数理统计的基本概念

数理统计的基本内容：

如何收集、整理所研究随机变量的数据

如何对这些数据进行分析

如何对所研究的随机变量做出种种推断

总体和样本

在数理统计中，我们把研究对象的全体称为总体，组成总体的每个成员称为个体.

特指：

研究对象的某项数量指标的全体称为总体，组成总体的每个成员的该项数量指标称为个体.

总体指标未知，可看作是一个随机变量，记为X.

(1)当总体X是离散型随机变量时，定义总体分布为

$$f(x,\theta)\triangleq P(X=x),$$ 即为总体X的概率函数.

(2)当总体X是连续型随机变量时，定义总体分布为

$$f(x,\theta)\triangleq f_X(x),$$ 即为总体X的概率密度函数.

例1　设总体 $X \sim P(\lambda)$，试写出总体分布 $f(x,\lambda)$.

解

$$f(x,\lambda) \triangleq P(X = x) = \frac{\lambda^x}{x!}e^{-\lambda}, x = 0,1,2,\cdots$$

例2　设总体 $X \sim R(0,\theta)$，试写出总体分布 $f(x,\theta)$.

解

$$f(x,\theta) \triangleq f_X(x) = \begin{cases} \theta^{-1} & 0 < x < \theta \\ 0 & \text{其余} \end{cases}$$

在数理统计中总体分布往往是未知的，有时虽然知道总体分布的类型，但分布中的参数却未知．所以我们希望从客观存在的总体中选取一些个体（即抽样），通过对这些个体作观察或测试来推断关于总体分布$f(x,\theta)$中的某些量如总体的均值或方差等．这些抽取的个体便称为是取自总体的一个样本，这些个体的观测值则称为样本观测值．

在试验前，样本的观测值是不确定的，为了体现随机性，在数理统计中样本记作$(X_1, X_2, \cdots, X_n)$，事实上是一个n维随机向量. 通过试验或观测得到的数值称为样本观测值，记作$(x_1, x_2, \cdots, x_n)$，其中n称为样本容量(样本大小). 也就是说样本是一组随机变量，而样本观测值是抽样完成以后所得到的这组随机变量的一次具体取值.

从总体中抽取样本的方法有很多， 我们主要采用简单随机抽样的方法，即有放回地重复独立抽取，这样得到的样本称为简单随机样本．即简单随机样本具有两个特点：

(1)独立性：$X_1, X_2, \cdots, X_n$ 是相互独立的；

(2)代表性：每个个体 X_i 的分布都和总体分布相同．

即 $X_i \sim f(x_i, \theta), i = 1, 2, \cdots, n$．以下都简称为样本．

问题

如何写出样本的联合密度（概率）函数?

样本的联合分布记为 $f^*\left(x_1,x_2,\cdots,x_n,\theta\right)$

事实上由简单随机样本的两个特点即可求得:

(1)设 X 为离散型随机变量，则 $X \sim f(x,\theta) \triangleq P(X=x)$

而样本 $(X_1, X_2, \cdots, X_n)$ 的联合概率函数为：

$$f^*(x_1, x_2, \cdots, x_n, \theta) \triangleq P(X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n)$$

$$= P(X_1 = x_1) P(X_2 = x_2) \cdots P(X_n = x_n)$$

$$= f_{X_1}(x_1, \theta) f_{X_2}(x_2, \theta) \cdots f_{X_n}(x_n, \theta)$$

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

例3 设总体 $X \sim B(1,p)$, $(X_1, X_2, \cdots, X_n)$ 是取自该总体的一个样本，试写出 $(X_1, X_2, \cdots, X_n)$ 的联合概率函数.

解

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

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

$$= p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i}, \quad x_i = 0,1, \quad i = 1,2,\cdots,n.$$

(2)设 X 为连续型随机变量，概率密度函数为 $f(x,\theta)$，则$(X_1,X_2,\cdots,X_n)$的联合概率密度函数为:

$$
\begin{aligned}
f^*(x_1,x_2,\cdots,x_n,\theta) &\triangleq f_{(X_1,X_2\cdots,X_n)}(x_1,x_2,\cdots,x_n) \\
&= f_{X_1}(x_1)\,f_{X_2}(x_2)\cdots f_{X_n}(x_n) \\
&= f(x_1,\theta)\,f(x_2,\theta)\cdots f(x_n,\theta) \\
&= \prod_{i=1}^{n} f(x_i,\theta).
\end{aligned}
$$

例4　设$(X_1, X_2, \cdots, X_n)$是取自总体 X 的一个样本, X 的概率密度函数为

$$f(x,\theta)=\begin{cases}\dfrac{2x}{\theta^2} & 0<x<\theta \\ 0 & \text{其余}\end{cases}$$

试写出$(X_1, X_2, \cdots, X_n)$的联合密度函数.

解　联合密度函数为

$$f^*(x_1,x_2,\cdots,x_n,\theta)=\begin{cases}\dfrac{2^n x_1\cdots x_n}{\theta^{2n}} & 0<x_i<\theta,\ i=1,2,\cdots,n \\ 0 & \text{其余}\end{cases}$$

课堂提问

例4　设$(X_1, X_2, \cdots, X_n)$是取自总体 X 的一个样本，X 的概率函数为

X	-1	0	1
Pr.	$\frac{\theta}{2}$	$1-\theta$	$\frac{\theta}{2}$

其中参数 $\theta(0<\theta<1)$，试写出$(X_1, X_2, \cdots, X_n)$的联合概率函数?

谢　谢

同济大学数学科学学院概率统计教学团队