm{s}).\mathrm{Y}]^2 = 170\mathrm{s}^2 + 65{,}6(1-\mathrm{s})^2 + 64(2s - 2s^2) = 107{,}6\mathrm{s}^2 - 3{,}2\mathrm{s} + 65{,}6$

Mức rủi ro: Var{s.X + (1 – s).Y} $= 106{,}6s^2 - 19{,}2\mathrm{s} + 1{,}6$

Mức rủi ro thấp nhất là: $\min\limits_{0\le s\le 1}$ Var{s.X + (1 – s).Y} $= \min\limits_{0\le s\le 1}(106{,}6s^2 - 19{,}2\mathrm{s} + 1{,}6)$

đạt tại $\mathrm{s} = 48/533$.