

专注于商业智能BI和大数据的垂直社区平台

数理统计的基本概念(二)

Allen

www.hellobi.com

课程目录

- 联合概率分布函数及联合概率密度
- •参数空间
- 统计量及抽样分布
- 次序统计量
- 小结



联合概率分布函数及联合概率密度

- 简单随机样本:从总体X中随机抽取一部分个体 $X_1, X_2, ..., X_n$,称 $X_1, X_2, ..., X_n$ 为取自X的容量为n的样本,n称作容量, $x_1, x_2, ..., x_n$ 称作样本观测值
 - 1.母体中的每一个体有同等机会被选入子样
 - 2.字样的分量 $X_1, X_2, ..., X_n$,是相互独立的随机变量,即子样的每一分量的观测结果并不影响其他分量的观测结果



联合概率分布函数及联合概率密度

• 样本作为随机变量,也有概率分布,称为样本分布

- 若总体X具有分布函数 F(x) , 则 $(X_1, X_2, ..., X_n)$ 的联合概率分布函数为 $F(X_1, X_2, ..., X_n) = \prod_{i=1}^n F(x_i)$
- 若总体X具有概率密度函数 f(x) , 则 $(X_1, X_2, ..., X_n)$ 的联合概率密度为 $f(X_1, X_2, ..., X_n) = \prod_{i=1}^n f(x_i)$



联合概率分布函数及联合概率密度

- 例:设某牌子的手表的寿命X服从指数分布 $f(x,\lambda) = \begin{cases} \lambda e^{-\lambda x}, x \geq 0, \\ 0, x < 0, \end{cases}$ 假设从一批产品中独立地抽取n块手表进行试验,测得寿命数据为 $X_1, X_2, ..., X_n$ 求样本 $(X_1, X_2, ..., X_n)$ 的概率分布
- 分析: $X_1, X_2, ..., X_n$ 是独立同分布,且 $X_i \sim f(x, \lambda)$,故所求的概率密度为

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

$$= \begin{cases} \lambda^n \exp\left\{-\lambda \sum_{i=1}^n x_i\right\}, x_1, x_2, ..., x_n \ge 0, \\ 0, \not \sqsubseteq \dot{\mathcal{C}}, \end{cases}$$



参数空间

- 上例中总体分布为指数分布 $f(x,\lambda)$, λ 是确定分布的常数 , 在数理统计中 , 出现在样本分布中的常数为参数
- 参数可能为已知的,也可能未知,把参数的所有可能取值所构成的集合称为参数空间,如上例中参数空间为 $\Theta = \{\lambda : \lambda > 0\}$



统计量及抽样分布

• 根据样本计算出的量称为统计量,实际上统计量是样本的某种函数

从该牌子手表中随机抽取n个手表的使用寿命

• 设 $X_1, X_2, ..., X_n$ 是总体X的一个简单随机样本, $T(X_1, X_2, ..., X_n)$ 为一个n元 连续函数,且T中不含任何关于总体的未知参数,则称 $T(X_1, X_2, ..., X_n)$ 为一个统计量,统计量的分布为抽样分布



统计量及抽样分布—常用统计量

- 样本均值:设 $X_1, X_2, ..., X_n$ 是总体X的一个简单随机样本称 $\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$ 为样本均值
- 样本方差:设 $X_1, X_2, ..., X_n$ 是总体X的一个简单随机样本 , \overline{X} 为样本均值 , $\pi \, S^2 = \frac{1}{n} \sum_{i=1}^n (X_i \overline{X})^2$ 或 $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i \overline{X})^2$ 为样本方差



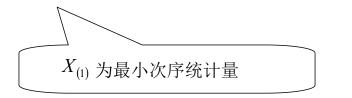
统计量及抽样分布—常用统计量

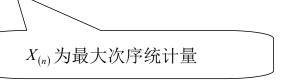
- 样本k阶原点矩:设 $X_1, X_2, ..., X_n$ 是总体X的一个简单随机样本称 $A_k = \frac{1}{n} \sum_{i=1}^n X_i^k$ 为样本的k阶原点矩
- 样本k阶中心矩:设 $X_1, X_2, ..., X_n$ 是总体X的一个简单随机样本, \overline{X} 为样本均值,称 $M_k = \frac{1}{n} \sum_{i=1}^{n} (X_i \overline{X})^k$ 为样本k阶中心矩



次序统计量

• 设 $X_1, X_2, ..., X_n$ 是总体X的一个简单随机样本, $x_1, x_2, ..., x_n$ 为样本观测值,将 $x_1, x_2, ..., x_n$ 按照从小到大的顺序排列为 $x_{(1)} \le x_{(2)} \le ... \le x_{(n)}$,这样,当样本 $x_1, x_2, ..., x_n$ 取值为 $x_1, x_2, ..., x_n$ 时,定义 $x_{(k)}$ 取值为 $x_{(k)}$,就称 $x_{(1)}, x_{(2)}, ..., x_{(n)}$ 为 $x_1, x_2, ..., x_n$ 的次序统计量







次序统计量

- 次序统计量之间是否相互独立?
- 设 X_1, X_2 是总体X的一个容量为2的样本,X的分布律如下

X	0	1
Р	1/2	1/2

• 所有的次序统计量 $X_{(1)}, X_{(2)}$ 如下



次序统计量

X_1	X_2	$X_{(1)}$	$X_{(2)}$
0	0	0	0
0	1	0	1
1	0	0	1
1	1	1	1

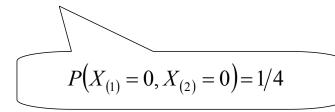
•X(1) 的分布律如下:

$X_{(1)}$	0	1
Р	3/4	1/4

• X(2)的分布律如下:

$X_{(2)}$	0	1
Р	1/4	3/4

$$P(X_{(1)} = 0)P(X_{(2)} = 0) = 3/4 \cdot 1/4 = 3/16$$







小结

- 联合概率分布函数及联合概率密度
- •参数空间
- 统计量及抽样分布
- 次序统计量
- 小结

