研究生现代工程数学
《数理统计》

主讲:李卫国

Mail:liwg@buaa.edu.cn

教材:数理统计(北航，孙海燕等)

学时:48 课时

第一章 概 论

1、什么是数理统计
2、数理统计的发展进程
3、数理统计的应用领域
4、数理统计的数学基础:概率论

第1节 数理统计学科简介

1、什么是数理统计

数理统计学 (Mathematical Statistics) 是数学的一个分支学科， 研究怎样有效地收集 、整理和分析带有随机性的数据,以对所观察的问题作出推断或预测， 直至为采取一定的决策和行动提供依据与建议。

以概率论为基础，对随机数据进行

收集	抽样理论方法、抽样调查
整理	描述性统计，常规统计
分析	建立统计模型
推断	假设检验、预测预报

数理统计内容

- 参数估计
  - 点 估 计
  - 区间估计
- 假设检验
  - 参数检验
  - 分布检验
- 多元分析
  - 主成分 因子分析
  - 判别、聚类
- 回归分析
  - 一元回归
  - 多元回归

2、数理统计的发展过程

- statistics 衍生于

  status国家, statista 政治家

- 萌芽于“国算学”
- 概率论:贝努利大数定律

  奠定数理统计基础

➢ 20世纪初至二战结束:快速发展

K.Pearson,1857-1936　矩估计,

1900 提出拟合优度卡方统计量

F.Galton,1822-1911　回归分析

R.A.Fisher 1890-1962　抽样理论,方差分析

W.S.Gosset 1928 t分布（里程碑）

J.Neyman 1894-1981　二人1928-1938论文

E.S.Pearson 1895-1980　假设检验基础

- 二战之后:规范深入

  U.W.Wishart,1928 发现“维希特”分布

  许宝禄 1940年代多元统计奠基工作

  G.U.Yule,1925-1930开始了时间序列分析

  A.wald,1950年代提出了统计决策理论

  1950年代后，贝叶斯学派兴起

- 计算机发展的推动

  统计学的广泛应用与普及

3、数理统计的应用

- 国民经济与企业管理
- 医药行业、医学统计
- 生物统计、社会统计
- 科学与工程实验数据处理
- 数据挖掘与数据仓库
- 云计算与大数据

4、数学基础：概率论回顾

一、概率空间

- 概率论是研究随机现象统计规律的数学分支。
- 对随机现象观察：随机试验
- 随机试验的结果：随机事件（掷骰子）
- 随机事件运算：交、并、余；样本点
- 要求运算可以实施 —— 运算封闭性

一、概率空间

1、事件域(运算封闭性，规定什么是事件)

- 定义：

设Ω是样本空间，F是Ω的子集构成的集合类

若满足

$$(1)\ \Omega \in F$$

$$(2)\ A \in F,\quad \overline{A} \in F$$

$$(3)\ A_n \in F,\quad \bigcup_{n=1}^{\infty} A_n \in F$$

则称 F为事件域(σ代数)。

一、概率空间

1、事件域（运算封闭性，规定什么是事件）

- 定义：
- 性质：可列交、并以及补的运算都封闭
- 可测空间 (Ω, F)
- 举例：

一、概率空间

2、概率

- 概率是随机事件发生可能性的度量
- P(A)表示 A 发生的可能性大小
- P(A) 是一个定义在事件域上的“集函数”
- P(A) 是对A的一种度量侧度: 概率侧度
- 规定了概率测度的空间:概率空间
- 概率的数学定义:“非负、归一、可加”

二、随机变量与分布函数

1、随机变量

- 把随机事件与数或数的集合对应起来
- 把随机事件映射到实数的集合
- 比如:
- 在比如:
- 随机变量是从样本空间到实数空间的可测映射
- 保证$\{w: X(w)\leq x\}$是事件，对应事件$\{X\leq x\}$

2、分布函数

- 分布函数的定义

$F(x)=P\{w: X(w)\leq x\}$，记为$P\{X\leq x\}$

- F(x)的性质

(1) 单调不减

(2) 归一化

(3) 右连续

- 离散分布: 分布列(律)
- 连续分布: 密度函数

4、随机向量的联合分布

- 二维随机向量(X,Y)
- 联合分布 $F(x,y) = P\{X\le x, Y\le y\}$
- 边际分布 $F_X(x)$, $F_Y(y)$
- 协方差 $Cov(X,Y) = E(X-EX)(Y-EY)$
- 相关系数
- X与Y独立 $F(x,y) = F_X(x)\, F_Y(y)$

  $f(x,y)=f_X(x)\, f_Y(y)$
- 推广到 n 维分布

三、常用分布（概率模型）

1、两点分布（0-1分布）

- 掷币是否正面, 射击是否命中, 考试是否通过
- A是否发生, 成功-失败模型
- 贝努利模型

X	0	1
P	1-p	p

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

2、二项分布 B(n, p)

- **掷n个币正面出现次数**
- **n次射击命中次数**
- **n次重复试验中A发生的次数**
- **n重贝努利模型**
- **n个两点分布的独立和**

$$P\{X=k\}=C_n^k p^k (1-p)^{n-k},$$
$$k=0,1,\cdots,n$$

3、离散均匀分布

- 等可能模型
- 掷骰子模型

$$P\{X=k\}=\frac{1}{n},\ k=1,2,\cdots,n$$

4、Poisson分布

- 随机点计数模型
- 灾害次数, 顾客流

$$P\{X=k\}=\frac{\lambda^k}{k!}e^{-\lambda},\ k=1,2,\cdots$$

5、几何分布

- “首中” 模型

$$P\{X=k\}=p(1-p)^{k-1},\ k=1,2,\cdots$$

6、多项分布

- n次足球比赛中“胜平负” 次数

$$P\{X_1=k_1,\cdots,X_r=k_r\}=\frac{n!}{k_1!\cdots k_r!}p_1^{k_1}\cdots p_r^{k_r},$$

$$k_1+\cdots+k_r=n$$

7、连续均匀分布 U(a,b)

- 区间[a,b]上的等可能模型

$$f(x)=\begin{cases}\dfrac{1}{b-a}, & a\le x\le b\\ 0\ , & □□\end{cases}$$

8、指数分布 E(λ)

- 寿命分布

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

9、正态分布 N(μ,σ²)

- 最常用的分布函数
- 最自然的分布形态
- 标准正态分布 N(0,1)

$$\phi(x)=\frac{1}{\sqrt{2\pi}}\exp(-\frac{x^2}{2})$$

正态分布 N(μ,σ²)

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$

10、对数正态分布 N(μ,σ²)

- **ln X 是正态分布**
- **寿命分布**

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma x}\exp\left\{-\frac{(\ln x-\mu)^2}{2\sigma^2}\right\},\quad x>0$$

11、伽玛分布 Γ(α,λ)

- **指数分布的推广**
- **寿命分布**

$$f(x)=\frac{\lambda^\alpha}{\Gamma(\alpha)}x^{\alpha-}e^{-\lambda x},\quad x>0$$

12、贝塔分布 Be(α,β)

$$f(x)=\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}x^{\alpha-1}(1-x)^{\beta-1},\quad 0<x<1$$

13、Weibull (韦伯, 威布尔）分布 W(m,λ,γ)

- 伽玛分布的推广
- 寿命分布，适应性强
- 三参数分布，形状、尺度、位置参数

$$f(x)=\lambda m(x-\gamma)^{m-1}\exp\left\{-\lambda(x-\gamma)^{m}\right\},\quad x>\lambda$$

14、统计三大分布

t 分布、卡方分布、F分布

四、极限与极限定理

1、随机变量序列的几种极限

- **以概率1收敛:**

$$P\{\lim_{n\to\infty} X_n = X\} = 1 \qquad X_n \xrightarrow{a.s} X$$

- **依概率收敛:** $X_n \xrightarrow{P} X$

$$\lim_{n\to\infty} P\{|X_n - X| \geq \varepsilon\} = 0 \quad □ \quad \lim_{n\to\infty} P\{|X_n - X| < \varepsilon\} = 1$$

- **r 阶收敛:**

$$\lim_{n\to\infty} E|X_n - X|^r = 0 \quad X_n \xrightarrow{r} X$$

- **均方收敛:** $\lim_{n\to\infty} E|X_n - X|^2 = 0$

- **依分布收敛:**

□ $X_n \sim F_n(x)$□$X \sim F(x)$.

□ $F(x)$□□□□□□ x□□ $\lim\limits_{n\to\infty} F_n(x) = F(x)$

□□ $F_n(x)$□□□□□ $F(x)$□□□

$$F_n(x) \xrightarrow{\text{W}} F(x)$$

□□ X_n□□□□□□ X.□□

$$X_n \xrightarrow{\text{L}} X$$

- **各种收敛性的相互关系**

$$X_n \xrightarrow{a.s} X \quad \Longrightarrow \quad X_n \xrightarrow{P} X \quad \Longrightarrow \quad X_n \xrightarrow{P} X$$

2、大数定律

- 伯努利大数定律：

□ $X_n \sim B(1,p)$□ □ □ □ □ □ □ □ □ □

$$\lim_{n\to\infty} P\left\{\left|\frac{1}{\mathrm{n}}\sum_{k=1}^{n} X_k - p\right| < \varepsilon\right\} = 1 \qquad \mu_n \xrightarrow{P} p$$

- 辛钦大数定律：

□ X_n□ □ □ □ □ □ □ □ □ $EX_n = \mu$,□

$$\lim_{n\to\infty} P\left\{\left|\frac{1}{\mathrm{n}}\sum_{k=1}^{n} X_k - \mu\right| < \varepsilon\right\} = 1 \qquad \overline{X}_n \xrightarrow{P} \mu$$

3、中心极限定理

- **德莫佛-拉普拉斯:**

□ $X_n \sim B(1,p)$□ □ □ □ □ □ □ □ □ □

$$\lim_{n\to\infty} P\{\frac{\mu_{\mathrm{n}} - np}{\sqrt{npq}} \le x\} = \Phi(x)$$

- **林德贝格-勒维:**

□ X_n□ □ □ □ □ □ □ □ □ $EX_n = \mu, Var(x) = \sigma^2 > 0$

□ $$\lim_{n\to\infty} P\left\{\frac{\sum_{k=1}^{n} X_k - \sum_{k=1}^{n} EX_k}{\sqrt{Var\left(\sum_{k=1}^{n} X_k\right)}} \le x\right\} = \Phi(x)$$