第 5 章　统计量及其抽样分布

内容提要

- 5.1 统计量
- 5.2 由正态分布导出的几个分布
- 5.3 样本均值的分布与中心极限定理

5.1 统计量

定义

设$X_1, X_2, \cdots, X_n$是从总体X中抽取的容量为n的一个样本，如果由此样本构造一个函数$T(X_1, X_2, \cdots, X_n)$，不依赖于任何未知参数，则称函数$T(X_1, X_2, \cdots, X_n)$是一个**统计量。**

说明

- 统计量是样本的一个**函数**
- 统计量是**统计推断的基础**

常见的统计量

- 样本均值
- 样本方差
- 变异系数
- 样本偏度
- 样本峰度

5.2 由正态分布导出的几个重要分布

统计推断的3个中心内容

- 抽样分布
- 参数估计
- 假设检验

抽样分布的重要性

- 抽样分布决定**统计量的性质**
- 抽样分布影响**统计推断的优良性**

正态分布相关的3大分布

- χ^2分布
- t分布
- F分布

χ^2分布

定义

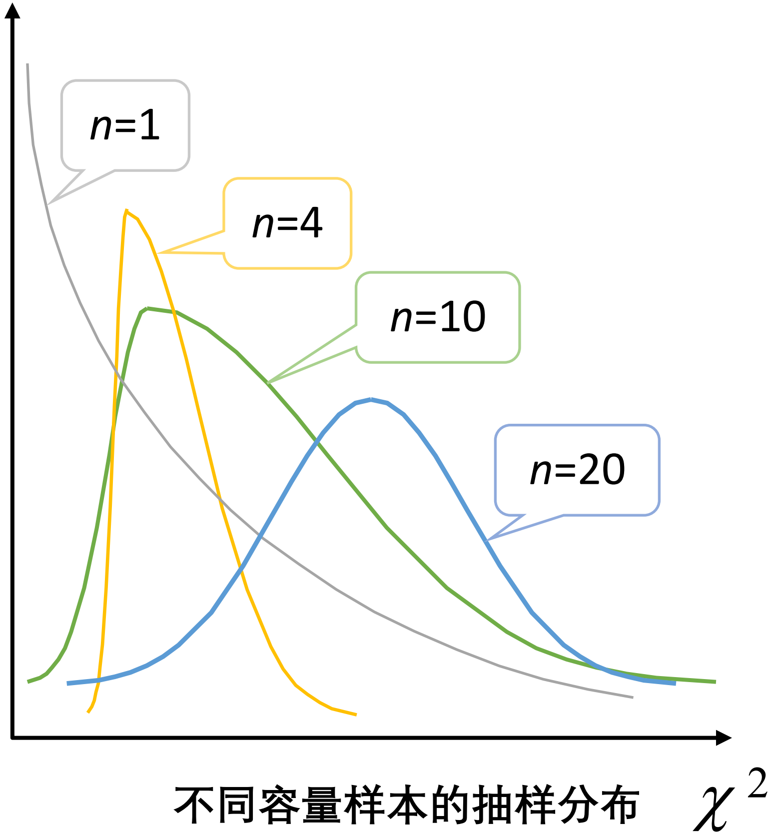
设随机变量X_1，X_2，…，X_n 相互独立，且$X_i(i=1，2，…，n)$服从标准正态分布 $N(0，1)$，则：它们的**平方和**$\sum_{i=1}^{n}X^2$ 服从自由度为n的$\boldsymbol{\chi^2}$**分布**。

记为： $\sum_{i=1}^{n}X^2 \sim \chi^2(n)$

提出： 赫尔默特（1875）和皮尔逊（1900）

性质

- 数学期望：$E(\chi^2)=n$
- 方差：$\mathrm{D}(\chi^2)=2n$
- 可加性：$X_1 \sim \chi^2(\boldsymbol{n_1})$，$X_2 \sim \chi^2(\boldsymbol{n_2})$，且相互独立，则$X_1+X_2 \sim \chi^2(\boldsymbol{n_1}+\boldsymbol{n_2})$
- 当$n\to+\infty$时，χ^2分布的极限分布是正态分布。



不同容量样本的抽样分布

t分布

定义

设随机变量$X \sim N(0，1)$，$Y \sim \chi^2(n)$，且X与Y独立，

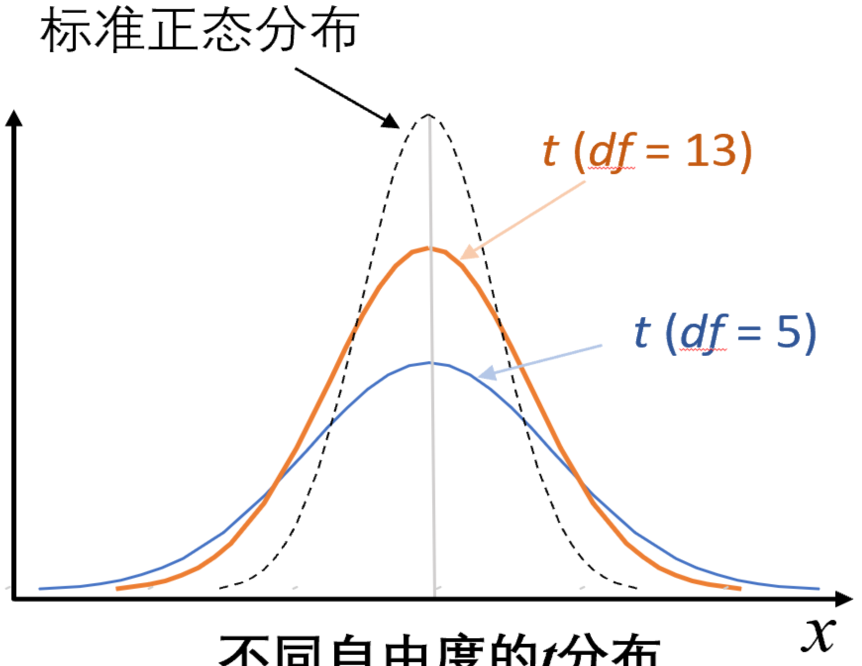
则$t = \frac{X}{\sqrt{Y/n}}$的分布为t分布（也称学生t分布）。

记为：$t \sim t(n)$

提出：戈塞特（1908）

性质

- 当$n = 1$时，t分布=柯西分布
- 当$n \geq 2$时，数学期望：$E(t) = 0$
- 当$n \geq 3$时，方差：$\mathrm{D}(t) = \frac{n}{n-2}$
- 当$n \geq 30$时，t分布**接近**标准正态分布
- 当$n \geq 120$时，t分布**非常接近**标准正态分布



不同自由度的*t*分布

F分布

定义

设随机变量Y与Z相互独立，且Y和Z分别服从自由度为m和n的χ^2分布，随机变量$X=\frac{Y/m}{Z/n}$，则X服从自由度为（m，n）的F分布。

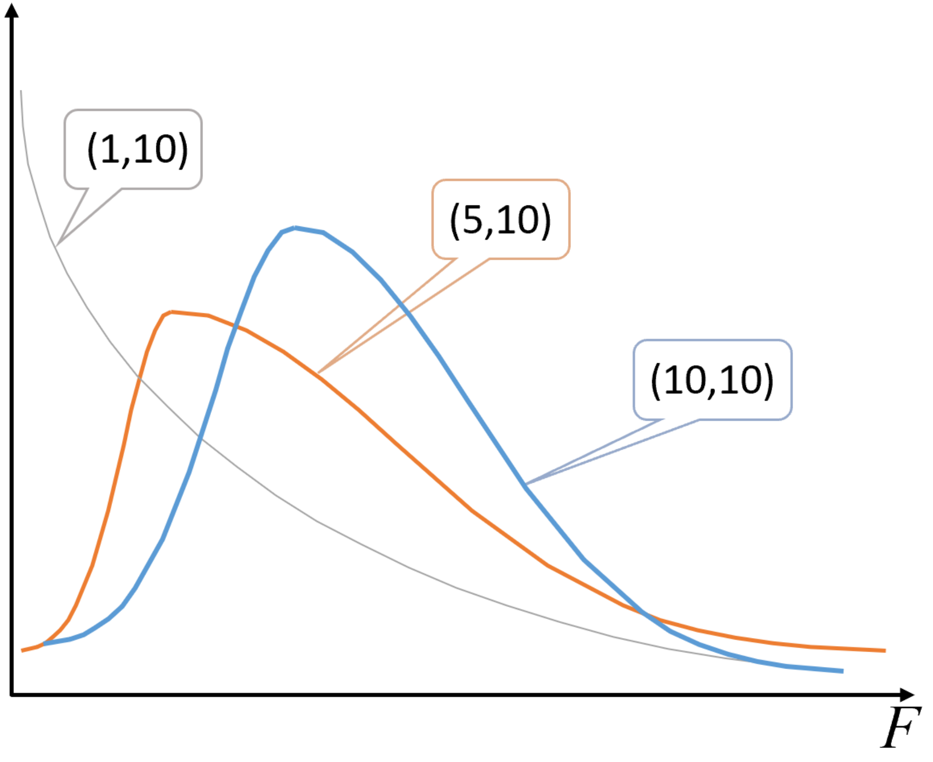
记为：$X \sim F(m,n)$

提出：费希尔

性质

- 当$n>2$时，数学期望：$E(X)=\frac{n}{n-2}$
- 当$n>4$时，方差：$D(X)=\frac{2n^2(m+n-2)}{m(n-2)(n-4)}$
- 与t分布关系：

  随机变量$t \sim t(n)$，则$t^2 \sim F(1,n)$



5.3 样本均值的分布与中心极限定理

样本均值的分布

当$n \to +\infty$时，$\bar{X} \sim N\left(\mu, \dfrac{\sigma^2}{n}\right)$

即：

当$n \to +\infty$时，$\dfrac{\bar{X}-\mu}{\sigma^2/n} \sim N(0，1)$

图示：[$\bar{X}$的抽样分布趋于正态分布的过程]

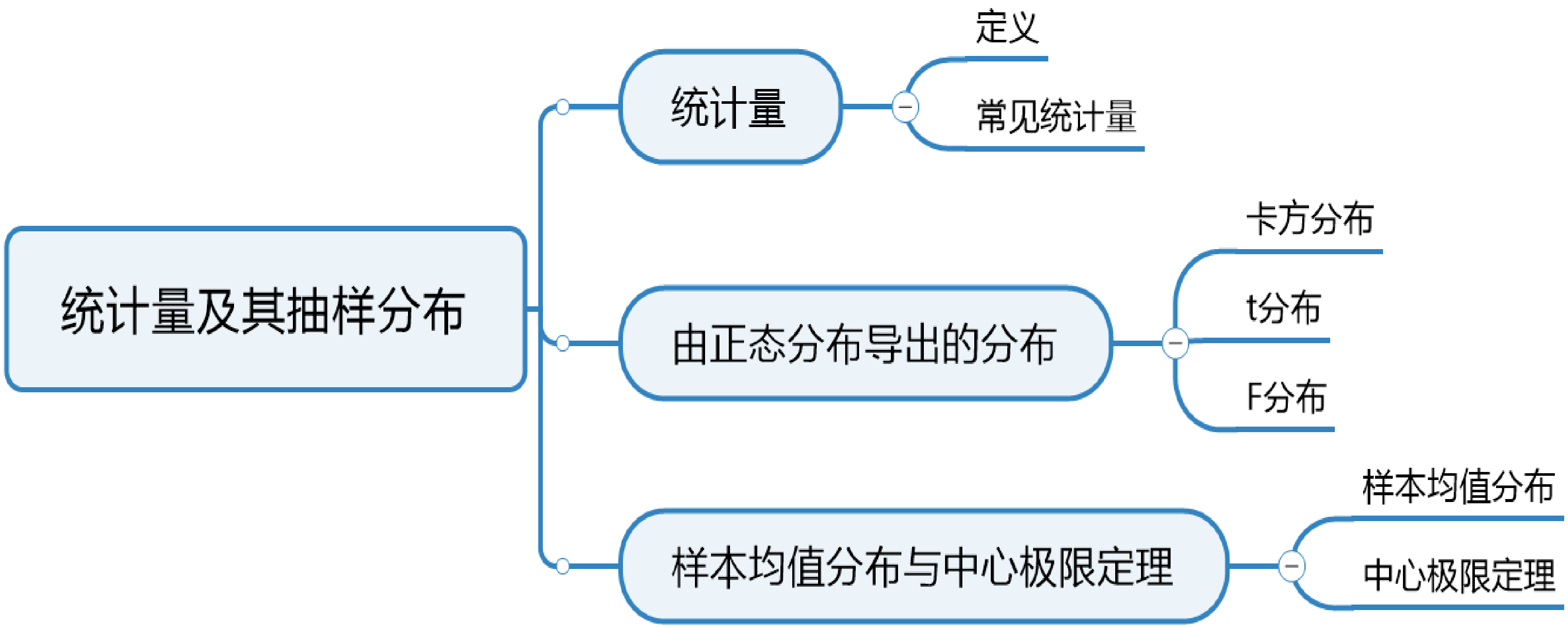
中心极限定理

从均值为μ，方差为σ^2的一个任意总体中抽取容量为n的样本，当n充分大时，样本均值的抽样分布近似服从均值为μ、方差为σ^2/n的正态分布。

问题：什么是"n充分大"?

- 小样本：样本量固定
- 大样本：$n \to +\infty$
- 经验判断：$n \geq 30$(参考值)

本章小结


统计量及其抽样分布
统计量
定义
常见统计量
由正态分布导出的分布
卡方分布
t分布
F分布
样本均值分布与中心极限定理
样本均值分布
中心极限定理