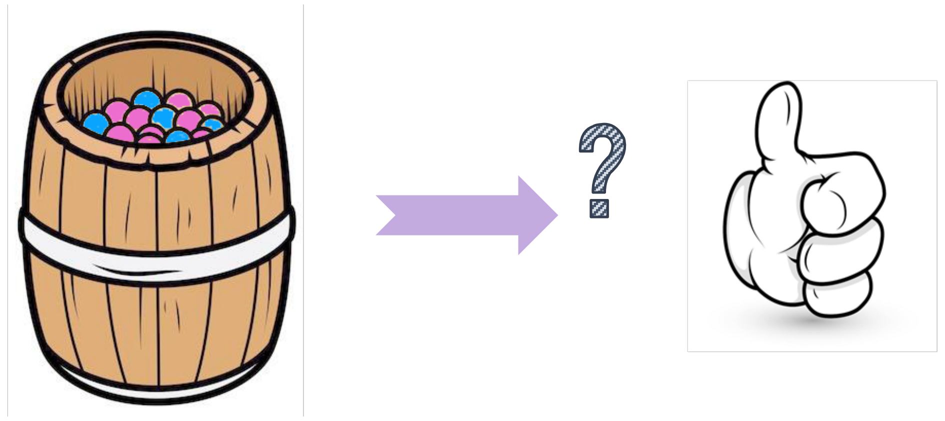
第6章 数理统计的基本概念

概率论与数理统计课程组

概率论与数理统计是研究随机现象统计规律性的一门学科。

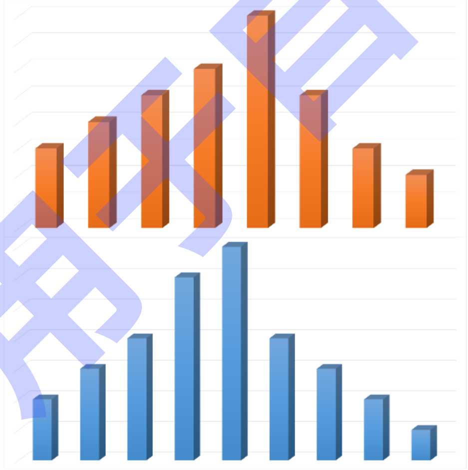
概率论



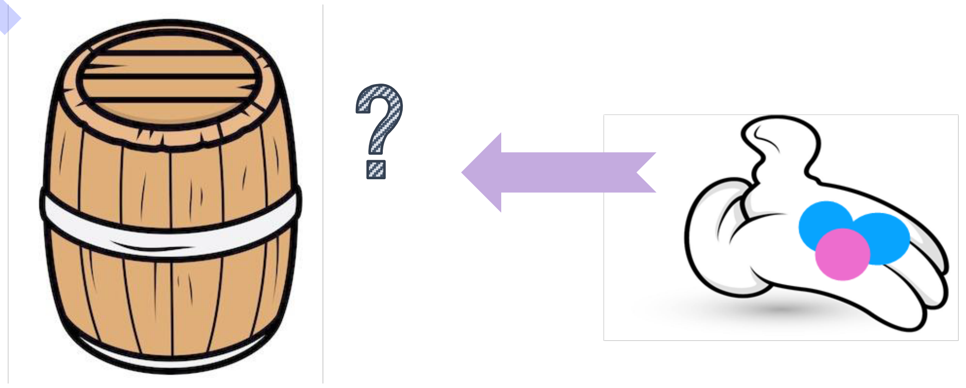
已知桶内球颜色比例，猜猜手中球的颜色？

F B Probability

Given model, predict data



数理统计



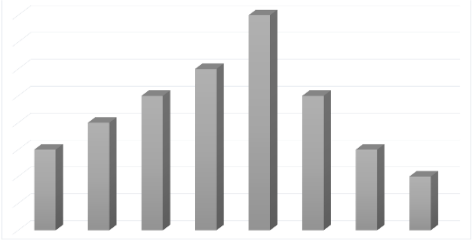
不断统计摸出球的颜色，推断：

* 桶内球颜色的比例（参数估计）

*是否可认为红蓝比例为1:2？（假设检验）

? Statistics

Given data, predict model



概率论、数理统计都是研究随机现象的统计规律性的数学分支，但两者研究角度不同

概率论：从随机变量X的已知分布出发，研究X的种种性质、规律、数字特征等

数理统计：随机变量X的分布未知或分布中含有未知参数，观察它的取值（采集数据），通过分析数据来推断X服从什么分布或确定未知参数

数理统计
- 收集、整理数据
- 统计推断

概率论

5 大数定律及中心极限定理

数理统计

1 概率论的基本概念
2 随机变量及其分布
3 多维随机变量及其分布
4 随机变量的数字特征

随机事件
概率
随机变量
分布函数
数字特征

大数定律
中心极限定理

样本
统计量
抽样分布
参数估计
假设检验

6 数理统计的基本概念
7 参数估计
8 假设检验

X

$g(X_1, X_2, \dots X_n)$

"随机"

"数据"

CHAPTER 6

数理统计的基本概念

§6.1 基本概念

§6.2 抽样分布

是后续两章理论和方法的基础

CHAPTER 6

数理统计的基本概念

6.1 基本概念

当研究考察对象的某项数量指标时，可针对这一指标进行试验（或观察），引入以下定义

❖ **总体** 试验**所有**可能的观察值（或数量指标）的**全体**

❖ **个体** 试验**每一个**可能的观察值

例

1. 检验灯泡厂生产的灯泡寿命

总体：全体灯泡寿命数值 个体：每个灯泡寿命数值

2. 调查某校男生的身高情况

总体：全校所有男生的身高数值构成的全体 个体：每个男生身高数值

❖ **总体的容量** 总体中所包含的个体的数目 **有限总体** **无限总体**

当有限总体包含的个体的总数很大时，可近似地将它看成是无限总体.

一般地，我们所研究的总体，即研究对象的某项数量指标X，其取值在客观上有一定的分布，X是一个随机变量.（总体是随机变量）

例如：研究某批灯泡的寿命时，关心的数量指标是其寿命，而寿命X可用某一概率分布$F(X)$来刻画，那么此总体 就可以用随机变量X或其分布函数$F(x)$表示.

随机变量X的分布函数和数字特征就称为总体的分布函数和数字特征. 今后将不区分总体与相应的随机变量 X

在实际中，总体的分布通常是未知的，或只知道它具有某种形式而其中包含未知参数. 那么如何对总体进行推断呢?

简单随机样本

抽样

在数理统计中，人们都是通过从总体中抽取一部分个体，根据获得的数据来推断总体的某些特征，这一抽取过程称为“**抽样**”，所抽取的部分个体称为“**样本**”. 样本中所包含的个体数目称为**样本容量**.

❖ 简单随机样本

设X是具有分布函数F的随机变量，若$X_1, X_2, \ldots, X_n$是相互独立、且与X具有相同分布函数F的随机变量，则称$X_1, X_2, \ldots, X_n$是一个来自总体X的容量为 n 的**简单随机样本，简称样本.**

❖ 样本值

一个样本$X_1, X_2, \ldots, X_n$的观察值$x_1, x_2, \ldots, x_n$，称为**样本值**

简单随机抽样的特点：

1. 代表性：$X_i\,(i=1,2\cdots,n)$ 与总体X有相同的分布

2. 独立性：$X_1, X_2, \ldots, X_n$ 是相互独立的随机变量

联合分布函数

设总体X的分布函数是$F(x)$，$X_1, X_2, \ldots,$ X_n是来自总体X的一个样本，则n维随机变量$(X_1, X_2, \ldots, X_n)$的联合分布函数为

$$F^*(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} F(x_i)$$

❖ **离散型**

若总体X的分布律为 $P(X=x)=p(x)$，$X_1, X_2, \ldots, X_n$是来自总体X的一个样本，则 n维随机向量$(X_1, X_2, \ldots, X_n)$的联合分布律为

$$p^*(X_1=x_1, X_2=x_2, \ldots, X_n=x_n) = \prod_{i=1}^{n} p(x_i)$$

❖ **连续型**

若总体X的概率密度为$f(x)$，$X_1, X_2, \ldots,$ X_n是来自总体X的一个样本，则 n维随机向量$(X_1,X_2, \ldots, X_n)$的联合概率密度为

$$f^*(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} f(x_i)$$

例 写出下列样本的联合概率函数

(1) $X_1, X_2, \dots, X_n$ 是来自总体 $N(\mu, \sigma^2)$ 的样本

(2) $X_1, X_2, \dots, X_n$ 是来自总体 $B(1, p)$ 的样本

解 (1) $X_1, X_2, \dots, X_n$ 的联合概率密度为

$$f^*(x_1,x_2,\cdots,x_n)=\prod_{i=1}^{n} f(x_i)=\prod_{i=1}^{n}\frac{1}{2\pi\sigma}\exp\left\{-\frac{(x_i-\mu)^2}{2\sigma^2}\right\}=\left(\frac{1}{2\pi\sigma}\right)^n\exp\left\{-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}\right\}$$

(2) $X_1, X_2, \dots, X_n$ 的联合分布律为

$$p^*(X_1=x_1,X_2=x_2,\cdots,X_n=x_n)=\prod_{i=1}^{n}P(X_i=x_i)=\prod_{i=1}^{n}p^{x_i}(1-p)^{1-x_i}$$

$$=p^{\sum_{i=1}^{n}x_i}(1-p)^{n-\sum_{i=1}^{n}x_i}$$

其中$x_1, x_2, \dots, x_n$在集合$\{0, 1\}$中取值

问题：用样本观察值推断总体，其结论可靠吗？

解决：根据抽样得到的样本观察值构造一个函数——样本分布函数（或称**经验分布函数**），再证明当n 很大时，经验分布函数近似于总体的分布函数.

经验分布函数

定义

设$X_1, X_2, \ldots, X_n$是来自总体X的一个样本，$x_1, x_2, \ldots, x_n$是样本$X_1, X_2, \ldots, X_n$的一组样本值，将其从小到大排列，并重新编号为$x_{(1)} \le x_{(2)} \le \ldots \le x_{(n)}$，则称函数

$$F_n(x)=\frac{x_1, x_2, \ldots, x_n\text{中小于等于}x\text{的样本值的个数}}{n}=\begin{cases}0, & x<x_{(1)}\\ \dfrac{k}{n}, & x_{(k)}<x<x_{(k+1)}\\ 1, & x\ge x_{(n)}\end{cases}$$

为总体X的经验分布函数

定理

对于任意实数x，当 $n\to\infty$ 时， $F_n(x)$以概率1收敛于总体X的分布函数$F(x)$，即

$$P\left(\lim_{n\to\infty}\sup_{-\infty<x<\infty}\left|F_n(x)-F(x)\right|=0\right)=1$$

对于任意实数x当n充分大时，经验分布函数的任意观察值$F_n(x)$ 与总体分布函数$F(x)$只有微小的差别，从而在实际中可当做$F(x)$使用.

例 设总体F具有一个样本值1,2,3，则经验分布函数$F_3(x)$的观察值为

$$F_3(x)=\begin{cases}0 & x<1\\ \dfrac{1}{3} & 1\le x<2\\ \dfrac{2}{3} & 2\le x<3\\ 1 & x\ge 3\end{cases}$$

统计量

样本是进行统计推断的依据. 但在实际应用中，通常需要针对具体问题对样本值进行整理和加工，构造出适当的样本的函数(即统计量)，利用这些函数来进行统计推断，揭示总体的统计特性.

定义

设$X_1, X_2, \dots, X_n$是来自总体X的一个样本，$g(X_1, X_2, \dots, X_n)$是$X_1, X_2, \dots, X_n$的连续函数，若g不含总体X的任何未知参数，则称$g(X_1, X_2, \dots, X_n)$是总体X的一个统计量，其观察值为$g(x_1, x_2, \dots, x_n)$称为统计量$g(X_1, X_2, \dots, X_n)$的样本值 .

注　统计量是不含任何未知参数的样本的函数，它完全依赖于样本, 故而也是随机变量， 有一定的分布，这个分布叫做统计量的抽样分布(下节学习).

例 设在总体 $N(\mu, \sigma^2)$中抽取样本(X_1, X_2, X_3)，其中 μ 已知， σ^2未知，指出下列哪些是统计量，哪些不是统计量

(i) $X_1 + X_2 + X_3$ (ii) $X_2 + 2\mu$ (iii) $\max(X_1, X_2, X_3)$

(iv) $\dfrac{1}{\sigma^2}\sum_{i=1}^{3} X_i^2$ (v) $|X_3 - X_1|$

解 第 (iv) 个不是，因为含有未知参数

常用统计量

它反映了总体均值的信息

❖ **样本均值** $\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$ 其样本值 $\overline{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$

❖ **样本方差** $S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2=\frac{1}{n-1}\left(\sum_{i=1}^{n}X_i^2-n\overline{X}^2\right)$

$E(S^2)=\sigma^2$

其样本值 $s^2=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x})^2=\frac{1}{n-1}\left(\sum_{i=1}^{n}x_i^2-n\overline{x}^2\right)$

❖ **样本标准差** $S=\sqrt{S^2}=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2}$

其样本值 $s=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i-\overline{x}\right)^2}$

❖ 样本k阶（原点）矩 $A_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k,\ k=1,2,\cdots$

其样本值 $a_k=\frac{1}{n}\sum_{i=1}^{n}x_i^k,\ k=1,2,\cdots$

❖ 样本k阶中心矩 $B_k=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^k,\ k=1,2,\cdots$

其样本值$b_k=\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^k,\ k=1,2,\cdots$

由大数定律可以得到下述结论：

定理 若总体X的 k 阶矩 $E(X^k)$ 存在，则当$n\to\infty$时，

$$A_k \xrightarrow{P} E\left(X^k\right) \qquad B_k \xrightarrow{P} E\left[\left(X-\mu\right)^k\right]$$

记 $E(X^k)=\mu_k$ ， 由第五章关于以概率收敛的序列的性质知

$$g(A_1, A_2, \dots, A_k) \xrightarrow{P} g(\mu_1, \mu_2, \dots, \mu_k)$$

其中g是连续函数

以上结论是下一章所要介绍的矩估计法的理论依据.

定理

设总体X均值为μ，方差为σ^2（不管服从什么分布），$X_1, X_2, \dots, X_n$是来自总体X的一个样本，$\overline{X}$ 和 S^2分别 是样本均值和样本方差，则有

$$E(\overline{X}) = \mu \qquad D(\overline{X}) = \sigma^2/n \qquad E(S^2) = \sigma^2$$

证明

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$E(\overline{X}) = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}E\left[\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n} E(X_i) \ = \mu$$

$$D(\overline{X}) = D\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n^2}D\left[\sum_{i=1}^{n} X_i\right]$$

$$= \frac{1}{n^2}\sum_{i=1}^{n} D(X_i) \ = \sigma^2/n$$

$$E(S^2)=E\left[\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2\right]$$

$$=\frac{1}{n-1}E\left[\sum_{i=1}^{n}(X_i^2-2\overline{X}X_i+\overline{X}^2)\right]$$

$$=\frac{1}{n-1}E\left[\sum_{i=1}^{n}X_i^2-2n\overline{X}\cdot\frac{1}{n}\sum_{i=1}^{n}X_i+n\overline{X}^2\right]$$

$$=E\left[\frac{1}{n-1}\left(\sum_{i=1}^{n}X_i^2-n\overline{X}^2\right)\right]$$

$$=\frac{1}{n-1}\left[\sum_{i=1}^{n}E(X_i^2)-nE(\overline{X}^2)\right]$$

$$=\frac{1}{n-1}\left[n(\sigma^2+\mu^2)-n(\sigma^2/n+\mu^2)\right]=\sigma^2$$

本节回顾

□ 简单随机样本

设X是具有分布函数F的随机变量，若$X_1, X_2, \ldots, X_n$是相互独立、且与X具有相同分布函数F的随机变量，则称$X_1, X_2, \ldots, X_n$是一个来自总体X的容量为 n 的**简单随机样本，简称样本.**

□ 常用统计量

❖ 样本均值 $\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$

❖ 样本方差 $S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2=\frac{1}{n-1}\left(\sum_{i=1}^{n}X_i^2-n\overline{X}^2\right)$

❖ 样本标准差 $S=\sqrt{S^2}=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2}$

❖ 样本k阶（原点）矩 $A_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k,\ k=1, 2, \cdots$

❖ 样本k阶中心矩 $B_k=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^k,\ k=1, 2, \cdots$

CHAPTER 6

数理统计的基本概念

§6.1 基本概念

§6.2 抽样分布

6.2 抽样分布

统计量是随机变量，它的分布称为抽样分布. 研究统计量的性质和评价一个统计推断的优良性，完全取决于其抽样分布的性质.下边介绍四大基础分布

常用统计量的分布（统计学四大分布）

1. 标准正态分布 $X \sim N(0, 1)$

$X \sim N(0, 1)$分布的概率密度为 $\varphi(x) = \dfrac{1}{\sqrt{2\pi}} e^{-x^2/2}$

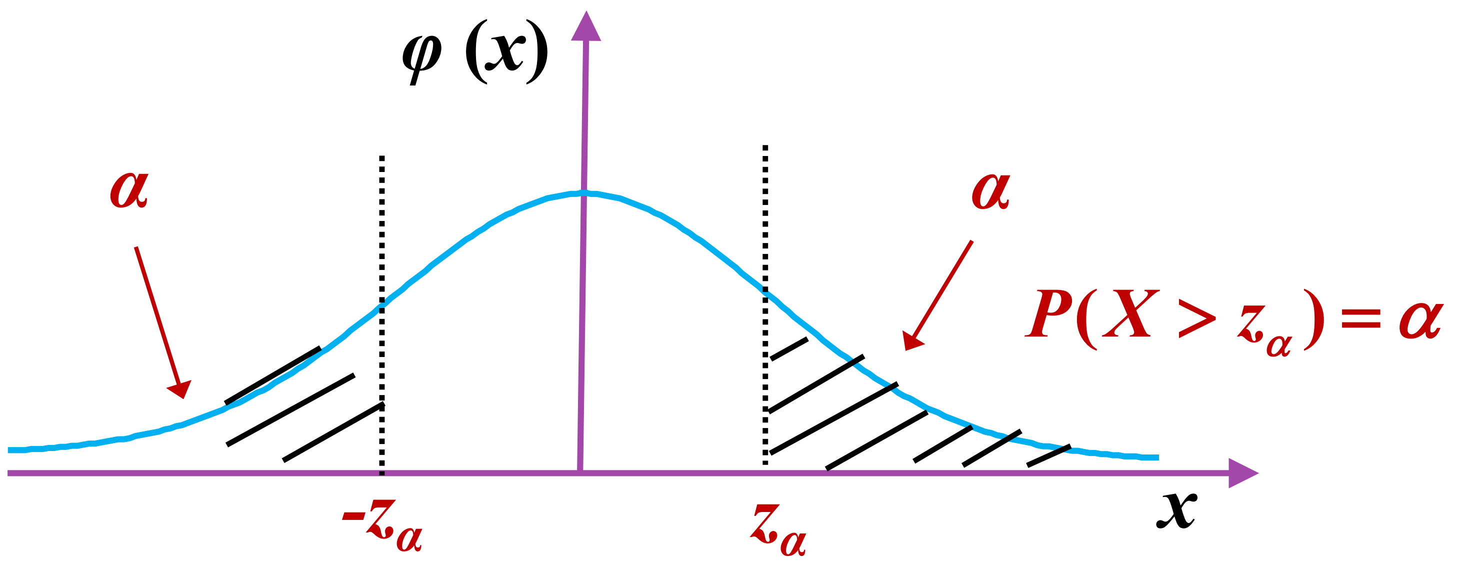
分布函数为 $\Phi(x) = \dfrac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt$

$N(0, 1)$分布的上α分位点

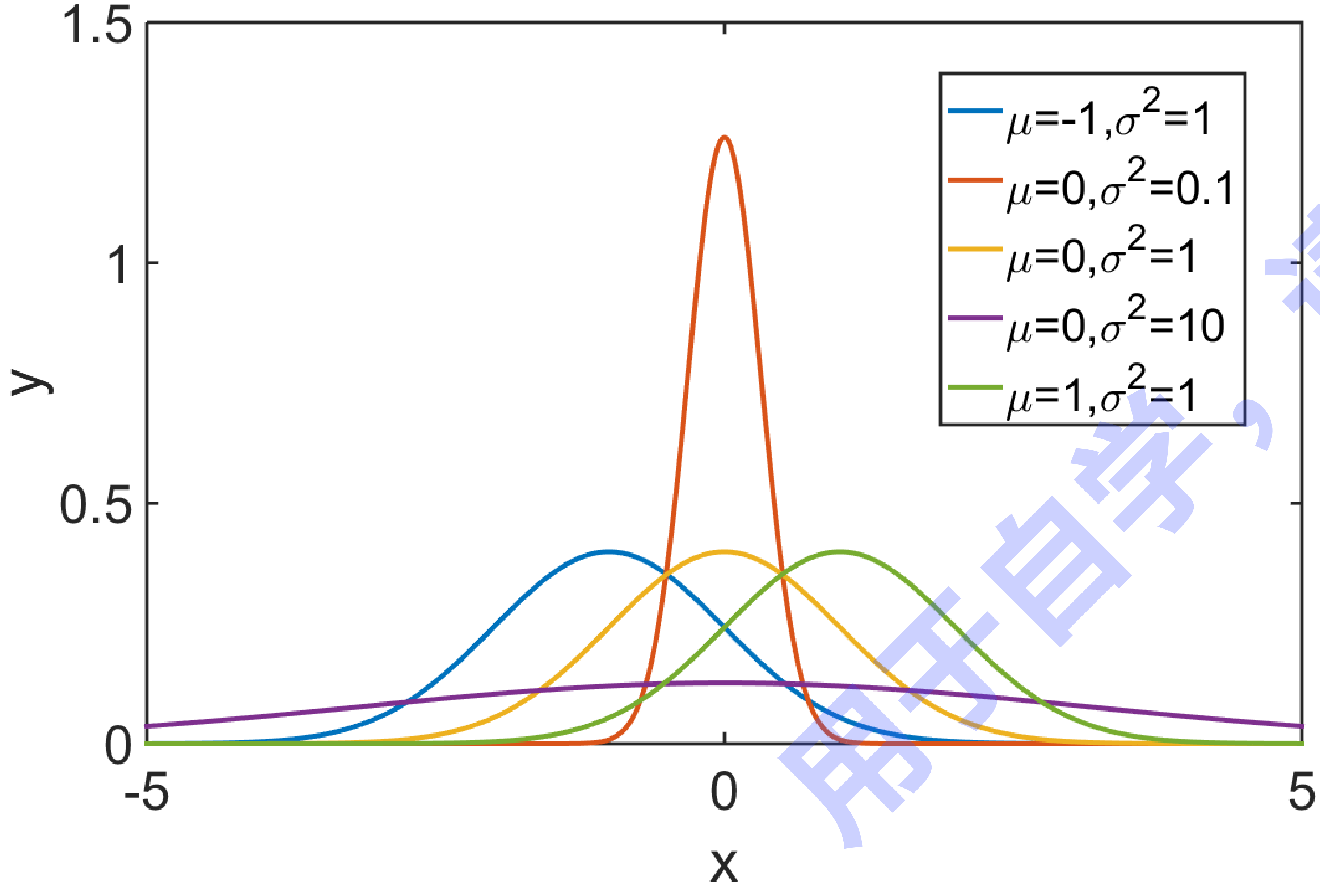
对于正数α($0< \alpha<1$)，满足 $P(X > z_\alpha) = \alpha$

的点z_α称是$N(0, 1)$分布的上α分位点

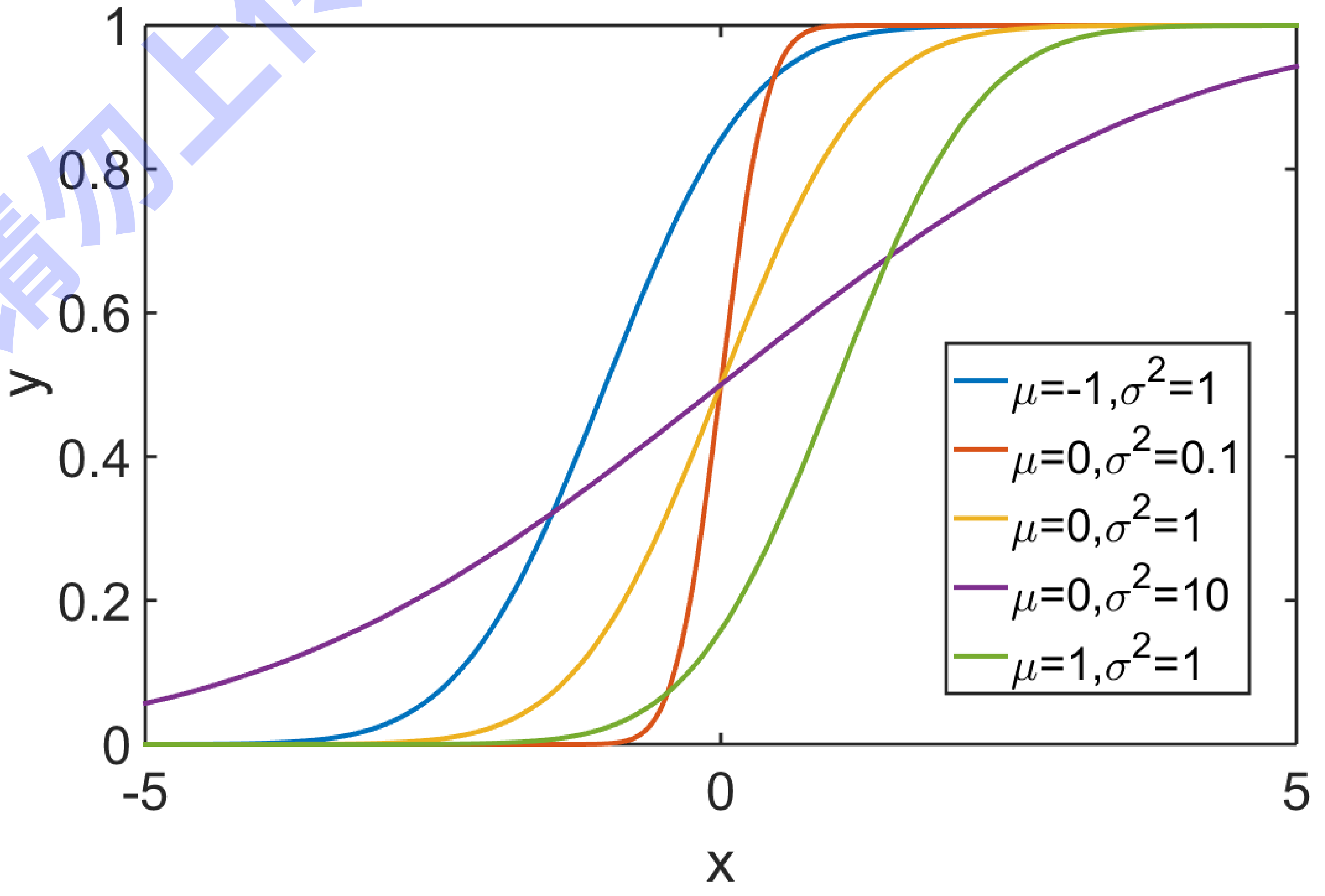
由标准正态分布的对称性可知 $z_{1-\alpha} = -z_\alpha$



正态分布的概率密度



正态分布的分布函数



2. χ^2分布$\chi^2 \sim \chi^2(n)$

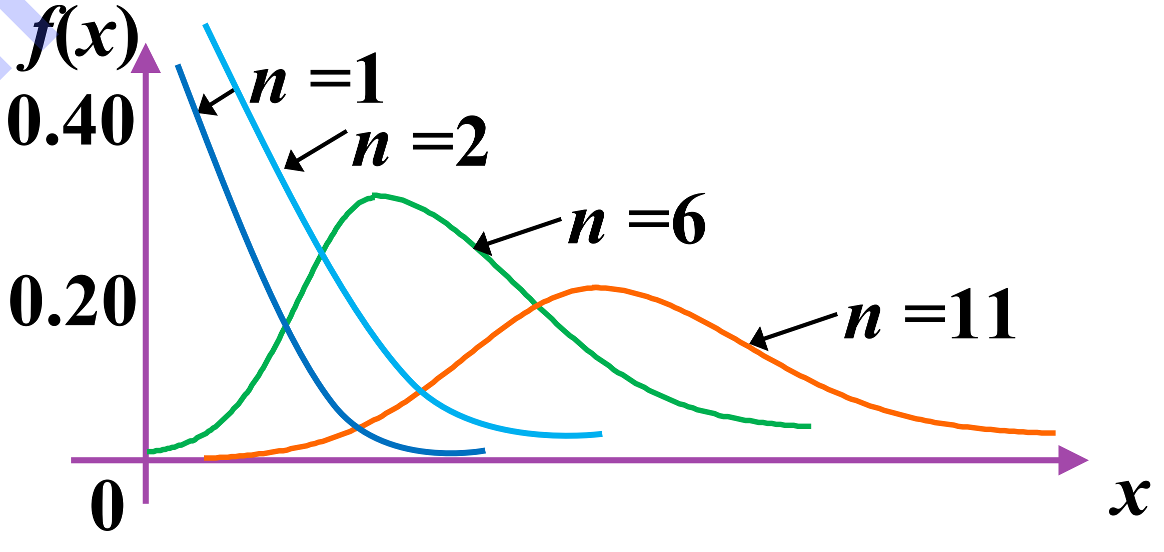
定义

设$X_1, X_2, \ldots, X_n$是来自总体$N(0,1)$的样本，则称统计量

$$\chi^2 = X_1^2 + X_2^2 + \ldots + X_n^2$$

服从自由度为n的χ^2分布，记为$\chi^2 \sim \chi^2(n)$

自由度是指右端包含的**独立变量**个数



$\chi^2 \sim \chi^2(n)$分布的概率密度为

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

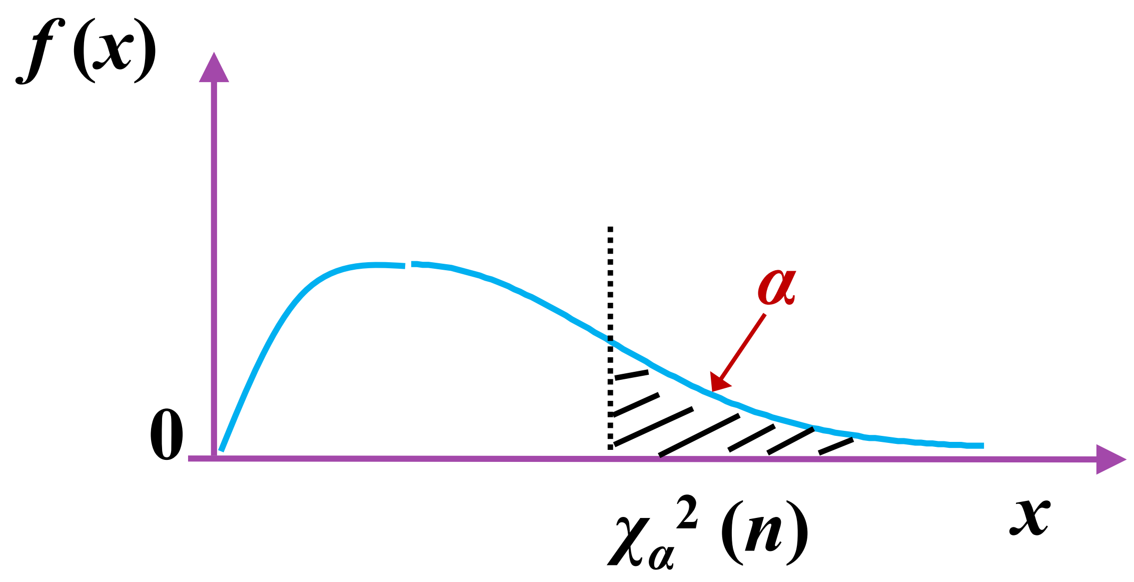
$$\Gamma(\alpha) = \int_0^{+\infty} x^{\alpha-1} e^{-x} dx$$

$$\chi^2 = \sum_{i=1}^{n} X_i^2 \sim \Gamma(\frac{n}{2}, 2)$$

χ^2 分布的上α分位点

对于正数α($0<\alpha<1$)，满足 $P\left(\chi^2>\chi_\alpha^2(n)\right)=\int_{\chi_\alpha^2(n)}^{+\infty} f(x)dx=\alpha$

的点 $\chi_\alpha^2(n)$ 称为$\chi^2(n)$分布的上α分位点



$\chi^2(n)$的上α分位点可以查表获得，当n充分大，如$n>40$时

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

其中z_α称是$N(0,1)$分布的上α分位点

χ^2 分布的可加性

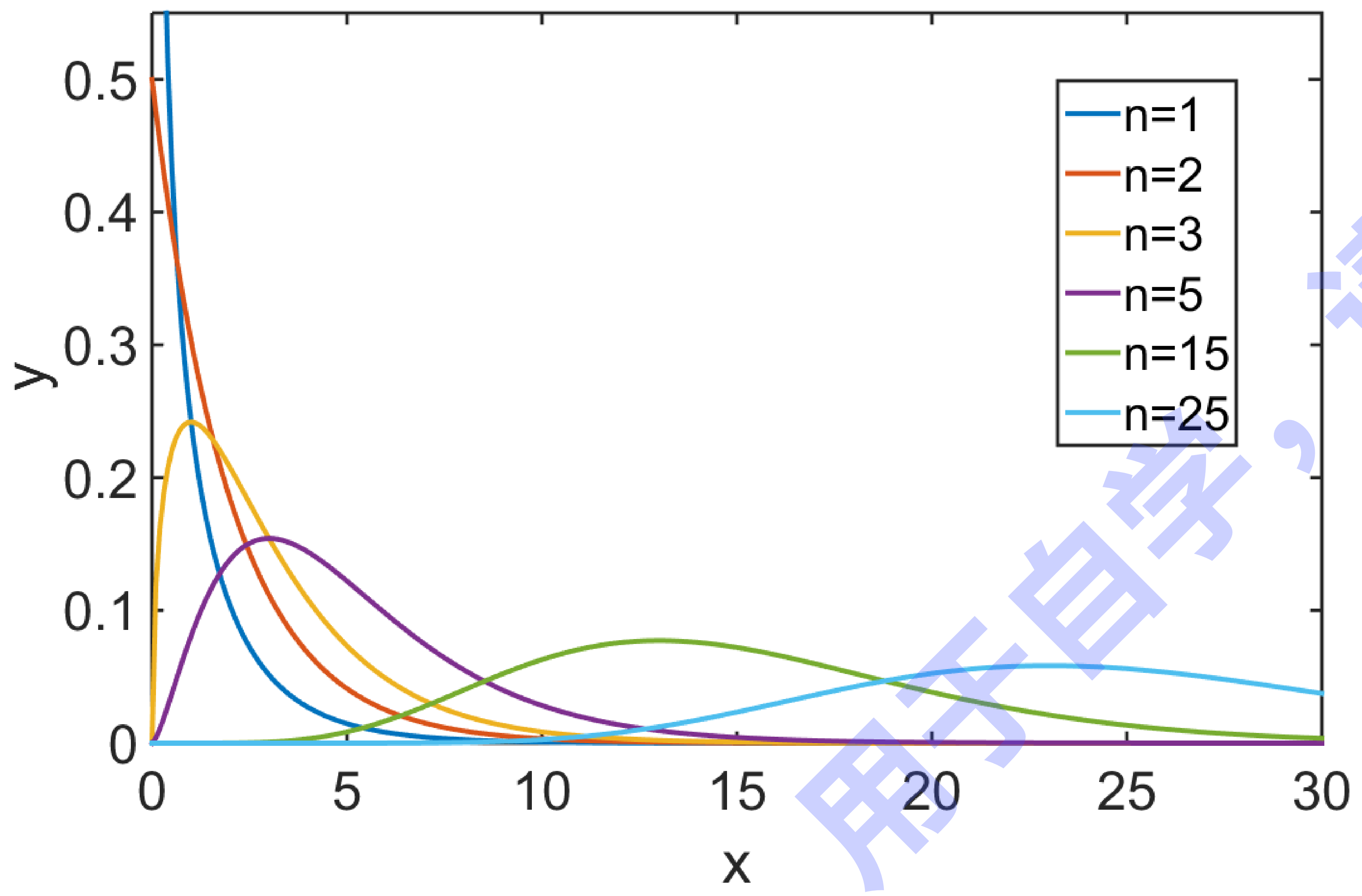
可推广至有限个的情形

设$\chi_1^2\sim\chi^2(n_1)$, $\chi_2^2\sim\chi^2(n_2)$, 且χ_1^2, χ_2^2相互独立，则有 $\chi_1^2+\chi_2^2\sim\chi^2(n_1+n_2)$

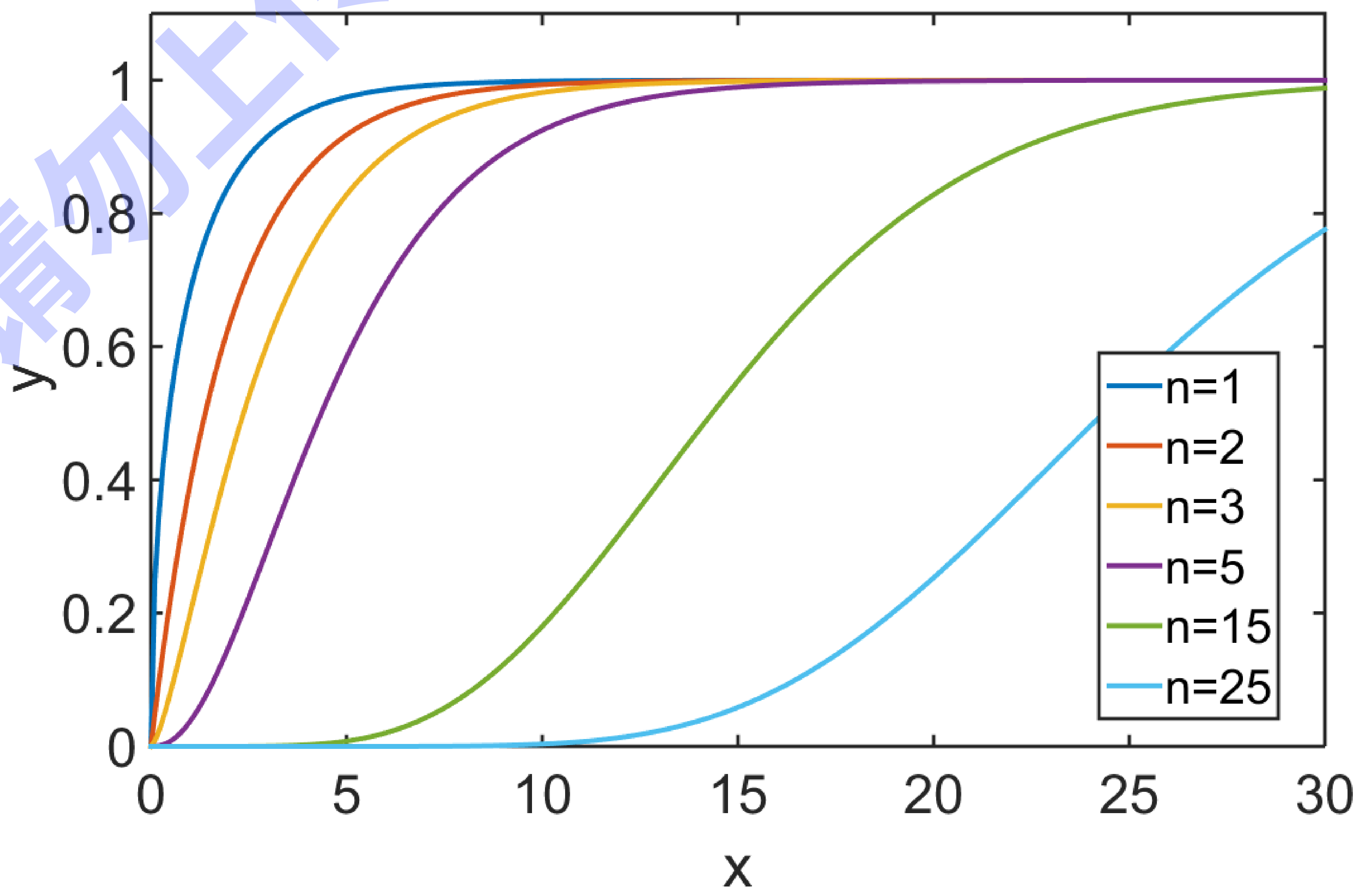
χ^2 分布的数学期望和方差

若$\chi^2\sim\chi^2(n)$，则有 $E(\chi^2)=n$, $D(\chi^2)=2n$

χ^2分布的概率密度



χ^2分布的分布函数



3. t 分布 $T \sim t(n)$

定义 又称学生氏分布

设 $X \sim N(0, 1)$, $Y \sim \chi^2(n)$，且 X, Y 相互独立，则称随机变量

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

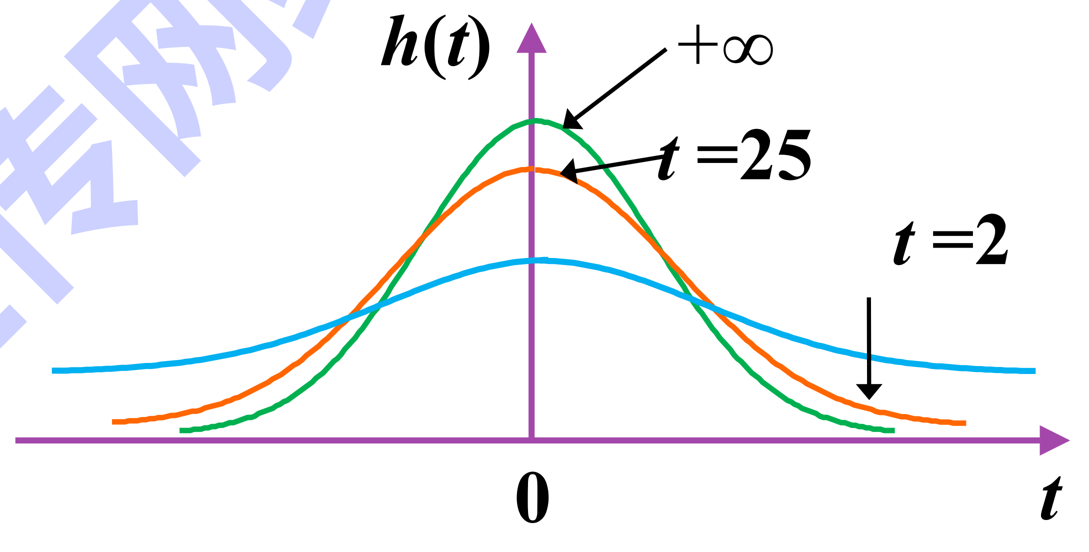
服从自由度为 n 的 t 分布，记为 $t \sim t(n)$

$t(n)$ 分布的概率密度为

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

较小 n 时， t 分布与 $N(0, 1)$ 分布相差较大；

但 n 充分大时， $\lim\limits_{n\to\infty} h(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$

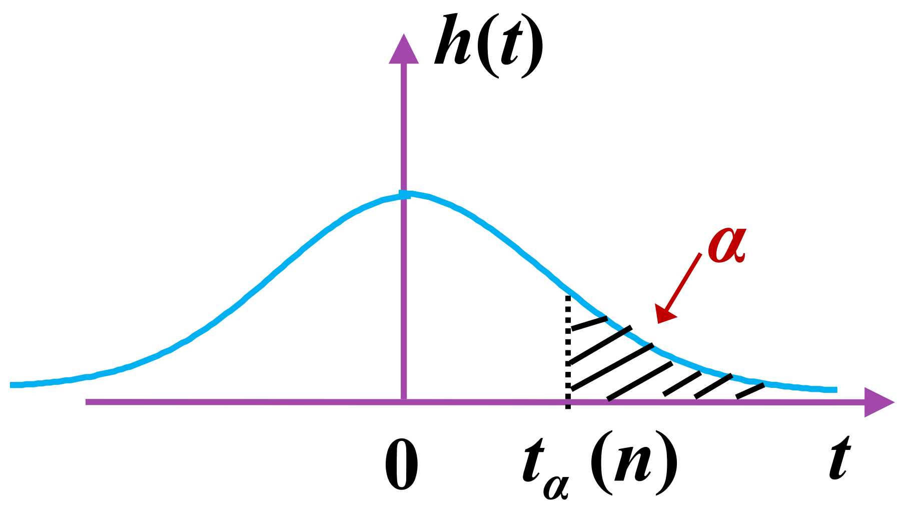


t 分布的上 α 分位点

对于正数 α，$0<\alpha<1$，满足

$$P(t > t_\alpha(n)) = \int_{t_\alpha(n)}^{+\infty} h(t)dt = \alpha$$

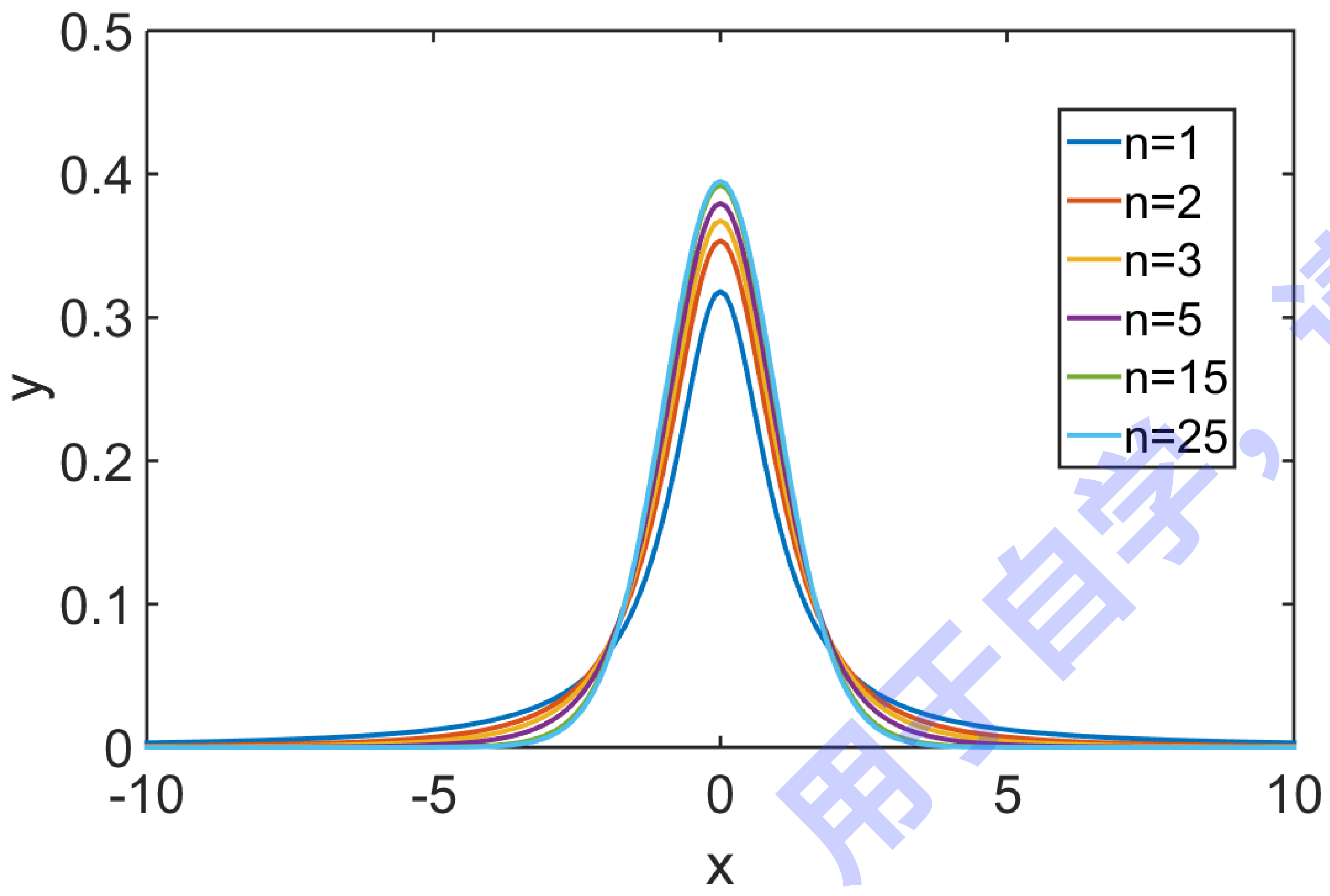
的点称是 $t(n)$ 的上 α 分位点



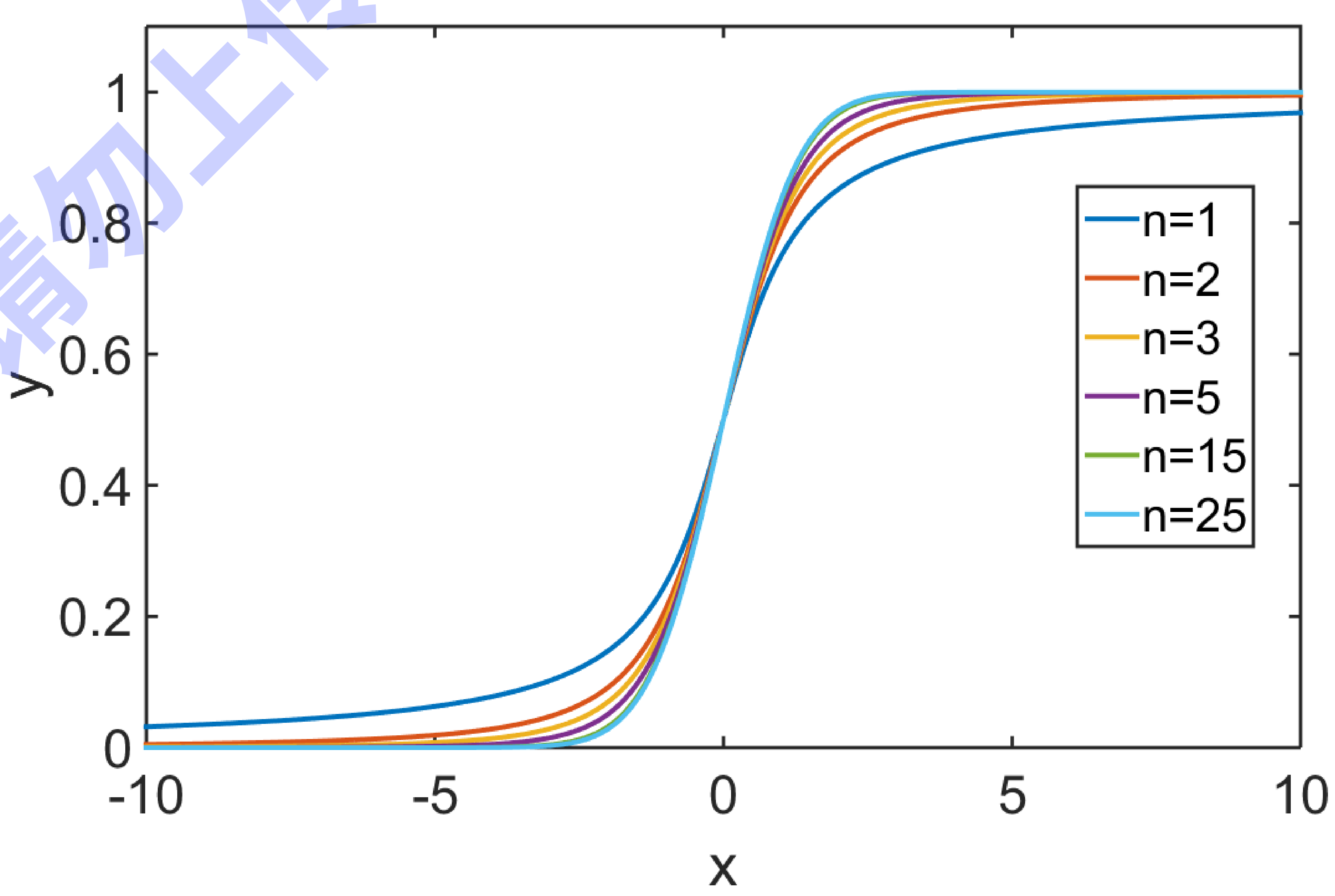
由对称性 $t_{1-\alpha}(n) = -t_\alpha(n)$

t 分布的上 α 分位点可查表获得，当 n 充分大，如 $n>45$ 时，可采用标准正态近似 $t_\alpha(n) \approx z_\alpha$

t 分布的概率密度



t 分布的分布函数



4. F 分布$F\sim F(n_1, n_2)$

定义

设$U\sim\chi^2(n_1)$, $V\sim\chi^2(n_2)$，且U, V相互独立，则称随机变量

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

服从自由度为(n_1, n_2)的F分布，记为$F\sim F(n_1, n_2)$

n_1为第一自由度，n_2为第二自由度

$F(n_1, n_2)$分布的概率密度为

$$f(y)=\begin{cases}\dfrac{\Gamma[(n_1+n_2)/2](n_1/n_2)^{n_1/2}y^{(n_1/2)-1}}{\Gamma(n_1/2)\Gamma(n_2/2)[1+(n_1y/n_2)]^{(n_1+n_2)/2}}, & y>0\\ 0, & \text{其他}\end{cases}$$

F分布的性质

若 $F\sim F(n_1,n_2)$, 则 $\dfrac{1}{F}\sim F(n_2,n_1)$

F分布的上分位点

对于正数α，$0<\alpha<1$，满足

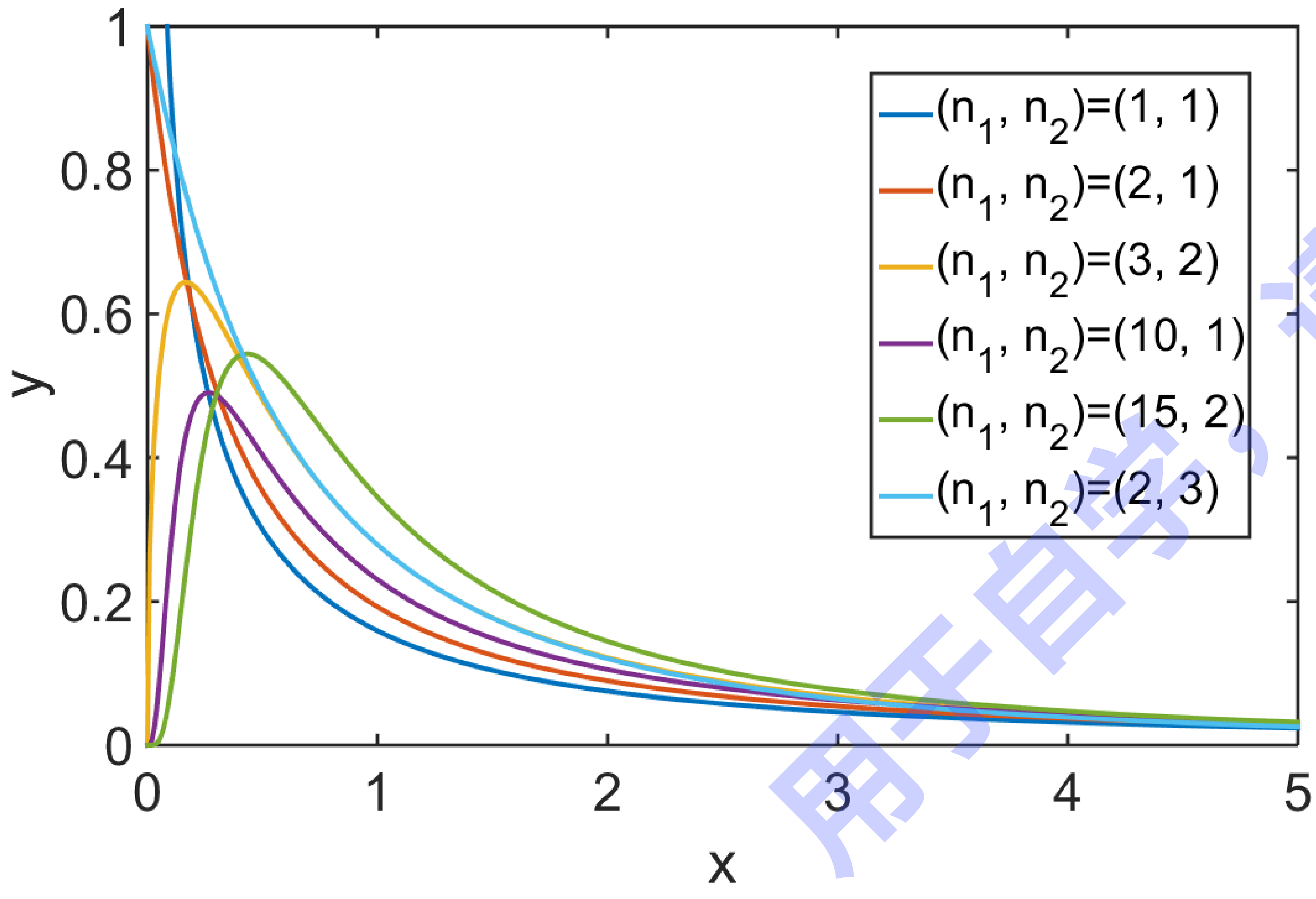
$$P\left(F>F_\alpha(n_1,n_2)\right)=\int_{F_\alpha(n_1,n_2)}^{+\infty}f(y)dy=\alpha$$

的点$F_\alpha(n_1,n_2)$称为$F(n_1, n_2)$的上α分位点

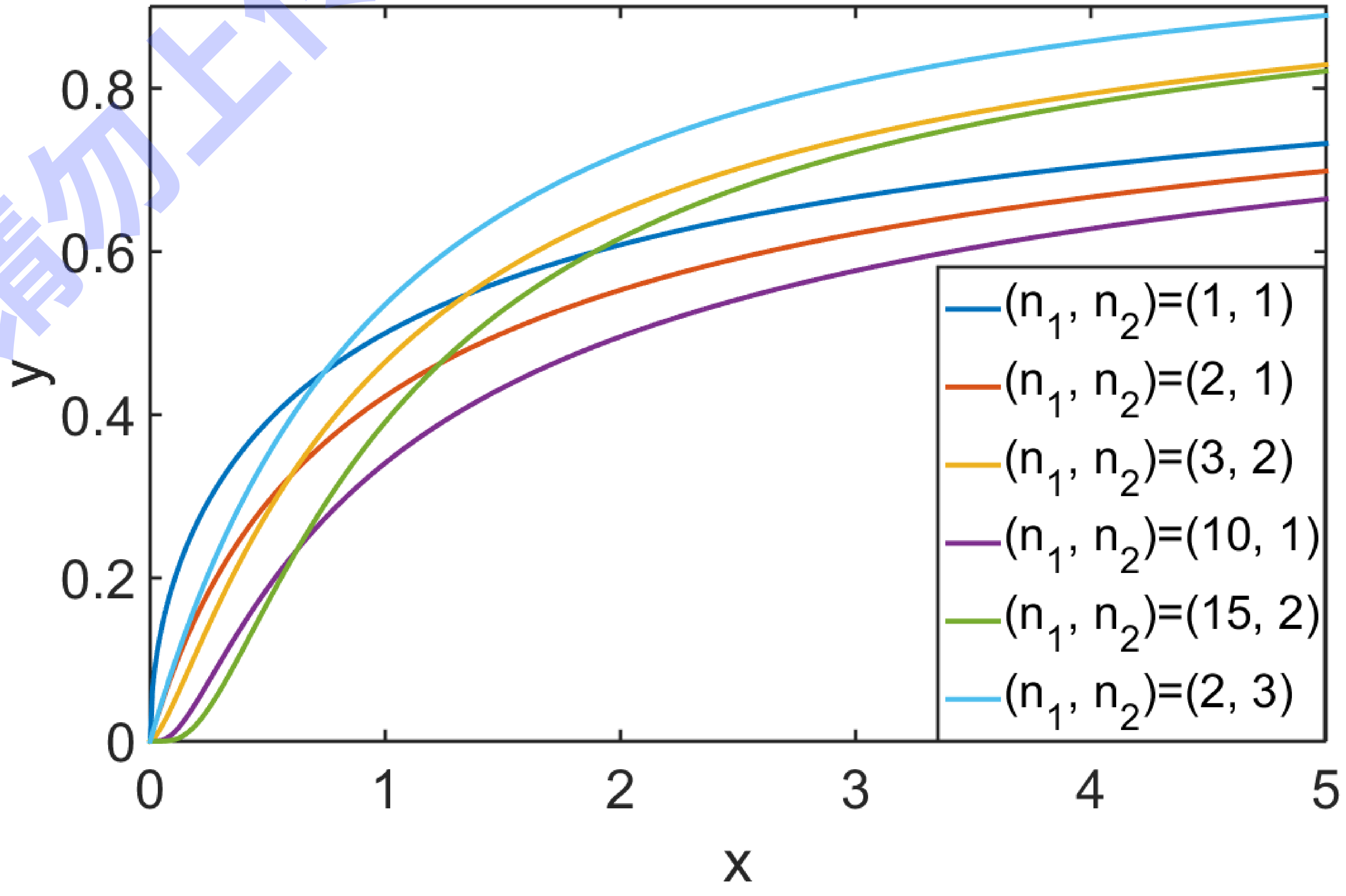
F分布的上α分位点可以查表获得，可用如下重要性质

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

F 分布的概率密度



F 分布的分布函数



例 设在总体 $X\sim N(\mu, \sigma^2)$ 中抽取样本$(X_1, X_2, \ldots, X_n)$，其中μ, σ^2未知，求:

(i) 统计量 $\chi^2=\dfrac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2$的分布

(ii) 设 $n=5$，若 $a(X_1-X_2)^2+b(2X_3-X_4-X_5)^2\sim\chi^2(k)$ 则a、b、k 各为多少？

解 (i) 作变换 $Y_i=\dfrac{X_i-\mu}{\sigma}$ $\quad i=1,2,\cdots,n$ 显然$Y_1, Y_2, \ldots, Y_n$相互独立，且来自总体$Y\sim N(0, 1)$

于是 $$\chi^2=\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2=\sum_{i=1}^{n}Y_i^2\sim\chi^2(n)$$

(ii) $X_1-X_2\sim N(0, 2\sigma^2)$, $\dfrac{(X_1-X_2)^2}{2\sigma^2}\sim\chi^2(1)$ $\quad 2X_3-X_4-X_5\sim N(0,6\sigma^2)$, $\dfrac{(2X_3-X_4-X_5)^2}{6\sigma^2}\sim\chi^2(1)$

X_1-X_2与 $2X_3-X_4-X_5$ 相互独立 $$\frac{(X_1-X_2)^2}{2\sigma^2}+\frac{(2X_3-X_4-X_5)^2}{6\sigma^2}\sim\chi^2(2)$$

$$a=\frac{1}{2\sigma^2},\quad b=\frac{1}{6\sigma^2}\qquad k=2$$

在实际应用中，四大分布在正态总体中，会演变出如下的八大分布

八大分布

正态总体：统计量服从的分布类型或规律总结

❖ 单正态总体

定理一 设$X\sim N(\mu,\sigma^2)$，而 $X_1, X_2, \ldots, X_n$ 是来自正态总体$N(\mu,\sigma^2)$的样本，$\overline{X}$ 和 S^2 分别是样本均值和样本方差，则有

1° $\overline{X}\sim N(\mu,\sigma^2/n)$　$\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$

2° $\dfrac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma^2}\sim\chi^2(n)$

3° $\dfrac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$　$\dfrac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{\sigma^2}\sim\chi^2(n-1)$ 且 $\overline{X}$ 与 S^2相互独立

4° $\dfrac{\overline{X}-\mu}{S/\sqrt{n}}\sim t(n-1)$

❖ 双正态总体

定理二 设 $(X_1,\cdots,X_{n_1})$和$(Y_1,\cdots,Y_{n_2})$ 分别是来自正态总体$N(\mu_1,\sigma_1^2)$和$N(\mu_2,\sigma_2^2)$ 的样本，

且这两个样本相互独立，设 $\overline{X}=\dfrac{1}{n_1}\sum_{i=1}^{n_1}X_i$，$\overline{Y}=\dfrac{1}{n_2}\sum_{i=1}^{n_2}Y_i$ 分别是样本均值，

$S_1^2=\dfrac{1}{n_1-1}\sum_{i=1}^{n_1}(X_i-\overline{X})^2$，$S_2^2=\dfrac{1}{n_2-1}\sum_{i=1}^{n_2}(Y_i-\overline{Y})^2$ 分别是样本方差，则有

5° $$\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{\sigma_1^2/n_1+\sigma_2^2/n_2}}\sim N(0,1)$$

6° 当$\sigma_1^2=\sigma_2^2$ 未知时 $$\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{S_W\sqrt{1/n_1+1/n_2}}\sim t(n_1+n_2-2)$$ 其中$S_W=\sqrt{\dfrac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}}$

7° $$\frac{n_2\sigma_2^2\sum_{i=1}^{n_1}(X_i-\mu_1)^2}{n_1\sigma_1^2\sum_{i=1}^{n_2}(Y_i-\mu_2)^2}\sim F(n_1,n_2)$$

8° $$\frac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2}=\frac{\sigma_2^2S_1^2}{\sigma_1^2S_2^2}\sim F(n_1-1,n_2-1)$$

例 设总体$X \sim N(40, 25)$

(1) 若从总体中抽取容量为36的样本，求 $P(38 \le \overline{X} \le 43)$

(2) 当样本容量 n 为多大时，$P(|\overline{X}-40|<1)=0.95$

解 (1) 由于$\mu=40, \sigma^2=5^2, n=36$ ，

因此 $\overline{X} \sim N\left(40, \frac{5^2}{36}\right)$ 从而

$$P(38 \le \overline{X} \le 43)=\Phi\left(\frac{43-40}{\sqrt{\frac{5^2}{36}}}\right)-\Phi\left(\frac{38-40}{\sqrt{\frac{5^2}{36}}}\right)$$

$$=\Phi(3.6)-\Phi(-2.4)=\Phi(3.6)+\Phi(2.4)-1=0.9916$$

例 设总体$X\sim N(40, 25)$

(1) 若从总体中抽取容量为36的样本，求 $P(38 \le \overline{X} \le 43)$

(2) 当样本容量 n 为多大时，$P(|\overline{X}-40|<1)=0.95$

解 (2) 由于$\mu=40, \sigma^2=5^2$， 因此 $\overline{X}\sim N\left(40,\dfrac{5^2}{n}\right)$ ，n为要确定的样本容量，所以

$$P(|\overline{X}-40|<1)=\Phi\left(\frac{40+1-40}{\sqrt{\frac{5^2}{n}}}\right)-\Phi\left(\frac{40-1-40}{\sqrt{\frac{5^2}{n}}}\right)=\Phi\left(\frac{\sqrt{n}}{5}\right)-\Phi\left(-\frac{\sqrt{n}}{5}\right)=2\Phi\left(\frac{\sqrt{n}}{5}\right)-1=0.95$$

从而可得 $\Phi\left(\dfrac{\sqrt{n}}{5}\right)=0.975, \dfrac{\sqrt{n}}{5}=1.96$ 故$n=96$

例 设$X_1, X_2, \dots, X_{14}$是来自正态总体$X\sim N(90,\sigma^2)$的一个样本，$\overline{X}$ 是样本均值，

(1) 若已知$\sigma^2=100$，求 $P\left(\sum_{i=1}^{14}(X_i-\overline{X})^2 \le 500\right)$

(2) 若σ^2未知，但已知样本方差$s^2=121$， 且 $P(|\overline{X}-90|\le k)=0.9$ ，求常数k

$\chi^2_{0.975}(13)=5, t_{0.05}(13)=1.7709$

解 (1) 由于 $\dfrac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{\sigma^2}\sim\chi^2(n-1)$ 且$n=14, \sigma^2=100$，因此 $\dfrac{\sum_{i=1}^{14}(X_i-\overline{X})^2}{100}\sim\chi^2(13)$

从而所求的概率为 $P\left(\sum_{i=1}^{14}(X_i-\overline{X})^2 \le 500\right)=P\left(\dfrac{\sum_{i=1}^{14}(X_i-\overline{X})^2}{100}\le\dfrac{500}{100}\right)=1-P\left(\dfrac{\sum_{i=1}^{14}(X_i-\overline{X})^2}{100}>5\right)$

$=1-0.975=0.025$

例 设$X_1, X_2, \ldots, X_{14}$是来自正态总体$X\sim N(90, \sigma^2)$的一个样本，$\overline{X}$ 是样本均值，

(1) 若已知$\sigma^2=100$，求 $P\left(\sum_{i=1}^{14}(X_i-\overline{X})^2\le 500\right)$

(2) 若σ^2未知，但已知样本方差$s^2=121$，且 $P(|\overline{X}-90|\le k)=0.9$ ，求常数k

$\chi^2_{0.975}(13)=5, t_{0.05}(13)=1.7709$

解 (2) 由于 $\dfrac{\overline{X}-\mu}{S/\sqrt{n}}\sim t(n-1)$ 且n=14，μ=90，s^2=121，因此 $\dfrac{\overline{X}-90}{\sqrt{121}/\sqrt{14}}\sim t(13)$

从而k的值取决于如下条件：$P(|\overline{X}-90|\le k)$ $=P\left(\left|\dfrac{\overline{X}-90}{11/\sqrt{14}}\right|\le\dfrac{k}{11/\sqrt{14}}\right)=0.9$

即 $P\left(\left|\dfrac{\overline{X}-90}{11/\sqrt{14}}\right|>\dfrac{k}{11/\sqrt{14}}\right)=0.1$ 由此可见，$\dfrac{k}{11/\sqrt{14}}=1.7709$ 从而k=5.2062

本节回顾

四大分布

1. 标准正态分布 $X \sim N(0, 1)$

2. χ^2分布$\chi^2 \sim \chi^2(n)$

3. t 分布$T \sim t(n)$

4. F 分布$F \sim F(n_1, n_2)$

八大分布

1° $\overline{X} \sim N(\mu, \sigma^2/n)$ $\quad \dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$

2° $\dfrac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma^2} \sim \chi^2(n)$

3° $\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$ $\quad \dfrac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{\sigma^2} \sim \chi^2(n-1)$ 且 $\overline{X}$ 与 S^2相互独立

4° $\dfrac{\overline{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$

复习思考题

1. 什么叫总体？什么叫简单随机样本？总体X的样本$X_1, X_2, \ldots, X_n$有哪两个主要性质？
2. 什么是统计量？什么是统计量的值？
3. 样本均值和样本方差如何计算？
4. $N(0, 1)$分布、t分布、χ^2分布以及F分布的双侧、下侧、上侧分位点是如何定义的？怎样利用附表查这些分位点的值？
5. 对一个正态总体的三个常用统计量及其分布是什么？
6. 对两个正态总体的三个常用统计量及其分布是什么？