

BÀI GIẢNG LÝ THUYẾT THỐNG KÊ

Bộ Môn Toán Ứng Dụng

Khoa Toán Tin - Trường đại học Sư phạm Hà Nội

Mục lục

1	Một số kiến thức xác suất cơ sở	4
1.1	Biến ngẫu nhiên rời rạc	4
1.1.1	Định nghĩa	4
1.1.2	Ví dụ	5
1.2	Hàm phân phối của biến ngẫu nhiên	7
1.2.1	Định nghĩa	7
1.2.2	Ví dụ	7
1.2.3	Kỳ vọng của biến có phân phối liên tục tuyệt đối	9
1.3	Một số bất đẳng thức	9
1.4	Biến ngẫu nhiên độc lập	10
1.5	Hệ số tương quan	11
1.6	Các dạng hội tụ của dãy biến ngẫu nhiên	12
1.7	Luật số lớn	12
1.7.1	Luật yếu số lớn	12
1.7.2	Luật mạnh số lớn	13
1.8	Hàm đặc trưng	13
1.9	Định lý giới hạn trung tâm	15
1.10	Phân phối chuẩn nhiều chiều	15
1.11	Phân phối Gamma, khi bình phương, student and F	16
1.11.1	Phân phối Gamma	16
1.11.2	Phân phối khi bình phương	17
1.11.3	Phân phối Student	18
1.11.4	Phân phối F	18
1.12	Phân phối của trung bình và phương sai mẫu	19
2	Ước lượng tham số	21
2.1	Mẫu và đặc trưng mẫu	21
2.2	Ước lượng điểm	23
2.2.1	Thống kê	23
2.2.2	Các ước lượng điểm	25

2.2.3	Khoảng tin cậy	25
2.3	Phương pháp tìm ước lượng	31
2.3.1	Ước lượng hợp lý cực đại	31
2.3.2	Phương pháp moment	32
2.4	Chặn dưới cho phương sai	33
3	Kiểm định giả thiết	36
3.1	Giới thiệu	36
3.2	Phương pháp tìm kiểm định	39
3.2.1	Kiểm định tỉ số hợp lý	39
3.3	Phương pháp tiêu chuẩn đánh giá	40
3.3.1	Tiêu chuẩn mạnh nhất	40
3.3.2	Tiêu chuẩn mạnh đều nhất	42
3.3.3	Tỉ số hợp lý đơn điệu	44
3.4	Một số bài toán kiểm định thường gặp	46
3.4.1	Kiểm định giả thiết cho giá trị trung bình của phân phối chuẩn với phương sai σ^2 đã biết	46
3.4.2	Kiểm định giả thiết cho giá trị trung bình của phân phối chuẩn với phương sai σ^2 chưa biết	49
3.4.3	Kiểm định giả thiết cho phương sai của một phân phối chuẩn	51
3.4.4	Kiểm định giả thiết cho tỉ lệ quần thể	52
3.5	Một số bài toán kiểm định giả thiết cho hai mẫu	53
3.5.1	Kiểm định cho hiệu giá trị trung bình của hai phân phối chuẩn với phương sai đã biết	53
3.5.2	Kiểm định cho hiệu giá trị trung bình của hai phân phối chuẩn với phương sai chưa biết	55
3.5.3	Tiêu chuẩn kiểm định t -test theo cặp	57
3.5.4	Kiểm định cho phương sai của hai quần thể có phân phối chuẩn	58
3.5.5	Kiểm định giả thiết cho hai tỉ lệ quần thể	59
3.6	Kiểm định khi bình phương	60
3.6.1	Tiêu chuẩn phù hợp	60
3.6.2	Kiểm định sự độc lập	65
3.6.3	Kiểm định tính thuần nhất	67
4	Hồi quy	69
4.1	Hồi quy tuyến tính đơn	69
4.1.1	Mô hình hồi quy tuyến tính đơn	69
4.1.2	Khoảng tin cậy cho σ^2	71
4.1.3	Khoảng tin cậy cho β_1	71
4.1.4	Khoảng tin cậy cho β_0	71

4.1.5	Khoảng dự báo	72
-------	-------------------------	----

TaiLieu.vn

Chương 1

Một số kiến thức xác suất cơ sở

1.1 Biến ngẫu nhiên rời rạc

1.1.1 Định nghĩa

Mỗi bnn là một quan sát về kết quả của phép thử. Giả sử X nhận các giá trị là $x_1, x_2, \dots$. Đặt $p_i^X = \mathbb{P}[X = x_i]$, $i \geq 1$. Ta có

$$\mathbb{P}^X(A) = \sum_{x_i \in A} \mathbb{P}[X = x_i] = \sum_{x_i \in A} p_i^X.$$

Ta viết phân phối của X dưới dạng bảng như sau

x	x_1	x_2	$\dots$	x_n	$\dots$
$\mathbb{P}[X = x]$	p_1^X	p_2^X	$\dots$	p_n^X	$\dots$

Định nghĩa 1.1.1. Giả sử X là bnn nhận các giá trị là $x_1, x_2, \dots$. Khi đó, ta gọi *kỳ vọng* của bnn X là đại lượng

$$\mathbb{E}[X] := \sum_i x_i \mathbb{P}[X = x_i] = \sum_i x_i p_i^X$$

nếu tổng trên có nghĩa, tức là khi một trong các trường hợp sau được thỏa mãn

- Ω có hữu hạn phần tử;
- chuỗi $\sum_i |x_i| p_i^X < \infty$;
- $x_i \geq 0$ với mọi i (trong trường hợp này $\mathbb{E}X$ có thể nhận giá trị $+\infty$).

Kí hiệu $\mathcal{L}^1$ là không gian tất cả các bnn có kỳ vọng hữu hạn trên $(\Omega, \mathcal{A}, \mathbb{P})$. Ta có các tính chất đơn giản sau

Định lý 1.1.2. Giả sử $X, Y \in \mathcal{L}^1$. Ta có các khẳng định sau

1. $\mathcal{L}^1$ là một không gian véc tơ và

$$\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y], \quad \forall a, b \in \mathbb{R}.$$

2. Nếu $X \geq 0$ thì $\mathbb{E}[X] \geq 0$. Hơn nữa, nếu $X \geq Y$ thì $\mathbb{E}[X] \geq \mathbb{E}[Y]$.
3. Nếu Z là bị chặn thì $Z \in \mathcal{L}^1$. Hơn nữa, nếu bnn Z' thỏa mãn $|Z'| \leq X$ thì $Z' \in \mathcal{L}^1$.
4. Nếu $X = \mathbb{I}_A$ là hàm chỉ tiêu của tập A thì $\mathbb{E}[X] = \mathbb{P}(A)$.
5. Giả sử $\varphi: \mathbb{R} \rightarrow \mathbb{R}$. Khi đó

$$\mathbb{E}[\varphi(X)] = \sum_i \varphi(x_i) p_i^X = \sum_{w \in \Omega} \varphi(X(w)) p_w$$

nếu tổng trên có nghĩa.

Nhận xét 1. Ta thấy nếu $\mathbb{E}[X^2] = \sum_i x_i^2 p_i^X < \infty$ thì

$$\mathbb{E}[|X|] = \sum_i |x_i| p_i^X \leq \frac{1}{2} \sum_i (|x_i|^2 + 1) p_i^X = \frac{1}{2} (\mathbb{E}[X^2] + 1) < \infty.$$

Định nghĩa 1.1.3. Phương sai của bnn X là

$$DX = \mathbb{E}[(X - \mathbb{E}[X])^2]$$

Áp dụng tính chất tuyến tính của kì vọng, ta có

$$DX = \mathbb{E}(X^2) - (\mathbb{E}X)^2,$$

và do đó

$$DX = \sum_i x_i^2 p_i^X - \left(\sum_i x_i p_i^X \right)^2.$$

1.1.2 Ví dụ

Phân phối nhị thức $B(n, p)$

Trong hộp có L bi trắng và M bi vàng. Thực hiện phép thử sau: lấy ra ngẫu nhiên từ hộp một bi, ghi lại màu của nó rồi trả lại hộp. Lập lại phép thử n lần và gọi X là số bi vàng trong n bi đã được lấy ra đó. Ta sẽ đi tìm phân phối xác suất của bnn X .

Trước hết ta xác định không gian xác suất của phép thử. Gọi các viên bi trắng là $T_1, T_2, \dots, T_L$ và các viên bi vàng là $V_1, \dots, V_M$. Không gian xác suất là

$$\Omega = \{(w_1, \dots, w_n) : w_i \in \{T_1, \dots, T_L, V_1, \dots, V_M\}, i = 1, \dots, n\}$$

cùng với độ đo xác suất đều. Ta có số các phần tử của Ω là $(L + M)^n$. Số lượng phần tử của tập $\{w \in \Omega : X(w) = k\}$ bằng $C_n^k M^k L^{n-k}$ với $k = 0, \dots, n$. Do vậy

$$\mathbb{P}[X = k] = C_n^k \left(\frac{M}{M+L} \right)^k \left(\frac{L}{M+L} \right)^{n-k}, \quad k = 0, \dots, n.$$

Đặt $p = \frac{M}{M+L}$ ta có phân phối của X là

$$\mathbb{P}[X = k] = C_n^k p^k (1-p)^{n-k}, \quad k = 0, \dots, n.$$

Ta gọi phân phối của X là *phân phối nhị thức* với cỡ n và xác suất p , kí hiệu là $X \sim B(n, p)$. Một cách tổng quát, $B(n, p)$ là phân phối của số lần thành công trong một dãy n phép thử độc lập liên tiếp, trong đó xác suất thành công của mỗi một phép thử đều bằng $p \in [0, 1]$.

Ta có

$$\begin{aligned} \mathbb{E}X &= \sum_{k=0}^n k \mathbb{P}[X = k] = \sum_{k=0}^n k C_n^k p^k (1-p)^{n-k} \\ &= np \sum_{k=1}^n C_{n-1}^{k-1} p^{k-1} (1-p)^{n-k} = np, \end{aligned}$$

và

$$\begin{aligned} \mathbb{E}(X^2) &= \sum_{k=0}^n k^2 \mathbb{P}[X = k] = \sum_{k=0}^n k^2 C_n^k p^k (1-p)^{n-k} \\ &= n(n-1)p^2 \sum_{k=2}^n C_{n-2}^{k-2} p^{k-2} (1-p)^{n-k} + np \sum_{k=1}^n C_{n-1}^{k-1} p^{k-1} (1-p)^{n-k} \\ &= n(n-1)p^2 + np, \end{aligned}$$

vậy nên $DX = np(1-p)$.

Phân phối hình học

Thực hiện nhiều lần một dãy các phép thử độc lập cho đến khi thành công thì dừng lại. Giả sử xác suất thành công ở mỗi lần thử đều bằng nhau và bằng p . Gọi X là số phép thử đã thực hiện cho đến khi có phép thử đầu tiên thành công. Bnn X được gọi là có phân phối hình học với tham số $p \in [0, 1]$ và

$$\mathbb{P}[X = n] = p(1-p)^{n-1}, \quad n = 1, 2, \dots$$

Tính toán đơn giản, ta được

$$\mathbb{E}X = \frac{1}{p}, \quad DX = \frac{1-p}{p^2}.$$

Phân phối Poisson $Poi(\lambda)$

Bnn X được gọi là có phân phối Poisson với tham số $\lambda > 0$, kí hiệu là $X \sim Poi(\lambda)$ nếu X nhận giá trị trong $\mathbb{N}$ và

$$\mathbb{P}[X = k] = \frac{e^{-\lambda} \lambda^k}{k!}, \quad k = 0, 1, \dots$$

Sử dụng khai triển Taylor của hàm e^x ta chứng minh được

$$\mathbb{E}X = DX = \lambda.$$

1.2 Hàm phân phối của biến ngẫu nhiên

1.2.1 Định nghĩa

Định nghĩa 1.2.1. Giả sử X là bnn nhận giá trị trên $\mathbb{R}$. Hàm số

$$F_X(x) = \mathbb{P}[X < x], \quad x \in \mathbb{R},$$

được gọi là *hàm phân phối* của bnn X .

Dễ thấy hàm $F = F_X$ thỏa mãn:

1. Nếu $x \leq y$ thì $F(y) - F(x) = \mathbb{P}[x \leq X < y]$;
2. F đơn điệu tăng;
3. F liên tục trái và có giới hạn phải tại mọi điểm;
4. $\lim_{x \rightarrow -\infty} F(x) = 0, \lim_{x \rightarrow +\infty} F(x) = 1$.

Ngược lại, với bất kì hàm số $F : \mathbb{R} \rightarrow [0, 1]$ thỏa mãn ba tính chất trên thì tồn tại một độ đo xác suất μ trên $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ sao cho $F(x) = \mu((-\infty, x))$, với mọi $x \in \mathbb{R}$ (Xem [13], mục 2.5.2).

Trong tiết trước ta đã làm quen với phân phối của các bnn rời rạc. Sau đây ta giới thiệu một loại bnn thường gặp khác là bnn có phân phối liên tục tuyệt đối.

Định nghĩa 1.2.2. Bnn X được gọi là có phân phối liên tục tuyệt đối với hàm mật độ f_X nếu

$$F_X(a) = \mathbb{P}[X < a] = \int_{-\infty}^a f_X(x) dx, \quad \forall a \in \mathbb{R}.$$

Hàm mật độ $f = f_X$ thỏa mãn các tính chất sau đây.

1. $f(x) \geq 0$ với mọi $x \in \mathbb{R}$ và $\int_{-\infty}^{+\infty} f(x) dx = 1$.
2. $\mathbb{P}[a < X < b] = \int_a^b f(x) dx$ với mọi số thực $a < b$. Hơn nữa, với mọi tập Borel $A \in \mathcal{B}(\mathbb{R})$ ta có

$$\mathbb{P}[X \in A] = \int_A f(x) dx. \quad (1.1)$$

Từ tính chất thứ hai suy ra nếu bnn X có phân phối liên tục tuyệt đối thì $\mathbb{P}[X = a] = 0$ với mọi $a \in \mathbb{R}$.

1.2.2 Ví dụ

Sau đây ta liệt kê một số phân phối liên tục tuyệt đối thường gặp nhất.

Phân phối đều $U[a, b]$

Với mỗi cặp số thực $a < b$. Bnn X được gọi là có *phân phối đều* trên đoạn $[a, b]$, kí hiệu là $X \sim U[a, b]$, nếu X nhận giá trị thuộc đoạn $[a, b]$ và có hàm mật độ

$$f_X(x) = \frac{1}{b-a} \mathbb{I}_{[a,b]}(x).$$

Hàm phân phối của X là

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

Phân phối mũ $Exp(\lambda)$

Giả sử $\lambda > 0$. Bnn X được gọi là có *phân phối mũ* với tham số λ , kí hiệu là $X \sim Exp(\lambda)$, nếu X nhận giá trị trên tập $(0, \infty)$ có hàm mật độ

$$f_X(x) = \lambda^{-1} e^{-x/\lambda} \mathbb{I}_{(0,\infty)}(x).$$

Hàm phân phối của X là

$$F_X(x) = (1 - \lambda^{-1} e^{-x/\lambda}) \mathbb{I}_{(0,\infty)}(x).$$

Phân phối chuẩn $\mathcal{N}(a, \sigma^2)$

Bnn X được gọi là có *phân phối chuẩn* với tham số a, σ^2 ($\sigma > 0$), kí hiệu là $X \sim \mathcal{N}(a, \sigma^2)$, nếu hàm mật độ của nó có dạng

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-a)^2}{2\sigma^2}}, \quad x \in \mathbb{R}.$$

Phân phối chuẩn $\mathcal{N}(0, 1)$ được gọi là *phân phối chuẩn tắc*.

Phân phối Gamma $\mathcal{G}(\alpha, p)$

Bnn X được gọi là có *phân phối Gamma* với tham số α, p ($\alpha, p > 0$), kí hiệu là $X \sim \mathcal{G}(\alpha, p)$, nếu hàm mật độ của nó có dạng

$$f_X(x) = \frac{\alpha^p x^{p-1} e^{-\alpha x}}{\Gamma(p)} \mathbb{I}_{(0,\infty)}(x),$$

trong đó $\Gamma(p) = \int_0^\infty x^{p-1} e^{-x} dx$.

Phân phối mũ là trường hợp đặc biệt của phân phối Gamma khi $p = 1$. Phân phối $\mathcal{G}(\frac{1}{2}, \frac{n}{2})$ được gọi là *phân phối khi bình phương* với n bậc tự do (viết gọi là $\chi^2(n)$).

1.2.3 Kỳ vọng của bnn có phân phối liên tục tuyệt đối

Định lý 1.2.3. Giả sử bnn X có hàm mật độ f và $h : \mathbb{R} \rightarrow \mathbb{R}$ là một hàm Borel. Khi đó nếu $\mathbb{E}(|h(X)|) < \infty$ hoặc h là không âm thì

$$\mathbb{E}(h(X)) = \int h(x)f(x)dx.$$

Ví dụ 1.2.4. Giả sử bnn $X \sim \text{Exp}(1)$. Khi đó áp dụng Định lý 1.2.3 cho $h(x) = x$ và $h(x) = x^2$ ta được

$$\mathbb{E}X = \int_0^\infty xe^{-x}dx = 1,$$

và

$$\mathbb{E}X^2 = \int_0^\infty x^2e^{-x}dx = 2.$$

1.3 Một số bất đẳng thức

Định lý 1.3.1. 1. (Bất đẳng thức Cauchy-Schwarz) Nếu $X, Y \in L^2$ thì $XY \in L^1$ và

$$|\mathbb{E}(XY)|^2 \leq \mathbb{E}(X^2)\mathbb{E}(Y^2). \quad (1.2)$$

2. $L^2 \subset L^1$ và nếu $X \in L^2$ thì $(\mathbb{E}X)^2 \leq \mathbb{E}(X^2)$.

3. L^2 là một không gian véc tơ: nếu $X, Y \in L^2$ và $\alpha, \beta \in \mathbb{R}$ thì $\alpha X + \beta Y \in L^2$.

Chứng minh. Nếu $\mathbb{E}(X^2)\mathbb{E}(Y^2) = 0$ thì $XY = 0$ h.c.c nên $|\mathbb{E}(XY)|^2 = \mathbb{E}(X^2)\mathbb{E}(Y^2) = 0$.

Nếu $\mathbb{E}(X^2)\mathbb{E}(Y^2) \neq 0$, áp dụng bất đẳng thức $2|ab| \leq a^2 + b^2$ với $a = X/\sqrt{\mathbb{E}(X^2)}$ và $b = Y/\sqrt{\mathbb{E}(Y^2)}$ rồi lấy kỳ vọng hai vế, ta được

$$2\mathbb{E}\left(\frac{XY}{\sqrt{\mathbb{E}(X^2)\mathbb{E}(Y^2)}}\right) \leq \mathbb{E}\left(\frac{X^2}{\mathbb{E}(X^2)}\right) + \mathbb{E}\left(\frac{Y^2}{\mathbb{E}(Y^2)}\right) = 2.$$

Từ đó ta được (1.2).

Áp dụng (1.2) cho $Y = 1$ ta được khẳng định thứ hai. Khẳng định thứ ba suy ra từ (1.2) và tính tuyến tính của kỳ vọng. $\square$

Nếu $X \in L^2$, ta kí hiệu

$$DX = \mathbb{E}[(X - \mathbb{E}X)^2].$$

DX được gọi là *phương sai* của bnn X . Nó đặc trưng cho độ phân tán của X xung quanh giá trị trung bình $\mathbb{E}X$. Sử dụng tính tuyến tính của kỳ vọng, ta có $DX = \mathbb{E}(X^2) - (\mathbb{E}X)^2$.

Định lý 1.3.2. 1. (Bất đẳng thức Markov) Giả sử $X \in L^1$, khi đó với mọi $a > 0$,

$$\mathbb{P}(|X| \geq a) \leq \frac{\mathbb{E}(|X|)}{a}.$$

2. (Bất đẳng thức Chebyshev) Giả sử $X \in L^2$, khi đó với mọi $a > 0$,

$$\mathbb{P}(|X - \mathbb{E}X| \geq a) \leq \frac{DX}{a^2}.$$

Chứng minh. 1) Do $a\mathbb{1}_{\{|X| \geq a\}}(w) \leq |X(w)|\mathbb{1}_{\{|X| \geq a\}}(w) \leq |X(w)|$ với mọi $w \in \Omega$. Lấy kỳ vọng hai vế ta được $a\mathbb{P}(|X| \geq a) \leq \mathbb{E}(|X|)$.

2) Áp dụng Bất đẳng thức Markov, ta có

$$\mathbb{P}(|X - \mathbb{E}X| \geq a) = \mathbb{P}(|X - \mathbb{E}X|^2 \geq a^2) \leq \frac{DX}{a^2}.$$

□

1.4 Biến ngẫu nhiên độc lập

Định nghĩa 1.4.1. 1. Các σ -đại số con $(\mathcal{A}_i)_{i \in I}$ của $\mathcal{A}$ được gọi là *độc lập* nếu với mọi tập con hữu hạn J của I và với mọi $A_i \in \mathcal{A}_i$, ta có

$$\mathbb{P}(\cap_{i \in J} A_i) = \prod_{i \in J} \mathbb{P}(A_i).$$

2. Các bnn $(X_i)_{i \in I}$ nhận giá trị trong $(E_i, \mathcal{E}_i)$ được gọi là *độc lập* nếu các σ -đại số $(X_i^{-1}(\mathcal{E}_i))_{i \in I}$ là độc lập.

Định lý 1.4.2. Giả sử X và Y là hai bnn bất kỳ. Các khẳng định sau là tương đương.

(i) X độc lập với Y ;

(ii) $F_{X,Y}(x,y) = F_X(x)F_Y(y)$ với mọi $x, y \in \mathbb{R}$;

(iii) $f(X)$ và $g(Y)$ là độc lập với mọi hàm Borel $f, g: \mathbb{R} \rightarrow \mathbb{R}$;

(iv) $\mathbb{E}[f(X)g(Y)] = \mathbb{E}[f(X)]\mathbb{E}[g(Y)]$ với mọi hàm Borel $f, g: \mathbb{R} \rightarrow \mathbb{R}$ cùng bị chặn hoặc cùng không âm.

Hệ quả 1.4.3. Giả sử (X, Y) là véc tơ ngẫu nhiên có phân phối liên tục tuyệt đối. Khi đó X và Y là độc lập khi và chỉ khi

$$f_{(X,Y)}(x,y) = f_X(x)f_Y(y), \quad \text{với mọi } x, y \in \mathbb{R}. \quad (1.3)$$

Ví dụ 1.4.4. Giả sử X và Y là hai bnn độc lập có phân phối Poisson với tham số lần lượt là λ và μ ($\lambda, \mu > 0$). Với mọi số tự nhiên n ,

$$\begin{aligned} \mathbb{P}(X + Y = n) &= \sum_{k=0}^n \mathbb{P}(X = k, Y = n - k) = \sum_{k=0}^n \mathbb{P}(X = k)\mathbb{P}(Y = n - k) \\ &= \sum_{k=0}^n \frac{e^{-\lambda} \lambda^k}{k!} \frac{e^{-\mu} \mu^{n-k}}{(n-k)!} = \frac{e^{-\lambda-\mu}}{n!} \sum_{k=0}^n C_n^k \lambda^k \mu^{n-k} = \frac{e^{-\lambda-\mu} (\lambda + \mu)^n}{n!}. \end{aligned}$$

Vậy nên $X + Y$ có phân phối Poisson với tham số $\lambda + \mu$.

1.5 Hệ số tương quan

Định nghĩa 1.5.1. Hiệp phương sai của hai bnn $X, Y \in L^2$ là

$$\text{cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}X)(Y - \mathbb{E}Y)].$$

Hệ số tương quan của hai bnn $X, Y \in L^2$ là

$$\rho(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{DXDY}}.$$

Sử dụng tính tuyến tính của kì vọng, ta có

$$\text{cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}X\mathbb{E}Y.$$

Mặt khác, sử dụng bất đẳng thức Cauchy-Schwarz ta có $|\text{cov}(X, Y)| \leq \sqrt{DXDY}$, nên suy ra

$$|\rho(X, Y)| \leq 1.$$

Hệ số tương quan $\rho(X, Y)$ đặc trưng cho độ phụ thuộc (tuyến tính) của hai bnn X và Y . Nếu $\rho(X, Y) \approx 1$ thì X và Y gần như tỉ lệ thuận với nhau, trong khi nếu $\rho(X, Y) \approx -1$ thì X và Y gần như tỉ lệ nghịch. Trong trường hợp $\rho(X, Y) = 0$ ta nói X và Y là *không tương quan*. Theo Định lí 1.4.2 ta có nếu X và Y là hai biến ngẫu nhiên độc lập thì $\text{cov}(X, Y) = 0$ nên chúng không tương quan. Điều ngược lại chưa chắc đã đúng.

Ví dụ 1.5.2. Giả sử X và Y là hai bnn độc lập có cùng phân phối chuẩn $N(0, 1)$. Đặt $Z = XY$ và $T = X - Y$. Ta có

$$\text{cov}(Z, T) = \mathbb{E}(XY(X - Y)) - \mathbb{E}(XY)\mathbb{E}(X - Y) = 0,$$

và

$$\text{cov}(Z, T^2) = \mathbb{E}(XY(X - Y)^2) - \mathbb{E}(XY)\mathbb{E}((X - Y)^2) = -2,$$

vì $\mathbb{E}(XY) = \mathbb{E}X\mathbb{E}Y = 0$, $\mathbb{E}(X^3Y) = \mathbb{E}(X^3)\mathbb{E}Y = 0$, $\mathbb{E}(XY^3) = \mathbb{E}X\mathbb{E}(Y^3) = 0$ và $\mathbb{E}(X^2Y^2) = \mathbb{E}(X^2)\mathbb{E}(Y^2) = 1$. Vậy Z và T là hai bnn không tương quan nhưng không độc lập.

Mệnh đề 1.5.3. Giả sử $(X_n)_{n \geq 1}$ là một dãy các bnn đôi một không tương quan. Khi đó

$$D(X_1 + \dots + X_n) = D(X_1) + \dots + D(X_n).$$

Chứng minh. Ta có

$$\begin{aligned} D(X_1 + \dots + X_n) &= \mathbb{E}\left[\left((X_1 - \mathbb{E}X_1) + \dots + (X_n - \mathbb{E}X_n)\right)^2\right] \\ &= \sum_{i=1}^n \mathbb{E}[(X_i - \mathbb{E}X_i)^2] + 2 \sum_{1 \leq i < j \leq n} \mathbb{E}[(X_i - \mathbb{E}X_i)(X_j - \mathbb{E}X_j)] \\ &= \sum_{i=1}^n \mathbb{E}[(X_i - \mathbb{E}X_i)^2] = \sum_{i=1}^n D(X_i), \end{aligned}$$

vì $\mathbb{E}[(X_i - \mathbb{E}X_i)(X_j - \mathbb{E}X_j)] = \mathbb{E}(X_i X_j) - \mathbb{E}(X_i)\mathbb{E}(X_j) = 0$ với mọi $1 \leq i < j \leq n$. □

1.6 Các dạng hội tụ của dãy biến ngẫu nhiên

Định nghĩa 1.6.1. Giả sử $(X_n)_{n \geq 1}$ là dãy bnn xác định trên không gian xác suất $(\Omega, \mathcal{A}, \mathbb{P})$. Dãy (X_n) được gọi là

- *hội tụ hầu chắc chắn* đến bnn X , kí hiệu là $X_n \xrightarrow{h.c.c} X$ hay $\lim_n X_n = X$ h.c.c, nếu

$$\mathbb{P}\left(w : \lim_{n \rightarrow \infty} X_n(w) = X(w)\right) = 1;$$

- *hội tụ theo xác suất* đến bnn X , kí hiệu là $X_n \xrightarrow{\mathbb{P}} X$, nếu với mọi $\epsilon > 0$,

$$\lim_{n \rightarrow \infty} \mathbb{P}(|X_n - X| > \epsilon) = 0;$$

- *hội tụ theo trung bình bậc p ($p > 0$)* đến bnn X , kí hiệu là $X_n \xrightarrow{L^p} X$ nếu $\mathbb{E}(|X_n|^p) < \infty$ với mọi n , $\mathbb{E}(|X|^p) < \infty$ và

$$\lim_{n \rightarrow \infty} \mathbb{E}(|X_n - X|^p) = 0.$$

- *hội tụ theo phân phối* đến bnn X , kí hiệu là $X_n \xrightarrow{d} X$ nếu

$$\lim_{n \rightarrow \infty} \mathbb{P}[X_n < x] = \mathbb{P}[X < x]$$

với mọi $x \in \mathbb{R}$ thoả mãn $\mathbb{P}[X = x] = 0$.

Mệnh đề 1.6.2. Giả sử $(X_n)_{n \geq 1}$ là dãy bnn.

1. Nếu $X_n \xrightarrow{L^p} X$ thì $X_n \xrightarrow{\mathbb{P}} X$ với mọi $p > 0$.
2. Nếu $X_n \xrightarrow{h.c.c} X$ thì $X_n \xrightarrow{\mathbb{P}} X$.

1.7 Luật số lớn

Trong tiết này ta sẽ giả sử $(X_n)_{n \geq 1}$ là dãy bnn xác định trên không gian xác suất $(\Omega, \mathcal{F}, \mathbb{P})$ và kí hiệu $S_n = X_1 + \dots + X_n$.

1.7.1 Luật yếu số lớn

Định lý 1.7.1. Giả sử $(X_n)_{n \geq 1}$ là dãy bnn thoả mãn

$$\lim_{n \rightarrow \infty} \frac{D(S_n)}{n^2} = 0.$$

Khi đó

$$\frac{S_n - \mathbb{E}S_n}{n} \xrightarrow{\mathbb{P}} 0, \text{ khi } n \rightarrow \infty.$$

Chứng minh. Với mọi $\epsilon > 0$, áp dụng bất đẳng thức Chebyshev, ta có

$$\mathbb{P}\left(\left|\frac{S_n - \mathbb{E}S_n}{n}\right| \geq \epsilon\right) \leq \frac{D(S_n)}{n^2\epsilon^2} \rightarrow 0 \text{ khi } n \rightarrow \infty.$$

Từ đây suy ra điều phải chứng minh. □

Sử dụng Mệnh đề 1.5.3, ta có hệ quả sau.

Hệ quả 1.7.2. Giả sử $(X_n)_{n \geq 1}$ là dãy các bnn độc lập và thỏa mãn

$$\lim_{n \rightarrow \infty} \frac{D(X_1) + \dots + D(X_n)}{n^2} = 0.$$

Khi đó

$$\frac{S_n - \mathbb{E}S_n}{n} \xrightarrow{\mathbb{P}} 0, \text{ khi } n \rightarrow \infty.$$

1.7.2 Luật mạnh số lớn

Định lý 1.7.3. Giả sử $(X_n)_{n \geq 1}$ là dãy các bnn độc lập cùng phân phối. Khi đó

$$\lim_{n \rightarrow \infty} \frac{S_n - \mathbb{E}S_n}{n} = 0 \quad \text{h.c.c khi và chỉ khi } \mathbb{E}[|X_1|] < +\infty.$$

1.8 Hàm đặc trưng

Định nghĩa 1.8.1. 1. Hàm đặc trưng của bnn X được xác định là ánh xạ $\varphi_X : \mathbb{R} \rightarrow \mathbb{C}$ cho bởi

$$\varphi_X(t) = \mathbb{E}[e^{itX}] = \int_{\mathbb{R}} e^{itx} dF_X(x).$$

2. Hàm đặc trưng của véc tơ ngẫu nhiên $X = (X_1, \dots, X_n)$ là ánh xạ $\varphi_X : \mathbb{R}^d \rightarrow \mathbb{C}$ cho bởi

$$\varphi_X(t_1, \dots, t_n) = \mathbb{E}\left[\exp\left(i \sum_{j=1}^n t_j X_j\right)\right].$$

Ví dụ 1.8.2. Giả sử X có phân phối Poisson với tham số $\lambda > 0$. Ta có

$$\varphi_t(X) = \sum_{k=0}^{\infty} \frac{e^{-\lambda} \lambda^k}{k!} e^{itk} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda e^{it})^k}{k!} = e^{\lambda(e^{it}-1)}.$$

Ví dụ 1.8.3. Giả sử X có phân phối chuẩn $N(0, 1)$. Khi đó

$$\varphi_X(t) = \int_{-\infty}^{\infty} e^{itx} \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx = \int_{-\infty}^{\infty} \frac{\cos tx}{\sqrt{2\pi}} e^{-x^2/2} dx + \int_{-\infty}^{\infty} \frac{\sin tx}{\sqrt{2\pi}} e^{-x^2/2} dx.$$

Do hàm $x \mapsto e^{-x^2/2} \sin tx$ là hàm khả tích và lẻ nên tích phân thứ hai bằng 0. Theo Định lý ?? ta có thể đạo hàm hai vế của đẳng thức trên theo t và thu được

$$\varphi'_X(t) = - \int_{-\infty}^{\infty} \frac{\sin tx}{\sqrt{2\pi}} x e^{-x^2/2} dx.$$

Áp dụng công thức tích phân từng phần, ta được

$$\varphi'_X(t) = - \int_{-\infty}^{\infty} \frac{\cos tx}{\sqrt{2\pi}} t e^{-x^2/2} dx = -t\varphi_X(t).$$

Giải phương trình vi phân $\frac{\varphi'_X}{\varphi_X} = -t$ với điều kiện $\varphi_X(0) = 1$, ta được

$$\varphi_X(t) = e^{-t^2/2}.$$

Nếu X có phân phối chuẩn $N(a, \sigma^2)$ thì

$$\varphi_X(t) = \frac{1}{\sqrt{2\pi\sigma^2}} \int e^{itx} e^{-\frac{(x-a)^2}{2\sigma^2}} dx.$$

Đổi biến $y = (x - a)/\sigma$, ta được

$$\varphi_X(t) = \frac{e^{ita}}{\sqrt{2\pi}} \int e^{it\sigma y} e^{-y^2/2} dy = e^{ita-t^2\sigma^2/2}.$$

Định lý 1.8.4. Hai véc tơ ngẫu nhiên có cùng phân phối khi và chỉ khi hàm đặc trưng của chúng bằng nhau.

Ví dụ 1.8.5. Giả sử X và Y là hai bnn độc lập có phân phối Poisson với tham số lần lượt là μ và λ . Khi đó theo Ví dụ 1.8.2, $X + Y$ có hàm đặc trưng là

$$\varphi_{X+Y}(t) = \mathbb{E}(e^{it(X+Y)}) = \mathbb{E}(e^{itX})\mathbb{E}(e^{itY}) = e^{(\lambda+\mu)(e^{it}-1)}.$$

Do đó $X + Y$ có phân phối Poisson với tham số $\mu + \lambda$.

Định lý 1.8.6. Các bnn $X_1, \dots, X_n$ là độc lập với nhau khi và chỉ khi

$$\varphi_{(X_1, \dots, X_n)}(t_1, \dots, t_n) = \varphi_{X_1}(t_1) \dots \varphi_{X_n}(t_n) \text{ với mọi } (t_1, \dots, t_n) \in \mathbb{R}^n.$$

Ví dụ 1.8.7. Giả sử X và Y là hai bnn độc lập có cùng phân phối chuẩn $N(0, 1)$. Theo Ví dụ 1.8.3

$$\varphi_{(X+Y, X-Y)}(t, s) = \mathbb{E}e^{it(X+Y)+is(X-Y)} = \mathbb{E}e^{i(t+s)X} \mathbb{E}e^{i(t-s)Y} = e^{-t^2-s^2}.$$

Lần lượt cho s và t bằng 0 ở đẳng thức trên, ta được $\varphi_{X+Y}(t) = e^{-t^2}$ và $\varphi_{X-Y}(s) = e^{-s^2}$. Do đó $X + Y$ và $X - Y$ đều có phân phối chuẩn $N(0, 2)$. Hơn nữa, chúng là độc lập vì

$$\varphi_{(X+Y, X-Y)}(t, s) = \varphi_{X+Y}(t)\varphi_{X-Y}(s) \text{ với mọi } t, s \in \mathbb{R}.$$

Định lí sau cho ta một tiêu chuẩn rất hữu hiệu để kiểm tra sự hội tụ theo phân phối của một dãy bnn. Chứng minh định lí có thể được tham khảo trong [13] trang 196-199.

Định lý 1.8.8. Dãy biến ngẫu nhiên X_n hội tụ theo phân phối đến X khi và chỉ khi dãy hàm đặc trưng của X_n hội tụ điểm đến hàm đặc trưng của X .

Kết quả sau thường được dùng trong thống kê

Định lý 1.8.9 (Định lý Slutsky). *Giả sử dãy biến ngẫu nhiên X_n hội tụ theo phân phối đến X và dãy biến ngẫu nhiên A_n hội tụ theo xác suất đến hằng số A . Khi đó*

- $A_n X_n \xrightarrow{d} AX$.
- $A_n + X_n \xrightarrow{d} A + X$.
- $X_n/A_n \xrightarrow{d} X/A$ nếu $A \neq 0$.

Ví dụ 1.8.10. Giả sử với mỗi $n \geq 1$, bnn X_n có phân phối chuẩn $N(a_n, \sigma_n^2)$ trong đó $a_n \rightarrow 0$ và $\sigma_n^2 \rightarrow 1$ khi $n \rightarrow \infty$. Khi đó dãy bnn (X_n) hội tụ yếu đến phân phối chuẩn tắc $N(0, 1)$ vì

$$\varphi_{X_n}(t) = e^{ita_n - \sigma_n^2 t^2/2} \rightarrow e^{-t^2/2}.$$

1.9 Định lí giới hạn trung tâm

Định lý 1.9.1. *Giả sử $(X_n)_{n \geq 1}$ là dãy bnn độc lập và cùng phân phối với $\mathbb{E}(X_n) = \mu$ và $DX_n = \sigma^2 \in (0, \infty)$. Đặt $S_n = X_1 + \dots + X_n$. Khi đó dãy bnn $Y_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}$ hội tụ theo phân phối tới bnn Y có phân phối chuẩn $N(0, 1)$.*

Ví dụ 1.9.2. Giả sử $(X_j)_{j \geq 1}$ là dãy bnn độc lập cùng phân phối với $\mathbb{P}(X_j = 1) = p$ và $\mathbb{P}(X_j = 0) = 1 - p$ với $p \in (0, 1)$. Ta có $\mu = \mathbb{E}(X_j) = p$ và $\sigma^2 = DX_j = p(1 - p)$. Theo Luật mạnh số lớn, ta có

$$\frac{S_n}{n} \xrightarrow{h.c.c} p,$$

và theo Định lí Giới hạn trung tâm 1.9.1, ta có

$$\frac{S_n - np}{\sqrt{np(1-p)}} \xrightarrow{w} Z \sim N(0, 1).$$

1.10 Phân phối chuẩn nhiều chiều

Định nghĩa 1.10.1. Một biến ngẫu nhiên $\mathbf{X} = (X_1, \dots, X_n)$ nhận giá trị trong $\mathbb{R}^n$ là Gauss (hoặc chuẩn nhiều chiều) nếu mọi tổ hợp tuyến tính $\sum_{j=1}^n a_j X_j$ hoặc là một hằng số hoặc là một biến ngẫu nhiên có phân phối chuẩn (1 chiều).

Hàm đặc trưng là một công cụ hữu ích để nghiên cứu các biến ngẫu nhiên Gauss.

Định lý 1.10.2. $\mathbf{X}$ là một biến ngẫu nhiên Gauss nhận giá trị trong $\mathbb{R}^d$ khi và chỉ khi hàm đặc trưng của nó có dạng

$$\varphi_{\mathbf{X}}(\mathbf{u}) = \exp\left(i\langle \mathbf{u}, \mu \rangle - \frac{1}{2}\langle \mathbf{u}, Q\mathbf{u} \rangle\right),$$

trong đó $\mu \in \mathbb{R}^n$ và Q là một ma trận $n \times n$ nửa xác định không âm đối xứng. Khi đó, Q là ma trận covarian của $\mathbf{X}$ và μ là kì vọng của $\mathbf{X}$, nghĩa là

$$\mu_j = \mathbb{E}[X_j], \quad Q_{k,j} = \text{cov}(X_k, X_j), \quad \text{với mọi } j \text{ và } k.$$

Ví dụ 1.10.3. Cho $X_1, \dots, X_n$ là các biến ngẫu nhiên độc lập nhận giá trị trong $\mathbb{R}$ có phân phối $\mathcal{N}(\mu_j, \sigma_j^2)$, khi đó $\mathbf{X} = (X_1, \dots, X_n)$ là Gauss. Hơn nữa, với bất kì ma trận hằng số $A \in \mathbb{R}^{m \times n}$ thì $\mathbf{Y} = Q\mathbf{X}^*$ là một biến ngẫu nhiên Gauss m chiều.

Mệnh đề 1.10.4. Cho $\mathbf{X}$ là một biến ngẫu nhiên Gauss nhận giá trị trong $\mathbb{R}^n$. Khi đó, các biến ngẫu nhiên X_j độc lập khi và chỉ khi ma trận covarian Q của $\mathbf{X}$ là ma trận đường chéo.

Mệnh đề 1.10.5. Cho $\mathbf{X}$ là một biến ngẫu nhiên Gauss nhận giá trị trong $\mathbb{R}^n$. $\mathbf{X}$ có một mật độ trên $\mathbb{R}^n$ khi và chỉ khi ma trận covarian Q không suy biến (nghĩa là, $\det Q \neq 0$), và hàm mật độ xác suất của $\mathbf{X}$ được cho bởi

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{2\pi^{n/2} \sqrt{\det Q}} e^{-\frac{1}{2}\langle \mathbf{x} - \mu, Q^{-1}(\mathbf{x} - \mu) \rangle}.$$

1.11 Phân phối Gamma, khi bình phương, student and F

1.11.1 Phân phối Gamma

Định nghĩa 1.11.1. Một biến ngẫu nhiên X được gọi là có phân phối Gamma $\mathcal{G}(\alpha, \lambda)$ nếu hàm mật độ xác suất của nó được xác định bởi

$$f_X(x) = \frac{x^{\alpha-1} e^{-x/\lambda}}{\Gamma(\alpha) \lambda^\alpha} I_{\{x>0\}},$$

trong đó, $\Gamma(\alpha) = \int_0^{+\infty} x^{\alpha-1} e^{-x} dx$ được gọi là hàm số Gamma.

Chú ý rằng $\mathcal{G}(1, \lambda) = \text{Exp}(\lambda)$.

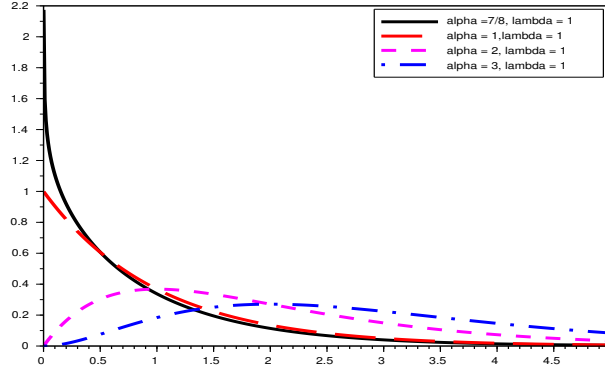
Mệnh đề 1.11.2. Nếu X có phân phối $\mathcal{G}(\alpha, \lambda)$ thì

$$\mathbb{E}[X] = \alpha \lambda, \quad DX = \alpha \lambda^2.$$

Hơn nữa, hàm đặc trưng của X được xác định bởi

$$\varphi_X(t) = \int_0^\infty e^{itx} \frac{x^{\alpha-1} e^{-x/\lambda}}{\Gamma(\alpha) \lambda^\alpha} dx = \left(\frac{1}{1 - i\lambda t} \right)^\alpha.$$

Hệ quả 1.11.3. Cho $(X_i)_{1 \leq i \leq n}$ là dãy các biến ngẫu nhiên độc lập. Giả sử với mỗi i , X_i có phân phối $\mathcal{G}(\alpha_i, \lambda)$. Khi đó, $S = X_1 + \dots + X_n$ có phân phối $\mathcal{G}(\alpha_1 + \dots + \alpha_n, \lambda)$.



Hình 1.1: Mật độ của phân phối Gamma

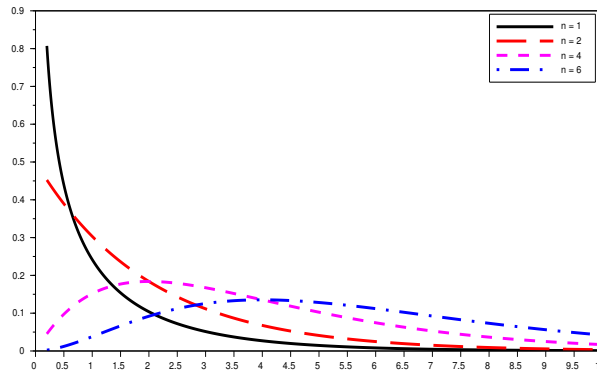
1.11.2 Phân phối khi bình phương

Định nghĩa 1.11.4. Cho $(Z_i)_{1 \leq i \leq n}$ là một dãy các biến ngẫu nhiên độc lập cùng phân phối chuẩn tắc. Khi đó, phân phối của $V = Z_1^2 + \dots + Z_n^2$ được gọi là *phân phối khi bình phương* với *bậc tự do* n và được kí hiệu là χ_n^2 .

Chú ý: vì Z_i^2 có phân phối $\mathcal{G}(\frac{1}{2}, 2)$ nên χ_n^2 có phân phối $\mathcal{G}(\frac{n}{2}, 2)$. Hơn nữa,

$$\mathbb{E}[\chi_n^2] = n, \quad D\chi_n^2 = 2n.$$

Một hệ quả cần chú ý từ định nghĩa của phân phối khi bình phương là nếu hai biến ngẫu nhiên U và V độc lập với $U \sim \chi_n^2$ và $V \sim \chi_m^2$ thì $U + V \sim \chi_{m+n}^2$.

Hình 1.2: Mật độ của phân phối χ^2

1.11.3 Phân phối Student

Định nghĩa 1.11.5. Nếu Z và U là hai biến ngẫu nhiên độc lập với $Z \sim N(0;1)$ và $U \sim \chi_n^2$ thì phân phối của $\frac{Z}{\sqrt{U/n}}$ được gọi là *phân phối Student với bậc tự do n* .

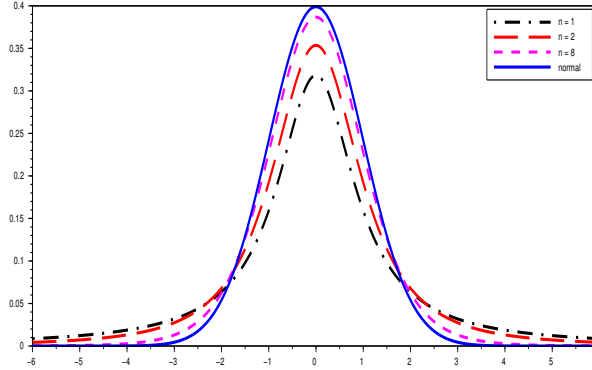
Phân phối Student còn được gọi là phân phối t .

Mệnh đề 1.11.6. Hàm mật độ xác suất của phân phối Student với bậc tự do n là

$$f_n(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{t^2}{n}\right)^{-(n+1)/2}.$$

Ngoài ra,

$$f_n(t) \xrightarrow{n \rightarrow \infty} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}.$$



Hình 1.3: Mật độ của phân phối Student

1.11.4 Phân phối F

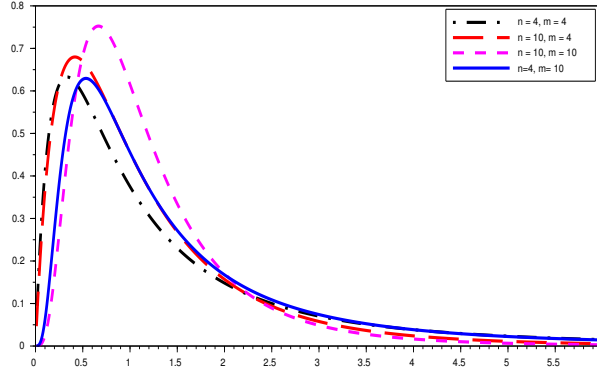
Định nghĩa 1.11.7. Cho U và V là hai biến ngẫu nhiên độc lập có cùng phân phối khi bình phương với bậc tự do lần lượt là m và n . Khi đó, phân phối của biến ngẫu nhiên

$$W = \frac{U/m}{V/n}$$

được gọi là *phân phối F với bậc tự do m và n* và được kí hiệu là $F_{m,n}$.

Hàm mật độ xác suất của W là

$$f(x) = \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \left(\frac{m}{n}\right)^{m/2} x^{m/2-1} \left(1 + \frac{m}{n}x\right)^{-(m+n)/2}, \quad x \geq 0.$$

Hình 1.4: Mật độ của phân phối F

1.12 Phân phối của trung bình và phương sai mẫu

Cho (X_n) là dãy các biến ngẫu nhiên độc lập có cùng phân phối chuẩn $\mathcal{N}(\mu, \sigma^2)$, và đặt

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i, \quad s_n^2 = \frac{1}{n-1} \sum_{i=1}^{n-1} (X_i - \bar{X}_n)^2$$

Mệnh đề 1.12.1. *Biến ngẫu nhiên $\bar{X}_n$ và vectơ của các biến ngẫu nhiên $(X_1 - \bar{X}_n, X_2 - \bar{X}_n, \dots, X_n - \bar{X}_n)$ độc lập với nhau.*

Chứng minh. Ta có:

$$s\bar{X}_n + \sum_{i=1}^n t_i(X_i - \bar{X}_n) = \sum_{i=1}^n a_i X_i$$

trong đó $a_i = \frac{s}{n} + (t_i - \bar{t})$. Chú ý rằng:

$$\sum_{i=1}^n a_i = s \quad \text{and} \quad \sum_{i=1}^n a_i^2 = \frac{s^2}{n} + \sum_{i=1}^n (t_i - \bar{t})^2.$$

Do đó, hàm đặc trưng của vectơ ngẫu nhiên $(\bar{X}_n, X_1 - \bar{X}_n, X_2 - \bar{X}_n, \dots, X_n - \bar{X}_n)$ là

$$\begin{aligned} &= \mathbb{E}[\exp(is\bar{X}_n + i \sum_{j=1}^n t_j(X_j - \bar{X}_n))] = \prod_{j=1}^n \exp\left(i\mu a_j - \frac{\sigma^2}{2} a_j^2\right) \\ &= \exp\left(i\mu s - \frac{\sigma^2}{2n} s^2\right) \exp\left(-\frac{\sigma^2}{2} \sum_{i=1}^n (t_i - \bar{t})^2\right). \end{aligned}$$

Khi $s = 0$, ta thấy số hạng đầu là hàm đặc trưng của $\bar{X}_n$ còn số hạng thứ hai là $(X_1 - \bar{X}_n, X_2 - \bar{X}_n, \dots, X_n - \bar{X}_n)$. Do đó, ta có điều phải chứng minh. $\square$

Hệ quả 1.12.2. $\bar{X}_n$ và s_n^2 độc lập với nhau.