问题的引入



在前面的章节中,介绍了概率论的基本内容,概率论中许多问题的讨论,常常是从已给的随机变量X出发来研究X的种种性质,那里X的概率分布都是已知的,或者假设是已知的.

但在实际问题中,一般说来,人们事先并不知道随机事件的概率、随机变量的概率分布和数字特征,而需要对它们进行估计或作某种推断,这就产生了数理统计的问题.



概率论和数理统计都是研究随机现象统计规律的学科, 但是它们在研究问题的方法上又有其自身的特点. 概率论由整体研究局部, 数理统计则是由局部研究整体.

概率论是数理统计的理论基础, 数理统计则是概率论的一种实际应用.

问题的引入



数理统计主要内容: 收集数据 统计推断

数据收集:研究如何对随机现象进行观察或试验, 以便获得 能够很好地反映整体情况的局部数据(抽样技术,试验设计 等).

统计推断:研究如何对收集到的局部数据进行整理分析,并对所考察的对象的整体特性作出尽可能准确可信的推测和判断(参数估计,假设检验,方差分析,回归分析等).



第六章

张小玲 安徽大学物质科学与信息技术研究院



概率论与数理统计

第六章 数理统计的 基本概念 第一节 总体与样本

第二节 统计量

第三节 抽样分布



第一节 总体与样本

一、总体、样本、统计量的概念

二、小结



总体:全体研究对象构成的集合.

个体: 总体中的每一个元素.

例如:从5000个产品中随机地抽检一个产品检查 是否合格的试验中,每个产品是个体,5000个产 品就是一个总体;

总体容量: 总体中包含个体的个数

有限总体

无限总体



抽样: 按照一定规则从总体抽取部分个体的过程.

样本: 按一定规则从总体中抽取的部分个体构成的集合.

样本容量: 样本中所含的个体(样品)的个数.

随机抽样(重复抽样, 非重复抽样), 判断抽样(分层抽样,系统

抽样,…)



总体与样本的表示

总体:X, Y, Z, ···

样本: X_1,\dots,X_n .

样本观测值: x_1, \dots, x_n .

简单随机样本:来自于同一个总体 X, 且相互独立的样本, 简称样本.



设总体X的分布函数为F(x),则取自X的简单随机

样本 $(X_1, X_2, ..., X_n)$ 的联合分布函数为:

$$F(x_1, x_2, ..., x_n)$$

= $P(X_1 \le x_1, X_2 \le x_2, ..., X_n \le x_n) = \prod_{i=1}^n F(x_i)$



1) 若离散型总体X有分布律 $\binom{x_1}{p_1} \binom{x_2}{p_2} \dots \binom{x_k}{p_k} \dots$,则样本的联合分布律为:

$$P(X_1 = x_{j1}, X_2 = x_{j2}, ..., X_n = x_{jn}) = \prod_{i=1}^n P(X_i = x_{ji}) = \prod_{i=1}^n p_{ji}$$

其中 x_{i1} 为 X_i 的一个可能取值,i = 1, 2, ..., n

2) 若连续型总体X有密度函数f(x),则样本的联合密度函数为:

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



例6.1.1 设总体X服从泊松分布 $P(\lambda)$,试求其样本 $(X_1, X_2, ..., X_n)$ 的联合概率

分布; (2) 设总体X服从指数分布 $E(\lambda)$,试求其样本 $(X_1,X_2,...,X_n)$ 的联合概率密度。

解 (1) 由于
$$X \sim P(\lambda)$$
,则有 $P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$, $k = 0, 1, ...$

由于样本 $X_1, X_2, ..., X_n$ 相互独立且与X同分布,则联合概率分布为

$$P(X_1 = k_1, X_2 = k_2, ..., X_n = k_n) = \prod_{i=1}^n P(X_i = k_i) = \prod_{i=1}^n \frac{\lambda^{k_i}}{k_i!} e^{-\lambda}$$

其中 $k_i = 0, 1, ..., i = 1, 2, ..., n$.



(2) 由于 $X \sim E(\lambda)$,则

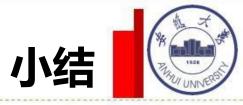
$$f(x) = \begin{cases} \lambda e^{-\lambda x}, x > 0, \\ 0, x \leq 0. \end{cases}$$

由于样本 $X_1, X_2, ..., X_n$ 相互独立且与X同分布,则样本的联合概率密度为

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

$$=\begin{cases} \lambda^n e^{-\lambda \sum_{i=1}^n x_i}, x_i > 0, i = 1, 2, ..., n \\ 0, & \text{ i.e.} \end{cases}$$

6.1. 总体与样本



总体: 全体研究对象构成的集合.

个体: 总体中的每一个元素.

总体容量: 总体中包含个体的个数 无限总体

样本: 按一定规则从总体中抽取的部分个体构成的集合.

样本容量: 样本中所含的个体(样品)的个数.

抽样: 按照一定规则从总体抽取部分个体的过程.