MÔN HỌC: XÁC SUẤT THỐNG KÊ

GV: Vũ Thị Hương

26th March 2020

2.1 Tập chính và mẫu

Khái niệm tập chính

Tập chính

Tập chính bao gồm tất cả các đối tượng mà ta cần nghiên cứu.

Ký hiệu tập chính: Ω
Đặc điểm tập chính:

- Kích thước lớn (hữu hạn hoặc vô hạn) khó khảo sát được toàn bộ.
- Các phần tử của tập chính thường có sự biến động theo thời gian.
- Việc khảo sát toàn bộ tập chính cần huy động nhiều người, tốn kém về thời gian và tiền bạc.

Các ví dụ

- Tập hợp các kỹ sư Việt Nam
- Tập hợp các điện thoại di động của hãng Samsung sản xuất.

Mẫu thực nghiệm

Là một tập con của tập chính

Mô tả một mẫu thực nghiệm:

$$x_1, x_2, \ldots, x_n$$

Kích thước mẫu: n

Các phương pháp lấy mẫu

- Lấy mẫu không hoàn lại.
- Lấy mẫu có hoàn lại.
- Lấy mẫu theo phân lớp.

2.2 Phân loại dữ liệu

Hai loại dữ liệu

- Dữ liệu định lượng.
  - Nhiệt độ cao nhất trong ngày tại Hà Nội.
  - Chiều cao của các sinh viên đại học.
  - Cường độ chịu nén của các mẫu bê tông.
- Dữ liệu định tính.
  - Dữ liệu danh nghĩa.
    - Quốc tịch hành khách.
    - Thành phần vật chất cấu thành bê tông.
  - Dữ liệu thứ bậc.
    - Đánh giá chất lượng (rất tốt, tốt, trung bình, không đạt).
    - Mức độ tín nhiệm (rất tín nhiệm, tín nhiệm, không tín nhiệm).

2.3 Phương pháp biểu diễn mẫu

Trường hợp kích thước mẫu nhỏ các giá trị quan sát có lặp lại

Ví dụ 2.4

Kiểm tra về tuổi của 30 học viên cao học ta được một mẫu và được sắp xếp lần lượt như sau:

28	31	29	27	30	29	29	26	30	28
28	29	27	26	32	28	32	31	25	30
27	30	29	30	28	29	31	27	28	28

Sắp xếp các giá trị từ nhỏ đến lớn ta được dãy số liệu sau:

25	26	26	27	27	27	27	28	28	28
28	28	28	28	29	29	29	29	29	29
30	30	30	30	30	31	31	31	32	32

Bảng phân phối tần số

x_i	25	26	27	28	29	30	31	32
n_i	1	2	4	7	6	5	3	2

Tổng quát: Mẫu quan sát có kích thước n khá nhỏ, các giá trị thuộc mẫu có lặp lại thì ta có mô tả mẫu theo bảng phân phối tần số

x_i	x_1	x_2	$\dots$	x_k
n_i	n_1	n_2	$\dots$	n_k

Trong bảng phân phối tần số các giá trị quan sát cũng có thể là giá trị định tính.

Ví dụ 2.5

Bảng dưới đây thống kê số lượng lao động được tuyển dụng (phân theo trình độ nghề) tại một hội chợ việc làm diễn ra tháng 4/2016 ở Hà Nội.

Trình độ lao động	Số lượng tuyển dụng
Trên đại học	2
Đại học	15
Cao đẳng	14
Trung cấp	21
Công nhân kỹ thuật lành nghề	7
Sơ cấp nghề	5
Lao động chưa qua đào tạo	38

2.3 Phương pháp biểu diễn mẫu

Trường hợp kích thước mẫu lớn

Trong trường hợp dữ liệu chứa quá nhiều giá trị phân biệt và số lần xuất hiện mỗi giá trị ít, người ta thường chia dữ liệu thành các lớp hay các khoảng rời nhau. Chú ý rằng, trong việc chia miền dữ liệu thành các khoảng, ta áp dụng nguyên tắc *cận dưới (trái) đúng*. Cụ thể, đối với mỗi khoảng $[a, b)$, ký hiệu là $a - b$, ta gọi a là cận dưới, b là cận trên, khoảng cách $c = b - a$ gọi là độ rộng của khoảng, điểm giữa $\dfrac{a+b}{2}$ gọi là giá trị trung bình của khoảng.

Trong mỗi khoảng ta đếm số lần các giá trị quan sát nằm trong khoảng $[a, b)$, bao gồm các giá trị quan sát x_i mà $a \leq x_i < b$.

2.3 Phương pháp biểu diễn mẫu

Trường hợp kích thước mẫu lớn

Miền giá trị quan sát được chia thành k khoảng rời nhau. Nhìn chung, nhưng không bắt buộc, các khoảng được chia với độ dài bằng nhau, ký hiệu là h. Khi đó mỗi khoảng có dạng

$$\Big[r + (i-1)h,\ r + ih\Big), \quad i = 1, 2, \ldots, m,$$

với r là một giá trị cụ thể, nhỏ hơn hoặc bằng giá trị nhỏ nhất của mẫu quan sát.

- ➢ Việc chia khoảng giúp cho việc biểu diễn dữ liệu đơn giản hơn, nhưng lại làm mất thông tin của dữ liệu gốc.
- ➢ Số khoảng chia m có thể chọn theo quy tắc Sturges, bằng số nguyên dương nhỏ nhất và lớn hơn $1 + \log_2(n)$.

Ví dụ 2.6

Thời gian làm việc (tính theo giờ) của công nhân trong một tuần tại một công ty được ghi lại như sau:

45	31	46	25	57	39	42	55	20	37
40	59	11	38	34	22	62	33	48	43
57	37	43	51	29	41	35	66	45	32
44	47	42	46	54	65	17	35	53	27
38	22	33	39	45	32	43	41	57	45

Quan sát thấy dữ liệu trên có giá trị nhỏ nhất là 11 và giá trị lớn nhất là 66. Ta chia miền dữ liệu thành 6 khoảng, mỗi khoảng có độ dài là 10 giờ, bắt đầu từ giá trị bằng 10. Khi đó ta có bảng phân phối tần số như sau:

x_i	10-20	20-30	30-40	40-50	50-60	60-70
n_i	2	6	16	15	8	3

2.3 Phương pháp biểu diễn mẫu

Khái niệm tần suất

Tần suất chính là tỉ lệ phần trăm một giá trị nào đó có trong dữ liệu. Tần suất của giá trị x_i được ký hiệu bởi

$$f_i = \frac{n_i}{n}.$$

Thay tần số bởi tần suất trong bảng phân phối tần số, ta thu được bảng phân phối tần suất.

x_i	x_1	x_2	$\ldots$	x_k
f_i	f_1	f_2	$\ldots$	f_k

Chú ý rằng

$$f_1 + f_2 + \ldots + f_k = 1.$$

Đa giác tần số, tần suất

Trên hệ trục tọa độ Đề-các, ta xác định các điểm (x_i, n_i) (hoặc (x_i, f_i)) sau đó vẽ đường gấp khúc nối các điểm được vẽ từ điểm đầu tới điểm cuối. Đường gấp khúc này được gọi là đa giác tần số (hoặc đa giác tần suất). Từ đồ thị này ta dễ dàng nhận ra được giá trị nào có tần số (tần suất) cao nhất hoặc thấp nhất.

Biểu đồ hình tròn

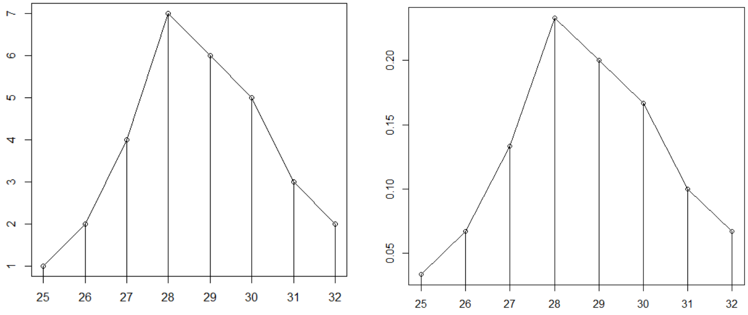
Biểu đồ này thường được sử dụng để biểu diễn tỉ lệ phần trăm khi các giá trị trong dữ liệu không phải ở dạng định lượng. Khi đó ta vẽ một hình tròn và chia thành các hình quạt theo tỉ lệ tương ứng với các giá trị định tính.

2.3 Phương pháp biểu diễn mẫu

Đa giác tần số, tần suất

Đa giác tần suất tần số cho Ví dụ 2.4

x_i	25	26	27	28	29	30	31	32
n_i	1	2	4	7	6	5	3	2
f_i	$1/30$	$2/30$	$4/30$	$7/30$	$6/30$	$5/30$	$3/30$	$2/30$

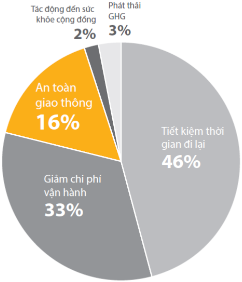


2.3 Phương pháp biểu diễn mẫu

Biểu đồ hình tròn

Ví dụ 2.10

Xe buýt nhanh và các hệ thống ưu tiên xe buýt đã trở thành một giải pháp hấp dẫn đối với nhu cầu lưu thông trong đô thị vì chi phí vốn tương đối thấp và thời gian thi công ngắn so với hệ thống giao thông đường sắt. Biểu đồ dưới đây cho thấy những nhân tố đóng góp vào lợi ích kinh tế của hệ thống xe buýt nhanh đang hoạt động ở các nước thuộc Mỹ Latinh.



Biểu đồ tần số, tần suất

Đối với dữ liệu được chia khoảng, ta biểu diễn chúng bởi biểu đồ hình cột, gồm các cột chữ nhật kề nhau có đáy trên trục hoành là các khoảng chia và chiều cao trên trục tung là tần số hoặc tần suất. Các biểu đồ này tương ứng gọi là biểu đồ tần số hoặc biểu đồ tần suất.

Ví dụ 2.11

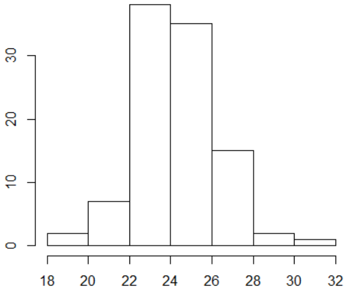
Chỉ số khối cơ thể (BMI - Body Mass Index), bằng cân nặng (tính theo kilogram) chia cho bình phương chiều cao (tính theo mét), là chỉ số dùng để đo độ béo phì của một người. Người bình thường có BMI $< 25\,(kg/m^2)$, $25 \leqslant$ BMI $< 30\,(kg/m^2)$ là quá cân và BMI $\geqslant 30\,(kg/m^2)$ là béo phì. Kiểm tra chỉ số BMI của 100 sinh viên, ta thu được bảng số liệu dưới đây.

2.3 Phương pháp biểu diễn mẫu

Biểu đồ tần số, tần suất

Hình bên dưới là biểu đồ tần số đối với bảng dữ liệu trên.

x_i	18-20	20-22	22-24	24-26	26-28	28-30	30-32
n_i	2	7	38	35	15	2	1



Nếu mẫu quan sát có bảng phân phối tần suất

x_i	x_1	x_2	$\ldots$	x_k
f_i	f_1	f_2	$\ldots$	f_k

Tương ứng, hàm phân phối thực nghiệm $F_n(x)$ được xác định bởi công thức

$$F_n(x) = \sum_{x_i \leqslant x} f_i = \begin{cases} 0 & x < x_1, \\ f_1 & x_1 \leqslant x < x_2, \\ f_1 + f_2 & x_2 \leqslant x < x_3, \\ \vdots & \vdots \\ f_1 + \cdots + f_{k-1} & x_{k-1} \leqslant x < x_k, \\ 1 & x \geqslant x_k. \end{cases}$$

Từ công thức trên, hàm phân phối thực nghiệm đôi khi còn được gọi là hàm tần suất cộng dồn.

2.4 Giá trị đặc trưng mẫu

Trung bình mẫu

Định nghĩa 2.3

Nếu n giá trị quan sát trong một mẫu được ký hiệu là $x_1, x_2, \ldots, x_n$ trung bình mẫu, kí hiệu $\bar{x}$, được xác định bởi

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Nếu mẫu cho bởi bảng phân phối tần số

x_i	x_1	x_2	$\ldots$	x_k
n_i	n_1	n_2	$\ldots$	n_k

thì giá trị trung bình được tính theo công thức

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} n_i x_i$$

Ví dụ 2.13

Một kỹ sư chuyên ngành vật liệu xây dựng nghiên cứu ảnh hưởng của một chất phụ gia làm từ tro trấu đối với sức chịu nén của một loại bê tông. Sau khi chế tạo thử 10 mẫu bê tông, người kỹ sư đo cường độ chịu nén của chúng (theo đơn vị $\mathrm{kgf/cm^2}$) và thu được kết quả sau:

$$95,\ 86,\ 102,\ 91,\ 97,\ 93,\ 101,\ 105,\ 88,\ 92$$

Hãy tính cường độ chịu nén trung bình của 10 mẫu bê tông trên.

Giải: Cường độ chịu nén trung bình của 10 mẫu bê tông trên là

$$\bar{x} = \frac{95+86+102+91+97+93+101+105+88+92}{10}$$
$$= \frac{950}{10} = 95\ (\mathrm{kgf/cm^2}).$$

2.4 Giá trị đặc trưng mẫu

Trung bình mẫu

Ví dụ 2.14

Tìm giá trị trung bình của các mẫu dữ liệu sau:

x_i	30	30,2	30,4	30,5	30,8	31,0
n_i	5	7	12	17	11	8

Giải: Ta có giá trị trung bình mẫu là

$$\bar{x} = \frac{30 \times 5 + 30,2 \times 7 + 30,4 \times 12 + 30,6 \times 17 + 30,8 \times 11 + 31 \times 8}{5 + 7 + 12 + 17 + 11 + 8}$$

$$= \frac{1833,2}{60} = 30,55.$$

2.4 Giá trị đặc trưng mẫu

Trung bình mẫu

Nếu dữ liệu cho ở dạng khoảng thì ta lấy trung điểm của các khoảng làm giá trị đại diện cho khoảng và tính giá trị trung bình như các trường hợp trên.

Ví dụ 2.15

Tìm giá trị trung bình của các mẫu dữ liệu sau:

x_i	12-14	14-16	16-18	18-20	20-22	22-24
n_i	2	6	9	8	4	1

Trước hết ta phải tính giá trị trung bình cho từng khoảng

$$x_1 = \frac{12+14}{2} = 13,\ x_2 = 15,\ x_3 = 17,\ x_4 = 19,\ x_5 = 21,\ x_6 = 23.$$

Ta có giá trị trung bình mẫu là

$$\begin{aligned} \bar{x} &= \frac{13\times 2 + 15\times 6 + 17\times 9 + 19\times 8 + 21\times 4 + 23\times 1}{2+6+9+8+4+1} \\ &= \frac{528}{30} = 17,6. \end{aligned}$$

Định nghĩa 2.6

Nếu $x_1, x_2, \ldots, x_n$ là mẫu có n quan sát, phương sai mẫu, kí hiệu là s^2, được xác định bởi

$$s^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{n-1} \tag{1}$$

và độ lệch tiêu chuẩn mẫu là căn bậc hai của phương sai mẫu

$$s = \sqrt{s^2} = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{n-1}}. \tag{2}$$

Ta có công thức

$$s^2 = \frac{n}{n-1}\left(\overline{x^2} - \bar{x}^2\right), \quad \text{với } \overline{x^2} = \frac{1}{n}\sum_{i=1}^{n} x_i^2$$

Ví dụ 2.25

Xét dữ liệu về độ dài (mm) chi tiết máy được gia công trên máy A và B

Gia công trên máy A:	28	30	31	29	32	30
Gia công trên máy B:	35	24	25	36	39	21

Trung bình mẫu và phương sai mẫu tính được như sau:

$$\overline{x_A} = \frac{28 + 30 + 31 + 29 + 32 + 30}{6} = 30,$$

$$\overline{x_B} = \frac{35 + 24 + 25 + 36 + 39 + 21}{6} = 30,$$

$$s_A^2 = \frac{1}{5}\Big[(28-30)^2 + (30-30)^2 + (31-30)^2 + (29-30)^2 + (32-30)^2 + (30-30)^2\Big] = 2$$

$$s_B^2 = \frac{1}{5}\Big[(35-30)^2 + (24-30)^2 + (25-30)^2 + (36-30)^2 + (39-30)^2 + (21-30)^2\Big] = 56,8.$$

Rõ ràng s_A^2 nhỏ hơn nhiều so với s_B^2 và do đó máy A chạy ổn định hơn máy B.

2.4 Giá trị đặc trưng mẫu

Cách tính trung bình và phương sai

Xét một mẫu cho bởi bảng phân phối tần số

x_i	x_1	x_2	$\dots$	x_k
n_i	n_1	n_2	$\dots$	n_k

Khi đó giá trị trung bình và giá trị phương sai mẫu được tính theo các công thức

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} n_i x_i,$$

$$\bar{x^2} = \frac{1}{n}\sum_{i=1}^{k} n_i x_i^2,$$

$$s^2 = \frac{n}{n-1}(\bar{x^2} - \bar{x}^2),$$

$$s = \sqrt{s^2}.$$

Tính toán theo bảng tính:

x_i	n_i	$n_i x_i$	$n_i x_i^2$
x_1	n_1	$n_1 x_1$	$n_1 x_1^2$
x_2	n_2	$n_2 x_2$	$n_2 x_2^2$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_k	n_k	$n_k x_k$	$n_k x_k^2$
$\sum$	n	$\sum n_i x_i$	$\sum n_i x_i^2$

Ví dụ 2.27

Tìm giá trị trung bình và phương sai của mẫu dữ liệu sau:

x_i	30	30,2	30,4	30,6	30,8	31,0
n_i	5	7	12	17	11	8

Ta lập bảng tính các tổng của dữ liệu

x_i	n_i	$n_i x_i$	$n_i x_i^2$
30	5	150	4500
30,2	7	211,4	6384,28
30,4	12	364,8	11089,92
30,6	17	520,2	15918,12
30,8	11	338,8	10435,04
31	8	248	7688
$\sum$	60	1833,2	56015,36

Giá trị trung bình và phương sai mẫu là

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} n_i x_i = \frac{1833,2}{60} = 30,5533;$$

$$\bar{x^2} = \frac{1}{n}\sum_{i=1}^{k} n_i x_i^2 = \frac{56015,36}{60} = 933,5893;$$

$$s^2 = \frac{n}{n-1}(\bar{x^2} - \bar{x}^2) = 0,0846.$$

2.5 Phân phối của các thống kê mẫu

Mẫu ngẫu nhiên và thống kê mẫu

- Lý do của việc lấy mẫu là do ta không thể kiểm tra tất cả các phần tử của tập chính.
- Để nghiên cứu các phần tử trên tập mẫu và tìm cách rút ra kết luận về toàn bộ tập chính, ta phải mô hình hóa mối liên hệ giữa tập chính và tập mẫu.
- Mỗi mẫu quan sát $(x_1, x_2, \ldots, x_n)$ sẽ sẽ thay đổi trong các lần chọn mẫu khác nhau. Nói riêng, ta xem x_i như là một thể hiện của một biến ngẫu nhiên X_i nào đó.
- Tổng quát hóa, tập mẫu $(x_1, x_2, \ldots, x_n)$ là một thể hiện cụ thể của biến ngẫu nhiên nhiều chiều $(X_1, X_2, \ldots, X_n)$.

Mẫu ngẫu nhiên

Cho $X_1, X_2, \ldots, X_n$ là các biến ngẫu nhiên độc lập và có cùng hàm phân phối $F(x)$. Ta gọi $(X_1, X_2, \ldots, X_n)$ là một mẫu ngẫu nhiên kích thước n từ tập chính có hàm phân phối $F(x)$.

Thống kê

Một thống kê là một hàm của mẫu ngẫu nhiên.

Một số thống kê quan trọng

Trung bình mẫu ngẫu nhiên

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

Phương sai mẫu ngẫu nhiên

$$S^2 = \frac{(X_1 - \overline{X})^2 + (X_2 - \overline{X})^2 + \cdots + (X_n - \overline{X})^2}{n-1}$$

Độ lệch tiêu chuẩn mẫu ngẫu nhiên $S = \sqrt{S^2}$.

Thống kê trung bình mẫu $\overline{X}$ là một thống kê quan trọng và được dùng trong các bài toán liên quan đến việc đánh giá trung bình tập chính thông qua trung bình mẫu.

Ký hiệu kỳ vọng của X là μ và phương sai của X là σ^2. Tương ứng $\overline{X}$ có các đặc trưng

- Kỳ vọng: $\mathbb{E}[\overline{X}] = \mathbb{E}X$ hay là $\mathbb{E}[\overline{X}] = \mu$.
- Phương sai: $\mathbb{V}[\overline{X}] = \dfrac{\mathbb{V}[X]}{n}$ hay là $\mathbb{V}[\overline{X}] = \dfrac{\sigma^2}{n}$.
- Nếu $X_1, X_2, \ldots, X_n$ có cùng phân phối chuẩn $N(\mu, \sigma^2)$ thì $\overline{X}$ sẽ có phân phối chuẩn với cùng kì vọng μ và phương sai σ^2/n.

2.5 Phân phối của các thống kê mẫu

Phân phối của trung bình mẫu

Định lý 2.1

a) Giả sử $X_1, X_2, \ldots, X_n$ là một mẫu ngẫu nhiên từ phân phối chuẩn $N(\mu, \sigma^2)$. Khi đó $\overline{X}$ có phân phối chuẩn $N(\mu, \dfrac{\sigma^2}{n})$ và thống kê

$$Z = \frac{(\overline{X} - \mu)\sqrt{n}}{\sigma}$$

có phân phối chuẩn tắc $N(0, 1)$.

b) (Định lý giới hạn trung tâm) Giả sử $X_1, X_2, \ldots, X_n$ là một mẫu ngẫu nhiên của biến ngẫu nhiên X với kì vọng μ và phương sai σ^2 hữu hạn thì phân phối của $\overline{X}$ có phân phối xấp xỉ phân phối chuẩn $N(\mu, \dfrac{\sigma^2}{n})$ và thống kê

$$Z = \frac{(\overline{X} - \mu)\sqrt{n}}{\sigma}$$

xấp xỉ phân phối chuẩn tắc $N(0, 1)$ khi n đủ lớn.

Ví dụ 2.30

Một công ty sản suất bóng đèn điện có tuổi thọ tuân theo luật phân phối chuẩn với tuổi thọ trung bình là 800 giờ và độ lệch tiêu chuẩn 40 giờ. Tính xác suất để một mẫu ngẫu nhiên 25 bóng đèn điện có tuổi thọ trung bình lớn hơn 780 giờ.

Giải: Ta gọi $\overline{X}$ là tuổi thọ trung bình của 25 bóng đèn điện. Khi đó $\overline{X}$ có phân phối chuẩn với giá trị trung bình là $\mu = 800$ và độ lệch tiêu chuẩn $\sigma = \dfrac{40}{\sqrt{25}} = 8$. Do đó $Z = \dfrac{\overline{X} - 800}{8}$ có phân phối chuẩn tắc $N(0, 1)$. Xác suất cần tính là

$$\mathbb{P}(\overline{X} > 780) = \mathbb{P}\left(\frac{\overline{X} - 800}{8} > \frac{780 - 800}{8}\right) = \mathbb{P}(Z > -2,5)$$
$$= 1 - \mathbb{P}(Z < -2,5) = 1 - \Phi(-2,5) = 1 - 0,0062 = 0,993$$

Ví dụ 2.31

Chi nhánh Hà Đông của công ty truyền hình cáp Việt Nam quản lý 70 nghìn khách hàng. Chi phí bảo hành cho mỗi khách hàng mỗi năm có thể coi là một biến ngẫu nhiên với trung bình 150 nghìn VND và độ lệch tiêu chuẩn 350 nghìn VND. Hãy ước tính xác suất để tổng chi phí bảo hành vượt quá 10,64 tỉ VND.

Giải: Ký hiệu X_i $(i = 1, 2, \dots, n = 7 \times 10^4)$ là chi phí bảo hành một năm cho khách hàng thứ i. Khi đó, chi phí bảo hành của toàn bộ hệ thống trong một năm là

$$n\bar{X} = \sum_{i=1}^{n} X_i.$$

Theo định lý giới hạn trung tâm, $\bar{X}$ là biến ngẫu nhiên có phân phối xấp xỉ chuẩn với trung bình 150 nghìn VND và độ lệch tiêu chuẩn $\dfrac{350 \times 10^3}{\sqrt{7 \times 10^4}}$.

Vậy xác suất để tổng chi phí bảo hành hàng năm vượt quá 10,64 tỉ VND là

$$\mathbb{P}(n\overline{X} > 10,64 \times 10^9) = \mathbb{P}\left(\overline{X} > \frac{10,64 \times 10^9}{7 \times 10^4}\right)$$

$$= \mathbb{P}\left(\frac{\overline{X} - 150 \times 10^3}{\dfrac{350 \times 10^3}{\sqrt{7 \times 10^4}}} > \frac{152 \times 10^3 - 150 \times 10^3}{\dfrac{350 \times 10^3}{\sqrt{7 \times 10^4}}}\right)$$

$$\approx \mathbb{P}(Z > 1,51) = 1 - \Phi(1,51) \approx 1 - 0,9345 = 0,0655.$$

Ví dụ 2.32

Một bệnh viện muốn tận dụng tầng 2 của một tòa nhà cũ gồm 2 tầng để làm khu vực khám bệnh. Để đảm bảo an toàn, người ta muốn ước tính sức chứa (bao nhiêu người?) của tầng 2 tòa nhà. Các kỹ sư xây dựng cho rằng kết cấu của tòa nhà có thể chịu được tải trọng 5 tấn. Giả sử rằng khối lượng của một người là biến ngẫu nhiên với trung bình 60 kg và độ lệch tiêu chuẩn 25 kg.

(a) Hãy ước tính sức chứa của tầng 2 tòa nhà sao cho xác suất để kết cấu tòa nhà bị phá hủy không vượt quá 5%.

(b) Sau khi gia cố tầng 2, các kỹ sư cho rằng sức chịu tải của tầng 2 tòa nhà là biến ngẫu nhiên có phân phối chuẩn với trung bình 7 tấn và độ lệch chuẩn 0,2 tấn. Hãy ước tính sức chứa của tầng 2 tòa nhà để xác suất kết cấu nhà bị phá hủy không vượt quá 5%.

Giải: a) Gọi X_i là khối lượng (tính theo kilogram) của người thứ $i \in \{1, 2, \ldots, n\}$. Ta phải tính xác suất để kết cấu của tòa nhà bị phá hủy

$$\mathbb{P}\Big(\sum_{i=1}^{n} X_i > 5 \times 10^3\Big).$$

Theo định lý giới hạn trung tâm, biến ngẫu nhiên $\overline{X} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$ có phân phối xấp xỉ chuẩn với trung bình 60 và phương sai $\dfrac{25^2}{n}$. Do đó

$$\mathbb{P}\Big(\sum_{i=1}^{n} X_i > 5 \times 10^3\Big) = \mathbb{P}(n\overline{X} > 5 \times 10^3)$$

$$= \mathbb{P}\left(\frac{\overline{X} - 60}{25/\sqrt{n}} > \frac{\dfrac{5 \times 10^3}{n} - 60}{25/\sqrt{n}}\right) = \mathbb{P}\left(Z > \frac{5 \times 10^3 - 60n}{25\sqrt{n}}\right).$$

Ta sẽ tìm n sao cho

$$\mathbb{P}\left(Z > \frac{5 \times 10^3 - 60n}{25\sqrt{n}}\right) = 1 - \Phi\left(\frac{5 \times 10^3 - 60n}{25\sqrt{n}}\right) \leq 0,05,$$

hay tương đương

$$\frac{5 \times 10^3 - 60n}{25\sqrt{n}} \geq 1,645.$$

Ta tìm được

$$n \leq 77.$$

Vậy nếu số người có mặt không vượt quá 77 người tại cùng một thời điểm trên tầng 2 của tòa nhà thì xác suất để kết cấu tòa nhà bị phá hủy sẽ ít hơn 5%. Việc ước tính này giúp cho bệnh viện có thể thiết kế các đơn vị khám bệnh sao cho vừa đảm bảo tận dụng các diện tích trống, vừa đảm bảo an toàn trong vận hành.

b) Gọi Y là biến ngẫu nhiên chỉ sức chịu tải của tầng 2 tòa nhà. Theo bài ra, Y có phân phối chuẩn với trung bình 7 tấn và độ lệch chuẩn 0,2 tấn. Xác suất để kết cấu bị phá hủy là

$$\begin{aligned}
\mathbb{P}\left(\sum_{i=1}^{n} X_i > Y\right) &= \mathbb{P}(n\overline{X} - Y > 0) \\
&= \mathbb{P}\left(\frac{(\overline{X} - Y/n) - (60 - 7 \times 10^3/n)}{\sqrt{25^2/n + 200^2/n^2}} > \frac{-(60 - 7 \times 10^3/n)}{\sqrt{25^2/n + 200^2/n^2}}\right) \\
&= \mathbb{P}\left(Z > \frac{7 \times 10^3 - 60n}{\sqrt{25^2 n + 200^2}}\right) = 1 - \Phi\left(\frac{7 \times 10^3 - 60n}{\sqrt{25^2 n + 200^2}}\right),
\end{aligned}$$

do biến ngẫu nhiên $\overline{X} - Y/n$ có phân phối xấp xỉ chuẩn với kỳ vọng

$$\mathbb{E}[\overline{X} - Y/n] = \mathbb{E}[\overline{X}] - \mathbb{E}[Y]/n = 60 - 7 \times 10^3/n$$

và phương sai

$$\mathbb{V}[\overline{X} - Y/n] = \mathbb{V}[\overline{X}] + \mathbb{V}[Y]/n^2 = 25^2/n + 200^2/n^2.$$

Cho xác suất này nhỏ hơn hoặc bằng 5%, ta thu được

$$\frac{7 \times 10^3 - 60n}{\sqrt{25^2 n + 200^2}} \geq 1,645,$$

tức là

$$n \leq 107.$$

Vậy sau khi gia cố tòa nhà, số người tối đa tầng 2 có thể chứa là 107 người, sao cho xác suất kết cấu bị phá hủy không lớn hơn 5%.

2.5 Phân phối của các thống kê mẫu

Phân phối của tỷ lệ mẫu

- Ta xét một dấu hiệu $\mathcal{T}$ nào đó trên các phần tử của tập chính. Gọi p là tỷ lệ phần tử của tập chính có dấu hiệu $\mathcal{T}$.
- Gọi X là biến ngẫu nhiên chỉ số lượng phần tử có dấu hiệu $\mathcal{T}$ khi kiểm tra một phần tử được chọn ngẫu nhiên từ tập chính, khi đó phân phối của tập chính được xác định bởi hàm khối xác suất
$$f(x) = \mathbb{P}(X = x) = \begin{cases} p & x = 1, \\ 1 - p & x = 0. \end{cases}$$
- Mẫu ngẫu nhiên $(X_1, X_2, \ldots, X_n)$ từ tập chính với hàm khối xác suất $f(x)$, nghĩa là X_i $(i = 1, \ldots, n)$ có cùng phân phối với biến ngẫu nhiên gốc X.
- Tổng $X_1 + X_2 + \cdots + X_n$ chính bằng số phần tử trong mẫu chứa dấu hiệu $\mathcal{T}$.

- Thống kê trung bình mẫu sẽ xác định tỉ lệ mẫu chứa dấu hiệu $\mathcal{T}$ và được gọi là tần suất mẫu ngẫu nhiên, ký hiệu bởi $\hat{p}$

$$\hat{p} = \frac{X_1 + X_2 + \cdots + X_n}{n}.$$

Khi n đủ lớn, tần suất mẫu ngẫu nhiên $\hat{p}$ sẽ có phân phối xấp xỉ chuẩn với trung bình p và phương sai $\dfrac{p(1-p)}{n}$.

Định lý 2.2

Nếu $n > 30$ và $np \geqslant 5,\ n(1-p) \geqslant 5$ thì thống kê tần suất mẫu $\hat{p}$ có phân phối xấp xỉ phân phối chuẩn $N\left(\mu = p, \sigma^2 = \dfrac{p(1-p)}{n}\right)$ và thống kê

$$Z = \frac{\hat{p} - p}{\sqrt{p(1-p)/n}}$$

sẽ có phân phối xấp xỉ phân phối chuẩn tắc $N(0, 1)$.

Ví dụ 2.33

Biết rằng tỷ lệ sinh viên Giao thông vận tải khi ra trường có việc làm đúng chuyên ngành là 80%. Kiểm tra ngẫu nhiên 50 sinh viên, tính xác suất để có ít nhất 45 sinh viên làm việc đúng chuyên ngành.

Giải: Ta gọi tần suất mẫu ngẫu nhiên có kích thước $n = 50$ là $\hat{p}$. Ta cần tính xác suất

$$\mathbb{P}\left(\hat{p} \geq \frac{45}{50}\right) = \mathbb{P}(\hat{p} \geq 0,9).$$

Kích thước mẫu $n = 50$ là đủ lớn ($np = 50 \times 0,8 = 40 > 5$ và $n(1-p) = 50 \times 0,2 = 10 > 5$), do đó tần suất mẫu $\hat{p}$ có phân phối xấp xỉ phân phối chuẩn với kì vọng $p = 0,8$ và phương sai $\dfrac{p(1-p)}{n} = 0,0032$. Vậy ta tính được

$$\mathbb{P}(\hat{p} \geq 0,9) = \mathbb{P}\left(\frac{\hat{p} - 0,8}{0,0566} > \frac{0,9 - 0,8}{0,0566}\right) = \mathbb{P}(Z > 1,77)$$
$$= 1 - \Phi(1,77) = 1 - 0,9616 = 0,0384.$$

Ví dụ 2.34

Một công ty vận tải hành khách bằng đường hàng không thường bán ra số lượng vé nhiều hơn số lượng ghế trên mỗi chuyến bay để tối đa hóa lợi nhuận khi có hành khách hủy chuyến. Số liệu trước đây cho thấy chỉ có 90% số hành khách đã mua vé và đến làm thủ tục bay. Giả sử với máy bay Airbus A330 có sức chứa 440 ghế, công ty bán ra 480 vé. Hãy tính xác suất để có nhiều hơn 440 khách mua vé và đến làm thủ tục bay.

Giải: Ta coi mỗi vé bán ra là một lần thực hiện phép thử. Biến ngẫu nhiên X_i nhận giá trị bằng 1 nếu hành khách thứ i mua vé và làm thủ tục bay, $X_i = 0$ nếu hành khách hủy chuyến. Như vậy với $n = 480$, ta có một mẫu ngẫu nhiên $(X_1, X_2, \ldots, X_n)$. Xác suất để có nhiều hơn 440 khách mua vé và đến làm thủ tục bay là

$$\mathbb{P}(X_1+X_2+\cdots+X_n > 440) = \mathbb{P}\left(\frac{X_1+X_2+\cdots+X_n}{n} > \frac{440}{480}\right) = \mathbb{P}\left(\hat{p} > \frac{11}{12}\right),$$

trong đó $\hat{p}$ là tần suất hành khách mua vé và đến làm thủ tục bay. Theo định lý giới hạn trung tâm, $\hat{p}$ có phân phối xấp xỉ phân phối chuẩn với trung bình $0,9$ và phương sai $\dfrac{0,9(1-0,9)}{480}$. Do đó

$$\mathbb{P}\left(\hat{p} > \frac{11}{12}\right) = \mathbb{P}\left(\frac{\hat{p}-0,9}{\sqrt{0,9(1-0,9)/480}} > \frac{11/12-0,9}{\sqrt{0,9(1-0,9)/480}}\right)$$
$$\approx \mathbb{P}(Z > 1,22) = 1 - \Phi(1,22) = 1 - 0,8888 = 0,1112.$$

Như vậy có thể nói, chỉ có khoảng 1 lần trong số 10 lần bán vượt số vé quy định, số hành khách mua vé và đến làm thủ tục bay mới vượt sức chứa của máy bay.

2.5 Phân phối của các thống kê mẫu

Phân phối của phương sai mẫu

Thống kê phương sai mẫu được xác định bởi

$$S^2 = \frac{(X_1 - \overline{X})^2 + (X_2 - \overline{X})^2 + \cdots + (X_n - \overline{X})^2}{n-1} = \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - n(\overline{X})^2\right)$$

Do đó ta có

$$\begin{aligned}
\mathbb{E}\left[S^2\right] &= \frac{1}{n-1}\left(\sum_{i=1}^{n}\mathbb{E}[X_i^2] - n\mathbb{E}[(\overline{X})^2]\right) \\
&= \frac{1}{n-1}\left(\sum_{i=1}^{n}(\mathbb{E}[X_i^2] - \mu^2) - n(\mathbb{E}[(\overline{X})^2] - \mu^2)\right) \\
&= \frac{1}{n-1}\left(n\sigma^2 - n\frac{\sigma^2}{n}\right) = \sigma^2.
\end{aligned}$$

2.5 Phân phối của các thống kê mẫu

Phân phối của phương sai mẫu

Định nghĩa 2.9

Giả sử $X_1, X_2, \ldots, X_n$ là các biến ngẫu nhiên có phân phối chuẩn tắc $N(0,1)$ và độc lập với nhau. Biến ngẫu nhiên

$$\chi_n^2 = X_1^2 + X_2^2 + \cdots + X_n^2$$

được gọi là biến ngẫu nhiên khi bình phương với n bậc tự do.

Hàm mật độ của biến χ_n^2 là

$$f(x) = \begin{cases} \dfrac{1}{2^{n/2}\Gamma(n/2)} x^{n/2-1} e^{-x/2} & x > 0 \\ 0 & x \leqslant 0 \end{cases}$$

trong đó $\Gamma(x) = \int\limits_0^{+\infty} e^{-t} t^{x-1} dt$ là kí hiệu của hàm Gamma.

Định lý 2.3

Giả sử X và Y độc lập có phân phối khi bình phương với bậc tự do m và n. Khi đó biến ngẫu nhiên $X + Y$ cũng có phân phối khi bình phương với bậc tự do $m + n$.

Định lý 2.4

Cho $X_1, X_2, \ldots, X_n$ là các biến ngẫu nhiên độc lập và có cùng phân phối chuẩn $N(\mu, \sigma^2)$. Khi đó, $\dfrac{(n-1)S^2}{\sigma^2}$ có phân phối khi bình phương với $n-1$ bậc tự do.

2.5 Phân phối của các thống kê mẫu

Phân phối của phương sai mẫu

Phân vị $\chi^2_{n,\alpha}$: Với các giá trị α gần 1 và α gần 0, tồn tại duy nhất một giá trị $\chi^2_{n,\alpha}$ thỏa mãn điều kiện

$$\mathbb{P}(\chi^2_n > \chi^2_{n,\alpha}) = \alpha.$$

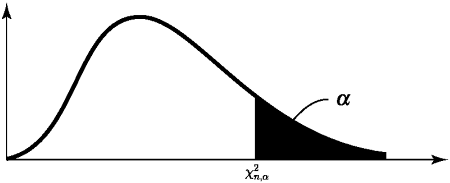


Figure: Hàm mật độ và phân vị $\chi^2_{n,\alpha}$ của phân phối χ^2_n.

Các giá trị phân vị $\chi^2_{n,\alpha}$ được tính sẵn (SGK).

Ví dụ 2.36

Cho χ^2 là biến ngẫu nhiên có phân phối khi bình phương với $n = 20$ bậc tự do. Tìm hai số $a < b$ sao cho

$$\mathbb{P}(a < \chi^2 < b) = 0,95, \quad \text{và} \quad \mathbb{P}(\chi^2 < a) = \mathbb{P}(\chi^2 > b).$$

Giải: Ta có

$$\mathbb{P}(\chi^2 < a) + \mathbb{P}(\chi^2 > b) = 1 - \mathbb{P}(a < \chi^2 < b) = 1 - 0,95 = 0,05.$$

Do $\mathbb{P}(\chi^2 < a) = \mathbb{P}(\chi^2 > b)$ nên

$$\mathbb{P}(\chi^2 < a) = \mathbb{P}(\chi^2 > b) = 0,05/2 = 0,025.$$

Tương ứng, ta có

$$\mathbb{P}(\chi^2 > b) = 0,025 = \mathbb{P}(\chi^2 > \chi^2_{20;0,025}).$$

Do đó tra bảng phân phối chi bình phương

$$b = \chi^2_{20;0,025} = 34,170.$$

Để tính giá trị của a ta sử dụng đẳng thức

$$\mathbb{P}(\chi^2 > a) = 1 - \mathbb{P}(\chi^2 < a) = 1 - 0,025 = 0,975$$

và tra bảng khi bình phương ta tính được

$$a = \chi^2_{20;0,975} = 9,591.$$

Ví dụ 2.37

Một cánh tay rô-bốt sử dụng một thiết bị điện tử để xác định vị trí đồ vật trong không gian ba chiều. Quá trình định vị luôn làm nảy sinh ra sai số. Giả sử các sai số X, Y, Z tương ứng theo ba chiều không gian là các biến ngẫu nhiên độc lập có cùng phân phối chuẩn với trung bình 0 mm và độ lệch tiêu chuẩn 5 mm. Hãy tính xác suất để điểm được định vị bởi cánh tay rô-bốt lệch ra khỏi vị trí thực tế của vật thể không vượt quá 12,5 mm.

Giải: Gọi W là khoảng cách từ điểm định vị bởi cánh tay rô-bốt đến vị trí thực tế của vật thể. Rõ ràng ta có

$$W = \sqrt{X^2 + Y^2 + Z^2}.$$

Xác suất cần tìm là

$$\mathbb{P}(W \leqslant 12,5) = \mathbb{P}(W^2 \leqslant 156,25) = \mathbb{P}(X^2 + Y^2 + Z^2 \leqslant 156,25).$$

Bằng cách đưa các biến ngẫu nhiên X, Y, Z có phân phối chuẩn $N(0,25)$ về biến chuẩn tắc $N(0,1)$, ta thu được

$$\begin{aligned}\mathbb{P}(X^2+Y^2+Z^2 \leqslant 156,25) &= \mathbb{P}(X^2/25+Y^2/25+Z^2/25 \leqslant 6,25)\\ &= \mathbb{P}(\chi_3^2 \leqslant 6,25) = 1-\mathbb{P}(\chi_3^2 > 6,25)\\ &= 1-0,1 = 0,9.\end{aligned}$$

Trong bài toán này, xác suất $\mathbb{P}(\chi_3^2 > 6,25)$ có thể tính được bằng cách tra ngược bảng phân vị khi bình phương như sau: Ta tìm trong hàng $n=3$ số gần với $6,25$ nhất, đó là số $6,251$, dò ngược lên trên ta thu được

$$\alpha = 0,1 = \mathbb{P}(\chi_3^2 > 6,25).$$

Định nghĩa 2.10

Giả sử Z và χ_n^2 là các biến ngẫu nhiên độc lập và tương ứng có phân phối chuẩn tắc $N(0, 1)$ và phân phối khi bình phương với n bậc tự do. Khi đó, biến ngẫu nhiên

$$T = \frac{Z}{\sqrt{\chi_n^2/n}}$$

được gọi là biến ngẫu nhiên Student với n bậc tự do.

Phân phối Student $T = T_n$ với n bậc tự do có hàm mật độ xác định bởi công thức

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

Một số trường hợp đặc biệt

a) Với $n = 1$ hàm mật độ là

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

b) Với $n = 2$ hàm mật độ là

$$f(x) = \frac{1}{\sqrt{(1+x^2)^3}}.$$

c) Với $n > 30$ hàm mật độ xấp xỉ phân phối chuẩn tắc $N(0,1)$, nghĩa là hàm mật độ xấp xỉ bởi

$$f(x) \approx \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

2.5 Phân phối của các thống kê mẫu

Phân phối student

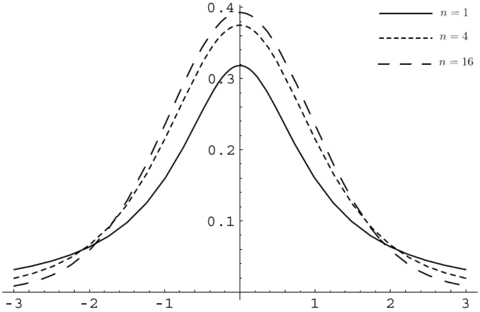


Figure: Hàm mật độ của phân phối Student đối với các bậc tự do khác nhau.

Hình trên là đồ thị của hàm mật độ phân phối Student với các bậc tự do $n = 1, 4, 16$. Hàm mật độ là hàm đối xứng có dạng hình chuông. Khi giá trị bậc tự do càng lớn thì hàm mật độ càng gần hàm mật độ của phân phối chuẩn tắc.

Định lý 2.5

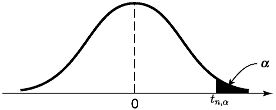
Cho $X_1, X_2, \ldots, X_n$ là các biến ngẫu nhiên độc lập cùng phân phối với phân phối chuẩn $N(\mu, \sigma^2)$. Khi đó, thống kê

$$T = \frac{(\overline{X} - \mu)\sqrt{n}}{S}$$

có phân phối Student với $n - 1$ bậc tự do.

Phân vị $t_{n,\alpha}$: Là giá trị duy nhất xác định theo đẳng thức

$$\mathbb{P}(T > t_{n,\alpha}) = \alpha$$



Một số công thức với biến student

1. $\mathbb{P}(T < -t_{n,\alpha}) = \mathbb{P}(T > t_{n,\alpha}) = \alpha$.
2. $\mathbb{P}(|T| > t_{n,\alpha}) = \mathbb{P}(T < -t_{n,\alpha}) + \mathbb{P}(T > t_{n,\alpha}) = 2\alpha$,
3. $\mathbb{P}(|T| < t_{n,\alpha}) = 1 - \mathbb{P}(|T| > t_{n,\alpha}) = 1 - 2\alpha$.
4. $t_{n,1-\alpha} = -t_{n,\alpha}$.

Ví dụ 2.40

Cho T là biến ngẫu nhiên có phân phối Student với $n = 20$ bậc tự do. Tìm hai số $a < b$ sao cho

$$\mathbb{P}(a < T < b) = 0,95, \quad \text{và} \quad \mathbb{P}(T < a) = \mathbb{P}(T > b).$$

Giải: Ta có $\mathbb{P}(T < a) + \mathbb{P}(T > b) = 1 - \mathbb{P}(a < T < b) = 1 - 0,95 = 0,05$. Do $\mathbb{P}(T < a) = \mathbb{P}(T > b)$ nên

$$\mathbb{P}(T < a) = \mathbb{P}(T > b) = 0,05/2 = 0,025.$$

Tra bảng phân phối Student ta có

$$\mathbb{P}(T > b) = 0,025 = \mathbb{P}(T > t_{20;0,025}).$$

Do đó

$$b = t_{20;0,025} = 2,086.$$

Ta biết rằng hàm mật độ của phân phối Student đối xứng qua trục tung nên $a = -b = -t_{20;0,025} = -2,086$.

Định nghĩa 2.11

Giả sử χ_n^2 và χ_m^2 là các biến ngẫu nhiên độc lập và có phân phối khi bình phương với n và m bậc tự do tương ứng. Khi đó biến ngẫu nhiên

$$F = \frac{\chi_n^2/n}{\chi_m^2/m}$$

được gọi là biến ngẫu nhiên Fisher với n và m bậc tự do, ký hiệu là $F_{n,m}$.

Hàm mật độ của phân phối $F_{n,m}$ được xác định bởi công thức

$$f(x) = \frac{\Gamma((n+m)/2)}{\Gamma(n/2)\Gamma(m/2)} n^{n/2} m^{m/2} x^{n/2-1} (nx+m)^{-(n+m)/2}, \ x > 0.$$

Định lý 2.6

Trên hai tập chính xét hai phân phối độc lập tương ứng tuân theo luật chuẩn lần lượt là $N(\mu_1, \sigma_1^2)$ và $N(\mu_2, \sigma_2^2)$. Giả sử S_1^2 và S_2^2 là các phương sai mẫu của hai mẫu ngẫu nhiên từ hai tập chính trên với kích thước lần lượt là n_1 và n_2. Khi đó

$$F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}$$

có phân phối F với số bậc tự do là $n = n_1 - 1,\ m = n_2 - 1$.

Phân vị $f_{\alpha,n,m}$: Là giá trị duy nhất xác định bởi
$\mathbb{P}(F_{n,m} > f_{n,m,\alpha}) = \alpha$

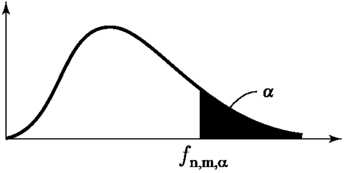


Figure: Giá trị phân vị $f_{\alpha,n,m}$ của phân phối $F_{n,m}$.

Ví dụ 2.41

Một số giá trị phân vị của phân phối Fisher

a) $f_{5;20;0,05} = 2,71$.

b) $f_{5;20;0,01} = 4,10$.

c) $f_{20;5;0,95} = \dfrac{1}{f_{5;20;0,05}} = \dfrac{1}{2,71} = 0,369$.

Chú ý rằng mức phân vị ở câu (b) trong ví dụ trên không thể tra được từ bảng tra phân phối Fisher $\alpha = 0,01$. Khi đó, ta phải dùng một trong các phần mềm hỗ trợ thống kê Spreadsheet hoặc Excel để tính.

Luyện tập

Bài 1.10 : Cường độ chịu nén của bê tông tông mác 200 được xem như biến ngẫu nhiên chuẩn với trung bình $\mu = 90kG/cm2$ và độ lệch tiêu chuẩn $\sigma = 4,5kG/cm2$. Nếu kiểm tra 7 mẫu bê tông mác 200 thì xác suất để cường độ chịu nén trung bình của các mẫu bê tông này nằm trong khoảng $(87,3;92,5)$ là bao nhiêu? Lời giải:

- Gọi X_i là biến ngẫu nhiên chỉ cường độ chịu nén của mẫu bê tông mác 200 thứ i. Theo giả thiết X_i có phân phối chuẩn với $\mathbb{E}X = 90, \mathbb{V}X = \sigma^2 = 4,5^2$.
- $\overline{X}$ là cường độ nén trung bình của 7 mẫu bê tông. Ta cần tính xác suất $\mathbb{P}(87,3 < \overline{X} < 92,5)$.

- Dẫn tới , $\overline{X}$ có phân phối chuẩn với giá trị trung bình là $\mu = 90$ và độ lệch tiêu chuẩn $\sigma = \dfrac{4,5}{\sqrt{7}}$.
- Dẫn đến xác suất cần tính là

$$\mathbb{P}(87,3 < \overline{X} < 92,5) = \Phi(\frac{92,5 - 90}{4,5/\sqrt{7}}) - \Phi(\frac{87,3 - 90}{4,5/\sqrt{7}}).$$

Bài 1.12: Một công ty điện lực quản lý một vùng dân cư gồm 20 nghìn hộ dân cư. Lượng điện tiêu thụ của mỗi hộ gia đình trong một tháng được xem như biến ngẫu nhiên với trung bình 370 kwh và độ lẹch tiêu chuẩn 350kwh. Hãy ước tính xác suất để tổng lượng điện tiêu thụ trong một tháng không vượt quá $7,5$ triệu kwh.

Lời giải:

- Gọi X_i là biến ngẫu nhiên chỉ lượng điện tiêu thụ trong 1 tháng của hộ gia đình thứ $i, i = 1, 2, ..., 20000$. Theo giả thiết X_i có phân phối tùy ý với $\mathbb{E}X = 370, \mathbb{V}X = \sigma^2 = 350^2$.
- $X_1 + X_2 + \ldots + X_{20000}$ là tổng lượng điện tiêu thụ của 20 nghìn hộ dân trong 1 tháng. Ta cần tính xác suất $\mathbb{P}(X_1 + X_2 + \ldots + X_{20000} \leq 7,5.10^6)$ hay chính

$$\mathbb{P}(\overline{X} = \frac{X_1 + X_2 + \ldots + X_{20000}}{20000} \leq \frac{7,5.10^6}{20000}$$

- Theo định lý giới hạn trung tâm, $\overline{X}$ có phân phối xấp xỉ phân phối chuẩn với giá trị trung bình là $\mu = 370$ và độ lệch tiêu chuẩn $\sigma = \dfrac{350}{\sqrt{20000}}$.
- Dẫn đến xác suất cần tính là

$$\mathbb{P}(\overline{X} \leq 375) = \Phi(\frac{375 - 350}{350/\sqrt{20000}})....$$

Bài 1.13: Một kết cấu thép có thể chịu được tải trọng tối đa $3,5$ tấn. giả sử trong lượng của mỗi kiện hàng là biến ngẫu nhiên có trung bình 59 kg và độ lệch tiêu chuẩn 15 kg. Hãy ước tính số lượng kiện hàng lớn nhất có thể đặt lên kết cấu để xác suất kết cấu an toàn cao hơn 99%. Lời giải:

- Gọi X_i là biến ngẫu nhiên chỉ .... thứ $i, i =$ .... Theo giả thiết X_i có phân phối tùy ý với $\mathbb{E}X = 59, \mathbb{V}X = \sigma^2 = 15^2$.
- $X_1 + X_2 + \ldots + X_n$ là trọng lượng đặt lên kết cấu. Kết cấu an toàn khi $X_1 + X_2 + \ldots + X_n \leq 3500$. Ta cần tìm n sao cho $\mathbb{P}(X_1 + X_2 + \ldots + X_n \leq 3500) > 0,99$ hay chính là tìm n sao cho

$$\mathbb{P}(\overline{X} = \frac{X_1 + X_2 + \ldots + X_n}{n} \leq \frac{3500}{n}) > 0,99.$$

- Theo định lý giới hạn trung tâm, $\overline{X}$ có phân phối xấp xỉ phân phối chuẩn chuẩn với giá trị trung bình là $\mu = 59$ và độ lệch tiêu chuẩn $\sigma = \frac{15}{\sqrt{n}}$.

- Dẫn đến

$$
\begin{aligned}
&\mathbb{P}(\overline{X} \leq \frac{3500}{n}) > 0,99 \\
&\Leftrightarrow \Phi(\frac{\frac{3500}{n} - 59}{15/\sqrt{n}}) > \Phi(2,33) \\
&\Leftrightarrow \frac{\frac{3500}{n} - 59}{15/\sqrt{n}} > 2,33
\end{aligned}
$$

Chú ý: Hàm $\Phi(z)$ là hàm đồng biến.

Bài 1.16: Thống kê cho thấy tỷ lệ sinh viên tốt nghiệp năm 2016 từ bỏ công việc đầu tiên sau 3 tháng thử việc là 60%. Hỏi phải tiến hành thăm dò bao nhiêu sinh viên tốt nghiệp năm 2016 để tỷ lệ sinh viên gắn bó với công việc đầu tiên sau 3 tháng nằm trong khoảng 40$\pm$5% với xác suất nhiều hơn 90%.

Giải:

- Gọi p là tỷ lệ sinh viên tốt nghiệp năm 2016 gắn bó với công việc đầu tiên sau 3 tháng thử việc. Theo giả thiết $p = 1 - 0,6 = 0,4$.
  Ta gọi $\hat{p}$ tần suất mẫu ngẫu nhiên có kích thước n tức là tỷ lệ sinh viên gắn bó với công việc đầu tiên sau 3 tháng thử việc trong số n sinh viên được chọn ngẫu nhiên. Ta cần tìm n sao cho
  $$\mathbb{P}(0,35 < \hat{p} < 0,45) > 0,9$$
  .

- Tần suất mẫu $\hat{p}$ có phân phối xấp xỉ phân phối chuẩn với kì vọng $p = 0,4$ và phương sai $\dfrac{p(1-p)}{n} = \dfrac{0,24}{n}$. Dẫn tới

$$\begin{aligned}
&\mathbb{P}(0,35 < \hat{p} < 0,45) > 0,9 \\
&\Leftrightarrow \Phi\Big(\frac{0,45-0,4}{\sqrt{0,24/n}}\Big) - \Phi\Big(\frac{0,35-0,4}{\sqrt{0,24/n}}\Big) > 0,9 \\
&\Leftrightarrow \Phi\Big(\frac{0,05}{\sqrt{0,24/n}}\Big) - \Phi\Big(\frac{-0,05}{\sqrt{0,24/n}}\Big) > 0,9 \\
&\Leftrightarrow \Phi\Big(\frac{0,05}{\sqrt{0,24/n}}\Big) - \Big(1 - \Phi\Big(\frac{0,05}{\sqrt{0,24/n}}\Big)\Big) > 0,9 \\
&\Leftrightarrow \Phi\Big(\frac{0,05}{\sqrt{0,24/n}}\Big) > 0,95 \\
&\Leftrightarrow \Phi\Big(\frac{0,05}{\sqrt{0,24/n}}\Big) > \Phi(1,96) \\
&\Leftrightarrow \frac{0,05}{\sqrt{0,24/n}} > 1,96
\end{aligned}$$

Chú ý: $\Phi(x) + \Phi(-x) = 1$.

Bài 1.17: Một nhà máy chế tạo một loại thiết bị điện tử có khả năng tự ngắt điện khi nhiệt độ tăng lên quá cao. Nghiên cứu trước đó cho thấy nhiệt độ làm cho thiết bị tự ngắt có phân phối chuẩn với phương sai σ^2. Hãy xác định giá trị a sao cho

$$\mathbb{P}(S^2/\sigma^2 < a) = 0,9$$

trong đó S^2 là phương sai của mẫu gồm 5 thiết bị được chọn ngẫu nhiên để kiểm tra. colorblue Lời giải

- Gọi X_i là biến ngẫu nhiên chỉ nhiệt độ làm cho thiết bị điện tử tự ngắt, $i = 1, 2, 3, 4, 5$.
  Theo giả thiết các X_i có phân phối chuẩn với phương sai σ^2.
- Ta có

$$\begin{aligned}
&\mathbb{P}(S^2/\sigma^2 < a) = 0,9 \\
\Leftrightarrow\ &\mathbb{P}\Big(\frac{(n-1)S^2}{\sigma^2} < (n-1)a\Big) = 0,9 \\
\Leftrightarrow\ &\mathbb{P}\Big(\frac{(n-1)S^2}{\sigma^2} > (n-1)a\Big) = 0,1
\end{aligned}$$

- Theo định lý về luật phân phối của phương sai mẫu thì $\dfrac{(n-1)S^2}{\sigma^2}$ có phân phối khi bình phương với $(n-1)$ bậc tự do. Phân vị $\chi^2_{n,\alpha}$: Với các giá trị α gần 1 và α gần 0, tồn tại duy nhất một giá trị $\chi^2_{n,\alpha}$ thỏa mãn điều kiện

$$\mathbb{P}(\chi^2_n > \chi^2_{n,\alpha}) = \alpha.$$

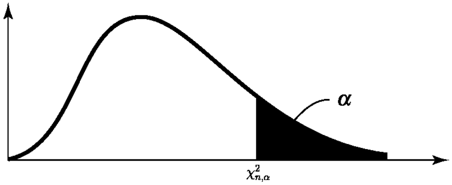


Figure: Hàm mật độ và phân vị $\chi^2_{n,\alpha}$ của phân phối χ^2_n.

- Dẫn đến $(n-1)a = \chi^2_{n-1;0,1}$. Thay $n = 5$ vào công thức tìm được a.

Bài 1.9/c) Cần tìm n sao cho

$$\mathbb{P}(173,5 < \bar{X} < 175,5) \geq 0,95$$

Giống bài 1.16