probability
probability

第六章
数理统计的基本概念

probability
probability

§6.1 总体、样本与统计量

§6.2 常用统计分布

§6.1 总体、样本与统计量

一、引言

引入

数理统计以概率论为理论基础,研究

1）研究如何以有效的方式收集和整理随机数据;

2) 研究如何合理地分析随机数据从而作出科学的推断 (称为**统计推断**).

两类工作有密切联系.

将主要介绍统计推断方面的内容.

二、总体

总体：研究对象的单位元素所组成的集合.

个体：组成总体的每个单位元素.

例1 要考察本校学生的身体情况，则将本校的所有学生视为一个总体，而每一位学生就是一个个体.

例2 考察某厂生产的电子元器件的质量，将全部产品视为总体，每一个元器件即为一个个体.

通常需要对总体的一项或几项**数量指标**进行研究.

如, 仅考虑学生的身高和体重(X, Y) ,不考虑学生的视力、成绩等.

如, 关心电子元件的寿命，则寿命 X 为其一个数量指标，且 X 是服从指数分布的随机变量.

由于上述数量指标往往是**随机变量**，具有一定的分布.

以后将(实际)总体和数量指标X等同起来.

总体分布是指数量指标 X的分布.

总 体 是 随 机 变 量

三、样本

一般，从总体中抽取一部分(取 n 个)进行观测，再依据这 n个个体的试验(或观察)的结果去推断总体的性质.

样本: 按照一定的规则从总体中抽取的一部分个体.

抽样：抽取样本的过程.

样本容量：样本中个体的数目 n .

将第 i 个个体的对应指标记为 X_i， i=1,2, …, n, 构成的随机向量 $(X_1, X_2, \cdots, X_n)$称为样本.

样本是一组随机变量，其具体试验(观察)数值记为： $x_1, x_2, \cdots, x_n$ ， 称为**样本观测值**，简称**样本值**.

为使样本具有代表性，抽样应满足什么条件?

（1）X_i 与总体同分布；

（2）$X_1, X_2, \cdots, X_n$ 相互独立.

定义6.1.1 设 $X_1, X_2, \cdots, X_n$ 是来自总体 X 的样本，如果<u>相互独立</u>且每个分量与总体<u>同分布</u>，称其为**简单随机样本**，简称样本.

若总体X的分布函数为 $F(x)$, 则样本X_1, X_2, ⋯, X_n的联合分布函数为

$$F(x_1, x_2, \cdots, x_n) = P\{X_1 \le x_1, X_2 \le x_2, \cdots, X_n \le x_n\}$$

$$= \prod_{k=1}^{n} F_{X_k}(x_k)$$

四、统计量

定义6.1.2 设X_1, X_2, ⋯, X_n是总体X的样本, T为n元实值函数，若样本的函数

$$T=T(X_1, X_2, \cdots, X_n)$$

是随机变量且不含未知参数，称 T为**统计量**.

对相应的样本值$(x_1, x_2, \dots, x_n)$，称

$$t = T(x_1, x_2, \dots, x_n)$$

为统计量的**统计值**.

TIPS

判断统计量

数理统计基本概念

总 体 是 随 机 变 量

样 本 是 随 机 向 量

统计量 是 随机变量(或向量)

常见统计量：

样本均值：

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

样本方差：

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

样本 k 阶原点矩：

$$A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^{\,k}$$

样本k阶中心矩：

$$M_k = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^k$$

统称样本矩

几个重要关系式:

$$\boxed{A_1 = \overline{X}}$$

$$\boxed{M_2 = \frac{n-1}{n}S^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2 = A_2 - A_1^2}$$

$\boxed{X,\ S^2,\ A_k,\ M_k}$ $\Longleftrightarrow$ $\boxed{x,\ s^2,\ a_k,\ m_k}$

统计量

统计值

思考 样本矩与总体矩 (即第四章中定义的矩) 的概念有什么区别?

样本矩 是 随机变量!
总体矩 是 数值!

例 6.1.1 设总体 $X \sim B(1, p)$，其中 p 是未知参数，$(X_1, X_2, \dots, X_5)$ 是来自 X 的简单随机样本，

1) 指出以下变量哪些是统计量，为什么？

$$X_1 + X_2, \quad \max_{1 \le i \le 5} X_i, \quad X_5 + 2p, \quad (X_5 - X_1)^2$$

2) 确定$(X_1, X_2, \dots, X_5)$的联合分布律？

解 1) 只有$X_5 + 2p$ 不是统计量,有未知参数p.

2) 因 $P\{X = x\} = p^x (1-p)^{1-x}, x = 0,1$

故 $(X_1, X_2, \dots, X_5)$ 的联合分布律为

$$P\{X_1=x_1, X_2=x_2, \dots, X_5=x_5\}$$

$$=\prod_{i=1}^{5} P\{X_i = x_i\} = \prod_{i=1}^{5} p^{x_i}(1-p)^{1-x_i}$$

$$= p^{\sum\limits_{i=1}^{5} x_i}(1-p)^{5-\sum\limits_{i=1}^{5} x_i}$$

$$x_i = 0, 1, \quad (i = 1,2,\cdots,5).$$

#

引例：某厂生产的一批产品中次品率为 p 。从中抽取10件产品装箱。

概

率

→

1）没有次品的概率

2）平均有几件次品

3）为以 0.95的概率保证箱中有10件正品，箱中至少要装多少件产品。

所有这些问题的关键是 p 是已知的！

如何获取 p ？ → 这就是数理统计的任务了！

一个很自然的想法就是：

首先从这批产品中随机抽取几件产品进行检验。

其次利用概率论的知识处理实测数据。

统计推断常解决的问题：

1）如何估计次品率 p？ 参数估计问题

2）如果以 $p<0.01$ 为出厂的标准，这批产品能否出厂？ 假设检验问题