天善智能
TIANSHAN SOFT

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（一）

Allen

www.hellobi.com

课程目录

- 概率分布函数
- 连续型随机变量及概率密度函数
- 引子
- 统计推断
- 总体、个体及简单随机样本
- 小结

概率分布函数

- 例：等可能的在[a,b]上投点，即所投的点在[a,b]中的任一子区间 $B=[c,d]$中的概率，与B的长度l_B成正比

$$P(B)=\frac{l_B}{b-a}=\frac{d-c}{b-a}$$

- 假设：如果投在[a,b]中的点的坐标为ω，令$\xi(\omega)=\omega,(a\le\omega\le b)$，这样就得到了随机变量$\xi(\omega)$

概率分布函数

上述$\xi(\omega)$，如果取[a,b]中任意一点值ω_0概率

$$P(\xi(\omega)=\omega_0)=P(\omega=\omega_0)=\frac{l_{\omega_0}}{b-a}=0$$

设$B=[c,d]\subset[a,b]$

$$P(c\le\xi\le d)=P(B)=\frac{d-c}{b-a}$$

$$P(c\le\xi\le d)=P(c<\xi\le d)$$

$$P(c<\xi\le d)=P(\xi\le d)-P(\xi\le c)$$

$$F(x)=P(\xi(\omega)\le x)=?$$

天善智能 TIANSHAN SOFT

概率分布函数

- 定义：定义在样本空间Ω上，取值于实数域的函数ξ(ω)，称为是样本空间Ω上的随机变量，并称$F(x)=P(\xi(\omega)\le x), x\in(-\infty,+\infty)$是随机变量ξ(ω)的概率分布函数
- 单调性：若$x_1 < x_2$，则$F(x_1)\le F(x_2)$
- 有界性：$F(-\infty)=0$，$F(+\infty)=1$
- 右连续性：$F(x+0)=F(x)$

天善智能 TIANSHAN SOFT

概率分布函数

- 例：若ξ只取一个值，即 $P(\xi=a)=1$ ，求ξ的分布函数

$$F(x)=P(\xi\le x)$$

- 当 $x\ge a$ ： 1
- 当 $x<a$ ： 0

连续型随机变量及概率密度函数

- 定义：若ξ(ω)是随机变量，$F(x)$ 是它的分布函数，如果存在函数 $p(x)$，对任意x,有 $F(x)=\int_{-\infty}^{x} P(y)dy$，则称ξ(ω) 为连续型随机变量

- 上式中 $F(x)$为连续型分布函数，$p(x)$是$F(x)$ 的概率密度函数
  - (1) $p(x)\geq 0$
  - (2) $\int_{-\infty}^{+\infty} p(x)dx = 1$

天善智能 TIANSHAN SOFT

引子

抛掷一枚硬币出现正面的概率 → 某一公路上车辆速度的分布

随机抽取编号为1，2，…，10的相同球，抽中编号为1的概率 → 手表寿命的长短服从的分布

盒子中有5个红球5个白球，随机抽取，抽中编号红球的概率 → 工厂检验产品是否合格所服从的两点分布的合格率p

统计推断

- **概率论**：随机变量及其所伴随的概率分布能够全面描述随机现象的统计规律，根据分布求事件的概率
- **数理统计**：从所要研究的对象全体中抽取一部分进行观察实验并取得相关信息，从而对整体及其分布进行推断
- **统计推断**：每个推断都伴随一定的概率以表明推断的可靠程度，这种伴随一定概率的推断称为统计推断

天善智能 TIANSHAN SOFT

总体、个体及简单随机样本

- 总体：把研究对象的全体所构成的集合称为母体或总体

研究某个牌子的手表的使用寿命，某个牌子的所有手表的寿命集合

- 个体：组成总体的每一成员

该牌子手表中的每一个手表的寿命

- 简单随机样本：从总体X中随机抽取一部分个体$X_1, X_2, \ldots, X_n$，称$X_1, X_2, \ldots, X_n$为取自X的容量为n的样本，n称作容量，$x_1, x_2, \ldots, x_n$称作样本观测值

从该牌子手表中随机抽取n个手表的使用寿命

天善智能
TIANSHAN SOFT

小结

- 概率分布函数
- 连续型随机变量及概率密度函数
- 引子
- 统计推断
- 总体、个体及简单随机样本
- 小结