Chapter 6 Statistics Principles and Sampling Distributions
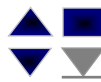
§1 Introduction and Basic principles

（引言和基本概念）

§2 Sampling distributions

（抽样分布）

Chapter 6 Statistics Principles and Sampling Distributions

§1 Introduction and Basic principles

(引言和基本概念)

§2 Sampling distributions

(抽样分布)

Sampling distribution

Samples $X_1, X_2, \cdots, X_n$ — Contain useful information

Statistics $g(X_1, X_2, \cdots, X_n)$ — Extract useful information from the data, e.g.:
$$\frac{1}{n}\sum_i^n X_i, \ \max_{1 \leq i \leq n} X_i$$

These are random variables. Need to determine their distributions which are called the sampling distributions.

Discussion ① Sampling distributions of the Standard Normal Distribution $N(0,1)$

$$\chi^2 - \text{distribution}, t - \text{distribution}, F - \text{distribution}$$

② Sampling distributions of general normal populations $N(\mu, \sigma^2)$

5 theorems about sampling distribution

(1) χ^2-distribution (卡方分布)

Let $X_1, X_2, \ldots, X_n$ be samples from the population $X \sim N(0,1)$, denote

$$\chi^2 = X_1^2 + X_2^2 + \cdots + X_n^2.$$

Then χ^2 follows the χ^2-distribution with degree of freedom n, denoted as $\chi^2 \sim \chi^2(n)$.

What is the degree of freedom (自由度)?

Straightforward meaning: degree of freedom is the number of variables that can change independently.

Strict meaning: In linear algebra, $\chi^2 = X_1^2 + X_2^2 + \cdots + X_n^2$ is a quadratic form. Degree of freedom is the rank of the quadratic form (二次型的秩).

(1) χ^2-distribution (卡方分布)

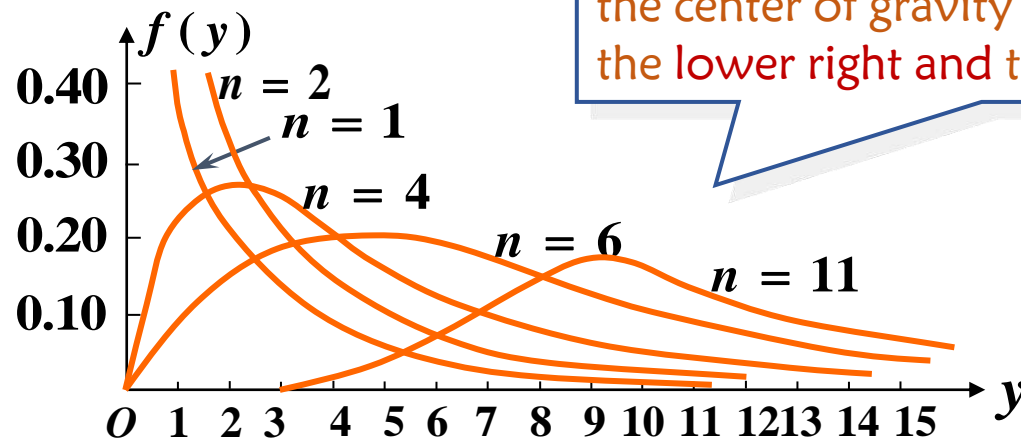Let $X_1, X_2, \dots, X_n$ be a sample from the population $X \sim N(0,1)$, denote

$$\chi^2 = X_1^2 + X_2^2 + \cdots + X_n^2.$$

Then χ^2 follows the χ^2-distribution with degree of freedom n, denoted as $\chi^2 \sim \chi^2(n)$.
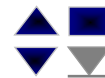
⬤ **Density function of the χ^2-distribution**

$$f(y) = \begin{cases} \dfrac{1}{2^{n/2}\Gamma(n/2)} y^{\frac{n}{2}-1} e^{-\frac{y}{2}}, & y > 0 \\ 0, & y \le 0 \end{cases}$$

The figure of χ^2-distribution

With the increases of the degree of freedom, the center of gravity of the curve moves to the lower right and the curve tends to be flat.

(1) χ^2-distribution (卡方分布)

Let $X_1, X_2, \dots, X_n$ be a sample from the population $X \sim N(0,1)$, denote

$$\chi^2 = X_1^2 + X_2^2 + \cdots + X_n^2.$$

Then χ^2 follows the χ^2-distribution with degree of freedom n, denoted as $\chi^2 \sim \chi^2(n)$.

● Additivity (可加性) of the χ^2-distribution

If $\chi_1^2 \sim \chi^2(n_1)$, $\chi_2^2 \sim \chi^2(n_2)$, and χ_1^2, χ_2^2 are independent, then

$$\chi_1^2 + \chi_2^2 \sim \chi^2(n_1 + n_2)$$

$$X_1^2 + X_2^2 + \cdots + X_{n_1}^2 \sim \chi^2(n_1) \qquad Y_1^2 + Y_2^2 + \cdots + Y_{n_2}^2 \sim \chi^2(n_2)$$

i.i.d. $N(0,1)$

$$\therefore X_1^2 + X_2^2 + \cdots + X_{n_1}^2 + Y_1^2 + Y_2^2 + \cdots Y_{n_2}^2 \sim \chi^2(n_1 + n_2)$$

Still i.i.d. $N(0,1)$

(1) χ^2-distribution (卡方分布)

Let $X_1, X_2, \ldots, X_n$ be a sample from the population $X \sim N(0,1)$, denote

$$\chi^2 = X_1^2 + X_2^2 + \cdots + X_n^2.$$

Then χ^2 follows the χ^2-distribution with degree of freedom n, denoted as $\chi^2 \sim \chi^2(n)$.

● **Numerical characteristics of the χ^2-distribution (卡方分布的数字特征)**

If $\chi^2 \sim \chi^2(n)$, then $\boxed{E(\chi^2) = n, \quad D(\chi^2) = 2n.}$

Proof: Pick n i.i.d. $N(0,1)$ r.v. $X_1, X_2, \ldots, X_n$, then χ^2 and $X_1^2 + X_2^2 + \cdots + X_n^2$ have the same distribution, so:

$$E(\chi^2) = E\left(\sum_{i=1}^{n} X_i^2\right) = \sum_{i=1}^{n} E(X_i^2) = \sum_{i=1}^{n} 1 = n$$

$$D(\chi^2) = \sum_{i=1}^{n} D(X_i^2) = nD(X_1^2) = n\left\{E(X_1^4) - \left[E(X_1^2)\right]^2\right\}$$

$$= n\left(\int_{-\infty}^{\infty} x^4 \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx - 1\right) = n(3-1) = 2n$$

Exercise： Suppose $X_1, X_2, \ldots, X_6$ are independent samples from $N(0,1)$ population. Find out C_1 and C_2 which make the below follows χ^2 distribution.

$$Y = C_1(X_1 + X_2)^2 + C_2(X_3 + X_4 + X_5 + X_6)^2$$

Answer:

1. $X_1, X_2, \ldots, X_n$ are i.i.d. random variables with their population. Based on the features, the linear combination of normal distribution still follows normal distribution

$$X_1 + X_2 \sim ? \qquad\qquad X_3 + X_4 + X_5 + X_6 \sim ?$$

2. Z transfer them: **?**

3. x^2-distribution

If $X_1, X_2, \ldots, X_n$ are the samples of the population $X \sim N(0,1)$, let

$$\boxed{\chi^2 = X_1^2 + X_2^2 + \cdots + X_n^2}$$

χ^2 follows x^2-distribution with degree of freedom of n, noted as $\boxed{\chi^2 \sim \chi^2(n)}$

4. The Additivity of x^2-distribution

If $X_1^2 \sim \chi^2(n_1)$, $X_2^2 \sim \chi^2(n_2)$, and $X_1^2 \ and \ X_2^2$ are independent, then

$$\boxed{\chi^2(n_1) + \chi^2(n_2) \sim \chi^2(n_1 + n_2)}$$

(2) *t*-distribution

If $X \sim N(0,1), Y \sim \chi^2(n)$, X and Y are independent, let

$$t = \frac{X}{\sqrt{Y/n}}$$

Then t follows the *t*-distribution with degree of freedom n, noted as $t \sim t(n)$.

⬤ **Density function of the *t*-distribution**

$$f(x) = \frac{\Gamma[(n+1)/2]}{\sqrt{n\pi}\,\Gamma(n/2)} \left(1 + \frac{x^2}{n}\right)^{-(n+1)/2} \quad , -\infty < x < \infty$$
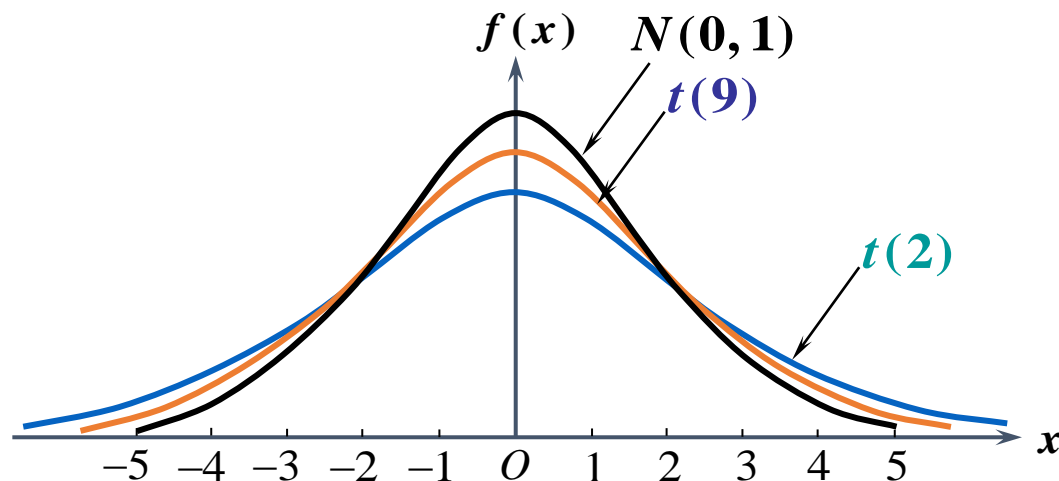
Thus $f(-x) = f(x),$

$f'(x) > 0$ for $x < 0,$

$f'(x) < 0$ for $x > 0$

$\lim\limits_{x \to -\infty} f(x) = 0 ,$

$\lim\limits_{x \to +\infty} f(x) = 0$

With the increase of the degree of freedom, the curve approaches $N(0,1)$
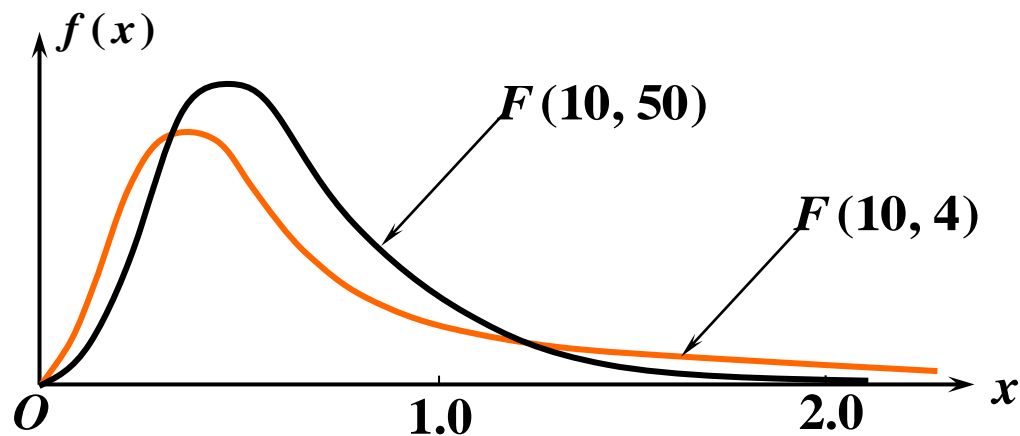
(3) F-distribution

If $U \sim \chi^2(n_1), V \sim \chi^2(n_2)$, U and V are independent, let

$$F = \frac{U/n_1}{V/n_2},$$

Then F follows the F-distribution with degree of freedom (n_1, n_2), denoted as $F \sim F(n_1, n_2)$.

● **Density function of the F-distribution**

$$f(x) = \begin{cases} \dfrac{\Gamma[(n_1 + n_2)/2]}{\Gamma(n_1/2)\Gamma(n_2/2)} n_1^{n_1/2} n_2^{n_2/2} \dfrac{x^{n_1/2-1}}{(n_1 x + n_2)^{(n_1+n_2)/2}}, & x > 0 \\ 0, & x \le 0 \end{cases}$$

$f(x)$

$F(10, 50)$

$F(10, 4)$

O　　　1.0　　　2.0　　　x

(3) F-distribution

If $U \sim \chi^2(n_1), V \sim \chi^2(n_2)$, U and V are independent, let

$$F = \frac{U/n_1}{V/n_2},$$

Then F follows the F-distribution with degree of freedom (n_1, n_2), denoted as $F \sim F(n_1, n_2)$.

● **Density function of the F-distribution**

$$f(x) = \begin{cases} \dfrac{\Gamma[(n_1+n_2)/2]}{\Gamma(n_1/2)\Gamma(n_2/2)} n_1^{n_1/2} n_2^{n_2/2} \dfrac{x^{n_1/2-1}}{(n_1 x + n_2)^{(n_1+n_2)/2}}, & x > 0 \\ 0, & x \le 0 \end{cases}$$

● **Important property of the F-distribution**

If $F \sim F(n_1, n_2)$, then $\dfrac{1}{F} \sim F(n_2, n_1)$.

Exercise：Suppose $X_1, X_2, \dots, X_5$ are independent samples from $N(0,1)$ population. Find out C which makes the below follows **t** distribution.

$$Y = \frac{C(X_1 + X_2)}{\sqrt{(X_3^2 + X_4^2 + X_5^2)}}$$

Answer:

1. $X_1, X_2, \dots, X_n$ are i.i.d. random variables with their population. Based on the features, the linear combination of normal distribution still follows normal distribution

$$X_1 + X_2 \sim N(0, 2)$$

Five Theorems about Sampling distributions
（抽样分布的5个定理）

The most important population: $X \sim N(\mu, \sigma^2)$

Question How to use samples $X_1, X_2, \ldots, X_n$ to infer μ, σ^2 ❓

Analysis: The inference of μ, σ^2 is implemented by constructing Statistics

   ①   How to construct "proper" statistics?

   ②   What distribution does the r.v. $g(X_1, X_2, \ldots, X_n)$ follow?

The most important conclusions in statistical inference are:

5 theorems about sampling distributions

All five theorems are based on the normal distribution

Theorem 1: Assume that $X_1, X_2, \ldots, X_n$ are samples from the population $X \sim N(\mu, \sigma^2)$, then

$$\boxed{\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)}$$

Proof: $\because X_1, X_2, \cdots, X_n$ are i.i.d. r.v.s following $N(\mu, \sigma^2)$.

$\therefore$ Based on the feature of normal distribution, the linear combination

$$\bar{X} = \frac{1}{n}(X_1 + X_2 + \cdots + X_n)$$

still follows a normal distribution.

$$\because E(\bar{X}) = \mu, D(\bar{X}) = \frac{\sigma^2}{n}$$

$$\therefore \bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

Theorem 2: If $X_1, X_2, \ldots, X_n$ are samples from the population $X \sim N(\mu, \sigma^2)$, $\bar{X}$ and S^2 are the sample mean and sample variance, then

① $\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$

② $\bar{X}$ and S^2 are independent

Idea of Proof:

$$\because (n-1)S^2 = \sum_{i=1}^{n}(X_i - \bar{X})^2 = \sum_{i=1}^{n}[(X_i - \mu) + (\mu - \bar{X})]^2$$

$$= \sum_{i=1}^{n}(X_i - \mu)^2 - n(\bar{X} - \mu)^2$$

$$\therefore \frac{(n-1)S^2}{\sigma^2} + \left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}\right)^2 = \sum_{i=1}^{n}\left(\frac{X_i - \mu}{\sigma}\right)^2$$

$$\chi^2(n-1) \qquad \chi^2(1) \qquad \chi^2(n)$$

Details of the proof are omitted here.

Theorem 2: If $X_1, X_2, \ldots, X_n$ are a sample from the population $X \sim N(\mu, \sigma^2)$, $\bar{X}$ and S^2 are the sample mean and sample variance, then

① $\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$

② $\bar{X}$ and S^2 are independent

Indication of the Theorem:

$$\because \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

$$\therefore E\left[\frac{(n-1)S^2}{\sigma^2}\right] = n-1, \qquad D\left[\frac{(n-1)S^2}{\sigma^2}\right] = 2(n-1)$$

$$\therefore E(S^2) = \sigma^2, \qquad D(S^2) = \frac{2\sigma^4}{n-1}$$

This indicates that the difference between S^2 and σ^2 is small on average, so we can use S^2 to estimate σ^2.

Theorem 3: If $X_1, X_2, \ldots, X_n$ are a sample from the population $X \sim N(\mu, \sigma^2)$, $\bar{X}$ and S^2 are the sample mean and sample variance, then

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n-1)$$

Proof: Based on Theorems 1 and 2, we have that

$$Y \triangleq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1) \quad, \chi^2 \triangleq \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

Moreover, Y and χ^2 are independent.

Based on the definition of the t-distribution:

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} = \frac{\dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}}}{\sqrt{\dfrac{(n-1)S^2/\sigma^2}{n-1}}} = \frac{Y}{\sqrt{\chi^2/(n-1)}} \sim t(n-1)$$

Two unknown parameters μ, σ

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

Theorem 1

One unknown parameter μ

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n-1)$$

Theorem 3

Theorem 4: Assume that $X_1, X_2, \ldots, X_{n_1}$ are a sample from the population $X \sim N(\mu_1, \sigma_1^2)$. $Y_1, Y_2, \ldots, Y_{n_2}$ are a sample from the population $Y \sim N(\mu_2, \sigma_2^2)$. The two samples are independent, and their sample means and sample variances are $\bar{X}, \bar{Y}, S_1^2, S_2^2$. Then

$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F(n_1 - 1, n_2 - 1)$$

Proof: Based on **Theorem 2**, $\boxed{\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)}$

$$\therefore U = \frac{(n_1 - 1)S_1^2}{\sigma_1^2} \sim \chi^2(n_1 - 1), V = \frac{(n_2 - 1)S_2^2}{\sigma_2^2} \sim \chi^2(n_2 - 1).$$

Since the two samples are independent, so S_1^2 and S_2^2 are independent. According to the definition of the F-distribution:

$$\boxed{F = \frac{U/n_1}{V/n_2}}$$

$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} = \frac{\dfrac{(n_1 - 1)S_1^2}{\sigma_1^2}/(n_1 - 1)}{\dfrac{(n_2 - 1)S_2^2}{\sigma_2^2}/(n_2 - 1)} = \frac{U/(n_1 - 1)}{V/(n_2 - 1)} \sim F(n_1 - 1, n_2 - 1)$$

Theorem 5: Assume that $X_1, X_2, \ldots, X_{n_1}$ are a sample from the population $X \sim N(\mu_1, \sigma^2)$. $Y_1, Y_2, \ldots, Y_{n_2}$ are a sample from the population $Y \sim N(\mu_2, \sigma^2)$. The two samples are independent, and their sample means and sample variances are $\bar{X}, \bar{Y}, S_1^2, S_2^2$. Then

$$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{S_\omega \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} \sim t(n_1 + n_2 - 2)$$

where

$$S_\omega^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}, S_\omega = \sqrt{S_\omega^2}$$

Proof: $\because \bar{X} \sim N\left(\mu_1, \dfrac{\sigma^2}{n_1}\right), \bar{Y} \sim N\left(\mu_2, \dfrac{\sigma^2}{n_2}\right)$ and $\bar{X}, \bar{Y}$ are independent

$\therefore \bar{X} - \bar{Y} \sim N\left(\mu_1 - \mu_2, \dfrac{\sigma^2}{n_1} + \dfrac{\sigma^2}{n_2}\right)$　$\therefore \dfrac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sigma \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} \sim N(0, 1)$

On the other hand, from Theorem 2, we have:

$$\frac{(n_1 - 1)S_1^2}{\sigma^2} \sim \chi^2(n_1 - 1), \frac{(n_2 - 1)S_2^2}{\sigma^2} \sim \chi^2(n_2 - 1)$$

Theorem 5: Assume that $X_1, X_2, \ldots, X_{n_1}$ are a sample from the population $X \sim N(\mu_1, \sigma^2)$. $Y_1, Y_2, \ldots, Y_{n_2}$ are a sample from the population $Y \sim N(\mu_2, \sigma^2)$. The two samples are independent, and their sample means and sample variances are $\bar{X}, \bar{Y}, S_1^2, S_2^2$. Then

$$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{S_\omega \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} \sim t(n_1 + n_2 - 2)$$

where

$$S_\omega^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}, S_\omega = \sqrt{S_\omega^2}$$

Proof: Based on the independency of S_1^2, S_2^2 and the additivity of the χ^2-distribution:

$$\frac{(n_1 - 1)S_1^2}{\sigma^2} + \frac{(n_2 - 1)S_2^2}{\sigma^2} = \frac{(n_1 + n_2 - 2)S_\omega^2}{\sigma^2} \sim \chi^2(n_1 + n_2 - 2)$$

Based on the independency of the samples and the definition of t-distribution:

$$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{S_\omega \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} = \frac{\dfrac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sigma \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}}{\sqrt{\dfrac{(n_1 + n_2 - 2)S_\omega^2}{\sigma^2} \Big/ (n_1 + n_2 - 2)}} \sim t(n_1 + n_2 - 2)$$
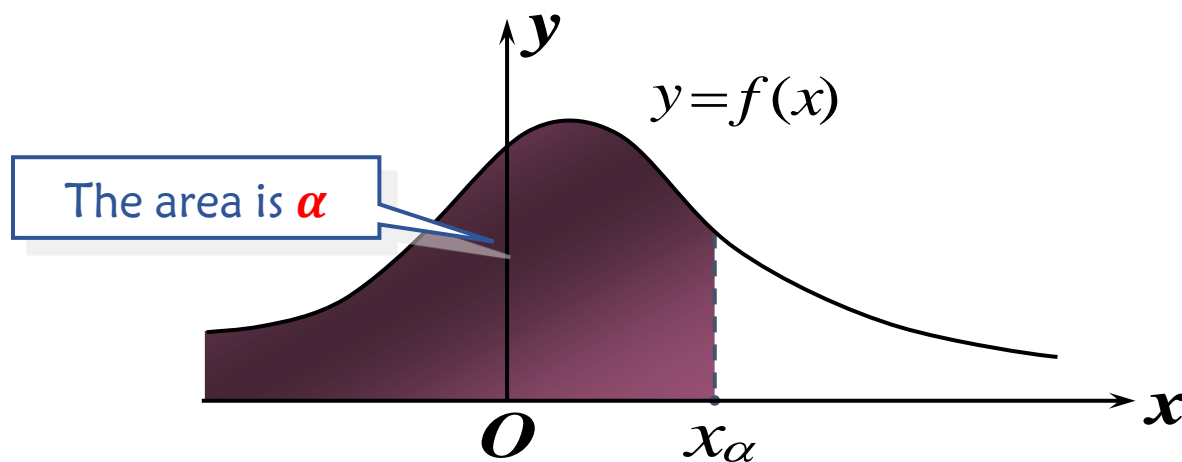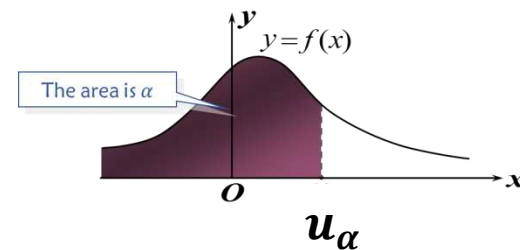
α-quantile (α-分位点)

Assume that X has density function $f(x)$, if for $\forall\, 0 < \alpha < 1$, constant x_α satisfies:

$$P\{X \le x_\alpha\} = \int_{-\infty}^{x_\alpha} f(x)dx = \alpha,$$

then x_α is called the α-quantile of density $f(x)$.

The area is α

$y = f(x)$

O x_α

α-quantile (α-分位点)

Assume that X has density function $f(x)$, if for $\forall\, 0 < \alpha < 1$, constant x_α satisfies:

$$P\{X \le x_\alpha\} = \int_{-\infty}^{x_\alpha} f(x)dx = \alpha,$$

then x_α is called the α-quantile of density $f(x)$.

● The α quantile of $N(0, 1)$ is denoted as u_α
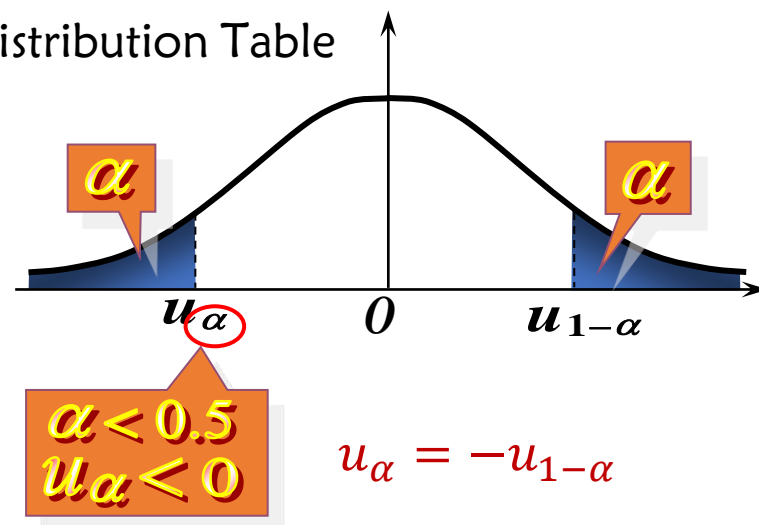
$$\boxed{\Phi(u_\alpha) = P\{X \le u_\alpha\} = \alpha}$$

Example: According to the Standard Normal Distribution Table

$$u_{0.975} = 1.96$$

$$u_{0.95} = \frac{1.64 + 1.65}{2} = 1.645$$

$$u_{0.05} = -u_{0.95} = -1.645$$

$\alpha < 0.5$
$u_\alpha < 0$

$$u_\alpha = -u_{1-\alpha}$$

α-quantile (α-分位点)

Assume that X has density function $f(x)$, if for $\forall\, 0 < \alpha < 1$, constant x_α satisfies:

$$P\{X \le x_\alpha\} = \int_{-\infty}^{x_\alpha} f(x)dx = \alpha,$$

then x_α is called the α-quantile of density $f(x)$.

● The α quantile of $t(n)$ is denoted as $t_\alpha(n)$

Example: According to the t-Distribution Table
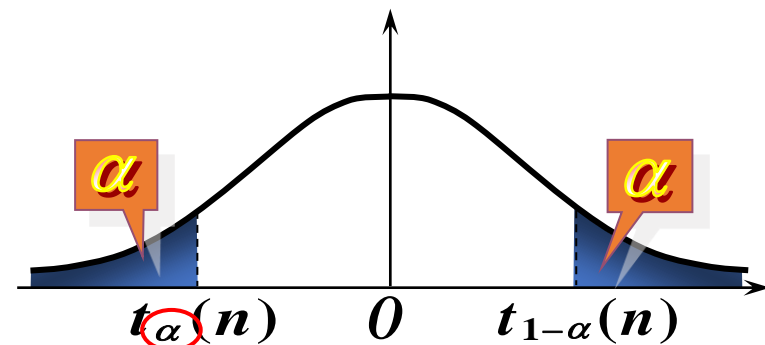
$t_{0.95}(6) = 1.9432$

$t_{0.10}(12) = -t_{0.90}(12) = -1.3562$

$t_{0.95}(55) \approx u_{0.95} = 1.645$

$$f(x) \to \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad (n \to \infty)$$

$$\therefore t_\alpha(n) \approx u_\alpha \quad (n > 45)$$

α

α

$t_\alpha(n)$ $\quad 0 \quad$ $t_{1-\alpha}(n)$

$\alpha < 0.5$
$t_\alpha(n) < 0$

$$t_\alpha(n) = -t_{1-\alpha}(n)$$

α-quantile (α-分位点)

Assume that X has density function $f(x)$, if for $\forall\ 0 < \alpha < 1$, constant x_α satisfies:

$$P\{X \le x_\alpha\} = \int_{-\infty}^{x_\alpha} f(x)dx = \alpha$$

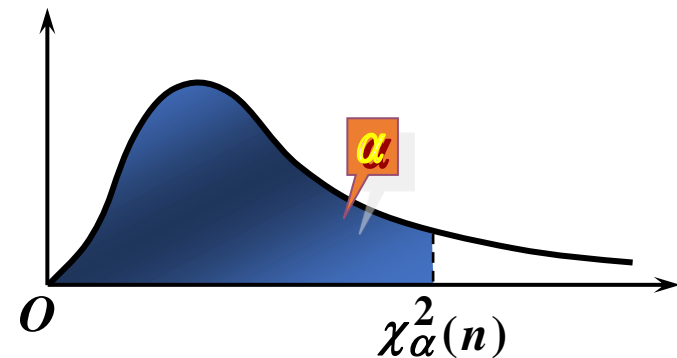Then x_α is called the α-quantile of density $f(x)$.

● The α quantile of $\chi^2(n)$ is denoted as $\chi_\alpha^2(n)$

Example: According to the χ^2-Distribution Table

$\chi_{0.95}^2(12) = 21.026$

$\chi_{0.05}^2(25) = 14.611$

$\chi_{0.95}^2(50) \approx \dfrac{1}{2}\left(1.645 + \sqrt{99}\right)^2 = 67.211$

Fisher showed that when n is large enough:

$$\chi_\alpha^2(n) \approx \frac{1}{2}\left(u_\alpha + \sqrt{2n-1}\right)^2$$

α-quantile (α-分位点)

Assume that X has density function $f(x)$, if for $\forall\, 0 < \alpha < 1$, constant x_α satisfies:

$$P\{X \le x_\alpha\} = \int_{-\infty}^{x_\alpha} f(x)dx = \alpha$$

Then x_α is called the α-quantile of density $f(x)$.

⬤ The α quantile of $F(n_1, n_2)$ is denoted as $F_\alpha(n_1, n_2)$
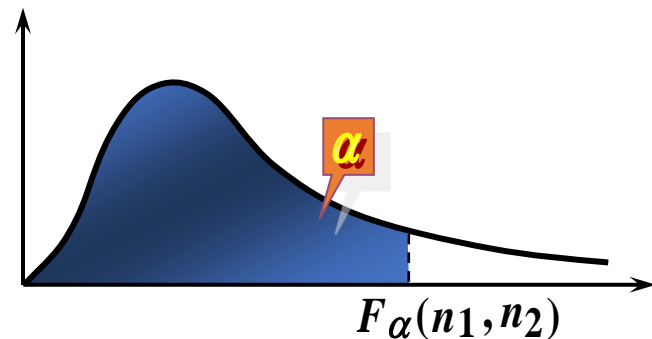
Example: According to the F-Distribution Table

$$F_{0.95}(5, 12) = 3.11$$

$$F_{0.90}(2, 25) = 2.53$$

$$F_{0.05}(6, 10) = \frac{1}{F_{0.95}(10,6)} = \frac{1}{4.06} = 0.246$$

If $F \sim F(n_1, n_2)$, then $F^{-1} \sim F(n_2, n_1)$. Thus

$$F_\alpha(n_1, n_2) = \frac{1}{F_{1-\alpha}(n_2, n_1)}$$

$F_\alpha(n_1, n_2)$

Triple-reverse formula
"三反" 公式

Homework

P198: questions 3、6、8

Supplementary questions:

1. Suppose that the population distribution is $N(240, 20^2)$. Draw a sample of size 36 and another of size 49 independently from the population. Compute the probability that the absolute value of the difference between the two sample means does not exceed 10.

2. Suppose that $X_1, X_2, \ldots, X_{10}$ are a sample from population $X \sim N(0, 0.3^2)$. Compute the constant C such that $P\left(\sum_{i=1}^{10} X_i^2 \leq C\right) = 0.95$.

3. Suppose that $X_1, X_2, \ldots, X_n$ are an independent sample from population $X \sim N(0, \sigma^2)$. Find the constant n such that $P(|\overline{X} - \mu| < 1) \geq 0.95$.

4. Suppose that X_1, X_2 are a sample from population $X \sim N(0, \sigma^2)$.

   a) Find the distribution of $\frac{(X_1 - X_2)^2}{(X_1 + X_2)^2}$

   b) Find the constant k such that $P\left\{\frac{(X_1 + X_2)^2}{(X_1 + X_2)^2 + (X_1 - X_2)^2} > k\right\} = 0.1$

5. Suppose that $X_1, X_2, \ldots, X_n, X_{n+1}$ are a sample from population $X \sim N(\mu, \sigma^2))$, $\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$, $S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X}_n)^2$. Compute the constant c such that $t_c = c \frac{X_{n+1} - \overline{X}_n}{S_n}$ follows the t-distribution and find the degree of freedom.

谢谢大家