数理统计

Mathematical Statistics

兰小红
科研管理楼1520
邮箱:xhlan@ustc.edu.cn

数学科学学院

2022 秋

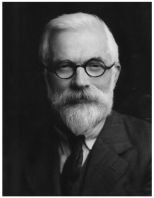
统计起源

- 统计学(Statistics)是一门科学,它研究怎样以有效的方式收集、整理、分析带随机性的数据，并在此基础上，对所研究的问题作出统计性的推断，直至对可能作出的决策提供依据或建议—《中国大百科全书・数学卷》
- 最早记录
  1. 中国夏朝– 人口调查统计
  2. 周朝– “司书”
  3. 《管子》– “调查”
- 统计与三种逻辑推理方法（演绎、归纳、诱导）
- 归纳推理：由观测的数据去匹配一个假设，从而由特殊推向一般的逻辑推理过程。
- “当我们不具备决定什么是真理的力量时，我们应遵从什么是最可能的，这是千真万确的真理。” – （法）笛卡尔

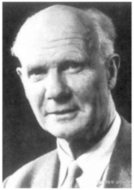
数理统计发展史

- 数理统计：研究统计学方法中的理论基础问题，是数学的一个分支
- 创立者：卡尔·皮尔逊 – 卡方统计量极限分布
  罗纳德·费希尔 – 统计学的数学基础
  埃贡·皮尔逊与乔治·内曼 — 假设检验Neyman-Pearson 理论

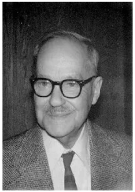
Karl Pearson（1857～1936）



Ronald Fisher（1890～1962



Egon Pearson (1895~1980)



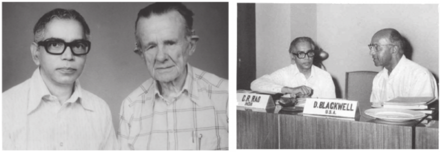
Jerzy Neyman (1894~1981)

许宝騄 (1910 ~ 1970) 参数估计和假设检验的优良性，多元统计分析数，统计量的极限分布，在数理统计和概率论方面第一个具有国际声望的中国数学家，是数理统计学奠基人之一。

-

数理统计发展史

- **哈拉尔德·克拉梅尔**(Harald Cramér, 1893 ~ 1985)与**C.R.劳**（Calyampudi Radhakrishna Rao，1920 ~ 至今）：Cramér-Rao 不等式，Rao-Blackwell 定理，《Mathematical methods of Statistics》(1946) 数理统计学成熟的标志



Rao with Cramér (left) and Blackwell (right)

亚伯拉罕.沃尔德 (Abraham Wald, 1902~1950)：序贯分析法，统计决策理论，改变二战进程的统计学家。

-
- Bayes 方法的兴起
- 直至如今，这门学科一直处于蓬勃发展与创新中...

- **本课程教材及参考资料**
  - 教材[0]：《数理统计（第二版）》，韦来生 编著，科学出版社，2015
  - 参考书[1]: 《Statistical Inference》2nd Edition，George Casella and Roger L. Berger，2002
  - 参考书[2]: 《高等数理统计（第二版）》，茆诗松 王静龙 濮小龙 编著，高等教育出版社，2006
  - 参考书[3]: 《数理统计讲义》，郑明 陈子毅 汪家冈 编著，复旦大学出版社，2006
  - 参考书[4]: 《数理统计与数据分析（原书第3 版）》，John A. Rice, 机械工业出版社，2013
  - 参考书[5]: 《Mathematical Statistics》2nd Edition，Jun Shao, Springer, 2003
- **主讲内容**
  1. 第一章 预备知识
  2. 第二章 点估计
  3. 第三章 参数假设检验
  4. 第四章 区间估计

1. 引论
2. 来自正态总体的抽样分布
3. 次序统计的分布
4. 指数族
5. 充分性原理

- 数理统计主要包括两部分：
  1. 有效地收集数据：强调“随机性”；两个重要分支–抽样调查方法，试验的设计和分析。举例说明见教材[0]例1.1.1 1.1.3。
  2. 有效地使用数据–统计推断。举例说明见教材[0]例1.1.4和如下例1.1
- **例1.1** 估计一个物体的重量a，在天平上称重5次获得数据$x_1, \cdots, x_5$。估计量$\hat{a}$ 采取下列三种方法：
  （1）$\hat{a} = \frac{x_1+\cdots+x_5}{5}$；
  （2）将$x_1, \cdots, x_5$ 按大小排列为$x_{(1)} \leq \cdots \leq x_{(5)}$，取$\hat{a} = x_{(3)}$；
  （3）$\hat{a} = \frac{x_{(1)}+x_{(5)}}{2}$。

问 哪个估计量最好？

答 在不同的假设条件和优劣评估标准下，有不同的最优估计量。

- **定义1.1** 一个统计问题所研究的对象的全体称为**总体**(Population)，而**样本**(Sample) 是从总体中抽取的一部分个体。在数理统计学中，总体可以用一个随机变量X 及其概率空间$\{\Omega, \mathcal{F}_0, P\}$ 来描述。
- **定义1.2** 一个问题的**统计模型**(Statistical Model)是指研究该问题时所抽样本的分布。具体而言：
  - 随机向量$\mathbf{X} = (X_1, \cdots, X_n)$ （样本）：观测数据$X_1, \cdots, X_n \sim X$.
  - $\mathbf{X}$ 的概率分布$(\mathcal{X}, \mathcal{F}, P)$：样本空间$\mathcal{X}$ (sample space) 即样本$\mathbf{X}$ 可能取值的全体；$\mathcal{F}$为样本所发生的所有可测事件组成的σ代数，P是在可测空间$(\mathcal{X}, \mathcal{F})$上定义的一个概率分布。
  - 统计模型$\mathcal{P} = \{P : P是(\mathcal{X}, \mathcal{F})上的一个**概率分布族**\}$。
- **定义1.3** 对于统计模型$\mathcal{P}$，假如分布族仅依赖于某个参数（或参数向量）θ，即

$$\mathcal{P} = \{P_\theta : \theta \in \Theta\}$$

其中Θ为参数空间，则称此模型为**参数（化）统计模型**。

注 如果统计模型$\mathcal{P}$中不含参数，则称之为**非参数统计模型**；**半参数统计模型**即既含有非参数部分又含有参数部分的统计模型。

- **例1.2** 测量误差模型：记第i个测量值为

$$X_i = \mu + \epsilon_i, \quad i = 1, \dots, n$$

  这里μ是一个实值参数，$\epsilon_1, \dots, \epsilon_n \quad i.i.d.$ 具有概率（累积）分布函数$G(\cdot)$。$G(\cdot)$不依赖于μ。
- $X_1, \dots, X_n \quad i.i.d.$ 具有概率（累积）分布函数$F(\cdot) = G(\cdot - \mu)$。
- 统计模型$\mathcal{P} = \{(\mu, G) : \mu \in \mathbb{R}, G \in \mathcal{G}\}$，这里$\mathcal{G}$ 是概率分布函数空间。
  - 参数模型：高斯测量误差$\epsilon_1, \dots, \epsilon_n \quad i.i.d. \sim N(0, \sigma^2)$，$\sigma^2 > 0$ 未知；
  - 半参数模型：均值为μ的对称测量误差，i.e. $\mathcal{G}$ 是关于0对称的分布函数族；

注 非参数模型举例：$X_1, \dots, X_n \quad i.i.d.$ 具有概率（累积）分布函数$G(\cdot)$, $G \in \mathcal{G}$。$\mathcal{G}$ 是定义在样本空间$\mathcal{X}$ 上的对称概率分布函数空间。

- **例1.3** 对一大批产品共N件进行抽样检查，其中废品$N\theta$ 件，θ 未知。抽样方式为一次抽取n 件，考虑这单个样本中的废品个数X。
  - 样本空间$\mathcal{X} = \{0, 1, \ldots, n\}$
  - 参数θ（废品率），其参数空间$\Theta = \{0, \frac{1}{N}, \ldots, \frac{N}{N}\}$
  - X的概率分布

$$P(X = x) = \frac{\begin{pmatrix} N\theta \\ k \end{pmatrix}\begin{pmatrix} N - N\theta \\ n - k \end{pmatrix}}{\begin{pmatrix} N \\ n \end{pmatrix}}$$

    即$X \sim Hypergeometric(N, N\theta, n)$，简记$H(N, N\theta, n)$。

- 参数统计模型$\mathcal{P} = \{H(N, N\theta, n) : \theta \in \Theta\}$。

注1 【0】例1.2.4抽样方式为：不放回抽样，一次抽一个，依次抽取，直至抽完n 个样本（$1 \leq n < N$），找出这个抽样相应的样本空间和参数统计模型。（注：这n 个样本之间并不相互独立）

注2 上述问题中如果抽样方式改为有放回抽样，则n个样本之间相互独立，见【0】例1.2.5。

- 离散型
  - 离散均匀分布 $\{DiscreteU(N_1, N_2) : N_1 < N_2, N_1, N_2 \in \mathbb{N}\}$；
  - 二项分布族 $\{b(n, p) : 0 < p < 1\}$；
  - Poisson分布族 $\{P(\lambda) : \lambda > 0\}$；
  - 多项分布族 $M(n, p_1, \cdots, p_r)$，这里所有$p_i > 0, \sum_{i=1}^r p_i = 1\}$；
  - 负二项分布 $\{NB(r, p) : r > 0, 0 < p < 1\}$；
  - 超几何分布 $H(N, M, n)$；等等。
- 连续型：
  - 均匀分布族 $\{U(a, b) : -\infty < a < b < \infty\}$；
  - （多元）正态分布族 $\{N(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 > 0\}$ $(\{N_n(\vec{\mu}, \Sigma)\})$；
  - Gamma分布族 $\{Gamma(\alpha, \beta) : \alpha > 0, \beta > 0\}$；
  - Beta分布族 $\{Beta(a, b) : a > 0, b > 0\}$；等等。
- 需要掌握的知识点：概率密度函数(p.d.f)的形状，期望，方差，特征函数等，参考【1】3.2~3.3节，【2】1.2节。

注1 三大抽样分布t分布，$\mathcal{X}^2$分布，F分布和指数族分布将在本章第2，5节中详细讲述。

注2 位置与尺度族（location and scale family）(ref. [1] sec.3.5)，混合分布（mixture distribution）(ref. [1] sec.4.4)等。

- **定义1.4** 设$X_1,\ldots,X_n$ 为从总体X中抽取的容量为n的样本，若$X_1,\ldots,X_n$ 相互独立，且具有相同分布F，则称$X_1,\ldots,X_n$ 为**简单（随机）样本**，简记$X_1,\ldots,X_n \quad i.i.d.$。仍记$\vec{X}=(X_1,\cdots,X_n)$可能取值的全体构成的样本空间为$\mathcal{X}$。
- **定义1.5** 设$X_1,\ldots,X_n$ 为从总体X中抽取的简单样本，$T(\vec{X})=T(X_1,\cdots,X_n)$是定义在样本空间$\mathcal{X}$ 上的实值或向量值函数，则称随机向量$T(\vec{X})$是一个**统计量**；T的概率分布称为T的抽样分布（Sampling distribution）。

注1 简言之，统计量T是关于$\vec{X}$的可测函数；

注2 统计量T只与样本$\vec{X}$有关，不能依赖于未知参数；

注3 统计量T是一个随机变量/向量，如果总体分布F未知，即使函数T已知，T的分布未必已知。统计推论和决策理论的一个主要问题即是找出T的分布！

注4 在具体问题中选用什么统计量以及判定所选用的统计量是否足够好地集中了样本中与所研究目标有关的信息，是本课程的一大内容。

两个常用统计量

- **定义1.6**
  1. 样本均值 (Sample mean)　随机样本值的算术平均
  $$\overline{X}=\frac{X_1+\cdots+X_n}{n}$$
  2. 样本方差 (Sample variance)
  $$S_X^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$$
  样本标准方差$S_X=\sqrt{S_X^2}$
- **问题**：能否找到$\overline{X}$和S_X^2的分布以及两者的联合分布？
- 取决于总体分布F的已知信息！！

两个常用统计量

- $\overline{X}$和S_X^2的基本性质
  1. $\sum_{i=1}^n (X_i - \overline{X}) = 0$；
  2. 令$Y_i = aX_i + b, i = 1, \ldots, n$，则$S_Y^2 = a^2 S_X^2$；
  3. $\min_{a\in\mathbb{R}} \sum_{i=1}^n (X_i - a)^2 = \sum_{i=1}^n (X_i - \overline{X})^2 = (n-1)S_X^2$；
  4. $S_X^2 = \frac{1}{n-1}[\sum_{i=1}^n X_i^2 - n\overline{X}^2]$。
- $\overline{X}$和S_X^2的样本矩
  1. 如果同分布F有有限均值μ，则$\mathbb{E}\overline{X} = \mu$;

  注 如$F = P_\theta$（含参数），则$\mathbb{E}\overline{X} = \int x dP_\theta = \mu(\theta)$.
  2. 如F有有限方差σ^2，则$Var(\overline{X}) = \frac{1}{n}\sigma^2$；

  注 如$F = P_\theta$（含参数），则$Var(\overline{X}) = \frac{1}{n}\sigma^2(\theta)$.
  3. 条件同(2)，则$\mathbb{E}S_X^2 = \sigma^2$；
  4. 如果$\mathbb{E}|X_1|^3$有限，则$\mathbb{E}(\overline{X})^3$ 和$Cov(\overline{X}, S_X^2)$有限；
  5. 如果$\mathbb{E}|X_1|^4$有限，则$Var(S_X^2)$有限。

- **例1.4**
  1. 如果$X_1, \ldots, X_n \quad i.i.d. \sim N(\mu, \sigma^2)$，则$\overline{X} \sim N(\mu, \frac{1}{n}\sigma^2)$；
  2. 如果$X_1, \ldots, X_n \quad i.i.d. \sim Exp(\lambda)$，则$\overline{X} \sim Gamma(n, n\lambda)$，这里$Exp(\lambda) = Gamma(1, \lambda)$，而$Gamma(\alpha, \beta) \quad p.d.f.$
  $$f(x|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, \quad x > 0.$$
  3. 如果$X_1, \ldots, X_n \quad i.i.d. \sim Cauchy(x_0, \gamma)$，则$\overline{X} \sim Cauchy(x_0, \gamma)$，这里$Cauchy(x_0, \gamma) \quad p.d.f.$
  $$f(x|x_0, \gamma) = \frac{1}{\pi\gamma[1 + \frac{(x-x_0)^2}{\gamma^2}]}, \quad -\infty < x < \infty.$$
- **证明（略）：** 考虑特征函数$\phi_{\overline{X}}(t) = \mathbb{E}[e^{i\overline{X}t}]$，$t \in \mathbb{R}$，

- **例1.4**（续）

分布	特征函数
$N(\mu, \sigma^2)$	$\exp\{it\mu - \frac{1}{2}\sigma^2 t^2\}$
$Gamma(\alpha, \beta)$	$(1 - \frac{it}{\beta})^{-\alpha}$
$Cauchy(x_0, \gamma)$	$\exp\{itx_0 - \gamma\lvert t\rvert\}$

注 对于(1)和(2)，当$n \to \infty$时，我们可利用(Lindeberg-Lévy)中心极限定理（CLT）获得大样本情形下$\overline{X}$的渐近分布，

$$\sqrt{n}(\overline{X} - \mu) \to N(0, \sigma^2), \quad n \to \infty;$$

但对于(3)，CLT不成立，因为$\mathbb{E}X_1$不存在，同时$\mathbb{E}X_1^2 = \infty$。

作业

习题1: Ex. 5(1)(2), 9, 10, 11;

习题2: Ex. 3(2).